Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 192 (2021) 3667–3676

Procedia Computer Science

www.elsevier.com/locate/procedia

25th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems

Selected approaches for decision rules construction-comparative study

Beata Zielosko[a,*], Krzysztof Żabiński[a]

[a] *Institute of Computer Science, University of Silesia in Katowice, Będzińska 39, 41-200 Sosnowiec, Poland*

Abstract

Decision rules are popular form of knowledge representation. From this point of view, length of such rules is an important factor since it influences on data understanding by experts. Unfortunately, the problem of construction of short rules is NP-hard, so different heuristics are discussed in the literature. The paper presents comparison of two selected methods for decision rules construction. The first one is connected with a new algorithm based on EAV model, the second one - with construction of rules based on reduct. Decision rules were induced for data sets from UCI ML Repository and compared from the point of view of length and support, and from the point of view of classification accuracy. Results of Wilcoxon test are also included.

Keywords: Rough sets; Decision rules; Length; Reduct; Classification; EAV model

1. Introduction

Data mining is a fast developing discipline of science mainly because of the constantly growing need to select, classify and deal with large amounts of data from different domains [16, 35]. The amounts are constantly growing, especially referring to switching a lot of disciplines of human activity to online mode recently. As the mentioned growth is practically uncontrollable and unclassifiable, new or improved data mining algorithms and methods are demanded.

Rough set theory is an extension within the classical set theory, suitable for describing concepts in the case of incomplete and uncertain data [22]. It has a variety of applications, e.g., bioinformatics, business and finance, decision analysis and systems, medicine, transport and many others. Methods and algorithms of rough sets are linked and used in other domains as machine learning, pattern recognition, data mining and knowledge discovery, feature selection and others [2, 5, 28, 32].

* Corresponding author
E-mail address: beata.zielosko@us.edu.pl

1877-0509
10.1016/j.procs.2021.09.140

Decision rules are popular and useful form of knowledge representation, mainly because of their simple form and easy interpretation by humans. From this point of view, length of such rules is an important evaluation measure studied in this paper. Decision rules can be induced directly from a data set or based on another object, which is called a reduct, in the rough sets. Reduct is a minimal subset of attributes that has the same classification power as the entire set of features. So, it allows to obtain the most relevant attributes from the whole set and is popular and useful tool for feature selection and knowledge discovery from data. Reduct cardinality is important evaluation measure, also from the point of view of knowledge representation.

Unfortunately, the problem of construction of rules and reducts with minimal length is NP-hard [18, 21]. In the paper, for decision rules construction, two approaches based on rough set theory are presented. The first one is based on greedy algorithm for partial reduct construction [17], the second one allows to construct decision rules directly from decision table [39] and can be considered as heuristic based on dynamic programming approach for decision rules minimization relative to length [1]. In [18] it was shown that under some natural assumptions on the class NP, considered greedy algorithm is close to the best polynomial approximate algorithms for the problem of partial reducts minimization. The aim of the paper is to compare length of decision rules constructed by the algorithm based on EAV (ang. Entity-Attribute-Value) model with decision rules constructed based on partial reduct. Results for classification accuracy for data sets from UCI Machine Learning Repository [8] are also provided.

The paper consists of six sections. Section 2 is devoted to the rough set theory, greedy algorithm for partial reduct construction and approaches for decision rules induction are presented. The algorithm which allows to construct decision rules directly from decision table and is based on EAV model was described in Section 3. Section 4 presents experimental results according to length, support and classification accuracy of rule-based models, for data sets from UCI ML Repository. Section 5 contains conclusions.

2. Rough sets

The rough set theory was developed by prof. Pawlak as a tool for dealing with incomplete and uncertain data [22]. A basic information unit for the rough sets is defined using indiscernibility relation defined relative to a given set of attributes. Objects characterized by the same values of attributes are indiscernible in view of the available knowledge about them. Sets of such indiscernible objects indicated by the same label (decision class) can be aggregated to form information granules about the universe and such sets are crisp. Otherwise, if sets of indiscernible objects are indicated by different decision classes, they are rough and such concepts cannot be characterized in the framework of knowledge available about their elements. So, in the rough sets theory there are used approximations of the rough concepts.

The lower approximation consists of all objects which surely belong to the concept and the upper approximation contains all objects which possibly belong to the concept. The difference between the upper and the lower approximation constitutes the boundary region of the rough concept. If the boundary region of a set is nonempty, it means that our knowledge about the set is not sufficient to define the concept precisely. So, rough set theory expresses imprecision by employing a boundary region of a set.

2.1. Reducts

In this section, some notions related to decision reduct will be introduced.

Data are presented in a tabular form which is defined as $T = (U, A \bigcup \{d\})$, where U is a nonempty, finite set of of objects and $A = \{a_1, \ldots, a_m\}$ is a nonempty, finite set of attributes, i.e., $a_i : U \to V_a$, where V_a is the set of values of attribute a_i called the domain of a_i. Attributes from the set A are called *conditional*. $d \notin A$ is a distinguished attribute called a *decision*.

For any subset of attributes $B \subset A$ and objects $x, y \in U$, an *indiscernibility relation* $IND(B)$ is defined as follows: $xIND(B)y \iff \forall a \in Ba(x) = a(y)$.

Based on indiscernibility relation, a notion of decision reduct can be defined: $B \subset A$ is a *decision reduct* if it is irreducible subset of attributes such that $IND(B) \subset IND(d)$. So, decision reduct is a minimal subset

of attributes which is sufficient to discern objects with different decision values. However, exact reducts can be overfitted, i.e., depend essentially on the noise or adjusted too much to the existing examples. In such situation, instead of an exact reduct with many attributes, it is more appropriate to work with a partial reduct containing smaller number of attributes which separate almost all pairs of rows with different decisions. Last years various kinds of reducts and their approximations are studied in rough set theory [10, 12, 17].

The popular evaluation measure connected with reduct is length, i.e., the number of attributes it contains. If reducts are considered as a way of knowledge representation, shorter ones are more preferable.

2.1.1. Partial reducts

There are different approaches for reduct construction. Since exact algorithms do not always allow to obtain reducts in reasonable time, especially for bigger data sets, approximate methods and different heuristics are very often used. Among others, we can distinguish genetic algorithms [29], heuristics based on discernibility matrix [15], construction of reducts based on sampling data from decision table [38], and many others [7, 20].

In [18, 21] it was shown that the problem of construcion of reduct with minimum cardinality is NP-hard. In the paper, greedy algorithm for partial reduct construction is presented [40]. It is based on the algorithm for partial cover construction. It was shown that there exist simple reduction of the problem of construction of an exact cover with minimal cardinality to the problem of construction of an exact reduct with minimal cardinality, the opposite reduction exists also. Similar situation is with approximate covers and reducts. Thus it was possible to use important results obtained for the set cover problem [27] for analysis of partial reducts. It was shown, that under some natural assumptions on the class NP, greedy algorithms are close to the best polynomial approximate algorithms for the problem of partial reducts minimization.

Algorithm 1 presents pseudo-code of a greedy algorithm for approximate test (super-reduct) construction. By $P(T)$ a set of unordered pairs of different rows of T with different values of decision attribute, is denoted. An attribute a_i *separates* a pair of rows from P(T) if they have different values at the intersection with the column a_i.

Let $0 \leq \alpha < 1$. A set of attributes Q is called an $\alpha - test$ ($\alpha - super - reduct$) for T if attributes from Q separate at least $(1 - \alpha)|P(T)|$ pairs of rows from the set $P(T)$. An $\alpha - test$ is called an $\alpha - reduct$ (or partial reduct) if each proper subset of the considered $\alpha - test$ is not an $\alpha - test$. Each $\alpha - test$ contains at least one $\alpha - reduct$ as a subset. For example, $0.01 - test$ means that at least 99% of pairs of rows from $P(T)$ should be separated by attributes from this partial test. If α is equal to 0 we have an exact test.

Algorithm 1 Greedy algorithm for partial test construction

Input: Decision table T with conditional attributes $a_1, \ldots, a_m$, and real number α, $0 \leq \alpha < 1$.
Output: α-test for T.

```
Q ⟵ ∅;
while Q is not an α-test for T do
    select a_i ∈ {a_1, ..., 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 ⟵ Q ∪ {a_i};
end while
return Q;
```

2.2. Decision rules

In this section, popular approaches for decision rules construction are presented.

Decision rules, considered in this paper, are induced directly or indirectly from decision table and are presented in the form:

$$(a_{i_1} = v_1) \wedge \ldots \wedge (a_{i_k} = v_k) \rightarrow d = v_d,$$

where $a_{i_1}, \ldots, a_{i_k} \in \{a_1, \ldots, a_m\}$, $v_i \in V_{a_i}$, $v_d \in V_d$. Pairs $(a_{i_1} = v_1)$ are called descriptors or simply conditions. The number of conditions in a premise part of rule is its length.

There are many methods for construction of decision rules encountered in the literature. They can be divided into two groups taking into account type of the algorithm used for their construction: exact or approximate ones. Approximate ones do not perfectly suit the learning data set whilst the exact ones do. On the other side, obtaining an exact solution within a reasonable time by using exact algorithms is not always achievable, especially for bigger data sets, so approximate algorithms are desirable.

In the first group we can distinguish brute-force approach applicable to relatively small decision tables, Boolean reasoning [23] and dynamic programming approach [1]. The second group consists of many different heuristics and modifications based on exact methods as based on Boolean reasoning [20], heuristics based on dynamic programming approach [39, 41], different kinds of greedy algorithms [19, 36], approaches based on sequential covering [6, 26] and many others [3, 9, 31].

There should be mentioned one more approach, which is used in rough set theory and allows for construction of decision rules based on reduct. In this case each rule has the same length equal to cardinality of reduct and each object from a decision table has assigned values corresponding to conditional attributes included in reduct only.

There are many measures for assessment of decision rules [4, 14, 37], which come from statistic, information theory and others domains, depending on the goal for which the rules are constructed. Two main perspectives are knowledge representation and knowledge discovery [33].

From the point of view of the first perspective length and support are important measures. According to length, the Minimum Description Length principle [24] states that:"the best hypothesis for a given set of data is the one that leads to the largest compression of data". Support allows to discover major patterns in data. It is a number of object from decision table such their attribute values satisfy the premise part of the rule and they have the same decision as the one attached to the rule.

From the point of view of the second perspective, classification accuracy is used. It is a number of object properly classified divided by the whole number of objects in test part of decision table.

3. Algorithm based on EAV model

The algorithm for construction of decision rules based on converting decision table to the entity-attribute-value form (EAV) has been introduced in [39]. It's general idea is to gather an attribute ranking basing on their distinguishability level. This level is determined by the attributes' standard deviation among decision classes. So, the mentioned approach allows to add feature selection step to the rule generation process. Described algorithm belongs to the group of heuristics as it constitutes approximate rules, but it grows from the dynamic programming approach for decision rules optimization relative to length [1].

The idea of dynamic programming approach for decision rules optimization is based on partitioning of a decision table into subtables which are created for each value of each conditional attribute. A directed acyclic graph is obtained which nodes correspond to subtables and edges are labeled by values of attributes. Based on the graph, it is possible to described decision rules with minimal length. However, if the number of attributes and their values is large, the size of the graph (the number of rows and edges) is huge. Therefore, obtaining an exact solution within a reasonable time is not always possible.

Thus, the algorithm based on EAV model was proposed. It uses the idea of partitioning of decision table into subtables, but only for the values of selected attributes. The process of partitioning subtables is finished when all rows in a given subtable have the same class label or all values of selected attributes were considered. Then, decision rules are created basing on corresponding values of selected attributes. Components of this new approach as transformation of decision table into EAV form and construction of ranking of attributes, turns out to simplify the process of generation of decision rules.

The first experimental results showed that the length of the rules constructed using proposed approach is not far from the optimal values obtained based on dynamic programming approach [39]. In the paper, the proposed algorithm will be compared with the reduct-based approach, from the point of view of knowledge representation and classification.

3.1. EAV Model

EAV model is a different way to represent the decision table. It has been proposed in [13]. The idea is that every attribute with its value forms a separate row in the EAV table. Moreover, the newly formed row is supplied with the row number and decision from the original form of input decision table.

The idea to convert to EAV form is to make the analysis more comfortable from the point of view of utilization of RDBMS (Relational Database Management System) for data processing. RDBMS engines are more optimized for dealing with large categorized data sets than standard programming languages. It allows to significantly decrease the computational time, especially for sets of numerous multi-valued attributes. The idea of applying SQL for rule generation is not new and has already been studied, for instance, for construction of association rules [25] and decision rules [34].

The example below demonstrates how an EAV table can be defined in RDBMS - PostgreSQL in this case:

Listing 1. EAV table

```
CREATE TABLE eav
(
  id serial primary key,
  attribute character varying,
  value character varying,
  decision character varying,
  row bigint
);
```

3.2. Selection of attributes

The idea behind the EAV-based rule generation algorithm is to use standard deviation as a level of distinguishability. The standard deviation is calculated for each attribute value and grouped per decision classes, it is based on Bayesian data analysis [30]. In order to calculate standard deviation of non-numerical attributes, their values' numerical equivalents are considered. The equivalents are typically just ordinal numbers.

With the growth of the standard deviation obtained, the distinguishability level of attributes also increases. It leads to the conclusion that the attributes with the highest values of the mentioned standard deviation are in the highest places in attributes ranking. This directly implies to the utilization of the feature selection in terms of rule generation.

3.3. Construction of decision rules

Before the algorithm will be presented, some notions are introduced.

A table obtained from T by the removal of some rows is called a subtable of the table T. Let T be nonempty, $a_{i_1}, \ldots, a_{i_k} \in \{a_1, \ldots, a_m\}$ and $b_1, \ldots, a_k$ be values of attributes. The subtable of the table T that contains only rows that have values $a_1, \ldots, a_k$ at the intersection with columns $a_{i_1}, \ldots, a_{i_k}$ is denoted by $T' = T(a_{i_1}, a_1) \ldots (a_{i_k}, a_k)$. Such nonempty subtable (including the table T) is called *separable subtable* of T.

A minimum decision value that is attached to the maximum number of rows in T is called the most common decision for T.

The table T is called *degenerate* if T is empty or all rows of T are labeled with the same decision's value, or all rows have the same values of conditional attributes.

Algorithm 2 presents pseudo-code of the algorithm for construction of decision rules based on EAV model.

Notation of symbols are the following: T is a decision table; p is the ceiling of then number of percentage of the selected best attributes in the formed ranking; R is the set of generated rules; v is the unique set of values from the T in EAV form grouped per rows of input table T; v_i and v_i' are temporary subsets of v for the sake of rule generating iteration, based on the values included in the set v and its subsets separable subtables are created.

Algorithm 2 Pseudo-code of algorithm generating decision rules for a decision table T.

Input: Input decision table T, number p of best attributes to be taken into consideration.
Output: Set of decision rules R

represent T as entity–attribute–value (EAV) form with separate decision;
represent each attribute's value in a discrete numerical form;
obtain attributes' standard deviation per decision class.
take p number of attributes of largest STD—in a descending order;
from T in EAV form select sets v of unique values (including decision) of attributes grouped per decision table's rows;
while there exist sets v_i in v not marked as processed **do**
  generate one-item v_i' set with initial value from v_i which corresponds to creation of separable subtable $T' = T(f_i, a_i)$;
  set v_i is not processed;
  while iterations number $< sizeof(v_i)$ OR separable subtable is not degenerate **do**
    extend v_i' by supplying it with the subsequent element from v_i which corresponds to next partition of T';
  end while
  generate decision rule basing on the values of attributes from v_i' (consequent is the most common decision for T' corresponding to v_i');
  supply the set R with the newly created rule;
  set v_i being processed.
end while

The Algorithm 2 starts its operation with choosing the attributes for rule generation. It is based on the idea of selection of attributes described in the previous section (taking into account standard deviation of attributes). The percentage of attributes taken into consideration should be empirically chosen as it is related to the training set's structure.

After choosing the attributes, the Algorithm 2 proceeds to the rule formulation step. To begin with, unique combinations of attributes with values from the training set are collected. The set of combinations consists of only the attributes selected in the previous step. The attributes order in the subsequent combinations is dependent on the place in the attributes ranking. As for now, the decision is not considered.

Next, for every of the attribute combination, the separable subtables of the input training table are formed. For every combination, the rule formulation begins with choosing the highest ranked attribute and its value. Separable subtable gets generated for this attribute. Then, it needs to be checked if this subtable is degenerate. If it is degenerate, then the rule can be already formed (from the chosen attribute and its value). It also contains degenerate subtable's decision. If it is not degenerate, the next attribute from the ranking is taken and the next separable subtable is created. If it is degenerate, the rule can be formed from these two mentioned attributes. The process is repeated iteratively, the stop criterion is either a degenerate table or all attributes already utilized. If all attributes are utilized and the subtable is not degenerate, the rule is formed from all the attributes and the most common decision from the separable subtable is taken as the decision. This feature allows the algorithm to be applied to inconsistent data sets.

4. Experimental results

Experiments were performed on data sets from UCI Machine Learning Repository [8]. Attributes of unique value for each row were removed. As for missing values, they were replaced with the most common values for a given attribute for which such a situation appears. When, in some of the decision tables, there were equal values of conditional attributes but different decisions, then each group of identical rows was replaced with a single row from the group with the most common decision for this group.

The goal of the conducted experiments was to compare decision rules generated by the EAV model-based algorithm and the approach based on partial reduct. Decision rules were assessed taking into account:

- length and support which are important from the point of view of knowledge representation,
- classification accuracy which is important from the point of view of knowledge discovery.

Table 1 presents length of decision rules created based on partial reducts, for $\alpha \in \{0.001, 0.01, 0.1\}$, and minimum, average and maximum length of decision rules constructed based on EAV model, for 100%, 80% and 60% of attributes from the whole set of features.

Table 1. Length of decision rules created based on partial reducts and EAV model-based algorithm

data set	rows	attr	α				100%			80%			60%		
			0.0	0.001	0.01	0.1	min	avg	max	min	avg	max	min	avg	max
breast-cancer	266	9	8	6	3	2	1	4.42	8	1	4.39	7	1	4.09	5
cars	1728	6	6	5	4	2	2	3.91	5	2	3.31	4	2	2.67	3
house-votes	279	16	11	7	4	2	2	5.61	15	2	5.56	12	2	5.38	9
kr-vs-kp	3196	36	29	10	6	3	2	7.28	34	2	7.17	27	2	6.86	20
mushroom	8124	22	4	3	2	1	1	1.98	7	1	1.98	7	1	1.98	7
soybean-small	47	35	2	2	2	1	1	3.40	8	1	3.40	8	1	3.40	8
spect-test	169	22	11	11	7	3	2	6.38	21	2	6.26	17	2	5.88	13
tic-tac-toe	958	9	8	6	5	3	2	4.71	7	2	4.71	7	2	4.29	5

As for rule lengths, the statistical analysis by means of the Wilcoxon two-tailed test [11] has been performed (comparison of results for $\alpha = 0$ and percentage of best attributes=100%). The average values of rule's length obtained for the EAV model-based algorithm for 100% attributes are significantly smaller than length of rules formed from reducts, for α equal to 0. It is graphically presented on the Figure 1. Taking the

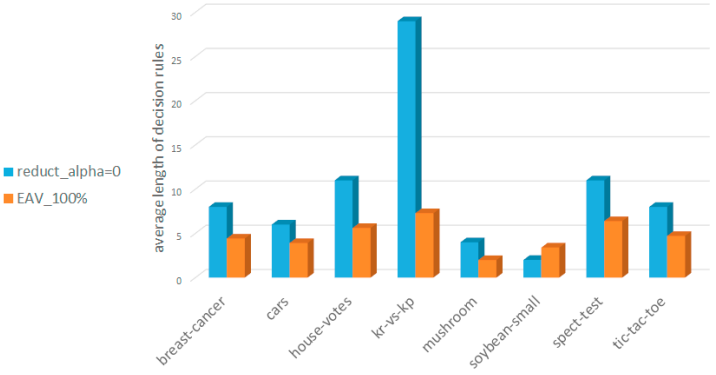

Figure 1. Length of decision rules constructed for the whole set of attributes in decision table, for both approaches

results into consideration, it can be seen, that reduct-based approach for rule generation, for small values of α constructs, on average, longer rules than the EAV model-based algorithm. When increasing α values, the results become comparable.

Tables 2 and 3 present minimum, average and maximum support of rules, for reduct-based approach and EAV model-based algorithm, respectively.

Table 2. Support of rules formed from reducts

data set	rows	attr	α=0.0			α=0.001			α=0.01			α=0.1		
			min	avg	max	min	avg	max	min	avg	max	min	avg	max
breast-cancer	266	9	1	1.04	2	1	1.06	2	1	2.59	6	1	6.71	15
cars	1728	6	1	1.00	1	1	3.81	4	1	10.52	12	10	64.21	108
house-votes	279	16	1	1.85	5	1	7.60	16	1	39.31	68	3	73.52	95
kr-vs-kp	3196	36	1	1.11	4	1	14.96	49	2	139.80	374	28	452.27	950
mushroom	8124	22	8	208.49	432	8	378.79	864	24	927.43	1584	36	2139.55	3408
soybean-small	47	35	4	7.51	10	4	7.51	10	4	7.51	10	4	11.51	17
spect-test	169	22	1	1.67	6	1	1.53	4	1	3.46	9	1	22.28	34
tic-tac-toe	958	9	1	1.00	1	1	2.53	7	1	5.38	17	2	32.08	78

Table 3. Support of rules generated by EAV model-based algorithm

data set	rows	attr	100%			80%			60%		
			min	avg	max	min	avg	max	min	avg	max
breast-cancer	266	9	1	3.73	22	1	3.73	22	1	4.14	22
cars	1728	6	1	22.52	64	2	23.32	64	6	27.69	64
house-votes	279	16	1	30.73	82	1	30.73	82	1	30.99	82
kr-vs-kp	3196	36	1	12.31	53	1	13.43	53	1	15.21	53
mushroom	8124	22	4	241.27	528	4	241.27	528	4	241.27	528
soybean-small	47	35	1	2.79	5	1	2.79	5	1	2.79	5
spect-test	169	22	1	14.22	35	1	14.20	35	1	14.30	35
tic-tac-toe	958	9	1	6.43	38	1	6.43	38	1	6.73	38

As for average values of rule support, the statistical analysis by means of the Wilcoxon two-tailed test has also been performed (comparison of results for $\alpha = 0$ and percentage of best attributes=100%). The values of rule support obtained for the EAV model-based algorithm for 100% attributes are significantly greater than values of support for rules formed from 0-reducts. However, for $\alpha = 0.1$ it can be observe in Tables 2 and 3 that results for reduct based-approach outperform results obtained for EAV model-based algorithm for 60% of best attributes.

Experiments connected with accuracy of rule-based classifiers were performed. Each set has been randomly divided into 10 equally-sized subsets then 10-fold cross validation procedure was applied. Tables 4 and 5 present average classification accuracy and standard deviation, for reduct-based approach and EAV model-based algorithm, respectively. Classification accuracy is the number of objects (rows) from the test part of decision table which are correctly classified divided by the number of all rows in the test part of decision table. In the case of conflicts a standard voting strategy was used, in which every rule votes with the weight proportional to the objects supporting it.

Table 4. Classification accuracy for rules formed from reducts

data set	α=0.0		α=0.001		α=0.01		α=0.1	
	accuracy	std	accuracy	std	accuracy	std	accuracy	std
breast-cancer	0.80	0.35	0.82	0.33	0.83	0.35	0.84	0.36
cars	0.73	0.05	0.75	0.07	0.77	0.08	0.76	0.05
house-votes	0.67	0.04	0.66	0.02	0.69	0.01	0.67	0.07
kr-vs-kp	0.60	0.40	0.61	0.38	0.57	0.41	0.57	0.41
mushroom	0.74	0.05	0.79	0.05	0.79	0.02	0.79	0.02
soybean-small	0.81	0.34	0.84	0.35	0.85	0.32	0.80	0.34
spect-test	0.99	0.02	0.99	0.02	0.99	0.02	0.99	0.02
tic-tac-toe	0.74	0.37	0.74	0.37	0.74	0.37	0.74	0.37

Table 5. Classification accuracy for rules generated by EAV model-based algorithm

data set	100%		80%		60%	
	accuracy	std	accuracy	std	accuracy	std
breast-cancer	0.78	0.32	0.79	0.34	0.81	0.34
cars	0.79	0.06	0.80	0.02	0.81	0.03
house-votes	0.74	0.07	0.77	0.10	0.78	0.10
kr-vs-kp	0.66	0.32	0.70	0.28	0.74	0.25
mushroom	0.87	0.13	0.87	0.12	0.87	0.12
soybean-small	0.81	0.31	0.82	0.31	0.84	0.32
spect-test	0.88	0.09	0.90	0.11	0.92	0.13
tic-tac-toe	0.71	0.39	0.75	0.39	0.77	0.39

As for classification results, the statistical analysis by means of the Wilcoxon two-tailed test has been performed, to verify the null hypothesis that there are none differences in the assessment of classifiers.

For Table 4 the analysis shows that $min(W_+, W_-) > W_{crit}$ (for all α values), so it can be concluded that there is no significant difference between classification results collected for different values of α. The null hypothesis has been confirmed.

Moreover, having repeated the Wilcoxon test for EAV model-based algorithm, it turns out that there are also no significant differences between classification results for 100% and 80% of best attributes. They are also comparable with the results obtained for reducts for all α values. Nevertheless, the classification results obtained for the EAV model-based algorithm for 60% of best attributes are significantly better than both results obtained for 100% and 80% of best attributes as well as for all α values of reduct-based approach ($min(W_+, W_-) < W_{crit}$).

It leads to the conclusion that EAV model-based algorithm can perform better from the point of view of knowledge discovery, for 60% of selected attributes than the reduct-based approach. When reducing the number of attributes to be taken into consideration, the experiments show that 60% of attributes is the number to go for - it is consistent with the results presented in [39].

5. Conclusions

In the paper, two selected approaches for decision rules construction were studied. They are different approximation methods however, both of them utilize selection of attribute idea. Besides, in the case of approach based on reduct it was proved that greedy algorithm is close to the best approximate algorithms for partial reduct minimization problem. It was the reason why this approach was chosen to compare the length of the rules obtained by algorithm based on EAV model.

Experiments were performed from the point of view of knowledge representation and knowledge discovery. Short rules are easier for understanding and interpreting by experts. Quality of the rule-based model describing a given data set is important component very often considered in different applications.

Obtained results show that despite similarities from the feature selection point of view, the EAV model-based approach allows to obtain shorter rules with greater support in the case of $\alpha = 0$ for reduct-based approach and 100% attributes for EAV-model based algorithm. For classification results it was statistically confirmed that 60% of selected attributes outperform results obtained for reduct-based approach, for each value of α.

Future works will be connected with improving the work of the heuristics proposed by authors, mainly related to the automatic selection of attributes considered during decision rules construction and comparison with other approaches.

References

[1] Alkhalid, A., Amin, T., Chikalov, I., Hussain, S., Moshkov, M., Zielosko, B., 2013. Optimization and analysis of decision trees and rules: dynamic programming approach. International Journal of General Systems 42, 614–634.
[2] Alsolami, F., Azad, M., Chikalov, I., Moshkov, M., 2020. Decision and Inhibitory Trees and Rules for Decision Tables with Many-valued Decisions. volume 156 of *Intelligent Systems Reference Library*. Springer.
[3] Alsolami, F., T.Amin, M.Moshkov, B.Zielosko, K.Żabiński, 2019. Comparison of heuristics for optimization of association rules. Fundamenta Informaticae 166, 1–14.
[4] An, A., Cercone, N., 2000. Rule quality measures improve the accuracy of rule induction: An experimental approach, in: Raś, Z.W., Ohsuga, S. (Eds.), ISMIS, Springer. pp. 119–129.
[5] Błaszczyński, J., de Almeida Filho, A.T., Matuszyk, A., Szeląg, M., Słowiński, R., 2021. Auto loan fraud detection using dominance-based rough set approach versus machine learning methods. Expert Systems with Applications 163, 113740.
[6] Błaszczyński, J., Słowiński, R., Szeląg, M., 2011. Sequential covering rule induction algorithm for variable consistency rough set approaches. Information Sciences 181, 987–1002.
[7] Cornelis, C., Martín, G.H., Jensen, R., Ślęzak, D., 2008. Feature selection with fuzzy decision reducts, in: Wang, G., Li, T., Grzymala-Busse, J.W., Miao, D., Skowron, A., Yao, Y. (Eds.), Rough Sets and Knowledge Technology, Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 284–291.
[8] Dua, D., Graff, C., 2017. UCI Machine Learning Repository. http://archive.ics.uci.edu/ml.
[9] Grzymała-Busse, J.W., 1997. A new version of the rule induction system lers. Fundamenta Informaticae 31, 27–39.
[10] Henzel, J., Janusz, A., Sikora, M., Ślęzak, D., 2020. On positive-correlation-promoting reducts, in: Bello, R., Miao, D., Falcon, R., Nakata, Michinoriand Rosete, A., Ciucci, D. (Eds.), Rough Sets, Springer International Publishing. pp. 213–221.
[11] Hole, G., 2011. Graham Hole Research Skills. University of Sussex. http://users.sussex.ac.uk/~grahamh/RM1web/WilcoxonHandoout2011.pdf.
[12] Janusz, A., Stawicki, S., 2011. Applications of approximate reducts to the feature selection problem, in: Yao, J., Ramanna, S., Wang, G., Suraj, Z. (Eds.), Rough Sets and Knowledge Technology, Springer Berlin Heidelberg. pp. 45–50.

[13] Kowalski, M., Stawicki, S., 2012. SQL-based heuristics for selected KDD tasks over large data sets, in: Proceedings of the Federated Conference on Computer Science and Information Systems, IEEE. pp. 303–310.
[14] Liu, H., Cocea, M., 2018. Induction of classification rules by gini-index based rule generation. Information Sciences 436-437, 227 – 246.
[15] Liu, Y., Zheng, L., Xiu, Y., Yin, H., Zhao, S., Wang, X., Chen, H., Li, C., 2020. Discernibility matrix based incremental feature selection on fused decision tables. International Journal of Approximate Reasoning 118, 1–26.
[16] Makarova, I., Khabibullin, R., Pashkevich, A., Shubenkova, K., 2017. Modeling as a method to improve road safety during mass events. Transportation Research Procedia 20, 430–435.
[17] Moshkov, M.J., Piliszczuk, M., Zielosko, B., 2007. On construction of partial reducts and irreducible partial decision rules. Fundamenta Informaticae 75, 357–374.
[18] Moshkov, M.J., Piliszczuk, M., Zielosko, B., 2008. Partial Covers, Reducts and Decision Rules in Rough Sets - Theory and Applications. volume 145 of *Studies in Computational Intelligence*. Springer.
[19] Moshkov, M.J., Piliszczuk, M., Zielosko, B., 2009. Greedy algorithm for construction of partial association rules. Fundamenta Informaticae 92, 259–277.
[20] Nguyen, H.S., 2006. Approximate Boolean reasoning: foundations and applications in data mining, in: Peters, J.F., Skowron, A. (Eds.), Transactions on Rough Sets V. Springer. volume 4100 of *Lecture Notes in Computer Science*, pp. 334–506.
[21] Nguyen, H.S., Ślęzak, D., 1999. Approximate reducts and association rules - correspondence and complexity results, in: Zhong, N., Skowron, A., Ohsuga, S. (Eds.), RSFDGrC. Springer. volume 1711 of *LNCS*, pp. 137–145.
[22] Pawlak, Z., 2002. Rough sets and intelligent data analysis. Information Sciences 147, 1–12.
[23] Pawlak, Z., Skowron, A., 2007. Rough sets and boolean reasoning. Information Sciences 177, 41–73.
[24] Rissanen, J., 1978. Modeling by shortest data description. Automatica 14, 465–471.
[25] Sarawagi, S., Thomas, S., Agrawal, R., 2004. Integrating association rule mining with relational database systems: Alternatives and implications. Data Mining and Knowledge Discovery 4, 89–125.
[26] Sikora, M., Wróbel, L., Gudyś, A., 2019. Guider: A guided separate-and-conquer rule learning in classification, regression, and survival settings. Knowledge-Based Systems 173, 1 – 14.
[27] Slavík, P., 1996. A tight analysis of the greedy algorithm for set cover, in: Proc. of the Twenty-Eighth Annual ACM Symposium on Theory of Computing, ACM New York. pp. 435–441.
[28] Ślęzak, D., Grzegorowski, M., Janusz, A., Kozielski, M., Nguyen, S.H., Sikora, M., Stawicki, S., Wróbel, L., 2018. A framework for learning and embedding multi-sensor forecasting models into a decision support system: A case study of methane concentration in coal mines. Information Sciences 451-452, 112–133.
[29] Ślęzak, D., Wróblewski, J., 2003. Order based genetic algorithms for the search of approximate entropy reducts, in: Wang, G., Liu, Q., Yao, Y., Skowron, A. (Eds.), RSFDGrC 2003. Springer. volume 2639 of *LNCS*, pp. 308–311.
[30] Ślęzak, D., 2005. Rough sets and bayes factor, in: Peters, J.F., Skowron, A. (Eds.), Transactions on Rough Sets III, Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 202–229.
[31] Stańczyk, U., Zielosko, B., 2020. Heuristic-based feature selection for rough set approach. International Journal of Approximate Reasoning 125, 187–202.
[32] Stańczyk, U., Zielosko, B., Żabiński, K., 2018. Application of greedy heuristics for feature characterisation and selection: A case study in stylometric domain, in: Nguyen, H., Ha, Q., Li, T., Przybyla-Kasperek, M. (Eds.), Proceedings of the International Joint Conference on Rough Sets, IJCRS 2018. Springer, Quy Nhon, Vietnam. volume 11103 of *Lecture Notes in Computer Science*, pp. 350–362.
[33] Stefanowski, J., Vanderpooten, D., 2001. Induction of decision rules in classification and discovery-oriented perspectives. Int. J. Intell. Syst. 16, 13–27.
[34] Świeboda, W., Nguyen, H.S., 2012. Rough set methods for large and spare data in eav format, in: IEEE RIVF International Conference on Computing Communication Technologies, Research, Innovation, and Vision for the Future, pp. 1–6.
[35] Tsybunov, E., Shubekova, K., Buyvol, P., Mukhametdinov, E., 2018. Interactive (intelligent) integrated system for the road vehicles' diagnostics. Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST 222, 195–204.
[36] Valmarska, A., Lavrač, N., Fürnkranz, J., Robnik-Šikonja, M., 2017. Refinement and selection heuristics in subgroup discovery and classification rule learning. Expert Systems with Applications 81, 147 – 162.
[37] Wróbel, L., Sikora, M., Michalak, M., 2016. Rule quality measures settings in classification, regression and survival rule induction - an empirical approach. Fundamenta Informaticae 149, 419 – 449.
[38] Yang, Y., Chen, D., Wang, H., Tsang, E.C., Zhang, D., 2017. Fuzzy rough set based incremental attribute reduction from dynamic data with sample arriving. Fuzzy Sets and Systems 312, 66–86.
[39] Żabiński, K., Zielosko, B., 2021. Decision rules construction: Algorithm based on eav model. Entropy 23, 1–18.
[40] Zielosko, B., Piliszczuk, M., 2008. Greedy algorithm for attribute reduction. Fundamenta Informaticae 85, 549–561.
[41] Zielosko, B., Żabiński, K., 2018. Optimization of decision rules relative to length based on modified dynamic programming approach, in: Stańczyk, U., Zielosko, B., Jain, L.C. (Eds.), Advances in Feature Selection for Data and Pattern Recognition. Springer. volume 138 of *Intelligent Systems Reference Library*, pp. 73–93.