Journal of Theoretical and Applied Information Technology
15 September 2012. Vol. 43 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

# AN IMPROVED ROUGH SET COMBINED WITH AGGLOMERATIVE CLUSTERING

**SHENG YAO, LONGSHU LI**
Department of Computer Science and Technology, Anhui University, Hefei 230601, Anhui, China

**ABSTRACT**

Rough set is a powerful mathematical tool that has been applied widely to extract knowledge from many databases. However, some drawbacks have been detected in rough set, such as inconsistency, lack of flexibility, excessive dependency on discretization of the initial attributes and so on. To overcome these drawbacks, we propose an improved rough set combined with agglomerative clustering. The concept of equivalence class is also incorporated to merge and divide subclass. The experimental results show the better performance of the proposed approach.

**Keywords:** *Rough Sets, Agglomerative Clustering, Automatic Extraction*

## 1. INTRODUCTION

Rough Set is a theory and method which was first proposed by Poland scholar Pawlak in 1982 .[1-3]It has been played an important role in the area of inductive machine learning to uncover hidden patterns in data. It is also capable of assigning uncertainty to the extracted knowledge, identifying partial or total dependencies (cause-effect relationships) in databases, and eliminating redundant data. Rough Set theory has been studied by many researchers, and has made great strides [4]. It has been applied in cases of medicine [5], engineering [6], finance [7] and others fields.

Cluster analysis is a data analysis tool used to group data with similar characteristics. These techniques have been used in many areas such as manufacturing, medicine, nuclear science, radar scanning, data mining, intrusion detection, bioinformatics, classification of statistical findings in social studies and so on.

Furthermore, hybrids have been created between rough set and other mathematical methods that improve the quality of decision rules induced by rough set method [8].Recently, there has been work in the area of applying rough set to deal with uncertainty in cluster analysis. For example, Mazlack et al. [9] proposes two techniques to select clustering attribute: i.e. bi-clustering (BC) technique based on bivalued. Huang [10], Gibson et al. [11], Guha et al. [12], Ganti et al. [13], and Dempster et al. [14] proposed many algorithms for clustering categorical data. While these methods make important contributions to the issue of clustering categorical data, they are not designed to handle uncertainty in the clustering process.

This paper presents an improved rough set combined with agglomerative clustering to improve knowledge extraction. The concept of equivalence class was also incorporated to merge and divide subclass. In addition, ChiMerge and the Chi2 methods were used for the necessary initial data discretization with some variations [15].

The organization of the paper is as follows: In section 2, the basic concepts of rough set are introduced. In section 3, the theoretical concepts of a novel improved rough set are presented, In section 4, the proposed algorithm is described, In section 5 ,the whole test is showed, and in section 6,the experimental analysis is presented. Finally the conclusion is given in section 7.

## 2. BASIC CONCEPTS OF ROUGH SETS

Rough set is a mathematical theory that is used to handle uncertainty problems. It classifies imprecise, uncertain or incomplete information expressed in terms of data acquired from experience. A rough set is represented by a pair of crisp sets, called the lower approximation, which comprises of elements belonging to it and upper approximation, which comprises of elements possibly in the set with respect to the available information.

Let U be the universe and let $R \subseteq U \times U$ be an equivalence relation on U, called an indiscernibility relation. The pair K = (U,R) is called an approximation space. The lower and upper approximation of set X with respect to R can be written as[2]

$$\underline{R}(X) = \bigcup \{ [x] \in U / R \| [x]_R \subseteq X \} \quad (1)$$

$$\overline{R}(X) = \bigcup \{ [x] \in U / R \| [x]_R \cap X \neq \varnothing \} \quad (2)$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Where $[x]_R = \{y \in U | xRy\}$ is the equivalence class of x.

## 3. STRATEGY ALGORITHM FOR THE IMPROVED ROUGH SET

When working with large data sets and very inconsistent, knowledge extraction based on rough set mainly suffers from some drawbacks. These drawbacks include lack of flexibility and excessive dependency on the intervals chosen in the discretization of the attributes[16]. We propose a new algorithm that tries to overcome by introducing two improvements: one is the equivalence classes obtained by the method of rough set and the other is the post made from new samples reserved in the data set.

The objective will primarily improve the learning of equivalence classes belonging to the boundary region of which has not been able to obtain any certain rule in the application of Variable Precision Rough Set Model[17]. To achieve this goal, it incorporates the concept of an equivalence class. These are composed after a process of clustering of the samples, and it will be useful in the partitioning of equivalence classes.

Thus, there are two possible separations of the equivalence classes, one made from the centers obtained in the clustering [18]and the other working with the new updated examples of knowledge. They have created "subclasses of equivalence" which will be defined by the discredited values of condition attributes and new attributes generated by mathematical equations involving attributes without discretization. These new subclasses may generate both positive and uncertain new rules.

## 4. PROPOSED IMPROVED ROUGH SET COMBINED WITH AGGLOMERATIVE CLUSTERING

To carry out the rough set combined with agglomerative clustering[19], an algorithm has been developed, and consists of the following steps:

Step 1: Create the table of decision---the examples are distributed in the data set to be discussed at a table.

Step 2: Remove initial knowledge---Variable Precision Rough Set Model is used to refine the results from a clustering of equivalence classes.

Step 3: Updating of knowledge---separate the examples closed to equivalence classes other than their own by hyperplanes.

Step 4: Test--final rules are tested with new examples obtained.

### 4.1 Creating The Decision Table

Our aim is to express the data set, in which knowledge is extracted, so that it can be treated in the following steps.

To do this, we select the attribute of decision, which classifies the examples, and the condition attributes, which are the factors able to perform this classification. The ultimate goal is to determine the decision attribute value from the information provided by the condition attributes.

In the method of rough set, the examples are provided to the algorithm in a decision table in which rows are distributed by the examples available for training and in which each column corresponds to one of the attributes considered. Each cell of the table shows the value of an item in one of these attributes. The value will be expressed both in discrete form and standard form.

### 4.2 Initial Extraction

The aim is to discover rules hidden in the data set. Variable Precision Rough Set Model is used, and then there will be a grouping or clustering process by DIANA method to obtain new knowledge with a greater number of certain rules. Thus, this step is composed of two phases :The first will apply the method of rough set with an acceptable error level classification, which is proposed for the Variable Precision Rough Set model. The second will be held on clustering in each of the resulting equivalence classes not included in the positive region and have generated certain rules.

#### 4.2.1 Phase 1

Briefly, the mathematical concepts of this phase correspond to the model of variable precision rough set .

Let $A = \{c_1, c_2, ..., c_n\}$ be the set of the condition attributes and $B = \{d\}$ be the set of decision attributes. The set A has an associated equivalence relation[2]:

$$A = \{(x, y) \in U \times U : f(x, c_a) = f(y, c_a), \forall c_a \in A\} \quad (3)$$

where U is the set of training examples, f (x, $c_a$) is the discretized value that takes the example for the attribute $c_a$. The relationship of equivalence

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$A$ induces the partition $A^* = \{X_1, X_2, ..., X_n\}$ on U, which divides the examples into equivalence classes. Similarly, we obtain the partition blocks $B^* = \{Y_1, Y_2, ..., Y_m\}$ composed and induced by the equivalence relation associated with the decision.

Given a permissible level of classification error β, 0 ≤ β ≤ 5, define the β-positive regions and β-limit partition[20]. The β-lower and upper approximation of $Y_j \in B^*$ in the space S = (U, A) are defined as follows:

$$POSS_\beta\left(B^*\right) = \cup_{Y_j \in B^*} \underline{S}_\beta(Y_j) = \cup_{Y_j \in B^*} (\cup_{X_i \overset{\beta}{\subseteq} X} X_i) \quad (4)$$

$$BNDS_\beta\left(B^*\right) = \cup_{Y_j \in B^*} (\overline{S}_\beta(Y_j) - \underline{S}_\beta(Y_j)) = \cup_{Y_j \in B^*} (\cup_{\beta < c(X_i, X) < 1-\beta} X_i) \quad (5)$$

Each equivalence class framed in the positive region generated a certain rule between condition attributes and decision. The equivalence classes included in the boundary region will lead to uncertain rules with a degree of uncertainty expressed by a confidence factor. Therefore, a result will be obtained in this phase:

1.Rules $X_i \rightarrow Y_j$ , where the condition is an equivalence class and the conclusion is a block of the partition induced by decision attribute.

2. Uncertain rules $X_i \rightarrow Y_j$ are computed :

$$\alpha_{Y_j}(X_i) = \frac{|X_i \cap Y_j|}{|X_i|} \quad (6)$$

where |A| indicates the cardinality of a finite set A.

Since it has a level of allowable error classification β> 0, certain rules[21] will have a classification error in the range [0, β]. Furthermore, no uncertain rule will satisfy $\alpha_{Y_j}(X_i) > 1 - \beta$ .

#### 4.2.2 Phase 2

The rough set model in the previous phase completes the extraction of knowledge from the available data. However, when working with very inconsistent decision tables, this method may be too rigid, despite the slight flexibility provided by the allowed β error level. To address this, we add a second phase which allows separation into "groups" of the equivalence classes belonging to the boundary region of the partition induced by decision attribute. We aim to obtain some new rules from the equivalence classes.

After this process, each example will have one more attribute, and its closest equivalence class. In addition, the center of each group will be useful for the next step "updating knowledge".

In the previous phase, equivalence classes $X_i$ were assigned the partition $A^*$ a positive or boundary regions of the partition $B^*$. Now, try each class according to the region to which they belong, as follows:

Case 1: The class $X_i$ is included in the positive region ( $POSS_\beta\left(B^*\right)$ )

Misestimate center half of all the examples make up the equivalence class. It is considered that the class is composed of a single group.

Case 2: The class $X_i$ class is included in the boundary region ( $BNDS_\beta\left(B^*\right)$).

One or more centers are calculated for the equivalence class clustering treated by Diana method. These centers will be used to divide into fictional groups $X_i$ .

The set of centers $P = \{p_1, p_2, ..., p_c\}$ is obtained, each $p_k$ is defined by its values on the attributes of condition. The class will be divided into as many groups as $X_i$ be the set P. The set of groups $G = \{G_1, G_2, ..., G_c\}$ is formed from $X_i$ , each containing a subset of $G_k$ examples and also taking the center point associated with $p_k$ .

To distribute all such cases belonging to $X_i$ in the different groups G, we will use the Euclidean distance measure. Thus, a sample $x \in X_i$ , part of the group $G_k$ is associated the center closest to x. Therefore, the index k of the group is calculated:

$$k(x) = \arg_k \min \|x - p_k\| \quad (7)$$

Once distribute all the examples of the set $X_i$ in the various groups $G_k$ , it holds that G is a set of nonempty subsets

$$G_1 \cup G_2 \cup \cdots \cup G_c = X_{Y_j}, G_i \cap G_j = \phi, i \neq j \quad (8)$$

Therefore, G has the characteristics necessary to partition the class $X_i$ , and may treat each $G_k$ as a block of the partition G. The elements belonging to one of these blocks are characterized by having the same values for all attributes.

Next, check each of the blocks of a partition G, if it can be incorporated into the positive region or boundary of the partition $B^*$. Thus, the blocks of G, belonging to the positive region, will generate certain rules, and they are framed in the boundary

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

region. The uncertain rules generate a degree of uncertainty expressed by a confidence factor, such that:

$$\alpha_{Y_j}(G_k) = \frac{|G_k \cap Y_j|}{|G_k|} \qquad (9)$$

However, the way of expressing the rule is more complex because the concept must be added that the example should be closer to the center corresponding to the rest. Consequently, a rule from the group $G_k$ is: "if the attribute $c_1$ is xxx, ..., and the attribute $c_n$ is zzz, and besides, the example is closer to the center than the rest of $p_k$ class centers equivalence, then the attribute is ddd".

Finally, if a new rule has created some, it has fulfilled the main objective of this phase, gain some knowledge from some examples of those who could not draw a clear conclusion to what decision they belonged. It was necessary to add attribute (nearest) for these certain rules.

### 4.3 Calculation Centers by DIANA

The method DIANA is an agglomerative clustering algorithm and starts with a single cluster containing all objects in the equivalence class $X_i$, and that subdividing will be successively smaller clusters, so that they will form a dendrogram.

To expedite the process of agglomerative clustering and not to waste time on excessive branching of the dendrogram, a new condition was imposed to stop the method, the final selection of a maximum number of clusters n is less than the number of examples that make $X_i$.

After running the method, we have n clusters and perform a selection of the most representative. To do this, we calculate the representation of a final cluster J:

$$REP_J = \frac{|J|}{|\max(d(i,j))|} \qquad (10)$$

where max (d (i, j)) is the maximum distance between two examples of the cluster.

Next, select the final cluster with the highest representation, being $\max\_REP_j$ and clusters that meet that: $REP_j > (\max\_REP_j / 2)$ . Finally, we calculated for each of these clusters with the middle half of all the examples, and these centers will become part of the final set of centers of P.

### 4.4 Updating Knowledge

This step doesn't aim to generate more knowledge as do in the previous step. This will discuss new update examples, areas of equivalence classes that are close to groups of neighbors.

Let the data set $Q = \{x_1, x_2, ..., x_n\}$ aside for updating the knowledge of those o do not know the decision. For each $x_i \in Q$ will be held the following process:

First, it checks what kind of equivalence class

$x_i \in A^*$ and $x_i$ belong to the example according to their discretized values of the attributes.

Furthermore, since the set of groups equivalence class $G = \{G_1, G_2, ..., G_c\}$ was created in Step 2 from the extraction of knowledge, looking which group associated with the group $G_k$ according to the center nearest the center, the index k of the selected group.

If the example does not belong to a group or 'underclass of equivalence "framed in the positive region then calculate their Euclidean distances to the centers of the groups of equivalence classes nearby. If the nearest center is closer than the same group that owns the instance, the case is recorded.

The original training examples of a group will be divided into the group by a certain percentage of examples of renovation that are closer to another group to own. This percentage, calculated on the number of examples that make up the group, is determined using the method of trial and error to be very dependent on the data set studied. The division will be made by the equation of hyperplane equidistant from the center of one set to which the update examples are closer.

This hyperplane is equivalent to that defined by a perceptron neuron. Thus, hereinafter, in the groups divided in this way to check which subset of the two created sample, will be presented to the perceptron partner, and as its output is assigned to a subset or the other.

Finally, if each one of the subsets generated in the separation belongs to the positive region or boundary region, and calculates the mean center for further separation.

## 5. TEST

Let the set of data $T = \{x_1, x_2, ..., x_n\}$ in reserve for test. For each test example $x_i \in T$ ,which kind of equivalence belong to the example is checked according to their discretized values of the

Journal of Theoretical and Applied Information Technology
15 September 2012. Vol. 43 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

attributes, $x_i \in A^*$. Also find to which group belong according to the nearest facility that is calculated from the phase 2. If this group has been divided by one or more hyperplanes, the group will be presented to the perceptron for example.

Once know the group owns the instance, if it is framed in the positive region, the decision will be assigned directly. On the contrary, if it belongs to the boundary region, more certainty factor will be assigned to the decision. For the test example does not belong to any of the equivalence classes formed in the extraction of knowledge, the group belonging to the positive region whose center is closest will be assigned to the decision.

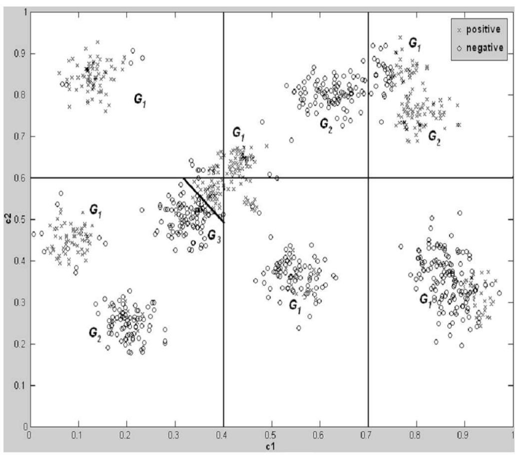


*Figure 1: Sets of sample data*

Figure 1 shows a sample data set characterized by two condition attributes ( $c_1$ and $c_2$ ) and one decision attribute with two possible values (positive and negative). About this group applies the described algorithm. The decision table obtained by executing step 1 of the algorithm is shown in Table 1.

*Table 1. Decision table of all sample data*

| | Attribute c1 | Attribute c2 | Decision |
|---|---|---|---|
| example 1 | 0.45 (high) | 0.63 (heavy) | positive |
| example 2 | 0.22 (low) | 0.34 (light) | negative |
| | ... (…) | ... (…) | ... |

In step 2 are generated partitions $A^* = \{X_1, X_2, X_3, X_4, X_5, X_6\}$, where $X_1$ contains the data $c_1$ and $c_2$, And $B^* = \{Y_1, Y_2\}$, where $Y_1$ is the data with a negative decision and $Y_2$ are t positive. Some rules are:

If $c_1$ and $c_2$ is high and light, then d is negative (completely true rule)

If $c_1$ and $c_2$ is low and heavy, then d is positive (completely true rule)

If $c_1$ and $c_2$ is low and light, then d is negative $\alpha_{Y_1}(X_1) = \dfrac{|X_1 \cap Y_1|}{|X_1|} = 0.35$

If $c_1$ and $c_2$ is low but non light, then d is positive $\alpha_{Y_2}(X_1) = \dfrac{|X_1 \cap Y_2|}{|X_1|} = 0.65$

For the equivalence class with $c_1$ and $c_2$ under light ( $X_1$ ), you can see three clusters obtained representative $(G_1, G_2, G_3)$, each with a center $(p_1, p_2, p_3)$. $G_1$ and $G_2$ will be incorporated into the positive region of $B^*$, generating two certain rules, while the $G_3$ will be part of the boundary region. The new rules are generated:

If $c_1$ is low, $c_2$ is light, and also the example is closer to the center that $p_1$ $p_2$ and $p_3$, then d is positive (rule completely true).

If $c_1$ is low, $c_2$ is light, and also the example is closer to the center that $p_2$ and $p_3$ $p_1$, then d is negative (rule completely true).

If $c_1$ is low, $c_2$ is light, and also the example is closer to the center than $p_3$ $p_1$ and $p_2$, then d is negative with a confidence factor = 0.70.

If $c_1$ is low, $c_2$ is light, and also the example is closer to the center than $p_3$ $p_1$ and $p_2$, then d is positive with a confidence factor = 0.30.

In step 3, there are examples of updates that equivalence class $X_1$ belong to group $G_3$, but the equivalence class with $c_1$ and $c_2$ are closer to the center of the group $G_1$. When the upgrade percentage of these examples is significant, we divided the group $G_3$ with the equation of hyperplane between the center of the group $G_3$ and the $G_1$. Thus, $G_3$ is divided into two subsets, each of which generates a certain rule since most of the examples are from a single class.

## 6. EXPERIMENTAL RESULTS

At this point, there will be a comparison between the methods of knowledge extraction. This will work with real data sets, very different from the UCI repository of databases for machine learning [22]. These sets are widely known and used in this area of knowledge, so that will deepen their description: "Iris" is the most simple, "BUPA" and

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

"Diabetes" are also simple sets but not as linearly separable "Glass "and" Carp " are less linear and more complex.

Each data set is divided into training set (2/3) and test set (1/3). In rough set divisible, 2/3 of the first form is used for initial training (step 2) and the rest is updated (step 3). In each set, parameters are chosen by trial and error, showed in Table 2.

For analyzing rough set, it is necessary for discretization of continuous variables in the first step. The modified Chi2 [23] with a variation, will lower the initial level of consistency , and more examples are needed to obtain more representative equivalence classes.

*Table 2. studied data sets and hit rates*

| Sets | Features | | | Parameters | | | Hit Rate | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Examples | Attributes | Classes | Lc final | n clusters | Update % | F1 | F2 | F3 | VPRS |
| Iris | 151 | 4 | 3 | 0.75 | 3 | 10% | 92.00% | 94.00% | 94.00% | 92.00% |
| BUP | 344 | 6 | 2 | 0.58 | 4 | 10% | 48.68% | 65.22% | 66.96% | 54.78% |
| Carp | 4177 | 4 | 9 | 0.53 | 8 | 15% | 54.66% | 54.60% | 54.74% | 55.68% |
| Glass | 213 | 7 | 6 | 0.40 | 3 | 20% | 56.35% | 67.61% | 69,01% | 57.75% |
| Diabetes | 768 | 7 | 2 | 0.66 | 4 | 20% | 55.47% | 67.58% | 67,98% | 59.77% |

*Table 3. Hit rates of different methods for knowledge extraction*

| Sets | Hit Rate | | | |
|---|---|---|---|---|
| | ID3 | CART | Multilayer Perceptron | RBFN |
| Iris | 94.00% | 94.00% | 96.00% | 96.00% |
| BUP | 60.86% | 64.36% | 70.42% | 60.01% |
| Carp | 55.03% | 54.24% | 56.03% | 5 5.75% |
| Glass | 67.61% | 70.41% | 54.94% | 54.93% |
| Diabetes | 74.21% | 69.52% | 73.05% | 65.63% |

As shown in Table 2, the proposed improved rough set combined with agglomerative clustering provides a progressive increase in the hit rate. This is mainly due to:

1. The examples of tests that do not belong to any equivalence class of forms in the training are awarded the group's decision included in the positive region with the nearest center. This can be done from the Phase 2 of the initial knowledge, when the centers are calculated.

2.A greater number of examples fall into certain rules (many in the following generated equivalence classes), which can be assigned with greater certainty than a decision as examples as true.

In Table 3, a comparison of the results is obtained for the studied data sets when using other methods of knowledge extraction following the different rough set strategies.

On the one hand, there are two types of decision trees (ID3 and CART ) and the other two types of neural networks (multilayer perceptron and radial basis networks [24]). In most cases, these techniques exceed the results obtained by the variable precision rough set, while observing the similar results of those rough set methods.

## 7. CONCLUSION

This paper proposes an improved rough set combined with agglomerative clustering. The concept of equivalence class was also incorporated to merge and divide subclass. This technique seeks to obtain certain rules from the uncertain rules made by the method of rough set. Thus, the number of new examples increases to be assigned to a considered decision. It has also allowed the assignment of a class to the examples that fall in uncertain rules or do not fall into any rules, because any such training was similar to them.

The technique requires the choice of several parameters such as the level of consistency achieved in the discretization, the final number of clusters up to each equivalence class or the minimum percentage of renovation examples

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

required to proceed to the division of an equivalence class. The values assigned to these parameters should be suitable for the instant case, considering the poor selection can cause the generation of rules less significant.

Finally, as a future extension to the presented method suggests the inclusion of neural networks to replace the perceptron RBFN used in the phase of updating knowledge.

## ACKNOWLEDGMENT

This paper is supported by National Natural Science Foundation of China (60273043), Natural Science Foundation of Anhui Province (090412054)and Natural Science Foundation of Anhui Higher Education Institutions (KJ2011Z020)

## REFRENCES:

[1] Pawlak, Z., "Rough Sets", *International Journal of Computer and Information Sciences*, vol. 11, No. 5, 1982, pp. 341-356.

[2] Pawlak, Z., "Rough Sets: Theoretical Aspects of Reasoning About Data", Dordrecht, Netherlands: Kluwer Academic, 1991.

[3] Pawlak, Z., Grzymala-Busse, J., Slowinski, R., Ziarko, W. "Rough Sets", *Communications of the ACM*, Vol. 38, No. 11, 1995, pp. 89-95.

[4] Slowinski, R., "Intelligent Decision Support. Handbook of Applications and Advances of the Rough Sets Theory", Kluwer Academic Publishers, Boston, 1992.

[5] Tsumoto, S., Tanaka, H., "Automated Discovery of Medical Expert System Rules from Clinical Databases Based on Rough Sets", *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 1996, pp.63-69.

[6] Nowicki, R., Slowinski, R., Stefanowski, J., "Evaluation of Vibroacoustic Diagnostic Symptoms by means of the Rough Sets Theory", *Computers in Industry*, Vol. 20, 1992, pp.141-152.

[7] Ziarko, W., Golan, Edwards, D., "An Application of Datalogic/R Knowledge Discovery Tool to Identify Strong Predictive Rules in Stock Market Data", *Proceedings of AAAI Workshop on Knowledge Discovery in Databases*, Washington, DC, 1993, pp.93-101.

[8] Malcolm J. Beynon, "Stability of continuous value discretisation: an application within rough set theory", *International Journal of Approximate Reasoning*, Vol. 35, No.1, pp.29-53.

[9] Mazlack, J., et al., "A rough set approach in choosing clustering attributes", *Proceedings of the ISCA 13th International Conference*, 2000, pp.1-6

[10] Z. Huang, "Extensions to the K-means algorithm for clustering large data sets with categorical values", *Data Mining and Knowledge Discovery*, Vol.2, No.3, pp. 283-304

[11] D. Gibson, J. Kleinberg, P. Raghavan, "clustering categorical data: an approach based on dynamical systems", *International Journal of Very Large Data Bases*, Vol. 8, No.3-4, 2000, pp.222-236

[12] S. Guha, R.Rastogi, K. Shim, "ROCK: a robust clustering algorithm for categorical attributes", *Proceedings of International Conference on Data Engineering*,1999, pp.512-521

[13] V. Ganti, J. Gehrke, R. Ramakrishman, "CACTUS-clustering categorical data using summaries", *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*,1999, pp.73-83

[14] A. Dempster, N. raird, D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm", *Journal of the Royal Statistical Society*, Vol. 39, No.1,1977, pp. 1-38

[15] Shen, L., Tay, F.E.H., "A discretization method for Rough Sets Theory", *Intelligent Data Analysis*, Vol. 5, No.5, 2001, pp. 431-438.

[16] Na Chen, Shaopu Yang, Cunzhi Pan and Erfu Guo, "Research on Discretization Algorithm of Continuous Attributes Based on PCNN in a Bridge Erecting Machine Safety Monitoring System", *Applied information and communication*, Vol.225, part 2, 2011, pp.708-714

[17] I.T.R.Yanto, "Applying variable precision rough set model for clustering student suffering study's anxiety", *Expert Systems with Applications*,Vol.39, No.1, 2012, pp.452-459

[18] E. Amigó et al., "A comparison of extrinsic clustering evaluation metrics based on formal constrains", *Information Retrieral*, Vol.12, No.4, 2009, pp.461-486

[19] J.T.Jeng et al., "Interval competitive agglomeration clustering algorithm", *Expert System with Applications*, Vol.37, No.9, 2010, pp.6567-6578

[20] C. J. Zhou, et al., "Discretizaiton of Attributes

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Based on Relative Positive Region of Decision Attribute", *Advanced Materials Research*, Vol. 271-273, 2011, pp.253-257

[21] T.P.Hong, "Learning cross-level certain and possible rules by rough sets", *Expert Systems with Applications*, Vol.34, No. 3, 2008, pp.1698-1706

[22] C.J Merz and P.M.Murphy, UCI repository of machine learning databases.http://www.ics.uci. Edu/~mlearn/MlRepository.

[23] Tay, F.E.H., "A modified Chi2 algorithm for discretization", *Knowledge and Data Engineering*, Vol.14, No.3, 2002, pp.666-670

[24] P. C. Kainen, V. Kurkora, M. Sanguineti, "Complexity of Gassian-radial-basis networks approximating smooth funciton", *Journal of Complexity*, Vol. 25, No.1, 2009, pp.63-74.