Visualization of Decision Rules – from the Cardiologist's Point of View

A Wlodyka[1], R Mlynarski[1], G Ilczuk[2], E Pilat[1], W Kargul[1]

[1]Upper Silesian Cardiology Centre, Katowice, Poland
[2]University of Silesia, Sosnowiec, Poland

Abstract

A lot of decision systems work internally using different forms of decision rules. In our experiments on large medical datasets, we found that when the number of conditions in a decision rule increases and the overall number of rules is greater than 20-50, it is really difficult to analyze and manage the stored knowledge. Our research concentrated on two methods of the visualization of decision rules: decision trees (AQDT-2 algorithm) and the so-called rule-diagrams, which present conditional parts of decision rules in a 3D matrix (2D layers are stacked in 3D cube). Experts agreed that decision trees present an attractive possibility of data visualization for small sets of rules (up to 100). They are intuitively understandable and with our own extensions, they provide a quick method for checking several what-if scenarios. For larger sets of rules (more then 100) rule-diagrams are definitely the better methods for analyzing patterns within data.

1. Introduction

A lot of decision systems work internally using different forms of decision rules [1,2]. In our experiments on large medical datasets, we found that when the number of conditions in a decision rule increases and the overall number of rules is greater than 20-50, it is really difficult to analyze and manage the stored knowledge. In this situation, methods that allow the visualization of the induced decision rules are of great importance.

Decision trees (Fig. 1) are an effective tool for describing a decision process but they also show some limitations if their structure must be adapted for new requirements [3]. This limitation is attributable to the fact, that decision structure (tree) stores information in form of procedural representation, which imposes an evaluation order of tests. In contrary declarative representation of knowledge such as decision rules can be evaluated in any order, so that it is possible to generate a large number of logically equivalent decision trees which differ in test ordering. This way decision rules may be easily modified and adapted for specified requirements and at the and this declarative knowledge representation may be transformed into procedural one (decision trees). In this paper we describe our transformation results achieved with an extended version of AQDT-2 method based on idea presented by Michalski [4].

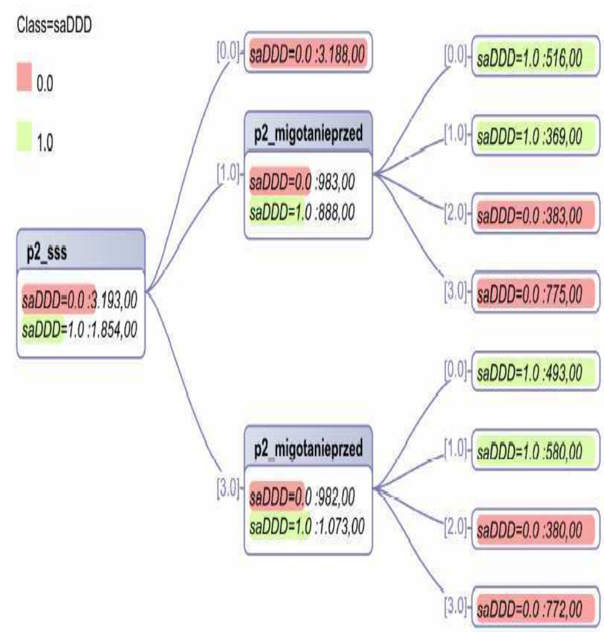

Figure 1. Example of a decision tree generated for DDD pacing depending on different forms of atrial fibrillation systems.
p2_migotanieprzed [PL]= atrial fibrillation - AF [ENG]

Alternative for decision trees are so-called rule-diagrams (Fig. 2) which present conditional parts of decision rules in a 3D matrix (2D layers are stacked in 3D cube). It is very advanced method of visualization very seldom used in medicine [3].

The question which of these methods is better from a cardiological point of view is still valid.

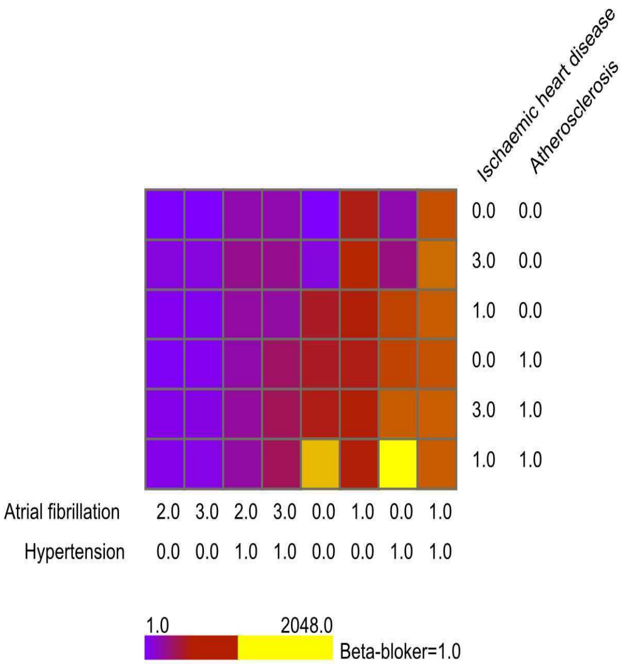

Figure 2. An example of rule-diagram generated by our system (1 layer, 4 attributes, decision Beta-blocker)

1.1. Aim of the study

The aim of the study was to evaluate which of the two earlier presented methods of visualization and presenting of decision rules by Rough Set algorithms is better from cardiological point of view.

2. Methods

Sets of decision rules from 5425 medical records were generated (our implementation of rough set based MLEM2) for 3 decision attributes:

1. decision about pacemaker implantation (15 rules),
2. decision about implantation of DDD pacemaker (21 rules)
3. decision about B-blocker treatment (70 rules).

These rules were generated and afterwards visualized using both decision trees and rule-diagrams. Our research concentrated on two methods of the visualization of decision rules: decision trees (AQDT-2 algorithm) and the so-called rule-diagrams, which present conditional parts of decision rules in a 3D matrix (2D layers are stacked in 3D cube).

The results were validated by three experienced cardiologists and one allied professional using the following two criteria: ease of understanding, speed of analysis and interactive reasoning. All data were cumulate in the relative database. Finally statistical analysis was performed using Statistica 6 (StatSoft Inc, USA, Polish version).

3. Results

Summary result for decision trees are presented on the figure 3, and for rule-diagrams on the figure 4.

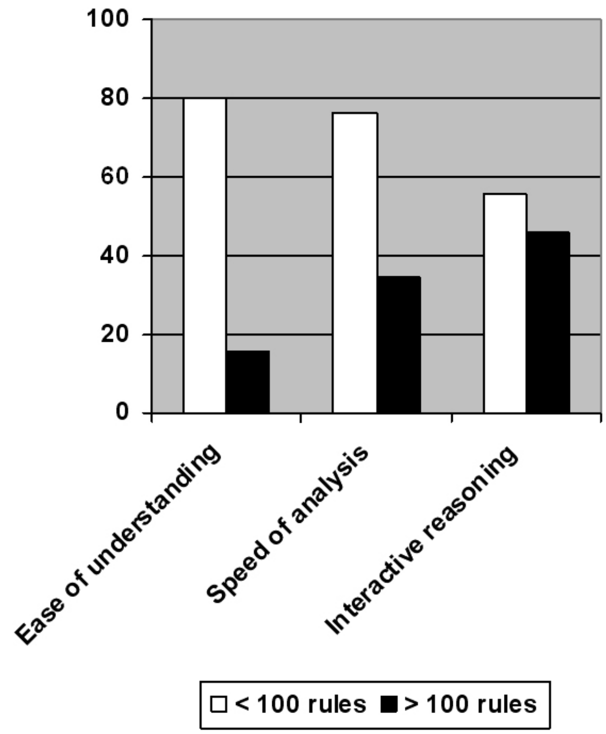

Figure 3. Average preferences in % for a decision tree with division into < or > 100 generated rules.

Experts agreed that decision trees present an attractive possibility of data visualization for small sets of rules (up to 100). They are intuitively understandable and with our own extensions, they provide a quick method for checking several what-if scenarios. For larger sets of rules (more then 100) rule-diagrams are definitely the better methods for analyzing patterns within data. The visualization of rules containing attributes with a large number of possible distinct values is an especially strong advantage of rule-diagrams. Nevertheless they require a learning process at the beginning.

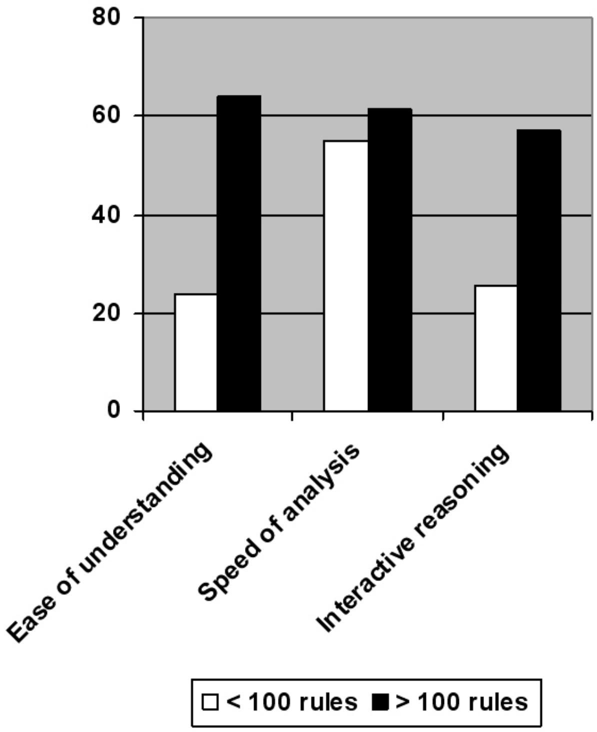

Figure 4. Average preferences in % for a rule-diagram with division into < or > 100 generated rules.

4. Discussion and conclusions

In our research we focus on decision systems based on the Rough Set developed by Pawlak and presented in 1982 theory is a mathematical approach to handle imprecision and uncertainty [5,6]. The main advantage of Rough Set theory in medical domain is a possibility of data analysis without any need of preliminary or additional information about analysed data. Main area of interest in our research is a complete vertical solution which is able to extract knowledge in form of decision rules from raw, medical data [7].

Our system was presented many times including Computers in Cardiology in Lyon, Valencia and Durham [8-11]. We achieved acceptable for clinical area accuracy, but still we had problems with presenting of data for cardiology practitioners. It seems that this is the most important factor in entrance these kind of methods into the clinical practise.

Two different methods of visualization were used in this research. None of them was preferred by doctors, however medical experts accordantly confirmed usefulness visualization in clinical practise.

Decision trees are an effective tool for describing a decision process but they also show some limitations if their structure must be adapted for new requirements. This limitation is attributable to the fact, that decision structure (tree) stores information in form of procedural representation, which imposes an evaluation order of tests. In contrary declarative representation of knowledge such as decision rules can be evaluated in any order, so that it is possible to generate a large number of logically equivalent decision trees which differ in test ordering. This way decision rules may be easily modified and adapted for specified requirements and at the end this declarative knowledge representation may be transformed into procedural one (decision trees). This step is realized in our research by an extended version of the original, presented by Michalski in [4], AQDT-2. For rendering a generated decision tree we implemented an algorithm which mostly follows the aesthetics defined in [12,13] for keeping a displayed tree as tight as possible without compromising its readability. This implementation in contrary to Bloesch uses a non-recursive algorithm which avoids stack overflows during a processing of large tree models [14].

Presenting decision rules in form of decision trees increases human understanding for generated decision rules. But in case of complex decision problems a number of generated decision rules rapidly increases what makes an analysis of a large decision tree. In such cases the other proposed form of knowledge presentation can be attractive – diagrams. Rule-diagrams are generated following this principle: each cell at X any Y axis represents a predefined combination of AV (attribute-value) pairs. If assigned to a single cell AV pairs match with AV pairs of a decision rule then a value of this cell will be changed. The change of the cell's value correlates with one or many of rule's properties (eg. strength, support or importance). This simple algorithm allows a fast but yet powerful data analysis of large sets of decision rules or even datasets. Our implementation of the described method supports several traditional extensions such as multiply layers translucent layers (separate layer pro decision class) or row/columns sorting based on a defined criteria.

Very important in this paper was tests which was performed by experienced cardiologists. They agree that methods of visualization significantly simplified understanding of rules and the chose of method strictly depend on numbers of rules which should be visualized. Very precious added value seems to be also incorporation to the research allied professional. Extension of this system into this medical group can be very important in introduction presented methods to the clinical practise [15]. In our opinion it was success. Allied professional (Registered Nurse during PhD stipend) in some rules even faster (if compare to the cardiologists) understand knowledge stored in decision rules.

4.1. Conclusions

In the case of mid-range and large medical databases, rule-diagrams seems to be more suitable for the task, whereas, in the case of small sets of rules when compared to the larger sets, decision trees were shown to have more advantages.

Acknowledgements

We would like to thank Iwona Grzegorczyk MS from the statistical department of the Upper Silesia Medical Center for her assistance in exporting the databases.

References

[1] Stefanowski J. On rough set based approaches to induction of decision rules. In: S.A. Polkowski L, editor, Rough Sets in Knowledge Discovery 1. Heidelberg, Germany: Physica-Verlag,1998,500-529.

[2] Ilczuk G, Wakulicz-Deja A. Attribute Selection and Rule Generation Techniques for Medical Diagnosis Systems. Lecture Notes in Computer Science 2005;3642:352-361.

[3] Ilczuk G, Wakulicz-Deja A. Visualization of Rough Set Decision Rules for Medical Diagnosis Systems. Lecture Notes in Computer Science 2007;4482:371-378.

[4] Michalski RS, Ibrahim FI. Learning Problem-Oriented Decision Structures from Decision Rule: The AQDT-2 System. Lecture Notes in Computer Science 1994;869: 416-426.

[5] Pawlak Z. Rough sets. International Journal of Computer and Information Sciences 1982;11:341–356.

[6] Pawlak Z, Skowron A. A Rough Set approach to decision rules generation. Inst. Comput. Sci. Research Report 1993; 23.

[7] Pawlak Z. Rough Sets: Theoretical aspects of reasoning about data. Boston, USA: Kluwer Academic Publishers 1991.

[8] Ilczuk G, Mlynarski R, Kargul W, Wakulicz-Deja A. New feature selection methods for qualification of the patients for cardiac pacemaker implantation. Computers in Cardiology 2007;34:423-426

[9] Mlynarski R, Ilczuk G, Wakulicz-Deja A, Kargul W. A New Method of Data Preparation for Cardiological Decision Support. Computers in Cardiology 2006;33:273-276.

[10] Mlynarski R, Ilczuk G, Pilat E, Wakulicz-Deja A, Kargul W. Automated Decision Support and Guideline Verification in Clinical Practice. Computers in Cardiology 2005;32:375-378

[11] Ilczuk G, Mlynarski R, Wakulicz-Deja A, Drzewiecka A, Kargul W. Rough Set Techniques for Medical Diagnosis Systems. Computers in Cardiology 2005;32:837-840

[12] John GH, Kohavi R, Pfleger K. Irrelevant features and the subset selection problem. In: ICML. 1994:121-129

[13] Viktor HL. Combining humans and machines into a cooperative multi-agent learning system. International Journal on Continuous Engineering and Life-Long Learning 12 (2002) 288-298

[14] Bloesch A. Aestetic Layout of Generalized Trees. Software-Practice and Experience 1993;23:817-827.

[15] Budihardjo A, Grzymała-Busse J, Woolery L. Program LERS-LB 2.5 as a tool for knowledge acquisition in nursing. In: Proceedings of the Fourth International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems. Kaloa, Kauai, Hawaii;1991:735-740.

Address for correspondence

Cardiological section
(Allied Professional)
Agnieszka Wlodyka
Oddział Elektrokardiologii
ul. Ziolowa 45/47
Katowice 40-635, Poland
email:
agnieszkawlodyka@o2.pl

Computer section
Grzegorz Ilczuk
Heuweg 12A
91334 Hemhofen, Germany
email:
Grzegorz.Ilczuk@Ilczuk.com

Cardiological section
Rafal Mlynarski
Oddział Elektrokardiologii
ul. Ziolowa 45/47
Katowice 40-635, Poland
email:
Rafal_Mlynarski@mp.pl