On Partial Covers, Reducts and Decision Rules with Weights

Mikhail Ju. Moshkov[1], Marcin Piliszczuk[2], and Beata Zielosko[3]

[1] Institute of Computer Science, University of Silesia
39, Będzińska St., Sosnowiec, 41-200, Poland
moshkov@us.edu.pl
[2] ING Bank Śląski S.A., 34, Sokolska St., Katowice, 40-086, Poland
marcin.piliszczuk@ingbank.pl
[3] Institute of Computer Science, University of Silesia
39, Będzińska St., Sosnowiec, 41-200, Poland
zielosko@us.edu.pl

Abstract. In the paper the accuracy of greedy algorithms with weights for construction of partial covers, reducts and decision rules is considered. Bounds on minimal weight of partial covers, reducts and decision rules based on an information on greedy algorithm work are studied. Results of experiments with software implementation of greedy algorithms are described.

Keywords: partial covers, partial reducts, partial decision rules, weights, greedy algorithms.

1 Introduction

The paper is devoted to consideration of partial decision-relative reducts (we will omit often words "decision-relative") and partial decision rules for decision tables on the basis of partial cover investigation.

Rough set theory [11,17] often deals with decision tables containing noisy data. In this case exact reducts and rules can be "overlearned" i.e. depend essentially on noise. If we see constructed reducts and rules as a way of knowledge representation [16] then instead of large exact reducts and rules it is more appropriate to work with relatively small partial ones. In [12] Zdzisław Pawlak wrote that "the idea of an approximate reduct can be useful in cases when a smaller number of condition attributes is preferred over accuracy of classification".

Last years in rough set theory partial reducts, partial decision rules and partial covers are studied intensively [6,7,8,9,10,13,19,20,21,22,23,24,27]. Approximate reducts are investigated also in extensions of rough set model such as VPRS (variable precision rough sets) [26] and α-RST (alpha rough set theory) [14].

We study the case where each subset, used for covering, has its own weight, and we must minimize the total weight of subsets in partial cover. The same situation is with partial reducts and decision rules: each conditional attribute has its own weight, and we must minimize the total weight of attributes in partial

J.F. Peters et al. (Eds.): Transactions on Rough Sets VI, LNCS 4374, pp. 211–246, 2007.

reduct or decision rule. If weight of each attribute characterizes time complexity of attribute value computation then we try to minimize total time complexity of computation of attributes from partial reduct or partial decision rule. If weight characterizes a risk of attribute value computation (as in medical or technical diagnosis) then we try to minimize total risk, etc.

In rough set theory various problems can be represented as set cover problems with weights:

- problem of construction of a reduct [16] or partial reduct with minimal total weight of attributes for an information system;
- problem of construction of a decision-relative reduct [16] or partial decision-relative reduct with minimal total weight of attributes for a decision table;
- problem of construction of a decision rule or partial decision rule with minimal total weight of attributes for a row of a decision table (note that this problem is closely connected with the problem of construction of a local reduct [16] or partial local reduct with minimal total weight of attributes);
- problem of construction of a subsystem of a given system of decision rules which "covers" the same set of rows and has minimal total weight of rules (in the capacity of a rule weight we can consider its length).

So the study of covers and partial covers is of some interest for rough set theory. In this paper we list some known results on set cover problems which can be useful in applications and obtain certain new results.

From results obtained in [20,22] it follows that the problem of construction of partial cover with minimal weight is NP-hard. Therefore we must consider polynomial approximate algorithms for minimization of weight of partial covers.

In [18] a greedy algorithm with weights for partial cover construction was investigated. This algorithm is a generalization of well known greedy algorithm with weights for exact cover construction [2]. The algorithm from [18] is a greedy algorithm with one threshold which gives the exactness of constructed partial cover.

Using results from [9] (based on results from [3,15] and technique created in [20,22]) on precision of polynomial approximate algorithms for construction of partial cover with minimal cardinality and results from [18] on precision of greedy algorithm with one threshold we show that under some natural assumptions on the class NP the greedy algorithm with one threshold is close to best polynomial approximate algorithms for construction of partial cover with minimal weight. However we can try to improve results of the work of greedy algorithm with one threshold for some part of set cover problems with weight.

We generalize greedy algorithm with one threshold [18], and consider greedy algorithm with two thresholds. First threshold gives the exactness of constructed partial cover, and the second one is an interior parameter of the considered algorithm. We prove that for the most part of set cover problems there exist a weight function and values of thresholds such that the weight of partial cover constructed by greedy algorithm with two thresholds is less than the weight of partial cover constructed by greedy algorithm with one threshold.

We describe two polynomial algorithms which always construct partial covers that are not worse than the one constructed by greedy algorithm with one threshold, and for the most part of set cover problems there exists a weight function and a value of first threshold such that the weight of partial covers constructed by the considered two algorithms is less than the weight of partial cover constructed by greedy algorithm with one threshold.

Information on greedy algorithm work can be used for obtaining lower bounds on minimal cardinality of partial covers [9]. We fix some kind of information on greedy algorithm work, and find unimprovable lower bound on minimal weight of partial cover depending on this information. Obtained results show that this bound is not trivial and can be useful for investigation of set cover problems.

There exist bounds on precision of greedy algorithm without weights for partial cover construction which do not depend on the cardinality of covered set [1,6,7,8]. We obtain similar bound for the case of weight.

The most part of the results obtained for partial covers is generalized on the case of partial decision-relative reducts and partial decision rules for decision tables which, in general case, are inconsistent (a decision table is inconsistent if it has equal rows with different decisions). In particular, we show that

– Under some natural assumptions on the class NP greedy algorithms with weights are close to best polynomial approximate algorithms for minimization of total weight of attributes in partial reducts and partial decision rules.
– Based on information receiving during greedy algorithm work it is possible to obtain nontrivial lower bounds on minimal total weight of attributes in partial reducts and partial decision rules.
– There exist polynomial modifications of greedy algorithms which for a part of decision tables give better results than usual greedy algorithms.

Obtained results will further to more wide use of greedy algorithms with weighs and their modifications in rough set theory and applications.

This paper is, in some sense, an extension of [9] on the case of weights which are not equal to 1. However, problems considered in this paper (and proofs of results) are more complicated than the ones considered in [9]. Bounds obtained in this paper are sometimes more weak than the corresponding bounds from [9]. We must note also that even if all weights are equal to 1 then results of the work of greedy algorithms considered in this paper can be different from the results of the work of greedy algorithms considered in [9]. For example, for case of reducts the number of chosen attributes is the same, but last attributes can differ.

The paper consists of five sections. In Sect. 2 partial covers are studied. In Sect. 3 partial tests (partial superreducts) and partial reducts are investigated. In Sect. 4 partial decision rules are considered. Sect. 5 contains short conclusions.

2 Partial Covers

2.1 Main Notions

Let $A = \{a_1, \ldots, a_n\}$ be a nonempty finite set. Elements of A are enumerated by numbers $1, \ldots, n$ (in fact we fix a linear order on A). Let $S = \{B_i\}_{i \in \{1,\ldots,m\}} =$

$\{B_1, \ldots, B_m\}$ be a family of subsets of A such that $B_1 \cup \ldots \cup B_m = A$. We will assume that S can contain equal subsets of A. The pair (A, S) will be called a *set cover problem*. Let w be a *weight function* which corresponds to each $B_i \in S$ a natural number $w(B_i)$. The triple (A, S, w) will be called a *set cover problem with weights*. Note that in fact weight function w is given on the set of indexes $\{1, \ldots, m\}$. But, for simplicity, we are writing $w(B_i)$ instead of $w(i)$.

Let I be a subset of $\{1, \ldots, m\}$. The family $P = \{B_i\}_{i \in I}$ will be called a *subfamily* of S. The number $|P| = |I|$ will be called the *cardinality* of P. Let $P = \{B_i\}_{i \in I}$ and $Q = \{B_i\}_{i \in J}$ be subfamilies of S. The notation $P \subseteq Q$ will mean that $I \subseteq J$. Let us denote $P \cup Q = \{B_i\}_{i \in I \cup J}$, $P \cap Q = \{B_i\}_{i \in I \cap J}$, and $P \setminus Q = \{B_i\}_{i \in I \setminus J}$.

A subfamily $Q = \{B_{i_1}, \ldots, B_{i_t}\}$ of the family S will be called a *partial cover for* (A, S). Let α be a real number such that $0 \leq \alpha < 1$. The subfamily Q will be called an *α-cover for* (A, S) if $|B_{i_1} \cup \ldots \cup B_{i_t}| \geq (1 - \alpha)|A|$. For example, 0.01-cover means that we must cover at least 99% of elements from A. Note that a 0-cover is usual (exact) cover. The number $w(Q) = \sum_{j=1}^{t} w(B_{i_j})$ will be called the *weight* of the partial cover Q. Let us denote by $C_{\min}(\alpha) = C_{\min}(\alpha, A, S, w)$ the minimal weight of α-cover for (A, S).

Let α and γ be real numbers such that $0 \leq \gamma \leq \alpha < 1$. Let us describe a *greedy algorithm with two thresholds* α *and* γ.

Let us denote $N = \lceil |A|(1 - \gamma) \rceil$ and $M = \lceil |A|(1 - \alpha) \rceil$. Let we make $i \geq 0$ steps and choose subsets $B_{j_1}, \ldots, B_{j_i}$. Let us describe the step number $i + 1$.

Let us denote $D = B_{j_1} \cup \ldots \cup B_{j_i}$ (if $i = 0$ then $D = \emptyset$). If $|D| \geq M$ then we finish the work of the algorithm. The family $\{B_{j_1}, \ldots, B_{j_i}\}$ is the constructed α-cover. Let $|D| < M$. Then we choose a subset $B_{j_{i+1}}$ from S with minimal number j_{i+1} for which $B_{j_{i+1}} \setminus D \neq \emptyset$ and the value

$$\frac{w(B_{j_{i+1}})}{\min\{|B_{j_{i+1}} \setminus D|, N - |D|\}}$$

is minimal. Pass to the step number $i + 2$.

Let us denote by $C^{\gamma}_{\text{greedy}}(\alpha) = C^{\gamma}_{\text{greedy}}(\alpha, A, S, w)$ the weight of α-cover constructed by the considered algorithm for the set cover problem with weights (A, S, w).

Note that greedy algorithm with two thresholds α and α coincides with the greedy algorithm with one threshold α considered in [18].

2.2 Some Known Results

In this subsection we assume that the weight function has values from the set of positive real numbers.

For natural m denote $H(m) = 1 + \ldots + 1/m$. It is known that

$$\ln m \leq H(m) \leq \ln m + 1 \ .$$

Consider some results for the case of exact covers where $\alpha = 0$. In this case $\gamma = 0$. First results belong to Chvátal.

Theorem 1. (Chvátal [2]) *For any set cover problem with weights* (A, S, w) *the inequality* $C^0_{\text{greedy}}(0) \leq C_{\min}(0)H(|A|)$ *holds.*

Theorem 2. (Chvátal [2]) *For any set cover problem with weights* (A, S, w) *the inequality* $C^0_{\text{greedy}}(0) \leq C_{\min}(0)H\left(\max_{B_i \in S} |B_i|\right)$ *holds.*

Chvátal proved in [2] that the bounds from Theorems 1 and 2 are almost unimprovable.

Consider now some results for the case where $\alpha \geq 0$ and $\gamma = \alpha$. First upper bound on $C^{\alpha}_{\text{greedy}}(\alpha)$ was obtained by Kearns.

Theorem 3. (Kearns [5]) *For any set cover problem with weights* (A, S, w) *and any* α, $0 \leq \alpha < 1$, *the inequality* $C^{\alpha}_{\text{greedy}}(\alpha) \leq C_{\min}(\alpha)(2H(|A|) + 3)$ *holds.*

This bound was improved by Slavík.

Theorem 4. (Slavík [18]) *For any set cover problem with weights* (A, S, w) *and any* α, $0 \leq \alpha < 1$, *the inequality* $C^{\alpha}_{\text{greedy}}(\alpha) \leq C_{\min}(\alpha)H\left(\lceil (1-\alpha)|A| \rceil\right)$ *holds.*

Theorem 5. (Slavík [18])) *For any set cover problem with weights* (A, S, w) *and any* α, $0 \leq \alpha < 1$, *the inequality* $C^{\alpha}_{\text{greedy}}(\alpha) \leq C_{\min}(\alpha)H\left(\max_{B_i \in S} |B_i|\right)$ *holds.*

Slavík proved in [18] that the bounds from Theorems 4 and 5 are unimprovable.

2.3 On Polynomial Approximate Algorithms for Minimization of Partial Cover Weight

In this subsection we consider three theorems which follow immediately from Theorems 13–15 [9].

Let $0 \leq \alpha < 1$. Consider the following problem: for given set cover problem with weights (A, S, w) it is required to find an α-cover for (A, S) with minimal weight.

Theorem 6. *Let* $0 \leq \alpha < 1$. *Then the problem of construction of* α*-cover with minimal weight is* NP*-hard.*

From this theorem it follows that we must consider polynomial approximate algorithms for minimization of α-cover weight.

Theorem 7. *Let* $\alpha \in \mathbb{R}$ *and* $0 \leq \alpha < 1$. *If* $NP \not\subseteq DTIME(n^{O(\log \log n)})$ *then for any* ε, $0 < \varepsilon < 1$, *there is no polynomial algorithm that for a given set cover problem with weights* (A, S, w) *constructs an* α*-cover for* (A, S) *which weight is at most* $(1 - \varepsilon)C_{\min}(\alpha, A, S, w) \ln |A|$.

Theorem 8. *Let* α *be a real number such that* $0 \leq \alpha < 1$. *If* $P \neq NP$ *then there exists* $\delta > 0$ *such that there is no polynomial algorithm that for a given set cover problem with weights* (A, S, w) *constructs an* α*-cover for* (A, S) *which weight is at most* $\delta C_{\min}(\alpha, A, S, w) \ln |A|$.

From Theorem 4 it follows that $C^{\alpha}_{\text{greedy}}(\alpha) \leq C_{\min}(\alpha)(1 + \ln |A|)$. From this inequality and from Theorem 7 it follows that under the assumption $NP \not\subseteq DTIME(n^{O(\log \log n)})$ greedy algorithm with two thresholds α and α (in fact greedy algorithm with one threshold α from [18]) is close to best polynomial approximate algorithms for minimization of partial cover weight. From the considered inequality and from Theorem 8 it follows that under the assumption $P \neq NP$ greedy algorithm with two thresholds α and α is not far from best polynomial approximate algorithms for minimization of partial cover weight.

However we can try to improve the results of the work of greedy algorithm with two thresholds α and α for some part of set cover problems with weights.

2.4 Comparison of Greedy Algorithms with One and Two Thresholds

The following example shows that if for greedy algorithm with two thresholds α and γ we will use γ such that $\gamma < \alpha$ we can obtain sometimes better results than in the case $\gamma = \alpha$.

Example 1. Consider a set cover problem (A, S, w) such that $A = \{1, 2, 3, 4, 5, 6\}$, $S = \{B_1, B_2\}$, $B_1 = \{1\}$, $B_2 = \{2, 3, 4, 5, 6\}$, $w(B_1) = 1$ and $w(B_2) = 4$. Let $\alpha = 0.5$. It means that we must cover at least $M = \lceil (1 - \alpha)|A| \rceil = 3$ elements from A. If $\gamma = \alpha = 0.5$ then the result of the work of greedy algorithm with thresholds α and γ is the 0.5-cover $\{B_1, B_2\}$ which weight is equal to 5. If $\gamma = 0 < \alpha$ then the result of the work of greedy algorithm with thresholds α and γ is the 0.5-cover $\{B_2\}$ which weight is equal to 4.

In this subsection we show that under some assumptions on $|A|$ and $|S|$ for the most part of set cover problems (A, S) there exist a weight function w and real numbers α, γ such that $0 \leq \gamma < \alpha < 1$ and $C^{\gamma}_{\text{greedy}}(\alpha, A, S, w) < C^{\alpha}_{\text{greedy}}(\alpha, A, S, w)$. First, we consider criterion of existence of such w, α and γ (see Theorem 9). First part of the proof of this criterion is based on a construction similar to considered in Example 1.

Let A be a finite nonempty set and $S = \{B_1, \ldots, B_m\}$ be a family of subsets of A. We will say that the family S is 1-*uniform* if there exists a natural number k such that $|B_i| = k$ or $|B_i| = k+1$ for any nonempty subset B_i from S. We will say that S is *strongly* 1-*uniform* if S is 1-uniform and for any subsets $B_{l_1}, \ldots, B_{l_t}$ from S the family $\{B_1 \setminus U, \ldots, B_m \setminus U\}$ is 1-uniform where $U = B_{l_1} \cup \ldots \cup B_{l_t}$.

Theorem 9. *Let (A, S) be a set cover problem. Then the following two statements are equivalent:*

1. *The family S is not strongly* 1-*uniform.*
2. *There exist a weight function w and real numbers α and γ such that* $0 \leq \gamma < \alpha < 1$ *and* $C^{\gamma}_{\text{greedy}}(\alpha, A, S, w) < C^{\alpha}_{\text{greedy}}(\alpha, A, S, w)$.

Proof. Let $S = \{B_1, \ldots, B_m\}$. Let the family S be not strongly 1-uniform. Let us choose minimal number of subsets $B_{l_1}, \ldots, B_{l_t}$ from the family S (it is possible that $t = 0$) such that the family $\{B_1 \setminus U, \ldots, B_m \setminus U\}$ is not 1-uniform where $U = B_{l_1} \cup \ldots \cup B_{l_t}$ (if $t = 0$ then $U = \emptyset$). Since $\{B_1 \setminus U, \ldots, B_m \setminus U\}$ is not 1-uniform, there exist two subsets B_i and B_j from S such that $|B_i \setminus U| > 0$ and $|B_j \setminus U| \geq |B_i \setminus U| + 2$. Let us choose real α and γ such that $M = \lceil |A|(1-\alpha) \rceil = |U| + |B_i \setminus U| + 1$ and $N = \lceil |A|(1-\gamma) \rceil = |U| + |B_i \setminus U| + 2$. It is clear that $0 \leq \gamma < \alpha < 1$. Let us define a weight function w as follows: $w(B_{l_1}) = \ldots = w(B_{l_t}) = 1$, $w(B_i) = |A| \cdot 2|B_i \setminus U|$, $w(B_j) = |A|(2|B_i \setminus U| + 3)$ and $w(B_r) = |A|(3|B_i \setminus U| + 6)$ for any B_r from S such that $r \notin \{i, j, l_1, \ldots, l_t\}$.

Let us consider the work of greedy algorithm with two thresholds α and α. One can show that during first t steps the greedy algorithm will choose subsets $B_{l_1}, \ldots, B_{l_t}$ (may be in an another order). It is clear that $|U| < M$. Therefore the greedy algorithm must make the step number $t + 1$. During this step the greedy algorithm will choose a subset B_k from S with minimal number k for which $B_k \setminus U \neq \emptyset$ and the value $p(k) = \frac{w(B_k)}{\min\{|B_k \setminus U|, M - |U|\}} = \frac{w(B_k)}{\min\{|B_k \setminus U|, |B_i \setminus U| + 1\}}$ is minimal.

It is clear that $p(i) = 2|A|$, $p(j) = (2 + \frac{1}{|B_i \setminus U| + 1})|A|$ and $p(k) > 3|A|$ for any subset B_k from S such that $B_k \setminus U \neq \emptyset$ and $k \notin \{i, j, l_1, \ldots, l_t\}$. Therefore during the step number $t + 1$ the greedy algorithm will choose the subset B_i. Since $|U| + |B_i \setminus U| = M - 1$, the greedy algorithm will make the step number $t + 2$ and will choose a subset from S which is different from $B_{l_1}, \ldots, B_{l_t}, B_i$. As the result we obtain $C^{\alpha}_{\text{greedy}}(\alpha, A, S, w) \geq t + |A| \cdot 2|B_i \setminus U| + |A|(2|B_i \setminus U| + 3)$.

Let us consider the work of greedy algorithm with two thresholds α and γ. One can show that during first t steps the greedy algorithm will choose subsets $B_{l_1}, \ldots, B_{l_t}$ (may be in an another order). It is clear that $|U| < M$. Therefore the greedy algorithm must make the step number $t + 1$. During this step the greedy algorithm will choose a subset B_k from S with minimal number k for which $B_k \setminus U \neq \emptyset$ and the value $q(k) = \frac{w(B_k)}{\min\{|B_k \setminus U|, N - |U|\}} = \frac{w(B_k)}{\min\{|B_k \setminus U|, |B_i \setminus U| + 2\}}$ is minimal.

It is clear that $q(i) = 2|A|$, $q(j) = (2 - \frac{1}{|B_i \setminus U| + 2})|A|$ and $q(k) \geq 3|A|$ for any subset B_k from S such that $B_k \setminus U \neq \emptyset$ and $k \notin \{i, j, l_1, \ldots, l_t\}$. Therefore during the step number $t + 1$ the greedy algorithm will choose the subset B_j. Since $|U| + |B_j \setminus U| > M$, the α-cover constructed by greedy algorithm will be equal to $\{B_{l_1}, \ldots, B_{l_t}, B_j\}$. As the result we obtain $C^{\gamma}_{\text{greedy}}(\alpha, A, S, w) = t + |A|(2|B_i \setminus U| + 3)$. Since $C^{\alpha}_{\text{greedy}}(\alpha, A, S, w) \geq t + |A| \cdot 2|B_i \setminus U| + |A|(2|B_i \setminus U| + 3)$ and $|B_i \setminus U| > 0$, we conclude that $C^{\alpha}_{\text{greedy}}(\alpha, A, S, w) > C^{\gamma}_{\text{greedy}}(\alpha, A, S, w)$.

Let the family S be strongly 1-uniform. Consider arbitrary weight function w for S and real numbers α and γ such that $0 \leq \gamma < \alpha < 1$. Let us show that $C^{\gamma}_{\text{greedy}}(\alpha, A, S, w) \geq C^{\alpha}_{\text{greedy}}(\alpha, A, S, w)$. Let us denote $M = \lceil |A|(1-\alpha) \rceil$ and $N = \lceil |A|(1-\gamma) \rceil$. If $M = N$ then $C^{\gamma}_{\text{greedy}}(\alpha, A, S, w) = C^{\alpha}_{\text{greedy}}(\alpha, A, S, w)$. Let $N > M$.

Let us apply the greedy algorithm with thresholds α and α to the set cover problem with weights (A, S, w). Let during the construction of α-cover this algorithm choose sequentially subsets $B_{g_1}, \ldots, B_{g_t}$. Let us apply now the greedy algorithm with thresholds α and γ to the set cover problem with weights (A, S, w). If during the construction of α-cover this algorithm chooses sequentially subsets

$B_{g_1},\ldots,B_{g_t}$ then $C^{\gamma}_{\text{greedy}}(\alpha,A,S,w)=C^{\alpha}_{\text{greedy}}(\alpha,A,S,w)$. Let there exist a nonnegative integer r, $0\le r\le t-1$, such that during first r steps the considered algorithm chooses subsets $B_{g_1},\ldots,B_{g_r}$, but at the step number $r+1$ the algorithm chooses a subset B_k such that $k\ne g_{r+1}$. Let us denote $B_{g_0}=\emptyset$, $D=B_{g_0}\cup\ldots\cup B_{g_r}$ and $J=\{i: i\in\{1,\ldots,m\}, B_i\setminus D\ne\emptyset\}$. It is clear that $g_{r+1},k\in J$. For any $i\in J$ denote $p(i)=\frac{w(B_i)}{\min\{|B_i\setminus D|,M-|D|\}}$, $q(i)=\frac{w(B_i)}{\min\{|B_i\setminus D|,N-|D|\}}$.

Since $k\ne g_{r+1}$, we conclude that there exists $i\in J$ such that $p(i)\ne q(i)$. Therefore $|B_i\setminus D|>M-|D|$. Since S is strongly 1-uniform family, we have $|B_j\setminus D|\ge M-|D|$ for any $j\in J$. From here it follows, in particular, that $r+1=t$, and $\{B_{g_1},\ldots,B_{g_{t-1}},B_k\}$ is an α-cover for (A,S).

It is clear that $p(g_t)\le p(k)$. Since $|B_k\setminus D|\ge M-|D|$ and $|B_{g_t}\setminus D|\ge M-|D|$, we have $p(k)=\frac{w(B_k)}{M-|D|}$, $p(g_t)=\frac{w(B_{g_t})}{M-|D|}$. Therefore $w(B_{g_t})\le w(B_k)$.

Taking into account that $C^{\gamma}_{\text{greedy}}(\alpha,A,S,w)=w(B_{g_1})+\ldots+w(B_{g_{t-1}})+w(B_k)$ and $C^{\alpha}_{\text{greedy}}(\alpha,A,S,w)=w(B_{g_1})+\ldots+w(B_{g_{t-1}})+w(B_{g_t})$ we obtain $C^{\gamma}_{\text{greedy}}(\alpha,A,S,w)\ge C^{\alpha}_{\text{greedy}}(\alpha,A,S,w)$. □

Let us show that under some assumptions on $|A|$ and $|S|$ the most part of set cover problems (A,S) is not 1-uniform, and therefore is not strongly 1-uniform.

There is one-to-one correspondence between set cover problems and tables filled by numbers from $\{0,1\}$ and having no rows filled by 0 only. Let $A=\{a_1,\ldots,a_n\}$ and $S=\{B_1,\ldots,B_m\}$. Then the problem (A,S) corresponds to the table with n rows and m columns which for $i=1,\ldots,n$ and $j=1,\ldots,m$ has 1 at the intersection of i-th row and j-th column if and only if $a_i\in B_j$.

A table filled by numbers from $\{0,1\}$ will be called *SC-table* if this table has no rows filled by 0 only. For completeness of the presentation we consider here a statement from [9] with proof.

Lemma 1. *The number of SC-tables with n rows and m columns is at least*

$$2^{mn}-2^{mn-m+\log_2 n}\ .$$

Proof. Let $i\in\{1,\ldots,n\}$. The number of tables in which the i-th row is filled by 0 only is equal to 2^{mn-m}. Therefore the number of tables which are not SC-tables is at most $n2^{mn-m}=2^{mn-m+\log_2 n}$. Thus, the number of SC-tables is at least $2^{mn}-2^{mn-m+\log_2 n}$. □

Lemma 2. *Let $n\in\mathbb{N}$, $n\ge 4$ and $k\in\{0,\ldots,n\}$. Then $C_n^k\le C_n^{\lfloor n/2\rfloor}<\frac{2^n}{\sqrt{n}}$.*

Proof. It is well known (see, for example, [25], p. 178) that $C_n^k\le C_n^{\lfloor n/2\rfloor}$. Let n be even and $n\ge 4$. It is known (see [4], p. 278) that $C_n^{\lfloor n/2\rfloor}\le\frac{2^n}{\sqrt{\frac{3n}{2}+1}}<\frac{2^n}{\sqrt{n}}$.

Let n be odd and $n\ge 5$. Using well known equality $C_n^{\lfloor n/2\rfloor}=C_{n-1}^{\lfloor n/2\rfloor}+C_{n-1}^{\lfloor n/2\rfloor-1}$ and the fact, that $C_{n-1}^{\lfloor (n-1)/2\rfloor}\ge C_{n-1}^k$ for any $k\in\{0,\ldots,n-1\}$, we obtain $C_n^{\lfloor n/2\rfloor}\le 2C_{n-1}^{\lfloor (n-1)/2\rfloor}$. Thus, $C_n^{\lfloor n/2\rfloor}\le\frac{2^n}{\sqrt{\frac{3(n-1)}{2}+1}}<\frac{2^n}{\sqrt{\frac{3(n-1)}{3}+1}}=\frac{2^n}{\sqrt{n}}$. Therefore for any $n\ge 4$ the inequality $C_n^{\lfloor n/2\rfloor}<\frac{2^n}{\sqrt{n}}$ holds. □

Theorem 10. *Consider set cover problems* (A, S) *such that* $A = \{a_1, \ldots, a_n\}$ *and* $S = \{B_1, \ldots, B_m\}$*. Let* $n \geq 4$ *and* $m \geq \log_2 n + 1$*. Then the fraction of set cover problems which are not 1-uniform is at least* $1 - \frac{9^{\frac{m}{2}+1}}{n^{\frac{m}{2}-1}}$.

Proof. The considered fraction is at least $\frac{q-p}{q}$ where q is the number of SC-tables with n rows and m columns, and p is the number of tables with n rows and m columns filled by 0 and 1 for each of which there exists $k \in \{1, \ldots, n-1\}$ such that the number of units in each column belongs to the set $\{0, k, k+1\}$.

From Lemma 1 it follows that $q \geq 2^{mn} - 2^{mn-m+\log_2 n}$. It is clear that $p \leq \sum_{k=1}^{n-1}(C_n^k + C_n^{k+1} + 1)^m$. From Lemma 2 it follows that $C_n^{\lfloor n/2 \rfloor} \geq C_n^k$ for any $k \in \{1, \ldots, n\}$. Therefore $p \leq (n-1)\left(3C_n^{\lfloor n/2 \rfloor}\right)^m$. Using Lemma 2 we conclude that $3C_n^{\lfloor n/2 \rfloor} < \frac{2^n}{\sqrt{\frac{n}{9}}}$ for any $n \geq 4$. Therefore $p < \frac{(n-1)2^{mn}}{\left(\frac{n}{9}\right)^{m/2}}$. Thus, $\frac{q-p}{q} = 1 - \frac{p}{q} > 1 - \frac{(n-1)2^{mn}}{\left(\frac{n}{9}\right)^{m/2}\left(2^{mn} - 2^{mn-m+\log_2 n}\right)}$. Taking into account that $m \geq \log_2 n + 1$ we obtain $\frac{q-p}{q} > 1 - \frac{2(n-1)}{\left(\frac{n}{9}\right)^{m/2}} > 1 - \frac{9^{\frac{m}{2}+1}}{n^{\frac{m}{2}-1}}$. □

So if n is large enough and $m \geq \log_2 n + 1$ then the most part of set cover problems (A, S) with $|A| = n$ and $|S| = m$ is not 1-uniform.

For example, the fraction of set cover problems (A, S) with $|A| = 81$ and $|S| = 20$ which are not 1-uniform is at least $1 - \frac{1}{9^7} = 1 - \frac{1}{4782969}$.

2.5 Two Modifications of Greedy Algorithm

Results obtained in the previous subsection show that the greedy algorithm with two thresholds is of some interest. In this subsection we consider two polynomial modifications of greedy algorithm which allow to use advantages of greedy algorithm with two thresholds.

Let (A, S, w) be a set cover problem with weights and α be a real number such that $0 \leq \alpha < 1$.

1. Of course, it is impossible to consider effectively all γ such that $0 \leq \gamma \leq \alpha$. Instead of this we can consider all natural N such that $M \leq N \leq |A|$ where $M = \lceil |A|(1-\alpha) \rceil$ (see the description of greedy algorithm with two thresholds). For each $N \in \{M, \ldots, |A|\}$ we apply greedy algorithm with parameters M and N to set cover problem with weights (A, S, w) and after that choose an α-cover with minimal weight among constructed α-covers.
2. There exists also an another way to construct an α-cover which is not worse than the one obtained under consideration of all N such that $M \leq N \leq |A|$. Let us apply greedy algorithm with thresholds α and α to set cover problem with weights (A, S, w). Let the algorithm choose sequentially subsets $B_{g_1}, \ldots, B_{g_t}$. For each $i \in \{0, \ldots, t-1\}$ we find (if it is possible) a subset B_{l_i} from S with minimal weight $w(B_{l_i})$ such that $|B_{g_1} \cup \ldots \cup B_{g_i} \cup B_{l_i}| \geq M$, and form an α-cover $\{B_{g_1}, \ldots, B_{g_i}, B_{l_i}\}$ (if $i = 0$ then it will be the family $\{B_{l_0}\}$). After that among constructed α-covers $\{B_{g_1}, \ldots, B_{g_t}\}$, ...,

$\{B_{g_1}, \ldots, B_{g_i}, B_{l_i}\}$, ... we choose an α-cover with minimal weight. From Proposition 1 it follows that the constructed α-cover is not worse than the one constructed under consideration of all γ, $0 \leq \gamma \leq \alpha$, or (which is the same) all N, $M \leq N \leq |A|$.

Proposition 1. *Let (A, S, w) be a set cover problem with weights and α, γ be real numbers such that $0 \leq \gamma < \alpha < 1$. Let the greedy algorithm with two thresholds α and α, which is applied to (A, S, w), choose sequentially subsets $B_{g_1}, \ldots, B_{g_t}$. Let the greedy algorithm with two thresholds α and γ, which is applied to (A, S, w), choose sequentially subsets $B_{l_1}, \ldots, B_{l_k}$. Then either $k = t$ and $(l_1, \ldots, l_k) = (g_1, \ldots, g_t)$ or $k \leq t$, $(l_1, \ldots, l_{k-1}) = (g_1, \ldots, g_{k-1})$ and $l_k \neq g_k$.*

Proof. Let $S = \{B_1, \ldots, B_m\}$. Let us denote $M = \lceil |A|(1-\alpha) \rceil$ and $N = \lceil |A|(1-\gamma) \rceil$.

Let $(l_1, \ldots, l_k) \neq (g_1, \ldots, g_t)$. Since $\{B_{g_1}, \ldots, B_{g_{t-1}}\}$ is not an α-cover for (A, S), it is impossible that $k < t$ and $(l_1, \ldots, l_k) = (g_1, \ldots, g_k)$. Since $\{B_{g_1}, \ldots, B_{g_t}\}$ is an α-cover for (A, S), it is impossible that $k > t$ and $(l_1, \ldots, l_t) = (g_1, \ldots, g_t)$. Therefore there exists $i \in \{0, \ldots, t-1\}$ such that during first i steps algorithm with thresholds α and α and algorithm with thresholds α and γ choose the same subsets from S, but during the step number $i+1$ the algorithm with threshold α and γ chooses a subset $B_{l_{i+1}}$ such that $l_{i+1} \neq g_{i+1}$.

Let us denote $B_{g_0} = \emptyset$, $D = B_{g_0} \cup \ldots \cup B_{g_i}$ and $J = \{j : j \in \{1, \ldots, m\}, B_j \setminus D \neq \emptyset\}$. It is clear that $g_{i+1}, l_{i+1} \in J$. For any $j \in J$ let $p(j) = \frac{w(B_j)}{\min\{|B_j \setminus D|, M - |D|\}}$ and $q(j) = \frac{w(B_j)}{\min\{|B_j \setminus D|, N - |D|\}}$. Since $N \geq M$, we have $p(j) \geq q(j)$ for any $j \in J$. Consider two cases.

Let $g_{i+1} < l_{i+1}$. In this case we have $p(g_{i+1}) \leq p(l_{i+1})$ and $q(g_{i+1}) > q(l_{i+1})$. Using inequality $p(g_{i+1}) \geq q(g_{i+1})$ we obtain $p(g_{i+1}) > q(l_{i+1})$ and $p(l_{i+1}) > q(l_{i+1})$. From last inequality it follows that $|B_{l_{i+1}} \setminus D| > M - |D|$.

Let $g_{i+1} > l_{i+1}$. In this case we have $p(g_{i+1}) < p(l_{i+1})$ and $q(g_{i+1}) \geq q(l_{i+1})$. Using inequality $p(g_{i+1}) \geq q(g_{i+1})$ we obtain $p(g_{i+1}) \geq q(l_{i+1})$ and $p(l_{i+1}) > q(l_{i+1})$. From last inequality it follows that $|B_{l_{i+1}} \setminus D| > M - |D|$.

So in any case we have $|B_{l_{i+1}} \setminus D| > M - |D|$. From this inequality it follows that after the step number $i+1$ the algorithm with thresholds α and γ must finish the work. Thus, $k = i+1$, $k \leq t$, $(l_1, \ldots, l_{k-1}) = (g_1, \ldots, g_{k-1})$ and $l_k \neq g_k$. □

2.6 Lower Bound on $C_{\min}(\alpha)$

In this subsection we fix some information about the work of greedy algorithm with two thresholds and find the best lower bound on the value $C_{\min}(\alpha)$ depending on this information.

Let (A, S, w) be a set cover problem with weights and α, γ be real numbers such that $0 \leq \gamma \leq \alpha < 1$. Let us apply the greedy algorithm with thresholds α and γ to the set cover problem with weights (A, S, w). Let during the construction of α-cover the greedy algorithm choose sequentially subsets $B_{g_1}, \ldots, B_{g_t}$.

Let us denote $B_{g_0} = \emptyset$ and $\delta_0 = 0$. For $i = 1, \ldots, t$ denote $\delta_i = |B_{g_i} \setminus (B_{g_0} \cup \ldots \cup B_{g_{i-1}})|$ and $w_i = w(B_{g_i})$.

As information on the greedy algorithm work we will use numbers $M_C = M_C(\alpha,\gamma,A,S,w) = \lceil |A|(1-\alpha)\rceil$ and $N_C = N_C(\alpha,\gamma,A,S,w) = \lceil |A|(1-\gamma)\rceil$, and tuples $\Delta_C = \Delta_C(\alpha,\gamma,A,S,w) = (\delta_1,\ldots,\delta_t)$ and $W_C = W_C(\alpha,\gamma,A,S,w) = (w_1,\ldots,w_t)$.

For $i = 0,\ldots,t-1$ denote

$$\rho_i = \left\lceil \frac{w_{i+1}(M_C - (\delta_0 + \ldots + \delta_i))}{\min\{\delta_{i+1}, N_C - (\delta_0 + \ldots + \delta_i)\}} \right\rceil \ .$$

Let us define parameter $\rho_C(\alpha,\gamma) = \rho_C(\alpha,\gamma,A,S,w)$ as follows:

$$\rho_C(\alpha,\gamma) = \max\{\rho_i : i = 0,\ldots,t-1\} \ .$$

We will prove that $\rho_C(\alpha,\gamma)$ is the best lower bound on $C_{\min}(\alpha)$ depending on M_C, N_C, Δ_C and W_C. This lower bound is based on a generalization of the following simple reasoning: if we must cover M elements and the maximal cardinality of a subset from S is δ then we must use at least $\lceil \frac{M}{\delta} \rceil$ subsets.

Theorem 11. *For any set cover problem with weights (A,S,w) and any real numbers α,γ, $0 \le \gamma \le \alpha < 1$, the inequality $C_{\min}(\alpha,A,S,w) \ge \rho_C(\alpha,\gamma,A,S,w)$ holds, and there exists a set cover problem with weights (A',S',w') such that*

$$M_C(\alpha,\gamma,A',S',w') = M_C(\alpha,\gamma,A,S,w), N_C(\alpha,\gamma,A',S',w') = N_C(\alpha,\gamma,A,S,w)$$
$$\Delta_C(\alpha,\gamma,A',S',w') = \Delta_C(\alpha,\gamma,A,S,w), W_C(\alpha,\gamma,A',S',w') = W_C(\alpha,\gamma,A,S,w)$$
$$\rho_C(\alpha,\gamma,A',S',w') = \rho_C(\alpha,\gamma,A,S,w), C_{\min}(\alpha,A',S',w') = \rho_C(\alpha,\gamma,A',S',w') \ .$$

Proof. Let (A,S,w) be a set cover problem with weights, $S = \{B_1,\ldots,B_m\}$, and α,γ be real numbers such that $0 \le \gamma \le \alpha < 1$. Let us denote $M = M_C(\alpha,\gamma,A,S,w) = \lceil |A|(1-\alpha)\rceil$ and $N = N_C(\alpha,\gamma,A,S,w) = \lceil |A|(1-\gamma)\rceil$. Let $\{B_{l_1},\ldots,B_{l_k}\}$ be an optimal α-cover for (A,S,w), i.e. $w(B_{l_1}) + \ldots + w(B_{l_k}) = C_{\min}(\alpha,A,S,w) = C_{\min}(\alpha)$ and $|B_{l_1} \cup \ldots \cup B_{l_k}| \ge M$.

Let us apply the greedy algorithm with thresholds α and γ to (A,S,w). Let during the construction of α-cover the greedy algorithm choose sequentially subsets $B_{g_1},\ldots,B_{g_t}$. Let us denote $B_{g_0} = \emptyset$.

Let $i \in \{0,\ldots,t-1\}$. Let us denote $D = B_{g_0} \cup \ldots \cup B_{g_i}$. It is clear that after i steps of greedy algorithm work in the set $B_{l_1} \cup \ldots \cup B_{l_k}$ at least $|B_{l_1} \cup \ldots \cup B_{l_k}| - |B_{g_0} \cup \ldots \cup B_{g_i}| \ge M - |D| > 0$ elements remained uncovered. After i-th step $p_1 = |B_{l_1} \setminus D|$ elements remained uncovered in the set B_{l_1}, ..., and $p_k = |B_{l_k} \setminus D|$ elements remained uncovered in the set B_{l_k}. We know that $p_1 + \ldots + p_k \ge M - |D| > 0$. Let, for the definiteness, $p_1 > 0,\ldots,p_r > 0, p_{r+1} = \ldots = p_k = 0$. For $j = 1,\ldots,r$ denote $q_j = \min\{p_j, N - |D|\}$. It is clear that $N - |D| \ge M - |D|$. Therefore $q_1 + \ldots + q_r \ge M - |D|$. Let us consider numbers $\frac{w(B_{l_1})}{q_1},\ldots,\frac{w(B_{l_r})}{q_r}$. Let us show that at least one of these numbers is at most $\beta = \frac{w(B_{l_1})+\ldots+w(B_{l_r})}{q_1+\ldots+q_r}$. Assume the contrary. Then $w(B_{l_1}) + \ldots + w(B_{l_r}) = \frac{w(B_{l_1})q_1}{q_1} + \ldots + \frac{w(B_{l_r})q_r}{q_r} > (q_1 + \ldots + q_r)\beta = w(B_{l_1}) + \ldots + w(B_{l_r})$ which is impossible.

We know that $q_1 + \ldots + q_r \geq M - |D|$ and $w(B_{l_1}) + \ldots + w(B_{l_r}) \leq C_{\min}(\alpha)$. Therefore $\beta \leq \frac{C_{\min}(\alpha)}{M-|D|}$, and there exists $j \in \{1, \ldots, k\}$ such that $B_{l_j} \setminus D \neq \emptyset$ and $\frac{w(B_{l_j})}{\min\{|B_{l_j} \setminus D|, N-|D|\}} \leq \beta$. Hence $\frac{w(B_{g_{i+1}})}{\min\{|B_{g_{i+1}} \setminus D|, N-|D|\}} \leq \beta \leq \frac{C_{\min}(\alpha)}{M-|D|}$ and $C_{\min}(\alpha) \geq \frac{w(B_{g_{i+1}})(M-|D|)}{\min\{|B_{g_{i+1}} \setminus D|, N-|D|\}}$.

Taking into account that $C_{\min}(\alpha)$ is a natural number we obtain $C_{\min}(\alpha) \geq \left\lceil \frac{w(B_{g_{i+1}})(M-|D|)}{\min\{|B_{g_{i+1}} \setminus D|, N-|D|\}} \right\rceil = \rho_i$. Since last inequality holds for any $i \in \{0, \ldots, t-1\}$ and $\rho_C(\alpha, \gamma) = \rho_C(\alpha, \gamma, A, S, w) = \max\{\rho_i : i = 0, \ldots, t-1\}$, we conclude that $C_{\min}(\alpha) \geq \rho_C(\alpha, \gamma)$.

Let us show that this bound is unimprovable depending on M_C, N_C, Δ_C and W_C. Let us consider a set cover problem with weights (A', S', w') where $A' = A$, $S' = \{B_1, \ldots, B_m, B_{m+1}\}$, $|B_{m+1}| = M$, $B_{g_1} \cup \ldots \cup B_{g_{t-1}} \subseteq B_{m+1} \subseteq B_{g_1} \cup \ldots \cup B_{g_t}$, $w'(B_1) = w(B_1), \ldots, w'(B_m) = w(B_m)$ and $w'(B_{m+1}) = \rho_C(\alpha, \gamma)$. It is clear that $M_C(\alpha, \gamma, A', S', w') = M_C(\alpha, \gamma, A, S, w) = M$ and $N_C(\alpha, \gamma, A', S', w') = N_C(\alpha, \gamma, A, S, w) = N$. We show $\Delta_C(\alpha, \gamma, A', S', w') = \Delta_C(\alpha, \gamma, A, S, w)$ and $W_C(\alpha, \gamma, A', S', w') = W_C(\alpha, \gamma, A, S, w)$.

Let us show by induction on $i \in \{1, \ldots, t\}$ that for the set cover problem with weights (A', S', w') at the step number i the greedy algorithm with two thresholds α and γ will choose the subset B_{g_i}. Let us consider the first step. Let us denote $D = \emptyset$. It is clear that $\frac{w'(B_{m+1})}{\min\{|B_{m+1} \setminus D|, N-|D|\}} = \frac{\rho_C(\alpha,\gamma)}{M-|D|}$. From the definition of $\rho_C(\alpha, \gamma)$ it follows that $\frac{w'(B_{g_1})}{\min\{|B_{g_1} \setminus D|, N-|D|\}} = \frac{w(B_{g_1})}{\min\{|B_{g_1} \setminus D|, N-|D|\}} \leq \frac{\rho_C(\alpha,\gamma)}{M-|D|}$. Using this fact and the inequality $g_1 < m + 1$ it is not difficult to prove that at the first step greedy algorithm will choose the subset B_{g_1}.

Let $i \in \{1, \ldots, t-1\}$. Let us assume that the greedy algorithm made i steps for (A', S', w') and chose subsets $B_{g_1}, \ldots, B_{g_i}$. Let us show that at the step $i+1$ the subset $B_{g_{i+1}}$ will be chosen. Let us denote $D = B_{g_1} \cup \ldots \cup B_{g_i}$. Since $B_{g_1} \cup \ldots \cup B_{g_i} \subseteq B_{m+1}$ and $|B_{m+1}| = M$, we have $|B_{m+1} \setminus D| = M - |D|$. Therefore $\frac{w'(B_{m+1})}{\min\{|B_{m+1} \setminus D|, N-|D|\}} = \frac{\rho_C(\alpha,\gamma)}{M-|D|}$. From the definition of the parameter $\rho_C(\alpha, \gamma)$ it follows that $\frac{w'(B_{g_{i+1}})}{\min\{|B_{g_{i+1}} \setminus D|, N-|D|\}} = \frac{w(B_{g_{i+1}})}{\min\{|B_{g_{i+1}} \setminus D|, N-|D|\}} \leq \frac{\rho_C(\alpha,\gamma)}{M-|D|}$. Using this fact and the inequality $g_{i+1} < m + 1$ it is not difficult to prove that at the step number $i+1$ greedy algorithm will choose the subset $B_{g_{i+1}}$.

Thus, $\Delta_C(\alpha, \gamma, A', S', w') = \Delta_C(\alpha, \gamma, A, S, w)$ and $W_C(\alpha, \gamma, A', S', w') = W_C(\alpha, \gamma, A, S, w)$. Therefore $\rho_C(\alpha, \gamma, A', S', w') = \rho_C(\alpha, \gamma, A, S, w) = \rho_C(\alpha, \gamma)$. From been proven it follows that $C_{\min}(\alpha, A', S', w') \geq \rho_C(\alpha, \gamma, A', S', w')$. It is clear that $\{B_{m+1}\}$ is an α-cover for (A', S') and the weight of $\{B_{m+1}\}$ is equal to $\rho_C(\alpha, \gamma, A', S', w')$. Hence $C_{\min}(\alpha, A', S', w') = \rho_C(\alpha, \gamma, A', S', w')$. □

Let us consider a property of the parameter $\rho_C(\alpha, \gamma)$ which is important for practical use of the bound from Theorem 11.

Proposition 2. *Let (A, S, w) be a set cover problem with weights and α, γ be real numbers such that $0 \leq \gamma \leq \alpha < 1$. Then $\rho_C(\alpha, \alpha, A, S, w) \geq \rho_C(\alpha, \gamma, A, S, w)$.*

Proof. Let $S = \{B_1, \ldots, B_m\}$, $M = \lceil |A|(1-\alpha) \rceil$, $N = \lceil |A|(1-\gamma) \rceil$, $\rho_C(\alpha, \alpha) = \rho_C(\alpha, \alpha, A, S, w)$ and $\rho_C(\alpha, \gamma) = \rho_C(\alpha, \gamma, A, S, w)$.

Let us apply the greedy algorithm with thresholds α and α to (A, S, w). Let during the construction of α-cover this algorithm choose sequentially subsets $B_{g_1}, \ldots, B_{g_t}$. Let us denote $B_{g_0} = \emptyset$. For $j = 0, \ldots, t-1$ denote $D_j = B_{g_0} \cup \ldots \cup B_{g_j}$ and $\rho_C(\alpha, \alpha, j) = \left\lceil \frac{w(B_{g_{j+1}})(M-|D_j|)}{\min\{|B_{g_{j+1}} \setminus D_j|, M-|D_j|\}} \right\rceil$. Then $\rho_C(\alpha, \alpha) = \max\{\rho_C(\alpha, \alpha, j) : j = 0, \ldots, t-1\}$.

Apply the greedy algorithm with thresholds α and γ to (A, S, w). Let during the construction of α-cover this algorithm choose sequentially subsets $B_{l_1}, \ldots, B_{l_k}$. From Proposition 1 it follows that either $k = t$ and $(l_1, \ldots, l_k) = (g_1, \ldots, g_t)$ or $k \le t$, $(l_1, \ldots, l_{k-1}) = (g_1, \ldots, g_{k-1})$ and $l_k \ne g_k$. Let us consider these two cases separately. Let $k = t$ and $(l_1, \ldots, l_k) = (g_1, \ldots, g_t)$. For $j = 0, \ldots, t-1$ denote $\rho_C(\alpha, \gamma, j) = \left\lceil \frac{w(B_{g_{j+1}})(M-|D_j|)}{\min\{|B_{g_{j+1}} \setminus D_j|, N-|D_j|\}} \right\rceil$. Then $\rho_C(\alpha, \gamma) = \max\{\rho_C(\alpha, \gamma, j) : j = 0, \ldots, t-1\}$. Since $N \ge M$, we have $\rho_C(\alpha, \gamma, j) \le \rho_C(\alpha, \alpha, j)$ for $j = 0, \ldots, t-1$. Hence $\rho_C(\alpha, \gamma) \le \rho_C(\alpha, \alpha)$. Let $k \le t$, $(l_1, \ldots, l_{k-1}) = (g_1, \ldots, g_{k-1})$ and $l_k \ne g_k$. Let us denote $\rho_C(\alpha, \gamma, k-1) = \left\lceil \frac{w(B_{l_k})(M-|D_{k-1}|)}{\min\{|B_{l_k} \setminus D_{k-1}|, N-|D_{k-1}|\}} \right\rceil$ and $\rho_C(\alpha, \gamma, j) = \left\lceil \frac{w(B_{g_{j+1}})(M-|D_j|)}{\min\{|B_{g_{j+1}} \setminus D_j|, N-|D_j|\}} \right\rceil$ for $j = 0, \ldots, k-2$. Then $\rho_C(\alpha, \gamma) = \max\{\rho_C(\alpha, \gamma, j) : j = 0, \ldots, k-1\}$. Since $N \ge M$, we have $\rho_C(\alpha, \gamma, j) \le \rho_C(\alpha, \alpha, j)$ for $j = 0, \ldots, k-2$. It is clear that $\frac{w(B_{l_k})}{\min\{|B_{l_k} \setminus D_{k-1}|, N-|D_{k-1}|\}} \le \frac{w(B_{g_k})}{\min\{|B_{g_k} \setminus D_{k-1}|, N-|D_{k-1}|\}} \le \frac{w(B_{g_k})}{\min\{|B_{g_k} \setminus D_{k-1}|, M-|D_{k-1}|\}}$. Thus, $\rho_C(\alpha, \gamma, k-1) \le \rho_C(\alpha, \alpha, k-1)$ and $\rho_C(\alpha, \gamma) \le \rho_C(\alpha, \alpha)$. □

2.7 Upper Bounds on $C^{\gamma}_{\text{greedy}}(\alpha)$

In this subsection we study some properties of parameter $\rho_C(\alpha, \gamma)$ and obtain two upper bounds on the value $C^{\gamma}_{\text{greedy}}(\alpha)$ which do not depend directly on cardinality of the set A and cardinalities of subsets B_i from S.

Theorem 12. *Let (A, S, w) be a set cover problem with weights and α, γ be real numbers such that $0 \le \gamma < \alpha < 1$. Then*

$$C^{\gamma}_{\text{greedy}}(\alpha, A, S, w) < \rho_C(\gamma, \gamma, A, S, w) \left(\ln \left(\frac{1-\gamma}{\alpha-\gamma} \right) + 1 \right) .$$

Proof. Let $S = \{B_1, \ldots, B_m\}$. Let us denote $M = \lceil |A|(1-\alpha) \rceil$ and $N = \lceil |A|(1-\gamma) \rceil$.

Let us apply the greedy algorithm with thresholds γ and γ to (A, S, w). Let during the construction of γ-cover the greedy algorithm choose sequentially subsets $B_{g_1}, \ldots, B_{g_t}$. Let us denote $B_{g_0} = \emptyset$, for $i = 0, \ldots, t-1$ denote $D_i = B_{g_0} \cup \ldots \cup B_{g_i}$, and denote $\rho = \rho_C(\gamma, \gamma, A, S, w)$. Immediately from the definition of the parameter ρ it follows that for $i = 0, \ldots, t-1$

$$\frac{w(B_{g_{i+1}})}{\min\{|B_{g_{i+1}} \setminus D_i|, N-|D_i|\}} \le \frac{\rho}{N-|D_i|} . \quad (1)$$

Note that $\min\{|B_{g_{i+1}} \setminus D_i|, N - |D_i|\} = |B_{g_{i+1}} \setminus D_i|$ for $i = 0, \ldots, t-2$ since $\{B_{g_0}, \ldots, B_{g_{i+1}}\}$ is not a γ-cover for (A, S). Therefore for $i = 0, \ldots, t-2$ we have $\frac{w(B_{g_{i+1}})}{|B_{g_{i+1}} \setminus D_i|} \leq \frac{\rho}{N-|D_i|}$ and $\frac{N-|D_i|}{\rho} \leq \frac{|B_{g_{i+1}} \setminus D_i|}{w(B_{g_{i+1}})}$. Thus, for $i = 1, \ldots, t-1$ during the step number i the greedy algorithm covers at least $\frac{N-|D_{i-1}|}{\rho}$ elements on each unit of weight. From (1) it follows that that for $i = 0, \ldots, t-1$

$$w(B_{g_{i+1}}) \leq \frac{\rho \min\{|B_{g_{i+1}} \setminus D_i|, N - |D_i|\}}{N - |D_i|} \leq \rho \ . \qquad (2)$$

Assume that $\rho = 1$. Using (2) we obtain $w(B_{g_1}) = 1$. From this equality and (1) it follows that $|B_{g_1}| \geq N$. Therefore $\{B_{g_1}\}$ is an α-cover for (A, S), and $C^{\gamma}_{\mathrm{greedy}}(\alpha) = 1$. It is clear that $\ln\left(\frac{1-\gamma}{\alpha-\gamma}\right) + 1 > 1$. Therefore the statement of the theorem holds if $\rho = 1$.

Assume now that $\rho \geq 2$. Let $|B_{g_1}| \geq M$. Then $\{B_{g_1}\}$ is an α-cover for (A, S). Using (2) we obtain $C^{\gamma}_{\mathrm{greedy}}(\alpha) \leq \rho$. Since $\ln\left(\frac{1-\gamma}{\alpha-\gamma}\right) + 1 > 1$, we conclude that the statement of the theorem holds if $|B_{g_1}| \geq M$. Let $|B_{g_1}| < M$. Then there exists $q \in \{1, \ldots, t-1\}$ such that $|B_{g_1} \cup \ldots \cup B_{g_q}| < M$ and $|B_{g_1} \cup \ldots \cup B_{g_{q+1}}| \geq M$.

Taking into account that for $i = 1, \ldots, q$ during the step number i the greedy algorithm covers at least $\frac{N-|D_{i-1}|}{\rho}$ elements on each unit of weight we obtain $N - |B_{g_1} \cup \ldots \cup B_{g_q}| \leq N\left(1 - \frac{1}{\rho}\right)^{w(B_{g_1}) + \ldots + w(B_{g_q})}$. Let us denote $k = w(B_{g_1}) + \ldots + w(B_{g_q})$. Then $N - N\left(1 - \frac{1}{\rho}\right)^k \leq |B_{g_1} \cup \ldots \cup B_{g_q}| \leq M - 1$. Therefore $|A|(1-\gamma) - |A|(1-\gamma)\left(1 - \frac{1}{\rho}\right)^k < |A|(1-\alpha)$, $1 - \gamma - 1 + \alpha < (1-\gamma)\left(\frac{\rho-1}{\rho}\right)^k$, $\left(\frac{\rho}{\rho-1}\right)^k < \frac{1-\gamma}{\alpha-\gamma}$, $\left(1 + \frac{1}{\rho-1}\right)^k < \frac{1-\gamma}{\alpha-\gamma}$, and $\frac{k}{\rho} < \ln\left(\frac{1-\gamma}{\alpha-\gamma}\right)$. To obtain last inequality we use known inequality $\ln\left(1 + \frac{1}{r}\right) > \frac{1}{r+1}$ which holds for any natural r. It is clear that $C^{\gamma}_{\mathrm{greedy}}(\alpha) = k + w(B_{q+1})$. Using (2) we conclude that $w(B_{q+1}) \leq \rho$. Therefore $C^{\gamma}_{\mathrm{greedy}}(\alpha) < \rho \ln\left(\frac{1-\gamma}{\alpha-\gamma}\right) + \rho$. □

Corollary 1. *Let ε be a real number, and $0 < \varepsilon < 1$. Then for any α such that $\varepsilon \leq \alpha < 1$ the following inequalities hold:*

$$\rho_C(\alpha, \alpha) \leq C_{\min}(\alpha) \leq C^{\alpha-\varepsilon}_{\mathrm{greedy}}(\alpha) < \rho_C(\alpha - \varepsilon, \alpha - \varepsilon)\left(\ln \frac{1}{\varepsilon} + 1\right) \ .$$

For example, if $\varepsilon = 0.01$ and $0.01 \leq \alpha < 1$ then $\rho_C(\alpha, \alpha) \leq C_{\min}(\alpha) \leq C^{\alpha-0.01}_{\mathrm{greedy}}(\alpha) < 5.61\rho_C(\alpha - 0.01, \alpha - 0.01)$, and if $\varepsilon = 0.1$ and $0.1 \leq \alpha < 1$ then $\rho_C(\alpha, \alpha) \leq C_{\min}(\alpha) \leq C^{\alpha-0.1}_{\mathrm{greedy}}(\alpha) < 3.31\rho_C(\alpha - 0.1, \alpha - 0.1)$.

The obtained results show that the lower bound $C_{\min}(\alpha) \geq \rho_C(\alpha, \alpha)$ is non-trivial.

Theorem 13. *Let (A, S, w) be a set cover problem with weights and α, γ be real numbers such that $0 \leq \gamma < \alpha < 1$. Then*

$$C^{\gamma}_{\mathrm{greedy}}(\alpha, A, S, w) < C_{\min}(\gamma, A, S, w)\left(\ln\left(\frac{1-\gamma}{\alpha-\gamma}\right) + 1\right) \ .$$

Proof. From Theorem 12 it follows that $C^{\gamma}_{\rm greedy}(\alpha, A, S, w) < \rho_C(\gamma, \gamma, A, S, w) \cdot \left(\ln\left(\frac{1-\gamma}{\alpha-\gamma}\right) + 1\right)$. The inequality $\rho_C(\gamma, \gamma, A, S, w) \leq C_{\min}(\gamma, A, S, w)$ follows from Theorem 11. □

Corollary 2. $C^{0}_{\rm greedy}(0.001) < 7.91C_{\min}(0)$, $C^{0.001}_{\rm greedy}(0.01) < 5.71C_{\min}(0.001)$, $C^{0.1}_{\rm greedy}(0.2) < 3.20C_{\min}(0.1)$, $C^{0.3}_{\rm greedy}(0.5) < 2.26C_{\min}(0.3)$.

Corollary 3. *Let* $0 < \alpha < 1$. *Then* $C^{0}_{\rm greedy}(\alpha) < C_{\min}(0)\left(\ln\frac{1}{\alpha} + 1\right)$.

Corollary 4. *Let* ε *be a real number, and* $0 < \varepsilon < 1$. *Then for any* α *such that* $\varepsilon \leq \alpha < 1$ *the inequalities* $C_{\min}(\alpha) \leq C^{\alpha-\varepsilon}_{\rm greedy}(\alpha) < C_{\min}(\alpha - \varepsilon)\left(\ln\frac{1}{\varepsilon} + 1\right)$ *hold.*

3 Partial Tests and Reducts

3.1 Main Notions

Let T be a table with n rows labeled by nonnegative integers (decisions) and m columns labeled by attributes (names of attributes) $f_1, \ldots, f_m$. This table is filled by nonnegative integers (values of attributes). The table T is called a *decision table*. Let w be a *weight function* for T which corresponds to each attribute f_i a natural number $w(f_i)$.

Let us denote by $P(T)$ the set of unordered pairs of different rows of T with different decisions. We will say that an attribute f_i *separates* a pair of rows $(r_1, r_2) \in P(T)$ if rows r_1 and r_2 have different numbers at the intersection with the column f_i. For $i = 1, \ldots, m$ denote by $P(T, f_j)$ the set of pairs from $P(T)$ which the attribute f_i separates.

Let α be a real number such that $0 \leq \alpha < 1$. A set of attributes $Q \subseteq \{f_1, \ldots, f_m\}$ will be called an α-*test* for T if attributes from Q separate at least $(1-\alpha)|P(T)|$ pairs from the set $P(T)$. An α-test is called an α-*reduct* if each proper subset of the considered α-test is not α-test. If $P(T) = \emptyset$ then each subset of $\{f_1, \ldots, f_m\}$ is an α-test, and only empty set is an α-reduct.

For example, 0.01-test means that we must separate at least 99% of pairs from $P(T)$.

Note that 0-reduct is usual (exact) reduct. It must be noted also that each α-test contains at least one α-reduct as a subset.

The number $w(Q) = \sum_{f_i \in Q} w(f_i)$ will be called the *weight* of the set Q. If $Q = \emptyset$ then $w(Q) = 0$.

Let us denote by $R_{\min}(\alpha) = R_{\min}(\alpha, T, w)$ the minimal weight of α-reduct for T. It is clear that $R_{\min}(\alpha, T, w)$ coincides with the minimal weight of α-test for T.

Let α, γ be real numbers such that $0 \leq \gamma \leq \alpha < 1$. Let us describe a greedy algorithm with thresholds α and γ which constructs an α-test for given decision table T and weight function w.

If $P(T) = \emptyset$ then the constructed α-test is empty set. Let $P(T) \neq \emptyset$. Let us denote $M = \lceil |P(T)|(1-\alpha) \rceil$ and $N = \lceil |P(T)|(1-\gamma) \rceil$. Let we make $i \geq 0$

steps and construct a set Q containing i attributes (if $i = 0$ then $Q = \emptyset$). Let us describe the step number $i + 1$.

Let us denote by D the set of pairs from $P(T)$ separated by attributes from Q (if $i = 0$ then $D = \emptyset$). If $|D| \geq M$ then we finish the work of the algorithm. The set of attributes Q is the constructed α-test. Let $|D| < M$. Then we choose an attribute f_j with minimal number j for which $P(T, f_j) \setminus D \neq \emptyset$ and the value

$$\frac{w(f_j)}{\min\{|P(T, f_j) \setminus D|, N - |D|\}}$$

is minimal. Add the attribute f_j to the set Q. Pass to the step number $i + 2$.

Let us denote by $R^{\gamma}_{\text{greedy}}(\alpha) = R^{\gamma}_{\text{greedy}}(\alpha, T, w)$ the weight of α-test constructed by greedy algorithm with thresholds α and γ for given decision table T and weight function w.

3.2 Relationships Between Partial Covers and Partial Tests

Let (A, S, w) be a set cover problem with weights and α, γ be real numbers such that $0 \leq \gamma \leq \alpha < 1$. Let us apply the greedy algorithm with thresholds α and γ to (A, S, w). Let during the construction of α-cover the greedy algorithm choose sequentially subsets $B_{j_1}, \ldots, B_{j_t}$ from the family S. Let us denote $O_C(\alpha, \gamma, A, S, w) = (j_1, \ldots, j_t)$.

Let T be a decision table with m columns labeled by attributes $f_1, \ldots, f_m$, and with a nonempty set $P(T)$. Let w be a weight function for T. We correspond a set cover problem with weights $(A(T), S(T), u_w)$ to the considered decision table T and weight function w in the following way: $A(T) = P(T)$, $S(T) = \{B_1(T), \ldots, B_m(T)\}$ where $B_1(T) = P(T, f_1), \ldots, B_m(T) = P(T, f_m)$, $u_w(B_1(T)) = w(f_1), \ldots, u_w(B_m(T)) = w(f_m)$.

Let α, γ be real numbers such that $0 \leq \gamma \leq \alpha < 1$. Let us apply the greedy algorithm with thresholds α and γ to decision table T and weight function w. Let during the construction of α-test the greedy algorithm choose sequentially attributes $f_{j_1}, \ldots, f_{j_t}$. Let us denote $O_R(\alpha, \gamma, T, w) = (j_1, \ldots, j_t)$.

Let us denote $P(T, f_{j_0}) = \emptyset$. For $i = 1, \ldots, t$ denote $w_i = w(f_{j_i})$ and

$$\delta_i = |P(T, f_{j_i}) \setminus (P(T, f_{j_0}) \cup \ldots \cup P(T, f_{j_{i-1}}))| \ .$$

Let us denote $M_R(\alpha, \gamma, T, w) = \lceil |P(T)|(1-\alpha) \rceil$, $N_R(\alpha, \gamma, T, w) = \lceil |P(T)|(1-\gamma) \rceil$, $\Delta_R(\alpha, \gamma, T, w) = (\delta_1, \ldots, \delta_t)$ and $W_R(\alpha, \gamma, T, w) = (w_1, \ldots, w_t)$.

It is not difficult to prove the following statement.

Proposition 3. *Let T be a decision table with m columns labeled by attributes $f_1, \ldots, f_m$, $P(T) \neq \emptyset$, w be a weight function for T, and α, γ be real numbers such that $0 \leq \gamma \leq \alpha < 1$. Then*

$$|P(T)| = |A(T)| \ ,$$
$$|P(T, f_i)| = |B_i(T)|, \ \ i = 1, \ldots, m \ ,$$
$$O_R(\alpha, \gamma, T, w) = O_C(\alpha, \gamma, A(T), S(T), u_w) \ ,$$

$$M_R(\alpha,\gamma,T,w) = M_C(\alpha,\gamma,A(T),S(T),u_w) \ ,$$
$$N_R(\alpha,\gamma,T,w) = N_C(\alpha,\gamma,A(T),S(T),u_w) \ ,$$
$$\Delta_R(\alpha,\gamma,T,w) = \Delta_C(\alpha,\gamma,A(T),S(T),u_w) \ ,$$
$$W_R(\alpha,\gamma,T,w) = W_C(\alpha,\gamma,A(T),S(T),u_w) \ ,$$
$$R_{\min}(\alpha,T,w) = C_{\min}(\alpha,A(T),S(T),u_w) \ ,$$
$$R^{\gamma}_{\text{greedy}}(\alpha,T,w) = C^{\gamma}_{\text{greedy}}(\alpha,A(T),S(T),u_w) \ .$$

Let (A,S,w) be a set cover problem with weights where $A = \{a_1,\ldots,a_n\}$ and $S = \{B_1,\ldots,B_m\}$. We correspond a decision table $T(A,S)$ and a weight function v_w for $T(A,S)$ to the set cover problem with weights (A,S,w) in the following way. The table $T(A,S)$ contains m columns labeled by attributes $f_1,\ldots,f_m$ and $n+1$ rows filled by numbers from $\{0,1\}$. For $i = 1,\ldots,n$ and $j = 1,\ldots,m$ at the intersection of i-th row and j-th column the number 1 stays if and only if $a_i \in B_j$. The row number $n+1$ is filled by 0. First n rows are labeled by the decision 0. Last row is labeled by the decision 1. Let $v_w(f_1) = w(B_1),\ldots,v_w(f_m) = w(B_m)$.

For $i = \{1,\ldots,n+1\}$ denote by r_i the i-th row. It is not difficult to see that $P(T(A,S)) = \{(r_1,r_{n+1}),\ldots,(r_n,r_{n+1})\}$. Let $i \in \{1,\ldots,n\}$ and $j \in \{1,\ldots,m\}$. One can show that the attribute f_j separates the pair (r_i,r_{n+1}) if and only if $a_i \in B_j$.

It is not difficult to prove the following statement.

Proposition 4. *Let (A,S,w) be a set cover problem with weights and α,γ be real numbers such that $0 \le \gamma \le \alpha < 1$. Then*

$$|P(T(A,S))| = |A| \ ,$$
$$O_R(\alpha,\gamma,T(A,S),v_w) = O_C(\alpha,\gamma,A,S,w) \ ,$$
$$M_R(\alpha,\gamma,T(A,S),v_w) = M_C(\alpha,\gamma,A,S,w) \ ,$$
$$N_R(\alpha,\gamma,T(A,S),v_w) = N_C(\alpha,\gamma,A,S,w) \ ,$$
$$\Delta_R(\alpha,\gamma,T(A,S),v_w) = \Delta_C(\alpha,\gamma,A,S,w) \ ,$$
$$W_R(\alpha,\gamma,T(A,S),v_w) = W_C(\alpha,\gamma,A,S,w) \ ,$$
$$R_{\min}(\alpha,T(A,S),v_w) = C_{\min}(\alpha,A,S,w) \ ,$$
$$R^{\gamma}_{\text{greedy}}(\alpha,T(A,S),v_w) = C^{\gamma}_{\text{greedy}}(\alpha,A,S,w) \ .$$

3.3 On Precision of Greedy Algorithm with Thresholds α and α

The following two statements are simple corollaries of results of Slavík (see Theorems 4 and 5) and Proposition 3.

Theorem 14. *Let T be a decision table, $P(T) \neq \emptyset$, w be a weight function for T, $\alpha \in \mathbb{R}$ and $0 \le \alpha < 1$. Then $R^{\alpha}_{\text{greedy}}(\alpha) \le R_{\min}(\alpha) H\left(\lceil (1-\alpha)|P(T)|\rceil\right)$.*

Theorem 15. *Let T be a decision table with m columns labeled by attributes $f_1,\ldots,f_m$, $P(T) \neq \emptyset$, w be a weight function for T, and α be a real number such that $0 \le \alpha < 1$. Then $R^{\alpha}_{\text{greedy}}(\alpha) \le R_{\min}(\alpha) H\left(\max_{i\in\{1,\ldots,m\}} |P(T,f_i)|\right)$.*

3.4 On Polynomial Approximate Algorithms

In this subsection we consider three theorems which follows immediately from Theorems 26–28 [9].

Let $0 \leq \alpha < 1$. Let us consider the following problem: for given decision table T and weight function w for T it is required to find an α-test (α-reduct) for T with minimal weight.

Theorem 16. *Let $0 \leq \alpha < 1$. Then the problem of construction of α-test (α-reduct) with minimal weight is NP-hard.*

So we must consider polynomial approximate algorithms for minimization of α-test (α-reduct) weight.

Theorem 17. *Let $\alpha \in \mathbb{R}$ and $0 \leq \alpha < 1$. If $NP \not\subseteq DTIME(n^{O(\log \log n)})$ then for any ε, $0 < \varepsilon < 1$, there is no polynomial algorithm that for given decision table T with $P(T) \neq \emptyset$ and weight function w for T constructs an α-test for T which weight is at most $(1-\varepsilon)R_{\min}(\alpha, T, w) \ln |P(T)|$.*

Theorem 18. *Let α be a real number such that $0 \leq \alpha < 1$. If $P \neq NP$ then there exists $\delta > 0$ such that there is no polynomial algorithm that for given decision table T with $P(T) \neq \emptyset$ and weight function w for T constructs an α-test for T which weight is at most $\delta R_{\min}(\alpha, T, w) \ln |P(T)|$.*

From Theorem 14 it follows that $R^{\alpha}_{\text{greedy}}(\alpha) \leq R_{\min}(\alpha)(1 + \ln |P(T)|)$. From this inequality and from Theorem 17 it follows that under the assumption $NP \not\subseteq DTIME(n^{O(\log \log n)})$ greedy algorithm with two thresholds α and α is close to best polynomial approximate algorithms for minimization of partial test weight. From the considered inequality and from Theorem 18 it follows that under the assumption $P \neq NP$ greedy algorithm with two thresholds α and α is not far from best polynomial approximate algorithms for minimization of partial test weight.

However we can try to improve the results of the work of greedy algorithm with two thresholds α and α for some part of decision tables.

3.5 Two Modifications of Greedy Algorithm

First, we consider binary diagnostic decision tables and prove that under some assumptions on the number of attributes and rows for the most part of tables there exist weight function w and numbers α, γ such that the weight of α-test constructed by greedy algorithm with thresholds α and γ is less than the weight of α-test constructed by greedy algorithm with thresholds α and α.

Binary means that the table is filled by numbers from the set $\{0, 1\}$ (all attributes have values from $\{0, 1\}$). *Diagnostic* means that rows of the table are labeled by pairwise different numbers (decisions). Let T be a binary diagnostic decision table with m columns labeled by attributes $f_1, \ldots, f_m$ and with n rows. We will assume that rows of T with numbers $1, \ldots, n$ are labeled by decisions $1, \ldots, n$ respectively. Therefore the number of considered tables is equal to 2^{mn}. Decision table will be called *simple* if it has no equal rows.

Theorem 19. *Let us consider binary diagnostic decision tables with m columns labeled by attributes $f_1, \ldots, f_m$ and $n \geq 4$ rows labeled by decisions $1, \ldots, n$. The fraction of decision tables T for each of which there exist a weight function w and numbers α, γ such that $0 \leq \gamma < \alpha < 1$ and $R^{\gamma}_{\text{greedy}}(\alpha, T, w) < R^{\alpha}_{\text{greedy}}(\alpha, T, w)$ is at least $1 - \frac{3^m}{n^{\frac{m}{2}-1}} - \frac{n^2}{2^m}$.*

Proof. We will say that a decision table T is not 1-uniform if there exist two attributes f_i and f_j of T such that $|P(T, f_i)| > 0$ and $|P(T, f_j)| \geq |P(T, f_i)| + 2$. Otherwise, we will say that T is 1-uniform. Using Theorem 9 and Proposition 3 we conclude that if T is not 1-uniform then there exist a weight function w and numbers α, γ such that $0 \leq \gamma < \alpha < 1$ and $R^{\gamma}_{\text{greedy}}(\alpha, T, w) < R^{\alpha}_{\text{greedy}}(\alpha, T, w)$.

We evaluate the number of simple decision tables which are 1-uniform.

Let us consider a simple decision table T which is 1-uniform. Let f_i be an attribute of T. It is clear that $|P(T, f_i)| = 0$ if and only if the number of units in the column f_i is equal to 0 or n. Let k, l be natural numbers such that $k, k+l \in \{1, \ldots, n-1\}$, and $i, j \in \{1, \ldots, m\}$, $i \neq j$. Let the decision table T have k units in the column f_i and $k+l$ units in the column f_j. Then $|P(T, f_i)| = k(n-k) = kn - k^2$ and $|P(T, f_j)| = (k+l)(n-k-l) = kn - k^2 + l(n-2k-l)$. Since T is 1-uniform, we have $l(n-2k-l) \in \{0, 1, -1\}$.

Let $l(n-2k-l) = 0$. Then $n-2k-l = 0$ and $l = n-2k$. Since l is a natural number, we have $k < n/2$.

Let $l(n-2k-l) = 1$. Since l, n and k are natural numbers, we have $l = 1$ and $n-2k-1 = 1$. Therefore $k = \frac{n}{2} - 1$. Since k is a natural number, we have n is even.

Let $l(n-2k-l) = -1$. Since l, n and k are natural numbers, we have $l = 1$ and $n-2k-1 = -1$. Therefore $k = \frac{n}{2}$. Since k is a natural number, we have n is even.

Let n be odd. Then there exists natural k such that $1 \leq k < \frac{n}{2}$ and the number of units in each column of T belongs to the set $\{0, n, k, n-k\}$. Therefore the number of considered tables is at most $\sum_{k=1}^{\lfloor n/2 \rfloor} (C_n^k + C_n^{n-k} + 2)^m$. Since $n \geq 4$, we have $2 \leq C_n^{\lfloor n/2 \rfloor}$. Using Lemma 2 we conclude that the number of 1-uniform simple tables is at most $\sum_{k=1}^{\lfloor n/2 \rfloor} \left(3C_n^{\lfloor n/2 \rfloor}\right)^m < n\left(\frac{3 \cdot 2^n}{\sqrt{n}}\right)^m$.

Let n be even. Then there exists natural k such that $1 \leq k < \frac{n}{2} - 1$ and the number of units in each column of T belongs to the set $\{0, n, k, n-k\}$, or the number of units in each column belongs to the set $\{0, n, \frac{n}{2}-1, \frac{n}{2}, \frac{n}{2}+1\}$. Therefore the number of considered tables is at most $\sum_{k=1}^{\lfloor n/2 \rfloor - 2} (C_n^k + C_n^{n-k} + 2)^m + (C_n^{n/2-1} + C_n^{n/2} + C_n^{n/2+1} + 2)^m$. It is well known (see, for example, [25], page 178) that $C_n^r < C_n^{n/2}$ for any $r \in \{1, \ldots, n\} \setminus \{n/2\}$. Therefore the number of 1-uniform tables is at most $n\left(3C_n^{n/2}\right)^m$. Using Lemma 2 we conclude that (as in the case of odd n) the number of 1-uniform simple tables is less than $n\left(\frac{3 \cdot 2^n}{\sqrt{n}}\right)^m = \frac{2^{mn} 3^m}{n^{\frac{m}{2}-1}}$. The number of tables which are not simple is at most $n^2 2^{mn-m}$. Therefore the number of tables which are not 1-uniform is at least

$2^{mn} - \frac{2^{mn}3^m}{n^{\frac{m}{2}-1}} - n^2 2^{mn-m}$. Thus, the fraction, considered in the statement of the theorem, is at least $1 - \frac{3^m}{n^{\frac{m}{2}-1}} - \frac{n^2}{2^m}$. □

So if $m \geq 4$ and n, $\frac{2^m}{n^2}$ are large enough then for the most part of binary diagnostic decision tables there exist weight function w and numbers α, γ such that the weight of α-test constructed by greedy algorithm with thresholds α and γ is less than the weight of α-test constructed by greedy algorithm with thresholds α and α.

The obtained results show that the greedy algorithm with two thresholds α and γ is of some interest. Now we consider two polynomial modifications of greedy algorithm which allow to use advantages of greedy algorithm with two thresholds α and γ.

Let T be a decision table, $P(T) \neq \emptyset$, w be a weight function for T and α be a real number such that $0 \leq \alpha < 1$.

1. It is impossible to consider effectively all γ such that $0 \leq \gamma \leq \alpha$. Instead of this we can consider all natural N such that $M \leq N \leq |P(T)|$ where $M = \lceil |P(T)|(1-\alpha) \rceil$ (see the description of greedy algorithm with two thresholds). For each $N \in \{M, \ldots, |P(T)|\}$ we apply greedy algorithm with parameters M and N to T and w and after that choose an α-test with minimal weight among constructed α-tests.
2. There exists also an another way to construct an α-test which is not worse than the one obtained under consideration of all N such that $M \leq N \leq |P(T)|$. Let us apply greedy algorithm with thresholds α and α to T and w. Let the algorithm choose sequentially attributes $f_{j_1}, \ldots, f_{j_t}$. For each $i \in \{0, \ldots, t-1\}$ we find (if it is possible) an attribute f_{l_i} of T with minimal weight $w(f_{l_i})$ such that the set $\{f_{j_1}, \ldots, f_{j_i}, f_{l_i}\}$ is an α-test for T (if $i = 0$ then it will be the set $\{f_{l_0}\}$). After that among constructed α-tests $\{f_{j_1}, \ldots, f_{j_t}\}$, ..., $\{f_{j_1}, \ldots, f_{j_i}, f_{l_i}\}$, ... we choose an α-test with minimal weight. From Proposition 5 it follows that the constructed α-test is not worse than the one constructed under consideration of all γ, $0 \leq \gamma \leq \alpha$, or (which is the same) all N, $M \leq N \leq |P(T)|$.

Next statement follows immediately from Propositions 1 and 3.

Proposition 5. *Let T be a decision table, $P(T) \neq \emptyset$, w be a weight function for T and α, γ be real numbers such that $0 \leq \gamma < \alpha < 1$. Let the greedy algorithm with two thresholds α and α, which is applied to T and w, choose sequentially attributes $f_{g_1}, \ldots, f_{g_t}$. Let the greedy algorithm with two thresholds α and γ, which is applied to T and w, choose sequentially attributes $f_{l_1}, \ldots, f_{l_k}$. Then either $k = t$ and $(l_1, \ldots, l_k) = (g_1, \ldots, g_t)$ or $k \leq t$, $(l_1, \ldots, l_{k-1}) = (g_1, \ldots, g_{k-1})$ and $l_k \neq g_k$.*

3.6 Bounds on $R_{\min}(\alpha)$ and $R^{\gamma}_{\text{greedy}}(\alpha)$

First, we fix some information about the work of greedy algorithm with two thresholds and find the best lower bound on the value $R_{\min}(\alpha)$ depending on this information.

Let T be a decision table such that $P(T) \neq \emptyset$, w be a weight function for T, and α, γ be real numbers such that $0 \leq \gamma \leq \alpha < 1$. Let us apply the greedy algorithm with thresholds α and γ to the decision table T and the weight function w. Let during the construction of α-test the greedy algorithm choose sequentially attributes $f_{g_1}, \ldots, f_{g_t}$.

Let us denote $P(T, f_{g_0}) = \emptyset$ and $\delta_0 = 0$. For $i = 1, \ldots, t$ denote $\delta_i = |P(T, f_{g_i}) \setminus (P(T, f_{g_0}) \cup \ldots \cup P(T, f_{g_{i-1}}))|$ and $w_i = w(f_{g_i})$.

As information on the greedy algorithm work we will use numbers $M_R = M_R(\alpha, \gamma, T, w) = \lceil |P(T)|(1-\alpha) \rceil$ and $N_R = N_R(\alpha, \gamma, T, w) = \lceil |P(T)|(1-\gamma) \rceil$, and tuples $\Delta_R = \Delta_R(\alpha, \gamma, T, w) = (\delta_1, \ldots, \delta_t)$ and $W_R = W_R(\alpha, \gamma, T, w) = (w_1, \ldots, w_t)$.

For $i = 0, \ldots, t-1$ denote

$$\rho_i = \left\lceil \frac{w_{i+1}(M_R - (\delta_0 + \ldots + \delta_i))}{\min\{\delta_{i+1}, N_R - (\delta_0 + \ldots + \delta_i)\}} \right\rceil \ .$$

Let us define parameter $\rho_R(\alpha, \gamma) = \rho_R(\alpha, \gamma, T, w)$ as follows:

$$\rho_R(\alpha, \gamma) = \max \{\rho_i : i = 0, \ldots, t-1\} \ .$$

We will show that $\rho_R(\alpha, \gamma)$ is the best lower bound on $R_{\min}(\alpha)$ depending on M_R, N_R, Δ_R and W_R. Next statement follows from Theorem 11 and Propositions 3 and 4.

Theorem 20. *For any decision table T with $P(T) \neq \emptyset$, any weight function w for T, and any real numbers α, γ, $0 \leq \gamma \leq \alpha < 1$, the inequality $R_{\min}(\alpha, T, w) \geq \rho_R(\alpha, \gamma, T, w)$ holds, and there exist a decision table T' and a weight function w' for T' such that*

$$M_R(\alpha, \gamma, T', w') = M_R(\alpha, \gamma, T, w), \ N_R(\alpha, \gamma, T', w') = N_R(\alpha, \gamma, T, w) \ ,$$
$$\Delta_R(\alpha, \gamma, T', w') = \Delta_R(\alpha, \gamma, T, w), \ W_R(\alpha, \gamma, T', w') = W_R(\alpha, \gamma, T, w) \ ,$$
$$\rho_R(\alpha, \gamma, T', w') = \rho_R(\alpha, \gamma, T, w), \ R_{\min}(\alpha, T', w') = \rho_R(\alpha, \gamma, T', w') \ .$$

Let us consider a property of the parameter $\rho_R(\alpha, \gamma)$ which is important for practical use of the bound from Theorem 20. Next statement follows from Propositions 2 and 3.

Proposition 6. *Let T be a decision table with $P(T) \neq \emptyset$, w be a weight function for T, $\alpha, \gamma \in \mathbb{R}$ and $0 \leq \gamma \leq \alpha < 1$. Then $\rho_R(\alpha, \alpha, T, w) \geq \rho_R(\alpha, \gamma, T, w)$.*

Now we study some properties of parameter $\rho_R(\alpha, \gamma)$ and obtain two upper bounds on the value $R^{\gamma}_{\text{greedy}}(\alpha)$ which do not depend directly on cardinality of the set $P(T)$ and cardinalities of subsets $P(T, f_i)$.

Next statement follows from Theorem 12 and Proposition 3.

Theorem 21. *Let T be a decision table with $P(T) \neq \emptyset$, w be a weight function for T and α, γ be real numbers such that $0 \leq \gamma < \alpha < 1$. Then*

$$R^{\gamma}_{\text{greedy}}(\alpha, T, w) < \rho_R(\gamma, \gamma, T, w) \left(\ln \left(\frac{1-\gamma}{\alpha-\gamma} \right) + 1 \right) \ .$$

Corollary 5. *Let $\varepsilon \in \mathbb{R}$ and $0 < \varepsilon < 1$. Then for any α, $\varepsilon \leq \alpha < 1$, the inequalities $\rho_C(\alpha,\alpha) \leq R_{\min}(\alpha) \leq R^{\alpha-\varepsilon}_{\text{greedy}}(\alpha) < \rho_R(\alpha-\varepsilon,\alpha-\varepsilon)\left(\ln\frac{1}{\varepsilon}+1\right)$ hold.*

For example, $\left(\ln\frac{1}{0.01}+1\right) < 5.61$ and $\left(\ln\frac{1}{0.1}+1\right) < 3.31$. The obtained results show that the lower bound $R_{\min}(\alpha) \geq \rho_R(\alpha,\alpha)$ is nontrivial.

Next statement follows from Theorem 13 and Proposition 3.

Theorem 22. *Let T be a decision table with $P(T) \neq \emptyset$, w be a weight function for T and α, γ be real numbers such that $0 \leq \gamma < \alpha < 1$. Then*

$$R^{\gamma}_{\text{greedy}}(\alpha,T,w) < R_{\min}(\gamma,T,w)\left(\ln\left(\frac{1-\gamma}{\alpha-\gamma}\right)+1\right) .$$

Corollary 6. *$R^0_{\text{greedy}}(0.001) < 7.91R_{\min}(0)$, $R^{0.001}_{\text{greedy}}(0.01) < 5.71R_{\min}(0.001)$, $R^{0.1}_{\text{greedy}}(0.2) < 3.20C_{\min}(0.1)$, $R^{0.3}_{\text{greedy}}(0.5) < 2.26R_{\min}(0.3)$.*

Corollary 7. *Let $0 < \alpha < 1$. Then $R^0_{\text{greedy}}(\alpha) < R_{\min}(0)\left(\ln\frac{1}{\alpha}+1\right)$.*

Corollary 8. *Let ε be a real number, and $0 < \varepsilon < 1$. Then for any α such that $\varepsilon \leq \alpha < 1$ the inequalities $R_{\min}(\alpha) \leq R^{\alpha-\varepsilon}_{\text{greedy}}(\alpha) < R_{\min}(\alpha-\varepsilon)\left(\ln\frac{1}{\varepsilon}+1\right)$ hold.*

3.7 Results of Experiments for α-Tests and α-Reducts

In this subsection we will consider only binary decision tables with binary decision attributes.

First Group of Experiments. First group of experiments is connected with study of quality of greedy algorithm with one threshold (where $\gamma = \alpha$ or, which is the same, $N = M$), and comparison of quality of greedy algorithm with one threshold and first modification of greedy algorithm (where for each $N \in \{M,\ldots,|P(T)|\}$ we apply greedy algorithm with parameters M and N to decision table and weight function and after that choose an α-test with minimal weight among constructed α-tests).

We generate randomly 1000 decision tables T and weight functions w such that T contains 10 rows and 10 conditional attributes $f_1,\ldots,f_{10}$, and $1 \leq w(f_i) \leq 1000$ for $i = 1,\ldots,10$.

For each $\alpha \in \{0.0, 0.1, \ldots, 0.9\}$ we find the number of pairs (T,w) for which greedy algorithm with one threshold constructs an α-test with minimal weight (an optimal α-test), i.e. $R^{\alpha}_{\text{greedy}}(\alpha,T,w) = R_{\min}(\alpha,T,w)$. This number is contained in the row of Table 1 labeled by "Opt".

We find the number of pairs (T,w) for which first modification of greedy algorithm constructs an α-test which weight is less than the weight of α-test constructed by greedy algorithm with one threshold, i.e. there exists γ such that $0 \leq \gamma < \alpha$ and $R^{\gamma}_{\text{greedy}}(\alpha,T,w) < R^{\alpha}_{\text{greedy}}(\alpha,T,w)$. This number is contained in the row of Table 1 labeled by "Impr".

Also we find the number of pairs (T,w) for which first modification of greedy algorithm constructs an optimal α-test which weight is less than the weight of

α-test constructed by greedy algorithm with one threshold, i.e. there exists γ such that $0 \leq \gamma < \alpha$ and $R^{\gamma}_{\text{greedy}}(\alpha, T, w) = R_{\min}(\alpha, T, w) < R^{\alpha}_{\text{greedy}}(\alpha, T, w)$. This number is contained in the row of Table 1 labeled by "Opt+".

Table 1. Results of first group of experiments with α-tests

α	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Opt	409	575	625	826	808	818	950	981	992	1000
Impr	0	42	47	33	24	8	6	5	2	0
Opt+	0	22	28	24	22	5	6	5	2	0

The obtained results show that the percentage of pairs for which greedy algorithm with one threshold finds an optimal α-test grows almost monotonically (with local minimum near to 0.4–0.5) from 40.9% up to 100%. The percentage of problems for which first modification of greedy algorithm can improve the result of the work of greedy algorithm with one threshold is less than 5%. However, sometimes (for example, if $\alpha = 0.3$ or $\alpha = 0.7$) the considered improvement is noticeable.

Second Group of Experiments. Second group of experiments is connected with comparison of quality of greedy algorithm with one threshold and first modification of greedy algorithm.

We make 25 experiments (row "Nr" in Table 2 contains the number of experiment). Each experiment includes the work with three randomly generated families of pairs (T, w) (1000 pairs in each family) such that T contains n rows and m conditional attributes, and w has values from the set $\{1, \ldots, v\}$.

If the column "n" contains one number, for example "40", it means that $n = 40$. If this row contains two numbers, for example "30–120", it means that for each of 1000 pairs we choose the number n randomly from the set $\{30, \ldots, 120\}$. The same situation is for the column "m".

If the column "α" contains one number, for example "0.1", it means that $\alpha = 0.1$. If this column contains two numbers, for example "0.2–0.4", it means that we choose randomly the value of α such that $0.2 \leq \alpha \leq 0.4$.

For each of the considered pairs (T, w) and number α we apply greedy algorithm with one threshold and first modification of greedy algorithm. Column "#i", $i = 1, 2, 3$, contains the number of pairs (T, w) from the family number i for each of which the weight of α-test, constructed by first modification of greedy algorithm, is less than the weight of α-test constructed by greedy algorithm with one threshold. In other words, in column "#i" we have the number of pairs (T, w) from the family number i such that there exists γ for which $0 \leq \gamma < \alpha$ and $R^{\gamma}_{\text{greedy}}(\alpha, T, w) < R^{\alpha}_{\text{greedy}}(\alpha, T, w)$. The column "avg" contains the number $\frac{\#1+\#2+\#3}{3}$.

In experiments 1–3 we consider the case where the parameter v increases. In experiments 4–8 the parameter α increases. In experiments 9–12 the parameter m increases. In experiments 13–16 the parameter n increases. In experiments

Table 2. Results of second group of experiments with α-tests

Nr	n	m	v	α	#1	#2	#3	avg
1	1–50	1–50	1–10	0–1	1	2	3	2.00
2	1–50	1–50	1–100	0–1	5	6	13	8.00
3	1–50	1–50	1–1000	0–1	10	8	11	9.67
4	1–50	1–50	1–1000	0–0.2	16	20	32	22.67
5	1–50	1–50	1–1000	0.2–0.4	23	8	12	14.33
6	1–50	1–50	1–1000	0.4–0.6	7	6	5	6.00
7	1–50	1–50	1–1000	0.6–0.8	3	5	3	3.67
8	1–50	1–50	1–1000	0.8–1	1	0	0	0.33
9	50	1–20	1–1000	0–0.2	19	11	22	17.33
10	50	20–40	1–1000	0–0.2	26	24	24	24.67
11	50	40–60	1–1000	0–0.2	21	18	23	20.67
12	50	60–80	1–1000	0–0.2	13	18	22	17.67
13	1–20	30	1–1000	0–0.2	27	26	39	30.67
14	20–40	30	1–1000	0–0.2	34	37	35	35.33
15	40–60	30	1–1000	0–0.2	22	26	23	23.67
16	60–80	30	1–1000	0–0.2	19	14	14	15.67
17	10	10	1–1000	0.1	36	42	50	42.67
18	10	10	1–1000	0.2	33	53	46	44.00
19	10	10	1–1000	0.3	43	25	45	37.67
20	10	10	1–1000	0.4	30	18	19	22.33
21	10	10	1–1000	0.5	10	10	13	11.00
22	10	10	1–1000	0.6	12	13	7	10.67
23	10	10	1–1000	0.7	3	13	6	7.33
24	10	10	1–1000	0.8	5	2	6	4.33
25	10	10	1–1000	0.9	0	0	0	0

17–25 the parameter α increases. The results of experiments show that the value of $\#i$ can change from 0 to 53. It means that the percentage of pairs for which first modification of greedy algorithm is better than the greedy algorithm with one threshold can change from 0% to 5.3%.

Third Group of Experiments. Third group of experiments is connected with investigation of quality of lower bound $R_{\min}(\alpha) \geq \rho_R(\alpha, \alpha)$.

We choose natural n, m, v and real α, $0 \leq \alpha < 1$. For each chosen tuple (n, m, v, α) we generate randomly 30 pairs (T, w) such that T contains n rows and m conditional attributes, and w has values from the set $\{1, ..., v\}$. After that we find values of $R^{\alpha}_{\text{greedy}}(\alpha, T, w)$ and $\rho_R(\alpha, \alpha, T, w)$ for each of generated 30 pairs. Note that $\rho_R(\alpha, \alpha, T, w) \leq R_{\min}(\alpha, T, w) \leq R^{\alpha}_{\text{greedy}}(\alpha, T, w)$. Finally, we find mean values of $R^{\alpha}_{\text{greedy}}(\alpha, T, w)$ and $\rho_R(\alpha, \alpha, T, w)$ for generated 30 pairs.

Results of experiments can be found in Figs. 1 and 2. In these figures mean values of $\rho_R(\alpha, \alpha, T, w)$ are called "average lower bound" and mean values of $R^{\alpha}_{\text{greedy}}(\alpha, T, w)$ are called "average upper bound".

In Fig. 1 (left-hand side) one can see the case when $n \in \{1000, 2000, \ldots, 5000\}$, $m = 30$, $v = 1000$ and $\alpha = 0.01$.

In Fig. 1 (right-hand side) one can see the case when $n = 1000$, $m \in \{10, 20, \ldots, 100\}$, $v = 1000$ and $\alpha = 0.01$.

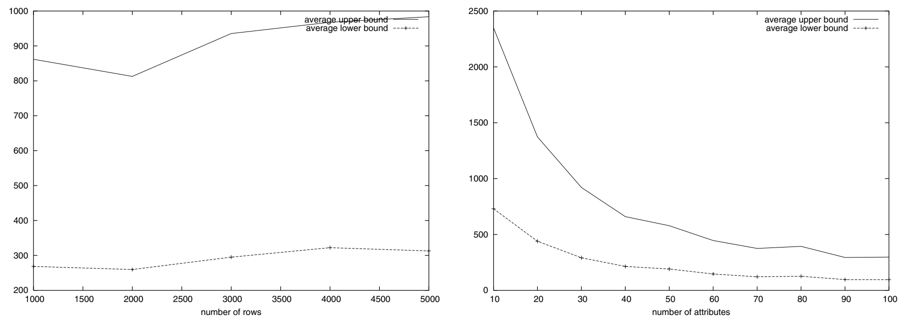

Fig. 1. Results of third group of experiments with α-tests (n and m are changing)

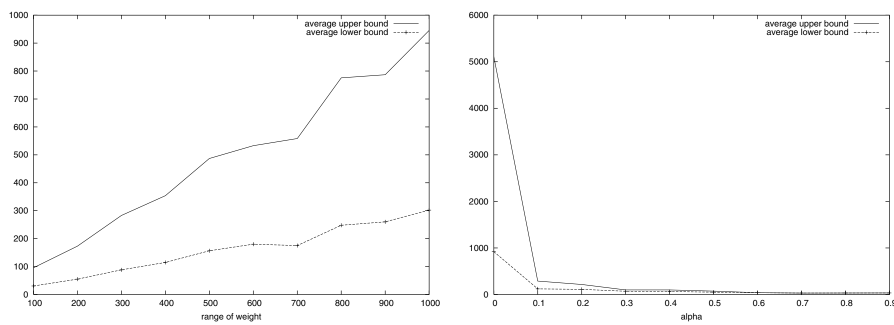

Fig. 2. Results of third group of experiments with α-tests (v and α are changing)

In Fig. 2 (left-hand side) one can see the case when $n = 1000$, $m = 30$, $v \in \{100, 200, \ldots, 1000\}$ and $\alpha = 0.01$.

In Fig. 2 (right-hand side) one can see the case when $n = 1000$, $m = 30$, $v = 1000$ and $\alpha \in \{0.0, 0.1, \ldots, 0.9\}$.

Results of experiments show that the considered lower bound is nontrivial and can be useful in investigations.

4 Partial Decision Rules

In this section we omit reasoning on relationships between partial covers and partial decision rules including reductions of one problem to another (description of such reductions can be found in [9]) and two propositions similar to Propositions 3 and 4.

4.1 Main Notions

Let T be a table with n rows labeled by nonnegative integers (decisions) and m columns labeled by attributes (names of attributes) $f_1, \dots, f_m$. This table is filled by nonnegative integers (values of attributes). The table T is called a *decision table*. Let w be a *weight function* for T which corresponds to each attribute f_i a natural number $w(f_i)$. Let $r = (b_1, \dots, b_m)$ be a row of T labeled by a decision d.

Let us denote by $U(T, r)$ the set of rows from T which are different from r and are labeled by decisions different from d. We will say that an attribute f_i *separates* rows r and $r' \in U(T, r)$ if rows r and r' have different numbers at the intersection with the column f_i. For $i = 1, \dots, m$ denote by $U(T, r, f_i)$ the set of rows from $U(T, r)$ which attribute f_i separates from the row r.

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

$$f_{i_1} = b_{i_1} \wedge \ldots \wedge f_{i_t} = b_{i_t} \rightarrow d \tag{3}$$

is called an α-*decision rule* for T and r if attributes $f_{i_1}, \dots, f_{i_t}$ separate from r at least $(1 - \alpha)|U(T, r)|$ rows from $U(T, r)$. The number $\sum_{j=1}^{t} w(f_{i_j})$ is called the *weight* of the considered decision rule.

If $U(T, r) = \emptyset$ then for any $f_{i_1}, \dots, f_{i_t} \in \{f_1, \dots, f_m\}$ the rule (3) is an α-decision rule for T and r. Also, the rule (3) with empty left-hand side (when $t = 0$) is an α-decision rule for T and r. The weight of this rule is equal to 0.

For example, 0.01-decision rule means that we must separate from r at least 99% of rows from $U(T, r)$. Note that 0-rule is usual (exact) rule. Let us denote by $L_{\min}(\alpha) = L_{\min}(\alpha, T, r, w)$ the minimal weight of α-decision rule for T and r.

Let α, γ be real numbers such that $0 \le \gamma \le \alpha < 0$. Let us describe a *greedy algorithm with thresholds* α *and* γ which constructs an α-decision rule for given T, r and weight function w. Let $r = (b_1, \dots, b_m)$, and r be labeled by the decision d.

The right-hand side of constructed α-decision rule is equal to d. If $U(T, r) = \emptyset$ then the left-hand side of constructed α-decision rule is empty. Let $U(T, r) \neq \emptyset$. Let us denote $M = \lceil |U(T, r)|(1 - \alpha) \rceil$ and $N = \lceil |U(T, r)|(1 - \gamma) \rceil$. Let we make $i \ge 0$ steps and construct a decision rule R with i conditions (if $i = 0$ then the left-hand side of R is empty). Let us describe the step number $i + 1$.

Let us denote by D the set of rows from $U(T, r)$ separated from r by attributes belonging to R (if $i = 0$ then $D = \emptyset$). If $|D| \ge M$ then we finish the work of the algorithm, and R is the constructed α-decision rule. Let $|D| < M$. Then we choose an attribute f_j with minimal number j for which $U(T, r, f_j) \setminus D \neq \emptyset$ and the value

$$\frac{w(f_j)}{\min\{|U(T, r, f_j) \setminus D|, N - |D|\}}$$

is minimal. Add the condition $f_j = b_j$ to R. Pass to the step number $i + 2$.

Let us denote by $L^{\gamma}_{\text{greedy}}(\alpha) = L^{\gamma}_{\text{greedy}}(\alpha, T, r, w)$ the weight of α-decision rule constructed by the considered algorithm for given table T, row r and weight function w.

4.2 On Precision of Greedy Algorithm with Thresholds α and α

The following two statements are simple corollaries of results of Slavík (see Theorems 4 and 5).

Theorem 23. *Let T be a decision table, r be a row of T, $U(T,r) \neq \emptyset$, w be a weight function for T, and α be a real number such that $0 \leq \alpha < 1$. Then $L^{\alpha}_{\mathrm{greedy}}(\alpha) \leq L_{\min}(\alpha) H\left(\lceil (1-\alpha)|U(T,r)| \rceil\right)$.*

Theorem 24. *Let T be a decision table with m columns labeled by attributes $f_1, \ldots, f_m$, r be a row of T, $U(T,r) \neq \emptyset$, w be a weight function for T, $\alpha \in \mathrm{I\!R}$ and $0 \leq \alpha < 1$. Then $L^{\alpha}_{\mathrm{greedy}}(\alpha) \leq L_{\min}(\alpha) H\left(\max_{i \in \{1,\ldots,m\}} |U(T,r,f_i)|\right)$.*

4.3 On Polynomial Approximate Algorithms

In this subsection we consider three theorems which follow immediately from Theorems 39–41 [9].

Let $0 \leq \alpha < 1$. Let us consider the following problem: for given decision table T, row r of T and weight function w for T it is required to find an α-decision rule for T and r with minimal weight.

Theorem 25. *Let $0 \leq \alpha < 1$. Then the problem of construction of α-decision rule with minimal weight is NP-hard.*

So we must consider polynomial approximate algorithms for minimization of α-decision rule weight.

Theorem 26. *Let $\alpha \in \mathrm{I\!R}$ and $0 \leq \alpha < 1$. If $NP \not\subseteq DTIME(n^{O(\log \log n)})$ then for any ε, $0 < \varepsilon < 1$, there is no polynomial algorithm that for given decision table T, row r of T with $U(T,r) \neq \emptyset$ and weight function w for T constructs α-decision rule for T and r which weight is at most $(1-\varepsilon) L_{\min}(\alpha, T, r, w) \ln |U(T,r)|$.*

Theorem 27. *Let α be a real number such that $0 \leq \alpha < 1$. If $P \neq NP$ then there exists $\delta > 0$ such that there is no polynomial algorithm that for given decision table T, row r of T with $U(T,r) \neq \emptyset$ and weight function w for T constructs α-decision rule for T and r which weight is at most $\delta L_{\min}(\alpha, T, r, w) \ln |U(T,r)|$.*

From Theorem 23 it follows that $L^{\alpha}_{\mathrm{greedy}}(\alpha) \leq L_{\min}(\alpha)(1 + \ln |U(T,r)|)$. From this inequality and from Theorem 26 it follows that under the assumption $NP \not\subseteq DTIME(n^{O(\log \log n)})$ greedy algorithm with two thresholds α and α is close to best polynomial approximate algorithms for minimization of partial decision rule weight. From the considered inequality and from Theorem 27 it follows that under the assumption $P \neq NP$ greedy algorithm with two thresholds α and α is not far from best polynomial approximate algorithms for minimization of partial decision rule weight.

However we can try to improve the results of the work of greedy algorithm with two thresholds α and α for some part of decision tables.

4.4 Two Modifications of Greedy Algorithm

First, we consider binary diagnostic decision tables and prove that under some assumptions on the number of attributes and rows for the most part of tables for each row there exist weight function w and numbers α, γ such that the weight of α-decision rule constructed by greedy algorithm with thresholds α and γ is less than the weight of α-decision rule constructed by greedy algorithm with thresholds α and α.

Binary means that the table is filled by numbers from the set $\{0,1\}$ (all attributes have values from $\{0,1\}$). *Diagnostic* means that rows of the table are labeled by pairwise different numbers (decisions). Let T be a binary diagnostic decision table with m columns labeled by attributes $f_1, \ldots, f_m$ and with n rows. We will assume that rows of T with numbers $1, \ldots, n$ are labeled by decisions $1, \ldots, n$ respectively. Therefore the number of considered tables is equal to 2^{mn}. Decision table will be called *simple* if it has no equal rows.

Theorem 28. *Let us consider binary diagnostic decision tables with m columns labeled by attributes $f_1, \ldots, f_m$ and $n \geq 5$ rows labeled by decisions $1, \ldots, n$. The fraction of decision tables T for each of which for each row r of T there exist a weight function w and numbers α, γ such that $0 \leq \gamma < \alpha < 1$ and $L^{\gamma}_{\text{greedy}}(\alpha, T, r, w) < L^{\alpha}_{\text{greedy}}(\alpha, T, r, w)$ is at least*

$$1 - \frac{n3^m}{(n-1)^{\frac{m}{2}-1}} - \frac{n^2}{2^m} \ .$$

Proof. Let T be a decision table and r be a row of T with number $s \in \{1, \ldots, n\}$.

We will say that a decision table T is 1-uniform relatively r if there exists natural p such that for any attribute f_i of T if $|U(T, r, f_i)| > 0$ then $|U(T, r, f_i)| \in \{p, p+1\}$. Using reasoning similar to the proof of Theorem 9 one can show that if T is not 1-uniform relatively r then there exist a weight function w and numbers α, γ such that $0 \leq \gamma < \alpha < 1$ and $L^{\gamma}_{\text{greedy}}(\alpha, T, r, w) < L^{\alpha}_{\text{greedy}}(\alpha, T, r, w)$.

We evaluate the number of decision tables which are not 1-uniform relatively each row. Let $(\delta_1, \ldots, \delta_m) \in \{0,1\}^m$. First, we evaluate the number of simple decision tables for which $r = (\delta_1, \ldots, \delta_m)$ and which are 1-uniform relatively r. Let us consider such a decision table T. It is clear that there exists $p \in \{1, \ldots, n-2\}$ such that for $i = 1, \ldots, m$ the column f_i contains exactly 0 or p or $p+1$ numbers $\neg\delta_i$. Therefore the number of considered decision tables is at most $\sum_{p=1}^{n-2} \left(C^p_{n-1} + C^{p+1}_{n-1} + 1 \right)^m$. Using Lemma 2 we conclude that this number is at most $(n-2)\left(3C^{\lfloor (n-1)/2 \rfloor}_{n-1}\right)^m < (n-1)\left(\frac{3 \cdot 2^{n-1}}{\sqrt{n-1}}\right)^m = \frac{2^{mn-m}3^m}{(n-1)^{\frac{m}{2}-1}}$. There are 2^m variants for the choice of the tuple $(\delta_1, \ldots, \delta_m)$ and n variants for the choice of the number s of row r. Therefore the number of simple decision tables which are 1-uniform relatively at least one row is at most $n2^m \frac{2^{mn-m}3^m}{(n-1)^{\frac{m}{2}-1}} =$

$\frac{n2^{mn}3^m}{(n-1)^{\frac{m}{2}-1}}$. The number of tables which are not simple is at most $n^2 2^{mn-m}$. Hence the number of tables which are not 1-uniform for each row is at least $2^{mn} - \frac{n2^{mn}3^m}{(n-1)^{\frac{m}{2}-1}} - n^2 2^{mn-m}$. Thus, the fraction, considered in the statement of the theorem, is at least $1 - \frac{n3^m}{(n-1)^{\frac{m}{2}-1}} - \frac{n^2}{2^m}$. □

So if $m \geq 6$ and n, $\frac{2^m}{n^2}$ are large enough then for the most part of binary diagnostic decision tables for each row there exist weight function w and numbers α, γ such that the weight of α-decision rule constructed by greedy algorithm with thresholds α and γ is less than the weight of α-decision rule constructed by greedy algorithm with thresholds α and α.

The obtained results show that the greedy algorithm with two thresholds α and γ is of some interest. Now we consider two polynomial modifications of greedy algorithm which allow to use advantages of greedy algorithm with two thresholds α and γ.

Let T be a decision table with m columns labeled by attributes $f_1, \dots, f_m$, $r = (b_1, \dots, b_m)$ be a row of T labeled by decision d, $U(T,r) \neq \emptyset$, w be a weight function for T and α be a real number such that $0 \leq \alpha < 1$.

1. It is impossible to consider effectively all γ such that $0 \leq \gamma \leq \alpha$. Instead of this we can consider all natural N such that $M \leq N \leq |U(T,r)|$ where $M = \lceil |U(T,r)|(1-\alpha) \rceil$ (see the description of greedy algorithm with two thresholds). For each $N \in \{M, \dots, |U(T,r)|\}$ we apply greedy algorithm with parameters M and N to T, r and w and after that choose an α-decision rule with minimal weight among constructed α-decision rules.
2. There exists also an another way to construct an α-decision rule which is not worse than the one obtained under consideration of all N such that $M \leq N \leq |U(T,r)|$. Let us apply greedy algorithm with thresholds α and α to T, r and w. Let the algorithm choose sequentially attributes $f_{j_1}, \dots, f_{j_t}$. For each $i \in \{0, \dots, t-1\}$ we find (if it is possible) an attribute f_{l_i} of T with minimal weight $w(f_{l_i})$ such that the rule $f_{j_1} = b_{j_1} \wedge \dots \wedge f_{j_i} = b_{j_i} \wedge f_{l_i} = b_{l_i} \rightarrow d$ is an α-decision rule for T and r (if $i = 0$ then it will be the rule $f_{l_0} = b_{l_0} \rightarrow d$). After that among constructed α-decision rules $f_{j_1} = b_{j_1} \wedge \dots \wedge f_{j_t} = b_{j_t} \rightarrow d$, ..., $f_{j_1} = b_{j_1} \wedge \dots \wedge f_{j_i} = b_{j_i} \wedge f_{l_i} = b_{l_i} \rightarrow d$, ... we choose an α-decision rule with minimal weight. From Proposition 7 it follows that the constructed α-decision rule is not worse than the one constructed under consideration of all γ, $0 \leq \gamma \leq \alpha$, or (which is the same) all N, $M \leq N \leq |U(T,r)|$.

Using Propositions 1 one can prove the following statement.

Proposition 7. *Let T be a decision table, r be a row of T, $U(T,r) \neq \emptyset$, w be a weight function for T and α, γ be real numbers such that $0 \leq \gamma < \alpha < 1$. Let the greedy algorithm with two thresholds α and α, which is applied to T, r and w, choose sequentially attributes $f_{g_1}, \dots, f_{g_t}$. Let the greedy algorithm with two thresholds α and γ, which is applied to T, r and w, choose sequentially attributes $f_{l_1}, \dots, f_{l_k}$. Then either $k = t$ and $(l_1, \dots, l_k) = (g_1, \dots, g_t)$ or $k \leq t$, $(l_1, \dots, l_{k-1}) = (g_1, \dots, g_{k-1})$ and $l_k \neq g_k$.*

4.5 Bounds on $L_{\min}(\alpha)$ and $L^{\gamma}_{\text{greedy}}(\alpha)$

First, we fix some information about the work of greedy algorithm with two thresholds and find the best lower bound on the value $L_{\min}(\alpha)$ depending on this information.

Let T be a decision table, r be a row of T such that $U(T, r) \neq \emptyset$, w be a weight function for T, and α, γ be real numbers such that $0 \leq \gamma \leq \alpha < 1$. Let us apply the greedy algorithm with thresholds α and γ to the decision table T, row r and the weight function w. Let during the construction of α-decision rule the greedy algorithm choose sequentially attributes $f_{g_1}, \ldots, f_{g_t}$.

Let us denote $U(T, r, f_{g_0}) = \emptyset$ and $\delta_0 = 0$. For $i = 1, \ldots, t$ denote $\delta_i = |U(T, r, f_{g_i}) \setminus (U(T, r, f_{g_0}) \cup \ldots \cup U(T, r, f_{g_{i-1}}))|$ and $w_i = w(f_{g_i})$. As information on the greedy algorithm work we will use numbers $M_L = M_L(\alpha, \gamma, T, r, w) = \lceil |U(T, r)|(1 - \alpha)\rceil$, $N_L = N_L(\alpha, \gamma, T, r, w) = \lceil |U(T, r)|(1 - \gamma)\rceil$ and tuples $\Delta_L = \Delta_L(\alpha, \gamma, T, r, w) = (\delta_1, \ldots, \delta_t)$, $W_L = W_L(\alpha, \gamma, T, r, w) = (w_1, \ldots, w_t)$.

For $i = 0, \ldots, t - 1$ denote

$$\rho_i = \left\lceil \frac{w_{i+1}(M_L - (\delta_0 + \ldots + \delta_i))}{\min\{\delta_{i+1}, N_L - (\delta_0 + \ldots + \delta_i)\}} \right\rceil .$$

Let us define parameter $\rho_L(\alpha, \gamma) = \rho_L(\alpha, \gamma, T, r, w)$ as follows:

$$\rho_L(\alpha, \gamma) = \max\{\rho_i : i = 0, \ldots, t - 1\} .$$

We will show that $\rho_L(\alpha, \gamma)$ is the best lower bound on $L_{\min}(\alpha)$ depending on M_L, N_L, Δ_L and W_L. Using Theorem 11 one can prove the following statement.

Theorem 29. *For any decision table T, any row r of T with $U(T, r) \neq \emptyset$, any weight function w for T, and any real numbers α, γ, $0 \leq \gamma \leq \alpha < 1$, the inequality $L_{\min}(\alpha, T, r, w) \geq \rho_L(\alpha, \gamma, T, r, w)$ holds, and there exist a decision table T', a row r' of T' and a weight function w' for T' such that*

$$M_L(\alpha, \gamma, T', r', w') = M_L(\alpha, \gamma, T, r, w),\ N_L(\alpha, \gamma, T', r', w') = N_L(\alpha, \gamma, T, r, w) ,$$
$$\Delta_L(\alpha, \gamma, T', r', w') = \Delta_L(\alpha, \gamma, T, r, w),\ W_L(\alpha, \gamma, T', r', w') = W_L(\alpha, \gamma, T, r, w) ,$$
$$\rho_L(\alpha, \gamma, T', r', w') = \rho_L(\alpha, \gamma, T, r, w),\ L_{\min}(\alpha, T', r', w') = \rho_L(\alpha, \gamma, T', r', w') .$$

Let us consider a property of the parameter $\rho_L(\alpha, \gamma)$ which is important for practical use of the bound from Theorem 29. Using Proposition 2 one can prove the following statement.

Proposition 8. *Let T be a decision table, r be a row of T with $U(T, r) \neq \emptyset$, w be a weight function for T, and α, γ be real numbers such that $0 \leq \gamma \leq \alpha < 1$. Then $\rho_L(\alpha, \alpha, T, r, w) \geq \rho_L(\alpha, \gamma, T, r, w)$.*

Now we study some properties of parameter $\rho_L(\alpha, \gamma)$ and obtain two upper bounds on the value $L^{\gamma}_{\text{greedy}}(\alpha)$ which do not depend directly on cardinality of the set $U(T, r)$ and cardinalities of subsets $U(T, r, f_i)$.

Using Theorem 12 one can prove the following statement.

Theorem 30. *Let T be a decision table, r be a row of T with $U(T,r) \neq \emptyset$, w be a weight function for T, $\alpha, \gamma \in \mathbb{R}$ and $0 \leq \gamma < \alpha < 1$. Then $L^{\gamma}_{\text{greedy}}(\alpha, T, r, w) < \rho_L(\gamma, \gamma, T, r, w) \left(\ln\left(\frac{1-\gamma}{\alpha-\gamma}\right) + 1\right)$.*

Corollary 9. *Let $\varepsilon \in \mathbb{R}$ and $0 < \varepsilon < 1$. Then for any α, $\varepsilon \leq \alpha < 1$, the inequalities $\rho_L(\alpha, \alpha) \leq L_{\min}(\alpha) \leq L^{\alpha-\varepsilon}_{\text{greedy}}(\alpha) < \rho_L(\alpha - \varepsilon, \alpha - \varepsilon) \left(\ln \frac{1}{\varepsilon} + 1\right)$ hold.*

For example, $\left(\ln \frac{1}{0.01} + 1\right) < 5.61$ and $\left(\ln \frac{1}{0.1} + 1\right) < 3.31$. The obtained results show that the lower bound $L_{\min}(\alpha) \geq \rho_L(\alpha, \alpha)$ is nontrivial.

Using Theorem 13 one can prove the following statement.

Theorem 31. *Let T be a decision table, r be a row of T with $U(T,r) \neq \emptyset$, w be a weight function for T, $\alpha, \gamma \in \mathbb{R}$ and $0 \leq \gamma < \alpha < 1$. Then $L^{\gamma}_{\text{greedy}}(\alpha, T, r, w) < L_{\min}(\gamma, T, r, w) \left(\ln\left(\frac{1-\gamma}{\alpha-\gamma}\right) + 1\right)$.*

Corollary 10. *$L^{0}_{\text{greedy}}(0.001) < 7.91 L_{\min}(0)$, $L^{0.001}_{\text{greedy}}(0.01) < 5.71 L_{\min}(0.001)$, $L^{0.1}_{\text{greedy}}(0.2) < 3.20 L_{\min}(0.1)$, $L^{0.3}_{\text{greedy}}(0.5) < 2.26 L_{\min}(0.3)$.*

Corollary 11. *Let $0 < \alpha < 1$. Then $L^{0}_{\text{greedy}}(\alpha) < L_{\min}(0) \left(\ln \frac{1}{\alpha} + 1\right)$.*

Corollary 12. *Let ε be a real number, and $0 < \varepsilon < 1$. Then for any α such that $\varepsilon \leq \alpha < 1$ the inequalities $L_{\min}(\alpha) \leq L^{\alpha-\varepsilon}_{\text{greedy}}(\alpha) < L_{\min}(\alpha - \varepsilon) \left(\ln \frac{1}{\varepsilon} + 1\right)$ hold.*

4.6 Results of Experiments for α-Decision Rules

In this subsection we will consider only binary decision tables T with binary decision attributes.

First Group of Experiments. First group of experiments is connected with study of quality of greedy algorithm with one threshold (where $\gamma = \alpha$ or, which is the same, $N = M$), and comparison of quality of greedy algorithm with one threshold and first modification of greedy algorithm (where for each $N \in \{M, \ldots, |U(T,r)|\}$ we apply greedy algorithm with parameters M and N to decision table, row and weight function and after that choose an α-decision rule with minimal weight among constructed α-decision rules).

We generate randomly 1000 decision tables T, rows r and weight functions w such that T contains 40 rows and 10 conditional attributes $f_1, \ldots, f_{10}$, r is the first row of T, and $1 \leq w(f_i) \leq 1000$ for $i = 1, \ldots, 10$.

For each $\alpha \in \{0.1, \ldots, 0.9\}$ we find the number of triples (T, r, w) for which greedy algorithm with one threshold constructs an α-decision rule with minimal weight (an optimal α-decision rule), i.e. $L^{\alpha}_{\text{greedy}}(\alpha, T, r, w) = L_{\min}(\alpha, T, r, w)$. This number is contained in the row of Table 3 labeled by "Opt".

We find the number of triples (T, r, w) for which first modification of greedy algorithm constructs an α-decision rule which weight is less than the weight of α-decision rule constructed by greedy algorithm with one threshold, i.e. there exists γ such that $0 \leq \gamma < \alpha$ and $L^{\gamma}_{\text{greedy}}(\alpha, T, r, w) < L^{\alpha}_{\text{greedy}}(\alpha, T, r, w)$. This number is contained in the row of Table 3 labeled by "Impr".

Also we find the number of triples (T, r, w) for which first modification of greedy algorithm constructs an optimal α-decision rule which weight is less than the weight of α-decision rule constructed by greedy algorithm with one threshold, i.e. there exists γ such that $0 \leq \gamma < \alpha$ and $L^{\gamma}_{\text{greedy}}(\alpha, T, r, w) = L_{\min}(\alpha, T, r, w) < L^{\alpha}_{\text{greedy}}(\alpha, T, r, w)$. This number is contained in the row of Table 3 labeled by "Opt+".

Table 3. Results of first group of experiments with α-decision rules

α	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Opt	434	559	672	800	751	733	866	966	998	1000
Impr	0	31	51	36	22	27	30	17	1	0
Opt+	0	16	35	28	17	26	25	13	1	0

The obtained results show that the percentage of triples for which greedy algorithm with one threshold finds an optimal α-decision rule grows almost monotonically (with local minimum near to 0.4–0.5) from 43.4% up to 100%. The percentage of problems for which first modification of greedy algorithm can improve the result of the work of greedy algorithm with one threshold is less than 6%. However, sometimes (for example, if $\alpha = 0.3$, $\alpha = 0.6$ or $\alpha = 0.7$) the considered improvement is noticeable.

Second Group of Experiments. Second group of experiments is connected with comparison of quality of greedy algorithm with one threshold and first modification of greedy algorithm.

We make 25 experiments (row "Nr" in Table 4 contains the number of experiment). Each experiment includes the work with three randomly generated families of triples (T, r, w) (1000 triples in each family) such that T contains n rows and m conditional attributes, r is the first row of T, and w has values from the set $\{1, \ldots, v\}$.

If the column "n" contains one number, for example "40", it means that $n = 40$. If this row contains two numbers, for example "30–120", it means that for each of 1000 triples we choose the number n randomly from the set $\{30, \ldots, 120\}$. The same situation is for the column "m".

If the column "α" contains one number, for example "0.1", it means that $\alpha = 0.1$. If this column contains two numbers, for example "0.2–0.4", it means that we choose randomly the value of α such that $0.2 \leq \alpha \leq 0.4$.

For each of the considered triples (T, r, w) and number α we apply greedy algorithm with one threshold and first modification of greedy algorithm. Column "#i", $i = 1, 2, 3$, contains the number of triples (T, r, w) from the family number i for each of which the weight of α-decision rule, constructed by first modification of greedy algorithm, is less than the weight of α-decision rule constructed by

Table 4. Results of second group of experiments with α-decision rules

Nr	n	m	v	α	#1	#2	#3	avg
1	1–100	1–100	1–10	0–1	4	2	4	3.33
2	1–100	1–100	1–100	0–1	7	14	13	11.33
3	1–100	1–100	1–1000	0–1	19	13	15	15.67
4	1–100	1–100	1–1000	0–0.2	20	39	22	27.00
5	1–100	1–100	1–1000	0.2–0.4	28	29	28	28.33
6	1–100	1–100	1–1000	0.4–0.6	22	23	34	26.33
7	1–100	1–100	1–1000	0.6–0.8	7	6	4	5.67
8	1–100	1–100	1–1000	0.8–1	0	1	0	0.33
9	100	1–30	1–1000	0–0.2	35	38	28	33.67
10	100	30–60	1–1000	0–0.2	47	43	31	40.33
11	100	60–90	1–1000	0–0.2	45	51	36	44.00
12	100	90–120	1–1000	0–0.2	37	40	55	44.00
13	1–30	30	1–1000	0–0.2	11	8	9	9.33
14	30–60	30	1–1000	0–0.2	20	22	35	25.67
15	60–90	30	1–1000	0–0.2	30	33	34	32.33
16	90–120	30	1–1000	0–0.2	40	48	38	42.00
17	40	10	1–1000	0.1	31	39	34	34.67
18	40	10	1–1000	0.2	37	39	47	41.00
19	40	10	1–1000	0.3	35	30	37	34.00
20	40	10	1–1000	0.4	27	20	27	24.67
21	40	10	1–1000	0.5	32	32	36	33.33
22	40	10	1–1000	0.6	28	26	24	26.00
23	40	10	1–1000	0.7	10	12	10	10.67
24	40	10	1–1000	0.8	0	2	0	0.67
25	40	10	1–1000	0.9	0	0	0	0

greedy algorithm with one threshold. In other words, in column "$\#i$" we have the number of triples (T, r, w) from the family number i such that there exists γ for which $0 \le \gamma < \alpha$ and $L^{\gamma}_{\text{greedy}}(\alpha, T, r, w) < L^{\alpha}_{\text{greedy}}(\alpha, T, r, w)$. The column "avg" contains the number $\frac{\#1+\#2+\#3}{3}$.

In experiments 1–3 we consider the case where the parameter v increases. In experiments 4–8 the parameter α increases. In experiments 9–12 the parameter m increases. In experiments 13–16 the parameter n increases. In experiments 17–25 the parameter α increases. The results of experiments show that the value of $\#i$ can change from 0 to 55. It means that the percentage of triples for which the first modification of greedy algorithm is better than the greedy algorithm with one threshold can change from 0% to 5.5%.

Third Group of Experiments. Third group of experiments is connected with investigation of quality of lower bound $L_{\min}(\alpha) \ge \rho_L(\alpha, \alpha)$.

We choose natural n, m, v and real α, $0 \le \alpha < 1$. For each chosen tuple (n, m, v, α) we generate randomly 30 triples (T, r, w) such that T contains

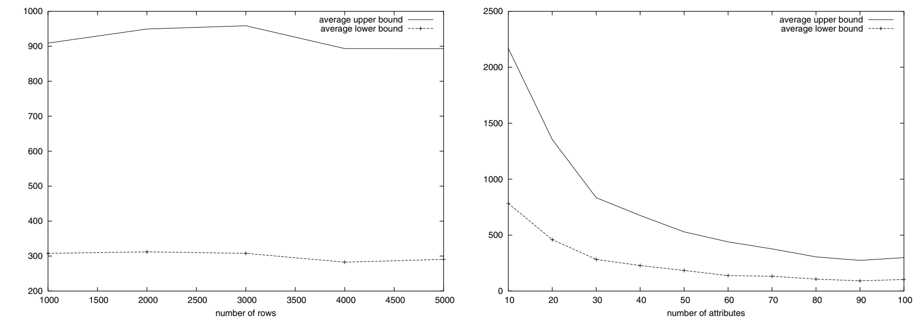

Fig. 3. Results of third group of experiments with rules (n and m are changing)

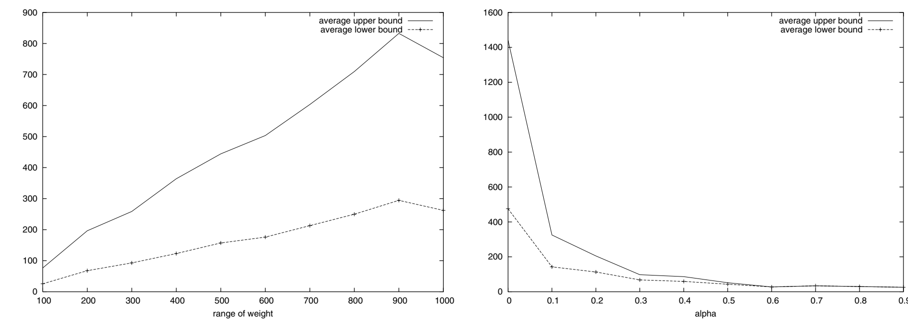

Fig. 4. Results of third group of experiments with rules (v and α are changing)

n rows and m conditional attributes, r is the first row of T, and w has values from the set $\{1, ..., v\}$. After that we find values of $L^{\alpha}_{\text{greedy}}(\alpha, T, r, w)$ and $\rho_L(\alpha, \alpha, T, r, w)$ for each of generated 30 triples. Note that $\rho_L(\alpha, \alpha, T, r, w) \leq L_{\min}(\alpha, T, r, w) \leq L^{\alpha}_{\text{greedy}}(\alpha, T, r, w)$. Finally, for generated 30 triples we find mean values of $L^{\alpha}_{\text{greedy}}(\alpha, T, r, w)$ and $\rho_L(\alpha, \alpha, T, r, w)$.

Results of experiments can be found in Figs. 3 and 4. In these figures mean values of $\rho_L(\alpha, \alpha, T, r, w)$ are called "average lower bound" and mean values of $L^{\alpha}_{\text{greedy}}(\alpha, T, r, w)$ are called "average upper bound".

In Fig. 3 (left-hand side) one can see the case when $n \in \{1000, 2000, \ldots, 5000\}$, $m = 30$, $v = 1000$ and $\alpha = 0.01$.

In Fig. 3 (right-hand side) one can see the case when $n = 1000$, $m \in \{10, 20, \ldots, 100\}$, $v = 1000$ and $\alpha = 0.01$.

In Fig. 4 (left-hand side) one can see the case when $n = 1000$, $m = 30$, $v \in \{100, 200, \ldots, 1000\}$ and $\alpha = 0.01$.

In Fig. 4 (right-hand side) one can see the case when $n = 1000$, $m = 30$, $v = 1000$ and $\alpha \in \{0.0, 0.1, \ldots, 0.9\}$.

Results of experiments show that the considered lower bound is nontrivial and can be useful in investigations.

5 Conclusions

The paper is devoted (mainly) to theoretical and experimental analysis of greedy algorithms with weights and their modifications for partial cover, reduct and decision rule construction. Obtained results will further to more wide use of such algorithms in rough set theory and its applications.

In the further investigations we are planning to generalize the obtained results to the case of decision tables which can contain missing values, continuous attributes, and discrete attributes with large number of values.

Acknowledgments

The authors are greatly indebted to Andrzej Skowron for stimulating discussions and to anonymous reviewers for helpful remarks and suggestions.

References

1. Cheriyan, J., Ravi, R.: Lecture Notes on Approximation Algorithms for Network Problems (1998) http://www.math.uwaterloo.ca/~jcheriya/lecnotes.html
2. Chvátal, V.: A greedy heuristic for the set-covering problem. *Mathematics of Operations Research* 4 (1979) 233–235.
3. Feige, U.: A threshold of $\ln n$ for approximating set cover (Preliminary version). Proceedings of 28th Annual ACM Symposium on the Theory of Computing (1996) 314–318.
4. Gavrilov G.P., Sapozhenko, A.A.: Problems and Exercises in Discrete Mathematics (third edition). Fizmatlit, Moscow, 2004 (in Russian).
5. Kearns, M.J.: The Computational Complexity of Machine Learning. MIT Press, Cambridge, Massachussetts, 1990.
6. Moshkov, M.Ju.: Greedy algorithm for set cover in context of knowledge discovery problems. In: Proceedings of the International Workshop on Rough Sets in Knowledge Discovery and Soft Computing (ETAPS 2003 Satellite Event). Warsaw, Poland. *Electronic Notes in Theoretical Computer Science* 82 (2003).
7. Moshkov, M.Ju.: On greedy algorithm for partial cover construction. In: Proceedings of the Fourteenth International Workshop Design and Complexity of Control Systems. Nizhny Novgorod, Russia (2003) 57 (in Russian).
8. Moshkov, M.Ju., Piliszczuk, M., Zielosko, B.: Greedy algorithm for construction of partial covers. In: Proceedings of the Fourteenth International Conference Problems of Theoretical Cybernetics. Penza, Russia (2005) 103 (in Russian).
9. Moshkov, M.Ju., Piliszczuk, M., Zielosko, B.: On partial covers, reducts and decision rules. LNCS Transactions on Rough Sets, Springer-Verlag (submitted).
10. Nguyen, H.S., Ślęzak, D.: Approximate reducts and association rules - correspondence and complexity results. In: Proceedings of the Seventh International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing. Yamaguchi, Japan. Lecture Notes in Artificial Intelligence 1711, Springer-Verlag (1999) 137–145.
11. Pawlak, Z.: Rough Sets – Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, Dordrecht, Boston, London, 1991.

12. Pawlak, Z.: Rough set elements. In: Polkowski, L., Skowron, A. (Eds.), Rough Sets in Knowledge Discovery 1. Methodology and Applications (Studies in Fuzziness and Soft Computing 18). Phisica-Verlag. A Springer-Verlag Company (1998) 10–30.
13. Piliszczuk, M.: On greedy algorithm for partial reduct construction. In: Proceedings of Concurrency, Specification and Programming Workshop 2. Ruciane-Nida, Poland (2005) 400–411.
14. Quafafou, M.: α-RST: a generalization of rough set theory. *Information Sciences* 124 (2000) 301–316.
15. Raz, R., Safra, S.: A sub-constant error-probability low-degree test, and sub-constant error-probability PCP characterization of NP. In: Proceedings of 29th Annual ACM Symposium on the Theory of Computing (1997) 475–484.
16. Skowron, A.: Rough sets in KDD. In: Proceedings of the 16-th World Computer Congress (IFIP'2000). Beijing, China (2000) 1–14.
17. Skowron, A., Rauszer, C.: The discernibility matrices and functions in information systems. In: Slowinski, R. (Ed.), Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory. Kluwer Academic Publishers, Dordrecht, Boston, London (1992) 331–362.
18. Slavík, P.: Approximation algorithms for set cover and related problems. Ph.D. thesis. University of New York at Buffalo (1998).
19. Ślęzak, D.: Approximate reducts in decision tables. In: Proceedings of the Congress Information Processing and Management of Uncertainty in Knowledge-based Systems 3. Granada, Spain (1996) 1159–1164.
20. Ślęzak, D.: Normalized decision functions and measures for inconsistent decision tables analysis. *Fundamenta Informaticae* 44 (2000) 291–319.
21. Ślęzak, D.: Approximate decision reducts. Ph.D. thesis. Warsaw University (2001) (in Polish).
22. Ślęzak, D.: Approximate entropy reducts. *Fundamenta Informaticae* 53 (2002) 365–390.
23. Ślęzak, D., Wróblewski, J.: Order-based genetic algorithms for the search of approximate entropy reducts. In: Proceedings of the International Conference Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing. Chongqing, China. Lecture Notes in Artificial Intelligence 2639, Springer-Verlag (2003) 308–311.
24. Wróblewski, J.: Ensembles of classifiers based on approximate reducts. *Fundamenta Informaticae* 47 (2001) 351–360.
25. Yablonskii, S.V.: Introduction into Discrete Mathematics (fourth edition). Vishaya Shkola, Moscow, 2003 (in Russian).
26. Ziarko, W.: Analysis of uncertain information in the framework of variable precision rough sets. *Foundations of Computing and Decision Sciences* 18 (1993) 381–396.
27. Zielosko, B.: On partial decision rules. In: Proceedings of Concurrency, Specification and Programming Workshop 2. Ruciane-Nida, Poland (2005) 598–609.