

## A Hybrid Feature Gene Selection Method based on Fuzzy Neighborhood Rough Set with Information Entropy

Tao Chen<sup>1,2\*</sup>, Zenglin Hong<sup>1</sup>, Fang-an Deng<sup>2</sup> and Man Cui<sup>1</sup>

<sup>1</sup> School of Automation, Northwestern Polytechnical University, Xi'an, Shaanxi, 710072, China

<sup>2</sup> School of Mathematics and Computer Science, Shaanxi University of Technology, Hanzhong, 723000, Shaanxi, China

\*Corresponding author: [ct79hz@126.com](mailto:ct79hz@126.com)

### Abstract

DNA microarray technique can detect tens of thousands of genes activity in cells and has been widely used in clinical diagnosis. However, microarray data has the characteristics of high dimension and small samples, moreover many irrelevant and redundant genes also decrease performance of classification algorithm. Feature gene selection is an effective method to solve this problem. This paper proposes a hybrid feature gene selection method. Firstly, a lot of irrelevant genes from original data were eliminated by using reliefF algorithm, and the candidate feature genes subset is obtained; Secondly, Fuzzy neighborhood rough set with information entropy which deals directly with continuous data is proposed to reduce redundant genes among genes subset above. Here, differential evolution algorithm is used to optimize radius before reduction by using fuzzy neighborhood rough set, because radius of neighborhood greatly affects reduction performance. The simulation results on six microarray datasets indicate that our method can obtain higher classification accuracy by using as few genes as possible, especially feature genes selected are important for understanding microarray data and identifying the pathogenic genes. The results demonstrated that this method is effective and efficient for feature genes selection.

**Keywords:** Feature gene selection; ReliefF algorithm; Rough set; Neighborhood rough set; Information entropy; Differential evolution algorithm

### 1. Introduction

DNA microarray technology is a major breakthrough in the field of molecular biology at the end of the last century, it can simultaneously detect expression levels of thousands of genes in cells in once experiment, and provide strong support for the gene diagnosis and pathogenic mechanism of disease related to genes, especially tumor [1, 2].

Microarray data often contains a small number of samples (tens or hundreds) and a large number of genes (tens of thousands). It leads to imbalance between the numbers of samples and genes and causes the curse of dimensionality. The curse of dimensionality will cause peak phenomenon of classification algorithm which mean the classification performance reaches the highest level when the dimensionality increases to a certain number, then classification performance slowly decreases with dimensionality increases. The reason is that the number of samples in the learning process is limited. Therefore, in order to ensure the classification performance, it is necessary to increase the number of samples. However the number of samples needed in the learning process increased with the increase of the number of features by the exponential growth mode. And the high cost of microarray experiments

limited the samples number. In addition, the microarray data contains many irrelevant and redundant genes, which not only increase the dimension of feature space and decrease the learning efficiency, but also increase the possibility of noise data, and interfere with the learning process of classification algorithm, ultimately affects the construction and the results of classification model. In order to decrease the influence of adverse factors and reduce the dimension of feature space, gene selection is an effective method to remove the irrelevant and redundant genes, and effectively improves the learning efficiency and the performance of algorithm to avoid over fitting phenomena.

Gene selection plays an important role in microarray data analysis. The most commonly gene selection approaches are based on gene ranking. Each gene is evaluated individually and assigned a score to reflect its correlation with the class according to certain criteria. And then genes are ranked according to their scores and the top-ranked genes are selected as feature genes. There are many feature gene selection methods ranking based, such as  $t$  statistic [3], information gain [4, 5],  $\chi^2$  statistic [6], the threshold number of misclassification score [7], concatenation of several feature filtering algorithms [8], relief [9], reliefF [10], mutual information [11-13] and neighborhood mutual information [14] *et al.*

The methods ranking-based are widely used in feature selection because of their advantages: simple principle, easy to calculate, linear time complexity and fast operation. However most of the methods often apply to data which has normal distribution. Moreover the interaction among genes was ignored and the redundant genes with high correlation cannot be eliminated in these methods. This correlation may be because these genes belong to the same signaling pathways related to the disease. If a gene is high-ranked, other genes that are highly correlated with it are also likely to be ranked high by the gene ranking method. The redundancy can also interfere with the results and lead to misclassifications. Finally these methods can't obtain higher classification performance because of no participate of classifier.

Relief [9] and reliefF [10] have been widely used in feature selection. The core idea is discernibility ability among "similar" samples with attribute as an indicator to evaluate the importance of attributes, and gives the classification weights of attributes. The algorithm has no special requirements for the distribution of the data and can deal with continuous and discrete data, and has less computational complexity. Therefore relief and reliefF are effective methods for reducing irrelevant attributes and have better classification performance compared with other methods ranking based.

Rough set, proposed by Pawlak, is a powerful mathematical tool to deal with uncertainty problem and has been widely applied to data mining, granularity computation and attribute reduction [15]. Reduct is a minimal subset of the original attribute set which is independent and has the same discernibility ability as all of the attributes in rough set. Rough set improves the computing speed, decreases storage space and enhances precision. But Pawlak's rough set cannot directly deal with continuous attributes, so continuous attributes must be discretized before attribute reduction. Qiang Shen pointed that discretization process may lead to some information loss from the original data, and then it affects classification accuracy because information is lost and not fully utilized [16]. Moreover, the selection of cut point in discretization process is extraordinarily sensitive to the result of attribute reduction, so it is very difficult to choose a suitable cut point in discretization process.

Neighborhood rough set [17-19] and Fuzzy rough set [20-23] were applied to deal with continuous attributes respectively. For neighborhood rough set, neighborhood relation can be used to generate a family of neighborhood granules from the universe characterized with numerical features, and then we can use these neighborhood granules to approximate decision class. Hu qinghua used neighborhood rough set to reduce attributes and got better performance. For fuzzy rough set, a fuzzy equivalence relation would be generated by a real

valued attribute, instead of crisp equivalence relation. Jensen and Shen [24-25] proposed a fuzzy rough quickreduct algorithm based on fuzzy dependency function to reduce attributes. Hu qinghua proposed an improved fuzzy rough set based on information entropy which measures the discernibility ability of continuous attributes better than fuzzy dependency function, that is fuzzy neighborhood rough set [26].

In this paper, we propose a hybrid feature gene selection method. Firstly, candidate gene subset which has high relevance with classification task is obtained by using reliefF algorithm. Then, differential evolution algorithm [27-28] is used to optimize radius of neighborhood, because radius of neighborhood of fuzzy neighborhood rough set highly affects reduction performance. Finally, feature genes are selected using optimized fuzzy neighborhood rough set.

The remainder of this paper is organized as follows: The reliefF algorithm is introduced in Section 2. Fuzzy neighborhood rough set based on information entropy is defined, and then a attribute reduction algorithm based on fuzzy neighborhood rough set and forward greedy search strategy is proposed in Section 3. Section 4 proposes a method to optimize radius of fuzzy neighborhood rough set using differential evolution algorithm. Section 5 gives ideas and steps of our proposed method. Section 6 gives the experimental results and analysis on six public benchmark microarray datasets. The conclusion is given in Section 7.

## 2. ReliefF Algorithm

Relief algorithm [9] was proposed by Kira and Rendell in 1994. The success of the algorithm is due to the fact that it is fast, easy to understand and implement and accurate even with dependent features and noisy data. However Relief can only be applied to two classification problems, so Kononenko proposed reliefF [10], which is the extension of relief. The key idea of reliefF is to estimate the quality of attributes according to how well their values distinguish between the instances that are near to each other. For that purpose, given a randomly selected instance  $x$ , reliefF searches for  $k$  nearest neighbors of  $x$  from the same class, called nearHist, and also  $k$  nearest neighbors of  $x$  from each of the different class, called nearMisses. The quality estimation  $w(g)$  for each attribute  $g$  is update formula, nearHist and nearMisses. In the update formula, the contribution of all the hits and misses are averaged. The process is repeated for  $n$  times to return weights of all features. ReliefF is fast, not limited by data types, fairly noise-tolerant, and unaffected by feature interaction, but it does not deal with redundant attributes.

---

**Algorithm1:** attribute reduction based on reliefF algorithm

---

**Input:** training set  $X = (x_1, x_2, \dots, x_n)$  and features set  $G = (g_1, g_2, \dots, g_m)$

**Output:** weight vector of the features  $W = (w_1, w_2, \dots, w_m)$

- 1) Initialize weight vector of the features:  $W = (0, 0, \dots, 0)$  ;
  - 2) For  $i = 1$  to  $n$ 
    - (1) randomly select a instance  $x_i$  from training set  $X$  ;
    - (2) search for  $k$  nearest neighbors of  $x_i$  from the same class, called nearHist  $H_j$  ;
    - (3) For each class  $C \neq class(x_i)$ 
      - (a) search for  $k$  nearest neighbors of  $x_i$  from each of the different class, called nearMisses  $M_j$
      - (b) For  $g = 1$  to  $m$
-

$$W(g) = W(g) - \left( \sum_{j=1}^k \text{diff}(g, x_j, H_j) \right) / kn + \sum_{C \neq \text{class}(x_i)} \left( \frac{P(C)}{1 - P(\text{class}(x_i))} \right) \left( \sum_{j=1}^k \text{diff}(g, x_j, M_j) / kn \right)$$

End;

End;

End.

where,  $P(C) = \text{num}(C)/n$ ,  $\text{num}(C)$  is instance of class  $C$  ;

$\text{diff}(g, x, y) = | \text{value}(g, x) - \text{value}(g, y) | / \max(g) - \min(g)$ ,  $\text{value}(g, x)$  is the value of instance  $x$  in feature  $g$ ,  $\max(g)$ ,  $\min(g)$  denote max and min value of all of instances in feature  $g$ .

### 3. Fuzzy Neighborhood Rough Set based on Information Entropy

#### 3.1. Neighborhood Rough Set (NRS)

For discrete attribute, the samples with the same attribute value are pooled into a set, called equivalence class. These samples are expected to belong to the same class; otherwise, they are inconsistent. It is easy to verify whether the decisions are consistent or not by analyzing their decisions. However, it is unfeasible to compute equivalence classes with continuous attributes because the probability of samples with the same numerical value is very small [17]. Intuitively speaking, the samples with the similar attribute values should be classified into a single class in this case; otherwise, the decision is not consistent. Based on this observation, neighborhood concept is introduced into the rough set, and then the model of neighborhood rough set was proposed [17, 19].

**Definition 1**  $U$  is a nonempty and finite set of samples  $\{x_1, x_2, \dots, x_n\}$  (called a universe),  $A$  is a set of attributes  $\{a_1, a_2, \dots, a_m\}$ .  $\langle U, A \rangle$  is called a information system. If  $A = C \cup D$ , where  $C$  is the set of condition attributes and  $D$  is the decision attributes,  $\langle U, A = C \cup D \rangle$  is called a decision system.

**Definition 2**  $\langle U, A = C \cup D \rangle$  is a decision system,  $\forall x_i \in U, B \subseteq C$ , the neighborhood  $\delta_B(x_i)$  of  $x_i$  in  $B$  is defined as:  $\delta_B(x_i) = \{x_j \mid \Delta(x_j, x_i) \leq \delta, x_j \in U\}$ . where  $\Delta$  is a distance function.  $\forall x_1, x_2, x_3 \in U$ ,  $\Delta$  satisfies:

- (1)  $\Delta(x_1, x_2) \geq 0$ ,  $\Delta(x_1, x_2) = 0$  if and only if  $x_1 = x_2$  ;
- (2)  $\Delta(x_1, x_2) = \Delta(x_2, x_1)$  ;
- (3)  $\Delta(x_1, x_2) + \Delta(x_2, x_3) \geq \Delta(x_1, x_3)$  .

A neighborhood relation  $N$  on the universe  $U$  can be written as a relation matrix  $M(N) = r_{ij}$ ,

$$\text{where } r_{ij} = \begin{cases} 1, \Delta(x_i, x_j) \leq \delta \\ 0, \Delta(x_i, x_j) > \delta \end{cases} .$$

**Definition 3**  $U$  is a universe and  $N$  is a neighborhood relation on  $U$ ,  $\langle U, N \rangle$  is called a neighborhood approximation space.  $\forall X \subseteq U$ , lower and upper approximation of  $X$  in  $\langle U, N \rangle$

is defined as:  $\underline{N}X = \{x_i \mid \delta(x_i) \subseteq X, x_i \in U\}$   $\overline{N}X = \{x_i \mid \delta(x_i) \cap X \neq \emptyset, x_i \in U\}$  .

The boundary region of  $X$  in  $\langle U, N \rangle$  is defined as:  $BN(X) = \overline{N}(X) - \underline{N}(X)$ .

**Definition 4**  $\langle U, A = C \cup D \rangle$  is a decision system and  $A$  can generate a neighborhood relation  $N$  on  $U$ ,  $\langle U, A = C \cup D, N \rangle$  is called a neighborhood decision system.

**Definition 5**  $\langle U, A = C \cup D, N \rangle$  is a neighborhood decision system.  $D$  divides  $U$  into  $N$  equivalence classes:  $X_1, X_2, \dots, X_N$ .  $\forall B \subseteq C$ , lower and upper approximation of decision  $D$  with respect to attributes  $B$  are defined as:  $\underline{N}_B D = \bigcup_{i=1}^N \underline{N}_B X_i$ ;  $\overline{N}_B D = \bigcup_{i=1}^N \overline{N}_B X_i$ .

where  $\underline{N}_B X = \{x_i \mid \delta_B(x_i) \subseteq X, x_i \in U\}$ ,  $\overline{N}_B X = \{x_i \mid \delta_B(x_i) \cap X \neq \emptyset, x_i \in U\}$ .

The decision boundary region of  $D$  with respect to attributes  $B$  is defined as:

$$BN(D) = \overline{N}_B(D) - \underline{N}_B(D).$$

### 3.2. Fuzzy Neighborhood Rough set (FNRS)

The concept of fuzzy neighborhood rough set is given as follows [26]. Suppose  $X$  is a nonempty set,  $R$  is a binary relation defined on  $X$ , denoted by a relation matrix  $M(R)$ :

$$M(R) = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \dots & r_{2n} \\ \dots & \dots & \dots & \dots \\ r_{n1} & r_{n2} & \dots & r_{nn} \end{bmatrix}, \text{ where } r_{ij} \in [0, 1] \text{ is the relation value of } x_i \text{ and } x_j.$$

**Definition 6**  $R$  is binary relation, and if  $R$  satisfies property as follows:

- (1)  $\forall x \in X, R(x, x) = 1$
- (2)  $\forall x, y \in X, R(x, y) = R(y, x)$
- (3)  $\forall x, y, z \in X, R(x, z) \geq \min\{R(x, y), R(y, z)\}$

$R$  is called a fuzzy equivalence relation.

**Definition 7**  $U$  is a universe and  $R$  is a fuzzy neighborhood equivalence relation on  $U$ . The fuzzy partition of  $U$  that generated by  $R$  is defined as:  $U/R = \{[x_i]_R\}_{i=1}^n$ , where

$$[x_i]_R = \frac{r_{i1}}{x_1} + \frac{r_{i2}}{x_2} + \dots + \frac{r_{in}}{x_n} \text{ is the fuzzy equivalence class generated by } x_i \text{ and } R.$$

**Definition 8**  $[x_i]_R$  is the fuzzy equivalence class generated by  $x_i$  and  $R$ . The cardinality of

$$[x_i]_R \text{ is defined as: } |[x_i]_R| = \sum_{j=1}^n r_{ij}.$$

**Definition 9**  $U$  is a universe and  $R$  is the fuzzy neighborhood equivalence relation on  $U$ ,

Information quality of approximation space  $\langle U, R \rangle$  is defined as:  $H(R) = -\frac{1}{n} \sum_{i=1}^n \log \lambda_i$ , where

$$\lambda_i = \frac{|[x_i]_R|}{n}.$$

The formula of information measure forms a map:  $H : R \rightarrow \mathfrak{R}^+$ , where  $R$  is a equivalence relation matrix,  $\mathfrak{R}^+$  is the nonnegative real number set. This map builds a foundation on that we can compare the discernibility ability, partition ability or approximating ability of multiple fuzzy equivalence relations. Entropy value increases monotonously with the discernibility ability or the knowledge's fineness. So the finer partition is, the greater entropy is, and the more significant attribute set is.

**Definition 10**  $\langle U, A, V, F \rangle$  is a fuzzy neighborhood information system, where  $A$  is the numeric attribute set,  $B$  and  $E$  are two subset of  $A$ .  $[x_i]_B$  and  $[x_i]_E$  are two fuzzy equivalence class containing  $x_i$  generated by  $B$  and  $E$  respectively.

The joint entropy of  $B$  and  $E$  is defined as:  $H(BE) = H(R_{E R_B}) = -\frac{1}{n} \sum_{i=1}^n \log \frac{|[x_i]_B \cap [x_i]_E|}{n}$ ;

The conditional entropy of  $E$  conditioned to  $B$  is defined as:  $H(E|B) = -\frac{1}{n} \sum_{i=1}^n \log \frac{|[x_i]_E \cap [x_i]_B|}{|[x_i]_B|}$ .

**Theorem 1**  $H(E|B) = H(BE) - H(B)$ .

**Definition 11**  $\langle U, A, V, F \rangle$  is a fuzzy neighborhood information system, where  $A$  is the condition attribute set,  $B \subseteq A$ ,  $a \in B$ .

- (1) If  $H(B) = H(B - a)$ , it says  $a$  is redundant.
- (2) If  $H(B) > H(B - a)$ , it says  $B$  is independent.
- (3) If  $B$  satisfies:  $H(B) = H(A)$  and  $\forall a \in B, H(B) > H(B - a)$ , it says  $B$  is a reduct of  $A$ .

**Definition 12**  $\langle U, A, V, F \rangle$  is a fuzzy neighborhood information system, where  $A = C \cup d$ ,  $C$  is condition attribute set and  $d$  is decision attribute.  $B \subseteq C$ ,  $\forall a \in B$ .

- (1) If  $H(d|B - a) = H(d|B)$ , it says  $a$  is redundant in  $B$  relative to  $d$ .
- (2) If  $H(d|B - a) > H(d|B)$ , it says  $B$  is independent.
- (3) If  $B$  satisfies  $H(d|B) = H(d|C)$  and  $\forall a \in B, H(d|B - a) > H(d|B)$ , it says  $B$  is a reduction of  $C$  relative to  $d$ .

Definition 11 and 12 proposed new definitions of dependence, reduction and relative reduction in terms of information theory.

**Definition 13**  $\langle U, A, V, F \rangle$  is a fuzzy neighborhood information system, where  $A$  is the condition attribute set,  $B \subseteq A$ ,  $a \in B$ , the significance of attribute  $a$  in  $B$  is defined as:

$$SIG(a, B) = H(B) - H(B - a).$$

**Definition 14**  $\langle U, A, V, F \rangle$  is a fuzzy neighborhood information system, where  $A = C \cup d$ ,  $C$  is condition attribute set and  $d$  is decision attribute.  $B \subseteq C$ ,  $\forall a \in B$ . the significance of attribute  $a$  in  $B$  relative to  $d$  is defined as:  $SIG(a, B, d) = H(d|B - a) - H(d|B)$ .

Definition 13 works in unsupervised feature selection, where  $SIG(a, B)$  measures the increment of discernibility ability introduced by attribute  $a$ . Definition 14 works in supervised feature selection.  $SIG(a, B)$  measures the increment of discernibility ability introduced by attribute  $a$ , where computes the increment of discernibility ability relative to the decision introduced by attribute  $a$ .

### 3.3. Attribute Reduction Algorithm based on Fuzzy Neighborhood Rough Set with Information Entropy and Forward Greedy Search Strategy

According to fuzzy neighborhood rough set based on information entropy measurement, an attribute reduction algorithm based on fuzzy neighborhood rough set with information entropy and forward greedy search strategy can be constructed.

---

**Algorithm2(FNRS\_FGSS):** Attribute reduction based on fuzzy neighborhood rough set with information entropy and forward greedy search strategy

---

**Input:** fuzzy neighborhood information system  $\langle U, A = C \cup d, V, F \rangle$ .

**Output:** a relative reduct  $D_{red}$ .

Step 1:  $\forall a \in A$ , compute the equivalence;

Step 2:  $D_{red} = \phi$ ;

Step 3: For each  $\forall a_i \in C - D_{red}$

Compute  $H_i = SIG(a_i, D_{red}, d)$

End;

Step 4: Choose attribute  $a_k$  which satisfies:  $SIG(a_k, D_{red}, d) = \max_i(H_i)$ ;

Step 5: If  $SIG(a_k, D_{red}, d) > 0$ , then  $D_{red} = D_{red} \cup a_k$ , goto Step 3;

Else return, output  $D_{red}$ ;

End.

---

## 4. Differential Evolution Algorithm

Differential evolution algorithm (DE), proposed in 1995, is a parallel and randomly search algorithm based on difference among groups [27-28]. Compared with the traditional evolutionary algorithm, DE applies real encoding to decrease complexity of genetic operation, especially its specific memory function has the ability of the dynamic tracking the current search condition. DE has many advantages of simple principle, less parameters, better global convergence. DE is widely used in the optimization fields because it has faster convergence speed than genetic algorithm, particle swarm algorithm and ant colony algorithm, and it is easy to obtain the global optimal solution as many studies showed.

For fuzzy neighborhood rough set, each point in real space generates a  $d$  neighborhood, and the  $d$  neighborhood becomes a basic information particle to describe arbitrary concept of the space. The size of classification boundary in fuzzy neighborhood rough set is not only connected with attribute space of classification problem, but also with analysis granularity, that is size of neighborhood, which is radius of neighborhood. The size of neighborhood reflects the size of classification granularity and determines the amount of training samples in classification boundary region. So the radius of neighborhood is an important factor to affect performance of fuzzy neighborhood rough set.

So far, there is no uniform standard for selecting radius of neighborhood, moreover its size is often associated with the research object, therefore radius also was obtained by experiment

method. But this method has low efficiency, and is often unable to obtain the optimal radius. This paper uses DE to optimize radius of neighborhood.

---

**Algorithm3:** Optimize radius of fuzzy neighborhood rough set based on DE

---

**Input:** training set  $x_{train}$ , testing set  $x_{test}$ , population size  $NP$ , upper boundary and lower boundary of individual  $x_{upper}$  and  $x_{low}$ , mutation operator  $F$ , crossover operator  $CR$  and the largest number of iterations  $T$ .

**Output:** optimal radius  $d$ .

Step 1: The initialization of population.

Randomly generate a initial population  $POP_0 = \{x_{i,0} \mid x_{i,0} = x_{low} + rand \cdot (x_{upper} - x_{low}), (i = 1, 2, \dots, NP)\}$ ,

where  $x_{i,0}$  is a real vector whose length is 1. Each vector represents an individual in the population, namely the radius  $d$  of neighborhood.

Step 2: Calculate individual fitness in current population.

Training set  $x_{train}$  is reduced using FNRS\_FGSS with  $x_{i,0}$  as radius, and then SVM is trained in training set reduced, and classification accuracy is known as the individual fitness, which is  $f_{i,0} = f(x_{i,0})$ .

Step 3: For  $G = 0$  to  $T-1$

(1) Mutation operation: individuals  $v_{i,G+1}$  are generated according to

$$v_{i,G+1} = x_{i1,G} + F \cdot (x_{i2,G} - x_{i3,G});$$

(2) Crossover operation: individuals  $u_{i,G+1}$  are generated according to

$$u_{i,G+1} = \begin{cases} v_{i,G+1}, & \text{if } rand \leq CR; \\ x_{i,G+1}, & \text{if } rand > CR; \end{cases}$$

(3) Calculate individual fitness  $f_{i,G+1}$  in temporary population, which is  $f_{i,G+1} = f(u_{i,G+1})$ ;

(4) Individuals  $x_{i,G+1}$  are selected based on greedy criterion from  $u_{i,G+1}$ , and then new generation population  $POP_{G+1} = \{x_{i,G+1}, i = 1, 2, \dots, NP\}$  is generated;

End

Step 4: Output the optimal individual, namely radius  $d$ .

---

## 5. Our Proposed Method

Microarray data contains many irrelevant and redundant genes which decrease classification performance, this paper designs a hybrid method to select feature genes from microarray data. The method includes two stages: (1) we first applied reliefF algorithms to select a gene subset of top-ranked that has high relevance with classification task; redundant genes were removed from the gene subset of top-ranked and feature genes were obtained by using fuzzy neighborhood rough set with information entropy. (2) In the second stage, differential evolution algorithm was used to optimize radius of neighborhood before reduction by using fuzzy neighborhood rough set, because radius of neighborhood greatly affects performance of fuzzy neighborhood rough set.



**Our proposed method: The hybrid feature gene selection method**

- Step1: standardize data;
- Step2: preselect a candidate subset of top-ranked genes by using reliefF;
- Step3: optimize radius of neighborhood using differential evolution algorithm;
- Step4: select feature genes by using FNRS\_FGSS with optimized radius;
- Step5: classify microarray datasets to verify effectiveness of our proposed method by using RBF-SVM as classifier.

**6. Experimental Results and Analysis**

**6.1. Experimental Datasets**

In order to evaluate the performance of our proposed method, six well-known benchmark cancer microarray datasets are selected and used in our experiments. The characteristics of the datasets are shown in Table 1.

**Table 1. Six Microarray Datasets**

Data set	classes	genes	samples	training samples	testing samples
Gliomas	2	12625	50	20	30
DLBCL	2	7129	77	32	45
LeukemiaGloub	3	7129	72	38	34
MLLLeukemia	3	12582	72	27	45
Lung	5	12600	203	142	61
ALL	6	12625	248	148	100

**6.2 Experimental Results and Analysis**

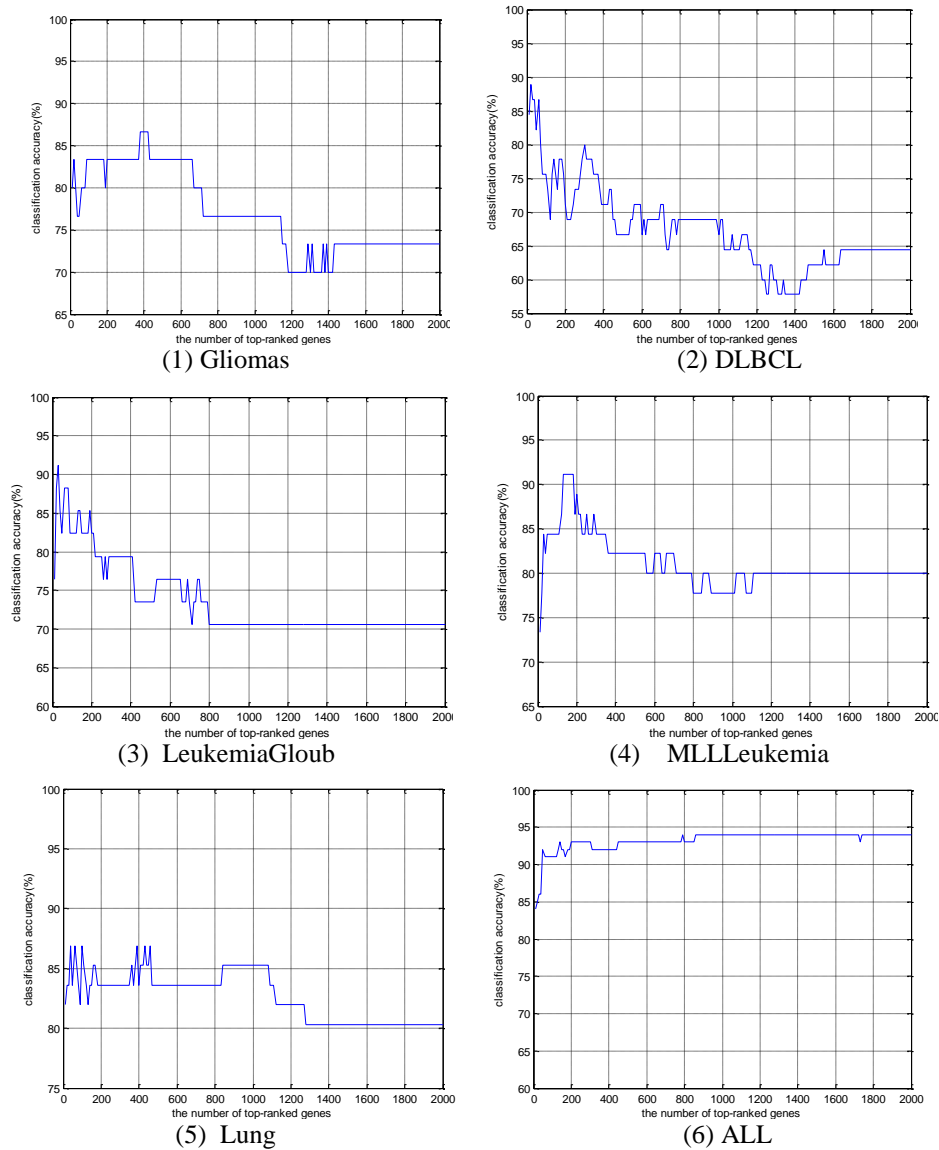
**6.2.1. Experiment 1:** The influence of the number of top-ranked genes on classification accuracy by using reliefF algorithm.

Research shows that the number of top-ranked genes by using reliefF can affect accuracy and efficiency of classification. And the number of selected genes has no uniform standard at present. In order to guarantee the performance of the algorithm, we investigate the influence of the number of top-ranked genes to classification accuracy by experiment method, and then determine the number of selected top-ranked genes.

Figure 1-3 show that the classification accuracy does not increase monotonously with the number of top-ranked genes, and the number of top-ranked genes by using reliefF, and this highly affects classification accuracy. For dataset ALL, classification accuracy gradually increases with the increase of the number of genes at the beginning, but its classification accuracy remains basically 93% when the number of genes reaches a certain extent (approximately 900). For the other datasets, at the beginning classification accuracy increases with the number of genes, its classification accuracy gradually reduces when the number of genes reaches a certain extent.

These results indicate that there are a large number of irrelevant genes in microarray data, and these genes affect the classification accuracy and time efficiency. It shows reliefF can not only significantly reduce the number of genes, but also improve the classification accuracy, it is an effective method for removing irrelevant genes with

classification tasks. In addition, because the number of top-ranked genes highly affects performance of reliefF algorithm, we give a method to roughly estimate number of top-ranked genes by experiment.



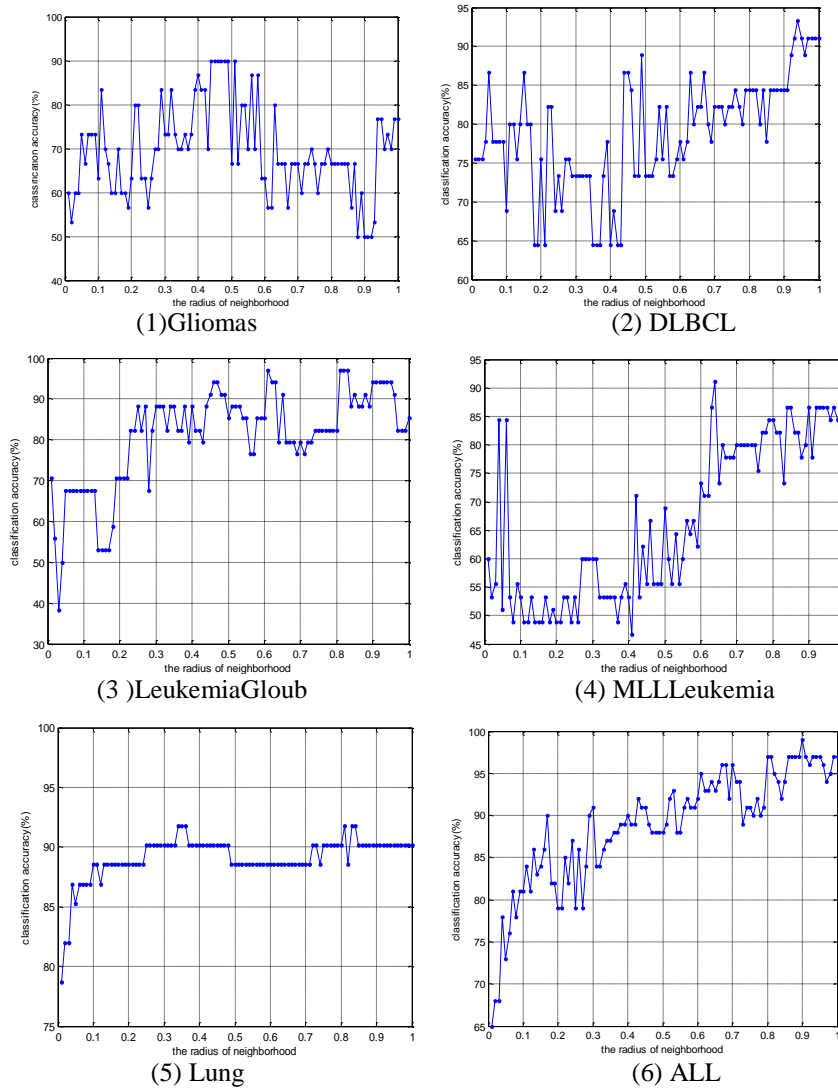
**Figure 1. The Influence of The Number of Top-ranked Genes to Classification Accuracy by using ReliefF Algorithm**

**6.2.2. Experiment 2:** The influence of radius of fuzzy neighborhood rough set on classification accuracy

The radius is an important factor for affecting the performance of fuzzy neighborhood rough set, and it will obtain different classification accuracy by using fuzzy neighborhood rough set with different radius. In order to explain this phenomenon and analyze the influence of radius of neighborhood on classification accuracy, we designed the experiment as follows:

when a radius is taken from [0,1] the step length is 0.01, then a gene subset will be got by using fuzzy neighborhood rough set, and then 100 genes subsets will be generated.

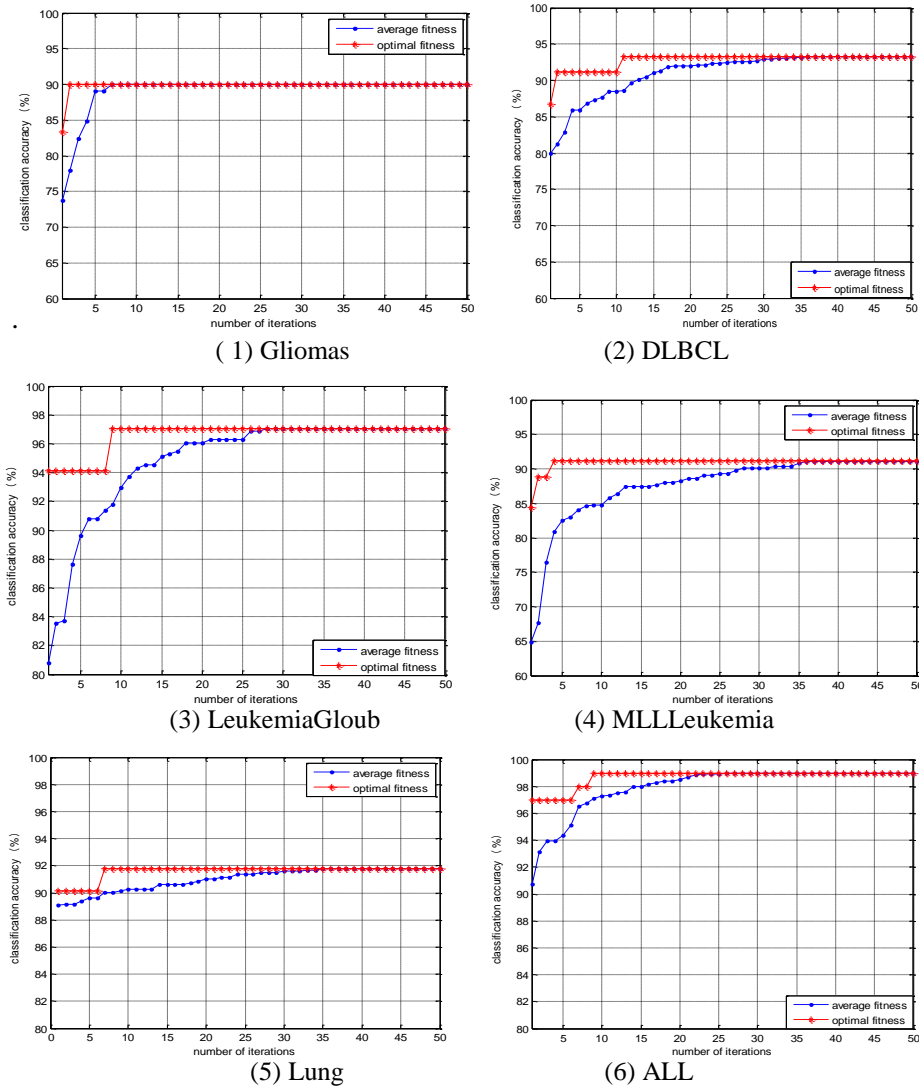
Figure 2 demonstrated that radius of fuzzy neighborhood rough set has the large influence of classification accuracy which is different from different radius by using fuzzy neighborhood rough set. Moreover classification accuracy will reduce greatly if radius is inappropriate. In addition, accurate optimal radius for the best classification accuracy can not be obtained by this experimental method, and it leads to a limitation for the widespread application of fuzzy neighborhood rough set. This paper applies differential evolution algorithm to optimize radius of neighborhood, in order to guarantee performance of fuzzy neighborhood rough set.



**Figure 2. The Influence of Radius of Fuzzy Neighborhood Rough Set on Classification Accuracy**

To optimize radius of DE, some parameters are set up in advance according to literature [27]: population size  $NP=15$ , mutation operator  $F=1$ , crossover operator  $CR=0.5$ , the largest number of iterations  $T=50$  and RBF-SVM is as classifier in experiment.

Compared with above method, we can obtain accurate optimal radius that will get the best classification accuracy using DE algorithm. Figure 3 shows DE is an effective method to find optimal radius, and the optimal radius can be obtained when iteration times is no more than 50. The results showed that DE is an effective and feasible method for optimizing the radius of fuzzy neighborhood rough set with strong global search ability and search efficiency.



**Figure 3. The Curve of Iterative Convergence based on DE**

**6.2.3. Experiment 3: Experimental results of different methods**

In order to compare our proposed method with classical rough set, MDL method [26] is used to discretize continuous attributes, because classical rough set cannot deal with continuous attributes directly.

(1) classification accuracy

Table 2 shows the classification accuracy of different methods on six datasets. Our proposed method has the highest classification accuracy in the five methods and is improved

at least 2% than other methods on all datasets except MLLLeukemia. To be specific, except MLLLeukemia, accuracy of our method is improved at least 2% than ReliefF+RS on the other five datasets. The results indicates that fuzzy neighborhood rough set is an effective method than classical rough set, because fuzzy neighborhood rough set can deal with continuous attributes to avoid information loss caused by discretization process.

Classification accuracy of our method is improved at least 3.5%, 3% and 24% relative to ReliefF, Kruskal-wallis and Raw, respectively. It indicates fuzzy neighborhood rough set is an effective attributes reduction method, because it can reduce redundant attributes to improve classification performance. For MLLLeukemia dataset, our proposed method, ReliefF+RS, ReliefF and Kruskal-wallis have the same result.

According to analysis, our proposed method is an effective attributes reduction method of mainly reducing irrelevant and redundant attributes from high dimensional dataset, especially it can reduce continuous attributes directly to avoid information loss, and it gains better performance than rough set.

**Table 2. Classification Accuracy of Different Methods**

Dataset	Raw	Kruskal-wallis	ReliefF	ReliefF+RS	Our method	Radius
Gliomas	66.7%	83.33%	86.7%	87.2%	<b>90%</b>	0.507
DLBCL	75.6%	73.33%	88.9%	88.9%	<b>93.3%</b>	0.924
LeukemiaGloub	55.9%	94.11%	91.17%	94.11%	<b>97.05%</b>	0.608
MLLLeukemia	68.9%	91.1%	91.1%	91.1%	<b>91.1%</b>	0.635
Lung	75.4%	91.8%	86.8%	88.2%	<b>91.8%</b>	0.338
ALL	68%	95%	94%	95.3%	<b>99%</b>	0.897

(2) the number of feature gene selected

Table 3 shows the number of feature genes selected with our proposed method is far less than the number of original data, ReliefF and Kruskal-wallis, and is slightly less than ReliefF+RS. It indicates fuzzy neighborhood rough set is an effective reduction algorithm than classical rough set that reduces redundant attributes. These results show our proposed method can reduce irrelevant and redundant attributes to gain smaller number of genes, and then decrease space dimensionality and the complexity of the data. The smaller number of genes can greatly improve the classification efficiency, and enhance the understanding of microarray data. Feature gene selected with our proposed method will provide more reliable basis and reference for clinical diagnosis and treatment.

Table 4 shows feature genes selected by using our proposed algorithm, and these selected feature genes can help to understand microarray data, especially provide support for research cancer from genomics.

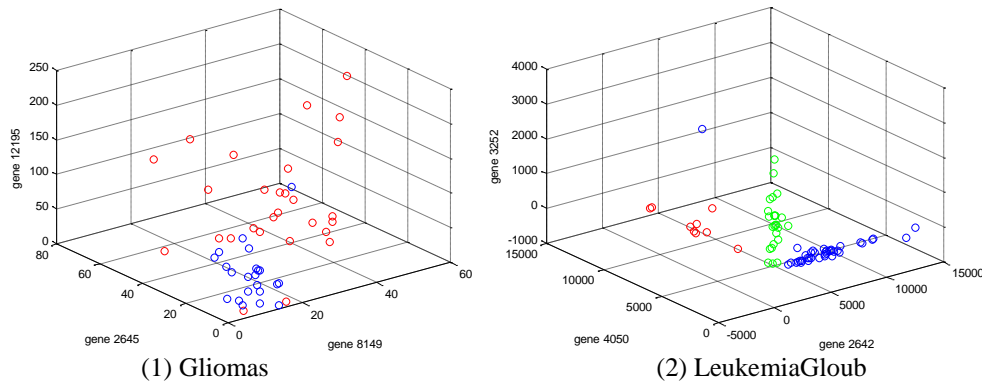
**Table 3. The Number of Feature Genes Selected of Different Methods**

Dataset	Raw	Kruskal-wallis	ReliefF	ReliefF+RS	Our method
Gliomas	12625	10	380	10	<b>3</b>
DLBCL	7129	20	20	7	<b>7</b>
LeukemiaGloub	7129	10	30	5	<b>4</b>
MLLLeukemia	12582	160	130	8	<b>4</b>
Lung	12600	30	40	15	<b>8</b>
ALL	12625	90	790	25	<b>23</b>

**Table 4. Feature Gene Selected by using Our Proposed Method**

DataSet	Feature gene ID
Gliomas	{8149, 2645, 12195}
DLBCL	{6179, 6575, 5077, 4546, 6361, 4028, 613}
LeukemiaGloub	{2642, 4050, 3252, 6613}
MLLLeukemia	{11325, 6019, 3768, 6716, 6979}
Lung	10175, 12097, 2478, 3845, 3919, 4786, 4977, 5920
ALL	{2610, 7715, 3331, 6387, 5478, 3010, 10228, 8556, 3340, 3102, 4842, 6738, 10145, 8868, 8827, 8750, 8808, 2479, 8063, 9456, 7477, 3295, 2693}

Figure 4 intuitive shows samples distribution in feature genes space on Gliomas and LeukemiaGloub. The results show that the boundary of different categories samples is clear, and it indicates only small number of feature genes can classify samples correctly. It verify feature gene selected is effective, meanwhile shows our proposed method is effective for selecting feature gene.



**Figure 4. Scatter Figure of Samples with Feature Genes Selected**

According to above analysis, our proposed method can improve classification accuracy while the number of feature genes is decreased. It shows that the method can improve classification accuracy by eliminating irrelevant and redundant genes of microarray gene data to reveal the essential characters and structure of the high dimensional data. The method is effective and efficient for feature gene selection. Especially, feature genes selected has an important meaning for understanding microarray data and finding pathogenic genes.

## 7. Conclusion

Because microarray data contains a lot of irrelevant and redundant genes, feature gene selection plays an important role. This paper presents a hybrid gene selection algorithm based on reliefF and fuzzy neighborhood rough set to improve classification performance. ReliefF can reduce irrelevant genes from original data and feature genes are obtained using fuzzy neighborhood rough set. Our proposed algorithm can deal with continuous attributes, without discretization requirement. Experiment results show our method obtains higher classification accuracy and fewer genes are selected. It indicates our algorithm is effective for the selection of feature gene in microarray data, and that has an important meaning for clinical research of cancer.

## Acknowledgements

This paper is supported by National Natural Science Foundation of China (81160183, 11305097), Scientific Research Program Funded by Shaanxi Provincial Education Department.

## References

- [1] M. B. Kursu, "Robustness of random forest-based gene selection methods", *BMC bioinformatics*, vol. 15, no. 1, (2014), pp. 1-8.
- [2] S. L. Wang, Y. H. Zhu, W. Jia and D. S. Huang, "Robust classification method of tumor subtype by using correlation filters", *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 9, no. 2, (2012), pp. 580-591.
- [3] S. Zhu, D. Wang and K. Yu, "Feature selection for gene expression using model-based entropy", *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 7, no. 1, (2010), pp. 25-36.
- [4] Y. Su, T. Murali, V. Pavlovic, M. Schaffer and S. Kasif, "RankGene: identification of diagnostic genes based on expression data", *Bioinformatics*, vol. 19, no. 12, (2003), pp. 1578-1579.
- [5] T. Li, C. Zhang and M. Ogihara, "A comparative study of feature selection and multiclass classification methods for tissue classification based on gene expression", *Bioinformatics*, vol. 20, no. 15, (2004), pp. 2429-2437.
- [6] H. Liu, J. Li and L. Wong, "A comparative study on feature selection and classification methods using gene expression profiles and proteomic patterns", *Genomics*, vol. 13, (2002), pp. 51-60.
- [7] A. Ben-Dor, L. Bruhn, N. Friedman, I. Nachman, M. Schummer and Z. Yakhini, "Tissue classification with gene expression profiles," *Journal of Computational biology*, vol. 7, nos. 3-4, (2000), pp. 559-583.
- [8] E. P. Xing, M. I. Jordan and R. M. Karp, "Feature selection for high-dimensional genomic microarray data," In *Proceedings of the Eighteenth International Conference on Machine Learning*, Morgan Kaufmann Publishers Inc, San Francisco, USA, (2001), pp. 601-608.
- [9] K. Kira and L. A. Rendell, "A practical approach to feature selection," *Proceedings of the ninth international workshop on Machine learning*, Morgan Kaufmann Publishers Inc., USA, (1992), pp. 249-256.
- [10] I. Kononenko, "Estimating attributes: analysis and extensions of RELIEF," *Proceedings of the European conference on machine learning*, Lecture notes in computer science, (1994), pp. 171-182.
- [11] H. Peng, F. Long and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and minredundancy", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, (2005), pp. 1226-1238.
- [12] X. Liu, A. Krishnan and A. Mondry, "An entropy-based gene selection method for cancer classification using microarray data," *BMC bioinformatics*, vol. 6, no.1, (2005), pp. 120-135.
- [13] R. Battiti, "Using mutual information for selecting features in supervised neural net learning," *Neural Networks*, *IEEE Transactions on*, vol. 5, no. 4, (1994), pp. 537-550.
- [14] Q. H. Hu, W. Pan, S. An, P. j. Ma and J. m. Wei, "An efficient gene selection technique for cancer recognition based on neighborhood mutual information," *International Journal of Machine Learning and Cybernetics*, vol.1, no. 4, (2010), pp. 63-74.
- [15] Z. Pawlak, "Rough sets", *Int'l Journal of Comp. & Information Sciences*, vol. 11, no. 5, (1982), pp. 341-356.
- [16] Q. Shen and A. Chouchoulas, "A rough-fuzzy approach for generating classification rules", *Pattern Recognition*, vol. 35, no. 11, (2002), pp. 2425-2438.
- [17] Q. H. Hu, D. R. Yu, J. F. Liu and C. X. Wu, "Neighborhood rough set based heterogeneous feature subset selection", *Information Sciences*, vol. 178, no. 18, (2008), pp. 3577-3594.
- [18] H. Zhao, "Intrusion detection ensemble algorithm based on bagging and neighborhood rough set", *International Journal of Security and Its Applications*, vol. 7, no. 5, (2013), pp. 193-204.
- [19] T. Chen, "Classification algorithm on gene expression profiles of tumor using neighborhood rough set and support vector machine", *Advanced Materials Research*, no. 850, (2014), pp. 1238-1242.
- [20] D. Dubois and H. Prade, "Putting fuzzy sets and rough sets together", *Intelligent Decision Support*, Springer Netherlands, (1992), pp. 203-232.
- [21] N. N. Morsi and M. M. Yakout, "Axiomatics for fuzzy-rough sets", *Fuzzy Set system*, vol. 100, no. 1-3, (1998), pp. 327-342.
- [22] W. Wu and W. Zhang, "Constructive and axiomatic approaches of fuzzy approximation operators," *Informat. Sci.*, vol. 159, nos. 3-4, (2004), pp. 233-254.
- [23] W. Wu, J. Mi and W. Zhang, "Generalized fuzzy-rough sets", *Informat. Sci.*, vol. 151, (2004), pp. 263-282.
- [24] Q. Shen and R. Jensen, "Selecting informative features with fuzzy-rough sets and its application for complex systems monitoring", *Pattern Recognition*, vol. 37, no.7, (2004), pp. 1351-1363.

- [25] R. Jensen and Q. Shen, "Fuzzy-rough attribute reduction with application to web categorization", *Fuzzy Sets System*, vol. 141, (2004), pp. 469-485.
- [26] Q. H. Hu, D. R. Yu and Z. X. Xie, "Information preserving hybrid data reduction based on fuzzy rough techniques", *Pattern Recognition Letters*, vol. 27, no. 5, (2006), pp. 414-423.
- [27] S. Das and P. N. Suganthan, "Differential evolution: a survey of the state-of-the-art", *Evolutionary Computation, IEEE Transactions on*, vol. 15, no. 1, (2011), pp. 4-31.
- [28] F. I. De, "Differential evolution for automatic rule extraction from medical databases", *Applied Soft Computing*, vol. 13, no. 2, (2013), pp. 1265-1283.

## Authors



**Tao Chen**, he received his Master from Lanzhou University (2007), and received his B.S. degree from Shaanxi University of Technology (2001), respectively. He is currently pursuing the Ph.D. degree in the School of Automation at Northwestern Polytechnical University. He is currently an associate professor in the school of mathematics and computer science, Shaanxi University of Technology. His current research interests are focused on data mining, pattern recognition and computational biology.



**Zenglin Hong**, he received his Ph.D. from School of Automation at Northwestern Polytechnical University. He is currently professor with School of Automation at Northwestern Polytechnical University. His current research interests are focused on systems engineering, land resources management and the regional economy.



**Fang-an Deng**, he received his Ph.D. from School of Science at Xidian University. He is currently professor of School of Mathematics and Computer Science at Shaanxi University of Technology. His current research interests are focused on intelligent information processing, rough set and soft algebra.



**Man Cui**, he received her Master in the school of communication engineering from Xi'an University of Science and Technology (2011), and received her B.E. degree from Xi'an University of Science and Technology (2008), respectively. She is currently pursuing the Ph.D. degree in the School of Automation at Northwestern Polytechnical University. Her current research interests are focused on decision support.