Proceedings of the Federated Conference on
Computer Science and Information Systems pp. 271–277
ISBN 978-83-60810-51-4

Tests for Decision Tables with Many-Valued Decisions—Comparative Study

Mohammad Azad[1], Igor Chikalov[1], Mikhail Moshkov[1], Beata Zielosko[1,2]
[1]Mathematical and Computer Sciences & Engineering Division
King Abdullah University of Science and Technology
Thuwal 23955-6900, Saudi Arabia
Email: {mohammad.azad, igor.chikalov, mikhail.moshkov, beata.zielosko}@kaust.edu.sa
[2]Institute of Computer Science
University of Silesia
39, Będzińska St., 41-200 Sosnowiec, Poland

***Abstract*—The paper is devoted to the study of a greedy algorithm for construction of approximate tests (super-reducts). 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. The idea of algorithm is connected with so-called boundary subtables. After constructing a test we use algorithm which tries to remove attributes from a test and obtain a reduct. We present experimental results connected with the cardinality of tests and reducts for randomly generated tables and data sets from UCI Machine Learning Repository which were converted to decision tables with many-valued decisions. To make some comparative study we presents also experimental results for greedy algorithm which constructs a test based on generalized decision approach.**

I. Introduction

In THE paper we study tests for decision tables with many-valued decisions. In such table, 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.

We can meet tables with many-valued decisions when we work with statistical or experimental data. In this case we often have groups of rows (objects) with equal values of conditional attributes but, probably, different values of the decision attribute. Instead of a group of rows, we can consider one row given by values of conditional attributes. We attach to this row a set of decisions: either all decisions for objects from the group, or k the most frequent decisions for objects from the group [1].

In the rough set theory [2], [3] decision tables are often considered that have equal rows labeled with different decisions. The set of decisions attached to equal rows is called the *generalized decision* for each of these equal rows [4], [5]. The usual way is to find for a given row its generalized decision. In the paper, this approach is called as approach based on generalized decision. However, the problem of finding an arbitrary decision or one of the most frequent decisions from the generalized decision (set of decisions) is interesting also. So, the study of decision tables with many-valued decisions can give us new tool for the rough set theory.

Reducts are minimal subsets of attributes (with respect to inclusion) which discern all pairs of objects with different decisions that are discernible by the whole set of attributes. Tests [6] (super-reducts) and reducts are used for feature selection and for construction of classifiers. In various applications, we often deal with decision tables which contain noisy data. In this case, exact tests and reducts can be "over-fitted", i.e., they depend essentially on the noise. So, instead of exact tests and reducts it is more appropriate often to work with approximate ones. In particular, they have smaller number of attributes.

Problem of construction of test with minimum cardinality is NP-hard [7], [8], [9]. Therefore, we consider approximate polynomial algorithm for test optimization.

In this paper, we present a greedy algorithm for construction of exact and approximate tests (super-reducts) for decision tables with many-valued decisions. Similar algorithm was presented in [10] but in this paper we also present algorithms for transformation a test into a reduct and some comparison with the approach based on generalized decision. Theoretical results for proposed approach were presented in [11] and in [1] for exact tests. We prove a bound on precision of the greedy algorithm for construction of tests. For an arbitrary natural t, the greedy algorithm for test construction has polynomial time complexity on tables which have at most t decisions in each set of decisions attached to rows. In this paper, we study binary decision tables with many-valued decisions. However, the obtained results can be extended to the decision tables filled by numbers from the set $\{0, \ldots, k-1\}$, where $k \geq 3$. We present experimental results for randomly generated decision tables and for data sets from UCI Machine Learning Repository [12]. We also make some comparative study for cardinality of tests based on proposed approach and based on generalized decision approach.

This paper consists of six sections. Section II, contains main notions. In Section III, we consider decision tables which have at most t decisions in each set of decisions attached to rows. In Section IV, we present a greedy algorithm for construction of approximate tests and algorithms for transformation of the test to the reduct. Section V contains results of experiments and Section VI – conclusions.

II. Main Notions

In this section, we consider definitions of notions 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 $f_1,\ldots,f_n$. 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 decision 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 a decision table T_0 with many-valued decisions can be found in Table I.

TABLE I
DECISION TABLE T_0 WITH MANY-VALUED DECISIONS

$T_0 =$

	f_1	f_2	f_3	d
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	$\{1,2\}$

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.

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.

A table obtained from T by removal of some rows is called a *subtable* of T. A subtable T' of T is called *boundary* subtable if T' is not degenerate but each proper subtable of T' is degenerate. We denote by $B(T)$ the number of boundary subtables of the table T. It is clear that T is a degenerate table if and only if $B(T)=0$. The value $B(T)$ will be interpreted as *uncertainty* of T.

All boundary subtables of the decision table T_0, for proposed approach, are depicted in Fig. 1.

$T_1 =$

	f_1	f_2	f_3	d
r_2	0	1	0	$\{1,3\}$
r_4	0	0	1	$\{2,3\}$
r_5	1	0	0	$\{1,2\}$

$T_2 =$

	f_1	f_2	f_3	d
r_1	1	1	1	$\{1\}$
r_4	0	0	1	$\{2,3\}$

$T_3 =$

	f_1	f_2	f_3	d
r_2	0	1	0	$\{1,3\}$
r_3	1	1	0	$\{2\}$

$T_4 =$

	f_1	f_2	f_3	d
r_1	1	1	1	$\{1\}$
r_3	1	1	0	$\{2\}$

Fig. 1. All boundary subtables of the decision table T_0

In case of generalized decision approach we consider sets of decisions attached to rows of the table T_0 as values of the decision attribute (see Fig. 2). The number of boundary subtables for such approach is equal to 10. Each boundary subtable of the table T_0 has 2 rows.

$T_0 =$

	f_1	f_2	f_3	d
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	$\{1,2\}$

$\Rightarrow$

d
1
2
3
4
5

Fig. 2. Transformation of decision attribute from the table T_0 for generalized decision approach

We will say that an attribute f_i *divides* a boundary subtable Θ of the table T if and only if this attribute is not constant on the rows of Θ (for example, for a binary decision table at the intersection with the column f_i we can find some rows which contain 1 and some rows which contain 0).

Let us define the notion of an α-test for the table T. Let α be a real number such that $0 \le \alpha < 1$.

An *α-test for the table* T is a subset of attributes $\{f_{i_1},\ldots,f_{i_m}\}$ such that these attributes divide at least $(1-\alpha)B(T)$ boundary subtables. Empty set is an α-test for T iff T is a degenerate table. An *α-reduct for the table* T is an α-test for T for which each proper subset is not an α-test. We denote by $R_{\min}(\alpha,T)$ the minimum cardinality of an α-test for the table T. It is clear that each α-test has an α-reduct as a subset. Therefore $R_{\min}(\alpha,T)$ is the minimum cardinality of an α-reduct.

III. Set $Tab(t)$ of Decision Tables

We denote by $Tab(t)$, where t is a natural number, the set of decision tables with many-valued decisions such that each row in the table has at most t decisions (is labeled with a set of decisions which cardinality is at most t).

The next statement was proved in [1]. For the completeness, we will give it with the proof.

Lemma 1: Let T' be a boundary subtable with m rows. Then each row of T' is labeled with a set of decisions whose cardinality is at least $m-1$.

Proof: Let rows of T' be labeled with sets of decisions $D_1,\ldots,D_m$ respectively. Then $D_1\cap\ldots\cap D_m=\emptyset$ and for any $i\in\{1,\ldots,m\}$, the set $D_1\cap\ldots\cap D_{i-1}\cap D_{i+1}\cap\ldots\cap D_m$ contains a number d_i. Assume that $i\ne j$ and $d_i=d_j$. Then $D_1\cap\ldots\cap D_m\ne\emptyset$ which is impossible. Therefore $d_1,\ldots,d_m$ are pairwise different numbers. It is clear that for $i=1,\ldots,m$, the set $\{d_1,\ldots,d_m\}\setminus\{d_i\}$ is a subset of the set D_i. ■

Corollary 1: Each boundary subtable of a table $T\in Tab(t)$ has at most $t+1$ rows.

Therefore, for tables from $Tab(t)$, there exists a polynomial algorithm for the finding of all boundary subtables and computation of parameter $B(T)$. For example, for any decision table T with one-valued decision (in fact, for any table from $Tab(1)$) the equality $B(T)=P(T)$ holds, where $P(T)$ is the number of unordered pairs of rows of T with different decisions.

IV. Algorithms for Test and Reduct Construction

Results presented in [10] show that problem of minimization of α-test cardinality, $0 \leq \alpha < 1$, is NP-hard. Based on [11] we know that under the assumption $NP \nsubseteq DTIME(n^{O(\log \log n)})$ the greedy algorithm is close to the best (from the point of view of precision) approximate polynomial algorithms for minimization of test cardinality. In this section, we present a greedy algorithm for construction of the test and algorithms for transformation of the test into the reduct.

We can describe the work with decision table with many-valued decisions, in the following way:

1) construction of all boundary subtables for a given decision table,
2) construction of an α-test using greedy algorithm,
3) transformation of the α-test into an α-reduct.

We use the greedy algorithm for set cover problem to construct an α-test (if $\alpha = 0$ we consider an exact test). Let T be a table with many-valued decisions containing n columns labeled with attributes $f_1, \ldots, f_n$. Let α be a real number such that $0 < \alpha < 1$, and $B(T)$ be the number of boundary subtables of the table T. Greedy algorithm at each iteration chooses an attribute which divides the maximum number of not divided boundary subtables. This algorithm stops if attributes from the test divide at least $(1 - \alpha)B(T)$ boundary subtables (see Algorithm 1). For example, if $\alpha = 0.1$ an α-test divides at least 90% of boundary subtables. By $R_{\mathrm{greedy}}(\alpha, T)$ we denote the cardinality of α-test constructed by the greedy algorithm. By $R_{\min}(\alpha, T)$ we denote the minimum cardinality of α-test.

Theorem 1: [11] Let T be a nondegenerate decision table with many-valued decisions and α be a real number such that $0 < \alpha < 1$. Then

$$R_{\mathrm{greedy}}(\alpha, T) \leq R_{\min}(0, T) \ln(1/\alpha) + 1.$$

If, for a fixed natural t, we apply the considered algorithm to decision tables from $Tab(t)$ then the time complexity of this algorithm (including construction of all boundary subtables) will be bounded from above by a polynomial depending on the length of decision table description.

Algorithm 1 Greedy algorithm for α-test construction

Require: $B(T)$, all boundary subtables of decision table T with attributes $f_1, \ldots, f_n$, $\alpha \in \mathbb{R}$, $0 \leq \alpha < 1$,
Ensure: an α-test for T

```
Q ← ∅;
while attributes from Q divide less than (1 − α)B(T) boundary subtables do
    select f_i ∈ {f_1, ..., f_n} with minimum index such that
    f_i divides the maximum number of boundary subtables
    not divided by attributes from Q
    Q ← Q ∪ {f_i};
end while
```

To transform an α-test into an α-reduct for T we consider three cases which are differ only in a way of choosing attributes from the test:

- LEFT: chooses attributes from an α-test from the left-hand-side (see Algorithm 2),
- RIGHT: chooses attributes from an α-test from the right-hand-side (see Algorithm 3),
- RANDOM: chooses attributes from an α-test randomly (see Algorithm 4).

In the following algorithms, we assume that Q is an ordered set.

Algorithm 2 Transformation of an α-test into an α-reduct: LEFT

Require: decision table T, an α-test $Q = \{f_1, \ldots, f_k\}$
Ensure: an α-reduct for T

```
Q′ ← ∅
for all f_i ∈ Q do
    for i = 1, ..., k
    Q′ ← Q \ {f_i}
    if Q′ is an α-test for T then
        Q ← Q′;
    end if
end for
```

Algorithm 3 Transformation of an α-test into an α-reduct: RIGHT

Require: decision table T, an α-test $Q = \{f_1, \ldots, f_k\}$
Ensure: an α-reduct for T

```
Q′ ← ∅
for all f_i ∈ Q do
    for i = k, ..., 1
    Q′ ← Q \ {f_i}
    if Q′ is an α-test for T then
        Q ← Q′;
    end if
end for
```

Algorithm 4 Transformation of an α-test into an α-reduct: RANDOM

Require: decision table T, an α-test $Q = \{f_1, \ldots, f_k\}$
Ensure: an α-reduct for T

```
Q* ← random permutation of Q
use the Algorithm 2 with Q*
```

V. Experimental Results

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 II. This table contains name of initial table, number of rows (column "Rows"), number of attributes (column "Attr"), spectrum of this table (column "Spectrum"), and number of removed attributes (column "Removed attributes"). Spectrum of decision table with many-valued decisions is a sequence $\#1$, $\#2$,..., where $\#i$, $i = 1, 2, \ldots$, is the number of rows labeled with sets of decision with the cardinality equal to i.

TABLE II
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				1
breast-cancer-1	193	8	169	24	0				1
breast-cancer-5	98	4	58	40					5
cars-1	432	5	258	161	13				1
flags-5	171	21	159	12					5
hayes-roth-data-1	39	3	22	13	4				1
kr-vs-kp-5	1987	31	1564	423					5
kr-vs-kp-4	2061	32	1652	409					4
lymphography-5	122	13	113	9					5
mushroom-5	4078	17	4048	30					5
nursery-4	240	4	97	96	47				4
nursery-1	4320	7	2858	1460	2				1
poker-hand-train-5	3324	5	156	1832	1140	188	7	1	5
poker-hand-train-5a	3323	5	130	1850	1137	199	6	1	5
poker-hand-train-5b	1024	5	0	246	444	286	44	4	5
spect-test-1	164	21	161	3					1
teeth-1	22	7	12	10					1
teeth-5	14	3	6	3	0	5	0	2	5
tic-tac-toe-4	231	5	102	129					4
tic-tac-toe-3	449	6	300	149					3
zoo-data-5	42	11	36	6					5

For decision tables described in Table II and $\alpha \in \{0.0, 0.001, 0.01, 0.1, 0.2, 0.5\}$, we constructed by the greedy algorithm α-tests. Table III presents the cardinality of constructed tests. Based on presented results we can see that the cardinality of α-test is decreasing when the value of α is increasing, till $\alpha = 0.5$ – in this case 0.5-tests for all datasets are equal to 1. For example, for data sets "kr-vs-kp-5" and "kr-vs-kp-4", the exact test contains 26 and 27 attributes respectively but 0.01-tests separates 99% of boundary subtables and contain only 7 attributes when the number of conditional attributes is 31 and 32 respectively.

Next, based on Algorithms 2, 3 and 4 we try to remove from an α-test some attributes and obtain an α-reduct. For all three cases (LEFT, RIGHT, RANDOM) we get the same results. Table IV, presents the cardinality of α-reducts. For four data sets ("cars-1", "kr-vs-kp-5", "kr-vs-kp-4", "spect-test-1") the difference in cardinality between tests and reducts is equal to 1, for "zoo-data-5" – 2 (such values are in bold).

We also considered randomly generated binary decision tables with many-valued decisions. Such tables contain: $n \in \{50, 80, 100\}$ rows, $m \in \{20, 30, 50\}$ conditional attributes, and $k \in \{2, 3\}$ values of decision attribute. For $k = 2$, values

TABLE III
CARDINALITY OF α-TESTS CONSTRUCTED BY THE GREEDY ALGORITHM

Decision table	α					
	0.0	0.001	0.01	0.1	0.2	0.5
balance-scale-1	2	2	2	1	1	1
breast-cancer-1	8	6	4	2	2	1
breast-cancer-5	4	4	3	1	1	1
cars-1	5	4	3	2	2	1
flags-5	13	9	4	2	1	1
hayes-roth-data-1	2	2	2	2	1	1
kr-vs-kp-5	26	12	7	4	3	1
kr-vs-kp-4	27	12	7	4	3	1
lymphography-5	11	9	5	2	2	1
mushroom-5	8	4	3	2	1	1
nursery-4	2	2	1	1	1	1
nursery-1	7	3	2	1	1	1
poker-hand-train-5	5	2	1	1	1	1
poker-hand-train-5a	5	2	1	1	1	1
poker-hand-train-5b	5	5	4	2	2	1
spect-test-1	10	10	6	3	3	1
teeth-1	5	5	4	2	2	1
teeth-5	3	3	3	2	2	1
tic-tac-toe-4	5	5	4	2	2	1
tic-tac-toe-3	6	5	4	2	2	1
zoo-data-5	9	9	7	3	2	1

TABLE IV
CARDINALITY OF α-REDUCTS OBTAINED FROM α-TESTS

Decision table	α					
	0.0	0.001	0.01	0.1	0.2	0.5
balance-scale-1	2	2	2	1	1	1
breast-cancer-1	8	6	4	2	2	1
breast-cancer-5	4	4	3	1	1	1
cars-1	**4**	4	3	2	2	1
flags-5	13	9	4	2	1	1
hayes-roth-data-1	2	2	2	2	1	1
kr-vs-kp-5	**25**	12	7	4	3	1
kr-vs-kp-4	**26**	12	7	4	3	1
lymphography-5	11	9	5	2	2	1
mushroom-5	8	4	3	2	1	1
nursery-4	2	2	1	1	1	1
nursery-1	7	3	2	1	1	1
poker-hand-train-5	5	2	1	1	1	1
poker-hand-train-5a	5	2	1	1	1	1
poker-hand-train-5b	5	5	4	2	2	1
spect-test-1	**9**	**9**	6	3	3	1
teeth-1	5	5	4	2	2	1
teeth-5	3	3	3	2	2	1
tic-tac-toe-4	5	5	4	2	2	1
tic-tac-toe-3	6	5	4	2	2	1
zoo-data-5	**7**	**7**	7	3	2	1

of decision attribute are from $\{1 - 10\}$ and $\{11 - 20\}$, for $k = 3$, values of decision attribute are from $\{1-10\}$, $\{11-20\}$ and $\{21 - 30\}$. For each triple (n, m, k) we generated 10 decision tables such that each value of conditional attribute is equal to b, $b \in \{0, 1\}$ with probability $\frac{1}{2}$. Probabilities of values of decision attribute are the same and equal to $\frac{1}{10}$.

Table V presents the average cardinality of constructed α-tests for randomly generated binary decision tables. For $\alpha \in \{0.0, 0.001, 0.01, 0.1\}$, we calculate the standard deviation (column "std") and denote the cardinality of test by R. Based on presented results we can say that the greedy algorithm constructs relatively short α-tests.

Using Algorithms 2, 3 and 4 we try to remove from the test some attributes and obtain the reduct. For all three cases (LEFT, RIGHT, RANDOM) we get the same results. Table VI, presents the average cardinality of α-reducts. For

TABLE V
AVERAGE CARDINALITY OF α-TESTS CONSTRUCTED BY THE GREEDY ALGORITHM FOR RANDOMLY GENERATED TABLES

Decision	$\alpha = 0.0$		$\alpha = 0.001$		$\alpha = 0.01$		$\alpha = 0.1$		α	α
table	R	std	R	std	R	std	R	std	0.2	0.5
T_100_20_2	9.9	0.5	8.8	0.4	6.5	0.5	4.0	0.0	3.0	1.0
T_100_20_3	9.9	0.5	8.8	0.4	6.0	0.0	3.1	0.3	3.0	1.0
T_100_30_2	9.5	0.5	8.5	0.5	6.4	0.5	4.0	0.0	3.0	1.0
T_100_30_3	9.1	0.3	8.1	0.3	6.0	0.0	3.1	0.3	3.0	1.0
T_100_50_2	9.1	0.3	8.2	0.4	6.0	0.0	4.0	0.0	3.0	1.0
T_100_50_3	9.0	0.0	8.0	0.0	6.0	0.0	3.0	0.0	2.9	1.0
T_50_20_2	7.7	0.5	7.6	0.5	6.0	0.0	3.6	0.5	3.0	1.0
T_50_20_3	7.4	0.5	7.4	0.5	6.0	0.0	3.0	0.0	3.0	1.0
T_50_30_2	7.7	0.5	7.4	0.7	5.9	0.3	3.7	0.5	3.0	1.0
T_50_30_3	7.2	0.4	7.2	0.4	6.0	0.0	3.0	0.0	3.0	1.0
T_50_50_2	7.0	0.0	7.0	0.0	6.0	0.0	3.2	0.4	3.0	1.0
T_50_50_3	7.0	0.0	7.0	0.0	5.8	0.4	3.0	0.0	2.8	1.0
T_80_20_2	9.2	0.9	8.3	0.5	6.3	0.5	4.0	0.0	3.0	1.0
T_80_20_3	9.2	0.4	8.2	0.4	6.0	0.0	3.0	0.0	3.0	1.0
T_80_30_2	8.8	0.4	8.1	0.3	6.0	0.0	4.0	0.0	3.0	1.0
T_80_30_3	8.7	0.5	7.9	0.3	6.0	0.0	3.2	0.4	3.0	1.0
T_80_50_2	8.4	0.5	8.0	0.0	6.0	0.0	4.0	0.0	3.0	1.0
T_80_50_3	8.1	0.3	8.0	0.0	6.0	0.0	3.0	0.0	3.0	1.0

$\alpha \in \{0.0, 0.001, 0.01, 0.1\}$, we calculate the standard deviation (column "std") and denote the cardinality of reduct by Rm. The difference in average cardinality between tests and reducts is very small (such values are in bold).

TABLE VI
AVERAGE CARDINALITY OF α-REDUCTS OBTAINED FROM α-TESTS FOR RANDOMLY GENERATED TABLES

Decision	$\alpha = 0.0$		$\alpha = 0.001$		$\alpha = 0.01$		$\alpha = 0.1$		α	α
table	Rm	std	Rm	std	Rm	std	Rm	std	0.2	0.5
T_100_20_2	**9.6**	0.5	8.8	0.4	6.4	0.5	4.0	0.0	3.0	1.0
T_100_20_3	**9.7**	0.5	**8.7**	0.5	6.0	0.0	3.1	0.3	3.0	1.0
T_100_30_2	9.5	0.5	**8.4**	0.5	**6.2**	0.4	4.0	0.0	3.0	1.0
T_100_30_3	9.1	0.3	8.1	0.3	6.0	0.0	**3.0**	0.0	3.0	1.0
T_100_50_2	**9.0**	0.0	**8.1**	0.3	6.0	0.0	4.0	0.0	3.0	1.0
T_100_50_3	9.0	0.0	8.0	0.0	6.0	0.0	3.0	0.0	2.9	1.0
T_50_20_2	**7.6**	0.5	**7.4**	0.5	6.0	0.0	3.6	0.5	3.0	1.0
T_50_20_3	7.4	0.5	7.4	0.5	6.0	0.0	3.0	0.0	3.0	1.0
T_50_30_2	7.7	0.5	7.4	0.7	5.9	0.3	3.7	0.5	3.0	1.0
T_50_30_3	7.2	0.4	7.2	0.4	6.0	0.0	3.0	0.0	3.0	1.0
T_50_50_2	7.0	0.0	7.0	0.0	6.0	0.0	3.2	0.4	3.0	1.0
T_50_50_3	7.0	0.0	7.0	0.0	5.8	0.4	3.0	0.0	**2.7**	1.0
T_80_20_2	**8.8**	0.6	8.3	0.5	**6.1**	0.3	4.0	0.0	3.0	1.0
T_80_20_3	**9.1**	0.3	8.2	0.4	6.0	0.0	3.0	0.0	3.0	1.0
T_80_30_2	8.8	0.4	8.1	0.3	6.0	0.0	4.0	0.0	3.0	1.0
T_80_30_3	8.7	0.5	7.9	0.3	6.0	0.0	**3.1**	0.3	3.0	1.0
T_80_50_2	8.4	0.5	8.0	0.0	6.0	0.0	4.0	0.0	3.0	1.0
T_80_50_3	8.1	0.3	8.0	0.0	6.0	0.0	3.0	0.0	3.0	1.0

To do some comparative study we also present results of experiments connected with cardinality of α-tests and α-reducts for approach based on generalized decision. For decision tables described in Table II and $\alpha \in \{0.0, 0.001, 0.01, 0.1, 0.2, 0.5\}$, we constructed by the greedy algorithm an α-test. Table VII presents cardinality of constructed α-tests.

Using Algorithms 2, 3 and 4 we try to remove from the test some attributes and obtain the reduct. For all three cases (LEFT, RIGHT, RANDOM) we get the same results. Table VIII presents the cardinality of α-reducts. Only for two data sets ("kr-vs-kp-5", "kr-vs-kp-4") we can find the difference in cardinality between tests and reducts, it is equal to 1.

Table IX, based on results from Tables III and VII, presents comparison of α-tests cardinality for proposed approach and approach based on generalized decision. Each input of this table is equal to the cardinality of α-test for generalized decision approach divided by the cardinality of α-test for proposed approach. For $\alpha = 0.2$ and $\alpha = 0.5$ we don't have any differences in cardinality of tests. For smaller values of α we can see that greedy algorithm for proposed approach sometimes construct two or three times shorter tests than for approach based on generalized decision (such values are in bold): "balance-scale-1" ($\alpha = 0.1$),"nursery-4" ($\alpha = 0.0$, $\alpha = 0.01$ and $\alpha = 0.1$), "nursery-1" ($\alpha = 0.1$), "poker-hand-training-true-5" and "poker-hand-training-true-5a" ($\alpha = 0.001$ and $\alpha = 0.01$).

We also considered randomly generated binary decision tables with many-valued decisions for generalized decision approach. Such tables contain: $n \in \{50, 80, 100\}$ rows, $m \in \{20, 30, 50\}$ conditional attributes, and $k \in \{2, 3\}$ values

TABLE VII
CARDINALITY OF α-TESTS - GENERALIZED DECISION APPROACH

Decision	α					
table	0.0	0.001	0.01	0.1	0.2	0.5
balance-scale-1	3	3	3	2	1	1
breast-cancer-1	8	6	4	2	2	1
breast-cancer-5	4	4	3	1	1	1
cars-1	5	5	3	2	2	1
flags-5	14	9	4	2	1	1
hayes-roth-data-1	3	3	3	2	1	1
kr-vs-kp-5	27	13	8	4	3	1
kr-vs-kp-4	28	13	8	4	3	1
lymphography-5	11	10	6	2	2	1
mushroom-5	9	4	3	2	1	1
nursery-4	4	3	2	2	1	1
nursery-1	7	5	3	2	1	1
poker-hand-training-true-5	5	4	3	1	1	1
poker-hand-training-true-5a	5	4	3	1	1	1
poker-hand-training-true-5b	5	5	4	2	2	1
spect-test-1	11	10	7	3	3	1
teeth-1	5	5	4	2	2	1
teeth-5	3	3	3	2	2	1
tic-tac-toe-4	5	5	4	2	2	1
tic-tac-toe-3	6	6	4	3	2	1
zoo-data-5	9	9	7	3	2	1

TABLE VIII
CARDINALITY OF α-REDUCTS - GENERALIZED DECISION APPROACH

Decision	α					
table	0.0	0.001	0.01	0.1	0.2	0.5
balance-scale-1	3	3	3	2	1	1
breast-cancer-1	8	6	4	2	2	1
breast-cancer-5	4	4	3	1	1	1
cars-1	5	5	3	2	2	1
flags-5	14	9	4	2	1	1
hayes-roth-data-1	3	3	3	2	1	1
kr-vs-kp-5	**26**	13	8	4	3	1
kr-vs-kp-4	**27**	13	8	4	3	1
lymphography-5	11	10	6	2	2	1
mushroom-5	9	4	3	2	1	1
nursery-4	4	3	2	2	1	1
nursery-1	7	5	3	2	1	1
poker-hand-training-true-5	5	4	3	1	1	1
poker-hand-training-true-5a	5	4	3	1	1	1
poker-hand-training-true-5b	5	5	4	2	2	1
spect-test-1	11	10	7	3	3	1
teeth-1	5	5	4	2	2	1
teeth-5	3	3	3	2	2	1
tic-tac-toe-4	5	5	4	2	2	1
tic-tac-toe-3	6	6	4	3	2	1
zoo-data-5	9	9	7	3	2	1

TABLE IX
COMPARISON OF α-TESTS CARDINALITY

Decision table	α 0.0	0.001	0.01	0.1	0.2	0.5
balance-scale-1	1.50	1.50	1.50	**2.00**	1.00	1.00
breast-cancer-1	1.00	1.00	1.00	1.00	1.00	1.00
breast-cancer-5	1.00	1.00	1.00	1.00	1.00	1.00
cars-1	1.00	1.25	1.00	1.00	1.00	1.00
flags-5	1.08	1.00	1.00	1.00	1.00	1.00
hayes-roth-data-1	1.50	1.50	1.50	1.00	1.00	1.00
kr-vs-kp-5	1.04	1.08	1.14	1.00	1.00	1.00
kr-vs-kp-4	1.04	1.08	1.14	1.00	1.00	1.00
lymphography-5	1.00	1.11	1.20	1.00	1.00	1.00
mushroom-5	1.13	1.00	1.00	1.00	1.00	1.00
nursery-4	**2.00**	1.50	**2.00**	**2.00**	1.00	1.00
nursery-1	1.00	1.67	1.50	**2.00**	1.00	1.00
poker-hand-training-true-5	1.00	**2.00**	**3.00**	1.00	1.00	1.00
poker-hand-training-true-5a	1.00	**2.00**	**3.00**	1.00	1.00	1.00
poker-hand-training-true-5b	1.00	1.00	1.00	1.00	1.00	1.00
spect-test-1	1.10	1.00	1.17	1.00	1.00	1.00
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	1.00	1.00	1.00	1.00	1.00	1.00
tic-tac-toe-3	1.00	1.20	1.00	1.50	1.00	1.00
zoo-data-5	1.00	1.00	1.00	1.00	1.00	1.00

of decision attributes. For $k = 2$, values of decision attribute are from $\{1-10\}$ and $\{11-20\}$, for $k = 3$, values of decision attribute are from $\{1 - 10\}$, $\{11 - 20\}$ and $\{21 - 30\}$. For each triple (n, m, k) we generated 10 decision tables such that each value of conditional attribute is equal to b, $b \in \{0, 1\}$ with probability $\frac{1}{2}$. Probabilities of values of decision attribute are the same and equal to $\frac{1}{10}$. Table X presents the average cardinality of constructed tests for randomly generated binary decision tables.

TABLE X
CARDINALITY OF α-TESTS FOR RANDOMLY GENERATED TABLES - GENERALIZED DECISION APPROACH

Decision table	α 0.0	0.001	0.01	0.1	0.2	0.5
T_100_20_2	10.00	9.00	7.00	4.00	3.00	1.00
T_100_20_3	10.20	9.00	6.90	4.00	3.00	1.00
T_100_30_2	10.00	8.90	6.90	4.00	3.00	1.00
T_100_30_3	9.90	9.00	7.00	4.00	3.00	1.00
T_100_50_2	9.30	8.30	6.50	4.00	3.00	1.00
T_100_50_3	9.60	9.00	6.70	4.00	3.00	1.00
T_50_20_2	8.10	7.90	6.00	4.00	3.00	1.00
T_50_20_3	8.00	7.70	6.00	4.00	3.00	1.00
T_50_30_2	7.90	7.50	6.00	4.00	3.00	1.00
T_50_30_3	8.00	7.50	6.00	4.00	3.00	1.00
T_50_50_2	7.70	7.20	6.00	4.00	3.00	1.00
T_50_50_3	8.00	7.10	6.00	4.00	3.00	1.00
T_80_20_2	9.70	8.70	6.50	4.00	3.00	1.00
T_80_20_3	9.70	8.70	6.50	4.00	3.00	1.00
T_80_30_2	9.20	8.10	6.10	4.00	3.00	1.00
T_80_30_3	9.20	8.40	6.10	4.00	3.00	1.00
T_80_50_2	9.00	8.00	6.00	4.00	3.00	1.00
T_80_50_3	8.90	8.00	6.00	4.00	3.00	1.00

Table XI, based on results from Tables V and X for randomly generated binary decision tables, presents comparison of α-tests cardinality for proposed approach and approach based on generalized decision. Each input of this table is equal to the cardinality of α-test for generalized decision approach divided by the cardinality of α-test for proposed approach. The difference in cardinality of tests is very small, even for $\alpha = 0$.

TABLE XI
COMPARISON OF α-TESTS CARDINALITY FOR RANDOMLY GENERATED TABLES

Decision table	α 0.0	0.001	0.01	0.1	0.2	0.5
T_100_20_2	1.01	1.02	1.08	1.00	1.00	1.00
T_100_20_3	1.03	1.02	1.15	1.29	1.00	1.00
T_100_30_2	1.05	1.05	1.08	1.00	1.00	1.00
T_100_30_3	1.09	1.11	1.17	1.29	1.00	1.00
T_100_50_2	1.02	1.01	1.08	1.00	1.00	1.00
T_100_50_3	1.07	1.13	1.12	1.33	1.03	1.00
T_50_20_2	1.05	1.04	1.00	1.11	1.00	1.00
T_50_20_3	1.08	1.04	1.00	1.33	1.00	1.00
T_50_30_2	1.03	1.01	1.02	1.08	1.00	1.00
T_50_30_3	1.11	1.04	1.00	1.33	1.00	1.00
T_50_50_2	1.10	1.03	1.00	1.25	1.00	1.00
T_50_50_3	1.14	1.01	1.03	1.33	1.07	1.00
T_80_20_2	1.05	1.05	1.03	1.00	1.00	1.00
T_80_20_3	1.05	1.06	1.08	1.33	1.00	1.00
T_80_30_2	1.05	1.00	1.02	1.00	1.00	1.00
T_80_30_3	1.06	1.06	1.02	1.25	1.00	1.00
T_80_50_2	1.07	1.00	1.00	1.00	1.00	1.00
T_80_50_3	1.10	1.00	1.00	1.33	1.00	1.00

VI. CONCLUSIONS

We studied the greedy algorithm for construction of α-tests. This algorithm requires the construction of all boundary subtables of the initial decision table. We proved that for an arbitrary natural t, the considered algorithm has polynomial time complexity on tables which have at most t decisions in each set of decisions attached to rows. We also presented algorithms which transform the test into the reduct. To work with decision tables with many-valued decisions we converted data sets from UCI ML Repository (by removal of some conditional attributes) into the form of decision tables described in Table II. Based on results contained in Sect. V, we can see that greedy algorithm constructs, for decision tables with many-valued decisions, short α-tests and improvements relative to cardinality of reducts are not significant. Based on results connected with comparative study, for tables from UCI ML Repository (see Table IX), we can observe that the cardinality of tests for proposed approach is 2 or 3 times smaller than for approach based on generalized decisions. In case of randomly generated binary decision tables (see Table XI) we don't have significant improvements.

Based on performed experiments we can't say which approach should be used because it depends on the aim of study. However, if we consider tests as a way of knowledge representation then we can see that tests constructed for proposed approach are shorter than tests constructed for the generalized decision approach. On the other hand, if we consider tests as a way for construction of classifier then we should compare the accuracy of classifiers for these two approaches. Such study will be considered in our future works.

REFERENCES

[1] M. Moshkov and B. Zielosko, *Combinatorial Machine Learning - A Rough Set Approach*. Heidelberg: Springer, 2011.
[2] Z. Pawlak, *Rough Sets - Theoretical Aspects of Reasoning about Data*. Dordrecht: Kluwer Academic Publishers, 1991.

[3] A. Skowron and C. Rauszer, "The discernibility matrices and functions in information systems," in *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory*. Dordrecht: Kluwer Academic Publishers, 1992, pp. 331–362.

[4] Z. Pawlak and A. Skowron, "Rudiments of rough sets," *Inf. Sci.*, vol. 177, no. 1, pp. 3–27, 2007.

[5] ——, "Rough sets and boolean reasoning," *Inf. Sci.*, vol. 177, no. 1, pp. 41–73, 2007.

[6] I. Chikalov, V. Lozin, I. Lozina, M. Moshkov, H. S. Nguyen, A. Skowron, and B. Zielosko, *Three Approaches to Data Analysis Test Theory, Rough Sets and Logical Analysis of Data*. Springer, 2012, in press.

[7] H. S. Nguyen and D. Ślęzak, "Approximate reducts and association rules - correspondence and complexity results," in *RSFDGrC 1999*, ser. LNCS, vol. 1711. Springer, 1999, pp. 137–145.

[8] M. Moshkov, M. Piliszczuk, and B. Zielosko, *Partial Covers, Reducts and Decision Rules in Rough Sets - Theory and Applications*. Heidelberg: Springer, 2008.

[9] D. Ślęzak, "Approximate entropy reducts," *Fundam. Inform.*, vol. 53, pp. 365–390, 2002.

[10] M. Azad, I. Chikalov, M. Moshkov, and B. Zielosko, "Greedy algorithms for construction of approximate tests," *Fundam. Inform.*, 2012, in press.

[11] M. Moshkov and B. Zielosko, "Construction of tests for tables with many-valued decisions," in *20th International Workshop Concurrency, Specification and Programming CS&P 2011, September 28–30, Pułtusk, Poland*. Białystok University of Technology, 2011, pp. 376–384.

[12] A. Asuncion and D. J. Newman, "UCI Machine Learning Repository," 2007. [Online]. Available: http://www.ics.uci.edu/~mlearn/