Identification and Characteristic Descriptions of Procedural Chunks

Case study on a game of chess

Jana Krivec
Department of Intelligent systems
Jožef Stefan Institute
Ljubljana, Slovenia
jana.krivec@ijs.si

Matej Guid, Ivan Bratko
Artificial Intelligence Laboratory
Faculty of computer and information science
Ljubljana, Slovenia
{matej.guid, ivan.bratko}@fri.uni-lj.si

Abstract
When dealing with cognitive architecture and behavior, chunks are one of the most well known and accepted constructs. Despite that, the nature of chunks still remains very elusive, especially with understanding chunks in procedural knowledge. Our attempt is to show the existence of chunks in procedural knowledge, define them, and describe their characteristics. With this purpose in mind, we use data mining techniques. We chose the game of chess as an experimental domain, due to its complexity, well defined rules, and a standardized measure of chess-players' knowledge. Results could contribute to the understanding of human information processing and cognitive architecture. They could be beneficial for tutoring and student modeling, and may serve as a framework for knowledge-based driven AI agents.

Keywords- procedural knowledge, information processing, chunk, machine learning, chess

I. INTRODUCTION

The cognitive sciences main questions are: How information is stored in memory, and how it is organized and retrieved when needed. In short: How information is processed. Most of the existing theories of human information processing have one concept in common - chunks (Chunking theory [10]; Template theory [5], ACT [1]). Chunks are pieces of meaningful information (stored in long-term memory) which enable easier learning and perception [9]. Although the concept of chunks is wide-spread and accepted, the nature of chunks is far from being clearly defined and understood. For the purpose of information processing exploration in humans, chess has proven to be a fertile ground for ideas and techniques of cognitive psychology, as well as in artificial intelligence (AI). It provides a complex reasoning problem from a manageable domain with a built-in performance criterion (ELO rating). Richness and adaptability of the chess environment provides many ways of mining for useful knowledge using several types of various descriptive features (or attributes). According to previous studies [5, 10], results obtained on a domain of chess can be generalized to other expert fields, thus meaning that chess has high external validity.

Most of experiments and theories of the chunking mechanisms in general and in chess specifically refer to chunks in declarative knowledge, *i.e.,* knowledge about facts [5, 10]. However, chess experts don't hesitate to confirm that chess moves (representing procedural knowledge) are also connected into meaningful units, so called chess motives. Procedural knowledge in general is the knowledge exercised in the performance of some task. In AI, procedural knowledge is one type of knowledge that can be possessed by an intelligent agent. Such knowledge is often represented as an algorithm, a finite-state machine or a computer program. In contrast, an AI system based on declarative knowledge might just contain a map of the building, together with information about the basic actions that can be done by the robot (like moving forward, turning, and stopping), and leave it to a domain-independent planning algorithm to discover how to use those actions to achieve the agent's goals [9]. In contrast to declarative knowledge, human procedural knowledge is more implicit, not easily articulated by the individual, since it is typically tacit and therefore more difficult to measure [11]. Consequently, there is lack in experiments and theories referring to chunks in procedural knowledge. However, chess players (and books) show chess variations all the time, that is, they operate with and communicate procedural knowledge all the time. Nevertheless, despite such a clearly visible role of procedural knowledge in chess, there has been surprising lack of study of this type of knowledge.

In our work, we intend to show the existence of chunks in chess variations (sequences of moves), which reflect the organization of procedural knowledge. To emphasize the difference with respect to well-known chunks in declarative knowledge, we will refer to these meaningful units in procedural knowledge as *procedural chunks.* We will try to extract chunks from chess games and explore the characteristics of obtained chunks. In order to investigate the nature of chunks in procedural knowledge, a reproduction experiment of chess variations will be conducted. Data will be processed with machine learning techniques, since they proved to be a powerful tool for discovering potential complex patterns [4].

II. METHOD

A. Participants

52 chess players and 50 non-chess players participated in the experiment. The average age of a chess player group was 26 years (ranging from 14-56), their average chess strength (measured in international “ELO” system) was 2275 ELO points (ranging from 1600 ELO, which represent the beginners, up to 2670 ELO which is a rating of top grandmasters).

B. Experimental procedure

Participants were asked to reproduce two kinds of tasks: chess variations (chess move sequences) as an experimental condition and irregular variations (move sequences, where chess pieces don’t move by chess rules) as a controlling condition (control of the memory capacity without any knowledge). 5 chess variations and 5 irregular variations were presented. The beginning of the variation was in the middle game, in order to prevent the possibility of participants being previously acquainted with the opening. When selecting different variations, attributes that might influence memorizing were taken into account, *e.g.*: number of pieces on the chess board at the beginning of the variation, number of different piece types included in the variation, number of squares occupied during the variation, style of the game (tactical/positional/mixed), number of pieces at the end of the variation, etc. Time interval between the consecutive moves was 2 seconds, as it is assumed to be long enough for individual to process the information to long-term memory. Length of the variations was 16 moves (or 32 half moves, *i.e.*, with white and black counted separately). All the experiments were conducted using specialized software designed for the purpose of this study. It uses classical and well-known WinBoard chessboard representation, and records exact times of user inputs in reproductions of the chess variations previously displayed by the program.

Before the beginning of the experiment participants were given the following instructions: “You will be presented a middle-game chess position. You can have a look at it and when ready you can press the button. At this point 16 moves-long variation will be shown to you. The time interval between two consecutive moves will be 2 seconds. Watch carefully and try to remember as many moves as possible. After the variation presentation is finished, you will be shown the beginning position. Your task will be to repeat the original variation as correctly as possible, while being careful to repeat it as fluently as possible, without pauses between moves.” They were warned that the time of their reconstruction will be measured and that the correction of moves is not possible.

C. Data evaluation procedure

Reconstructed moves were evaluated in two ways. Firstly, we took into consideration only moves that were reconstructed correctly (“correct moves”). Secondly, we were evaluating all the reconstructed moves, which were in the original variation no matter if the move-order was correct or not (“included moves”).

1) Defining chunks in procedural knowledge

When defining/extracting chunks in procedural knowledge, we first considered a method for defining chunks in declarative knowledge, developed by Chase and Simon [10]. It is based on the assumption that longer time interval during the reconstruction of a meaningful material (*i.e.*, the material about which we have relevant knowledge) reveals the recall of a new structure/chunk from the long-term memory. We adapted this method for reconstruction of each chess variation. In our case, longer time interval between two consecutive reconstructed moves reflects the recall of a new motive/chunk from the memory storage. “Longer time interval” was defined in the following way: Times of each participant were normalized; one way of normalization was calculating the percentage of time used for a certain move with regard to time spent for the whole reconstruction while the second way was conversion into z-values. The quantity z represents the distance between the raw score and the population mean in units of the standard deviation. z is negative when the raw score is below the mean, positive when above [14]. Further on, for each move a median of the normalized times of all the participants who reconstructed the particular move was calculated. Next, for all the moves in one game an average value and standard deviation of normalized time medians was calculated. All the moves that exceeded the boundary of the average value plus one standard deviation were considered as a “long time interval” and as such the beginning of a new procedural chunk.

Our second method for defining procedural chunks involved the number of reconstructions for the particular move, or so called collective reconstruction (an idea introduced by Bratko *et al.* [2]). In this case, we counted the number of participants that reconstructed a particular move. Big declines in the number of reconstructions (top 25%) indicated the beginning of a new chunk.

The third version of chunks definition is based on association rules. Association rule mining finds interesting associations and/or correlation relationships among large sets of data items. They show attribute value conditions that occur frequently together in a given dataset [13]. In our case, they join together the moves that are closely connected with each other, *i.e.*, many participants reconstruct them together. When such a unit is detected, the moves constructing it are eliminated from further analysis. Using this procedure, we divided a chess variation into a number of chunks, *i.e.*, units that are closely connected among each other and weakly connected with other moves.

Final chunks are defined with a comparison of all three above mentioned methods.

2) Exploring the characteristics of the procedural knowledge chunks

Our second major goal was to describe the characteristics of chunks in procedural knowledge. With this purpose in mind, we extracted many different attributes that might affect the reconstruction of chunks and treated them with machine learning (ML) techniques in order to clarify the nature of procedural chunks. A major focus of machine learning research is to automatically recognize complex patterns in data. The constructed knowledge is often in the form of readable, understandable trees, rules, and other representations thus enabling further study and fine tuning. Two examples of successful scientific and engineering ML tools are Weka [12] and Orange [3].

Data for machine learning and data mining are most commonly presented in attribute-class form, *i.e.*, a "learning matrix", where rows represent examples, and columns attributes [3].

In our case, an example is a move of the variation, described with more than 100 different attributes. Table I lists some of them.

TABLE I. DESCRIPTIONS OF ATTRIBUTES ("RYBKA" STANDS FOR "CHESS PROGRAM RYBKA 3 AT 10-PLY SEARCH DEPTH").

Name	Description
MoveNo	move sequent number
MoveValue	evaluation of the position after move played (Rybka)
MoveRank	rank of the move played in the original variation (Rybka)
MoveDiff	difference between the evaluation of the best move and the move in the variation (Rybka)
Nodes	number of nodes evaluated (Rybka)
AllMoves	number of all possible moves in the position
SimMoves	number of moves of similar quality in the position (+/-0.50 in Rybka evaluation)
PieceType	moved piece type
SqPassed	number of squares passed over by the pieces in particular variation
SqOccupied	number of squares occupied by pieces in the variation
AllPieces	number of pieces in the position
PlayType	type of play (positional, tactical, mixed)
MoveLength	move length (measured in squares passed by the moved piece)
Crafty's attributes	Crafty's evaluation function attributes (around 100 attributes)

The values of these attributes were defined for each move in a variation.

Class values were normalized times and number of reconstructions. Class values were discretized in different ways; most commonly into three groups (lower than M-1SD, from M-1SD to M+1SD and above M+1SD).

Typically, data mining involves repeated experimentation with data, using different methods, parameters, and data to find most meaningful relations. In our case, data was processed on different levels: games were once treated unified, while other time each of the five games was treated separately. Furthermore, we processed (a) each move separately and (b) moves joined into chunks as exposed by our chunk defining methods. All of these different types of data combinations were performed for "included" and "correct" moves.

III. Results

A. General findings

Chess players reconstruct "chess variations" significantly better than "illegal variations". This difference does not appear in the case of non-chess players. We considered this as a proof of a relevant knowledge contribution to better reconstruction success. This finding is similar to that of Simon and Chase regarding reconstructing legal vs. illegal chess positions [10].

Furthermore, it was shown that better players (players with higher ELO rating) have better reconstruction success rate of legal variations (see Figure 1). This is another indicator that more expert knowledge results in better success rate in perception and memorizing. The results are accordant with previous researches of Simon [10] and Gobet [5].

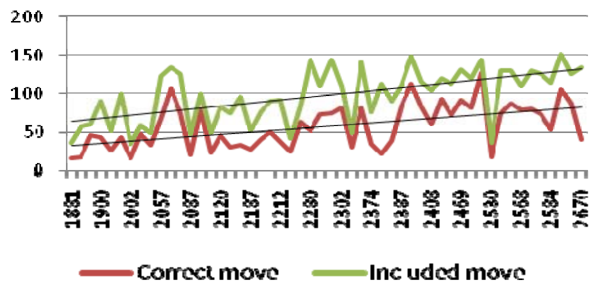

Figure 1. Reconstruction success according to ELO in terms of the number of "correct" and "included" moves reconstructed.

By further data analyses and exploration, with the help of clustering analysis on reconstruction success rate of "correct" and "included" moves, chess players were divided into 2 groups regarding their chess strength. Group A: Average ELO=2407 (SD=178, N=25); Group B: Average ELO=2135 (SD=179, N=27). The separation ELO point, which divides the two groups, was calculated on the following way:

$$\frac{(AvgELOgroupA\text{-}1SD)+(AvgELOgroupB+1SD)}{2} = 2271{,}41$$

The ELO of 2271,41 is concidered to ba a splitting point between better and worse chess players.

All the procedures and methods were separately applied (a) on different ELO groups and (b) for all participants together.

Table II shows how many ("correct" and "included") moves on average were correctly reconstructed by differently strong chess players (divided into the two groups using the before mentioned clustering analysis), and average times spent for the reconstruction ("correct", "included", and all reconstructed moves being treated separately).

TABLE II. NUMBER OF RECONSTRUCTED MOVES AND TIMES SPENT FOR THE RECONSTRUCTION BY GROUPS OF DIFFERENTLY STRONG CHESS PLAYERS (IN LEGAL VARIATIONS)

		No. ”Correct” (moves)	No. ”Included” (moves)	Avg Time ”Correct” (sec.)	Avg Time “Included” (sec.)	Avg Time (sec.)
ELO>2271	**M**	**71,46**	**117,50**	**3,07**	**3,34**	**3,48**
	SD	26,93	27,62	1,45	1,20	1,29
ELO≤2271	**M**	**42,33**	**75,63**	**4,07**	**5,08**	**5,75**
	SD	22,79	29,07	1,31	1,60	1,80

The results clearly show that:

1. stronger chess players have better reconstruction success,
2. stronger chess players have shorter reconstruction times,
3. both weaker and stronger chess players on average spent less time for reconstruction of correctly reconstructed moves (it is worth noting also that the times spent for “correct” moves is shorter than for merely “included” moves).

First and second results are congruent with previous findings, stating that better relevant knowledge results not only in more accurate, but also in quicker responses in new situations [15]. The third result confirms our observations during the conduction of the experiments that the time spent for the reconstruction of a certain move appears to be a good indicator of the mistakes in the reconstruction process: the correct moves were often reconstructed quicker.

Further general findings of information processing referred to description of the context influence on the reconstruction success. One way of finding meaningful relations between reconstruction success and the context characteristics was the application of C4.5 [12], a ML method used for induction of classification trees. This method is most commonly used when the emphasis is on transparency of the constructed knowledge. An example of a decision tree acquired is shown in Figure 2. From that tree we can conclude that the sequent move number in connection with the dispersion of the variation (number of squares occupied by pieces in the variation) has an influence on the reconstruction success. The further the sequent move number and the higher the dispersion, the lower is the reconstruction success.

The most powerful classification trees are those with best classification accuracy. To estimate the accuracy of the trees, we used 10-fold cross-validation. The estimated accuracy of a classification tree corresponds to the probability that a new example will be correctly classified. In the presented tree, classification accuracy is very high. It should be noted, however, that the predicted classes are only

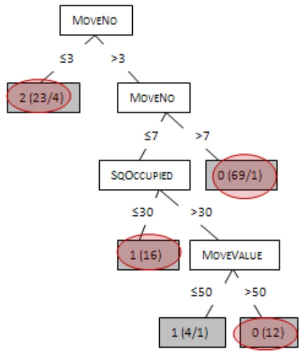

Figure 2. Decision tree constructed from conditions when all the games and all the participants were taken into account together, moves were treated separately and the number of the correct moves was set as a class value; 0=lower than M-1SD, 1=from M-1SD to M+1SD, 2=higher than M+1SD . The classification accuracy is 91.94%.

qualitative predictions of numerical values (*e.g.*, within one standard deviation from the median).

Similarly, we can conclude that the fewer equally good moves there are in a position, and fewer possible moves there are in a position, the shorter is time needed for the reconstruction of the move. Another example would be that reconstruction is better in tactical than in positional chess variations, probably due to smaller move dispersion and fewer equally good moves in each position.

Our next aim is to evaluate whether moves by stronger chess pieces are easier to reconstruct than those by the weaker pieces, and if pieces closer to the participant are easier to reconstruct than the more distant ones as proposed by Grimbergen [6]. Besides, we will explore whether the move length (number of squares the piece passed over during its way from the initial to the final square) impact the reconstruction of the moves.

Last but not least, we concluded that reconstruction success does not differ regarding gender and age differences.

B. Chunk extraction

Chunks were extracted for every chess game in the experiments. In Figure 3 we can see an example of an extracted procedural chunk. Moves 1.Rxg6 Rxh4 2.Rg8 Rg4 3.g6 (1-5 in Fig. 3) are found to be more closely associated together and represent a chunk, while 3…Kb6 (6) is a move indicating a new chunk, which is by its context significantly different from the previous one.

It is particularly interesting that all the participants that correctly reconstructed moves up to the diagramed position also correctly (and relatively quickly) reconstructed every single move of the extracted procedural chunk, while many failed to reconstruct the next move in the sequence (Kc5-b6), also spending considerably more time immediately after the moves in the chunk were reconstructed.

Figure 3. An example of a procedural chunk.

C. *Chunk characteristics*

When considering procedural chunks characteristics, we have two main questions in mind: What influences the chunks length, and what influences their difficulty. Difficulty is defined by reconstruction success rate and average time spent for the reconstruction of a chunk. The before mentioned attributes will be taken into account. In this case, a class will be defined by (a) an average number of "correct/included" moves in a chunk, (b) time of the moves in a chunk, and (c) the chunk length.

Answering these questions belongs to future work.

IV. PRACTICAL APPLICATION AND BENEFITS

In this paper, we presented our work that aims at the following three contributions to the understanding of human information processing and cognitive architecture:

1. We introduced the notion of procedural chunks.
2. We proposed methods for extraction of procedural chunks and identification of their characteristics.
3. We presented some general findings about information processing and some initial results that emerged from the early phase of our research.

In general, our findings could be applied when teaching procedural knowledge. The obtained results could also be useful as a theoretical framework when trying to develop knowledge-based AI agent. Furthermore, with the ever more extensive e-tutoring development, the results of our study could be used in student modeling. Levinson *et al.* [8] argued that almost no competitive chess programs use AI language or knowledge representation methods, since they are too slow for a real time, high performance applications. Despite of enormous progress in the power of chess programs, their capabilities to *explain* why certain moves are good or bad in a language understandable to humans are very limited. The computer chess community has done embarrassingly little research in the areas of intelligent chess tutoring and automatic annotation of chess games, where knowledge representation and acquisition are of considerable importance, and human information processing and cognitive architecture may prove to be important as well. In our specific case, the results could be beneficial in building a knowledge base for an automated chess tutor [7], a computer program which automatically annotates chess games in a humanlike way. It might help with suggestions such as when to comment, how to treat users with different chess strength, etc.

REFERENCES

[1] J. R. Anderson., C. B. Boyle, and B. J. Reiser, "Intelligent tutoring systems". Science, vol. 228, April 1985, pp. 456–462, doi:10.1126/science.228.4698.456. 17746875.

[2] I. Bratko, P. Tancig and S. Tancig, "Detection of Positional Patterns in Chess", International Computer Chess Associaton Journal, Vol.7 (2), 1984, p.p.63-73

[3] J. Demsar, and B. Zupan, "From Experimental Machine Learning to Interactive Data Mining", White Paper (www.ailab.si/orange), 2005, Faculty of Computer and Information science, University of Ljubljana.

[4] J. Furnkranz, Machine lerning in computer chess: the next generation." International Computer Chess Association Journal, Vol 19, 1996, pp. 147-161

[5] F., Gobet, P. C. R. Lane, S. Croker, P. C-H. Cheng, G. Jones, I. Oliver and J. M. Pine, "Chunking mechanisms in human learning". *TRENDS in Cognitive Sciences, vol. 5,* 2001, pp. 236-243

[6] R. Grinbergen, "Cognitive Modeling of Knowledge-Guided Information Acquisition in Games" *Proc. Computers and Games, Lecture Notes in Computer Science,* Springer, *Vol. 5131*, Oct. 2008, pp. 169-179.

[7] M. Guid, M. Možina, J. Krivec, A. Sadikov, and I. Bratko, "Learning Positional Features for Annotating Chess Games: A Case Study." *Proc. Computers and Games, Lecture Notes in Computer Science,* Springer, *Vol. 5131*, Oct. 2008, pp. 192-204.

[8] R. Levinson, H. Feng-hsiung, T. A. Marsland, J. Schaeffer, D. E. Wilkins, "The Role of Chess in Artificial Intelligence Research". IJCAI, vol.2, 1991, pp. 547-552.

[9] G. A. Miller, "The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information." Psychological Review, Vol. 63, 1956, pp. 81-97.

[10] H. A. Simon, W. G. Chase, "Perception in chess," Cognitive Psychology, vol.4, 1973, pp.55-81.

[11] http://en.wikipedia.org/wiki/Procedural_knowledge

[12] I. H. Witten, E. Frank, "Data Mining – Practical Machine Learning Tools and Techniques (sec. ed.)," Morgan Kaufmann, 2005.

[13] http://www.resample.com/xlminer/help/Assocrules/associationrules_intro.htm

[14] http://en.wikipedia.org/wiki/Standard_score

[15] http://lesonexchange.net/factors_affecting_perception.htm