Rough Set Approach to the Survival Analysis

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

[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

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. The rough set framework is combined with the Kaplan-Meier product estimation and the Cox's proportional hazard modeling.

1 Introduction

Analysis of medical data requires decision models well understandable by medical experts. The theory of rough sets [4] provides knowledge representation which is well understandable for the medical experts. A number of valuable rough set based applications to medical domain is known from the literature [1,2].

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. Thus, this data set seems to be perfect for learning how complex decision semantics can influence the algorithmic framework and results of its performance.

The main topics of the paper are the following: *(i)* Given statistical methods used in the medical survival analysis, like the Kaplan-Meier's product-limit estimate and the Cox's proportional hazard model [3], find descriptions of patient groups with different survival estimates. *(ii)* 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.

J.J. Alpigini et al. (Eds.): RSCTC 2002, LNAI 2475, pp. 522–529, 2002.

2 Rough Set Framework

In the rough set theory [4] the sample of data takes the form of an information system $\mathbb{A} = (U, A)$, where each attribute $a \in A$ is a function $a : U \to V_a$ into the set of all possible values on a. Given arbitrary $a \in A$ and $v_a \in V_a$, we say that object $u \in U$ *supports descriptor* $a = v_a$ iff $a(u) = v_a$.

One can regard descriptors as *boolean variables* and use them to construct logical formulas as their boolean combinations. The sets of objects supporting such combinations are obtained by using standard semantics of logical operators. The domain of the rough set theory is to *approximate concepts* $X \subseteq U$ by means of supports of boolean formulas constructed over $\mathbb{A} = (U, A)$ [4,7].

One often specifies a distinguished decision d to be predicted under the rest of attributes A. Let us consider *decision table* $\mathbb{A} = (U, A \cup \{d\})$, $d \notin A$. Concepts to be approximated are *decision classes* of objects supporting descriptors $d = v_d$, $v_d \in V_d$. The most widely applied formulas are conjunctions of *conditional descriptors* $a = v_a$, $a \in A$, $v_a \in V_a$. They correspond to *decision rules*

$$\bigwedge_{a \in B}(a = v_a) \Rightarrow (d = v_d) \tag{1}$$

where (*almost*) all objects, which support $a = v_a$, $a \in B$, should support $d = v_d$.

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 by introducing *the set approximations* [4], *generalized decision functions* and, e.g., *rough membership functions* [5]. In some applications decision can be expressed as a continuous value, function plot or a compound decision scheme (like, e.g., *rough membership distribution* – probabilistic distribution spanned over the set of original decision values [9]). Then there is a need for measuring how close two values of decision are. Such measures may be devised in a manner supporting the particular goal we want to achieve.

Regardless of the decision value semantics, the rough set principle of searching for approximations of decision concepts remains the same. It corresponds to the problem of extraction of the optimal set of (*approximate*) decision rules from data; In other words – the problem of construction of an *approximation space* [7, 10], within which the classes of objects with similar decision behavior will be well described by conditional boolean formulas. Given decision table $\mathbb{A} = (U, A \cup \{d\})$, we would thus like to search for such *indiscernibility classes*

$$[u]_B = \{u' \in U : \forall_{a \in B}(a(u) = a(u'))\} \tag{2}$$

that objects $u' \in [u]_B$ have decision values (sets, vectors, distributions, function plots, estimates, etc.) similar to $d(u)$, and objects $u' \notin [u]_B$ have decision values being far from $d(u)$, according to specified decision distance semantics. Moreover, we would like to optimize and simplify the structure of such classes (clusters, neighborhoods), to obtain possibly general description of approximate conditions→decision dependencies (cf. [7,9,10]).

3 Medical Data

We consider the data table gathering 557 patients, labeled with values over the columns described in Fig. 1, selected by the 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)

Fig. 1. The selected attributes of medical data

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 three decision values. Hence, a kind of probabilistic analysis should be applied (cf. [1,2,9]). 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? How to relate it to the others? How to adjust the discernibility criteria in order to get well-founded results? – We try to answer to these questions by handling decision values having a compound type, partially symbolic (for the *Success* class), partially numeric (since one may be interested e.g. in averages over the survival periods in the *Defeat* class), and partially undefined (unknown), needing perhaps a kind of indirect estimation.

4 Survival Analysis

4.1 The Kaplan-Meier Product-Limit Estimate

In the studies concerned with survival analysis, especially in medical domain, we consider two types of observations (patient records) – *complete* and *censored.* The observation is complete if it has lasted for the period of 5 years and in that time a repeating cancer was recorded. Otherwise, the observation is censored – In our case the set of censored objects coincides with the sum of the *Success* and *Unknown* classes of patients.

The Kaplan-Meier product-limit estimate (cf. [3]) is a method providing the means for construction of so called *survival function* (or *survivorship*) $S(t)$, given the complete and censored observations ordered in time. It returns the cumulative proportion of cases surviving up to the time t:

$$S(t) = \prod_{j=1}^{t} \left(\frac{N-j}{N-j+1}\right)^{\delta(j)} \qquad (3)$$

where N denotes the total number of patients and $\delta(j) = 0$ if the j-th case is censored, and 1 otherwise (complete case). The advantage of this method is that it does not depend on grouping of cases into intervals.

In Fig. 2 we present the plot of Kaplan-Meier product-limit estimate for our data set (a) and estimates for groups corresponding to two considered types of surgery (b). One can see that operations (Radical/Modified) form the groups of patients with different survival characteristics. However, we need an automated method of searching for decision rules pointing at possibly more distinguishable characteristics. According to discussion at the end of Section 2, we need an algorithm which is able to search for minimal conjunctions of descriptors enabling to discern objects with local Kaplan-Meier curves that are *far enough* one to the others. Since it's difficult to design an appropriate decision attribute enabling such a performance, we follow a well known statistical method for approximating such curves, described in the next subsection.

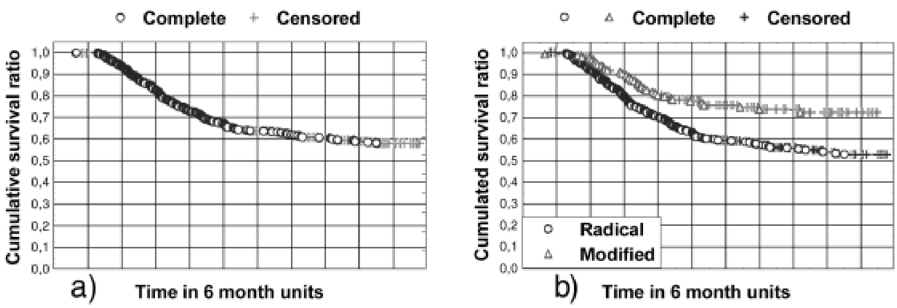

Fig. 2. Kaplan-Meier estimate: a) global b) with respect to the surgery type.

4.2 The Cox's Proportional Hazard Model

This is a semi-parametric method essentially based on assumption that the survival time distribution is exponential (cf. [6]). It attempts to estimate the instantaneous risk of death (hazard) of each particular patient $u \in U$ at time t, by using formula:

$$h(t, a_1(u), \ldots, a_n(u)) = h_0(t) e^{\beta_1 a_1(u) + \cdots + \beta_n a_n(u)} \tag{4}$$

where $a_1, \ldots, a_n$ are attributes (called *indicator variables*), $\beta_1, \ldots, \beta_n$ are *regression coefficients*, and function $h_0(t)$ is so called *baseline hazard.* This model can be linearized by considering a new attribute:

$$\mathsf{PI}(u) = \ln \frac{h(t, a_1(u), \ldots, a_n(u))}{h_0(t)} = \beta_1 a_1(u) + \cdots + \beta_m a_n(u) \tag{5}$$

Prognostic Index PI is a linear combination of existing attributes. Coefficients $\beta_1, \ldots, \beta_n$ are retrieved by using linear regression with the goal of reflecting the actual survival pattern present in the data (expressed as, e.g., Kaplan-Meier estimate) as closely as possible. The task is to find a collection of time-independent variables (attributes) that, at the same time, have clear clinical interpretation and allow the PI to be applied efficiently to making prognosis. In our experiments we used 5 variables (attributes) to form PI – The first 5 listed in Fig. 1. The choice of attributes was suggested by medical experts. By comparing the prediction from the Cox's model with Kaplan-Meier estimate in Fig. 2a, we noticed that the choice of attributes is proper as those two plots were very close (practically indistinguishable at a picture).

If the choice of attributes is proper, one can identify the intervals of values of PI, which correspond to the patient groups with different survivorship patterns. By dividing the data set into three groups with respect to the values of PI we obtain a significant diversification of survivorship patterns, as shown in Fig. 3a. We follow [6], where this method is considered for other type of medical data.

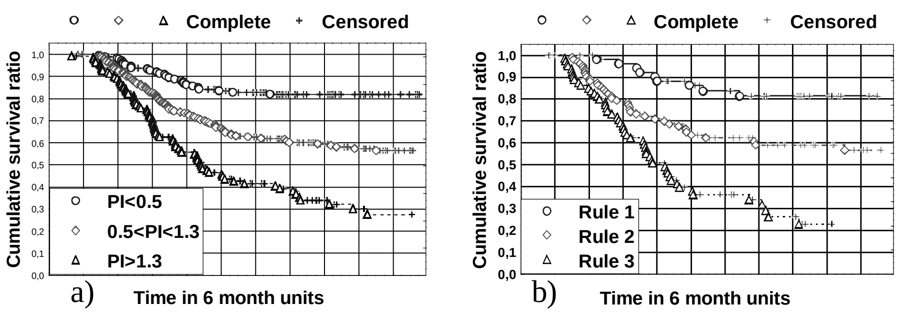

Fig. 3. Application of results from the Cox's proportional hazard model: a) Kaplan-Meier estimate for three groups of patients, defined with respect to the values of PI; b) Kaplan-Meier estimate for objects matching decision rules targeting these intervals.

4.3 Prognostic Rules

Given PI-based patterns shown in Fig. 3a, we can describe efficiently particular tendencies. We define a new decision table, where conditional attributes are, as before, the first 5 in Fig. 1, and values of decision correspond to the above found intervals of values of PI. By searching for decision rules within this table, we can construct approximations of different survival tendencies. We show just three exemplary decision rules (abbreviations are consistent with Fig. 1):

$$\begin{array}{ll} \text{Rule 1:} & O = m \wedge E = 0 \wedge S = 1 \wedge L = 1 \Rightarrow PI < 0.5 \\ \text{Rule 2:} & O = r \wedge T = ur \wedge S = 1 \Rightarrow PI \in [0.5, 1.3) \\ \text{Rule 3:} & O = r \wedge T = ur \wedge E = 1 \wedge S = 2 \wedge L = 1 \Rightarrow PI \geq 1.3 \end{array} \quad (6)$$

The above rules are exact, with matching at the level of 54, 106 and 82 objects, respectively.[1] The right sides can be recalculated in terms of Kaplan-Meier curves corresponding to subsets of objects (patients) defined by the left sides. We illustrate them in Fig. 3b. It leads to the following *prognostic rules*:

$$\begin{array}{l} O = m \wedge E = 0 \wedge S = 1 \wedge L = 1 \Rightarrow \textit{Upper tendency} \text{ in Fig. 3b} \\ O = r \wedge T = ur \wedge S = 1 \Rightarrow \textit{Middle tendency} \text{ in Fig. 3b} \\ O = r \wedge T = ur \wedge E = 1 \wedge S = 2 \wedge L = 1 \Rightarrow \textit{Lower tendency} \text{ in Fig. 3b} \end{array} \quad (7)$$

The proposed method is a two-step process, providing rules related to the complex, Kaplan-Meier-related decision. In the first step we encode the actual decision in terms of PI. Then we recalculate the actual survivorship estimates.

5 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. According to Fig. 2, modified operations seem to provide better results. However, this type is generally not applied to more serious cases. 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.

A solution would be to search for pairs of subsets with Kaplan-Meier estimates as distant to each other as possible. Such estimates are difficult to compare in a direct manner. Hence, one may consider approximations of their behavior, like, e.g., probabilistic distributions over the values of Prognostic Index. Another approach is to focus on percentages of successful operations. Let us explain it by basing on an exemplary rule derived from data:

$$T = wr \wedge E = 0 \wedge S = 1 \Rightarrow \left| \begin{array}{l} P(\text{success after } \textit{radical}) = 0.36 \\ P(\text{success after } \textit{modified}) = 0.626 \end{array} \right. \quad (8)$$

[1] Obviously, one can search also for inexact PI-related rules, which are regarded as better fitting medical phenomena than exact ones (cf. [1,2]). However, the resulting rules are not exact by means of application to survival analysis anyway.

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 [11].

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 extend the distance-based approach to approximate discernibility between probabilistic distributions [9]. 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:

$$R(u) = \frac{|\{u' \in [u]_A : O(u') = r \wedge Suc(u')\}|}{|\{u' \in [u]_A : O(u') = r\}|} \quad M(u) = \frac{|\{u' \in [u]_A : O(u') = m \wedge Suc(u')\}|}{|\{u' \in [u]_A : O(u') = m\}|} \tag{9}$$

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

$$DD(u_1, u_2) = |D(u_1) - D(u_2)| > \varepsilon \tag{10}$$

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 (10).

Fig. 4 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* [8]. 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 in terms of (10) for some thresholds $\varepsilon > 0.2$.

6 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. To cope with it, the rough set framework for generating optimal decision rules was combined with the Kaplan-Meier product estimation and the Cox's proportional hazard modeling. Moreover, we proposed how to search for and visualize so-called cross-decision rules, helpful while comparing the considered surgeries in terms of their results.

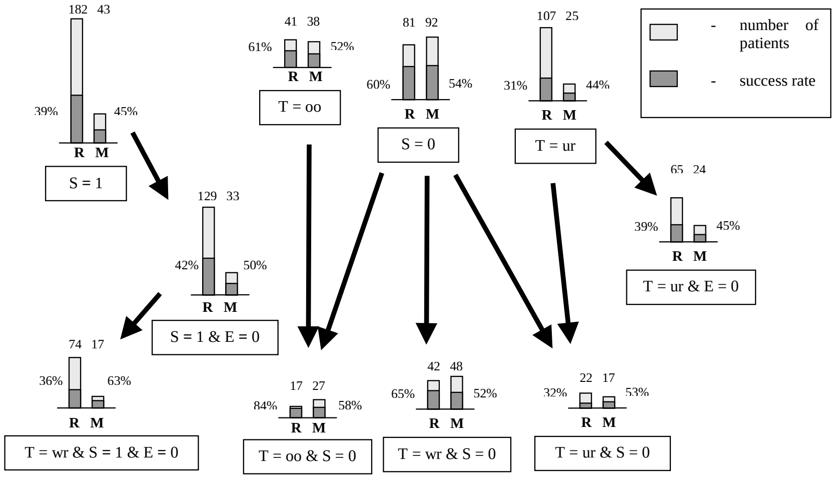

Fig. 4. Visualization of cross-decision rules.

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. Grzymala-Busse J.P., Grzymala-Busse J.W., Hippe Z.S.: Prediction of melanoma using rule induction based on rough sets. In: Proc. of SCI'01 (2001) 7, pp. 523–527.
3. Hosmer D.W. Jr., Lemeshow S.: Applied Survival Analysis: Regression Modeling of Time to Event Data. John Wiley & Sons, Chichester (1999).
4. Pawlak Z.: Rough sets – Theoretical aspects of reasoning about data. Kluwer Academic Publishers (1991).
5. Pawlak Z., Skowron A.: Rough membership functions. In: R.R. Yaeger, M. Fedrizzi, J. Kacprzyk (eds), Advances in the Dempster Shafer Theory of Evidence. Wiley (1994) pp. 251–271.
6. Schlichting P. et al.: Prognostic Factors in Cirrhosis Identified by Cox's Regression Model. Hepatology, 3(6) (1983) pp. 889–895.
7. Skowron A., Pawlak Z., Komorowski J., Polkowski L.: A rough set perspective on data and knowledge. In: W. Kloesgen, J. Żytkow (eds), Handbook of KDD. Oxford University Press (2002) pp. 134–149.
8. Skowron A., Synak P.: Patterns in Information Maps. In Proc. of RSCTC'02 (2002).
9. Ślęzak D.: Approximate decision reducts (in Polish). Ph.D. thesis, Institute of Mathematics, Warsaw University (2001).
10. Wróblewski, J.: Adaptive methods of object classification (in Polish). Ph.D. thesis, Institute of Mathematics, Warsaw University (2001).
11. Zytkow J.M, Zembowicz R.: Contingency Tables as the Foundation for Concepts, Concept Hierarchies, and Rules: The 49er System Approach. Fundamenta Informatice 30, IOS Press (1997) pp. 383–399