International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:5, No:3, 2011

Heterogeneous Attribute Reduction in Noisy System based on a Generalized Neighborhood Rough Sets Model

Siyuan Jing, Kun She

***Abstract*—**Neighborhood Rough Sets (NRS) has been proven to be an efficient tool for heterogeneous attribute reduction. However, most of researches are focused on dealing with complete and noiseless data. Factually, most of the information systems are noisy, namely, filled with incomplete data and inconsistent data. In this paper, we introduce a generalized neighborhood rough sets model, called VPTNRS, to deal with the problem of heterogeneous attribute reduction in noisy system. We generalize classical NRS model with tolerance neighborhood relation and the probabilistic theory. Furthermore, we use the neighborhood dependency to evaluate the significance of a subset of heterogeneous attributes and construct a forward greedy algorithm for attribute reduction based on it. Experimental results show that the model is efficient to deal with noisy data.

***Keywords*—**attribute reduction, incomplete data, inconsistent data, tolerance neighborhood relation, rough sets

I. Introduction

ROUGH set theory, proposed by Pawlak [1], has been proven to be an efficient tool for feature selection, rule extraction and knowledge discovery from uncertain information. The basic idea of rough set theory is to classify objects of discourse, contained in a finite universe U , into equivalence classes with respect to some attributes. The objects in each class are indiscernible, and this indiscernible relation induces a partition of the universe into some blocks, called knowledge granules or elemental concepts. Any attribute of attribute set P can induce a partition Π_P of the universe. An arbitrary subset X of the universe U can be approximated by two sets $\{\underline{P}(X), \overline{P}(X)\}$, called the lower approximation and the upper approximation, respectively. If $\underline{P}(X) = \overline{P}(X)$, we say that X can be precisely approximated by knowledge Π_P , and the set is called a definable set; otherwise, we say X is a rough set. The approximation ability of an information system depends on the knowledge Π_P, the finer Π_P is, the more accurately objects can be approximated. This way of processing is consistent with the cognition of human.

Siyuan Jing is with the Department of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China (e-mail: jingsiyuan_628@126.com).

Kun She., is with the Department of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China (e-mail:kunshe@126.com).

As the classical rough sets model can just be used to evaluate categorical attributes, Hu generalized the classical model with neighborhood relations and proposed a neighborhood rough sets model [8]. This model can deal with the problems of heterogeneous attribute reduction, namely, categorical attribute and numerical attribute. As this research is mainly focused on the problem of heterogeneous attribute, the noise-tolerant ability, which is important in real world, is not strong.

In this paper, we propose a noise-tolerant neighborhood rough sets model by generalizing existing NRS model with a tolerance neighborhood relation and the probabilistic theory. The proposed model can induce a family of much more comprehensive information granules to characterize arbitrary concepts in complex universe. The properties of the model are discussed and some important theorems are also introduced and proven. Furthermore, we use the neighborhood dependency to evaluate the significance of a subset of heterogeneous attributes and construct a forward greedy algorithm for attribute reduction based on it. We compare the model with two popular noise-tolerant rough sets model, i.e. VQRS and fuzzy VPRS. Numerical experiments are presented and experimental results show that the proposed model is efficient to deal with noisy data.

II. Neighbor Theory

Neighbor theory proposed by T.Y. Lin in 1990 has been an important tool for many artificial intelligence tasks [4]. Yao and Wu extended the theory in 1998 and 2002 respectively [5], [6]. In this section, we review some knowledge of neighbor theory.

Definition 2.1: U is a non-empty finite set of objects, Δ is a given function. We say $NAS = (U, \Delta)$ is a neighbor approximation space where:

1). $\Delta(x_1, x_2) \geq 0, \Delta(x_1, x_2) = 0$, if and only if $x_1 = x_2, \forall x_1, x_2 \in U$

2). $\Delta(x_1, x_2) = \Delta(x_2, x_1), \forall x_1, x_2 \in U$

3). $\Delta(x_1, x_3) \leq \Delta(x_1, x_2) + \Delta(x_2, x_3), \forall x_1, x_2, x_3 \in U$

We say Δ is the distance function in this neighbor approximation space.

A suitable distance function is the key to a successful application of neighbor theory. Here, two useful distance function, Euclidean distance function and Minkowski distance function are reviewed.

Definition 2.2: Given two points $x_i = \{x_{1i}, x_{2i}, \cdots, x_{ni}\}$ and $x_j = \{x_{1j}, x_{2j}, \cdots, x_{nj}\}$ in N-dimensions Euclidean space, the distance of them can be computed as:

$$\Delta(x_i, x_j) = \left(\sum_{k=1}^{n} (x_{ki} - x_{kj})^2 \right)^{\frac{1}{2}}$$

Generally, Minkowski distance function is defined as:

$$\Delta_P(x_i, x_j) = \left(\sum_{l=1}^{n} (x_{li} - x_{lj})^P \right)^{\frac{1}{P}}$$

To construct a neighborhood rough sets model for universe granulation on the numerical attribute, Hu proposed a δ neighbor [8].

Definition 2.3: Given a neighbor space (U, Δ), $\forall x \in U, \delta \geq 0$, we say $\delta(x)$ is a δ neighbor of x whose centre is x and radius is δ, where:

$$\delta(x) = \{y | \Delta(x, y) \leq \delta, y \in U\} \quad (1)$$

Given $A_1, A_2 \subset A$, represent numerical attribute set and nominal attribute set respectively. The δ neighbor of x is defined as:

1). $\delta_{B_1}(x) = \{y | \Delta_{B_1}(x, y) \leq \delta, y \in U\}$

2). $\delta_{B_2}(x) = \{y | \Delta_{B_2}(x, y) = 0, y \in U\}$

3). $\delta_{B_1 \cup B_2}(x) = \{y | \Delta_{B_1}(x, y) \leq \delta \wedge \Delta_{B_2}(x, y) = 0, y \in U\}$

III. A Tolerance Neighborhood Relation and a Distance Function

In this section, we introduce a tolerance relation to neighbor theory to construct a new approximation relation. The relation can handle the incomplete data.

A. Incomplete information system and tolerance relation

Information system is an important knowledge representation tool in rough sets. It can be defined as follows:

Definition 2.4: An information system is a pair $IS = (U, A)$, where

1). U is a non-empty finite set of objects;

2). A is a non-empty finite set of attributes;

3). For every $a \in A$, there is a mapping $a: U \to V_a$, where V_a is called the value set of a.

If V_a contains a null value for at least one attribute $a \in A$, then *IS* is called an incomplete information system, otherwise it is complete. Furthermore, we will denote the null value by $*$.

Let $IS = (U, A)$ is an information system, $P \subseteq A$ is an attribute set. Literature [5] defines a binary relation on U as follows:

$$SIM(P) = \{(x, y) \in U \times U | \forall a \in P, a(x) = a(y) \text{ or } x = * \text{ or } y = *\} \quad (2)$$

The expression $SIM(P)$ is a tolerance relation on U. We can easily find that $SIM(P) = \cap_{a \in A} SIM(\{a\})$. Let $S_P(x)$ denote the set $\{y \in U | (x, y) \in SIM(P)\}$. It induces a covering of U and it's easy to get $S_P(x) \neq \varnothing$ for every $x \in U$, and $\cup_{x \in U} S_P(x) = U$. This work can be found in [2].

B. A distance function

To deal with categorical attribute, numerical attribute and set-valued attribute by neighbor theory at one time, we construct a new distance function Δ^T with tolerance ability.

$$\Delta^T(x, y) = \sqrt{\sum_{l=1}^{N} d_{a_l}(x, y)^2} \quad (3)$$

$$d_{a_l}(x, y) = \begin{cases} nom\_diff_{a_l}(x, y) & \text{attribute l is nominal} \\ & \text{or set-valued} \\ num\_diff_{a_l}(x, y) & \text{attribute l is numerical} \end{cases} \quad (4)$$

Meanwhile, $nom\_diff_{a_l}(x, y)$ and $num\_diff_{a_l}(x, y)$ are defined as follows.

$$nom\_diff_{a_l}(x, y) = \begin{cases} 0 & if\ x = y\ or\ x = *\ or\ y = * \\ 1 & if\ x \neq y \end{cases} \quad (5)$$

$$num\_diff_{a_l}(x, y) = \begin{cases} 0 & if\ x = *\ or\ y = * \\ \dfrac{|x_l - y_l|}{\max(a_l) - \min(a_l)} & if\ x \neq *\ and\ y \neq * \end{cases} \quad (6)$$

Especially, $\Delta^T(x, x) = 0$.

C. Five Rules for Consolidation

Based on the tolerance relation and distance function, we can improve the δ neighbor to a new neighbor with tolerance ability.

Definition 2.5: A neighborhood information system is a triple $NIS = (U, A, \Delta)$, where U is a non-empty finite set of objects; A is a non-empty finite set of attributes; Δ represents distance function in A; A and Δ form a family of neighborhood relation on U.

If there is null value for at least one object a in each attributes, *NIS* is called an incomplete neighborhood information system, otherwise it is complete. Similarly, we denote the null value by $*$.

Definition 2.6: Given a neighborhood information system $NIS = (U, A, \Delta)$, $\Delta = \Delta^T$, threshold $\delta \geq 0$, we can define a binary relation on U, called tolerance neighborhood relation, as follows:

$$SIM_N^{\delta}(P) = \{(x, y) | \forall a \in P, \Delta_a^T(x, y) \leq \delta\} \quad (7)$$

We can easily find that $SIM_N^{\delta}(P)$ is reflexive and symmetrical, but it's not transitive. It induces a covering on U. We define the information granules induced by $SIM_N^{\delta}(P)$ as

$$SN_P^{\delta}(x)=\left\{y\middle|(x,y)\in SIM_N^{\delta}(P),y\in U\right\} \quad (8)$$

Properties 2.1: Given a tolerance neighborhood relation $SIM_N^{\delta}(P)$ on neighborhood information system $NIS=(U,A,\Delta)$, $P\subseteq A$, it includes some properties as follows:

1). $SIM_N^{\delta}(P)=\bigcap_{a\in P}SIM_N^{\delta}(\{a\})$

2). $SN_P^{\delta}(x)\neq\varnothing$, because $x\in\delta(x)$

3). $\forall x\in U,\bigcup_{x\in U}SN_P^{\delta}(x)=U$

The proof procedures are omitted because it's obvious.

Furthermore, we can represent the neighborhood relation by a matrix as $M(N)=(r_{ij})_{n\times n}$, r_{ij} represents the neighborhood relation of x_i and x_j. If $x_j\in SN_P^{\delta}(x), r_{ij}=1, otherwise, r_{ij}=0$

Theorem 2.1: Given a neighborhood information system $NIS=(U,A,\Delta)$ and threshold δ, $B_1,B_2\subseteq C$, if $B_1\subseteq B_2$, then $\forall x\in U, SN_{B_2}^{\delta}(x)\subseteq SN_{B_1}^{\delta}(x)$.

Proof. $B_1\subseteq B_2\Rightarrow\bigcap_{a\in B_2}SIM_N^{\delta}(\{a\})\subseteq\bigcap_{a\in B_1}SIM_N^{\delta}(\{a\})$;

$\Rightarrow SN_{B_2}^{\delta}(x)=\left\{y\in U\middle|(x,y)\in\bigcap_{a\in B_2}SIM_N^{\delta}(\{a\})\right\}$

$\subset SN_{B_1}^{\delta}(x)=\left\{y\in U\middle|(x,y)\in\bigcap_{a\in B_1}SIM_N^{\delta}(\{a\})\right\}$

This completes the proof.

Theorem 2.1 states that the granules, induced by tolerance neighborhood relation, would be fined when new attributes add to former attribute set.

D.An illustrating example

Table 1 shows an information system about students. $S=(S_1,S_2,S_3,S_4,S_5,S_6,S_7,S_8,S_9)$ is a set of students, attribute A1, A2 and A3 represent the score, the evaluation and the familiar language of each students, respectively. Table 2 shows the results of distance computation from table 1 by formula (3).

TABLE I
STUDENTS INFORMATION

	S1	*S2*	*S3*	*S4*	*S5*	*S6*	*S7*	*S8*	*S9*
A1	78	88	98	61	90	100	0	55	68
A2	Good	Good	Good	Bad	Good	Good	Bad	Bad	*
A3	{E,J}	*	{E}	{E}	*	{E,J}	{J}	*	*

TABLE II
DISTANCE

	S1	*S2*	*S3*	*S4*	*S5*	*S6*	*S7*	*S8*	*S9*
S1	0	0.1	1.02	1.42	0.12	0.22	1.62	1.02	0.1
S2	0.1	0	0.1	1.04	0.02	0.12	1.4	1.05	0.2
S3	1.02	0.1	0	1.07	0.08	1	1.72	1.09	0.3
S4	1.42	1.04	1.07	0	1.04	1.47	1.17	0.06	0.07
S5	0.12	0.02	0.08	1.04	0	0.1	1.35	1.06	0.22
S6	0.22	0.12	1	1.47	0.1	0	1.73	1.1	0.32
S7	1.62	1.4	1.72	1.17	1.35	1.73	0	0.55	0.68
S8	1.02	1.05	1.09	0.06	1.06	1.1	0.55	0	0.13
S9	0.1	0.2	0.3	0.07	0.22	0.32	0.68	0.13	0

If $\delta=0.1$, we can get a family of information granules like follows:

$SN_A^{0.1}(S_1)=\{S_1,S_2,S_9\}$, $SN_A^{0.1}(S_2)=\{S_1,S_2,S_3,S_5\}$,

$SN_A^{0.1}(S_3)=\{S_2,S_3,S_5\}$, $SN_A^{0.1}(S_4)=\{S_4,S_8,S_9\}$,

$SN_A^{0.1}(S_5)=\{S_2,S_3,S_5,S_6\}$, $SN_A^{0.1}(S_6)=\{S_5,S_6\}$,

$SN_A^{0.1}(S_7)=\{S_7\}$, $SN_A^{0.1}(S_8)=\{S_4,S_8\}$,

$SN_A^{0.1}(S_9)=\{S_1,S_4,S_9\}$

If $\delta=0.2$, we can get another family of information granules like follows:

$SN_A^{0.2}(S_1)=\{S_1,S_2,S_9\}$, $SN_A^{0.2}(S_2)=\{S_1,S_2,S_3,S_5,S_6,S_9\}$,

$SN_A^{0.2}(S_3)=\{S_2,S_3,S_5\}$, $SN_A^{0.2}(S_4)=\{S_4,S_8,S_9\}$,

$SN_A^{0.2}(S_5)=\{S_1,S_2,S_3,S_5,S_6\}$, $SN_A^{0.2}(S_6)=\{S_2,S_5,S_6\}$,

$SN_A^{0.2}(S_7)=\{S_7\}$, $SN_A^{0.2}(S_8)=\{S_4,S_8,S_9\}$,

$SN_A^{0.2}(S_9)=\{S_1,S_2,S_4,S_8,S_9\}$

Here, $A=A_1\cup A_2\cup A_3$

IV. VARIABLE PRECISION TOLERANCE NEIGHBORHOOD ROUGH SETS MODEL

In this section, we propose the VPTNRS.

Definition 2.7: Given a neighborhood approximation space $NAS=(U,\Delta)$ and $X\subseteq U$, $\Delta=\Delta^T$ and threshold $\delta\geq 0$, the lower approximation and upper approximation about X can be defined as:

$$\begin{cases}\underline{UN}(X)=\left\{x\middle|SN^{\delta}(x)\subseteq X,\forall x\in U\right\}\\ \overline{UN}(X)=\left\{x\middle|SN^{\delta}(x)\cap X\neq\varnothing,\forall x\in U\right\}\end{cases} \quad (9)$$

And $\forall X\subseteq U,\underline{UN}(X)\subseteq X\subseteq\overline{UN}(X)$.

The boundary domain, positive domain and negative domain can be defined as:

$$BN(X)=\overline{UN}(X)-\underline{UN}(X) \quad (10)$$

$$POS(X) = \underline{UN}(X) \quad (11)$$

$$NEG(X) = U - \overline{UN}(X) \quad (12)$$

$POS(X)$ represents the granules contained by X; $NEG(X)$ represents the granules not contained by X; $BN(X)$ represents the granules partially contained by X. If $\overline{UN}(X) = \underline{UN}(X)$, we say X in this neighborhood approximation space is definable; otherwise, it's indefinable, it's rough.

In fact, some data are noisy, namely, are incomplete and inconsistent. The above method can just handle incomplete data. Ziarko proposed a probabilistic method to improve the classical rough sets with a tolerance threshold, called variable precision rough sets (VPRS) [7], [11]. Here, we study this technique and introduce it to our model.

Definition 2.8: VPTNRS is defined as follows:

$$\begin{cases} \underline{UN}^k(X) = \left\{ x \middle| \left(\dfrac{card\left(SN^\delta(x) \cap X\right)}{card\left(SN^\delta(x)\right)} \right) \geq k, \forall x \in U \right\} \\ \overline{UN}^k(X) = \left\{ x \middle| \left(\dfrac{card\left(SN^\delta(x) \cap X\right)}{card\left(SN^\delta(x)\right)} \right) < 1-k, \forall x \in U \right\} \end{cases} \quad (13)$$

Where $1 \geq k \geq 0.5$.

V. VPTNRS in Decision Table

Decision table is an important tool for knowledge representation. The difference between information system and decision table is the former does not contain decision attribute. In this section, we will introduce the VPTNRS to decision table.

Definition 2.9: Given a neighborhood information system $NIS = (U, A, \Delta)$, we say it's a neighborhood decision table $NDT = (U, C \cup D, \Delta)$ if $A = C \cup D$, C represents the condition attribute set, D represents the decision attribute, and C forms a family of neighborhood relation on U.

Definition 2.10: Given a neighborhood decision table $NDT = (U, C \cup D, \Delta)$, $\forall P \subseteq C$, $\Delta = \Delta^T$ and threshold $\delta \geq 0$, the decision attribute D divide the universe U into some equivalence classes, $X_1, X_2, \cdots, X_n$, the lower approximation and upper approximation about D relative to P can be defined as:

$$\begin{cases} \underline{UN}_P^k(D) = \bigcup_{i=1}^n \underline{UN}_P^k(X_i) \\ \overline{UN}_P^k(D) = \bigcup_{i=1}^n \overline{UN}_P^k(X_i) \end{cases} \quad (14)$$

The boundary domain, positive domain and negative domain can be defined the same as formula (10), (11) and (12).

Next, a useful digital feature of VPTNRS called dependency is discussed.

Definition 2.11: Given a neighborhood decision table $NDT = (U, C \cup D, \Delta)$, $\forall P \subseteq C$, we can define the dependency relation between P and D as:

$$\gamma_P(D) = \frac{card\left(POS_P(D)\right)}{card(U)} \quad (15)$$

It can be denoted as $P \Rightarrow_\gamma D$ for short. Obviously, $0 \leq \gamma_P(D) \leq 1$

The dependency $\gamma_P(D)$ reflects the dependence degree of D relative to P. This definition is in accordance with the classical rough set theory. It can be used to evaluate the importance of a subset of attributes.

Theorem 2.2: Given a neighborhood decision table $NDT = (U, C \cup D, \Delta)$, $\Delta = \Delta^T$, threshold $\delta \geq 0$ and $B_1, B_2 \subseteq C$, if $B_1 \subseteq B_2$, then $x \in POS_{B_1}(D) \Rightarrow x \in POS_{B_2}(D)$.

Proof. Without loss of generality, let $\forall x \in POS_{B_1}(D)$, D_j represents the equivalence class of j divided by decision D. According to theorem 2.1, if $B_1 \subseteq B_2$, then $SN_{B_2}^\delta(x) \subseteq SN_{B_1}^\delta(x)$. So, $x \in POS_{B_2}(D)$

This completes the proof.

Theorem 2.3: $\gamma_P(D)$ is monotonous, that's to say, if $B_1 \subseteq B_2 \subseteq \cdots \subseteq C$, then $\gamma_{B_1}(D) \leq \gamma_{B_2}(D) \leq \cdots \leq \gamma_C(D)$.

Proof. According to theorem 2.2, if $B_1 \subseteq B_2 \subseteq \cdots \subseteq C$, then $POS_{B_1}(D) \leq POS_{B_2}(D) \leq \cdots \leq POS_C(D)$. According to formula 4.15, we can get $\gamma_{B_1}(D) \leq \gamma_{B_2}(D) \leq \cdots \leq \gamma_C(D)$.

This completes the proof.

Definition 2.12: Given a neighborhood decision table $NDT = (U, C \cup D, \Delta)$, $P \subseteq C$, $\Delta = \Delta^T$ and threshold $\delta \geq 0$. The significance of attribute $a \in C - P$ relative to P can be defined as:

$$sig(a, P, D) = \gamma_{P \cup a}(D) - \gamma_P(D) \quad (16)$$

Based on the significance function, the relative reduction and the relative core can be defined in the classical way.

VI. Algorithm for Attribute Reduction

As mentioned above, the dependency $\gamma_P(D)$ reflects the approximating power of a condition attribute set. It can be used to measure the significance of a subset of attributes. The aim of attribute selection is to search a subset of attributes such that the classification problem has the maximal consistency in the selected feature spaces.

There are four key steps in a feature selection algorithm: subset generation, subset evaluation, stopping criterion and result validation. In algorithm, we begin with an empty set red of attribute, and we add one feature which makes the increment of dependency maximal into the set red in each round. This is

the strategy of subset generation. We embed the subset evaluation in this strategy by maximizing the increment of dependency. The algorithm does not stop until the dependency increase equals to zero by adding any new feature into the attribute subset red or all of the attributes have been added to the reduction. The algorithm is given as below:

Algorithm:
Input: $NDT = (U, C \cup D, \Delta)$ and threshold δ
Output: a reduction red
1: $\varnothing \rightarrow red$
2: for each $a_i \in C - red$
3: compute $sig(a_i, red, D)$
4: end
5: select a_k if $a_k = MAX_i(SIG(a_i, red, D))$
6: if $SIG(a_i, red, D) \geq \varepsilon$ // ε is a small positive number to control the convergence
7: $red \cup a_k \rightarrow red$
8: if $C - red = \varnothing$,go to step 9
9: else return red
10: end

There are two key steps in algorithm above. One is to compute the neighborhood of samples, the other is to analyze whether the neighborhood of a sample is consistent. With sorting technique, we can find the neighborhoods of samples in time complexity $O(n\log(n))$, while time complexity of the second step is $O(n)$. So the worst case of computational complexity of reduction is $O(N^2 n\log(n))$, here N and n are the numbers of features and samples.

VII. Experiments and Discussion

In this section, we will perform experiments to evaluate the proposed method. First, we compare three methods with results of attribute reduction, i.e. VPTNRS, VQRS [3] and fuzzy VPRS [9]. Second, we use two frequently-used classification techniques, i.e. kNN and CART, to evaluate the accuracy of classification in terms of the selected features of each method.

In our experiments, 8 choice UCI standard data sets from the machine learning data repository, University of California at Irvine [10], are used. These data sets are described in table 3. All of the selected data sets are including numerical features and categorical features. The numbers of samples are between 108 and 1000. Moreover, *all data sets are incomplete***.**

For VQRS can not be directly applied in numerical attribute, MDL [12] technique was used to discretize the data. In experiments, for each data set, we randomly divide the samples into 10 subsets, and use nine of them as training set and the rest one as the test set. After 10 rounds, we compute the average value and variation as the final performance.

The results of feature selection are presented in table 4. VPTNRS obtains 5.8 attributes on average, compared to 6.75 of VQRS and 6.75 of fuzzy VPRS. It explains that the proposed method selects much less features than other two methods. This can help reduce the time of classification. Besides, accuracy of classification computed based on these selected subsets of features is also important and must be discussed.

CART and kNN are introduced to test the quality of the reduction. The classifying accuracy of raw data and the reduced feature subsets based on 10-fold cross validation are shown in tables 5 and 6. It's easy to see that (1) the accuracies computed based on selected features of VPTNRS are better than the other two methods, both in kNN and CART; (2) the average accuracy of two different algorithms computed based on selected features of VPTNRS decreases only by 2.0% and 1.0% respectively, compared with that computed based on the raw data; (3) moreover, it's easy to see that, in data set Aus, Hep and Horse, the selected features by VPTNRS is less than other two methods, meanwhile induce much better classification results. In the results of other data sets, the number of feature selected by three methods are very close, but VPTNRS still shows a little better than other two methods.

TABLE III
DATA SETS DESCRIPTION

	Data Set	*Sample*	*Attribute*	*Numerical*	*Categorical*	*Class*
1	Annealing(Anne)	798	38	9	29	6
2	Australian(Aus)	690	14	6	8	2
3	Auto	205	26	16	9	6
4	Breast-Cancer(BC)	286	9	1	8	2
5	Bridges1(Bridg1)	108	13	9	3	6
6	Credit Screening(Credit)	690	15	6	9	2
7	Hepatitis(Hep)	155	19	6	13	2
8	Horse-Colic(Horse)	368	27	7	20	2

TABLE IV
THE RESULTS OF FEATURE SELECTION

Data Set	*Attribute*	*VPTNRS*	*VQRS+MDL*	*Fuzzy VPRS*
Anne	38	3	3	3
Aus	14	6	7	7
Auto	26	11	13	12
BC	9	4	5	4
Bridg1	13	3	3	2
Credit	15	7	6	8
Hep	19	8	10	11
Horse	27	4	7	7
Average	19.1	5.8	6.75	6.75

TABLE V
CLASSIFICATION ACCURACY OF DIFFERENT FEATURE SUBSETS WITH CART (%)

Data Set	*Raw Data*	*VPTNRS*	*VQRS+MDL*	*Fuzzy VPRS*
Anne	92.7	89.2	89.2	86.1
Aus	91.2	89.3	88.1	88.7
Auto	84.6	82.8	84.8	80.4
BC	89.2	88.5	83.3	88.5
Bridg1	79.2	76.2	76.2	74.8
Credit	88.4	87.4	85.2	87.1
Hep	85.2	83.1	82.7	81.0
Horse	87.1	85.1	83.9	84.8
Average	87.2	85.2	84.2	83.9

TABLE VI
CLASSIFICATION ACCURACY OF DIFFERENT FEATURE SUBSETS WITH KNN (%)

Data Set	*Raw Data*	*VPTNRS*	*VQRS+MDL*	*Fuzzy VPRS*
Anne	99.8	100	100	96.1
Aus	93.3	92.3	87.3	91.9
Auto	86.8	85.3	85.7	84.5
BC	95.3	92.4	89.5	92.4
Bridg1	84.2	80.2	80.2	76.3
Credit	82.7	85.5	83.0	84.1
Hep	94.2	91.5	88.3	87.5
Horse	95.9	96.5	88.5	89.3
Average	91.5	90.5	87.8	87.8

VIII. CONCLUSION

Reducing redundant or irrelevant features can improve classification performance in most of cases and decrease cost of classification. The NRS is an efficient tool to do this work for it can deal with heterogeneous attributes. But the traditional NRS can not handle both incomplete data and inconsistent data. The proposed VPTNRS solve this problem very well by generalizing NRS with a tolerance neighborhood relation and the probabilistic theory. The experimental results, compared to VQRS and fuzzy VPRS, prove that the model has strong power in handling noisy data.

REFERENCES

[1] Pawlak Z., "Rough sets.", Theoretical Aspects of Reasoning about Data, Kluwer, 1991.
[2] G. Y. Wang, "Rough Set Theory and Knowledge Discovery.", Xi'an:Xi'an Jiaotong University Press, 2001.
[3] C. Cornelis, M. De Cock, A. Radzikowska, "Vaguely Quantified Rough Sets,", Proc. 11th Int. Conf. on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing (RSFDGrC2007), Lecture Notes in Artificial Intelligence 4482, 2007, pp: 87–94.
[4] T. Y. Lin, Q. Liu, K J Huang, "Rough sets neighborhood systems and approximation.", In 15th International Symposium on Methodologies of Intelligent Systems, 1990.
[5] Y. Y. Yao, "Relational interpretation of neighborhood operators and rough set approximation operators.", Information sciences, vol. 111, no.198, pp: 239-259, 1998.
[6] W. Z. Wu, W. X. Zhang, "Neighborhood operator systems and approximations.", Information sciences, vol. 144, no.14, pp: 201-217, 2002.
[7] W. Ziarko, "Set approximation quality measures in the variable precision rough set model.", Soft Computing Systems, Management and Applications, pp: 442-452, 2001.
[8] Q. H. Hu, D. R. Yu, Z. X. Xie, "Numerical attribute reduction based on neighborhood granulation and rough approximation.", Chinese Journal of software, vol. 19, no.3, pp.640−649, 2008.
[9] M. R. Alicja, L. Rolka, "Variable Precision Fuzzy Rough Sets", Transaction on Rough Sets, LNCS, 144-160, 2004.
[10] D. J. Newman, S. Hettich, C. L. Blake, C. J. Merz, "UCI Repository of Machine Learning Databases.", University of California, Department of Information and Computer Science, Irvine, CA, 1998. <http://www.ics.uci.edu/~mlearn/MLRepository.html>.
[11] W. Ziarko, " Variable precision rough set model" , Journal of Computer and System Sciences, vol. 46, pp: 39-59, 1993.
[12] U. Fayyad, K. Irani, "Discrediting continuous attributes while learning Bayesian networks.", in 13th International Conference on Machine Learning, Morgan Kaufmann, 1996, pp: 157– 165.

Siyuan Jing received the BS and MS degree in computer science and engineering from University of Electronic Science and Technology of China, Chengdu, China, in 2004 and 2008 respectively. He is now third-year PhD student in University of Electronic Science and Technology of China. His research interests include rough sets, multi-criteria decision making technique and cloud computing.

Kun She received the BS degree in applied mathematics from University of Electronic Science and Technology of China, in 1989, and the MS degree of electronics and communications from southwest communication research institute of China, in 1992, and the PhD degree of applied computer from University of Electronic Science and Technology of China, in 2006. He is now the professor at University of Electronic Science and Technology of China. His interests include computer network, information security and artificial intelligence.