decision rules, user-driven rule induction,
rule quality, survival analysis,
bone marrow transplantation

Marek SIKORA[1], Łukasz WRÓBEL[1], Monika MIELCAREK[2], Krzysztof KAŁWAK[2]

APPLICATION OF RULE INDUCTION TO DISCOVER SURVIVAL FACTORS OF PATIENTS AFTER BONE MARROW TRANSPLANTATION

Decision rules are commonly used tool for classification and knowledge discovery in data. The aim of this paper is to provide decision rule-based framework for analysis of survival data and apply it in mining of data describing patients after bone marrow transplantation. The paper presents a rule induction algorithm which uses sequential covering strategy and rule quality measures. An extended version of the algorithm gives the possibility of taking into account user's requirements in the form of predefined rules and attributes which should be included in the final rule set. Additionally, in order to summarize the knowledge expressed by rule-based model, we propose the rule filtration algorithm which consists in selection of statistically significant rules describing the most disjoint parts of the entire data set. Selected rules are identified with so-called survival patterns. The survival patterns are rules which conclusions contain Kaplan-Meier estimates of survival function. In this way, the paper combines rule-based data classification and description with survival analysis. The efficiency of our method is illustrated with the analysis of data describing patients after bone marrow transplantation.

1. INTRODUCTION

Frequent task in medical research studies is an analysis of gathered research material in terms of searching for some patterns. In addition to statistical methods, data mining and knowledge discovery are becoming more popular for such studies [6]. In this paper, we apply the decision rule induction algorithm to analyze survival data. Decision rule induction [14],[18],[20],[26],[31],[45],[52] is widely used knowledge discovery method applied for pattern identification in data. We use decision rules for two basic aims. One is the prediction of whether the patient would survive longer than given time. The other is the description of patient groups with different survival characteristics. In particular, we present a way of associating decision rules with Kaplan-Meier analysis [30]. We also improve readability of the rule-based model with the use of rule filtration algorithm which selects the rules describing the most disjoint groups of patients.

Rule induction algorithms automatically select attributes and ranges of attributes which occur in their premises. This causes the situations in which induced rules not always contain the most important information for a user. We propose a modification of the rule induction algorithm which allows taking into account user's requirements. These requirements are defined as the initial set of rules and attributes which should be included in the final rule set.

The analysis of results obtained by various rule induction algorithms leads to the statement that description and classification abilities of induced rules depend, among others, on measures applied to

[1]Institute of Computer Sciences, Silesian University of Technology, Akademicka 16, 44-100 Gliwice, Poland.
[2]Department of Pediatric Hematology, Oncology and Bone Marrow Transplantation, Wrocław Medical University, Bujwida 44, 50-345 Wrocław, Poland.

control the process of induction [3],[8],[21],[29],[45],[46],[48]. Quality measures are applied in the rule building phases (growing and pruning) as well as during the classification conflicts resolution. Depending on characteristics of the analyzed data set, various measures may lead to various rule sets characterized by different classification and description abilities. In the presented studies the classifier with the best classification abilities was obtained based on the Correlation [21] measure. In order to obtain the best possible classifier, the efficiency of 34 measures known as rule quality and/or rule interestingness measures, was tested [46],[48].

The efficiency of our proposals is illustrated with the analysis of data describing patients after bone marrow transplantation. Factors affecting 5-year survival rate since the transplantation date are searched in the data set. The data analysis shows that the best results in the construction of rule-based data models used to classification and description are obtained in a semi-automatic way, through interaction with a domain expert.

The paper is organized as follows. In Section 2, we present a survey of rule-based methods in medical applications. In Section 3, we discuss the problems of rule induction, rule evaluation, rule filtration and survival models selection. The characteristics of analyzed data set describing patients after bone marrow transplantation and the results of experiments carried out on this data are presented in Section 4. Section 5 is a summary of the paper and the presentation of the directions of further research.

2. RELATED WORK - RULE-BASED METHODS IN MEDICAL APPLICATIONS

In medical data analysis rule induction was applied mainly to problems concerning diseases' diagnosis (in expert systems diagnosing head diseases [55],[56],[57], liver disorders [58], obesity problems with children [13], early diagnosis of rheumatic diseases [16], in verification of magnetic resonance diagnosis [49], etc). Recently, works making attempts at applying the rule induction and rough set theory [40] for identification of features which have the most major influence on patient's survival time [5],[41] have also appeared.

In the UCI Repository [19] which is benchmark data set repository for machine learning problems, medical databases are sizable group. In papers [15],[16],[36] efficiency of various rule induction methods is illustrated by the analysis of benchmark medical databases (Pima Indians Diabetes, Liver Disorders, Lymphography etc.).

Association rule induction algorithms or their modifications are frequently used to obtain rule-based description of gathered data [2],[32],[47],[53]. The rough set theory and rules determined by so-called relative reducts [26],[40],[60] are also often applied. Since a number of rules generated in such a way is usually very big, the various types of postprocessing methods are used in order to select the most interesting rules. The postprocessing techniques apply, among others, so-called interestingness measures [24],[44] for the assessment of rule quality. In the paper [38] it was presented a wide set of interestingness measures which are applied for evaluation of association rules describing various medical data sets. In the paper, rankings obtained by particular measures were compared with rankings established by a domain expert. Thus, the analysis concerned a subjective agreement between expert's assessments and assessments made by particular interestingness measures.

Until now relatively small number of publications concerning relations between survival analysis and rule inductions algorithms has appeared [5],[34],[41]. In the paper [41] the rough sets were used for identification of main factors that affect survival time of patients. The survival time was considered as a discrete variable with predefined values (e.g. survival-time$\in$[56;73] months).

In the paper [5] the rough sets were applied for rule induction. Before the rule induction, a value of the so-called prognostic index coefficient (PI) is calculated for each example. The prognostic index coefficient calculated based on the Cox's proportional hazard model was applied. A range of PI values is divided so that survival curves determined for observations with values of the PI coefficient belong to different intervals are statistically different. For ranges of decision classes defined in such way decision rules were determined.

The paper [34] presents LASD algorithm for induction of rules from survival data. For this purpose,

the authors adopted the Logical Analysis of Data (LAD) algorithm – the combinatorial approach to rule induction. Induction of survival trees [7] can be also mentioned as correlated works.

It is worth to mention that in medical applications it is very important to determine rules containing information useful for a user [28]. Until now, a way of extending the rule induction algorithm by the possibility of taking into account user's requirements is not described in so many papers. Examples of rules interactive building [43] and rule induction based on user's defined patterns that contain preferences concerning attributes or premises of which rules are composed [11],[39],[53], are presented. Gamberger and Lavrac [22] present a similar proposal for rule-based subgroup discovery.

At the end of this review, it is worth mentioning that medical data are usually unbalanced. For unbalanced data one class (a minority class) is under-represented in comparison to the remaining majority classes. The minority class is usually of primary interest. In the machine learning algorithms there are used various techniques to balance the distribution between classes [10] or adjust the induction algorithm to the unbalanced data [46],[48]. In particular, the use of rule quality measures taking into account information about the distribution of examples allows for improving sensitivity and specificity of the rule-based classifier [46],[48].

3. METHODS

3.1. DECISION RULES

Let us assume that a finite training set Tr of examples is given. Each training example is described by a set $A \cup \{d\}$ of features, i.e. $a : Tr \to Va$ for each $a \in A \cup \{d\}$. The set Va is called the range of the attribute a. Elements of A are called conditional attributes, the variable d is called a decision attribute, and its value is identified with assignment of an example to a specific concept (decision class). Conditional attributes can be of symbolic (discrete-valued) or of numeric (real-valued) type. The decision attribute is of symbolic type. Each example x which belongs to the Tr set can be written as the vector $x = (x_1, x_2, .., x_{|A|}, y)$, where $x_i = a_i(x)$ for each $i \in \{1, 2, .., |A|\}$ and $y = d(x)$. The conditional expression of the form (1) is called a decision rule:

$$\textbf{if } w_1 \textbf{ and } w_2 \textbf{ and } \ldots \textbf{ and } w_k \textbf{ then } d = v. \tag{1}$$

An example which is covered (i.e. which satisfies all elementary conditions w_i) by the rule is assigned to the concept indicated in the rule conclusion. The positive examples are those which belong to decision class pointed in the rule conclusion. The negative examples are all the others. Construction of elementary conditions w_i may be various and depends on a rule induction algorithm. Any elementary condition is most often the expression of the form $w_i \equiv a_i \ op \ Z_i$, where a_i is the name of the conditional attribute, op is one of the relation operators $\{=, \epsilon, \leq, \geq, >, <\}$, Z_i is the range of elementary condition satisfying $Z_i \subset Va_i$. The decision rule reflects relationship between values of conditional attributes and assignment to the decision class. However, decision attribute is often absent in the case of survival data. The examples of survival data are usually described by a set $A \cup \{t, s\}$ of attributes, where A is a set of conditional attributes, t is survival time and s is survival status. The survival time attribute contains information about the time that has elapsed since the beginning of observation (for example, bone marrow transplantation). The survival status specifies whether the event (death) occurred or not. In this paper, we assume right-censored model of data which is the most common one. Right-censored examples are those for which event was not observed and their lifetime is greater than recorded time. We are interested in which factors influence survival time shorter and longer than given time T. Therefore, the examples of entire data set are categorized into decision classes according to their survival time and status by the following rules:

- if survival time of example is less than T and event occurred then the example is assigned to the decision class which is denoted in this paper as $dead$,
- if survival time of example is less than T and event did not occur (i.e. the example is censored) then the example is assigned to the $alive$-T class,

- if survival time of example is greater than or equal to T then example is assigned to the $alive$ class.

Note that the $alive$ decision class contains both censored and not censored observations.

The training set which is used for rule induction does not contain survival time and status attributes (they are replaced by decision attribute). Moreover, it consists only of examples from the $dead$ and $alive$ decision classes, because it is not known whether examples from the $alive$-T class will survive (or not) longer than time T. However, the $alive$-T examples are used in further part of analysis – during the filtration of rules (section 3.5). In the literature there are described the methods of the inclusion of examples from the $alive$-T class to the set of training examples. Generally one can say that these methods assign examples from the $alive$-T class to the $alive$ class. This assignment is usually made by giving to each example a weight reflecting its opportunity (not necessarily understood as the probability) to be found in the $alive$ set [51],[63]. However, for the data set considered in the paper the attempt at assignment weights to the examples failed. The obtained results were worse than in the case of usage only examples belonging to the $alive$ and $dead$ classes during the rule induction.

3.2. RULE AND RULE SET QUALITY EVALUATION

Let r be a decision rule induced from the training set Tr. Let p denote the number of positive examples from Tr covered by the rule r (let P stand for all positive examples in the training set), and n denote the number of negative examples covered the rule r (N stands for all negative examples). The aim of most rule induction algorithms is to generate rules characterized by as high values of p (i.e. with high coverage) and as small values of n (i.e. with high consistency) as possible. However, the increase in value of p usually goes hand-in-hand with the increase in value of n, therefore, the rule induction algorithms usually use some rule learning heuristics which allow finding trade-off between coverage and consistency. In most cases, this trade-off is determined by the rule quality measures. Roughly speaking, a rule induction algorithm can be seen as the method of generating rules which are characterized by high values of selected quality measure. The rule quality measure decides about the form of generated rules and therefore affects the number of induced rules as well as their classification abilities.

Two basic rule quality measures are $precision(r) = p/(p+n)$ and $coverage(r) = p/P$. $Precision$ is the fraction of positive examples among all examples covered by the rule. $Coverage$ expresses what percentage of all positives examples are covered by the rule. These two measures are appropriate to reflect the quality of the rule (the higher the values of $precision$ and $coverage$ are the better is the rule), however, they are not very suitable for rule induction. $Precision$ tends to generate a large number of rules that overfit training data [29],[45],[46],[48], and $coverage$ does not take into account the number of negative examples covered by the rule. For this reason, most rule induction algorithms use more sophisticated rule quality measures.

At present, in literature on the subject over 50 objective well-known rule evaluation measures can be found [3],[8],[21],[24],[25],[35],[44],[59]. For decision rules the measures function as quality measures [3],[8],[45],[46],[48],[59] and evaluation or search heuristics [21],[35]. A lot of measures used for evaluation of decision rules function also more widespread as rule interestingness measures (attractiveness measures) [24],[25],[38],[44]. Interestingness measures are used for evaluation of already induced rules, both decision and association rules to be evaluated. In the next part of the paper we will use the term quality measures.

The quality measure applied in a rule induction algorithm has big importance for quality of the induced rule set. This is confirmed by numerous empirical researches [3],[8],[29],[45],[46],[48]. To date, no consistent recommendation allowing us to indicate what quality measure guarantees obtaining the best rule classifier based on characteristics of a training data set has been presented. Studies on this problem were undertaken [3],[29],[45] but results obtained up till now are not better than results got by arbitrary selection of the quality measure.

In our rule induction algorithm we use [46],[48] an adaptive mechanism of the quality measure selection that is matched with characteristics of the analyzed data set. The adaptation consists in

application of internal cross-validation that enables us to verify efficiency of various measures on a data set not influenced by a test set. Due to increased computational overhead, application of all known quality measures was impossible. On the basis of our previous studies [45],[46],48] and other works [3],[8],[29] we selected four quality measures (presented in Table 1) that led to obtaining the rule classifier with good quality. For two-class classification problems, classifiers obtained based on measures contained in Table 1 were characterized by high sensitivity and specificity.

Table 1. Rule quality measures applied in rule induction.

$Correlation(r) = \frac{pN - Pn}{\sqrt{PN(p+n)(P-p+N-n)}}$	$RSS(r) = \frac{p}{P} - \frac{n}{N}$
$CN2(r) = 2\left(p \ln\left(\frac{(P+N)p}{(p+n)P}\right) + n \ln\left(\frac{(P+N)(n+1)}{(p+n)N}\right)\right)$	$s(r) = \frac{p}{p+n} - \frac{P-p}{P-p+N-n}$

The $Correlation$ measure computes the correlation between the predicted and the target labels. It was applied to rule induction algorithms as well as to subgroup discovery and association rule evaluation [21],[24],[29].

The $CN2$ measure is also known as the $J - measure$. Originally, this measure was proposed by Smyth and Goldman [50], by describing the rule induction algorithm ITRULE. The measure is derived from information theory. It can be decomposed into two components, one of which evaluates the rule complexity, and the second measures the difference between the distribution of the number of positive and negative examples in the whole training set, and in the set of examples covered by the rule. In this way, the idea of the $J - measure$ is consistent with the minimum description length principle, because it measures both complexity and $precision$ of the rule. In the $CN2$ algorithm the measure is used to eliminate irrelevant rules.

The RSS measure is a measure equivalent, in terms of the rules order, to the well-known Weighted Relative Accuracy measure used by Lavrac and Flach [35] both in rule induction and subgroup discovery. However, RSS and WRA have different ranges of values. The use of RSS for classification conflicts resolving leads to better results than the use of the WRA measure [46,48].

The $s\text{-}Bayesian\ confirmation\ measure$ has been proposed by Christensen [12] as a confirmation measure. The first component of the measure evaluates the rule $precision$, the second is responsible for decrease of the quality of rules that cover small number of examples. Its application to evaluation of decision rules obtained by the rough set theory was considered in papers [25].

In this paper we introduce one more quality measure for the rule – the $Survival\ Difference$ $(SurvDiff)$. This measure relates rule-based analysis with two fundamental methods of survival data analysis: the estimation of survival function by Kaplan-Meier (KM) [30] method and the comparison of KM survival curves between two groups of observations. In our study these groups are determined by each of the induced rules. As each of the rules divides the analyzed data set into two groups of examples - covered by the rule and not covered by the rule - it is interesting from the survival analysis point of view to answer the question whether the difference between KM curves of such groups is statistically significant. Therefore, we associate each rule with the value of $SurvDiff$ measure which is equal to a confidence level of log-rank test between the KM curve of observations covered by the rule and the KM curve of observations not covered by the rule. The higher value of $SurvDiff$ is the greater difference between these two KM curves.

The quality of rule-based classifier is usually measured with its performance achieved on a data set independent of the training set. Two main criteria for evaluating the classifier are: overall accuracy and balanced accuracy. The overall accuracy is the ratio of the number of correctly classified examples to the number of all examples. This is one of the most common criteria for assessing a classifier. However, in the case of unbalanced distribution of examples between decision classes, higher value of overall accuracy is often achieved at the cost of low accuracy of minority class (-es), therefore in such case balanced accuracy is more appropriate [4]. It calculates classification accuracy of each decision class

and then takes an average over all classes. In the case of the balanced accuracy, the proper classification of examples is equally important for each decision class, no matter how large it is. For two-class problems and the discrete rule-based classifier the maximization of the balanced accuracy is equivalent to maximization of the Area Under the ROC Curve (AUC) [4],[17].

3.3. RULE INDUCTION

Finding the minimal set of classification rules which cover the given set of examples and correctly predict their decision classes is a computationally expensive task, therefore most of the rule induction algorithms use some heuristics. One of the most common approaches is sequential covering (known also as separate-and-conquer) strategy [20],[37]. To put it briefly, this strategy consists in learning a rule which covers some part of training examples, next, the examples covered by the learned rule are removed from the training set and the rule learning process starts recursively for remaining examples.

Our implementation of the rule induction algorithm also works in the separate and conquer fashion. The outcome of the algorithm is a set of rules describing each decision class of the training set. The process of induction of a single rule consists of two phases: growing and pruning. In the growing phase, the elementary conditions are added one by one to the premise of the rule. In the case of nominal attributes the elementary condition can take the form of $(a = v)$, and for the numerical attributes it can take one of two forms: $(a < v)$ or $(a > v)$. For the numerical attributes the value v is the arithmetic mean between two successive values from the range of attribute a. The set of all the possible elementary conditions which might be added to the rule is created on the basis of examples currently covered by the rule. It means that the domains of the attributes within which the ranges of elementary conditions are determined are narrowed to the values which are taken by the examples covered by the currently formed rule. Moreover, the elementary condition is tested only if its addition to the rule causes that such a refined rule covers at least one positive example which is not covered by rules generated so far. The refinement which has the highest value of the specified rule quality measure among all refinements possible in a single step is selected as the final one.

During the rule induction it should be taken into account that not always all attribute values of individual examples are known. In literature several proposals of strategies for handling unknown values of the attributes in rule learning algorithms can be found [27],[62]. In our implementation it is assumed that the elementary condition always returns the value $false$ for the unknown value of the attribute. It means that if the example has unknown values then it can be covered only by the rule which does not contain attributes which values are unknown for this example.

The process of rule growing is terminated if the rule is accurate (i.e. it covers none of the negative examples) or if the addition of the next conditions to the rule no longer increase its precision. After the rule growing phase, the rule is pruned. There are deleted the elementary conditions if their removal does not cause decrease in the current value of rule quality measure. The rule pruning algorithm uses a hill climbing strategy. The pruned rule is added to the final set of rules, and then the process of rule induction is started again for the rest of the uncovered positive examples.

The workflow of the algorithm can be summarized as follows:

```
RuleInduction(examples, ruleQualityMeasure)

# the set of generated rules (initially empty)
ruleSet := ∅

# decisionClass is a set of examples which have the same value of decision attribute
foreach (decisionClass in examples)

   uncoveredPositives := decisionClass
   while (uncoveredPositives ≠ ∅)
```

```
    # rule with empty antecedent and consequent pointing current decision class
    rule := ∅

    # the set of examples covered by the rule (initially empty rule covers all examples)
    covered := examples

    # rule growing phase
    do
     conditions := PossibleElementaryConditions(covered)

     bestQuality := −∞
     bestCondition := ∅
     foreach (c in conditions)
      if (|Covered(rule ∪ c, uncoveredPositives)| > 0)
       # evaluate the quality of rule with c condition added to its antecedent
       quality := Evaluate(rule ∪ c, ruleQualityMeasure)
       if (quality > bestQuality)
        bestQuality := quality
        bestCondition := c
       end if
      end if
     end foreach

     # add selected elementary condition to the antecedent of the rule
     rule := rule ∪ bestCondition
     covered := Covered(rule, examples)
    until (stop criterion)

    # rule pruning phase
    rule := Prune(rule, examples, ruleQualityMeasure)

    covered := Covered(rule, examples)
    uncoveredPositives := uncoveredPositives \covered

    ruleSet := ruleSet ∪ rule
   end while
end foreach

return ruleSet
```

Depending on implementation, the induced rules can be unordered or ordered within a decision list. The described rule induction algorithm returns the set of unordered rules. The main difference between unordered rules and the decision list lies in the method of classification of examples. In the case of the decision list the rules are tested one by one according to the established order. The decision class pointed by the first rule which covers test example is assigned. During the classification of examples with the use of unordered set of rules it may happen that the test example is covered by rules describing different decision classes. In that case a strategy for resolving such conflicts has to be chosen. The most popular one is known as the weighted voting scheme and consists in assigning a numeric value called the confidence degree to each rule. Confidence degrees of the rules covering the test example are summed up for each decision class and then a class with the maximal confidence degree is picked. In this paper the weighted voting scheme is also used during classification. In the classification mechanism, which

we used, the confidence degree of each rule is equal to its value of quality measure used during the rule induction. This allows achieving good results of the classification [48],[54]. If the test example is not covered by any of the rules, then it is always treated as wrongly classified (in the case of the decision list such an unrecognized example is often classified to the majority class).

3.4. USER-DRIVEN RULE INDUCTION

The user does not have any influence on which attributes and their values will be used by the rule induction algorithm. It is not difficult to imagine the situation when according to the user an important attribute will not appear in any of generated rules, therefore there is a need of extending the rule induction algorithm by the possibility of taking into account user's requirements according to which attributes or elementary conditions should be included in the final rule set. Due to that we have extended the rule induction algorithm by the possibility of taking into account user's preferences, which may take one of the following forms:

O1. The user gives the form of expert rules:
   (a) the algorithm conducts their evaluation on the whole training set or in cross-validation mode; there are neither added any elementary conditions to the rules, nor generated any new rules;
   (b) the algorithm redefines the rules given by the user (which means that it tries to add elementary conditions to each of them, so as to maximize the value of specified rule quality measure) and then it conducts their evaluation on the whole training set or in cross-validation mode,
   (c) the algorithm specifies the rules given by the user and then it generates additional rules for the rest, uncovered examples.

O2. The user gives the form of elementary conditions which should appear in at least one rule from each decision class.

O3. The user gives the attributes which should appear in at least one rule from each decision class.

The option O1a is aimed at verification and evaluation of user's hypotheses. The hypotheses have to be expressed in the language of decision rules. The O1b option additionally allows for redefinition of given by the expert rules in order to maximize the value of specified rule quality measure. If a set of rules obtained with the option O1a or O1b is evaluated, then the overall and balanced classification accuracies are calculated on the basis of examples covered by the rules only. In other words, if the test example is not covered by any of the rules, then it is neither counted as correctly, nor wrongly classified. In this case, the important information is how many test examples were unrecognized by the classifier. High classification accuracy may be achieved at the expense of low coverage of the test set. For the options O1c, O2 and O3, the overall and balanced classification accuracies count unrecognized examples as wrongly classified.

For the options O2 and O3 the user may additional impose the following requirements for the rules of each decision class:

- all user's attributes/elementary conditions should appear in the rule simultaneously,
- all user's attributes/elementary conditions should appear in the rule set,
- at least one of user's attributes/elementary conditions should appear in the rules.

User's requirements are examined in the initial phase of rule induction. The algorithm first tries to generate rules which meet user's requirements, and then (except for the options O1a and O1b continues rule induction for the rest of uncovered positive examples. The rule induction algorithm takes into account simple requirements (i.e. defined only by one the option O1, O2 or O3, as well as complex requirements, which are a combination of the options O1c, O2 and O3. An example of a complex requirement is a combination of the options O1c with O2 and/or O3. In this case, the algorithm starts induction with a redefinition of rules, and if redefined rules do not meet requirements O2/O3 then additional rules satisfying O2/O3 are generated. If the options O2 and O3 are used together, then the algorithm first tries to generate rules satisfying O2, and then if such rules do not meet requirements O3, then the algorithm generates additional rules satisfying O3.

3.5. RULE FILTRATION

The sequential covering strategy does not prevent the occurrence of rules which cover similar subsets of training examples and therefore generated sets often contain redundant rules. In order to reduce the rule number, the filtration algorithms [1],[45],[48] can be applied. During filtration, rules which are unnecessary on the grounds of some criterion are removed from the input rule set. Reduction of the number of rules is of great importance for the clarity of the rule-based model.

In the case considered here, the application of the decision rule filtration algorithms [1],[45],[48] is not possible because of the algorithm's inability to analyze examples from the $alive{-}T$ class. Therefore, in order to summarize knowledge expressed by the rule-based model of survival data, a new filtration algorithm has been proposed with the following assumptions about the output rule set (i.e. rules after filtration):

C1. The rule set contains rules which value of $SurvDiff$ measure is greater than α.
C2. The rule set covers at least $\beta \cdot 100\%$ of all available examples.
C3. Removal of any rule from the set causes the C2 criterion not to be met.

The subset of rules satisfying criteria C1, C2 and C3 will be called the $\alpha/\beta\ Survival\ Model$. The $\alpha \in [0, 1)$ and $\beta \in (0, 1]$ parameters have to be defined by the user. The criterion C1 narrows the search space of induced rules to these which are statistically significant according to the log-rank test. The criterion C2 reduces the number of rules at the expense of decreased coverage of the analyzed data set – the lower the required coverage, the lower the number of output rules. The criterion C3 means that we are interested in minimal sets of rules. In view of these assumptions, depending on the parameters specified by a user, we can encounter the following situations:

- The $\alpha/\beta\ Survival\ Model$ does not exist.
- There is exactly one $\alpha/\beta\ Survival\ Model$.
- There are many different $\alpha/\beta\ Survival\ Model$.

In the first case, the user can reduce the requirements for minimum coverage and/or statistical significance of the rules. In the third case, it is necessary to take additional criterion into account in order to choose one model. Among all $\alpha/\beta\ Survival\ Model$, the model which maximizes the value of criterion (2) is chosen.

$$\sum_{r \in SM} |Cov(\{r\}, E) - Cov(SM - \{r\}, E)| \tag{2}$$

The expression $Cov(SM, E)$ in the formula (2) is equal to $\bigcup_{r \in SM} Cov(\{r\}, E)$, and it denotes the set of examples from E covered by the rules belonging to the survival model SM. The formula (2) can be treated as an additional criterion (C4) for rule selection. The criterion C4 allows for selection of rules which describe the most disjoined groups of examples.

Finding the optimal set of survival patterns satisfying criteria C1-C4 is computationally expensive and it comes down, inter alia, to solving classical NP-complete task which is the minimum set cover problem. Therefore, the following heuristics of $\alpha/\beta\ Survival\ Model$ building has been proposed:

```
SurvivalModelSelection(rules, α, β, q, E)

    # q - rule quality measure
    # E - the set of examples

    # a-rule means the rule pointing to the alive decision class
    # d-rule means the rule pointing to the dead decision class

    αrules := decision rules with SurvDiff greater than α

    if |Cov(αrules,E)|/|E| < β return ∅
```

```
# the set of selected survival rules

SM := initialize with a-rule and d-rule with the highest value of quality measure q among
all αrules

αrules := αrules - SM

while |Cov(SM,E)|/|E| < β
r := find in αrules such a rule r that |Cov({r},E) − Cov(SM,E)| is the largest
SM := SM + r
αrules := αrules - r
end while
SM := remove rules r for which |Cov(SM−{r},E)|/|E| ≥ β
return SM
```

The algorithm starts with the removal of these rules from the input set of rules which are not statistically significant. If the coverage of such selected rules is less than specified by a user, then empty rule set is returned. Otherwise, the output set of rules is initialized with the best (according to the chosen rule quality measure) rules from each decision class (one rule for each class). Then, rules which maximize the value of (2) (i.e. rules which cover the highest number of examples yet uncovered by the created so far set of rules) are successively added to the output set of rules. Rules are added until they will not obtain the coverage specified by a user. It is worth noticing that in the set of rules obtained in this way may be found redundant rules, i.e. their removal does not cause that the criterion C2 is not satisfied. In the last stage of the presented filtration algorithm such rules are removed. As the final step, the KM curves are calculated for each rule from the obtained α/β $Survival$ $Model$.

3.6. DECISION RULE-BASED FRAMEWORK FOR SURVIVAL DATA

At the end of this section we briefly summarize previous subsections by the workflow of proposed methodology (Figure 1).

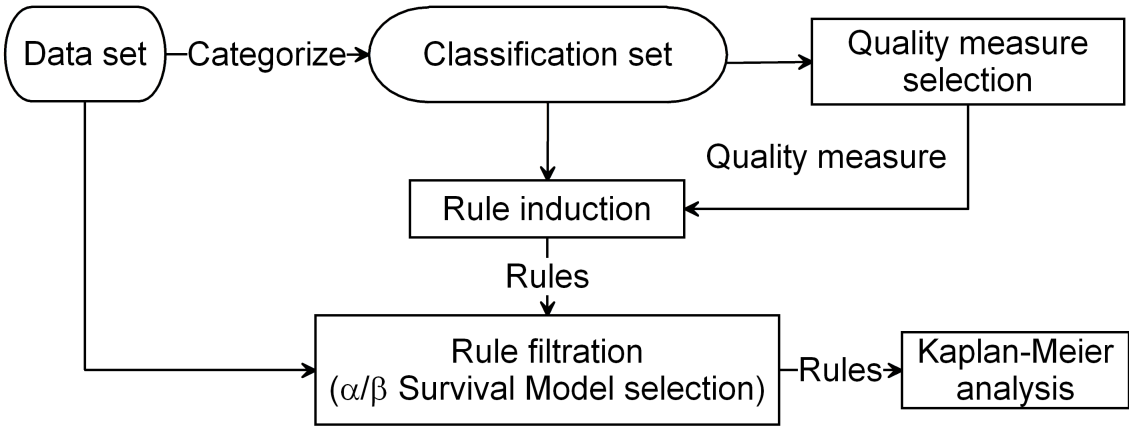

Fig. 1. The workflow of decision-based analysis of survival data.

The examples of input data set are categorized into decision classes: $dead$, $alive$-T and $alive$. The examples from $dead$ and $alive$ classes create classification set. In order to choose suitable rule quality measure for analyzed data, the rule induction is preceded by the adaptive selection of quality measure [46],[48]. The selection consists in estimating the performance of rule-based model with the use of cross-validation for each quality measure under consideration. In cross-validation method the classification data set is divided in a random way into k disjoint subsets of almost the same cardinality. Next, k

experiments are carried out. In each experiment, one of k subsets is a test set and the union of the remaining $k-1$ subsets is the training set. After each experiment, the performance of model is evaluated by the formula (3).

$$MeanSD \cdot BAC \tag{3}$$

In the expression (3) the $MeanSD$ is an average value of $SurvDiff$ measure of induced rules. BAC is the value of balanced accuracy achieved by the classifier on a test set. The criterion defined in such a way takes into account both classification abilities and survival importance of the rule-based model.

After carrying out k experiments, the final value of classifier performance is computed as an arithmetic mean of the performances from all experiments. The measure which obtains the highest performance is selected as the final one and used for the rule induction on the whole data set. The next step is filtration of the obtained rules. The filtration uses also examples form the $alive$-T decision class. Additionally, the final rule set is supplemented with the Kaplan-Meier analysis.

If the form of rules obtained automatically does not contain information useful from the diagnostic perspective, the next step of the analysis is the user-driven rule induction. The classification accuracy of the rules obtained automatically will be a reference to the rules defined by the expert or redefined by the user-driven rule induction algorithm. Useful information obtained on the basis of fully automatic induction is also a set of optimal parameter values (in the case of our algorithm it is a quality measure) supervising the process of rule induction and redefinition.

4. DATA ANALYSIS

The analyzed data set describes 187 patients (75 females and 112 males) at the age of 0.6 to 20.2 years (median 9.6) admitted to the Department of Pediatric Bone Marrow Transplantation, Oncology and Hematology, Wrocław Medical University, Poland. Disease spectrum included 155 malignant disorders (i.a. 67 patients with acute lymphoblastic leukemia, 33 with acute myelogenous leukemia, 25 with chronic myelogenous leukemia, 18 with myelodysplastic syndrome) and 32 nonmalignant cases (i.a. 13 patients with severe aplastic anemia, 5 with Fanconi anemia, 4 with X-linked adrenoleukodystrophy). The procedure of unmanipulated allogeneic unrelated donor hematopoietic stem cell transplantation was performed in each case according to the European protocols or the guidelines of the European Blood and Marrow Transplant Inborn Errors Working Party with worldwide accepted modifications based on disease and/or patient's condition status prior transplantation. Each patient was characterized by a set of 42 conditional attributes. Table 2 presents interpretations of selected ones. The motivation of this study was to identify the most important factors influencing the success or failure of the transplantation procedure. In particular, verification of the research hypothesis that increased dosage of CD34+ cells / kg expands overall survival time without simultaneous occurrence of adverse events affecting patients' quality of life. For this purpose, the data set was categorized into three classes: the patients for whom at least 5 years have passed since the transplantation (the class $alive$), the patients who died within 5 years after transplantation (the class $dead$), and the patients who are still alive but their survival time is less than 5 years (the class $alive$-5). The distribution of examples between decision classes was the following: 36 examples from the class $alive$, 85 from the class $dead$, and 66 other examples.

4.1. CLASSIFICATION RESULTS

The data set used for construction of rule-based classifier consists of examples from $alive$ and $dead$ decision classes (121 examples in total). Obtained results are presented in Table 3. The results are an average from 100 experiments (i.e. stratified 10-fold cross validation repeated 10 times). The first column is the name of quality measure which was used during rule induction. The next columns are: the number of generated rules, the mean number of elementary conditions per rule ($Avg\ cond$), the mean

Table 2. Selected conditional attributes of analyzed data set.

Name	Description
Recipient_Rh	Presence of the Rh factor on recipient's red blood cells
Recipient_age	Age of the recipient of hematopoietic stem cells at the time of transplantation
Recipient_body_mass	Body mass of the recipient of hematopoietic stem cells at the time of transplantation
Recipient_CMV	Presence of cytomegalovirus infection in the recipient of hematopoietic stem cells prior to transplantation
Donor_CMV	Presence of cytomegalovirus infection in the donor of hematopoietic stem cells prior to transplantation
CMV_status	Serological compatibility of the donor and the recipient of hematopoietic stem cells according to cytomegalovirus infection prior to transplantation
Recipient_ABO	ABO blood group of the recipient of hematopoietic stem cells
Donor_ABO	ABO blood group of the donor of hematopoietic stem cells
ABO_match	Compatibility of the donor and the recipient of hematopoietic stem cells according to ABO blood group
Gender_match	Compatibility of the donor and recipient according to their gender
Donor_age	Age of the donor at the time of hematopoietic stem cells apheresis
HLA_match	Compatibility of antigens of the main histocompatibility complex of the donor and the recipient of hematopoietic stem cells (10/10, 9/10, 8/10, 7/10 allele/antigens) according to ALL international BFM SCT 2008 criteria
Stem_cell_source	Source of hematopoietic stem cells (bone marrow or peripheral blood)
Relapse	Reoccurrence of the disease
PLT_recovery	Time to platelet recovery defined as platelet count > 50000/mm3
ANC_recovery	Time to neutrophils recovery defined as neutrophils count > 0.5 x 10^9/L
T_aGvHD_III_IV	Time to development of acute graft versus host disease stage III or IV
extcGvHD	Extensive chronic graft versus host disease
CD34 (10^6/kg)	CD34+ cell dose per kg of recipient body weight
CD3 (10^8/ kg)	CD3+ cell dose per kg of recipient body weight
CD3/CD34	CD3+ cell to CD34+ cell ratio

value of the rule precision ($Avg\ prec$) and rule coverage ($Avg\ cov$), the classification accuracies of each decision class ($Acc\ alive$, $Acc\ dead$) in percentages, the mean value of $MeanSD \cdot BAC$ criterion.

Table 3. Characteristics and classification accuracy of obtained rules.

Measure	Rules	Avg cond	Avg prec	Avg cov	Acc alive	Acc dead	MeanSD · BAC
$Correlation$	19.5	4.4	0.903	0.539	61.8	71.8	0.65
$CN2$	32.0	3.0	0.995	0.266	62.3	68.4	0.63
RSS	22.8	3.9	0.888	0.501	41.7	82.0	0.60
s	24.5	2.6	0.988	0.154	57.3	64.2	0.55

The results presented in Table 3 show that classifiers composed of various numbers of rules and with different classification abilities can be obtained depending on applied rule quality measure. The rule quality measure influences the form of induced rules (what is reflected by different number of rules, elementary conditions and difference in mean precision and coverage of rules) as well as classification accuracy of $dead$ and $alive$ decision classes. The best result according to $MeanSD \cdot BAC$ criterion was obtained by the $Correlation$ measure and therefore it was selected as the final measure for further analysis. For comparison purposes the classification accuracies obtained by well-known rule-based ($MODLEM$ [52], $JRip$ [14], $PART$ [61]) and tree-based ($J48, SimpleCart$) algorithms from $Weka$ [61] package are given in Table 4. The columns BAC and Acc respectively present the values of balanced and overall classification accuracies (in percentages).

Table 4. Classification accuracy of selected algorithms from the $Weka$ package.

Algorithm	Acc alive	Acc dead	BAC	Acc
$MODLEM$	$29.6 \pm 26^{(*)}$	$85.6 \pm 11.8^{(+)}$	57.6 ± 14.2	68.9 ± 11.6
$JRip$	49.8 ± 31.3	82.1 ± 13.7	66.0 ± 13.0	72.6 ± 8.9
$PART$	$39 \pm 22.7^{(*)}$	75.6 ± 13.3	57.3 ± 11.9	64.6 ± 10.5
$J48$	$30.9 \pm 22.8^{(*)}$	75.3 ± 14.1	$53.1 \pm 12.6^{(*)}$	62.1 ± 11.4
$SimpleCart$	$1.3 \pm 8.5^{(*)}$	$98.8 \pm 6.6^{(+)}$	$50.1 \pm 2.8^{(*)}$	69.8 ± 5.1
$Correlation$	61.8 ± 28.7	71.8 ± 15.1	66.8 ± 14.4	68.6 ± 12.0

Results marked with $^{(*)}/^{(+)}$ show statistically significant degradation/improvement over results obtained by our algorithm in which $Correlation$ measure was used. The results were compared according to a paired two-sided corrected resampled t-test [61] at 0.05 significance level. In only two cases our approach was significantly worse – in comparison with $MODLEM$ and $SimpleCart$ algorithm for Acc $dead$ criterion. However, both the algorithms are characterized by low accuracy of the $alive$ class what makes them not suitable for this data. The classification results obtained by our approach are one of the best results and they are comparable to the results of the $JRip$ algorithm. However, $JRip$ seems to prefer higher accuracy of the $dead$ class at the cost of lower accuracy of the $alive$ class. The results for the $Correlation$ measure are characterized by smaller difference in accuracy between decision classes, which is more preferable in our case when misclassification costs are not possible to determine.

4.2. RULE FILTRATION AND KAPLAN-MEIER ANALYSIS

The $Correlation$ measure, selected in previous section, was applied for rule induction on the whole available data set. Eighteen rules (7 for $alive$ class and 11 for $dead$ class) which classify the training set with the accuracy of 92.6% (the class $dead$ 89.4%, the class $alive$ 100%) were obtained. In order to summarize knowledge expressed by these rules the $0.999/0.8$ $Survival$ $Model$ was generated. The rules included in the model were as follows:

R1. **if** $(extcGvHD = No)$ **and** $(Relapse = No)$ **and** $(Donor\_age \in [26.4; 46.1))$ **and** $(PLT\_recovery = patient\ achieved\ platelet\ recovery)$ **and** $(T\_aGvHD\_III\_IV \geq 14.5\ or\ acute\ graft\ was\ not\ developed)$ **and** $(ANC\_recovery < 24)$ **and** $(Recipient\_body\_mass < 73.5)$ **then** alive
precision=0.8621, coverage=0.6944, SurvDiff=1.0, 5-year mean=4.71, #covered=64(4)

R2. **if** $(PLT\_recovery \geq 13.5)$ **and** $(Relapse = No)$ **and** $(Recipient\_age < 17.6)$ **and** $(Recipient\_body\_mass < 72)$ **and** $(Donor\_age < 45.5)$ **and** $(T\_aGvHD\_III\_IV \geq 14.5\ or\ acute\ graft\ was\ not\ developed)$ **and** $(Gender\_match = not(F\ to\ M))$ **then** alive
precision=0.7073, coverage=0.8056, SurvDiff=1.0, 5-year mean=4.31, #covered=76(12)

R3. **if** $(CD34 \in [1.265; 10.815))$ **and** $(Recipient\_age \geq 11.6)$ **and** $(Donor\_age \geq 20.5)$ **then** dead
precision=0.95, coverage=0.4471, SurvDiff=0.999996, 5-year mean=1.93, #covered=56 (38)

R4. **if** $(Relapse = Yes)$ **and** $(Recipient\_Rh = Rh+)$ **then** dead
precision=1.0, coverage=0.2235, SurvDiff=0.999864, 5-year mean=1.36, #covered=20 (19)

At the bottom of each rule the following values are given: the $precision$ and $coverage$ of the rule, the value of $SurvDiff$ measure (rounded to six significant digits), the value of restricted mean with upper limit set to 5 years (i.e. expected number of years, out of the first 5 years, that would be experienced by group covered by the rule), the number of examples covered by the rule with the number of deaths in the parentheses. In the calculation of $SurvDiff$, 5-year mean and the number of covered examples the entire data set was used (i.e. examples from the $alive$-5 class are also taken into account); the values of $precision$ and $coverage$ measures were evaluated on the whole available data set.

Let us illustrate briefly how to interpret knowledge expressed by decision rules. For example, the rule R2 describes patients who achieved platelet recovery after 13 days, until now they did not have disease relapse, they are less than 17.6 years old, their weight is less than 72 kilograms, the donor age is less than 45.5 years, an acute graft versus host disease stage III or IV was not developed or it was developed no earlier than the 15^{th} day after transplantation and presumptive sex incompatibility between a donor and recipient was not the incompatibility of the type Female to Male.

Figure 2 presents the KM survival curves for the groups of observations covered by the rules R1-R4. Additionally, the graph also shows the KM survival curve (named as 'default') for the whole data set. Significant differences can be observed between the curves of observations covered by the rules from the $alive$ class and the curves determined by the rules from the $dead$ class. The observations covered by the rules R1 and R2 are characterized by higher survival rate than observations covered by the rules R3 and R4. These differences are also confirmed by the value of 5-year restricted mean which is about twice higher for the rules R1 and R2 than for the rules R3 and R4.

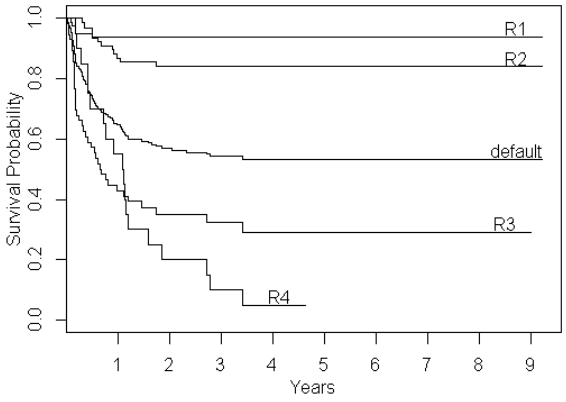

Fig. 2. KM survival curves for observations covered by the rules R1-R4.

For each rule the value of $SurvDiff$ measure is close to 1. It indicates significant differences between KM survival curves of two groups determined by each of the rule. In order to illustrate these differences, the survival curves for R1 and R3 rules are presented in Figure 3. The curves for observations which are not covered by the rule are labeled with overlined identifier of the rule. The censored observations are marked with +.

The KM survival curve for observations covered by the rule R1 is characterized by a slower decrease than for observations which are not covered by this rule. The patients meeting criteria of R1 rule are also more likely to survive the early post-transplant period. Similar conclusions can be drawn for groups determined by the rule R3, but in this case the rule R3 characterizes the group of patients with lower survival rate than the group which is not covered by R3.

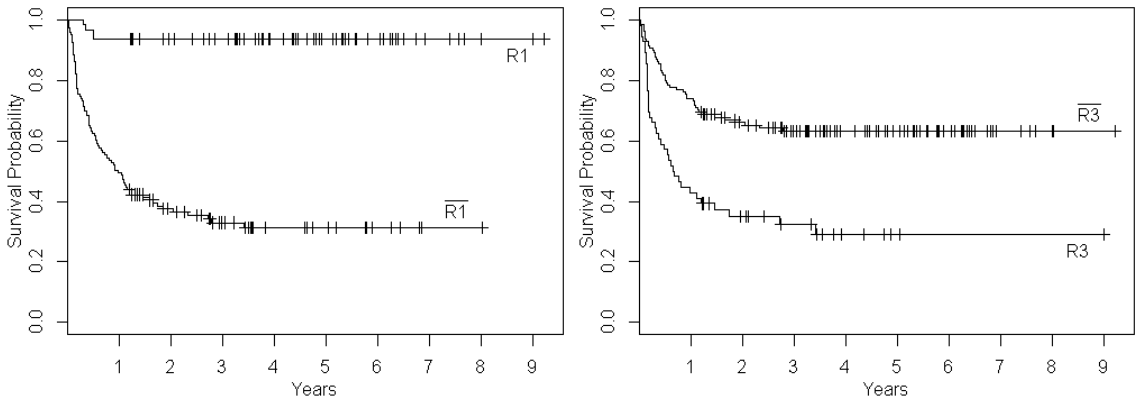

Fig. 3. KM survival curves determined by the rules R1, $\overline{R1}$, and R3, $\overline{R3}$.

The survival curves sketched on the basis of examples covering and not covering the induced rules are characterized by high values of the $SurvDiff$ measure. Additionally, the integrated Brier score was used in order to measure prediction error of survival models. The Brier score measures the error with which survival models predict the probability of patient's survival for a specified time period[61]. In order to use the Brier score to assess the quality of the survival models, these models must be determined on the basis of a training set, and verified on the basis of a testing set. It should be noted that the Brier score is a cost-type criterion i.e. the lower value the more accurate the prediction. Table 5 contains values of the Integrated Brier Score (IBS) [23] obtained after analysis of the considered data set. For comparison purposes, the methods dedicated especially to the analysis of the survival data were also applied. In the analysis there were used: the Kaplan-Meier method [30] and survival tree induction algorithms, included in the R environment [42] for statistical computing. The experiment was on the entire data set (the $alive$, $dead$ and $alive$-T classes) using a stratified 10-fold stratified cross-validation repeated 10 times.

Table 5. Comparison of the Brier score values for survival models obtained based on the survival tree induction and decision rule-based model.

Algorithm	IBS
Kaplan-Meier	0.264± 0.010
RPART	0.364± 0.104
CTREE	0.226± 0.070
Decision rule-based survival model	0.256± 0.028

The obtained results show that although the rule induction did not proceed directly towards the minimization of the Integrated Brier Score, the survival model established by the determined rules is also characterized by the low value of this index.

4.3. THE RESULTS OF USER-DRIVEN RULE INDUCTION

The results presented in the previous sections were obtained by the algorithm which induces rules in fully automatic way. Such generated rules may not always be satisfactory for an expert. In this section the results of rule induction which takes into account user's requirements are presented. The user's requirements were created on the basis of the paper [33] where the hypothesis that increased dosage of CD34+ cells/kg expands overall survival time of patients was verified; the CD34+ cells/kg dosage was divided into smaller and larger than $10 \cdot 10^6$ CD34+ cells/kg dosages (value close to the median). Admittedly, the rules generated with the use of $Correlation$ measure contained CD34+ cells/kg attribute, however it occurred only in the rules for the $dead$ decision class. Therefore, the experiments in which the set of rules was initialized by a user and it contained interesting variable CD34+ cells/kg for both decision classes were performed.

The aim of first experiment was the verification of classification abilities of the following decision rules:

R5. **if** $(CD34 \geq 10)$ **then** alive
precision=0.3922, coverage=0.5556, SurvDiff=0.9873, 5-year mean=3.43, #covered=86 (31)

R6. **if** $(CD34 < 10)$ **then** dead
precision=0.7714, coverage=0.6353, SurvDiff=0.9873, 5-year mean=2.61, #covered=101 (54)

The rules R5 and R6 will be referenced as the base rules in further parts of the analysis. The classification accuracy of decision classes of whole classification set was as follows: 55.6% of the $alive$ class and 63.5% of the $dead$. Interestingly enough, there is no statistical difference between accuracies obtained by the $Correlation$ measure and the base rules (according to paired two-sided corrected resampled t-test at 0.05 significance level for stratified 10-fold cross-validation repeated 10 times). It means that hypotheses formulated by Kałwak [33] et al. have also, beside strong survival importance, good classification abilities which are comparable to hypotheses obtained in automatic way. However, the precision of base rules are much lower than precision of rules generated by the $Correlation$ measure. The rule R5 covers more negative than positive examples, and the rule R6 covers 45.7% of examples from counter-decision class. Therefore, the second experiment consisted in the redefinition of the base rules. The following rules were obtained:

R7. **if** $(CD34 \geq 10.815)$ **and** $(PLT\_recovery \geq 14)$ **and** $(Relapse = No)$ **and** $(Donor\_age < 45.7)$ **and** $(T\_aGvHD\_III\_IV = acute\ graft\ versus\ host\ disease\ stage\ III\ or\ IV\ was\ not\ developed)$ **then** alive
precision=1.0, coverage=0.4167, SurvDiff=1.0, 5-year mean=5.0, #covered=31(0)

R8. **if** $(CD34 \in [1.265; 10.815))$ **and** $(Recipient\_age \geq 11.6)$ **and** $(Donor\_age \geq 20.5)$ **and** $(Recipient\_body\_mass \geq 31.5)$ **then** dead
precision=0.9737, coverage=0.4353, SurvDiff=1.0, 5-year mean=1.84, #covered=53(37)

The redefinition leads to significantly more precise rules. The classification accuracy calculated for the whole available data set also improved: 93.8% of the $alive$ class and 100% of $dead$. However,

56.2% of examples were unrecognized by the classifier built with the rules R7 and R8.

Let us notice the rules R7 and R8 were induced from the whole available data set - rules induced in the cross-validation mode, that is for a changed set of training examples, may differ from ones determined on the whole set. As can be seen in the rules R7 and R8, the ranges of CD34+ cells/kg attribute were set in slightly different way. Moreover, the attribute $CD34$ does not occur alone in any of the rules. Deeper analysis shows that the most important attributes in the rule R7 are: $CD34$, $PLT\_recovery$ and $Relapse$. In other words, the removal of $Donor\_age$ and $T\_aGvHD\_III\_IV$ attributes from the rule R7 does not cause significant changes in a set of examples covered by this rule. Similarly, the removal of $Donor\_age$ and $Recipient\_body\_mass$ attributes from the rule R8 leads to the rule with similar parameters. Additionally, the replacement of the condition $(CD34 \in [1.265; 10.815))$, occurring in the rule R8 without $Donor\_age$ and $Recipient\_body\_mass$ attributes, with the condition $(CD34 < 10.815)$ causes that such modified rule covers 3 examples more - two examples from the $alive$ class and one from the $dead$ class. As the results of above considerations the simplified version of R7 and R8 rules can be presented:

R9. **if** $(CD34 \geq 10.815)$ **and** $(PLT\_recovery \geq 14)$ **and** $(Relapse = No)$ **then** alive
precision=0.8, coverage=0.4444, SurvDiff=1.0, 5-year mean= 4.59, #covered=42(4)

R10. **if** $(CD34 < 10.815)$ **and** $(Recipient\_age \geq 11.6)$ **then** dead
precision=0.8889, coverage=0.4706, SurvDiff= 0.99995, 5-year mean= 2.1, #covered=62(40)

The above examples show that with the use of decision rules it is possible to refine the conditions in which the dosage of CD34+ cells/kg has an impact on overall survival of the patients. Higher dosages of CD34+ are related to longer survival, and lower dosages to shorter survival. The success of the therapy is to be expected among patients with first occurrence of the disease. The relationship between low doses of CD34+, and the transplantation failure has been observed in teenage patients. An important information is also that in the set of examples covered by the rule R9, concentration of patients which has acute graft versus host disease is the same as in the set of examples satisfying only the condition $(PLT\_recovery \geq 14 \; and \; Relapse = No)$. This means that the use of higher doses of CD34+ does not affect the appearance of the negative, acute form of the graft versus host disease. An interesting and strongly confirming the desirability of a higher dose of CD34+ is also the observation that in the set satisfying the condition $(CD34 \geq 10.815 \; and \; PLT\_recovery \geq 14 \; and \; Relapse = No)$ there is about 50% fewer patients with the chronic form of the graft versus host disease than in the set satisfying only the condition $(PLT\_recovery \geq 14 \; and \; Relapse = No)$.

5. CONCLUSIONS

The paper presents an application of the decision rule induction algorithm in knowledge discovery in survival data. The developed methodology was illustrated with the analysis of the data describing patients after bone marrow transplantation. A major role in rule induction and rule set evaluation were played by quality measures. The $SurvDiff$ measure was introduced in order to relate decision rules with survival analysis. The $SurvDiff(MeanSD)$ measure with balanced accuracy (BAC) creates a criterion which allows us to select a model describing significant survival patterns with good classification abilities. The process of knowledge discovery is supported by rule filtration algorithm and by extension of rule induction which enables user to impose requirements on induced rules.

Conducted experiments point out that a properly supervised rule induction process can be the basis not only to obtain interpretable models as a ground for classification but also for the analysis and prediction of survival time. It is able to find accurate rules which divide patients into groups with similar survival rate. However, the output number of rules might be relatively big what makes the interpretation of obtained knowledge difficult. Additionally, the rules obtained within one class may cover similar sets of examples. The proposed rule filtration algorithm helps to overcome these problems. Experiments showed also one more disadvantage of rule-based analysis. In the rule there may appear the elementary conditions removal of which does not cause significant changes in the set of examples covered by the rule. In other words, covering, for example, two more negative examples by the rule may be irrelevant

for a user, but may be relevant for the rule induction algorithm which usually tries to increase rule quality at the cost of greater number of elementary conditions in the rule premise. The user-driven induction combined with manual analysis of such conditions can bring more readable rules.

The obtained results enabled us to identify the most significant survival factors in the form of decision rules. In particular, we refined the conditions in which the dosage of CD34+ cells/kg has an impact on overall survival of the patients. This shows that the procedure presented in this paper may be a source of new and useful medical knowledge.

The experiments described in the paper were carried out using our own software implemented in the C#. Our further works will concentrate on developing an interactive system which will allow for rule-based analysis of survival data. The authors plan to make the system available to a wide circle of users. The first part of the software comprising the rule induction algorithm is integrated with the R package [42], and is available at http://crules.r-forge.r-project.org/. In addition to sequential covering strategy of rule induction, we plan to use the algorithm which generates rules on the basis of so-called quasi-shortest object-related relative reducts [45],[60]. Such a system will lead to more advanced study of pre-transplantation and transplant-related data.

ACKNOWLEDGEMENT

The research of the first author was supported by the National Science Centre based on the decision DEC-2011/01/D/ST6/07007. The second author is a scholarship holder of project "SWIFT" POKL.08.02.01-24-005/10 co-financed by the European Union within the European Social Fund. The authors would like to thank the anonymous reviewer for his/her comments and suggestions to improve the quality of the paper.

BIBLIOGRAPHY

[1] AGOTENES T., KOMOROWSKI J., LOKEN T., Taming large rule models in rough set approaches, Lecture Notes in Artificial Intelligence, 1999, Vol. 1704, pp. 193-203.
[2] AGRAWAL R., SRIKANT R., Fast Algorithms for Mining Association Rules, Proc. of the 20th VLDB Conference, Santiago, Chile, 2004.
[3] AN A., CERCONE N., Rule quality measures for rule induction systems – description and evaluation, Computational Intelligence, 2001, Vol. 17, No. 3, pp. 409-424.
[4] BAIRAGI R., SUCHINDRAN C. M., An estimation of the cut-off point maximizing sum of sensitivity and specificity, Sankhya, The Indian Journal of Statistics, 1989, Vol. 51, No. B-2, pp. 263-26.
[5] BAZAN J., SKOWRON A., ŚLĘZAK D., WRÓBLEWSKI J., Searching for the Complex Decision Reducts: The Case Study of the Survival Analysis, Lecture Notes in Artificial Intelligence, 2003, Vol. 2871, pp. 160-168.
[6] BELLAZZI R., DIOMIDOUS M., SARKAR I. N., TAKABAYASHI K., ZIEGLER A., MCCRAY A. T., Data analysis and data mining: current issues in biomedical informatics, Methods of Information in Medicine, 2011, Vol. 50, No. 6, pp. 536–544.
[7] BOU-HAMAD I., LAROCQUE D., BEN-AMEUR H., A review of survival trees, Statistics Surveys 5, 2011, pp. 44–71.
[8] BRUHA I., TKADLEC J., Rule quality for multiple-rules classifier: Empirical expertise and theoretical methodology, Intelligent Data Analysis, 2003, Vol. 7, pp. 99-124.
[9] CHAVES R., RAMIREZ J., GORRIZ J. M., PUNTONET C. G., Association rule-based feature selection method for Alzheimer's disease diagnosos, Expert Systems with Applications, 2012, Vol. 39, pp. 11766-11774.
[10] CHAWLA N. V., BOWYER K. W., HALL L. O., KEGELMEYER W. P., SMOTE: Synthetic minority over-sampling technique, Journal of Artificial Intelligence Research, 2012, Vol. 16, pp. 321-357.
[11] CHEN S., LIU B., Generating Classification Rules According to User's Existing Knowledge, Proc. of the SIAM International Conference on Data Mining SDM-01, 2001.
[12] CHRISTENSEN D., Measuring confirmation, Journal of Philosophy, 1999, Vol. XCVI, pp. 437-461.
[13] CLARKE E. J., WACLAWIW M. A., Probabilistic rule induction from a medical research study database, Computers and Biomedical Research, 1996, Vol. 29, pp. 271–83.
[14] COHEN W. W., Fast effective rule induction, Proc. of the twelfth Int. Conference ICML95, 1995, pp.115-123.
[15] DAUD N. R., CORNE D. W., Human readable rule induction in medical data mining, Lecture Notes in Electrical Engineering, 2009, Vol. 27, No. 7, pp. 787-798.
[16] DZEROSKI S., LAVRAC N., Rule induction and instance-based learning applied in medical diagnosis, Technol. Health Care, 1996, Vol. 4, No. 2, pp. 203-221.
[17] FAWCETT T., An introduction to ROC analysis, Pattern Recognition Letters, 2006, Vol. 27, pp. 861-874.
[18] FAWCETT T., PRIE a system for generation rulelist to maximize ROC performance, Data Mining and Knowledge Discovery, 2008, Vol. 17, pp. 207–224.

[19] FRANK A., ASUNCION A., UCI Machine Learning Repository [http://archive.ics.uci.edu/ml], 2010.
[20] FÜRNKRANZ J., Separate-and-conquer rule learning, Artificial Intelligence Review, 1999, Vol. 13, pp. 3-54.
[21] FÜRNKRANZ J., FLACH P. A., ROC 'n' Rule Learning – Towards a Better Understanding of Covering Algorithms, Machine Learning, 2005, Vol. 58, pp. 39-77.
[22] GAMBERGER D., LAVRAC N., Expert-guided subgroup discovery: methodology and application, Journal of Artificial Intelligence Research, 2002, Vol. 17, pp. 501-527.
[23] GRAF E., SCHMOOR C., SAUERBREI W., SCHUMACHER M., Assessment and Comparisons of Prognostic Classification Schemes for Survival Data, Statistics in Medicine, 1999, Vol. 18, pp. 2529–2545.
[24] GENG L., HAMILTON H. J., Interestingness measures for data mining: A survey, ACM Computing Surveys, 2006, Vol. 39, No. 3.
[25] GRECO S., PAWLAK Z., SŁOWIŃSKI R., Can Bayesian confirmation measures be useful for rough set decision rules? Engineering Applications of Artificial Intelligence, 2004, Vol. 17, pp. 345-361.
[26] GRZYMAŁA-BUSSE J. W., ZIARKO W., Data mining based on rough sets, in Wang J. (Ed.) Data Mining Opportunities and Challenges, IGI Publishing, Hershey, PA, USA, 2003, pp. 142-173.
[27] GRZYMAŁA-BUSSE J. W., Characteristic relations for incomplete data: A generalization of the indiscernibility relation, Transactions on rough sets IV, LNCS, 2005, Vol. 3700, pp. 58–68.
[28] HAZZANI M. J., MANI S., SHANKLE W. R., Acceptance of rules generated by machine learning among medical experts, Methods of Information in Medicine, 2001, Vol. 40, No. 5, pp. 380–385.
[29] JANSSEN F., FÜRNKRANZ J., On the quest for optimal rule learning heuristics, Machine Learning, 2010, Vol. 78, pp. 343-379.
[30] KAPLAN E. L., MEIER P., Nonparametric estimation from incomplete observations, J. Am. Stat. Assoc., 1958, Vol. 53, pp. 457-481.
[31] KAUFMAN K. A., MICHALSKI R. S., Learning in Inconsistent World, Rule Selection in STAR/AQ18, Machine Learning and Inference Laboratory, 1999, Report No. P99-2.
[32] KAVSEK B., LAVRAC N., APRIORI-SD: Adapting association rule learning to subgroup discovery, Applied Artificial Intelligence, 2006, Vol. 20, pp. 543–583.
[33] KAŁWAK K., PORWOLIK J., MIELCAREK M., GORCZYŃSKA E., Higher CD34+ and CD3+ cell doses in the graft promote long-term survival, and have no impact on the incidence of serve acute or chronic Graft-versus-host disease after in vivo T cell-depleted unrelated donor hematopoietic stem cell transplantation in children, American Society for Blood and Marrow Transplantation. Biology of Blood Marrow Transplantation, 2010, Vol. 16, pp. 1388-1401.
[34] KRONEK L. P., REDDY A., Logical analysis of survival data: prognostic survival models by detecting high-degree interactions in right-censored data, Bioinformatics, 2008, Vol. 24, pp. 248-253.
[35] LAVRAC N., FLACH P. A., ZUPAN B., Rule Evaluation Measures: A Unifying View, Lecture Notes in Artificial Intelligence, 1999, Vol. 1634, pp. 174-185.
[36] LAVRAC N., KAVSEK B., FLACH P. A., TODOROVSKI L., Subgroup discovery with CN2-SD, Journal of Machine Learning Research, 2004, Vol. 5, pp. 153-188.
[37] MICHALSKI R. S., Discovering Classification Rules Using variable-Valued Logic System VL_1, Proc. of the 3rd Int. Joint Conf. on Artificial Intelligence, 1973, pp. 162-172.
[38] OHSAKI M., ABE H., TSUMOTO S., YOKOI H., YAMAGUCHI T., Evaluation of rule interestingness measures in medical knowledge discovery in databases, Artificial Intelligence in Medicine, 2007, Vol. 41, pp. 177-196.
[39] PADMANABHAN B., TUZHILIN A., A belief-driven method for discovering unexpected patterns, Proc. of the Fourth Int.l Conference on Knowledge Discovery and Data Mining, 1998, pp. 94–100.
[40] PAWLAK Z., Rough sets: Theoretical aspects of reasoning about data, Kluwer, Dordrecht, 1991.
[41] PATTARAINTAKORN P., CERCONE N., A foundation of rough sets theoretical and computational hybrid intelligent system for survival analysis, Computers and Mathematics with Applications, 2008, Vol. 56, pp. 1699-1708.
[42] R DEVELOPMENT CORE TEAM: R, A language and environment for statistical computing, R Foundation for Statistical Computing, http://www.R-project.org/, Vienna, Austria, 2011.
[43] RAFEA A. A., SHAFLIK S. S., SHAALAN K. F., An interactive system for association rule discovery for life assurance, Proc. of the Int. Conference on Computer, Communication and Control Technologies CCCT '04, 2004, pp. 32-37.
[44] SAHAR S., Interestingness measures – On determining what is interesting, Data Mining and Knowledge Discovery Handbook, Springer-Verlag, Berlin, 2010, pp. 603-612.
[45] SIKORA M., Decision rule-based data models using TRS and NetTRS – methods and algorithms, Transaction on Rough Sets IX, LNCS, 2010, Vol. 5946, pp. 130-160.
[46] SIKORA M., WRÓBEL Ł., Data-driven Adaptive Selection of Rule Quality Measures for Improving the Rule Induction Algorithm, Lecture Notes in Artificial Intelligence, 2011, Vol. 6743, pp. 279-287.
[47] SIKORA M., GRUCA A., Induction and selection of the most interesting Gene Ontology based multiattribute rules for descriptions of gene groups, Pattern Recognition Letters, 2011, Vol. 32, pp. 258-269.
[48] SIKORA M., WRÓBEL Ł., Data-driven Adaptive Selection of Rule Quality Measures for Improving Rule Induction and Filtration Algorithms, International Journal of General Systems, 2013, Vol. 42, No. 4.
[49] SŁOWIŃSKI K., STEFANOWSKI J., SIWIŃSKI D., Application of rule induction and rough sets to verification of megnetic resonance diagnosis, Fundamenta Informaticae, 2002, Vol. 53, pp. 345-363.
[50] SMYTH P., GOODMAN R. M., Rule induction using information theory, Proc. of Knowledge Discovery in Databases, MIT Press, Boston, 1991, pp.159-175.
[51] STAJDUHAR I., DALBELO-BASIC B., Uncensoring censored data for machine learning: A likelihood-based approach, Expert Systems with Applications, 2012, Vol. 39, pp. 7226-7234.
[52] STEFANOWSKI J., Rough set based rule induction techniques for classification problems, Proc. of the 6th European Congress of Intelligent Techniques and Soft Computing, 1998, pp. 107-119.
[53] STEFANOWSKI J., VANDERPOOTEN D., Induction of Decision Rules in Classification and Discovery-Oriented Perspectives, International Journal of Intelligent Systems, 2001, Vol. 16, No. 1, pp. 13-27.
[54] ŚLĘZAK D., WIDZ S., Is It Important Which Rough-Set-Based Classifier Extraction and Voting Criteria Are Applied Together? Lecture Notes in Artificial Intelligence, 2010, Vol. 6086, pp. 187-196.

[55] TSUMOTO S., TANAKA H., Automated discovery of medical expert system rules form clinical databases based on rough sets, Proc. of the Second Int. Conf. on Knowledge Discovery & Data Mining, AAAI Press, 1996, pp. 63-69.
[56] TSUMOTO S., Automated Discovery of Positive and Negative Knowledge in Medical Databases, IEEE Engineering in medicine and biology, 2000, Vol. 19, No. 4, pp. 56-62.
[57] TSUMOTO S., Mining diagnostic rules from clinical databases using rough sets and medical diagnostic model, Information Sciences, 2004, Vol. 162, pp. 65-80.
[58] ULUTASDEMIR N., DAGLI O., Evaluation of risk of death in hepatitis by rule induction algorithms, Scientific Research and Essays, 2010, Vol. 20, No. 5, pp. 3059-3062.
[59] YAO Y. Y., ZHONG N., An analysis of quantitative measures associated with rules, Lecture Notes in Artificial Intelligence, 1999, Vol. 1574, pp. 479-488.
[60] YAO Y., ZHAO Y., WANG J., On reduct construction algorithms, Lecture Notes in Computer Sciences, 2008, Vol. 5150, pp. 100-117.
[61] WITTEN I. H., FRANK E., Data mining: practical machine learning tools and techniques, Morgan Kaufmann, 2005.
[62] WOHLARB L., FÜRNKRANZ J., A review and comparison of strategies for handling missing values in separate-and-conquer rule lerarning, Journal of Intelligent Information Systems, 2011, Vol. 36, No. 1, pp. 73-9.
[63] ZUPAN B., DEMSAR J., KATTAN M. W., BECK R., BRATKO I., Machine Learning for Survival Analysis: A Case Study on Recurrence of Prostate Cancer, Lecture Notes in Artificial Intelligence, 1990, Vol. 1620, pp. 346-355.