

## Classifiers Based on Partial Reducts and Partial Decision Rules

Beata Zielosko<sup>1</sup>, Marcin Piliszczuk<sup>2</sup>, and Arkadiusz Kocjan<sup>1</sup>

<sup>1</sup> Institute of Computer Science, University of Silesia, 39, Będzińska St., Sosnowiec,  
41-200, Poland

<sup>2</sup> ING Bank Śląski S.A., 34, Sokolska St., Katowice, 40-086, Poland

### Abstract

In the paper we study accuracy of classifiers based on (i) decision rules corresponding to partial (approximate) reducts and (ii) partial decision rules constructed directly for rows of decision table. We use greedy algorithms which generate partial reducts and partial decision rules. Based on them we prepare sets of rules, which are used in the classification process. We present results of experiments with twelve decision tables from UCI Machine Learning Repository.

**Keywords:** Partial tests, partial reducts, partial decision rules, greedy algorithm, classifiers.

### 1 Introduction

The paper is devoted to the consideration of accuracy for classifiers based on partial reducts and partial decision rules. The following approaches to decision rule construction are well known: (i) for a given decision table we construct a decision reduct and later for each row we construct a decision rule based on attributes from this reduct (Komorowski et al., 1999), (ii) for rows of a given decision table we construct decision rules directly.

In this paper we study generalizations of these approaches. Instead of exact decision reducts we use partial (approximate) decision reducts, and instead of exact decision rules we use partial (approximate) decision rules.

Last years partial reducts and partial decision rules were studied intensively (Moshkov et al., 2007a,b; Nguyen and Ślęzak, 1999; Piliszczuk, 2005; Ślęzak, 2000, 2002; Wróblewski, 2001; Zielosko and Piliszczuk, 2008; Zielosko, 2005).

The most important reasons to use partial reducts and rules are the following: such reducts and rules depend on the noise in less degree than exact reducts and rules; based on partial reducts and rules we can often construct more compact classifiers.

We try to construct partial reduct with minimal or close to minimal cardinality, and partial decision rules with minimal or close to minimal length. To this end we use greedy algorithms. In Moshkov et al. 2008, we show that under some natural assumptions on the class NP greedy algorithms are close to best approximate

polynomial algorithms for minimization of partial superreduct cardinality and for minimization of partial decision rule length.

We made some experiments with classifiers based on partial decision rules corresponding to partial reducts, and classifiers based on partial decision rules constructed for rows of decision table directly. The results of experiments show that the accuracy of classifiers based on partial reducts and partial decision rules is often better than the accuracy of classifiers based on exact reducts and exact decision rules.

The paper consists of five sections. In Section 2 main notions of partial reducts, tests and decision rules are considered. In Section 3 we describe classifiers based on decision rules corresponding to partial reducts and classifiers based on partial decision rules constructed directly for rows of decision table. In Section 4 results of experiments with real-life decision tables are presented. Section 5 contains short conclusions.

## 2 Main Notions

In this section main notions for partial reducts, tests (Section 2.1), and partial decision rules (Section 2.2) are described.

Let  $T$  be a table with  $n$  rows labeled by decisions and  $m$  columns labeled by attributes (names of attributes)  $a_1, \dots, a_m$ . This table is filled by values of attributes. The table  $T$  is called a *decision table* (Pawlak, 1991). Two rows are called *different* if they have different values of attributes at the intersection with at least one column  $a_i$ .

### 2.1 Partial Tests and Reducts

Let us denote by  $P(T)$  the set of unordered pairs of different rows from  $T$  with different decisions.

We will use the symbol  $\alpha$  to denote a real number such that  $0 \leq \alpha < 1$ . A set of attributes  $Q$  is called an  $\alpha$ -test for  $T$ , if attributes from  $Q$  separate at least  $(1 - \alpha)|P(T)|$  pairs from the set  $P(T)$ .

An  $\alpha$ -test is called an  $\alpha$ -reduct if each proper subset of the considered  $\alpha$ -test is not an  $\alpha$ -test. If  $P(T) = \emptyset$ , then each subset of  $\{a_1, \dots, a_m\}$  is an  $\alpha$ -test, and only empty set is an  $\alpha$ -reduct.

For example, 0.01-test means that we should separate at least 99% pairs of rows from  $P(T)$ . Remark that a 0-reduct is an exact reduct. Note also that each  $\alpha$ -test contains at least one  $\alpha$ -reduct as a subset.

We will denote by  $R_{\min}(\alpha) = R_{\min}(\alpha, T)$  the minimal cardinality of  $\alpha$ -reduct for  $T$ . It is clear that  $R_{\min}(\alpha, T)$  is equal to the minimal cardinality of  $\alpha$ -test for  $T$ .

Let us describe the greedy algorithm with threshold  $\alpha$  (Algorithm 1) which constructs an  $\alpha$ -test for  $T$ .

By  $R_{\text{greedy}}(\alpha) = R_{\text{greedy}}(\alpha, T)$  we denote the cardinality of constructed  $\alpha$ -test for  $T$ .

The following Theorem 1 and Theorem 2 were proved in Moshkov et al. 2008.

---

**Algorithm 1:** Greedy algorithm for partial test construction.
 

---

**Input** : Decision table  $T$  with conditional attributes  $a_1, \dots, a_m$ , and real number  $\alpha$ ,  $0 \leq \alpha < 1$ .

**Output:**  $\alpha$ -test for  $T$ .

$Q \leftarrow \emptyset$ ;

**while**  $Q$  is not an  $\alpha$ -test for  $T$  **do**

select  $a_i \in \{a_1, \dots, a_m\}$  with minimal index  $i$  such that  $a_i$  separates the maximal number of pairs from  $P(T)$  unseparated by attributes from  $Q$ ;

$Q \leftarrow Q \cup \{a_i\}$ ;

**end**

return  $Q$ ;

---

**Theorem 1** Let  $0 \leq \alpha < 1$  and  $P(T) \neq \emptyset$ . Then

$$R_{\text{greedy}}(\alpha) \leq R_{\min}(\alpha)(1 + \ln |P(T)|).$$

**Theorem 2** Let  $\alpha \in \mathbf{R}$  and  $0 \leq \alpha < 1$ . If  $NP \not\subseteq DTIME(n^{O(\log \log n)})$ , then for any  $\varepsilon$ ,  $0 < \varepsilon < 1$ , there is no polynomial algorithm that, for a given decision table  $T$  with  $P(T) \neq \emptyset$ , constructs an  $\alpha$ -test for  $T$  which cardinality is at most  $(1 - \varepsilon)R_{\min}(\alpha, T) \ln |P(T)|$ .

From Theorems 1 and 2 it follows that, under the assumption

$$NP \not\subseteq DTIME(n^{O(\log \log n)}),$$

the greedy algorithm is close to the best polynomial approximate algorithms for partial test minimization.

## 2.2 Partial Decision Rules

Let  $r = (b_1, \dots, b_m)$  be a row from  $T$  labeled by a decision  $d$ . We will denote by  $U(T, r)$  the set of rows from  $T$  which are different (in at least one column  $a_i$ ) from  $r$  and are labeled by decisions different from  $d$ . We will say that an attribute  $a_i$  separates a row  $r' \in U(T, r)$  from the row  $r$ , if the rows  $r$  and  $r'$  have different values at the intersection with the column  $a_i$ .

We will use the symbol  $\alpha$  to denote a real number such that  $0 \leq \alpha < 1$ . A decision rule

$$(a_{i_1} = b_{i_1}) \wedge \dots \wedge (a_{i_t} = b_{i_t}) \rightarrow d$$

is called an  $\alpha$ -decision rule for  $(T, r)$  if attributes  $a_{i_1}, \dots, a_{i_t}$  separate from  $r$  at least  $(1 - \alpha)|U(T, r)|$  rows from  $U(T, r)$ . The number  $t$  is called the length of this decision rule. If  $U(T, r) = \emptyset$  then for any  $a_{i_1}, \dots, a_{i_t} \in \{a_1, \dots, a_m\}$  the considered rule is an  $\alpha$ -decision rule for  $(T, r)$ .

For example, 0.01-decision rule means that we should separate from  $r$  at least 99% of rows from  $U(T, r)$ . Note that 0-decision rule is an exact decision rule.

Let us describe the greedy algorithm with threshold  $\alpha$  (Algorithm 2) which constructs an  $\alpha$ -decision rule for  $(T, r)$ .

Results similar to Theorems 1 and 2 can be obtained also for partial decision rules (Moshkov et al., 2008).

---

**Algorithm 2:** Greedy algorithm for partial decision rule construction.

---

**Input** : Decision table  $T$  with conditional attributes  $a_1, \dots, a_m$ , row  $r = (b_1, \dots, b_m)$  of  $T$  labeled by the decision  $d$  and real number  $\alpha$ ,  $0 \leq \alpha < 1$ .

**Output:**  $\alpha$ -decision rule for  $(T, r)$ .

$Q \leftarrow \emptyset$ ;

**while** attributes from  $Q$  separate from  $r$  less than  $(1 - \alpha)|U(T, r)|$  rows from  $U(T, r)$  **do**

select  $a_i \in \{a_1, \dots, a_m\}$  with minimal index  $i$  such that  $a_i$  separates from  $r$  the maximal number of rows from  $U(T, r)$  unseparated by attributes from  $Q$ ;

$Q \leftarrow Q \cup \{a_i\}$ ;

**end**

return  $\bigwedge_{a_i \in Q} (a_i = b_i) \rightarrow d$ ;

---

### 3 Classification

In this section we present two classifiers: based on decision rules corresponding to partial reducts (Section 3.1) and classifiers based on partial decision rules constructed directly for rows (Section 3.2).

We consider the following problem of classification (prediction): for a given decision table  $T$  and a given new object  $v$  it is required to generate a decision corresponding to  $v$  using only values of conditional attributes from  $T$  on  $v$  (Koronacki and Ćwik, 2005; Hand et al., 2005; Skowron, 2000)

#### 3.1 Classifiers based on decision rules corresponding to partial reduct

For the decision table  $T$  and given  $\alpha$ ,  $0 \leq \alpha < 1$ , we construct an  $\alpha$ -test by Algorithm 1. After that by removing some attributes from this  $\alpha$ -test we obtain an  $\alpha$ -reduct for the table  $T$ . Let this  $\alpha$ -reduct be equal to  $Q = \{a_{i_1}, \dots, a_{i_t}\}$ . Then for each row  $r$  of  $T$  we construct a corresponding decision rule. Let  $r = (b_1, \dots, b_m)$  and  $r$  be labeled by the decision  $d$ . Then the considered rule is equal to

$$(a_{i_1} = b_{i_1}) \wedge \dots \wedge (a_{i_t} = b_{i_t}) \Rightarrow d$$

Let us denote by  $C(T, Q)$  the set of decision rules corresponding to all rows of  $T$ . The set  $C(T, Q)$  can be considered as a classifier which for a given new object  $v$  creates a decision for this object using only values of attributes from  $Q$  on  $v$ . For each possible decision  $d$  we find the number  $N_d(v)$  of rules from  $C(T, Q)$  such that (i) the left hand side of the considered rule is true for  $v$ , and (ii) the right hand side of the rule is equal to  $d$ . If  $N_d(v) > 0$  for at least one decision  $d$  then we choose a decision  $d$  for which  $N_d(v)$  has maximal value. Otherwise we choose some fixed decision  $d_0$ .

### 3.2 Classifiers based on partial decision rules constructed directly for rows of $T$

For every row of the decision table  $T$  and given  $\alpha$ ,  $0 \leq \alpha < 1$ , we construct an  $\alpha$ -decision rule by Algorithm 2. From the obtained set of rules we remove repeating rules.

We denote the obtained set by  $C(T)$ . For each rule from  $C(T)$  we compute the support of this rule. It is the number of rows from the training set such that (i) the left hand side of the rule is true for the considered rule; (ii) the decision attached to the row is equals the decision from the right hand side of the rule.

The set  $C(T)$  can be considered as a classifier which for a given new object  $v$  creates a decision for this object using only values of conditional attributes on  $v$ . For each possible decision  $d$  we find the total sum of support  $M_d(v)$  of rules from  $C(T)$ , such that (i) the left hand side of the considered rule is true for  $v$ , and (ii) the right hand side of the rule is equal to  $d$ . If  $M_d(v) > 0$  for at least one decision  $d$  then we choose a decision  $d$  for which  $M_d(v)$  has maximal value. Otherwise we choose some fixed decision  $d_0$ .

## 4 Experiments

We made experiments for the case of partial reducts and for the case of partial decision rules, with 12 decision tables from UCI Machine Learning Repository (Newman et al., 1998).

To evaluate the accuracy of classifiers we use train-and-test method. We randomly split datasets with proportion 70% for training set and 30% for testing set. For 'monks3' and 'spect' we use existing training and testing sets.

Accuracy of classification is the number of objects properly classified divided by the number of all objects in testing set.

For each table we choose minimal  $\alpha \in \{0.00, 0.01, 0.02, \dots, 0.50\}$  for which the accuracy of constructed classifier is maximal. This value of  $\alpha$  is denoted by  $\alpha_{opt}$ . The results of experiments can be found in Table 1 for the case of partial reducts and in Table 2 for the case of partial decision rules.

Results presented in Table 1 show that the use of partial reducts ( $\alpha$ -reducts with  $\alpha > 0$ ) leads to improvement of accuracy of classification for 8 decision tables.

Based on results presented in Table 2 we can observe that for 5 decision tables partial decision rules ( $\alpha$ -decision rules with  $\alpha > 0$ ) give better results than exact ( $\alpha = 0$ ) decision rules.

We can also notice that classifiers based on partial decision rules presented in Table 2 are often better than classifiers based on partial reducts presented in Table 1. However for 2 decision tables we have the opposite situation.

## 5 Conclusions

Results of experiments show that classifiers based on partial reducts are often better than the classifier based on exact reducts. For the case of classifiers based on partial decision rules we have the similar situation.

TABLE 1: Classifiers based on partial reducts

Table	Accuracy $\alpha = 0$	Accuracy $\alpha = \alpha_{opt}$	$\alpha_{opt}$
balance	0.064	0.668	0.03
car	0.713	0.848	0.01
krkopt	0.099	0.246	0.01
kr-vs-kp	0.734	0.956	0.01
letter-recognition	0.221	0.221	0.00
lymphography	0.644	0.733	0.01
nursery	0.000	0.894	0.01
soybean-small	1.000	1.000	0.00
tic-tac-toe	0.635	0.767	0.04
zoo	0.935	0.935	0.00
spect	0.872	0.872	0.00
monks3	0.861	0.972	0.01

TABLE 2: Classifiers based on partial decision rules

Table	Accuracy $\alpha = 0$	Accuracy $\alpha = \alpha_{opt}$	$\alpha_{opt}$
balance	0.658	0.856	0.18
car	0.890	0.892	0.01
krkopt	0.386	0.386	0.00
kr-vs-kp	0.989	0.989	0.00
letter-recognition	0.710	0.710	0.00
lymphography	0.733	0.822	0.22
nursery	0.974	0.974	0.00
soybean-small	1.000	1.000	0.00
tic-tac-toe	0.931	0.931	0.00
zoo	0.968	0.968	0.00
spect	0.818	0.834	0.12
monks3	0.931	0.963	0.05

In our future work we are planning to study accuracy of such classifiers with  $\alpha \in \{0.0, \dots, 0.3\}$  with step equals 0.001. To evaluate the accuracy we are planning to use cross validation method.

## Acknowledgements

The authors wishes to thank prof. Mikhail Ju. Moshkov for consultations. The research has been partially supported by the grant N N206 400434 “Greedy algorithms for construction of partial decision reducts” from Ministry of Science and Higher Education of the Republic of Poland and by the project “Decision support — new generation systems” of Innovative Economy Operational Programme 2007-2013 (Priority Axis 1. Research and development of new technologies) managed by Ministry of Regional Development of the Republic of Poland.

## References

- David HAND, Hekki MANNILA, Padhraic SMYTH, (2005), Principles of Data Mining, WNT, Warsaw (in Polish).
- Jan KOMOROWSKI, Zdzisław PAWLAK, Lech POLKOWSKI, Andrzej SKOWRON, (1999), Rough sets: A tutorial, in *S.K. Pal and A. Skowron (eds.), Rough-Fuzzy Hybridization: A New Trend in Decision-Making, Springer-Verlag, Singapor*, pp. 3–98.
- Jacek KORONACKI, Jan ĆWIK, (2005), Statistical Learning Systems, WNT, Warsaw (in Polish).
- Mikhail MOSHKOV, Marcin PILISZCZUK, Beata ZIELOSKO, (2007a), On construction of partial reducts and irreducible partial decision rules, *Fundamenta Informaticae 75(1-4)*, pp. 357–374.
- Mikhail MOSHKOV, Marcin PILISZCZUK, Beata ZIELOSKO, (2007b), Universal attribute reduction problem, in *Proceedings of the International Conference Rough Sets and Intelligent Systems Paradigms, LNCS 4585, Springer-Verlag*, pp. 417–426.
- Mikhail MOSHKOV, Marcin PILISZCZUK, Beata ZIELOSKO, (2008), On partial covers, reducts and decision rules, *LNCS Transactions on Rough Sets, Springer-Verlag* (to appear).
- Son H.NGUYEN, Dominik ŚLĘZAK, (1999), 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, LNAI 1711, Springer-Verlag*, pp. 137–145
- David NEWMAN, Seth HETTICH, Christopher MERZ, Cathy BLAKE, (1998), Repository of machine learning databases, <http://www.ics.uci.edu/~mlearn/MLRepository.html>, University of California, Irvine, Department of Information and Computer Sciences.
- Zdzisław PAWLAK (1991), Rough Sets – Theoretical Aspects of Reasoning about Data *Kluwer Academic Publishers, Dordrecht, Boston, London*.
- Marcin PILISZCZUK (2005), On greedy algorithm for partial reduct construction, in *Proceedings of Concurrency, Specification and Programming Workshop, Ruciane-Nida, Poland*, pp. 400–411.
- Andrzej SKOWRON (2000), Rough sets in KDD, in *Proceedings of the 16-th World Computer Congress (IFIP'2000), Beijing, China*, pp. 1–14.

- Dominik ŚLĘZAK (2000), Normalized decision functions and measures for inconsistent decision tables analysis, *Fundamenta Informaticae* 44, pp. 291–319.
- Dominik ŚLĘZAK (2002), Approximate entropy reducts, *Fundamenta Informaticae* 53, pp. 365–390.
- Jakub WRÓBLEWSKI (2001), Ensembles of classifiers based on approximate reducts, *Fundamenta Informaticae* 47, pp. 351–360.
- Beata ZIEŁOSKO, Marcin PILISZCZUK (2008), Greedy Algorithm for Attribute Reduction *Fundamenta Informaticae* (to appear).
- Beata ZIEŁOSKO (2005), On partial decision rules, in *Proceedings of Concurrency, Specification and Programming Workshop, Ruciane-Nida, Poland*, pp. 598–609.