Commonsense Inference in Dynamic Spatial Systems
“Phenomenal and Reasoning Requirements”

Mehul Bhatt

SFB/TR 8 Spatial Cognition
Universität Bremen, Germany
bhatt@informatik.uni-bremen.de

Abstract

Spatial changes within an environment are typically a result of interaction— actions and events — occurring within. Reasoning about such changes, when dealt with formally within the context of qualitative spatial calculi and logics of action and change, poses several difficulties along multiple dimensions: (a) phenomenal requirements stemming from the dynamic nature of the spatial system (e.g., appearing and disappearing objects), (b) reasoning requirements (e.g., abductive explanation), (c) domain-independent or epistemological (e.g., persistence, ramification), and (d) aspects concerning the need to satisfy the intrinsic (axiomatic) properties of the spatial calculi (e.g., compositional consistency) being modelled. This paper, encompassing the phenomenal and reasoning aspects in (a) and (b) respectively, presents some instances that demonstrate the role of commonsense reasoning and the non-monotonic inference patterns it necessitates whilst representing and reasoning about dynamic spatial systems in general.

1 Motivation

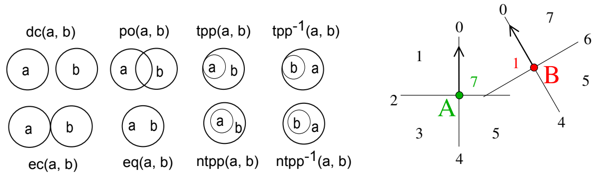
Dynamic Spatial Systems ($\mathcal{DSS}$) are systems where spatial configurations, denoted by sets of qualitative spatial relations, undergo transformations as a result of actions and events occurring within the environment [Bhatt and Loke, 2008]. The $\mathcal{DSS}$ approach is applicable in a wide-range of application domains as diverse as cognitive robotics, diagrammatic reasoning, architecture design, geographical information systems and even the new generation of ambient intelligence systems involving behaviour or activity monitoring. From the viewpoint of such applications, the basic functionality required from the $\mathcal{DSS}$ approach remains the same, namely, the capability to serve either a predictive (i.e., projection, planning) or an explanatory (e.g., causal explanation) function in the context of high-level qualitative models of space and spatial change. This in turn requires that change in general and spatial change in specific, and its relationship to action, events and other aspects such as causality be taken seriously.

Existing qualitative spatial modelling techniques have primarily remained focused on reasoning with static spatial configurations. In general, research in the qualitative spatial reasoning domain has remained focused on the representational aspects of spatial information conceptualization and the construction of efficient computational apparatus for reasoning over those by the application of constraint-based techniques [Cohn and Renz, 2007, Renz and Nebel, 2007]. For instance, given a qualitative description of a spatial scene, it is possible to check for its consistency along arbitrary spatial domains (e.g., topology, orientation and so forth) in an efficient manner by considering the general properties of a qualitative calculus [Ligozat and Renz, 2004]. However, for applications such as the ones aforementioned, these methods require a realistic interpretation, such as the one provided by the stated $\mathcal{DSS}$ perspective, where sets of spatial relations undergo change as a result of named occurrences in the environment, or broadly, reasoning about space and reasoning about actions and change are consilidate into a ‘Reasoning about Space, Actions and Change’ (RSAC) paradigm [Bhatt, 2009]. Consequently, the formal embedding of arbitrary spatial calculi – whilst preserving their high-level axiomatic semantics and if necessary, their low-level algebraic properties too – has to be investigated from the viewpoint of formalisms that deal with action and change in general.

In this paper, we illustrate the utility of commonsense reasoning, and the non-monotonicity it entails, toward realising the suggested embedding of spatial calculi within general formalisms of action and change. Note that the embedding per se is extensive, and not the object of this paper. Rather, we solely focus on *some* commonsense inference patterns that occur whilst achieving the said embedding. These patterns pertain to the following aspects:

AI existential consistency of complete spatial situation descriptions given that fact that the domain of discourse of primitive spatial entities may be incompletely known, i.e., unknown objects may have appeared or known objects may have either temporarily disappeared or may have permanently ceased to exist

AII modelling causal explanation tasks, where given a set of temporally-ordered observations, the objective is to derive an explanation in terms of the (spatial and non-spatial) actions and events that may have caused the observations. Here, modelling causal explanation abductively necesssitates the use of a circumscriptive non-



(a) $\mathcal{RCC}-8$ Primitives (b) The $\mathcal{OPRA}_2$ relation $\vec{A}\ 2\angle^1_7\ \vec{B}$

Figure 1: Topological and Orientation Calculi

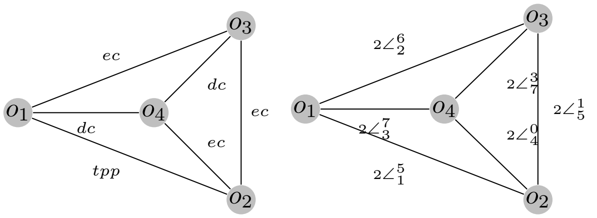
monotonic approach.

In comparision to the other epistemological and intrinsic spatial calculi related commonsense patterns, which are excluded from this paper, the aspects in (AI–AII) are extrinsic to the process of embedding and concern phenomenal aspects (section 3) of dynamic spatial systems and the computational or reasoning requirements (section 4) expected from an operationalization of the $\mathcal{DSS}$ perspective.

2 Qualitative Spatial Primitives

The objective of this paper is to intuitively present the nature of commonsense reasoning as relevant to aspects (AI–AII; section 1). As such, we do not go into the details of the formal axiomatisation of a theory of change or the details pertaining to the constitution of a qualitative spatial calculus. However, a basic overview of the ontological setup is needed to make the paper self-contained.

The spatial ontology that is required depends on the nature of the spatial calculus that is being modeled. In general, spatial calculi can be classified into two groups: topological and positional calculi. When a topological calculus such as the Region Connection Calculus (RCC) [Randell et al., 1992] is being modeled, the primitive entities are spatially extended and could possibly even be 4D spatio-temporal histories (e.g., in a domain involving the analyses of motion-patterns). Alternately, within a dynamic domain involving translational motion in a plane, a point-based (e.g., Double Cross Calculus [Freksa, 1992], $\mathcal{OPRA}_m$ [Moratz, 2006]) or line-segment based (e.g., Dipole Calculus [Schlieder, 1995, Moratz et al., 2000]) abstraction with orientation calculi suffices. Figure 1(a) is a 2D illustration of relations of the RCC-8 fragment of the region connection calculus. This fragment consists of eight relations: disconnected (dc), externally connected (ec), partial overlap (po), equal (eq), tangential proper-part (tpp) and non-tangential proper-part ($ntpp$), and the inverse of the latter two tpp^{-1} and $ntpp^{-1}$. Similarly, Fig. 1(b) illustrates one primitive relationship for the Oriented Point Relation Algebra (OPRA) [Moratz, 2006], which is a spatial calculus consisting of oriented points (i.e., points with a direction parameter) as primitive entities. The granularity parameter m determines the number of angular sectors, i.e., the number of base relations. Applying a granularity of $m = 2$ results in 4 planar and 4 linear regions (Fig. 1(b)), numbered from 0 to 7, where region 0 coincides with the orientation of the point. The family of $\mathcal{OPRA}_m$ calculi are designed for reasoning about the relative orientation relations between oriented points and are well-suited for dealing with objects that have an intrinsic front or move in a particular direction.



(a) $\mathcal{RCC}-8$ (b) $\mathcal{OPRA}_2$

Figure 2: Complete N-Clique Descriptions

Spatial Situation Descriptions

Spatial situation descriptions consist of a complete n-clique graph for a domain of n objects. Further, there is one such clique for every type of spatial domain (e.g., topology, orientation) that is modelled. Precisely, for a spatial scene description with n domain objects and k spatial calculi being modeled, the scene description involving n objects requires a complete n-clique specification with $[n(n-1)/2]$ spatial relationships for each of the respective calculi (Fig. 2). Given such spatial scene descriptions, the following notion of existential consistency is definable:

Definition 2.1 ($\mathcal{E}$-Consistency). *A spatial scene description is $\mathcal{E}$-Consistent, i.e., existentially consistent, if there exists at least one spatial relationship of any spatial domain (i.e., topology, orientation etc) that every existing spatial object participates in with other existing object(s).* ☐

From the viewpoint of planning and explanation tasks, $\mathcal{E}$-Consistency is necessary and useful toward maintaining the consistency of spatial scene descriptions given the fact that appearance of new objects and disappearance of existing ones may have occurred within the system. In the context of such phenomenal requirements, the significance and use of $\mathcal{E}$-Consistency from Definition 2.1 is further elaborated on in section 3.

3 Phenomenal Commonsense: Appearance and Disappearance of Objects

Appearance of new objects and disappearance of existing ones, either abruptly or explicitly formulated in the domain theory, is characteristic of non-trivial dynamic spatial systems. In robotic applications, it is necessary to introduce new objects into the model, since it is unlikely that a complete description of the robot's environment is either specifiable or even available. Similarly, it is also typical for a mobile robot operating in a dynamic environment, with limited perceptual or sensory capability, to lose track of certain objects because of issues such as noisy sensors or a limited field-of-vision. As an example, consider a '*delivery scenario*' in which a vehicle/robot is assigned the task of delivering '*object*(*s*)' from one '*way-station*' to another. In the initial situation description, the domain consists of a finite number of '*way-stations*' and deliverable '*objects*' (see Fig.

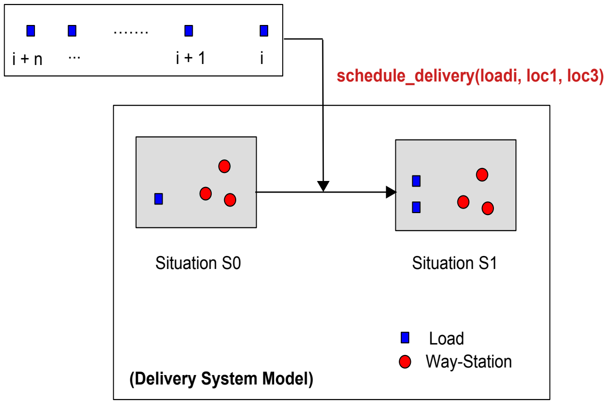


Figure 3: Appearance Events - Delivery Example

3). However, the scheduling of new objects for delivery in future situations will involve introducing new ‘objects’ into the domain theory. For example, an external event[1] such as ‘$schedule\_delivery(new\_load,\ loc_1,\ loc_3)$’ introduces a new object, namely ‘*new_load*’, into the domain.

Appearance and disappearance events involving the modification of the domain of discourse are not unique to applications in robotics. Even within the projected next-generation of event-based and temporal geographic information systems, appearance and disappearance events are regarded to be an important typological element for the modelling of dynamic geospatial processes [Claramunt and Thériault, 1995, Worboys, 2005]. For instance, Claramunt and Thériault [1995] identify the basic processes used to define a set of low-order spatio-temporal events which, among other things, include appearance and disappearance events as fundamental. Similarly, toward event-based models of dynamic geographic phenomena, Worboys [2005] suggests the use of appearance and disappearance events at least in so far as single object behaviours are concerned. We regard that such phenomena, being intrinsic to a typical dynamic spatial system, merit systematic treatment.

Maintaining and Propagating Existential Facts

The case of disappearance is not problematic, however, for the case of appearance and re-appearance, some questions that need to be addressed include:

- what is the spatial relationship (topological, directional etc) of the newly appearing object with other existing objects?
- given the fact that a newly appearing object is, from a model-theoretic viewpoint, *unknown* in the past, how to make it ‘known’ and ‘not exist’ in the past? (this scenario is illustrated model-theoretically in Fig. 5)
- how to make past and present situation descriptions ‘compositionally consistent’?[2]

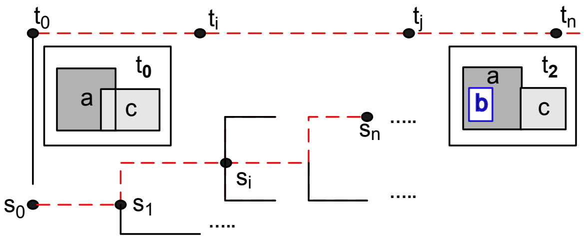


Figure 4: Branching-time Situation History

From a representational viewpoint, introducing new objects in the domain poses a problem since there is no general way to deal with an incompletely known domain of discourse. For instance, let $< s_0,\ s_1,\ s_2, \ldots,\ s_n >$ denote a situation-based linear history or one branch within the branching-tree structure of the overall situation space (see Fig. 4). From a dynamic spatial system perspective, each state corresponding to every situation with this history is primarily a set denoting the spatial configuration of objects in that situation. Further assume that an object ‘b’, that is unknown or not a part of the dynamic ‘spatial configuration set’ in the initial situation ‘s_0’, comes into existence (by an appearance event) in a later situation, say ‘s_2’. At this point, it is necessary to incorporate the non-existence of ‘b’ in the situations preceding ‘s_2’ by (non-monotonically) propagating its non-existence backwards into the situation-based history. In fact, appearance of previously unknown objects is the only reason ‘*existential facts*’ about objects need to be included as propositional fluents / dynamic properties at a domain-independent level. The case of disappearing objects is trivial and simply involves negating and object’s existential status upon the occurrence of disappearance events. Indeed, an object that is known but has disappeared may not participate in spatial relationships with other objects, until such a time when it reappears. The following steps summarise the solution approach for the case where an object’s identity is maintained upon reappearance:

S1 firstly, maintain existential facts about objects by way of the propositional fluent $exists(o,\ s_1)$

S2 add special ‘*appearance*’ and ‘*disappearance*’ events that act on the existential fluent through direct effect axioms (i.e., disappearance *causes* an object to *not exist* and so forth)

S3 maintain ‘*null*’ spatial relationships between non-existing objects and all other existing objects. Indeed, this also implies that such null relationships acquire a special status in the situation calculus being modelled. For instance, a calculus such as RCC-8 with eight spatial relationships becomes a calculus with nine primitive relationships.

S4 add a constraint that newly appearing objects must participate in at least one ‘*non-null*’ spatial relationship with an already existing object. The precise relationship is specifiable in domain specific ways.

S5 finally, either apply predicate completion for $exists(\ldots)$ or minimizing it (to close its extensionality) on a situation-by-situation basis. This ensures that newly ap-

---

[1] External events are those occurrences that do not have an associated occurrence criteria and may therefore occur abruptly.

[2] Compositional consistency refers to the satisfaction of the global constraints formulated by composition theorems relevant to every spatial calculus that is modelled.

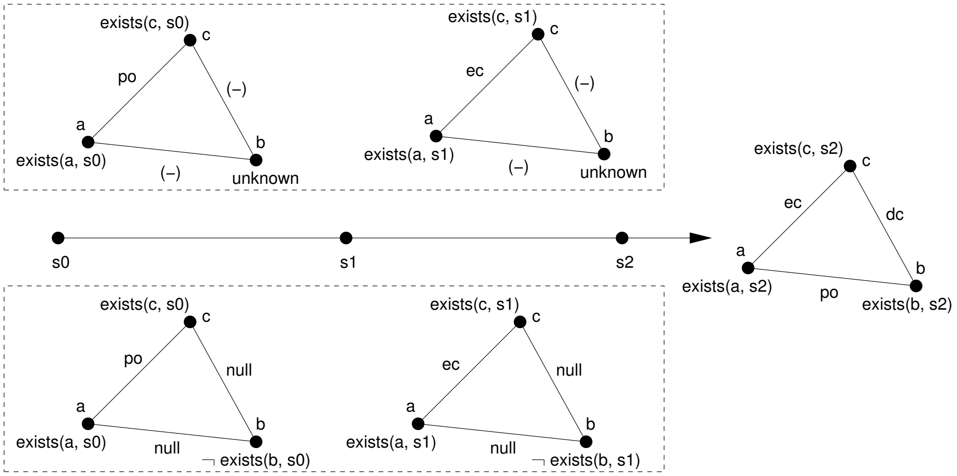


Figure 5: Appearance and Scene Descriptions

pearing objects are by default assumed to 'not exist' in the entire history of the system

Model-Theoretic Implications

In a strictly model-theoretic sense, *appearances* and *disappearances* should respectively correspond to the addition and removal of typed-entities, which in this case are spatial objects, from the underlying domain of discourse.[3] Whereas this is true in the case of the manner in which we model appearances, the same does not hold for disappearances. Strictly speaking, a *disappearance* does not lead to the removal of the entity from the domain in a model-theoretic sense. It simply modifies the object's existential property in a way such that:

- its spatial relationships with all other objects assume the value of $null$. This is easily achievable via the inclusion appropriate effect axioms for the propositional fluent *exists*.
- it cannot participate in subsequent spatial transformations that the system undergoes. This restriction is enforced by compiling the relevant existential preconditions for all relevant/potential occurrences

Note that it is also possible that a previously-disappeared object may re-appear in a later situation. In this case, instead of introducing a new object into the model, this is merely a case of modifying the existential fact about the concerned object. Again, this is achieved via the effect axioms. However, the spatial relationship of the new or re-appearing object with other existing objects cannot be $null$. An alternative approach could have been to not make any assumption with regard to the relationship of the new object with other objects until such information becomes available in a future situation and for as long as such information does not become available, its relationship with other objects will continue to be $null$ because of the default assumption of *inertia*. However, allowing this behaviour leads to '*existential inconsistencies*' where there exist situations in which a object *exists* and does not participate in any (qualitative) spatial relationship with any other object, which is clearly a situation that cannot arise in reality, i.e., the coming into existence of an object has to be based on some real observation (e.g., in robotics applications) or from some othe rsource of data (e.g., GIS dataset).

4 Explanatory Commonsense: Reasoning Requirement

Explanation tasks constitute a basic reasoning requirement in many application domains. Here, given a set of time-stamped observations or snap-shots (e.g., observation of a mobile-robot or time-stamped GIS data), the objective is to explain which events and/or actions may have caused the resulting state-of-affairs. From a rather general viewpoint, explanatory reasoning encompasses all problems resembling the classic '*stolen-car scenario*'. Explanation, in general, is regarded as a converse operation to temporal projection essentially involving reasoning from effects to causes, i.e., reasoning about the past [Shanahan, 1989]. In the context of the situation calculus formalism [McCarthy, 1977], which is a general formalism for modelling dynamic domains, Shanahan [1993, 1997] proposes a non-monotonic approach that utilises circumscription as a basis of minimization (of effects) and explanation derivation (in terms of potential occurrences). We have specialised this approach toward the formulation of an abductive occurrence-driven causal explanation task, where a set of time-ordered observations (e.g., pertaining to spatial configurations) may be explained in terms of the spatial actions and events that may have caused the observed state-of-affairs.

Let $\mathcal{L}$ denote a first-order many-sorted language with equality and the usual alphabet of logical symbols $\{\neg, \wedge, \vee, \forall, \exists, \supset, \equiv\}$.[4] With $\mathcal{L}$ as a basis, a situation calculus meta-theory Σ_{sit} required from the viewpoint of the causal explanation task in [Bhatt and Loke, 2008] is adopted:

---

[3] Whether such a truly general solution is achievable model-theoretically remains doubtful. The approach we suggest is at least applicable in the present context of modelling dynamic spatial systems and modelling the *appearances* and *disappearances* thereof.

[4] Although the $\mathcal{L}$ requires additional predicates, such details are not relevant here any may be found in [Bhatt and Loke, 2008].

Definition 4.1 (Theory of Space & Change: $\Sigma_{sit} \cup \Sigma_{space}$). *The foundational theory Σ_{sit} of the situation calculus formalism consists of the following set of formulae: the property causation axiom determining the relationship between being 'caused' and being 'true', a generic frame axiom in order to incorporate the assumption of inertia, uniqueness of names axioms for the fluents, occurrences and fluent denotations, and domain closure axioms for propositional and functional fluents. Σ_{space} constitutes a formalisation of the general aspects pertaining to the static and dynamic aspects of spatial calculi. Σ_{space} essentially denotes a general spatial theory that can be re-used in arbitrary dynamic spatial domains.* ☐

With respect to a basic theory of space and change in Definition 4.1 that accounts for causation, inertia and ramification, and a qualitative spatial theory, we present the general structure of commonsense reasoning involved in abducing an object's appearance for a simple scenario.

Structure of Causal Explanation WRT. $[\Sigma_{sit} \cup \Sigma_{space}]$
We outline the structure of the causal explanation task without going into the details of the underlying/supporting axiomatisation: 'consider again the illustration in Fig. 4 – the situation-based history $< s_0, s_1, \ldots, s_n >$ represents one path, corresponding to a actual time-line $< t_0, t_1, \ldots, t_n >$, within the overall branching-tree structured situational space. Furthermore, assume a simple system consisting of objects 'a', 'b' and 'c' and also that the state of the system is available at time-point t_0 and t_2. Note that the situational-path and the time-line represent an actual as opposed to a hypothetical evolution of the system. From the viewpoint of this discussion, two auxiliary predicates, namely $HoldsAt(\phi, t)$ and $Happens(\theta, t)$, that range over '*time-points*' instead of '*situations*' are needed to accommodate the temporal extensions required to map a path in the situation-space to an actual time-line; complete definitions can be found in Pinto [1994]. Given an initial situation description as in Φ_1 (see (1)), where 'b' is unknown and 'a' and 'c' are partially overlapping, in order to explain an observation sentence such as Φ_2, a formula of the form in Δ needs to be derived'.

$$\left\{\begin{array}{l}
\Phi_1 \equiv HoldsAt(\phi_{top}(a, c), po, t_1) \\
\Phi_2 \equiv HoldsAt(\phi_{top}(a, c), ec, t_2) \wedge HoldsAt(exists(b), true, t_2) \\
\qquad \wedge HoldsAt(\phi_{top}(b, a), ntpp, t_2) \\
[\Sigma_{sit} \wedge \Sigma_{space} \wedge \Phi_1 \wedge \Delta] \models \Phi_2, where \\
\Delta \equiv (\exists\, t_i, t_j, t_k).[t_1 \leq t_i < t_2 \wedge Happens(appearance(b), t_i)] \\
\wedge [t_i < t_j < t_2 \wedge Happens(tran(b, a, tpp), t_j)] \wedge \\
[t_k < t_2 \wedge Happens(tran(a, c, ec), t_k)] \wedge [t_k \neq t_i \wedge t_k \neq t_j]
\end{array}\right\} \tag{1}$$

The derivation of Δ primarily involves non-monotonic reasoning in the form of minimising change and abducing appearance, in addition to making the usual default assumptions about inertia; the details of the derivation may be found in [Bhatt and Loke, 2008].

Domain-Specific Heuristics in Abduction
The non-monotonicity required in modelling explanation tasks is characteristic to modelling explanation problems abductively in general, rather than being peculiar to spatial reasoning tasks. However, one aspect of this non-monotonicity is characteristic to a spatial reasoning task – in deriving minimal models or explanations of observations consisting of changing spatial configurations, it is possible that the derived explanations may be *inadequate*, i.e., may not include domain-specific occurrences that have caused the observed changes. For instance, consider a geographic information system domain / scenario as depicted in Fig. 6. At a domain-independent level (i.e., at the level of a general spatial theory), the scene may be described using topological and qualitative size relationships. Consequently, the only changes that are identifiable at the level of the spatial theory are *shrinkage* and eventual *disappearance* – this is because a domain-independent spatial theory may only include a generic typology (appearance, disappearance, growth, shrinkage, deformation, splitting, merging etc) of spatial change at the most. However, at a domain-specific level, these changes could characterize a specific event (or process) such as, for instance, *deforestation*. The hypotheses or explanations that are generated during a explanation process should necessarily consist of the domain-level occurrences in addition to the underlying (associated) spatial changes (as per the generic typology) that are identifiable. That is to say, that the derived explanations be '*adequate*' and more or less take a form such as: '*Between time-points t_i and t_i, the process of deforestation is abducible as one potential hypothesis*'. To achieve this adequacy, a model-filtration heuristic that disregards those models (i.e., explanations) that do not include any domain-specific (spatial) occurrences (actions or events) leads to explanations that are adequate, if such explanation exists per se – this is because minimal models that only consist of a domain-independent explanation (e.g., in the form of *shrinkage*, *disappearance* and a temporal-order between these two) would be excluded by such a filtration heuristic.

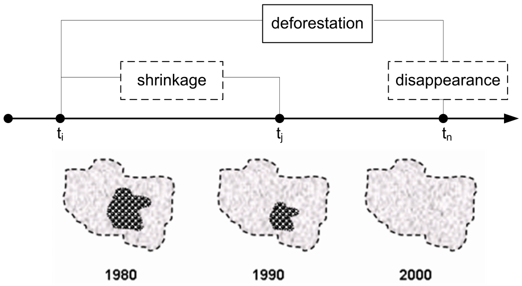


Figure 6: Domain Specific and Independent Abduction

Other potential solution to achieve adequacy is to include high-level or domain-specific predicates that relate the domain-independent occurrences (as per the typology) to arbitrary high-level processes that have a domain-dependent interpretation. Notwithstanding the fact that we regard both potential solutions to the problem of achieving adequacy to be rather rudimentary or ad-hoc solutions, it must be pointed out that the model-filtration approach is more general and does not presuppose any information of the domain-independent typology on the part of a domain modeler.

5 Discussion and Outlook

Qualitative spatial methods have primarily remained focused on reasoning with static spatial configurations. However, for applications such as cognitive robotics, these methods require a more realistic interpretation, where sets of spatial relations undergo change as a result of named occurrences in the environment. Consequently, the formal embedding of arbitrary spatial calculi – whilst preserving their high-level axiomatic semantics and low-level algebraic properties – has to be investigated from the viewpoint of formalisms such as the situation calculus, event calculus and fluent calculus. At a higher level of abstraction, this will result in the (native) incorporation of commonsense notions of space and spatial change within languages such as GOLOG and FLUX for their use in arbitrary robot control domains. In general, the areas of commonsense reasoning, and action and change are mature and established tools, formalisms and languages from therein are general enough to be applied to the case of dynamic spatial systems, where relational spatial models undergo change as a result of interaction in the environment.

The commonsense reasoning patterns pertaining to spatial reasoning illustrated in this paper have been investigated in the context of operationalizing the $\mathcal{DSS}$ perspective within situation calculus [Bhatt and Loke, 2008]. This constitutes one approach to operationalize the reasoning about space, actions and change paradigm [Bhatt, 2009]. Closely related is the work of Davis [2008, 2009] that investigates the use of commonsense reasoning about the physical properties of objects within a first-order logical framework. The key highlight of this work is that it combines commonsense qualitative reasoning about 'continuous time, Euclidean space, commonsense dynamics of solid objects, and semantics of partially specified plans' Davis [2009]. Other formalizations such as within a belief revision framework [Alchourrón et al., 1985], nonmonotonic causal formalizations in the manner of [Giunchiglia et al., 2004] are possible and the subject of ongoing study. Additionally, the suitability of event calculus [Kowalski and Sergot, 1986] and fluent calculus [Thielscher, 1998], vis-à-vis the situation calculus at least for specific reasoning tasks or scenarios is also a topic worth investigating.

Acknowledgements: The author acknowledges the financial support of the ALEXANDER VON HUMBOLDT STIFTUNG, GERMANY toward this research.

References

C. E. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change: Partial meet contraction and revision functions. *J. Symb. Log.*, 50(2):510–530, 1985.

M. Bhatt. Reasoning about space, actions and change: A paradigm for applications of spatial reasoning. *Spatial Cognition and Computation*, 2009. (in review).

M. Bhatt and S. Loke. Modelling dynamic spatial systems in the situation calculus. *Spatial Cognition and Computation*, 8(1):86–130, 2008. ISSN 1387-5868.

C. Claramunt and M. Thériault. Managing Time in GIS: An Event-Oriented Approach. In *Temporal Databases*, pages 23–42, 1995.

A. G. Cohn and J. Renz. Qualitative spatial reasoning. In F. van Harmelen, V. Lifschitz, and B. Porter, editors, *Handbook of Knowledge Representation*. Elsevier, 2007.

E. Davis. Pouring liquids: A study in commonsense physical reasoning. *Artif. Intell.*, 172(12-13):1540–1578, 2008. ISSN 0004-3702.

E. Davis. How does a box work? a study in the qualitative dynamics of solid objects. *Artif. Intell.*, 2009. to appear.

C. Freksa. Using orientation information for qualitative spatial reasoning. In *Proceedings of the Intl. Conf. GIS, From Space to Territory: Theories and Methods of Spatio-Temporal Reasoning in Geographic Space*, pages 162–178, London, UK, 1992. Springer-Verlag. ISBN 3-540-55966-3.

E. Giunchiglia, J. Lee, V. Lifschitz, N. McCain, and H. Turner. Nonmonotonic causal theories. *Artif. Intell.*, 153(1-2):49–104, 2004.

R. Kowalski and M. Sergot. A logic-based calculus of events. *New Gen. Comput.*, 4(1):67–95, 1986. ISSN 0288-3635.

G. Ligozat and J. Renz. What is a qualitative calculus? a general framework. In *PRICAI*, pages 53–64, 2004.

J. McCarthy. Epistemological problems of artificial intelligence. In *IJCAI*, pages 1038–1044, 1977. Re-printed in Readings in Non-Monotonic Reasoning, 1987, pages 46–52.

R. Moratz. Representing relative direction as a binary relation of oriented points. In *ECAI*, pages 407–411, 2006.

R. Moratz, J. Renz, and D. Wolter. Qualitative spatial reasoning about line segments. In *ECAI*, pages 234–238, 2000.

J. A. Pinto. *Temporal reasoning in the situation calculus*. PhD thesis, 1994. Supervisor: Raymond Reiter.

D. A. Randell, Z. Cui, and A. Cohn. A spatial logic based on regions and connection. In *KR'92. Principles of Knowledge Representation and Reasoning: Proceedings of the Third International Conference*, pages 165–176. Morgan Kaufmann, San Mateo, California, 1992.

J. Renz and B. Nebel. Qualitative spatial reasoning using constraint calculi. In *Handbook of Spatial Logics*, pages 161–215. 2007.

C. Schlieder. Reasoning about ordering. In *Proc. of COSIT'95*, volume 988 of *Lecture Notes in Computer Science*, pages 341–349. Springer, Berlin, Heidelberg, 1995.

M. Shanahan. Prediction is deduction but explanation is abduction. In *IJCAI*, pages 1055–1060, 1989.

M. Shanahan. Explanation in the situation calculus. In *IJCAI*, pages 160–165, 1993.

M. Shanahan. *Solving the frame problem: a mathematical investigation of the common sense law of inertia*. MIT Press, 1997. ISBN 0-262-19384-1.

M. Thielscher. Introduction to the fluent calculus. *Electron. Trans. Artif. Intell.*, 2:179–192, 1998.

M. F. Worboys. Event-oriented approaches to geographic phenomena. *Int. Journal of Geographical Inf. Science*, 19 (1):1–28, 2005.

Order-of-Magnitude Based Link Analysis for False Identity Detection

Tossapon Boongoen and Qiang Shen
Department of Computer Science, Aberystwyth University, UK

Abstract

Combating identity fraud is crucial and urgent as false identity has become the common denominator of all serious crime, including mafia trafficking and terrorism. Typical approaches to detecting the use of false identity rely on the similarity measure of textual and other content-based characteristics, which are usually not applicable in the case of deceptive and erroneous description. This barrier can be overcome through link information presented in communication behaviors, financial interactions and social networks. Quantitative link-based similarity measures have proven effective for identifying similar problems in the Internet and publication domains. However, these numerical methods only concentrate on link structures, and fail to achieve accurate and coherent interpretation of the information. Inspired by this observation, this paper presents a novel qualitative similarity measure that makes use of multiple link properties to refine the underlying similarity estimation process and consequently derive semantic-rich similarity descriptors. The approach is based on order-of-magnitude reasoning. Its applicability and performance are experimentally evaluated over a terrorism-related dataset, and further generalized with publication data.

Introduction

False identity has become the common denominator of all serious crime such as mafia trafficking, fraud and money laundering. Particularly in the UK, financial losses due to such cause are reported to be around 1.3 billion pounds each year (Wang *et al.* 2006). Holders of false identity are determined to avoid accountability and traces for law enforcement authority. In essence, such offence is intentionally committed with a view to perpetrating another crime from the most trivial to the most dreadful imaginable. Organized criminals make use of counterfeit identity to cover up illicit activities and illicitly gained capital. Especially in the case of terrorism, it is widely utilized to provide financial and logistical support to terrorist networks that have set up and encourage criminal activities to undermine civil society. Tracking and preventing terrorist activities undoubtedly requires authentic identification of criminals and terrorists who typically possess multiple fraud and deceptive names, addresses, telephone numbers and email accounts.

With present high-quality off-the-shelf equipment, it is almost effortless to obtain false identity documents. Conversely, it requires a great deal of time and experience to distinguish between genuine and forged copies. However, a successful detection can prevent the revolting consequence like that of shocking September-11 terrorist attacks. In particular to this tragedy, US authorities seriously failed to discover the use of false identities by nineteen terrorists, who were all able to enter the United States without any problem, in the very morning of the attacks. Most of them typically possess several dates of birth and multiple aliases (Boongoen & Shen 2008). For instance, *Mohamed Atta*, alleged ringleader of the September 11 attacks, has exploited several different aliases of *Mehan Atta, Mohammad El Amir, Muhammad Atta* and *Muhammad Al Amir Awad Al Sayad*. Identity verification and name variation detection systems (Wang *et al.* 2006) that rely solely on the inexact search of textual attributes may be effective in some cases. However, these methods would fail drastically to disclose unconventional truth of highly deceptive identity like that between *Osamabin Laden* and *The Prince* (Hsiung *et al.* 2005).

The aforementioned dilemma may be overcome through link analysis, which seeks to discover knowledge based on the relationships in data about people, places, things, and events. Intuitively, despite using distinct false identities, each terrorist normally exhibits unique relations with other entities involving in legitimate activities found in any open or modern society, making use of mobile phones, public transportation and financial systems. Link analysis techniques have proven effective for identity problems (Badia & Kantardzic 2005), (Hsiung *et al.* 2005) by exploiting link information instead of content-based information, which is typically unreliable due to intentional deception, translation and data-entry errors (Wang *et al.* 2005). Recently, link analysis is also employed by Argentine intelligence organizations to analyzing Iranian-Embassy telephone records in such a way to make a circumstantial case that the Iranian Embassy had been involved in the July 18, 1994, terror bombing of a Jewish community centre (Porter 2008).

Essentially, to justify the similarity between entities (e.g. names, publications and web pages) in a link network, many well-known algorithms like SimRank (Jeh & Widom 2002), PageSim (Lin, King, & Lyu 2006) and Connected-Triple (Klink *et al.* 2006) analogously concentrate only on the car-

---

dinality of joint neighbors to which they are directly linked, without taking into account the characteristics of a link itself. As such, the quality of the similarity evaluation may be enhanced by including uniqueness measure of links (Boongoen & Shen 2008) within the overlapping neighbor context. However, a definite precaution to combining multiple measures is the inaccuracy of quantitative descriptions, which are usually caused by a few link patterns with unduly high values. As a result, the measures of other patterns are very small and their interpretations become rather misleading.

In light of such shortcoming, this paper presents a novel link-based similarity measure that derives a qualitative similarity description from multiple link characteristics each expressed using the absolute order-of-magnitude model (Piera 1995). In essence, these properties are perceived at different precision levels, and hence being gauged in accordance to distinct orders of magnitude spaces. With different sets of measurement labels (i.e. landmarks), these scales differ by at least one qualitatively important order of magnitude. Particularly, a semi-supervised method is introduced to select data-driven landmarks, which are more reliable than those human-directed ones. In order to combine measures of multiple link properties, the homogenization of such references (Agell, Rovira, & Ansotegui 2000) is required to realize the ultimate similarity description, where relevance of properties is proficiently blended within the aggregation process.

The rest of this paper is organized as follows. Section 2 introduces the absolute order-of-magnitude model upon which the present research is developed. Following that, Section 3 describes link properties and order-of-magnitude based similarity evaluation. Section 4 presents the semi-supervised method for designing landmarks, which is data-driven and more robust than the human-directed counterpart. The experimental evaluation of this qualitative link-based similarity measure to detecting the use of false identity is detailed in Section 5. The paper is concluded in Section 6, with the perspective of further work.

Absolute Order of Magnitude Model

The absolute order of magnitude (AOM) model (Piera 1995) operates on a finite set of ordered labels or qualitative descriptors achieved via a partition of the real number line $\mathcal{R}$. Each element of the partition represents a basic qualitative class to which a label is associated. The number of labels selected to express each variable of a real problem is subject to both the characteristics and the precision level required to support comprehension and communication. In practice, multiple label sets with dissimilar granularities are typically utilized to define domain attributes qualitatively.

Despite the intuition that the number of labels is not fixed, the most conventional partitions are symmetric. That is, the partition of the underlying domain typically has n positive and n negative labels, which is formally represented by $OM(n)$, and referred to as the AOM model of granularity n. The real-line partition into $2n+1$ labels is dictated by the set of $2n-1$ landmarks. In essence, landmarks are domain dependent and determined by either subjective justification of human experts or learning from data. For instance, the $OM(3)$ model is built on the following set of landmarks: $\{-\beta, -\alpha, 0, \alpha, \beta\}$. Figure 1 illustrates the resulting partition into seven qualitatively distinct order-of-magnitude labels, which are the most commonly used: Negative Large (NL), Negative Medium (NM), Negative Small (NS), Zero (0), Positive Small (PS), Positive Medium (PM) and Positive Large (PL) (Olmo *et al.* 2007).

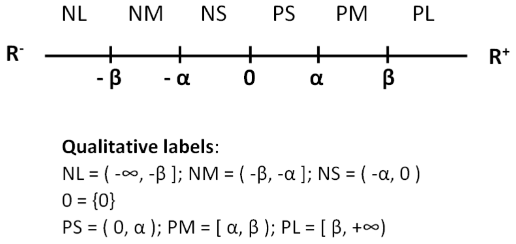


Figure 1: The $OM(3)$ absolute partition.

Order of Magnitude Space

An order of magnitude (OM) space S defined for a qualitative variable is the combination of the ordered label set S_l and the interval-like treatment of such labels. For instance, the value of one variable is expressed by the set of basic labels $S_l = \{B_1, \ldots, B_n\}$ with $B_1 < \ldots < B_n$ denoting its qualitative order, meaning that $\alpha < \beta, \forall \alpha \in B_i, \beta \in B_j, i < j$. The corresponding OM space S is formally described as $S = S_l \cup \{[B_i, B_j] | B_i, B_j \in S_l, i < j\}$. In effect, the label $[B_i, B_j]$ with $i < j$ is defined as the union of the elements within the set $\{B_i, B_{i+1}, \ldots, B_j\}$. In addition, the order in S_l induces the partial order $\leq_p$ in S, which represents *being more precise than* or *being less general than*:

$$[B_i, B_j] \leq_p [B_p, B_q] \iff [B_i, B_j] \subset [B_p, B_q] \qquad (1)$$

where $[B_i, B_i] = \{B_i\}$. According to Figure 2, the least precise label is $[B_1, B_n]$, denoted by ?. This manipulation of ordered labels allows reasoning and analysis with single or combined labels that may reflect uncertainty of one agent on another agent's judgement.

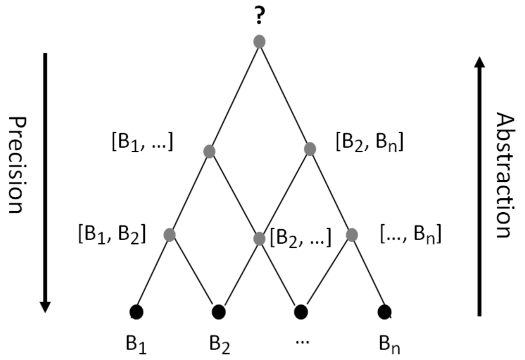


Figure 2: The graphical illustration of the partial order relation $\leq_p$ in an order-of-magnitude space S.

It is possible to define qualitative equality, termed q-equal, in an $OM(n)$ space S. Given $O, P \in S$, O and P are q-equal or $O \approx P$, if there is a $Q \in S$ such that $Q \leq_p O$ and $Q \leq_p P$. This effectively implies that O and P encompass, in part or in full, common basic elements. In addition, for presentational simplicity, $\forall O \in S$, the sets $B_O = \{B \in S_l - \{0\}, B \leq_p O\}$ and $B^*_O = \{B \in S_l, B \leq_p O\}$ are termed the *base of O* and the *enlarged base of O*, respectively.

Qualitative Algebra of AOM

At the outset, the mathematical structure of the AOM model, called Qualitative Algebra or Q-algebra, was initially defined as the unification of sign and interval algebra over a continuum of qualitative partitions of the real line (Travé-Massuyès & Piera 1989). However, although being superior to the sign algebra, such qualitative operators usually produce ambiguous and indeterminate outcomes. Accordingly, this barrier has been tackled via the notion of *qualitative expression of a real operator* (Agell, Rovira, & Ansotegui 2000). In particular, qualitative operators are considered as multidimensional functions defined in an AOM space. The Cartesian product of $S^1, S^2, \ldots, S^k$ (where k is the number of variables of a given problem domain, S^i is an $OM(n)$ space, $i = 1 \ldots k$) is adopted to express the outcome of a real operator in $\mathcal{R}^k$ qualitatively, which is reflected onto the resulting qualitative space $S^{'}$.

Given a real operator ω defined on $\mathcal{R}^k$ involving k real variables with each taking values in $\mathcal{R}$, the corresponding qualitative abstraction of ω, denoted as $[\omega]$, is specified on S^k with values in $S^{'}$ as follows:

$$[\omega](X_1, X_2, \ldots, X_k) = [\omega(X_1, X_2, \ldots, X_k)]_{S'} \qquad (2)$$

where $X_i \in S^i, i = 1 \ldots k$ and $\omega(X_1, X_2, \ldots, X_k) = \{\omega(x_1, x_2, \ldots, x_k), x_i \in X_i\}$. Inherently, $[\omega]$ assigns to each k-tuple element of $(X_1, X_2, \ldots, X_k)$ a qualitative description of the subset enclosing all underlying numerical results of applying ω over all real values in $X_1, X_2, \ldots, X_k$.

To simplify this, it is feasible to generate the qualitative operator, $[\omega]$, from the basic ordered labels of an OM space, S, $S^i = S, \forall i = 1 \ldots k$. For any $[\omega]$ and $X_1, X_2, \ldots, X_k \in S$:

$$[\omega](X_1, X_2, \ldots, X_k) = \bigcup_{B_i \in B^*_{X_i}} [\omega](B_1, B_2, \ldots, B_k) \qquad (3)$$

According to Equation 2, the qualitative operator $[\omega]$ can be generalized as follows:

$$[\omega](X_1, \ldots, X_k) = \bigcup_{B_i \in B^*_{X_i}} [\omega(B_1, \ldots, B_k)]_{S'} \qquad (4)$$

It is noteworthy that the $[\omega]$ operator presented above is compatible only to variables specified in the same order of magnitude space. To enhance the applicability of this terminology, the utilization of this qualitative operator is further introduced to multi-granularity domains via the homogenization of references, which has been successfully applied to realistic problems like credit risk prediction (Agell, Rovira, & Ansotegui 2000) and marketing segmentation (Olmo *et al.* 2007). This intuitive technique is extensively used in the current research, which will be thoroughly discussed below.

Order-of-Magnitude Based Link Analysis

This section introduces a novel order-of-magnitude based link analysis in which multiple link properties are combined to improve the quality of estimated link-based similarity measures.

Link Properties

Link analysis is based on examining relation patterns amongst references of real-world entities, which can be formally specified as an undirected graph $G(V, E)$. It is composed of two sets, the set of vertices V and that of edges E, respectively. Let X and R be the sets of all references and their relations in the dataset. Then, vertex $v_i \in V$ denotes reference $x_i \in X$ and each edge $e_{ij} \in E$ linking vertices $v_i \in V$ and $v_j \in V$ corresponds to a relation $r_{ij} \in R$ between references $x_i \in X$ and $x_j \in X$. Each edge $e_{ij} \in E$ possess statistical information $f_{ij} \in \{1, \ldots, \infty\}$, representing the frequency of any relation occurring between references x_i and x_j within the underlying dataset. With this terminology, several methods have been introduced to evaluate the similarity between information objects: SimRank (Jeh & Widom 2002), Connected-Triple (Klink *et al.* 2006), PageSim (Lin, King, & Lyu 2006) and a variety of random walk methods (Minkov, Cohen, & Ng 2006) (see more details in (Getoor & Diehl 2005) and (Liben-Nowell & Kleinberg 2007)).

Cardinality Property (CT) In essence, existing techniques, such as SimRank and Connected-Triple, have concentrated exclusively on the numerical count of shared neighboring objects. Let $v_i \in V$ be an entity of interest (e.g. a terrorist name in intelligence data or a paper in a publication database) and $N_{v_i} \subset V$ be a set of entities directly linked to v_i, called neighbors of v_i. The similarity between entities v_i and v_j is then determined by the cardinality of $N_{v_i} \cap N_{v_j}$, the set of shared neighbors where N_{v_i} and N_{v_j} are sets of neighbors of entities v_i and v_j, respectively. Effectively, the higher the cardinality is, the greater the similarity of these entities becomes.

Uniqueness Property (UQ) Despite their simplicity, cardinality based methods are greatly sensitive to noise and often generate a large proportion of false positives (Klink *et al.* 2006). This shortcoming emerges because these methods exclusively concern with the cardinality property of link patterns without taking into account the underlying characteristics of a link itself. As the first attempt to extend this approach by addressing such characteristics, the *uniqueness measure* of link patterns has been suggested as the additional criterion to CT to refine the estimation of similarity values (Boongoen & Shen 2008).

Given a graph $G(V, E)$ in which objects and their relations are represented with members of the sets of vertices V and edges E, respectively, a uniqueness measure UQ_{ij}^k of any two objects i and j (denoted by vertices $v_i, v_j \in V$) can be approximated from each joint neighbor k (denoted by the vertex $v_k \in V$) as follows:

$$UQ_{ij}^k = \frac{f_{ik} + f_{jk}}{\sum_m f_{mk}} \tag{5}$$

where f_{ik} is the frequency of the link between objects i and k occurring in data, f_{jk} is the frequency of the link between objects j and k, and f_{mk} is the frequency of the link between object k and any object m.

To summarize the uniqueness of joint link patterns UQ_{ij} between objects i and j, the ratios estimated for each shared neighbor are aggregated as

$$UQ_{ij} = \frac{1}{n} \sum_{k=1}^{n} UQ_{ij}^k \tag{6}$$

where n is the number of overlapping neighbor objects that objects i and j are commonly linked to.

Link Based Similarity Evaluation

A common drawback of those numerical measures previously presented is the inability to achieve coherent and natural interpretation through existing seemingly fine-grained scales. Exploring a link network with crisp numerically-valued criteria is typically considered inflexible comparing to the use of interval and linguistic descriptors. Specifically, a wrong interpretation of a property measure may occur if there exists a unduly high property value within a link network. A more accurate and naturally expressive measure is to exploit qualitative labels like highly, moderately or poorly certain.

In order to overcome this important shortcoming, measures of link properties like cardinality and uniqueness are gauged in accordance with property-specific order-of-magnitude (OM) spaces. Subsequently, the link-based similarity value is derived by combining these qualitative descriptors each assigned with a possibly different degree of relevance. Homogenizing of references in multi-granularity OM spaces (Agell, Rovira, & Ansotegui 2000) is applied to this aggregation process in such a way that values measured in distinct scales can be analogously manipulated.

OM Spaces for Link Properties At the outset, measures of link properties, originally in quantitative terms, are translated into elements of ordered label sets. Formally, let P^i and L^i be the set of intervals partitioned on the real line and that of the corresponding qualitative labels, defined for measures of the link property i on the discourse U^i. That is, $P^i = \{p_1^i, \ldots, p_{n^i}^i\}$ and $L^i = \{l_1^i \ldots l_{n^i}^i\}$, where n^i is the number of intervals/labels and $l_1^i < \ldots < l_{n^i}^i$ denotes the qualitative orders of magnitude specified for property i. Without causing confusion, for simplicity, intervals partitioned on real number line are termed partitions. They are non-overlapped over the discourse U^i, and their crisp boundaries are determined by one or two members of the landmark set $M^i = \{m_1^i, \ldots, m_{n^i-1}^i\}$. Each partition p_j^i is qualitatively expressed by the label $l_j^i, \forall j = 1 \ldots n^i$, and its interval is defined by lower bound α_j^i and/or upper bound β_j^i such that $\alpha_j^i, \beta_j^i \in M^i$ and $\alpha_j^i \leq \beta_j^i$.

Intuitively, the number of labels should be small enough so as not to impose useless precision onto analysts, but it must be rich enough to allow meaningful assessment and discrimination of measurement (Herrera & Herrera-Viedma 2000). In fact, average human beings can reasonably manage to bear in mind seven or so items/labels (Miller 1956).

For the current research with $i \in \{CT, UQ\}$, as a simple example, measures of the cardinality property over the discourse $U^{CT} = [0, \infty)$ may be described using a member of the label set of three qualitative labels ($n^{CT} = 3$), $L^{CT} = \{l_1^{CT} = Small, l_2^{CT} = Medium, l_3^{CT} = Large\}$. In particular, if the landmark set $M^{CT} = \{m_1^{CT} = 2, m_2^{CT} = 6\}$, members of the partition set are specified as $P^{CT} = \{p_1^{CT} = [0, 2], p_2^{CT} = (2, 6], p_3^{CT} = (6, \infty)\}$. Likewise, the uniqueness measure, whose values can be defined on the universe of discourse $U^{UQ} = [0, 1]$, which may be expressed using the ordered set of five qualitative descriptors ($n^{UQ} = 5$), $L^{UQ} = \{l_1^{UQ} = VeryLow, l_2^{UQ} = Low, l_3^{UQ} = Moderate, l_4^{UQ} = High, l_5^{UQ} = VeryHigh\}$. Using the set of landmarks ($M^{UQ} = \{m_1^{UQ} = 0.1, m_2^{UQ} = 0.3, m_3^{UQ} = 0.6, m_4^{UQ} = 0.8\}$), the corresponding partition set can be defined as $P^{UQ} = \{p_1^{UQ} = [0, 0.1], p_2^{UQ} = (0.1, 0.3], p_3^{UQ} = (0.3, 0.6], p_4^{UQ} = (0.6, 0.8], p_5^{UQ} = (0.8, 1]\}$.

Similarity Measure via Aggregation of Properties Relying on one particular link property, as with existing link-based methods, for justifying the similarity between any two objects in a link network may lead to false interpretation and perhaps revolting consequences. The more rational alternative is to integrate all available link properties in order to refine the similarity measure. Fortunately, the link-based similarity between any two vertices $v_a, v_b \in V$ in the link network can be estimated through the aggregation of qualitative descriptors each corresponding to a particular link property i. In particular, each property i can be assigned with a different degree of relevance (e.g. importance) RV^i, which may be given by domain experts in according with their past experiences or estimated from past data if such expertise is not readily available. Similar to measures of link properties previously emphasized, relevance can be naturally expressed using the order-of-magnitude label set L^{RV}, such as $L^{RV} = \{None, +, ++, +++\}$ or $L^{RV} = \{0, 1, 2, 3\}$. In the discussion above, the relevance degrees of cardinality $RV^{CT} \in L^{RV}$ and uniqueness properties $RV^{UQ} \in L^{RV}$ are subjectively set to 2 and 1, respectively.

However, since label sets defined for different properties are usually of unequal granularity, they have to be homogenized onto a common scale on which references of distinct label sets can be uniformly manipulated and integrated. Following the work of (Agell, Rovira, & Ansotegui 2000), the

Table 1: Homogenized landmarks.

Landmarks	CT	UQ
Original	2, 6	0.1, 0.3, 0.6, 0.8
Step1	0, 4	-0.2, 0, 0.3, 0.5
Step2	-4, 0, 4	-0.5, -0.3, -0.2, 0, 0.2, 0.3, 0.5
Step3	-4, -2, -1, 0, 1, 2, 4	-0.5, -0.3, -0.2, 0, 0.2, 0.3, 0.5
Homogenized	-3, -2, -1, 0, 1, 2, 3	-3, -2, -1, 0, 1, 2, 3
Irrelevant	-3, -2, -1, 1, 2	-3, -2, 1

procedure below will be used here:

- *Step1*: Convert each set of landmarks M^i into a symmetric arrangement. Given a central landmark $m_c^i \in M^i$, translate each landmark $m_t^i, t = 1 \dots n^i - 1$ to the new landmark sm_t^i in the symmetric scale using $sm_t^i = m_t^i - m_c^i$. Note that the central landmark is now 0 in the new scale.
- *Step2*: Landmarks appearing on both positive and negative sides may be dissimilar in general. A fully symmetric pattern can be achieved by adding missing landmarks, so that one absolute landmark can be found on both positive and negative sides of 0. Obviously, these newly added elements are of balancing purpose only, therefore they will not be used to represent values and will be deliberately marked as irrelevant.
- *Step3*: The landmark sets for each property are further modified by adding new landmarks on both side of 0, in such a way that all landmark sets have the same cardinality. Similar to Step 2, new elements are irrelevant with respect to each particular property and are simply to support the unification mechanism.

In accordance to the landmarks of two link properties given earlier, Table 1 summarizes the results achieved at each step of the homogenization process.

Following the terminology of AOM algebra, with the property-specific relevance degrees previously clarified, order-of-magnitude based similarity measure (OMS) can be estimated from measures of any n properties using the qualitative expression of a real weighted summation $[\omega]$:

$$\begin{aligned} OMS &= [\omega](X_1, \dots, X_n, RV_1, \dots, RV_n) \\ &= [\omega(X_1, \dots, X_n, RV_1, \dots, RV_n)]_{S^{Sum}} \end{aligned} \quad (7)$$

where $X_i \in S^H$ is the qualitative measure of link property $i, i = 1 \dots n$, expressed on the homogenized scale S^H, RV_i is its corresponding relevance degree, S^{Sum} is the resulting order-of-magnitude space of this summarization and ω is defined as

$$\begin{aligned} \omega(X_1, \dots, X_n, RV_1, \dots, RV_n) &= \omega(x_1, \dots, x_n, rv_1, \dots, rv_n) \\ &= x_1 rv_1 + \dots + x_n rv_n \end{aligned} \quad (8)$$

where $x_i \in X_i, rv_i \in RV_i, i = 1 \dots n$.

Specific to the two link property measures used herein: CT and UQ, with their relevance degrees being RV^{CT} and RV^{UQ} and the homogenized scale S^H being {-3, -2, -1, 0, 1, 2, 3}, the previous equations can be employed as follows:

$$\begin{aligned} OMS &= [\omega](CT, UQ, RV^{CT}, RV^{UQ}) \\ &= [\omega(CT, UQ, RV^{CT}, RV^{UQ})]_{S^{Sum}} \end{aligned} \quad (9)$$

Following that

$$OMS = [\omega(2ct + uq)]_{S^{Sum}} \quad (10)$$

where $ct \in M^{CT}$, and M^{CT} is the set of relevant landmarks of CT in the homogenized scale S^H: $M^{CT} = \{0, 3\}$. Likewise, uq is a member of M^{UQ}, with M^{UQ} = {-1, 0, 2, 3}. Effectively, the resulting order-of-magnitude space S^{Sum} is established upon landmark values of this qualitative operation, which are $\{-1, 0, 2, 3, 5, 6, 8, 9\}$. To obtain a coherent interpretation of similarity measures within the S^{Sum} space, a set of qualitative labels L^{OMS}, as partitions of S^{Sum}, is chosen to express the different orders of magnitude of the similarity values. For instance, $L^{OMS} = \{Low\ (OMS < 2), Medium\ (2 \leq OMS \leq 6), High\ (OMS > 6)\}$. Note that a more or less refined label sets can be used depending on the precision level required.

Semi-Supervised Method to Designing Landmarks

Designing an appropriate set of landmarks M^i for a link property i is non-trivial and proves to be critical towards the quality of generated similarity measures. A simple approach is to rely on human experts, who select suitable landmark values in accordance with their personal intuition and judgment. This is not usually effective regarding the availability of experts and the diverse nature of different problem domains. Besides, human input may be rather subjective and inconsistent. As a result, a data-driven mechanism that can be used to obtain an appropriate M^i is specifically discussed herein.

For a link property i, a density graph is formulated to represent the proportion of entity pairs (i.e. (v_x, v_y), $v_x, v_y \in V$), each with different property measure i_{xy}. Let $D : [0, i_{max}] \rightarrow [0, 1]$ be the density function (where i_{max} denotes the maximum value of i_{xy}), which is formally defined as

$$D(t) = \frac{N(t)}{\sum\limits_{\forall r \in [0, i_{max}]} N(r)} \tag{11}$$

where $N(t)$ denotes a number of entity pairs (v_x, v_y) whose property measure $i_{xy} \geq t, t \in [0, i_{max}]$. Figure 3 presents the density function of cardinality property (i.e. $i = CT$) derived from the Terrorist dataset (Hsiung *et al.* 2005), where $CT_{max} = 113$ (and the magnified presentation of $D(t), t \in \{7, 113\}$ is included for better interpretation).

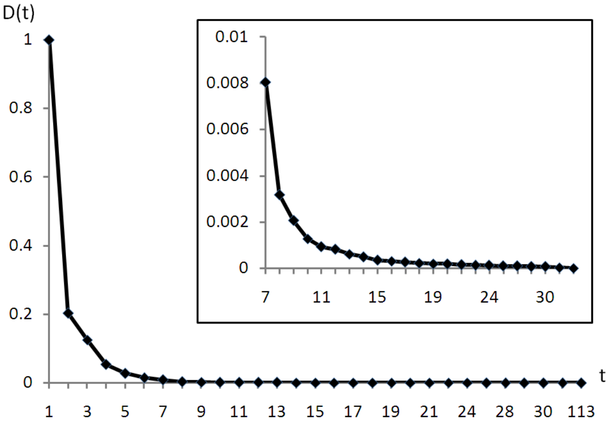


Figure 3: Example of density function derived from Terrorist dataset.

With this function, the following set of heuristics can be articulated especially to help data analysts to assess a proper set of landmarks M^i for link property i:

- Let $M^i = \{m_1^i, m_2^i, \ldots, m_{n^i}^i\}$ be an appropriate landmark set for property i, where $m_g^i \leq i_{max}, \forall g \in \{1 \ldots n^i\}$ and $m_h^i \leq m_{h+1}^i, \forall h \in \{1 \ldots n^i - 1\}$.
- Each pair of adjacent landmarks (i.e. m_h^i and m_{h+1}^i) encapsulates all property values $i_{xy} \in [m_h^i, m_{h+1}^i)$ whose density $D(i_{xy})$ can be perceived at a particular order of magnitude. Note that orders of magnitude utilized in this research are of $\alpha \times 10^z$, where $z \in \{-1, -2, \ldots, -\infty\}$ and $\alpha \in (0, 10)$. According to Figure 3, M^{CT} of the Terrorist dataset is $\{4, 7, 10, 23\}$ such that $D(CT_{xy})$ is expressed at five different orders of magnitude of
  - 10^{-1} where $CT_{xy} < 4$
  - 10^{-2} where $4 \leq CT_{xy} < 7$
  - 10^{-3} where $7 \leq CT_{xy} < 10$
  - 10^{-4} where $10 \leq CT_{xy} < 23$
  - 10^{-5} where $CT_{xy} \geq 23$

This semi-supervised method is effective to assist analysts to design appropriate landmarks and descriptive labels, based on quality measures of the particular link network being studied. Unlike human-directed alternatives, it is data oriented and capable of being adapted to a variety of problems.

Application to False Identity Detection

This section presents the application of the order-of-magnitude link-based similarity evaluation to detecting the use of false identities. Particularly, its performance is empirically evaluated over the terrorism-related dataset, and further generalized with a publication data collection.

False Identity Detection

To battle false identity, an exact-match query to a law enforcement computer system is simply ineffective. A better approach extensively studied in (Bilenko & Mooney 2003) and (Wang *et al.* 2006) is to exploit the similarity measure of names obtained from one or several string-matching techniques. Despite their reported success, these *content-based* methods can not handle cases where completely different names are deployed. For instance, they would fail to recognize the association between these pairs of terrorists' name, whose overlapping text content is void.

- (*ashraf refaat nabith henin, salem ali*)
- (*bin laden, the prince*)
- (*bin laden, the emir*)
- (*abu mohammed nur al-deen, the doctor*)

Accordingly, the *link-based* approach, which has proven effective for similar problems in a wide range of domains (e.g. publication (Klink *et al.* 2006), online resources (Hou & Zhang 2003), (Lin, King, & Lyu 2006), email (Minkov, Cohen, & Ng 2006) and intelligence data analysis (Hsiung *et al.* 2005)), has been put forward to underpin the accountability for unstructured information.

Let O be the set of real-world entities each being referred to by at least one member of another set X, which is a collection of names or references. A pair of names (x_i, x_j) are aliases when both names correspond to the same real-world entity: $(x_i \equiv o_k) \wedge (x_j \equiv o_k), o_k \in O$. In practice, disclosing an alias pair in graph G is to find a couple of vertices (v_i, v_j), whose similarity $s(v_i, v_j)$ is significantly high. Intuitively, the higher $s(v_i, v_j)$ the greater the possibility that vertices v_i and v_j, and hence corresponding names x_i and x_j, constitute the actual alias pair.

Datasets

The performance and applicability of the proposed approach is evaluated over the following distinct datasets: Terrorist (Hsiung *et al.* 2005) and DBLP (Klink *et al.* 2006). Terrorist is a link dataset manually extracted from web pages and news stories related to terrorism. Each node presented in this link network is a name of person, place or organization, while a link denotes a co-occurrence association between objects through reported events. Figure 4 presents an example of this link network where names *Bin laden* and *Abu abdallah* refer to the same real-world person.

DBLP (Digital Bibliography and Library Project) is the dataset containing co-authoring information extracted from different bibliographical databases. In this link network, each node represents a reference name of an author and a link denotes the fact that two names appear as the co-authors

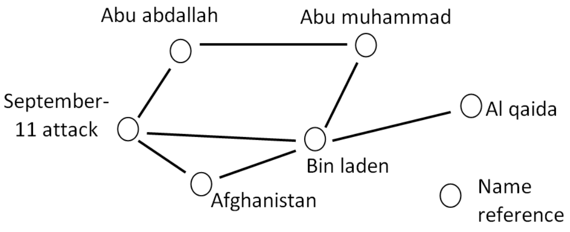


Figure 4: An example of Terrorist dataset.

of a paper (or papers). Table 2 summarizes the number of links, objects and alias pairs included in these datasets.

Table 2: Dataset details (number of objects, links and alias pairs).

Dataset	Objects	Links	Alias Pairs
Terrorist	4088	5581	919
DBLP	2796	8157	23

Performance Evaluation

Efficiency of Semi-Supervised Method Initially, it is important to examine the effectiveness of the proposed semi-supervised method for modeling a landmark set. By following the heuristics previously prescribed, appropriate landmark values are:

- For Terrorist dataset, $M^{CT} = \{4, 7, 10, 23\}$ and $M^{UQ} = \{0.05, 0.12, 0.27, 0.43, 1\}$.
- For DBLP dataset, $M^{CT} = \{2, 5, 9, 15\}$ and $M^{UQ} = \{0.008, 0.04, 0.17, 0.31, 1\}$.

With these data-oriented landmarks, Table 3 compares the number of disclosed alias pairs successfully detected by different methods, where K denotes the number of entity pairs with highest similarity measures (details of homogenization for semi-supervised landmarks are not included due to space limitation). Note that OMS and OMS^H represent order-of-magnitude based similarity measures, with semi-supervised and human-directed landmarks, respectively. In addition, QT denotes a simple integration of numerical CT and UQ measures, where relevance degrees RV^{CT} and RV^{UQ} (2 and 1, respectively) similar to those of OMS and OMS^H are employed.

These results indicate that the OMS measure with semi-supervised landmarks usually outperforms both human-directed landmarks OMS^H and the quantitative evaluation QT, especially over Terrorist dataset.

Comparison with Alternative Link-Based Methods The performance of the OMS method is further generalized by evaluating against the following two state-of-the-art link-based measures: SimRank (SR) and PageSim (PS), respectively.

Table 3: Number of alias pairs disclosed by each method.

K	OMS	OMS^H	QT
Terrorist			
200	43	9	8
400	80	57	41
600	115	91	60
800	146	110	75
1000	180	138	102
DBLP			
100	4	1	1
200	5	2	1
300	5	3	2
400	6	5	4
500	10	6	5

- Principally, with the objective of finding similar publications given their citation relations, SimRank relies on the cardinality of shared neighbors that are iteratively refined to a fixed point (Jeh & Widom 2002). In each iteration, the similarity of any pair of vertices $v_i, v_j \in V$, $s(v_i, v_j)$, is approximated as

$$s(v_i, v_j) = \frac{C \sum_{p=1}^{|N_{v_i}|} \sum_{q=1}^{|N_{v_j}|} s(N^p_{v_i}, N^q_{v_j})}{|N_{v_i}||N_{v_j}|} \tag{12}$$

  where $N_{v_i}, N_{v_j} \subset V$ are sets of neighboring vertices to which vertices v_i and v_j are linked, respectively. Individual neighbors of both vertices are denoted as $N^p_{v_i}$ and $N^q_{v_j}$, for $1 \leq p \leq |N_{v_i}|$ and $1 \leq q \leq |N_{v_j}|$. The constant $C \in [0, 1]$ is a decay factor that represents the confidence level of accepting two non-identical entities to be similar. Note that $s(v_i, v_j) = 0$ when $N_{v_i} = \emptyset$ or $N_{v_j} = \emptyset$.

- Within a different domain, PageSim (Lin, King, & Lyu 2006) was developed to capture similar web pages based on associations implied by their hyperlinks. In essence, the similarity measure $ps(v_i, v_j)$ between vertices v_i and v_j is dictated by the coherence of ranking scores $R(v_g, v_i)$ and $R(v_g, v_j)$ propagated to them from any other vertex $v_g \in V$. It is noteworthy that ranking scores are explicitly generated using the page ranking scheme, PageRank (Brin & Page 1998), of the most developed Google search engine (with detailed computational mechanism for the ranking scores omitted here). Formally, PageSim can be defined as

$$ps(v_i, v_j) = \sum_{\forall v_g \in V, v_g \notin \{v_i, v_j\}} \frac{\min(R(v_g, v_i), R(v_g, v_j))^2}{\max(R(v_g, v_i), R(v_g, v_j))} \tag{13}$$

According to Table 4, the OMS measure consistently outperforms other link-based methods over both datasets. In spite of its low performance, the SimRank measure, which has been recognized as a benchmark link analysis technique for publication (Getoor & Diehl 2005) and Internet (Calado *et al.* 2006) domains, is included in this evaluation as to reflect the difficulty of this task. Based on the results presented in Tables 3-4, the proposed method does encounter the problem of false positives. However, its performance with respect to this difficulty has been substantially improved as compared to other link-based similarity methods.

Table 4: Number of alias pairs disclosed by each method.

K	OMS	SR	PS
Terrorist			
200	43	0	7
400	80	0	36
600	115	1	63
800	146	1	79
1000	180	2	92
DBLP			
100	4	0	1
200	5	1	1
300	5	2	1
400	6	2	2
500	10	3	4

Computational Complexity In addition to evaluating these methods in terms of discovered alias pairs, it is important to investigate the computational complexity that would determine or even limit their actual real-world applications. Let a link network consist of n distinct entities, each averagely linked to other m entities. The time complexity for the OMS approach to generate all pair-wise similarity values is $O(n^2m^2)$. With f iterations of similarity refinement, the time complexity of SimRank is $O(n^2m^2f)$. Note that the results shown in Table 4 are obtained using $f = 3$ (with its usual range being 3-5).

In contrast, the PageSim is rather complex compared to the others as it begins with ranking all entities using the PageRank technique, whose time complexity is $O(nmt)$ where t is the number of iterations for refining the ranking values (with t being 3 in this experiment). Having accomplished the ranking process, the similarity of two entities is estimated on the ranking values propagated from their shared neighbors, with the maximum connecting-path length of r (r set to 3 for the results given in Table 4). As a result, the overall time complexity of PageSim method is $O(n^2m^{2r} + nmt)$.

Hence, the OMS method introduced in this paper not only performs well in terms of precision, but also proves to be practical for alias detection, with efficient time consumption.

Conclusion

This paper has presented a novel qualitative link-based similarity measure, which can be efficiently employed for intelligence data analysis and disclosing the use of false identity typically appearing in terrorists and criminals' activities. Unlike initial numerical similarity estimation that concentrates solely on the link structures, the qualitative method also includes underlying link properties such as uniqueness in order to purify the similarity description. In addition, qualitatively distinct order-of-magnitude labels incorporate semantics towards similarity justification and allow coherent interpretation and reasoning that is hardly feasible with pure numerical terms.

Technically, measures of link properties are gauged in accordance with property-specific order of magnitude spaces, whose dissimilar scales are subsequently homogenized to permit the unification of their values. In essence, the similarity descriptor is achieved via aggregating property values regarding to their relative degrees of relevance. Empirically, this qualitative approach consistently outperforms numerical similarity measures over terrorism-related and publication datasets. However, in order to generalize its performance and applicability, it is crucial to evaluate this method with more relevant data. Also, relevance degrees allocated for distinct link properties may be better learned from data, instead of relying on human-directed ones.

Acknowledgments

This work is sponsored by the UK EPSRC grant EP/D057086. The authors are grateful to team members for their contribution, whilst taking full responsibility for the views expressed in this paper.

References

Agell, N.; Rovira, X.; and Ansotegui, C. 2000. Homogenising references in orders of magnitude spaces: An application to credit risk prediction. In *Proceedings of International Workshop on Qualitative Reasoning*, 1–8.

Badia, A., and Kantardzic, M. M. 2005. Link analysis tools for intelligence and counterterrorism. In *Proceedings of IEEE International Conference on Intelligence and Security Informatics, Atlanta*, 49–59.

Bilenko, M., and Mooney, R. J. 2003. Adaptive duplicate detection using learnable string similarity measures. In *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 39–48.

Boongoen, T., and Shen, Q. 2008. Detecting false identity through behavioural patterns. In *Proceedings of Int. Crime Science Conference, London*.

Brin, S., and Page, L. 1998. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems* 30(1-7):107–117.

Calado, P.; Cristo, M.; Gonçalves, M. A.; de Moura, E. S.; Ribeiro-Neto, B. A.; and Ziviani, N. 2006. Link based similarity measures for the classification of web documents. *Journal of American Society for Information Science and Technology* 57(2):208–221.

Getoor, L., and Diehl, C. P. 2005. Link mining: a survey. *ACM SIGKDD Explorations Newsletter* 7(2):3–12.

Herrera, F., and Herrera-Viedma, E. 2000. Linguistic decision analysis: steps for solving decision problems under linguistic information. *Fuzzy Sets and Systems* 115:67–82.

Hou, J., and Zhang, Y. 2003. Effectively finding relevant web pages from linkage information. *IEEE Transactions on Knowledge and Data Engineering* 15(4):940–951.

Hsiung, P.; Moore, A.; Neill, D.; and Schneider, J. 2005. Alias detection in link data sets. In *Proceedings of International Conference on Intelligence Analysis*.

Jeh, G., and Widom, J. 2002. Simrank: A measure of structural-context similarity. In *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 538–543.

Klink, S.; Reuther, P.; Weber, A.; Walter, B.; and Ley, M. 2006. Analysing social networks within bibliographical data. In *Proceedings of International Conference on Database and Expert Systems Applications*, 234–243.

Liben-Nowell, D., and Kleinberg, J. 2007. The link-prediction problem for social networks. *Journal of the American Society for Information Science and Technology* 58(7):1019–1031.

Lin, Z.; King, I.; and Lyu, M. R. 2006. Pagesim: A novel link-based similarity measure for the world wide web. In *Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence*, 687–693.

Miller, G. 1956. The magical number seven, plus or minus two: Some limits on our capacity for information processing. *Psychological Review* 63(2):81–97.

Minkov, E.; Cohen, W. W.; and Ng, A. Y. 2006. Contextual search and name disambiguation in email using graphs. In *Proceedings of Int. Conference on Research and Development in IR*, 27–34.

Olmo, C.; Sánchez, G.; Prats, F.; Agell, N.; and Sánchez, M. 2007. Using orders of magnitude and nominal variables to construct fuzzy partitions. In *Proceedings of IEEE International Conference on Fuzzy Systems*, 1–6.

Piera, N. 1995. Current trends in qualitative reasoning and applications. *Monografia CIMNE, 33. International Center for Numerical Methods in Engineering, Barcelona.*

Porter, G. 2008. Crying (iranian) wolf in argentina. *Asia Times Online (www.atimes.com).*

Travé-Massuyès, L., and Piera, N. 1989. The orders of magnitude models as qualitative algebras. In *Proceedings of 11th International Joint Conference on Artificial Intelligence*, 1261–1266.

Wang, A. G.; Atabakhsh, H.; Petersen, T.; and Chen, H. 2005. Discovering identity problems: A case study. In *Proceedings of IEEE International Conference on Intelligence and Security Informatics, Atlanta*, 368–373.

Wang, G. A.; Chen, H.; Xu, J. J.; and Atabakhsh, H. 2006. Automatically detecting criminal identity deception: an adaptive detection algorithm. *IEEE Transactions on Systems, Man and Cybernetics, Part A* 36(5):988–999.

QCM: A QP-Based Concept Map System

Morteza Dehghani and Ken Forbus

Qualitative Reasoning Group, Northwestern University
2133 Sheridan Road, Evanston, IL 60201 USA
{morteza, forbus}@northwestern.edu

Abstract

Qualitative representations have proven to be useful formalisms for capturing human mental models. As a result, qualitative modeling could become an important tool for cognitive science. Specifically, an environment in which qualitative representations can be used to explore mental models and different type of reasoning and simulations can be performed on these models can be a useful tool for cognitive scientists. In this paper, we introduce the Qualitative Concept Map system, designed for cognitive scientists, for building and simulating qualitative and Bayesian models using qualitative process theory and Bayesian inference.

Introduction

Qualitative representations capture the intuitive, causal aspects of many human mental models (Forbus & Gentner 1997). This includes aspects of modeling not handled by traditional formalisms, such as conditions of applicability and other types of modeling knowledge. Qualitative modeling could become an important tool for cognitive science, by providing formal languages for expressing human mental models. The qualitative reasoning community has explored a wide range of representations and techniques, pursuing its goal to capture the breadth of qualitative reasoning, ranging from the person in the street to the expertise of scientists (Forbus et al. 2004). A unified platform in which cognitive scientists can apply qualitative representations, explore mental models and be able to integrate these models with other forms of reasoning, can become a useful tool for cognitive scientists.

In this paper, we present the Qualitative Concept Map system (QCM) which provides a cognitive scientist friendly environment that allows modelers to explore qualitative models, incorporate them into probabilistic models and output them in formats usable in other forms of reasoning (e.g. analogical reasoning). An earlier version of this system was used to build models of transcript data (Dehghani, Unsworth, Lovett, & Forbus, 2007). These models were exported as predicated calculus statements which were used via analogical generalization to classify the models based on the culture and level of expertise of the participants. Since then, we have expanded the model in several ways. First, we integrated our qualitative simulator (Gizmo), to provide a complementary first-principles simulation engine. Second, we added a probabilistic reasoning mode. Finally, we enhanced the user interface functionality to provide easier access to reasoning features.

We first introduce our system, discuss its different features and describe some real-world cognitive science examples modeled in it. Next, we describe the qualitative mode of the system and Gizmo. We then describe the probabilistic mode and how information available in the qualitative mode can be integrated into the probabilistic mode. We close by discussing related and future work.

Qualitative Concept Map System

QCM is the first modeling tool which has been specifically designed for cognitive scientists. It provides a unified reasoning platform in which mental models can be constructed and analyzed using Qualitative Process theory (Forbus, 1984) and Bayesian Networks (Pearl 1988). QCM is connected to Gizmo, a full implementation of QP theory, for providing qualitative simulations, including envisionment. QCM also uses a Bayesian inference algorithm for calculating probabilities of evidence and posterior probabilities.

QCM uses a concept map interface (Novak & Gowin, 1984). For example, Figure 1 shows how QCM can be used to model the effects of fear on different properties of the self, and effects of external processes on these properties, as described in Jami ‘al Sa’adat (The Collector of Felicities) (al-Naraqi, 18th Century), an Islamic book of ethics written in the 18th century. QCM automatically checks for any modeling errors which violate the laws of QP theory and probability theory, providing detailed error messages. QCM can import and export models via GraphML (Brandes, Eiglsperger, Herman, Himsolt, & Marshall, 2001), allowing graphs drawn in QCM to be easily viewed in other graph drawing programs. This facilitates collaboration between modelers. More importantly, for cognitive simulation purposes, models can be exported as predicate calculus statements. This enables QCM models to be used in a variety of types of reasoning, such as analogical reasoning.

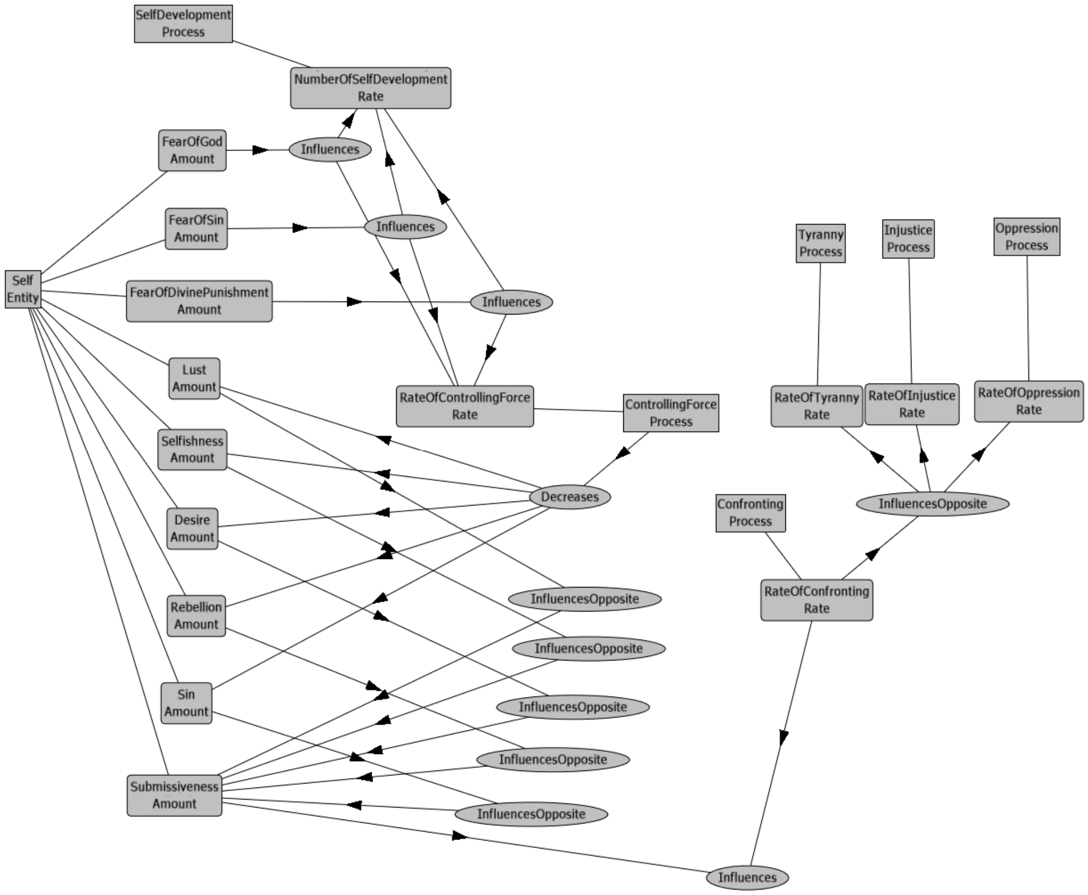


Figure 1: The Effects of Fear on Different Properties of the Self

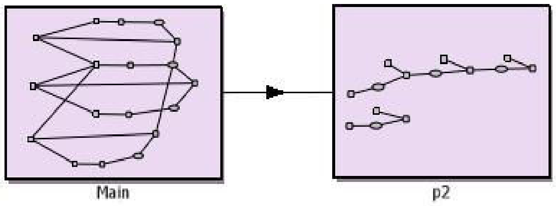


Figure 2: The Meta-Pane

QCM utilizes multiple panes to represent distinct qualitative states. This is important for capturing changes over time. For example, often modelers need to discuss immediate effects of a change followed by long-term effects of a change. The meta-pane (Figure 2) allows modelers to see all the states at once. Modelers can easily extend the vocabulary of specific processes and quantities used in the models, to expedite model creation.

QCM has been used for modeling a variety of different phenomena, from abstract models of religious beliefs to concrete qualitative reasoning scenarios. Figure 3 illustrates one pane from a model for the Bears Disappearing scenario modeled from transcript data gathered by psychologists from a native American group (Dehghani et al. 2007). Figure 4 shows the initial state of a heat transfer scenario and figure 6 is an example of Bayesian reasoning in QCM.

QP Modeling

QP theory as a representation language for physical phenomena includes:

- Continuous parameters (quantities)
- Causal relationships between them (influences)
- Mechanisms underlying physical causality (physical processes)

Systems and phenomena are modeled via sets of entities with continuous parameters, whose relationships are expressed using a causal, qualitative mathematics, where processes provide an explicit notion of mechanism. In QP

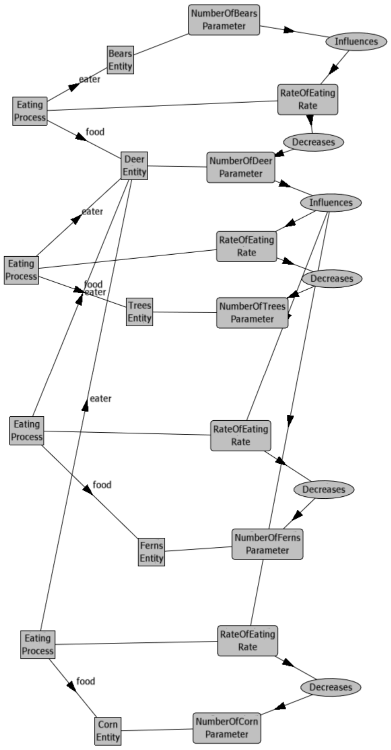


Figure 3: The Bears Disappearing Scenario Modeled from Transcript Data

theory direct influences are modeled using I+ (≡ *Increases*) and I- (≡ *Decreases*) which indicate an integral connection between two parameters, i.e., heat flow decreases the heat of its source and increases the heat of its destination. Indirect influences are modeled by $\propto_{Q+}$ (≡ *Influences*) and $\propto_{Q-}$ (≡ *InfluencesOpposite*) which indicate functional dependence between two parameters, i.e., the heat of something determines its temperature. Gizmo Mk2 is a full

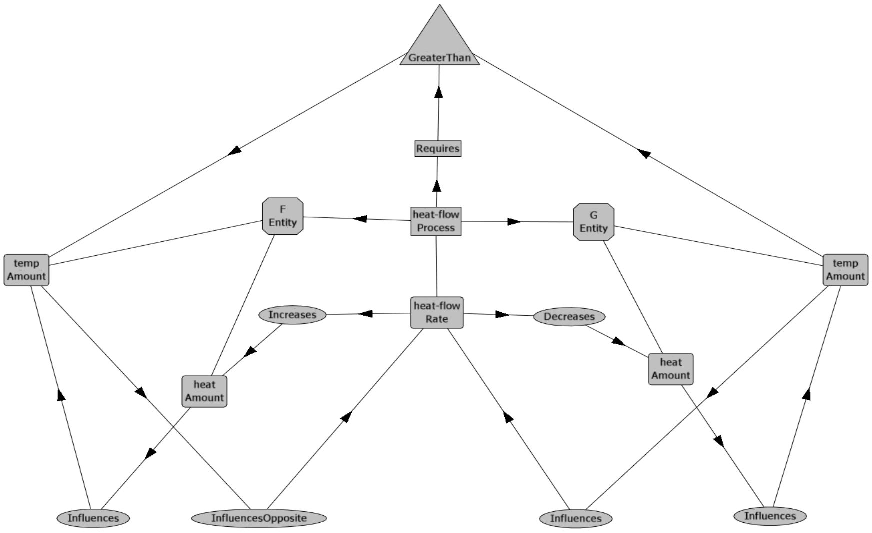


Figure 4: Heat-Transfer Scenario

```
;;; Quantity Functions
(defquantityfunction Rate (?thing))
(defquantityfunction heat-flow-rate (?Rate))
(defquantityfunction heat (?Amount))
(defquantityfunction Amount (?thing))
(defquantityfunction temp (?Amount))

;;; Entities
(defentity G-type
  :quantities    ((heat :type Amount)
                  (temp :type Amount))
  :consequences ((qprop (temp G-type)
                        (heat G-type)))
  :documentation "finite-thermal-physob")

(defentity F-type
  :quantities    ((heat :type Amount)
                  (temp :type Amount))
  :consequences ((qprop (temp F-type)
                        (heat F-type)))
  :documentation "finite-thermal-physob")

;;; Processes
(defprocess heat-flow
 :participants ((the-G :type G-type)
                (the-F :type F-type))
 :conditions   ((> (temp the-G) (temp the-F)))
 :quantities   ((heat-flow-rate :type Rate))
 :consequences ((i- (heat the-G) heat-flow-rate)
                (i+ (heat the-F) heat-flow-rate)
                (qprop (heat-flow-rate heat-flow)
                        (temp the-G))
                (qprop- (heat-flow-rate heat-flow)
                         (temp the-F))))
```

Figure 1: Domain Theory generated from the Heat-Transfer Scenario

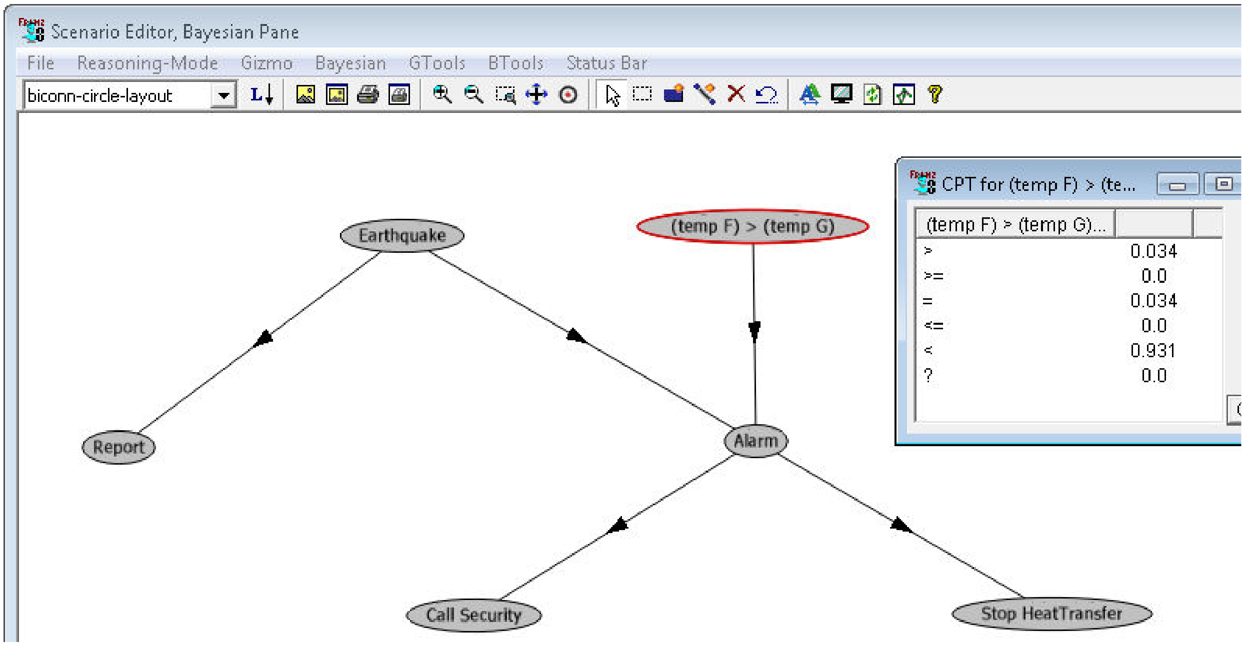
implementation of QP theory and works as the qualitative reasoning engine of QCM. Gizmo has been designed to be lightweight and incremental to be used as a module in larger systems. The user has tight control over the process of qualitative simulation in Gizmo. Algorithms for both total and attainable envisioning are included as well.

In order to provide support for novice modelers, the domain theory and the scenario of the model are automatically extracted from the graph and sent to Gizmo. This is extraction is performed by going over all the nodes in the graph and, for each node, determining the type of node it is (e.g. Entity, Process, Quantity). Based on this information, QCM automatically obtains the required information for that type of node from the graph and sends the information to Gizmo. The domain theory extracted for the heat-transfer model of Figure 4 is presented in Figure 5. If the system determines that the model is missing some required information, a detailed error message is presented to the modeler.

The automatic extraction of the domain theory and the scenario file is, we believe, a major boon to novice modelers. While many of the ideas of qualitative modeling come naturally to scientists, outside of computer science, experience in writing logically quantified formulae is rare. Modelers need motivation, and being able to get results without having to first write a general domain theory helps reduce the entry barrier. As their models become more complex, the automatically produced models can become a starting point for writing standard QP theory domain models.

Bayesian Modeling

Agents continually update their beliefs using different



(temp F) > (temp G)...	
>	0.034
>=	0.0
=	0.034
<=	0.0
<	0.931
?	0.0

Figure 6: A Bayesian Network

types of new information. These updates affect their causal beliefs about the uncertainties in the world. In order to model this process, we need a rich causal representation and a method for capturing and updating uncertain beliefs about the world. QP theory provides us with a high level of expressiveness needed to capture many intuitive, causal aspects of human cognition. One can use the QP framework to reason about relations between things and the effect of these relations on the state of the world. However, QP theory does not provide the mechanism necessary for capturing probabilities. Bayesian networks (Pearl, 1988) are the most widely used approach for probabilistic reasoning. This formalism provides a succinct representation for probabilities, where conditional probabilities can be represented and reasoned with in an efficient manner. Providing an interface in which both QP and Bayesian formalisms can be used in parallel can potentially be helpful for cognitive scientists.

QCM provides a framework in which the agent's knowledge about the causal structure of the world can be captured using the QP formalism, while the agent's uncertain knowledge and expectations about the outcomes of his/her actions can be captured by subjective probabilities and represented by a Bayesian Network. Modelers can switch the mode of reasoning from QP to Bayesian and make probabilistic models. This feature allows cognitive scientists to take advantage of different types of reasoning available in both formalisms. In the Bayesian mode, modelers can perform exact inference on the network and calculate the probabilities using Recursive Conditioning (RC) (Darwiche, 2001). RC is an any-space algorithm which works by recursively partitioning the network into smaller networks using conditioning and solving each subnetwork as an independent problem. Networks created in the Bayesian mode are saved in the Hugin format, which is the standard format for many data mining and machine learning programs. This again helps modelers who use QCM collaborate more easily with other scientists using other modeling programs.

Determining a Priori Probabilities using Qualitative Simulations

One of the main obstacles in probabilistic reasoning is finding the a priori probabilities of variables in the model. One approach to overcome this obstacle is to use qualitative simulations. QCM uses the information available in the QP mode to calculate a priori probabilities of quantities used in the qualitative model. In this framework, the probability distribution is defined over a set of possible worlds determined by the constraints of the qualitative model. If the modeler chooses to include a qualitative parameter, such as a quantity or a derivative, as a node in the probabilistic model, QCM can determine the probabilistic distribution of the values of that parameter by model counting. The idea is to calculate the degree of belief in that statement over all the possible worlds determined by qualitative envisionment. For example, if `(temp F) > (temp G)` relationship from the heat-transfer scenario of Figure 4 needs to be included as a node in the model, QCM performs an attainable envisionment determining in how many possible worlds `(temp F) β (temp F)` where `β={<,<=,=,>=,>,?}` hold to be true. Based on this measure a probability value can be assigned to `(temp F) > (temp G)` (see Figure 6 for an example of a Bayesian network which uses this relationship). In other words, we are saying that under the current constraints in *n* of *m* possible worlds `(temp F) > (temp G)`, therefore the probability of `(temp F) > (temp G)` is *n*/*m*. We believe this method can provide a robust way of calculating a priori probabilities for physical phenomena for which we can define a QP model for.

Related Work

QCM is a successor to VModel (Forbus et al. 2001). VModel was developed to help middle-school students learn science. Like QCM, it uses a subset of QP theory to provide strong semantics. However, VModel was limited to single-state reasoning, whereas QCM can be used to model continuous causal phenomena with multiple states. Similar differences hold with Betty's Brain (Biswas et al 2001), which provides a domain-specific concept map environment that students can use in learning stream ecology.

The closest other qualitative modeling tools are MOBUM (Machado & Bredeweg, 2001) and VISIGARP (Bouwer & Bredeweg 2001) which have lead to Garp3 (Bredweg et al 2006, Bredweg et al 2007). Like QCM, these environments are aimed at researchers, but their focus is on constructing models for qualitative simulation, including generic, first-principles domain theories. QCM focuses instead on helping capture concrete, situation-specific qualitative explanations of phenomena. Thus, it provides a useful tool for scientists working with interview data.

Different approaches for qualitative Bayesian inference have been proposed. These methods include: qualitative probabilistic networks (Wellman 1990), qualitative certainty networks (Parsons and Mamdani 1993) and a method which incorporates order of magnitude reasoning in qualitative probabilistic networks (Parsons 1995). Keppens (2007a, 2007b) employs some of these methods for qualitative Bayesian evidential reasoning in the domain of crime investigation. QCM integrates information available from qualitative simulations in probabilistic networks, whereas other approaches mostly use qualitative techniques in performing inference on Bayesian networks.

Conclusions

QCM provides the basic functionality needed for cognitive scientists to build, simulate and explore qualitative mental models. This system has been expanded in several ways since the version used in Dehghani et al. (2007). First, it now uses Gizmo as its qualitative reasoning engine, offering a full range of qualitative simulation abilities. Second, modelers can now work in a probabilistic mode and use RC to perform exact inference on their models. Third, QCM automatically integrates qualitative information for calculating a priori probabilities of quantities used in the qualitative mode. Fourth, the interface of the system has been enhanced offering easier access to reasoning capabilities. Finally, models can now be exported in different formats facilitating collaboration between modelers. We believe that QCM provides the formalism and the functionality necessary for automatic evaluation of psychological data. Moreover, it can potentially be a helpful tool for teaching undergraduate cognitive science courses.

Acknowledgments

This research was supported by a grant from the US Air Force Office of Scientific Research.

References

Biswas, G., Schwartz, D., Bransford, J. & The Teachable Agents Group at Vanderbilt. (2001). Technology Support for Complex Problem Solving: From SAD Environments to AI. In Forbus, K. and Feltovich, P. (eds). *Smart Machines in Education: The coming revolution in educational technology*. AAAI Press.

Brandes, U., Eiglsperger, M., Herman, I., Himsolt, M. and Marshall, M. S. (2001). "GraphML Progress Report: Structural Layer Proposal." *Proceedings of the 9th International Symposium on Graph Drawing (GD-01).* Vienna, Austria.

Bredeweg, B., Bouwer, A., Jellema, J., Bertels, D., Linnebank, F. and Liem, J. (2007). Garp3 - A New Workbench for Qualitative Reasoning and Modelling. *4th International Conference on Knowledge Capture* (K-CAP 2007), D. Sleeman and K. Barker (eds), pages 183-184, Whistler, BC, Canada, October 2007.

Bredeweg, B., Salles, P., Bouwer, A., and Liem, J. (2006). Garp3 – A new Workbench for Qualitative Reasoning and Modelling. *Proceedings of the 20th International Workshop on Qualitative Reasoning*. Hanover, NH.

Bouwer A. and Bredeweg B. (2001): VISIGARP: graphical representation of qualitative simulation models, in: *Artificial Intelligence in Education – AI-ED in the Wired and Wireless Future*, J.D. Moore, G. Luckhardt Redfield and J.L. Johnson, eds, IOS Press/Ohmsha, Osaka, Japan, 2001, pp. 294–305.

Darwiche, A. (2001). "Recursive Conditioning." *Artificial Intelligence* 126, no. 1-2: 5-41.

Dehghani, M., Unsworth, S., Lovett, A., Forbus, K. (2007). Capturing and Categorizing Mental Models of Food Webs using QCM. 21st International Workshop on Qualitative Reasoning (QR). Aberystwyth, U.K.

Forbus, K. (1984). Qualitative Process Theory. Artificial Intelligence, 24, 85-168.

Forbus, K., Carney, K., Harris, R., and Sherin, B. (2001) Modeling Environment for Middle-School Students: a Progress Report. In G. Biswas ed. *Papers from the 2001 Qualitative Reasoning Workshop* Stoughton, WI, The Printing House.

Forbus, K., Carney, K., Sherin, B. and Ureel, L. (2004). Qualitative modeling for middle-school students. *Proceedings of the 18th International Qualitative Reasoning Workshop*, Evanston, Illinois, August.

Forbus, K. and Gentner, D. (1997). Qualitative mental models: Simulations or memories? *Proceedings of the Eleventh International Workshop on Qualitative Reasoning*, Cortona, Italy, June 3-6, pp. 97-104.

Keppens, J. (2007). "On Qualitative Probabilities for Legal Reasoning about Evidence." *Proceedings of the 21st International Workshop on Qualitative Reasoning.*

Keppens, J. (2007). "Towards Qualitative Approaches to Bayesian Evidential Reasoning." *Proceedings of the 11th International Conference on Artificial Intelligence and Law.*

Machado, V. and Bredeweg, B. 2001. Towards Interactive Tools for Constructing Articulate Simulations. *Proceedings of QR01*

Novak J. D., & Gowin D. B. (1984). *Learning How To Learn*. New York: Cambridge University Press.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems.* San Francisco: Morgan Kaufmann.

Parsons, S. (1995). "Refining reasoning in qualitative probabilistic networks." *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence.*

Parsons, S. & E Mamdani, E. (1993). "On reasoning in networks with qualitative uncertainty." *Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence.*

Wellman, M. (1990). "Fundamental concepts of qualitative probabilistic networks." *Artificial Intelligence* 44 (1990): 257-303.

Factored Envisioning[1]

Johan de Kleer
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304 USA
dekleer@parc.com

Kenneth D. Forbus
Qualitative Reasoning Group
Northwestern University
Evanston, IL, 60201, USA
forbus@northwestern.edu

Tom Hinrichs
Qualitative Reasoning Group
Northwestern University
Evanston, IL, 60201, USA
hinrichs@cs.northwestern.edu

Sungwook Yoon
Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304 USA
syoon@parc.com

Eric K. Jones
BAE Systems
6 New England Executive Park, Burlington, MA 01803, USA
eric.k.jones@baesystems.com

Abstract

Envisioning has been used extensively to model behavior of physical systems. Envisioning generates the qualitatively distinct possible behaviors without numerically simulating every possible set of input conditions and model parameters. This paper applies envisioning to analyze course of action (COA) diagrams to determine the qualitatively distinct outcomes of military operations. In order to avoid the combinatorial explosion of possible states, this envisioner factors non-interacting units into separate envisionment threads. The envisioner uses Assumption-Based Truth Maintenance to further limit combinatorial explosion and estimate probability of outcomes. We illustrate the performance of the factored envisioner on a variety of examples provided by military experts. We analyze its scaling performance and demonstrate its ability to track operations from sparse observations.

1 Introduction

Military planners generate courses of action (COAs) to describe how they intend to achieve their goals. COAs are described using a combination of text and graphics (Figure 1). Ideally, in the US Army, a commander generates several significantly distinct COAs, and wargames them against multiple COAs hypothesized for enemy forces. This wargaming process has several benefits. First, it helps find weaknesses in COAs. Second, it forces commanders and their staffs to think about what the other side might be planning, which sets up expectations that can be useful during operations. Unfortunately, this process is currently carried out by hand, making it time-consuming. Planning time is often at a premium, so shortcuts are often taken, degrading the quality of the results. Having automated support for envisioning possible futures could potentially offer valuable assistance to comman-

---

[1] Sponsored by Defense Advanced Research Projects Agency (DARPA)Information Processing Technology Office (IPTO) Program: Deep Green Issued by CECOM.

The views and conclusions in this document are those of the authors and should not be interpreted as representing the official policies, either expressly or implied, of the Defense Advanced Research Projects Agency or the U.S. Government.

ders and their staffs. By rapidly generating possible futures, subtle advantages or "black swan" disasters could be more easily found.

We believe that, with the right advances, the qualitative reasoning technique of envisioning could provide such automated assistance. Qualitative representations provide a natural fit to the mental models of military commanders. Commanders divide terrain up into functionally significant pieces, and in the early stages of planning, focus only on the actions that directly support achieving their goals, without worrying about logistics or other supporting concerns. Wargaming in military decision-making processes focuses on discrete, distinct possible categories of outcomes — in other words, qualitative states. However, the military domain is more challenging than any domain in which envisioning has been previously applied. The number of "moving parts" is high, as are the actions they can participate in. Unlike most engineered systems, where a schematic can be developed to define in advance possible interactions, potential interactions in military reasoning must be detected dynamically. To overcome these problems, this paper describes the idea of factored envisioning, where we dynamically identify collections of entities whose behaviors must be reasoned about together.

Section 2 describes our assumed architecture and summarizes aspects of terrain reasoning that are relevant for this paper. Section 3 discusses the rule and COA language we use, and Section 4 applies classic envisioning to the military domain. Section 5 illustrates why factored envisioning is necessary. Section 6 defines the key ideas of factored envisioning, and Section 7 shows how an ATMS is used to achieve scale-up. Section 8 shows how large envisionments can be represented compactly. Section 9 briefly discussed the use of probabilities in tracking possible states, and Section 10 discusses related and future work.

2 Conceptual architecture and terrain

A simple concept for a battlespace reasoner consists of three parts: (1) an interface which supports COA entry, using the standard graphical language used by militaries for units, tasks, and the features they impose on terrain (e.g., the unit boundaries in Figure 1), (2) an envisioner which takes a set of Blue (the friendly side) COAs and a set of Red (the other side) COAs, and generates a set of qualitative states indicating all the qualitatively distinct ways that things might turn out, and (3) a tracker which, given observations during an operation, assesses which of these states the battle is in, and what COA Red is following. Our focus here is only on the envisioner, and how to compute probabilities that a tracker would need.

One of the key factors in military reasoning is terrain. We use qualitative spatial representations of terrain, based on a formalization of military terrain analysis techniques [Donlon & Forbus, 1999]. Qualitative regions are defined both within the COA and as regions implied by the COA. Examples of specifically defined COA regions include engagement areas and avenues of advance. Examples of implied COA regions include the regions where visibility and/or weapons range envelopes intersect along movements specified by the combination of Blue and Red COAs. Implied COA regions are crucial to identify because they constitute regions where interactions can occur. That is, our strategy for detecting interactions involves first finding spatial intersections, filtering those using temporal constraints to see if relevant units can be in the same place at the same time, and then considering the nature of those units and their goals (as assigned within their COAs) to ascertain what sort of interaction, if any, takes place. We exploit this strategy below, but otherwise, the details of the qualitative spatial reasoning we use lie outside the scope of this paper.

We model military actions using qualitative rules, using a PDDL-like [McDermott, 1982] rule language for durative actions [Do & Kambhampati, 2002]. All actions happen over time. Each action has a distinct beginning, duration and end. For example, Figure 2 illustrates the action of `unit` moving from location `from` through path `path` to location `to`. At the beginning of the action the unit is located at location `from` and at end of the action it is located at location `to`.

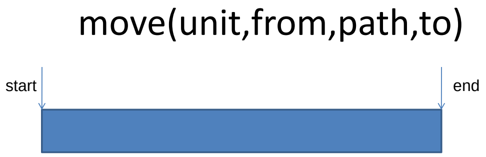


Figure 2: Move action.

```
(:action move
  :parameters (?unit - unit ?from - location ?path - path
               ?to - location)
  :condition (and (at start (location ?unit ?from))
                  (at start (trafficable ?unit ?path))
                  (at start (path ?from ?path ?to))
                  (over all (not (underfire ?unit))))
  :effect (and (over all (location ?unit ?path))
               (over all (decreasing (distance ?unit ?to)))
               (at end (location ?unit ?to)))
  :duration :definite)
```

`unit, location, path` and `location` are distinct types. The `:parameters` slot declares all the variables of the action and their types. `:condition` indicates properties which must hold. At the beginning of the action the moving `?unit` must be located at location `?from`, the unit must be able to traverse the path (e.g., not too heavy or too wide), and the path must connect location `?unit` to location `?to`. Movement is severely restricted if a unit is under fire. The `:effect` slot indicates that the unit is on that path for the entire duration, the distance from the destination is constantly decreasing and (if the action is not interrupted) at the end the unit will be at location `?to`. `:definite` indicates the action has a definite end time.

`attack-by-fire` models an attack on a location where the enemy unit(s) may not be known. A slight extension to PDDL allows this rule to identify the enemy unit(s) `enemy`.

```
(:action attack-by-fire
  :parameters (?u - unit ?from - location
               ?enemy-location - location)
  :condition (and (at start (EnemyAtLocation ?u ?enemy-location
                                             ?enemy))
                  (over all (> (strength ?u) 0))
                  (over all (location ?enemy ?enemy-location))
                  (over all (location ?u ?from)))
  :effect (and (over all (decreasing (strength ?enemy)))
               (over all (underfire ?enemy))
               (at end (assign (posture ?enemy) defeated)))
  :duration :definite)
```

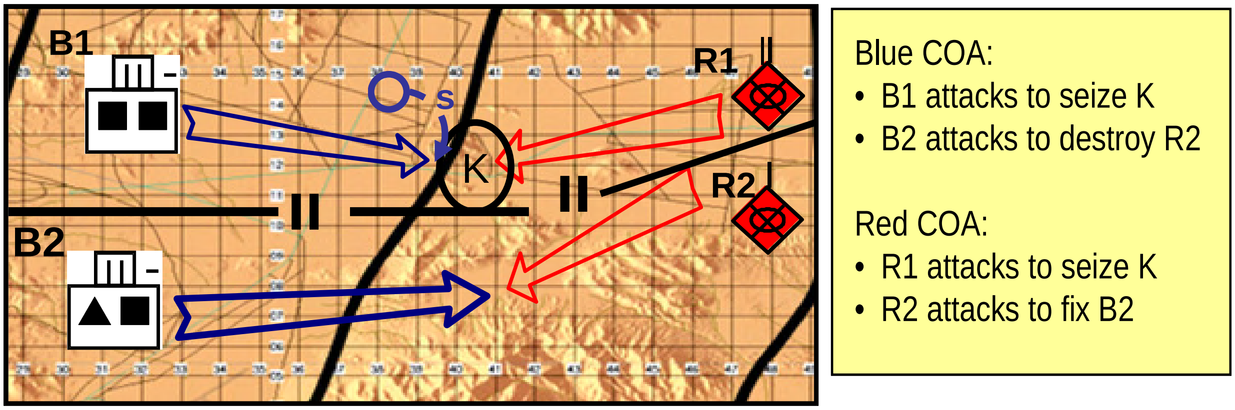


Figure 1: This COA describes two independent interactions: (1) Both Blue and Red are trying to seize K, and (2) and Red is trying to prevent Blue from moving further east. The horizontal line with vertical strokes specifies a unit boundary. It's Blue's intent that there will be no interaction across this boundary. Red may have other plans.

We have implemented rules for military tasks frequently used in COA's. This includes a set of basic tasks (e.g., movement and firing) that are commonly used in defining more complex tasks.

3 COA language

COA's are described graphically. In the complete system, commanders enter COA's graphically on top of terrain maps. We can also use nuSketch Battlespace [Forbus, Usher, & Chapman, 2003] to input COAs graphically. For the purposes of exposition we utilize an extremely simple language for COA's which includes:

- A ground action instance such as `(move B1 initialB1 AxisB K)`. Such items are executable only if their preconditions apply.
- A sequence of COA items which will be executed in order.
- `(cease <action> <actor>)` to explicitly terminate an ongoing action.
- `(if <condition> <coa-items>)` for a decision point.

In our simple COA language the top half of Figure 1 is described by:

```
(move B1 initialB1 AxisB K)
(seize B1 K)
(move R1 initialR1 AxisR K)
(seize R1 K)
```

4 Classical envisioning

In qualitative reasoning one of the most common ways to represent time is as instants, separated by open intervals, much like the real line. Each action has a distinct beginning and end. Many actions can take place simultaneously. A *situation* is a bundle of ongoing actions. The bundle is minimal: no action can stop and start within the temporal interval. The start time of a situation is the latest of all the start times of all its actions. The end time of a situation is the earliest end time of all its actions. Predicates (except `location`) are constant over the duration of a process. Quantities are presumed to change monotonically over time.

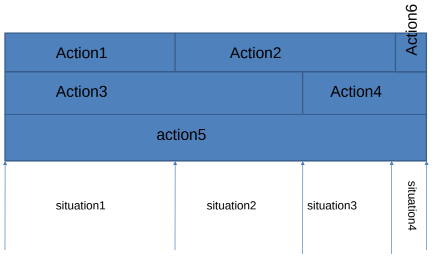


Figure 3: 6 actions and 4 situations.

The envisioning process [de Kleer & Brown, 1984; Forbus, 1984; Kuipers, 1986] generates a graph of situations which describes all possible qualitatively distinct possible evolutions of a system. Classical envisioning operates as follows:

1. Determine the combined influences on each quantity.
2. Identify all quantities that are changing towards their limit points.
3. Find all legal possible orderings for those quantities to reach their limit points. In worst case if there are n changing quantities there may be 2^n possible endings. Typically only a small subset of the combinations will satisfy the conditions.
4. For each possible ending, compute the next possible situation by (1) terminating actions which naturally end

or whose preconditions no longer hold (interrupted actions), (2) starting any new actions whose preconditions now hold, and (3) adding the new situation to the envisionment.

Consider envisioning Figure 1. Figure 4 describes the resulting envisionment. This envisionment is constructed as follows:

1. Two new actions start in situation 1: B1 and R1 simultaneously start moving to location K (by decreasing their distance from their endpoints).
2. In situation 2 R1 and B1 are moving along their respective avenues of advance. This situation can end in three possible ways:
3. Situation 4 describes the case when $B1$ arrives at K first.
4. Situation 5 describes the case when $R1$ arrives at K first.
5. Situation 3 describes the case when $R1$ and $B1$ arrive at K simultaneously.
6. Situations 2, 4 and 5 all lead to a common situation 3 where both $B1$ and $R1$ fight. As both are reducing the strengths of the other, there are two possible outcomes: either Red or Blue's strength reduces to 0 (in many cases units disengage before at some limit point greater than 0). The probability of an outcome depends on many factors, including the arrival time. If B1 arrives early, then its probability of winning would be higher.
7. Situation 6 where Red wins.
8. Situation 7 where Blue wins.

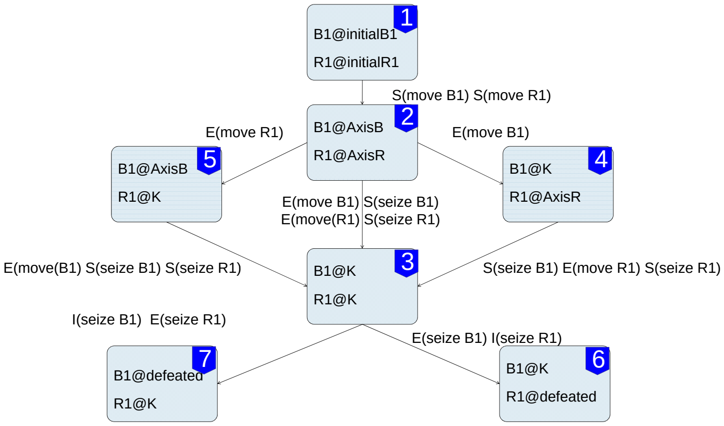


Figure 4: Envisionment of R1 and B1 generated by our envisioner. Nodes are labeled by their id and operating unit locations (or "defeated") and edges are labeled by actions starting (S), ending (E) or interrupted (I).

5 Why factored envisioning is needed

Consider the COA illustrated in Figure 5:

```
(move RF1 initialRF1 Axis1 Hill1)
(move RF2 initialRF2 Axis2 Hill2)
(move RF3 initialRF3 Axis3 Hill3)
```

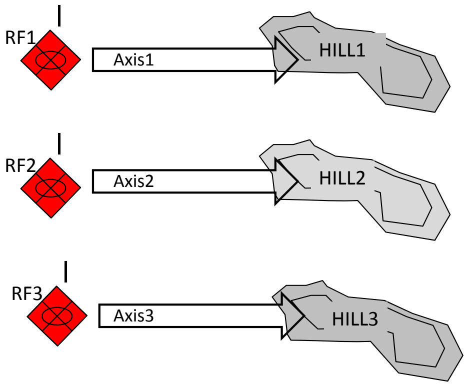


Figure 5: Simple COA to demonstrate factored envisioning.

The resulting envisionment consists of the 26 situations in Figure 6. In situation 1 all actions start: RF1, RF2, RF3 start moving to their destinations. In situation 2, all actions are ongoing and the question is only when each will end, or put another way, which reaches its destination first. Given n independent actions, there are $2^n - 1$ possible combinations of ending options.

One of the central tenets of qualitative reasoning is to only make distinctions which matter. This applies to envisionments as well. As RF1, RF2 and RF3 do not interact, the envisionment of Figure 6 makes many needless distinctions. The key idea of factored envisioning is avoid grouping actions that do not interact. In factored envisioning, each situation describes a partial description of the world, and each set of actions is grouped into situations which only interact with each other. Figure 8 illustrates a factored envisionment. The top node is a full situation and the three branches represent a partition into partial situations.

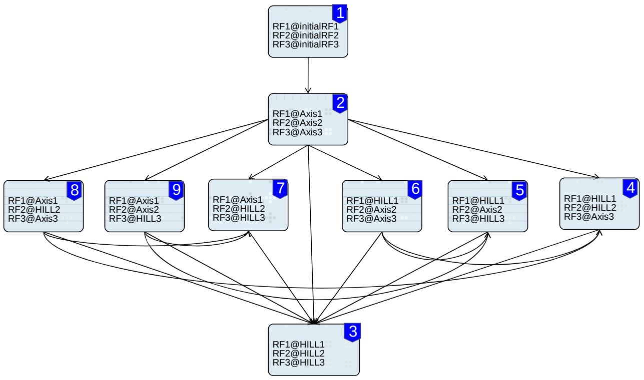


Figure 7: With situation merging 26 situations are reduced to 9.

6 Factored Envisioning

The main purpose of factored envisioning is to avoid the irrelevant overspecificity and needless exponential explosion in

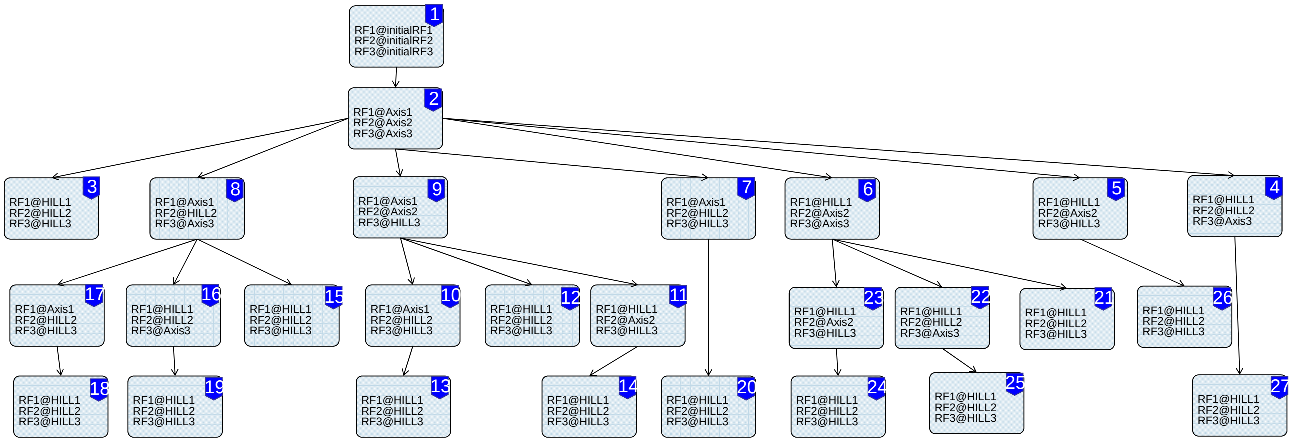


Figure 6: 3 non-interacting actions leads to 26 situations without situation merging. For clarity, edge labels have been omitted.

situations. We introduce a notion of *kernel* situation, as opposed to a *full* situation which we have been using so far. A full situation describes the positions and actions of all the units on the battlefield. A kernel situation describes the positions and actions of *some* of the units on the battlefield, but with one additional condition: every unit within the kernel interacts with every other (perhaps transitively). Intuitively, a kernel situation is the smallest set of interacting units possible. In the envisionment in Figure 4 situation is 3 is a kernel situation as both units are interacting. All the other situations are full situations. None of the situations in Figure 6 are kernel.

Factored envisioning uses full envisioning as a subprocedure. Intuitively, factored envisioning proceeds as follows. Any full situation is partitioned into its non-interacting kernel situations. The full envisioner is invoked on each of those kernel situations (where every other unit is hidden). This will produce a set of space-time tubes or histories [Hayes, 1990]. For every possible space-time intersection, the factored envisioner constructs a new kernel situation and invokes the envisioner on this combined situation to see if new possible interactions result (this may result in the construction of a new location). Our algorithm intersects first by space and then by time. Figure 8 depicts the envisionment of Figure 5. The ellipitcal top node depicts a non-situation node comprised of three kernel situations. Figure 9 describes the factored envisionment of the COA from the introduction. Node 1 is comprised of two kernels:2 and 3. Kernel situations 2-6-7 describe the movement of R1. Kernel situations 3-4-5 describe the movement of B1. Node 8 depicts the joining of the two situations and kernel nodel 9 depicts the battle. The battle has two outcomes one in which Red wins and another in which Blue wins. Nodes 11 and 10 contain two kernel situations each. Although graphically this envisionment appears more complex, each node in a factored envisionment only describes a small local state of affairs and this produces dramatic improvements in envisioning performance and subsequent analysis (as discussed in Section 8). The triangle node is a non-situation to describe that kernel situations 7 and 5 interact. There are no other interactions. The three elipitical nodes are non-situation nodes are comprised of multiple kernel situations.

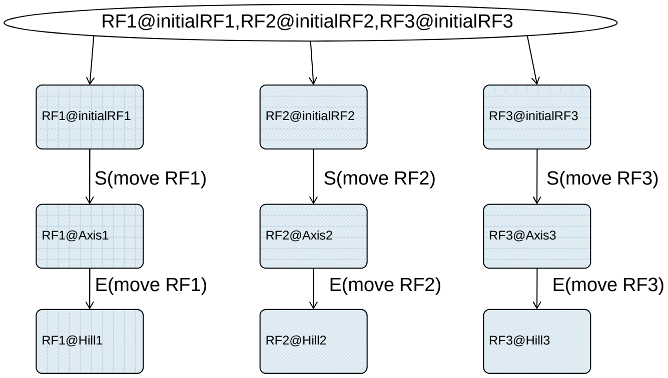


Figure 8: Simple factored envisionment of Figure 5.

7 Using an ATMS

The envisionment uses a probabilistic Assumption-Based Truth Maintenance System to represent ambiguities and perform all the needed evidential reasoning [de Kleer, 2008]. Every situation and transition is represented by an ATMS node. For example, situation 1 of Figure 4 is the initial state. ATMS nodes $s1$, $t1$ and $s2$ are created to represent the start situation and its transition to the next situation. The following two justifications are added to the ATMS:

$$s1 \rightarrow t1,$$

$$t1 \rightarrow s2.$$

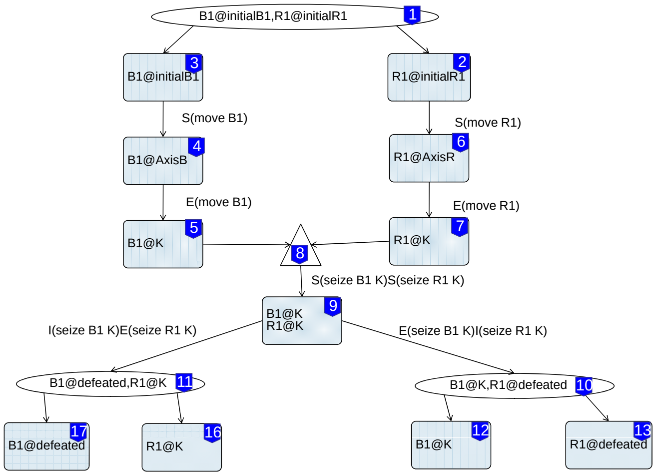


Figure 9: Kernel envisionment.

However, there are 3 possible transitions from $s2$ corresponding to the cases where B1 or R1 reaches K first or together. These possibilities are represented as follows. The usual ATMS nodes are created to represent the three outcomes, but three assumptions ($A2$, $A3$, $A4$) are created to encode the exclusive disjunction of the alternatives:

$$s2 \wedge A3 \rightarrow t3,$$

$$s2 \wedge A2 \rightarrow t2,$$

$$s2 \wedge A5 \rightarrow t5,$$

$$t3 \rightarrow s4,$$

$$t4 \rightarrow s5,$$

$$t5 \rightarrow s3,$$

$$oneof(A2, A3, A5).$$

In addition, a probability is assigned to each assumption which is computed using a more detailed model which considers speed of the units and the terrain they have to cover to reach K. The outcome of the battle (situation 3) depends both on both the path taken to reach K, and the properties of the units. Again two assumptions are created to represent both outcomes. The probabilities of these assumptions are derived from more detailed military models.

As a result of this justification structure the ATMS constructs a label for each node. This label consists of minimum sets of assumptions that can be used to derive that node given the justifications. This label takes the form of prime implicates and is a d-DNNF expression [Darwiche & Marquis, 2002]. The probability of any node can be directly derived from its label:

$$p(x) = \sum_{e \in label(x)} p(e),$$

and,

$$p(A_1, \ldots, A_n) = \prod A_i.$$

Thus the PATMS directly computes the prior probability of every situation.

Of far greater importance for planning is the conditional probability of reaching some objective B from situation A. This can be directly computed from the PATMS by:

$$P(B|A) = \frac{P(A \wedge B)}{P(A)}.$$

There may be multiple situations which achieve a commander's intent. The most useful measure of a situation's desirability is its expected utility:

$$EU(S) = \sum_F U(F)P(F|S).$$

(U is usually only non-zero for end-states.) Although probabilities are well-defined for both kernel and full situations, utility is only well-defined for full situations. Blind alleys or "black swan" events are situations with significant conditional probability but with very low expected utility.

ATMS assumptions are also used to keep results of different COA pairs distinct while eliminating redundant envisioning. An assumption is created for every COA to represent "This COA is being executed." Thus, if there are 3x3 COAs, 6 assumptions are created. The three assumptions for each side are mutex. These assumptions have the prior probability of the particular COA. (However, in most cases the commander is interested in the conditional properties so the prior on a root is not that relevant.)

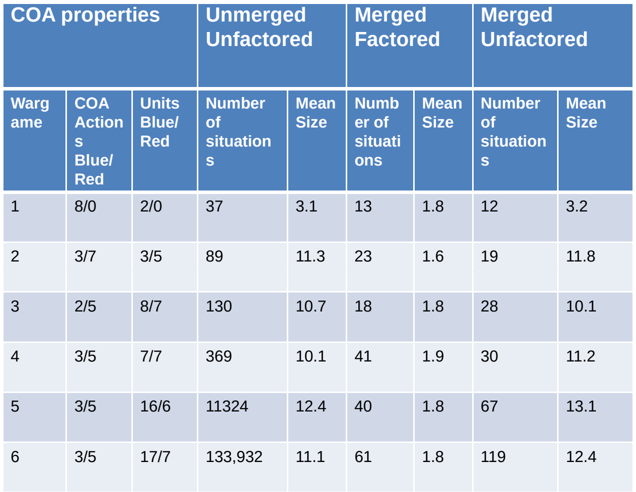
8 Packing

In order to avoid combinatorial explosion in situations it is important to detect qualitatively similar situations. There will often be multiple paths to reach a particular situation. Every full and kernel situation will have an ATMS node. Figure 10 makes the case for merged factored futures graphs. On the vertical axis are 6 war games and their characteristics. "Unmerged Unfactored" is the number of (full) situations generated and their mean size. "Merged Factored" is the number of (kernel) situations and their mean size. The final column "Merged Unfactored" is the number of (full) situations and their mean size. The envisioner includes utilities to move back and forth from kernel and full situation descriptions of an envisionment. The envisioner can move fluidly between kernel and full situations as needed.

9 Tracker Precision

The objective of the tracker is to identify the actual COA-pair (and of particular concern the enemy's COA) during operations. The commander must be signaled as soon as possible when there are future situations with low expected utility (typically < 0.5) and high expected probability (typically > 0.25). When blind alleys arise, the commander must develop new COAs — the aim of this project is to support commander decision making, not do it for him.

One rarely has full information during operations. The incomming observations will be scattered and partial. DARPA provides sample data which is very noisy along all dimensions (time, position, strength, ammunition, fuel, etc.) The



COA properties			Unmerged Unfactored		Merged Factored		Merged Unfactored	
Wargame	COA Actions Blue/ Red	Units Blue/ Red	Number of situations	Mean Size	Number of situations	Mean Size	Number of situations	Mean Size
1	8/0	2/0	37	3.1	13	1.8	12	3.2
2	3/7	3/5	89	11.3	23	1.6	19	11.8
3	2/5	8/7	130	10.7	18	1.8	28	10.1
4	3/5	7/7	369	10.1	41	1.9	30	11.2
5	3/5	16/6	11324	12.4	40	1.8	67	13.1
6	3/5	17/7	133,932	11.1	61	1.8	119	12.4

Figure 10: Futures graph sizes.

unreliability of time stamps is particularly challenging. Figure 11 illustrates the quality of the data we work with. A typical battle may produce as many as $10,000$ such messages over a couple of hours. What makes tracking even plausible with such poor data is that the tracker need only distinguish amongst paths in the futures graphs and typically there are only a handful to distinguish among at a given time.

A basic Bayesian tracker provides good results on the data and futures graphs we have tested. We associate a probability with each possible path and update it with Bayes' rule after every message m:

$$p(P|\{m\} \cup M) = \alpha L(P,m)p(P|M).$$

As time is as noisy as other quantities it has no special status. The likelihood $L(P,m)$ that message m corresponds to path P is computed with a simple linear function of the likelihood scores of the parameters (including time). To compute the likelihoods we need to translate qualitative ranges into quantities to numerical values. We simply use the mean value of each range. We use the PATMS probabilities for the transition probabilities. This approach more than achieves DARPA's desired metrics for blind alley detection.

10 Related and Future Work

Recently there has been an upsurge in research on adversarial reasoning [Kott & McEneasney, 2007], but we are aware of no prior approaches which use qualitative representations extensively or perform envisioning. Cohen's Abstract Force Simulator [King *et al.*, 2002] uses numerical Monte Carlo simulation to identify qualitative regions in parameter spaces.

Within the qualitative reasoning community, [Clancy & Kuipers, 1997] describes a qualiative simulator, DecSIM which partions a system into non-interacting collections a priori using causal ordering. In our approach, the partitions are determined dynamically because all possible interactions cannot be determined a priori. Although later versions of DecSIM identify non-interacting collections dynamically it focusses only on eliminating "chatter" when all interactions are known a priori.

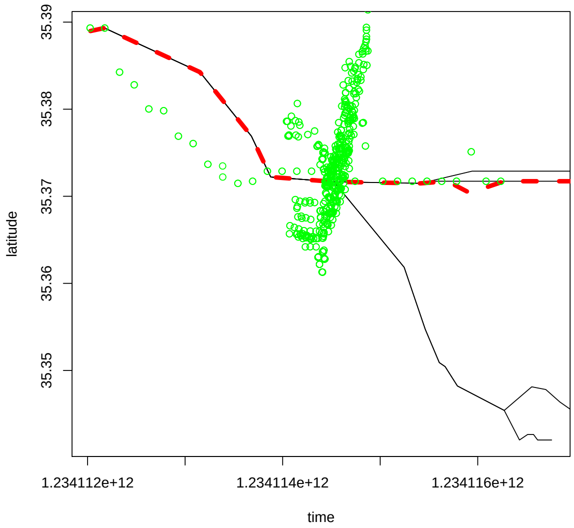


Figure 11: This figure illustrates the challenges the tracker faces on a very simple example. The horizontal axis is reported time and the vertical axis is latitude of a unit. The lines describe the 5 paths through the futures graph (interpolated as they are all qualitative states). The circles are the incomming observation data. The correct path is indicated by the dashed lines.

The work described here is part of a larger DARPA-sponsored effort called Deep Green to develop a system that helps Army commanders and their staff develop robust plans that can handle a wide range of foreseeable contingencies and rapidly update them during plan execution as the situation evolves. Our system is being developed and tested using a collection of realistic army scenarios created by a small team of highly regarded subject-matter experts, including a former commander of the Army's National Training Center. Significant interest in transitioning the results of Deep Green have already been expressed by the Army.

We plan to explore three avenues in future work. First, it is unclear that PDDL rules are the best representation language for military tasks. Some combinations of actions (to represent decision-making by subordinate commanders) and processes (to model continuous effects in a composable way) may provide a more natural way to represent these phenomena. Second, the range of military tasks needs to be further expanded, to handle a wider range of COAs. Finally, in collaboration with military experts, we need to develop the fine-grained level of models for probability estimation, providing the input probabilities for particular outcomes which will then be propagated through the envisionment by the ATMS.

References

[Clancy & Kuipers, 1997] Clancy, D., and Kuipers, B. 1997. Model decomposition and simulation: A component based qualitative simulation algorithm. In *Proc. AAAI 97*, 118–124.

[Darwiche & Marquis, 2002] Darwiche, A., and Marquis, P. 2002. A knowledge compilation map. *Journal of Artificial Intelligence Research* 17:229–264.

[de Kleer & Brown, 1984] de Kleer, J., and Brown, J. S. 1984. A qualitative physics based on confluences. *Artificial Intelligence* 24(1):7–84. Also in: Bobrow, D. (ed.) *Qualitative Reasoning about Physical Systems* (North-Holland, Amsterdam, 1984 / MIT Press, Cambridge, Mass., 1985).

[de Kleer, 2008] de Kleer, J. 2008. Using model-based diagnosis for bayesian inference. In *The 19th International Workshop on Principles of Diagnosis*, 55–62.

[Do & Kambhampati, 2002] Do, M. B., and Kambhampati, S. 2002. Sapa: a multi-objective metric temporal planer. *Journal of Artificial Intelligence Research* 20:155–194.

[Donlon & Forbus, 1999] Donlon, J., and Forbus, K. 1999. Using a geographic information system for qualitative spatial reasoning about trafficability. In *Proceedings 13th international workshop on Qualitative Reasoning*.

[Forbus, Usher, & Chapman, 2003] Forbus, K.; Usher, J.; and Chapman, V. 2003. Qualitative spatial reasoning about sketch maps. In *Proceedings of the Fifteenth Annual Conference on Innovative Applications of Artificial Intelligence*.

[Forbus, 1984] Forbus, K. D. 1984. Qualitative process theory. *Artificial Intelligence* 24(1):85–168. Also in: Bobrow, D. (ed.) *Qualitative Reasoning about Physical Systems* (North-Holland, Amsterdam, 1984 / MIT Press, Cambridge, Mass., 1985).

[Hayes, 1990] Hayes, P. J. 1990. The second naive physics manifesto. In *Readings in qualitative reasoning about physical systems*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. 46–63.

[King *et al.*, 2002] King, G.; Heeringa, B.; Catalano, J.; and Cohen, P. 2002. Models of defeat. 85–90. Proceedings of the 2nd International Conference on Knowledge Systems for Coalition Operations.

[Kott & McEneasney, 2007] Kott, A., and McEneasney, W. 2007. *Adversarial Reasoning: Computational Approaches to Reading the Opponent's Mind*. Chapman and Hall/CRC.

[Kuipers, 1986] Kuipers, B. J. 1986. Qualitative simulation. *Artificial Intelligence* 29(3):289–338. Also in: D. S. Weld and J. de Kleer (eds) *Readings in Qualitative Reasoning about Physical Systems* (Morgan Kaufmann, Los Altos, Calif., 1990).

[McDermott, 1982] McDermott, D. V. 1982. A temporal logic for reasoning about processes and plans. *Cognitive Science* 6(2):101–155.

Dark Knowledge in Qualitative Reasoning: A Call to Arms

Kenneth D. Forbus[1], Dedre Gentner[2]

[1]Qualitative Reasoning Group, Northwestern University
2133 Sheridan Road, Evanston, IL, 60208, USA
[2]Psychology Department, Northwestern University
2029 Sheridan Road, Evanston, IL, 60208, USA
{forbus, gentner}@northwestern.edu

Abstract

While people do qualitative reasoning, there is ample evidence that they do not always do it well. Two current crises, human-induced climate change and the financial meltdown, can be traced in part to faulty mental models. The QR community has formalisms that can potentially help with public education about such problems, but so far we have not been very successful in doing so. We claim that part of the reason is that current QR accounts do not adequately incorporate experiential knowledge. We argue that it is important to find better ways to improve public qualitative reasoning abilities, in part by helping people enlist their experience-based models via analogy.

Introduction

It is a truism that democracy works best with informed citizens. Alas, there is ample evidence that people do not have adequate mental models for many relevant areas. Consider two current crises: Human-induced climate change and the financial meltdown. In both cases, to be sure, there are people and organizations that are deliberately trying to obscure what is happening for their own reasons. But there is ample evidence that a fundamental failure of education has taken place. One key problem in understanding climate change is what Cronin *et al* (2008) call *stock-flow failure*. In system dynamics terms, a stock is something that accumulates, i.e., something that would be modeled in QP theory (Forbus, 1984) as a directly influenced parameter. A system with only flows has no accumulation, and its outputs are basically a function of its inputs. Surprisingly, when people are given graphs indicating the inputs to a system with accumulation, they often ignore the accumulation, and sketch the output as if it were simply a function of the inputs. This failure occurs even in highly educated people with technical backgrounds. Cronin *et al* further showed through a series of experiments that this could not be explained by problems in interpreting graphs, misunderstanding of context, lack of motivation, or lack of cognitive capacity. It is, quite simply, a failure of mental models reasoning.

What should be striking for our community is that we have what may potentially be some of the best ideas for helping people overcome these problems. It is difficult to teach ideas without accessible formalisms. The formalisms of QR, which factor out traditional mathematics and make causality explicit, could be of great value in education. But we have not been very successful in spreading these ideas more broadly.

This paper argues that to change this situation, we need to expand our models to be more psychologically oriented. Physicists postulated dark matter in order to explain the missing mass in their observations. By analogy, "dark knowledge" is the knowledge for which we lack elegant formalisms in QR, but which nevertheless is a major factor in human qualitative reasoning. Dark knowledge is concrete knowledge: specific facts and cases, derived from first-hand experience or via culture, that are remembered and used for many daily reasoning tasks via analogy. In terms of its size, we believe it far outweighs fully general first-principles knowledge, providing the "missing mass" that holds our conceptual universe together.

Understanding how people reason with dark knowledge is important for QR to reach its full potential. Moreover, we believe that understanding how human qualitative reasoning works is crucial for finding better ways to teach people to reason well about the complex problems we all face. This paper also argues that harnessing experiential knowledge through analogy is one potential way to transform education, making it better able to prepare people for the challenges ahead.

We begin by looking at the broad picture that QR and psychology paint of qualitative knowledge and reasoning. Next we look at how to improve human mental models, using a favored explanatory device for climate change, and one of QR's favorite examples, the humble bathtub (Kuipers, 1994), to highlight how the use of analogy in explanations can be improved. After that, we discuss some ideas for tools and techniques for both understanding experiential knowledge better, and for improving education. We end with a call to arms.

Human mental models: The big picture

The study of qualitative reasoning was originally motivated by observing human reasoning: People who do not know differential equations reason about many physical phenomena perfectly well, and even scientists and engineers rely on simpler, qualitative models when framing problems and interpreting data. The "standard model" in QR explains this in terms of general, broad-coverage domain theories, expressed exclusively in first-principles terms. Given a particular scenario or problem, these general concepts are applied via instantiation to create a scenario model that can then be reasoned with.

This standard model has been remarkably successful in building a variety of useful systems. However, we believe it has strong limitations as a psychological account of human qualitative reasoning. We have proposed that much of human knowledge about the physical world is concrete (Forbus & Gentner 1997). In some sense it's obvious: people have episodic memories, whereas most of today's QR models do not. But there are reasons to believe that the use of experiential knowledge has profound consequences for human qualitative reasoning.

When someone starts learning about a novel phenomenon, they accumulate experiences. Even concrete experiences can be used directly in very similar situations, via within-domain analogies. We take experience quite broadly here: We include cultural influences such as language and education, as well as first-hand interaction with the world. While many of us have read about carbon sequestration and credit default swaps, for example, few of us have actually experienced these processes first-hand.

As experiences accumulate, they are used to construct generalizations, at first prototypical behaviors (*protohistories*, in Forbus & Gentner (1986)) and later causal fragments that can be turned into model fragments (the *causal corpus* in Forbus & Gentner (1986)). These generalizations are one source of misconceptions. Importantly, earlier forms of knowledge are added to, but not replaced by, later, more refined models. Once someone learns differential equations, for example, they still use simpler models, learned earlier, to throw balls, estimate stopping distances while driving, and other tasks where differential equations are in principle relevant.

Our hypothesis is that much of the knowledge people use in qualitative reasoning is concrete, at the level of protohistories and causal corpus. To be sure, we believe that something like first-principles domain theories are learned, either via analogical generalization or via direct instruction. In experts they are especially rich, including a tight integration with mathematical models. But even experts rely on experience-based models in their professional reasoning. For example, analogy seems to play an important role in model formulation (Falkenhainer, 1992; Klenk *et al* 2005). In non-experts, or even in experts, knowledge in many domains can be thought of as "pastiche models" (Collins & Gentner, 1987) or "in pieces" (diSessa, 1993), i.e., local, context-specific models.

Ideally, knowledge learned in school becomes tightly integrated with knowledge learned from experience, reorganizing it in ways that make correct reasoning more likely. Unfortunately, there is ample evidence that this integration is difficult, often leading to accumulation of multiple conflicting models. For example, Clement (1982) and McCloskey (1983) both showed that even students who did well in physics classes often continue to have and use incorrect qualitative models of force and motion. These misconceptions are uncorrelated with mathematical knowledge, and even honors students are susceptible to them (Halloun & Hestenes 1985). New misconceptions can arise during instruction as well (Spiro *et al* 1989;Vosniadou, 1994). Moreover, students sometimes actively work to protect and maintain their misconceptions, erecting "mental shields" when they are threatened by new information (Feltovich *et al* 2001).

Let us reexamine the stock-flow failure identified by Cronin *et al* (2008) in light of this model. They argue that people use a *correlation heuristic* in reasoning about systems with multiple continuous inputs and outputs. That is, when given the task of controlling a system which accumulates something, they tend to believe that the shape of the output should look something like the shape of the input, but delayed in time. This is the sort of heuristic that could very easily be derived from everyday experience, where the preponderance of input/output pairs we see are more often correlated in their behavior than not. If we turn the faucet in the sink or bathtub higher, water comes out faster, perhaps after some delay. The same thing happens when we turn on the tap on a garden hose. If this heuristic works in many situations, it is natural to apply it to new problems.

How can we improve mental models?

How can we improve people's mental models? Simply handing them a modeling language, even in student-friendly terms (e.g. Betty's Brain, Biswas *et al* 2001; VModel, Forbus *et al* 2004) is not enough. Showing them qualitative simulations (e.g., Bredeweg *et al* 2008) is not enough. These both are good starts, but unless we work on ways to integrate what they learn from these experiences into their prior knowledge, such interventions will not have as much impact as desired.

We believe that analogy is an excellent mechanism for integrating knowledge. Understanding the connections between experiences and/or models requires comparing them and understanding "what goes with what", which is exactly what the structural alignment process at the heart of analogy does (Gentner, 1983). Further evidence indicating that analogy can be used to rapidly learn mental models (Kotovsky & Gentner, 1996; Gentner *et al* 2009) Combining the conceptual clarity of qualitative representations with the integrative power of analogy is, we suspect, exactly what we need to create new ways to help people reason better about complex situations. Showing how to think formally and qualitatively about systems that someone has experienced first-hand provides a solid base domain that can then be projected by analogy to other target domains that need to be understood. Leveraging everyday experience provides solidity to conclusions that might not otherwise be plausible.

Research in psychology and learning sciences provides some insights for the effective use of analogy. For example, it is important to have learners work through correspondences in detail, so that they get the most out of the analogy (Kurtz *et al* 2001). Ensuring that the base domain is well-understood, and learners are focused on the relevant aspects of it, helps them apprehend the analogy (Richland *et al* 2007). We illustrate via an example next.

The bathtub analogy for climate change

Understanding climate change has proven to be quite difficult. Part of the problem is how counter-intuitive it is: For most of human history, people were at the mercy of weather and climate, and our impact seemed extremely small compared to the vastness of the planet. But as we grew in number and the planet did not, this changed. Now the modeling assumption of endless resources is clearly not accurate. One analogy that has been used to communicate the problem (e.g. Sterman 2008) is a favorite QR example, the humble bathtub.

Bathtubs have a faucet (or faucets) which can be opened to let water in, a drain which can be opened to let water out, and some capacity for holding water plus one or more people. Overflowing is something to be avoided. Our experience teaches us that for some level of water, it is likely that when we sit down in the tub it will overflow. In this analogy, the atmosphere is like the volume of the bathtub. The accumulation of carbon in the atmosphere is like the accumulation of water in the bathtub. Just as there is a level at which overflows are likely in the bathtub, there is some level at which accumulated carbon causes problems on a massive scale (countries going underwater, starvation, etc.).

In explaining this analogy so far, we have been very explicit about what aspects of the base domain should be considered, so explicit as to cross the line into belaboring the obvious for the already-informed. This degree of elaboration is useful to provide a solid foundation for extending an analogy into new areas, or using it to help understand new ways of reasoning. Having students work through the correspondences explicitly and in detail, by constructing a table for instance, helps ground the mapping.

This analogical model provides considerable value in reasoning. If the inflow is larger than the outflow, then the level will be rising. This is what is happening in the atmosphere, with CO_2, methane, and other greenhouse gasses being produced faster than natural processes can absorb them, and hence they accumulate in the atmosphere. Opening the tap wider in the bathtub will cause the level to rise faster, and increasing carbon emissions will lead to disaster more quickly. In public policy terms, a "conservative" strategy often proposed is to keep carbon emissions at their current level. But, mapping this strategy to the bathtub, one can easily see that this is not enough: The level will continue to rise inexorably to overflow, unless emissions are reduced below their current levels.

A good analogy provides a framework that can be expanded to incorporate additional ideas. For example, suppose we cannot or will not turn down the faucet. The only way to prevent an overflow is to increase the rate of draining – with buckets, if need be. In the case of the atmosphere, planting new forests is one way to improve its "drainage". Unfortunately, a recent result about the oceans absorbing less carbon due to increases in atmospheric temperature[1] can be understood as one of the "drains" becoming less effective, and thus leading to a higher rate of carbon accumulation – a potentially nasty positive feedback cycle.

Modeling bathtubs, and other everyday examples, is a common practice in QR because it allows us to compare formalisms more easily. It is also a useful exercise for someone learning a new modeling language because it helps integrate the new language into their experiences. It is important to walk through everyday behaviors, and show how they can indeed be derived from the consequences of the primitives. For instance, the relative rate of the inflow and outflow determines whether the amount of water, and hence the level, is increasing. The idea that one can get a stable balance between inflow and outflow for a range of levels can also be examined, although this will take more work since people are less likely to be familiar with this notion. To see how important elaborating the everyday example is, consider this: In some of the experiments exploring the stock-flow failure, the simulated system being controlled was a bathtub! When entering a technical problem, people often check their intuitions at the door. Tightly coupling abstract models and everyday experience seems central to the challenge we face.

Promoting transfer

How can we help people apply new ideas when they are potentially relevant? Research on analogy in instruction suggests that having learners compare cases can double the odds of them applying concepts to new situations where they are relevant (Gentner *et al* 2003). Re-using the bathtub as an analog to credit card debt provides an example.

Bathtub	**Credit Card Usage**
Faucet setting	Monthly charges
Drain setting	Monthly payment
Level of water	Amount of debt
???	Interest rate

This table of correspondences helps us understand that we are missing something in the analogy: What is the

---

[1] http://www.guardian.co.uk/environment/2009/jan/12/sea-co2-climate-japan-environment

bathtub equivalent of the interest rate on a credit balance? This is like a second faucet, whose setting is determined by the level of water and the interest rate. So even if there are no new monthly charges, debt will continue to accumulate, thanks to this second faucet. Again, this may seem obvious, but it is interesting just how many people in the US economy behave as if they do not believe this is true.

Mental models and the financial crisis

Untangling the causes of the current financial crisis is an ongoing process, being undertaken from a variety of perspectives by a number of disciplines. Consequently, the evidence here is less well worked out than in the case of climate change. However, even at this stage of understanding, some reasonable conjectures can be made.

One factor appears to be the seduction of mathematical models, especially embedded in software, over historical experience. Markets go down as well as up. The history of economic bubbles provides ample evidence that people tend to ignore this fact (Mackay, 1841). Coming up with an accurate and clear model of this debacle is itself a tough qualitative modeling challenge. For example, one of the factors that has made the current crisis so widespread is the dangerous process of "repackaging" mortgages as if they were securities. If qualitative models of causal factors affecting risk were included, and propagated through the multiple levels of repackaging, they might have helped alert investors to the potential dangers. Such models will require reasoning about distributions – if the economic climate becomes tougher, business will lay off employees. If many people are laid off, then they will be unable to pay their mortgages[2], leading to the collapse of these "securities." This in turn makes the economic climate tougher still, by drying up credit. Being able to systematically examine worst-case, as well as best-case, possible outcomes might help mitigate the "herd thinking" that underlies bubbles.

What is to be done?

We believe that the QR community has unique contributions to make in helping to improve public education on climate change, financial problems, and other issues raised by our more complex and more tightly interconnected world. We see a three-pronged approach as necessary: (1) more research on the nature of human mental models, including experiential knowledge, (2) more research on how to improve human learning and reasoning, and (3) construction of tools that help people reason and learn, based on the best available results from cognitive science (including learning sciences). We consider each in turn.

Understanding human mental models

In the 1980s and early 1990s, much of the energy in the QR research community was spent on developing formalisms for qualitative dynamics. While the accounts developed have been shown to be robust, by being used in a wide variety of problems and domains, the climate and financial crises illustrate that either (a) these formalisms are not being used by people or (b) there are other representations and processes being used in human reasoning as well. The evidence against (a) mostly comes from protocol analyses, and more research establishing that people do in fact use ideas like qualitative proportionality to organize causal models is needed. The evidence for (b) is strong, e.g. the misconception literature in science education. We believe that the nature of experience-based knowledge must be better understood, and that no account of qualitative reasoning and its place in common sense will be complete without it.

Another reason for strengthening qualitative reasoning skills is to overcome the blind acceptance of the authority of mathematical models. In the financial crisis particularly, executives relied on models produced by their "quants" without fully understanding their implications. Better articulation of the underlying assumptions and causal factors assumed might have led to more caution.

Psychologists have an easier time exploring experience-based knowledge because they can study systems that have plenty of it (i.e., people). For computational modeling, the situation is more complicated. Most QR systems are either fed their knowledge by hand, or are processing information from a specific set of numerical sensors. Hand-feeding systems descriptions expressed in their internal representations does not scale very far. Exploring the role of experience in qualitative reasoning requires finding reasonable approximations to the representations that people build up by interacting with the world. Importantly, by "world" we mean both the physical world and the cultural world: Many physical phenomena are only experienced at best indirectly, with our models of them gleaned from our culture, via reading, lectures, and conversations.

As progress in vision and robotics continues, there will be platforms where experience can be directly gathered by interacting with the physical world. But we need not wait for such platforms, especially given the importance of cultural inputs in human learning. It is already possible to create systems that semi-automatically produce formal representations from simplified natural language (e.g. Kuehne & Forbus, 2004; Tomai & Forbus 2009) and sketches (e.g. Forbus *et al* 2008). These media are relatively easy to produce, and can be used to experiment with learning experience-based models (e.g., Friedman & Forbus 2008; Friedman & Forbus 2009).

Improving human learning and reasoning

The misconception literature in science education shows that helping people achieve accurate models of physical

---

[2] As of 3/5/09, 48% of Americans with subprime mortgages are behind on their payments or are in foreclosure. Source: http://www.newsvine.com/_news/2009/03/05/2508945-mortgage-woes-break-records-again-in-4q

phenomena is quite difficult. For many problems that become matters of public policy there are two additional sources of difficulty: (1) They are more complex, in terms of the number and variety of causal influences and (2) There can be vested interests actively attempting to sow confusion, to better achieve their own ends. When education becomes an adversarial game, it becomes much harder.

Here is an example: George Will, in the Washington Post on 2/15/09, wrote

> "As global levels of sea ice declined last year, many experts said this was evidence of man-made global warming. Since September, however, the increase in sea ice has been the fastest change, either up or down, since 1979, when satellite record-keeping began. According to the University of Illinois' Arctic Climate Research Center, global sea ice levels now equal those of 1979."

There are a number of misstatements here. The first is a misrepresentation of how data are evaluated. As Andrew Revkin, in a New York Times blog posting[3], puts it,

> "No single year marks a trend or holds evidence of long-term climate change."

He quotes Jennifer Francis, from Rutgers, who responds to one of Will's assertions with an excellent qualitative explanation of why warming has contributed to the speed of ice recovery:

> "At the end of summer each year, the sea ice refreezes and continues to do so until late spring. Thin ice and open water generate new ice faster than thick ice, as the heat from the ocean below is able to escape more easily to the atmosphere. In the autumns of 2007 and 2008, the rate of ice production was very large because there was so much open water and thin ice – the rapid growth is completely expected."

Mr. Will's confusion is symptomatic of a major problem we have in our culture. When journalists and opinion-makers have trouble understanding scientific evidence and arguments, the effect of their confusion is multiplied by decreasing the clarity of public debate.

Building tools to support reasoning and learning

The QR community has already invested substantial effort into making tools that use qualitative modeling to help students learn and to help inform the public about the possible consequences of policy choices (e.g., Sallas & Bredeweg, 2001). There is certainly much more to be done in this area.

The importance of experiential knowledge in human qualitative reasoning suggests that we need to incorporate ways to exploit it into our tools. For example, our QCM system (Dehghani & Forbus, 2009) is a new qualitative modeling tool aimed at cognitive scientists, to help them model data that they have collected. It deliberately allows users to create situation-specific descriptions of physical processes, rather than forcing them to first create and then instantiate a first-principles domain theory. The idea is that situation-specific models may be all that they need for particular investigations, and that even if their goal is to construct a robust, broadly-applicable first-principles domain theory for some area of human knowledge, building concrete, specific models is a better way to start. In other words, contemplating multiple specific models may be a better way to formulate general domain theories. We suspect that the same approach could be useful for students as well, given the success of Betty's Brain and VModel.

Another way to incorporate experiential knowledge in our learning and reasoning tools is to enable them to work with their users' analogies, and to supply their own. Explicitly helping people work through correspondences and seeing what analogical inferences follow, for example, could be a valuable service in a learning environment. A system could propose new analogies, drawing upon interesting examples it has formally represented as part of its world knowledge. People can often work through an analogy once it is proposed, but they find it much harder to retrieve distant (as opposed to close) analogs (Gentner, Rattermann, & Forbus 1993). Support software can potentially have an easier time retrieving distant analogs, since they have fewer distracting experiences, fewer distracting perceptual representations, and can encode experiences thoroughly.

A Call to Arms

We believe that the ideas and formalisms developed by the qualitative reasoning community can play an important role in public education. Democracies require informed citizens. In today's world, citizens are faced with the need to understand quite subtle arguments about very complex interlocking systems, and have to sift through both honestly conflicting evidence and special-interest induced hazes. We believe that the ability to do robust, sound qualitative reasoning is an important part of meeting this need. But to succeed, we must take into account experiential knowledge, the "dark knowledge" of QR, because it seems to play a central role in human mental models. Our models of reasoning and learning need to incorporate it, and our designs for educational systems and interventions need to take it into account.

Advances in natural interaction modalities (natural language, sketch understanding, vision, robotics) provide new tools by which we can accumulate in digital form knowledge about experience. We hope that this will facilitate research on the roles of experience in qualitative reasoning. And we hope that this, in turn, will help us develop a new generation of QR techniques and systems to help with these crucial matters of public education.

---

[3] http://dotearth.blogs.nytimes.com/2009/02/27/expers-big-flaw-in-wills-ice-assertions/

Acknowledgements

We thank Morteza Dehghani, Scott Friedman, Andrew Lovett, and Jon Wetzel for helpful suggestions.

References

Biswas, G., Schwartz, D., Bransford, J. & The Teachable Agents Group at Vanderbilt. (2001) Technology Support for Complex Problem Solving: From SAD Environments to AI. In K. Forbus and P. Feltovich (Eds.). *Smart Machines in Education*, 71-98. AAAI Press/MIT Press

Bredeweg, B., Salles, P., Bouwer, A., Liem, J., Nuttle, T., Cioaca, E., Nakova, E., Noble, ., Caldas, A., Uzonov, Y., Varadinova, E., and Zitek, A. 2008. Towards a structured approach to building qualitative reasoning models and simulations. *Ecological Informatics*, 3(1),1-12. Elsevier.

Clement, J. 1982. Students' preconceptions in introductory mechanics. *American Journal of Physics* 50(1):66-71.

Cronin, M., Gonzalez, C., & Sterman, J. 2009. Why don't well-educated adults understand accumulation? A challenge to researchers, educators, and citizens. *Organizational Behavior and Human Decision Processes* 108,116-130.

Deghani, M. and Forbus, K. 2009. QCM: A QP-based concept map system. *Proceedings of QR09.*

diSessa, A. (1993). Toward an Epistemology of Physics. *Cognition and Instruction*, 10(2-3), 105-225.

Falkenhainer, B. 1992. Modeling without amnesia: Making experience-sanctioned approximations. *Proceedings of QR92*, 44-55.

Feltovich, P., Coulson, R., and Spiro, R. 2001. Learners' (Mis)understanding of important and difficult concepts. In Forbus, K., and Feltovich, P. (Eds.) *Smart Machines in Education,* 349-375, AAAI Press/MIT Press.

Forbus, K. (1984). Qualitative process theory. *Artificial Intelligence*, 24, 85-168.

Forbus, K., Carney, K., Sherin, B. and Ureel, L. 2004. VModel: A visual qualitative modeling environment for middle-school students. *Proceedings of the 16th Innovative Applications of Artificial Intelligence Conference*, San Jose.

Forbus, K. and Gentner, D. (1986). Learning Physical Domains: Towards a Theoretical Framework. In Michalski, R., Carbonell, J. and Mitchell, T. (Eds.), *Machine Learning: An Artificial Intelligence Approach*.

Forbus, K., Usher, J., Lovett, A., Lockwood, K., & Wetzel, J. (2008). CogSketch: Open-domain sketch understanding for cognitive science research and for education. *Proceedings of the Fifth Eurographics Workshop on Sketch-Based Interfaces and Modeling*. Annecy, France.

Friedman, S. & Forbus, K. (2008). Learning Causal Models via Progressive Alignment & Qualitative Modeling: A Simulation. Proceedings of CogSci 2008.

Friedman, S., Forbus, K., 2009. Learning Naïve Physics models and misconceptions. *Proceedings of CogSci09.*

Gentner, D. 1983. Structure-Mapping: A theoretical framework for analogy, *Cognitive Science* 7(2):155-170.

Gentner, D., Levine, S., Dhillon, S. & Poltermann, A., (in press) Using structural alignment to facilitate learning of spatial concepts in an informal setting. To appear in *Proceedings of the Second International Workshop on Analogy*, Sofia, Bulgaria, 2009

Gentner, D., Loewenstein, J., Thompson, L. 2003. Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology*, 95, 393-408.

Gentner, D., Rattermann, M.J., & Forbus, K. 1993 The roles of similarity in transfer: Separating retrievability from inferential soundness. *Cognitive Psychology* 25, 524-575.

Halloun, I. & Hestenes, D. 1985 Common-sense concepts about motion. *Am. J. Phys.* 53, 1056.

Klenk, M., Forbus, K., Tomai, E., Kim,H., and Kyckelhahn, B. (2005). Solving Everyday Physical Reasoning Problems by Analogy using Sketches. *Proceedings of 20th National Conference on Artificial Intelligence (AAAI-05),* Pittsburgh, PA.

Kotovsky, L., & Gentner, D. 1996. Comparison and categorization in the development of relational similarity. *Child Development*, 67, 2797-2822.

Kuehne, S. and Forbus, K. (2004). Capturing QP-relevant information from natural language text. *Proceedings of the 18th International Qualitative Reasoning Workshop*, Evanston, Illinois, August.

Kuipers, B. 1994. *Qualitative Reasoning: Modeling and simulation with incomplete knowledge.* Cambridge, Mass.: MIT Press.

Kurtz, K., Miao, C., & Gentner, D. 2001. Learning by analogical bootstrapping. *Journal of the Learning Sciences*, 10(4):417-446.

Mackay, C. 1841. *Extraordinary Popular Delusions & the Madness of Crowds*. Three Rivers Press.

McCloskey, M. 1983. Naïve theories of motion. In Gentner, D. and Stevens, A., (Eds) *Mental Models*. 299-324. Erlbaum.

Richland, L., Zur, O., and Holyoak, K. 2007. Cognitive supports for analogies in the mathematics classroom. *Science* 316:1128-1129, May 25th.

Salles, P. & Bredeweg, B. 2001. Constructing Progressive Learning Routes through Qualitative Simulation Models in Ecology. *Artificial Intelligence in Education: AI-ED in the Wired and Wireless Future*, J.D. Moore, G. Luckhardt Redfield, and J.L. Johnson (eds), pages 595-597, IOS-Press/Ohmsha, Japan, Osaka.

Spiro, R., Feltovich, P., Coulson, R., & Anderson, D. 1989. Multiple analogies for complex concepts: Antidotes for analogy-induced misconception in advanced knowledge acquisition. In *Similarity and Analogical Reasoning*, eds. S. Vosniadou & A. Ortony, 498-531. Cambridge University Press.

Sterman, J. 2008. Risk Communication on Climate: Mental Models and Mass Balance. *Science*, 322(532-533), Oct 24.

Tomai, E. and Forbus, K. (2009). EA NLU: Practical Language Understanding for Cognitive Modeling. *Proceedings of the 22nd International Florida Artificial Intelligence Research Society Conference*. Sanibel Island, Florida.

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction*, 4, 45-69.

Learning and Reasoning with Qualitative Models of Physical Behavior

Scott E. Friedman, Kenneth D. Forbus, & Jason Taylor
friedman@northwestern.edu, forbus@northwestern.edu, jason.taylor@eecs.northwestern.edu
Qualitative Reasoning Group, Northwestern University
2133 Sheridan Road, Evanston, IL USA

Abstract

Building models of the physical world from examples is an important challenge for qualitative reasoning systems. We describe a system that can learn intuitive models of physical behaviors from a corpus of multimodal, multi-state stimuli, consisting of sketches and text. The system extracts and temporally encodes exemplars from the stimuli and uses analogical generalization to abstract prototypical behaviors. Using statistical analysis, the system parameterizes these abstractions into qualitative representations for reasoning. We show that the explanations the system provides for new situations are consistent with those given by naïve students.

Keywords: Cognitive modeling; conceptual change; misconceptions; naïve physics; qualitative reasoning

Introduction

Many people have intuitive models of physical domains that are at odds with scientific models (Smith, diSessa, & Roschelle, 1994; diSessa, 1993; Brown, 1994; Vosniadou, 1994). While productive for reasoning about everyday physical phenomena, these naïve models cause patterns of misconceptions. These misconceptions may result from improperly generalizing or contextualizing experience (Smith, diSessa, & Roschelle, 1994) or from incorporating instruction into a flawed intuitive framework (Vosniadou, 1994). Understanding how such intuitive models come about is an important problem for understanding conceptual change (Forbus & Gentner, 1986).

Computational models of conceptual change (e.g. Esposito et al., 2000; Ram, 1993) tend to describe how existing concepts are changed, but not how those initial concepts are learned. We believe it is important for such models to encompass the learning of the initial concepts, to reduce tailorability. This paper describes a simulation of learning intuitive physics models from experience. Experiences are provided as combinations of sketches and natural language, which are automatically processed to produce symbolic representations for learning. The encoding process is centered on the concepts to be learned, and it constructs qualitative representations of behavior across time as exemplars. Analogical generalization is used with a statistical criterion to induce abstract models of typical patterns of behavior, which constitutes our representation of intuitive models. These models can be used to make predictions and perform simple counterfactual reasoning. We compare the system's explanations to those of human students on reasoning tasks from Brown (1994) and the Force Concept Inventory (Hestenes et al., 1992).

We next briefly summarize the relevant aspects of qualitative process theory and structure-mapping theory used in the simulation. Then we describe how our stimuli are represented and encoded, motivated by results and ideas from the cognitive science literature. The learning process itself is described next, followed by how these models are used in reasoning. We show that the system's explanations of two physical situations are compatible with student explanations. We close with related and future work.

Qualitative Process Theory

People's intuitive physical knowledge appears to rely heavily on qualitative representations (Forbus & Gentner, 1986; Baillargeon, 1998). Consequently, we use qualitative process theory (Forbus, 1984) as part of our model. The learning we model here is what provides the foundation for ultimately learning physical processes; in the framework of Forbus & Gentner (1986), we are modeling the construction of *protohistories* to describe typical patterns of behavior from experience, and building on those a *causal corpus* consisting of causal relationships between those typical patterns. To represent these patterns of behavior, we use the concept of *encapsulated history* (EH) from QP theory.

An encapsulated history represents a category of abstracted behavior, over some span of time. Unlike model fragments, EHs can mention time explicitly, referring to multiple qualitative states and events. The *participants* are the entities over which an EH is instantiated. The *conditions* are statements which must hold for an instance of the EH to be *active.* When an instance of an EH is active, the statements in its *consequences* are assumed to be true. We use encapsulated histories as explanatory schemata: When instantiated, they provide an explanation for a behavior via recognizing it as an instance of a typical pattern. Furthermore, they can predict possible causes and consequences of a behavior, and hypothesize hidden conditions when a behavior is known to be active.

Since EHs can include multiple qualitative states, they can be used for learning causal relationships between behaviors and properties of the world. In naïve mechanics, for example, the models of movement, pushing, and blocking learned by the simulation are represented by EHs.

Figure 1 illustrates an EH learned by the simulation. This can be read as: *P1* pushes *P2* while *P1* and *P2* touch; the direction *dir1* from the pusher *P1* to the pushed *P2* matches the direction of the push; and pushed *P2* consequently moves (*M1*) in the direction *dir1* of the push. When given a test scenario, the system checks its learned EHs to determine whether its participants match entities in the scenario. If so, instances of those EHs are created. Each EH instance is active only if the statements in its conditions hold in the scenario. If the consequences fail to hold, that is a prediction failure of an active EH.

Encapsulated history consequences may contain typicality expressions, such as the `Normal-Usual` attribute in Figure 1. Inferring this consequence in a scenario context indicates that the phenomenon (here, the `PushingAnObject` event) has been explained by an encapsulated history.

```
define-encapsulated-history Push05
Participants:
Entity(?P1), Entity(?P2), PushingAnObject(?P3),
Direction(?dir1), Direction(?dir2)

Conditions:
providerOfMotiveForce(?P3, ?P1),
objectActedOn(?P3, ?P2),
dir-Pointing(?P3, ?dir1),
touches(?P1, ?P2),
dirBetween(?P1, ?P2, ?dir1),
dirBetween(?P2, ?P1, ?dir2)

Consequences:
Normal-Usual(and(PushingAnObject(?P3),
                 providerOfMotiveForce(?P3, ?P1),
                 objectActedOn(?P3, ?P2)))
causes-SitProp(Push05,
                (exists ?M1
                  (and MovementEvent(?M1),
                       objectMoving(?M1, ?P1),
                       motionPathway(?M1, ?dir1)))
```

Figure 1: **An encapsulated history relating pushing and movement.**

Analogical Generalization

Our hypothesis is that people use analogical generalization to construct encapsulated histories. To model this process, we use SEQL (Keuhne et al., 2000). SEQL is grounded in structure-mapping theory (Gentner, 1983), and uses the Structure-Mapping Engine, SME (Falkenhainer et al., 1989). Given two representations, a base and a target, SME computes a set of mappings that describe how they can be aligned (i.e. correspondences), candidate inferences that might be projected from one description to the other, and a structural evaluation score that provides a numerical measure of similarity. SEQL uses SME as follows. SEQL maintains a list of exemplars and generalizations. Given a new exemplar, it is first compared against each generalization using SME. If the score is over the *assimilation threshold*, they are combined to update the generalization. Otherwise, the new exemplar is compared with the unassimilated exemplars. Again, if the score is high enough, the two exemplars are combined to form a new generalization. Otherwise, the exemplar is added to the list of unassimilated exemplars. The combination process maintains a probability for each statement in a generalization, based on how frequently it occurred in the exemplars generalized within (Halstead & Forbus, 2005). These probabilities are used in our simulation for doing statistical tests.

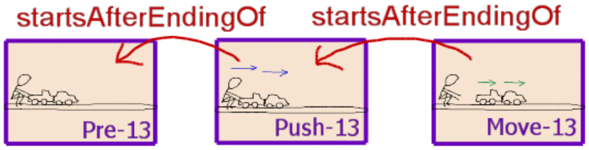


Figure 2: A sketched behavior

Multimodal Stimuli

To reduce tailorability, we provide experiences to the simulation in the form of sketches (e.g. Figure 2) accompanied by natural language text. This serves as an approximation to what learners might perceive and hear in the world. The sketches are created in CogSketch[1] (Forbus et al., 2008), an open-domain sketch understanding system. In CogSketch, users draw and label *glyphs*, objects in the sketch, to link the content of the sketches to concepts in CogSketch's knowledge base[2]. CogSketch automatically computes qualitative spatial relations between the glyphs such as topological relations (e.g. touching), relative size, and positional relationships (e.g. above).

Sketched behaviors are segmented into distinct states according to qualitative differences in behavior (e.g. changes in contact and actions of agents) to accord with findings in psychological event segmentation (Zacks, Tversky, & Iyer, 2001). Each state is drawn as a separate sketch. The sequential relationships between them are drawn as arrows on the *metalayer*, where other sub-sketches are treated as glyphs, as Figure 2 illustrates. The child, truck, and car are glyphs in the sketched states. The two right-pointing arrows in state *Push-13* are *pushing* annotations, and the two right-pointing arrows in state *Move-13* are *velocity* annotations.

Two lines of evidence motivate our encoding of the physical phenomena of pushing, movement, and blocking as separate concepts. diSessa (1993) notes that people are unlikely to confuse successful resistance (i.e. a wall blocking a person's push) from nonsuccess (i.e. a ball moving due to tugging a string) in recalling events, and that these phenomena are encoded separately. Talmy (1988) attributes this separation of success and nonsuccess encoding to varying language schemata between the two conditions.

For information not easily communicated via sketching, we use simplified English, which is converted to predicate calculus via a natural language understanding system (Tomai & Forbus, 2009). One sentence used in conjunction with the sketch in Figure 2 is, "The child child-13 is playing with the truck truck-13." The special names **`child-13`** and **`truck-13`** are the internal tokens used in the sketch for the child and the truck respectively, so that linguistically

---

[1] CogSketch is available online at http://spatiallearning.org/projects/cogsketch_index.html

[2] CogSketch uses a combination of knowledge extracted from OpenCyc (www.opencyc.org) and our own extensions for qualitative, analogical, and spatial reasoning.

expressed information is linked with information expressed via the sketch. This sentence leads to these assertions being added to the exemplar:

```
(isa truck-13 Truck)
(isa play1733 RecreationalActivity)
(performedBy play1733 child-13)
(with-UnderspecifiedAgent play1733 truck-13)
```

If the NLU system finds an ambiguity it cannot handle, it displays alternate interpretations for the experimenter to choose. No hand-coded predicate calculus statements are included in the stimuli.

This method of simulation input has limitations: Sketches are less visually rich than images, and they do not provide opportunities for the learner to autonomously experiment. Nevertheless, we believe that this is a significant advance over the hand-coded stimuli typically used by other systems, given the reduction in tailorability. These multimodal stimuli are used by our system as examples for learning and as scenarios for reasoning.

Learning

The system is provided with a set of target phenomena to learn, here *pushing*, *movement*, and *blocking*. We assume that for a truly novice learner, words used in contexts of behaviors that they do not understand are clues that there is something worth modeling.

Given a new stimulus, the system finds all instances of target phenomena that it describes, and generates an exemplar for each instance. Since an instance of a particular phenomenon may continue across state boundaries, these occurrences can span multiple states. Temporal relationships between these occurrences are derived to support learning of preconditions and consequences. For example, consider a series of states S_1-S_3, where a man is pushing a crate in S_1-S_2 and not in S_3, and the crate moves in S_2-S_3 but not in S_1. The motion would have a `startsDuring` relationship with the pushing. Each stimulus observed by the simulation is automatically temporally encoded into exemplars using this strategy.

Generalizing behaviors

For each target phenomenon, the system maintains a separate instance of SEQL, a *generalization context* (Friedman & Forbus, 2008). A generalization context has an entry pattern that is used to determine when an exemplar is relevant. For example, the entry pattern for *pushing* is:

```
(and (isa ?x PushingAnObject)
     (providerOfMotiveForce ?x ?y)
     (objectActedOn ?x ?z))
```

Figure 3 shows the generalization contexts and their contents after the learning experiment described below. Our system currently operates in batch mode, not attempting to construct models until after all of the stimuli have been processed.

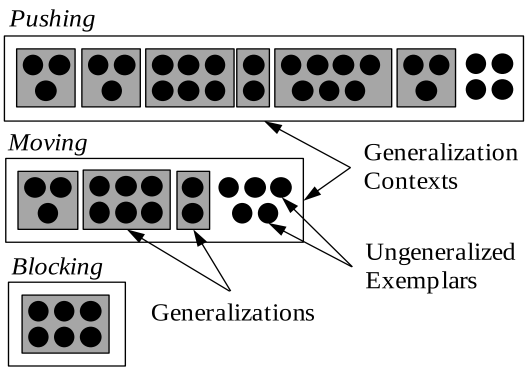


Figure 3: Generalization contexts after learning

Constructing intuitive models

The system creates encapsulated histories from generalizations in two steps: (1) Statistics are used to determine which generalizations are worth modeling with EHs, and (2) worthwhile generalizations are parameterized to create EHs. We discuss each step in turn.

Filtering generalizations

Not all SEQL generalizations can be parameterized into useful encapsulated histories. Some generalizations are overly broad, and would result in EHs that make inaccurate predictions. Consequently, the system filters out overly broad generalizations using the probability information constructed during generalization.

Generalizations are filtered by identifying correlated phenomena within generalizations and measuring the phenomena's correlation across generalizations. We assume a probability threshold t (here, 0.9) for correlation. That is, if any target phenomenon p is in a generalization with probability $P(p) \geq t$, then p is considered a *correlated phenomenon* within that generalization's context. A generalization is *decisive* if the binary entropy of all correlated phenomena p are less than the binary entropy of t, or $H(P(p)) \leq H(t)$. Entropy is the appropriate criterion to use because it measures information gain (i.e., low entropy implies high gain). Only decisive generalizations are parameterized into encapsulated histories.

Extracting Causal Models from Generalizations

The system creates one encapsulated history per decisive generalization. Expressions whose probability is lower than the probability threshold t (here, 0.9) are excluded from the EH, thus reducing contingent phenomena. Expressions that remain are analyzed to determine what role they should play in the encapsulated history.

An expression is held to be either (a) a *cause* of the state, (b) a *consequence* of the state, or (c) a *condition* that holds during the state, based on analyzing the temporal relationships involved. If an expression begins with the current state, ends with the start of the current state, or ends during the current state, it is a possible cause. If it temporally subsumes or coincides with the state, it is a possible condition. Otherwise, if it begins at any point during or immediately following the current state, it is a possible consequence.

Probabilities and temporal relationships are used to hypothesize causality. For instance, in one generalization, movement starts *with* a pushing event with P = 0.5, and starts *after* a pushing event with P = 0.5. In this case, movement is not a likely condition for pushing because it only satisfies the temporal requirement half the time, P(*starts-with*) < *t*. Conversely, movement is a likely consequence, because starting *with* and starting *after* are both permissible temporal relations of consequences, and P(*starting-with*) + P(*starting-after*) > *t*.

After the causes, conditions, and consequences are determined, the system defines an encapsulated history by introducing variables for entities that appear in the conditions, creating existence statements for the entities that appear only in the consequences, and using the generalization's attribute information to construct the participants information. Figure 1 and Figure 5 illustrate. Notice that, while the learning process removes most irrelevancies, in **`Block00`** the entity **`?P1`** is included even though it is not causally relevant. It is there because the examples involving pushing all involve the pushing agent standing or sitting on a surface – so to the system, blocking must involve touching something else.

Reasoning with Encapsulated Histories

Given a new scenario, the system attempts to make sense of it by instantiating its encapsulated histories. For each EH, it finds instances within the scenario. When an instance's conditions hold, it is active, and the statements in its **`Consequences`** are assumed to hold. This can include predicting new phenomena, as illustrated by the movement *M1* consequence in Figure 1. When constraints are violated, or consequences are not satisfied, the EH instance can be used to generate counterfactual explanations, as explained below.

To illustrate, consider a scenario used by Brown (1994) and others, illustrated in Figure 4. The sketch shows a book on a table. Gravity pushes down on the book and the table.

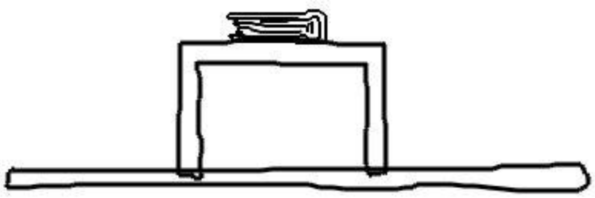


Figure 4: An example from Brown (1994) for testing learned knowledge

The scenario description includes two occurrences of pushing: gravity pushing the book and gravity pushing the table. The encapsulated history in Figure 5 can be instantiated sufficiently to be considered for inference by the simulation, since the criterion is that all non-event participants be identifiable in the scenario. Some event participants, such as pushing and blocking, need not be identified because these can be instantiated as predictions.

```
define-encapsulated-history Block00
Participants:
Entity(?P1), Entity(?P2), Entity(?P3), Entity(?P4),
PushingAnObject(?P5), PushingAnObject(?P6),
Blocking(?P7)

Conditions:
providerOfMotiveForce(?P5, ?P2),
objectActedOn(?P5, ?P3),
dir-Pointing(?P5, ?dir1),
providerOfMotiveForce(?P6, ?P3),
objectActedOn(?P6, ?P4),
dir-Pointing(?P6, ?dir1),
doneBy(?P7, ?P4),
objectActedOn(?P7, ?P3),
dirBetween(?P2, ?P3, ?dir1),
dirBetween(?P3, ?P4, ?dir1),
dirBetween(?P3, ?P2, ?dir2),
dirBetween(?P4, ?P3, ?dir2),
touches(?P2, ?P3),
touches(?P3, ?P4),
touches(?P2, ?P1)

Consequences:
Normal-Usual(and(PushingAnObject(?P5),
                providerOfMotiveForce(?P5, ?P2),
                objectActedOn(?P5, ?P3)))
Normal-Usual(and(PushingAnObject(?P6),
                providerOfMotiveForce(?P6, ?P3),
                objectActedOn(?P6, ?P4)))
Normal-Usual(and(Blocking(?P7), doneBy(?P7, ?P4),
                objectActedOn(?P7, ?P3)))
```

Figure 5: An encapsulated history relating pushing and blocking phenomena

Specifically, activating **`Block00`** to explain gravity pushing the book requires assuming two additional events, per the conditions in Figure 5: (1) gravity ?*P2* pushes the book ?*P3* in the direction *?dir1* of the initial push, and (2) an entity ?*P4* blocks the book ?*P3*. The table alone satisfies the constraints on *?P4*, binding the last of the non-event participants. This is sufficient grounds for the simulation to instantiate new pushing and blocking events, binding them to *?P6* and *?P7*, respectively.

The simulation has two strategies for answering questions about a scenario. If the question concerns a phenomenon that is predicted by the EH instances it has created for the scenario, it answers based on that information, including any causal argument provided as part of the EH. If the question concerns some phenomenon that is not predicted, it assumes that phenomenon occurs and looks at what new EHs could be instantiated to explain it. The instantiation failures for those EH instances are provided as the reasons for the phenomenon not occurring, as shown below.

Experiment

To test whether this model can learn psychologically plausible encapsulated histories from multimodal stimuli, we observe the explanations it provides for a question from Brown's (1994) assessment of student mental models and a question from Hestenes et al.'s (1992) Force Concept Inventory. We start by summarizing human results, then describe the conditions used for the simulation, and compare the human and simulation results.

Brown's results

A question about the scenario in Figure 5 was asked of high school students: *Does the table exert a force against the book?*

Brown reported that 33 of 73 students agreed that it must, in order to counteract the downward force of the book. This is the physically correct answer. However, the 40-student majority denied that the table exerted a force. Their reasons fell into five categories:

1. Gravity pushes the book flat, and the book exerts a force on the table. The table merely supports the book (19 students)
2. The table requires energy to push (7 students)
3. The table is not pushing or pulling (5 students)
4. The table is just blocking the book (4 students)
5. The book would move up if the table exerted a force (4 students)

We query our simulation similarly, to determine whether it can reproduce some of the reasons that students gave.

Force Concept Inventory

The Force Concept Inventory (FCI) (Hestenes et al., 1992) is an assessment designed to identify student misconceptions about force. Many FCI questions involve the relationships between force, mass, and velocity, and the composition of forces to determine direction of motion. Figure 6 illustrates our sketch of question 6 from the FCI. The scenario describes a puck on a frictionless surface, moving with constant velocity, until it receives an instantaneous kick. The student must decide along which of the five paths (labeled choice-27-a/b/c/d/e below) the puck will move after receiving the kick.

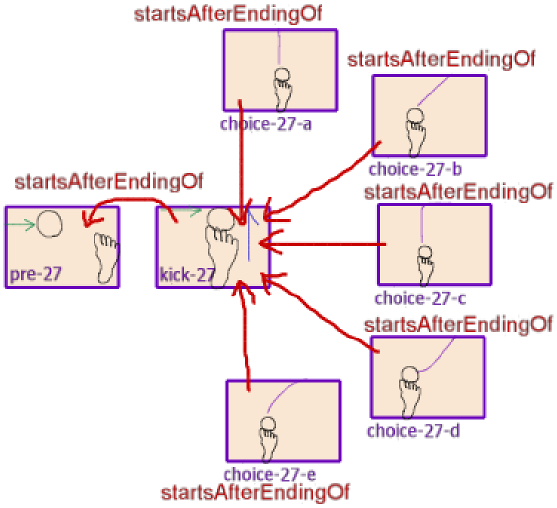


Figure 6: An example from the Force Concept Inventory (Hestenes et al, 1992)

Five pre-physics-instructed student populations, ranging from high school to college, predicted the puck would, on average:

(a) 34% - move upward, in the direction of the kick.
(b) 38% - per Newtonian principles, move diagonally.
(c) 3% - move upward and then curve to the right.
(d) 6% - gradually curve in the direction of the kick.
(e) 18% - curve in the direction of initial motion.

Other FCI questions concerned the relationships between velocity, mass, and acceleration, which were not target concepts of our simulation.

Simulation setup

Our simulation was implemented using the Companion Cognitive Systems architecture (Forbus et al., 2008), using semi-independent asynchronous agents. The Session Reasoner (the Companions agent responsible for domain reasoning) begins with 17 sketches with accompanied natural language as learning stimuli. Like Figure 2, all stimuli include pushing phenomena, and either movement or blocking phenomena. The learning stimuli did not include the test scenarios.

For each stimulus, the Session Reasoner first encodes it into exemplars, resulting in a total of 28 pushing exemplars, 16 moving exemplars, and 6 blocking exemplars. Before encoding the next stimulus, the Session Reasoner contacts the *Analogical Tickler* agent to generalize the exemplars using SEQL. The SEQL assimilation threshold was set to 0.5, which results in ten generalizations across the three generalization contexts, as illustrated in Figure 3.

After all of the learning stimuli are encoded and the exemplars are generalized, the Session Reasoner generates EHs from the resulting SEQL generalizations. The EH probability threshold was set to 0.9. Consequently, six of the generalizations were decisive, leading to the push→move model of Figure 1, the push→block model in Figure 5, and four additional models.

The four additional models learned by the system were not activated during problem solving. Three EHs describe movement behaviors caused by pushing, with minor variations in the conditions. The fourth EH describes classic "billiard ball" causality, with a push causing motion, which then causes another push and setting another entity into motion.

Both problem solving scenarios are conducted by the Session Reasoner, which tries to activate its learned EHs within the scenario contexts.

Comparison with human results

Given these EHs, how does the system perform? When given Brown's (1994) test scenario, the system activates EHs to infer the additional events of the book pushing down against the table and the table pushing down against the ground.

For Brown's query, since the simulation does not have the event of the table pushing upward against the book as a current prediction, it uses the counterfactual strategy. Only the EH of Figure 1 can provide a possible explanation. Assuming this EH is active, the simulation gets a new prediction: The book should move upward as a result of the table's push. This prediction contradicts the book's lack of motion in the scenario. Consequently, it answers that the table does not push up on the book. This is essentially the same as answer 5, given by four students.

After the proof by contradiction, the system identifies activated EHs in which the book and table jointly participate to explain their behavior in the scenario. Consequently, it uses the EH in Figure 5 to explain that gravity pushes down on the book, that the book pushes down on the table, and that the table blocks the book. This is similar to answer 4, given by four students. This explanation also resembles answer 1, given by 19 students, though the students cite the concept of support, which was not among the simulation's target phenomena. Could the system learn models corresponding to the other explanations for this scenario? If the target phenomena and corpus included the concept of support and energy, it seems likely to us that it could, but this is an empirical question. With a different corpus of examples – perhaps including examples like those used by Camp & Clement (1994) and the rest of Brown (1994) – the simulation may be capable of coming to the correct model. Answer 3 may rest on an interpretation of events being mutually exclusive, i.e., if the table is blocking, then it cannot be doing the other actions. Further experiments should clarify this.

When given the FCI scenario, the system activates the EH from Figure 1 within the "kick" state and predicts that the puck will translate in the direction of the kick during or immediately after the kick. Upon evaluating all possible following states, the system concludes that *choice-27-a* is the only successor state that fulfills this prediction. The system predicts this path for the puck, as do 34% of the FCI-assessed students in Hestenes et al. (1992), which represents the most popular misconception. The results from both scenarios support the hypothesis that the models learned by the system are like those used by physics-naïve students.

Related Work

The closest simulations are the COBWEB (Fisher, 1987) model of conceptual clustering and INTHELEX (Esposito et al., 2000), which develops and revises prolog-style theories. COBWEB does unsupervised learning of hierarchical relationships between concepts, in contrast with our use of supervised learning (via entry patterns in generalization contexts) of causal models. COBWEB calculated probabilities of features, whereas SEQL provides probabilities of structured relations. INTHELEX uses refinement operators to model multiple steps in a trajectory of learned models, whereas we focus only on one transition, the first. Both COBWEB and INTHELEX used hand-represented input stimuli, whereas ours is derived by the simulation from sketches and natural language. Ram (1993) discusses SINS, a robot navigation system that retrieves cases, adapts control parameters, and learns new associations incrementally. Both our system and SINS develop concepts incrementally from experience; however, our system learns models of physical behaviors and causal laws, while SINS learns associations between environmental conditions and control parameters.

Lockwood et al. (2005) used CogSketch and SEQL to model the learning of spatial prepositions, using single sketches labeled with words, in contrast to the sequences of sketches labeled with sentences used here.

Discussion & Future Work

We have described how analogical generalization and qualitative modeling can be used to simulate the process of learning initial intuitive models. To reduce tailorability, the simulation inputs were combinations of sketches and simplified English. The resulting answers match a subset of those of given by human students on the same scenarios.

While we believe that this is a significant first step, there is much more to be done. Other domains and physical phenomena must be incorporated, to provide more evidence as to generality. Second, we need to conduct statistical tests to determine how order-sensitive the simulation is, and how the quality of models learned varies with the number of examples provided. Additionally, modeling the induction of physical process models from the encapsulated histories learned by the system is an important step in learning intuitive physics (Forbus & Gentner, 1986).

Finally, we plan to incorporate these ideas in a larger-scale learning model, where the quality and content of its predictions guide future learning. The Companion Cognitive Systems architecture is an ideal platform for this endeavor because one of its primary goals is ubiquitous learning over an extended lifetime. With our learning and reasoning methodologies integrated into Companion Cognitive Systems, agents can use multimodal stimuli to learn new models and evaluate the productivity of existing models. These are important characteristics of a larger model of conceptual change.

Acknowledgments

This work was funded by the Office of Naval Research under grant N00014-08-1-0040. We wish to thank Dedre Gentner for discussions of concept formation with SEQL.

References

Baillargeon, R. (1998). A model of physical reasoning in infancy. *Advances in infancy research*, 3, 305-371.

Brown, D. (1994). Facilitating conceptual change using analogies and explanatory models. *International Journal of Science Education*, 16(2), 201-214.

Camp, C. W., & Clement, J. J. (1994). *Preconceptions in mechanics. Lessons dealing with students' conceptual difficulties.* Dubuque, IA: Kendall/Hunt Publishing Company.

diSessa, A. (1993). Toward an Epistemology of Physics. *Cognition and Instruction*, 10(2-3), 105-225.

Esposito, F., Semeraro, G., Fanizzi, N., & Ferilli., S. (2000). Conceptual Change in Learning Naive Physics: The Computational Model as a Theory Revision Process. In E. Lamma and P. Mello (Eds.), AI*IA99: Advances in Artificial Intelligence, Lecture Notes in Artificial Intelligence 1792, 214-225, Springer: Berlin.

Falkenhainer, B., Forbus, K. & Gentner, D. (1989). The Structure Mapping Engine: Algorithm and examples. *Artificial Intelligence*, 41, 1-63.

Fisher, D. H. (1987). Knowledge acquisition via incremental concept clustering. *Machine Learning*, 2,139-172.

Forbus, K. (1984). Qualitative process theory. *Artificial Intelligence*, 24, 85-168.

Forbus, K. & Gentner, D. (1986). Learning Physical Domains: Towards a Theoretical Framework. In Michalski, R., Carbonell, J. and Mitchell, T. (Eds.), *Machine Learning: An Artificial Intelligence Approach.*

Forbus, K., Usher, J., Lovett, A., Lockwood, K., & Wetzel, J. (2008). CogSketch: Open-domain sketch understanding for cognitive science research and for education. *Proceedings of the Fifth Eurographics Workshop on Sketch-Based Interfaces and Modeling*. Annecy, France.

Forbus, K., Klenk, M. and Hinrichs, T. (2008). Companion Cognitive Systems: Design Goals and Some Lessons Learned. *AAAI Fall Symposium on Naturally Inspired Artificial Intelligence.*

Friedman, S. & Forbus, K. (2008). Learning Causal Models via Progressive Alignment & Qualitative Modeling: A Simulation. *Proceedings of CogSci 2008.*

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7(2).

Halstead, D. and Forbus, K. (2005). Transforming between Propositions and Features: Bridging the Gap. *Proceedings of AAAI-2005*. Pittsburgh, PA.

Hestenes, D., Wells, M., Swackhamer, G. (1992). Force Concept Inventory. *The Physics Teacher*, 30, 141-158.

Kuehne, S., Forbus, K., Gentner, D. and Quinn, B. (2000). SEQL: Category learning as progressive abstraction using structure mapping. *Proceedings of CogSci 2000*.

Lockwood, K., Forbus, K., & Usher, J. (2005). SpaceCase: A model of spatial preposition use. *Proceedings of the 27th Annual Conference of the Cognitive Science Society*.

Ram, A. (1993). Creative conceptual change. *Proceedings of CogSci 1993.*

Smith, J., diSessa, A., & Roschelle, J. (1994). Misconceptions Reconceived: A Constructivist Analysis of Knowledge in Transition. *Journal of the Learning Sciences*, 3(2), 115-163.

Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science*, 12(1), 49-100.

Tomai, E. and Forbus, K. (2009). EA NLU: Practical Language Understanding for Cognitive Modeling. *Proceedings of the 22nd International Florida Artificial Intelligence Research Society Conference*.

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction*, 4, 45-69.

Zacks, J., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology. General.* 130(1), 29-58.

Assessing the Ecological Impacts of Agriculture Intensification Through Qualitative Reasoning

F.F. Goulart [a], P. Salles [b] & C.H. Saito [c]

Institute of Biological Science, Graduate Program in Ecology, University of Brasília
Campus Darcy Ribeiro, Brasília, 70.910-900, Brazil
[a] goulart.ff@gmail.com; [b] psalles@unb.br; [c] saito@unb.br

Abstract

How to feed the world without loosing what is left of biodiversity? Two answers for this question are found in the literature. On the one hand, the "Land Sparing" paradigm suggests that increasing yield by means of intensive agricultural systems would fulfill the needs of human population and save natural landscapes. On the other hand, "Biodiversity Friendly Farming" argues that agricultural intensification has deep impacts on both biodiversity and ecosystem properties and suggests that non-intensive farming practices keep the ecological balance and still may produce large quantities of high quality food (food security). This work presents a Qualitative Reasoning (QR) model that compares the impacts of intensive and non-intensive agriculture on water resources, biodiversity and productivity. The simulations show the inefficiency of intensive agriculture in protecting water resources and biodiversity, and the efficiency of non-intensive approach in terms of food production and ecosystem conservation.

Introduction

How to conserve biodiversity in a world with increasing food demand? Some authors suggest that by increasing the productivity of agricultural systems the demand of converting unfarmed areas into productive ones would decrease, leaving more space to conserve wildlife (Green *et al*. 2005; Balmford, Green and Schalermann 2005). However, Perfecto and Vandermeer (2005), among others, argue that ecological impacts of agriculture intensification can go far beyond farmed areas. The debate on whether agriculture intensification can or cannot prevent further biodiversity loss is now polarized between two opposite paradigms: "Land-Sparing", based on the idea that intensification could spare land for biodiversity conservation, and "Biodiversity-Friendly Farming" that suggests less intensive farming practices may combine food production and biodiversity conservation.

Land Sparing X Biodiversity Friendly Farming

Agriculture intensification is known to be one of the main causes of extinction all over the world (Benton, Vickery and Jeremy 2003). Despite these negative effects, the Green Revolution, an intensification process that since 1945 raised the world's gross yield in 106% and contributed to population growth and relative increase of well-being worldwide (Cassaman 1999). Defenders of the "Land Sparing" paradigm (Green *et al*. 2005) claim that productivity of existing farmed systems should increase in order to leave more space for conservation purposes. However, intensive agriculture may cause serious harm to native habitats in many ways. The use of pesticides can seriously threat non-target organisms, including human beings. Intensification also decreases agriculture matrix permeability by isolating populations living in natural habitat patches. Ecological theories (McArthur and Wilson 1967, Levins 1970) predict that no population or community can be maintained if it is not connected to others. Finally, as pointed out by the "Biodiversity-Friendly Farming", many studies show that less intense managed systems (eg. agroforests) can support high levels of biodiversity and yet have high productivity (Perfecto and Vandermeer 2008). In this context, the use of QR techniques (Weld and de Kleer 1990) may be useful to compare assumptions and consequences for the environment of these two approaches.

A model to express the relationships between farming and environment services

The model has been built following the Qualitative Process Theory (Forbus 1984) and the compositional modeling approach (Falkenhainer and Forbus 1991). Accordingly, processes are the initial cause of changes in the system, modeled by direct influences (I+ and I-) they put on state variables. Such changes may propagate to other quantities via qualitative proportionalities (P+ and P-). The model was implemented in the Garp3 workbench (Bredeweg *et al.* 2006) and consists of 53 model fragments involving 7 entities and 18 quantities. It holds, in the current version, 57 simulations. Entities and configurations are shown in Table 1.

Tabel 1: Source entity, configuration and target entity of the model

Source Entity	**Configuration**	**Target Entity**
Investor	invests in	Agriculture
Agriculture	occurs in	Farmed area
Farmed area	contains	Natural area
Farmed area	Has	Source
Farmed area	Has	Water resources
Source	affect	Unfarmed area
Emigration	Emit	Source
Agriculture	Uses	Water resources
Agriculture	impacts	Natural area

The model describes a landscape composed by many relatively small natural patches (natural area) and few large ones (sources), embedded in an agricultural matrix (the farmed area). It is known that the maintenance of species diversity in small natural areas depends on the colonization by individuals coming from a large area, the species source. Therefore the rate of species variation in an isolated natural area is the balance between colonization from external sources and extinction rates caused by the insular nature of small habitats.

The colonization process depends on the permeability of the farmed area. Permeability is defined in terms of physical and biological characteristics that facilitate or render the flux of propagules (fruits, seeds, larvae or individuals) through it. For instance, if an animal have to cross a large area of pasture (low permeability) before colonizing a forest fragment, it probably would suffer some harm before reaching its destination. In the model described here, permeability should be equal or greater than value medium as a condition for propagules to cause influence (P+) on species variation rate of unfarmed areas, as shown in Figure 1.

What happens if intensification takes place in a non-intensified landscape? Agriculture intensification main characteristics are mechanization, the use of artificial fertilizers and pesticides, irrigation and loss of spatial heterogeneity. Heterogeneity is considered here as the physical structure of the ecosystem. Intensified systems are characterized as homogenous (as they hold monocultures). Non-intensified systems have high spatial heterogeneity (vertical and horizontal) as they are composed by a mosaic of associations between different cultures. Water resources have fundamental importance for both the survival of natural ecosystems and the productive system. In this model, irrigation is a main factor that may impact water quantity in farmed areas and changes in water quality are determined by the quantity of fertilizers.

Productivity is influenced by the intensification parameters mentioned above and by ecological factors and by biodiversity, both in farmed and unfarmed areas. Biodiversity and environmental services are important for agricultural production such as they provide, among others, climate stability, water and nutrient cycling, pollination and protection against pest outbreak (Matson *et al.* 1997).

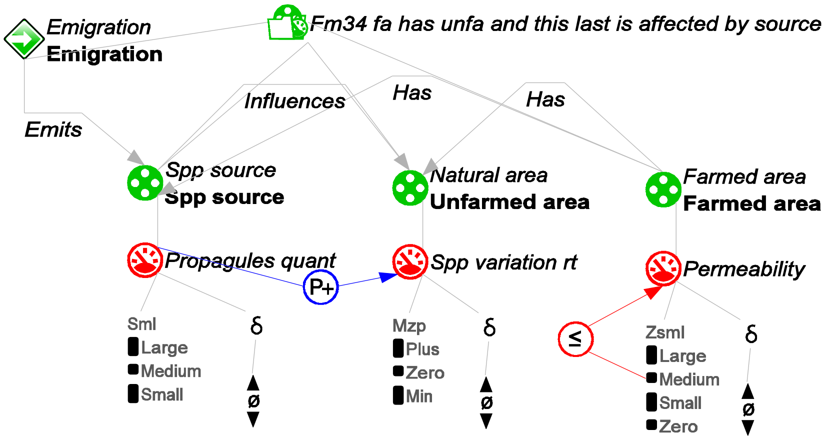


Figure 1: Model fragment showing that *Permeability* value should be equal or greater than medium for *Propagules* to influence *Species variation rate*.

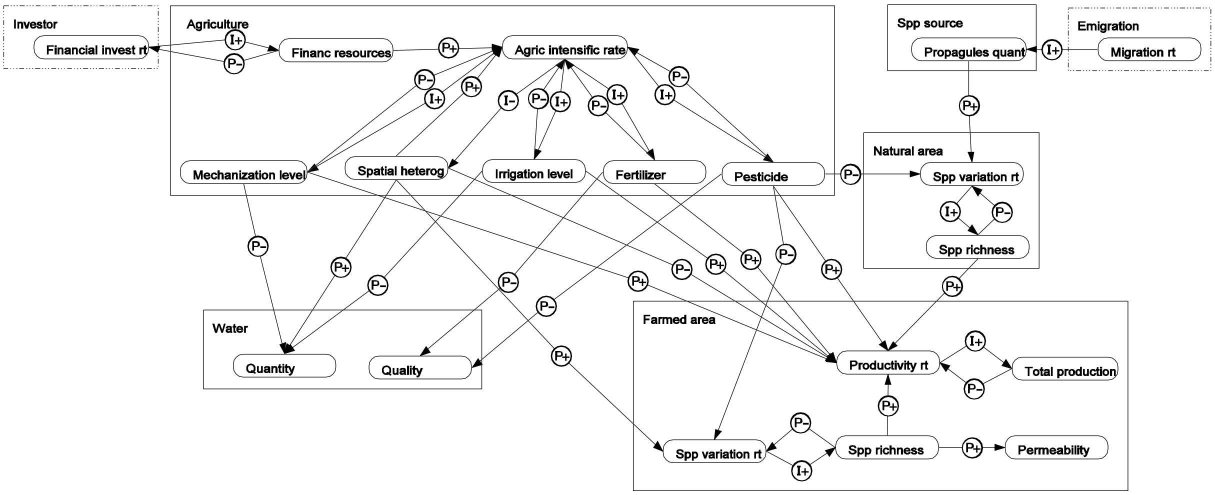


Figure 2: Causal model obtained in state 1 in simulations starting with both intensive and non-intensive agriculture scenarios.

Causal explanations for the effects of intensive and non-intensive agriculture

Intensive agriculture. The more complex simulation supported by the model starts with a scenario showing a landscape with non-intensified agriculture changed by the intensification process, which is triggered by investments on mechanization, fertilizers, pesticides and irrigation. Initially water quality, water quantity and spatial heterogeneity are in the highest values of their quantity spaces. Species richness in both farmed and natural areas have medium value, with propagules coming in large quantities into the natural area from a source area. Production is medium and stable.

The simulation produces one initial state, and the full simulation produces 43 states. The causal model obtained in state 1 (figure 2) reads as follows: a positive *Investment rate* causes *Financial resources* to increase and this change activates the intensification process (*Agriculture intensification rate* becomes positive). This process causes the quantities *Mechanization level*, *Irrigation level*, *Fertilizer* and *Pesticide* to increase, and *Spatial heterogeneity* to decrease. Influenced by these changes, *Water quantity* and *Water quality* decrease. *Species variation rate* in natural areas receives opposite influences from *Propagules quantity* and *Pesticides.* Considering that these two quantities are increasing, the result is ambiguous and *Species variation rate* may increase or decrease. This way *Species richness* in natural areas also may increase, stabilize or decrease. In farmed area, *Species variation rate* decreases due to the influences from *Spatial heterogeneity* and *Pesticides*, and as a consequence *Species richness* decreases. This change causes *Permeability* to decrease, making the propagule movement harder. *Productivity rate* in farmed area is influenced by *Species richness* both in natural and farmed areas, and by the five quantities affected by the intensification process. The final result is ambiguous, and the production may decrease, when the negative forces are greater than the positive ones, or increase, when environmental services provided by biodiversity have stronger influence on the farmed area.

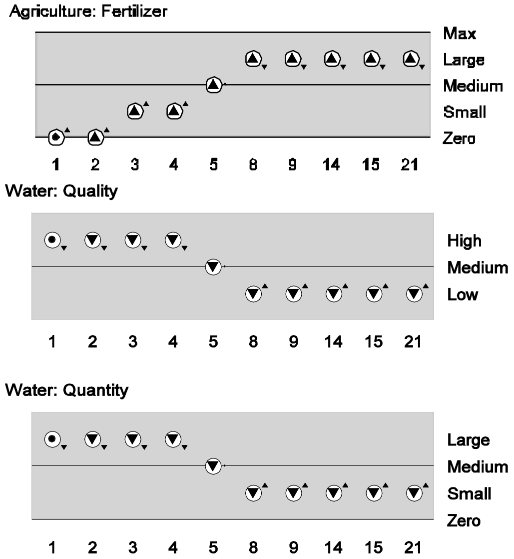


Figure 3. Value diagrams showing the effects of agriculture intensification on water quality and quantity.

The behaviour path [1 → 2 → 3 → 4 → 5 → 8 → 9 → 14 → 15 → 21] illustrates some of the consequences of intensive agriculture. *Fertilizers* and *Irrigation level* increase up to value large in state 8, and keep increasing within this interval until the end state 21, causing water quantity and quality to decrease (Figure 3).

Besides that the key for understanding the system behaviour can be found in the values *of Permeability*. As *Spatial heterogeneity* is decreasing, it eventually causes *Species richness* in farmed area to decrease, which in turn causes *Permeability* to decrease too. As soon as *Permeability* became smaller than medium in state 8, the influence from *Propagules* on *Species variation rate* is no longer active (see model fragment in Figure 1). The balance between the influences of *Propagules* and *Pesticides* on *Species variation rate* in the natural area was changing already and in state 8 the rate starts to decrease. As a consequence, *Species richness* in natural area starts to decrease in state 9 and eventually reaches the value small in state 21 (Figure 4).

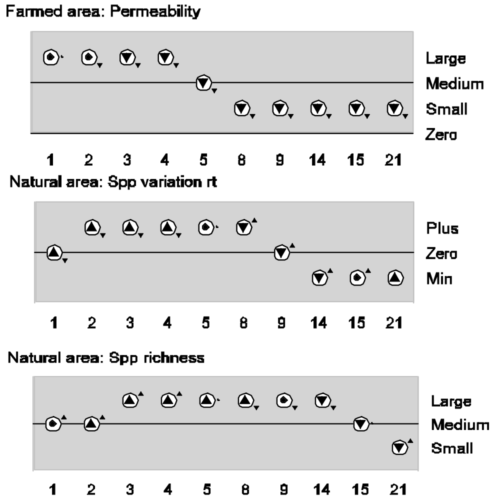


Figure 4. Value history diagrams of the quantities of permeability, species richness and species variation rate showing the effects of agriculture intensification.

The decline of productivity when ecosystem services collapse in intensive agriculture is shown in Figure 5. The *Productivity rate* is increasing until state 4. The opposite forces become equal in state 5, and the negative forces become stronger in state 8, causing the rate to decrease. *Total production* stabilizes in state 9 and decreases until the end of the simulation, when spatial heterogeneity, permeability, biodiversity, water quality and quantity, and production also have the lowest values.

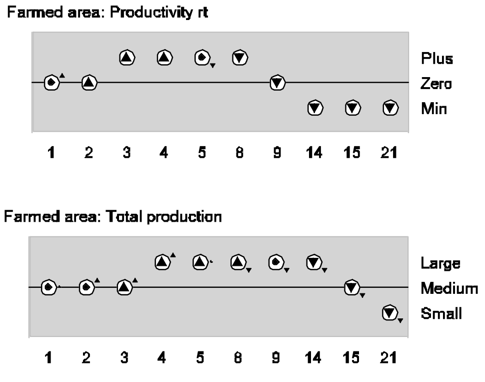


Figure 5. Value diagrams of the quantities showing the effects of agriculture intensification in the total production.

Non-intensive agriculture. In the initial scenario the rates of both processes, finance investment and agriculture intensification, are zero. As a consequence, the quantities that represent the main features of intensive agriculture have values zero too, as if they don't exist. *Spatial heterogeneity*, *Water quality* and *Water quantity* are also constant, at their maximum values. The other quantities have the same value as in the intensive agriculture simulation. The simulation produces one initial state and 6 states in total, being the causal model the same as the one shown in Figure 2. Water quality and quantity and spatial heterogeneity do not change during the simulation, and the biodiversity of both natural and farmed areas increase. Despite the low-input characteristics of this approach, total production increases and the environmental services are kept functioning. This pattern is known to happen empirically in sustainable agricultural systems (Perfecto and Vandermeer 2005).

Discussion and final remarks

There is a growing concern about the fact every day millions of people go to bed hungry. Apparently the dilemma is the following: shall we use what is left of natural land to produce food or to conserve biodiversity? From the point of view of the work described here, the question is conceptually wrong. There are alternative agricultural practices that can harbor high levels of biodiversity with satisfactory productivity (Vandermeer and Perfecto 2005). Also, agriculture intensification involves high ecological, social, cultural, public health and economic costs (Perfecto and Vandermeer 2008, Matson *et al.* 1997). The contribution brought by the model described here to the intensification *versus* conservation debate is to ground the simulation results on explicitly represented causal relations. The simulations showed the superiority of agro-ecological practices for both community (species) and ecosystems (environment services), keeping the productivity at a low cost.

References

Alexandros, N. 1999. World Food and Agriculture: outlook for the medium and longer term. *Proceedings of National Academy of Science*, 96:5908-5914.

Balmford, A.; Green, E. and Scharlemann, J.P.W. 2005. Sparing land for nature: exploring the potential impact of changes in agricultural yield on the area needed for crop production. *Global Change Biology*, 11: 1594–1605.

Benton, T.; Vickery, J. and Jeremy, D. 2003. Farmland biodiversity: is habitat heterogeneity the key? *Trends in Ecology and Evolution*, 18 (4).

Bredeweg, B.; Bouwer, A.; Jellema, J.; Bertels, D.; Linnebank, F. and Liem, J. 2006. Garp3 – A new Workbench for Qualitative Reasoning and Modelling. In Bailey-Kellog, C., Kuipers, B. (eds.) *Proceedings of the 20th International Workshop on Qualitative Reasoning*, pp. 21-28.

Cassman, K. 1999. Ecological intensification of cereal production systems: yield potential, soil quality, and precision agriculture. *Proceedings of National Academy Science*, 96: 5952–5959.

Falkenhainer, B. and Forbus, K. 1991. Compositional modeling- finding the right model for the job. *Artificial Intelligence*, 51, 95-143.

Forbus, K. 1984. Qualitative process theory. *Artificial Intelligence*, 24, 85-168.

Green, R.; Cornell, S.; Schalermann, J. and Balmford, A. 2005. Farming and the fate of wild nature. *Science*, 307: 550-555.

Levins, R. 1970. Extinction. *In* Gesternhaber, M. (ed.) *Some Mathematical Problems in Biology*. Providence, Rhode Island, American Mathematical Society, pp. 77–107.

MacArthur, R.H. and Wilson, E.O. 1967. *The theory of island biogeography*. Princeton: Princeton University Press.

Matson, P.; Parton, W.; Power, A. and Swift, M. 1997. Agricultural intensification and ecosystem properties. *Science*, 277: 504–509.

Perfecto, I. and Vandermeer, J. 2008. Biodiversity Conservation in Tropical Agroecosystems A New Conservation Paradigm. *New York Academy of Sciences*, 1134: 173–200.

Vandermeer, J. and Perfecto, I. 2005. The future of farming and conservation. *Science*, 308(308): 1257.

Weld, D. and de Kleer, J. 1990. *Readings in Qualitative Reasoning about Physical Systems*, 1–8. San Francisco, Calif.: Morgan Kaufmann.

Intelligent Authoring of 'Graph of Microworlds' for Adaptive Learning with Microworlds

Tomoya Horiguchi
Graduate School of Maritime Sciences,
Kobe University
5-1-1, Fukaeminami, Higashinada, Kobe,
Hyogo, 658-0022 Japan
horiguti@maritime.kobe-u.ac.jp

Tsukasa Hirashima
Department of Information Engineering,
Hiroshima University
1-4-1, Kagamiyama, Higashihiroshima,
Hiroshima 739-8527 Japan
tsukasa@isl.hiroshima-u.ac.jp

Abstract

In science education, it is important to sequence a set of *microworlds* (which means *a system and its model* limited from educational viewpoint) of various complexity adaptively to the context of learning. We previously proposed *Graph of Microworlds* (GMW), a framework for indexing a set of microworlds based on their models. By using GMW, it is possible to adaptively select the microworld a student should learn next, and to assist him in transferring between microworlds. However, it isn't easy to describe GMW because an author must have the expertise in the process of modeling. In this research, we propose a method for semi-automating the description of GMW by introducing the *compositional modeling* mechanism. Our method assists an author in generating a set of indexed microworlds and also in considering educational meanings of the relations between them. We present how to design such a function and also illustrate how it works. A preliminary test with a prototype system showed the effectiveness of our method.

Introduction

In physics education, it is important for a student to acquire the ability to make appropriate models of various phenomena in the domain. For this purpose, a set of problems are provided in which he/she must think about some physical systems and their behaviors. In each problem, the range of systems and their behaviors are usually limited from some educational viewpoint in order for him/her to be able to understand the laws/principles behind the phenomena. This is called a *microworld* [1]. For the systematic understanding of the domain theory, therefore, it is necessary to sequence a set of microworlds of various complexity (from relatively simple systems/phenomena to more complicated ones) adaptively to the context of learning.

In designing ITSs (Intelligent Tutoring Systems) with such a function, it is essential to appropriately index a set of microworlds. Especially, it is important to explain why, in the situation given by a microworld, the laws/principles are applicable and why the model is valid. It is also important to explain why/how the model changes if the situation is changed. In order to make such explanations, it is necessary to index a set of microworlds based on their models and the process of modeling.

Therefore, we proposed a *Graph of Microworlds* (GMW), which is a framework for indexing the microworlds and the relations between them based on their models and the process of modeling (Horiguchi and Hirashima, 2005). We also showed, by using GMW, it becomes possible to design a function for adaptively selecting the microworld which a student should learn next, and a function for assisting a student in transferring between microworlds. However, it isn't easy to describe a GMW because an author must make a lot of indices in a model-based way. He/She must have the expertise in the process of modeling. In this research, therefore, we propose a method for semi-automating the description of GMW by introducing an automatic modeling mechanism (i.e., *compositional modeling* (Falkenhainer and Forbus, 1991; Rickel and Porter, 1994; Levy et al., 1997)).

Adaptive Learning Support with GMW

An example of GMW for elementary mechanics is shown in Fig. 1. Each microworld is indexed with the situation it deals with, the model of the situation and the process of modeling. A student can learn the physical law(s)/principle(s) necessary for the modeling and the skill(s) for the model-based problem solving in each microworld (they are called a *learning item*). Two microworlds which deal with similar situations but different models (i.e., different law(s)/principle(s) is(are) necessary) are linked to each other with an edge. *Parameter-change rules* (Addanki et al., 1991) are attached to such an edge which relate the difference between the situations of two microworlds to the difference between the behaviors of their models. This means one model is the necessary evolution of the other (with the perturbation of situation). Such a relation between two microworlds is called an *educationally meaningful* relation. In order to make a student learn the domain theory progressively, a GMW should include as many such relations as possible.

In Fig. 1, when a student learned linear uniform motion in MW-1, MW-2 and MW-4 are identified as the candidates he should learn next because they are adjacent to MW-1. Additionally, for assisting a student in transferring from MW-1

---

[1] Though this term usually indicates simulation-based interactive learning environments, we, in this paper, use it for indicating a system and its model made by limiting its structure and behavior from some (educational) viewpoint.

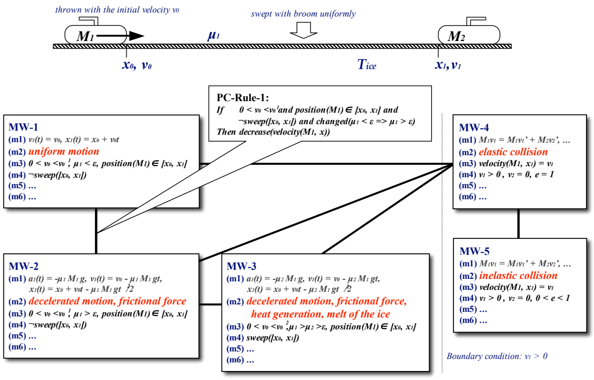


Figure 1: An example of Graph of Microworlds

to MW-2, a task is generated by using the parameter change rule, such as: *derive the velocity of M_1 when the value of* μ_1 *becomes greater and the friction becomes not negligible*. In this task, the necessity of the model of MW-2 is strongly suggested because the difference between the velocities of M_1 before/after the change of μ_1 can't be explained only by the model of MW-1.

In GMW, the situations and the differences between them are represented with a set of *modeling assumptions* (Falkenhainer and Forbus, 1991; Rickel and Porter, 1994; Levy et al., 1997) which constrain the viewpoint in modeling the system, the behavioral range of the system to be considered. Modeling assumptions represent the conditions about the system's structure and its state under which the model is valid. They are, however, not merely the applicable conditions of laws/principles, but the conceptualization of decision making in modeling the system. Therefore, an instance of a modeling assumption usually has its alternative(s). They are exclusive, and the model based on the latter is qualitatively different from the one based on the former. Modeling assumptions, therefore, can be a useful conceptual tool for describing the qualitative differences between various models.

Method for Assisting Authors in Describing GMW

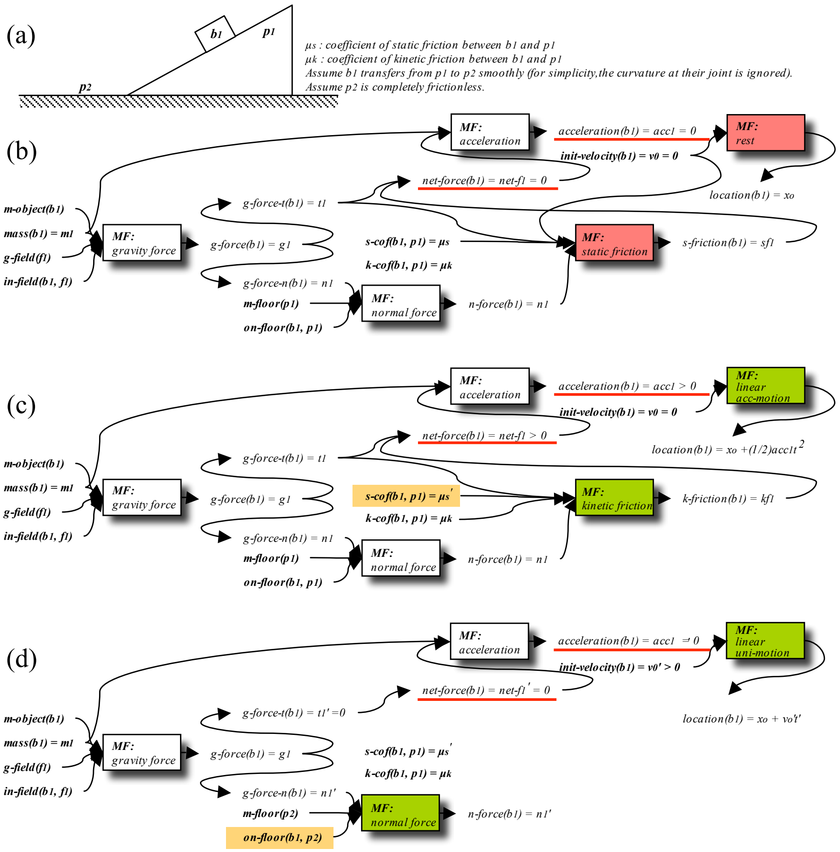
It is not easy for (non-programmer) authors to describe a GMW. First, (1) it needs the expertise in the process of modeling to index the models with their modeling assumptions, especially because modeling assumptions are usually implicit information in models. Second, (2) it is difficult to find the various situations which embodies the law(s)/principle(s) covering the given set of learning items of the domain because its search space becomes vast. Lastly, (3) the set of microworlds must have as many educationally meaningful relations between them as possible.

We, therefore, propose a method for assisting an author in describing GMW by a *generation-test method*, in which he/she semi-automatically generates the models of various situations one after another, and judges whether each of them is appropriate to the GMW from an educational viewpoint. By using *compositional modeling* mechanism (Falkenhainer and Forbus, 1991; Rickel and Porter, 1994; Levy et al., 1997), this method is implemented as follows: First, (1) an author finds *a* situation which embodies *a* learning item (i.e., law(s)/principle(s)). The compositional modeler automatically generates the model and indexes it by its modeling assumptions. Second, (2) he/she perturbs this situation. The compositional modeler automatically generates the model of this new situation and indexes it by its modeling assumptions. Third, (3) if the new model embodies *another* learning item which is appropriate as a neighbor of the former learning item, he/she decides whether it is added to the GMW or not. If he/she judges that the difference between these two models is educationally meaningful, he/she adds the new one and the new edge between them. (4) By repeating (2) and (3) to grow the GMW, the author would finally get the whole GMW which embodies the set of learning items to be covered.

In this procedure, the work an author should do is to identify the relation between two models based on the perturbation of situation (i.e., the difference of modeling assumptions) and to judge whether it is educationally meaningful or not. In order to assist him/her, therefore, the function is desirable which makes advice on what physical meaning a difference of modeling assumptions has. In the next section, therefore, we describe a method for designing such a function by classifying the modeling assumptions based on their physical meanings and by grouping the exclusive ones which can't be made simultaneously (Horiguchi and Hirashima, 2008).

Relations between Models based on the Difference of Modeling Assumptions

We classify the modeling assumptions made in modeling physical systems into *constraints of physical structure* (CPS) and *constraints of operating range* (COR). Constraint of physical structure (CPS) is the assumption which specifies what kind of objects, relations and their attributes in a physical system are considered. CPS represents the decisions about perspectives and granularity. On the other hand, physical phenomena occur assuming a physical system is



μs : coefficient of static friction between b1 and p1
μk : coefficient of kinetic friction between b1 and p1
Assume b1 transfers from p1 to p2 smoothly (for simplicity,the curvature at their joint is ignored).
Assume p2 is completely frictionless.

Figure 2: An example of difference between models

in a specific state. When the state changes, the model may become invalid. Therefore, a model must have the specification about the range (in its state space) within which it is valid. It is called constraint of operating range (COR).

In each type of these modeling assumptions, there are usually the sets of exclusive ones which can't be made simultaneously. For example, in a physical system, it isn't allowed to make assumptions *consider friction between two blocks* and *not consider friction between them* simultaneously as CPS. Therefore, by grouping the model fragments each of which has exclusive modeling assumption(s), it is possible to design the function for suggesting the relation between the models in two microworlds before and after the perturbation of situation. That is, first, the two sets of model fragments are compared, each of which composes each model. Then, if a pair of model fragments each of which belongs to each model and matches the same/similar partial situation has exclusive modeling assumption(s), the relation between the models is inferred from the type (i.e., physical meaning) of the assumption. Referring the relations between two models thus enumerated by the system, an author identifies the most appropriate relation and judges its educational meaning.

[Example-1] Fig. 2a shows the physical system in which an object b_1 is put on an inclined plane p_1 (to which a horizontal plane p_2 is connected). Fig. 2b is a model (i.e., a set of instantiated model fragments) of a situation of this system in which b_1 remains at rest on p_1 because the tangential component of b_1's gravity on p_1 is smaller than the maximum static friction between b_1 and p_1. It (called model-1) consists of 5 model fragments, including *static friction* and *rest*. If the coefficient of static friction is decreased in this situation, another situation may occur in which b_1 moves downward accelerated by its gravity (and the kinetic friction). The model of this situation (called model-2) is Fig. 2c and it consists of 5 model fragments, including *kinetic friction* and *linear acc-motion*.

The model fragments *static friction* in model-1 and *kinetic friction* in model-2 correspond to each other because they are instantiated by matching with the same physical structure in these models. Their CORs are exclusively different only in the modeling assumption which constrains the range of the value of the coefficient of static friction. It is, therefore, inferred that there is a difference between these models in 'the change from static friction to kinetic friction because of the change in the value of the coefficient of static friction.' The model fragments rest in model-1 and linear acc-motion in model-2 also correspond to each other because of the same reason. Their CORs are exclusively different only

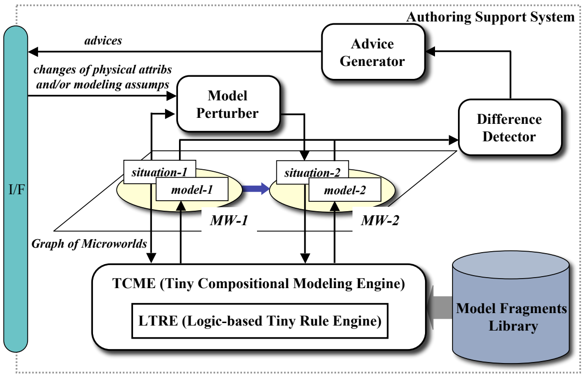


Figure 3: Architechture of the system

```
(defModelFragment (static-friction ?blk ?flr ?s-cof)
    :Individuals
    ((?blk :conditions (m-block ?blk))
     (?flr :conditions (m-floor ?flr))
     (?s-cof :conditions (static-cof ?blk ?flr ?s-cof)))
    :Assumptions
    ((on-floor ?blk ?flr)
     (applied-force ?blk (normal-force ?blk ?flr)))
    :Conditions
    ((= (v-mag (velocity ?blk)) 0.0)
     (< (mag (net-force ?blk))
        (* (static-cof ?blk ?flr)
           (mag (normal-force ?blk ?flr)))))
    :Relations
    ((Quantity ?self)
     (= (v-mag ?self) (mag (net-force ?blk)))
     (= (v-dir ?self) (+ (dir (net-force ?blk)) 180))
     (applied-force ?blk ?self)))
```

Figure 4: An example of model fragment

in the modeling assumption which constrains the range of the value of b_1's acceleration. It is, therefore, inferred that there is a difference between these models in 'the change from rest to linear accelerated motion because of the change in the value of b_1's acceleration.'

Design of a Prototype System

We developed a prototype system for GMW-authoring with our method. Note that it currently implements only basic functions: *situation interpreter/perturber*, *compositional modeler* and *difference detector*, except for (GUI-based) *user interface*. The architecture of the system is shown in Fig. 3.

Compositional modeler (we call this implementation TCME: Tiny Compositional Modeling Engine) generates the model of a given situation (i.e., a set of modeling assumptions) by applying the domain knowledge (i.e., the library of model fragments) to it. In the library of model fragments, model fragments written in the form shown in Fig. 4 are stored. They are translated into a set of *clauses* and used for the inference in LTRE. LTRE, which is a Logic-based Truth maintenance system (LTMS) coupled to a forward-chaining Rule Engine (Forbus and deKleer, 1993), maintains the dependency network of constraints of the generated model and guarantee the consistency of it.

Situation interpreter/perturber translates a given set of physical attributes and their values into modeling assumptions which are used for the inference in TCME (e.g., quantitative representation of relative position of mechanical objects are translated into its qualitative ones). If the value(s) of physical attribute(s) of a situation is(are) changed, a set of modeling assumptions of the new situation is output. An author perturbs the situation of a model by changing the value(s) of its physical attribute(s) or by changing its modeling assumptions directly to make a new model.

Difference detector detects and enumerates the differences between two given models (which are generated by TCME) with the method explained in the previous section. The differences are shown to an author with the explanations of why they appeared by the advice generator.

We developed a set of model fragments for TCME and the rules for situation interpreter/perturber which cover the basic examples of elementary mechanics. In a preliminary test, the prototype system could output the differences of models in several examples correctly. For example, in Fig. 2, when the coefficient of static friction in model-1 was decreased, model-2 was generated and the differences explained in Example-1 were output. When the friction between b_1 and p_1 was neglected in model-2 (its modeling assumption *Consider(friction(*b_1, p_1*))* was directly changed), the model of the new situation was generated (called model-3) in which b_1 moves downward accelerated by only its gravity without kinetic friction. As for the differences between model-2 and model-3, 'the disappearance of kinetic friction because of the neglection of friction (specialization of the model)' was output. Additionaly, when the time variable was increased in model-2, the model of the new situation was generated (called model-4) in which b_1 moves on p_2 at a constant velocity. As for the differences between model-2 and model-4, 'the change of relative position among b_1, p_1 and p_2 because of the evolution of time,' 'the disappearance of kinetic friction and normal force between b_1 and p_1, and the appearance of normal force between b_1 and p_2 because of the change of relative position among b_1, p_1 and p_2' and 'the change from linear accelerated motion to linear uniform motion because of the change of the value of b_1's acceleration' were correctly output.

Conclusion

In this paper, we presented a method for assisting an author in indexing a set of microworlds based on their models. Currently, it has been tested with only very small prototype. It is necessary to scale up our method by elaborating the classification of modeling assumptions for developing the larger library of model fragments.

References

Addanki, S., Cremonini, R. and Penberthy, J.S. 1991. Graphs of models. *Artificial Intelligence* 51: 145–177.

Falkenhainer, B. and Forbus, K.D. 1991. Compositional Modeling: Finding the Right Model for the Job. *Artificial Intelligence* 51: 95–143.

Forbus, K.D. and deKleer, J. 1993. *Building Problem Solvers.* MIT Press.

Horiguchi, T. and Hirashima, T. 2008. Intelligent Support for Authoring 'Graph of Microworlds' based on Compositional Modeling Technique. In Proceedings of QR2008, 49–57.

Horiguchi, T. and Hirashima, T. 2005. Graph of Microworlds: A Framework for Assisting Progressive Knowledge Acquisition in Simulation-based Learning Environments. In Proceedings of AIED2005, 670–677.

Levy, A.Y., Iwasaki, Y. and Fikes, R. 1997. Automated model selection for simulation based on relevance reasoning. *Artificial Intelligence* 96: 351–394.

Rickel, J. and Porter, B. 1994. Automated Modeling for Answering Prediction Questions: Selecting the Time Scale and System Boundary. In Proceedings of AAAI-94, 1191–1198.

Automated detection of electrocardiographic diagnostic features through an interplay between Spatial Aggregation and Computational Geometry

Liliana Ironi and **Stefania Tentoni**
IMATI - CNR
via Ferrata 1, 27100 Pavia, Italy
{ironi,tentoni}@imati.cnr.it

Abstract

Within the medical domain, Functional Imaging provides methods for effectual visualization of diagnostically relevant numeric fields, i.e. of spatially referenced measurements of variables related to organ functions. Unveiling the salient physical events that underly a functional image is most appropriately addressed by feature extraction methods that exploit the domain-specific knowledge combined with spatial relations at multiple abstraction levels and scales. The identification of specific patterns that are known to characterize classes of pathologies provides an important support to the diagnosis of disturbances, and the assessment of organ functions. In this work we focus on Electrocardiographic diagnosis based on epicardial activation fields. This kind of data, which can now be obtained non invasively from body surface data through mathematical model-based reconstruction methods, can hit electrical conduction pathologies that routine surface ECGs may miss. However, their analysis/interpretation still requires highly specialized skills that belong to few experts. Given an epicardial activation field, the automated detection of salient patterns in it, grounded on the existing interpretation rationale, would represent a major contribution towards the clinical use of such valuable tools whose diagnostic potential is still largely unexplored. We focus on epicardial activation isochronal maps, which convey information about the heart electric function in terms of the depolarization wavefront kinematics. An approach grounded on the integration of a Spatial Aggregation (SA) method with concepts borrowed from Computational Geometry provides a computational framework to extract, from the given activation data, a few basic features that characterize the wavefront propagation, as well as a more specific set of diagnostic features that identify an important class of heart rhythm pathologies, namely reentry arrhythmias due to block of conduction.

Keywords: Biomedical imaging; functional imaging; image based diagnosis; spatial aggregation; computational geometry; electrocardiography; cardiac electrical function.

Introduction

One of the most important application domains where imaging has proved extremely useful is Medical Diagnosis. The process of identifying a pathological condition can be greatly supported by signs of deviations from normality that can be drawn from images. Within this context the term “imaging” usually refers to techniques to build images of anatomical districts of the human body (e.g. radiographies, CAT, NMR); more broadly, it can also include methods aimed at providing graphical representations of temporally/spatially referenced measurements of variables related to specific organ functions (e.g. EEG, ECG signals, activation maps). In this latter case the term “functional” imaging is to be preferred.

Many functional images are graphical representations of a physical field: a potential contour map, for instance, is the spatial representation of a potential field. Thereby, the task of analyzing such images is not adequately tackled by traditional Image Processing methods, which have been designed for raster images. The issue of unveiling the salient physical events underlying a functional image is more appropriately and effectively addressed through feature extraction methods that can exploit the domain-specific knowledge at different abstraction levels. Such an issue is particularly relevant in view of performing explanation and reasoning tasks.

Within the field of Qualitative Spatial Reasoning, Spatial Aggregation (SA) (Yip and Zhao 1996) provides the most appropriate conceptual framework for feature extraction at multiple levels, according to a powerful hierarchical abstraction strategy. In the direction of making the approach more robust and of integrating within the basic SA framework methods borrowed from quantitative research fields, several works have contributed to make it an even more attractive framework for the development of functional imaging tools (Ironi and Tentoni 2003a; 2003b). Any such tool would ground on domain-specific knowledge, as the inference mechanisms rely on a network of relations that, besides dealing with spatial properties, explicitly encode such knowledge.

With the present work we continue our research effort aimed at delivering novel tools to support the assessment of the electric cardiac function (Ironi and Tentoni 2007). Diagnosing the cardiac electric function has always been a hard task for the difficulty met in the identification of salient electrical events and their spatial association with specific epicardial sites. In the clinical context, diagnosis of conduction pathologies is still carried out on the ECG signals for which the interpretative rationale is well-established. Several tools exist for automated ECG segmentation and classification. Most of these tools are based on the integration of wavelet transforms with neural/fuzzy-neural networks, to deal respectively with the signal decomposition and classification tasks (see for example (Clifford, Azuaje, and McSharry 2006)). Within AI, Qualitative Reasoning has also played an important role in providing a number of automated ECG interpretation tools (Bratko, Mozetic, and Lavrac 1989; Weng et al. 2001; Kundu, Nasipuri, and Basu 1998). Unfortunately some important rhythm disturbances may be incorrectly located or even missed by routine ECGs. Even body surface high resolution mapping may fail because signs of cardiac electrical events on the torso surface are weak.

In recent years, model-based numerical inverse procedures have made it possible to obtain non-invasively the epicardial activation field from body surface data. That has engaged researchers in the effort towards novel methods for electrocardiographic imaging (Oster et al. 1997; Ramanathan et al. 2004). However, the interpretative rationale for cardiac maps is only partially defined, and the ability to abstract the most salient visual features from a map and relate them to the complex underlying phenomena still belongs to few experts. Due to the extreme complexity of the physical system the task of automating diagnosis of conduction disturbances from a 2D/3D activation field is therefore hard, and necessarily limited to the current interpretation rationale. Within this field functional image-based diagnosis is at its beginning, and, in accordance with the available rationale, currently regards only a few classes of conduction disturbances. The potential of Qualitative Spatial Reasoning in contributing to its development is high: a tool for the automated extraction of spatially referenced features of the cardiac electrical function would bridge the gap between established research outcomes and clinical practice.

Our work fits into a long-term research project aimed at delivering an automated electrocardiac map interpretation tool. To detect salient spatiotemporal features in the epicardial activation field, we exploit the inference mechanisms provided by a computational tool grounded on Spatial Aggregation and on Computational Geometry concepts: from a given numeric field we extract spatial objects that, at different abstraction levels, qualitatively characterize spatiotemporal phenomena, and discover and abstract the skeleton of patterns relevant to the diagnostic task.

In this paper we focus on epicardial activation maps, which convey information about the heart electric function in terms of the depolarization wavefront kinematics. These kind of maps are very useful to diagnose rhythm disturbances. We describe how some basic spatiotemporal features that characterize the propagation of the electrical excitation can be abstracted from the given activation data, and in particular we define a set of distinctive features that identify an important class of rhythm disturbances due to blocks of conduction.

Feature abstraction from a numeric field

The comprehension of physical phenomena benefits from the visualization of the spatial course of relevant variables. A visual representation obtained from a given numeric field can be further inspected, and searched for homogeneities and specific patterns that have a physical meaning. This "imagistic" reasoning activity, that goes beyond mere visualization, is performed at multiple levels through a sequence of abstractions and manipulations of spatial objects that capture key physical properties.

Spatial Aggregation

Spatial Aggregation (SA) is a general-purpose framework that provides a suitable ground to capture spatiotemporal adjacencies at multiple scales in spatially distributed data. It was designed to derive and manipulate qualitative spatial representations that abstract important features of the underlying data, for their use in automated reasoning tasks (Yip and Zhao 1996; Ironi and Tentoni 2003a; 2003b).

In outline, SA transforms a numeric input field into a multi-layered symbolic description of the structure and behavior of the physical variables associated with it. This results from iterating transformations of lower-level objects into more abstract ones through the exploitation of qualitative equivalence properties shared by neighbor objects.

SA abstraction mechanisms are based on three main steps, namely *Aggregation*, *Classification*, and *Redescription*, that exploit domain-specific knowledge and spatial adjacencies (see Fig.1):

1. *Aggregation.* Spatial adjacency of low-level objects is encoded within a neighborhood graph.
2. *Classification.* Neighbor objects are grouped by similarity, according to a domain-specific equivalence predicate that defines a feature of interest.
3. *Redescription.* Similarity classes are singled out as new high-level objects that provide an abstract representation of the feature.

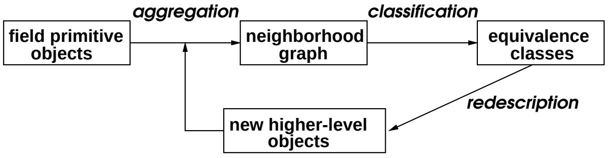


Figure 1: Basic inference steps in Spatial Aggregation.

Step 1 mostly exploits geometrical properties, either metrical or topological. However, to ensure the robustness of the *Classification* step, it can also take into account available non-geometrical knowledge, associated with the objects to be aggregated and related to the physical context (Ironi and Tentoni 2003a). For example, with respect to the contouring task, the values of the variable associated with each field point can be exploited to generate an appropriate neighborhood graph that guarantees the abstraction outcome against both artifactual curve entanglement and segmentation (Ironi and Tentoni 2003a). Step 3 is crucial in that a non-effectual redescription of new objects may jeopardize subsequent abstractions stemming therefrom.

Such operators can be iterated over and over until the behavioral and structural information about the underlying physical phenomenon, that is required to perform a specific task, is extracted from the data set. The hierarchical structure of the whole set of the so-built objects defines a bi-directional mapping between higher and lower-level aggregates, and, consequently, it facilitates the identification of the pieces of information relevant for a specific task.

The role of Computational Geometry

Within the SA abstraction mechanism, *Redescription* instantiates visual features that may play a role in the spatial reasoning process. The geometric representation of the new objects must convey a meaningful effectual visual synthesis of the underlying similarity class.

Computational Geometry methods and concepts can play an important role in providing algorithms for the redescription of newly abstracted objects. An important class of objects for which the chosen representation format particularly needs to suit the reasoning task is that of 2D bounded regions. These latter ones can result, for example, from the application of a similarity relation grounded on interval values to a set of contiguous isopoints. The similarity classes correspond to regions that need to be instantiated as new geometrical objects for further treatment. Sometimes a centroid can do, while in other situations a more articulated - though compact - redescription may be necessary, for example when the qualitative topological structure of the region needs to be captured at multiple scales. The choice of the most appropriate format and scale for the redescribed object is task-driven.

For qualitative reasoning tasks, a region descriptor should be:

i) robust and stable with respect to noise and small perturbations of the boundary,

ii) capable to roughly capture the location and global extent of the region,

iii) capable to capture the topological structure of the region at an appropriate scale of details with respect to the task, and of course

iv) computationally feasible.

An effectual representation of a region can be provided by its "gross skeleton", as defined in the following.

Gross skeleton of a region

The concept of "gross skeleton", that we are going to introduce, is derived from the "medial axis" of a bounded region. The medial axis is geometrically defined as the locus of the centers of circles that are internally tangent to the region's boundary. That results in a set of curves roughly running along the middle of the region. Unfortunately, the medial axis is very sensitive to small perturbations of the boundary: noisy contours produce a number of secondary branches.

The medial axis can be thought of as a geometric skeleton of the figure, whose complexity, given by the number of branches, corresponds to the boundary complexity, defined as the number of its curvature extrema. For its instability the medial axis is not immediately suitable as a figure descriptor in contexts affected by noise, and as such it is also inappropriate where finer scale details are irrelevant and to be ignored.

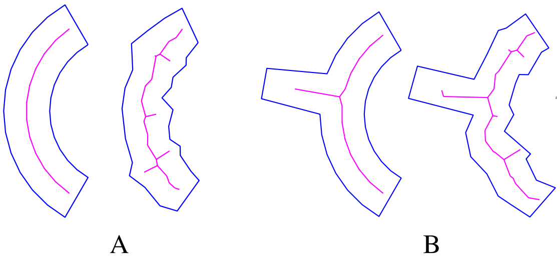


Figure 2: Panels A,B refer to two sample regions. For each panel, the Voronoi based medial axis (magenta) is shown when the region's boundary is relatively smooth (left), and when it is affected by noise (right)

Exact computation of the medial axis is difficult in general. It is well-known that an approximation of the medial axis of a region can be obtained from the Voronoi diagram related to a finite set of points that sample the region's boundary (Brandt and Algazi 1992): it consists in the subset of Voronoi edges that lie completely within the region's interior. Such approximation, though, still suffers from the cited instability problem as it is illustrated in Fig.2.

The following algorithm builds a robust simplified topological skeleton of a given polygonal region, namely the gross skeleton, by exploiting a relevance measure (Sakai and Sugihara 2006), and selectively pruning the approximated Voronoi medial axis. Instability is removed by dropping spurious/irrelevant branches that correspond to unneeded information about finer contour's details.

Algorithm (*gross skeleton construction*).

Given $\{P_1, ..P_n\}$, vertices of a closed polyline bounding a connected region $\mathcal{L}$,

1. Compute $\mathcal{M}$, Voronoi approximation of the medial axis of $\mathcal{L}$, as follows:
   (a) Build the Voronoi diagram related to the set of vertices $\{P_1, ..P_n\}$,
   (b) Retain only the edges that are completely internal to $\mathcal{L}$.
2. Compute the "index of relevance" $\beta(E)$ of each edge $E \in \mathcal{M}$, as
$$\beta : \mathcal{M} \rightarrow (0,1) \qquad \beta(E) = 2|l|/|\partial\mathcal{L}|$$
   where if P_i, P_k are the generators of Voronoi edge E, $|l|$ is the length of shortest path connecting P_i with P_k along the region's boundary $\partial\mathcal{L}$, and $|\partial\mathcal{L}|$ is the length of the regions's boundary (Fig.3).
3. (*Selective pruning*) Starting with $\mathcal{L}^* = \mathcal{M}$,
$$\forall E \in \mathcal{M}, \ \beta(E) < \beta^* \ \Rightarrow \ \mathcal{L}^* := \mathcal{L}^* \backslash \{E\}$$
   where β^* is a given relevance threshold.

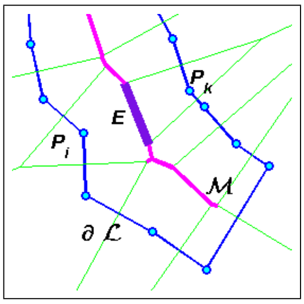


Figure 3: Steps in the construction of the gross skeleton of a polygonal region. Vertices P_i, P_k of the region's boundary (partially shown, blue line) generate Voronoi edge E (thicker line). Part of the Voronoi tessellation (green thin lines), and of the approximated medial axis $\mathcal{M}$ (magenta thick line) are also shown.

Selective pruning of the medial axis $\mathcal{M}$ is performed according to an edge relevance criterion by which irrelevant boundary details are dropped: edges with a very low β value have a negligible effect on the region's boundary. The result

$$\mathcal{L}^* = \{E \in \mathcal{M} \mid \beta(E) \geq \beta^*\}$$

is a connected linear structure that reflects the global topological structure of the region, as well as its rough location and spatial extent.

The choice of the relevance threshold β^* affects the complexity of the resulting gross skeleton $\mathcal{L}^*$, and adjusts the descriptor to the scale required by the reasoning task: as greater β^* is, as more simplification is required.

In Fig. 4 a series of perturbations of the smooth sample region shown in Fig. 2A are reported: in each case both the Voronoi medial axis and the gross skeleton are computed. The figure clearly shows how more robust the gross skeleton is with respect to the Voronoi medial axis approximation, and how the global shape of the region is captured.

Functional Imaging of the cardiac electric function

The heart is site of cyclic electrical activity which causes the muscle to rhythmically contract. The propagation of the electric excitation within the myocardium is a quite complex 4D spatiotemporal process that electrocardiologists explore on reference surfaces (epicardial, endocardial) by means of relevant variables, such as the electric potential, the activation time and the wavefront propagation velocity. Due to the difficulty of combining spatial and temporal aspects, exploring the potential $u(\mathbf{x}, t)$, a function of space and time, is a hard task. A more global and synthetic view on the spatiotemporal process of excitation is provided by the epicardial representation of the *activation time* $\tau(\mathbf{x})$, defined as the instant at which an epicardial site $\mathbf{x}$ changes its electric state from resting to activated. Such an instant is commonly estimated as the point of minimum (time) derivative extracted from the electrogram $t \to u(\mathbf{x}, t)$. Therefore, the activation time is a sort of landmark variable which embeds a qualitatively significant event in the electric potential time course, and, when spatially represented on the whole epicardial surface, it holds a powerful diagnostic potential.

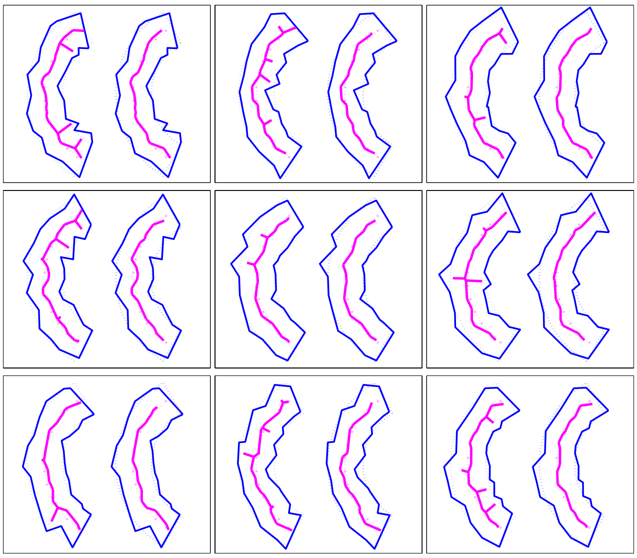


Figure 4: Each panel refers to a distinct perturbation of the smooth region shown in Figure 2A. In each panel: the Voronoi based medial axis (left), and the gross skeleton obtained by pruning with $\beta^* = 0.25$ (right).

In imaging of the cardiac electric function, an important role is played by activation maps: such maps are contour maps of the activation time that convey information about the wavefront structure and propagation. In a previous work (Ironi and Tentoni 2007), in accordance with the existing rationale of interpretation, we tackled the problem of defining and abstracting, within the SA framework, a set of spatial objects that capture a few important basic features of activation: isochrones, whose spatial sequence depicts the spread of excitation by snapshots, wavefront breakthrough and exit locations, fast propagation pathways.

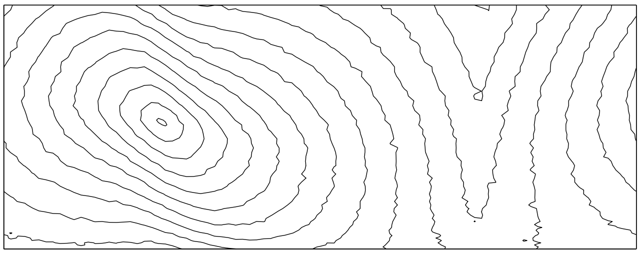


Figure 5: Activation map as obtained from noisy data.

As an example, Fig. 5 shows an activation map obtained from noisy simulated data related to a case of normal propagation elicited by single site pacing. Let us remark that the activation time field is actually related to a 3D model of the epicardium; in order to have a unique global planar

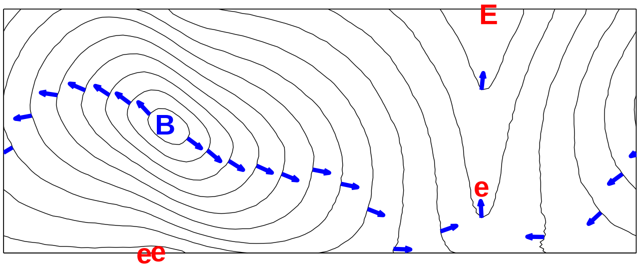


Figure 6: Main wavefront propagation features abstracted from the sample data of Fig. 5: activation isochrones, breakthrough (B) and exit sites (e/E), and fast propagation pathways (thick vectors).

view with minimal spatial distortion, we considered an axial cylindrical projection of this map. After preliminary noise removal, from the activation field the main wavefront propagation features are detected: the sequence of isochrones, the breakthrough and extinction sites, which respectively mark where excitation starts and ends on the epicardial surface, and the fast propagation pathways (Fig. 6).

In this work we focus on an important class of pathological conditions, namely *reentry ventricular tachycardia* (VT), and provide SA-based definitions and algorithms for the abstraction and spatial redescription of the features involved. Reentry VT usually follows a myocardial infarction, as the presence of scar tissue enhances resistivity and modifies the patterns of wavefront propagation. When conduction is abnormally slow across a region ($\leq$ 0.1 m/sec, (Cranefield 1975)), an anomalous activation pattern, called "reentry", can be triggered: the excitation wavefront travels in single/multiple circular patterns, and reenters the area where it arose from. Much research effort has been devoted to the study and characterization of this disorder (Burnes, Taccardi, and Rudy 2000; Burnes et al. 2001; de Bakker et al. 1993).

The key components of the reentrant VT pattern, in terms of wavefront kinematics, are (Fig. 7):

1. a *cul-de-sac*-like region (isthmus), bounded by lines of block;
2. a breakthrough site in the isthmus area;
3. a single/multiple-loop reentry propagation pattern;
4. an excitation ending site located proximal to the breakthrough, but outside the blocked area.

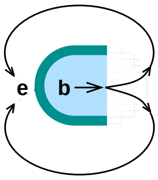


Figure 7: Schematic VT reentry circuit components.

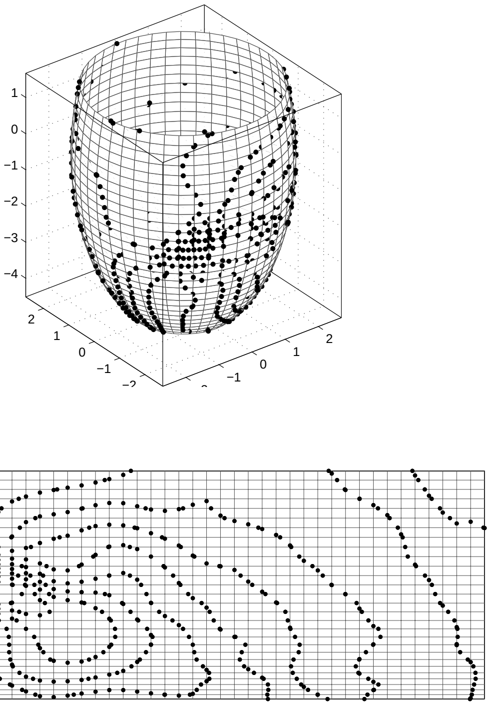


Figure 8: Isopoints (black dots) on the epicardial surface: the surface mesh is shown. Top panel: 3D geometry. Bottom panel: 2D cylindrical projection.

Given the discretized epicardial geometry Ω_h and the activation time field $\tau = \tau(\mathbf{x_i})$, $\mathbf{x_i} \in \Omega_h$, the main steps carried out to map it to a structural spatial representation of the salient propagation features, including the possible presence of a reentry VT pattern, are here very briefly summarized:

1. Breakthrough and exit sites, isopoints, and the time ordered sequence of the isochrones are first obtained (Ironi and Tentoni 2007); a planar projection of the 3D geometry is used to provide an overall representation of the features on the epicardial surface (Fig. 8: isopoints as they are mapped onto the projection).
2. The velocity field is computed as (Colli Franzone, Guerri, and Pennacchio 1998)

   $$\mathbf{v}(\mathbf{x}) = \nabla\tau(\mathbf{x})/|\nabla\tau(\mathbf{x})|^2$$

   where ∇ is the gradient operator. By mapping the velocity module range into a small set of qualitative values, e.g. *very-slow, slow, medium, high, very-high*, in accordance with threshold values suggested by the experts, the epicardial surface gets partitioned into homogeneous subregions, each of them labeled by the same qualitative value of the velocity module. In this context, the value *very-slow* corresponds to a pathological condition. Then:
3. If the region labeled *very-slow* , $\mathcal{L}$, is not empty,

(a) it gets redescribed by its gross skeleton, $\mathcal{L}^*$, which represents the abstracted *"conduction block"* line;

(b) a set of propagation lines, obtained as stream lines of the vector field, is generated from a neighborhood of the ends of the block line. They get classified into *"main propagation patterns"* according to their ending site, which can only be either a wavefront exit site or the upper/lower border of the map;

(c) check whether, among the ending sites associated with the main propagation paths, at least one is located close to the breakthrough site located nearest to the isthmus area (loop pattern).

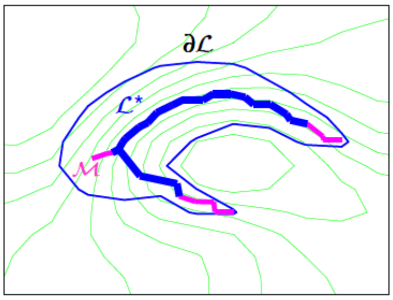


Figure 9: The approximated medial axis $\mathcal{M}$ (magenta thick line), and its pruned version $\mathcal{L}^*$ (blue thick line) are shown within the very-low-velocity area bounded by $\partial\mathcal{L}$.

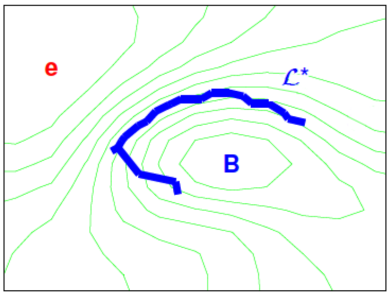


Figure 10: The conduction block is extracted as a complex of features: a line of block (simplified skeleton of the critical velocity area) which leaves a breakthrough and an extinction site at opposite sides.

Let us remark that very noisy data should be properly pre-processed to reduce noise to an acceptable level and allow reliable and robust feature extraction. Data smoothing actually corresponds to the way the expert approaches the visual reasoning task: getting rid of minor or spurious details to catch the main patterns.

Step 3 is aimed at discovering and abstracting a possible reentry circuit by singling out its key components.

Figure 9 shows, for the data set corresponding to Fig. 8, a detail of the area where isochrones are spatially denser: the boundary $\partial\mathcal{L}$ of a critical *very-slow* region is shown, as well as the Voronoi based medial axis $\mathcal{M}$, and the gross skeleton $\mathcal{L}^*$. Figure 10 shows the abstracted conduction block complex: a cul-de-sac region where isochrones get more crowded, bounded by a line of block which leaves a breakthrough and an extinction sites, spatially close to each other, at opposite sides. The line of block, so extracted as gross skeleton of the *very-slow* area, corresponds to merging the locally crowded isochrones.

Figure 11 shows the global outcome of the abstraction processes. It consists of: the sequence of activation isochrones, the breakthrough and exit sites, the discovered block of conduction, and the reentrant propagation patterns, starting at the ends of the block arc.

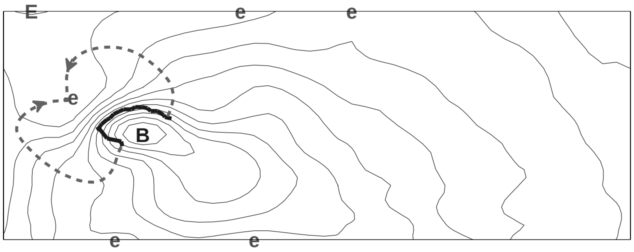


Figure 11: Outcome of the abstraction processes: activation isochrones (thin solid lines), breakthrough/exit sites (B/E labels), and the block of conduction (thick solid line). A couple of wavefront propagation lines, starting at the ends of the block arc, are shown (dashed thick lines).

Discussion

The approach herein proposed to automatically capture specific aspects of cardiac electrical activity is of broad methodological interest to electrocardiography, and more in general, to medical imaging. It results from the integration of standard computational geometry concepts with a spatial aggregation methodology. This latter, that aims at interpreting a numeric input field, allows us to capture structural information about the underlying physical phenomenon, and to identify its global patterns and the causal relations between them. Thanks to its hierarchical strategy in extracting objects at different scales, it facilitates the definition of inference rules that favor automated reasoning on spatiotemporal phenomena to perform a specific task.

In this work we focussed on algorithms for automated detection of diagnostic features from activation time fields. For the diagnosis of rhythm disturbances, activation maps are most appropriate as they provide information about the spatiotemporal course of the excitation wavefront. We showed how spatiotemporal features that characterize an important class of arrhythmias can be extracted from the given activation field.

As for the realization of a complete diagnostic tool for cardiac electric activity, further insight into the electric function could be drawn from the analysis of temporal sequences of potential data $t_i \rightarrow u(\mathbf{x}, t_i) \quad \mathbf{x} \in \Omega_h \; , \quad i = 1, ..n \; ,$ through the search for local current inflows/outflows identified by typical patterns of potential maxima and minima within isopotential maps. From potential data, especially intramural measurements, we could derive information about the electrical activity prior to its surface breakthrough that is complementary with respect to that obtainable from surface activation data. That would allow us to locate intramural components of reentry pathways associated with arrhythmogenic activity. However, the challenge of combining spatial and temporal aspects in a full 4D analysis goes with the still incomplete rationale of interpretation of such maps, and makes advances in this direction more remote.

At any rate, even if we limit our attention to epicardial activation data, additional work needs to be done:

(i) To validate the robustness of the proposed methodology on measured data, and then clearly delimit its weaknesses and strengths when applied in a clinical context. To this regard, sensitivity to noise should be more deeply investigated.

(ii) To deal with more complex phenomena, such as those involving the Purkinje network or multiple stimuli, and properly characterize and capture all of their propagation aspects. An example of such a complex situation is illustrated in Fig. 12: multiple wavefronts originate at distinct sites, after a few milliseconds they collide and merge into two fronts that advance in opposite directions. Wavefront collisions, which mark abrupt changes in the front topology and discontinuities in the velocity field, deserve both diagnostic attention and computational care.

(iii) To define a strategy for the comparison of the features of a given map against those of a *nominal* one, with the aim to detect and explain possible deviations from the expected patterns.

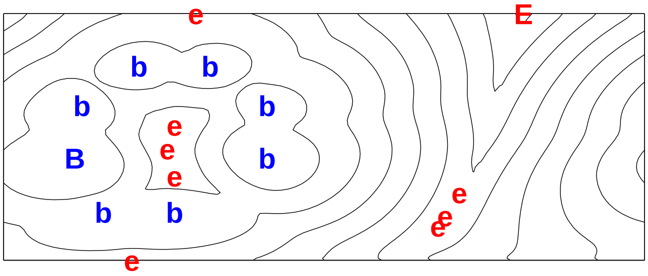


Figure 12: Activation isochrones (thin solid lines), and breakthrough/exit sites (B/E labels) in a case of simulated Purkinjie involvement. Multiple wavefronts break through at distinct sites, and then collide.

From a broader application perspective, the methodology we propose could contribute to a diagnostic tool specifically designed for arrhythmias, and could be used in a therapeutical context to evaluate the efficacy of a drug therapy aimed at normalizing the rhythm, through the detection of its effects on the spatial activation patterns.

References

Brandt, J. W., and Algazi, V. R. 1992. Continuous skeleton computation by Voronoi diagram. 55(3):329–338.

Bratko, I.; Mozetic, I.; and Lavrac, N. 1989. *Kardio: A Study in Deep and Qualitative Knowledge for Expert Systems*. Cambridge, MA: MIT Press.

Burnes, J. E.; Taccardi, B.; Ershler, P. R.; and Rudy, Y. 2001. Noninvasive electrocardiogram imaging of substrate and intramural ventricular tachycardia in infarcted hearts. *Journal of the American College of Cardiology* 38(7):2071–2078.

Burnes, J. E.; Taccardi, B.; and Rudy, Y. 2000. A noninvasive imaging modality for cardiac arrhythmias. *Circulation* 102(17):2152–2158.

Clifford, G. D.; Azuaje, F.; and McSharry, P. E., eds. 2006. *Advanced Methods and Tools for ECG Analysis*. Boston/London: Artech House Publishing.

Colli Franzone, P.; Guerri, L.; and Pennacchio, M. 1998. Spreading of excitation in 3-D models of the anisotropic cardiac tissue. II. Effect of geometry and fiber architecture of the ventricular wall. *Mathematical Biosciences* 147:131–171.

Cranefield, P. F. 1975. *The Conduction of the Cardiac Impulse: the Slow Response and Cardiac Arrhythmias*. Mount Kisco NY: Futura Publishing Co.

de Bakker, J. M.; van Capelle, F. J.; Janse, M. J.; Tasseron, S.; Vermeulen, J. T.; de Jonge, N.; and Lahpor, J. R. 1993. A noninvasive imaging modality for cardiac arrhythmias. *Circulation* 88(3):915–926.

Ironi, L., and Tentoni, S. 2003a. On the problem of adjacency relations in the spatial aggregation approach. In Salles, P., and Bredeweg, B., eds., *Seventheen International Workshop on Qualitative Reasoning (QR2003)*, 111–118.

Ironi, L., and Tentoni, S. 2003b. Towards automated electrocardiac map interpretation: an intelligent contouring tool based on spatial aggregation. In Berthold, M. R.; Lenz, H.-J.; Bradley, E.; Kruse, R.; and Borgelt, C., eds., *Advances in Intelligent Data Analysis V*, 397–417. Berlin: Springer.

Ironi, L., and Tentoni, S. 2007. Automated detection of qualitative spatio-temporal features in electrocardiac activation maps. *Artificial Intelligence in Medicine* 39:99–111.

Kundu, M.; Nasipuri, M.; and Basu, D. K. 1998. A knowledge based approach to ECG interpretation using fuzzy logic. *IEEE Trans. Systems, Man, and Cybernetics* 28(2):237–243.

Oster, H. S.; Taccardi, B.; Lux, R. L.; Ershler, P. R.; and Rudy, Y. 1997. Noninvasive electrocardiographic imaging: reconstruction of epicardial potentials, electrograms, and isochrones and localization of single and multiple electrocardiac events. *Circulation* 96:1012–1024.

Ramanathan, C.; Ghanem, R. N.; Jia, P.; Ryu, K.; and Rudy, Y. 2004. Noninvasive electrocardiographic imaging for cardiac electrophysiology and arrythmia. *Nature Medicine* 1–7.

Sakai, H., and Sugihara, K. 2006. A method for stable construction of medial axes in figures. *Electronics and Communications in Japan - Part II* 89(7):48–55.

Weng, F.; Quiniou, R.; Carrault, G.; and Cordier, M.-O. 2001. Learning structural knowledge from the ECG. In *ISMDA-2001*, volume 2199, 288–294. Berlin: Springer.

Yip, K., and Zhao, F. 1996. Spatial aggregation: Theory and applications. *Journal of Artificial Intelligence Research* 5:1–26.

Using Contextual Graphs for Supporting Qualitative Simulation Explanation

Mohamed El Habib LARABA

LIRE Laboratory- University Mentouri of Constantine
Route de Ain El Bey, 25000 Algeria
hlaraba@umc.edu.dz

Patrick BREZILLON

LIP6, case 169, University Paris 6
104 av. du Président Kennedy, 75016, France
Patrick.Brezillon@lip6.fr

Abstract

We proposed in a previous research an explanatory dialogical agentbased tool for explaining a qualitative simulation algorithm. The main limitation of an agent in our explanatory system was its incapacity to adapt itself to a changing context. The main reason concerns the agent's inability to share and understand, through its cognitive component, new contextual information not directly accessible for reasoning on it. In this paper, we present the basis on a new functionality of agents that allows contextual information to be freely distributed among agents and we model agent activity by using contextual graphs, a context-based formalism of representation allowing a uniform representation of elements of knowledge, reasoning and contexts.

Keywords. Qualitative Simulation, Explanation, Context,.Contextual graphs

Introduction

Laraba (2006) proposed a framework for explaining a qualitative simulation algorithm. Explanation was viewed as a problem solving process with its own reasoning and knowledge. An explanatory tool was then proposed and described at a high level of abstraction resulting on a dialogical agent-based This explanatory system cooperated with the end-user to provide him with the best explanation enhancing his comprehension of the QSIM algorithm. Explanations depend essentially on the context in which the user and the explanatory system interact. Such contextualized explanations are the result of a process and constitute a medium of communication between the user and the system

The main limitation of an agent in our explanatory system is its incapacity to adapt itself to changes of the context. The reason comes from the agent's inability to share and understand, through its cognitive component, a new contextual information that is not directly accessible for reasoning on it. In our explanatory system, an agent needs to handling a context representation for developing a shared context with other agents cooperating to generate the best explanation.

In this paper, we present a new functionality of agents that allows contextual information to be freely exchanged among agents, facilitating the generation and understanding of relevant explanations. Hereafter, Section 2 introduces contextual graphs and their relation with explanations. Section 3 recalls our previous work. Section 4 presents the revision we made of the system for including a model of context and Section 5 proposes modeling revised agent activity using contextual graphs.

Contextualized Explanations

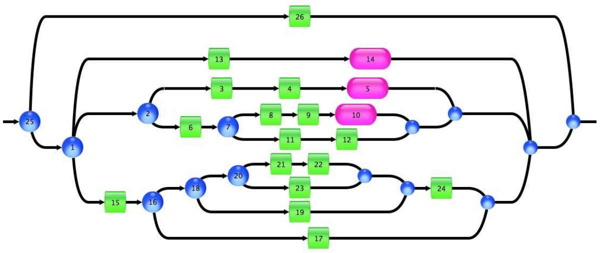
Introducing Context

Context has always played an important, if little understood, role in human intelligence. This is especially true in human communication and decision making. Context awareness allows an agent to develop his/her mental representation of the world and of others with which s/he interacts. Contextual elements come from different sources: each agent, the task accomplishment, the situation in which the task is realized, the environment, etc. A shared context allows many important aspects of human interaction to remain implicit when agents interact.

At least, there is now a consensus around the following definition "context is what constrains reasoning without intervening in it explicitly" (Brezillon and Pomerol, 1999)

Introducing Contextual graphs

A contextual graph represents the different ways to solve a problem. It is a directed graph, acyclic with one input and one output and a general structure of spindle (Brezillon, 2005). Figure 1 gives an example of contextual graph. A path in a contextual graph corresponds to a specific way (i.e. a practice) for the problem solving represented by the contextual graph. It is composed of elements of reasoning and of contexts, the latter being instantiated on the path followed (i.e. the values of the contextual elements are required for selecting an element of reasoning among several ones). Elements in a contextual graph are actions (square boxes in Figure 1), activities (complex actions like subgraphs), contextual elements (couples C-R in Figure 1) and parallel action groupings (a kind of complex contextual elements). A contextual element is a pair composed of a contextual node (e.g. C4 in Figure 1) and a recombination node (e.g. R4).



25: Is the site already known?
Yes 26: Look for new stuffs
No 1: What is the link target?
Html page 13: Open the target in a new window
14: Activity-1
PDF, DOC or PS page
2: Is there a html version?
Yes 3: Open target in new window
4: Look for the keywords
5: Activity-1
No 6: Download the document
7: Have I time now?
Yes 8: Open the document
9: Look for the keywords
10: Activity-1
No 11: Record document
12: Close the window
PPT page
15: Open the target in a new window
16: Duration of the download?
Short 18: Is it for a course?
Yes 20: Can page content be found?
Yes 21: Copy the slide
22: Paste it in a ppt doc
No 23: Note idea for later
No 19: Explore the presentation
24: Go to the next slide
Long 17: Close the window

Fig. 1. Activity exploitation of a Web page (from Brézillon, 2005)

Contextual graphs and Explanations

The acquisition of a new practice in a contextual graph corresponds to the addition of actions and contextual elements justifying the addition if the action(s). Moreover, several other contextual information pieces either are recorded automatically (date of creation, author, the practice-parent) or provide by the user (a definition and comments on the item that is introduced, etc.). An explanation is generated from the whole set of these contextual elements, thanks the formalism of representation allowing this. Thus, the expressiveness of an explanation depends essentially on the richness of contextual-graph formalism.

Our previous work

We considered an end-user observing QSIM progress on a particular physical phenomenon, say, the trajectory of a ball thrown in the air (Kuipers, 2001). The end-user wishes to have more ample information, and asks the explanatory system a query in natural language. For example, a query may be "why does such qualitative state appear after this number of transitions?" This may concerns the behaviour tree that is produced by the qualitative simulator, and represented by a qualitative table of state transitions (supposing that the user is well introduced in qualitative simulation). Another user that is novice in qualitative simulation could ask a query like "why does the ball change trajectory at that time?"

Laraba (2006, 2007) discusses some interesting points about explanatory reasoning and knowledge models. The first point is that some explanatory tasks need particular knowledge from different sources and feed by different subtasks executing simultaneously different sub-queries. For example, the task “Why-not-know” can be replaced by the sub-tasks “Why-not-know-C” for collecting constructive knowledge, “Why-not-know-D” for gathering domain knowledge and “Why-not-know-CC” for seeking cooperative and contextual knowledge. The interest is that other tasks of high level such as “Why-how-know” can be decomposed on the same basis of sub-tasks “Why-how-know-C”, “Why-how-know-D” and “Why-how-know-CC”. It is easy to establish a kind of library of such sub-tasks and to allocate them to agents (Laraba, 2007).

The second point is that interaction between the explanatory tool and the user also can be managed by a set of specific tasks. It is the case of the tasks “Analque”, “Consexp” and “Genexp” (Sansonnet et al., 2002). Again, such tasks can be allocated to specific agents.

Thus, it looks natural to design and develop the architecture of the explanatory tool in an agent-based formalism of representation to express the required distribution characteristics that we discuss in the following. For space constraints, our discussion will be limited to two tasks, namely “Why-how-key” and “Why-not-key”.

Agent Activity

The agent “Anque” introduces the explanatory process when it receives a user’s query. After checking the syntactic and semantic validity of the query, it will detect its object by identifying the type of adverb that is used. Finally, the needed knowledge is determined and, eventually, other agents are sollicitated either to confirm the detected interrogation by choosing one of the agents “Why-how-key” and “Why-not-key”, or to extract the knowledge necessary for the production of the explanatory text by opting for one of the following agents: “Whow-know-C”, “Whow-know-D”, “Whow-know-CC”, “Whot-know-C”, “Whot-know-D”, “Whot-know-CC”.

Then, the agent “Conex” takes over the construction of the explanatory text that the agent “Genex” will generate in naturel language and transmit to the end user. The end-user may be satisfied with it and the explanatory process is then interrupted, or not satisfied and the system is required to provide another explanation. The explanatory process is then either boosted such as described previously for a new request or relaunched after the explanatory knowledge updated otherwise. Both tasks are taken over by agent “Anque”.

Agent Model

To consider the cognitive processes operated during the various explanatory activities of an agent, we propose a dialogical agent modular architecture including four components represented in Figure 2.

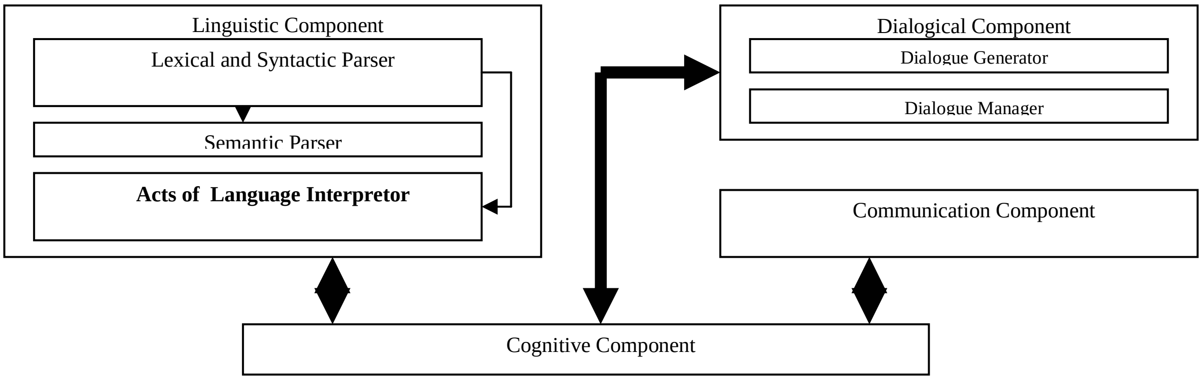


Fig. 2. Agent model

Discussion

A better understanding of each end-user’s needs and of the appropriate agents for our explanatory system needs context-specific information. Contextual knowledge intervenes in an implicit way in the explanation production process such as the knowledge elaborated during explanatory reasoning. Contextual knowledge appears at different levels from the knowledge retrieved from sources to the knowledge needed in the building of the explanation and its generation to the end-user.

This supposes that an agent in our explanatory system needs context not only to being explicitly represented but also shared and understood among agents cooperating to provide end-user with the best explanation. This is the goal of the new functionality that we plan for allowing contextual information to be freely distributed among agents. It will provide agent with the ability to capture context and to reason on it.

The introduction of the new functionality supposes an extension of our explanatory tool by adding a Context-Aware component to it, including:

- A context-capture sub-component: which acts when a new end-user request is received, to gather end-user personal information, his skills, his intervention location and time and some surroundings information that it transmits to the context-reasoning sub-component.
- A context-reasoning sub-component: which gathers the end-user profile according to the information transmitted by the context-capture sub-component, and transmits this information to the cognitive agent.

Then, the following revised agent model is obtained.

:

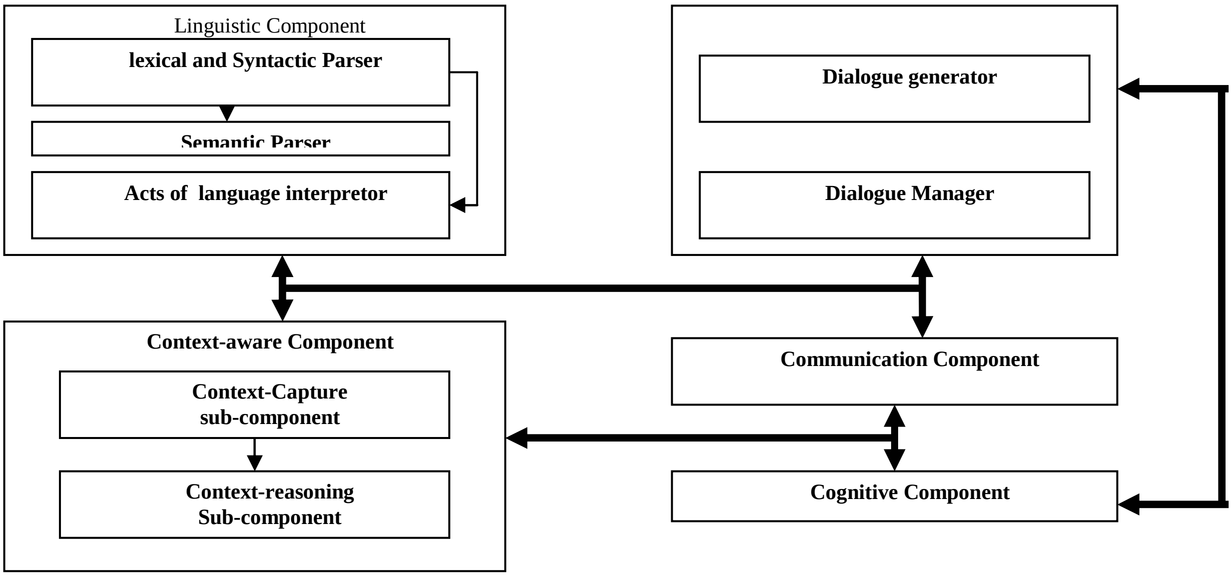
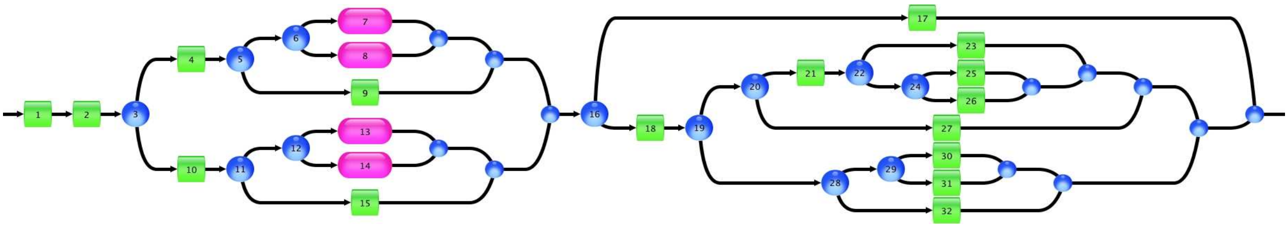


Fig. 3. : Revised Agent-model

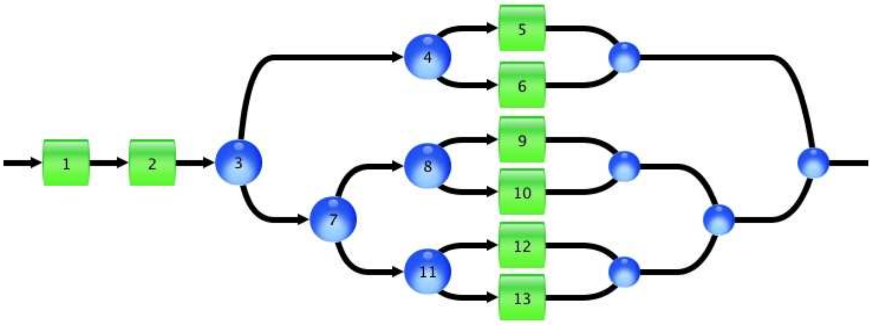
Activity modeling in the revised agent

New agent activity is modeled using contextual graphs. Figure 4 shows the contextual graph for the “throwing ball” example. It shows how the explanatory system determines the type of simulation (4: Initiate a classical simulation or 10: Initiating a qualitative simulation) that is needed according to user’s preferences (2: get user information) and how deciding in the example to present to user according to his profile (5: type of user).



1: Thrown ball movement Simulation
2: Get user information
3: User preference?
Classical simulation 4: Initiate a classical simulation
5: Type of user?
First user case 6: An Expert-user?
Yes 7: First Example
No 8: Second Example
Other user case 9: Deal with other user case
Qualitative simulation 10: Initiating a qualitative simulation
11: Type of user?
First user case 6: An Expert-user?
Yes 7: First Example
No 8: Second Example
Other user case 9: Deal with other user case
16: Are explanations needed?
No 17: Trigger an explanation by explanatory agent
Algorithm EXPLIQSIM18: Analyze user intervention
19: Request?
Analyze user request
20: valid request?
Yes 21: Analyze user question
22: Why-How?
Yes 23: Explain type_1
No 24: Why-not?
Yes 25:Explain type_2
No 26:Conclude on a failure
No 27: Conclude on a failure
Other intervention
28: Knowledge?
Yes 29: Compatible?
No 30: Conclude on a failure
Yes 31: Update
No 32: Conclude on a failure

Fig. 4 : Contextual graph for “throwing ball” example with its legend



1: Throwing-ball
2: Reasoning
3: Reasoning type?
First type 4: Manually ?
Ball thrown horizontally 5: Apply reasoning_1
Ball thrown vertically 6: Apply reasoning_2
Second type 7: With a software?
Case_1 8: Integral calculus?
Ball thrown horizontally 9: Apply reasoning_3
Ball thrown vertically 10: Apply reasoning_4
Case_2 11: Differential equations calculus?
Ball thrown horizontally 12: Apply reasoning_5
Ball thrown vertically 13: Apply reasoning_6

Fig. 5 : Contextual graph for “Activity example” with the definition of the elements

The contextual graph shows different ways to simulate the throwing ball phenomenon. The first two paths in the contextual graph correspond to two specific ways (i.e. two practices) for simulating that phenomenon, namely classical simulation and qualitative simulation. When a path is selected (Action 3 or Action 5 in the contextual graph) according to the information collected about the user (preferences and knowledge), first, the corresponding elements of context are instanciated and, second, an element of reasoning is selected. This information and other information pieces that deal with some practice changes (the user is responsible of) in the contextual graph are used by the explanatory agent for generating an explanation, and to tailor its explanation by detailing parts unknown of the user and sum up parts developed by the user. Such an explanation might be asked by the user after observing the simulation process (Action 8 in the contextual graph) or triggered by an explanatory agent (Action 7 in the contextual graph) that anticipates user’s reasoning from the contextual graph and then providing him with suggestions or explanations. In both cases explanatory agent may fail to match the user’s practice with its recorded practices. Then, the system needs to acquire incrementally new knowledge and learning the corresponding practice developed by the user (generally due to specific values of contextual elements not taken into account before). This is an explanation from the user to the system.

Conclusion

This study relies on the realization of an explanatory tool that we developed earlier. The important step in the evolution of the explanatory tool concerns, first, the use of contextual knowledge as a part of the body of explanatory knowledge, in the realm of Karsenty and Brezillon’s claim (1995) and, second, the use of Contextual Graphs for modeling agent activity. This allows the generation of two types of explanation (user-based explanation and real-time explanation) among those that Brezillon (2008) discussed.

The next step would be to integrate more intimately context modeling within this architecture. We think that making context explicit in an explanatory system will have positive consequences: first a better management of the knowledge upstream, and second a better management of interaction with end-users.

References

Brézillon, P. (2005). Task-realization models in Contextual Graphs. In 5th International and Interdisciplinary Conference on Modeling and Using Context, Lectures Notes in Artificial Intelligence, Vol 3554, pp. 55--68, Springer-Verlag.

Brézillon, P. (2008). Explaining for Contextualizing and Contextualizing for Explaining. Invited talk at Exact 2008 workshop on Explanation-Aware Computing.

Brézillon, P. (1994). Contextualized explanations, Proceedings of International Conference on Expert Systems for Development, Volume, Issue, 28-31 Mars 1994 Page(s):119 – 124

Brézillon, P. and Pomerol, J.-Ch. (1999) Contextual knowledge sharing and cooperation in intelligent assistant systems. Le Travail Humain, 62(3): 223-246.

Karsenty L. and Brézillon P. (1995). Cooperative problem solving and explanation, International Journal of Expert Systems With Applications, 8(4): 445-462.

Kuipers, B.J (2001). Qualitative simulation, *In“Encyclopedia of Physical Science and Technology”, Troisième Edition*, NY: Academic Press, pp 287-300.

Laraba, M.E.H (2006). A Dialogical Agent-based Framework for Explaining the Results of QSIM Algorithm to the End-User, Qualitative reasoning workshop QR 2006, Hannover, USA.

Laraba, M.E.H. (2007). Sur le Raisonnement Qualitatif: Une Approche à base d’Agents Dialogiques pour l’Explication de la Simulation Qualitative, Thèse de Doctorat, Université Mentouri de Constantine, Algeria.

Sansonnet, J.-P., Pitel, G., Sabouret, N. (2002). Un Langage de Description d'Agents dédié à l'Interaction Dialogique, Journées Francophones pour l'Intelligence Artificielle Distribuée et les Systèmes Multi-Agents, .JFIADSMA'02, Lille, France.

Wooldridge, M., Jennings., N.R. (1995). Agent theories, architectures, and languages. Wooldridge, Jennings (ed), Intelligent Agents, Springer Verlag, pp.1-22.

Model Building Experiences using Garp3: Problems, Patterns and Debugging

Jochem Liem and **Floris E. Linnebank** and **Bert Bredeweg**
Human-Computer Studies, Informatics Institute, University of Amsterdam, The Netherlands
Email: {J.Liem, F.E.Linnebank, B.Bredeweg}@UvA.nl

Abstract

Capturing conceptual knowledge in QR models is becoming of interest to a larger audience of domain experts. Consequently, we have been training several groups to effectively create QR models during the last few years. In this paper we describe our teaching experiences, the issues the modellers encountered and the solutions to solve them in the form of reusable patterns, and finally a structured way to debug models.

Introduction

Domain experts have been making more complex QR models the last few years. The models capture several processes and their interactions. However, different modellers seem to be reinventing modelling patterns to solve certain problems. This paper is meant to raise awareness in the model building community about the frequently encountered representational issues and possible solutions. The suggestions described in this paper are not prescriptive, but describe patterns that other modellers have found useful. As such, this paper is different than usual QR papers, since it is not focussed on algorithms, but instead offers modelling advice.

The paper introduces the groups we have trained the last few years, explains the representation used in the Garp3 QR modelling and simulation workbench, describes modelling and debugging issues and proposes solution patterns.

Modeller training

Three groups of modellers we have trained the last few years are particularly interesting.

- The first group are the (PhD-level) researchers who participated in the NaturNet-Redime EU project. These (non-computer scientist) domain-experts created models about ecology topics that they are actively researching.
- The second group are BSc. Future Planet Studies students who started their first semester of college education. As such, they had neither advanced computer-science knowledge, nor detailed knowledge about particular domains.
- The final group are PhD-students doing the SIKS research school Knowledge Modelling course[1]. Most of them have backgrounds in fields close to computer-science.

The researchers have been working with the Garp3 software for about 2,5 years. Their training started at the second project meeting of the NaturNet-Redime project in Amsterdam, which consisted of 2 full days of hands-on practical sessions (including a 2 hour lecture), followed by a single day of working through the structured approach to building QR models (Bredeweg et al. 2008) (including a 2 hour lecture). Extra training was given during each following project meeting, which included a day of assignments and a day of debugging models in Sofia, a day of using the Sketch environment and the sharing and reuse functionality (Liem, Bouwer, and Bredeweg 2007) in Latvia, and a day of collaboratively improving the case study models in Germany. Additionally, the researchers were supported via bi-weekly Skype/Flashmeetings to discuss modelling issues. The results are complex models involving multiple interacting processes (Sànchez-Marrè et al. 2008).

The BSc. students doing an 8 week conceptual modelling course were divided in pairs. They spent the first 4 weeks learning to make Concept Maps (Cañas et al. 2004) and creating ontologies using Protégé (Knublauch et al. 2004). In the last 4 weeks the student pairs were learning to create QR models. The main goal for them was to create a small model about at least 2 processes relevant to the carbon cycle (and global warming). In addition to learning the QR technology, the students were asked to work towards this goal during these 4 weeks. Each week the students gave a 10-15 minute presentation about their current modelling progress towards the carbon cycle models. The students were supposed to spend about 8-10 hours a week on the course (including the weekly 3 hour practical session).

In first QR week, a 1 hour lecture was given contrasting QR models with concept maps and ontologies, explaining the general ideas of QR, and the applications of QR models. Following the lecture, the students worked on the Tree & Shade model (Bredeweg et al. 2006b). Nearly all students finished this exercise within the practical session. In the second QR lecture, the communicating vessels model (Bredeweg et al. 2006b) was used to explain the key representational aspects of QR models such as structure (entities and configurations), causality (proportionalities and influences), inequalities, correspondences and model fragments. In the rest of the session the students worked on recreating the population interaction model (Bredeweg et al. 2006b),

---

[1] http://hcs.science.uva.nl/SIKS/Siks2008

which took them most of the assigned time. In the last two weeks, feedback was given on the models presented by students. In addition, we gave two extra 2 hour practical sessions to accommodate requests by students. Most students created excellent models focussing on two processes.[2]

The PhD students of the SIKS research school doing the Knowledge Modelling course[3] got a 1 hour lecture on QR, followed by a two hour practical session in which they had to recreate the Tree & Shade model. The PhD students required on average about half an hour less time compared to the BSc. students to finish. This seems mostly to be due to their computer skills. However, those who also had some knowledge modelling skills were able to finish the modelling task the fastest (up to 30 minutes faster compared to the students without modelling skills).

QR Modelling and Simulation using Garp3

The introductory QR lectures are supposed to give the audience enough basic knowledge to allow some hands-on experience creating QR models using the Garp3 workbench[4] (Bredeweg et al. 2006a). Garp3 allows modellers to capture their knowledge about the structure and the important processes in their system of interest as *model fragments*. These can be considered formalisations of the knowledge that applies in certain general situations. Model fragments can be considered rules indicating that if certain model ingredients are present (conditions), certain other model ingredients must also apply (consequences). The can be represented as: $conditions \Rightarrow consequences$.

Next to model fragments, different *scenarios* can be modelled. These represent specific start states of a system. Garp3 can run simulations of models based on a particular scenario. The result of such a simulation is a state graph, in which each state represents a particular possible situation of the system, and the transitions represent the possible ways a situation can change into another.

The state graph is generated by the simulation engine roughly as follows. The engine takes a scenario as input, and finds all the model fragments that apply to that scenario. The consequences of the matching model fragments are added to the scenario. The result is a state description including all consequences. Based on this augmented state description new knowledge can be inferred, such as the derivatives of quantities. Given a completed state description, the possible successor states are inferred. The complete state graph is generated by applying the reasoning to the new states.

The QR representation has a strict separation between structure and behaviour. The structure of a system is represented using *entities* (objects), *agents* and *configurations* (relations). For example, a lion hunting a zebra can be represented as two entities (lion and zebra) and a configuration (hunts). If the food web is considered to be the system, a hunter disturbing this system could be represented as an agent.

One of the key behavioural model ingredients are *quantities*. Quantities represent the features of entities and agents that change during simulation. A quantity has a magnitude and a derivative, which represent its current value and trend. The magnitude and derivative are each defined by a quantity space that represent the possible values the magnitude and the derivative can have. Such quantity spaces are defined by a set of alternating *point* and *interval* values.

$M_v(Q_1)$ is used to refer to the current value of the magnitude of a quantity. $M_s(Q_1)$, the sign of the magnitude, indicates whether the magnitude is positive, zero or negative ($M_s(Q_1) \in \{+, 0, -\}$). $D_v(Q_1)$ is used to refer to the current value of the derivative of a quantity, which has to be a value from the predefined derivative quantity space ($D_v(Q_1) \in \{-, 0, +\}$). $D_s(Q_1)$ is used to refer to the current sign of a derivative.

As a shorthand to refer to the current magnitude and current derivative value of a quantity at the same time, we use the notation Q[X,Y], where Q is the quantity, X is the current magnitude value and Y is the current derivative value. For example, Size[+,-] indicates that the current size is positive and decreasing. This combination of the current magnitude and current derivative value is called the *quantity value*.

Causality

Important for QR models is the explicit notion of causality between different quantities. Garp3 represents the causal dependencies using direct and indirect influences (Forbus 1984). Direct influences, called *influences* for short, are represented as $Q_1 \overset{I+}{\rightarrow} Q_2$. Influences can be either positive (as above) or negative. The positive influence will increase $D_v(Q_2)$ if $M_s(Q1) = +$, decrease it if $M_s(Q1) = -$, and have no effect when $M_s(Q1) = 0$. For a negative influence, it is the other way around.

The indirect influences, called *proportionalities*, are represented as $Q_1 \overset{P+}{\rightarrow} Q2$. Similar to influences, proportionalities can be either positive or negative. The positive proportionality will increase $D_v(Q_2)$ if $D_s(Q_1) = +$, have no effect if it is stable, and decrease if it is below zero. For a negative proportionality, it is the other way around.

Other Behavioural Ingredients

Other behavioural ingredients essential for qualitative simulations in Garp3 are operators, inequalities, value assignments and correspondences. Operators (+ and -) are used to calculate the magnitude value of quantities (e.g. $Q_1 - Q_2 = Q_3$, to indicate $M_v(Q_1) - M_v(Q_2) = M_v(Q_3)$). Inequalities can be placed between different model ingredient types: (1) magnitudes ($M_v(Q_1) = M_v(Q_2)$)[5], (2) derivatives ($D_v(Q_1) < D_v(Q_2)$, (3) values $Q_1(point(Max)) = Q_2(point(Max))$[6], (4) operator rela-

---

[2] This type of learning through modelling will be the main topic of research in the recently started DynaLearn EU project.

[3] http://hcs.science.uva.nl/SIKS/Siks2008

[4] http://www.garp3.org

[5] Even if two quantities have the same qualitative value, they can still be quantitatively different (different points in the interval). An inequality can be used to indicate that they also have the same quantitative values.

[6] Values with the same name associated with different quantities do not necessarily have the same value. Points can represent to different quantitative values (e.g. the *maximum* heights of two

tions ($M_v(Q_1) - M_v(Q_2) < M_v(Q_3) - M_v(Q_4)$), (5) combinations of 1, 2, 3 and 4 (although magnitude and derivative items cannot be combined in a single expression). Value assignments simply indicate that a quantity has a certain qualitative value ($M_v(Q_1) = Q_1(Plus)$). Finally, correspondences are used to indicate that for certain values of one quantity, values of another quantity can be inferred. There are quantity correspondences ($Q_1 \overset{Q}{\leftrightarrow} Q_2$) and value correspondences ($Q_1(Plus) \overset{V}{\rightarrow} Q_2(Plus)$), which can both be either directed or undirected. The value correspondence indicates that if $M_v(Q_1) = Q_1(Plus)$ then $M_v(Q_2) = Q_2(Plus)$. If the value correspondence is bidirectional, the reverse inference is also possible. Quantity correspondences can be considered a set of value correspondences between each consecutive pair of the values of both quantities. There are also inverse quantity space correspondences ($Q_1 \overset{Q\hookleftarrow}{\leftrightarrow} Q_2$) that indicate that the first value in Q_1 corresponds to the last value in Q_2, the second to the one before last, and so on.

Modelling Issues

Representing Structure

Entities or quantities? One of the main purposes of conceptual models is communication. QR models make an explicit distinction between structure and behaviour of a system to make models easier to understand. The quantities describing the behaviour of the system are attached to entities that describe the structure of the system. A balanced distribution between the number of quantities and the number of entities (i.e. only a few quantities per entity) improves the communicative value of a model.

The number of entities in a model should depend on the importance of those entities in the system. Otherwise they could be represented as quantities. For example, in the river restoration models (Sànchez-Marrè et al. 2008) we frequently noticed the use of quantities such as *oxygen concentration* and *Particulate Organic Matter (POM) concentration* as properties of an entity *river*. Since the POM and oxygen do not have important properties of their own for purposes of this model, they are modelled as quantities (the concentrations are properties of the river).

However, if we consider algae in the river, there is a modelling choice to be made. Algae can be modelled as an entity in the system (living in the river), or as a quantity of the river (Algae concentration). This choice depends on the importance of the Algae for the processes modelled in the system. For example, if the photosynthesis or biomass of the algae is important, Algae should become an entity with these features as quantities, since these quantities are features of the Algae and not of the river.

Configuration naming and direction In the investigation of the models created the last few years, it became apparent that in the modelling of the structure of a system, naming the configurations and choosing a direction is often experienced as being an issue. For example, when population A is preying on population B, is it better to formalise this as Population A $\xrightarrow{\text{preys on}}$ Population B, or as Population B $\xrightarrow{\text{is preyed on by}}$ Population A?

This issue is analogous to writing in either active or passive voice. In our experience, the passive voice is frequently used. We propose that the active form should be consistently used for the naming of configurations. This shortens the configuration names, making the diagrams easier to read. Furthermore, if text is generated based on the contents of a model (e.g. a question generator or virtual character explaining the model), the quality of the text will be better.

Relationship reification There are relationships in systems that are difficult to formalise as configurations, since there are no verbs to describe them. For example, the Ants' Garden model (Salles, Bredeweg, and Bensusan 2006) describes the different interactions that populations can have with each other, such as commensalism, parasitism, and symbiosis. For parasitism a configuration *parasitises* could be defined, however no such verbs are available for commensalism and symbiosis. Using a configuration *lives in symbiosis with* seems suboptimal, since it has a long name and the direction seems arbitrary since the inverse is also true. Adding a second configuration to remedy this would only make the diagram more complex.

Another related issue is representing the speed of these processes, such as the parasitism rate (or other properties of the relationship). Assigning this rate (formalised as a quantity) to either of the populations participating in this relationship seems incorrect, as it is determined by the interaction of these populations, and not one particular population alone.

As a solution to these issues the relationship can be reified, i.e. represented as an entity. In the Ants' Garden model, the symbiosis relationship is described as an entity with *symbiont 1* and *symbiont 2* configuration relationships to the two populations. Although not in the Ants' garden model, the speeds at which these processes operate can be formalised as quantities attached to the reified relationships.

Representing causality

Choosing a proportionality or influence An important difficulty we encountered with all three groups is conveying the difference between influences and proportionalities. Moreover, even after having hands-on experience with creating models based on exercises, modellers still have trouble choosing whether to use an influence or a proportionality.

The key concept to understand is that only influences initiate change in a system and that proportionalities only propagate change. Specifically, the magnitude of the source quantity of an influence determines the derivative of the target quantity. As such, influences only cause change when the source quantity has a non-zero magnitude value. Proportionalities on the other hand determine the derivative of the target quantity based on the derivative of the source quantity, and thus only change the derivative of the target quantity when the source quantity is not stable.

We propose the following rule of thumb to decide whether an influence or a proportionality should be used when a modeller is sure that two quantities are causally linked. First, assume that the source quantity has a positive magnitude

---

container do not have to be equal).

value, but is stable (i.e. the derivative is zero). If the target quantity is supposed to change an influence should be used. Otherwise, a proportionality should be used. The reason this rule works is that a proportionality does not have an effect in this setting (since the derivative of the source quantity is stable), while an influence does since the magnitude of the source quantity is non-zero.

For example, consider water flowing from a tab into a bucket. The flow causes the amount of water in the bucket to increase. Should an influence or proportionality be used? Consider that the flow is positive but not changing. Since the amount of water should still increase, an influence should be used. The same rule of thumb can be used when considering the causal relation between the amount of water and the height of the water in the bucket. Consider that the amount of water is positive but stable. Since the height of the water should also remain stable, a proportionality should be used.

Causal chains Causal chains often start with an influence followed by several proportionalities that propagate the effect. Chains of proportionalities following each other occur quite often. These kind of causal chains are seen in many models. In contrast, it is unlikely that there is another influence in a causal chain. As such, causal chains with influences in them are more likely to be incorrect. Conceptually a causal chain should be seen as a process that affects several causally linked quantities. Other influences should therefore be part of other causal chains.

A special case of a causal chain is one that contains a loop of proportionalities. For example, $A \xrightarrow{P+} B, B \xrightarrow{P+} C, C \xrightarrow{P+} A$. These loops of proportionalities should be avoided, as the value of the derivatives of these quantities can never be derived. The reason is that to derive the derivative of one of the quantities, the derivative of the quantity before it has to be known. However, to determine that derivative, the derivative of the quantity before that has to be known, etcetera.

Feedback loops A frequently asked question about QR models is whether feedback loops are supported by QR models. A feedback loop in a system is a situation in which the effect of a process will influence this same process. For example, the growth of a tree increases the size of the tree, but the size of the tree also increases the growth rate. Feedback loops frequently occur in QR models and are one of the most basic patterns that occur in most models. The mentioned tree example can be seen in the Tree & Shade model.

A feedback loop is represented in Garp3 by specifying an influence from a process quantity to a target quantity and a proportionality from the target quantity to the process quantity. For example, in the Tree & Shade model there is a positive influence from the growth rate of the tree to the size of the tree and a positive proportionality from the size of the tree to the growth rate. This pattern exactly captures the feedback loop in the system.

Such feedback loops do not have to be direct. There can be a causal chain from the process quantity through several quantities with the final quantity providing the feedback to the process quantity. One such example can be seen in the communicating vessels model. The flow in the pipe between two containers has a negative influence on the volume of the liquid in the container (i.e. the flow reduces the amount of water). There are positive proportionalities from Volume to Height and from Height to Pressure to indicate that if volume changes, height will change in the same direction and if the height changes, the pressure will also change in the same direction. The feedback is represented in the form of a positive proportionality from the pressure to the flow. This proportionality indicates that the flow will increase if the pressure increases and decrease if the pressure decreases (as it will if water is flowing out of the container).

Causal Interactions As part of each introductory QR lecture we present the audience with a set of exercises in which two causal dependencies affect the same quantity. A member of the audience is asked what the resulting derivative value will be for the affected quantity. Each of the three groups of modellers had difficulty in deriving the correct derivative and explaining the result.

The exercises start with an exercise that tests whether the audience has understood the semantics of the causal dependencies. An example exercise is $Q1[-,0] \xrightarrow{I-} Q2[+,?]$. The audience has to indicate that Q2 will decrease, since the magnitude of Q1 is negative and it affects Q2 though a negative influence. Several of the people in the audience are able to correctly derive the correct result and explain it to the rest of the audience.

In the following exercises the audience has to derive the derivative of the quantity that is affected by two causal dependencies. For example, $Q1[-,-] \xrightarrow{I+} Q2[+,?] \xleftarrow{I-} Q3[-,-]$. The correct answer here is that the derivative of Q2 is ambiguous. The reasoning is as follows. Q1 has a negative magnitude which results in a negative effect on Q2 through the positive influence. Q3 has a negative magnitude value but influences Q2 through a negative influence. As a result the effect on Q2 is positive. Given that there is a positive and a negative result on Q2 the result is ambiguous.

Although not explained during the lecture, in more advanced modelling the ambiguity of this kinds of examples can be resolved by adding inequality knowledge. For example, knowing that $M_v(Q1) > M_v(Q3)$ allows us to derive a unique derivative value. Since the negative effect of the positive influence from Q1 is smaller (less negative) than the positive effect of the negative influence from Q3 (more negative), Q2 will increase.

Dealing with multiple competing causal dependencies Many real-world problems involve multiple processes affecting single quantities. Although two competing influences of different types can be determined through a single inequality (see previous section), the more general case with multiple causal dependencies is more intricate. Consider two influences of the same type (both positive or both negative) affecting a single quantity, for example, the effects of release of CO_2 from the ocean (which can be negative to model the absorption of CO_2) and the burning of fossils fuels on the CO_2 concentration in the air. Given $release[-,+] \xrightarrow{I+} concentration[+,?] \xleftarrow{I+} burning[+,+]$, the derivative value of concentration is ambiguous.

Inequality knowledge between concentration and burning with not resolve the ambiguity. The knowledge that is needed is whether the absolute magnitude value of release is bigger or smaller than the absolute magnitude value of burning. In the former case, $D_v(concentration) = +$, while in the latter case $D_v(concentration) = -$. However, such representing absolute values and reasoning with them has not been solved in Garp3 yet.

A general pattern that can be used instead is specifying an inequality between the sum of all quantities with positive effect and the sum of all quantities with negative effects. In the example, the knowledge $release + burning < 0$ allow us to infer $D_v(concentration) = -$. If we also consider the effects of photosynthesis ($photosynthesis \xrightarrow{I-} concentration$), we can again resolve the ambiguity by specifying the inequality $release + burning < photosynthesis$ ($D_v(concentration) = -$). This pattern allows modellers to specify what the result on the influenced quantity will be given a set of magnitude values (as conditions) of the processes.[7]

Correspondences as causality Correspondences are used to ensure that magnitude values always occur together. They are often paired with proportionalities. The correspondence assures that the magnitude values of the quantities always co-occur, while the proportionality ensures that the derivative values are equivalent (assuming there are no other causal dependencies on the quantity). In the case that there are multiple causal dependencies on the target quantity, a correspondence might be to strict.

Correspondences can be either directional or bidirectional. Directional correspondences are important when the magnitude value of one quantity can be inferred from another quantity, but not the other way around. For example, when the size of a population is zero, the birth rate should be zero, but the birth rate can be zero with positive population size. An example of a bidirectional correspondence is between the size of a population and their biomass.

Establishing Quantity Spaces

The selection of suitable quantity spaces for quantities is experienced as being a difficult task even by experienced modellers. A quantity space should contain just the right amount of distinctive values necessary to model the behaviour of that particular quantity in a qualitatively meaningful way. Inherent in this choice is the purpose of the model. For example, when modelling the effect of phytoplankton concentration on the amount of light it absorbs, modelling the critical concentration value when the other primary producers underneath them become significantly deprived of light in the quantity space is important. However, when modelling the effects of global warming on phytoplankton, this value is less important and could be left out of the model.

When choosing a quantity space it is important to determine the qualitative distinct values a quantity can take that might cause a change in behaviour. This means thinking of particular value ranges in which a certain behaviour of the system takes place. These ranges are bounded by particular points that represent the thresholds between these ranges. This is the reason that quantity spaces in the QR representation consist of consecutively intervals (ranges) and points. In the above example about light, the concentration of phytoplankton could either be: no plankton, some plankton, a threshold representing the critical amount of plankton, and more than the critical concentration of plankton (e.g. {point(zero), positive, point(critical), hazardous}.

Note that the choice of the quantity space {zero, low, point(medium), high} for, for example, the size of a population is not ideal if medium is thought of as an interval (like low and high). Firstly, there seems to be no clear reason why this distinction is important from a behavioural point of view. Secondly, medium becomes a point value in this quantity space, and the behavioural properties of points are quite different than those of intervals. As a result certain behaviours of the system will not be simulated. The main reason for this is the epsilon ordering concept (de Kleer and Brown 1984), which indicates that changes from a point to an interval always have precedence over a change from an interval to a point. This means that a changing quantity can remain having the same interval magnitude value in consecutive states, but a changing quantity that is in a point value must change to the next magnitude value in the next state.

Consider two growing populations with size low. Given the quantity space discussed above there are only three possible behaviour paths. From the first state, there are three possible options, either the first population reaches medium first, the second population reaches medium first, or they reach medium simultaneously. Since medium is a point value, it is not possible for the population that is still low to reach medium before the other population has reached high, due to the epsilon ordering rule. As such, this possible behaviour is not captured in the model. Consequently, we argue that the choice for this quantity space should be avoided, and that in general modellers should make sure not to model intervals as points.

Actuator Patterns

Although conceptually changes in systems should either be caused by processes active in the system or by forces outside the system, there are several technical ways to initiate change within a QR model. For instance, it is possible to indicate that a certain quantity is always increasing. Several frequently occurring patterns can be used to initiate change in a QR model. We call these *actuator patterns*, since they put the system into action.

Process actuator Processes represent the causes of change within a system. Consider the Growth process ($Growth \xrightarrow{I+} Size$ in a process model fragment *Growth*) represented in the Tree & Shade model. There are three variations of this actuator that are commonly used. In the simplest variation, the growth rate is simply assigned a positive mag-

---

[7] When a quantity is affected by both a proportionality and an influence with opposing effects, the result is always ambiguous. Consequently, we argue that mixing different types of causal dependencies should be avoided. This rule is known as the *homogeneous influences adequacy constraint* (Rickel and Porter 1997).

nitude and a stable derivative through consequential value assignments. A drawback is that the growth of the tree cannot change and can never become zero. To resolve this issue, in the second variation a feedback is added between the size of the tree and its growth rate ($Size \xrightarrow{P+} Growth$). Consequently, no value assignments are needed in the model fragment, except a start value for the growth rate in the scenario. The third variation removes the need for the value assignment in the scenario. In the Growth model fragment a correspondence is added to indicate that a non-existing tree does not grow ($Size(zero) \xrightarrow{V} Growth(zero)$). Furthermore, a child model fragment is created that indicates that all trees grow ($M_v(Size) > zero \Rightarrow M_v(Growth) > zero$).

External Actuator Pattern and Exogenous Behaviour The *external actuator* pattern models processes or effects of processes from outside the system. The patterns consists of an agent representing the source of the effect, and an associated quantity which represents an exogenous variable. *"Human modelers treat a variable as exogenous only if it is approximately independent of the other variables in the model."* (Rickel and Porter 1997). Garp3 allows exogenous behaviour to be specified for exogenous quantities (Bredeweg, Salles, and Nuttle 2007), which allows modellers to indicate that a quantity remains constant, is increasing, decreasing or steady, or has sinusoidal or random behaviour. Sinusoidal behaviour is used for cycles, such as day-night cycles, tides (monthly), and precipitation (yearly), while random behaviour is used for quantity behaviour that a modeller is unsure of and might unexpectedly change (e.g. rainfall over a shorter period of time).

There are two variants of the external actuator pattern. To model an external process (fully determined by forces outside the system) a quantity is combined with an influence. The influencing exogenous quantity tends to be set using a value assignment (as in the process actuator pattern), with either the derivative being set or determined by a feedback relationship. The second variant models the effects of external processes using an exogenous quantity and a proportionality. These external processes are often determined by giving the quantity an exogenous behaviour.

The choice between the two variants depends on what the exogenous quantity should do. For example, when the exogenous quantity fully determines a quantity in the system (e.g. with two corresponding large quantity spaces), this is modelled using a proportionality. For example, the nutrient run-off caused by farming fully determines the nutrient level in the Danube river and delta, and the average ambient temperature of the surrounding land determines the temperature in the river and delta (Sànchez-Marrè et al. 2008). In contrast, an exogenous process is used when an important process has to be modelled. For example, a fishery manager stocking young salmon in a river, or economical development activities increasing the number of anglers (Sànchez-Marrè et al. 2008).

Equilibrium Seeking Mechanisms The equilibrium seeking mechanism pattern models equalizing flows due to a potential difference. For example, energy exchange between two objects with different temperatures, or a liquid flow equalizing the pressures in the communicating vessels system. Key in this pattern is the flow, which is determined by the difference between two state variables, e.g. of the temperatures of two objects ($Temperature1 - Temperature2 = Heat\ flow$). The heat flow reduces the heat from one object, and transfers it to the other ($Flow \xrightarrow{I-} Heat1, Flow \xrightarrow{I+} Heat2$). Finally, the two state variables determining the flow also determine the derivative of the flow. If the temperature of the first object increases, the flow will increase ($Temperature1 \xrightarrow{P+} Heat\ Flow$), while if the temperature of the second object would increase, the flow would decrease ($Temperature2 \xrightarrow{P-} Heat\ flow$). In the communicating vessels model, the pressure quantities determine the flow, while the flow changes the volumes of the water in the containers through influences.

Competing Processes The *competing processes* pattern consists of multiple interacting influences that model competing processes. There are at least two processes, such as the birth and death rate of a population, or more such as its immigration and emigration rates. The processes influence a single quantity, in this case the size of the population ($Birthrate \xrightarrow{I+} Size, Deathrate \xrightarrow{I-} Size$). There are also feedbacks: a larger population means a larger birth rate ($Size \xrightarrow{P+} Birthrate$) and a larger death rate ($Size \xrightarrow{P+} Deathrate$). More details on how to deal with these kind of interactions is explained in the Sections *Causal interactions* and *Dealing with multiple competing causal dependencies*. The mentioned examples come from the 'Single population model with basic processes' model which is provided with the Ants' Garden model.

Issues when Running Simulations

Maximum simulation result Modellers often ask why their QR models generate so many states. One of the main reasons this question is asked is because modellers tend to underestimate the number of states a QR model can potentially generate. The maximum number of states that a model can generate is equal to the Cartesian product of all the quantity spaces of all the quantities in a model. So a model with 10 quantities with three possible magnitude values can generate at most $(3x3)^{10} = 3486784401$ states, which is the number of magnitude values times the number of derivative values raised to the power of the number of quantities. As such, adding one quantity more to a model can potentially mean almost an order of magnitude more states (number of potential magnitude values times the number of derivative values). Note that this number includes only the possible different states due to different magnitude and derivative values and does not include different states due to different inequalities. So even more states are possible.

Successor states without correspondences A frequently seen reason for a large number of states is non-corresponding quantities. Consider that all changing quantities in a point value will change to an interval value in the next state due to the epsilon ordering rule (which states that changes from a point value to an interval are immediate).

Given a state in which quantities all have interval values, often a large number of successor states result if the quantities do not correspond in certain way. The reason is that for each quantity it is possible for it to either change or remain the same. Consequently, there is a successor state for each combination of changing or not changing quantities. The number of combinations for such binary variables is 2^n, however the combination in which no quantities change is not a successor but the state itself. As such, for a single state the number of successors s given a number on non-corresponding changing quantities q can be calculated though $s = 2^q - 1$.

Constraining behaviour Given that a model potentially results in an unusable large state graph, it is essential that its behaviour is constrained. Technically all behavioural relationships between quantities constraint behaviour, however correspondences and proportionalities are especially appropriate. Given two non-corresponding quantities, each combination of magnitude values is possible. Adding a correspondence assures that only each corresponding pair of values is possible. Also adding a proportionality removes the potential of the two quantities changing independently of each other (given that there is no other causal dependency on the targeted quantity). This combination of ingredients makes quantities behave equivalently, and thus allows them to be counted as a single quantity for purposes of determining the maximum number of states.

Inequality statements also help constrain the behaviour.[8] For example, specifying that the birth and death rates are above zero when the population size is above zero removes behaviour. For purposes of simulation it might also be insightful to specify fixed values or ranges for quantities. These are modelled by adding new model fragments that indicate that if a specific assumption holds certain (in)equalities hold for quantities. For example, in the R-star model (Nuttle, Bredeweg, and Salles 2005), when the assumption 'Limited resource build-up' holds, the resources available to the plant population are smaller or equal to medium.

Inactive model fragments Modellers frequently ask why certain model fragments do not become active during their simulations. Modellers usually know that the reason is that certain conditions in their model fragment are not fulfilled by the state. However, their real question is how they can determine which conditions are not fulfilled. In many cases we encountered that there is a mismatch between the model fragment and the state (or scenario). For example, the direction of a configuration is reversed. Our advice is to rebuild the state as a scenario and try to run the simulation. Usually the inconsistency is detected in this process. In the other cases the scenario can be changed to determine what the inconsistency is.

No states Beginning modellers often find it difficult to solve simulation results with no states. Having no states always means at least one model fragment was considered. Otherwise the simulation result would consist of at least one state which corresponds to the scenario. The inconsistency is caused either by the scenario and a matching model fragment, or by a combination of matching model fragments. The easiest way to find the inconsistency is by making all model fragments inactive. Consequently, there should be at least one state corresponding to the scenario. Model fragments can be activated one by one to detect the model fragment that causes the inconsistency.

Not all expected states Sometimes simulations do not generate all the expected states. Our advice to improve the model is creating a scenario that corresponds to the expected state. If the simulation results in no states there is an inconsistency that has to be resolved.Then, modellers should create a scenario that represents a predecessor state and determine if both these desired states are generated[9]. By working backwards in this way towards an already generated state allows the desired branch of behaviour to be simulated.

Inconsistencies Inconsistencies are caused by inconsistent inequalities. Determining which inequalities are inconsistent is a difficult issue and a topic on its own. The following is a list of sources of inequalities that should be checked when searching for the reason of an inconsistency.

- Magnitude (or derivative) value assignments in model fragments (or scenarios).
- Inequalities explicitly represented in model fragments or scenarios.
- Inequalities resulting from the calculations of operators (plus or minus). These calculations result in an inequality indicating that a quantity has a certain magnitude (or derivative) greater, smaller or equal to a specific value.
- Value assignments caused by correspondences. When a quantity A has a certain value, the corresponding quantity B also has to have that specific value.
- Value assignments resulting from influence resolution. The result of resolving of influences and proportionalities is a set of value assignments indicating whether quantities are increasing, stable or decreasing.
- Value assignments resulting from advanced (exogenous) quantity behaviour. It is possible to specify advanced quantity behaviour for specific quantities in scenarios. For example, a quantity can increase, change randomly, or move as a sinus. This behaviour sets the derivative of the quantity. It is also possible to generate all the possible magnitudes of a quantity.
- Another source of facts are engine rules. These rules indicate what is possible in a simulation results, and always apply. Engine rules impose these constraints by imposing inequalities. The most important rules to consider are the quantity constraints and the continuity constraints.

---

[8] Modellers should take note that constraints should make sense for a domain perspective. For example, when a heater heats a pan, the heat of the heater cannot be set to stable, as this would make it impossible for the heat flow process to take heat from the heater. To make the stable heat possible, there should be at least another competing process that adds heat to the heater.

[9] Showing termination nodes (potential successor states) is helpful here.

- The quantity constraints simply specifies that each quantity space has to have a value within its quantity space. This is usually represented by two inequalities. The first indicates that it has a value greater or equal to its top value, and the second indicates that it has a value smaller or equal to its top value.
- The continuity constraints is a transition rule that indicates that a magnitude or derivative has to gradually change, e.g. a derivative cannot change from increasing to decreasing without passing though stable. For a derivative this would result in an inequality that indicates that the derivative in smaller or equal to zero when a quantity is decreasing. An example of when the continuity rule can cause conflicts is when one of a pair of opposing influences disappears.
- A special source of inequalities are the simulation preferences. These simulation preferences can be changed in the simulation preferences window. The most notable to consider are the two extreme values rules.
  - The 'Apply quantity space constraints on extreme values' rule indicates that the derivative of quantities has to be smaller or equal to zero (cannot increase) in their top magnitude value (if it is a point), and is greater or equal to zero (cannot decrease) in their bottom magnitude value (if it is a point). This rule applies to all extreme point values except zero.
  - The 'Apply quantity space constrains on zero as extreme value' applies the 'Apply quantity space constrains on extreme values' rule for zero as an extreme point value.

Conclusions and Future Work

This paper identifies frequently occurring model building issues, misconceptions and suboptimal modelling, and provides solutions, patterns and modelling advice. The issues and patterns originate from well-established models made by the groups we have trained over the last few years. We aim to contribute to the building of qualitative models raising awareness about the issues with model builders and providing them with the means to resolve them. In the coming years we will focus on providing better software support on resolving the presented issues and making frequently used patterns easier to represent.

Acknowledgements

The research presented in this paper is co-funded by the EC within FP6 (2002-2006) (project NaturNet-Redime, number 004074, www.NaturNet.org), and FP7 (2007-2013) (project DynaLearn, 231526, www.DynaLearn.eu). We thank our project partners, the students participating in our courses, and our reviewers for their insightful comments.

References

Bredeweg, B.; Bouwer, A.; Jellema, J.; Bertels, D.; Linnebank, F.; and Liem, J. 2006a. Garp3 - a new workbench for qualitative reasoning and modelling. In Bailey-Kellogg, C., and Kuipers, B., eds., *20th International Workshop on Qualitative Reasoning (QR-06)*, 21–28.

Bredeweg, B.; Bouwer, A.; Liem, J.; and Salles, P. 2006b. Curriculum for learning about QR modelling. Technical Report Deliverable D6.9, STREP project FP6 no. 004074, Naturnet-Redime.

Bredeweg, B.; Salles, P.; Bouwer, A.; Liem, J.; Nuttle, T.; Cioaca, E.; Nakova, E.; Noble, R.; Caldas, A. L. R.; Uzunov, Y.; Varadinova, E.; and Zitek, A. 2008. Towards a structured approach to building qualitative reasoning models and simulations. *Ecological Informatics* 3(1):1–12.

Bredeweg, B.; Salles, P.; and Nuttle, T. 2007. Using exogenous quantities in qualitative models about environmental sustainability. *AI Communications* 20(1):49–58.

Cañas, A. J.; Hill, G.; Carff, R.; Suri, N.; Lott, J.; Gómez, G.; Eskridge, T. C.; Arroyo, M.; and Carvajal, R. 2004. Cmaptools: A knowledge modeling and sharing environment. In Cañas, A. J.; Novak, J. D.; and González, F. M., eds., *Concept maps: Theory, methodology, technology. Proceedings of the first international conference on concept mapping*, 125–133.

de Kleer, J., and Brown, J. S. 1984. A qualitative physics based on confluences. *Artificial Intelligence* 24(1-3):7–83.

Forbus, K. D. 1984. Qualitative process theory. *Artificial Intelligence* 24(1-3):85–168.

Knublauch, H.; Fergerson, R.; Noy, N.; and Musen, M. 2004. The protege owl plugin: An open development environment for semantic web applications. In McIlraith, S. A.; Plexousakis, D.; and van Harmelen, F., eds., *International Semantic Web Conference*, 229–243. Hiroshima, Japan: Springer.

Liem, J.; Bouwer, A.; and Bredeweg, B. 2007. Collaborative conceptual modeling: Share, search and reuse. In Price, C., ed., *21st International Workshop on Qualitative Reasoning (QR'07)*, 71–78.

Nuttle, T.; Bredeweg, B.; and Salles, P. 2005. R-star - a qualitative model of plant growth based on exploitation of resources. In Hofbaur, M.; Rinner, B.; and Wotawa, F., eds., *19th International Workshop on Qualitative Reasoning (QR'05)*, 47–53.

Rickel, J., and Porter, B. 1997. Automated modeling of complex systems to answer prediction questions. Technical report.

Salles, P.; Bredeweg, B.; and Bensusan, N. 2006. The ants garden: Qualitative models of complex interactions between populations. *Ecological Modelling* 194(1-3):90–101.

Sànchez-Marrè, M.; Béjar, J.; Comas, J.; Rizzoli, A.; ; and Guariso, G., eds. 2008. *Proceedins of iEMSs 2008: International Congress on Environmental Modelling and Software Integrating Sciences and Information Technology for Environmental Assessment and Decision Making: Workshop 5 - Qualitative models for sustainability.* p1971-2005.

Computing Human-Like Qualitative Topological Relations via Visual Routines

Andrew Lovett **Kenneth Forbus**

Qualitative Reasoning Group, Northwestern University
2133 Sheridan Rd, Evanston, IL 60208 USA
Andrew-Lovett@northwestern.edu
Forbus@northwestern.edu

Abstract

A core problem in spatial reasoning is finding an appropriate set of relationships to compute. This paper proposes that humans represent topological relationships between 2D regions using three basic, qualitative relations: **contains, intersects,** and **overlaps-with**. We show how these relations can be computed from sketched inputs using a model of mid-level perception. Results from a pilot experiment indicate that these three relationships suffice to explain people's judgments on four English spatial terms ("intersects", "overlaps", "connects to", and "contains"), although a combination of the three is generally required for each term.

Introduction

A major problem in building systems that reason about space is determining the correct set of spatial relations to represent. In the QR community, the Region Connection Calculus (RCC8) (Cohn 1996; Cohn et al. 1997; see Figure 1) is a prominent and effective way of representing topological relations between two-dimensional shapes. RCC8 includes 8 qualitative terms which exhaustively describe the set of possible topological relations between two shapes. RCC8 relations have been used in a number of applications, from qualitative spatial simulation (Randell et al. 1992) to sketch understanding (Forbus et al. 2008).

While representational schemes like RCC8 are useful for building formal AI reasoning systems, it is not clear how closely they align with human spatial representations. Reasoning systems which use human-like representations are better equipped for both interacting with humans in cooperative endeavors and modeling human thought processes in cognitive modeling studies. However, there have been few attempts by AI researchers to look at how humans compute and represent topological relations.

In one notable exception from Geographic Information Systems, Xu and Mark (1997) conducted a study in which they showed participants scenes containing pairs of linear objects (such as roads and rivers). Participants were instructed to indicate how well various predicates described the scenes (predicates included "X crosses Y," "X connects with Y," "X merges with Y," etc). By studying their results, the authors were able to get a better idea of the various factors that determined which predicate people might use in describing a geographic scene.

While the Xu and Mark results are helpful, we believe there is a more general question of what are the topological primitives computed and represented by humans when they examine a visual scene. By *primitives*, we mean a small set of relations from which all (or at least most) other topological relations can be computed. These primitives should meet the following requirements:

1) They should be easily computable by humans using low- or mid-level visual operations.
2) They should not be tied to any particular domain, such as geography.
3) While the individual primitives may not correspond to topological terms in the English language, such as "contains" or "intersects with," it should be possible to explain how humans can use the primitives together to compute and assess those terms.

In this paper, we propose that people use three topological primitives for representing two-dimensional visual scenes: **contains**, **intersects**, and **overlaps-with**. We show how these primitives can be computed using *visual routines* (Ullman 1984), a general approach to modeling

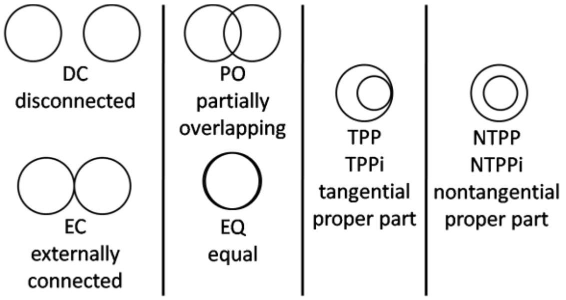


Figure 1. The Region Connection Calculus (RCC8) relations for describing topology. The TPP and NTPP relations each have inverse relations, TPPi and NTPPi.

Operation Type	**Operations**
Covert Attention	Curve Tracing, Scanning, Region Coloring
Working with Elements or Objects	Attribute Access, Activation, Inhibition/Excitation, Deletion
Working with Objects	Object Creation, Binding to Elements
Maintenance	Marking Locations

Table 1. Elementary operations in VRS.

mid-level visual processing. We then evaluate the primitives by examining how well they explain human assessments of four topological terms from English: "connects," "intersects," "overlaps," and "contains." Note that while the primitives and the English terms look quite similar, our results show that there is by no means a one-to-one mapping between primitives and English terms.

We begin by presenting Visual Routines for Sketching (VRS), an implementation of Ullman's visual routines proposal which we are developing. We then summarize the psychological literature on topological relations and show how it motivates the use of our three topological primitives. Then, we describe the visual routines written in VRS to compute our three primitives. After this, we present the results from a preliminary psychological study conducted to evaluate our primitives. We conclude by discussing related and future work.

Visual Routines for Sketching

Ullman (1984) proposed that people have access to a set of *elementary operations*, operations we can run over our visual working memory to extract information. This finite set of operations can be combined in different ways to create a near-infinite set of *visual routines* for computing different spatial features and relations.

We are developing Visual Routines for Sketching (VRS), a computer implementation of visual routines, as a platform for experimenting with computational models of perception. It provides a set of low-level elementary operations, supported by the psychophysics and cognitive psychology literature. Using these operations, researchers can construct visual routines based on their theories for how a particular spatial feature is computed. These routines can be run and evaluated on two-dimensional sketches or line drawings in CogSketch[1] (Forbus et al. 2008), an open-domain sketch understanding system.

CogSketch users can create sketches either by drawing with a pen or by importing shapes built in PowerPoint. VRS works directly with the *ink* of the sketch, the lines representing the edges of each object. Thus, it avoids edge detection issues.

VRS's current vocabulary of operations is given in Table 1. As we describe each of the levels of representation in the system, we refer to operations listed in this table.

Basic Representation

Ullman (1984) suggested that the human perceptual system uses a bottom-up, parallel approach to build an initial *basic representation* of the visual world. VRS computes a basic representation via two steps: First, the ink is projected onto a retinotopic map, a simplification of V1 in the primary visual cortex which represents the orientation of any edges at each location in the image. This produces a set of *edge activations* at various locations.

Second, edge activations are grouped together to form *contours*. This step is based on the contour integration literature (e.g., Yen and Finkel 1998; Li 1998), which suggests that there is a parallel process in which people group edges together based on the Gestalt grouping principles of good continuation and closedness. To these principles we add the hard constraint of uniform connectedness (Palmer and Rock, 1994). That is, edge activations will only be grouped together in a contour if they are the same color and they lie directly adjacent to each other in the visual representation. In the future, we plan to relax the connectedness constraint partially to allow the system fill in gaps between parts of a line (e.g., Saund, 2003).

Incremental Representation

Ullman proposed that there is a set of elementary operations that can be applied serially to the basic representation. By combining these operations into visual routines, an individual can both gather information and update the representation, thus producing an *incremental representation*. In VRS there are three key elementary operations, inspired by Ullman's proposal, which gather data and add *visual elements* to the incremental representation:

1) *Curve Tracing* traces along consecutive edge activations. It produces a *curve*, a new grouping of activations which may lie along one or multiple contours.

2) *Scanning* begins at one location and moves forward in a fixed direction. It produces a straight curve representing the line scanned over.

3) *Region Coloring* fills in the area between curves and contours, creating a new *region*.

All three operations take optional arguments that allow them to be constrained in several ways, e.g., curve tracing along a region, region coloring along a curve, or scanning between two points. The operations can be used to gather information, such as detecting what other elements lie along a curve or within a region.

---

[1] Available for download at: http://silccenter.org/projects/cogsketch_index.html

The visual elements in the incremental representation can be queried via the *Attribute Access* operation to access data such as the size of an element, the center of an element, the curvedness of a curve, or the orientation of a straight curve. Elements can also be *Inhibited*, causing them to be ignored by future operations.

Objects

The *Object Creation* operation sets up *object files* (Kahneman et al., 1992). Object files serve as a bridge between the visual representation and higher-level, conceptual representations. Each object file contains indices (Pylyshyn, 2001), which point to the curves and regions that make up the object in the incremental representation. Because objects can share regions or curves in the incremental representation (as when two shapes overlap), it is possible for multiple object files to point down to the same visual elements in the incremental representation. However, these elements can only point up to one object file at a time. To ensure that a particular object file's visual elements are pointing up to it, a routine must *Activate* that object.

Universal Routine

Different visual routines may be relevant to studying different images. However, there needs to be some type of routine to run on the basic representation and gain enough information to determine what follow-up routine to use. Thus, Ullman suggested that there might be a *universal routine* which is applied by default to visual stimuli. The following is a universal routine written using the elementary operations described above. This routine identifies the objects in a visual scene.

Universal Routine: Finding objects in the visual scene
1) **Region Coloring**: Color the ground, locate any contours in it. 2) **Curve Tracing**: Trace each contour to determine whether it is a closed shape. Produces a curve. 3) **Object Creation**: Make an object file for each curve. 4) **Region Coloring**: If an object is a closed shape, color the area inside it to identify its interior. Produces one or more regions, which will be bound to the object. May also locate new contours located within the object. 5) **Recursion:** For any new contours located, repeat steps 2-5.

Current State of VRS

At present, VRS contains the elementary operations listed in Table 1. However, we are still in the process of determining the full set of operations and the ways they can interact. Eventually, we hope to develop a simple coding language which will allow other researchers to build their own visual routines by combining elementary operations in novel ways.

Psychological Motivation

Much of the psychological work on topological relations has been related to linguistic terms and how they vary across languages and cultures. Landau and Jackendoff (1993) analyzed the full set of spatial prepositions in the English language—several of which describe topological relations—and determined the various factors that determined which preposition is used to describe a scenario. One important factor was distance. Different distances resulted in the use of different prepositions for describing the relative positions of two objects:

Inside: “in,” “inside,” “throughout”
Contact: “on,” “all over”
Proximal: “near,” “all around”
Distal: “far”

Here, both **inside** and **contact** could be seen as topological primitives that determine which preposition should be used. Landau and Jackendoff further found that the preposition used was only rarely affected by the form of the objects being related to each other. They suggested that our mental representations of relative location are separate from our representations of shape and identity, and they predicted that other languages would similarly use spatial prepositions that were not related to the objects' forms.

A number of studies have found fault with this prediction (see Kemmerer 2006 for a review). There are languages that base the preposition used on the form of the objects being related (e.g., relative tightness of an object in a container for Korean: Hespos and Spelke 2004).

However, there may still be some set of domain-independent topological primitives that are universally computed. These primitives might be combined with object shape and object identity in determining which spatial preposition should be used, with the appropriate combinations varying across languages. Levinson and Meira (2003) conducted a survey of nine highly different languages in which speakers of each language were shown the same set of pictures depicting topological relations and asked to describe those pictures. While there were major differences in how each language grouped the pictures, there appeared to be correlations across languages. Multidimensional scaling revealed that many languages group together pictures relating to *in* (e.g., an animal in a cage), *attachment* (clothes on a clothesline), *on/over* (an object on a table), *on-top* (a tablecloth covering a table), and *near/under*. These groups align with the commonly discussed topological concepts of *containment* and *attachment*, and the physical concept of *support*.

While the distinction between these concepts clearly depends upon the forms of the objects being related, and the distinctions tend to be even more fine grained in many languages, it seems reasonable to propose that Landau and Jackendoff's primitives, **inside** and **contact**, likely aid in distinguishing between *containment*, when one object is

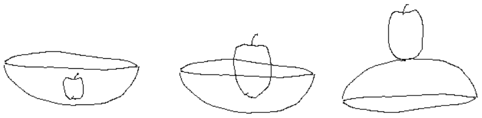


Figure 2. From left to right, the apple is **inside** the bowl, the apple **overlaps** with the bowl, and the apple **intersects** the bowl.

located entirely within another object, and *attachment* or *support*, when the objects are merely touching.

However, we believe that these two primitives are not sufficiently detailed. There are multiple possible forms of contact between two objects in a visual scene. In the simplest form, **intersection**, the edges of the objects simply touch each other in some way. In another form, **overlap**, there is space in the visual scene which is occupied by both the objects. For example, in Figure 2 both the apple inside the bowl and the apple that overlaps with the bowl would be labeled as "in the bowl," whereas the apple that merely intersects the bowl would be labeled as "on the bowl." In this paper, we will be testing the hypothesis that people use both the **intersection** and **overlap** primitives, along with **containment,** to compute and assess topological relations.

The Primitives

We have chosen to use three topological primitives: **contains**, **intersects**, and **overlaps-with**. Each of these primitives describes the location of one object, the *target*, relative to another target, the *ground.* In this section, we describe what these primitives mean and give the visual routines for computing them. All visual routines are computed over objects which can be identified in the visual scene using the universal routine described above.

Intersects

This relationship holds whenever some part of one shape's edge intersects some part of the other shape's edge. The visual routine for computing this is given below.

Intersects (Target, Referent)
1) **Activation**: Activate the Referent object, causing all its associated edges to point up to it. 2) **Curve Tracing**: Trace along the Target object's curve, checking whether any of the Referent object's edges are encountered.

Overlaps-with

This relation is defined only for pairs of closed shapes (although variations might apply to other shape types). Two shapes are overlapping if their interiors share some region. That is, there is some area that lies within both closed shapes. However, the shapes must also both have regions that are not shared: one shape cannot lie entirely inside the other shape. Note that if one shape **overlaps-with** another shape, it necessarily also **intersects** the other shape. The visual routine is given below.

Overlaps-with (Target, Referent)
1) **Attribute Access**: Check whether the Referent and Target objects share any regions. 2) **Attribute Access**: Check whether the Referent has regions not shared by the Target. 3) **Attribute Access**: Check whether the Target has regions not shared by the Referent. 4) **Combine Data**: If the objects share regions but they both have regions not shared with the other, then they overlap.

Contains

This relation is defined only when one object, the referent, is a closed shape. **Contains** holds when the other object, the target, lies entirely within the referent. The visual routine is given below. Note that this routine actually calls the **overlaps** routine.

Contains (Referent, Target)
1) **Activation**: Activate the Target object, causing all its associated edges to point up to it. 2) **Region Coloring:** Color in the Referent's regions, checking to see whether any of the Target's edges lie within the Referent. 3) **Visual Routine Call:** Check whether Overlaps-with(Target, Referent) is false. 4) **Combine Data**: If part of the Target lies within the Referent, and the Target and Referent do not overlap, then the Referent contains the Target.

Relation to RCC8

Recall that RCC8 (Figure 1) consists of six topological relations, plus two inverse relations, whereas our approach uses only three relations. Nonetheless, all of the RCC8 relations except EQ[2] (equal) can be easily computed from our three relations (see Table 2). We believe this supports our argument that our relations are more basic, or more fundamental. In particular, RCC8 distinguishes between "Tangential Proper Part" (TPP) and "Non-Tangential Proper Part" (NTPP). It seems unlikely that humans make this distinction, at least in their initial representations. The more primitive **contains** relationship captures the important commonalities across TPP and NTPP.

Experiment

We conducted a pilot psychological study to evaluate our primitives. In this study, participants saw basic visual

---

[2] It would be relatively straightforward to write a visual routine to compute the **equal** relationship. However, we think it unlikely that humans encode such a relationship, since two objects whose regions and edges are identical will be indistinguishable from each other.

RCC8 Relation	**Primitives to Compute it**
DC	**!intersects ^ !contains**
EC	**intersects ^ !contains ^ !overlaps**
PO	**overlaps**
TPP/TPPi	**contains ^ intersects**
NTPP/NTPPi	**contains ^ !intersects**

Table 2. The topological primitives which can be used to compute seven of the eight RCC8 relations.

scenes consisting of a large red circle and a small green circle (see Figure 3). These scenes were accompanied by a statement such as "Red intersects with green." Participants were instructed to rate the appropriateness of the statement as a description of the scene, using a scale from 0 to 10.

We evaluated our model by examining whether the topological primitives **contains**, **intersects**, and **overlaps-with** could explain individuals' ratings for English terms. We assume that an individual might assess a statement such as "Red overlaps with green" by computing some linear combination of the three primitives.

We hypothesized that, if our model was accurate, it should explain both average and individual performance. That is, (1) For each English term, there should be a set of weights for the primitives that correlates highly with average human ratings for that term. This means that the weights are expressing the degree to which individuals consider each of the primitives *on average* in assessing that term. (2) For each English term and each participant, there should be a set of weights for the primitives that correlates highly with that individual's ratings. This set of weights describes what that particular person considers when assessing the English term. Note that there might be high inter-individual differences in the weights. However, if all individuals are basing their assessments on the primitives, then there should be some appropriate set of weights for all individuals.

Methods

Stimuli consisted of a red circle with radius .5 inches and a green circle with radius .2 inches. There were nine possible distances between the green circle and the red circle, which varied from the two circles being entirely disconnected to the circles overlapping to the green circle being located entirely within the red circle (see Figure 3). There were also four possible directions between the red circle's center and the green circle's center (up, down, left, and right). Thus, there were 36 total images.

Each image was accompanied by one of the following sentences:

"Red intersects green."
"Red overlaps with green."
"Red connects to green."
"Red contains green."

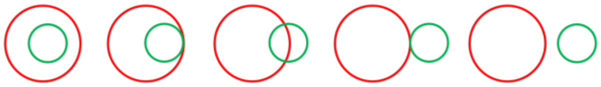


Figure 3. Five of the nine total distances between the large red and small green circles.

A given participant saw each sentence paired with each image, for a total of 36 x 4 = 144 trials. The trials were presented in a random order for each participant.

Participants chose a rating from 0 to 10 for each image/sentence pair by selecting a value from a pop-out menu. Participants were given as much time as they desired to choose the ratings. However, participants chose ratings relatively fast, typically going through the 144 trials in about ten minutes.

The study was run using 10 participants, five male and five female. Nine spoke English as a first language, while the other had learned English at an early age.

Analysis

Our primary question in analyzing the results was whether participants' ratings could be explained using the primitives **contains**, **intersects**, and **overlaps-with**. We used CogSketch and VRS to compute these qualitative relations for each of the 36 images.

In evaluating whether the primitives could explain either average or individual performance, our system performed an exhaustive search for the set of weights for the primitives which maximized the Pearson correlation coefficient with the human data.

Results

Table 3 shows the correlations between the model and human ratings. As the table shows, the model correlated quite high (.98 or above) with the average human ratings for each of the four English terms. The model also correlated well with the ratings of individuals. The median correlations with individuals were all above .9. "Overlaps" was the only term for which any of the individual correlations fell below .85.

	Average	Individual Median	Individual Minimum	Individual Maximum
"Intersects"	.995	.966	.881	.999
"Overlaps"	.994	.932	.790	.999
"Connects"	.993	.95	.850	1.0
"Contains"	.981	.953	.890	.994

Table 3. Correlations between the model and human ratings of the four English terms.

There are at least two alternative explanations for the high performance of the model. One is that there is nothing special about our primitives. Perhaps any three randomly generated factors could correlate highly with human data, after performing an exhaustive search for the optimal set of weights for those factors. The other is that our model does not require all three primitives. Perhaps two of the primitives are doing all the work, and the third primitive is extraneous.

To rule out either of these possibilities, we compared our model against four other possible models (see Table 4). Three were constructed by leaving one of the three primitives out of the model and determining weights for only two primitives. The last model, Random-3 was constructed by building three random primitives (simply by randomly computing a value of *true* or *false* for each primitive's presence in each of the 36 stimuli) and then searching for an optimal set of weights for the three random primitives. Because of the randomness involved, we constructed triplets of random primitives 40 times for each English term and averaged the results.

	C,I,O	C,I	C,O	I,O	Random-3
"Intersects"	.995	.901	.906	.994	.227
"Overlaps"	.994	.781	.956	.934	.213
"Connects"	.993	.980	.475	.993	.189
"Contains"	.981	.917	.981	.068	.217

Table 4. Several models' correlations with the average human ratings. Letters indicate which primitives were used in each model. Random-3 uses three randomly generated factors.

As Table 4 shows, the complete model, C,I,O easily outperforms all other models. The models containing only two primitives typically perform slightly worse for three of the English terms, but each performs significantly worse on at least one term. Thus, clearly all three primitives are required to model the human rating data. The random model performs far worse than any of the other models, indicating that the particular primitives chosen for our model are much better than random factors.

Table 5 gives the optimal weights for each of the three primitives in explaining the average human ratings of the four English terms. As the table shows, there was by no means a one-to-one mapping between primitives and English terms. Participants considered at least two primitives in assessing each of the four English terms. In assessing the trickier "Overlaps" term, they appear to have considered all three primitives, on average.

	Contains	**Intersects**	**Overlaps-with**
"Intersects"	.048	.435	.516
"Overlaps"	.245	.209	.546
"Connects"	.021	.979	- .206
"Contains"	.695	0	.305

Table 5. Optimal weights of the three primitives in explaining the average human ratings of the four English terms.

Figure 4 shows the performance of the model for each individual in greater detail. In addition to showing the correlations, the figure shows the amount of weight given to each primitive by each individual. As the figure shows, there was a great deal of variation across individuals.

Discussion

As the results show, we can vary the weight assigned to each of the three primitives to create models that correlate well with either average or individual assessments of different topological terms. We can also examine the weights to see how different individuals are performing their assessments. For example, participant 7 apparently assessed "Red connects to green" entirely based on the **intersects** primitive (see Figure 4). That is, the participant believed the shapes were "connected" any time their edges intersected. On the other hand, participant 1's model of "connects" had a strongly negative weight for **overlaps-with**. That is, the participant believed the shapes were "connected" when their edges intersected without their areas overlapping.

The results for "contains" were also quite interesting. Participant 3's model of "contains" consists almost entirely of **contains**, indicating that the participant thought one shape contained another when the other shape was located entirely within it. However, other participants' models of "contains" also give some weight to **overlaps-with**. This suggests that when the two circles merely overlapped, many participants believed it was somewhat appropriate to say one shape "contained" the other.

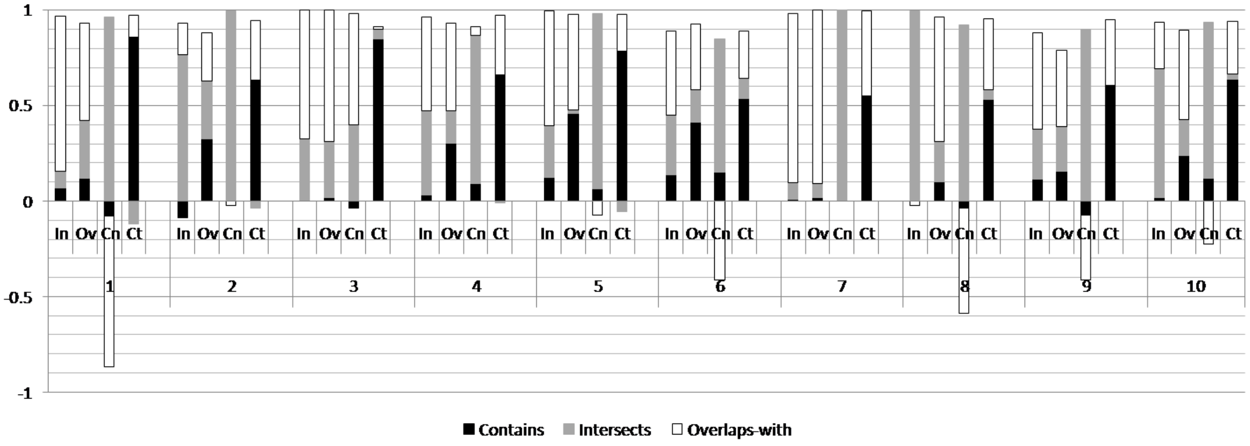


Figure 4. Model correlations with each of the 10 participants. The colors of the bars show the relative weight of each of the three primitives. In = "Intersects" Ov = "Overlaps" Cn = "Connects" Ct = "Contains"

Related Work

Lockwood and colleagues (Lockwood et al., 2006; Lockwood, Lovett, and Forbus, 2008) have used CogSketch and its predecessor along with a model of analogical generalization (Halstead and Forbus, 2005) to automatically learn representations of spatial prepositions like “on” and “in.” They have demonstrated that in both English and Dutch, the topological relation between the figure and ground plays an important role in determining which linguistic term should be used to describe the objects, although other relations like relative position are also important for some terms. They used the full set of RCC8 relations to represent topological relationships in their work.

A number of researchers have built computer models based on the idea of visual routines. However, many of these models are designed only to solve a particular problem (e.g., Chapman, 1992; Horswill, 1995), and thus they miss out on the generality promised by the original idea. Rao (1998) constructed a system for both learning and performing visual routines for solving different spatial problems. However, because his focus was on controlling a robot in the real world, the elementary operations in his system are in many cases more complex and higher-level than the simple operations proposed by Ullman.

Conclusions and Future Work

Thus far, our results support our hypothesis that individuals use three topological primitives in assessing two-dimensional topological relations. However, we have only the tested the hypothesis using a small set of stimuli. In the future, we would like to expand the stimuli set to include a greater range of shapes. In particular, how do individuals assess topological relations between open shapes, e.g., lines, or between one open and one closed shape? We would also like to expand the range of English terms being assessed. However, we suspect that it will be difficult to come up with many more topological terms that can be assessed in a domain-general manner, that is, between abstract shapes. Finally, a more distant goal would be to look at how well the primitives explain topological terms from other languages.

We would also like to assess our hypothesis using a similarity rating task. In such a task, participants would see pairs of stimuli and rate their similarity on a scale from 0 to 10. Previous work has shown that people perceive stimuli as more similar or closer together when they are located in the same qualitative categories (e.g., color names: Winawer et al. 2007; or regions of a room: Newcombe and Liben 1982). Thus, by identifying similarity clusters we can better determine individuals’ qualitative categories.

One long-term question is how our two-dimensional topological primitives relate to topological relations between three-dimensional objects. We suspect that topological relations between real-world objects like those explored cross-culturally by Levinson and Meira (2003) require integrating 2D topological primitives with both 3D depth cues and conceptual information. However, an exploration of the factors used in assessing spatial relations in three-dimensional visual scenes lies outside the scope of the present body of work.

Finally, we plan to continue developing VRS as a testbed for building cognitive models of perceptions. At present, we are concurrently evaluating a model of positional relations with VRS. Eventually, we hope to make VRS publicly available so that other researchers can use it to evaluate their own theories.

Acknowledgements

This work was supported by NSF SLC Grant SBE-0541957, the Spatial Intelligence and Learning Center (SILC). We thank our participants for volunteering to take part in our study.

References

Chapman, D. 1992. Intermediate Vision: Architecture, Implementation, and Use. *Cognitive Science* 16: 491-537.

Cohn, A. 1996. Calculi for Qualitative Spatial Reasoning. In J. Calmut, J. A. Campbell, and J. Pfalzgraph. eds. *Artificial Intelligence and Symbolic Mathematic Computation,* LNCS 1138, 124-143. Springer Verlag.

Cohn, A. G., Bennet, B., Gooday, J., and Gotts, N. M. 1997. Qualitative Spatial Representation and Reasoning with the Region Connection Calculus. *GeoInformatica* 1: 275-316.

Forbus, K., Usher, J., Lovett, A., Lockwood, K., and Wetzel, J. 2008. CogSketch: Open-Domain Sketch Understanding for Cognitive Science Research and for Education. In Proceedings of the Fifth Eurographics Workshop on SBIM.

Halstead, D., and Forbus, K. 2005. Transforming between Propositions and Features: Bridging the Gap. Proceedings of AAAI. Pittsburgh, PA.

Hespos, S. J., and Spelke, E. S. 2004. Conceptual precursors to language. *Nature* 430: 453-456.

Horswill, I. 1995. Visual Routines and Visual Search: A Real-Time Implementation and an Automata-Theoretical Analysis. In Proceedings of *IJCAI ’95*.

Kahneman, D., Treisman, A., and Gibbs, B. J. 1992. The Reviewing of Object Files: Object-Specific Integration of Information. *Cognitive Psychology* 24: 175-219.

Kemmerer, D. 2006. The Semantics of Space: Integrating Linguistic Typology and Cognitive Neuroscience. *Neurolopsychologia* 44: 1607-1621.

Landau, B., and Jackendoff, R. 2003. “What” and “Where” in Spatial Language and Spatial Cognition. *Behavioral and Brain Sciences* 16: 217-265.

Levinson, S., and Meira, S. 1993. ‘Natural Concepts’ in the Spatial Topological Domain—Adpositional Meanings in Crosslinguistic Perspective: An Exercise in Semantic Typology. *Language* 79: 485-516.

Li, Z. 1998. A Neural Model of Contour Integration in the Primary Visual Cortex. *Neural Computation* 10(4): 903-940.

Lockwood, K., Lovett, A., and Forbus, K. 2008. Automatic Classification of Containment and Support Spatial Relations in English and Dutch. In Proceedings of Spatial Cognition.

Lockwood, K., Forbus, K., Halstead, D., and Usher, J. 2006. Automatic Categorization of Spatial Prepositions. In Proceedings of the 28th Annual Conference of the Cognitive Science Society. Vancouver, Canada.

Newcombe, N., and Liben, L. S. 1982. Barrier Effects in the Cognitive Maps of Children and Adults. *Journal of Experimental Child Psychology* 34: 46-58.

Palmer, S., and Rock, I. 1994. Rethinking Perceptual Organization: The Role of Uniform Connectedness. *Psychonomic Bulletin and Review* 1(1): 29-55.

Pylyshyn, Z. W. 2001. Visual Indexes, Preconceptual Objects, and Situated Vision. *Cognition, 80*, 127-158.

Randell, D. A., Cui, Z., and Cohn, A. G. 1992. Naïve Topology: Modelling the Force Pump. In P. Struss and B. Faltings. eds. *Advances in Qualitative Physics*, 177-192. MIT Press.

Rao, S. 1998. Visual Routines and Attention. Ph.D. diss., Dept. of Electrical Engineering and Computer Science, MIT, Cambridge, MA.

Saund, E. 2003. Finding Perceptually Closed Paths in Sketches and Drawings. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25(4): 475-491.

Ullman, S. 1984. Visual Routines. *Cognition* 18: 97-159.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., and Boroditsky, L. 2007. Russian Blues Reveal Effects of Language on Color Discrimination. *Proceedings of the National Academy of Sciences* 104(19): 7780-7785.

Xu, J., and Mark, D. M. 2007. Natural Language Understanding of Spatial Relations between Linear Geographic Objects. *Spatial Cognition & Computation* 7:4, 311-347.

Yen, S. C., and Finkel, L. H. 1998. Extraction of Perceptually Salient Contours by Striate Cortical Networks. *Vision Research* 38: 719-741.

A qualitative model of the salmon life cycle in the context of river rehabilitation

R. A. A. Noble [a, *], B. Bredeweg [b], F. Linnebank [b], P. Salles [c] & I. G. Cowx [a]

[a] The University of Hull International Fisheries Institute, Hull, HU6 7RX UK (r.a.noble@hull.ac.uk and i.g.cowx@hull.ac.uk)
[b] Human Computer Studies, University of Amsterdam, Kruislaan 419 (matrix I), 1098 VA Amsterdam, The Netherlands (B.Bredeweg@uva.nl)
[c] Institute of Biological Sciences, University of Brasilia, Campus Darcy Ribeiro, Brasilia, 70.910-900, Brazil (psalles@unb.br)

Abstract

A qualitative model was developed in Garp3 to capture and formalise knowledge about river rehabilitation and the management of an Atlantic salmon population. The model integrates information about the ecology of the salmon life cycle, the environmental factors that may limit the survival of key life stages and links with human activities such as agriculture, habitat rehabilitation and fishing. The overall aim of the model was to explore the effects of rehabilitation in the context of a complete life cycle scenario. The scenarios and simulations produced were able to explore these processes in the context of a complete life cycle, but at this scale the simulations were time consuming. Therefore, in addition to these scenarios a series of smaller demonstrator scenarios were developed that succinctly explored individual concepts within the system.

Introduction

River rehabilitation projects often target economically valuable and/or threatened fish species (e.g. Atlantic salmon, *Salmo salar* L.). Conservation of these species is often based around quantitative life-cycle models (e.g. Faivre *et al.* 1997, Aprahamian, Wyatt & Shields 2006) that examine the recruitment of individuals to each consecutive life stage to either identify the factors that impinge on the size of the population, or to set targets for conservation (Hendry *et al.* 2007, Milner *et al.* 2000). Hence, planning of rehabilitation activities often focus on the key human activities that impact on different life stages of the fish populations/community (Cowx & Welcome 1998). As such, models that are able to integrate concepts in ecology, river rehabilitation and socio-economic elements, could be useful for knowledge communication of the requirements for rehabilitation and the potential outcomes of measures. However, quantitative information concerning the effects of rehabilitation measures is often incomplete and difficult to predict (Cowx & Gerdeaux 2004, Cowx & Van Zyll de Jong 2004).

Computer-based Artificial-Intelligence (AI) approaches have been promoted for use in conceptualising and integrating qualitative and incomplete information in ecology and natural resource management (Rykiel 1989). Qualitative Reasoning (QR) modelling is an example of an AI approach that has been promoted for use in modelling ecological systems. (Salles & Bredeweg 2006 and Salles *et al.* 2006a,b) because much ecological knowledge is incomplete, uncertain, qualitative and fuzzy, expressed verbally and diagrammatically, making analytical or numerical solutions difficult or impossible to achieve (Rykiel 1989). For example, QR modelling has been previously used to examine the functioning of Atlantic salmon redds (spawning “nests”) to model the factors and processes that control mortality at this critical life stage for recruitment success (Guerrin & Dumas 2001a, b). In addition, Tetzlaff *et al.* (2008) highlighted the need for transferable tools in catchment based hydrological modelling that conceptualize system behaviour by integrating theoretical perspectives and empirical studies.

The model developed here followed the compositional modelling approach (Bredeweg *et al.* 2008, Falkenhainer & Forbus 1991) using the Garp3 software. The ultimate aim was to simulate the whole life cycle scenario by considering each individual life stage in the salmon life history and the influence of human activities on the particular river/habitats they occupy. A compositional approach to scenario building was also used in the modelling process to test specific model fragments and to act as final scenarios within the model to demonstrate specific concepts.

Life Cycle Concepts

Salmon life history

Atlantic salmon exhibits an anadromous life history. The fundamentals of anadromy are that spawning and early development occurs within freshwater habitats whilst adult growth occurs in the marine environment (Figure 1). Returning adult salmon migrate to the upper reaches of their natal rivers to spawn, cutting redds (nests) in coarse gravel substrate to provide protection and adequate flow through of clean water and oxygen to the fertilised eggs. Eggs and early larval stages occupy these interstitial habitats until they develop to juveniles and emerge from the gravels to occupy riffle/pool habitats. After two to four years in fresh water, maturing juvenile salmon undergo physiological changes, which allow them to tolerate saline water and prompts their migration, as smolts, to sea. Given this, the model considers four key stages in the life history of the salmon in rivers: within-gravel phase (eggs); juvenile phase; smolt phase and adult phase (Mills 1989, Crisp 1993, Crisp 2000).

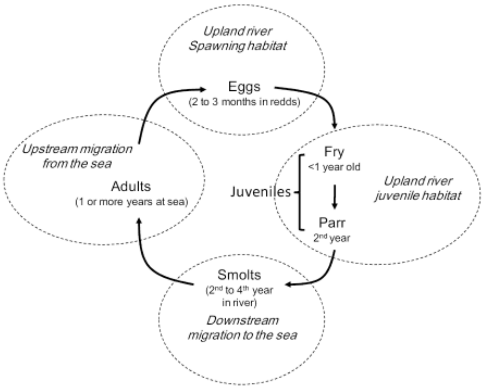


Figure 1 A schematic representation of the key life stages, behaviours and habitats involved in the life cycle of the Atlantic salmon.

Life stages

The key concepts within the life cycle are the different life stages and their survival from one life stage to the next. Therefore, survival is the fundamental process represented by the system. In this context each stage is considered to be an independent (sub) population within the model. This allowed simple model fragments to be developed that apply to all life stages. The basic model fragments "Population" describe that populations are entities that are characterised by the quantities *Recruitment* and *Survival* (Figure 2). This representation allows the modelling of the survival process within a life stage, denoting the numbers that start in the life stage (*Recruitment*) and the numbers that survive to the next life stage (*Survival*). In all cases, both these quantities were represented using the same quantity space (QS): {Zero, Low, Medium, High, Max}. The implementation of ordinal quantity spaces for the number of individuals in each life stage gave a semi-quantitative aspect to the model enabling greater levels of understanding and interpretation of behaviours. The values chosen were designed to be easily understood and give information pertaining to the population/conservation status of the life stage/population as a whole.

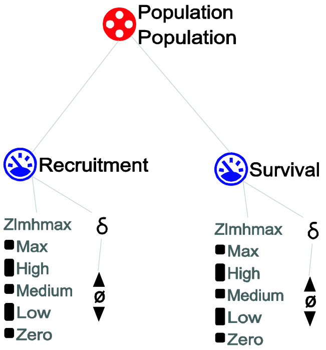


Figure 2 Model fragment "Population" describing the *Population* entity and qualities of *Recruitment* and *Survival*. Each quantity has a QS of {zero, low, medium, high, max}. Each quantity also has a derivative quantity space, denoted by δ; increasing (▲), steady (ø) or decreasing (▼).

Survival and recruitment

Within the salmon life history, the transition of individuals from one life stage to the next (hereafter termed recruitment) is governed by a combination of processes relating to growth, survival and maturation. The number of individuals of each life stage in a salmon population decreases from eggs through the juvenile and sub-adult stages to adults due to factors influencing mortality (e.g. predation, food availability, habitat quality, individual viability and exploitation) (Mills 1989, Crisp 1993, Crisp 2000). In general, fish life histories are typified by adult populations that deposit large numbers of eggs, which are subject to very high mortality in early life stages. Indeed reported values of survival from egg to smolt are around 2-4% (Aprahamian *et al.* 2006).

Most models used to assess the status of salmon populations use life-history models to determine the numbers of spawning adults required to maintain the population given the impact of mortality on different life stages (Aprahamian *et al.* 2006, Milner *et al.* 2000). This is enabled by the relatively distinct life stages and because they either occupy relatively distinct habitats or undergo specific migrations that are themselves potentially characterised by discrete sources/causes of mortality. This model implements this in a qualitative manner.

Although *Recruitment* and *Survival* for all life stages have the same quantity space representation, and hence qualitative equality, this does not necessarily represent quantitative equality. In the case of recruitment quantitative equality is assumed given the instantaneous transition from *Survival* of life stage n to the *Recruitment* to life stage n+1. Within a life stage there is only qualitative equality between *Recruitment* and *Survival* given that due to mortality the number surviving a life stage is always much less than the number at the start of the life stage. However, the qualitative equality of the quantity spaces within a life stage is used to represent the concept that, even though the actual number of individuals in a life stage may be far less than the numbers in the preceding life stage, the numbers in that succeeding life stage can still be considered high or low for that life stage. Therefore, this QS model, in which quantitative and qualitative equality between the QS depended on the concept, was implemented to give some semi-quantitative information without potentially increasing complexity in a model that had inherent complexity due to the number of life stages considered.

Whilst recruitment, mortality and survival are inherently linked (for example successful recruitment to the next life stage is defined by survival through a life stage), the use of sub-populations for each life stage necessitated the isolation of survival and recruitment within the representation. Therefore, recruitment was represented as the process linking one life stage to the next (Figure 3).

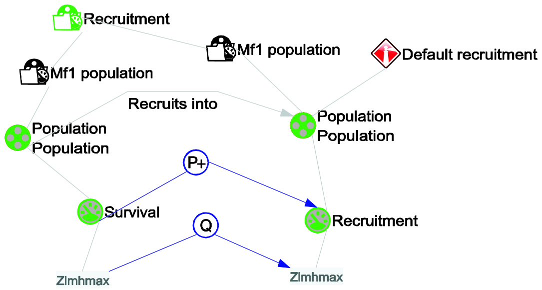


Figure 3 Model fragment “Default recruitment” describing the general recruitment relationship between life stages. In this representation each life stage is modelled as an individual population.

In general this was represented as a simple correspondence (Q) and positive proportionality (P+) between the *Survival* in life stage n to the *Recruitment* in life stage n+1. This survival/recruitment relationship between life stages was represented in different ways depending on the life stages and scenarios being considered. This was implemented using related model fragments, made independent using assumption labels related to the *Population* entity. The assumption “*Default recruitment*” implemented the strict correspondence (Q) and positive proportionality (P+) between the *Survival* in life stage n to the *Recruitment* in life stage n+1 (Figure 3). The assumption “*Spawning recruitment*” implemented a less strict interpretation of this relationship just using a proportionality P+ [*Recruitment* egg, *Survival* adult], zero-zero/max-max value correspondences (V) between their quantity spaces and an equality statement determining that the *Recruitment* of eggs must be greater than or equal to the *Survival* of adults. This denoted the possibility that adults have a high fecundity that may give the potential for the spawning event to regenerate a population and result in a relatively higher number of individuals than the initial number of adults present.

Factors limiting survival

The representation of the within life stage survival modelled the concept that the numbers surviving a life stage is determined by a combination of the starting size of the population (*Recruitment*) and the level of mortality during the life stage. Given the purpose of the model was to represent the effects of human activities on salmon populations, and the fact that human activities generally act through impacts of the habitat (or water) quality within a river, then the number potentially surviving a life stage can be limited by both the level of recruitment and the quality of the habitat they inhabit (Mills 1989, Crisp 1993, Crisp 2000). This representation contains concepts that are similar to the context of carrying capacity in ecological systems. As such the representation considers three basic situations. Firstly, the number recruited is less than the habitat quality and the population is below that which the habitat could support and hence the number surviving is limited by the number recruited (carrying capacity exceeds recruitment) (Figure 4). Secondly, the number recruited exceeds the habitat quality and the numbers surviving is limited by the higher mortality induced by low habitat quality and hence the population is limited by the habitat available (recruitment exceeds carrying capacity). Thirdly, the number recruited and the habitat quality are in balance and the number surviving is limited by both and no increase in recruitment or habitat quality would improve the numbers surviving (system is in balance with carrying capacity). This was modelled using a conceptual quantity, *Potential*, which is a combination of the *Recruitment* and the *Habitat quality*. These two limiting factors act through the *Potential*, which can be viewed as the maximum size limit of the *Survival* in any situation. This was implemented in the model using complex value correspondences (Q) and proportionalities (P+) between the controlling variables and the *Potential* (where the controlling quantity, either *Recruitment* or *Habitat quality*, was the quantity with the lesser magnitude). This necessitated three model fragments where 1) *Habitat quality* > *Recruitment*, 2) *Habitat quality* < *Recruitment*, and 3) *Habitat quality* = *Recruitment*.

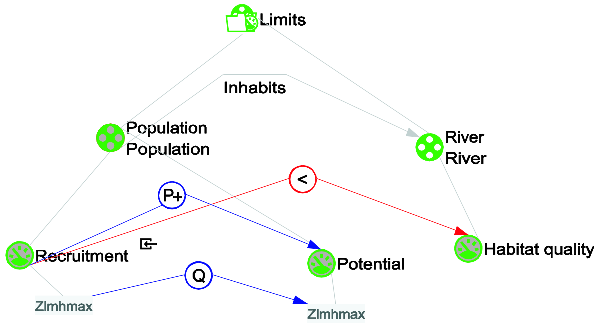


Figure 4 Model fragment “Recruitment limiting” describing the information used to define the value of *Potential* when *Recruitment* < *Habitat quality*.

Recruitment and mortality

The *Survival* is limited by the *Potential* and changes in response to being >, < or = to the *Potential*. The regulation in the *Survival* (due to an imbalance with *Potential*) conceptually results from changes in the balance of the level of recruitment and the mortality/survival rates. In situations where the numbers surviving is less than the potential the numbers surviving can increase due to the effect of recruitment exceeding that of mortality. Conversely, when the *Survival* is greater than the *Potential* the *Survival* decreases due to the effects of higher mortality exceeding the effects of recruitment. To minimise complexity in the model, the net effect of this was modelled as a single abstract quantity, the *Difference* (with QS {extreme min, minus, zero, plus, extreme plus}), which itself was derived as a calculus (Figure 5):

$$\mathit{Difference} = \mathit{Potential} - \mathit{Survival}$$

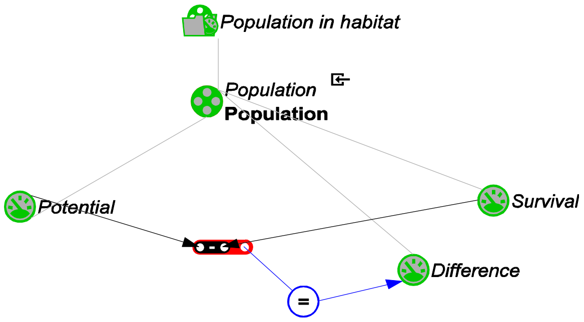


Figure 5 Model fragments “Population difference” that describe the calculation of the *Difference* value which controls the *Survival* of a life stage in relation to the *Potential*.

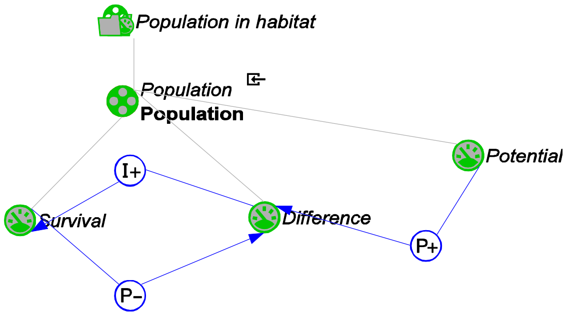


Figure 6 Model fragments “Difference regulates Survival” that describe the relationship I+ [*Survival*, *Difference*] which causes the *Survival* value to increase or decrease towards becoming equal to the *Potential*.

The effect of *Difference* on *Survival* is then modelled as a dependency I+ [*Survival*, *Difference*] (Figure 6). Feedbacks in the calculus are also modelled as P+ [*Difference*, *Potential*] and P- [*Difference*, *Survival*] to determine how the value of *Difference* changes with dynamic behaviours in *Potential* (from the behaviour of *Recruitment* and *Habitat quality*) and *Survival* (caused by the dependency from *Difference*).

Habitat Quality and Human Activities

Catchment concepts

Rivers can be seen as a habitat that integrates a number of physical processes that occur within the catchment of a river (e.g. catchment drainage) and, as such, the quality of a river can be integrated from the quality of these catchment characteristics/processes (e.g. Tetzlaff *et al.* 2007). This is a paradigm within fisheries management that recognises the effects of human activities, such as forestry, agriculture and urbanisation on the quality of the riverine environment (Collares-Pereira & Cowx 2004, Cowx & Welcomme 1998, Cowx 1994). This link is represented in the model by the conceptual chain of reasoning that human activities in a catchment can impact on natural catchment processes; these then impact on some specific quality of the catchment that reduces the integrity of the catchment. This reduced integrity then has an impact on the quality of a specific habitat within a river. This simple conceptual chain, linking both specific factors and conceptual quantities (e.g. catchment integrity) allowed a common approach to modelling different human activities and their effects on different habitats and life stages.

Human influences over habitat

Within the model, humans and human activities were modelled using the notion of “Agent” fragments, which in Garp3 model information about elements of the model which are defined as “external impact”. This gave an explicit representation of humans as agents having an effect on the river/salmon life cycle that was external to the fundamental ecological system being modelled. The chain of reasoning from human agents to catchment integrity through to river habitat quality was modelled using two main groups of model fragments. Firstly, each individual human activity was modelled in a specific Agent model fragment that represented the link between the intensity of the human activity, the quality of the specific catchment characteristic and the catchment’s integrity. Secondly, a general static fragment described the link between the catchment integrity and the river’s habitat quality (Figure 7).

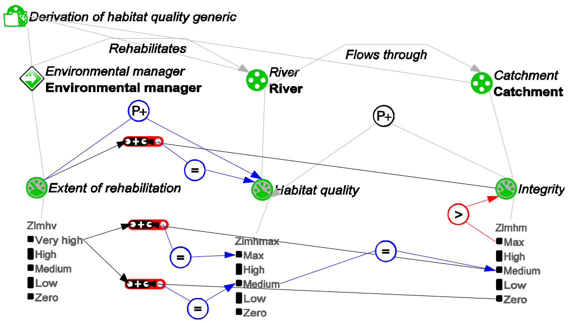


Figure 7 Model fragment "Derivation of habitat quality - generic" describing determination of the magnitude and derivative of *Habitat quality* based on the values and behaviour of catchment *Integrity* and the *Extent of rehabilitation* undertaken by an agent *Environmental manager*.

In each situation the Habitat quality was determined as a calculus between the catchment integrity and the extent of rehabilitation undertaken by an environmental manager (Agent). These model fragments represented the concepts that if catchment integrity was max (the highest value in the QS, equivalent to zero human impacts) then habitat quality was max and that when integrity was less than max then habitat quality could be improved by rehabilitation. Specific calculus statements were made to determine that whilst rehabilitation adds to habitat quality the total effect of rehabilitation may be limited to improving habitat quality through only one quantity space interval.

Modelling issues and solutions

Control of the *Potential*

Whilst the control of *Potential* using directed correspondences and proportionalities (Q and P+ [*Potential*, *Habitat quality*] or Q and P+ [*Potential*, *Number recruited*]) is a simple and successful representation of the system when *Habitat quality* and *Number recruited* are unequal (e.g. Figure 4), difficulties were observed when *Habitat quality* was equal to *Number recruited*. This was especially the case when these two controlling factors had differing derivative behaviours (the values were moving in opposite directions). Essentially, these were dynamic situations where at some point both *Potential* was limited by both *Habitat quality* and *Recruitment* and either one or both of these controlling variables was changing so that one of the variables then became the single controlling factor, e.g. *Potential* switches from being determined by the recruitment to being determined by the habitat quality. In this situation, reasoning produced behaviour paths that terminated in states when the reasoning engine had insufficient information to make suitable influence resolution or the next state would be inconsistent and contain conflicting information. This related to reasoning paths that required *Potential* to change derivative and/or value in an inconsistent way, needing to switch derivative behaviour without first attaining a steady derivative (a behaviour which is terminated by Garp3 as being inconsistent with logical reasoning). To continue representing the system using P+ proportionalities and determining quantity values using directed correspondences a suite of 9 model fragments was developed to control reasoning in situations when *Habitat quality* was equal to *Number recruited*. These 9 fragments (summarised in Table 1) were implemented to consider all 9 possible conditions considering the derivative behaviours of *Habitat quality* and *Recruitment*. In each of these model fragments the consequences for the derivative of *Potential* was determined, together with which factor controlled the value of *Potential* through a directed correspondence (Q). The exclusion of P+ proportionalities and the explicit statement of the resulting behaviour of *Potential* when *Habitat quality* equalled *Recruitment* simplified the reasoning to give the explicitly desired consistent and logical behaviours for further, more complicated, scenarios.

Table 1 Definitions of correspondences (Q) and derivatives (δ) used to define the conditions and consequences in the 9 model fragments used to define *Potential* (P) in the different conditions for the combination of derivatives for *Habitat quality* (Hq) and *Recruitment* (R) when those two quantities are equal.

		Recruitment (R)		
Habitat quality (Hq)	Derivative conditions	δ+	δø	δ-
	δ+	Q [P, R]; Q [P, Hq] P δ+	Q [P, R] P δø	Q [P, R] P δø
	δø	Q [P, Hq] P δø	Q [P, R]; Q [P, Hq] P δø	Q [P, R] P δø
	δ-	Q [P, Hq] P δø	Q [P, Hq] P δø	Q [P, R]; Q [P, Hq] P δ-

Derivative behaviour of *Difference*

Interrogation of the behaviour paths and dependency diagrams generated by Garp3 during the model development indicated inconsistent behaviour relating to the derivatives (δ) of *Difference* when both *Recruitment* and *Habitat quality* resulted in a dynamic behaviour of *Potential*. In particular the inconsistent behaviours were caused in situations when either:

Potential > *Survival* (i.e. *Difference* is plus), δ *Potential* is plus and is bigger than δ *Survival*, which is also plus (due to I+ from *Difference*) OR
Potential < *Survival* (i.e. *Difference* is minus), δ *Potential* is minus and is less than δ *Survival*, which is also minus (due to I+ from *Difference*).

In these situations the result is that *Difference* is either 1) plus and increasing or 2) minus and decreasing. The behaviour paths in this situation become inconsistent in a situation where the derivative of *Potential* becomes steady. In this state the configurations of model fragments indicate that in:

Situation (1) *Difference* should be plus and decreasing (as the difference between *Potential* and *Survival* is now getting smaller because the value of *Potential* is steady and the value of *Survival* is increasing due to the I+ from *Difference*), and in;

Situation (2) *Difference* should be minus and increasing.

In both cases this is an inconsistent behaviour as logically the derivative of *Difference* must pass through a zero derivative (δØ) to move from increase (δ+) to decrease (δ-) or vice versa. These inconsistent behaviours relate to problems in modelling simplistic qualitative calculus of the form:

Potential - Survival = Difference

Potential and *Survival* have dynamic behaviours, especially as in this case where the relationship I+ [*Survival*, *Difference*] gives complex derivative behaviours to both *Difference* and *Survival*. Current modelling in Garp3 only allows modelling with primary derivative information, although to model this calculus behaviour requires information concerning secondary derivatives to produce consistent transitions for the primary derivative of *Difference*. One solution to this problem was to model the quantity space of *Difference* using extreme point values (extreme minus and extreme plus) and then restrict the model simulation to allow the value of *Difference* to change derivative only in the point values rather than in intervals. For example, in situation (1) this allows the value of *Difference* to go from Plus (δ+) to extreme plus (δØ) and then to extreme plus (δ-) to complete a consistent change in derivative behaviour. This modelling approach can be seen as a fix in a situation where information about secondary derivative behaviour is explicitly required.

Scenarios and behaviours

Simple concept scenarios

The compositional modelling approach used by Garp3 allows for scenarios to be built with different levels of complexity exploring either a specific component of the system in question (hereafter referred to as "concept scenarios") or the system as a whole (the full life cycle in this model). The use of many diverse concept scenarios provides a basis both for building and testing model fragments during the model building process and for exploring important concepts and behaviours within sub-components of the system once the total model is implemented. The use of such an approach, providing "building blocks" that go towards explaining the overall life cycle scenarios, is almost certainly an important step in educational settings to aid interpretation of such a large system that may at first seem complicated and daunting to explore.

The concept scenarios were controlled using exogenous derivative behaviours which can be assigned to any quantity in Garp3 (Bredeweg *et al.* 2007). These exogenous controls (indicated by "*!*" next to the quantity under exogenous control in the scenario diagrams (see Figure 8)) can be used to trigger simulations and behaviours in isolated components of the system or to trigger simulations for scenarios considering the whole life cycle.

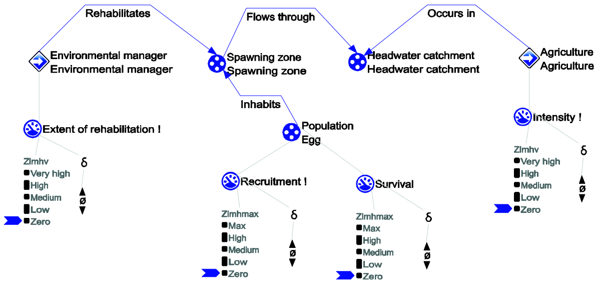


Figure 8 An example concept scenario considering the effect of degradation of spawning habitats due to agriculture on a recovering egg population. Exogenous controls (*!*) are applied to *Recruitment*, *Intensity* and *Extent of rehabilitation* to generate behaviours during a simulation.

Although the concept scenarios only considered a small component of the overall system, the simulations and behaviours they could produce were still large (essentially related to the number of quantities considered and the range of values in their QS). This is due to the potential for the reasoning engine to consider all possible orderings of potential changes in the values of dynamic quantities and produce different behaviour paths accordingly. To reduce this potential complexity the "fastest path heuristic" option in the Garp3 simulation settings was used. Essentially this option allows the reasoning engine to consider that "if a quantity can change value in the next step it will" and as such all quantities that can change value do so in the same reasoning step instead of the engine considering all possible sequences and ordering of quantity value changes. As such, although this option may remove some potential behaviour, it produces simulations of a smaller more manageable size that retain the key behaviours of interest.

Example concept scenario

An example concept scenario detailing the effect of agricultural impacts on the quality of spawning habitat is shown in Figures 8 to 11. This scenario is designed to explore the effects agricultural practises can have on sedimentation processes in a catchment and the amount of fine sediments that enter an upland river reducing its suitability as a spawning habitat (Soulsby *et al*. 2007, Crisp 2000, Crisp 1993, Mills 1989). The outputs of the simulation include the initial scenario (Figure 8) and exogenous controls, the causal model (available for each state transition, Figure 9) behind the behaviour/simulation (Figure 10) and the value history of states and behaviour paths (Figure 11). In the scenario described here, a single egg population inhabits a spawning habitat that occurs within a catchment. Initial value and exogenous behaviour statements are made to determine that the extent of rehabilitation is zero and unchanging, the egg population has zero recruitment (although it is increasing through an exogenous control), zero survival and that the intensity of agriculture in the catchment is initially zero but increasing through an exogenous behaviour. This scenario represents a system with an initially pristine habitat but without a population of eggs (Figure 8).

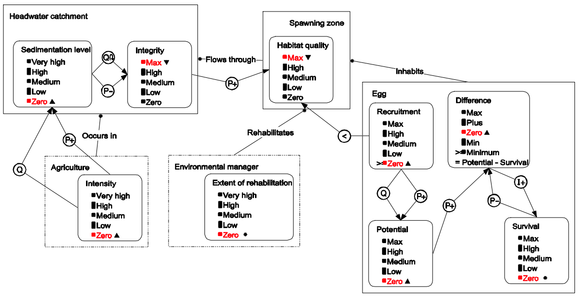


Figure 9 Example causal model indicating both the current state and what is causing the system to change. In this example the exogenous increase in *Intensity* of agriculture (zero ▲), is propagating through the system (P+) causing increase in *Sedimentation level* in the *Headwater catchment* and decreases in *Integrity* and *Habitat quality* (both Max ▼).

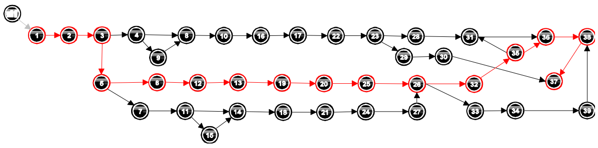


Figure 10 Simulation behaviour paths for the concept scenario (Figure 8). The full simulation for the scenario generated a total of 39 states and one end state, state [37] (a behaviour path [1 → 2 → 3 → 5 → 6 → 12 → 13 → 19 → 20 → 25 → 26 → 32 → 35 → 36 → 38 → 37] is highlighted).

The exogenous behaviour of recruitment represents the creation and establishment of a population although this is happening at the same time as the intensity of agriculture increases causing an impact on the quality of the spawning habitat. The simulation identified a behaviour comprising 39 states and one possible end state (state 37) with a number of possible behaviour paths to the end state (Figure 10). In this case all behaviours include an initial increase in the survival of eggs (due to the increase in *Potential* caused by its link to the increase in *Recruitment*) followed by a period of decline (due to the switch in the potential when it becomes controlled by the declining *Habitat quality*) and then a final state of zero *Survival* when *Habitat quality* becomes zero (Figure 11). In this simulation the different behaviour paths are caused by the potentially different rates in the exogenous derivatives of *Recruitment* and *Habitat quality* and the possibility of *Survival* reaching the low or high interval before the switch in the population behaviour.

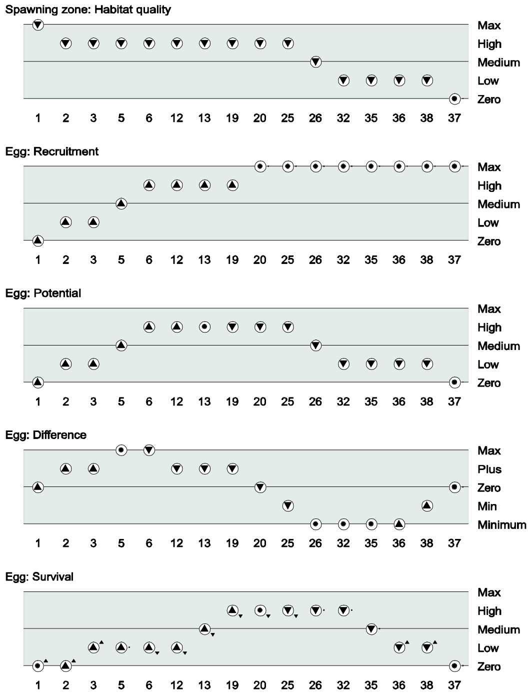


Figure 11 Value history for the behaviour path for the simulation (Figure 10) of a concept scenario (Figure 8).

Life cycle scenarios

The overall aim of this model was to implement scenarios that integrated all the concepts explored in the simple concept demonstrators into single life cycle models that explored human impacts at the whole population level and the link to basic socio-economic elements linked to the system (e.g. costs of rehabilitation). An example of such a scenario is given in Figure 12, which considers the re-establishment of salmon population that previously became extinct, a common situation in systems that have been heavily impacted by agricultural and industrial activities. In this scenario salmon are reintroduced through stocking and the impacts of high intensity agriculture in the catchment surrounding the spawning habitats of the river is rehabilitated. The scenario considers the consequences of these actions for the salmon population.

Unfortunately, the amount of logical reasoning processing required to run a full simulation successfully for such a complex model often took a long time on a desktop PC. The amount of time to obtain simulations from these life cycle scenarios ranged from a couple of hours to a couple of days depending on the contents. This was despite the use of the fastest path heuristic and successfully simulated life cycle scenarios often not generating a huge number of states (for example the simulation above only generated 136 states). This long processing time limited the use and development of this type of cyclical scenario.

Discussion

Design of Quantity spaces – semi-quantitative models

The model presented here aimed to capture and formalise domain knowledge concerning the salmon life cycle and river rehabilitation for use to enhance education about sustainability issues. As such it was developed from knowledge that has been obtained from both qualitative and quantitative sources. For such a model to be easily understood and interpreted within the Garp3 software this qualitative and quantitative information was fused into information concerning quantities and quantity spaces that inherently became semi-quantitative with QS for the main entities and quantities that are ordinal in nature and reflect some key values in the system. This approach is used in QR models to aid their interpretation beyond the basic qualitative concepts of zero; > or < zero; <, > or = and increasing, decreasing or steady. This is achieved using QS with a number of interval and point values that reflect key values and thresholds within the system of study. Whilst this approach is common and fairly straightforward in physical systems, it is less common and less easy to implement for ecological models. For example the ecological models published using Garp3 (e.g. Salles *et al.* 2006a,b, Salles & Bredeweg 2006) have tended to concentrate on exploring and modelling the processes and have used a simple {zero, plus, max} QS to represent the number of individuals in a population. In such systems the interest is generally in whether the population is present/absent or at its maximum and how it is behaving. Further development of QS to include differentiation in the abundance of a population using some key values (e.g. low, medium, high) has had limited use in ecological modelling. In the five state QS used by Salles & Bredeweg (2006) to represent population abundance in a model to explore succession processes in Cerrado vegetation, the max point value in a QS was used as a landmark and related to the concept of carrying capacity. Whereas, in this model carrying capacity was not represented as a fixed point in the QS but could occur at any value at which the *Number surviving* was in balance with the *Potential*. Hence it can be considered that the max point in the QS only reflects the carrying capacity of pristine habitats.

The use of detailed QS has potential to allow for quicker and easier interpretation of a simulation through interrogation of the value history alone, whereas simple QS requires close interrogation of the equation history, something that may be harder for inexperienced learners to comprehend. However, the use of detailed QS does make modelling and model reasoning more complicated, resulting in more complex simulations and larger behaviour paths to represent the semi-quantitative knowledge. The choice of QS used here {zero, low, medium, high, max} only reflects four states of interest; that of zero, low, high and max. The medium point merely reflects an instantaneous transition from low to high. As such it has no real interpretation value to the model in itself. However, the medium point is very important for model development and was fundamental in the implementation of calculus and value correspondences. The difficulties identified in the model implementation together with the solutions used, highlight that determination of QS and model complexity is a fundamental issue in qualitative modelling in ecology.

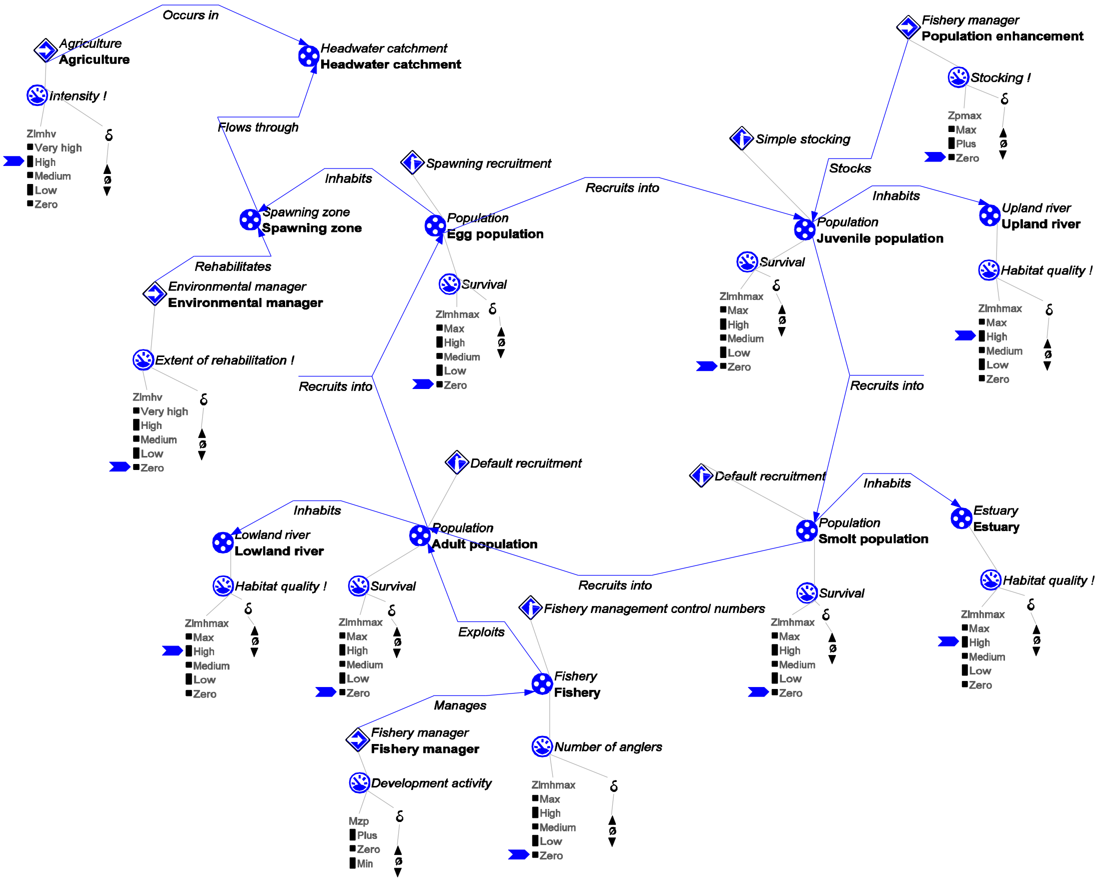


Figure 12 An example of a life cycle scenario detailing the sequence of life stages in the salmon life history that are included in the model and their relationship with specific river habitats and human activities (agents). Scenarios contain details of the entities involved, their structural relations (configurations), the starting values of some of their associated quantities (blue arrows) and modelling assumptions that apply to entities (e.g. Default recruitment). The scenario may also contain exogenous controls (*!*) over some of the quantities. In this scenario the initial values of population size (*Survival* quantity) is set to zero.

Modelling solutions

During model implementation a number of solutions were required to control ambiguity and inconsistency. In some cases these resulted from the choice of the QS used, and the logical conditions and consequences under which processes were modelled. The model fragments used to make these solutions included these approaches:

- a suite of model fragments that fully describe and limit possible behaviours through limiting consequences to conditions;
- a model fragment that removes a possible behaviour.

Examples of the first are found in the model fragments that specify the behaviour of *Potential* when it is switching from being controlled by either *Habitat quality* or *Recruitment* to the other. These model fragments make logical statements about the outcomes of situations and thus ensure behaviours in given situations. Examples of the second are two of the model fragments that control the derivative behaviour of *Difference*. The two model fragments make logical statements that when the *Difference* value is minus or plus then the derivative value cannot be stable. This acts to restrict the reversing of the *Difference* derivative to the extreme point values. This was used to overcome the current limitations in Garp3 for modelling information regarding second order derivatives.

Ideally, such solutions should be, either not required or, kept to a minimum and only act in the same way modelling assumptions are currently used. However, as highlighted by Salles *et al.* (2006), ambiguities and inconsistencies will arise in QR modelling in ecology due to the use of incomplete knowledge. Knowledge representation is likely to be even more incomplete when modelling up scales from models concerning fundamental small scale processes to large scale models that represent abstracted versions of fundamental concepts. For example, the model here does not include the whole suite of biological processes that act to control population size (e.g. Salles & Bredeweg 2006, Salles *et al.* 2006) but represents an abstract version designed to capture the key ideas. In these cases it is likely that such studies will deliberately model incomplete knowledge and thus create ambiguities and inconsistencies that need technical modelling solutions to overcome them. Fundamentally, the choice between the level of complexity used and the use of technical fixes to control ambiguities and inconsistencies must depend on the objectives and final use/users of the model, i.e. what level of causal explanation for behaviours is required by the end user and what knowledge the modeller is attempting to formalise and communicate.

Complexity

The concept scenarios developed here showed that for even simple scenarios complex and variable behaviours could be generated, including when the fastest path heuristic and some modelling assumptions and behaviour limiting model fragments were used. The majority of this complexity was because the modelling approach allows some flexibility in the rate of changes for the *Survival* in each life stage relative to the rate of changes in the *Potential* that is generated by the exogenous behaviours of rehabilitation activities, and their effect on habitat quality. This level of ambiguity for one life stage was multiplied when multiple life stages were considered. The life cycle scenarios became very complex and resulted in large and time consuming simulations, even when only a single human activity was active on a single life stage.

Complex models are to a great extent necessary for large and complicated systems. It is likely that such models necessitate a large number of entities and quantities to convey the required information and concepts. It is also likely that the questions asked of such a model will require complex scenarios with a multitude of active interacting entities, or exogenous factors acting on simulations (Bredeweg *et al.* 2007). Furthermore, additional complexity can be introduced by the use of large quantity spaces that may be used to convey semi-quantitative information to aid interpretation or describe critical points. Therefore, QR models can easily become complex. Given this it is likely that modellers will be interested in controlling complexity when dealing with large systems. Developing model components (such as the solutions used here) and processes (such as fastest path heuristic) will allow modellers to control the levels of complexity in the model, which may allow them to generate relatively simple simulations from complex scenarios, i.e. isolating only the key behaviours whilst still retaining all the elements of the system.

Given the current processing requirements of the life cycle scenarios, the use and exploration of such complex cyclical reasoning scenarios is still limited, especially in an educational setting where time may be critical. However, the results obtained here are positive and indicate there is great potential in such cyclical models, especially when the larger scenarios can be associated with smaller concept demonstrators, which allow the larger scenario to be broken down into more manageable components. In addition to this, software and hardware developments allow for faster reasoning and simulation resolution and thus make these sorts of complex models more manageable in the future.

Conclusions

The salmon life-cycle model was easily able to explore scenarios related to single or pairs of life stages. These small concept scenarios are useful to allow model users to explore and understand fundamental parts of the overall system without having to isolate information contained in models simulating the whole system. The complexity of the system limited the exploration of scenarios considering the whole life cycle. Additional complexity resulted from ambiguity and inconsistency in the abstract representation of some of the concepts. These ambiguities and inconsistencies were controlled using a number of modelling solutions. These solutions, together with developing diverse concept scenarios and using some of the newer simulation options in Garp3, provide a basis for modellers begin to handle complexity in large models.

Acknowledgments

This work was funded by the European Commission’s Sixth Framework Programme for Research and Development (project number 004074, Naturnet-Redime - http://www.naturnet.org). Thanks must also go to Dr Tim Nuttle (University of Jena, Germany), Jochem Liem and Anders Bouwer (University of Amsterdam) for their assistance in developing the case study above. The authors would also like acknowledge funding from the DynaLearn Project (EC 7th Framework STREP – http://www.DynaLearn.eu).

This is a working version of a full paper currently submitted to *Ecological Informatics*.

References

Armstrong, J.D., Kemp, P.S., Kennedy, G.J.A., Ladle, M. and Milner, N.J., 2003. Habitat requirements of Atlantic salmon and brown trout in rivers and streams. *Fisheries Research* 62, 143–170.

Aprahamian, M.W., Wyatt, R.J., Shields, B.A., 2006. Use of biological reference points for the conservation of Atlantic salmon, *Salmo salar*, in the River Lune, North West England. *Fisheries Management and Ecology* 13, 21-30.

Bredeweg, B., Bouwer, A., Jellema, J., Bertels, D., Linnebank, F., Liem, J., 2006. Garp3 – *A new Workbench for Qualitative Reasoning and Modelling*. In Bailey-Kellog, C., Kuipers, B. (eds.) *Proceedings of the 20th International Workshop on Qualitative Reasoning*, pp. 21-28 (conference held in Hanover, New Hampshire, US, 10-12 July, 2006).

Bredeweg, B., Forbus, K., 2003. Qualitative modelling in education. *AI Mag*. 24, 35–46.

Bredeweg, B., Salles, P., Bouwer, A., Liem, J., Nuttle, T., Cioaca, E., Nakova, E., Noble, R., Caldas, A.L.R., Uzunov, Y., Varadinova, E., Zitek, A., 2008. Towards a structured approach to building qualitative reasoning models and simulations. *Ecological Informatics* 3, 1-12.

Bredeweg, B., Salles, P., Nuttle, T., 2007. Using exogenous quantities in qualitative models about environmental sustainability. *AI Communications* 20(1), 49-58.

Clark, M.J., 2002. Dealing with uncertainty: adaptive approaches to sustainable river management. *Aquatic Conservation: Marine and Freshwater Ecosystems* 12(4), 347-364.

Collare-Pereira, M.J. & Cowx, I.G., 2004. The role of catchment scale environmental management in freshwater fish conservation. *Fisheries Management and Ecology* 11, 303-312.

Cowx, I.G., (ed.) 1994. *Rehabilitation of Freshwater Fisheries*. Oxford: Fishing News Books, Blackwell Science, 496pp.

Cowx, I.G., Gerdeaux, D., 2004. The effects of fisheries management practises on freshwater ecosystems. *Fisheries Management and Ecology* 11, 145-152.

Cowx, I.G., Van Zyll de Jong, M., 2004. Rehabilitation of freshwater fisheries: tales of the unexpected? *Fisheries Management and Ecology* 11, 243-250.

Cowx, I.G., Welcomme, R.L., (eds.) 1998. *Rehabilitation of rivers for fish*. Oxford: Fishing News Books, Blackwell Science, 260pp.

Crisp, D.T., 1993. The environmental requirements of salmon and trout in fresh water. *Freshwater Forum* 3, 176–202.

Crisp T.J., 2000. *Trout and Salmon. Ecology, Conservation and Rehabilitation*. Blackwell Science, Oxford. 212 pp.

Faivre, R., Dumas, J., Charron, M., Badia, J., Prouzet, P., 1997. *River basin management using a stochastic model of the salmon life cycle*. In: *Proceedings of International Congress of Modeling and Simulation* (MODSIM'97), Hobart, Tasmania, pp. 1536–1541.

Falkenhainer, B., Forbus, K., 1991. Compositional modelling: finding the right model for the job. *Artificial Intelligence* 51, 95 - 143.

Guerrin F., Dumas J., 2001a. Knowledge representation and qualitative simulation of salmon redd functioning. Part I: qualitative modelling and simulation. *Biosystems* 59, 75-84.

Guerrin F., Dumas J., 2001b. Knowledge representation and qualitative simulation of salmon redd functioning. Part II: qualitative model of redds. *Biosystems* 59, 85-108.

Hendry, K., Sambrook, H., Waterfall, R., 2007. Assessment of salmon stocks and the use of management targets; a case study of the River Tamar, England. *Fisheries Management and Ecology* 14, 7-19.

Mills, D., 1989. *Ecology and Management of Atlantic Salmon*. Chapman and Hall, London.

Milner N.J., Davidson I., Wyatt R.J., Aprahamian M.W., 2000. *The application of biological targets to the management of salmon recreational fisheries in England and Wales*. In: I.G. Cowx (ed.) *Management and Ecology of River Fisheries.* Oxford: Fishing News Books, Blackwell Science, pp. 361–372.

Rykiel, E.J., 1989. Artificial-Intelligence and Expert Systems in Ecology and Natural-Resource Management. *Ecological Modelling* 46, 3-8.

Salles, P., Bredeweg., B. 2003. Qualitative reasoning about population and community ecology. *AI Magazine* 24, 77-90.

Salles, P., Bredeweg, B., 2006. Modelling population and community dynamics with qualitative reasoning. *Ecological Modelling* 195 (1-2), 114-128.

Salles, P., Bredeweg, B., Araujo, S., 2006. Qualitative models about stream ecosystem recovery: Exploratory studies. *Ecological Modelling* 194, 80-89.

Salles, P., Bredeweg, B., Bensusan, N., 2006. The ants' garden: qualitative models of complex interactions between populations. *Ecological Modelling* 194 (1-3), 90-101.

Soulsby, C., Youngson, A.F. Moir, H.J., Malcolm, I.A., 2001. Fine sediment influence on salmonid spawning habitat in a lowland agricultural stream: a preliminary assessment. *The Science of the Total Environment* 265, 295-307.

Tetzlaff, D., Malcolm, I.A., Soulsby, C., 2007. Influence of forestry, environmental change and climatic variability on the hydrology, hydrochemistry and residence times of upland catchments. *Journal of Hydrology* 346, 93-111.

Tetzlaff, D., McDonnell, J.J., Uhlenbrook, S., McGuire, K.J., Bogaart, P.W., Naef, F., Baird, A.J., Dunn, S.M., Soulsby, C., 2008. Conceptualizing catchment processes: simply too complex? *Hydrological Processes* 22, 1727-1730.

Application of qualitative reasoning models in the scientific education of deaf students

Paulo Salles[a], Gisele M. Feltrini[b], Isabella G. de Sá[a],
Mônica M.P. Resende[b] and Heloisa Lima-Salles[c]

[a] Institute of Biological Sciences (psalles@unb.br; isabellagontijo@gmail.com)
[b] Graduate program in Science Teaching (gisele_morisson@yahoo.com.br; monicamresende@terra.com.br)
[c] Department of Linguistics, Portuguese and Classical Languages (hsalles@unb.br)
University of Brasilia, Brazil

Abstract

Regarding the education of deaf students (in Brazil), three conditions have to be met in order to bring qualitative reasoning (QR) models into the classroom: (a) a bilingual education should be provided, the Brazilian Sign Language (LIBRAS) being the first and Portuguese the second language; (b) in the absence of scientific vocabulary in LIBRAS, it has to be created; (c) given the aural impairment, which is cognitively compensated through an over-developed visual ability, a visually oriented pedagogy is needed. This paper describes how qualitative reasoning may provide an adequate scenario to create a vocabulary in sign language for representing scientific concepts while offering support for the integration of visually-oriented models and simulations, and written Portuguese in educational activities.

Key words: qualitative models, deaf, science education

1. Introduction

The Brazilian educational system is nowadays faced with the legal determination of promoting the education of deaf students along with hearing students in the so-called inclusive classrooms. In this context, it is important to understand the requirements for a successful inclusion of the deaf. Previous work [4; 6] has shown that QR models [7] are powerful tools for the education of deaf students, as they have interesting features for accomplishing this task: they articulate knowledge about different physical and social systems in conceptual models, presented with a graphical interface. A concise vocabulary is used to describe the phenomena represented in the models, and a restrict set of modeling primitives is enough to represent a wide class of scientific concepts. Finally, explicit representation of causal relations makes it is possible to ground predictions and explanations about the system behavior. In this context, the present work seeks to answer the following question: *What are the requirements to bring qualitative models into the classroom as useful tools for science education of deaf students?*

2. Sign language representation of QR models

Education is a well established area of application for QR models (Bredeweg and Forbus, 2003). This work explores these models as a tool for acquiring scientific concepts, improvement of linguistic skills and of inferential reasoning, already worked out with deaf students (Lima-Salles *et al.*, 2004; Salles *et al.*, 2005). Two qualitative models were used, ‘tree and shade’, already used and validated in (Lima-Salles *et al.*, 2004), and ‘global warming’, created to be the testbed for this study. The models were built in the QR engine Garp3 (Bredeweg *et al.*, 2006), following the Qualitative Process Theory (Forbus, 1984).

The causality chain shown in Figure 1 reads as follows. With investments, industry produces residues, including greenhouse gases. Besides that, in order to develop agricultural activities, farmers remove natural forest and burn residues of biomass, also releasing greenhouse gases. Both processes positively influence the Gross Domestic Product (GDP), and pollutant concentration influences the Earth temperature. Above a certain threshold, a positive influence establishes

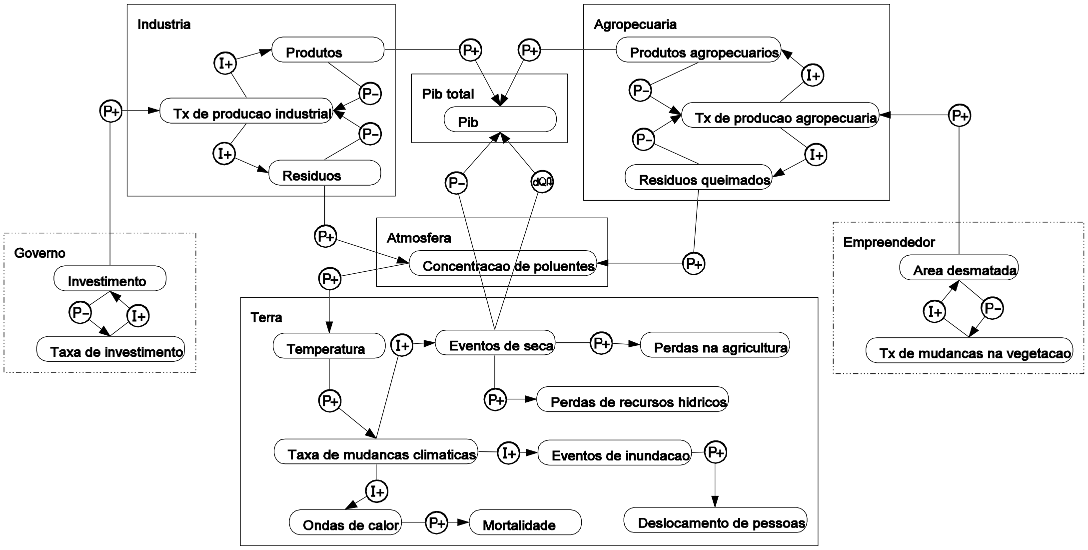


Figure 1. Causal model obtained in a simulation with the 'global warming' model.

a causal link between temperature and climate change rate, triggering the climate change process. The results include events of drought, flood and heat waves, which in turn cause, respectively, losses in agriculture and water resources, population displacement and human mortality. It is assumed that events of drought have a negative influence on GDP. Lack of scientific vocabulary in LIBRAS render difficult the development and understanding concepts by deaf students. This paper proposes a procedure for creating a set of signs to represent scientific concepts based on reusable QR modeling primitives. Creating signs is a complex process, and has to be carried out by the deaf community. First, a deep understanding of the topic to be represented is required. Such understanding has to be achieved at the community level, because a single person cannot impose to the community a sign for a (scientific) concept. In the work described here, a study group of 6 undergraduate deaf students first acquired understanding of models and modeling primitives, and subsequently created the signs. The study group produced a glossary of 32 lexical items in LIBRAS with terms used in qualitative models. Besides that, assignments for the models were created in a collaborative work carried out by the researchers and 8 secondary school teachers during a course on science education.

3. Model and sign validation

The 'global warming' model was conceptually and operationally validated by an expert, 8 secondary school teachers and 4 deaf undergraduate students. The expert concluded that the representation of causality in the model is acceptable, on the basis of (scientific) knowledge available (Rykiel, 1995). The teachers recognized its potential for the development of cognitive competences and abilities in science education. The deaf students were able to explain causal models, in written Portuguese, an evidence that they understood the concepts. Validation of the signs started with the presentation of modeling primitives and models to 17 deaf undergraduate students. Next, answers to a questionnaire and suggestions were collected, with the study group closing the loop reviewing each sign. Models, glossary and assignments were compiled into a DVD to be distributed to schools. This material is unique, as it presents the models in LIBRAS and uses written Portuguese in the assignments dedicated to explore causal reasoning and written skills in Portuguese (Lima-Salles *et al*., 2004; Salles *et al.*, 2005). Ongoing work includes the application of the DVD into the classroom, with both deaf and hearing students. The vocabulary in LIBRAS will be expanded and used to describe new scientific concepts and models.

4. Discussion and final remarks

How to adequately handle QR models in order to have them brought into the classroom as useful tools for science education of deaf students? The answer can be summarized as follows: focus on bilingual education, which has the potential to fulfill the needs of both deaf and hearing students; create a vocabulary for expressing scientific concepts in sign language, following a procedure that includes the representation of recurrent categories of scientific concepts (so that the signs may be reused in different contexts) and the participation of the deaf community, teachers and experts; and produce didactic material based on qualitative models and in a visual pedagogy in which a diagrammatic approach is integrated with written texts in Portuguese to explore concept acquisition, and the development of language skills and of logical reasoning. The didactic material produced in the project may become the basis for the creation of a community of practice of deaf and hearing students that learn scientific concepts with the support of QR models and modern AI technologies (cf. www.dynalearn.eu).

Acknowledgements

We would like to thank the deaf students, teachers, Aline Mesquita and Maylena Gonçalves for their contributions, and to PROESP/ CAPES/ MEC grant 1553/2003, for the financial support to the project ‘Português como segunda língua na educação científica de surdos’ (Portuguese as a second language in scientific education of deaf students).

References

Bredeweg, B. and Forbus, K. 2003. Qualitative modeling in education. *AI Magazine*, 24 (2003), 4, 35-46.

Bredeweg, B.; Bouwer, A.; Jellema, J.; Bertels, D.; Linnebank, F. and Liem, J. 2006. Garp3 - A new Workbench for Qualitative Reasoning and Modelling. In C.Bailey-Kellogg and B.Kuipers (eds.) *Proceedings of the 20th International Workshop on Qualitative Reasoning (QR06)*, 21-28.

Forbus, K. 1984. Qualitative process theory. *Artificial Intelligence*, 24:85–168.

Lima-Salles, H.; Salles, P. and Bredeweg, B. 2004. Qualitative Reasoning in the Education of Deaf Students: scientific education and acquisition of Portuguese as a second language. In K. Forbus and de Kleer, J. (eds.) *Proceedings of the 18th International Workshop on Qualitative Reasoning (QR04)*, 97-104.

Rykiel, E. 1996. Testing Ecological Models: the Meaning of Validation. *Ecological Modelling*, 90(1996), 229-244.

Salles, P.; Lima-Salles, H. and Bredeweg, B. 2005. The Use of Qualitative Reasoning Models of Interactions Between Populations to Support Causal Reasoning of Deaf Students. In C.-K.Looi, G.McCalla, B.Bredeweg and J.Breuker (eds.) *Artificial Intelligence and Education: Supporting learning through Intelligent and Socially Informed Technology*. Amsterdam, IOS Press / Omasha, 579-586.

Weld, D. and de Kleer, J. (eds.) 1990. *Readings in Qualitative Reasoning about Physical Systems*. San Mateo, CA, Morgan Kaufmann.

Using Qualitative Reasoning in Modelling Consensus in Group Decision-Making

Llorenç Roselló and **Francesc Prats** and **Mónica Sánchez**
Polytechnical University of Catalonia, Barcelona, Spain
e-mail: {llorenc.rosello, francesc.prats, monica.sanchez}@upc.edu

Núria Agell *
Esade, Ramon Llull University, Barcelona, Spain
e-mail: nuria.agell@esade.edu

Abstract

Ordinal scales are commonly used in rating and evaluation processes. These processes usually involve group decision making by means of an experts' committee. In this paper a mathematical framework based on the qualitative model of the absolute orders of magnitude is considered. The entropy of a qualitatively described system is defined in this framework. On the one hand, this enables us to measure the amount of information provided by each evaluator and, on the other hand, the coherence of the evaluation committee. The new approach is capable of managing situations where the assessment given by experts involves different levels of precision. The use of the proposed measures within an automatic system for group decision making will contribute towards avoiding the potential subjectivity caused by conflicts of interests of the evaluators in the group.

Introduction

Nowadays, accreditation, audit, or rating agencies are dealing with a huge problem. Most committees are unable to ensure their legitimacy. Recent events have questioned the integrity of the rating agencies and their processes, and scandal stories about them have appeared in press and media.

This work is intended to be a first step towards the definition of evaluation measures in the group decision processes. To this end we introduce an approach based on qualitative reasoning models and the concept of entropy in order to measure the degree of coherence reached by an evaluation group.

Qualitative Reasoning (QR) is a sub-area of Artificial Intelligence that seeks to understand and explain human beings' ability for qualitative reasoning (Forbus 1996), (Kuipers 2004). The main objective is to develop systems that permit operating in conditions of insufficient numerical data or in the absence of such data. As indicated in (Travé-Massuyès and Dague 2003), this could be due to both a lack of information as well as to an information overload. A main goal of Qualitative Reasoning is to tackle problems in such a way that the principle of relevance is preserved; that is to say, each variable has to be valued with the level of precision required (Forbus 1984). It is not unusual for a situation to arise in which it is necessary to work simultaneously with different levels of precision, depending on the available information. To this end, the mathematical structures of Orders of Magnitude Qualitative Spaces (OM) were introduced.

The concept of entropy has its origins in the nineteenth century, particularly in thermodynamics and statistics. This theory has been developed from two aspects: the macroscopic, as introduced by Carnot, Clausius, Gibbs, Planck and Caratheodory; and the microscopic, developed by Maxwell and Boltzmann (Rokhlin 1967). The statistical concept of Shannon's entropy, related to the microscopic aspect, is a measure of the amount of information (Shannon 1948), (Cover and Thomas 1991).

Starting from the adaptation of the basic principles of Measure Theory (Halmos 1974), (Folland 1999) to the structure of OM (Roselló et al. 2008), this paper defines the concept of entropy within the QR framework.

Taking into account that entropy can be used to measure the amount of information, this work presents a way of measuring the amount of information given by an evaluator when describing a system by means of orders of magnitude. On the other hand, the defined entropy is applied to analyse the coherence degree of an evaluation committee in group decision making.

Section 2 presents the theoretical framework. In Section 3, the qualitative description induced by an evaluator is studied. Two operations for information aggregation and the concept of entropy in the absolute orders of magnitude spaces are defined in Section 4 and 5 respectively, and Section 6 introduces a coherence degree in group decision. The paper ends with several conclusions and outlines some proposals for future research.

Theoretical Framework

Order of magnitude models are an essential piece among the theoretical tools available for qualitative reasoning about physical systems ((Kalagnanam, Simon, and Iwasaki 1991), (Struss 1988). They aim at capturing order of magnitude commonsense ((Travé-Massuyès 1997)) inferences, such as used in the engineering world. Order of magnitude knowl-

---

[*] This work has been partly funded by MEC (Spanish Ministry of Education and Science) AURA project (TIN2005-08873-C02). Authors would like to thank their colleagues of GREC research group of knowledge engineering for helpful discussions and suggestions.

edge may be of two types: absolute or relative. The absolute order of magnitudes are represented by a partition of $\mathbb{R}$, each element of the partition standing for a basic qualitative class. A general algebraic structure, called Qualitative Algebra or Q-algebra, was defined on this framework ((Travé-Massuyès and Piera 1989)), providing a mathematical structure which unifies sign algebra and interval algebra through a continuum of qualitative structures built from the rougher to the finest partition of the real line. The most referenced order of magnitude Q-algebra partitions the real line into 7 classes, corresponding to the labels: Negative Large(NL), Negative Medium(NM), Negative Small(NS), Zero(0), Positive Small(PS), Positive Medium(PM) and Positive Large(PL). Q-algebras and their algebraic properties have been extensively studied ((Missier, Piera, and Travé 1989), (Travé-Massuyès and Dague 2003))

Order of magnitude knowledge may also be of relative type, in the sense that a quantity is qualified with respect to another quantity by means of a set of binary order-of-magnitude relations. The seminal relative orders of magnitude model was the formal system FOG ((Raiman 1986)), based on three basic relations, used to represent the intuitive concepts of ”negligible with respect to” (Ne), ”close to” (Vo) and ”comparable to” (Co), and described by 32 intuition-based inference rules. The relative orders of magnitude models that were proposed later improved FOG not only in the necessary aspect of a rigorous formalisation, but also permitting the incorporation of quantitative information when available and the control of the inference process, in order to obtain valid results in the real world ((Mavrovouniotis and Stephanopoulos 1987), (Dague 1993a), (Dague 1993b)).

In ((Travé-Massuyès et al. 2002), (Travé-Massuyès and Dague 2003)) the conditions under which an absolute orders of magnitude and a relative orders of magnitude model are consistent is analysed and the constraints that consistency implies are determined and interpreted.

In (Roselló et al. 2008) a generalization of qualitative orders of magnitude was proposed to provide the theoretical basis on which to develop a Measure Theory in this context.

The *classical orders of magnitude qualitative spaces* (Travé-Massuyès and Dague 2003) verify the conditions of the generalized model introduced in (Roselló et al. 2008). These models are built from a set of ordered basic qualitative labels determined by a partition of the real line.

Let X be the real interval $[a_1, a_n)$, and a partition of this set given by $\{a_2, \dots, a_{n-1}\}$, with $a_1 < a_2 < \dots < a_{n-1} < a_n$. The set of basic labels is

$$\mathcal{S} = \{B_1, \dots, B_{n-1}\},$$

where, for $1 \leq i \leq n-1$, B_i is the real interval $[a_i, a_{i+1})$. The set of indexes is $I = \{1, 2, \dots, n-1\}$.

For $1 \leq i < j \leq n-1$ the non-basic label $[B_i, B_j)$ is:

$$[B_i, B_j) = \{B_i, B_{i+1}, \dots, B_{j-1}\},$$

and it is interpreted as the real interval $[a_i, a_j)$.

For $1 \leq i \leq n-1$ the non-basic label $[B_i, B_\infty)$ is:

$$[B_i, B_\infty) = \{B_i, B_{i+1}, \dots, B_{n-1}\},$$

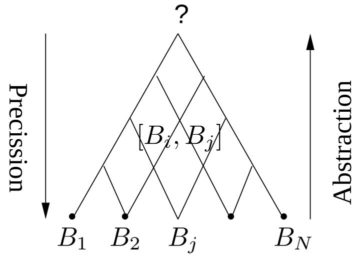


Figure 1: The space $\mathbb{S}_n$

and it is interpreted as the real interval $[a_i, a_n)$.

The complete universe of description for the Orders of Magnitude Space is the set

$$\mathbb{S}_n = \{\, [B_i, B_j) \mid B_i, B_j \in \mathcal{S}, i \leq j\} \cup \{\, [B_i, B_\infty) \mid B_i \in \mathcal{S}\},$$

which is called the absolute orders of magnitude qualitative space with granularity n, also denoted $OM(n)$.

There is a partial order relation $\leq_P$ in $\mathbb{S}_n$ “to be more precisely than”, given by:

$$L_1 \leq_P L_2 \Longleftrightarrow L_1 \subset L_2. \qquad (1)$$

The least precise label is denoted by ? and it is the label $[B_1, B_\infty)$, which corresponds to the interval $[a_1, a_n)$.

This structure permits working with all different levels of precision from the label ? to the basic labels.

In some theoretical works, orders of magnitude qualitative spaces are constructed by partitioning the whole real line $(-\infty, +\infty)$ instead of a bounded real interval $[a_1, a_n)$. However, in most real world applications involved variables do have a lower bound a_1 and an upper bound a_n, and then values less than a_1 or greater than a_n are considered as outliers and they are not treated like any other. To introduce the classical concept of entropy by means of qualitative orders of magnitude spaces, Measure Theory is required. This theory seeks to generalize the concept of “length”, “area”and “volume”, understanding that these quantities need not necessarily correspond to their physical counterparts, but may in fact represent others. The main use of the measure is to define the concept of integration for orders of magnitude spaces. In (Roselló et al. 2008) measures on the generalized qualitative orders of magnitude spaces are defined.

Qualitativization induced by an evaluator

To introduce the concept of entropy by means of qualitative orders of magnitude, it is necessary to consider the qualitativization function between the set to be qualitatively described and the space of qualitative labels, $\mathbb{S}_n$.

To simplify the notation, let us express with a calligraphic letter the elements in $\mathbb{S}_n$; thus, for example, elements $[B_i, B_j)$ or $[B_i, B_\infty)$ shall be denoted as $\mathcal{E}$.

Let Λ be the set that represents a magnitude or a feature that is qualitatively described by means of the labels of $\mathbb{S}_n$. Since Λ can represent both a continuous magnitude such as position and temperature, etc., and a discrete feature such as salary and colour, etc., Λ could be considered as the range of a function

$$a : I \subset \mathbb{R} \to Y,$$

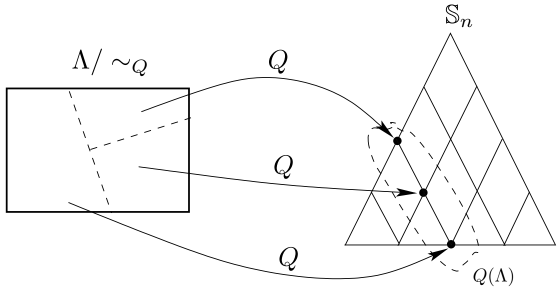


Figure 2: The qualitativization of a set Λ by means of Q.

where Y is a convenient set. For instance, if a is a room temperature during a period of time $I = [t_0, t_1]$, Λ is the range of temperatures during this period of time. Another example can be considered when $I = \{1, \dots, n\}$ and $\Lambda = \{a(1), \dots, a(n)\}$ are n people whose eye colour we aim to describe. In general, $\Lambda = \{a(t) = a_t \,|\, t \in I\}$.

The process of qualitativization is given by a function

$$Q : \Lambda \to \mathbb{S}_n,$$

where $a_t \mapsto Q(a_t) = \mathcal{E}_t =$ minimum label (with respect to the inclusion $\subset$) which describes a_t, i.e. the most precise qualitative label describing a_t. All the elements of the set $Q^{-1}(\mathcal{E}_t)$ are "representatives" of the label $\mathcal{E}_t$ or "are qualitatively described" by $\mathcal{E}_t$. They are qualitatively equal.

The function Q induces a partition in Λ by means of the equivalence relation:

$$a \sim_Q b \Longleftrightarrow Q(a) = Q(b).$$

This partition will be denoted by $\Lambda/\sim_Q$, and its equivalence classes are the sets $Q^{-1}(Q(a_j)) = Q^{-1}(\mathcal{E}_j)$, $\forall j \in J \subset I$. Each of these classes contains all the elements of Λ which are described by the same qualitative label.

Information aggregation

Given two qualitativizations Q and Q' of the set Λ over a space $\mathbb{S}_n$ it is natural to define two different operations between them. Intuitively speaking, one is the result of *mix* the two knowledges in a new knowledge that includes everything known about each element of Λ, and the other one is the result of taking what is *common* between the two knowledges.

The operation mix $\vee$

Definition 1 *Given two qualitativizations Q and Q', the operation $Q \vee Q'$ is a new qualitativization function $Q \vee Q' : \Lambda \to \mathbb{S}_n$ such that*

$$(Q \vee Q')(a_t) = Q(a_t) \sqcup Q'(a_t),$$

where $\sqcup$ is the connex union of labels i.e. the minimum label describing the elements of $Q^{-1}(Q(a_t))$ and the elements of $Q'^{-1}(Q'(a_t))$.

The partition

$$(\Lambda/\sim_Q) \cap (\Lambda/\sim_{Q'}) = \{X_i \cap Y_j | X_i \in \Lambda/\sim_Q, Y_j \in \Lambda/\sim_{Q'}\}.$$

is not the partition $\Lambda/\sim_{Q\vee Q'}$ because there may be $a_{t_0} \in X_{i_0} \cap Y_{j_0}$ and $a_{t_1} \in X_{i_1} \cap Y_{j_1}$ such that $(Q \vee Q')(a_{t_0}) = (Q \vee Q')(a_{t_1})$. The relation between these partitions is given by the next proposition.

Proposition 1 *Given a set Λ, the space $\mathbb{S}_n$ and two qualitativizations Q and Q', then each class of $\Lambda/\sim_{Q\vee Q'}$ is a (disjoint) union of classes of $(\Lambda/\sim_Q) \cap (\Lambda/\sim_{Q'})$:*

$$\mathrm{Class}_{Q\vee Q'}(x) = \bigcup_{y \in \mathrm{Class}_{Q\vee Q'}(x)} (\mathrm{Class}_Q(y) \cap \mathrm{Class}_{Q'}(y))$$

Proof: This set equality will be proven by double inclusion:

$\subset$) If $z \in \mathrm{Class}_{Q\vee Q'}(x)$ then it is trivial that $z \in \bigcup_{y\in \mathrm{Class}_{Q\vee Q'}(x)} (\mathrm{Class}_Q(y) \cap \mathrm{Class}_{Q'}(y))$.

$\supset$) If $z \in \bigcup_{y\in \mathrm{Class}_{Q\vee Q'}(x)} (\mathrm{Class}_Q(y) \cap \mathrm{Class}_{Q'}(y))$ then there exists $y \in \mathrm{Class}_{Q\vee Q'}(x)$ such that $Q(z) = Q(y)$ and $Q'(z) = Q'(y)$, then $(Q \vee Q')(z) = (Q \vee Q')(y) = (Q \vee Q')(x)$, whence $z \in \mathrm{Class}_{Q\vee Q'}(x)$

The last step is the proof that it is a disjoint union: let be $y, z \in \mathrm{Class}_{Q\vee Q'}(x)$, then $\mathrm{Class}_Q(y) \cap \mathrm{Class}_{Q'}(y) \cap \mathrm{Class}_Q(z) \cap \mathrm{Class}_{Q'}(z) = \emptyset$ or $\mathrm{Class}_Q(y) \cap \mathrm{Class}_{Q'}(y) = \mathrm{Class}_Q(z) \cap \mathrm{Class}_{Q'}(z)$. In effect:

$$t \in \mathrm{Class}_Q(y) \cap \mathrm{Class}_{Q'}(y) \cap \mathrm{Class}_Q(z) \cap \mathrm{Class}_{Q'}(z) \Rightarrow$$
$$\Rightarrow Q(t) = Q(y), Q'(t) = Q'(y), Q(t) = Q(z), Q'(t) = Q'(z) \Rightarrow$$
$$\Rightarrow Q(y) = Q(z), Q'(y) = Q'(z) \Rightarrow$$
$$\Rightarrow \mathrm{Class}_Q(y) = \mathrm{Class}_Q(z), \mathrm{Class}_{Q'}(y) = \mathrm{Class}_{Q'}(z).$$

□

The operation common $\wedge$

The concept of coherence is required in order to introduce the operation common:

Definition 2 *Given a set Λ and a qualitative space $\mathbb{S}_n$, two qualitativizations of Λ, Q, Q' are* coherent, *$Q \rightleftarrows Q'$, iff*

$$Q(a_t) \cap Q'(a_t) \neq \emptyset, \quad \forall a_t \in \Lambda. \tag{2}$$

This last condition is equivalent to say that $Q(a_t) \approx Q'(a_t), \forall a_t \in \Lambda$.[1]

It is clear that the relation $\rightleftarrows$ is symmetric and reflexive.

Definition 3 *Given a set Λ and a qualitative space $\mathbb{S}_n$, the* set of coherent qualitativizations *of a qualitativization Q,* $\mathrm{Cohe}(Q)$, *is*

$$\mathrm{Cohe}(Q) = \{Q' \textit{qualitativization of } \Lambda \mid Q \rightleftarrows Q'\} \tag{3}$$

---

[1] In the theory of absolute orders of magnitude, two labels $\mathcal{E}, \mathcal{F}$ are qualitative equal, $\mathcal{E} \approx \mathcal{F}$, iff $\mathcal{E} \cap \mathcal{F} \neq \emptyset$.

Intuitively speaking, $\mathrm{Cohe}(Q)$ are all the qualitativizations having "some agreement" when they assign labels to all the elements of Λ.

Definition 4 *Given two qualitativizations Q and Q', such that $Q \rightleftarrows Q'$, the operation $Q \wedge Q'$ is a new qualitativization function $Q \wedge Q' : \Lambda \to \mathbb{S}_n$ such that*

$$(Q \wedge Q')(a_t) = Q(a_t) \cap Q'(a_t).$$

It is not difficult to check that the operations mix and common are commutative and associative, so it can be considered the mix and common operation of any number of qualitativizations $Q_1, \ldots, Q_n$.

An order relation can be defined from the operation common and mix:

Definition 5 *Given two qualitativizations Q and Q' of a set Λ over a qualitative space $\mathbb{S}_n$, Q* is less accurate *than Q', or $Q \leq Q'$, when $Q \vee Q' = Q$. That is to say that $\forall a_t \in \Lambda$ then $Q'(a_t) \subset Q(a_t)$, i.e. each element of the set Λ is more precise described by Q' than by Q.*

Entropy

The information of a label

The information of a label $\mathcal{E}$ will be a positive continuous real function on the measure of the label, and will be denoted by $I(\mathcal{E})$. It also will be assumed that if a label $\mathcal{E}$ is more precise than a label $\mathcal{E}'$, then there is more information in $\mathcal{E}$ than in $\mathcal{E}'$:

$$\mathcal{E} \leq_P \mathcal{E}' \Rightarrow I(\mathcal{E}) \geq I(\mathcal{E}').$$

Another assumption about the function I is that the information of the label ? is zero.

The following definition of I inspired in the Shannon theory of information ((Shannon 1948)) verifies these assumptions:

Definition 6 *The information of a label $\mathcal{E} \in \mathbb{S}_n$ is*

$$I(\mathcal{E}) = \log \frac{1}{\mu(\mathcal{E})},$$

where μ is a normalized measure defined in $\mathbb{S}_n$ and $\mu(\mathcal{E}) \neq 0$.

It is trivial to check that it is positive and continuous, and decreases with respect to $\leq_P$:
From the definition of $\leq_P$ in expression (1) from the section 2:

$$\mathcal{E} \leq_P \mathcal{F} \Rightarrow \mathcal{E} \subset \mathcal{F} \Rightarrow \mu(\mathcal{E}) \leq \mu(\mathcal{F}) \Rightarrow \log \frac{1}{\mu(\mathcal{E})} \geq \log \frac{1}{\mu(\mathcal{F})}$$

Morover, $I(?) = \log 1 = 0$.
Example: In the classical $\mathbb{S}_n$ model, defining a measure $\mu([a_i, a_{i+1}]) = (a_{i+1} - a_i)/(a_n - a_1)$, the information of a label is $I([a_i, a_{i+1}]) = \log \left(\frac{a_n - a_1}{a_{i+1} - a_i} \right)$.

Entropy of a qualitativization in $\mathbb{S}_n$

Let us suppose a normalized measure $\overline{\mu}$ in the set Λ.

Definition 7 *The entropy H of a qualitativization Q is defined as:*

$$H(Q) = \sum_{\mathcal{E} \in \mathbb{S}_n} \overline{\mu}(Q^{-1}(\mathcal{E})) I(\mathcal{E}). \tag{4}$$

If $\Lambda / \sim_Q = \{X_i, i \in J\}$, that is, the set of equivalence classes of $\sim_Q$, then the expression 4 can be expressed as

$$H(Q) = \sum_{i \in J} \overline{\mu}(X_i) I(Q(X_i)). \tag{5}$$

The expression of entropy in the definition (7) defines the entropy as a weighted average of the information of the elements of the set Λ given by Q.

Proposition 2 *Given a set Λ and the space $\mathbb{S}_n$, each with its own measure, the maximum entropy, $H(\tilde{Q})$, is achieved when $Q(\Lambda) = \{\mathcal{E}_*\}$ where $\mathcal{E}_*$ is the shortest label with respect to μ. In other words, the maximum entropy is reached when Q maps the whole set Λ to the most precise label:*

$$H(\tilde{Q}) = \max_{Q} H(Q) = I(\min_{\mathcal{E}} \mu(\mathcal{E})) = \log \frac{1}{\mu(\mathcal{E}_*)}.$$

Proof: It is clear that the label with maximum information is the shortest label with respect to μ; if this label is called $\mathcal{E}_*$, then

$$H(Q) \leq \sum_{\mathcal{E} \in \mathbb{S}_n} \overline{\mu}(Q^{-1}(\mathcal{E})) I(\mathcal{E}_*) =$$

$$= I(\mathcal{E}_*) \sum_{\mathcal{E} \in \mathbb{S}_n} \overline{\mu}(Q^{-1}(\mathcal{E})) = I(\mathcal{E}_*),$$

because the measure $\overline{\mu}$ is normalized and the set $Q^{-1}(\mathcal{E})$ is a partition of Λ. □

According to this proposition it is possible to define the precision of a qualitativization:

Definition 8 *The* precision *of a qualitativization Q of a set Λ, $h(Q)$, is the relative entropy respect the maximum entropy $H(\tilde{Q})$ for the set Λ in $\mathbb{S}_n$*

$$h(Q) = \frac{H(Q)}{H(\tilde{Q})} \tag{6}$$

This quantity is a real number between 0 and 1, the closer it is to 1, the more accurate the evaluator is.

Lemma 1 *For all labels $\mathcal{E}, \mathcal{F} \in \mathbb{S}_n$ it is hold that $I(\mathcal{E} \sqcup \mathcal{F}) \leq I(\mathcal{E}) + I(\mathcal{F})$.*

Proof: Since $\mathcal{E} \leq_P \mathcal{E} \sqcup \mathcal{F}$ then $I(\mathcal{E}) \geq I(\mathcal{E} \sqcup \mathcal{F})$ and then $I(\mathcal{E} \sqcup \mathcal{F}) \leq I(\mathcal{E}) + I(\mathcal{F})$ □

From lemma 1 the next result with respect the operation mix of two qualitativizations is presented:

Theorem 1 *Given a set Λ, the space $\mathbb{S}_n$, and two qualitativizations Q and Q',then*

$$H(Q \vee Q') \leq H(Q) + H(Q').$$

Proof: From equation 4

$$H(Q \vee Q') = \sum_{\mathcal{F} \in (Q \vee Q')(\Lambda)} \overline{\mu}((Q \vee Q')^{-1}(\mathcal{F})) I(\mathcal{F}), \quad (7)$$

and using the proposition 1

$$(Q \vee Q')^{-1}(\mathcal{F}) = \bigcup_{X_i \in (\Lambda/\sim_Q) \cap (\Lambda/\sim_{Q'})} X_i.$$

Since this union is a disjoint union and $\overline{\mu}$ is a measure

$$\overline{\mu}((Q \vee Q')^{-1}(\mathcal{F})) = \sum_{i \in J} \overline{\mu}(X_i),$$

where J is an index set. Taking into account that $X_i \in (\Lambda/ \sim_Q) \cap (\Lambda/ \sim_{Q'})$, it can be expressed as $X_i = M_{j_i} \cap N_{k_i}$ where $M_{j_i} \in \Lambda/ \sim_Q$ and $N_{k_i} \in \Lambda/ \sim_{Q'}$. By construction of $\Lambda/_{Q \vee Q'}$, each label is $\mathcal{F} = Q(M_{j_i}) \sqcup Q'(N_{j_i})$, then

$$\overline{\mu}((Q \vee Q')^{-1}(\mathcal{F})) I(\mathcal{F}) =$$

$$= \sum_{i \in J} \overline{\mu}(M_{j_i} \cap N_{k_i}) I(Q(M_{j_i}) \sqcup Q'(N_{j_i})),$$

from the lemma 1:

$$\overline{\mu}((Q \vee Q')^{-1}(\mathcal{F})) I(\mathcal{F}) \leq$$

$$\leq \sum_{i \in J} \overline{\mu}(M_{j_i} \cap N_{k_i}) I(Q(M_{j_i}) + I(Q'(N_{j_i})),$$

Putting it all together into 7

$$H(Q \vee Q') \leq \sum_{M \in \Lambda/\sim_Q, N \in \Lambda/\sim_{Q'}} \overline{\mu}(M \cap N)(I(Q(M)) + I(Q(N))),$$

On the other hand $M \cap N \subset M, N$ so $\overline{\mu}(M \cap N) \leq \overline{\mu}(M), \overline{\mu}(N)$ whence the inequality is inferred. □

The next proposition shows that the entropy respects the accuracy relation between qualitativizations:

Proposition 3 *Given a set Λ, the space $\mathbb{S}_n$, and two qualitativizations Q and Q' such that $Q \leq Q'$ then $H(Q) \leq H(Q')$.*

Proof: Lets write $\Lambda/ \sim_Q = \cup_{i \in M} X_i$, $\Lambda/ \sim_Q = \cup_{j \in N} Y_j$, and $(\Lambda/ \sim_Q) \cap (\Lambda/ \sim_{Q'}) = \cup_{i,j}(X_i \cap Y_j)$. For each $X_i \in \Lambda/ \sim_Q$ there exist a subset of index $N_i \subset N$ such that $X_i = \cup_{j \in N_i}(X_i \cap Y_j)$ and vice-versa, there exist a subset of index $M_j \subset M$ such that $Y_j = \cup_{i \in M_j}(X_i \cap Y_j)$ (all unions are disjoint unions). If $X_i \cap Y_j \neq \emptyset$ then from definition 5:

$$Q'(Y_j) \subset Q(X_i) \Rightarrow I(Q(X_i)) \leq I(Q'(Y_j)) \quad (8)$$

The entropy of Q is

$$H(Q) = \sum_{i \in M} \overline{\mu}(X_i) I(Q(X_i)) =$$

$$= \sum_{i \in M} \overline{\mu}\left(\cup_{j \in N_i}(X_i \cap Y_j)\right) I(Q(X_i)) =$$

$$= \sum_{i \in M} \left(\sum_{j \in N_i} \overline{\mu}(X_i \cap Y_j) I(Q(X_i)) \right) \leq$$

from the inequality in (8)

$$\leq \sum_{i \in M} \left(\sum_{j \in N_i} \overline{\mu}(X_i \cap Y_j) I(Q(Y_j)) \right) =$$

$$= \sum_{j \in N_i} \left(\sum_{i \in M} \overline{\mu}(X_i \cap Y_j) I(Q(X_i)) \right) =$$

$$= \sum_{j \in N} \overline{\mu}(Y_j) I(Q(Yj)) = H(Q').$$

□

Coherence degree in group decision

The measure of the precision and coherence in group decision evaluation problems is one of the main applications of the theory presented in this paper. The underlying idea on the next definition stands on the need to measure the precision of a set of evaluators and the coherence degree of its evaluations when they are evaluating a set by means of labels belonging to a $\mathbb{S}_n$.

First of all there is a formalization of the problem of the group evaluation of a set: Given a space $\mathbb{S}_n$, a finite non empty set $\Lambda = \{a_1, \ldots, a_N\}$ and set $\mathbb{E} = \{\alpha_1, \ldots, \alpha_M\}$, (it is the set of group evaluators), a *group evaluation* of Λ is the pair $(\Lambda, \mathcal{Q}_{\mathbb{E}})$, where $\mathcal{Q}_{\mathbb{E}} = \{Q_i : \Lambda \to \mathbb{S}_n \mid i \in \mathbb{E}\}$.

There exists *coherence* in the group, if and only if, the group is coherent, i.e. iff $\forall Q \in \mathcal{Q}_{\mathbb{E}}, \mathrm{Cohe}(Q) = \mathcal{Q}_{\mathbb{E}}$. Notice that it is evident that the last condition is satisfied if there exists a Q such that $\mathrm{Cohe}(Q) = \mathcal{Q}_{\mathbb{E}}$. Assuming that the group is in coherence, the next definition of coherence degree measures the relation between the entropy of operations mix and common in the qualitativizations of the group:

Definition 9 *Given a group evaluation $(\Lambda, \mathcal{Q}_{\mathbb{E}})$ in coherence, the* coherence degree *of the group, $\kappa(\mathcal{Q}_{\mathbb{E}})$, is*

$$\kappa(\mathcal{Q}_{\mathbb{E}}) = \frac{H(\bigvee_{i \in \mathbb{E}} Q_i)}{H(\bigwedge_{i \in \mathbb{E}} Q_i)} \quad (9)$$

When the whole group qualitativizes the set Λ in the same way, i.e., when $Q_i = Q_j, \forall i, j \in \mathbb{E}$, then $\kappa(\mathcal{Q}_{\mathbb{E}}) = 1$, and if $\kappa(\mathcal{Q}_{\mathbb{E}}) = 1$ then $Q_i = Q_j, \forall i, j \in \mathbb{E}$. On the other hand, the spread with Q_i, implies a small $H(\bigvee_{i \in \mathbb{E}} Q_i)$ and a big $H(\bigwedge_{i \in \mathbb{E}} Q_i)$. The given degree of coherence will give us a global index with respect to the whole group of evaluators. The key point on this definition is that the closer this degree is to 1, the closer the group is to be in a consensus relation.. When the coherence degree is not satisfactory, an iterative process will start to increase this degree.

The next property shows that the coherence degree of a group evaluation problem cannot increase by adding to the group a new evaluator.

Proposition 4 *Consider a group evaluation* $(\Lambda, \mathcal{Q}_{\mathbb{E}})$ *in coherence. Let be* Q_{new} *a new evaluator of* Λ *such that* $Q_{\text{new}} \notin \mathcal{Q}_{\mathbb{E}}$, *then*

$$\kappa(\mathcal{Q}_{\mathbb{E}} \cup \{Q_{\text{new}}\}) \leq \kappa(\mathcal{Q}_{\mathbb{E}}).$$

Proof:: From the definitions 4 and 5 can be deduced the inequalities $Q \vee Q' \leq Q, Q' \leq Q \wedge Q'$ whence can be deduced that if a new evaluator joints the group of evaluators then:

$$(\vee_{i\in\mathbb{E}} Q_i) \vee Q_{\text{new}} \leq \vee_{i\in\mathbb{E}} Q_i,$$
$$\wedge_{i\in\mathbb{E}} Q_i \leq (\wedge_{i\in\mathbb{E}} Q_i) \wedge Q_{\text{new}}.$$

From proposition

$$H(\vee_{i\in\mathbb{E}} Q_i) \vee Q_{\text{new}}) \leq H(\vee_{i\in\mathbb{E}} Q_i),$$
$$H(\wedge_{i\in\mathbb{E}} Q_i) \leq H((\wedge_{i\in\mathbb{E}} Q_i) \wedge Q_{\text{new}}),$$

whence $\kappa(\mathcal{Q}_{\mathbb{E}} \cup \{Q_{\text{new}}\}) \leq \kappa(\mathcal{Q}_{\mathbb{E}})$. □

Therefore, the only way to increase the coherence degree in a group is that the evaluators in the group reconsider the problem.

Conclusions and future research

A mathematical framework is presented to define group decision techniques to measure precision and coherence based on a qualitative structure of orders of magnitude.

This paper introduces the concept of entropy by means of absolute orders of magnitude qualitative spaces to measure the amount of information of a system when using orders of magnitude descriptions to represent it. On the other hand, entropy makes it possible to introduce a measure of coherence in group decision-making problems.

The obtained results can be applied to tackle evaluation and ranking problems which require an ordinal set of labels to qualify decision alternatives.

A coherence degree is introduced in order to obtain an objective measure of reliability in group decision making to detect incoherencies and avoid potential subjectivity caused by conflicts of interest regarding evaluators.

From a theoretical point of view, future research could focus on two lines. On the one hand, it could focus on the analysis of the given structure of the qualitative descriptions of a system to define a lattice using mix and common operations. On the other hand a distance between qualitative descriptions will be defined by means of conditioned entropy.

Within the framework of applications, this work and its related methodology will be orientated towards the development of techniques to detect malfunctioning within an evaluation committee, and to analyse whether it can reflect a corruption or a lack of knowledge in a part of the committee.

References

Cover, T. M., and Thomas, J. A. 1991. *Elements of Information Theory*. Wiley Series in Telecomunications.

Dague, P. 1993a. Numeric reasoning with relative orders of magnitude. AAAI Conference, Washington.

Dague, P. 1993b. Symbolic reasoning with relative orders of magnitude. 13th IJCAI, Chambéry.

Folland, G. 1999. *Real Analysis: Modern Techniques and Their Applications*. Pure and Applied Mathematics: A Wiley-Interscience Series of Texts, Monographs, and Tracks. John Wiley & Sons, Inc.

Forbus, K. 1984. Qualitative process theory. *Artificial Intelligence* 24:85–158.

Forbus, K. 1996. *Qualitative Reasoning*. CRC Hand-book of Computer Science and Engineering. CRC Press.

Halmos, P. R. 1974. *Measure Theory*. Springer-Verlag.

Kalagnanam, J.; Simon, H.; and Iwasaki, Y. 1991. The mathematical bases for qualitative reasoning. *IEEE Expert.*

Kuipers, B. 2004. Making sense of common sense knowledge. *Ubiquity* 4(45).

Mavrovouniotis, M., and Stephanopoulos, G. 1987. Reasoning with orders of magnitude and approximate relations. AAAI Conference, Seattle.

Missier, A.; Piera, N.; and Travé, L. 1989. Order of magnitude algebras: a survey. *Revue d'Intelligence Artificielle* 3(4):95–109.

Raiman, O. 1986. Order of magnitude reasoning. *Artificial Intelligence* (24):11–38.

Rokhlin, V. 1967. Lectures on the entropy of eeasure ereserving eransformations. *Russian Math. Surveys* 22:1 – 52.

Roselló, L.; Prats, F.; Sánchez, M.; and Agell, N. 2008. A definition of entropy based on qualitative descriptions. In L. Bradley, L. T.-M., ed., *22nd International Workshop on Qualitative Reasoning (QR 2008)*. University of Boulder, Colorado, USA.

Shannon, C. E. 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27:379 – 423.

Struss, P. 1988. Mathematical aspects of qualitative reasoning. *AI in Engineering* 3(3):156–169.

Travé-Massuyès, L., and Dague, P., eds. 2003. *Modèles et raisonaments qualitatifs*. Hermes Science (París).

Travé-Massuyès, L., and Piera, N. 1989. The orders of magnitude models as qualitative algebras. Number 11th. IJCAI.

Travé-Massuyès, L.; Prats, F.; Sánchez, M.; and Agell, N. 2002. Consistent relative and absolute order-of-magnitude models. 16th International Workshop on Qualitative Reasoning.

Travé-Massuyès, L. e. a. 1997. *Le raisonnement qualitatif pour les sciences de l'ingénieur*. Ed. Hermès.

Qualitative approximation to Dynamic Time Warping similarity between time series data

Blaž Strle, Martin Možina, Ivan Bratko
Faculty of Computer and Information Science
University of Ljubljana
Slovenia

Abstract

Dynamic time warping (DTW) is a method for calculating the similarity between two time series which can occur at different times or speeds. Although its effectiveness made it very popular in several disciplines, its time complexity of $O(N^2)$ makes it useful only for relatively short time series. In this paper, we propose a qualitative approximation Qualitative Dynamic Time Warping (QDTW) to DTW. QDTW reduces a time series length by transforming it to qualitative time series. DTW is later calculated between qualitative time series. As qualitative time series are normally much shorter than their corresponding numerical time series, time to compute their similarity is significantly reduced. Experimental results have shown improved running time of up to three orders of magnitude, while prediction accuracy only slightly decreased.

1. Introduction

Time series is a form of data that is present in virtually every scientific discipline and business application. It can be described as a sequence of observations, measured at successive times, spaced at (often uniform) time intervals. Dynamic Time Warping (DTW) (Sakoe and Chiba 1978) is a method for calculating the similarity between two time series which can occur at different times or speeds. Its ability to warp time axis and find optimal alignment between two time series has made it very popular. DTW has been used in several disciplines (Keogh and Pazzani 2001), such as: speech recognition, gesture recognition, data mining, robotics, manufacturing and medicine. In spite of its effectiveness, its time complexity of $O(N^2)$ makes it useful only for relatively short time series. This limitation can be overcome by reducing time series length. In qualitative modeling, numerical models can be seen as an abstraction of the real world and qualitative models are often viewed as a further abstraction of numerical models (Bratko 2000). In this abstraction, some quantitative information is abstracted away while keeping information that is relevant to the problem.

In this paper, we introduce a qualitative approximation Qualitative Dynamic Time Warping (QDTW) to DTW. QDTW reduces time series size by transforming it to a qualitative time series. As qualitative time series are usually much simpler and shorter than numerical time series, savings in running time are large.

The rest of this paper is structured as follows. Section 2 briefly reviews classic Dynamic Time Warping, including several techniques that make it more time efficient. In Section 3 we introduce and describe our modification to classic DTW. In Section 4, DTW and QDTW are experimentally evaluated on three domains and the results are discussed. Section 5 gives conclusions and future work.

2. Dynamic time warping

2.1 Dynamic Time Warping

In this section we briefly describe classic Dynamic Time Warping method. Dynamic Time Warping aligns two time series in the way some distance measure is minimized (usually Euclidean distance is used). Optimal alignment (minimum distance warp path) is obtained by allowing assignment of multiple successive values of one time series to a single value of the other time series and therefore DTW can also be calculated on time series of different lengths. Figure 1 shows examples of two time series and value alignment between them for Euclidean distance (left) and DTW similarity measure (right). Notice that the time series have similar shapes, but are not aligned in time. While Euclidean distance measure does not align time series, DTW does address the problem of time difference. By using DTW, optimal alignment is found among several different warp paths. This can be easily represented if two time series $A = (a_1, a_2, ..., a_n)$ and $B = (b_1, b_2, ..., b_m)$, $a_i, b_j \in \mathbb{R}$ are arranged to form a n-by-m grid. Each grid point corresponds to an alignment between elements $a_i \in A$ and $b_j \in B$. A warp path $W = w_1, w_2, ..., w_k, ..., w_K$ is a sequence of grid points, where each w_k corresponds to a point $(i, j)_k$ - warp

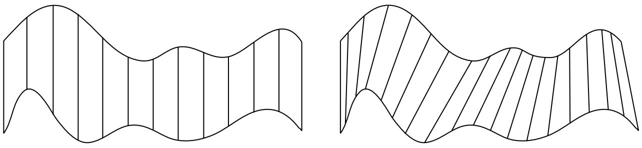


Figure 1: Example of two time series. Lines between time series show value alignment used by Euclidean distance (left) and Dynamic Time Warping similarity measure (right).

path W maps elements of sequences A and B. A warp path is typically subject to several constraints:

- **Boundary conditions:** $w_1 = (1, 1)$ and $w_K = (n, m)$. This requires the warping path to start in first point of both sequences and end in last point of both sequences.
- **Continuity:** Let $w_k = (a, b)$ then $w_k - 1 = (a', b')$ where $a - a' \leq 1$ and $b - b' \leq 1$. This restricts the allowable steps in the warping path to adjacent cells.
- **Monotonicity:** Let $w_k = (a, b)$ then $w_k - 1 = (a', b')$ where $a - a' \geq 0$ and $b - b' \geq 0$. This forces the points in W to be monotonically spaced in time.

From all possible warp paths DTW finds the optimal one:

$$DTW(A, B) = min_W[\sum_{k=1}^{K} d(w_k)]$$

Here $d(w_k)$ is the distance between elements of time series.

Algorithm The goal of DTW is to find minimal distance warp path between two time series. Dynamic programming can be used for this task. Instead of solving the entire problem all at once, solutions to sub problems (sub-series) are found and used to repeatedly find the solution to a slightly larger problem. Let $DTW(A, B)$ be the distance of the optimal warp path between time series $A = (a_1, a_2, ..., a_n)$ and $B = (b_1, b_2, ..., b_m)$ and let $D(i, j) = DTW(A', B')$ be the distance of the optimal warp path between the prefixes of the time series A and B:

$$D(0, 0) = 0$$

$$A' = (a_1, a_2, ..., a_i), B' = (b_1, b_2, ..., b_j)$$

$$0 \leq i \leq n, 0 \leq j \leq m$$

Then $DTW(A, B)$ can be calculated using the following recursive equations:

$$D(0, 0) = 0$$

$$D(i, j) = min(D(i - 1, j), D(i, j - 1), D(i - 1, j - 1)) + d(a_i, b_j)$$

Here $d(a_i, b_j)$ is the distance between two values of the two time series (usually Euclidean distance is used).

The most common way of calculating $DTW(A, B)$ is to construct a $n*m$ cost matrix M, where each cell corresponds to the distance of the minimal distance warp path between the prefixes of the time series A and B (Figure 2):

$$M(i, j) = D(i, j)$$

$$1 \leq i \leq n$$

$$1 \leq j \leq m$$

We start by calculating all the fields with small indexes and then progressively continue to calculate fields with higher indexes:

```
for i = 1...n
    for j = 1...m
        M(i,j) = min(M(i-1,j), M(i,j-1), M(i,j)) + dst($a_i$,$b_j$)
```

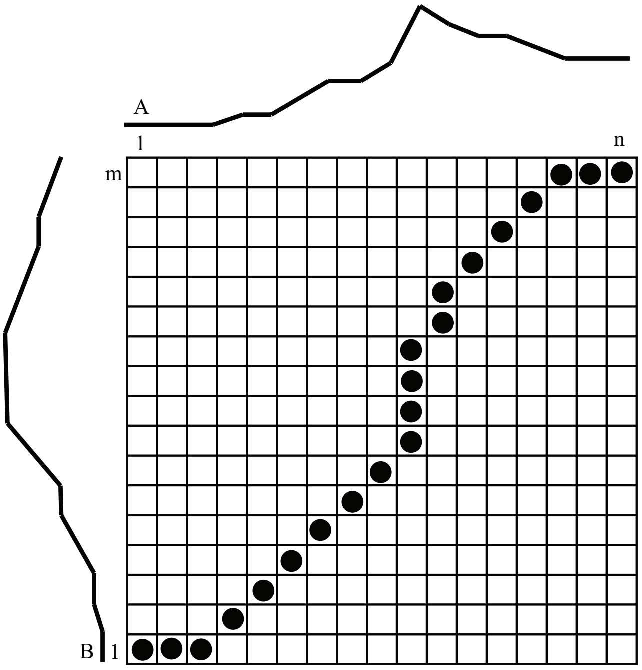


Figure 2: Minimal distance warp path between time series A and B.

The distance corresponding to the minimal distance warp path equals the value in the cell of a matrix M with the highest indexes $M(n, m)$. A minimal distance warp path can be obtained by following cells with the smallest values from $M(n, m)$ to $M(1, 1)$ (in Figure 2 the minimal distance warp path is marked with dots).

2.2 Improvements of Dynamic Time Warping

Although DTW's ability to find minimal distance warp path between time series makes it superior to simpler measures like Euclidean or Manhattan distance, its time complexity of $O(N^2)$ makes it useful only for relatively short time series. Many attempts to solve this issue have been proposed (Keogh and Pazzani 1999; Salvador and Chan 2007) which can be categorized as (Salvador and Chan 2007):

- constraints,
- data abstraction,

Constraints limit a minimum distance warp path search space by reducing allowed warp along time axis. Two most commonly used constraints are Sakoe-Chiba Band (Sakoe and Chiba 1978) and Itakura Parallelogram (Itakura 1975) which are shown in Figure 3.

Data abstraction speeds up the DTW algorithm by reducing the size of the input time series. Usually this technique speeds up DTW by a large constant factor for the price of a lower accuracy (Salvador and Chan 2007).

In this paper we are only interested in the data abstraction category. The data abstraction approach has already been used in (Keogh and Pazzani 1999) and (Salvador and Chan 2007). In (Salvador and Chan 2007), time series is reduced several times and warp path found by DTW on lower resolution time series is used to calculate DTW on higher resolution time series.

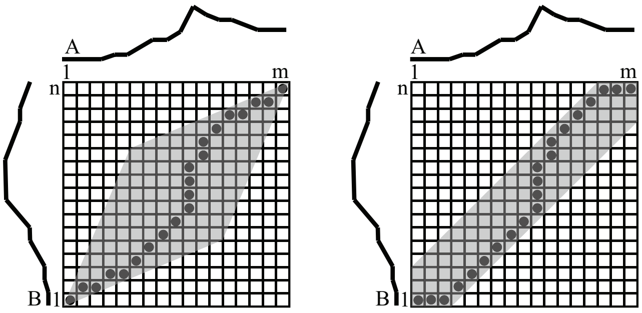


Figure 3: Itakura Parallelogram (left) and Sakoe-Chiba Band (right) constraints. Only shaded cells are used by DTW algorithm.

Data reduction is done by averaging adjacent pairs of points (data size is reduced by the factor of 2 every time resolution is decreased). In (Keogh and Pazzani 1999) a time series is approximated by a set of piecewise linear segments. The distance between segments is defined as the square of the distances of their means. Both of these approaches reduce time series size at the price of a lower accuracy. (Salvador and Chan 2007) compensate lower accuracy by calculating DTW several times on different resolution data, but data reduction part is still done at the price of information loss. Figure 4 shows a minimal distance warping path between sequences $(1, 2, 3, 4, 5)$ and $(5, 4, 3, 2, 1)$. Although they are very dissimilar, their mean values (shown as circles) are the same. This clearly shows drawbacks of data reduction by averaging, since the distance between these two segments would be 0.

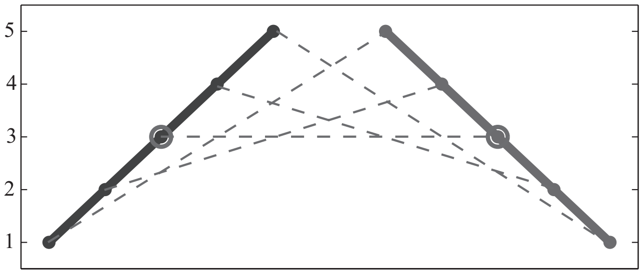


Figure 4: DTW between two time series. Circles represent mean time series value. Although the time series are not similar, their mean values are the same.

3. Qualitative Dynamic Time Warping (QDTW)

In our approach we would like to reduce time series size by removing information that is irrelevant for DTW. Our approach is based upon following theorem:

Theorem 1 *If two sequences A and B are qualitatively equal then*

$$DTW(A, B) \leq \varepsilon,$$

where

$$\varepsilon = \min(n * \text{maxdiff}(A)/2, m * \text{maxdiff}(B)/2).$$

Term $\text{maxdiff}(S)$ *is the maximal absolute difference between two adjacent elements in a time series S.*

We define two sequences to be qualitatively equal if both sequences are monotonic and their start and end values are equal. Figure 5 shows several examples of qualitatively equal sequences.

The theorem is based on the fact that in monotonic time series, the order in time (which a warp path has to respect) also corresponds to the order in the values. The theorem enables an approximation of $DTW(A, B)$ by qualitative DTW, described in the sequel. Suppose that time series A and B are samplings in time of two monotonic continuous functions of time. Then ε can be made arbitrarily small by increasing the density of sampling. Note that the sampling should be sufficiently dense w.r.t. the changes in the function value (not w.r.t. time). Consequently, if the "density approaches infinity" for any of the sequences A or B in Theorem 1, then $DTW(A, B)$ approaches 0.

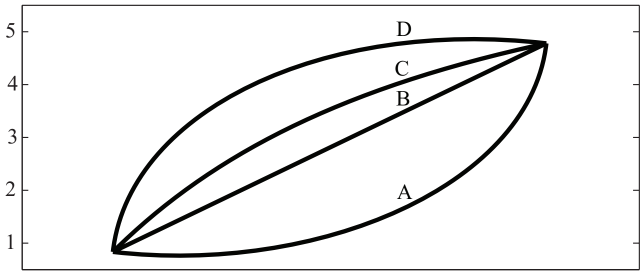


Figure 5: Four qualitatively equal sequences. DTW between any pair of them is 0.

QDTW transforms the original, numerical sequence to a qualitative sequence and then calculates DTW on the new sequence. Similar approach, where sequence is first transformed to a sequence of segments and their mean value is latter used to calculate DTW, was already proposed in (Keogh and Pazzani 1999). Main differences between approaches are in how segments are obtained and how this segments are latter used as input to the DTW. In our approach input sequences to the DTW consists of extreme points, that is the border points between the monotonic segments of the original curve (Figure 6). All monotonic segments are bound between two adjacent extreme points in the original sequence.

In our implementation, the program Qing (Žabkar et al. 2007) was used to extract the extreme points. Qing takes a sequence and a "persistence" parameter as input and returns a sequence of extreme points as output. Persistence parameter defines a minimal distance between extreme points (only extremes that differ more than persistence are returned).

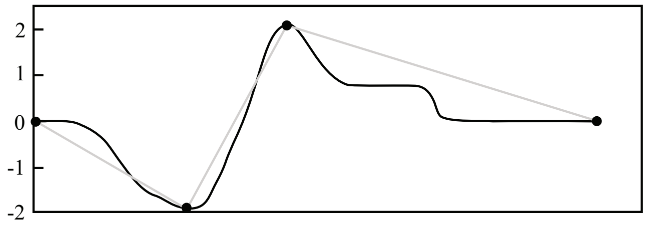


Figure 6: Example of a numerical sequence and its corresponding qualitative sequence where the black curve represents the original time series and the dots represent the extreme points - the border points between the three monotonic segments of the original curve: $(0,-2)$, $(-2,2)$, $(2,0)$. Sequence size is reduced from several points to only four points.

Consider two monotonic sequences $A = (a_1, a_2, ..., a_n)$, and $B = (b_1, b_2, ..., b_m)$. Then:

$$QDTW(A, B) = DTW((a_1, a_n), (b_1, b_m)),$$

where a_1, a_n, b_1, b_m are the extreme points. If $a_1 = b_1$ and $a_n = b_m$ then from the Theorem 1 following holds:

$$|QDTW(A, B) - DTW(A, B)| \leq \varepsilon.$$

When sequences are qualitatively equal, QDTW and DTW are almost equal (Theorem 1), otherwise problems can arise. There are two possible ways of violating the conditions for the applicability of Theorem 1:

- Extreme points do not coincide.
- Sequences are not monotonic.

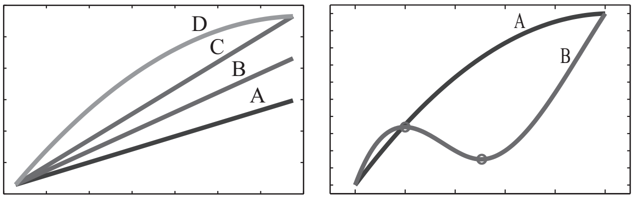


Figure 7: Possible violations of the conditions for the applicability of Theorem 1.

An example of monotonic sequences where the extreme points do not coincide is shown on the left side of Figure 7. It is obvious that DTW distance between base sequence A and any of the target sequences B, C, D is not necessarily the same as QDTW distance. More than in the actual values, we are interested in the distance order of target sequences B, C, D, when compared to base sequence A:

$$DTW(A, D) > DTW(A, C) > DTW(A, B),$$

$$QDTW(A, D) = QDTW(A, C) > QDTW(A, B).$$

When sequences with different extreme points (B, C) are compared to the base sequence (A), the order is preserved. In the case that target sequences have the same extreme points (C, D), QDTW cannot distinguish between them, when compared to the base sequence (A).

On the right hand side of Figure 7, a monotonic sequence is compared to a sequence that is not monotonic. If non monotonic part of sequence B (segment between two dots) is not detected (this can be due to high persistence parameter in the Qing algorithm), then both sequences have the same extreme points and $QDTW(A, B) = 0$, while $DTW(A, B) > 0$. On the other hand, if the decreasing part of sequence B is detected (small persistence), then sequence B is split into three segments by four extreme points. $QDTW(A, B)$ is calculated between the sequence of two extreme points from A and the sequence of four extreme points from B. As inner extreme points from B (bounding monotonically decreasing segment) have to map to extreme points from A, $QDTW(A, B)$ distance between A and B is quite large. With increasing number of short segments that map to one long segment, QDTW distance quickly increases. For now this represents the biggest problem of QDTW approach and should be solved in the future work.

Although, as we have shown, QDTW is not completely insensitive to information loss due to data reduction, we believe this will not significantly influence classification accuracy, and improved running time over DTW will more than compensate for slightly lower accuracy. The experimental evaluation that follows investigates this expectation.

4. Experimental evaluation

DTW is commonly used in time series classification domains. In these domains similarity or dissimilarity between time series determine whether time series belong to the same class or not. Therefore, similarity measure between time series is crucial part of the classification algorithm. Theorem 1 ensures that QDTW performs nearly the same as DTW if time series consist of qualitatively equal segments. This condition is rather strong. True applicability of QDTW can only be revealed with experimental evaluation on real world domains where conditions of Theorem 1 are not necessarily satisfied. With experimental evaluation, we would like to investigate how well QDTW performs in comparison to classic DTW in classification tasks. We are mostly interested in classification accuracy and execution time. The method was evaluated on three domains with different time series characteristics. Following data sets were used:

- **Australian Sign Language signs (High Quality) Data Set** (Kadous and Sammut 2002): The data set consists of the readings from 22 sensors that measure native signer hand position (11 sensors per hand) in time while signing one of 95 Auslan signs. For each Auslan sign 27 examples were recorded (total of 2565 examples). Due to DTWs high time complexity, only a subset of the original dataset was used. The subset consists of examples of the following ten signs: spend, lose, forget, innocent, Norway, happy, later, eat, cold, crazy.
- **Character Trajectories Data Set** (Asuncion and Newman 2007): The data set consists of 3-dimensional pen tip velocity trajectories which were recorded whilst writing individual characters. There are 20 different characters

in the data set. All of 2858 examples were captured by the same person using WACOM tablet. Due to the DTW time complexity only one seventh of the original examples were used (every seventh example from the original data set was included in the subset without changing the order of examples in the original dataset). All of the character labels (20) were included in the subset.

- **Character Recognition Data Set**: The data set consists of data from three sensors that measure the subject's hand acceleration while writing individual characters. There are 26 different characters in the data set. All of the 391 examples were obtained by the same person using tri-axis accelerometer.

4.1 Accuracy

In this section we are interested in how well QDTW performs in comparison to DTW and how different persistence settings effect classification accuracy. Classification was done using weighted k-nearest neighbor (k=3) algorithm using DTW or QDTW as similarity measure. The leave one out approach was used to estimate classification accuracy. QDTW method was evaluated using several relative persistence settings: 0.1, 0.2, 0.4 and 0.6. For each time series, persistence is obtained by multiplying relative persistence with the difference between time series maximum and minimum value.

As all the datasets consist of several variables (multivariate time series domains), any of these variables can be used for evaluation. Some of these variables are highly informative (similar examples belong to the same class while dissimilar examples belong to different classes) while others may not correlate with the class (random variables). On random variables, any similarity measure will behave similarly to a random similarity measure, so it makes sense to evaluate similarity measures only on highly informative variables. For this reason one variable, where DTW performs best, is used from each dataset to compare QDTW to DTW. These variables are: 'ryaw', 'y' and 'accY' from Australian Sign Language signs, Character Trajectories and Character Recognition datasets respectively. Classification accuracies using DTW and QDTW with different persistence settings are shown in Figure 8.

In comparison to DTW, QDTW (p=0.1) performed best on Australian Sign Language signs dataset where the difference between classification accuracies is only 0.01 (1.3%). QDTW performed worst on Character Recognition dataset where classification accuracy dropped by nearly 16% in comparison to DTW (from 0.88 for DTW to 0.74 for QDTW with persistence setting 0.1).

To evaluate how persistence affects classification accuracy, DTW and QDTW results for different relative persistence values (0.1, 0.2, 0.4 and 0.6) are ranked from best (1) to worst (5). For each dataset, average rank over all variables is calculated. Results are summarized in Table 1.

Table 1 confirms, as expected, that classification accuracy decreases with increasing relative persistence. The only domain where in some cases accuracy improved with increased relative persistence is Australian Sign Language domain.

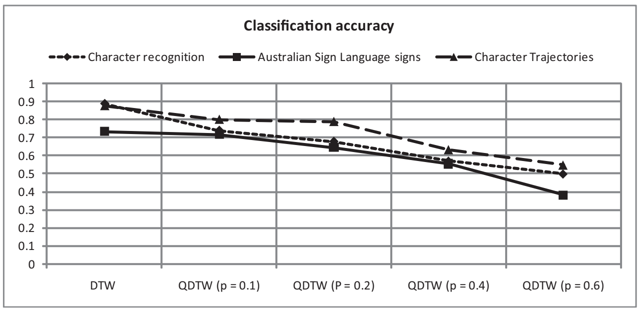


Figure 8: Classification accuracies for DTW and QDTW similarity measures where p denotes different relative persistence settings. Classification accuracies (shown from left to right) are for Australian Sign Language signs dataset: 0.73, 0.72, 0.64, 0.56, 0.38, for Character Trajectories dataset: 0.88, 0.80, 0.79, 0.63, 0.55 and for Character Recognition dataset: 0.88, 0.74, 0.68, 0.57, 0.50.

Table 1: Average rank for different relative persistence settings.

Method	Australian	Character Trajec.	Character Recog.
DTW	2.09	1	1
QDTW p=0.1	2.20	2	2
QDTW p=0.2	2.98	3	3
QDTW p=0.4	3.68	4.17	4
QDTW p=0.6	4.05	4.83	5

This can happen due to the presence of noise in some of its attributes, which can be removed only by more robust qualitative models. Overall, smaller relative persistence means larger classification accuracy in all evaluated datasets.

4.2 Efficiency

In this section we are interested in time efficiency of QDTW algorithm. Time efficiency is estimated with the number of distance calculations between two values of time series (size of the cost matrix M) which are needed for calculating DTW or QDTW similarity between two time series. Before calculating similarity, QDTW needs to transform time series to qualitative representation. As Qing is very efficient for qualitative modeling of time series, time to build qualitative models is insignificant in comparison to the time needed to calculate similarity and is thus omitted.

Time efficiency was estimated on all three datasets using variables 'ryaw', 'y' and 'accY' from Australian Sign Language signs, Character Trajectories and Character Recognition dataset respectively. For each dataset, similarity between all pairs of examples was calculated and average size of the cost matrix M ($M = m * n$, where m and n are time series lengths) is returned as a result. Figure 9 shows average size of the cost matrix M for calculating DTW and QDTW for all three domains.

From Figure 9, it is evident that QDTW was much faster than DTW on all three domains. Even for small persistence

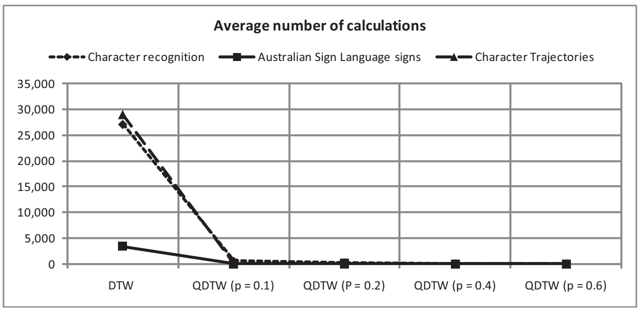


Figure 9: Average number of performed distance calculations between two values of time series when calculating similarity using DTW or QDTW where p is relative persistence setting. Average number of distance calculations (shown from left to right) are for Australian Sign Language signs dataset: 3382, 35, 25, 16, 9, for Character Trajectories dataset: 28893, 34, 30, 21, 13 and for Character Recognition dataset: 27058, 528, 182, 42, 19.

values, the savings in the number of distance calculations between two values of time series (size of the cost matrix M) are enormous (speed up by factor of nearly 100 on Australian Sign Language signs dataset , to nearly 850 on Character Trajectories dataset).

Besides comparison of QDTW to DTW, we are also interested in how different persistence settings effect time efficiency. Figure 10 shows average number of performed distance calculations between two values of time series for different relative persistence values. It can be seen from Figure 10 that the average number of performed distance calculations between two values of time series is decreasing with higher persistence values. The results also show that similar persistence values on different domains do not necessarily mean similar savings in time. This means QDTWs performance is not only persistence dependent but also domain dependent.

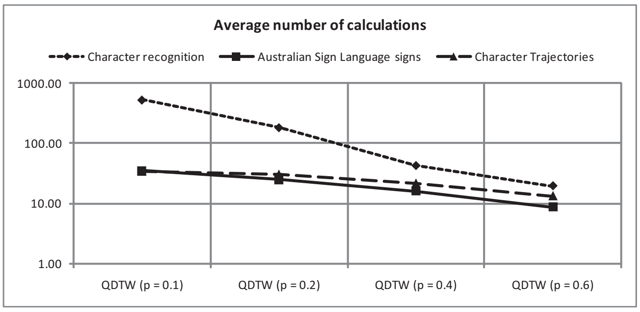


Figure 10: Average number of performed distance calculations between two values of time series (shown on logarithmic scale) when calculating QDTW similarity with different relative persistence settings (p).

5. Conclusions and future work

In this paper, we have stated a new theorem (Theorem 1), which explains when time series data can be reduced without loss of information relevant to DTW. Shortcomings of data reduction by averaging have been explained and new algorithm QDTW (Qualitative Dynamic Time Warping) have been introduced. QDTW is a modification of DTW algorithm, which is based on Theorem 1. It transforms time series data into qualitative series and thus significantly reduces data size. Experimental results have shown up to 1000 times speed-up with respect to the DTW algorithm. These significant improvements in efficiency are often obtained at acceptable loss in classification accuracy. QDTW major drawbacks are its inability to guarantee bounds on deviations from the optimal warp path solution, and its domain dependent efficiency. In future work, we will try to improve QDTW accuracy by reducing errors due to violations of the conditions for the applicability of Theorem 1. Special attention will be devoted to problems which arise due to non-monotonicity of segments, which is sometimes discovered by QING, while sometimes it is not. In these cases, we are comparing sequences with large number of short segments and sequences with small number of long segments, which usually results in a poor estimation of distance given by QDTW.

Acknowledgments

This work was partly funded by the X-Media project (www.x-media-project.org) sponsored by the European Commission as part of the Information Society Technologies (IST) programme under EC grant number IST-FP6-026978, and Slovene Agency for Research and Development (ARRS).

References

Asuncion, A., and Newman, D. 2007. UCI machine learning repository.

Bratko, I. 2000. *Prolog Programming for Artificial Intelligence*. Addison Wesley.

Itakura, F. 1975. Minimum prediction residual principle applied to speech recognition. *Acoustics, Speech and Signal Processing, IEEE Transactions on* 23(1):67–72.

Kadous, M. W., and Sammut, S. C. 2002. Temporal classification: Extending the classification paradigm to multivariate time series. Technical report, University of New South Wales, School of Computer Science and Engineering.

Keogh, E. J., and Pazzani, M. J. 1999. Scaling up dynamic time warping to massive dataset. In *PKDD '99: Proceedings of the Third European Conference on Principles of Data Mining and Knowledge Discovery*, 1–11. London, UK: Springer-Verlag.

Keogh, E. J., and Pazzani, M. J. 2001. Derivative dynamic time warping. In *In First SIAM International Conference on Data Mining (SDM2001*.

Sakoe, H., and Chiba, S. 1978. Dynamic programming algorithm optimization for spoken word recognition. *Acoustics, Speech and Signal Processing, IEEE Transactions on* 26(1):43–49.

Salvador, S., and Chan, P. 2007. Toward accurate dynamic time warping in linear time and space. *Intell. Data Anal.* 11(5):561–580.

Žabkar, J.; Jerše, G.; Mramor, N.; and Bratko, I. 2007. Induction of qualitative models using discrete morse theory. In *Proceedings of the 21st Workshop on Qualitative Reasoning*.

Modeling for Fault Localization in Data Warehouse Applications

[1]Peter Struss, [2]Vikas Shivashankar, [3]Mohamed Zahoor

[1]Technische Universität München, Germany, struss@in.tum.de, [2]IIT Madras, Chennai, India,
[3]Ericsson India Private Limited, Chennai, India

Abstract

The paper describes first results of an attempt to develop a general tool for localizing faults in applications of data warehouse technology. Genericity is achieved by a model-based approach: a model of the application is configured from a library of models of standard (types of) modules and exploited by a consistency-based diagnosis algorithm, originally used for diagnosing physical devices. In order to obtain discriminating interdependencies, the behavior description in the models is stratified according to different roles and processing of the various types of the data and captures the potential impact of faults of process steps and data transfer on the data as well as on sets of data. Reflecting the nature of the initial symptoms and of the potential checks, these descriptions are stated at a qualitative level. In the current solution, the symptoms are assumed to stem from human assessment of reports generated from the data ware house, while checks can be inspection of the data base or other persistent data and rerunning certain process steps. The solution has been validated in customer report generation of a provider of mobile phone services.

1. Introduction

One of the most urgent needs these days is to effectively support debugging of software, which becomes an ever increasing factor to determine both the industrial and commercial sphere and our personal lives. One of the most successful techniques of model-based problem solving is component-oriented consistency-based diagnosis (see [Struss 08]). Exploiting this technology, which has helped to localize and identify faults in devices, for software debugging has been pursued for quite some time (see [Struss 08] for some references).

There are a number of obstacles that hamper a straightforward transfer of consistency-based diagnosis techniques to software debugging. The most fundamental one is the difference between **diagnosis** of (well-designed) artifacts and **debugging** of software: while the former aims at identifying or localizing the deviation of a faulty realization from a **correct design**, the latter is concerned with identifying or localizing the reason why an **incorrect design** fails to meet the specification.

The second obstacle is modeling itself: at the code level, a component-oriented model becomes too complex and prevents a solution to scaling up to interesting programs, whereas at a very high level of software modules, the models tend to become very specific and are not reusable across different problem instances, which results in a (usually inhibitive) high cost of modeling.

Thirdly, while modeling the possible faults is often straightforward for physical systems (a shorted resistor is consistent with an increased current, but an open one is not), modeling faults in software is usually infeasible, because the space of programmers' faults is infinite.

The work we presented here is guided by the idea that classes of certain standardized software applications may help to overcome the abovementioned obstacles by providing an intermediate level of abstraction that allows for reusable models of standard software modules and, especially for generic fault models – an approach we have not encountered in the existing literature on model-based software debugging.

In this paper, we address fault localization in data warehouse applications as an instance of such a class of standardized software applications.

The next section introduces the foundations of this application area and describes our specific project: a data warehouse application of a communication network provider in India. After a brief characterization of component-oriented consistency-based diagnosis, section 4 presents the core contribution of this paper, the foundations and examples of generic models for debugging of data warehouse applications. We then present the specialization to an application in customer report generation of a provider of mobile phone services (section 6) and discuss the results of an initial validation of the approach and future work.

2. Application Domain: Data Warehousing

2.1 General Background

Data Warehousing and On-Line Analytical Processing (OLAP) are essential elements in decision support systems. Nowadays, there is a need to not only manage huge amounts of data, but also an equally, if not more, important requirement of analyzing this data and extracting useful information, and data warehousing technologies support this. Many commercial products and tools in this area are now available, aiming at enabling faster and more informed decision making.

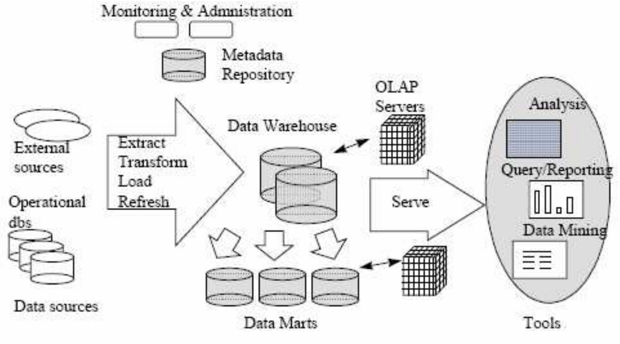


Figure 1 Architecture of a generic data warehouse system

A data warehouse is a "subject-oriented, integrated, time-varying, non-volatile collection of data that is used primarily in organizational decision making."[Inmon 92]. The aim of data warehousing technologies is different from that of operational databases, which typically take care of day-to-day transactions. Unlike the latter, the focus in data warehousing is decision support, and, hence, summarized and consolidated data are more important than individual records. Data warehouses are **orders of magnitude larger** than typical databases and their main bottleneck is in answering complex ad-hoc queries involving scans, joins and aggregations typically over millions of records. Therefore, data warehousing technologies are becoming more sophisticated, complex and, as a result, more fault-prone, as well.

The general architecture of a data warehousing system is as shown in **Figure 1** [Chaudhuri 97]. The major modules in such a system are:

- **Pre-processing** – This set of modules deals with the cleaning of data, normalization of certain fields and other pre-processing methods needed to bring the data to a common standard format.
- **Loading of the data warehouse** – This deals with the loading of the pre-processed data appropriately into the warehouse.
- **Summarization and consolidation using data marts** – This includes aggregating and consolidating the warehouse data and storing it into customized databases called data marts.

Therefore, a typical cycle in a Data Warehousing application is:

- Arrival of new data
- Pre-processing of the data
- Loading into the data warehouse
- Consolidation of new data with old data
- Storing consolidated data into data marts

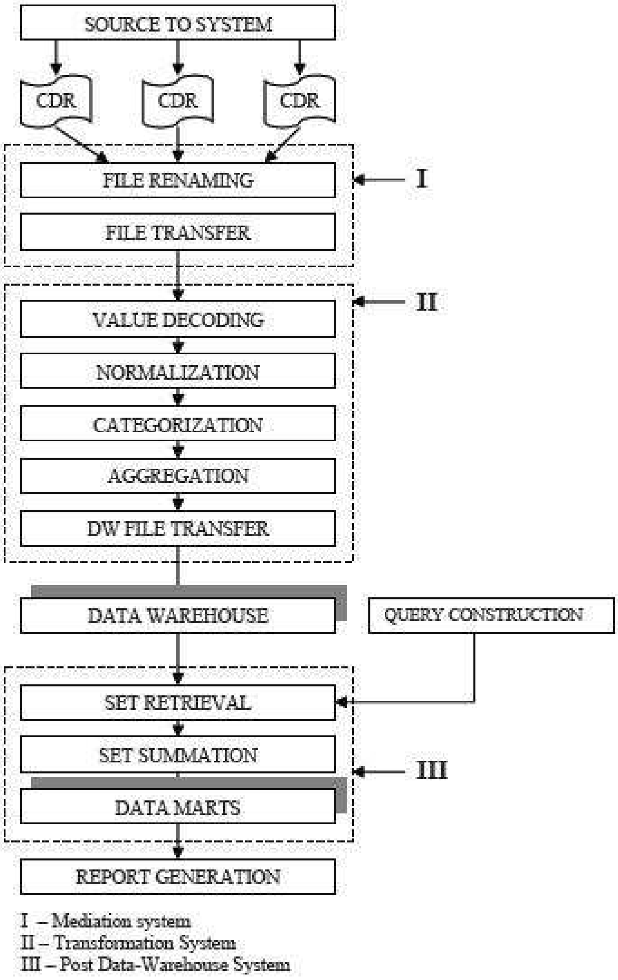


Figure 2: Process diagram of a report generation system based on call data

2.2 Report Generation Based on Call Data

The report generation tool is a system (**Figure 2**) used to generate useful information from consumer usage records known as Call Data Records (CDR). The CDRs are generated by a number of network nodes operating in different regions and contain data such as duration of the call (in case of normal calls), data volume transferred (in case of a GPRS call), source and destination numbers, cost of the call, location identifiers of source and destination regions. The data is subject to various pre-processing steps in the Data Warehousing System (DWS) and then loaded into the data warehouse.

Extract-transform-and-load operations are then applied to the warehouse data to obtain customized figures, such as countrywide aggregate revenue for a given time period (e.g. a month), total revenue from a particular region, number of active subscribers in a given region, the liability of the service providers to the customers, the region-wise distribution of network usage etc., which are then stored in specialized data warehouses known as **data marts**.

Updates to the data marts are typically done on a daily basis. From the data marts customized reports are generated. For instance, the **balance report** shows the total account balance of the subscriber base on a given date, thus used for reporting the operator's liability.

Another example is the **customer usage report**, which gives information about the usage statistics of the customer base for a given period of time, both for voice as well as GPRS calls.

A detailed process diagram is as shown in Figure 2. Once a CDR file is received from the source nodes, the **mediation** module processes it and renames the CDR file, assigning it a unique sequence number. After this, the CDR file is transferred via FTP to the **transformation system** for further pre-processing.

The **collection** engine of the transformation system monitors the directories for any incoming CDRs from the mediation system. Once a file of CDRs is received, the engine transforms each CDR into an internal data structure in the **value decoding** module. The **processing engine** checks the CDR for mandatory fields, the **normalization** module normalizes all numbers to a uniform format, and the **categorization** stage attaches tags to the CDR based on values of certain fields, such as tagging the records as local, national or international according to the source and destination numbers. The **aggregation** step performs the combination of multiple fields into one, deriving a new field based on certain existing fields etc. For instance, this step combines the local timestamp field and the time zone information in the CDR to generate a UTC timestamp. The CDR is now stored in another data structure and passed onto the **distribution** engine, which transfers all processed CDRs from the data structure to an output file. Once this is done for all the CDRs, they are stored into the data warehouse.

Extract-transform-and-load operations are carried out periodically on the data warehouse to populate customized consolidated values into the data marts. There are different kinds of data marts based on their functionality, such as financial, usage-level and subscriber-life-cycle data marts. The consolidated values in the data marts are then visualized using a customized **report generation** system as shown in **Figure 2**.

During the various processing and transfer steps, data can be corrupted in many ways and lead to missing or wrong data stored in the data warehouse and/or the data marts or appearing in the reports. For instance, a breakdown in the network connectivity during the transfer of CDRs into the warehouse might lead to incomplete data in the warehouse, thus leading to faults downstream. Usually, such defects are not detected until some results in the reports are identified as obviously incorrect, e.g. the total revenue for a time period being orders of magnitude smaller or larger than expected. Localizing the cause for this deviation in the entire process chain can be a tedious and time-consuming task for the staff. Some reasons for this are frequent changes in the structure and modules of the system, the fact that most intermediate results are not persistent and high efforts to rerun parts of the process.

The following is a typical fault scenario encountered in the application where the total **number of active subscribers** in the system according to a generated report was not matching the expected value. To check whether the fault was produced during report generation, the data marts were inspected. When the same error was found in the data marts (thus implying that the fault was created upstream in the process), the warehouse data was then checked for errors. When the warehouse data was found to be **OK** (and yet the value in the data marts was wrong), it was concluded that there is an error with the **set retrieval module logic**. The code, after being checked, was indeed found to be buggy.

3. Component-oriented Consistency-based Diagnosis

The description of the system and the task suggests a perspective of "Localizing the fault in one component of the system as the possible cause of its misbehavior". Component-oriented consistency-based diagnosis (see [Struss 08]) has been developed as a solution to diagnosis of a broad class of physical artifacts. In a nutshell, it can be informally described as follows: the behavior of each component (type) of a system is modeled in a context-independent manner. Each component C_j can be in one of different behavior modes $\text{mode}_i(C_j)$. The correct or intended behavior mode (OK) is one of them, and others are either simply its negation or a list of specific (classes of) misbehaviors (such as "open" or "shorted" for a resistor). An overall system model is (automatically) configured according to the system structure (i.e. the interconnectivity of the components) for a mode assignment

$$\text{MA} = \{\text{mode}_i(C_j)\},$$

which specifies a unique behavior mode for each component.

A diagnosis is obtained as a mode assignment MA whose model is consistent with the observations:

$$\text{MODEL(MA)} \cup \text{OBS} \nvdash \bot .$$

Even if only the OK modes have an associated model, this yields fault localization. If models of the various fault modes exist, then fault identification can be performed and fault localization can be more confined.

Despite a number of obstacles, that were mentioned in the introduction, the principles and techniques of component-oriented consistency-based diagnosis can be exploited for fault localization in programs under certain conditions.

4. Diagnostic Model of Data Warehouse Applications

4.1 The Main Ideas

The overall process described in section 2.2 is a sequence of steps all data have to go through to ultimately yield a result in a report. If a wrong result is detected, each of these steps may be suspected to have caused it. A straightforward application of consistency-based diagnosis as described in section 3 (with each step modeled as a component in a linear structure) will produce exactly this result. Both for a human and a (semi-)automatic debugging aid, there are three basic ways to reduce the set of diagnostic candidates and finally obtain a fault localization:

- Collect **more observations**. In our application, this means checking intermediate data. Besides the data warehouse and the data marts, the only persistent data are **the output of the mediation system**. Inspecting more intermediate results requires re-running the steps, which is time-consuming and should be done only after having confined the location of the fault as precisely as possible by the following means.
- Use **fault models**. In contrast to physical systems, it is impossible to find a small set of models covering the abnormal behavior of pieces of software in the general case. However, at the abstract level of the functional description of a data warehouse application, it becomes feasible to describe some plausible improper behaviors of a module. This becomes even more powerful together with the third step.
- **Refine** the **structure**. This is achieved by stratifying the data according to their type and role in the process. Different steps affect different fields of the record, and so do faults in these steps. For instance, a bug in normalization of a temporal representation may corrupt the time information, but leaves location information unchanged. And an incomplete transmission of data truncates a set of records, but leaves the content unmodified.

The last example illustrates the need to not only model the manipulation of the content of records, but explicitly represent and propagate properties of **record sets**. If the record, say, for a particular day is incomplete, then summing up some numerical information will yield a number which is too small.

This in turn motivates the modeling principle chosen: the models capture the **deviation** of properties of **data fields** or **sets** from those that would have been obtained if everything had worked as planned. Starting from an observed deviation of some report result, the system is going to identify models of the entire process that are consistent with this deviation. In this abstract representation, the references for the deviations remain implicit and dependent on the context: they are given by whatever are the outputs of the various steps that the respective report result depends on.

4.2 Partitioning of the Data

In this section, we present a general principle for partitioning the data for the debugging purpose. The rationale behind this is the fact that software modules only refer to certain parts of the data and also modify only certain fields on the data. Therefore, each module **induces** a **partition** of the data fields, basically into **relevant** and **irrelevant** to the function of the module. Relevant fields are those that are either referred to or modified by the module. Our strategy is, therefore, to construct a global partitioning that respects all local partitions.

This can be formalized as follows: For each module M_i and fields $f_j \subset F$ from the data records:

A_i is the set of fields $f_j \in F$ of the input whose content may **affect** the result, both under normal and abnormal behavior,

E_i is the set of fields $f_j \in F$ of the output that are **effects** of the processing of the module under normal and abnormal behavior.

In addition, each field $f_j \in F$ has a type $T(f_j)$ which influences the (description of the) potential deviations that it can exhibit such as **Numerical**, **String** etc. (see following subsection).

Based on the local partitioning are found, the global partitioning is defined as the one that respects all local partitions and the type, with the partitions being maximal:

$$\begin{aligned} \exists\, k, f_{l,} f_m \in P_k \Leftrightarrow (\forall i \quad & (f_l \in A_i \Leftrightarrow f_m \in A_i) \\ & \wedge (f_l \in E_i \Leftrightarrow f_m \in E_i)\ ) \\ & \wedge\ (T(f_l) = T(f_m))) \end{aligned}$$

For example, in case of the aggregation module, A_i represents the fields that are aggregated and E_i the aggregated field. Similarly, for the retrieval module, A_i are the keys to the query while E_i comprises the selected output fields.

4.3 Types of Fields and their Domains

The data fields and the data occurring in the query and report generation steps are categorized into **numerical** (such as duration of a call in our application), **categorical** (such as source and destination phone numbers), and **string** (such as a database query). We use the following domains, which capture the deviation of an actual value of a variable, X, from some reference value, X_{ref}:

Numerical = {Ok, -, --, +, ++, oppSign}, where

- **Ok** if $X = X_{ref}$
- **oppSign** if $(X * X_{ref} < 0)$
- **-** if $(X * X_{ref} >= 0) \wedge (X < X_{ref}) \wedge \neg (X << X_{ref})$
- **--** if $(X * X_{ref} >= 0) \wedge (X << X_{ref})$
- **+** if $(X * X_{ref} >= 0) \wedge (X > X_{ref}) \wedge \neg (X >> X_{ref})$
- ++ if $(X * X_{ref} >= 0) \wedge (X >> X_{ref})$

Categorical = {Ok, Wrong}, where
- **Ok** if $X = X_{ref}$
- **Wrong** if $X \neq X_{ref}$

String = {Ok, Null, Wrong, SynWrong},where
- **Ok** if $X = X_{ref}$
- **Null** if $(X = null) \wedge \neg (X = X_{ref})$
- **Wrong** if $\neg (X = null) \wedge \neg (X = X_{ref}) \wedge (X \text{ is valid})$
- **SynWrong** if $\neg (X = null) \wedge \neg (X = X_{ref}) \wedge \neg (X \text{ is valid})$

The motivation for valid, invalid and null strings is predominantly to capture features of database queries: **valid** strings are those which are syntactically correct (i.e. which will execute without an exception on a database), whereas **invalid** strings are those which will result in an error when executed on a database. **Null** strings are also used to handle the case when the string construction module failed **completely,** resulting in an empty string.

As explained above, the model also captures explicitly how a set of data, DS, which is processed, is related to the data that should be processed in the proper process, DS_{ref}. The domain of the respective variable is

Set = {Ok, Empty, Subset, Superset, Wrong}, where
- **Ok** if $DS = DS_{ref}$
- **Empty** if $(DS = \{\}) \wedge \neg (DS = DS_{ref})$
- **Subset** if $\neg (DS = \{\}) \wedge (DS \subset DS_{ref})$
- **Superset** if $\neg (DS = \{\}) \wedge (DS \supset DS_{ref})$
- **Wrong** if $\neg (DS \subset DS_{ref}) \wedge \neg (DS_{ref} \subset DS) \wedge \neg (DS = DS_{ref})$

4.4 Models

Once the stratification of data into appropriate groups is established, models of individual components capturing both the desired and possible faulty behaviors can be designed, capturing the information about how a component treats the abovementioned **partitions** of a record. In the following, we present some examples from the model library.

File transfer component. If we consider the File Transfer component (which, in our application, handles the transfer of files containing CDRs across a network), we know that only the 'record set' property can be affected, i.e. if the transfer is not successful, either the file transfer was incomplete (nevertheless preserving the integrity of an individual record) or nothing at all was transferred, resulting in a completely unsuccessful transfer. A full description of the model of this component is shown in **Table 1**.

As can be observed from the table, in the OK mode of the component, the set property of the CDR file is simply propagated, i.e. output of the component is identical to its input.

Table 1 : Model of the File transfer Component

STATUS	Input.set	Output.set
OK	Ok	Ok
	Wrong	Wrong
	Empty	Empty
	Subset	Subset
	Superset	Superset
CONNECTION DISRUPTED	*	Subset
	*	Empty
	Superset	Wrong

Table 2: Model of the Query construction Component

STATUS	qStrTemplate	qCriteria	qString
OK	Ok	Ok	Ok
	Ok	Wrong	Wrong
	Wrong	*	Wrong
	Wrong	*	SynWrong
FAULTY	*	*	Wrong
	*	*	SynWrong

However, in the fault mode when the FTP connection is broken, the model captures the fact that no matter what the nature of the input, the output could be either a **Subset** of the original data (resulting from a partial loss in connectivity) or an **Empty** set (resulting from a complete loss of connectivity). In addition, if the input is a **Superset**, the output after truncation can be a **Wrong** set (which means, we ignore the highly unlikely case that transaction incidentally produces the proper set).

However, an assumption made while building this model is that the file transfer component never spoils the integrity of the data and only can disrupt the set property, which is indeed true in our case study.

In our application,this model is used in different places in the process: the data transfer to the transformation system and the transfer into the data warehouse.

Query construction component. This takes as input a query template, **qStringTemplate**, with placeholders for variables and categorical variables, **qCriteria** containing values for these placeholders, and produces a query string, **qString**. It is used to construct queries automatically in order to retrieve desired information from the data warehouse. The model of this component is described in **Table 2**. In the **OK** mode of operation, if both inputs are Ok, the output is Ok. If not, the output takes appropriate values for different input cases as shown in the table.

In the **FAULTY** mode of operation, no matter what the values of the input are, the output string can take the values Wrong or SynWrong.

Set retrieval component. As a final example, we consider the component that retrieves relevant data from the data warehouse for a particular operation (e.g. to calculate total revenue for a particular period, this module extracts the per-CDR revenue data) which then may be given as input to a module that performs an operation on this data (such as the summation component). The inputs to this component are the query string for the actual retrieval, **qString**, the data set on which the query operates, **inputSet**, and **selectKey**, which determines the required field (e.g. the revenue per CDR) and generates the relevant subset of data, **outputSet**. A complete description of the model is given in **Table 3**.

In a similar manner, the other components are modeled, capturing both the normal and deviant behavior with appropriate fault modes.

It should be noted as an important disadvantage that the global partitioning, being dependent on the local ones, may have to be changed if new modules are introduced or the records are modified. In order to obtain truly generic models, in a future solution, they should be stated in abstract terms of their sets A_i, E_i, $F\backslash(A_i \quad E_i)$ and the mapping to the record fields should be represented separately.

5. Structuring the Call Data

Based the principles of section 4.2, the fields of the CDR were grouped into the following 9 groups:

- **CDR Information** – this group deals with CDR-specific information such as CDR identifier.
- **Account Information** – this deals with the account information of the subscriber, such as the plan being used, the base location of the subscriber etc.
- **Call-Information** – this gives information about the source and destination phone numbers, whether they are roaming or not etc
- **Cost-Information** – this gives information about the rates that the subscriber will be charged for this call
- **Duration of Call** – gives the duration of the call
- **Location-Information** – gives the location identifiers of the subscribers
- **Data Volume** – gives the data volume transferred in case of a GPRS call
- **Timestamp of call** – gives the time at which the call began
- **Final-charge of call** – gives the final amount that the subscribers are charged.

In addition, the models propagate

- **Set Information** - dealing with the set property of a file of CDRs.

6. Validation of the Diagnostic Model

So far, the models were validated against a small set of typical and representative scenarios (motivated by real cases), and the fault localization of the diagnosis tool under the available observation was compared to the manual debugging steps. We present two of these cases in the following.

Table 3: Model of the Set retrieval Component

STATUS	qString	inputSet	selectKey	outputSet
OK	Wrong	*	*	Wrong
	*	Wrong	*	Wrong
	*	*	Wrong	Wrong
	Wrong	*	*	Subset
	*	Subset	*	Subset
	*	Wrong	*	Subset
	*	*	Wrong	Subset
	Wrong	*	*	Superset
	*	Superset	*	Superset
	*	*	Wrong	Superset
	Wrong	*	*	Empty
	*	Empty	*	Empty
	*	Wrong	*	Empty
	*	*	Wrong	Empty
	Ok	Ok	Ok	Ok
FAULTY	*	*	*	Empty
	*	*	*	Subset
	*	*	*	Wrong
	*	*	*	Superset

6.1 Scenario One: Consumer Usage Amount Less than Expected Value.

In this scenario, it was observed that the customer usage amount displayed in the report generated by the system is less than the expected value.

The steps taken to manually localize the fault were as follows:

1. **Generate report** – erroneous value present in report
2. **Probe data marts** – erroneous value present in data mart (implying that the cause for the fault is upstream)
3. **Query data warehouse –** correct **duration** values are present in the data warehouse (implying that something is wrong with the selection criteria in the query or selectKeys, in this case, the timestamps)
4. **Analyze the number of CDRs in result set** – does not match with expected value
5. **Analyze timestamp of a CDR and compare with output of mediation system** – does not match

Therefore, the diagnosis was '**Erroneous timestamp calculation**' and indeed, the **aggregation** component containing the timestamp calculation code was found to be buggy.

The steps taken to localize the fault using the model-based diagnosis system (summarized in Table **4**) were:

1. **Initialize given evidence**, i.e. **Total duration** as observed in data marts is '–'(step number 2 in the manual debugging). With this evidence as input, the diagnosis algorithm outputs all consistent diagnoses as shown in the first column of Table **4.**
2. **Output of Set retrieval module is Wrong** (step number 4 in the manual debugging) **-** exonerates the **Set Summation** module (since the fault has occurred before this component was used).
3. **Time Info in the data warehouse is Wrong** (step number 3 in the manual debugging) **-** eliminates a number of candidate diagnoses leaving the 4 diagnoses in column 3 of the table.
4. **Time Info at output of Mediation module is Ok** (step number 5 in the manual debugging) **-** exonerates the 'Source to System' component.

This leaves us with three suspect modules for more detailed probing and debugging, including the component that was actually found to be faulty, namely the **aggregation** component.

6.2 Scenario Two: Number of Active Subscribers not Matching Expected Value.

In this scenario, the starting point is an error in the report summarizing the active subscriber statistics. The manual debugging procedure required **4** probes to narrow down onto the module causing the fault, the **Set retrieval** component, which are:

1. **Generate report** – erroneous value in report
2. **Probe data marts** – erroneous value present in data mart (implying that the cause for the fault is upstream)
3. **Run query on data warehouse** – correct value is obtained, indicating the problem is downstream from the data warehouse.
4. **Analyze the Set Retrieval component –** found to be buggy.

With the help of the diagnosis engine the faulty module is sequentially localized as shown in **Table 5**.

The cases provide some evidence that component-oriented consistency-based diagnosis provides the basis for a useful debugging aid. More specifically, the level of abstraction of the component models appears to be expressive enough for the task. This indicates that the tool may indeed successfully guide a human debugger without requiring him to have deep detailed knowledge about the system structure, the modules, recent modifications etc. any more**.** This is possible since this domain knowledge about the system is now incorporated into the model. Therefore, at least for a set of common sources of errors, a person not too experienced with the data warehouse system can perform debugging, which was previously impossible.

7. Future Work

In this paper, we described the models for consistency-based debugging of a data warehouse application and its validation. So far, only the diagnostic part has been realized. For a real debugging aid, a module has to be integrated that proposes "probes", i.e. inspection of persistent data and rerunning process steps. More scenarios will be treated to establish the basis for making a business case that justifies the development of a tool for everyday use in this area.

Table 4: Debugging Trace for Scenario 1. The evidence is incrementally added in order to obtain focused diagnoses as is shown by the monotonically shrinking diagnosis set. "X" means that the respective module is no longer a (minimal) diagnosis

Evidence 1: Output duration total is **less** than expected	**Evidence 2**: Set property of Result Set output by Set Retrieval is **Wrong**	**Evidence 3**: TimeInfo of CDRs present in data warehouse is **Wrong**	**Evidence 4:** TimeInfo of CDRs output of mediation module is **Ok**
Source to system	Source to system	Source to system	**X**
File transfer	File transfer	**X**	**X**
Value decoding	Value decoding	Value decoding	Value decoding
Normalization	Normalization	Normalization	Normalization
Aggregation	Aggregation	Aggregation	Aggregation
DW file transfer	DW file transfer	**X**	**X**
Data warehouse	Data warehouse	**X**	**X**
Query construction	Query construction	**X**	**X**
Set retrieval	Set retrieval	**X**	**X**
Set summation	**X**	**X**	**X**
Explanation: With the initial symptom, all components are candidates for fault localization.	**Explanation:** Since output of Set Retrieval itself is Wrong, it means fault has occurred at or before this component.	**Explanation:** All components downstream of this new observation are exonerated.	**Explanation:** All components upstream of this observation are exonerated since, till this point, the values are Ok.

Table 5: Diagnosis Sequence for Scenario 2 showing the monotonically decreasing diagnosis set size, ultimately narrowing down to the faulty component. "X" means that the respective module is no longer a (minimal) diagnosis

Evidence 1: Output subscriber count is **less** than expected	**Evidence 2**: Set property of output by Set Retrieval is **Wrong**	**Evidence 3**: Set property of result set output by data warehouse is **Ok**	**Evidence 4: TimeInfo** of CDR in data warehouse is **Ok**	**Evidence 5: AcctInfo** of CDR in data warehouse is **Ok**
Source to system	Source to system	Source to System	Source to system	**X**
File transfer	File transfer	**X**	**X**	**X**
Value decoding	Value decoding	**X**	**X**	**X**
Normalization	Normalization	**X**	**X**	**X**
Aggregation	Aggregation	Aggregation	**X**	**X**
DW file transfer	DW file transfer	**X**	**X**	**X**
Data warehouse	Data warehouse	Data warehouse	Data warehouse	**X**
Query construction	Query construction	**X**	**X**	**X**
Set retrieval	Set retrieval	Set retrieval	Set retrieval	Set retrieval
Set summation	**X**	**X**	**X**	**X**
Explanation: With the initial symptom, all components are candidates for fault localization.	**Explanation:** Since output of Set Retrieval itself is Wrong, it means fault has occurred at or before this component.	**Explanation:** All components modifying the set property of the CDRs upstream are exonerated.	**Explanation:** All components modifying the timeInfo property of the CDRs upstream are exonerated.	**Explanation:** All components modifying the acctInfo property of the CDRs upstream are exonerated, thus narrowing down to the correct fault localization.

References

[Chaudhuri 97] S. Chaudhuri, U. Dayal: An overview of data warehousing and OLAP Technology. In: ACM SIGMOD Record, 1997.

[Inmon 92] W.H. Inmon: Building the Data Warehouse. John Wiley, 1992.

[Struss 08] P. Struss: Model-based Problem Solving. In: van Harmelen, F., Lifschitz, V., and Porter, B. (eds.). Handbook of Knowledge Representation, Elsevier, 2008

Automated Critique of Sketched Designs in Engineering

Jon Wetzel and Ken Forbus
Qualitative Reasoning Group, Northwestern University
2133 Sheridan Road, Evanston, IL, 60201, USA
jw@northwestern.edu, forbus@northwestern.edu

Abstract

Designers often use a series of sketches to explain how their design goes through different states or modes to achieve its intended function. Learning how to create such explanations turns out to be a difficult problem for engineering students. An automated "crash test dummy" to let students practice explanations would be desirable. This paper describes how to carry out a core piece of the reasoning needed in such system. We show how an open-domain sketch understanding system can be used to enter many aspects of such explanations, and how qualitative mechanics can be used to check the plausibility of the intended state transitions. The system is evaluated using a corpus of sketches based on designs from an engineering school design & communications course.

1 Introduction

One of the cornerstones of engineering education is learning to design. In the early stages of design, sketches dominate. A complex mechanism can go through multiple states or have multiple modes to achieve its intended function. To communicate how their design works, designers typically use a series of sketches, plus verbal or written information (depending on circumstance) to express information not easily sketched. According to instructors, learning how to communicate with sketches can be quite difficult for students. We are working with Northwestern's Engineering Design and Communication course (EDC) to improve students' ability to communicate using sketches. The idea is to create a *Design Buddy* for students to use in practicing explanations via sketching. The input to Design Buddy will be a sketched explanation of how their design is supposed to operate. The software's job is to scrutinize the design, and see if their explanation is plausible.

The Design Buddy is an ambitious project, and currently it is far from complete. This paper focuses on a key problem in this task: Providing feedback on explanations of intended mechanical behavior of multi-state mechanisms, entered via sketching. This problem is key because (as explained below) many designs predominantly involve forces and motion. It is a good starting point because it factors out other aspects of intent which are more open-ended (e.g., using traction pads for a device normally used in a bathroom, where surfaces are often wet) and will require additional interface modalities (e.g. text or speech) to convey.

Section 2 describes how we handle sketched input and the spatial reasoning required. Section 3 describes the qualitative mechanics reasoning involved. Section 4 describes the explanation critiquing algorithm, and Section 5 describes the evaluation on student projects[1] like the one-handed fingernail clipper in Figure 1. We close by discussing other related work and future work.

2 Sketching multi-state explanations

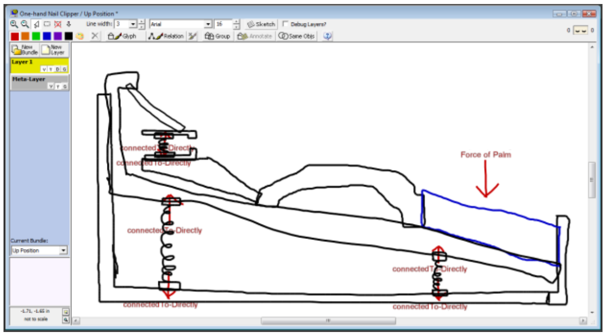


Figure 1: One-handed fingernail clipper, an EDC Project, in the up position. The hand is laid horizontally across the top, fingers pointing left, and the palm presses down to close the clipper.

We use CogSketch [Forbus *et al.*, 2008], an open-domain sketch understanding system[2], for entering and analyzing sketches. CogSketch enables users to draw *glyphs* that represent entities. A glyph is drawn by pressing a button, drawing whatever strokes constitute it, then pressing another button. This manual segmentation method is better suited for complex drawings than pen-up or time-out constraints (cf. [Cohen *et al* 1997]), because the parts of a complex

---

[1] Student projects are typically done for real customers, including patients at the Chicago Rehabilitation Institute. For instance, stroke victims often only have one working hand, which motivates several of the design tasks in the corpus.

[2] CogSketch is publicly available at http://www.silccenter.org/projects/cogsketch_index.html

design are often best drawn by multiple strokes, not always connected, and designers need to be able to take their time and think while sketching (e.g. Figure 1). What a glyph represents is indicated by labeling it with a concept from CogSketch's knowledge base (KB). This KB uses OpenCyc-derived knowledge [OpenCyC] as a starting point, so it is extremely broad (i.e., over 58,000 concepts). For example, the springs in Figure 1 are given the conceptual label **`Spring-Device`**, a concept from the KB. This is in contrast with recognition-based approaches, which require the system designer to identify in advance a small collection of entity types that can be sketched, and train recognizers for each type (cf. [Hammond & Davis, 2005]). While such systems can be useful in many circumstances, the open-ended nature of general engineering design tasks involves many more types than there are distinct visual symbols for, hence the need for another means to conceptually label them. In human to human sketching, conceptual labeling is typically accomplished via natural language. In CogSketch, a specialized interface enables users to attach KB concepts to glyphs after they are drawn. This approach means that users are never distracted by recognition errors, which tend to break their train of thought. However, it does expose them to more of the KB internals than is appropriate for a fielded system, an issue we return to in Section 7.

In addition to glyphs representing entities, CogSketch also supports *annotation glyphs* to describe an object's properties, and *relation glyphs* to describe relationships between entities. We use annotation glyphs to describe applied forces and directions of motion, using arrows. In Figure 1, for example, the force applied by the user's palm is indicated by the downward arrow on the right. Relation glyphs are used to provide a way of describing the relationships between different objects in a sketch or the different states explaining a design (see below).

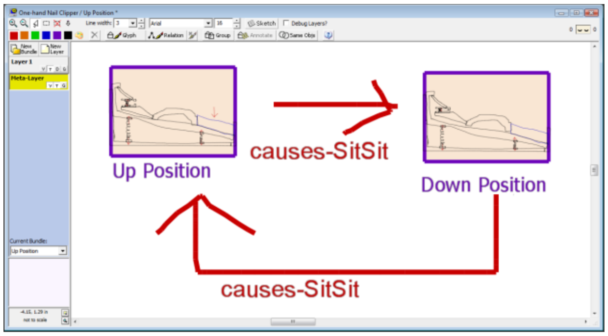


Figure 2: The metalayer provides a way to sketch multi-state explanations. Relation glyphs describe intended causal relationships between states.

CogSketch performs a variety of visual analyses on the digital ink that makes up a glyph, using techniques motivated by studies of human visual and spatial reasoning [Forbus *et al* 2008]. For example, CogSketch computes qualitative topological relationships (RCC8, [Cohn, 1996]), which we use to analyze the connectivity of parts. It also segments the ink of a glyph into lines and corners, which are used here to identify surface normals at points of contact.

In CogSketch, a sketch consists of multiple *subsketches*, each of which describes some coherent aspect of a sketch. Here subsketches are used to represent the distinct states of a design. CogSketch includes a *metalayer*, a special pane on which every subsketch of the sketch appears as an automatically-generated glyph. Multi-state explanations are entered via creating subsketches corresponding to each state, and then linking them via relationship glyphs on the metalayer. Figure 2 illustrates the explanation for the states of the one-handed fingernail clipper, the first state of which was depicted in Figure 1. The relation glyphs, each labeled with the KB relation **`causes-SitSit`** (situation causes situation), indicate that the first state will lead to the second state, and the second state will lead to a return to the first state. The second state was created by cloning the first state on the metalayer (depicted in Figure 2), then editing it by moving and resizing parts to indicate the changes therein. This can greatly simplify the sketching process, compared to pencil and paper.

3 Qualitative Mechanics

As described in [Wetzel and Forbus, 2008], we have adapted existing qualitative physics representations [Nielsen 1988][Kim 1993] for analyzing mechanisms. These representations include forces, motion, rigid objects, and the transmission of forces and movement via surface contacts. Our subsequent analysis of a corpus of student designs (see Section 5) motivated several extensions, including how forces and motion transfer across direct, rigid connections between objects, and models of springs and gears.

We use qualitative mechanics (QM) for two purposes. The first is to predict how the objects depicted in a state will behave. The second is to verify that the necessary requirements are met for each state transition to be possible. That is, given the forces that are occurring in an initial state, will the motions required to reach its proposed causal consequent actually occur?

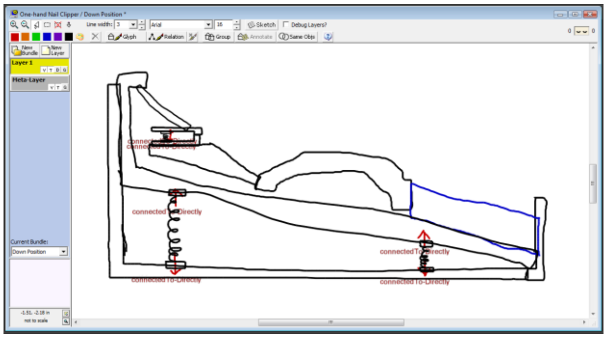


Figure 3: The "Down Position" subsketch captures the state after the clipper has been closed and the force of the palm is removed. The parts move back upward due to the compressed springs.

The connection between the entities in the sketch and QM concepts is made via conceptual labeling. For example, in Figure 1, parts which will not move relative to the sketched view are labeled with the concept **`FixedRigidObject`**. Parts which are free to move are labeled **`RigidObject`**, and the three springs are labeled as **`Spring-Device`**. The sketch also contains relation glyphs that indicate a direct connection (in the sense of glued or welded together) between objects. These relation glyphs are labeled with the relationship **`connectedTo-Directly`**. CogSketch also provides an interface for applying this relation directly to the pair of glyphs without drawing a relation glyph—we have drawn them here for illustrative purposes. An annotation glyph applied to the actuating palm rest and labeled with the concept **`forceArrow`** represents the force of the palm pressing down on the device.

As noted above, the user creates the second state (Figure 3) initially by cloning the first state on the metalayer. In the second state the palm rest is depressed, moving a latch running through the mechanism downwards that pulls the clippers closed. The springs are resized to fit the new location of the parts they are attached to, making them smaller. In order for the system to know the springs are no longer in a neutral position (currently the default) an additional conceptual label is added to the spring objects, **`CompressedSubstance`**. Finally, since the palm is no longer pressing down on the palm rest, the force annotation glyph is removed.

4. Critiquing explanations

```
CheckSketchTransitions(sketch)
For each subsketch in GetSubsketches(sketch)
 UpdateSurfaceContactKnowledge(subsketch)
For each subsketch-pair in
        GetTransitionPairs(sketch)
  For each requirement in DeduceReqs(subsketch-pair)
   For each verification in VerifyReqs(requirement)
    If verification = requirement
       then PrintSuccess(requirement, verification)
       else PrintFailure(requirement, verification)
```

Figure 4: The critique algorithm precomputes surface contact knowledge before deducing and verifying the requirements of each state transition pair (derived from the causes-SitSit relationships).

The algorithm for critiquing explanations (Figure 4) begins by using the spatial knowledge in each state to derive the set of surface contact relationships, including surface normals, between the objects in that state, using techniques from [Klenk *et al.*, 2005]. It then takes each pair of states that are linked by a causal relationship and uses an inference engine to determine what is required to transition from the antecedent state to the consequent state (**`DeduceReqs`** step, Figure 4). Currently these rules only look for motion-related differences, i.e. the appearance or lack of translation or rotation. To determine if an object has moved, the objects of type **`fixedRigidObject`** are used as reference points. For example, the glyph representing the palm rest in State 2 is lower than it was in State 1, relative to the outer frame of the device. This creates a state transition requirement that, in order for State 2 to follow from State 1, it is necessary for the palm rest to translate downwards. Similar facts are created for the other moving parts, and the same analysis is done for the transition from State 2 back to State 1.

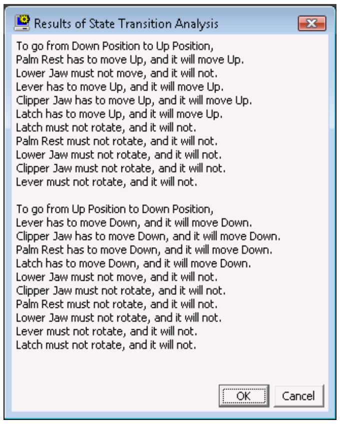
Rotations of objects between subsketches are detected in two ways. First, CogSketch automatically computes the qualitative orientation (e.g. right, up, quadrant 1, etc.) for each object in each subsketch. Looking this up is fast, but if the rotation is small the difference may not appear. If this fails, we use a cognitive model of mental rotation [Lovett *et al* 2007] to find the corresponding edges of the glyphs in each subsketch. The resulting mapping of edges is then used to calculate the angle of rotation between the glyphs. In the nail clipper example none of the parts change their orientation from state to state, so for each object the rotational requirement is that no rotation occurs.

Once the requirements for each transition have been computed, the system checks to see if they are satisfied (**`VerifyReqs`** step, Figure 4) by using qualitative mechanics to predict the next translation and rotation of the object in the antecedent state. Translation is inferred based on the constraints on the movement of the objects and the net force acting on the object. The movement constraints come from being a fixed object or being in direct contact with, or being directly connected (e.g. glue) to, another object with a constraint. The net force is found by finding all the forces acting on an object and resolving them to find the net force. The vectors used here are qualitative [Nielsen 1988], using quadrants and their edges. To help resolve ambiguities with opposing forces, the user can input a force's magnitude when creating force arrows. Both the net force and the movement constraints require the surface contact information from the sketch, which are computed at the beginning of the transition checking algorithm (Figure 4). Once they are found, if the object is free to move in a direction indicated by the net force, it will do so, otherwise it will not move. In the nail clipper sketch (Figure 2), going from State 1 (up position) to State 2 (down position), the qualitative analysis derives that the initial force will move all the free parts—from the palm rest to the upper jaw of the clipper—as drawn. For the reverse transition, the spring representation predicts that the compressed springs will provide upward forces on the other parts, causing all the parts to move upward toward their original State 1 positions. Note that the forces in State 2 did not have to be explicitly drawn as annotations by the user, as the external force in State 1 did. Instead, this force was inferred from the fact that the springs are labeled as compressed in State 2[3].

Rotation is verified in a way analogous to translation using one extra piece of knowledge: the center of rotation. Finding the center of rotation for an arbitrary object with arbitrary qualitative surface contacts and forces acting on it was beyond the current scope of this research; for now we require the user to label it with an annotation glyph. Once this is known, the torques on an object can be derived via knowing the forces on it and their relative position to the center of rotation. Similarly, rotational constraints can be derived based on surface contacts. If the object is free to rotate in a direction indicated by the net torque it will do so. Eight examples in Section 5 include instances of rotation.

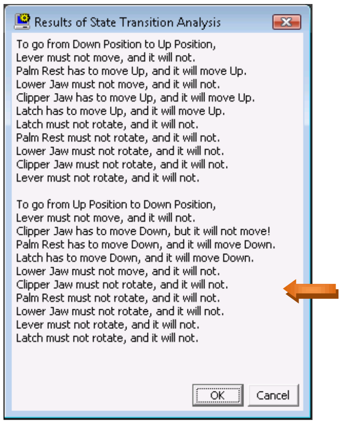
---

[3] Automatically deducing that the shorter spring in State 2 implies that it is compressed, given that the spring in State 1 is neutral, is an example of reasoning about depiction that we intend to incorporate in later versions (e.g. [Lockwood *et al* 2008]).



```
To go from Down Position to Up Position,
Palm Rest has to move Up, and it will move Up.
Lower Jaw must not move, and it will not.
Lever has to move Up, and it will move Up.
Clipper Jaw has to move Up, and it will move Up.
Latch has to move Up, and it will move Up.
Latch must not rotate, and it will not.
Palm Rest must not rotate, and it will not.
Lower Jaw must not rotate, and it will not.
Clipper Jaw must not rotate, and it will not.
Lever must not rotate, and it will not.

To go from Up Position to Down Position,
Lever has to move Down, and it will move Down.
Clipper Jaw has to move Down, and it will move Down.
Palm Rest has to move Down, and it will move Down.
Latch has to move Down, and it will move Down.
Lower Jaw must not move, and it will not.
Clipper Jaw must not rotate, and it will not.
Palm Rest must not rotate, and it will not.
Lower Jaw must not rotate, and it will not.
Lever must not rotate, and it will not.
Latch must not rotate, and it will not.
```



```
To go from Down Position to Up Position,
Lever must not move, and it will not.
Palm Rest has to move Up, and it will move Up.
Lower Jaw must not move, and it will not.
Clipper Jaw has to move Up, and it will move Up.
Latch has to move Up, and it will move Up.
Latch must not rotate, and it will not.
Palm Rest must not rotate, and it will not.
Lower Jaw must not rotate, and it will not.
Clipper Jaw must not rotate, and it will not.
Lever must not rotate, and it will not.

To go from Up Position to Down Position,
Lever must not move, and it will not.
Clipper Jaw has to move Down, but it will not move!
Palm Rest has to move Down, and it will move Down.
Latch has to move Down, and it will move Down.
Lower Jaw must not move, and it will not.
Clipper Jaw must not rotate, and it will not.
Palm Rest must not rotate, and it will not.
Lower Jaw must not rotate, and it will not.
Lever must not rotate, and it will not.
Latch must not rotate, and it will not.
```

Figure 5a(left): The explanation checks out.
Figure 5b(right): With the lever moved off to the side, the sketch violates an expectation (denoted with "!").

Finally, the system compares the results of verification with the requirements and outputs a list indicating whether they were successfully met or not. As Figure 5a illustrates, the requirements are translated into English using a simple set of templates. Figure 5b shows the output of the same system if the lever on top of the nail clipper is disconnected from the rest of the mechanism. Without it, there is nothing to exert force on the upper jaw and it will no longer move down. Violated requirements are denoted with an "!". These summaries are intended for development purposes; the NL generation for student feedback will focus on places where the system finds problems with their explanations.

5. Evaluation

The system was evaluated on examples derived from EDC projects, such as the running example of the one-handed fingernail clipper. A corpus of 39 projects was collected. 19 of these were deemed not mechanically interesting, lacking moving parts or being mainly electrical (e.g. circuits) or flow-centered (e.g. pumps). Of the 20 remaining examples, sixteen were suitable for the system. Four of them were beyond the spatial reasoning capabilities of CogSketch (mostly three-dimensional). Six of the remaining sixteen were redundant or very similar to other designs, so we performed the evaluation using only the ten designs (including the nail clipper in the earlier sections) that describe the space of problems which the system could handle.

Since the original student designs were on posters or pencil and paper, we sketched them using CogSketch ourselves. The remainder of this section highlights some of the strengths and weaknesses of the system as shown by its performance on the ten evaluation examples.

5.1 Example 1: Book Holder

Not every system in the EDC projects was intended to be a chain or sequence of states. Many projects are made to contain or stabilize something. Figure 6 shows a device designed to hold open a book. To convey this intention, we made this sketch of the desired state, cloned it and then asserted that the first state causes its copy. The system then infers that we mean for all parts in the sketch to stay stationary. The exposed page of the book has a rotational force arrow on it denoting the natural tendency for that page to flip upwards, but the clamp holds the page firmly in place. The system sees this constraint and agrees with our assertion that nothing will move.

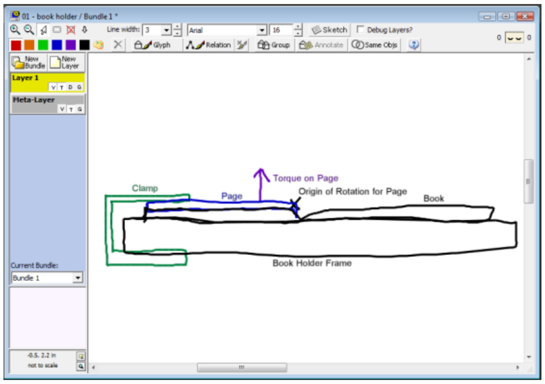


Figure 6: A book holder, viewed from the book's edge. The open page experiences an upward force, but is clamped from the left.

To test the alternate case we made another state, this time with the clamp disconnected. In this case the system warns that while the page stays stationary in our sketch, it will in reality rotate clockwise.

5.2 Example 2: Baja Mini

The Baja Mini in Figure 7 is representative of several projects that involve vehicles like go carts or solar cars. Torque on the wheels will cause it to move to the left. Without friction, the system predicts (correctly) that it will not move. When force arrows were added to represent friction, the system inferred that the whole cart could move, with the wheels pushing the frame along with them via surface contact.

Figure 7: An all-terrain vehicle in motion. Assumed torque on wheels and ground friction are required to infer motion.

5.3 Example 3: Finger Trainer

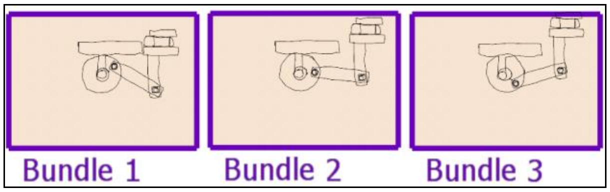


Figure 8: A device for re-training precision finger movements. The palm rests on the top with the finger stuck through a "key". The up and down movement simulates typing.

The Finger trainer (Figure 8) was difficult for the system for a couple of reasons. First, there were a number of places where parts overlapped but were not necessarily in direct contact with each other. We could draw the attachment as going around the end of the finger on the right, but in the actual project, the finger socket is a glove with the end cut off, making that drawing inaccurate. Similarly, bolts connect the different beams and the wheel but the beams and wheel have no contact with each other.

To solve this problem we will need to formally describe a three-dimensional attribute such as "inside" or "behind" to help describe the relationship between the finger and the finger slot. Also, describing the motion of the bar between the wheel and the vertical bar is difficult for the current QM because it is constrained by two different axes of rotation. Its motion will turn out to be a translation plus a rotation about some point on neither axis—we could find this point manually, but it would be laborious to require the user to do so. We plan to use Kim's [1993] work on linkages as a basis for representing these kinds of connections between objects in the future.

5.4 Example 4: One-handed egg cracker

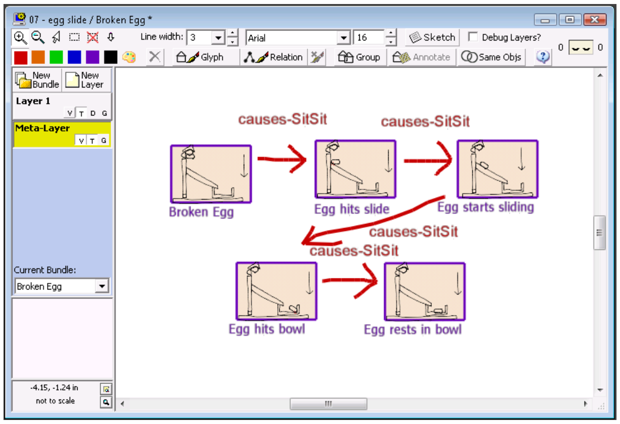


Figure 9: Device for cracking an egg with one hand. The egg shell remains in the hand (upper left) and the egg yolk slides to a bowl at the bottom of the structure.

Sketching the one-handed egg cracker (Figure 9) involved showing how the egg yolk moves down a slide and lands in a bowl at the base. While representing the process of cracking an egg is beyond the level of our QM currently, the system successfully understood the motion of the egg yolk falling, making contact with the slide, turning and sliding down the slide, making contact with the bowl, rotating and coming to rest. This example demonstrates that our system can handle a variety of translations and rotations. However, it illustrates a current weakness: it cannot reason about states which have not been drawn. There are more states here than a human partner would have required to understand the explanation, which places an extra burden on the student. We plan to investigate automatically generating new subsketches in the sketch via constrained qualitative simulation to "fill in" the implied intermediate states, to ensure that they can indeed be consistently created.

5.5 Example 5: Recliner with Shock-Absorber

To handle a non-rigid body (like the human body), the system does not try to infer what will happen to the body itself but does pays attention to any forces attributed as coming from that body. In example 5 (Figure 10) the system reasons about the behavior of the seat back, correctly predicting that it will rotate clockwise and compress the shock absorber, but it has nothing to say about the human sitting in the chair, for whom there is no QM representation yet.

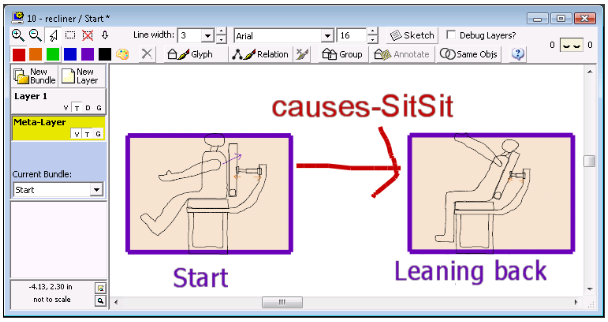


Figure 10: A reclining chair for people suffering from involuntary muscle spasms.

5.6 Example 6: Paint Roller

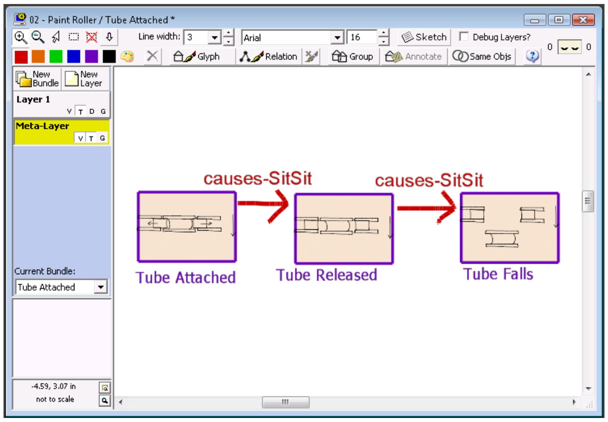


Figure 11: A quick-release paint roller. When the clamps are pulled outwards the tube falls under gravity.

Figure 11 shows a paint roller with a quick-release mechanism for changing the roll. It is drawn from a head-on perspective but might be better understood by a human if it was drawn from top down. Currently the system is limited by a lack of understanding of the conventions for illustrating depth in a drawing. If this were a top down sketch, the tube would get smaller in the third state. Work continues on interpreting these kinds of conventions.

5.7 Example 7: Ab Machine

Figure 12 shows another example of a non-rigid body at work in a sketch. This example shows a case in which our primary, qualitative method of detecting rotation is insufficient to detect a required change. The middle panel starts at about 135° and rotates counter-clockwise a little but not enough to be near 180°, the next distinct qualitative direction (i.e. left rather than quadrant 2). As mentioned in Section 4, we use a model of mental rotation to confirm that this piece has actually rotated.

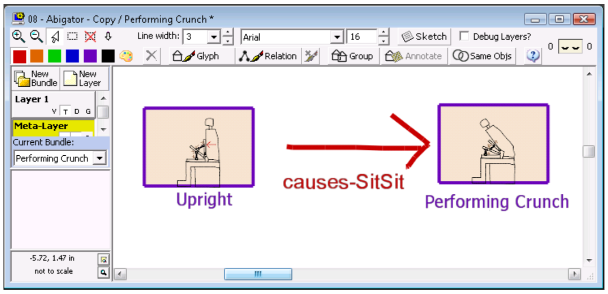


Figure 12: The device is for helping people in a wheel chair exercise their core muscles. It contains three fixed axis panels separated by springs.

5.8 Example 8: Dual-action Switch

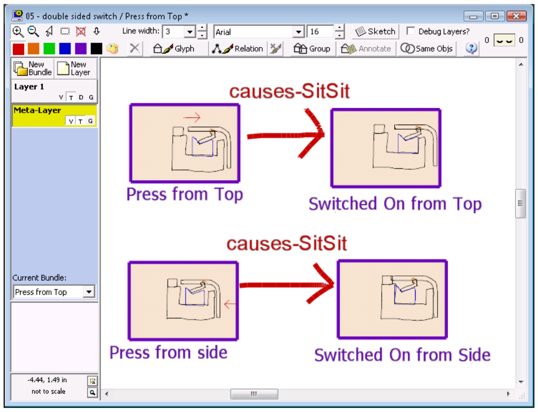


Figure 13: An electrical switch that can be activated by pressing the side or top.

The system successfully understood the mechanical aspects of this electrical switch in Figure 13. However, it is missing the greater context. There is no representation for the electrical aspects, e.g. how surface contact can transfer electric current, and the difference between a conducting surface and a non-conducting surface. The complete design sketch for this device would show these details, and while a human can infer them from looking at this sketch, it is lost on the system at this point. Many designs involve multiple domains, but we believe our state transition requirement representation is general enough to extend to those, given appropriate extensions to our knowledge base and qualitative reasoning capabilities. As we continue to work with EDC, these representations will be added and eventually be used by the `DeduceReqs` step of our algorithm (Figure 4).

5.9 Example 9: Wheelchair Softball

The wheelchair example (Figure 14) is unique in that it is drawn top-down. Students in EDC are often expected to draw their designs from side, top, and oblique perspectives. CogSketch is currently able to handle side view and top view sketches. In this case, the surface contact and force inferences worked without any extra additions to the QM knowledge. However, as discussed below, oblique perspectives are the subject of future work.

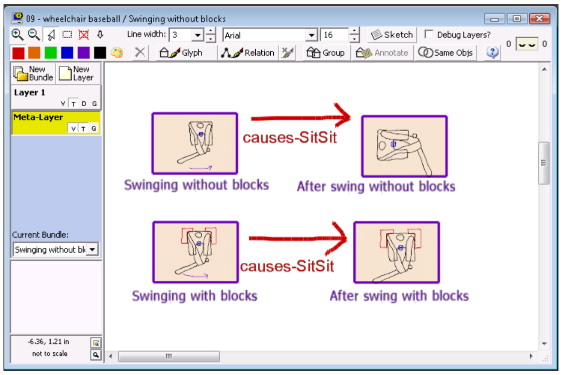


Figure 14: Rigid blocks prevent a wheelchair from rotating under the influence of swinging a baseball bat.

5.10 Example 10: Retractable Stacking Mechanism

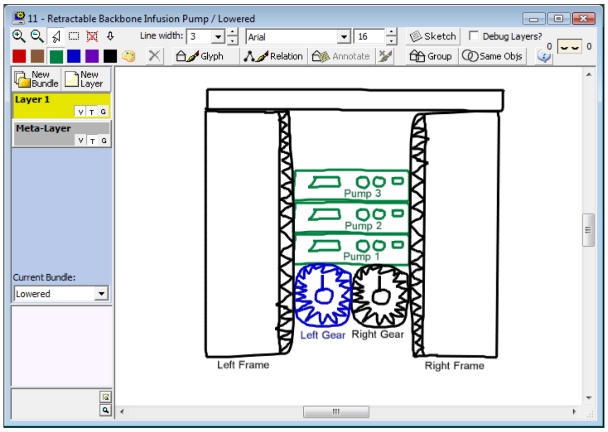


Figure 15: This retractable stacking mechanism allows pieces of medical equipment to be swapped in and out easily.

The retractable stacking mechanism in Figure 15 can be mounted on a cart for easily transporting interchangeable medical devices (in this case, backbone infusion pumps). It also demonstrates our representation of gears and toothed surfaces (drawn with zig-zag lines). When two toothed surfaces are in contact their objects are considered to be enmeshed, enabling certain behaviors. For example, when a counterclockwise torque is applied to the left gear, it rolls upwards along the fixed frame to the left. The right gear, also enmeshed with the left gear, rotates clockwise and likewise moves upward along the right frame. Together they lift up the stack of equipment until it is snug against the top of the case, preventing them from falling out.

Currently we must draw straight edges around the toothed surfaces to improve the performance of our surface-contact detection, which would otherwise have to deal with many small edges. One approach to simplifying this would be to add a perceptual model of textured edges to CogSketch, allowing it directly produce a simpler edge representation.

6. Related Work

SketchIt [Stahovich *et al* 1998] used multiple sketches linked by state transition diagrams to generate new concrete designs of fixed-axis devices, mediated by qualitative representations. Our use of sketches linked by state transitions to describe multi-state behavior is similar, but we also use them for describing alternate modes, and given the nature of our task, cannot assume that they are correct. Our qualitative mechanics reasoning is not limited to fixed-axis devices, but stays entirely at the level of sketched representations. In SketchIt users were required to identify important surface contacts, which is not unreasonable for its intended use by expert designers. Since we are dealing with novices, we must identify them automatically when possible.

Most work on sketch understanding has focused on glyph recognition, e.g., [Alvarado and Davis, 2004; Hammond & Davis, 2005; Kurtoglu and Stahovich, 2002]. Human to human sketching demonstrably does not require recognition, as anyone looking at sketches made by others without knowing the context can attest. However, recognition can act as an important catalyst, making the interaction more natural, so we would like to incorporate such techniques if further analysis indicates they could help. Recognition-based systems typically act as an interface to some traditional software system (e.g., simulation setup in [Cohen *et al* 1997] or a physics simulator [Alvarado and Davis, 2001]). Quantitative mechanical simulation would not be wise for our task, since we are focused on conceptual design, before enough information is known to support accurate numerical simulation, and inaccurate simulation would be misleading. Our use of qualitative reasoning to operate at the same conceptual level that the student is working at enables us to provide natural feedback on their explanations.

7. Future Work

The critique system described here will provide the core reasoning capability for the Design Buddy. We briefly summarize five areas where additional research is needed: extended spatial reasoning, extended qualitative mechanics, adding factors in critiquing, intent understanding, and controlled natural language processing.

Extended visual and spatial reasoning: The current techniques for computing surface contacts and axes of rotation are incomplete. Consequently, we currently use annotations to identify axes of rotation. Automating this requires improved qualitative representations of curves. Research on 3D reasoning in CogSketch is underway [Lovett *et al* 2008], which will allow us to handle perspective sketches.

Extended qualitative mechanics. The system currently only handles rigid objects plus springs and gears. We plan to use techniques from [Kim 1993] to incorporate liquids and gasses, but new theories will be needed to handle pliable solids, strings, and elastic materials. Incorporation of defaults and using broader world knowledge in model formulation is a key step. Friction is a prime example. By default one should consider friction, but choices of specific materials can be made to reduce or enhance friction, depending on the designer's intent. Adding more knowledge about materials to the KB, and appropriate default reasoning to challenge a student's explanation, will be useful steps. Our representation of the interaction of toothed surfaces could also be generalized to explain how friction causes rolling behavior.

Adding critique factors: As noted above, the state transition analysis used in generating critiques only looks at motion. There are many other relevant differences that could be included, such as changes in connection or the introduction and removal of forces. Resource consumption across paths of states can be worth monitoring for some designs. These will be added incrementally, driven by what is needed by student design projects.

Intent understanding: The current explanation input system only allows simple descriptions of intent, i.e., whether or not something moves. For the near term, we intend to continue to focus on behavioral constraints, since those can be expressed in qualitative mechanics. For the longer term, incorporating real-world motivations requires broadening of the knowledge base (e.g., that bathrooms often have wet surfaces) and more natural language input. Even then, breadth can be somewhat controlled, since those factors are often best critiqued by the student's teammates, customers for the design, and instructors.

Controlled natural language processing: While CogSketch has the ability to accept unprocessed natural language strings as labels for concepts, it currently does not provide any facility for suggesting interpretations of them in the underlying knowledge base. For conceptual labeling, we plan on using simple phrase-level techniques for inferring appropriate concepts (e.g., "spring" is the canonical pretty name for `Spring-Device` in the KB). For intent input, we plan on using a menu-based system for constructing phrases with drag & drop of sketch items for deictic reference [Forbus *et al* 2003].

Importantly, we do not have to achieve all of the above goals to start experiments with students. As our evaluation indicates, our system can already handle 25% of the typical class designs, and our collaborating instructors are willing to work with us to focus on pedagogically interesting designs within that space. Consequently, we are next focusing on automating center of rotation detection and natural language concept labeling, which should be enough for initial "pull-out" studies with EDC students in 2009. Our hope is that the work described here is a major step towards our goal, that by a combination of techniques from AI and cognitive science, engineering students will, in the long run, be able to receive help from software anytime, anyplace, in a reasonably natural way.

Acknowledgments

We thank our EDC collaborators, Ed Colgate, Bruce Ankenman, Ann McKenna, John Anderson, and Stacy Benjamin, for their valuable conversations and access to students and their portfolios. Emmet Tomai and Matthew Klenk provided valuable help with surface contact detection. This research was supported by the Spatial Intelligence and Learning Center, NSF grant SBE0541957.

References

[Alvarado and Davis, 2001] Alvarado, C., and Davis, R. 2001. Resolving ambiguities to create a natural computer-based sketching environment. *Proceedings of IJCAI01.*

[Alvarado and Davis, 2004] Alvarado, C., Davis R. 2004. Multi-domain sketch understanding. Massachusetts Institute of Technology, Cambridge, MA.

[Cohen *et al* 1997] Cohen, P., Johnston, M., McGee, D., Oviatt, S., Pittman, J., Smith, I. Chen, L., and Clow, J. 1997. QuickSet: Multimodal interaction for simulation setup and control. *Proceedings of the 5th Conference on Applied Natural Language.*

[Cohn 1996] Cohn, A. 1996. Calculi for qualitative spatial reasoning. In *Artificial Intelligence and Symbolic Mathematical Computation*, LNCS 1138, Eds: J. Calmet, J.A. Campbell, & J. Pfalzgraf, Springer Verlag, 124-143.

[Forbus *et al* 2003] Forbus, K., Usher, J., and Chapman, V. 2003. Sketching for military courses of action diagrams. *Proceedings of IUI'03*, January, Miami, FL.

[Forbus *et al.*, 2008] Forbus, K., Lovett, A., Lockwood, K., Wetzel, J., Matuk, C., Jee, B., and Usher, J. (2008). CogSketch. *Proceedings of AAAI 2008*.

[Hammond & Davis, 2005] Hammond, T. & Davis, R. 2005 LADDER, a sketching language for user interface developers. *Computers & Graphics* 29 (518-532).

[Kim 1993] Kim, H. (1993). Qualitative reasoning about fluids and mechanics. Ph.D. dissertation and ILS Technical Report, Northwestern University. Evanston, IL.

[Klenk *et al.*, 2005] Klenk, M., Forbus, K., Tomai, E., Kim,H., and Kyckelhahn, B. (2005). Solving Everyday Physical Reasoning Problems by Analogy using Sketches. *Proceedings of 20th National Conference on Artificial Intelligence (AAAI-05)*, Pittsburgh, PA.

[Kurtoglu and Stahovich, 2002] Kurtoglu T., Stahovich T.F., 2002. Interpreting Schematic Sketches Using Physical Reasoning, In *AAAI Spring*

[Lockwood *et al* 2008] Lockwood, K., Lovett, A., Forbus, K., Dehghani, M., and Usher, J. (2008). A Theory of Depiction for Sketches of Physical Systems. *Proceedings of QR 2008*.

[Lovett *et al* 2007] Lovett, A., Forbus, K., and Usher, J. 2007. Analogy with qualitative spatial representations can simulate solving Ravens' Progressive Matrices. *Proceedings of CogSci07*. Nashville, TN.

[Lovett *et al* 2008] Lovett, A., Dehghani, M., and Forbus, K. 2008. Building and comparing qualitative descriptions of three-dimensional design sketches. *Proceedings of QR 2008*. Boulder, CO.

[Nielsen 1988] Nielsen, P.E. (1988). A qualitative approach to rigid body mechanics. (Tech. Rep. No. UIUCDCS-R-88-1469; UILU-ENG-88-1775). Urbana, Illinois: University of Illinois at Urbana-Champaign, Department of Computer Science.

[OpenCyC] CyCorp, available at http://www.opencyc.org.

[Stahovich *et al.*, 1998]Stahovich T.F., Davis R., Shrobe H. 1998. Generating multiple new designs from a sketch. In *Artificial Intelligence* 104 (1998) 211–264.

[Wetzel and Forbus, 2008] Wetzel, J., Forbus, K. (2008). Integrating Open-Domain Sketch Understanding with Qualitative Two-Dimensional Rigid-Body Mechanics. *Proceedings of the 22nd International Workshop on Qualitative Reasoning*. Boulder, CO.

Evaluating the potential of Qualitative Reasoning to capture and communicate knowledge on sustainable catchment management

Andreas Zitek, Sabine Preis, Michaela Poppe, Susanne Muhar

Institute of Hydrobiology and Aquatic Ecosystem Management, University of Applied Life Sciences, Vienna;
Max-Emanuelstraße 17, 1180 Vienna
andreas.zitek@boku.ac.at, sabine.preis@boku.ac.at, michaela.poppe@boku.ac.at, susanne.muhar@boku.ac.at

Abstract

This paper presents the potential use of Qualitative Reasoning (QR) to capture and communicate knowledge on sustainable catchment management. Based on a case study, qualitative models dealing with issues of a sustainable development of riverine landscapes were developed and implemented using the Garp3 software following a general modeling framework. The evaluation of the models and the QR approach by students and experts revealed the high potential of QR models to capture and communicate complex knowledge in an understandable and interesting manner, mainly due to the ability of the presented approach to capture qualitative system dynamics and integrate 'hard' and 'soft' facts in a structured way. In the future a library of expert models might serve as an important source of information for both, education and management.

The issue of worldwide impaired river catchments

World wide river systems with their related catchments have been substantially altered due to the pressures of human populations with severe consequences for the ecological integrity and health of riverine landscapes (Dynesius and Nilsson 1994; Boon et al. 2000; Jungwirth et al. 2002). Furthermore the past lack of considering environmental variability and potential catastrophic events in an adequate manner, e.g. catastrophic flood events, increasingly causes avoidable damages to humans and human infrastructures globally (Singh 1996). Especially participatory approaches to natural resource use planning and management sustaining adequate communication and the integration of scientific knowledge with stakeholder needs are needed to achieve a sustainable development. Communication can be therefore seen as a central process to achieve integrated environmental management. To establish modeling approaches in the catchment management processes, the education of a new generation of students, managers, planners, scientists and politicians is needed being capable of dealing with this complex issue. Modeling approaches dealing with system dynamics (quantitatively and qualitatively) offered to interested students, scientists, managers, planners and politicians could significantly contribute to the peoples capability to deal with this complexity (Grant 1998). After Sterman (1994) effective methods for learning in and about complex dynamic systems must include:

(1) Tools to elicit participant knowledge, articulate and reframe perceptions, and create maps of the feedback structure of a problem from those perceptions.
(2) Simulation tools to assess the dynamics of those maps and test new policies.
(3) Methods to improve scientific reasoning skills, strengthen group process and overcome defensive routines for individuals and teams.

The use of QR in aquatic ecoscience and management

Besides traditional numerical approaches for mediated and integrated modeling (Van den Belt 2004), more recently 'Qualitative Reasoning' has become a new frontier for structuring and integrating qualitative knowledge (Bredeweg et al. 2007a,b) with increasing use in aquatic ecoscience and integrated management (Salles et al. 2006). For example QR models have been successfully used to capture the effects of anthropogenic activities on benthic macroinvertebrate communities in watersheds (Tullos and Neumann 2006), to describe general sustainability issues in river catchments (Salles et al. 2007) and to qualitatively representing the cause effects relationships related to the indicators of environmental sustainability of the millennium development goals (Salles 2005). Furthermore the application of QR modelling in social learning environments has been assessed (Bredeweg and Salles 2002) and it has been realized, that especially in complex systems integrating a variety of disciplines and viewpoints, the use of QR models and simulations as decision-support tools has significant potential (Lee 2000; Tullos and Neumann 2006). However, as the Garp3 software tool (http://www.garp3.org) for allowing a broader application of this modeling approach has become available only recently (Bredeweg et al. 2007a), the acceptance (Yearley

1999) and the future potential of the modeling approach and the models developed need to be assessed, as this has been done also for other approaches (Stavredes 2001; Van den Belt 2004) and more recently also for QR models on water quality (Araújo et al. 2008).

The river Kamp case study

Catastrophic floods and inundations in August 2002, a nearly 2000-annual event, set new conditions for life and economy in the Kamp-valley, Austria, facing flood control management, landscape architecture and land use planning with essential and future challenges. Consequently, the high water event finally represented a chance to develop the riverine landscape together with the local population as well as with the concerned scientific disciplines considering social, economic and ecological claims, especially with regard to the EU-WFD. On this basis an overall integrated concept towards the sustainable development of the River Kamp landscape has been developed at the University of Natural Resources and Applied Life Sciences, Vienna (Preis et al. 2006). Besides the consideration of the spatial scale (from catchment level up to planning onto municipalities) the interdisciplinary work of the different disciplines biology/nature conservation, landscape planning, water resources management, regional planning, agriculture and forestry and hydropower production was considered. Moreover, planning was conducted in participation with authorities, stakeholders and the local population to achieve sustainability. The integration of the population into the planning activities exceeded pure information policy with the possibility for the local population to actively participate in developing the future scenarios for their valley. The experiences and knowledge gained within the project provided the essential basis for the development of the models that were primarily developed as learning material for students and to inform managers on the system structure as a basis for decision making. Based on the data and experiences from the river Kamp case study, two models describing the basic issues for a sustainable development and management of the riverine landscape were developed. Besides a model representing the essential of entities and processes involved in the implementation and development of a sustainable management of the riverine landscape ('model A', Fig. 1), a second model describing the effect of hydropower production (water storage and release and water abstraction) on sensitive fish populations ('model B', Fig. 2). Following a general modeling framework (Bredeweg et al. 2007b) models were developed and implemented using the Garp3 software (Bredeweg et al. 2007a). After capturing the general system structure of the Kamp valley, setting the system boundaries for the modeling approach the causal models were set up in the modeling workbench of Garp3 in an interactive and collective modeling effort.

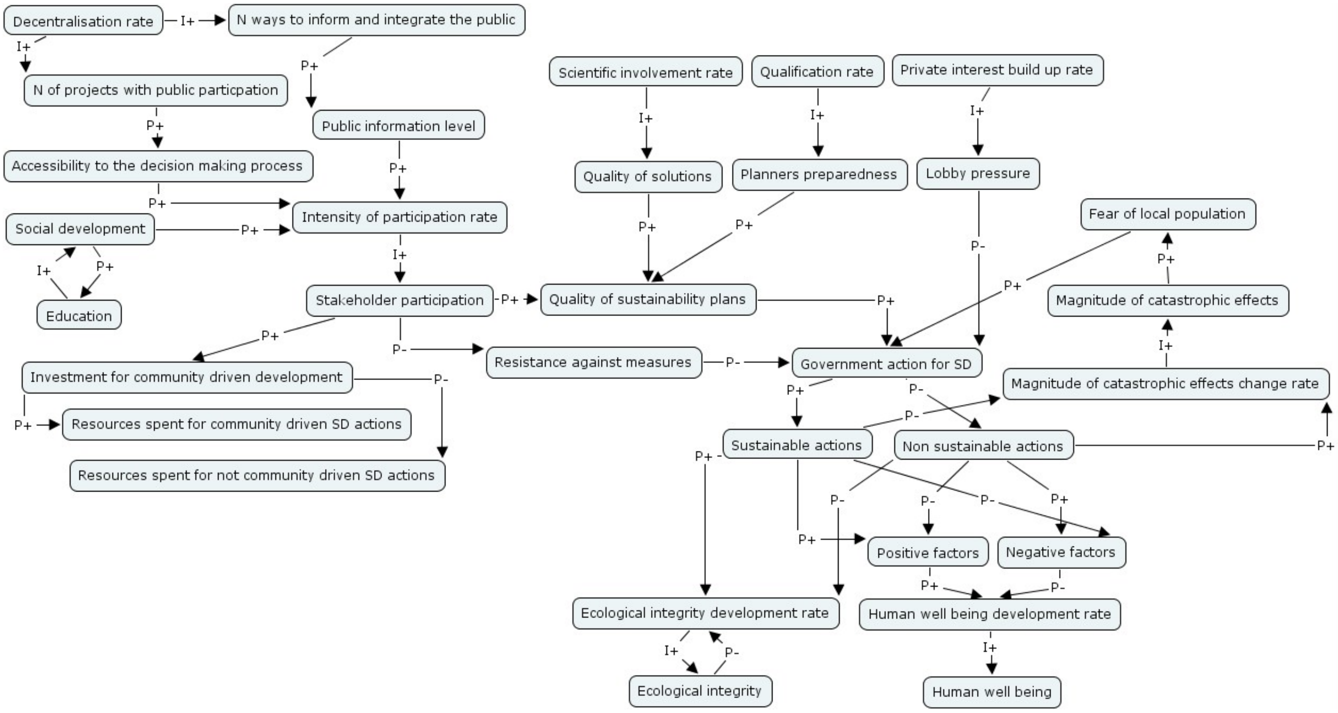


Figure 1:Causal model representing the process of the development and implementation of sustainable catchment management plans in the Kamp valley ('model A').

Finally the model with different scenarios was implemented in a compositional modeling approach based on semi-independent model fragments describing various aspects of objects and processes. Based on the full causal model, several smaller sub-models that could be linked via their different simulation outcomes were implemented. Besides a general description of both models, only the one scenario of 'model A' will be presented here in more detail, to show the basic principle of model building and simulation with the Garp3 software.

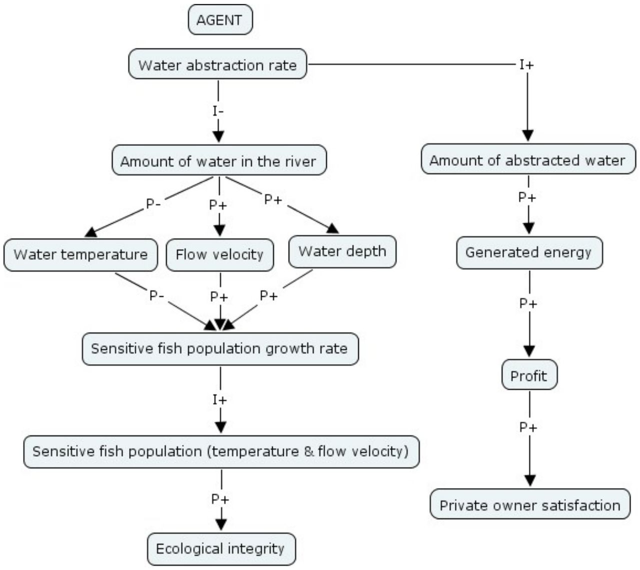


Figure 2: Causal model representing the effect of water abstraction on fish and stakeholder satisfaction ('model B').

Model A

The entities of 'model A' are divided into 5 groups 'Biological entity', 'Culture', 'Development plan', 'Environment' and 'Set of entities'. The main entities involved are 'Planners', 'Politicians' and 'Stakeholders' as biological entities (here we tried to capture the idea of the hierarchical structure of biological systems), the 'Community', which lives in the valley (can be seen as a set of entities – e.g. all people living there together with stakeholders), 'Education', 'Government' and 'Science' as expression of the culture of a country, the 'Development plan' as a basis for the implementation of sustainability issues and the 'River basin' (the 'Kamp valley') as the relevant environment. The entities are related by 'configurations' defining the basic system structure and describing mainly the direction and type of influences. Out of seven sub-models that were developed to simulate the full causal model presented in Fig. 1 (Zitek et al. 2006), only the sub-model 1 'Community fear influences government action for sustainable development (SD)' will be presented here. The 'sub-model 'Community fear affects government action for sustainable development (SD)' consists only of one model fragment that captures the basic processes, triggering the government to become active in the Kamp valley reducing 'Non-sustainable actions' and increasing 'Sustainable actions' (Fig. 3).

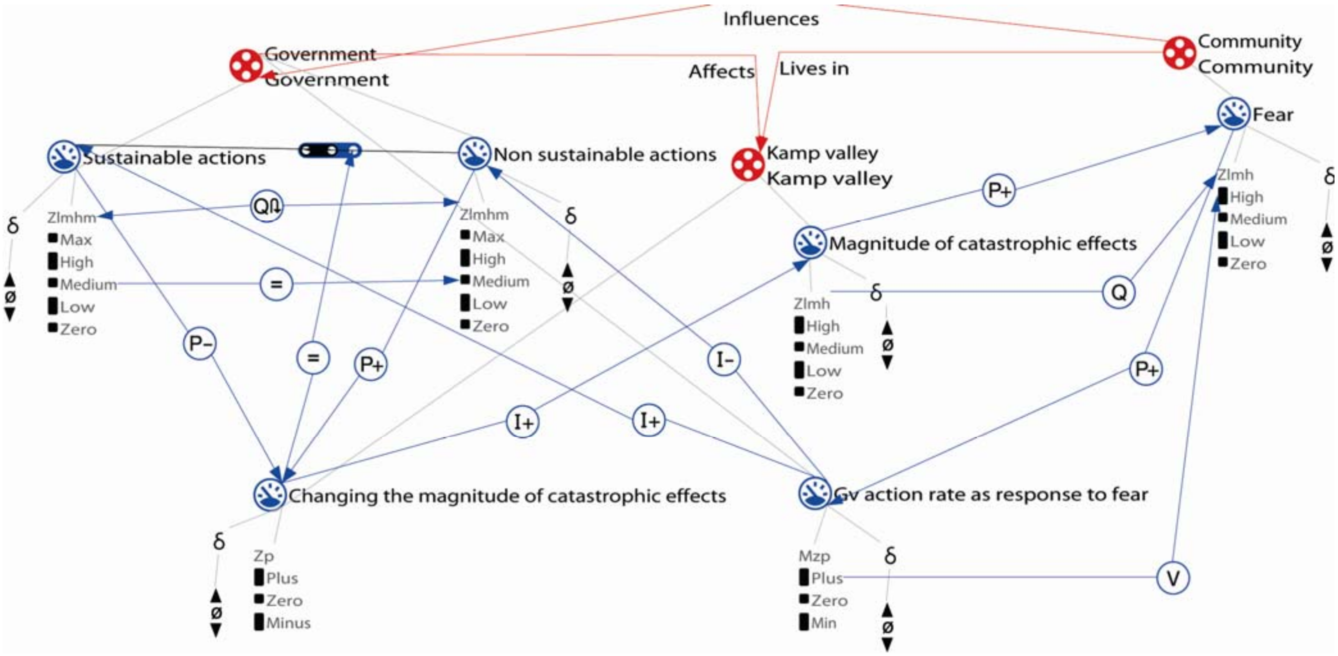


Figure 3: Model fragment 'Community fear affects government action for sustainable development (SD)' representing the whole sub-model 1 of 'model A'.

This sub-model shows how the ‘Magnitude of catastrophic effects’ is influenced by ‘Non-sustainable actions’ in the ‘Kamp valley’. When the ‘Magnitude of catastrophic effects’ is <High>, the ‘Fear’of the community from future catastrophic events is also <High>; this influences the government to force ‘Sustainable actions’ and a decrease ‘Non-sustainable actions’. Fig. 4 shows the behavior graph obtained in the simulation of sub-model 1 starting with low magnitude of catastrophic effects and low fear of the population, but a maximum of non-sustainable actions (see also the value history in Fig. 5). The model tries to capture the idea, that non-sustainable actions cause an increase of potential catastrophic effects, which then frightens the local population which lives in continuous fear from future catastrophic events creating pressure on the government; usually after a certain time people forget catastrophic events, which decreases the fear, and increasing the probability of new unsustainable actions to be implemented starting the reaction circle again. This leads to a circular behavior of the simulation.

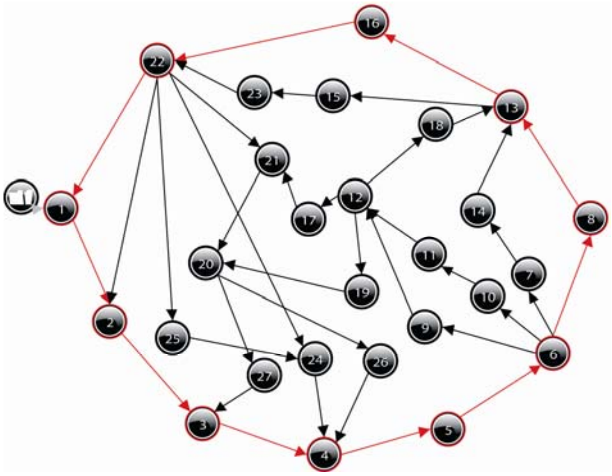


Figure 4: Behaviour graph obtained in a simulation of the sub-model 1 ‘Community fear affects government action for sustainable development (SD)’ of ‘model A’.

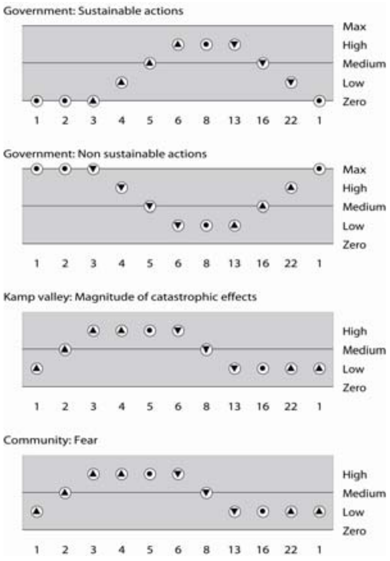


Figure 5: Value history diagram of relevant quantities in one selected behaviour path [1→22→1] of the simulation of the sub-model 1 of ‘model A’; rates are not shown.

Model B

Model B, ‘Hydropower production and sensitive fish species’, explores important problems related to hydropower use in the Kamp valley and its effect on fish (see the causal model in Fig. 3). Additionally the aspect of energy production, consumed energy, and energy sold is modeled together with stakeholder satisfaction to represent the causal principle behind the tendency of the owners of hydropower plants to maximize the amount of abstracted water. There are mainly two ways of influencing a river by hydropower use: (1) water abstraction and the creation of a residual or minimum flow stretch with the related effects to the physical environment (loss of water, loss of flow velocity, reduction of depth and increase of water temperature), and (2) the storage of water in a reservoir and a constant or peaking release of water from hypolimnetic parts of the reservoir leading to decreased temperatures below the reservoir. The decreased temperatures generally favor cold water species and repress the reproduction of warmwater species. If the water is on the one hand released at a constant rate this destroys mainly the natural flow regime of a river, if released in a peaking mode (‘hydropeaking’) it affects fish mainly due to the frequent changes of habitat conditions. Therefore model B focuses on the exploration of the two ways of hydropower use and its effects on fish and representations are developed that describe the effects of a reduced amount of water in the river (reduced flow velocity and increased temperature) on the fish fauna. Different effects the changed physical environment on different types of fish species (favoring fish with low requirements to flow velocity, the so called indifferent species, or suppressing species with high flow velocity needs, the so called rheophilous species; favoring fish due to temperature increase or suppressing them) are captured in model fragments and assumptions. This allows for a comprehensive representation of the effects of the different modes of hydropower production on different guilds of the river type specific fish community of the river Kamp. The entities are defined according the main perspectives we wanted to represent in ‘model B’: ‘Energy source’ (‘Hydropower plant’), ‘Fish’ (‘Flow velocity sensitive fish’, ‘Temperature sensitive fish’) representing the river type specific fish fauna, ‘Stakeholder (‘Private owner’) which run hydropower plants and try to maximize their economical benefit, ‘Water and water body’ (‘Reservoir’, ‘River’) as a basis for aquatic live and energy production.

Model evaluation

The evaluation of models is an important step in the model building process (Rykiel 1996). Validation proves if the scientific and conceptual contents of the model are acceptable for its intended use, verification proves that the model is correctly implemented by a demonstration of its use. Proving the acceptance of the QR approach and the software mainly evaluates the potential of the model and the modeling approach for broader use. The qualitative simulation models related to the sustainable development of the Kamp valley were generally intended to be used by stakeholders, decision makers and students to learn about the complex interactions between human use and natural resources in river catchments. To evaluate the models a two steps approach was chosen. A general evaluation was mainly focusing on the 'acceptance of the chosen approach and model' by students and scientists of different domains and an expert evaluation was more focusing on "validation and verification" of the models. The general evaluation was based on a power point presentation and a collective exploration of parts of the model using Garp3 on personal Laptops. Six students and five experts of different aquatic resource domains participated in the event, which lasted for about 2 hours. After the presentation and collective and interactive inspection of important scenarios and model fragments the participants were asked to fill in pre-prepared questionnaires. At the beginning of the evaluation process, the attendees were asked, whether they are an expert in a specific scientific field or a student. Next the participants were asked to rate a statement given with the following options: 'I fully disagree', 'I largely disagree', 'I somewhat disagree/agree', 'I largely agree', 'I fully agree'. They also were asked for additional statements. Furthermore separate expert evaluations were run with one domain expert per model as face to face discussions based on the printed causal maps and a conjoint exploration of important model fragments and simulations using Garp3 on one Laptop.

The following statements and questions were used for the general evaluation process:

1) QR models present complex knowledge in an understandable manner.
2) The QR approach allows for a clear representation of real world phenomena like a sustainable development of the riverine landscape "Kamp".
3) QR and Garp3 can be seen as a valuable learning tool for real world causal relationships related to a sustainable development of riverine landscapes.
4) The presented QR model might significantly contribute to the understanding of students and stakeholders which entities and processes drive a sustainable development of a riverine landscape and therefore enhances their capability of making decisions.
5) The causal map of the model reflects important information related to a sustainable development of the Kamp valley.
6) Which part of the model was most interesting for you?
7) Which part of the model most should be enhanced?
8) The model can be used for the targeted purpose of teaching students and other interested stakeholders on sustainability issues on a catchment level.
9) For which purpose do you think the presented QR approach is most suited?
    a. Stakeholder integration
    b. University lectures
    c. Decision making
    d. Others (to be added e.g. technical staff from the government, researchers, secondary school students).
10) Additional comments?

For the separate expert evaluations the following statements and questions were additionally used with the same questions being used re-verbalized for both expert evaluations:

11) The entities and configurations are relevant and sufficient to support a representation of the system structure.
12) The quantities used capture the most interesting properties of the entities.
13) The quantity spaces and values capture the most interesting qualitative states of the entities.
14) The (important) model fragments are conceptually correct and clear.
15) The presented scenarios describe a real situation that it is good enough to trigger an interesting/good simulation.
16) The general behavior (how it develops through the simulation) of the presented model is in accordance to what is already known (or accepted).

Results

General results

Both evaluations, the general evaluation and the expert evaluations yielded a very positive feedback with regard to the QR approach, the Garp3 software used to build models and the models themselves representing important issues related to the sustainable development of the riverine landscape Kamp. For example most people 'largely or fully agreed' that QR models represent complex knowledge in an understandable manner and that QR and Garp3 can be seen as a valuable learning tool for understanding real world causal relationships related to a sustainable development of riverine landscapes. Also most people 'largely or fully agreed' that the presented QR models might significantly contribute to the understanding of students and stakeholders which entities and processes drive a sustainable development of a riverine landscape and therefore enhances their capability of making decisions. So the produced software and models in QR language clearly allow students to interact with and learn about sustainable catchment management and to inform managers on the system structure as a basis for decision making. A high potential of an application of QR models in various fields, mainly in education but also in decision making and research was suggested by many participants. The potential of the Garp3 software and the QR approach to sustain collective, interactive social learning, also in a mediated modeling approach, was pointed out. Mainly the identification of dependencies and causal relationships was seen as a prerequisite for understanding a system and therefore also for learning and decision making. With regard to a broader use of QR models in society especially for decision making it was stated, that it might take some time and engagement to establish approaches like that in society. University education using and teaching such approaches was seen as an important basis for a further application.

Evaluation results of model A

Parts of 'model A', that were most interesting for the evaluators were:

- To see the causal interrelatedness of the involved entities of the Kamp management system.
- That private interest might negatively influence the sustainability process.
- Furthermore that the combined influence of planners, science and local population (stakeholders) defines the quality of sustainability plans and the whole sustainability process; this understanding opens up the possibility of different potential intervention options to reach the goal of a sustainable development.
- To see that both, ecological integrity and human well being are represented in the sustainability model.
- Identification of the catastrophic event as trigger for government action for sustainable development.
- The idea that money spent for measures can only be treated as money spent for a community driven development, if the community is involved in the process of developing and implementing measures (otherwise the money is suggested not to be spent for a community driven investment).

Parts of 'model A', that should be enhanced in the eyes of the evaluators were:

- Private interests should be better represented, as a basis to minimize them and achieve sustainable development
- The government action for sustainable development should be better described, as in reality this is of high complexity, being also driven by the general political structure, difficulties between different organization units with regard to their competences (personal behavior) and differences in financial resources; additionally very often policies with complementary aims exist, as policies often lack behind the social development. That means, a more detailed study and representation of the internal political structures determining the implementation process is needed.
- Generally it was noted, that it is of crucial importance to use a well agreed terminology and to well define the terms in use.

With regard to 'model A' it was noted that it could be of relevance, to think about which to degree each of the three known pillars for sustainable development (ecology, society, economy, Pope et al. 2004) is contributing to a sustainable development; in other words probably existing paradigms preferring one of the pillars might prevent a sustainable development (Lackey 1998).

Evaluation results of model B

Parts of the 'model B', that were most interesting for the evaluators were:

- That it is easy to change the content of a scenario by using and exchanging different assumptions allowing for a simplified modeling the effects of the same human pressure on different guilds of fish (positive and negative effects of flow velocity and water temperature on different guilds).

Parts of 'model B', that should be enhanced in the eyes of the evaluators were:

- A more realistic representation of the natural variability of the river discharge (probably by using the random function in the scenario editor) and the amount of abstracted water related to mean annual flow as this defines the frequency of water overflow events at weirs that are suspected to have a significant effect on fish.
- A more realistic representation of the influence of the length of the water abstraction stretch on the temperature development within the river (at the moment the river stretch is treated as a 'container' with the same abiotic factors everywhere).
- Integration of the effect of river morphology on fish and on water temperature.

Additionally collected interesting statements

With regard to the presented models but also to the QR approach some further interesting statements were collected. For example it was stated, that some behaviors of simulations might not be true in real world systems (e.g. that they stay within an interval for a certain time steps before they change). This should be avoided, when not explicitly defined as model target, although there are still QR domain specific ingredients, semantics and behaviors (e.g. the quantity spaces as points and intervals), that might conflict with the intuitive way of stakeholders to express things. Simulation behaviors of presented final models should be restricted as much as needed to avoid outcomes that are not intended (although one also might significantly learn from unwanted outcomes of a simulation). Therefore it is suggested that the end user should (1) only be confronted with simulations & scenarios that exactly show the intended behavior and (2) as less as possible confronted with QR domain specific features not to irritate an intuitive modeling building practice by domain specific restrictions. There were also some suggestions specific to the software (Garp3) produced within the project. With regard to the software packages available for building QR models prior to the project, Garp3 can now be used very intuitively to build QR models representing a prerequisite for the target, to motivate stakeholders and students to use the software and put their conceptual knowledge in causal models. Some specific comments on future developments of Garp3 to make the modeling process easier were also collected. Finally it was stated that a linkage of the causal models to a GIS would open a new field of promising applications.

Summary & Discussion

Integrated catchment management is becoming a central issue for sustainable management of aquatic resources world wide. Although many approaches have been developed, successful implementation of integrated and sustainable management strategies heavily depend on individuals being capable to guide this process. Managers, planners, politicians and scientists are faced with new and complex tasks of the integration of different fields of science with social and political and economical stakeholders. Modeling has been recognized as an important tool that could be used within integrated catchment management processes for various tasks and as the ability of humans to process information and deal with complexity is relatively weak, generalizations are necessary for human (Flood and Carson 1993). Both at the individual and collective levels, coping with complexity requires the ability to strategically filter the vast quantity of available information, and to integrate the key information into some sort of implicit or explicit predictive model. (Beratan 2007). With the qualitative approach presented here, mainly the integration of results from different scientific fields, and soft knowledge from political and social sciences with stakeholder preferences was achieved. At the beginning the modeling process itself turned out to be a challenging task, especially the identification of the essential rates and entities to characterize a system when developing dynamic QR models. The following three questions might guide this definition process:
(1) Which entities should be included?
(2) Which quantities are related to this entity?
(3) Which are the main processes in the system of interest?
The evaluation of the QR approach, the software and the models within the present study gained promising results related to a broader application of QR models in an integrated catchment management. Especially the possibility to run dynamic simulations on conceptual knowledge offers a variety of applications in research and management. Although the presented models were found to be generally suited for the proposed use as learning material for students and to inform managers on the system structure as a basis for decision making, also improvements of the models for a more realistic reflection of the modelled systems were suggested. These suggestions could be easily implemented into the models, and themselves could be treated as results of the modelling process. To establish modeling approaches in the catchment management processes, the education of a new generation of students, managers, planners, scientists and politicians is needed being capable of dealing with this complex issue. The creation of a library of model fragments dealing with all aspects of sustainable catchment management might help to educate students and provide essential information to managers and politicians and scientists. A standard evaluation procedure should be developed to assure the quality of the models and simulations. Only certified models should be re-used, although also other models might represent interesting starting points for various modeling purposes. Qualitative reasoning due to its potential to integrate 'hard' and 'soft' facts, to build causal models and to run dynamic simulations has great potential to become an important contribution to integrated catchment management at multiple levels of the implementation process (Fig. 6).

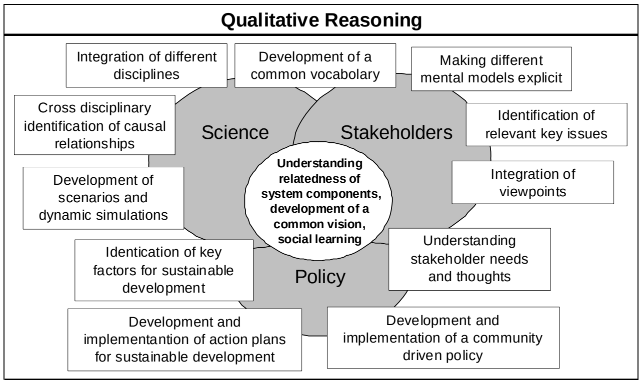


Figure 6: The potential of QR models to frame the process of integrated catchment management.

Acknowledgements. This research is co-funded by the European Commission within the Sixth (project number 004074, project acronym Naturnet-Redime, http://www.naturnet.org.) and Seventh Programme for Research and Development (project number 231526, project acronym DynaLearn, www.DynaLearn.eu). Furthermore the authors would like to thank Bert Bredeweg for helpful comments and suggestions to this manuscript.

References

Araújo, S. C. S., Salles P., and Saito, C. H. 2008. A case study on qualitative model evaluation using data about river water quality. *Ecological Informatics* **3,** 13-25.

Beratan, K. K. 2007. A cognition-based view of decision processes in complex social–ecological systems. *Ecology and Society* **12,** 27.

Boon, P. J., Davies, B. R., and Petts, G. E., Eds. 2000. *Global Perspectives on River Conservation: Science, Policy and Practice.* Chichester: John Wiley.

Bredeweg, B., Bouwer, A., Jellema, J., Bertels, D., Linnebank, F., and Liem, J. 2007a. Garp3: a new workbench for qualitative reasoning and modelling. In: *Proceedings of the 4th international conference on Knowledge capture*, Whistler, BC, Canada, ACM, New York, NY, USA.

Bredeweg, B., and Salles, P. 2002. A case study of collaborative modelling: building qualitative models in ecology. In: B. BREDEWEG (Ed.) Proceedings of the International workshop on Model-based Systems and Qualitative Reasoning for Intelligent Tutoring Systems, San Sebastian, Spain.

Bredeweg, B., Salles, P., Bouwer, A., Liem, J., Nuttle, T., Cioaca, E., Nakova, E., Noble, R., Rios Caldas, A.-L., Uzunov, Y., Varadinova, E., and Zitek, A. 2007b. Towards a Structured Approach to Building Qualitative Reasoning Models and Simulations. *Ecological Informatics* **3,** 1-12.

Dynesius, M., and Nilsson, C. 1994. Fragmentation and flow regulation of river systems in the northern third of the world. *Science* **266,** 753-762.

Grant, W. E. 1998. Ecology and natural resource management: reflections from a systems perspective. *Ecological Modelling* **108,** 67–76.

Jungwirth, M., Muhar, S., and Schmutz, S. 2002. Re-establishing and assessing ecological integrity in riverine landscapes. *Freshwater Biology* **47,** 867-887.

Lackey, R. T. 1998. Seven pillars of ecosystem management. *Landscape and Urban Planning* **40,** 21-30.

Lee, M. H. 2000. Model-Based Reasoning: A Principled Approach for Software Engineering. *Software - Concepts and Tools* **19,** 179-189.

Pope, J., Annandale, D., and Morrison-Saunders, A. 2004. Conceptualising sustainability assessment. *Environmental Impact Assessment Review* **24,** 595-616.

Preis, S., Muhar, S., Habersack, H., Hauer, C., Hofbauer, S., and Jungwirth, M. 2006. Sustainable development of the river landscape Kamp: Presentation of a management process with regard to the requirements of the EU water framework directive (EU-WRRL) *(in German). Österreichische Wasser- und Abfallwirtschaft* **11-12,** 159-167.

Rykiel, E. J. 1996. Testing ecological models: the meaning of validation. *Ecological Modelling* **90,** 229–244.

Salles, P. 2005. Qualitative Representations of Indicators of Environmental Sustainability of the Millennium Development Goals. In: M. HOFBAUR, B. RINNER and F. WOTAWA (Eds.). Proceedings of the Nineteenth International Workshop on Qualitative Reasoning (QR05), Graz, Austria.

Salles, P., Bredeweg, B., and Araujo, S. 2006. Qualitative models about stream ecosystem recovery: exploratory studies. *Ecological modelling* **194,** 80-89.

Salles, P., Bredeweg, B., Caldas, A. L. R., and Nuttle, T. 2007. Modelling sustainability in the Riacho Fundo water basin (Brasília, Brazil). In: C. PRICE (Ed. Proceedings of the 21st International Workshop on Qualitative Reasoning (QR'07), Aberystwyth, Wales, U.K.

Singh, V. P. 1996. *Hydrology of Disasters*. Water Science and Technology Library. Kluwer Academic Publishers.

Stavredes, T. 2000. A system dynamics evaluation model and methodology for instructional technology support. *Computers in Human Behavior* **17,** 409-419.

Sterman, J. 1994. Learning in and about complex systems. *System Dynamics Review* **10,** 291-330.

Tullos, D., and Neumann, M. 2006. A qualitative model for analyzing the effects of anthropogenic activities in the watershed on benthic macroinvertebrate communities. *Ecological Modelling* **196,** 209-220.

Van den Belt, M. 2004. *Mediated modeling: a system dynamics approach to environmental consenus building*. Washington-Covelo-London: Island Press, 339 pp.

Yearley, S. 1999. Computer models and the public`s understanding of science: a case study analysis. *Social Studies of Science* **29,** 845-866.

Zitek, A., Muhar, S., Preis, S., Schmutz, S., Bredeweg, B., and Salles, P. 2006. QR models and documentation for learning about sustainable development, focusing on basic ecological and socio-economic features for an integrative and sustainable development of the riverine landscape of the Kamp valley. Deliverable D6.6.2, STREP project FP6 no. 004074, Naturnet-Redime.