Proceedings of 2013 International Conference on
Technology Innovation and Industrial Management
29-31 May 2013, Phuket, Thailand

RULE GENERATION BASED ON NOVEL TWO-STAGE MODEL

Kuo-Ping Lin, Department of Information Management, Lunghwa University of Science and Technology, Taoyuan 333, Taiwan
Email: kplin@mail.lhu.edu.tw

Ching-Lin Lin, Department of Information Management, Lunghwa University of Science and Technology, Taoyuan 333, Taiwan

Yu-Ming Lu, Department of Information Management, Lunghwa University of Science and Technology, Taoyuan 333, Taiwan

Ping-Feng Pai, Department of Information Management, National Chi Nan University,University Rd., Puli, Nantou 545, Taiwan

ABSTRACT

Purpose *– The purpose of this paper is to develop a novel two-stage model for promoting the effect of rule generation based on rough set. In order to improve traditional rough set method, the novel two-stage model adopts new kernel intuitionistic fuzzy clustering (KIFCM) to promote performance of rough set theory. Moreover, the e-learning customer data set in Taiwan is also examined for demonstrate the effectiveness and practicality of model.*

Design/methodology/approach *– In this paper, the authors present a new kernel intuitionistic fuzzy rough set model which combines novel KIFCM with rough set. The rule generation can divide to two stages for effective rule generation. In the first stage, KIFCM can utilize the advantages of kernel function and intuitionistic fuzzy sets to cluster raw data into similarity groups. In the second stage, the rough set theory is employed to generate rules with different groups. Finally, based on decision rules of rough set with different groups the results of system can be obtained and analyzed for users.*

Findings *– The novel rule generation model adopts pre-process, which is KIFCM clustering technique, can effectively assist traditional rough set in promoting the performance. In analysis of e-learning data set, the empirical result indicates that proposed novel rule generation model can outperform traditional decision models.*

Practical implications *–This novel two-stage model can provide a new and effective technique for data mining, database system, …, etc. Furthermore, in the research, proposed model also practically was applied to analyze and model customer's tendency in e-learning platform with proper decision rules.*

Originality/value *– The rough set theory has widely used in dealing with data mining and classification problems. This research proposes a construct of novel two-stage model which can effectively improve traditional rough set theory by using KIFCM clustering technology in the first stage. Real e-learning data set also is employed for demonstrate the effectiveness and practical. Based on the empirical result, the novel two-stage model can be evidenced that can actually apply in real information platform.*

Keywords: Rule generation, rough set, kernel intuitionistic fuzzy clustering, e-learning

INTRODUCTION

Rough set theory (RST) (Pawlak, 1982; Pawlak, 1991) is useful in analysis of inconsistent decision tables composed of attribute value data about many objects. The RST is a mathematical tool to deal with vagueness and uncertainty. In RST we can extract the minimal attribute sets without deterioration of quality of approximation, and minimal length decision rules corresponding to lower or upper approximation. This approach has been one of fundamental important in artificial intelligent (AI) especially in the fields of machine learning, knowledge acquisition, decision support system, medical information, and pattern recognition. Table 1 summarized some developments of rough set-based rule generation technology since 2000.

Beynon et al. (2000) used variable precision rough set (VPRS), which suggested by Ziarko (1993a, 1993b), to analyze expenditure on elementary and secondary education at the US state level. The standard linear discriminant analysis (LDA) also was employed in the same example. The result show the VPRS has good performance in classification accuracy index and indicates that rules based on rough set may be evaluated in various ways, some of which are analogous to certainty factors in standard rule-based method. Khoo and Zhai (2001) proposed an integrated prototype system (RClass-Plus) that combines rough set theory, genetic algorithms and Boolean algebra, for inductive learning. The prototype system was validated using the state of machine case that result shows the RClass-Plus has better capabilities, which are respectively dealing with uncertainty and inconsistency, simple and concise rules induction, and extracting complete rules, than iterative dichotomiser 3 (ID3) (Quinlan, 1986), learning from examples based on rough sets (LERS) (Grzymala-Busse, 1992), and rough-set-based approach for classification (RClass) (Khoo et al., 1999). Mak and Munakata (2002) reviewed and compared the rule extraction capabilities of rough sets with neural networks and ID3. This research indicated the rough set method offers much better explanatory capability than the neural network method and distills the data into a set of simple and usable rules. In new Product entry decision case the result also shows the rough set can obtain better performance in predictive accuracy index than neural networks and ID3. Pal et al. (2003) proposed a Rough-fuzzy multilayer perceptron (Rough-fuzzy MLP) with modular concept using a genetic algorithm to obtain a structure network suitable for both classification and rule extraction. Four compared methods, which are *Subset* method (Fu, 1993), *M* of *N* method (Towell and Shavlik, 1993), X2R (Liu and Tan, 1995), and decision tree-based method C4.5 (Ross Quinlan, 1993), were also examined in vowel, pat, medical data sets. The effectiveness of the Rough-fuzzy MLP model is extensively demonstrated in these data sets. Hou and Huang (2004) used fuzzy variable precision rough set approach to manufacturing process. The induced rules are compared with the rules by the traditional rough set. The fuzzy rough set approach, being less sensitive to noisy data, induces better rules than traditional rough set approach. Tseng et al. (2005) used rough set to predicting quality characteristic in machining operations. The results indicate a higher accuracy over traditional the tradition multi-nominal regression and general discriminant methods. Pattaraintakorn et al. (2006) proposed an ordinal prediction based on rough sets and soft computing which illustrates the formulation of more meaningful rules using the notion of ordinal prediction. The results are the rules that are constructed from the interval antecedents and are able to predict intervals rather than unique values of the target function. Wang et al. (2006) used rough set attribute reduction algorithm that employed particle swarm optimization (PSO) to find minimal rough set reducts, and compared with neural networks, decision trees and a fuzzy rule extraction algorithm based on Fuzzy Min-Max Neural Networks (FRE-FMMNN). The result shows that reducts found by rough set attribute reduction algorithm were more efficient and generated decision rules with better classification performance. Inuigchi and Miyajima (2007) proposed a rule induction based on rough set from two decision tables. The research pointed out decision rule is positively supported by a decision table and does not have any conflict with the other decision table and a second level decision rule is positively supported by both decision tables. Hong et al. (2007) proposed a hybrid fuzzy rough set system using fuzzy lower and upper approximations transform each quantitative value into a fuzzy set of linguistic term. The certain and possible rules are then generated based on these fuzzy approximations. Teoh et al. (2008) proposed a hybrid fuzzy time series model for stock index forecasting based on cumulative probability distribution approach (CPDA) and rough set rule induction. Two empirical stock markets (TAIEX and NYSE) were used as evaluating database, and two methodologies (Chen, 1996; Yu, 2005) are used as comparison models. The result shows hybrid fuzzy time series model has outperformance in the two examples. Qian et al. (2008) proposed a rule extracting algorithm based on the converse approximation which can be named REBCA is designed to extract decision rules from a decision table. Fan et al. (2009) proposed an incremental rule-extraction algorithm based on the previous rule extraction to resolve there aforementioned issues, and successfully applied in lager database. Ma et al. (2009) proposed a classification algorithm named Rule Generation based on Classification Attribute (RGCA) to deduct negative and positive rules. The results show the classification accuracy of RGCA algorithm is better than traditional positive based algorithm in examples of Machine learning databases. Luo and Zhong (2010) used rough set data mining method for knowledge acquisition and decision-making reasoning method, and applied in Medical Expert Diagnosis System (MEDS). Othman et al. (2011) developed a data mining approach of integrated-rough-set-and-genetic-algorithm based rule discover from digital protective relay's resident event report. This integrated approach has been proven to be an exact manifestation of real operation characteristics hidden in the event report, and has better performance than decision tree based analysis. Shi et al. (2012) adopted rough sets theory to reduce features and associate rule mining in KANSEI Engineering. The case of mobile phone was examined and indicated that rough sets theory can obtain better performance than orthogonal method. Huang et al. (2013) proposed incremental rough set theory based rule induction and applied to customer preferred estimation in the cell phone purchase.

Observation of past research of RST can be concluded some phenomenons (1) Recently RST can effectively apply in various areas. (2) RST can obtain better performance than some popular method such as decision tree (ID3 and C4.5), and neural networks. (3) Some researches proposed hybrid RST-based methods which can successfully improve performance of RST.

Table 1. Some developments of rough set-based rule generation technology since 2000.

Author(s)	Year	Technology	Applied problem	No. of condition attribute	Compared method(s)
Beynon et al.	2000	VPRS	Expenditure on elementary and secondary education	8	LDA
Khoo and Zhai	2001	RClass-Plus	State of machine	3	ID3, LERS, and RClass
Mak and Munakata	2002	Rough set	New Product entry decision	3	ID3, and neural networks
Pal et al.	2003	Rough-fuzzy MLP	Vowel/ Pat/ Medical	6/ 2/ 11	Subset method, M of N method, X2R, and C4.5
Hou and Huang	2004	Fuzzy rough set	Manufacturing process	8	Rough set
Tseng et al.	2005	Rough set	Machining operations	8	Multi-nominal regression, and general discriminant
Pattaraintakorn et al.	2006	Rough set with notion of ordinal prediction	Melanoma data set	7	—
Wang et al.	2006	Rough set	Brain glioma	14	Neural networks, decision trees and FRE-FMMNN
Inuigchi and Miyajima	2007	Rough set	Evaluation of television set	3	—
Hong et al.	2007	Rough set and fuzzy set theory	Fisher’s Iris Data	4	—
Teoh et al.	2008	Rough set	Trading stock	1/1	Chen’s method

Author(s)	Year	Technology	Applied problem	No. of condition attribute	Compared method(s)
		and CPDA	index/trading database		(1996) and Yu's method (2005)
Qian et al.	2008	REBCA	Numerical example	4	—
Fan et al.	2009	Incremental rough set	Information system	4	—
Ma et al.	2009	RGCA	Lens/ Balloons/ Iris/ Wine (Machine learning databases)	6/4/4/4	—
Luo and Zhong	2010	Rough set	MEDS	9	—
Othman et al.	2011	Rough set and genetic algorithm	Power protection system maintenance	30	Decision tree
Shi et al.	2012	Rough set	KANSEI Engineering	9	Orthogonal method
Huang et al.	2013	Rough set	Customer preferred estimation	4	—

—: means non.

This paper therefore aims to develop a novel two-stage model for promoting the effect of rule generation based on novel KIFCM and RST because accurately generating decision rules can effectively discover and construct domain's knowledge. In the first stage, clustering technology is utilized to classify the pattern datasets. This research develops a novel KIFCM which can be evidenced obtains good performance in numerical example. The results of this novel clustering can roughly class similar pattern into the same group. This stage can effectively reduce complexity of pattern datasets. Based on the results of stage one, the RST, which has been widely applied in the rule generation, is employed as the main rule generation technology in stage 2. With different groups, RST can further obtain better performance. Furthermore, this research adopts two types numerical example to examine proposed novel two-stage model.

The rest of this paper is organized as follows. Section 2 presents the main construct of novel two-stage model for rule generation, and introduces novel KIFCM and RST respectively. Section 3 depicts machine learning datasets and real e-learning case, experimental results, and compares with various models. Section 4 presents the research summary and conclusions.

NOVEL TWO-STAGE MODEL

The design construct of novel **two-stage** model. This model main adopts kernel intuitionistic fuzzy clustering and rough set in stages 1 and 2, respectively. Firstly, KIFCM can utilize the advantages of kernel function and intuitionistic fuzzy sets to cluster raw data into similarity groups. Traditional fuzzy clustering techniques could not effectively solve multi-variables or nonlinear problems. Kernel-based methods have attracted great attention and have been applied in many fields (Vapnic, 1998; Graves and Pedrycz, 2010; Lin et al., 2011; Zhang and Chen, 2004). The kernel-based method involves performing an arbitrary nonlinear mapping data from the original dimensional feature space to a higher dimensional space, called kernel space. In the kernel space the original data may apply classifiers. The use of kernels has received considerable attention because kernels make it possible to map data onto a high-dimensional feature space in order to increase the representation capability of linear machines. Moreover, KIFCM incorporates another uncertainty which is the hesitation degree that arises while defining the membership function. When hesitation degrees are used for determining precisely membership function because Yager class of IFSs (hesitation degrees) can assist to amplify differentiation of clusters, this is done to maximize the good points in the class. The KIFCM can obtain the advantages of both.

Secondly, the rough set theory is employed to generate rules with different groups. Rough set is a powerful tool to solve vagueness and uncertainty problems, which is suitable to analyze quantitative attributes and qualitative ones. A principle task, in the rough set of rule generation, is to compute reducing relative to a particular kind of information system, the decision system. Finally, based on decision rules of rough set with different groups the results of system can be obtained and analyzed for users. The illustration of novel kernel intuitionistic fuzzy rough set model is shown in Figure 1:

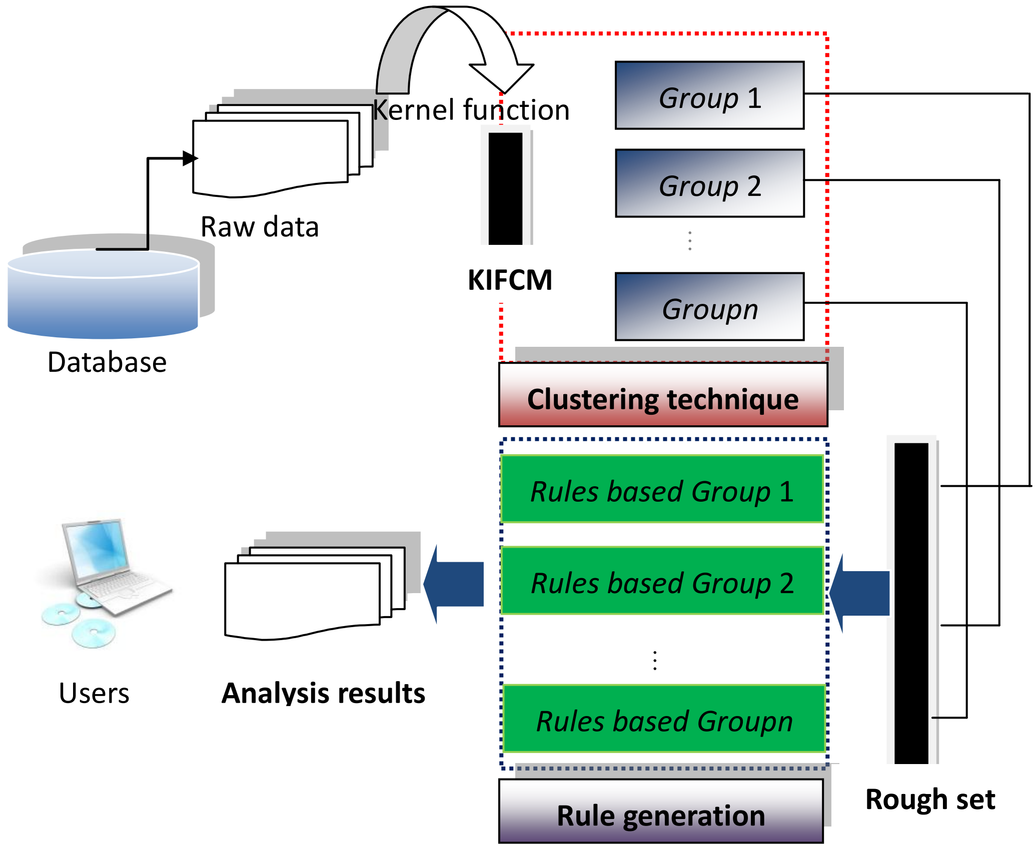


Figure 1. The structure of two-stage model for rule generation and extraction

The procedures can be described as follows.

Step 1: The data of selected variables can be obtained from a database. These selected data are based on expert opinion.

Step 2: (Stage one) In the first stage, the fuzzy clustering methods: novel KIFCM can be employed, help to define similar pattern into groups. These groups approximately represent the pattern that may have common variables and similarity characters; based on the groups, the pattern datasets can reduce interactive complexity.

Step 3: (Stage two) In the second stage, the two-stage model uses the RST to rule generation in various groups. The RST model can generate and discover decision rules in various groups, and subsequently provide more creditable rules and analysis.

Step 4: This research adopts three measuring index. Firstly, accuracy rate (J_1) can be introduced as following:

$$J_1 = 100 \times \frac{N_{correct}}{Data_{total}} \quad (1)$$

where $N_{correct}$ is number of correct classification of data, and $Data_{total}$ is total number of data. Second, coverage rate (J_2), and a comprehensive index (J), which is equal to $J_1 + J_2$, to measure the performance. The coverage rate can be formulated as following:

$$J_2 = \frac{N_{coverage}}{Data_{total}}, \quad (2)$$

where $N_{coverage}$ is number of coverage data by generating rules.

1. Kernel intuitionistic fuzzy *c*-means clustering algorithm

Clustering analysis involves the discovery of a data structure and partitions a data set into a number of subsets with correlated data. Clustering was widely applied in several fields, such as taxonomy, geology, business, engineering systems, medicine, and image processing (for example, Bezdek, 1981; Yang, 1993; Honda and Ichihashi, 2004; Kohonen, 1997; Liu and Wang, 2007). FCM is one of the most widely employed techniques. In FCM and fuzzy clustering concepts, the membership function of clusters can be defined based on a distance function, and thus, the degrees of memberships may express the proximity of the data to the multi-cluster centers. The proposed KIFCM considers the kernel function and IFSs that the KIFCM attempts to maximize the good points in the kernel space. The procedures of KIFCM can be described as following:

***Step*1 (Input parameters)**: The input parameters are number of clusters to classify (c), parameter in updating the clustering membership functions (m), parameter of Yager's intuitionistic fuzzy complement (α), number of epochs to carry out (k), parameter of Gaussian kernel (σ), and tolerance for the solution accuracy (δ) respectively.

***Step* 2 (Generating initial membership and center):** By the KIFCM, the initial memberships of the data $\mathbf{x}_i$, i = 1, …, N, with the crisp input-output to the clusters j (j = 1, …, c) are generated randomly and denoted as $U(k) = [\mu_{ij}^{(k)}]_{N\times c}$ (k = 0) under $\sum_{\forall j} \mu_{ij}^{(k)} = 1 \forall i$. By generating randomly memberships of the data in KIFCM, the center $C_j^{(k)}$ can be easily calculated.

***Step* 3 (Calculating objective function)** : With a data set $\boldsymbol{x}_i$, KIFCM, the modified objective function of the fuzzy c-means using the kernel space and IFSs need to be minimized. Therefore, KIFCM minimizes the following designed objective function:

$$J_{KIFCM} = \sum_{j=1}^{c}\sum_{i=1}^{N}\left(\mu_{ij}^{(k)}\right)^m \times \left\|\varphi(\boldsymbol{x}_i), \varphi(C_j^{(k)})\right\|^2 \quad (3)$$

where $\left\|\varphi(\boldsymbol{x}_i), \varphi(C_j^{(k)})\right\|^2$ is the Euclidean distance between $\varphi(\boldsymbol{x}_i)$ and $\varphi(C_j^{(k)})$, and $\varphi(\boldsymbol{x}_i)$ and $\varphi(C_j^{(k)})$ are the kernel spaces of $\boldsymbol{x}_i$and $C_j^{(k)}$, respectively. The squared distance is computed in the kernel space using a kernel function, as follows (Shen et al., 2006; Zhang and Chen, 2003):

$$\left\|\varphi(\boldsymbol{x}_i), \varphi(C_j^{(k)})\right\|^2 = K(\boldsymbol{x}_i, \boldsymbol{x}_i) + K(C_j^{(k)}, C_j^{(k)}) - 2K(\boldsymbol{x}_i, C_j^{(k)}) \quad (4)$$

In previous research, the Gaussian kernel was employed almost exclusively because the kernel function can obtain better performance with a Gaussian kernel function (Graves and Pedrycz, 2010). The Gaussian kernel can be expressed as $K(x_i, x_j) = \exp(-\left\|x_i - x_j\right\|_2^2 / \sigma^2)$. Therefore, with a Gaussian kernel function, the designed objective function can be written as

$$J_{KIFCM} = 2\sum_{j=1}^{c}\sum_{i=1}^{N}\left(\mu_{ij}^{(k)}\right)^m \times (1 - K(\boldsymbol{x}_i, C_j^{(k)})) \quad (5)$$

***Step* 4 (Updating membership and center)** : The optimization of the partition matrix U involves the use of Lagrangian multipliers, leading to the expression

$$\mu_{ij}^{(k+1)} = \frac{(1/(1-K(\boldsymbol{x}_i, C_j^{(k+1)})))^{(1/(m-1))}}{\sum_{h=1}^{c}(1/(1-K(\boldsymbol{x}_i, C_h^{(k+1)})))^{(1/(m-1))}} \quad (6)$$

To incorporate intuitionistic fuzzy properties in the kernel fuzzy c-means algorithm, the hesitation degree is calculated initially by $\pi_A(x) = 1 - \mu_A(x) - (1-\mu_A(x)^{\alpha})^{1/\alpha}$., and the intuitionistic fuzzy membership values are obtained as follows:

$$\mu_{ij}^{(l+1)*} = \mu_{ij}^{(l+1)} + \pi_{ij}^{(l+1)} \tag{7}$$

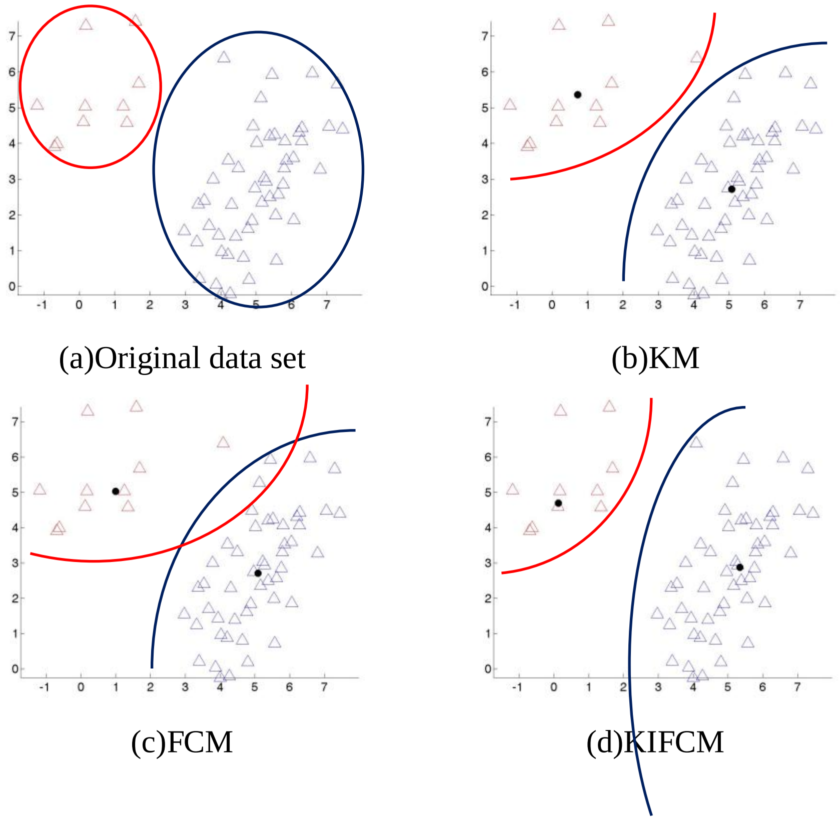
where $\mu_{ij}^{(l+1)*}$ denotes the intuitionistic kernel fuzzy membership of the *i*th data in the *j*th class.

By Eq.(7) in KIFCM, the prototypes $C_j^{(k+1)*}$ with the Gaussian kernel function can be written as

$$C_j^{(k+1)*} = \frac{\sum_{i=1}^{N} \mu_{ij}^{(k+1)} K(x_i, C_j^{(k+1)}) x_i}{\sum_{i=1}^{N} \mu_{ij}^{(k+1)} K(x_i, C_j^{(k+1)})} \tag{8}$$

***Step* 5 (Terminational condition)** : The cluster center is updated simultaneously with the U matrix. As in traditional FCM, KIFCM optimizes the objective function by continuously updating the membership functions and centers of clusters until $\left\|U^{(k+1)} - U^{(k)}\right\|$ is less than the threshold δ.

A numerical example is examined in the section for comparison of various clustering algorithms. In numerical example, the study generated 10 random points for *Cluster* 1 and 50 random points for C*luster* 2. The *Cluster* 1 multivariate normal distribution with mean vector was [1 5], and the covariance matrix sigma was $\begin{bmatrix} 1.500 & 0.866 \\ 0.866 & 2.500 \end{bmatrix}$. The *Cluster* 2 multivariate normal distribution with mean vector was [5 3], and the covariance matrix sigma was identical to that of *Cluster* 1. Figure 2(a) shows the original random data points of example. The remainder of Figure 2 displays the cluster results of KM, FCM, and KIFCM with the cluster centers displayed in Figure 2. Figure 2 also labels key data points by the red cycle in example. Table 2 shows the comparison of the accurate rate and center points of KM, FCM, and KIFCM. KIFCM obtain the best accurate rates which is equal to 1. The KIFCM algorithms could correctly class the data set. The cluster centers C_1 and C_2 of KIFCM algorithm can be searched lower and more right than traditional KM and FCM. These phenomena may be evidence that the results of example can obtain better performance when using kernel functions and IFs techniques.



(a)Original data set (b)KM

(c)FCM (d)KIFCM

Figure 2. Illustration of the original data set and results of examplewith various clustering algorithms.

Table 3. Comparison of various clustering algorithms in synthesis data set.

Method	Parameters	Accuracy rate (J_1)	Center points
KM	c=2	0.98	C_1=(0.71, 5.36), C_2=(5.06, 2.71)
FCM	c=2, m=2	0.98	C_1=(0.98, 5.03), C_2=(5.09, 2.70)
KIFCM	c=2, m=2.33, σ=165.76, α= 4.95	1.00	C_1=(0.12, 4.68), C_2=(5.33, 2.87)

2. Rough set theory for rule generation with KIFCM clustering

Rough set theoryis a mathematical approach for handling vagueness and uncertainty in data analysis. Rough sets aim at forming an approximate definition for a target set in terms of some definable sets, especially, when the target set is uncertain or imprecise. A rough set is characterized by a pair of precise concepts, called lower and upper approximations, generated using object indiscernibilities. Some basic concepts in the rough set theory can be found in Pawlak (1982) and Pawlak (1991). In RST, the most important problems are the reduction of attributes and the generation of decision rules. In rough set theory of rule generation, principal task is to compute reducts relative to a particular kind of information system. Instead the lower and upper approximations of all decision concepts are computed and rules are induced. The rules are categorized into certain and approximate (possible) rules depending on the lower and upper approximations, respectively. This RST of rule generation can be described below.

Let $E=\langle U, A\rangle$ be a decision table with F and $D=\{d_1, d_2, \ldots, d_l\}$ its sets of condition and decision attributes respectively, where U is a nonempty finite set called the universe and A is a nonempty finite set of attributes. An attribute a can be regarded as a function from domain U to some value set V_a. Divided the decision table $E = \langle U, A\rangle$ into l tables $E_i = \langle U_i, A_i\rangle$, i=1, …,l, corresponding to the l decision attributes $d_1, \ldots, d_l$, where $U = U_1 \cup \ldots \cup U_l$ and $A_i = F\cup\{d_i\}$ where $d\notin F$ is the decision variable D.

Let $\{x_{i1},\ldots, x_{ij}\}$ be the set of those objects of U_i that occur in E_i, i=1,…,l, and in this research the set, which can be defined $X_c = \{x_{ci1},\ldots, x_{cij}\}$, has been divided c clusters by novel KIFCM. Based on results of KIFCM, the same group objects of X_c have similar characters which can effectively assistance rule generation of RST.

***Step* 1. (Determining indiscernibility relation)** Let $a\in A$, $P\subseteq A$. A binary relation I_P, which can be called the indiscernibility relation, can be defined as following:

$I_P = \{(x, y) \in U$: for every $a \in P$, $a(x) = a(y)\}$.

Then, $I_P = \cap_{a\in P} I_a$. If $X_c \subseteq U$, the set $\{x\in U: [x]_P \subseteq X_c\}$ and $\{x\in U: [x]_P \cap X_c \neq \emptyset\}$, where $[x]_P$ denotes the equivalence class of the object $x\in U$ relative to I_P, are called *P-lower* and *P-upper* approximation of X_c in E and denote by $\underline{I}X_c$ and $\overline{I}X_c$ respectively. If X_c is *P-definable* then $\underline{I}X_c = \overline{I}X_c$; otherwise X_c is *P-rough*.

Example I: Consider the university of discourse $U = \{x_{c1}, x_{c2}, x_{c3}, x_{c4}, x_{c5}, x_{c6}\}$ and I is any equivalence relation in I_P which partitions U into $\{\{x_{c1}, x_{c2}, x_{c4}\}, \{x_{c3}, x_{c6}\}, \{x_{c5}\}\}$. The for any subset $X_c = \{x_{c1}, x_{c2}, x_{c3}, x_{c4}\}$ of U, $\underline{I}X_c = \{x_{c1}, x_{c2}, x_{c4}\}$ and $\overline{I}X_c = \{x_{c1}, x_{c2}, x_{c3}, x_{c4}, x_{c6}\}$. Figure

3 displays the lower and upper approximations of set X_c.

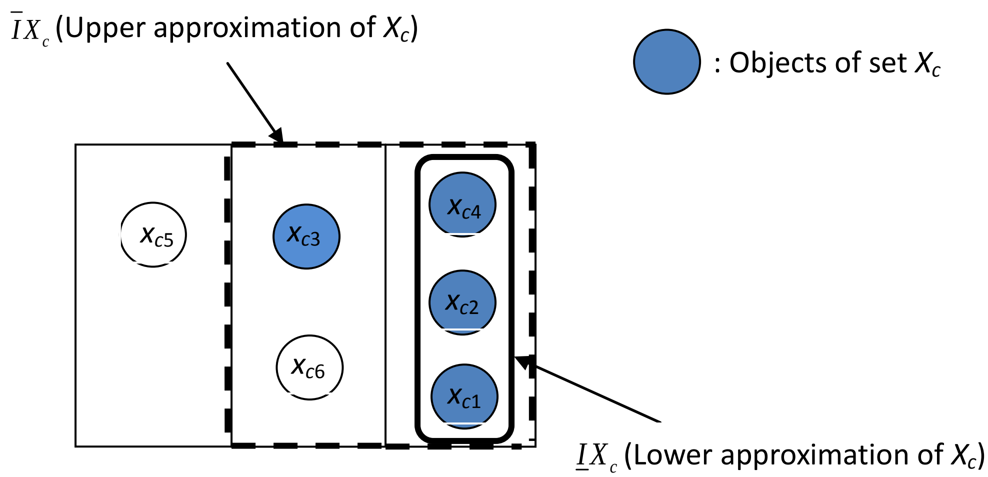


Figure 3 Illustration of the lower and upper approximations of set X_c. (Pawlak et al., 1995)

The *P-positive* region of X_c can be defined $\{x_{c1}, x_{c2}, x_{c4}\}$ and the *P-negative* region of X_c can be defined $\{x_{c5}\}$. On the other hand, consider a subset $Y_c = \{x_{c3}, x_{c5}, x_{c6}\}$, $\underline{I}Y_c = \{x_{c3}, x_{c5}, x_{c6}\}$ and $\overline{I}Y_c = \{x_{c3}, x_{c5}, x_{c6}\}$. The $\underline{I}Y_c = \overline{I}Y_c$, hence Y_c can be defined as *P-definable*.

***Step* 2. (Dispensable and indispensable attributes)** Let $f \in F$. A attribute f is dispensable in E, if $POS_{(F-(f))}(D) = POS_F(D)$; otherwise attribute f is indispensable in E. Where $POS_F(D)$ is $\bigcup_{X_c \in I_{d_i}} \underline{F}X_c$ and $\underline{F}X_c$ is the lower approximation. If all $f \in F$ are indispensable, f is an independent.

***Step* 3. (*Reduct* and *CORE*)** A *reduct* is the minimal attribute subset preserving the condition. If the E is independent and $POS_I(D)$, a set of attributes $I \subseteq F$ can be called a *reduct* of F. $CORE(F)$ can be defined the set of all attributes indispensable in F.

***Step* 4. (Discernibility matrix)** Now for each *reductP* = $\{p_1, \ldots, p_k\}$, a discernibility matrix $M_{d_i}(P)$ can be defined as follows (Banerjee et al., 1998):

$$f_{ij} = \{a \in P: a(x_{ci}) \neq a(x_{cj})\}, \qquad (9) \qquad \text{for } i, j = 1, \ldots, n.$$

For each object $x_{cj} \in x_{ci1}, \ldots, x_{cib}$, the discernibility function $k_{d_i}^{x_{cj}}$ can be defined as following:

$$k_{d_i}^{x_{cj}} = \bigcap\{\bigcup(f_{ij}) : 1 \le i, j \le n, j < i, f_{ij} \ne \varnothing\} \quad (10)$$

Where $\bigcup(f_{ij})$ is the disjunction of all members of c_{ij}. Then $k_{d_i}^{x_{cj}}$ is brought to its conjunctive norm form (c.n.f).

Example II (Ganesan et al., 2007) Consider the knowledge representation system given $F = \{a, b, c, d\}$ and $D = \{E\}$ which are condition and decision attributes respectively, and based on KIFCM the clustering results of objects $\{\{x_{11}, x_{12}, x_{13}, x_{14}\}, \{x_{25}, x_{26}, x_{27}\}\}$ can be obtained. The discernibility matrix can be shown as Figure 4.

	a	b	c	d	E
x_{11}	1	0	2	1	1
x_{12}	1	0	2	0	1
x_{13}	1	2	0	0	2
x_{14}	1	2	2	1	0
x_{25}	2	1	0	0	2
x_{26}	2	1	1	0	2
x_{27}	2	1	2	1	1

Group 1

Group 2

	x_{11}	x_{12}	x_{13}
x_{12}	—		
x_{13}	b, c, d	b, c	
x_{14}	b	b, d	c, d

	x_{25}	x_{26}
x_{26}	—	
x_{27}	c, d	c, d

Figure 4. Illustration of the discernibility matrix based on clustering technique KIFCM.

Here, the elements of discernibility matrix can be shown in "OR" which means the element $\{b, c, d\}$ can be shown $b \vee c \vee d$. Further, the entire matrix can be written by using the connective AND.

Two-stage model for e-learning data set

In this section a real dataset of enterprise was examined. The customer data set of e-learning comes from one of famous enterprise in Taiwan. Figure 5 displays the process of real information platform with novel two-stage model in E-learning enterprise. Six conditional attributes 'Sex', 'Age', 'Academic', 'Department', 'Location', and 'Occupation' were selected, and number of instances are 669. The all conditional attributes has been coded by valued for effective executing generation rules which can see in Table 7, for instance, minimum of 'Academic' is '2' which means 'National Tsing Hua University'. Table 4 shows the dataset extensively included various instances. Decision attributes can be divided six categories which are 'Civil service examinations', 'Matriculation examinations', 'Tying examinations', 'Certification examinations', 'Learning lifelong', and 'employment examinations' respectively. These determines of categories were based on database of E-learning enterprise.

Table 5 displays the measuring indexes of the KM+RST, FCM+RST and KIFCM+RST models with different *Cluster* values in E-learning dataset. The results show that the KIFCM+RST model with c=3 can derive higher comprehensive index values of 0.71 in E-learning dataset. From the sensitivity analysis provided Table 5, the KIFCM+RST model alsoagain demonstrates good, stable performance. Therefore, KIFCM+RST model can also be recommended as an alternative rule generation model in the E-learning dataset.

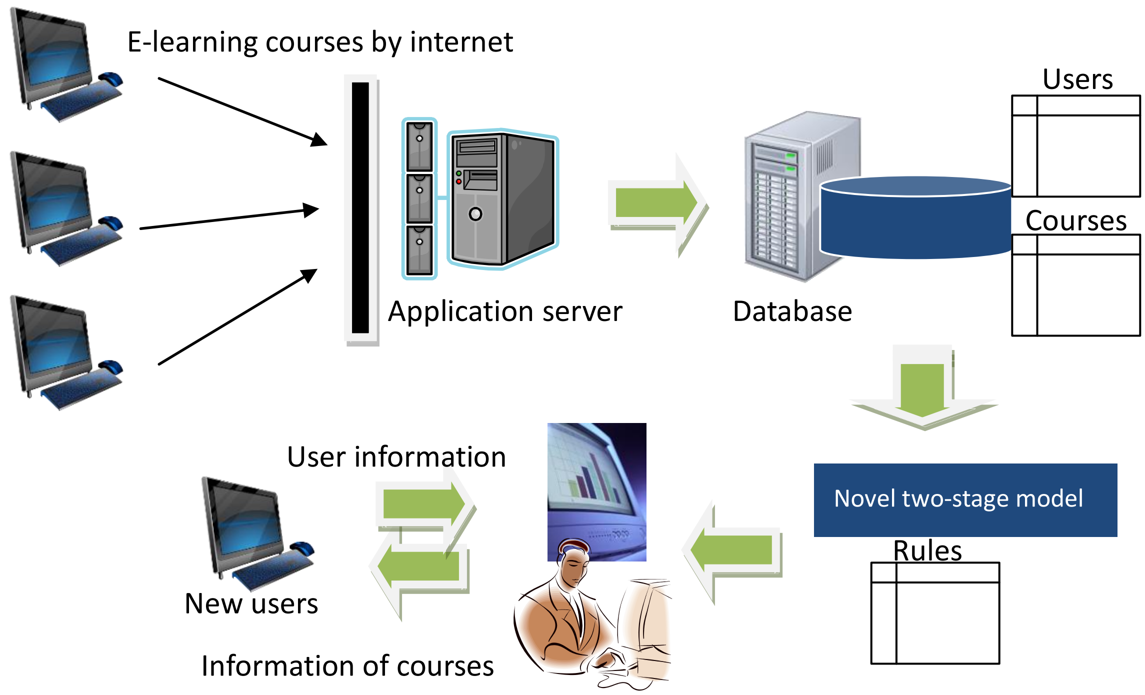


Figure 5. The process of real information platform with novel two-stage model in E-learning enterprise.

Table 4. Summary statistics and information of attributes with E-learning data sets.

Title	NO. of instances	Predictive attributes	Attributes Statistics				NO. of class
			Minimum	Maximum	Mean	Standard deviation	
E-learning	669	Sex	1: Male 2:Female				6
		Age	11	50	32.2	5.81	
		Academic	2	190	47.91	35.63	
		Department	1	178	99.31	49.81	
		Location	100	932	488.44	267.59	
		Occupation	1:Student, 2:Nonstudent				

Table 5. Testing measuring indexes of the KM+RST, FCM+RST and KIFCM+RST models with

different *Cluster* values in E-learning dataset.

Date set	Method: The 1st stage (KM, FCM, and KIFCM)+ The 2nd stage (Rough set)		Accuracy rate (J_1)	Cover rate (J_2)	Comprehensive index (J)	Average of comprehensive index	No. of rules	CPU time (Minute)
E-learning	KM	c=2	0.15	0.09	0.24	0.31	487	25.95
		c=3	0.18	0.20	0.38		486	21.68
		c=4	0.12	0.21	0.33		482	22.23
	FCM	c=2, m=2	0.16	0.10	0.26	0.31	487	27.85
		c=3, m=2	0.16	0.14	0.30		487	27.85
		c=4, m=2	0.18	0.21	0.39		483	24.05
	KIFCM	c=2, m=2, σ=150,α=11	0.33	0.31	0.64	0.68	519	45.29
		c=3, m=2, σ=150,α=11	0.33	0.38	0.71		522	54.40
		c=4, m=2, σ=150,α=11	0.32	0.38	0.70		535	111.59

Table 6 also shows the performances of the three models in the E-learning dataset. Proposed novel two-stage model can be proved has better performance than ID3 and standard RST with measuring indexes in the E-learning dataset. As results of Machine learning datasets proposed two-stage forecasting outperformed the other models in E-learning dataset, and the preprocess clustering mechanism (KIFCM) can effectively promote performance of RST.

Table 6. Comparison of the testing measuring indexes with various methods in E-learning dataset.

Date set	Method	Accuracy (J_1)	Cover rate (J_2)	Comprehensive index (J)	No. of rules	CPU time (Minute)
E-learning	ID3	0.14	0.17	0.31	104	1.65
	Rough set	0.16	0.19	0.35	486	27.51
	Two-stage model	0.33	0.38	0.71	522	90.98

CONCLUSIONS

This study developed a novel two-stage model for promoting the effect of rule generation based on rough set. The real e-learning customer dataset in Taiwan was examined. This real case can evidence the novel two-stage model can actually apply to real information platform. The success of the novel two-stage model can be attributed to the fact that the novel fuzzy clustering technology can effectively improve performance. Future studies may consider using various data preprocessing techniques to improve the performance of the proposed model.

ACKNOWLEDGEMENTS

The authors would like to thank the National Science Council of the Republic of China for financially supporting this research under Contract No. NSC 100-2410-H-262-002-MY2.

REFERENCES

1. Asuncion, A., and Newman, D.J. (2007), UCI Machine Learning Repository, University of California, School of Information and Computer Science, Irvine, CA, (Online) Available: (http://archive.ics.uci.edu/beta/).
2. Atanassov, K.T. (1986), “Intuitionistic fuzzy sets,” *Fuzzy Sets and Systems*, Nol. 20, pp. 87–96.
3. Atanassov, K.T. (1989), “More on intuitionistic fuzzy sets,” *Fuzzy Sets and Systems*, vol. 33, pp. 37–46.
4. Atanassov, K.T. (1999), Intuitionistic Fuzzy Sets: Theory and Applications, Heidelberg: Physica-Verlag.
5. Beynon, M., Curry, B., and Morgan, P. (2000), “Classification and rule induction using rough set theory”, *Expert System*, Vol. 17, pp. 136–147.
6. Bezdek, J.C. (1981), Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum

Press, New York.

7. Burillo, P., and Bustince, H. (1996), "Vague sets are intuitionistic fuzzy sets," *Fuzzy Sets and Systems*, Vol. 79, pp. 403–405.
8. Chen, S.M. (1996), "Forecasting enrollments based on fuzzy time series", Fuzzy Sets and Systems, Vol. 81, pp. 311−319.
9. Fan, Y.-N., Tseng, T.-L., Chern, C.-C., and Huang, C.-C. (2009), "Rule induction based on an incremental rough set", Expert Systems with Applications, Vol. 36, pp. 11439−11450.
10. Fisher, R.A. (1936), "The use of multiple measurements in taxonomic problems," *Annual Eugenics*, Vol. 7 Part II, pp. 179–188.
11. Fu, L.M. (1993), "Knowledge-based connectionism for revising domain theories", *IEEE Transaction on System, Man, and Cybernetics*, Vol. 7, pp. 173−182.
12. Ganesan, G., Latha, D., and Raghavendra Rao C. (2007), "Reduct generation in information systems", *Engineering Letters*, Vol. 14, pp. EL_14_2_5.
13. Graves, D., and Pedrycz, W. (2010), "Kernel-based fuzzy clustering and fuzzy clustering: A comparative experimental study," *Fuzzy Sets and Systems*, Vol. 161, pp. 522–543.
14. Grzymala-Busse, J.W. (1992), "LERS—a system for learning from examples based on rough sets, Intelligent Decision Support," in Slowinski, R. (Ed.), *Handbook of Applications and Advances of the Rough Sets Theory*, Kluwer, Dordrecht, Boston, London, pp. 3−18.
15. Haberman, S.J. (1976), "Generalized Residuals for Log-Linear Models," *Proceedings of the 9th International Biometrics Conference*, Boston, pp. 104–122.
16. Honda, K., and Ichihashi, H. (2004), "Linear fuzzy clustering techniques with missing values and their application to local principal component analysis", *IEEE Transaction on Fuzzy Systems*, Vol. 12, pp. 183–193.
17. Hong, T.-P., Wang T.-T., and Wang, S.-L. (2007), "Mining fuzzy β-certain and β-possible rules from quantitative data based on the variable precision rough-set model", *Expert Systems with Applications*, Vol. 32, pp. 223–232.
18. Hou, T.-H., and Huang, C.-C. (2004), "Application of fuzzy logic and variable precision rough set approach in a remote monitoring manufacturing process for diagnosis rule induction", *Journal of Intelligent Manufacturing*, Vol. 15, pp. 395−408.
19. Huang, C.C., Tseng, T.L., Fan, Y.N., Hsu, C.H. (2013), "Alternative rule induction methods based on incremental object using rough set theory", *Applied Soft Computing*, Vol. 13, pp. 372−389.
20. Inuiguchi, M., and Miyajima, T. (2007), "Rough set based rule induction from two decision tables", *European Journal of Operational Research*, Vol. 181, pp. 1540−1553.

21. Kohonen, T. (1997), Self-Organizing Maps, second ed. Springer-Verlag, Berlin.
22. Liu, H., and Tan, S.T. (1995), “X2R: A fast rule generator”, *Proceedings of 1995 IEEE International Conference on Systems, Man, and Cybernetics*, Vancouver, British Columbia, Canada, pp. 215–220.
23. Liu, X., and Wang, L. (2007), “Computing the maximum similarity bi-clusters of gene expression data”, *Bioinformatics*, Vol. 23, pp. 50–56.
24. Luo, H.C., and Zhong, Y.B. (2010), “Application and design of MEDS based on rough set data mining rule”, *Fuzzy Information and Engineering*, Vol. AISC 78, pp. 683–692.
25. Ma, T., Leong, J., Cui, M., and Tian, W. (2009), “Inducing positive and negative rules based on rough set”, *Information technology Journal*, Vol. 8, pp. 1039–1043.
26. Othman, M.L., Aris, I., Othman, M.R., and Osman, H. (2011), “Rough-Set-and-Genetic-Algorithm based data mining and Rule Quality Measure to hypothesize distance protective relay operation characteristics from relay event report”, *Electrical Power and Energy Systems*, Vol. 33, pp. 1437–1456.
27. Pal, S.K., Mitra, S., and Mitra, P. (2003), “Rough-fuzzy MLP: Modular evolution rule generation, and evaluation”, *IEEE Transaction on Knowledge and Data Engineering*, Vol. 15, pp. 14–25.
28. Pattaraintakorn, P., Cercone, N., and Naruedomkul, K. (2006), “Rule learning: Ordinal prediction based on rough sets and soft-computing ”, Applied Mathematics Letters, Vol. 19, pp. 1300–1307.
29. Pawlak, Z. (1982), “Rough sets”, *International Journal of Information Computer Science*, Vol. 11, pp. 341–356.
30. Pawlak, Z. (1991), Rough sets: Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, Dordrecht.
31. Pawlak, Z., Grzymala-Busse, J., Slowinski, R., and Ziarko, W. (1995), “Rough sets”, *Communications of the ACM*, Vol. 38, pp. 89–95.
32. Pham, D.T., and Aksoy, M.S. (1995), “A new algorithm for inductive learning”, *Journal of Systems Engineering*, Vol. 5, pp. 115–122.
33. Shen, H., Yang, J., Wang, S., and Liu, X. (2006), “Attribute weighted Mercer kernel based fuzzy clustering algorithm for general non-spherical datasets,” *Soft Computing*, Vol. 10, pp. 1061–1073.
34. Teoh, H.J., Cheng, C.-H., Chu, H.-H., and Chen, J.-S. (2008), “Fuzzy time series model based on probabilistic approach and rough set rule induction for empirical research in stock market”, *Data & Knowledge Engineering*, Vol. 67, pp. 103–117.

35. Tseng, T.-L., Kwon, Y., and Ertekin, Y.M. (2005), "Feature-based rule induction in machining operation using rough set theory for quality assurance", *Robotics and Computer-Integrated Manufacturing*, Vol. 21, pp. 559–567.
36. Qian, Y., Liang, J., and Dang, C. (2008), "Converse approximation and rule extraction from decision tables in rough set theory", *Computers and mathematics with Applications*, Vol. 55, pp. 1754–1765.
37. Quinlan, J.R. (1986), "Induction of Decision Trees",*Machine Learning*, Vol. 1, pp. 81–106.
38. Quinlan, J.R. (1993), C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, Inc., San Mateo, Calif.
39. Shi, F., Sun, S., Xu, J. (2012), "Employing rough sets and association rule mining in KANSEI knowledge extraction", *Information Sciences*, Vol. 196, pp. 118–128.
40. Towell, G.G., and Shavlik, J.W. (1993), "Extracting refined rules from knowledge-based neural networks", *Machine Learning*, Vol. 13, pp. 71–101.
41. Wang, X., Yang, J., Jensen, R., Liu, X. (2006), "Rough set feature selection and rule induction for prediction of malignancy degree in brain glioma", *Computer Methods and Programs in Biomedicine*, Vol. 83, pp. 147–156.
42. Yang, M.S. (1993), "A survey of fuzzy clustering", *Mathematical and Computer Modelling*, Vol. 18, pp. 1–16.
43. Yu, H.K. (2005), "Weighted fuzzy time series models for TAIEX forecasting", *Physica A*, Vol. 349, pp. 609–624.
44. Zadeh, L.A. (1965), "Fuzzy sets," *Information and Control*, Vol. 8, pp. 338–353.
45. Zhang, D.Q., and Chen, S.C. (2003), "Clustering incomplete data using kernel-based fuzzy C-means algorithm," *Neural Processing Letters*, Vol. 18, pp. 155–162.
46. Ziarko, W. (1993a), "Variable precision rough set model", *Journal of Computer and System Sciences*, Vol. 46, pp. 36–59.
47. Ziarko, W. (1993b), "Analysis of uncertain information in the framework of variable precision rough sets", *Foundations of Computing Decision Sciences*, Vol. 18, pp. 381–396.