

# Action Rules Approach to Solving Diagnostic Problems in Clinical Medicine

Hanna A. Wasyluk<sup>1</sup> and Zbigniew W. Raś<sup>2,3</sup>

<sup>1</sup>*The Medical Center of Postgraduate Education, Warszawa,*

<sup>2</sup>*University of North Carolina, Dept. of Comp. Science, Charlotte, N.C. 28223,*

<sup>3</sup>*Polish-Japanese Institute of Information Technology, Warszawa,*

<sup>1,3</sup>*Poland*

<sup>2</sup>*USA*

## 1. Introduction

The beginning of the twenty first century have brought considerable advances in the field of computer-based medical systems. It resulted from noticeable improvements in medical care, from ease of storage and access of digital imaging through gathering of computerized medical data to accessing on-line literature, patient monitoring, therapy planning and computer support for medical diagnosis. Similarly to other domains, decision-support systems have proven to be valuable tools that help practitioners in facing challenging medical problems, such as diagnosis and therapy. Given that stakes happen to be extremely high, support of decision-making plays a particularly important role in the field of medicine. As an example, in the domain of hepatology, inexperienced clinicians have been found to make a correct diagnosis in jaundiced patients in less than 45% of cases [1,2]. Incorrect diagnosis leads to suboptimal treatment that means waste of resources, time and sometimes of human life. Automating those processes that can be captured by focus on important signs and symptoms, eliminates human error and leads to overall improvements in the quality of medical care.

## 2. Liver disorders as particularly recommended area for computer-aided diagnostics

The domain of liver diseases is especially suited to application of computer-based methods and there are several reasons for high expectations from computer-aided diagnosis. Firstly, the number of cases of liver disorders is on the rise. In Poland, the number of new acute and chronic hepatitis cases is roughly half a million per year. Secondly, correct diagnosis, especially in early stages of disease, is difficult. There is variety of diseases that manifests with similar symptoms. Finally, early diagnosis is critical, as in some cases damage to the liver caused by an untreated disorder may be irreversible.

Typically, a patient suffering from symptoms suggesting of abdominal disorders seeks help of a family doctor. Primary care physicians face the daunting task of determining the source of discomfort based on patient-reported data and physical examination, possibly enhanced

Source: Decision Support Systems, Advances in, Book edited by: Ger Devlin,  
ISBN 978-953-307-069-8, pp. 342, March 2010, INTECH, Croatia, downloaded from SCIYO.COM

with the results of basic medical tests. Correct diagnosis under these circumstances is difficult and accuracy can be low. Based on our observations, we estimate that at this stage only 40-60% of the cases are diagnosed correctly.

This rather low diagnostic performance is caused by several etiophysiological and organizational factors. These include the nature of the liver, e.g. its high productive reserves that often make the abnormalities noticeable only when the disease reaches an advanced stage. Growth of liver diseases is often slow and unspecific and the symptoms may be hardly noticeable. Liver disorders are influenced by environmental factors, such as alcohol intake, medications and dietary habits. There is still insufficient knowledge of immunological factors that cause certain pathologies of liver. Undiagnosed or misdiagnosed viral hepatitis often leads to irreversible defects of the liver that may have bearing on the manifestation of the possible later disorders. Reliable diagnosis of a liver disorder can be often established only based on the results of liver biopsy. However, biopsy is an examination that is performed only in specialized clinics and may not be available to primary care physician (1).

### 3. The HEPAR computer-aided diagnosis system

The HEPAR system was designed and built in collaboration between the Institute of Biocybernetics and Biomedical Engineering of the Polish Academy of Sciences and Medical Center of Postgraduate Education in Warsaw. The system, operating in the Gastroenterological Clinic of the Institute of Food and Feeding in Warsaw, allows for systematic collection and processing of clinical data of hepatological patients diagnosed and treated in the clinic. The system is equipped with a diagnostic module built around several statistical methods that have been found useful in diagnosis. An integral part of the HEPAR is database, created in 1990 and thoroughly maintained since then. The current database contains roughly 860 patients records - with the ultimate diagnosis verified by means of biopsy, laparoscopy and often longitudinal follow-up. Each hepatological case is described by over 150 different medical findings, such as patient self-reported data, results of physical examination, laboratory tests and finally a histopathologically verified diagnosis. The system includes the module supporting the physicians in making a diagnosis.

Computer-based systems accompanying a medical diagnosis can be based on variety of principles. Generally, diagnosis supporting rules used in such systems may reflect knowledge provided by medical doctors (experts), textbooks and regularities extracted from clinical databases. In accordance with this, we distinguish two kinds of systems: knowledge based systems and database systems (3).

HEPAR could be treated as a database system because decision rules used there relay mainly on information extracted from the hepatological database.

The patients from this database have been classified by clinicians into several adequate liver diseases. The support of diagnosis in HEPAR system is based on a comparison of a new patient case with similar cases from its database.

Graphical representation of the data on a plane in a form of *diagnostic* maps is particularly important in the HEPAR. The diagnostic maps represent the data on a plane in such a way that differences between selected diseases are enhanced. The diagnostic maps used in the system result from linear transformations of multivariate data sets into two dimensions. These transformations are obtained through minimization of the perceptron criterion functions, that have links to early research on neural networks [3].

Medical doctors willingly accept data mapping because it gives them an opportunity for a better understanding of a diagnostic problem.

The nearest neighbors techniques originated as one of the pattern recognition methods. They have many practical applications related to a classification problem. Generally, the nearest neighbors classification or diagnostic rules are based on a comparison of a new object with the most similar objects that have been previously classified. The cases similar to presently diagnosing patients are found in the system either directly in the database or through their representation on *diagnostic map*.

This paper - presenting a new strategy of constructing action rules (called ARAS) and its application to the HEPAR database - is implying that such an approach might improve medical reasoning associated with the computer-aided diagnosis of patients with liver diseases.

#### 4. Material and methods

Finding useful rules is an important task of a knowledge discovery process. Most researchers focus on techniques for generating patterns from a data set such as classification rules, association rules...etc. They assume that it is user's responsibility to analyze these patterns in order to infer solutions for specific problems within a given domain. The classical knowledge discovery algorithms have the potential to identify enormous number of significant patterns from data. Therefore, people are overwhelmed by a large number of uninteresting patterns and it is very difficult for a human being to analyze them in order to form timely solutions. Therefore, a significant need exists for a new generation of techniques and tools with the ability to assist users in analyzing a large number of rules for a useful knowledge [4].

Action rules can be seen as logical terms describing knowledge about possible actions associated with objects which is hidden in a database. Classical strategy for discovering action rules requires prior extraction of classification rules which next are evaluated pair by pair with a goal to build a strategy of action based on condition features in order to get a desired effect on a decision feature. An actionable strategy is represented as a term  $r = [\omega \wedge (\alpha \rightarrow \beta)] \rightarrow (\phi \rightarrow \psi)$ , where  $\omega$ ,  $\alpha$ ,  $\beta$ ,  $\phi$ , and  $\psi$  are descriptions of objects, in our case seen as patients. The term  $r$  states that when a fixed condition  $\omega$  is satisfied and the changeable behavior  $(\alpha \rightarrow \beta)$  occurs in patients registered in a database so does the expectation  $(\phi \rightarrow \psi)$ . This paper proposes a new strategy of agglomerative type, called ARAS, for constructing action rules directly from single classification rules.

The ARAS strategy has been implemented as one of the modules in HEPAR Clinical Decision Support System which consists of a clinical database and a set of tests and procedures, facilitating decision-making process for patients with liver disorders.

#### 5. Constructing action rules

Action rules, introduced in [5] and investigated further in [6], [7], [8] are constructed from certain pairs of classification rules. Interventions, defined in [8], are conceptually very similar to action rules [10], [11].

The previous process of constructing action rules from pairs of classification rules is not only unnecessarily expensive but also gives too much freedom in constructing their classification parts. In [8] it was shown that action rules do not have to be built from pairs of classification

rules and that single classification rules are sufficient to achieve the same goal. However, this paper only proposed a theoretical lattice-theory type framework without giving any detailed algorithm for action rules construction. In this paper we propose a very simple LERS-type algorithm for constructing action rules from a single classification rule. LERS is a classical example of a bottom-up strategy which constructs rules with a conditional part of the length  $k+1$  after all rules with a conditional part of the length  $k$  have been constructed. Relations representing rules produced by LERS are marked. System ARAS assumes that LERS is used to extract classification rules. This way system ARAS instead of verifying the validity of certain relations only has to check if these relations are marked by LERS. The same, if we use LERS as a pre-processing module for ARAS, then the overall complexity of the algorithm is decreased [12], [13].

System ARAS differs from the tree-based strategies for action rules discovery (for instance from DEAR [6]) because clusters generated by its second module are formed around target classification rules. An action rule can be constructed in ARAS from two classification rules only if both of them belong to the same cluster and one of them is a target classification rule. So, the complexity of the second module of ARAS is  $O(k \cdot n)$ , where  $n$  is the number of classification rules extracted by LERS and  $k$  is the number of clusters. The time complexity of the second module of DEAR is equal to  $O(n \cdot n)$ , where  $n$  is the same as in ARAS. The first module of ARAS is the same as the first module of DEAR, so their complexities are the same.

## 6. Application of action rules to HEPAR database

HEPAR database contains information about 758 patients described by 106 attributes (including 31 laboratory tests with values discretized to: below normal, normal, above normal). It has 14 stable attributes. Two laboratory tests are invasive: HBsAg [in tissue] and HBeAg [in tissue]. The decision attribute has 7 values: I (acute hepatitis), IIa (subacute hepatitis [types B and C]), IIb (subacute hepatitis [alcohol-abuse]), IIIa (chronic hepatitis [curable]), IIIb (chronic hepatitis [non-curable]), IV (cirrhosis-hepatitis), V (liver-cancer).

The diagnosis of liver disease depends on a combination of patient's history, physical examinations, laboratory tests, radiological tests, and frequently a liver biopsy. Blood tests play an important role in the diagnosis of liver diseases. However, their results should be analyzed along with the patient's history and physical examination. The most common radiological examinations used in the assessment of liver diseases are ultrasound and sonography. Ultrasound is a good test for the detection of liver masses, assessment of bile ducts, and detection of gallstones presence. However, it does not detect the degree of inflammation or fibrosis of the liver. Ultrasound is a noninvasive procedure and there are no risks associated with it. Liver biopsy enables the doctor to examine how much inflammation and how much scarring has occurred. Liver biopsy is an example of invasive procedure that carries certain risks to the patient. Therefore, despite of the importance of its results to the diagnosis, clinicians try to avoid biopsy as often as possible. However, liver biopsy is often the only way to establish correct diagnosis in patients with chronic liver disorders [14].

A medical treatment is naturally associated with reclassification of patients from one decision class into another one. In our research we are mainly interested in the reclassification of patients from the class IIb into class I and from the class IIIa into class I but without referring to any invasive tests results in action rules.

Database HEPAR has many missing values so we decided to remove all its attributes with more than 90% of null values under the assumption that these attributes are not related to

invasive tests. Also, subjective attributes (like history of alcohol abuse) and basic medical tests have been removed. Finally, we used classical null value imputation techniques to make the resulting database complete [15].

The next step of our strategy is to apply RSES software to find d-reducts [16].

The set  $R = \{m, n, q, u, y, aa, ah, ai, am, an, aw, bb, bg, bm, by, cj, cm\}$  is one of them and it does not contain invasive tests. The description of its values is given below:

|  |                                |
|--|--------------------------------|
| m - Bleeding                             | an - Jaundice in pregnancy     |
| n - Subjaundice symptoms                 | aw - Erythematous dermatitis   |
| q - Eructation                           | bb - Cysts                     |
| u - Obstruction                          | bg - Sharp liver edge (stable) |
| y - Weight loss                          | bm - Blood cell plaque         |
| aa - Smoking                             | by - Alkaline phosphatase      |
| ah - History of viral hepatitis (stable) | cj - Prothrombin index         |
| ai - Surgeries in the past (stable)      | cm - Total cholesterol         |
| am - History of hospitalization (stable) | dd -Decision attribute         |

## 7. Results and conclusion

Two action rules from among over eighty, discovered from the HEPAR database which was reduced to d-reduct R in this investigation, are given below:

I.

$$[(ah = 1) \wedge (ai = 2) \wedge (am = 2) \wedge (bg = 1)] \wedge [(cm = 1) \wedge (aw = 1) \wedge (u; \rightarrow 1) \wedge (bb = 1) \wedge (aa = 1) \wedge (q; \rightarrow 1) \wedge (m; \rightarrow 1) \wedge (n = 1) \wedge (bm; \rightarrow \text{down}) \wedge (y = 1) \wedge (by; \text{norm} \rightarrow \text{up})] \Rightarrow (dd; IIIa \rightarrow I)$$

II.

$$[(ah = 1) \wedge (ai = 2) \wedge (am = 2) \wedge (bg = 1)] \wedge [(cm = 1) \wedge (aw = 1) \wedge (u; \rightarrow 1) \wedge (bb = 1) \wedge (aa = 1) \wedge (q; \rightarrow 1) \wedge (m; \rightarrow 1) \wedge (n = 1) \wedge (bm; \rightarrow \text{down}) \wedge (y = 1) \wedge (by; \text{norm} \rightarrow \text{down})] \Rightarrow (dd; IIIa \rightarrow I)$$

Both presented above action rules are applicable to patients, who previously underwent surgery, but did not have viral hepatitis. By excluding certain symptoms, such as obstruction, bleeding etc., physicians might be able to re-classify the patient from class III A (advanced stadium of hepatopathy) to class I (initial phase of the disease). Under some circumstances, it might be considered as a substantial gain in problem-solving reasoning in the diagnostic/patients management process in liver diseases [17].

## 8. Perspectives for the future

In this chapter a new strategy of constructing action rules has been presented; it may help a medical doctor to state the final diagnosis easier using advice from the clinical decision support system. We tested this new strategy, called ARAS, on the medical database HEPAR

and the results confirmed its clinical usefulness. What would be the next step in this promising strategy?

Generally speaking, action rules can be seen as logical expressions describing knowledge about actions associated with objects of interest. They are hidden in a database and a researcher's role is to discover them and use them to enhance a decision support system. Based on construction methods, action rules discovery can be divided into two types: rule-based and object-based approach.

In rule-based approach, actionable patterns [6], [7], [20], [21] are built on the foundations of pre-existing rules. It consists of two main steps: (1) in the first step, a standard learning method is used to detect interesting patterns in the form of classification rules, association rules, or clusters and (2) the second step is to use an automatic or semi-automatic strategy to inspect such results and derive possible action strategies. These strategies provide an insight of how values of some attributes need to be changed - so the desirable objects can be shifted to a desirable group. However, the standard data mining methods such as *LERS* [12] or *ERID* [10] (which is used for incomplete data sets) extract only the shortest or close to the shortest rules. Therefore, by following this approach, some meaningful action patterns can be easily missed.

In the object-based approach, action rules are discovered directly from a database [8], [18], [19] and the same their number is usually much larger than the number of rules discovered by following a rule-based approach. System ARAS is an exception - it follows rule-based approach, but because it does not require to construct pairs of rules in its discovery process, the number of action rules it discovers is comparable to numbers obtained when we follow object-based approaches.

Meta-actions [22] are defined as actions which trigger changes of flexible attributes either directly or indirectly because of correlations among certain attributes in the system. Links between meta-actions and changes they trigger within the values of flexible attributes can be defined by a mapping linking meta-actions with changes of attribute values used to describe objects in the decision system. In medical domain, taking a drug is a classical example of a meta-action. For instance, Lamivudine is used for treatment of chronic hepatitis B. It improves the seroconversion of e-antigen positive hepatitis B and also improves histology staging of the liver, but at the same time it can cause a number of other symptoms. This is why doctors have to order certain lab tests to check patient's response to that drug. We are planning to add the meta-actions module to HEPAR decision support system in the near future.

## 9. References

- [1] Wasyluk H.: Second Evaluation of the Diagnostic Computer System HEPAR. Proceedings 6<sup>th</sup> World Multiconference of Systemics, Cybernetics and Informatics Orlando, USA, ed. N. Callaos, Vol. 2; 2002. p. 242-246.
- [2] Wasyluk H., Oniško A., Drużdżel M.: Support of diagnosis of liver disorders based on a causal Bayesian network model. *Med. Sci Monitor* No 7, suppl. 1; 2001. p. 327- 332.
- [3] Bobrowski L., Wasyluk H.: Diagnosis supporting rules of the HEPAR system, Proc. MEDINFO 2001, part 2, p. 1309-1313, IOS Press, Amsterdam.
- [4] Pawlak Z.: Information systems - theoretical foundations. In: *Information Systems Journal*; 1991. Vol. 6. p. 205-218.

- [5] Adomavicius G., Tuzhilin A.: Discovery of actionable patterns in databases: the action hierarchy approach. Proceedings of KDD097 Conference, Newport Beach, CA, AAAI Press; 1997.
- [6] Tsay LS, Raś Z.W.: Action rules discovery system DEAR, method and experiments, in Journal of Experimental and Theoretical Artificial Intelligence, Taylor & Francis, Vol. 17, No. 1-2; 2005. p. 119-128.
- [7] Tzacheva A., Raś Z.W.: Action rules mining, in: International Journal of Intelligent Systems, Wiley, Vol. 20, No. 7; 2005. p. 719-736.
- [8] Raś Z.W., Dardzińska A.: Action rules discovery - a new simplified strategy. In: Foundations of Intelligent Systems, Esposito F. et al. (Eds.), LNAI, No. 4203, Springer; 2006. p. 445-453.
- [9] Greco S., Matarazzo B., Pappalardo N., Słowiński R.: Measuring expected effects of interventions based on decision rules. In: Journal of Experimental and Theoretical Artificial Intelligence, Taylor Francis; 2005. Vol. 17. No. 1-2.
- [10] Dardzińska A., Raś Z.W.: "Extracting Rules from Incomplete Decision Systems: System ERID". In: Foundations and Novel Approaches in Data Mining, (Eds. T.Y. Lin, S. Ohsuga, C.J. Liau, X. Hu), Advances in Soft Computing, Vol. 9, Springer; 2006. p. 143-154.
- [11] Silberschatz A., Tuzhilin A.: On subjective measures of interestingness in knowledge discovery. Proceedings of KDD095 Conference, AAAI Press; 1995.
- [12] Grzymała-Busse J.: A new version of the rule induction system LERS, in: Fundamenta Informaticae Journal., Vol. 31. No. 1; 1997. p. 27-39.
- [13] Tsay LS., Raś Z.W.: Action rules discovery system DEAR3. In: Foundations of Intelligent Systems, LNAI, No. 4203, Springer; 2006. p. 483-492.
- [14] Oniśko A., Drużdżel M., Wasyluk H.: Extension of the HEPAR II model to multiple-disorder diagnosis. In: Intelligent Information Systems, Advances in Soft Computing, Springer; 2000. p. 303-313.
- [15] Raś Z.W., Wyrzykowska E., Wasyluk H.: ARAS: Action rules discovery based on agglomerative strategy, in: Mining Complex Data, Post-Proceedings of the ECML/PKDD'07 Third International Workshop, MCD 2007, LNAI, Springer (in press)
- [16] RSES - <http://logic.mimuw.edu.pl/rses/>
- [17] Wasyluk H., Raś Z.W., Wyrzykowska E.: Application of action rules to HEPAR clinical decision support system Exper&Clin. Hepatology 2008, 4(2), pp. 46-48.
- [18] Im S., Ras Z.W., Wasyluk H.: Action Rule Discovery From Incomplete Data, in: Knowledge and Information Systems Journal, Springer, 2009, (will appear)
- [19] Ras Z.W., Dardzinska A., Tsay L.S., Wasyluk H.: Association Action Rules, IEEE/ICDM Workshop on Mining Complex Data (MCD 2008), Pisa, Italy, ICDM Workshops Proceedings, IEEE Computer Society, 2008, 283-290.
- [20] Yang Q., Chen H.: Mining case for action recommendation, in Proceedings of ICDM 2002, IEEE Computer Society, 2002, 522-529.
- [21] Yang Q., Yin J., Lin C.X., Chen T.: Postprocessing decision trees to extract actionable knowledge, in Proceedings of ICDM 2003, IEEE Computer Society, 2003, 685-688.

- [22] Wang K., Jiang Y., Tuzhilin A.: Mining Actionable Patterns by Role Models, in Proceedings of the 22nd International Conference on Data Engineering, IEEE Computer Society, 2006, 16-25.



## **Decision Support Systems Advances in**

Edited by Ger Devlin

ISBN 978-953-307-069-8

Hard cover, 342 pages

**Publisher** InTech

**Published online** 01, March, 2010

**Published in print edition** March, 2010

This book by In-Tech publishing helps the reader understand the power of informed decision making by covering a broad range of DSS (Decision Support Systems) applications in the fields of medical, environmental, transport and business. The expertise of the chapter writers spans an equally extensive spectrum of researchers from around the globe including universities in Canada, Mexico, Brazil and the United States, to institutes and universities in Italy, Germany, Poland, France, United Kingdom, Romania, Turkey and Ireland to as far east as Malaysia and Singapore and as far north as Finland. Decision Support Systems are not a new technology but they have evolved and developed with the ever demanding necessity to analyse a large number of options for decision makers (DM) for specific situations, where there is an increasing level of uncertainty about the problem at hand and where there is a high impact relative to the correct decisions to be made. DSS's offer decision makers a more stable solution to solving the semi-structured and unstructured problem. This is exactly what the reader will see in this book.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Hanna A. Wasyluk and, Zbigniew W. Ras (2010). Action Rules Approach to Solving Diagnostic Problems in Clinical Medicine, Decision Support Systems Advances in, Ger Devlin (Ed.), ISBN: 978-953-307-069-8, InTech, Available from: <http://www.intechopen.com/books/decision-support-systems-advances-in/action-rules-approach-to-solving-diagnostic-problems-in-clinical-medicine>

# **INTECH**

open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821

© 2010 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](#), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.