ABML Knowledge Refinement Loop: A Case Study

Matej Guid[1], Martin Možina[1], Vida Groznik[1], Dejan Georgiev[2], Aleksander Sadikov[1], Zvezdan Pirtošek[2], and Ivan Bratko[1]

[1] Faculty of Computer and Information Science, University of Ljubljana, Slovenia
[2] Department of Neurology, University Medical Centre Ljubljana, Slovenia

Abstract. Argument Based Machine Learning (ABML) was recently demonstrated to offer significant benefits for knowledge elicitation. In knowledge acquisition, ABML is used by a domain expert in the so-called ABML *knowledge refinement loop*. This draws the expert's attention to the most critical parts of the current knowledge base, and helps the expert to argue about critical concrete cases in terms of the expert's own understanding of such cases. Knowledge elicited through ABML refinement loop is therefore more consistent with expert's knowledge and thus leads to more comprehensible models in comparison with other ways of knowledge acquisition with machine learning from examples. Whereas the ABML learning method has been described elsewhere, in this paper we concentrate on detailed mechanisms of the ABML knowledge refinement loop. We illustrate these mechanisms with examples from a case study in the acquisition of neurological knowledge, and provide quantitative results that demonstrate how the model evolving through the ABML loop becomes increasingly more consistent with the expert's knowledge during the process.

1 Introduction

Machine learning has long ago been proposed as a way of addressing the problem of knowledge acquisition [1]. While it was shown that it can be successful in building knowledge bases [2], the major problem with this approach is that automatically induced models rarely conform to the way an expert wants the knowledge organized and expressed. Models that are incomprehensible have less chance to be trusted by experts and users alike [3]. In striving for better accuracy, modern trends in machine learning pay only limited attention to the comprehensibility and the intuitiveness of prediction models [4].

A common view is that a combination of a domain expert and machine learning would yield the best results [5]. Argumentation Based Machine Learning (ABML) [6] naturally fuses argumentation and machine learning. One of the advantages over traditional machine learning methods is better comprehensibility of the obtained models. Improvement in comprehensibility is especially important in the light of knowledge acquisition. Through argumentation, ABML enables the expert to articulate his or her knowledge easily and in a very natural way. Moreover, it prompts the expert to share exactly the knowledge that is most useful for the machine to learn, thus significantly saving the time of the expert.

ABML is comprised of two main parts: the modified machine learning algorithm that can handle and use arguments, and the iterative ABML loop that manages the interaction between the expert(s) and the machine. The algorithm usually takes all the credit for successful results, however, the iterative loop is at least as important. After all, the loop is what the expert and the knowledge engineer actually use during the process of knowledge elicitation, while the inner workings of the algorithm remain hidden in the background. Therefore, in this paper, the focus is (solely) on the loop part of the ABML process.

The paper represents a continuation of the work presented in [7] and [8]. Here we provide a step-by-step presentation of the knowledge elicitation process with ABML. Through a case study of acquiring knowledge for a neurological decision support system, we analyze and clearly demonstrate the benefits of this particular style of the interaction between the experts and the machine learning algorithm. Along the way, the reader is alerted to some typical and atypical situations and is shown how to deal with them. The paper illustrates what is expected from the experts and knowledge engineers alike, demonstrates the required level of their involvement, and conveys the natural feel of the human-computer interaction. We also demonstrate quantitatively how the model obtained with the ABML knowledge elicitation process becomes more and more consistent with the expert's knowledge.

The organization of the paper is as follows. Chapter 2 describes the case study domain and experimental setup, and Chapter 3 shortly describes the ABML algorithm, focusing mainly on its modifications in view of the current application. The ABML iterative process, the main part of the paper, is in Chapter 4. We finish with an evaluation of the knowledge elicitation process and conclusions.

2 Domain Description and Experimental Setup

We are developing a neurological decision support system (DSS) to help the neurologists differentiate between three types of tremors: essential, Parkinsonian, and mixed tremor (comorbidity, see [8] for more detail). The system is intended to act as a second opinion and a teaching tool for the neurologists. Although several sets of guidelines for diagnosing both essential and Parkinsonian tremor do exist [9], none of them enjoys general consensus in the neurological community.

The data set consisted of 114 patients. These were divided into a learning set with 47 examples and a test set with 67 examples. The class distribution was: 50 patients diagnosed with essential tremor (ET), 23 patients with Parkinsonian tremor (PT), and 41 patients with a mixed-type tremor (MT). The patients were described by 45 attributes.

According to the domain experts, some of the characteristics reflected in these attributes speak in favour of a particular tremor type as follows.

Essential tremor is characterized by *postural tremor*, *kinetic tremor*, *harmonics*, *essential spiral drawings*, positive family *anamnesis* etc.

Parkinsonian tremor is characterized by *resting tremor*, *bradykinesia*, *rigidity*, *Parkinsonian spiral drawings* etc.

Mixed tremor implies presence of both essential tremor and Parkinsonian tremor.

However, according to domain experts it is difficult to combine these characteristics into sensible rules for successful diagnosis.

3 Argument Based Machine Learning (ABML)

Argument Based Machine Learning (ABML)[6] is machine learning extended with concepts from argumentation. In ABML, arguments are used as means for experts to elicit some of their knowledge through explanations of the learning examples. The experts need to focus on one specific case at the time only and provide knowledge that seems relevant for this case. We use the ABCN2 [6] method, an argument based extension of the well-known CN2 method, that learns a set of unordered probabilistic rules from examples with attached arguments, also called *argumented examples*.

The problem domain described in this paper contains a class variable with three values. According to the domain expert opinion, it was appropriate to translate our three-class problem into two binary-class problems solved by two binary classifiers. The first binary classifier distinguishes between ET and non-ET, the second between PT and non-PT. A new case is then probabilistically classified roughly as follows. The first classifier assigns probability p(ET) to class ET, the second p(PT) to class PT. The predicted probability of MT is then p(MT) = 1 – p(ET) – p(PT). ET and MT are merged into EMT class, while PT and MT are merged into PMT class. The two binary classifiers are independent, so it may happen that p(MT) < 0. In such cases the three probabilities are adjusted to satisfy the formal properties of probabilities (see [8] for details).

3.1 ABML knowledge refinement loop

The ABML knowledge refinement loop consists of the following steps:

Step 1: Learn a hypothesis with ABCN2 using given data.
Step 2: Find the "most critical" example and present it to the expert. If a critical example can not be found, stop the procedure.
Step 3: Expert explains the example; the explanation is encoded in arguments and attached to the learning example.
Step 4: Return to step 1.

In the sequel, we explain (1) how we select critical examples, and (2) how we obtain all necessary information for the chosen example.

Identifying critical examples. A critical example is an example the current hypothesis can not explain very well. As our method gives probabilistic class prediction, we first identify the most problematic example as one with highest probabilistic error. To estimate the probabilistic error we used a k-fold cross-validation repeated n times (e.g. $n = 4, k = 5$), so that each example is tested n times. The critical example is thus selected according to the following two rules.

1. If the problematic example is from class MT, it becomes the critical example.
2. Otherwise, the method will seek out which of the rules is the culprit for example's misclassification. As the problematic rule is likely to be bad, since it covers our problematic example, the critical example will become an example from PT or MT class (or ET or MT) covered by the problematic rule. Then, the expert will be asked to explain what are the reasons for the patient's diagnosis. Domain expert's explanations should result in replacing the problematic rule with a better one for the PMT (or EMT) class, which will not cover the problematic example.

Are the expert's arguments good enough or should they be improved? Here we describe in details the third step of the above algorithm:

Step 3a: Explaining a critical example. If the example is from the MT class, the expert can be asked to explain its Parkinsonian and essential signs (which happens when the problematic example is from MT) or to explain only one of the diseases. In the other two cases (ET or PT), the expert always explains only signs relevant to the example's class. The expert then articulates a set of reasons confirming the example's class value. The provided argument should contain a minimal number of reasons to avoid overspecified arguments.

Step 3b: Adding arguments to an example. The argument is given in natural language and needs to be translated into domain description language (attributes). If the argument mentions concepts currently not present in the domain, these concepts need to be included in the domain (as new attributes) before the argument can be added to the example.

Step 3c: Discovering counter examples. Counter examples are used to spot if an argument is sufficient to successfully explain the critical example or not. If not, ABCN2 will select a counter example. A counter example has the opposite class of the critical example, however it is covered by the rule induced from the given arguments.

Step 3d: Improving arguments with counter examples. The expert needs to revise his initial argument with respect to the counter example.

Step 3e: Return to step 3c if counter example found.

4 Knowledge Elicitation Process for Differentiating Tremors

In this section, we analyze the complete knowledge elicitation process for differentiating between essential and Parkinsonian tremors. We identify the main effects of each iteration on the process.

Iteration 1 Example E.65 (classified as MT) was the first critical example selected by our algorithm. The expert was asked to describe which features are in favor of ET *and* which features are in favor of PT. He explained that the presence of harmonics speaks in favor of ET, while the presence of bradykinesia speaks in favor of PT. Both features were selected as the most influential ones.

The presence of harmonics was represented by four attributes in the data set, each one with possible values of *true* and *false*. The expert explained that just one of these feature values being *true* already suffices to decide in favor of ET. Similarly, the presence of bradykinesia was indicated with two attributes, one for the left side and one for the right side, with possible values in range from 0 (not indicated) to 5 (high). The expert explained that the side does not play any particular role for differentiating between ET and PT, and that any value higher than zero already speaks in favor of PT.

The expert's explanation served the knowledge engineer to induce two new attributes: (1) HARMONICS, with possible values *true* (indicating the presence of harmonics) and *false*, and (2) BRADYKINESIA, with possible values *true* (bradykinesia is present on the left side *or* on the right side) and *false* (bradykinesia was not indicated on *either* side). At the same time the original six attributes (indicating harmonics and bradykinesia) were excluded from the domain, since it is their *combination* (reflected in the expert's argument) that provides relevant information according to the expert.

Based on the expert's explanation, the reasons (1) "HARMONICS is *true*" and (2) "BRADYKINESIA is *true*" were added as the arguments for ET and PT, respectively, to the critical example E.65.

The method selected E.67 as a counter example for the expert's argument in favor of ET, and E.12 as the counter example for his argument in favor of PT. The expert was now asked to compare the counter example E.67 with the critical example E.65, and to explain what is the most important feature in favor of ET in E.65 that does not apply to E.67. Similarly, he was asked to explain what is the most important feature in favor of ET in E.65 that does not apply to E.12.

It turned out that both counter examples occurred as consequences of the following errors in the data set. In case of E.67, one of the original four attributes for harmonics was set to *true*, although the actual value was discovered to be *false* upon examination. Consequently, the value of the newly added attribute HARMONICS in E.67 had to be corrected from *true* to *false*. In case of E.12, upon the examination of the feature values, the expert realized that some strong arguments in favor of PT were overlooked at the time of diagnosis. After careful deliberation, the class of E.12 was modified from ET to MT by the expert.

In this case, the method actually helped to discover errors in the data set. Improving the arguments turned out to be unnecessary: the correction of the aforementioned errors in the data set resulted in no further counter examples. Thus, Iteration 1 was concluded.

Iteration 2 Upon entering into Iteration 2, E.61 (MT) was selected as a critical example. The expert was asked to describe which features are in favor of ET. He gave two features as an explanation: the presence of postural tremor and the presence of resting tremor. Similarly as in Iteration 1, these two features were each represented by two attributes. Again, the expert explained that neither the side nor the magnitude of a non-zero value play any particular role in differentiating between ET and PT.

The expert's explanation served to induce two derived attributes: (1) POSTURAL, with possible values *true* (indicating the presence of postural tremor) and *false*, and (2) RESTING, with possible values *true* (indicating the presence of resting tremor) and *false*. The original four attributes (indicating the presence of postural tremor and resting tremor) were excluded from the domain. The reason "POSTURAL is *true* and RESTING is *true*" was added as the argument for ET to the critical example E.61.

The expert's argument did not prove to be sufficient to produce a rule with pure distribution: the method selected E.32 as the counter example. The expert was asked to compare the counter example E.32 with the critical example E.61.

The expert spotted the presence of bradykinesia in E.32 as the most important difference, and thus extended his argument to "POSTURAL is *true* and RESTING is *true* and BRADYKINESIA is *false*."

A new counter example was found by the method: E.51. It turned out that this counter example occurred as a consequence of another misdiagnosis. After reviewing the feature values describing the patient's conditions, the expert modified the class of E.51 from MT to PT. No further counter examples were found.

Iteration 3 Critical example E.55 (MT) was presented to the expert. The expert gave two features in favor of ET: the presence of postural tremor and the presence of kinetic tremor. He also gave two features in favor of PT: the presence of bradykinesia and the presence of rigidity in upper extremities. Kinetic tremor and rigidity were each represented by two attributes, similarly as in some aforementioned cases.

The knowledge engineer induced two attributes: (1) KINETIC, with possible values *true* (indicating the presence of kinetic tremor) and *false*, and (2) RIGIDITY, with possible values *true* (indicating rigidity in upper extremities) and *false*. The original four attributes (indicating the presence of kinetic tremor and rigidity in upper extremities) were excluded from the domain.

Expert's explanation lead to (1) "POSTURAL is *true* and KINETIC is *true*," and (2) "BRADYKINESIA is *true* and RIGIDITY is *true*" as the arguments for ET and PT, respectively, to the critical example E.55.

The method selected E.63 as a counter example for the argument in favor of ET. The expert explained what is the most important feature in favor of ET in the critical example E.55 that does not apply to E.63. He contemplated that ET typically occurs much earlier than PT, and advocated that if tremor occurs before the age of 50 (as in E.55), it is usually ET. There were no counter examples for the argument in favor of PT.

The knowledge engineer realized that there was no suitable attribute in the domain that would express exactly what the expert had just explained. There were similar attributes AGE (indicating the age of the patient) and TREMOR.PERIOD (indicating the number of years since the tremor was diagnosed). They were used to construct a new attribute TREMOR.START, indicating the patient's age when the tremor was diagnosed.

The argument in favor of ET was extended to "POSTURAL is *true* and KINETIC is *true* and TREMOR.START < 50." No more counter examples were found.

Iteration 4 Critical example E.51 (MT) was presented to the expert. The expert explained that positive anamnesis and postural tremor are in favor of ET, while qualitative assessment (given by the neurologist at the time of the examination of a patient) is in favor of PT. The arguments to E.51 therefore became (1) "ANAMNESIS is *positive* and POSTURAL is *true*" for ET, and (2) "QUALITATIVE.ASSESSMENT is *PT*" for PT.

With the help of the counter example E.62 the argument for ET was extended to "ANAMNESIS is *positive* and POSTURAL is *true* and BRADYKINESIA is *false*." Another counter example E.21, for argument in favor of PT, turned out to be misdiagnosed. The class of E.21 was modified from ET to MT. There were no further counter examples.

Iteration 5 Critical example E.42 (MT) was presented to the expert. The qualitative assessment by the neurologist was given as sufficient argument in favor of ET. The assessment of free-hand spiral drawings were in favor of PT, as can be seen from the following explanation by the expert: "The assessment of the free-hand spiral in some of the four observations in the original data is Parkinsonian, and none of them indicative of essential tremor."

This explanation lead to a new attribute SPIRO.FREE.PT.ONLY. By analogy, another attribute, SPIRO.FREE.ET.ONLY, was introduced, while the four original attributes were excluded from the domain upon consultation with the expert. The arguments to E.42 became "QUALITATIVE.ASSESSMENT is *ET*" for ET, and "SPIRO.FREE.PT.ONLY is *true*" for PT.

No counter examples opposing the argument for ET were found, while E.33 was given as the counter example against the argument for PT. The expert mentioned template-based spiral drawings: “The assessment of the template-based spiral in some of the four observations (attributes) in the original data are essential in E.42, and none of them is Parkinsonian. This does not apply to E.33.”

Similarly as in the above case, new attributes SPIRO.TEMPLATE.PT.ONLY and SPIRO.TEMPLATE.ET.ONLY were introduced, and the four original attributes were excluded from the domain. The argument attached to E.42 for PT was extended to “SPIRO.FREE.PT.ONLY is *true* and SPIRO.TEMPLATE.ET.ONLY is *false*.” This time no counter examples were found by the algorithm.

Iteration 6 Critical example was E.39 (MT). The expert’s argument for ET were postural tremor and the qualitative assessment in favor of ET. The presence of resting tremor was the argument for PT.

No counter examples opposing the argument in favor of ET were found. Counter example against the argument in favor of PT became E.45. When comparing this counter example with the critical example, the expert spotted an important difference: the lack of harmonics in the critical example, and their presence in the counter example. The argument in favor of PT was thus extended to “RESTING is *true* and ANY.HARMONICS is *false*.” No new counter examples were found.

Iteration 7 Iteration 7 turned out to be exceptionally short. When the expert was asked to give arguments in favor of ET and PT for the critical example E.26 (MT), he realized that there were no valid arguments in favor of ET. The class was therefore changed from MT to PT.

Iteration 8 Critical example E.30 (ET) was presented to the expert. The expert was asked to describe which features are in favor of ET. He gave two features as an explanation: the presence of kinetic tremor and the lack of bradykinesia. No counter examples were found.

Iteration 9 Iteration 9 also demanded very little time from the expert. He was presented with critical example E.36 (MT). The expert gave two features in favor of ET: the presence of kinetic tremor and the presence of postural tremor. No counter examples were found to the expert’s arguments.

Iteration 10 The expert observed critical example E.21 (MT). The expert gave the following explanation in favor of ET: “Qualitative assessment of both free-hand and template-based spiral in some of the observations (attributes) are essential, and none of them is Parkinsonian.” The knowledge engineers thus attached the following argument to the critical example: “SPIRO.FREE.ET.ONLY is *true* and SPIRO.TEMPLATE.ET.ONLY is *true*.” No counter examples and no further critical examples were found.

Review of the model At the end of the iterative procedure, the expert is asked to review the final model. Upon examination of the rules, the expert noticed the following rule that was in contradiction with his general knowledge about the domain.

IF TREMOR.START $>$ 61 THEN class = EMT; [16,0]

The newly added attribute TREMOR.START occurred in this counter-intuitive rule (according to the expert) and now became the subject of careful examination.

The expert realized that the values of the attribute TREMOR.PERIOD from which TREMOR.START was calculated, may indeed reflect the number of years since the tremor was *diagnosed*, but this attribute actually does not reflect the age when the tremor actually *started*. The reason for this was pointed out by the expert: the patients with ET tend to visit the neurologist only when the tremor starts to cause them problems in everyday life, and this is usually several years *after* it actually first occurred. While it is commonly accepted that ET typically occurs much earlier than PT, the attribute TREMOR.START simply cannot reflect the time of its occurrence. The expert and the knowledge engineer decided to exclude the attribute TREMOR.START from the domain.

As a consequence, another critical example emerged. The expert was now asked to improve the argument "POSTURAL is *true* and KINETIC is *true*," given to critical example E.55 in Iteration 3, again having E.63 as a counter example. He realized that there is another important difference between E.63 and E.55: the presence of resting tremor in E.63. He extended his argument to "POSTURAL is *true* and KINETIC is *true* and RESTING is *false*." The method found no counter examples to the expert's argument.

The expert revised the newly induced rules and found them all to be acceptable. No further critical examples were found, and the knowledge elicitation process concluded.

5 Results

The number of rules after each iteration varied from 12 to 15. Table 1 shows the final model, *i.e.*, rules obtained after the end of the knowledge elicitation process. The domain expert evaluated each rule according to the following criteria.

Counter-intuitive: an illogical rule that is in contradiction with expert knowledge.
Reasonable: a rule consistent with expert knowledge, but insufficient to decide in favor of ET or PT on its basis alone.
Adequate: a rule consistent with expert knowledge, ready to be used as a strong argument in favor of ET or PT.

Figure 1 demonstrates how the model became increasingly more consistent with the expert's knowledge during the knowledge elicitation process. The expert's evaluations of the initial and final rules are significantly different (p=0.026 using Mann-Whitney-Wilcoxon non-parametric test). All rules in the final model are consistent with the domain knowledge, and five of them were marked by the expert as sufficiently meaningful to determine the type of tremor by themselves. The other nine rules (marked as *reasonable* by the expert) could alone not be improved by the method to *adequate* rules, as there were no counter examples in the data set, and therefore the method did not have any reason to further specialize these rules. A larger number of learning examples might, however, result in a greater number of *adequate* rules.

During the process of knowledge elicitation, 15 arguments were given by the expert, 14 new attributes were included into the domain, and 21 attributes were excluded from the domain. After each iteration, the obtained model was evaluated on the test data set. If all the rules that triggered were for the class EMT (PMT), then the example was classified as ET (PT). In cases where the rules for both classes triggered, the example was classified as MT.

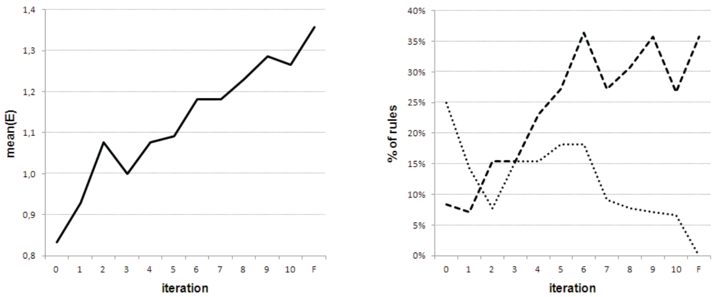

Fig. 1: The graph on the left side shows the average of expert's evaluations of the rules (0 - counter-intuitive, 1 - reasonable, 2 - adequate) obtained after each iteration of the knowledge elicitation process. The graph on the right side shows the percentage of counter-intuitive (the lower curve) and adequate (the upper curve) rules among all rules obtained after each iteration.

We compared classification accuracies improvements for ABCN2, Naive Bayes (NB), and kNN.[3] The ABCN2's classification accuracy on the test set improved from the initial 52% (NB: 63%; kNN: 58%) to the final 82% (NB: 81%; kNN: 74%). This result shows that the higher consistency with expert's knowledge is *not* obtained at the expense of classification accuracy.

The overall expert's time involvement was about 20 hours, which is rather low considering the high complexity of the presented domain. The relevance of the critical and counter examples shown to the expert is also reflected by the fact that they assisted the expert to spot occasional mistakes in the data.

6 Conclusions

We described a complete knowledge elicitation process with Argument Based Machine Learning (ABML) for a neurological decision support system. A neurological domain, namely differentiating between essential, Parkinsonian, and mixed tremor, served as a case study to demonstrate the following benefits of ABML for knowledge elicitation.

1. It is easier for domain experts to articulate knowledge; the expert only needs to explain a single example at the time.
2. It enables the expert to provide only relevant knowledge by giving him or her critical examples.
3. It helps the expert to detect deficiencies in his or her explanations by providing counter examples.

Our step-by-step presentation of the knowledge elicitation process with ABML can serve the reader as a guideline on how to effectively use the argument-based approach to knowledge elicitation. In addition, the main result of this paper is a quantitative demonstration of how the rules become increasingly more consistent with expert's knowledge during the ABML knowledge elicitation process. It is also interesting to note that ABML loop resulted in a simplification of the original set of attributes.

[3] In Naive Bayes, the conditional probabilities were estimated by relative frequencies for discrete attributes and by LOESS for continuous attributes. In kNN, the Euclidian distance was selected and k was set to 5.

Table 1: The rules after the end of the knowledge elicitation process. The condition and class columns show the condition and the consequent parts of a rule. Columns $+$ and $-$ stand for the number of positive and negative examples covered, respectively. All rules have pure distributions – they do not cover any examples from the opposite class. Column E stands for the expert's evaluation (0 - counter-intuitive, 1 - reasonable, 2 - adequate) of a rule.

#	Condition	Class	$+$	$-$	E
1	IF QUALITATIVE.ASSESSMENT = *ET*	EMT	21	0	1
2	IF BRADYKINESIA = *false*	EMT	18	0	1
3	IF BRADYKINESIA = *true* AND RIGIDITY = *true*	PMT	17	0	2
4	IF QUALITATIVE.ASSESSMENT = *ET* AND POSTURAL = *true*	EMT	16	0	1
5	IF RIGIDITY = *false* AND KINETIC = *true*	EMT	15	0	1
6	IF KINETIC = *true* AND BRADYKINESIA = *false*	EMT	13	0	1
7	IF SPIRO.FREE.PT.ONLY = *true* AND SPIRO.TEMPLATE.ET.ONLY = *false*	PMT	13	0	1
8	IF HARMONICS = *true*	EMT	12	0	2
9	IF RESTING = *true* AND HARMONICS = *false* AND RIGIDITY = *true*	PMT	12	0	2
10	IF POSTURAL = *true* AND KINETIC = *true* AND RESTING = *false*	EMT	10	0	1
11	IF QUALITATIVE.ASSESSMENT = *PT*	PMT	10	0	1
12	IF RESTING = *false* AND POSTURAL = *true* AND BRADYKINESIA = *false*	EMT	8	0	2
13	IF POSTURAL = *true* AND ANAMNESIS = *positive* AND BRADYKINESIA = *false*	EMT	8	0	2
14	IF SPIRO.FREE.ET.ONLY = *true* AND SPIRO.TEMPLATE.ET.ONLY = *true*	EMT	7	0	1

References

1. Forsyth, R., Rada, R.: Machine learning: applications in expert systems and information retrieval. Halsted Press, New York, NY, USA (1986)
2. Langley, P., Simon, H.A.: Applications of machine learning and rule induction. Commun. ACM **38**(11) (1995) 54–64
3. Martens, D., Baesens, B., Gestel, T.V., Vanthienen, J.: Comprehensible credit scoring models using rule extraction from support vector machines. European Journal of Operational Research **183**(3) (2007) 1466 – 1476
4. Verbeke, W., Martens, D., Mues, C., Baesens, B.: Building comprehensible customer churn prediction models with advanced rule induction techniques. Expert Systems with Applications **38**(3) (2011) 2354 – 2364
5. Webb, G.I., Wells, J., Zheng, Z.: An experimental evaluation of integrating machine learning with knowledge acquisition. Mach. Learn. **35**(1) (1999) 5–23
6. Možina, M., Žabkar, J., Bratko, I.: Argument based machine learning. Artificial Intelligence **171**(10/15) (2007) 922–937
7. Možina, M., Guid, M., Krivec, J., Sadikov, A., Bratko, I.: Fighting knowledge acquisition bottleneck with argument based machine learning. In: The 18th European Conference on Artificial Intelligence (ECAI). (2008) 234–238
8. Groznik, V., Guid, M., Sadikov, A., Možina, M., Georgiev, D., Kragelj, V., Ribarič, S., Pirtošek, Z., Bratko, I.: Elicitation of neurological knowledge with ABML. In Peleg, M., Lavrač, N., Combi, C., eds.: Artificial Intelligence in Medicine. Volume 6747 of Lecture Notes in Computer Science., Springer Berlin / Heidelberg (2011) 14–23
9. Pahwa, R., Lyons, K.E.: Essential tremor: Differential diagnosis and current therapy. American Journal of Medicine **115** (2003) 134–142