Searching for the complex decision reducts The case study of the survival analysis

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

[1] Institute of Mathematics, University of Rzeszów
Rejtana 16A, 35-959 Rzeszów, Poland
[2] Institute of Mathematics, University of Warsaw
Banacha 2, 02-097 Warsaw, Poland
[3] Department of Computer Science, University of Regina
Regina, SK, S4S 0A2 Canada
[4] Polish-Japanese Institute of Information Technology
Koszykowa 86, 02-008 Warsaw, Poland

Abstract. Generalization of the fundamental rough set discernibility tools aiming at searching for relevant patterns for complex decisions is discussed. As an example of application, there is considered the post-surgery survival analysis problem for the head and neck cancer cases. The goal is to express dissimilarity between different survival tendencies by means of clinical information. It requires handling decision values in form of plots representing the Kaplan-Meier product estimates for the groups of patients.

1 Introduction

In many rough set approaches to data analysis, especially these dedicated to (strongly) inconsistent data tables, where the decision class approximations cannot be determined to a satisfactory degree, decisions can take more complex forms, e.g., the collections or probabilistic distributions of the original decision values (cf. [6, 10]). In the same way, one could consider, e.g., statistical estimates, plots, etc., definable using the original attributes, in a way appropriate for a particular decision problem. Then, one should search for relevant patterns for approximation of such decision structures. We study how complex decision semantics can influence the algorithmic framework and results of its performance. We show that quite unusual structures can be still handled using just slightly modified rough set algorithms based on discernibility and Boolean reasoning [9].

Complex attribute values occur often in the medical domain, while analyzing heterogeneous data gathering series of measurements, images, texts, etc. [2]. We illustrate our approach using data representing medical treatment of patients with the head and neck cancer cases. The data table, collected for years by Medical Center of Postgraduate Education in Warsaw, Poland, consists of 557 patient records described by 29 attributes. The most important attributes are well-defined qualitative features. The decision problem, however, requires approximation of especially designed complex decision attribute, corresponding to the needs of the survival analysis [3].

One may conclude that the proposed methodology is applicable not only to the presented case study but also to other medical, as well as, e.g., multimedia or robotics problems. The results can also be treated as a step towards hybridization of case-based reasoning with the rough set approach [4, 7].

2 Illustrative Example

In rough set theory the sample of data takes the form of an information system $\mathbb{A} = (U, A)$, where each attribute $a \in A$ is a function $a : U \to V_a$ from the universe U into the set V_a of all possible values of a. Figure 1 illustrates the meaning of the attribute values for the information system $\mathbb{A} = (U, A)$, where $A = \{\#, ttr, st_l, st_{cr}, loc, gap, rec\}$. U gathers 557 patients labelled with their values for the elements of A. For instance, the object with the vector of values $(1, after, T2, cN3, throat, 3.5, 1)$ corresponds to a patient, who was treated with one-sided operation, after unsuccessful radiotherapy, with the local stage of cancer classified as $T2$, the regional stage clinically (before the operation) classified as $cN3$, with the cancer recognized in the throat, the last notification done after 3.5 years, during which the cancer recurrence was observed.

Column with description	Values with description
# – No. of Sides Operated	1 – operation needed at one side; 2 – at both sides
ttr – Type of Treatment	*only* – only operation applied; *radio* – together with radiotherapy; *after* – after unsuccessful radiotherapy
st_l – Local Stage	*T1, T2, T3, T4*
st_{cr} – Clinical Regional Stage	*cN0, cN1, cN2, cN3*
loc – Localization	*larynx, throat, other*
gap – Time Interval Gap	the gap between operation and the last notification
rec – Recurrence Notification	1 – recurrence observed; 0 – otherwise

Fig. 1. The selected attributes of medical data

Object $u \in U$ *supports descriptor* $a = v_a$ iff $a(u) = v_a$. Descriptors, treated as *boolean unary predicates*, are atomic logical formulas. Descriptors for quantitative attributes can be built using also, e.g., inequalities. According to the experts, a person who survives more than 5 years after surgery is regarded as the *success* case, even if the same type of cancer repeats after. A person who dies within 5 years can be the *defeat* or *unknown* case due to the reason of death. We obtain the following *decision classes* of patients, described by means of conjunctions of descriptors built over qualitative *rec* and quantitative *gap*:

1. ***defeat***: the set of objects, which support conjunction $rec = 1 \wedge gap < 5$
2. ***unknown***: the set of objects, which support conjunction $rec = 0 \wedge gap < 5$
3. ***success***: the set of objects, which support descriptor $gap \geq 5$

One of the aims of rough set theory is to approximate decision classes by means of *conditional attributes* [5]. We want to approximate *defeat*, *unknown*, and *success* using clinical information. We consider *decision table* $\mathbb{A} = (U, C \cup \{d\})$ with conditional attributes $C = \{\#, ttr, st_l, st_{cr}, loc\}$ and distinguished *decision attribute* $d \notin C$, which indicates decision classes defined above.

Decision approximation is usually stated by means of *"if .. then .." rules*, such that (*almost*) all objects that support the conditional part of the rule, drop into the specified decision class. *Inconsistency* of $\mathbb{A} = (U, C \cup \{d\})$ can be expressed by, e.g., *boundary regions* of decision classes [5], *generalized decision sets* [8, 9], or *rough memberships* [6, 10], which label each $u \in U$ with distribution of its *indiscernibility class* $[u]_C = \{u' \in U : \forall_{a \in C}(a(u) = a(u'))\}$ among decisions:

$$\overrightarrow{\mu}_{d/C}(u) = \left\langle \frac{|[u]_C \cap defeat|}{|[u]_C|}, \frac{|[u]_C \cap unknown|}{|[u]_C|}, \frac{|[u]_C \cap success|}{|[u]_C|} \right\rangle \quad (1)$$

Inconsistency of $\mathbb{A}$ corresponds to distributions, which do not specify a unique decision class for some $u \in U$. Figure 2 illustrates such distributions for a couple of elements of U. In this case one can expect difficulties in constructing reasonable decision rules. We discuss a solution of this problem in the next section.

u	#	ttr	st_l	st_{cr}	loc	$\lvert[u]_C\rvert$	$\lvert[u]_C \cap def\rvert$	$\lvert[u]_C \cap unk\rvert$	$\lvert[u]_C \cap suc\rvert$
0	1	*only*	$T3$	$cN1$	*larynx*	25	15	4	6
4	1	*after*	$T3$	$cN1$	*larynx*	38	8	18	12
24	1	*radio*	$T3$	$cN1$	*larynx*	23	6	7	10
28	1	*after*	$T3$	$cN0$	*throat*	18	4	8	6
57	1	*after*	$T4$	$cN1$	*larynx*	32	12	14	6
91	1	*after*	$T3$	$cN1$	*throat*	35	5	16	14
152	1	*only*	$T3$	$cN0$	*larynx*	27	9	14	4
255	1	*after*	$T3$	$cN0$	*larynx*	15	2	6	7
493	1	*after*	$T3$	$cN1$	*other*	19	6	7	6
552	2	*after*	$T4$	$cN2$	*larynx*	14	6	3	5

Fig. 2. Statistics for randomly selected objects. The first column contains the object's ordinal number. The next five columns contain the attribute values. The last four columns contain cardinalities enabling calculation of the rough membership coefficients.

3 Discernibility-based reduction

Approximation of decision classes corresponds to the construction of an *approximation space* [8], where objects with similar decisions are well described by conditional formulas. Given $\mathbb{A} = (U, C \cup \{d\})$, we search for indiscernibility classes $[u]_C$, such that if $u' \in [u]_C$, then $d(u')$ is *close to* $d(u)$. We also try to generalize such classes by reducing the number of needed attributes (cf. [8, 9]). Let us consider *discernibility matrix* $\mathbb{M}_{\mathbb{A}}$ (cf. [9]), where:

1. columns correspond to attributes $a \in C$
2. rows correspond to the pairs of objects (u, u') such that $d(u) \neq d(u')$
3. for row (u, u') and column $a \in C$ we put 1, if $a(u) \neq a(u')$ and 0 otherwise

Any *irreducible covering* $B \subseteq C$ of $\mathbb{M}_{\mathbb{A}}$ [1] corresponds to a *decision reduct* – an irreducible subset of attributes providing consistent subtable $\mathbb{B} = (U, B \cup \{d\})$.

[1] The covering of binary $\mathbb{M}_{\mathbb{A}}$ takes the form of any subset of columns $B \subseteq C$ such that for any row we have at least one $a \in B$ with value 1 on this row.

For inconsistent $\mathbb{A} = (U, C \cup \{d\})$ there is impossible to cover $\mathbb{M}_{\mathbb{A}}$ at all. Still, one can search for reducts as the *approximate* coverings of $\mathbb{M}_{\mathbb{A}}$ or as the coverings of *modified* matrices (cf. [8–10]). For instance, the rows of a discernibility matrix can correspond to the pairs of objects with different rough membership distributions. Any irreducible covering of such a matrix corresponds to a decision reduct for consistent decision table $\mathbb{A} = (U, C \cup \{\overrightarrow{\mu}_{d/C}\})$. Then, however, we cannot group the objects with *very similar* distributions. A solution ([10]) is to consider only the pairs (u, u') with *enough distant* distributions, i.e. such that

$$\varrho\left(\overrightarrow{\mu}_{d/C}(u), \overrightarrow{\mu}_{d/C}(u')\right) \geq \alpha \tag{2}$$

for a specified function ϱ and threshold $\alpha > 0$. Irreducible coverings of such obtained matrix, further denoted by $\mathbb{M}^{\alpha}_{\mathbb{A}}$, provide *$\alpha$-approximate decision reducts* $B \subseteq C$, which *approximately* preserve information induced by C about d. Condition (2) can be applied with *any other* function ϱ, which measures distances between *any other* complex decision values calculated for classes $[u]_C$, $u \in U$.

u	u'	$\#$	ttr	st_l	st_{cr}	loc
0	255	0	1	0	1	0
0	91	0	1	0	0	1
0	4	0	1	0	0	0
0	28	0	1	0	1	1
0	152	0	0	0	1	0
0	24	0	1	0	0	0

u	u'	$\#$	ttr	st_l	st_{cr}	loc
152	255	0	1	0	0	0
152	552	1	1	1	1	0
91	552	1	0	1	1	1
57	255	0	0	1	1	0
255	552	1	0	1	1	0
24	152	0	1	0	1	0

Fig. 3. $\mathbb{M}^{\alpha}_{\mathbb{A}}$ for $U = \{0, 4, 24, 28, 57, 91, 152, 255, 493, 552\}$. The rows correspond to the pairs, for which Euclidean distance between distributions is not lower than $\alpha = 0.365$. One can see that the only irreducible covering takes the form of the set $B = \{ttr, st_{cr}\}$. This is the only α-approximate decision reduct in this case.

The main theoretical contribution of this paper is the reduction methodology based on conditions similar to (2), but at the level of *local decision reducts* [9]. An *α-approximate* local reduct is built by discerning a given $u \in U$ from all $u' \in U$ such that inequality (2) holds. It corresponds to operations on matrix $M^{\alpha}_{\mathbb{A}}(u)$, which is $M^{\alpha}_{\mathbb{A}}$ restricted to the rows related to u. A covering $B \subseteq C$ of $M^{\alpha}_{\mathbb{A}}(u)$ generates the rule described by the u's values on B: if $u' \in U$ supports descriptors $a = a(u)$ for all $a \in B$, then the decision of u' is *close to* that of u.

Just like at the level of decision reducts, we can consider arbitrary criteria for measuring the decision distances. A general problem is that sometimes there can be objects $u', u'' \in U$, which do not need to be discerned from u but their decision characteristics are too distant to each other to put both of them to the same class.[2] Therefore, we should add to $M^{\alpha}_{\mathbb{A}}(u)$ also the rows encoding the need of keeping at least one of objects u', u'' outside the support of any local reduct derived at the basis of u. It is illustrated in Figure 4 for criterion (2). This is a novel approach, which can be extended to other types of complex decisions, assuming a distance measure between the decision values is given.

[2] This problem does not occur in the classical case, where objects $u', u'' \in U$ need not to be discerned from u, iff $d(u) = d(u')$ and $d(u) = d(u'')$, what implies $d(u') = d(u'')$.

u	u'	#	ttr	st_l	st_{cr}	loc
0	91	0	1	0	0	1
152	255	0	1	0	1	1
91	552	1	0	1	1	1
57	255	0	0	1	1	1
24	152	0	1	0	1	1

Fig. 4. Matrix $M^\alpha_{\mathbb{A}}(u)$ for $u = 91$ and $\alpha = 0.365$. Its coverings correspond to *α-approximate local reducts* $B1 = \{ttr, st_l\}$, $B2 = \{st_l, st_{cr}\}$, and $B3 = \{loc\}$. Their supports are equal $[u]_{B1} = \{4, 28, 91, 255, 493\}$, $[u]_{B2} = \{4, 57, 91, 493\}$, and $[u]_{B3} = \{28, 91\}$. They correspond to patterns $ttr = after \wedge st_l = T3$, $ttr = after \wedge st_{cr} = cN1$, and $loc = throat$.

Another problem corresponds to the task of case-based reasoning, aiming at deriving new decisions from the clusters of objects with similar decision characteristics (cf. [4, 7]). As an example, let us consider the case of handling rough membership distributions. We can label a given cluster $[u]_B$, obtained as the support of α-approximate local reduct $B \subseteq C$ obtained at the basis of $u \in U$, with distribution $\overrightarrow{\mu}_{d/B}(u)$ calculated as (1), for $[u]_B$ instead of $[u]_C$. One can rewrite $\overrightarrow{\mu}_{d/B}(u)$ as the average of distributions $\overrightarrow{\mu}_{d/C}(u')$, $u' \in U$.

It is shown in [10] that if Euclidean distances between distributions $\overrightarrow{\mu}_{d/C}$ of all elements of $[u]_B$ are lower than α, what is the case for α-approximate local reducts, then the same can be said about distances between $\overrightarrow{\mu}_{d/C}(u')$ and $\overrightarrow{\mu}_{d/B}(u)$, for any $u' \in [u]_B$. Therefore, we can talk about a *complex decision rule* saying that if a given object supports descriptors $a = a(u)$, for any $a \in B$, then its decision distribution is *close to* $\overrightarrow{\mu}_{d/B}(u)$.

The above kind of case-based reasoning analysis should be reconsidered for any other applied decision and distance semantics. For instance, in the following sections we discuss the local reduct patterns grouping *similar* plots representing the Kaplan-Meier product estimates (cf. [3]). Although we can search for the groups of such plots using the same discernibility procedure as above, further research is needed to examine to what extent the estimate calculated for some class $[u]_B$, $u \in U$, $B \subseteq C$, can be regarded as a *representative* for the collection of mutually similar estimates calculated locally for classes $[u']_C$, $u' \in [u]_B$.

4 Discernibility approach to the survival analysis

In the survival analysis, one distinguishes *complete* and *censored* objects. In case of the considered medical data, the set of complete objects coincides with the *defeat* decision class. The Kaplan-Meier product-limit estimate (cf. [3]) provides the means for construction of *survival function* $S(t)$, which returns cumulative proportion of cases surviving up to the time t after operation. We define it as

$$S(t) = \prod_{u \in defeat: gap(u) \leq t} \frac{|U| - |\|gap \leq gap(u)\|_{\mathbb{A}}|}{|U| - |\|gap \leq gap(u)\|_{\mathbb{A}}| + 1} \tag{3}$$

where $\|gap \leq gap(u)\|_{\mathbb{A}} = \{u' \in U : gap(u') \leq gap(u)\}$. $S(t)$ can be recalculated for any subset of U as illustrated in Figure 5. Since all the cases $u \in U$, such that inequality $gap(u) > 5$ is satisfied, are censored (because they are *successes*), we restrict ourselves to the survival plots within the range of $[0, 5]$ years.

One used to assume that a given attribute provided more information, if it split data onto subtables with less similar survival plots. For instance, Figure 5 illustrates the meaning of *ttr*. The patients not treated with radiotherapy seem to have more chances for survival, because the corresponding plot has the highest level of chances after 5 years. It does not mean, however, that radiotherapy should not be applied. The type of treatment $ttr = only$ is applied to relatively less severe cases, which makes the corresponding survival characteristics more optimistic "by definition". This is an example why the decision behaviors corresponding to the values of single attributes should be analyzed in the context of other attributes. It was a motivation for applying to this data the proposed generalization of the rough set approach to searching for approximate reducts.

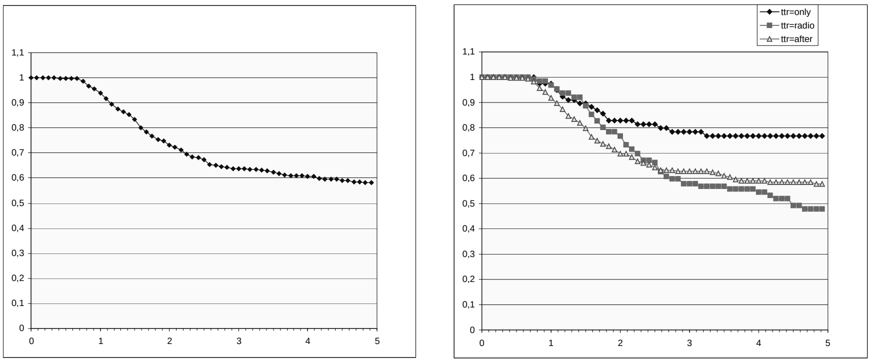

Fig. 5. The plots of function $S(t)$. The left plot corresponds to the whole U. The right plots correspond to subtables filtered with respect to foregoing values of attribute *ttr*.

Let us label each $u \in U$ with the survival function $S_u^C(t)$ obtained for the indiscernibility class $[u]_C$. Since functions $S_u^C(t)$ seem to contain more adequate information than $\overrightarrow{\mu}_{d/C}(u)$, we repeat the discernibility reduction process described in the previous section for the survival plot decisions. To do this, we can follow the same procedures of creating and analyzing α-approximate discernibility matrices $M_{\mathbb{A}}^\alpha$ and $M_{\mathbb{A}}^\alpha(u)$, but for the decision distances calculated between the survival plots instead of Euclidean distances between distributions.

We considered two examples of distances enabling to compare pairs of objects $u, u' \in U$. The first one, $\varrho_{area}(u, u')$, equals the area between the plots of $S_u^C(t)$ and $S_{u'}^C(t)$. The second one, $\varrho_{merged}(u, u')$, refers more to particular cases than to the plots. It averages the differences between the survival estimates for the *defeat* objects in $[u]_C \cup [u']_C$ before and after merging $[u]_C$ and $[u']_C$ within a more general cluster. A broader family of distances should be analyzed in future.

Given functions ϱ_{area} or ϱ_{merged}, we can search for clusters of objects with α-approximately similar Kaplan-Meier characteristics. Procedure based on discernibility matrices $M_{\mathbb{A}}^\alpha(u)$ assures that any local reduct $B \subseteq C$ obtained for a given $u \in U$ forms the cluster $[u]_B$ of α-approximately similar objects. A question is whether the plots $S_u^B(t)$ corresponding to such clusters can be regarded as their *representatives* in the same way as discussed for distributions $\overrightarrow{\mu}_{d/B}(u)$. As mentioned before, it must be analyzed, whether mutual closeness of estimates $S_{u'}^C(t)$ for all $u' \in [u]_B$ assures the same kind of closeness to $S_u^B(t)$. Although the experiments confirm this tendency, further theoretical studies are needed.

5 Selected experimental results

We performed experiments for the data table described by conditional attributes $C = \{\#, ttr, st_l, st_{cr}, loc\}$. Just like in Section 3, we were generating $\mathbb{M}_{\mathbb{A}}^{\alpha}$ and $\mathbb{M}_{\mathbb{A}}^{\alpha}(u)$ basing on criterion (2), now applied for distances ϱ_{area} and ϱ_{merged}.

α	ϱ_{area}	ϱ_{merged}	ϱ_{area}	ϱ_{merged}
0.3	$\{\#, ttr, st_l, st_{cr}, loc\}$	$\{ttr, st_l, st_{cr}, loc\}$	57	37
0.4	$\{ttr, st_l, st_{cr}, loc\}$	$\{ttr, st_l, loc\}$	39	43
0.5	$\{ttr, st_l, st_{cr}, loc\}$	$\{st_l, st_{cr}, loc\}, \{\#, st_l, st_{cr}\}, \{ttr, st_l\}$	34	43
0.6	$\{st_l, st_{cr}\}$	$\{st_l, st_{cr}\}, \{ttr, st_l\}$	25	27
0.7	$\{st_l, st_{cr}\}, \{ttr, st_l, loc\}$	$\{st_{cr}, loc\}, \{st_l, st_{cr}\}, \{ttr\}$	13	18

Fig. 6. The reduction results for various approximation thresholds, where the first column contains the threshold value α, two next columns contain α-approximate decision reducts, and two last columns contain numbers of α-approximate local decision reducts.

Because of the lack of space, we focus just on two observations. The first one concerns the type of treatment ttr. In Figure 6 we can see that ttr occurs in majority of reducts, for various approximation thresholds. In case of ϱ_{merged} and $\alpha = 0.7$ it even begins to be an approximate reduct itself. On the other hand, the occurrence of ttr together with other attributes for lower thresholds suggests that it should not be considered totally independently, as noticed in Section 4.

$\lvert[u]_B\rvert$	B	#	ttr	st_l	st_{cr}	loc
79	$\{ttr\}$	$*$	$only$	$*$	$*$	$*$
132	$\{ttr\}$	$*$	$radio$	$*$	$*$	$*$
346	$\{ttr\}$	$*$	$after$	$*$	$*$	$*$
82	$\{st_l\}$	$*$	$*$	$T2$	$*$	$*$
361	$\{st_l\}$	$*$	$*$	$T3$	$*$	$*$
276	$\{st_{cr}\}$	$*$	$*$	$*$	$cN1$	$*$
174	$\{loc\}$	$*$	$*$	$*$	$*$	$throat$
46	$\{\#, loc\}$	1	$*$	$*$	$*$	$other$
62	$\{\#, loc\}$	2	$*$	$*$	$*$	$larynx$

$\lvert[u]_B\rvert$	B	#	ttr	st_l	st_{cr}	loc
25	$\{st_l, st_{cr}\}$	$*$	$*$	$T4$	$cN0$	$*$
27	$\{st_l, st_{cr}\}$	$*$	$*$	$T4$	$cN2$	$*$
6	$\{st_l, st_{cr}\}$	$*$	$*$	$T4$	$cN3$	$*$
102	$\{st_{cr}, loc\}$	$*$	$*$	$*$	$cN0$	$larynx$
49	$\{st_{cr}, loc\}$	$*$	$*$	$*$	$cN2$	$larynx$
14	$\{st_{cr}, loc\}$	$*$	$*$	$*$	$cN3$	$larynx$
11	$\{st_{cr}, loc\}$	$*$	$*$	$*$	$cN0$	$other$
6	$\{st_{cr}, loc\}$	$*$	$*$	$*$	$cN2$	$other$
5	$\{st_{cr}, loc\}$	$*$	$*$	$*$	$cN3$	$other$

Fig. 7. α-approximate local reducts obtained for $\alpha = 0.7$ and ϱ_{merged}. The first column contains the number of supporting objects. The second column contains attributes defining the reduct. The last five columns describe the reduct patterns. For instance, $*, *, T4, cN3, *$ corresponds to pattern $st_l = T4 \wedge st_{cr} = cN3$, supported by 6 objects.

The most frequent attribute occurring in Figures 6, 7 is the local cancer stage st_l. This is very surprising because the initial hypothesis formulated by medical experts was that st_l can be reduced given the rest of considered clinical features. Actually, st_l does seem to provide less amount of information than the other attributes while comparing the Kaplan-Meier estimates for their particular values. However, as often happens in the rough set reduction processes, potentially least valuable attributes turn out to be crucial for discerning important cases.

6 Conclusions

We discussed a rough set approach to extraction of relevant patterns for compound decisions. We considered decision values modeled by rough membership distributions (cf. [6, 10]) and the Kaplan-Meier's product-limit survival estimates (cf. [3]). We focused on searching for possibly minimal subsets of attributes approximately preserving the decision information, as well as the clusters of objects with approximately similar decision characteristics, described by possibly general patterns. The solutions were presented as approximate reducts derived using appropriately modified rough set discernibility procedures (cf. [9]).

In future we plan to develop a general approach to visualization of the chains of patterns characterized by various thresholds for decision distances, as initiated in [1]. We are going to strengthen the correspondence between the issue of the unified representation of the classes of objects with similar decision characteristics and the case-based reasoning challenges (cf. [7]). We also plan to continue the experiments concerning the considered medical data, in purpose of extending the results of this paper, as well as our previous experiences in this area.

Acknowledgements: Supported by Polish National Committee for Scientific Research (KBN) grant No. 8T11C02519. Special thanks to Dr. Antoni Osmólski from the Medical Center of Postgraduate Education in Warsaw, Poland.

References

1. Bazan J., Osmólski A., Skowron A., Ślęzak D., Szczuka M., Wróblewski J.: Rough set approach to the survival analysis. In Proc. of RSCTC'02 (2002).
2. Cios K.J., Kacprzyk J. (eds): Medical Data Mining and Knowledge Discovery. Studies in Fuzziness and Soft Computing 60, Physica Verlag, Heidelberg (2001).
3. Hosmer D.W. Jr., Lemeshow S.: Applied Survival Analysis: Regression Modeling of Time to Event Data. John Wiley & Sons, Chichester (1999).
4. Lenz M., Bartsch-Spoerl B., Burkhard H.-D., Wess S. (eds): Case-Based Reasoning Technology – From Foundations to Applications. LNAI 1400, Springer (1998).
5. Pawlak Z.: Rough sets – Theoretical aspects of reasoning about data. Kluwer Academic Publishers (1991).
6. 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.
7. Polkowski L., Skowron A., Komorowski J.: Approximate case-based reasoning: A rough mereological approach. In: Proc. of the 4-th German Workshop on Case-Based Reasoning, System Developments and Evaluation (1996) pp. 144–151.
8. 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.
9. 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.
10. Ślęzak D.: Various approaches to reasoning with frequency-based decision reducts: a survey. In: L. Polkowski, S. Tsumoto, T.Y. Lin (eds.), Rough Sets in Soft Computing and Knowledge Discovery: New Developments. Physica Verlag (2000).