Rough Set Approach to the Survival Analysis

Jan Bazan [1], Antoni Osmólski[2], Andrzej Skowron[3]
Dominik Ślęzak[4], Marcin Szczuka[3], Jakub Wróblewski[4]

Abstract
Application of rough set based tools to the post-surgery survival analysis is discussed. Decision problem is defined over data related to the head and neck cancer cases, for two types of medical surgeries. The task is to express the differences between expected results of these surgeries and to search for rules discerning different survival tendencies.

1. Introduction

Analysis of medical data requires decision models well understandable by medical experts. The theory of rough sets [3] seems to provide such models. A number of rough set based applications to medical domain is known from the literature [1].
We analyze data about medical treatment of patients with various kinds of the head and neck cancer cases. The data, collected for years by Medical Center of Postgraduate Education in Warsaw, consists of 557 patient records described by 29 attributes, reduced – after consultation with medical experts – to 7 columns. Except the dates, important conditional attributes are well-defined symbolic attributes. On the other hand, decision problems are defined over especially designed attributes, which are of a complex structure. It enables focusing in the foregoing analysis on the complex decision semantics, having a clear interpretation of the conditional part of decision rules.
The main topic of the paper is the following: Given information about the type of surgery applied to each particular patient, describe patient groups with different comparative statistics of survivals versus the surgery type chosen. In the full version of the paper we also consider statistical methods used in the medical survival analysis, like the Kaplan-Meier's product-limit estimate and the Cox's proportional hazard model [2], to find descriptions of groups with various survival estimates.

2. Rough Set Framework

In the rough set theory [3] sample of data takes the form of an information system $\boldsymbol{A}=(U,A)$, where each attribute $a\in A$ is a function $a:U\rightarrow V_a$ into the set of all possible values on a. Reasoning about data often requires a distinguished decision to be predicted under information over the rest of attributes. In this case, we consider decision table $\boldsymbol{A}=(U,A\cup\{d\})$, where $d\notin A$.
Definition: Let $\boldsymbol{A}=(U,A\cup\{d\})$ and $B\subseteq A$ be given. Consider arbitrary $a\in B$ and $v_a\in V_a$. We say that object $u\in U$ satisfies *descriptor* $a=v_a$ iff $a(u)=v_a$. Let us consider collection of such descriptors, one for each element of B, and decision descriptor $d=v_d$. We say that $\boldsymbol{A}$ satisfies *decision rule*

$$\bigwedge_{a\in B}(a = v_a) \Rightarrow (d = v_d)$$

iff each $u\in U$, which satisfies all descriptors $a=v_a$, $a\in B$, at the left side, also satisfies descriptor $d=v_d$ at the right side. We say that the above decision rule is *minimal (irreducible)* iff there is no $a\in B$ such that after removal of its descriptor from the left part the rule is still satisfied by $\boldsymbol{A}$.
In case of many real-life decision problems, in particular the one we are dealing with, there is an issue of data inconsistency, where construction of the above decision rules is difficult or impossible. In the rough set theory this problem is addressed, e.g., by introducing *B-generalized decision function* $\partial_{d/B}:U\rightarrow P(V_d)$ [3],[4] such that $\partial_{d/B}(u) = \{d(u'): u'\in[u]_B\}$, where

$$[u]_B = \{u'\in U: \forall_{a\in B}\,(a(u)=a(u'))\}$$

is *B-discernibility class* of u. Generalized decision functions correspond to inexact decision rules

[1] Institute of Mathematics, University of Rzeszów, Rejtana 16A, 35-959 Rzeszów, Poland
[2] Medical Center of Postgraduate Education, Marymoncka 99, 01-813 Warsaw, Poland
[3] Institute of Mathematics, Warsaw University, Banacha 2, 02-097 Warsaw, Poland
[4] Polish-Japanese Institute of Information Technology, Koszykowa 86, 02-008 Warsaw, Poland

$$\bigwedge a \in B \, (a = a(u)) \Rightarrow \bigvee v \in \partial_{d/B}(u) \, (d = v)$$

which should be optimized in terms of the tradeoff between the number of descriptors at their left and right sides. One can base such an optimization process onto discerning objects with different generalized decisions, by using descriptors defined over conditional attributes [4]. This is just an example of a more general methodology of constructing rules by discerning pairs of objects with different values of a specially designed complex decision.

In some applications decision can be expressed not as a single value but rather as a continuous value, function plot or compound decision scheme. Then additional problem arises – we cannot simply say that two decision values are different. There is a need for measuring how close two values of decision are (e.g., measuring distances between probabilistic distributions spanned over the set of decision values [6]). Such measures may be devised in a manner supporting particular goal we want to achieve.

3. Medical Data

We consider the data table gathering 557 patients, labeled with values over columns described in the table below, selected by medical experts as of special importance while analyzing the surgery results.

Operation (*O*)	Radical (r) Modified (m)
Treatment (*T*)	Operation Only (oo) With Radiotherapy (wr) Unsuccessful Radiotherapy (ur)
Ext. Spread (*E*)	1 iff extracapsullar spread is observed, 0 otherwise
Stage (*S*)	Pathological stages, denoted by 0,1,2
Localization (*L*)	Integer codes of the cancer localization
Time Interval (*I*)	Measured between the date of operation and the date of the last notification
Notification (*N*)	Dead (d) Alive (a) No information (n)

The size of data, understood in terms of the number of objects and attributes, is relatively small and thus, any – even exhaustive – approach to searching for appropriate solutions could be applied. A question is, however, not about complexity of the search but about the definition of the problem itself. The main task is to show, whether the risk of modified operation is not greater than in case of radical operation. It is an important factor, since modified operation is less invasive and giving less side effects. According to the experts' knowledge, a person who survives more than 5 years after surgery is regarded as a positively supporting case, even if the same type of cancer repeats after. We have three classes of patients: **Success:** those who survived more than 5 years after surgery, **Defeat:** those who died because of the same cancer as that previously treated, **Unknown:** those who died within 5 years but because of the other reasons and those with no data about the last notification provided.

Let us consider a new decision attribute with three values, corresponding to the above classes. Technically, this attribute can be created by basing on *Time Interval* and *Notification* columns. Searching for rules pointing at the *Success* and *Defeat* decision values may provide the wanted results. Such decision table is very inconsistent: all conditional indiscernibility classes contain objects with all decision values. Hence, the probabilistic analysis should be applied. While analyzing probabilistic decision distributions we should, however, remember that the only thing we know for sure is that objects from *Success* and *Defeat* classes should be discerned. What about the *Unknown* class? In the full version of the paper we deal with this problem by referring to statistical methods used in the medical survival analysis, like the Kaplan-Meier's product-limit estimate and the Cox's proportional hazard model [2]. Now, let us consider the simplified model with just two classes; where the *Unknown* class is merged with the *Success* class, and *Success* means that we have no evidence about *Defeat*.

4. Cross-Decision Rules

The task is to compare the risk of modified and radical operations. Descriptions of groups of patients who should be treated with the particular kind of operation are especially worth finding. We should search for groups of patients treated with both types and compare survival characteristics between such obtained subgroups. If we are able to find description of a set of patients, which splits onto reasonably large subsets in terms of type of operation applied, then knowledge resulting from the

differences in survival characteristics can be informative for medical experts. Let us explain it by basing on an exemplary rule derived from data:

$$T=\text{wr} \wedge E=0 \wedge S=1 \Rightarrow \left| \begin{array}{l} \text{Prob(success after } \textit{radical}\text{)} = 0.36 \\ \text{Prob(success after } \textit{modified}\text{)} = 0.626 \end{array} \right.$$

It means that within the set of patients with surgeries performed after unsuccessful radiotherapy, without extracapsullar spread observed and with the middle level of pathological stage, only 36% patients treated with radical operation survives successfully while modified operation provides success in 62.6%. Let us call this kind of knowledge representation a *cross-decision rule*. It describes two-dimensional statistics related to a pair of features – the type of operation and successful patient's survival – similarly as in case of contingency tables [7].

The question is how to express the criteria enabling automatic extraction of descriptors discerning between different behaviors of such two-dimensional statistics. We propose to follow the distance-based approach to approximate discernibility between probabilistic distributions [6]. For a given $u \in U$, let us consider probabilities of success of radical and modified operations over its A-indiscernibility class $[u]_A$, where $A = \{ T, E, S, L \}$. They take the following form:

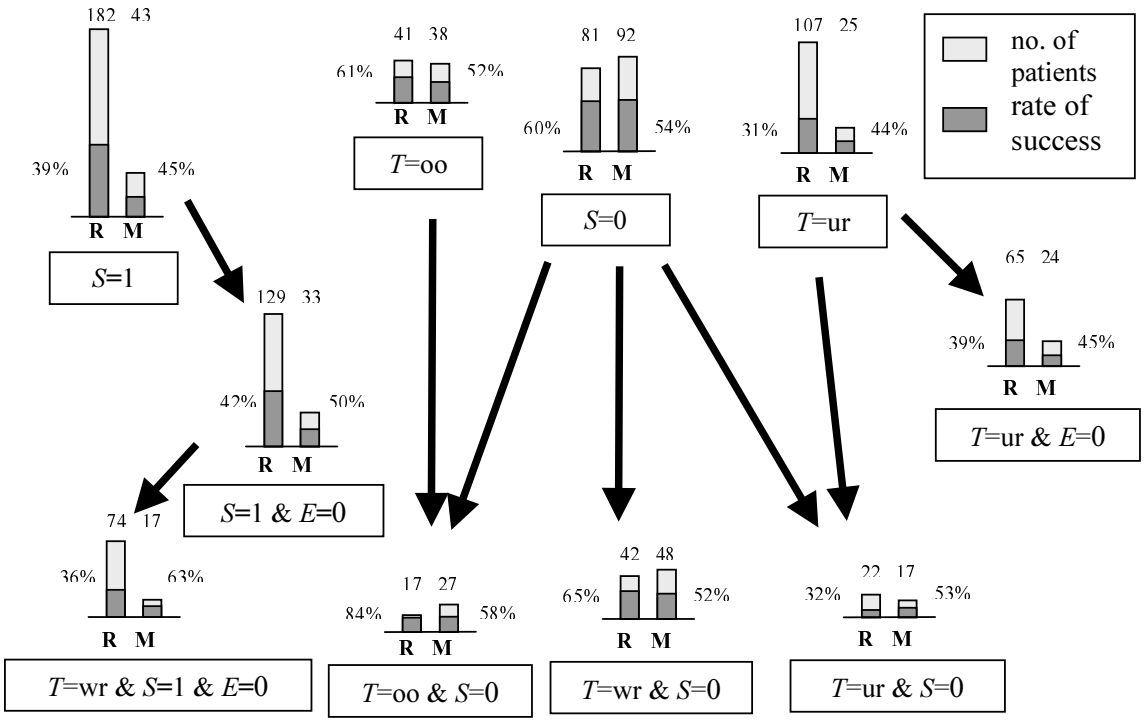
$$R(u) = \frac{|\{u' \in [u]_A : O(u') = r \wedge \text{Suc}(u')\}|}{|\{u' \in [u]_A : O(u') = r\}|} \qquad M(u) = \frac{|\{u' \in [u]_A : O(u') = m \wedge \text{Suc}(u')\}|}{|\{u' \in [u]_A : O(u') = m\}|}$$

where Suc(u) denotes that surgery turned out to be successful for a given $u \in U$. Let us consider the difference $D(u)=R(u)-M(u)$ and approximation threshold $\varepsilon \in [0,1)$. We regard objects $u_1, u_2 \in U$ as necessary to be discerned, iff

$$DD(u_1,u_2) = |D(u_1) - D(u_2)| > \varepsilon$$

Otherwise, we allow putting objects together, as matching the same rule. It provides a method for searching for object-based rules by keeping only these descriptors $a=a(u)$, which are necessary for discerning pairs satisfying the above inequality.

Figure presented in the next page provides characteristics of groups of patients in terms of decision rules. It relates these rules to each other, as being generalizations for higher and counterexamples for lower approximation thresholds. This way of visualization remains analogous to methodology based on *information maps* [5]. We show only these combinations of descriptors, which are irreducible at the level at most $\varepsilon=0.2$ and – moreover – such that the subsets of objects supporting each of both types of surgeries are at least of cardinality 10. There are 11 rules satisfying such requirements. Each rule is labeled with: description of its premise, number of patients treated with two considered surgery types within the set of objects matching the rule, and probabilities of success of these surgeries, conditioned by the premise. Rules are connected with arrows, leading to more specified descriptions, irreducible for some thresholds $\varepsilon>0.2$.



5. Conclusions

Application of rough set based tools to the post-surgery survival analysis of cancer data was discussed. The task was to express the differences between expected results of applications of two types of medical operations and to discover rules, which discern different tendencies in survival statistics. We proposed how to search for and visualize so-called cross-decision rules, helpful while comparing the considered surgeries in terms of their results.

Acknowledgements

Supported by Polish National Committee for Scientific Research (KBN) grant No. 8T11C02519. Special thanks to Medical Center of Postgraduate Education.

References

[1] Cios, K.J., Kacprzyk, J. (Eds): Medical Data Mining and Knowledge Discovery. *Studies in Fuzziness and Soft Computing 60*, Physica Verlag, Heidelberg 2001.
[2] Hosmer, D.W. Jr., Lemeshow, S.: *Applied Survival Analysis: Regression Modeling of Time to Event Data*. John Wiley & Sons, Chichester 1999.
[3] Pawlak, Z.: *Rough sets – Theoretical aspects of reasoning about data*. Kluwer Academic Publishers, Dordrecht 1991.
[4] Skowron, A., Rauszer, C.: The discernibility matrices and functions in information systems. In: R. Słowiński (Ed.), *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory*, Kluwer Academic Publishers, Dordrecht, 1992, pp. 311–362.
[5] Skowron, A., Synak, P.: Patterns in Information Maps. In: Proc. of RSCTC'02, 2002.
[6] Ślęzak, D.: *Approximate decision reducts* (in Polish). Ph.D. thesis, Institute of Mathematics, Warsaw University, 2001.
[7] Żytkow, J., Zembowicz, R.: From contingency table to other forms of knowledge. In: U. Fayyad, G. Shapiro, P. Smyth and R. Uthurswamy (Eds.), *Advances in Knowledge Discovery and Data Mining*, 1997.