

Greedy Algorithm for the Construction of Approximate Decision Rules for Decision Tables with Many-Valued Decisions

Mohammad Azad¹(✉), Mikhail Moshkov¹, and Beata Zielosko²

¹ Computer, Electrical and Mathematical Sciences and Engineering Division,
King Abdullah University of Science and Technology,
Thuwal 23955-6900, Saudi Arabia

{mohammad.azad,mikhail.moshkov}@kaust.edu.sa

² Institute of Computer Science, University of Silesia,
39, Będzińska St., 41-200 Sosnowiec, Poland
beata.zielosko@us.edu.pl

Abstract. The paper is devoted to the study of a greedy algorithm for construction of approximate decision rules. This algorithm is applicable to decision tables with many-valued decisions where each row is labeled with a set of decisions. For a given row, we should find a decision from the set attached to this row. We consider bounds on the precision of this algorithm relative to the length of rules. To illustrate proposed approach we study a problem of recognition of labels of points in the plain. This paper contains also results of experiments with modified decision tables from UCI Machine Learning Repository.

1 Introduction

In this paper, we consider one more extension of the notion of decision table – decision table with many-valued decisions. In a table with many-valued decisions, each row is labeled with a nonempty finite set of decisions, and for a given row, we should find a decision from the set of decisions attached to this row.

Such tables arise in problems of discrete optimization, pattern recognition, computational geometry, decision making etc. [10, 17]. However, the main sources of decision tables with many-valued decisions are datasets filled by statistical or experimental data. In such datasets, we often have groups of objects with equal values of conditional attributes but, probably, different values of the decision attribute. Instead of a group of objects, we can consider one object given by values of conditional attributes. We attach to this object a set of decisions: either all decisions for objects from the group, or k the most frequent decisions for objects from the group, etc. As a result we obtain a decision table with many-valued decisions. In real life applications we can meet multi-label data when we study, e.g., problem of semantic annotation of images [4], music categorization into emotions [35], functional genomics [3], and text categorization [36].

In the rough set theory [22, 30, 31], decision tables are considered often that have equal rows labeled with different decisions. The set of decisions attached

to equal rows is called the *generalized decision* for that rows [23–25]. Here our aim is to find the generalized decision for a given row. In the paper, we will call this approach the generalized decision approach. However, the problem of finding an arbitrary decision or one of the most frequent decisions from the generalized decision is interesting also. Such study of decision tables with many-valued decisions can give a new tool for the rough set theory. In [2] and [18] we considered problem of construction of tests (super-reducts) and decision trees for decision tables with many-valued decisions. To choose one of the attributes we used uncertainty measure that is the number of boundary subtables.

Decision table with many-valued decisions can be considered as a decision table with an incomplete information because we don't know which decision should be chosen from the set of decisions. Incomplete information exists also in decision tables where instead of a single value of conditional attribute we have a subset of values of the attribute domain. In [13,14] approaches to interpreting queries in a database with such incomplete information were discussed. Z. Pawlak [22] and E. Orłowska [21] proposed Non-deterministic Information System for dealing with an incomplete information. Information incompleteness is connected also with missing values of attributes or intervals on values of attributes. M. Kryszkiewicz in [11] proposed method for computing all optimal generalized rules from decision table with missing values. In [27–29] authors proposed rule generation system, based on Apriori algorithm, where incomplete information was considered as nondeterministic information.

In literature, often, problems connected with multi-label data are considered from the point of view of classification (multi-label classifications problems) [7,8,15,19,33,34,37]. Here our aim is not to deal with classification but to show that proposed approach for construction of decision rules for decision tables with many-valued decisions can be useful when we deal with knowledge representation. In various applications, we often deal with decision tables which contain noisy data. In this case, exact rules can be “over-fitted”, i.e., depend essentially on the noise. So, instead of exact rules with many attributes, it is more appropriate to work with approximate rules with smaller number of attributes. Besides, classifiers based on approximate decision rules have often better accuracy than classifiers based on exact decision rules.

In the proposed approach a greedy algorithm constructs α -decision rules (α is a degree of rule uncertainty), and the number of rules for a given row is equal to the cardinality of set of decisions attached to this row. Then we choose for each row in a decision table a rule with the minimum length. The choice of shorter rules is connected with the Minimum Description Length principle [26].

The problem of construction of rules with minimum length is NP-hard. Therefore, we consider approximate polynomial algorithm for rule optimization. Based on results of U. Feige [9] it was proved in [16], for decision tables with one-valued decision, that greedy algorithm under some natural assumptions on the class NP, is close to the best polynomial approximate algorithms for partial decision rule minimization. It is natural to use these results in our approach. Note that each decision table with one-valued decision can be interpreted also as a decision table where each row is labeled with a set of decisions which has one element.

The paper, extending a conference publication [6] and some results presented in [17], is devoted to the study of a greedy algorithm for construction of approximate decision rules for decision tables with many-valued decisions. The greedy algorithm for rule construction has polynomial time complexity for the whole set of decision tables with many-valued decisions.

We discuss also a problem of recognition of labels of points in the plain which illustrates the considered approach and the obtained bounds on precision of this algorithm relative to the length of rules.

In this paper, we study only binary decision tables with many-valued decisions. However, the obtained results can be extended to the decision tables filled by numbers from the set $\{0, \dots, k-1\}$, where $k \geq 3$. We present experimental results based on modified data sets from UCI Machine Learning Repository [12] (by removal of some conditional attributes) into the form of decision tables with many-valued decisions. Experiments are connected with length of α -decision rules, number of different rules, lower and upper bounds on the minimum length of α -decision rules and 0.5-hypothesis for α -decision rules. We also present experimental results for the generalized decision approach. It allows us to make some comparative study of length and number of different rules, based on the proposed approach and the generalized decision approach.

The paper consists of eight sections. In Sect. 2, main notions are discussed. In Sect. 3, a parameter $M(T)$ and auxiliary statement are presented. This parameter is used for analysis of a greedy algorithm. Section 4 is devoted to the consideration of a set cover problem. In Sect. 5, the greedy algorithm for construction of approximate decision rules is studied. In this section we also present a lower and upper bounds on the minimum rule length based on the information about greedy algorithm work, and 0.5-hypothesis for tables with many-valued decisions. In Sect. 6, we discuss the problem of recognition of labels of points in the plain. In Sect. 7, experimental results are presented. Section 8 contains conclusions.

2 Main Definitions

In this section, we consider definitions corresponding to decision tables with many-valued decisions.

A *(binary) decision table with many-valued decisions* is a rectangular table T filled by numbers from the set $\{0, 1\}$. Columns of this table are labeled with attributes f_1, \dots, f_n . Rows of the table are pairwise different, and each row is labeled with a nonempty finite set of natural numbers (set of decisions). Note that each decision table with one-valued decisions can be interpreted also as a decision table with many-valued decisions. In such table, each row is labeled with a set of decisions which has one element. An example of decision table with many-valued decisions T_0 is presented in Table 1.

We will say that T is a *degenerate* table if either T is empty (has no rows), or the intersection of sets of decisions attached to rows of T is nonempty.

Table 1. Decision table T_0 with many-valued decisions
$$T_0 = \begin{array}{c|ccc|c} & f_1 & f_2 & f_3 & \\ \hline r_1 & 1 & 1 & 1 & \{1\} \\ r_2 & 0 & 1 & 0 & \{1, 3\} \\ r_3 & 1 & 1 & 0 & \{2\} \\ r_4 & 0 & 0 & 1 & \{2, 3\} \\ r_5 & 1 & 0 & 0 & \{1, 2\} \end{array}$$

A decision which belongs to the maximum number of sets of decisions attached to rows in T is called the *most common decision for T* . If we have more than one such decision, we choose the minimum one. If T is empty then 1 is the most common decision for T .

Let $r = (b_1, \dots, b_n)$ be a row of T labeled with a set of decisions $D(r)$ and $d \in D(r)$. By $U(T, r, d)$ we denote the set of rows r' from T for which $d \notin D(r')$. We will say that an attribute f_i separates a row $r' \in U(T, r, d)$ from the row r if the rows r and r' have different values at the intersection with the column f_i . The pair (T, r) will be called a *decision rule problem*.

Let α be a real number such that $0 \leq \alpha < 1$. A decision rule

$$f_{i_1} = b_1 \wedge \dots \wedge f_{i_m} = b_m \rightarrow d \quad (1)$$

is called an α -*decision rule for the pair (T, r) and decision $d \in D(r)$* if attributes f_{i_1}, \dots, f_{i_m} separate from r at least $(1 - \alpha)|U(T, r, d)|$ rows from $U(T, r, d)$. The number m is called the *length* of the rule (1). For example, 0.01-decision rule means that attributes contained in the rule should separate from the row r at least 99% of rows from $U(T, r, d)$. If α is equal to 0 we have an exact decision rule (0-decision rule) for (T, r) and d . If $U(T, r, d) = \emptyset$ then for any $f_{i_1}, \dots, f_{i_m} \in \{f_1, \dots, f_n\}$ the rule (1) is an α -decision rule for (T, r) and d . The rule (1) with empty left-hand side (when $m = 0$) is an α -decision rule for (T, r) and d if $U(T, r, d) = \emptyset$.

We will say that a decision rule is an α -*decision rule for the pair (T, r)* if this rule is an α -decision rule for the pair (T, r) and a decision $d \in D(r)$. We denote by $L_{\min}(\alpha, T, r, d)$ the minimum length of an α -decision rule for the pair (T, r) and decision $d \in D(r)$. We denote by $L_{\min}(\alpha, T, r)$ the minimum length of an α -decision rule for the pair (T, r) . It is clear that

$$L_{\min}(\alpha, T, r) = \min\{L_{\min}(\alpha, T, r, d) : d \in D(r)\}.$$

Let α, β be real numbers such that $0 \leq \alpha \leq \beta < 1$. One can show that $L_{\min}(\alpha, T, r, d) \geq L_{\min}(\beta, T, r, d)$ and $L_{\min}(\alpha, T, r) \geq L_{\min}(\beta, T, r)$.

3 Parameter $M(T)$

In this section, we consider definition of a parameter $M(T)$ and auxiliary statement from [17]. For the completeness, we will give this statement with proof.

We will use the parameter $M(T)$ to evaluate precision of a greedy algorithm relative to the length of rules.

Let T be a decision table with many-valued decisions, which has n columns labeled with attributes $\{f_1, \dots, f_n\}$.

Now, we define the parameter $M(T)$ of the table T . If T is a degenerate table then $M(T) = 0$. Let now T be a nondegenerate table. Let

$$\bar{\delta} = (\delta_1, \dots, \delta_n) \in \{0, 1\}^n.$$

Then $M(T, \bar{\delta})$ is the minimum natural m such that there exist attributes $f_{i_1}, \dots, f_{i_m} \in \{f_1, \dots, f_n\}$ for which $T(f_{i_1}, \delta_{i_1}) \dots (f_{i_m}, \delta_{i_m})$ is a degenerate table. Here $T(f_{i_1}, \delta_{i_1}) \dots (f_{i_m}, \delta_{i_m})$ is a subtable of the table T consisting only rows that have numbers $\delta_{i_1}, \dots, \delta_{i_m}$ at the intersection with the columns f_{i_1}, \dots, f_{i_m} . We denote

$$M(T) = \max\{M(T, \bar{\delta}) : \bar{\delta} \in \{0, 1\}^n\}.$$

Lemma 1. *Let T be a nondegenerate decision table with many-valued decisions which have n columns labeled with attributes f_1, \dots, f_n , $\bar{\delta} = (\delta_1, \dots, \delta_n) \in \{0, 1\}^n$, and $\bar{\delta}$ be a row of T . Then*

$$L_{\min}(0, T, \bar{\delta}) \leq M(T, \bar{\delta}) \leq M(T).$$

Proof. By definition, $M(T, \bar{\delta})$ is the minimum natural m such that there exist attributes $f_{i_1}, \dots, f_{i_m} \in \{f_1, \dots, f_n\}$ for which subtable

$$T' = T(f_{i_1}, \delta_{i_1}) \dots (f_{i_m}, \delta_{i_m})$$

is a degenerate table. The subtable T' is nonempty since $\bar{\delta}$ is a row of this subtable. Therefore there is a decision d which, for each row of T' , belongs to the set of decisions attached to this row.

One can show that a decision rule

$$f_{i_1} = \delta_{i_1} \wedge \dots \wedge f_{i_m} = \delta_{i_m} \rightarrow d$$

is a 0-decision rule for the pair $(T, \bar{\delta})$ and decision d . Therefore $L_{\min}(0, T, \bar{\delta}) \leq m = M(T, \bar{\delta})$. By definition, $M(T, \bar{\delta}) \leq M(T)$. \square

4 Set Cover Problem

In this section, we consider a greedy algorithm for construction of an approximate cover (an α -cover).

Let α be a real number such that $0 \leq \alpha < 1$. Let A be a set containing $N > 0$ elements, and $F = \{S_1, \dots, S_p\}$ be a family of subsets of the set A such that $A = \bigcup_{i=1}^p S_i$. We will say about the pair (A, F) as about a *set cover problem*. A subfamily $\{S_{i_1}, \dots, S_{i_t}\}$ of the family F will be called an α -cover for (A, F) if $|\bigcup_{j=1}^t S_{i_j}| \geq (1 - \alpha)|A|$. The problem of searching for an α -cover with minimum cardinality for a given set cover problem (A, F) is NP-hard [20, 32].

We consider now a greedy algorithm for construction of an α -cover (see Algorithm 1). At each step, this algorithm chooses a subset from F which covers the maximum number of uncovered elements from A . This algorithm stops when the constructed subfamily is an α -cover for (A, F) .

Algorithm 1. Greedy algorithm for approximate set cover problem

Input: a set cover problem (A, F) and real number α , $0 \leq \alpha < 1$.

Output: α -cover for (A, F) .

$G := \emptyset$, and $COVER := \emptyset$;

while $|G| < (1 - \alpha)|A|$ **do**

In the family F we find a set S_i with minimum index i such that

$$|S_i \cap (A \setminus G)| = \max\{|S_j \cap (A \setminus G)| : S_j \in F\}.$$

$G := G \cup S_i$ and $COVER := COVER \cup \{S_i\}$;

end while

return $COVER$

We denote by $C_{\text{greedy}}(\alpha, A, F)$ the cardinality of the constructed α -cover for (A, F) , and by $C_{\text{min}}(\alpha, A, F)$ we denote the minimum cardinality of an α -cover for (A, F) .

The following statement was obtained by J. Cheriyan and R. Ravi in [5]. We present it with our own proof.

Theorem 1. *Let $0 < \alpha < 1$ and (A, F) be a set cover problem. Then*

$$C_{\text{greedy}}(\alpha, A, F) \leq C_{\text{min}}(0, A, F) \ln(1/\alpha) + 1.$$

Proof. We denote $m = C_{\text{min}}(0, A, F)$. If $m = 1$ then, as it is not difficult to show, $C_{\text{greedy}}(\alpha, A, F) = 1$ and the considered inequality holds. Let $m \geq 2$ and S_i be a subset of maximum cardinality in F . It is clear that $|S_i| \geq N/m$. So, after the first step we will have at most $N - N/m = N(1 - 1/m)$ uncovered elements in the set A . After the first step we have the following set cover problem: the set $A \setminus S_i$ and the family $\{S_1 \setminus S_i, \dots, S_p \setminus S_i\}$. For this problem, the minimum cardinality of a cover is at most m . So, after the second step, when we choose a set $S_j \setminus S_i$ with maximum cardinality, the number of uncovered elements in the set A will be at most $N(1 - 1/m)^2$, etc.

Let the greedy algorithm in the process of α -cover construction make g steps and construct an α -cover of cardinality g . Then after the step number $g - 1$ more than αN elements in A are uncovered. Therefore $N(1 - 1/m)^{g-1} > \alpha N$ and $1/\alpha > (1 + 1/(m - 1))^{g-1}$. If we take the natural logarithm of both sides of this inequality we obtain $\ln 1/\alpha > (g - 1) \ln(1 + 1/(m - 1))$. It is known that for any natural p , the inequality $\ln(1 + 1/p) > 1/(p + 1)$ holds. Therefore $\ln(1/\alpha) > (g - 1)/m$ and $g < m \ln(1/\alpha) + 1$. Since $m = C_{\text{min}}(0, A, F)$ and $g = C_{\text{greedy}}(\alpha, A, F)$, we have $C_{\text{greedy}}(\alpha, A, F) < C_{\text{min}}(0, A, F) \ln(1/\alpha) + 1$. \square

5 Greedy Algorithm for α -Decision Rule Construction

In this section, we present a greedy algorithm for α -decision rule construction, lower and upper bounds on the minimum length of α -decision rules (Sect. 5.1) and 0.5-hypothesis connected with the work of a greedy algorithm (Sect. 5.2).

We use the greedy algorithm for construction of α -covers to construct α -decision rules. Let T be a table with many-valued decisions containing n columns labeled with attributes f_1, \dots, f_n , $r = (b_1, \dots, b_n)$ be a row of T , $D(r)$ be the set of decisions attached to r , $d \in D(r)$, and α be a real number such that $0 < \alpha < 1$.

We consider a set cover problem $(A(T, r, d), F(T, r, d))$ where $A(T, r, d) = U(T, r, d)$ is the set of all rows r' of T such that $d \notin D(r')$ and $F(T, r, d) = \{S_1, \dots, S_n\}$. For $i = 1, \dots, n$, the set S_i coincides with the set of all rows from $A(T, r, d)$ which are different from r in the column f_i . One can show that the decision rule

$$f_{i_1} = b_{i_1} \wedge \dots \wedge f_{i_m} = b_{i_m} \rightarrow d$$

is an α -decision rule for (T, r) and decision $d \in D(r)$ if and only if $\{S_{i_1}, \dots, S_{i_m}\}$ is an α -cover for the set cover problem $(A(T, r, d), F(T, r, d))$. Evidently, for the considered set cover problem, $C_{\min}(0, A(T, r, d), F(T, r, d)) = L_{\min}(0, T, r, d)$, where $L_{\min}(0, T, r, d)$ is the minimum length of 0-decision rule for (T, r) and decision $d \in D(r)$.

Let us apply the greedy algorithm (see Algorithm 1) to the considered set cover problem. This algorithm constructs an α -cover which corresponds to an α -decision rule $rule(\alpha, T, r, d)$ for (T, r) and decision $d \in D(r)$. From Theorem 1 it follows that the length of this rule is at most

$$L_{\min}(0, T, r, d) \ln(1/\alpha) + 1.$$

We denote by $L_{\text{greedy}}(\alpha, T, r)$ the length of the rule constructed by the following polynomial time algorithm: for a given α , $0 < \alpha < 1$, decision table T , row r of T and decision $d \in D(r)$, we construct the set cover problem $(A(T, r, d), F(T, r, d))$ and then apply to this problem the greedy algorithm for construction of an α -cover. We transform the obtained α -cover into an α -decision rule $rule(\alpha, T, r, d)$. Among the α -decision rules $rule(\alpha, T, r, d)$, $d \in D(r)$, we choose a rule with the minimum length. This rule is the output of the considered algorithm. We denote by $L_{\min}(\alpha, T, r)$ the minimum length of an α -decision rule for (T, r) . According to what has been said above we have the following statement.

Theorem 2. *Let T be a non-degenerate decision table with many-valued decisions, r be a row of T , and α be a real number such that $0 < \alpha < 1$. Then*

$$L_{\text{greedy}}(\alpha, T, r) \leq L_{\min}(0, T, r) \ln(1/\alpha) + 1.$$

Note that the considered algorithm is a generalization of an algorithm studied in [16].

Example 1. Let us apply the considered greedy algorithm to $\alpha = 0.1$, decision table T_0 (see Table 1) and the second row r_2 of this table.

For each $d \in D(r_2) = \{1, 3\}$ we construct the set cover problem $(A(T, r_2, d), F(T, r_2, d))$, where $A(T, r_2, d)$ is the set of all rows r' of T such that $d \notin D(r')$, $F(T, r_2, d) = \{S_1, S_2, S_3\}$, and S_i coincides with the set of rows from $A(T, r_2, d)$ which are different from r_2 in the column f_i , $i = 1, 2, 3$. We have:

- $A(T, r_2, 1) = \{r_3, r_4\}$, $F(T, r_2, 1) = \{S_1 = \{r_3\}, S_2 = \{r_4\}, S_3 = \{r_4\}\}$,
- $A(T, r_2, 3) = \{r_1, r_3, r_5\}$, $F(T, r_2, 3) = \{S_1 = \{r_1, r_3, r_5\}, S_2 = \{r_5\}, S_3 = \{r_1\}\}$.

Now, we apply the greedy algorithm for the set cover problem (with $\alpha = 0.1$) to each of the constructed set cover problems, and transform the obtained 0.1-covers into 0.1-decision rules.

For the case $d = 1$, we obtain the 0.1-cover $\{S_1, S_2\}$ and corresponding 0.1-decision rule $f_1 = 0 \wedge f_2 = 1 \rightarrow 1$.

For the case $d = 3$, we obtain the 0.1-cover $\{S_1\}$ and corresponding 0.1-decision rule $f_1 = 0 \rightarrow 3$. We choose the shortest rule $f_1 = 0 \rightarrow 3$ which is the result of our algorithm work.

In order to show that the problem of minimization of α -decision rule length is NP-hard, let us consider a set cover problem (A, F) where $A = \{a_1, \dots, a_N\}$ and $F = \{S_1, \dots, S_m\}$. We define the decision table as $T(A, F)$, this table has m columns corresponding to the sets S_1, \dots, S_m respectively, and $N + 1$ rows. For $j = 1, \dots, N$, the j -th row corresponds to the element a_j . The last $(N + 1)$ -th row is filled by 0. For $j = 1, \dots, N$ and $i = 1, \dots, m$, at the intersection of j -th row and i -th column 1 stays if and only if $a_j \in S_i$. The set of decisions corresponding to the last row is equal to $\{2\}$. All other rows are labeled with the set of decisions $\{1\}$.

One can show that, for any α , $0 \leq \alpha < 1$, a subfamily $\{S_{i_1}, \dots, S_{i_t}\}$ is an α -cover for (A, F) if and only if the decision rule

$$f_{i_1} = 0 \wedge \dots \wedge f_{i_t} = 0 \rightarrow 2$$

is an α -decision rule for $T(A, F)$ and the last row of $T(A, F)$.

So, we have a polynomial time reduction of the problem of minimization of α -cover cardinality to the problem of minimization of α -decision rule length for decision tables with many-valued decisions. Since the first problem is NP-hard [20, 32], we have

Proposition 1. *For any α , $0 \leq \alpha < 1$, the problem of minimization of α -decision rule length for decision tables with many-valued decisions is NP-hard.*

5.1 Upper and Lower Bounds on $L_{\min}(\alpha, T, r)$

In this section, we present some results connected with lower and upper bounds on the minimum length of α -decision rules, based on the information obtained during the greedy algorithm work.

Let T be a decision table with many-valued decisions and r be a row of T . Let α be a real number such that $0 \leq \alpha < 1$. We apply the greedy algorithm to T, r, α and each $d \in D(r)$ (really to the corresponding set cover problem) and obtain for every $d \in D(r)$ an α -decision rule for the pair (T, r) and decision d . Among these rules we choose a rule with the minimum length, and denote this length by $u(\alpha, T, r)$. It is clear that

$$L_{\min}(\alpha, T, r) \leq u(\alpha, T, r).$$

Let $d \in D(r)$. We apply the greedy algorithm to T, r, α and d , and construct the α -decision rule $rule(\alpha, T, r, d)$. Let the length of this rule be equal to t , and $\delta_i, i = 1, \dots, t$, be the number of rows from $U(T, r, d)$ separated from row r at the i -th step of the greedy algorithm work. We denote

$$l(\alpha, T, r, d) = \max \left\{ \left\lceil \frac{[(1 - \alpha)|U(T, r, d)|] - (\delta_0 + \dots + \delta_i)}{\delta_{i+1}} \right\rceil : i = 0, \dots, t - 1 \right\},$$

where $\delta_0 = 0$. Let us denote

$$l(\alpha, T, r) = \min_{d \in D(r)} l(\alpha, T, r, d).$$

We can almost repeat the first part of the proof of Theorem 1.67 from [16] to obtain the following lower bound:

$$L_{\min}(\alpha, T, r, d) \geq l(\alpha, T, r, d),$$

where $L_{\min}(\alpha, T, r, d)$ is the minimum length of α -decision rule for (T, r) and d . From this inequality it follows that

$$L_{\min}(\alpha, T, r) \geq l(\alpha, T, r).$$

5.2 0.5-Hypothesis

In the book [16], the following 0.5-hypothesis was formulated for decision tables with one-valued decisions: for the most part of decision tables for each row r under the construction of decision rule, during each step the greedy algorithm chooses an attribute which separates from r at least one-half of unseparated rows with decisions other than decision attached to the row r .

Let T be a decision table with many-valued decisions and r be a row of T . We will say that 0.5-hypothesis is true for T and r if for any decision $d \in D(r)$ under the construction of decision rule for the pair (T, r) and decision d , during each step the greedy algorithm chooses an attribute which separates from r at least 50% of unseparated rows from $U(T, r, d)$.

We will say that 0.5-hypothesis is true for T if it is true for each row of T .

Now we consider some theoretical results regarding to 0.5-hypothesis for decision tables with many-valued decisions.

A *binary information system* I is a table with n rows (corresponding to objects) and m columns labeled with attributes f_1, \dots, f_m . This table is filled by numbers from $\{0, 1\}$ (values of attributes). For $j = 1, \dots, n$, we denote by r_j the j -th row of the table I .

The information system I will be called *strongly saturated* if, for any row $r_j = (b_1, \dots, b_n)$ of I , for any $k \in \{0, \dots, n - 1\}$ and for any k rows with numbers different from j , there exists a column f_i which has at least $\frac{k}{2}$ numbers $-b_i$ (b_i is the value of the f_i column for the row r_j) at the intersection with the considered k rows.

First, we evaluate the number of strongly saturated binary information systems. After that, we study the work of the greedy algorithm on a decision table with many-valued decisions obtained from a strongly saturated binary information system by adding a set of decisions to each row. It is clear that 0.5-hypothesis holds for every such table.

Theorem 3 [16]. *Let us consider binary information systems with n rows and $m \geq n + \log_2 n$ columns labeled with attributes f_1, \dots, f_m . Then the fraction of strongly saturated information systems is at least $1 - 1/2^{m-n-\log_2 n+1}$.*

For example, if $m \geq n + \log_2 n + 6$, than at least 99% of binary information systems are strongly saturated.

Let us consider the work of the greedy algorithm on an arbitrary decision table T with many-valued decisions obtained from the strongly saturated binary information system. Let r be an arbitrary row of table T and $d \in D(r)$. For $i = 1, 2, \dots$, after the step number i at most $|U(T, r, d)|/2^i$ rows from $U(T, r, d)$ are unseparated from r . It is not difficult to show that $L_{\text{greedy}}(\alpha, T, r) \leq \lceil \log_2(1/\alpha) \rceil$ for any real α , $0 < \alpha < 1$, where $L_{\text{greedy}}(\alpha, T, r)$ is the length of α -decision rules constructed by the greedy algorithm for (T, r) . One can prove that $L_{\text{greedy}}(0, T, r) \leq \log_2 |U(T, r, d)| + 1$. It is easy to check that $l(0, T, r) \leq 2$.

6 Problem of Recognition of Labels of Points in the Plain

In this section, we present a problem of recognition of colors of points in the plain (note that, we recognize labels attached to the points, and labels are named as colors), which illustrates the considered approach and the obtained bounds on precision of the greedy algorithm relative to the length of α -decision rules.

Let we have a finite set $S = \{(a_1, b_1), \dots, (a_n, b_n)\}$ of points in the plane and a mapping μ which corresponds to each point (a_p, b_p) a nonempty subset $\mu(a_p, b_p)$ of the set $\{green, yellow, red\}$. Colors are interpreted as decisions, and for each point from S we need to find a decision (color) from the set of decisions attached to this point. We denote this problem by (S, μ) .

For the problem (S, μ) solving, we use attributes corresponding to straight lines which are given by equations of the kind $x = \beta$ or $y = \gamma$. These attributes

are defined on the set S and take values from the set $\{0, 1\}$. Consider the line given by equation $x = \beta$. Then the value of corresponding attribute is equal to 0 on a point $(a, b) \in S$ if and only if $a < \beta$. Consider the line given by equation $y = \gamma$. Then the value of corresponding attribute is equal to 0 if and only if $b < \gamma$.

We now choose a finite set of straight lines which allow us to construct a decision rule with the minimum length for the problem (S, μ) . It is possible that $a_i = a_j$ or $b_i = b_j$ for different i and j . Let a_{i_1}, \dots, a_{i_m} be all pairwise different numbers from the set $\{a_1, \dots, a_n\}$ which are ordered such that $a_{i_1} < \dots < a_{i_m}$. Let b_{j_1}, \dots, b_{j_t} be all pairwise different numbers from the set $\{b_1, \dots, b_n\}$ which are ordered such that $b_{j_1} < \dots < b_{j_t}$.

One can show that there exists a decision rule with minimum length which use only attributes corresponding to the straight lines defined by equations $x = a_{i_1} - 1, x = (a_{i_1} + a_{i_2})/2, \dots, x = (a_{i_{m-1}} + a_{i_m})/2, x = a_{i_m} + 1, y = b_{j_1} - 1, y = (b_{j_1} + b_{j_2})/2, \dots, y = (b_{j_{t-1}} + b_{j_t})/2, y = b_{j_t} + 1$.

Now, we describe a decision table $T(S, \mu)$ with $m + t + 2$ columns and n rows. Columns of this table are labeled with attributes f_1, \dots, f_{m+t+2} , corresponding to the considered $m + t + 2$ lines. Attributes f_1, \dots, f_{m+1} correspond to lines defined by equations $x = a_{i_1} - 1, x = (a_{i_1} + a_{i_2})/2, \dots, x = (a_{i_{m-1}} + a_{i_m})/2, x = a_{i_m} + 1$ respectively. Attributes $f_{m+2}, \dots, f_{m+t+2}$ correspond to lines defined by equations $y = b_{j_1} - 1, y = (b_{j_1} + b_{j_2})/2, \dots, y = (b_{j_{t-1}} + b_{j_t})/2, y = b_{j_t} + 1$ respectively. Rows of the table $T(S, \mu)$ correspond to points $(a_1, b_1), \dots, (a_n, b_n)$. At the intersection of the column f_l and row (a_p, b_p) the value $f_l(a_p, b_p)$ stays. For $p = 1, \dots, n$, the row (a_p, b_p) is labeled with the set of decisions $\mu(a_p, b_p)$.

Example 2. A problem (S, μ) with four points and corresponding decision table $T(S, \mu)$ is depicted in Fig. 1. We write “g” instead of “green”, “r” instead of “red”, and “y” instead of “yellow”.

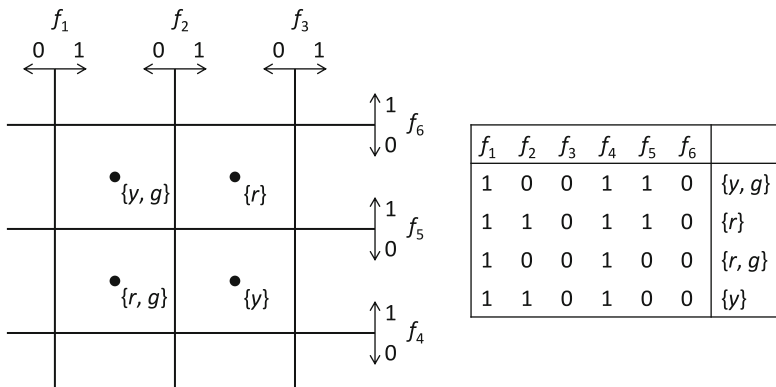


Fig. 1. Problem (S, μ) and corresponding decision table $T(S, \mu)$

Let us evaluate the parameter $M(T(S, \mu))$.

Proposition 2. $M(T(S, \mu)) \leq 4$.

Proof. We denote $T = T(S, \mu)$. Let $\bar{\delta} = (\delta_1, \dots, \delta_{m+t+2}) \in \{0, 1\}^{m+t+2}$. If $\delta_1 = 0$, or $\delta_{m+1} = 1$, or $\delta_{m+2} = 0$, or $\delta_{m+t+2} = 1$, then $T(f_1, \delta_1)$, or $T(f_{m+1}, \delta_{m+1})$, or $T(f_{m+2}, \delta_{m+2})$, or $T(f_{m+t+2}, \delta_{m+t+2})$ is empty table and $M(T, \bar{\delta}) \leq 1$. Let $\delta_1 = 1$, $\delta_{m+1} = 0$, $\delta_{m+2} = 1$ and $\delta_{m+t+2} = 0$. One can show that in this case there exist $i \in \{1, \dots, m\}$ and $j \in \{m+2, \dots, m+t+1\}$ such that $\delta_i = 1$, $\delta_{i+1} = 0$, $\delta_j = 1$, and $\delta_{j+1} = 0$. It is clear that the table $T(f_i, \delta_i)(f_{i+1}, \delta_{i+1})(f_j, \delta_j)(f_{j+1}, \delta_{j+1})$ contains exactly one row. So $M(T, \bar{\delta}) \leq 4$ and $M(T) \leq 4$. \square

From Lemma 1, Theorem 2 and Proposition 2 the next statement follows:

Corollary 1. For any real α , $0 < \alpha < 1$, and any row r of the table $T(S, \mu)$,

$$L_{\text{greedy}}(\alpha, T(S, \mu), r) < 4 \ln(1/\alpha) + 1.$$

Note that $4 \ln(1/0.01) + 1 < 19.43$, $4 \ln(1/0.1) + 1 < 10.22$, $4 \ln(1/0.2) + 1 < 7.44$, and $4 \ln(1/0.5) + 1 < 3.78$.

7 Results of Experiments

This section consists of three parts:

- experimental results for the many-valued decisions approach (Sect. 7.1),
- experimental results for the generalized decision approach (Sect. 7.2),
- comparative study (Sect. 7.3).

We consider a number of decision tables from UCI Machine Learning Repository [12]. In some tables there were missing values. Each such value was replaced with the most common value of the corresponding attribute. Some decision tables contain conditional attributes that take unique value for each row. Such attributes were removed. In some tables there were equal rows with, possibly, different decisions.

In this case each group of identical rows was replaced with a single row from the group which is labeled with the set of decisions attached to rows from the group. To obtain rows which are labeled with sets containing more than one decision we removed from decision tables more conditional attributes. The information about such decision tables can be found in Table 2. This table contains name of initial table, number of rows (column “Rows”), number of attributes (column “Attr”), spectrum of this table (column “Spectrum”), and list of names of removed attributes (column “Removed attributes”). Spectrum of a decision table with many-valued decisions is a sequence $\#1, \#2, \dots$, where $\#i$, $i = 1, 2, \dots$, is the number of rows labeled with sets of decision with cardinality equals to i . All experiments are performed using DAGGER software tool [1] in C++.

Table 2. Characteristics of decision tables with many-valued decisions

Decision table	Rows	Attr	Spectrum						Removed attributes
			#1	#2	#3	#4	#5	#6	
balance-scale-1	125	3	45	50	30				left-weight
breast-cancer-1	193	8	169	24					tumor-size
breast-cancer-5	98	4	58	40					inv-nodes,node-caps,deg- malig,breast-quad,irradiat
cars-1	432	5	258	161	13				buying
flags-5	171	21	159	12					zone,language,religion,circles, sunstars
hayes-roth-data-1	39	3	22	13	4				marital_status
kr-vs-kp-5	1987	31	1564	423					katri,mulch,rimmx,skrxp,wknc
kr-vs-kp-4	2061	32	1652	409					katri,mulch,rimmx,wknc
lymphography-5	122	13	113	9					lymphatics,changes_in_node, changes_in_stru, special_forms,no_of_nodes_in
mushroom-5	4078	17	4048	30					odor,gill-size,stalk-root,stalk- surface-below-ring,habitat
nursery-4	240	4	97	96	47				parents,housing,finance,social
nursery-1	4320	7	2858	1460	2				parents
poker-hand-train-5	3324	5	156	1832	1140	188	7	1	S1,C1,C2,C4,C5
poker-hand-train-5a	3323	5	130	1850	1137	199	6	1	C1,S2,C2,C3,C4
poker-hand-train-5b	1024	5	0	246	444	286	44	4	C1,C2,C3,C4,C5
spect-test-1	164	21	161	3					F3
teeth-1	22	7	12	10					top incisors
teeth-5	14	3	6	3	0	5	0	2	bottom incisors,top canines,bottom canines,top premolars,bottom molars
tic-tac-toe-4	231	5	102	129					top-right-square,middle-middle- square,bottom-left- square,bottom-right-square
tic-tac-toe-3	449	6	300	149					middle-middle-square,bottom- left-square,bottom-right- square
zoo-data-5	42	11	36	6					feathers,backbone,breathes, legs,tail

7.1 Proposed Approach

We made four groups of experiments which are connected with:

- length of constructed α -decision rules,
- number of different α -decision rules,
- lower and upper bounds on the minimum length of α -decision rules,
- 0.5-hypothesis.

The first group of experiments is the following. For decision tables described in Table 2 and $\alpha \in \{0.0, 0.001, 0.01, 0.1, 0.2, 0.3\}$, we apply to each row of table the greedy algorithm. After that, among the constructed rules we find minimum (column “min”), average (column “avg”) and maximum (column “max”) length of such rules. Results can be found in Tables 3 and 4.

One can see that the length of constructed α -decision rules is decreasing when the value of α is increasing, and the greedy algorithm constructs relatively short α -decision rules.

Table 3. Length of α -decision rules for $\alpha \in \{0.0, 0.001, 0.01\}$ constructed by greedy algorithm

Decision table	$\alpha = 0.0$			$\alpha = 0.001$			$\alpha = 0.01$		
	min	avg	max	min	avg	max	min	avg	max
balance-scale-1	2	2.00	2	2	2.00	2	2	2.00	2
breast-cancer-1	1	2.94	5	1	2.94	5	1	2.81	5
breast-cancer-5	1	1.72	3	1	1.72	3	1	1.72	3
cars-1	1	1.38	4	1	1.38	4	1	1.36	3
flags-5	1	2.43	5	1	2.43	5	1	2.10	4
hayes-roth-data-1	1	1.59	2	1	1.59	2	1	1.59	2
kr-vs-kp-5	1	4.11	11	1	4.11	11	1	2.97	8
kr-vs-kp-4	1	4.14	11	1	4.14	11	1	2.98	8
lymphography-5	1	2.68	5	1	2.68	5	1	2.66	5
mushroom-5	1	1.52	5	1	1.52	5	1	1.46	4
nursery-4	1	1.33	2	1	1.33	2	1	1.33	2
nursery-1	1	2.05	5	1	2.00	5	1	1.80	3
poker-hand-train-5	2	2.35	5	2	2.35	5	2	2.27	4
poker-hand-train-5a	2	2.24	5	2	2.24	4	2	2.09	4
poker-hand-train-5b	2	2.01	5	2	2.01	5	2	2.01	4
spect-test-1	1	1.32	5	1	1.32	5	1	1.29	4
teeth-1	1	2.27	3	1	2.27	3	1	2.27	3
teeth-5	1	1.93	3	1	1.93	3	1	1.93	3
tic-tac-toe-4	2	2.24	4	2	2.24	4	2	2.24	4
tic-tac-toe-3	3	3.29	6	3	3.29	6	3	3.14	5
zoo-data-5	1	2.19	5	1	2.19	5	1	2.19	5

Table 4. Length of α -decision rules for $\alpha \in \{0.1, 0.2, 0.3\}$ constructed by greedy algorithm

Decision table	$\alpha = 0.1$			$\alpha = 0.2$			$\alpha = 0.3$		
	min	avg	max	min	avg	max	min	avg	max
balance-scale-1	1	1.06	2	1	1.00	1	1	1.00	1
breast-cancer-1	1	1.75	3	1	1.28	2	1	1.01	2
breast-cancer-5	1	1.15	2	1	1.00	1	1	1.00	1
cars-1	1	1.22	2	1	1.17	2	1	1.00	1
flags-5	1	1.20	2	1	1.02	2	1	1.00	1
hayes-roth-data-1	1	1.00	1	1	1.00	1	1	1.00	1
kr-vs-kp-5	1	1.77	4	1	1.17	2	1	1.04	2
kr-vs-kp-4	1	1.73	4	1	1.16	3	1	1.03	2
lymphography-5	1	1.80	3	1	1.37	2	1	1.01	2
mushroom-5	1	1.12	2	1	1.01	2	1	1.00	1
nursery-4	1	1.01	2	1	1.00	1	1	1.00	1
nursery-1	1	1.38	2	1	1.00	2	1	1.00	1
poker-hand-train-5	1	1.06	2	1	1.00	1	1	1.00	1
poker-hand-train-5a	1	1.01	2	1	1.00	1	1	1.00	1
poker-hand-train-5b	2	2.00	2	2	2.00	2	1	1.00	1
spect-test-1	1	1.25	3	1	1.08	2	1	1.07	2
teeth-1	1	1.50	2	1	1.14	2	1	1.05	2
teeth-5	1	1.71	3	1	1.29	2	1	1.07	2
tic-tac-toe-4	1	1.47	3	1	1.04	2	1	1.01	2
tic-tac-toe-3	2	2.00	3	1	1.50	2	1	1.01	2
zoo-data-5	1	1.69	4	1	1.26	3	1	1.12	2

Table 5. Number of different rules constructed by greedy algorithm

Decision table	values of α					
	0.0	0.001	0.01	0.1	0.2	0.3
balance-scale-1	51	51	51	45	27	23
breast-cancer-1	164	164	164	112	49	35
breast-cancer-5	61	61	61	38	25	25
cars-1	23	23	19	12	11	10
flags-5	159	159	155	111	101	98
hayes-roth-data-1	14	14	14	7	6	6
kr-vs-kp-5	856	856	577	188	93	65
kr-vs-kp-4	914	914	621	210	105	72
lymphography-5	59	59	58	44	32	23
mushroom-5	493	508	448	165	73	58
nursery-4	7	7	7	13	5	5
nursery-1	89	74	33	23	10	10
poker-hand-train-5	392	392	402	136	36	36
poker-hand-train-5a	333	333	255	93	34	34
poker-hand-train-5b	52	52	52	52	52	8
spect-test-1	29	29	29	28	22	22
teeth-1	22	22	22	22	22	22
teeth-5	14	14	14	14	14	14
tic-tac-toe-4	52	52	52	30	15	14
tic-tac-toe-3	157	157	137	63	57	16
zoo-data-5	16	16	16	16	15	15

Table 5 presents the number of different rules constructed by the greedy algorithm for $\alpha \in \{0.0, 0.001, 0.01, 0.1, 0.2, 0.3\}$. In the worst case, the number of different rules can be equal to the number of rows in decision table T . One can see that with the exception of three tables, the number of different rules is non-increasing when the value of α is increasing.

Next group of experimental results is connected with lower and upper bounds on the minimum length of α -decision rules. Figures 2 and 3 present average values of bounds $l(\alpha, T, r)$ and $u(\alpha, T, r)$ among all rows r of T for α , $0 \leq \alpha < 1$, with the step 0.01.

The last group of experiments is connected with 0.5-hypothesis. Table 6 contains, for $i = 1, 2, \dots$, the average percentage of rows separated at i -th step of the greedy algorithm (average among all rows r and decisions $d \in D(r)$).

For decision tables described in Table 2 we find the number of rows for which 0.5-hypothesis is true. Table 7 contains name of decision table, number of rows and number of rows for which 0.5-hypothesis is true.

Results in Table 6 show that average percentage of rows separated at i -th step of the greedy algorithm during the exact decision rule construction is more than or equal to 50% (7-th step of the greedy algorithm for “spect-test-1”). We say that 0.5-hypothesis is true for T if it is true for each row of T . Based on results in Table 7 we can see that 0.5-hypothesis is true for 12 decision tables

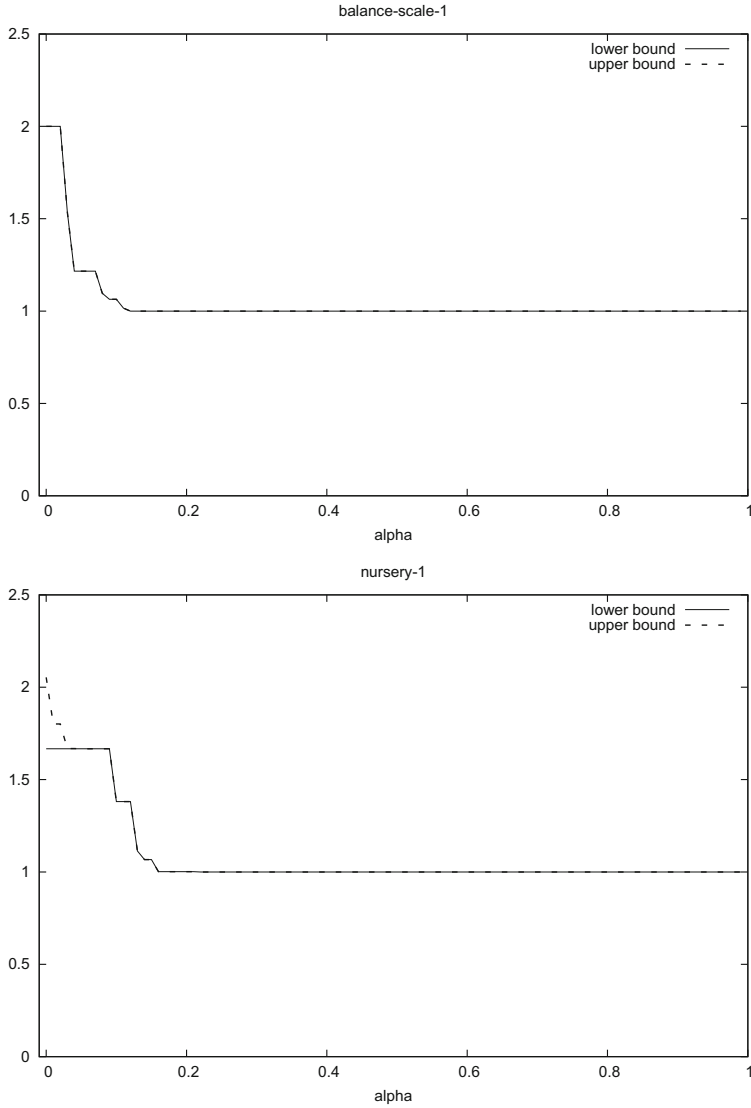


Fig. 2. Lower and upper bounds on $L_{\min}(\alpha, T, r)$ (“balance-scale-1” and “nursery-1”)

and is not true for 9 decision tables: “breast-cancer-1”, “kr-vs-kp-5”, “kr-vs-kp-4”, “lymphography-5”, “poker-hand-train-5”, “poker-hand-train-5a”, “spect-test-1”, “tic-tac-toe-3”, and “zoo-data-5”.

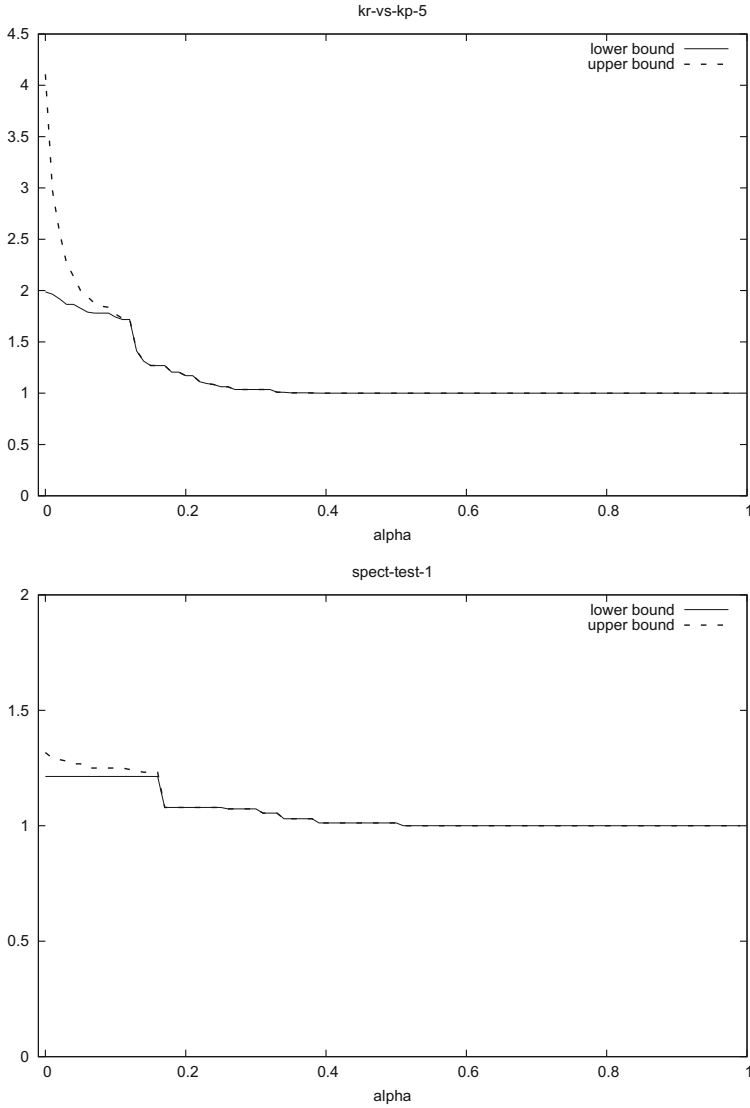


Fig. 3. Average values of lower and upper bounds on $L_{\min}(\alpha, T, r)$ (“kr-vs-kp-5” and “spect-test-1”)

7.2 Generalized Decision Approach

In this section, we present experimental results for α -decision rules relative to:

- length of constructed α -decision rules,
- number of different α -decision rules,
- 0.5-hypothesis.

Table 6. Average percentage of rows separated at i -th step of the greedy algorithm work

Decision table	Number of step i									
	1	2	3	4	5	6	7	8	9	10
balance-scale-1	91.5	97.2	100							
breast-cancer-1	84.5	82.8	87.6	92.9	84.8	100				
breast-cancer-5	91.8	94.1	97.6	100						
cars-1	90.2	86.9	89.1	85.7	100					
flags-5	92.9	92.4	91.9	92.5	100					
hayes-roth-data-1	89.7	96.4	100							
kr-vs-kp-5	85.7	79.9	78.2	80.2	81.6	81.4	85.7	80.2	80.8	78.3
kr-vs-kp-4	85.8	79.5	78.4	79.8	81.9	82.1	85.3	82.1	82.3	80.0
lymphography-5	83.9	86.2	92.1	94.2	100					
mushroom-5	97.0	97.3	91.2	92.6	96.8	100				
nursery-4	94.8	99.8	90.9	100						
nursery-1	90.2	92.3	93.3	93.6	92.5	92.7	100			
poker-hand-train-5	92.3	87.2	91.0	86.3	100					
poker-hand-train-5a	92.3	88.7	91.1	84.9	100					
poker-hand-train-5b	75.5	92.2	81.5	87.5	100					
spect-test-1	94.3	89.4	80.5	85.6	77.4	75.0	50.0	100		
teeth-1	88.2	89.9	100							
teeth-5	87.0	92.0	100							
tic-tac-toe-4	86.7	91.6	90.6	94.9	100					
tic-tac-toe-3	78.5	83.6	88.8	88.1	91.2	100				
zoo-data-5	83.8	83.4	80.0	80.6	100					

Table 7. Number of rows in decision tables for which 0.5-hypothesis is true

Decision table	rows	rows with 0.5-hypothesis
balance-scale-1	125	125
breast-cancer-1	193	191
breast-cancer-5	98	98
cars-1	432	432
flags-5	171	171
hayes-roth-data-1	39	39
kr-vs-kp-5	1987	1873
kr-vs-kp-4	2061	1949
lymphography-5	122	121
mushroom-5	4078	4078
nursery-4	240	240
nursery-1	4320	4320
poker-hand-train-5	3324	3321
poker-hand-train-5a	3323	3322
poker-hand-train-5b	1024	1024
spect-test-1	164	163
teeth-1	22	22
teeth-5	14	14
tic-tac-toe-4	231	231
tic-tac-toe-3	449	445
zoo-data-5	42	40

Table 8. Transformation of the set of decisions for the generalized decision approach

$$T_0 = \begin{array}{c|ccc|c} & f_1 & f_2 & f_3 & d \\ \hline r_1 & 1 & 1 & 1 & \{1\} \\ r_2 & 0 & 1 & 0 & \{1, 3\} \\ r_3 & 1 & 1 & 0 & \{2\} \\ r_4 & 0 & 0 & 1 & \{2, 3\} \\ r_5 & 1 & 0 & 0 & \{1, 2\} \end{array} \Rightarrow \begin{array}{|c|} \hline d \\ \hline 1 \\ \hline 2 \\ \hline 3 \\ \hline 4 \\ \hline 5 \\ \hline \end{array}$$

In the generalized decision approach [23–25], the greedy algorithm constructs for each row one α -decision rule which has on the right-hand side the generalized decision (a number encoding the set of decisions attached to a given row) see Table 8.

For decision tables described in Table 2 and $\alpha \in \{0.0, 0.001, 0.01, 0.1, 0.2, 0.3\}$ we apply to each row of table the greedy algorithm. After that, among the constructed rules we find minimum (column “min”), average (column “avg”) and maximum (column “max”) length of such rules. Results can be found in Tables 9 and 10.

We can say that for this approach the greedy algorithm constructs relatively short α -decision rules.

Table 9. Length of α -decision rules for $\alpha \in \{0.0, 0.001, 0.01\}$ —generalized decision approach

Decision table	$\alpha = 0.0$			$\alpha = 0.001$			$\alpha = 0.01$		
	min	avg	max	min	avg	max	min	avg	max
balance-scale-1	3	3.00	3	3	3.00	3	2	2.92	3
breast-cancer-1	1	3.29	6	1	3.29	6	1	3.08	6
breast-cancer-5	1	2.38	4	1	2.38	4	1	2.38	4
cars-1	1	2.51	5	1	2.51	5	1	2.16	4
flags-5	1	2.49	5	1	2.49	5	1	2.15	4
hayes-roth-data-1	1	2.13	3	1	2.13	3	1	2.13	3
kr-vs-kp-5	1	4.81	12	1	4.42	11	1	3.14	8
kr-vs-kp-4	1	4.81	12	1	4.42	11	1	3.11	8
lymphography-5	1	2.94	6	1	2.94	6	1	2.89	6
mushroom-5	1	1.54	8	1	1.53	6	1	1.47	4
nursery-4	1	2.02	4	1	2.02	4	1	1.69	4
nursery-1	1	3.32	7	1	2.98	6	1	2.35	4
poker-hand-train-5	4	4.78	5	4	4.09	5	3	3.00	3
poker-hand-train-5a	4	4.78	5	4	4.08	5	3	3.00	3
poker-hand-train-5b	4	4.96	5	4	4.96	5	3	3.86	4
spect-test-1	1	1.72	10	1	1.72	10	1	1.68	9
teeth-1	1	2.27	3	1	2.27	3	1	2.27	3
teeth-5	1	1.93	3	1	1.93	3	1	1.93	3
tic-tac-toe-4	3	3.79	5	3	3.79	5	3	3.41	5
tic-tac-toe-3	3	4.55	6	3	4.55	6	3	3.61	5
zoo-data-5	1	2.55	6	1	2.55	6	1	2.55	6

Table 10. Length of α -decision rules for $\alpha \in \{0.1, 0.2, 0.3\}$ —generalized decision approach

Decision table	$\alpha = 0.1$			$\alpha = 0.2$			$\alpha = 0.3$		
	min	avg	max	min	avg	max	min	avg	max
balance-scale-1	2	2.00	2	1	1.02	2	1	1.00	1
breast-cancer-1	1	1.70	3	1	1.30	2	1	1.08	2
breast-cancer-5	1	1.30	2	1	1.00	1	1	1.00	1
cars-1	1	1.44	2	1	1.44	2	1	1.00	1
flags-5	1	1.22	2	1	1.01	2	1	1.00	1
hayes-roth-data-1	1	1.59	2	1	1.26	2	1	1.00	1
kr-vs-kp-5	1	1.71	4	1	1.24	3	1	1.04	2
kr-vs-kp-4	1	1.68	4	1	1.23	3	1	1.04	2
lymphography-5	1	1.82	3	1	1.33	2	1	1.07	2
mushroom-5	1	1.12	2	1	1.01	2	1	1.00	1
nursery-4	1	1.67	2	1	1.07	2	1	1.00	1
nursery-1	1	1.67	2	1	1.15	2	1	1.00	1
poker-hand-train-5	1	1.00	1	1	1.00	1	1	1.00	1
poker-hand-train-5a	1	1.00	1	1	1.00	1	1	1.00	1
poker-hand-train-5b	2	2.00	2	2	2.00	2	1	1.00	1
spect-test-1	1	1.59	3	1	1.14	2	1	1.05	2
teeth-1	1	1.50	2	1	1.14	2	1	1.05	2
teeth-5	1	1.71	3	1	1.29	2	1	1.07	2
tic-tac-toe-4	2	2.03	3	2	2.00	2	1	1.15	2
tic-tac-toe-3	2	2.03	3	2	2.00	2	1	1.27	2
zoo-data-5	1	1.81	4	1	1.33	3	1	1.21	2

Table 11. Number of different rules—generalized decision approach

Decision table	values of α					
	0.0	0.001	0.01	0.1	0.2	0.3
balance-scale-1	125	125	118	83	27	27
breast-cancer-1	169	169	164	117	65	48
breast-cancer-5	82	82	82	44	33	33
cars-1	141	141	94	40	40	15
flags-5	159	159	154	115	106	106
hayes-roth-data-1	25	25	25	22	19	18
kr-vs-kp-5	1133	1073	783	256	140	87
kr-vs-kp-4	1205	1141	817	265	142	90
lymphography-5	73	73	72	58	43	33
mushroom-5	543	556	481	214	76	64
nursery-4	41	41	16	16	12	12
nursery-1	572	406	202	79	38	16
poker-hand-train-5	2883	2388	1459	213	213	213
poker-hand-train-5a	2904	2351	1439	205	205	205
poker-hand-train-5b	998	998	743	280	280	94
spect-test-1	42	42	42	41	29	23
teeth-1	22	22	22	22	22	22
teeth-5	14	14	14	14	14	14
tic-tac-toe-4	131	131	117	53	53	28
tic-tac-toe-3	347	347	262	96	95	49
zoo-data-5	24	24	24	24	21	20

Table 12. Average percentage of rows separated at i -th step of the greedy algorithm work-generalized decision approach

Decision table	Number of step i									
	1	2	3	4	5	6	7	8	9	10
balance-scale-1	83.9	86.0	100							
breast-cancer-1	85.5	82.6	83.8	88.0	87.0	100				
breast-cancer-5	93.3	90.5	92.1	100						
cars-1	90.1	77.4	82.8	83.8	100					
flags-5	92.9	92.4	93.6	91.2	100					
hayes-roth-data-1	87.7	85.3	100							
kr-vs-kp-5	86.1	80.5	77.1	77.9	75.1	73.4	76.9	76.6	74.4	87.1
kr-vs-kp-4	86.4	80.4	77.4	78.2	75.2	72.6	75.9	76.2	73.7	81.1
lymphography-5	83.5	85.5	88.8	86.4	88.9	100				
mushroom-5	96.9	97.4	91.2	94.4	90.5	48.9	77.8	100		
nursery-4	90.3	97.3	98.6	100						
nursery-1	89.4	83.5	86.7	89.3	91.3	93.0	100			
poker-hand-train-5	92.4	77.5	79.2	84.3	100					
poker-hand-train-5a	92.3	77.2	79.4	84.6	100					
poker-hand-train-5b	75.7	77.0	78.9	82.5	100					
spect-test-1	91.9	94.9	81.0	84.9	75.0	66.7	25.0	33.3	50.0	100
teeth-1	87.4	89.8	100							
teeth-5	86.3	92.4	100							
tic-tac-toe-4	73.7	79.8	84.7	89.5	100					
tic-tac-toe-3	73.4	76.2	79.2	82.1	85.8	100				
zoo-data-5	83.9	80.7	72.6	81.1	90.0	100				

Table 13. Number of rows in decision tables for which 0.5-hypothesis is true-generalized decision approach

Decision table	rows	rows with 0.5-hypothesis true
balance-scale-1	125	125
breast-cancer-1	193	192
breast-cancer-5	98	98
cars-1	432	432
flags-5	171	171
hayes-roth-data-1	39	39
kr-vs-kp-5	1987	1799
kr-vs-kp-4	2061	1871
lymphography-5	122	120
mushroom-5	4078	4072
nursery-4	240	240
nursery-1	4320	4320
poker-hand-train-5	3324	3324
poker-hand-train-5a	3323	3323
poker-hand-train-5b	1024	1024
spect-test-1	164	163
teeth-1	22	22
teeth-5	14	14
tic-tac-toe-4	231	231
tic-tac-toe-3	449	446
zoo-data-5	42	39

We computed the number of different rules constructed by the greedy algorithm for $\alpha \in \{0.0, 0.001, 0.01, 0.1, 0.2, 0.3\}$. Results can be found in Table 11. For the generalized decision approach, in the worst case, the number of different rules can be equal to the number of rows in decision table T . With the exception of one table, the number of different rules is nonincreasing with the growth of α .

The last group of experiments is connected with 0.5-hypothesis. Table 12 contains, for $i = 1, 2, \dots$, the average percentage of rows separated at i -th step of the greedy algorithm work (average among all rows r). For two decision tables, the average percentage of separated rows is less than 50%: for “spect-test-1” – at the 7-th and the 8-th step of the greedy algorithm work, for “mushroom-5” – at the 6-th step of the greedy algorithm work.

We say that 0.5-hypothesis is true for T , if it is true for each row of T . Table 13 contains, for decision tables described in Table 2, number of rows for which 0.5-hypothesis is true. From 21 decision tables, the 0.5-hypothesis is not true for 8 of them: “breast-cancer-1”, “kr-vs-kp-5”, “kr-vs-kp-4”, “lymphography-5”, “mushroom-5”, “spect-test-1”, “tic-tac-toe-3” and “zoo-data-5”.

7.3 Comparative Study

In this section, we make comparative study of α -decision rules for the proposed approach and the generalized decision approach, relative to:

- length of constructed α -decision rules,
- number of different α -decision rules,
- 0.5-hypothesis.

Table 14, based on results from Tables 3 and 9 presents, for $\alpha \in \{0.0, 0.001, 0.01\}$, a comparison of minimum (column “min”), average (column “avg”) and maximum (column “max”) length of α -decision rules for both approaches. Each input of Table 14 is equal to the (min, avg, max) length of α -decision rules for the generalized decision approach divided by the (min, avg, max) length of α -decision rules for proposed approach.

We can find decision tables for which minimum, average and maximum length of α -decision rules constructed using the proposed approach is two or more times shorter than minimum, average and maximum length of α -decision rules constructed using generalized decision approach. However, for the maximum values of length of 0.01-decision rules for decision tables “poker-hand-train-5” and “poker-hand-train-5a” we have an opposite situation.

Table 15, based on results from Tables 4 and 10 presents, for $\alpha \in \{0.1, 0.2, 0.3\}$, comparison of minimum (column “min”), average (column “avg”) and maximum (column “max”) length of α -decision rules for both approaches. Each input of Table 15 is equal to the corresponding input of Table 10 divided by the input of Table 4.

Results are similar to the results from Table 14.

Table 16, based on results from Tables 5 and 11 presents, for $\alpha \in \{0.0, 0.001, 0.01, 0.1, 0.2, 0.3\}$, a comparison of the number of different α -decision rules

Table 14. Comparison of length of α -decision rules for $\alpha \in \{0.0, 0.001, 0.01\}$

Decision table	$\alpha = 0.0$			$\alpha = 0.001$			$\alpha = 0.01$		
	min	avg	max	min	avg	max	min	avg	max
balance-scale-1	1.50	1.50	1.50	1.50	1.50	1.50	1.00	1.46	1.50
breast-cancer-1	1.00	1.12	1.20	1.00	1.12	1.20	1.00	1.10	1.20
breast-cancer-5	1.00	1.38	1.33	1.00	1.38	1.33	1.00	1.38	1.33
cars-1	1.00	1.82	1.25	1.00	1.82	1.25	1.00	1.59	1.33
flags-5	1.00	1.02	1.00	1.00	1.02	1.00	1.00	1.02	1.00
hayes-roth-data-1	1.00	1.34	1.50	1.00	1.34	1.50	1.00	1.34	1.50
kr-vs-kp-5	1.00	1.17	1.09	1.00	1.08	1.00	1.00	1.06	1.00
kr-vs-kp-4	1.00	1.16	1.09	1.00	1.07	1.00	1.00	1.04	1.00
lymphography-5	1.00	1.10	1.20	1.00	1.10	1.20	1.00	1.09	1.20
mushroom-5	1.00	1.01	1.60	1.00	1.01	1.20	1.00	1.01	1.00
nursery-4	1.00	1.52	2.00	1.00	1.52	2.00	1.00	1.27	2.00
nursery-1	1.00	1.62	1.40	1.00	1.49	1.20	1.00	1.31	1.33
poker-hand-train-5	2.00	2.03	1.00	2.00	1.74	1.00	1.50	1.32	0.75
poker-hand-train-5a	2.00	2.13	1.00	2.00	1.82	1.25	1.50	1.44	0.75
poker-hand-train-5b	2.00	2.47	1.00	2.00	2.47	1.00	1.50	1.92	1.00
spect-test-1	1.00	1.30	2.00	1.00	1.30	2.00	1.00	1.30	2.25
teeth-1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
teeth-5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
tic-tac-toe-4	1.50	1.69	1.25	1.50	1.69	1.25	1.50	1.52	1.25
tic-tac-toe-3	1.00	1.38	1.00	1.00	1.38	1.00	1.00	1.15	1.00
zoo-data-5	1.00	1.16	1.20	1.00	1.16	1.20	1.00	1.16	1.20

Table 15. Comparison of length of α -decision rules for $\alpha \in \{0.1, 0.2, 0.3\}$

Decision table	$\alpha = 0.1$			$\alpha = 0.2$			$\alpha = 0.3$		
	min	avg	max	min	avg	max	min	avg	max
balance-scale-1	2.00	1.89	1.00	1.00	1.02	2.00	1.00	1.00	1.00
breast-cancer-1	1.00	0.97	1.00	1.00	1.02	1.00	1.00	1.07	1.00
breast-cancer-5	1.00	1.13	1.00	1.00	1.00	1.00	1.00	1.00	1.00
cars-1	1.00	1.18	1.00	1.00	1.23	1.00	1.00	1.00	1.00
flags-5	1.00	1.02	1.00	1.00	0.99	1.00	1.00	1.00	1.00
hayes-roth-data-1	1.00	1.59	2.00	1.00	1.26	2.00	1.00	1.00	1.00
kr-vs-kp-5	1.00	0.97	1.00	1.00	1.06	1.50	1.00	1.00	1.00
kr-vs-kp-4	1.00	0.97	1.00	1.00	1.06	1.00	1.00	1.01	1.00
lymphography-5	1.00	1.01	1.00	1.00	0.97	1.00	1.00	1.06	1.00
mushroom-5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
nursery-4	1.00	1.65	1.00	1.00	1.07	2.00	1.00	1.00	1.00
nursery-1	1.00	1.21	1.00	1.00	1.15	1.00	1.00	1.00	1.00
poker-hand-train-5	1.00	0.94	0.50	1.00	1.00	1.00	1.00	1.00	1.00
poker-hand-train-5a	1.00	0.99	0.50	1.00	1.00	1.00	1.00	1.00	1.00
poker-hand-train-5b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
spect-test-1	1.00	1.27	1.00	1.00	1.06	1.00	1.00	0.98	1.00
teeth-1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
teeth-5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
tic-tac-toe-4	2.00	1.38	1.00	2.00	1.92	1.00	1.00	1.14	1.00
tic-tac-toe-3	1.00	1.02	1.00	2.00	1.33	1.00	1.00	1.26	1.00
zoo-data-5	1.00	1.07	1.00	1.00	1.06	1.00	1.00	1.08	1.00

Table 16. Comparison of number of different rules

Decision table	values of α					
	0.0	0.001	0.01	0.1	0.2	0.3
balance-scale-1	2.45	2.45	2.31	1.84	1.00	1.17
breast-cancer-1	1.03	1.03	1.00	1.04	1.33	1.37
breast-cancer-5	1.34	1.34	1.34	1.16	1.32	1.32
cars-1	6.13	6.13	4.95	3.33	3.64	1.50
flags-5	1.00	1.00	0.99	1.04	1.05	1.08
hayes-roth-data-1	1.79	1.79	1.79	3.14	3.17	3.00
kr-vs-kp-5	1.32	1.25	1.36	1.36	1.51	1.34
kr-vs-kp-4	1.32	1.25	1.32	1.26	1.35	1.25
lymphography-5	1.24	1.24	1.24	1.32	1.34	1.43
mushroom-5	1.10	1.09	1.07	1.30	1.04	1.10
nursery-4	5.86	5.86	2.29	1.23	2.40	2.40
nursery-1	6.43	5.49	6.12	3.43	3.80	1.60
poker-hand-train-5	7.35	6.09	3.63	1.57	5.92	5.92
poker-hand-train-5a	8.72	7.06	5.64	2.20	6.03	6.03
poker-hand-train-5b	19.19	19.19	14.29	5.38	5.38	11.75
spect-test-1	1.45	1.45	1.45	1.46	1.32	1.05
teeth-1	1.00	1.00	1.00	1.00	1.00	1.00
teeth-5	1.00	1.00	1.00	1.00	1.00	1.00
tic-tac-toe-4	2.52	2.52	2.25	1.77	3.53	2.00
tic-tac-toe-3	2.21	2.21	1.91	1.52	1.67	3.06
zoo-data-5	1.50	1.50	1.50	1.50	1.40	1.33

for both approaches. Each input of Table 16 is equal to the number of different α -decision rules for the generalized decision approach divided by the number of different α -decision rules for the proposed approach. We can see that often the number of different α -decision rules for the generalized decision approach is two or more times greater than the number of different rules for the proposed approach.

The last group of results is connected with 0.5-hypothesis. Based on results from Tables 7 and 13 we can see that, for the proposed approach, the 0.5-hypothesis is not true for 9 decision tables, for generalized decision approach, the 0.5-hypothesis is not true for 8 decision tables. So, the difference is not significant.

8 Conclusions

We studied the greedy algorithm for construction of approximate decision rules. This algorithm has polynomial time complexity for the whole set of decision tables with many-valued decisions. We obtained a bound on precision of this algorithm relative to the length of rules, and considered lower and upper bounds on the minimum length of α -decision rules. We studied binary decision tables with many-valued decisions but the considered approach can be used also for decision tables with more than two values of attributes, as presented in Sect. 7.

Experimental results are connected with the construction of exact and approximate decision rules. Based on them, we can see, that the greedy algorithm constructs relatively short α -decision rules. We also presented results relative to length, number of different α -decision rules and 0.5-hypothesis for the approach based on generalized decision.

Based on results connected with comparison of two approaches we can see that the length and number of different rules constructed in the framework of our approach (one decision from the set of decisions attached to row) are usually smaller than the length and number of different rules constructed in the framework of the generalized decision approach (all decisions from the set of decisions attached to row).

Future investigations will be connected with the study of other greedy algorithms and construction of classifiers for decision tables with many-valued decisions.

Acknowledgements. Research reported in this publication was supported by the King Abdullah University of Science and Technology (KAUST).

The authors wish to express their gratitude to anonymous reviewers for useful comments.

References

1. Alkhalid, A., Amin, T., Chikalov, I., Hussain, S., Moshkov, M., Zielosko, B.: Dagger: a tool for analysis and optimization of decision trees and rules. *Comput. Inf. Soc. Factors New Inf. Technol. Hypermedia Perspect. Avant-Garde Experiences Era of Communicability Expansion*, 29–39 (2011)
2. Azad, M., Chikalov, I., Moshkov, M., Zielosko, B.: Greedy algorithms for construction of approximate tests for decision tables with many-valued decisions. *Fundamenta Informaticae* **120**(3–4), 231–242 (2012)
3. Blockeel, H., Schietgat, L., Struyf, J., Džeroski, S., Clare, A.: Decision trees for hierarchical multilabel classification: a case study in functional genomics. In: Fürnkranz, J., Scheffer, T., Spiliopoulou, M. (eds.) *PKDD 2006*. LNCS (LNAI), vol. 4213, pp. 18–29. Springer, Heidelberg (2006). doi:[10.1007/11871637_7](https://doi.org/10.1007/11871637_7)
4. Boutell, M.R., Luo, J., Shen, X., Brown, C.M.: Learning multi-label scene classification. *Pattern Recogn.* **37**(9), 1757–1771 (2004)
5. Cheriyan, J., Ravi, R.: Lecture notes on approximation algorithms for network problems (1998). <http://www.math.uwaterloo.ca/~jcheriya/lecnotes.html>
6. Chikalov, I., Zielosko, B.: Decision rules for decision tables with many-valued decisions. In: Yao, J.T., Ramanna, S., Wang, G., Suraj, Z. (eds.) *RSKT 2011*. LNCS (LNAI), vol. 6954, pp. 763–768. Springer, Heidelberg (2011). doi:[10.1007/978-3-642-24425-4_95](https://doi.org/10.1007/978-3-642-24425-4_95)
7. Clare, A., King, R.D.: Knowledge discovery in multi-label phenotype data. In: Raedt, L., Siebes, A. (eds.) *PKDD 2001*. LNCS (LNAI), vol. 2168, pp. 42–53. Springer, Heidelberg (2001). doi:[10.1007/3-540-44794-6_4](https://doi.org/10.1007/3-540-44794-6_4)
8. Comité, F., Gilleron, R., Tommasi, M.: Learning multi-label alternating decision trees from texts and data. In: Perner, P., Rosenfeld, A. (eds.) *MLDM 2003*. LNCS, vol. 2734, pp. 35–49. Springer, Heidelberg (2003). doi:[10.1007/3-540-45065-3_4](https://doi.org/10.1007/3-540-45065-3_4)

9. Feige, U.: A threshold of $\ln n$ for approximating set cover. *J. ACM (JACM)* **45**(4), 634–652 (1998)
10. Greco, S., Matarazzo, B., Słowiński, R.: Rough sets theory for multicriteria decision analysis. *Eur. J. Oper. Res.* **129**(1), 1–47 (2001)
11. Kryszkiewicz, M.: Rules in incomplete information systems. *Inf. Sci.* **113**(34), 271–292 (1999)
12. Lichman, M.: UCI Machine Learning Repository (2013)
13. Lipski, W.: On databases with incomplete information. *J. ACM (JACM)* **28**(1), 41–70 (1981)
14. Lipski Jr., W.: On semantic issues connected with incomplete information databases. *ACM Trans. Database Syst.* **4**(3), 262–296 (1979)
15. Mencia, E.L., Furnkranz, J.: Pairwise learning of multilabel classifications with perceptrons. In: *IEEE International Joint Conference on Neural Networks, 2008, IJCNN 2008 (IEEE World Congress on Computational Intelligence)*, pp. 2899–2906 (2008)
16. Moshkov, M.J., Piliszczuk, M., Zielosko, B.: *Partial Covers, Reducts and Decision Rules in Rough Sets—Theory and Applications*. SCI, vol. 145. Springer, Heidelberg (2008)
17. Moshkov, M., Zielosko, B.: *Combinatorial Machine Learning—A Rough Set Approach*. SCI, vol. 360. Springer, Heidelberg (2011)
18. Moshkov, M., Zielosko, B.: Construction of α -decision trees for tables with many-valued decisions. In: Yao, J.T., Ramanna, S., Wang, G., Suraj, Z. (eds.) *RSKT 2011. LNCS (LNAI)*, vol. 6954, pp. 486–494. Springer, Heidelberg (2011). doi:[10.1007/978-3-642-24425-4_63](https://doi.org/10.1007/978-3-642-24425-4_63)
19. Moshkov, M.J.: Greedy algorithm for decision tree construction in context of knowledge discovery problems. In: Tsumoto, S., Słowiński, R., Komorowski, J., Grzymała-Busse, J.W. (eds.) *RSCTC 2004. LNCS (LNAI)*, vol. 3066, pp. 192–197. Springer, Heidelberg (2004). doi:[10.1007/978-3-540-25929-9_22](https://doi.org/10.1007/978-3-540-25929-9_22)
20. Nguyen, H.S., Slezak, D.: Approximate reducts and association rules - correspondence and complexity results. In: *Proceedings of the 7th International Workshop on New Directions in Rough Sets, Data Mining, and Granular-Soft Computing. RSFDGrC 1999*, pp. 137–145. Springer, London (1999)
21. Orowska, E., Pawlak, Z.: Representation of nondeterministic information. *Theoret. Comput. Sci.* **29**(12), 27–39 (1984)
22. Pawlak, Z.: *Rough Sets—Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishers, Dordrecht (1991)
23. Pawlak, Z., Skowron, A.: Rough sets and boolean reasoning. *Inf. Sci.* **177**(1), 41–73 (2007)
24. Pawlak, Z., Skowron, A.: Rough sets: some extensions. *Inf. Sci.* **177**(1), 28–40 (2007)
25. Pawlak, Z., Skowron, A.: Rudiments of rough sets. *Inf. Sci.* **177**(1), 3–27 (2007)
26. Rissanen, J.: Modeling by shortest data description. *Automatica* **14**(5), 465–471 (1978)
27. Sakai, H., Ishibashi, R., Koba, K., Nakata, M.: Rules and apriori algorithm in non-deterministic information systems. In: Peters, J.F., Skowron, A., Rybiński, H. (eds.) *Transactions on Rough Sets IX. LNCS*, vol. 5390, pp. 328–350. Springer, Heidelberg (2008). doi:[10.1007/978-3-540-89876-4_18](https://doi.org/10.1007/978-3-540-89876-4_18)
28. Sakai, H., Nakata, M., Ślęzak, D.: Rule generation in lipski’s incomplete information databases. In: Szczuka, M., Kryszkiewicz, M., Ramanna, S., Jensen, R., Hu, Q. (eds.) *RSCTC 2010. LNCS (LNAI)*, vol. 6086, pp. 376–385. Springer, Heidelberg (2010). doi:[10.1007/978-3-642-13529-3_40](https://doi.org/10.1007/978-3-642-13529-3_40)

29. Sakai, H., Nakata, M., Ślęzak, D.: A prototype system for rule generation in lip-ski's incomplete information databases. In: Kuznetsov, S.O., Ślęzak, D., Hepting, D.H., Mirkin, B.G. (eds.) RSFDGrC 2011. LNCS (LNAI), vol. 6743, pp. 175–182. Springer, Heidelberg (2011). doi:[10.1007/978-3-642-21881-1_29](https://doi.org/10.1007/978-3-642-21881-1_29)
30. Skowron, A., Rauszer, C.: The discernibility matrices and functions in information systems. In: Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory, pp. 331–362. Kluwer Academic Publishers (1992)
31. Ślęzak, D.: Normalized decision functions and measures for inconsistent decision tables analysis. *Fundamenta Informaticae* **44**(3), 291–319 (2000)
32. Ślęzak, D.: Approximate entropy reducts. *Fundamenta Informaticae* **53**(3–4), 365–390 (2002)
33. Tsoumakas, G., Katakis, I.: Multi-label classification: an overview. *Int. J. Data Warehouse. Min.* **3**(3), 1–13 (2007)
34. Tsoumakas, G., Katakis, I., Vlahavas, I.: Mining multi-label data. In: Data Mining and Knowledge Discovery Handbook, pp. 667–685. Springer, US (2010)
35. Wiczorkowska, A., Synak, P., Lewis, R., Raś, Z.W.: Extracting emotions from music data. In: Hacid, M.-S., Murray, N.V., Raś, Z.W., Tsumoto, S. (eds.) ISMIS 2005. LNCS (LNAI), vol. 3488, pp. 456–465. Springer, Heidelberg (2005). doi:[10.1007/11425274_47](https://doi.org/10.1007/11425274_47)
36. Zhou, Z.H., Jiang, K., Li, M.: Multi-instance learning based web mining. *Appl. Intell.* **22**(2), 135–147 (2005)
37. Zhou, Z.H., Zhang, M.L., Huang, S.J., Li, Y.F.: Multi-instance multi-label learning. *Artif. Intell.* **176**(1), 2291–2320 (2012)