Supervised Discretization for Rough Sets – a Neighborhood Graph Approach

Visnja Ognjenovic, Vladimir Brtka
Technical Faculty "Mihajlo Pupin"
University of Novi Sad
Djure Djakovica bb
23000 Zrenjanin, Serbia
{visnjao, vbrtka}@tfzr.uns.ac.rs

Martin Jovanovic
Faculty of Electronic Engineering
University of Nis
Aleksandra Medvedeva 14
18000 Nis, Serbia
martin.jovanovic@elfak.ni.ac.rs

Ivana Berkovic
Technical Faculty "Mihajlo Pupin"
University of Novi Sad
Djure Djakovica bb
23000 Zrenjanin, Serbia
berkovic@tfzr.uns.ac.rs

Abstract. *Rough set theory has become an important mathematical tool for dealing with uncertainty in data. The data discretization is one of the main problems to be solved in the process of synthesis of decision rules from table-organized data. In this paper, we present a new discretization method in the context of supervised training. This method is based on the neighborhood graph. To evaluate supervised discretization, we used data sets obtained from the UCI Machine Learning Repository. We have used the Rosetta system and proposed SSCO system. The experimental results show that our method is effective.*

Keywords: Rough sets, supervised discretization, neighborhood graph.

1 Introduction

Rough set theory (RST) was introduced by Pawlak [14] in 1980s. The goal of the rough set theory is to synthesize approximation of concepts from the data tables. This approximation could be expressed in form of decision rules, which can be used for classification tasks, (so called classification rules). The set of rules contains the rules in the If Then form [5, 6, 9]. Most real life data sets consist of, not only discrete attribute's values, but also continuous attribute's values. To utilize approaches based on RST more effectively, we should replace continuous attribute's values by discrete attribute's values.

The discretization step determines how coarsely we want to view the world. For instance, temperature, which is usually measured in real numbers, can be discretized into two, three or more intervals. Discretization is not specific to the rough set approach but is a pre-required step and is often performed implicitly, behind the scene, using expert knowledge [9, 4]. Many traditional discretization algorithms have been applied to rough sets. Some new discretization algorithms have also been proposed from the viewpoint of rough sets [7, 10, 11, 16].

Discretization algorithms can be divided into two categories: unsupervised and supervised [10]. As in [7] some unsupervised algorithms, such as the equal width (EW) and equal frequency binning (EF), do not take advantage of class information to increase their performance, so, the resulting discretization schemes do not provide much efficiency when used in the classification process, e.g. they contain more intervals than necessary. Many different supervised algorithms have been proposed: statistics-based, entropy-based, naive scaler, seminaive scaler, etc.

Methods, which do not consider the specificity of RST decision system, may yield inconsistency and reduce extension of classification rules after using discretization methods [7]. That is the reason for development of some special discretization methods for RST [1, 11].

This paper deals with a possibility of presenting a set of rules by a graph (decision tree). Presentation by the graph enables analysis of adjacency matrix. By the approach based on the matrix analysis, it is possible to observe single child nodes. These nodes show attributes which should be re-discretized in order to achieve a set of rules with greater classification capabilities. This is very important for domain experts because they can have introspection to attribute's importance. Unlike some existing rough sets based system such as Rosetta, it is possible to obtain a fully automated discretization and re-discretization algorithm without any user interference. The re-discretization of selected attributes also means that a kind of self-learning automated process is achieved.

This paper is organized as follows. The second section contains a brief introduction to the rough set theory and some preliminaries that are relevant to this paper. In section 3, we present our discretization method. Experimental results are given in section 4. The fifth section of this work contains conclusions and remarks.

2 Rough sets theory

The Rough sets theory as an original approach proved to be very useful for the analysis of data in various domains [13]. In Rough set terminology, a data table is also called an information system.

Let U be a universe (finite set of objects), $Q = \{q_1, q_2, ..., q_m\}$ is a finite set of attributes, V_q is the domain of attribute q and $V = \bigcup_{q \in Q} V_q$. [15].

An information system is defined as the quadruple $S = \langle U, Q, V, f \rangle$ where $f = U \times Q \to V$ is a total function such that $f(x, q) \in V_q$ for each $q \in Q, x \in U$, called information function.

If some of the attributes are interpreted as outcomes of classification, the information system $S = \langle U, Q, V, f \rangle$ can be defined as a decision system by $DS = \langle U, C, D, V, f \rangle$, where $C \cup D = Q$, $C \cap D = \varnothing$. C is called the set of condition attributes and set D is called the set of decision attributes [15]. Usually, there is one decision attribute.

Definition 2. Indiscernibility Relation. To every non–empty subset of attributes P is associated an indiscernibility relation on U, denoted by I_P:

$$I_P = \{(x, y) \in U \times U | \forall q \in P, f(x, q) = f(y, q)\} \tag{1}$$

The relation (1) is an equivalence relation – reflexive, symmetric and transitive. The family of all the equivalence classes of the I_P is denoted by $U|I_P$ and class containing an element x by $I_P(x)$ [9]:

$$U|I_P = \{[x]_{Ip} | x \in U\} \tag{2}$$

Where [x] I_P is the equivalence class:

$$[x]_{Ip} = \{y \in U | (x, y) \in I_p\} \tag{3}$$

If $(x, y) \in I_P$, then x and y are indiscernible (or indistinguishable) by attributes from P .

Definition 3. Set approximations. Let X be a non–empty set of U and $\varnothing \neq P \subseteq Q$.

Set X is approximated by means of P–lower (4) and P–upper (5) [9] approximations of X :

$$\underline{P}(X) = \{x \in U : I_P(x) \subseteq X\} \tag{4}$$

$$\overline{P}(X) = \bigcup_{x \in X} I_P(x) \tag{5}$$

The P–boundary of X is denoted by $Bn(X)$:

$$Bn(X) = \overline{P}(X) - \underline{P}(X) \tag{6}$$

Example 1. In the Table 1, there is a universe of six objects $U = \{x_1, x_2, x_3, x_4, x_5, x_6\}$ and each object is described by means of four attributes:

- Age
- Body Mass
- Fat%
- Blood Sugar Level

(Age, Body Mass and Fat% are conditional attributes, and Blood Sugar Level is decision attribute).

Object	***Age***	***Body Mass Index (BMI)***	***Fat%***	***Blood Sugar Level (BSL)***
x_1	young	good	low	low
x_2	middle-age	medium	low	high
x_3	middle-age	medium	low	low
x_4	old	medium	low	high
x_5	middle-age	good	high	high
x_6	young	medium	high	low

Table 1. Simple example of information system

In this particular case, the object x_1 is described by: Age=young, BMI=good, Fat%=low, BSL=low and so on. If $P = \{Age, BMI, Fat\%\}$ then, by (1), we have:

$$I_P = \{\{x_1,x_1\},\{x_2,x_2\},\{x_2,x_3\},\{x_3,x_2\},$$
$$\{x_3,x_3\},\{x_4,x_4\},\{x_5,x_5\},\{x_6,x_6\}\}$$

$$U|I_p = \{\{x_1\},\{x_2,x_3\},\{x_4\},\{x_5\},\{x_6\}\}.$$

Let us consider a case when set X contains only those elements where Blood Sugar Level is low: $X = \{x_1, x_3, x_6\}$, (see Table 1). Now, we can approximate set X using only the information contained in P by constructing the P–lower (4) and P–upper (5) approximations of X:

$$\underline{P}(X) = \{x_1, x_6\},$$

$$\overline{P}(X) = \{x_1, x_2, x_3, x_6\}.$$

The P–boundary (6) of X is:

$$Bn(X) = \{x_2, x_3\}.$$

2.1 Data reduction

One natural dimension of reducing data is to identify equivalence classes. This could be achieved by keeping only those attributes that preserve the indiscernibility relation and consequently, set approximation. The rejected attributes are redundant (superfluous) since their removal cannot worsen the classification. Let $\varnothing \neq P \subseteq Q$ and $a \in P$. Attribute a is superfluous in P, if $I_P = I_{P-\{a\}}$, otherwise it is indispensable attribute. The set P is orthogonal if all its attributes are indispensable. The set $P - \{a\}$ is a reduct of P if it is orthogonal and $I_P = I_{P-\{a\}}$ [3, 8].

From Example 1, one can notice that objects x_2 and x_3 (P–boundary of X) have exactly the same values of condition attributes but different value of the decision attribute. If R={Age, BMI}, S={Age, Fat%}, and T={BMI, Fat%}, then it is obvious that $I_R=I_P$ and $I_S=I_P$ while $I_T \neq I_P$. This means that R and S are reducts of P, while T is not. Attribute Age is indispensable, but attributes BMI and Fat% may be mutually exchanged. There are usually several subsets of reducts. After computing reducts the rules are easily constructed by overlaying the reducts over originating decision table and reading the values.

2.2 Data discretization

The reduct is direct consequent of data discretization. That is why it is important how to find the effective heuristics for discretization of the real values.

Since rough set theory is a logically founded approach, which is based on indiscernibility, the discretization of continuous attributes is a key transformation in rough sets. Discretization is usually performed prior to the learning process, which aims to partition continuous attribute's values into a finite set of adjacent intervals in order to generate attributes with a small number of distinct values. A good discretization algorithm can produce a concise summarization of continuous attributes, to help the experts and users to understand the data more easily, but also can make learning more accurate and faster [7, 11].

2.3 Decision rules

The expression $a = v$, where a is attribute and v is attribute value, is called descriptor. Now, it is possible to investigate rules of the form: IF α THEN β. Here α (rule's antecedent) denotes a conjunction (AND logical operator) of descriptors that only involve attributes of some reduct and let β (rule's consequent) denote a descriptor $d = v$, where d is decision attribute and v is allowed decision value [3].

It is essential to notice that shorter reduct set means shorter decision rules in the rule set generated from that reduct.

The length of generated rules is linked with the used reduct: all attributes from reduct are used in the IF part of each rule [5].

3 Supervised discretization by neighborhood graph

Decision rules can be represented using graphs, see Example 2.

Example 2. Lets define a set of rules:

1. IF A=20 AND C=3 THEN D=x
2. IF A=20 AND C=4 THEN D=x
3. IF A=30 AND C=3 THEN D=y
4. IF A=30 AND C=4 THEN D=z
5. IF A=30 AND C=5 THEN D=y
6. IF A=40 AND C=4 THEN D=w
7. IF A=40 AND C=5 THEN D=z
8. IF A=50 AND B=14 AND C=5 THEN D=z
9. IF A=50 AND B=16 AND C=6 THEN D=y
10. IF A=60 AND C=6 THEN D=w

Where A, B and C are the condition attributes while D is the decision attribute. The If parts of the rules could be shown by the graph (decision tree), see Figure 1:

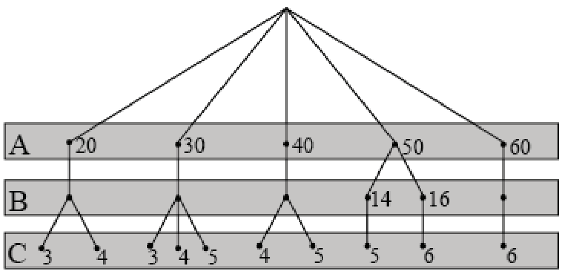

Figure 1. *Decision rules represented by the graph*

The initial set of rules is fragmented by attribute A so that each first-level node is associated with an attribute value. Second level of the graph is formed by attribute B, and finally third level of the graph is formed by attribute C. The graph from Figure 1 fully corresponds to set of rules from Example 2. The nodes which are not marked with an attribute value indicate the attribute which is not included in the If part of the corresponding rule, these nodes are only-child nodes.

By analyzing a set of rules shown in Example 2 as well as the graph shown in Figure 1, it is noticed that the attribute B occurs rarely (only in rules 8 and 9), which means that it has less significant effect on the decision. We propose that influence of the attribute B should be checked by re-discretization (an alternative way of discretization). This way, an expert has a choice between shorter rules and rules in which he can observe an influence of the attribute B.

Also, the adjacency matrix can observe the attribute B. By analysis of the adjacency matrix it is possible to observe an attributes, which occurs rarely (like attribute B from Example 2). This is the case when nodes have only one child (only-child). Thanks to this, it is possible to observe a sub matrix of adjacency matrix, which describes the frequency of occurrences of certain attributes. This makes it possible to identify attributes, whose frequency of occurrences in rules will be checked by re-discretization. The analysis is done by usage of the adjacency matrix of the graph in order to identify the only-child nodes.

Finally, method of discretization based on neighborhood graph can be formalized through the following steps:

I The first step is the standard discretization of attributes (the discretization method is selected by an expert), including the reduct computation and rule synthesis. This can be done by Rosetta system, RSES or some other similar system.

II The second step is the analysis of rules using the graph (decision tree). Based on the adjacency matrix of the graph we can find those attributes that have a higher number of consecutive only-child nodes.

III Only for the attributes that have a higher number of consecutive nodes that are only-child, the discretization is applied again, but more precise (a larger number of intervals is formed), the reducts are computed and new set of rules is generated.

IV By comparing the confusion matrices obtained by two different sets of rules, which were generated in previous steps, expert decides which set is better.

This kind of the supervised discretization is suitable for rough sets theory based systems for decision rules synthesis because it determines the attributes set for re-discretization. This is a standard step in rough sets based systems when supervised discretization is used [7].

4 Experiments and results

To present our method for discretization, we use well-known data sets from the UCI Machine Learning Repository [2]: Iris data set (iris). “This is perhaps the best known database to be found in the pattern recognition literature” [2]. The data set contains 3 classes of 50 instances each, where each class refers to a type of iris plant.

Number of Attributes: 4 numeric predictive attributes and one class attribute.

Attribute Information:

a_1: sepal length in cm
a_2: sepal width in cm
a_3: petal length in cm
a_4: petal width in cm
class attribute values:
-- Iris Setosa (1)
-- Iris Versicolour (2)
-- Iris Virginica (3)

This dataset is very suitable because of small number of objects (150) and small number of attributes involved, so that the discretization calculations can be done manually.

To generate rules, Rosetta system [12] is frequently used, but, in this research, we have used the SSCO system [3, 4, 5]. The SSCO system has been developed at the Technical Faculty "Mihajlo Pupin" in Zrenjanin, Serbia. This system is used for automated decision rules synthesis and is based on systematic syntax state space search. This section presents the comparison of the rules obtained by the SSCO system when different discretization algorithms are used. The comparison of the rule sets is done by the comparison of the corresponding confusion matrices.

4.1 The experiment with the discretization of the attributes

At the beginning, discretization was done manually in order to gain better opportunities for observation the only-child nodes in a matrix of neighborhood graph (decision tree). First, we have used the EW algorithm in order to discretize the set through generating a set of cuts, and then rules were generated by SSCO system. The output of the SSCO system includes If Then rules led by pair [*n*, *m*] where *n* corresponds to number of objects which support the If part of the rule while *m* is the ratio s/n, $s \leq n$ where *s* is a number of rules with different Then part for corresponding If part of the rule. In other words, if $m = 1$ then rule is generated by the objects from lower approximation of the rough set (exact rule), but if $m < 1$ then rule is generated by objects from boundary region of the rough set (inexact rule).

For example the rule: [4, 1] IF (a1,80), (a2,40), (a3,70) THEN (a5, Iris-virginica), means that if the sepal length equals 8 cm, sepal width equals 4 cm and petal length equals 7 cm then, it is Iris-virginica. There are four objects that support If part of the rule and there is no other rule with exactly the same If part but different Then part, so this is an exact rule. The previously described rule set is accompanied by so called confusion matrix. Confusion matrix *C* is a $|V_d| \times |V_d|$ matrix, where V_d is the set of possible values of decision attribute. This matrix with integer entries summarizes the performance of rule set while classifying the set of objects. Entry:

$$C_{i,j} = \left|\{x \in U : d(x) = i, \bar{d}(x) = j\}\right|,$$

Where $d(x)$ is the actual decision and $\bar{d}(x)$ is the predicted decision, which counts the number of objects that really belong to class *i*, but were classified to class *j*. Obviously, it is desirable for the diagonal entries to be as large as possible.

It is important to notice that confusion matrix was formed without any voting system so that one object could be classified to more than one class by some inexact rules.

In the case when EW discretization algorithm is used for iris database, SSCO output is as follows:

```
62 rules are added:

IF THEN Form
[1,1] IF (a1,80), (a2,40), (a3,70)
THEN (a5, Iris-virginica)
[1,1] IF (a1,80), (a2,40), (a3,60)
THEN (a5, Iris-virginica)
[1,1]  IF (a1,80), (a2,30),
(a3,70), (a4,24) THEN (a5, Iris-
virginica)
...
[2,1] IF (a1,50), (a2,30), (a3,20),
(a4,4) THEN (a5, Iris-setosa)
```

```
[12,1] IF (a1,50), (a2,30),
(a3,20), (a4,2) THEN (a5, Iris-
setosa)
[2,1] IF (a1,50), (a2,30), (a3,10),
(a4,4) THEN (a5, Iris-setosa)
[7,1] IF (a1,50), (a2,30), (a3,10),
(a4,2) THEN (a5, Iris-setosa)
[1,1] IF (a1,50), (a2,20), (a3,40)
THEN (a5, Iris-versicolor)
[2,1] IF (a1,50), (a2,20), (a3,30)
THEN (a5, Iris-versicolor)
[1,1] IF (a1,50), (a2,20), (a3,10)
THEN (a5, Iris-setosa)
[4,1] IF (a1,40) THEN (a5, Iris-
setosa)
```

Confusion matrix:

	1	2	3
1	50	0	0
2	0	50	9
3	0	7	50

134 objects are classified by exact rules, 16 objects are classified by inexact rules, 0 objects were not classified.

Based on the adjacency matrix the graph was formed. It was observed that the attribute a_2: sepal width and attribute a_4: petal width appear less frequently in a set of rules, and that rule `IF (a1,40) THEN (a5, Iris-setosa)` is very short. It is very important to notice that there is a frequent consecutive occurrence of the only-child nodes for attributes a_2 and a_4 in the graph. Further step of the experiment involves re-discretization process applied to attribute a_4.

4.2 The experiment with the re-discretization based on an analysis of the graph adjacency matrix

In this experiment the preparation of data was done so that the attribute a_4: petal width is discretized again (intervals are halved). Furthermore, we obtained 16 more rules, and the difference refers mainly to the rules for just one type of iris (Iris-setosa).

SSCO output is as follows:

```
78 rules are added:

IF THEN Form
```

```
[1,1] IF (a1,80), (a2,40), (a3,70)
THEN (a5, Iris-virginica)
[1,1] IF (a1,80), (a2,40), (a3,60)
THEN (a5, Iris-virginica)
[1,1] IF (a1,80), (a2,30), (a3,70),
(a4,23) THEN (a5, Iris-virginica)
...
[1,1] IF (a1,50), (a2,30), (a3,20),
(a4,5) THEN (a5, Iris-setosa)
[2,1] IF (a1,50), (a2,30), (a3,20),
(a4,4) THEN (a5, Iris-setosa)
[9,1] IF (a1,50), (a2,30), (a3,20),
(a4,2) THEN (a5, Iris-setosa)
[3,1] IF (a1,50), (a2,30), (a3,20),
(a4,1) THEN (a5, Iris-setosa)
[2,1] IF (a1,50), (a2,30), (a3,10),
(a4,3) THEN (a5, Iris-setosa)
[6,1] IF (a1,50), (a2,30), (a3,10),
(a4,2) THEN (a5, Iris-setosa)
[1,1] IF (a1,50), (a2,30), (a3,10),
(a4,1) THEN (a5, Iris-setosa)
[1,1] IF (a1,50), (a2,20), (a3,40)
THEN (a5, Iris-versicolor)
[2,1] IF (a1,50), (a2,20), (a3,30)
THEN (a5, Iris-versicolor)
[1,1] IF (a1,50), (a2,20), (a3,10)
THEN (a5, Iris-setosa)
[3,1] IF (a1,40), (a4,2) THEN (a5,
Iris-setosa)
[1,1] IF (a1,40), (a4,1) THEN (a5,
Iris-setosa)
```

Confusion matrix:

	1	2	3
1	50	0	0
2	0	50	6
3	0	7	50

137 objects are classified by exact rules, 13 objects are classified by inexact rules, 0 objects were not classified.

Here, class marker 1 refers to Iris Setosa, class marker 2 refers to Iris Versicolour, while class marker 3 refers to Iris Virginica.

By more precise discretization of the attribute a_4: petal width we have obtained slightly better objects classification. In the second experiment more objects (137) were classified by exact rules (134 object in the first experiment). Potentially, with every further re-discretization of the attributes that have a high number of consecutive only-child nodes, even better classification could be achieved.

5 Conclusions and future work

There are various methods and techniques for decision rules extraction from data. The rough sets theory (RST) is a good theoretical background for If Then rule synthesis from table-organized data. The classification power of these rules is often checked by calculation of so called confusion matrix. This matrix with integer entries summarizes the performance of rule set while classifying the set of objects. In this paper we presented the If Then rules generating process by so called SSCO system which has been developed at the Technical Faculty "Mihajlo Pupin" in Zrenjanin, Serbia. This system is used for automated decision rules synthesis and it is based on systematic syntax state space search. As a previous step to rule generating procedure there is a discretization step as a pre-required step, so that continuous attribute's values are divided into a finite set of adjacent intervals in order to generate attributes with a small number of distinct values. The analysis of generated decision rules was done by usage of the graph approach. Based on the adjacency matrix, it is possible to extract the attributes that appear less frequently in a set of rules. It is shown how to reconsider the influence of less frequent attributes to the value of the decision attribute. This is very important for human experts because they could have the deeper insight to generated rule sets. We have used UCI Machine Learning Repository "Iris" data set to conduct the experiments. First experiment is conducted so that equal width (EW) discretization algorithm is used and rules are formed by SSCO system. In the second experiment, the analysis of previously generated rules is done by usage of the graph (decision tree). Based on the adjacency matrix of the graph we can find those attributes that have a higher number of consecutive only-child nodes. Only for the attributes that have a higher number of consecutive nodes that are only-child, the discretization is applied again, but more precise (a larger number of intervals is formed), and new set of rules is generated. The result shows that re-discretization obtains better rules (a better confusion matrix is achieved). Also, this method can be incorporated into SSCO algorithm so that re-discretization process will be done automatically.

After incorporation of re-discretization algorithm into SSCO system, future work will include the test on a larger data set as well as the comparison with other supervised discretization methods. Then the system will be tested on real life data from various domains.

Acknowledgment Ministry of Science and Technological Development, Republic of Serbia financially support this research; under the project number TR32044 "The development of software tools for business process analysis and improvement".

References:

[1] Blajdo P., Grzymala-Busse W. J., Hippe S. Z., Knap M., Mroczek T. and Piatek L., A Comparison of Six Approaches to Discretization—A Rough Set Perspective, Lecture Notes in Computer Science, 2008, Volume 5009/2008

[2] Blake, C.L., Merz, C.J.: UCI Machine Learning Repository, http://archive.ics.uci.edu/ml/

[3] Brtka V. Berkovic I., Stokic E., Srdic B., A comparison of rule sets generated from Databases by indiscernibility relation - A rough sets approach, ICCP 2007: IEEE 3rd International Conference on Intelligent Computer Communication and Processing, Proceedings, (2007), vol. br. , str. 279-282

[4] Brtka V., Berkovic I., Brtka E., Jevtic V., A Comparison of Rule Sets Induced by Techniques Based on Rough Set Theory, SISY 2008, 6th International Symposium on Intelligent Systems and Informatics, September 26-27, 2008 Subotica,Serbia, IEEE Catalog Number: CFP0884C-CDR, ISBN: 978-1-4244-2407-8, Library of Congress: 2008903275.

[5] Brtka V., Berkovic I., Stokic E., Srdic B., Automated extraction of decision rules from medical databases - A rough sets approach (Proceedings Paper), 2007 5TH INTERNATIONAL SYMPOSIUM ON INTELLIGENT SYSTEMS & INFORMATICS, (2007), vol. br. , str. 27-31

[6] Brtka V., Stokic E., Srdic B., "Automated extraction of decision rules for leptin dynamics—A rough sets approach", Journal of Biomedical Informatics 41, pp. 667 – 674, 2008.

[7] Feng Jiang, Zhixi Zhao and Yan Ge, A Supervised and Multivariate Discretization Algorithm for Rough Sets, Lecture Notes in Computer Science, 2010, Volume 6401/2010, 596-603

[8] Greco S., Benedetto M., Slowinski R., "New Dewelopments in the Rough Set Approach to Multi – Attribute Decision Analysis", *Bulletin of International Rough Set Society*, Volume 2, Number 2/3, 1998, pp. 57–87.

[9] Komorowski J., Pawlak Z., Polkowski L., Skowron A., "Rough Sets: A Tutorial", http://citeseer.ist.psu.edu/komorowski98rough.html, 1998.

[10] Nguyen, H.S., Nguyen, S.H.: Discretization Methods in Data Mining. In: Rough Sets in Knowledge Discovery, Physica, pp. 451–482 (1998)

[11] Nguyen, H.S.: Discretization problem for rough sets methods. In: Polkowski, L., Skowron, A. (eds.) RSCTC 1998. LNCS (LNAI), vol. 1424, pp. 545–555. Springer, Heidelberg (1998)

[12] Øhrn, A.: Rosetta Technical Reference Manual (1999), http://www.idi.ntnu.no/_aleks/rosetta

[13] Pawlak Z., Grzymala-Busse J., Slowinski R. and Ziarko W., "Rough sets", Association for Computing Machinery. *Communications of the ACM*, 38, 1995.

[14] Pawlak, Z.: Rough sets. International Journal of Computer and Information Sciences 11, 341–356 (1982)

[15] Pawlak, Z.: Rough sets: Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, Dordrecht (1991)

[16] Zhao J., Zhou Y., New heuristic method for data discretization based on rough set theory, The Journal of China Universities of Posts and Telecommunications, Volume 16, Issue 6, December 2009, Pages 113-120