Research Article

Feature Genes Selection Using Fuzzy Rough Uncertainty Metric for Tumor Diagnosis

Jiucheng Xu,1,2 Yun Wang,1 Keqiang Xu,1 and Tianli Zhang3

1College of Computer and Information Engineering, Henan Normal University, Xinxiang, Henan, China
2Engineering Technology Research Center for Computing Intelligence and Data Mining, Xinxiang, Henan, China
3College of Political Science and Public Administration, Henan Normal University, Xinxiang, Henan, China

Correspondence should be addressed to Yun Wang; 121084@htu.edu.cn

Received 20 September 2018; Revised 6 December 2018; Accepted 2 January 2019; Published 27 January 2019

Academic Editor: Markos G. Tsipouras

Copyright ©2019 Jiucheng Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To select more effective feature genes, many existing algorithms focus on the selection and study of evaluation methods for feature genes, ignoring the accurate mapping of original information in data processing. Therefore, for solving this problem, a new model is proposed in this paper: rough uncertainty metric model. First, the fuzzy neighborhood granule of the sample is constructed by combining the fuzzy similarity relation with the neighborhood radius in the rough set, and the rough decision is defined by using the fuzzy similarity relation and the decision equivalence class. Then, the fuzzy neighborhood granule and the rough decision are introduced into the conditional entropy, and the rough uncertainty metric model is proposed; in the meantime, the definition of measuring the significance of feature genes and the proof of some related theorems are given. To make this model tolerate noises in data, this paper introduces a variable precision model and discusses the selection of parameters. Finally, based on the rough uncertainty metric model, we design a feature genes selection algorithm and compare it with some existing similar algorithms. The experimental results show that the proposed algorithm can select the smaller feature genes subset with higher classification accuracy and verify that the model proposed in this paper is more effective.

1. Introduction

Nowadays, with the continuous changes of human lifestyle and environment, the incidence and mortality of cancers are rising. Therefore, how to improve the analysis, identification, and treatment of tumors has become one of the research hotspots of scholars [1]. Gene expression profiling data can approximately reflect the expression information of the entire genome of biological cells. With the development of gene chip technology, the accurate acquisition of gene expression profiling data becomes possible, which provides an important basis for clinical tumor diagnosis and tumor pathogenesis research [2–4]. However, among the large number of genes included in the gene expression profiling data, there are only a few important genes that can be used as information genes to track diseases [5, 6]. Therefore, when scholars process gene expression profiling data, they will be transformed into information genes selection problems, that is, feature genes selection, whose purpose is to reduce noises and redundant data in gene expression profiles and to obtain feature genes subsets with strong disease-recognition ability [5].

The gene expression profiling data contain complex and specific information. If the original information of the data can be applied to the calculation as accurately as possible, the result of feature genes selection will be improved to a large extent. The classical rough set proposed by Pawlak [7] has been extensively developed and studied [8, 9]. Its theory is based on the equivalence relation and can only deal with the discrete data. The processing of continuous data needs to be discretized, and it will face problems just like information loss. To solve this problem, neighborhood rough set [10–13] and fuzzy rough set [14–18] are successively proposed as two important models. The neighborhood rough set can directly process the continuous data, which overcomes the shortcomings of classical rough set, but it cannot accurately describe the fuzziness of samples under the fuzzy background.
In the fuzzy rough set, the description of a sample is usually depicted by its relationship with the neighbor samples, so the data noise will increase the risk of the calculation result and increase the classification error rate [19]. For this problem, the concept of fuzzy neighborhood is proposed in the literature [20], which overcomes the above deficiencies to some extent and constructs a fitting feature selection model based on the fuzzy rough set. It is crucial to find superior feature evaluation functions in the feature selection process, such as the dependence [19, 21, 22] and information entropy [23–25] methods that have already been proposed. As a kind of knowledge acquisition tool, rough set theory is gradually used in the analysis of gene expression profiling data, which uses the dependency function to evaluate the classification ability of feature genes subset, and has achieved good research results. However, the dependency function mainly depends on the positive domain and the boundary domain, which will lead to inaccurate measurement. At present, some scholars have introduced information entropy into rough sets, such as rough entropy [26] and conditional entropy [27, 28]. What some scholars mentioned in the literatures [29, 30] is that the positive domain and the boundary domain, which will lead to inaccurate measurement. At present, some scholars have introduced information entropy into rough sets, such as rough entropy [26] and conditional entropy [27, 28].

Let the universe of discourse \( U = \{x_1, x_2, \ldots, x_n\} \) be an \( n \)-dimensional real-value space, \( L = R^n \times R^n \rightarrow R \). Then, \( L \) is a measure on \( R^n \) and satisfies the following conditions:

1. \( L(x_1, x_2) \geq 0 \), and the equality holds up if and only if \( x_1 = x_2, \forall x_1, x_2 \in R^n \)
2. \( L(x_1, x_2) = L(x_2, x_1) \), \( \forall x_1, x_2 \in R^n \)
3. \( L(x_1, x_2) \leq L(x_1, x_3) + L(x_2, x_3) \), \( \forall x_1, x_2, x_3 \in R^n \)

For \( \forall x \subseteq U \), its neighborhood is expressed as \( \varepsilon(x_i) = \{ x \mid x \in U, L(x, x_i) \leq \varepsilon \} \) and \( \varepsilon > 0 \). According to the nature of metric,

1. \( \varepsilon(x_i) \neq \emptyset \)
2. \( x_i \in \varepsilon(x_i) \implies x_i \in \varepsilon(x_j), \forall x_j \subseteq U \)
3. \( \bigcup_{i=1}^{n} \varepsilon(x_i) = U \)

It can be seen from (3) that \( U \) is the full coverage of \( \{\varepsilon(x_i), i = 1, 2, \ldots, n\} \).

2.2. Fuzzy Relation. The fuzzy relation is described approximately by the membership degree of a sample on a set about an attribute. This expression is a good mathematical representation of clear and unclear concepts. Dubois and Prade [14] mentioned the basic fuzzy relations.

Let \( U = \{x_1, x_2, \ldots, x_n\} \) and a mapping be \( S(\cdot): U \rightarrow [0, 1] \) on the universe of discourse \( U \), then \( S \) is a fuzzy set. For \( \forall x \in U \), \( S(x) \) is the membership of \( x \) on \( S \), and \( F(U) \) is recorded as a fuzzy set on \( U \). \( B \) is an attribute set of the sample, which can induce a fuzzy binary relationship \( R_B \) on \( U \).

If \( \forall x, y \in F(U) \) are two fuzzy sets, their intersection, union, and complement are computed as follows:

1. \( (M \cap N)(x) = M(x) \cap N(x) = \min(M(x), N(x)) \)
2. \( (M \cup N)(x) = M(x) \cup N(x) = \max(M(x), N(x)) \)
3. \( M'(x) = 1 - M(x) \)

If \( R_B \) satisfies,

1. reflexivity: \( R_B(x, x) = 1, \forall x \in U \)
2. symmetry: \( R_B(x, y) = R_B(y, x), \forall x, y \in U \)

then, \( R_B \) is the fuzzy similarity relation on \( U \). For \( \forall x, y \in U \), the fuzzy neighborhood of \( x \) about \( R_B \) can be expressed as \( [x]_{R_B}(y) = R_B(x, y), y \in U \), which is a fuzzy set on \( U \).

2.3. Fuzzy Neighborhood Granule. The fuzzy relation can accurately describe the relationship between samples, but the strength of association between samples is different. To reduce the influence of redundant and noise in the data, the neighborhood radius can be set to filter out some weak correlation data to improve the computational efficiency. Wang et al. [20] proposed the concept of fuzzy neighborhood:
Let \( \langle U, A, D \rangle \) be a decision information system. For \( U = \{x_1, x_2, \ldots, x_n\} \), \( B \subseteq A \), \( \forall x, y \in U \), \( R_B \) is the fuzzy similarity relation on \( U \). For \( \forall b \in B \), the fuzzy neighborhood of \( x \) with respect to \( R_B \) is defined as

\[
[x]_{R_B}^a(y) = R_B(x, y) = \min_{b \in B} R_b(x, y),
\]

where \( R_b(x, y) = 1 - |x_b - y_b| \) and \( x_b \) and \( y_b \), respectively, represent the value of the corresponding attribute \( b \). For \( \forall B \in A \), the fuzzy neighborhood granule of \( x \) with respect to \( R_B \) is defined as

\[
[x]^a_{R_B}(y) = \begin{cases} R_B(x, y), & R_B(x, y) \geq \alpha, \\ 0, & R_B(x, y) < \alpha, \end{cases}
\]

where \( \alpha \) is the fuzzy neighborhood radius, and when \( \alpha_1 \geq \alpha_2 \), \( [x]_{R_B}^{\alpha_2}(y) \) represents the value of the corresponding attribute \( \alpha \).

3. Rough Uncertainty Metric

In this section, the paper introduces the rough decision by using the fuzzy similarity relation between samples to further determine the inclusion of equivalence classes accurately. Combining the rough decision with the fuzzy neighborhood granule of the sample, the definition of the conditional entropy with respect to the attribute set and the proof of its related theorems are given. Then, the rough uncertainty metric model is proposed, and the corresponding feature genes selection algorithm is designed according to the proposed model.

3.1. Rough Decision. Generally, the decision equivalence classes of the general domain are divided by decision attributes, but there is always a correlation between samples. For example, a fuzzy neighborhood granule of a sample contains samples with different equivalent decision classes. Therefore, a sample cannot be completely categorized into a decision equivalence class. In this paper, the fuzzy similarity relation is used to calculate the membership degree of each sample for different decision equivalence classes, and a more accurate rough decision is proposed.

Definition 1. Let \( \langle U, A, D \rangle \) be a decision information system. \( U = \{x_1, x_2, \ldots, x_n\} ; U/D = \{D_1, D_2, \ldots, D_r\} \), \( R_A \) is the fuzzy similarity relation on \( U \), and the rough decision \( RD \) is defined as follows:

\[
RD = \{RD_1^T, RD_2^T, \ldots, RD_r^T\},
\]

\[
RD_j = \{RD_j(x_1), RD_j(x_2), \ldots, RD_j(x_n)\},
\]

\[
RD_j(x_i) = \sum_{d \in D_j} \left[ \sum_{[x]_{R_A}^a(y)} x_i \right],
\]

where \( d \in D_j, i = 1, 2, \ldots, n \), and \( j = 1, 2, \ldots, r \).

3.2. Rough Uncertainty Metric Model. In the data characterization stage, the original information of the data is restored as much as possible. Next, fuzzy neighborhood granule and rough decision are used to realize the uncertainty metric of feature genes, and the evaluation method of the significance of feature genes is given.

Definition 2. Let \( \langle U, A, D \rangle \) be a decision information system. \( \forall B, C \subseteq A \) are two conditional attribute subsets, and \( [x]_{R_B}^{\alpha}(y) \) is the fuzzy neighborhood granule of \( x \) with radius \( \alpha \). The rough entropy with respect to \( B \) is defined as

\[
E_c(B) = - \frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \left( \frac{1}{|x]_{R_B}^\alpha(y)|} \right).
\]

The joint entropy with respect to \( B, C \) is defined as

\[
E_c(B, C) = - \frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \left( \frac{1}{|x]_{R_B}^\alpha(y) \cap [x]_{R_C}^\alpha(y)|} \right).
\]

where \( \forall x, y \in U \), \( |[x]_{R_B}^\alpha(y)| \) represents the number of nonzero values in the fuzzy neighborhood granule of the object \( x \), and \( 1/|[x]_{R_B}^\alpha(y) \cap [x]_{R_C}^\alpha(y)| \) represents the probability of an element in the fuzzy neighborhood granule \( [x]_{R_B}^\alpha(y) \cap [x]_{R_C}^\alpha(y) \).

Definition 3. Let \( S \) and \( T \) be two fuzzy sets. \( |S \cap T| \) is defined as the number of nonzero values of objects whose membership degree in \( S \) is not greater than that of \( T \).

Example 1. Given a sample set \( U = \{x_1, x_2, \ldots, x_{10}\} \), \( S \) and \( T \) are two fuzzy sets on \( U \), the membership of the samples on them are assumed as follows:

\[
S = \begin{bmatrix} 0.7 & 0.4 & 0.5 & 0.7 & 0.8 & 0.3 & 0.5 & 0.1 & 0.2 & 0.6 \\ x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 & x_9 & x_{10} \end{bmatrix},
\]

\[
T = \begin{bmatrix} 0.5 & 0.4 & 0.3 & 0.6 & 0.4 & 0.7 & 0.4 & 0.9 & 0.2 & 0.3 \\ x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 & x_9 & x_{10} \end{bmatrix},
\]

\( |S \cap T| \) represents the number of nonzero values of objects whose membership degree in \( S \) is not greater than that of \( T \). Thus,

\[
|S \cap T| = \left| \{x_2, x_4, x_9, x_{10}\} \right| = 4.
\]

Similarly, \( |T \cap S| \) represents the number of nonzero values of objects whose membership degree in \( T \) is not greater than that of \( S \):

\[
|T \cap S| = \left| \{x_1, x_2, x_3, x_4, x_5, x_7, x_9, x_{10}\} \right| = 8.
\]

Definition 4. Let \( \langle U, A, D \rangle \) be a decision information system. \( U/D = \{D_1, D_2, \ldots, D_r\} \), \( B \subseteq A \) is a conditional attribute subset, fuzzy neighborhood granule \([x]_{R_B}^\alpha(y)\), and rough decision \( RD \) are two fuzzy matrices. Then, the conditional entropy of the decision attribute set \( D \) with respect to \( B \) on \( U \) is defined as

\[
E_c(RD | B) = - \frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \left( \frac{|[x]_{R_B}^\alpha(y) \cap [x]_{RD}^\alpha(y)|}{|[x]_{R_B}^\alpha(y)|} \right).
\]

where \([x]_{RD}^\alpha(y)\) represents the rough decision \( RD_j^T \) corresponding to the equivalence class \( D_j \) to which the sample \( x_i \) belongs.
Example 2. Given a decision table \( \langle U, A, D \rangle \), where \( U = \{x_1, x_2, \ldots, x_5\} \), \( U/D = \{D_1, D_2\} \), \( D_1 = \{x_1, x_2, x_3\} \), and \( D_2 = \{x_4, x_5\} \), \([x_i]_{R_A}(y)\) is the fuzzy similarity relation matrix with respect to attribute set \( A \) and

\[
[x]_{R_A}(y) = \begin{bmatrix}
R_A(x_1, x_1) & \cdots & R_A(x_1, x_5) \\
\vdots & \ddots & \vdots \\
R_A(x_5, x_1) & \cdots & R_A(x_5, x_5)
\end{bmatrix}
\]

where \( R_A \) is a function of the fuzzy similarity relation. Then, \( RD_1 = \{0, 0.2689, 0.2027, 0.8458, 0.6028\} \). So, the rough decision is as follows:

\[
\begin{bmatrix}
1 & 0 \\
0.7311 & 0.2689 \\
0.1542 & 0.8458 \\
0.3972 & 0.6028
\end{bmatrix}
\]

Here, the neighborhood radius \( \alpha = 0 \) (the specific experiment will discuss the parameters according to different data sets). According to Definition 2, \( \|x_i\|_{R_A}^\alpha(y)\) represents the number of nonzero values in the fuzzy neighborhood granule of the object \( x_i \), and then,

\[
\|x_1\|_{R_A}^\alpha(y) = \|1, 0.53, 0.7, 0, 0\| = 3,
\]

\[
\|x_2\|_{R_A}^\alpha(y) = \|0.53, 1, 0.7, 0.2, 0.62\| = 5,
\]

\[
\|x_3\|_{R_A}^\alpha(y) = \|0.7, 0.7, 1.11, 0.5\| = 5,
\]

\[
\|x_4\|_{R_A}^\alpha(y) = \|0, 0.2, 1.11, 1.0.7\| = 4,
\]

\[
\|x_5\|_{R_A}^\alpha(y) = \|0, 0.6.2, 0.5, 0.7, 1\| = 4,
\]

where \( \|x_i\|_{RD}^\alpha(y) \) represents the rough decision \( RD_T^\alpha \) corresponding to the equivalence class \( D_r \) to which the sample \( x_i \) belongs, where \( r = 1, 2 \). Since \( x_1, x_2, x_3 \in D_1 \) and \( x_4, x_5 \in D_2 \),

\[
\begin{align*}
[x_1]_{RD}^\alpha(y) & = \{1, 0.7311, 0.7973, 0.1542, 0.3972\}, \\
[x_2]_{RD}^\alpha(y) & = \{1, 0.7311, 0.7973, 0.1542, 0.3972\}, \\
[x_3]_{RD}^\alpha(y) & = \{1, 0.7311, 0.7973, 0.1542, 0.3972\}, \\
[x_4]_{RD}^\alpha(y) & = \{0, 0.2689, 0.2027, 0.8458, 0.6028\}, \\
[x_5]_{RD}^\alpha(y) & = \{0, 0.2689, 0.2027, 0.8458, 0.6028\}.
\end{align*}
\]

According to Definition 3, \( [x_i]_{RD}^\alpha(y) \) is similar to \( S \) and \( [x_i]_{RD}^\alpha(y) \) is similar to \( T \) in it. Then,

\[
[x_1]_{R_A}^\alpha(y) \cap [x_1]_{RD}^\alpha(y) = 3,
\]

\[
[x_2]_{R_A}^\alpha(y) \cap [x_2]_{RD}^\alpha(y) = 2,
\]

\[
[x_3]_{R_A}^\alpha(y) \cap [x_3]_{RD}^\alpha(y) = 3,
\]

\[
[x_4]_{R_A}^\alpha(y) \cap [x_4]_{RD}^\alpha(y) = 2,
\]

\[
[x_5]_{R_A}^\alpha(y) \cap [x_5]_{RD}^\alpha(y) = 1.
\]

The remaining samples are equally available. Therefore, the conditional entropy of the decision attribute set \( D \) with respect to \( A \) on \( U \) is
\[ E_c(\text{RD} \mid A) = \frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \left[ \frac{|x_i|_{RD}^a(y)}{|x_i|_{RD}^a(y) \cap |x_i|_{RD}(y)} \right] = \frac{1}{5} \left( \log_2^3 + \log_2^5 + \log_2^5 + \log_2^4 + \log_2^4 \right) = 1.0118. \]  

**Theorem 1.** Let \( \{U, A, D\} \) be a decision information system, \( U/D = \{\{x_1\}_U, \ldots, \{x_U\}_U\} \), \( B \subseteq A \) is a conditional attribute subset, RD is a rough decision defined on decision attribute set D. Then,  

\[ E_c(\text{RD} \mid B) = E_c(B) - E_c(\text{RD}, B). \]  

**Proof.** According to Definitions 4 and 2, the derivation process is as follows:  

\[ E_c(\text{RD} \mid B) = \frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \left[ \frac{|x_i|_{RD}^a(y)}{|x_i|_{RD}^a(y) \cap |x_i|_{RD}(y)} \right] - \log_2 \left[ \frac{|x_i|_{RD}^a(y) \cap |x_i|_{RD}(y)}{|x_i|_{RD}^a(y) \cap |x_i|_{RD}(y)} \right] \]

\[ = \frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \left[ \frac{|x_i|_{RD}^a(y)}{|x_i|_{RD}^a(y) \cap |x_i|_{RD}(y)} \right] - \frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \left[ \frac{|x_i|_{RD}^a(y) \cap |x_i|_{RD}(y)}{|x_i|_{RD}^a(y) \cap |x_i|_{RD}(y)} \right] \]

\[ = E_c(B) - E_c(\text{RD}, B). \]  

Hence,  

**Theorem 2.** Let \( \{U, A, D\} \) be a decision information system, \( \forall B \subseteq A, E_c(\text{RD} \mid B) \geq 0 \).  

**Proof.** Assume \( E_c(\text{RD} \mid B) < 0 \) equivalent to \( \log_2 \left( \frac{|x_i|_{RD}^a(y)}{|x_i|_{RD}^a(y) \cap |x_i|_{RD}(y)} \right) < 0 \), that is, \( |x_i|_{RD}^a(y) \cap |x_i|_{RD}(y)) \), obviously not established. So \( \forall B \subseteq A, E_c(\text{RD} \mid B) \geq 0 \).  

**Theorem 3.** Let \( \{U, A, D\} \) be a decision information system, \( \forall M, N \subseteq A \) are two conditional attribute subsets. If \( |x_i|_{RD}^a(y) \subseteq |x_i|_{RD}^a(y) \), then \( E_c(\text{RD} \mid M) \leq E_c(\text{RD} \mid N) \); the equation is true if and only if \( |x_i|_{RD}^a(y) = |x_i|_{RD}^a(y) \).
Definition 5. Let \( \langle U, A, D \rangle \) be a decision information system, \( \forall a \in A \), the significance of condition attribute \( a \) relative to \( A \) is defined as
\[
\text{SIG}(a, A, D) = E_c(\text{RD} | A - \{a\}) - E_c(\text{RD} | A).
\] (21)

\( \forall r \in A - \text{red} \), the significance of condition attribute \( r \) relative to \( \text{red} \) is defined as
\[
\text{SIG}(r, \text{red}, D) = E_c(\text{RD} | \text{red}) - E_c(\text{RD} | \text{red} \cup \{r\}).
\] (22)

3.3. Feature Genes Selection Algorithm Based on Rough Uncertainty Metric Model. For the above theory, this paper designs a feature genes selection algorithm based on the rough uncertainty metric model. As shown in Algorithm 1, the application of the new model in feature genes selection is realized.

4. Experimental Results and Analysis

In addition to the richness of the theory, a good model needs to have a good practical effect. Therefore, the experimental contrast analysis is set up in this part. Under the same conditions, the same genetic data are used to compare our algorithm with other existing similar algorithms. The specific experimental results data are used to illustrate the advantages of the proposed model.

4.1. Experiment Preparation. In order to verify the validity of the proposed model, four data sets are selected from the public data sources as experimental objects. The specific information is shown in Table 1. The data sets WPBC, WDBC, and Heart-Cle are selected from the UCI Machine Learning Repository, and Colon is selected from http://datam.ir.a-star.edu.sg/datasets/krbd/. At the beginning of the data processing stage, in order to eliminate the influence of value dimension inconsistency among features, all numerical experimental data will be normalized and mapped to \([0, 1]\) by the formula \( f(x_j) = (x_j - x_{\text{min}}) / (x_{\text{max}} - x_{\text{min}}) \). In addition, the fuzzy similarity relation of samples \( x_i \) and \( x_j \) with respect to an attribute is calculated as
\[
\begin{align*}
r_{ij} = \begin{cases} 
1 - |x_i - x_j|, & |x_i - x_j| \leq 1 - \alpha, \\
0, & |x_i - x_j| > 1 - \alpha.
\end{cases}
\end{align*}
\] (23)

4.2. Parameter Discussion. In order to reduce the influence of noise and redundancy in the data used in the experiment, this paper sets the parameter \( \alpha \) as the neighborhood radius to calculate the fuzzy neighborhood granule of the sample, which will filter out the data that are less relevant to the sample and consumes computation time as much as possible, improving the efficiency of the experiment to some extent. Due to the objective conditions in reality, these noise and redundant data can only be minimized but cannot be completely avoided. Therefore, the experimental results need to tolerate the error effects caused by these data. So, the parameter \( \beta \) is set in this paper to control the size of the error in the experiment; thus, a smaller feature genes subset with higher classification accuracy is selected. Since different data sets have different correlation strengths, the parameters need to be set separately according to different data sets.

This paper sets the parameters \( \alpha \) and \( \beta \) to vary from 0 to 0.5, respectively, with an interval of 0.05. For different data sets, the experiment compares the number of feature genes and the corresponding classification accuracy obtained by different parameters. Two classifiers, support vector machine (linear SVM) and K-nearest neighbor (KNN, \( K = 3 \)), are used to evaluate the classification accuracy of feature genes subset by 10-fold cross-validation. The comparison process is shown in Figures 1–4 (the classification accuracy in the figures is based on the linear-SVM classifier, which is the same as the experimental result under the 3NN classifier). Finally, the selected feature genes subset and corresponding suitable parameters under different data sets are shown in Table 2.

4.3. Experimental Comparison. In this section, in order to verify the validity of the proposed model, the designed algorithm is compared with the existing similar algorithms. The experimental objects include original data without algorithm processing (raw data), feature selection algorithm based on fuzzy entropy (FISEN) [32], feature selection algorithm based on fuzzy neighborhood rough set (FNRS) [19], and algorithm proposed in this paper (RUM). The comparison includes two aspects: the number of selected feature genes and the classification accuracy.

A good feature genes selection algorithm aims at finding a subset of smaller feature genes that make classification more accurate. First, the number of original data and selected feature genes under different algorithms are shown in Table 3. It is not difficult to find from the table that the number of feature genes in the original data is more, and significant reduction is achieved by different algorithms. Comparing the average number of feature genes, the FNRS algorithm is lower than the FISEN algorithm, and the RUM algorithm is the least among all the compared algorithms. In different data sets, the RUM algorithm is also the least compared to other comparison algorithms. Therefore, the RUM algorithm proposed in this paper is superior in terms of the number of selected feature genes.

Certainly, only relying on the minimum number of selected feature genes is not enough to illustrate the advantages of an algorithm. If the experimental results can achieve equal or even higher classification accuracy on a relatively small subset of feature genes, the algorithm can be proved to be excellent. Table 4 shows the classification accuracy of the original data and the data obtained from the different algorithms under classifier linear-SVM. In general, the RUM algorithm proposed in this paper is the highest among the four experimental objects in terms of average classification accuracy. From the perspective of a single data set, the two classification accuracies obtained by the RUM algorithm and the FNRS algorithm are the same on the WDBC data set, which are higher than others, and the same
Input: rough decision information system \( \langle U, A, D \rangle \), parameters \( \alpha \) and \( \beta \) (\( \alpha \) is fuzzy neighborhood radius and \( \beta \) as constraint to tolerate noises in data)

Output: one set of feature genes \( \text{red} \)

Step 1. Initialize \( \text{red} = \emptyset \)

Step 2. \( \forall a \in A \), calculate the fuzzy similarity matrix \( R_a \)

Step 3. According to Definition 1, calculate the rough decision \( \text{RD} \)

Step 4. According to the definition of fuzzy neighborhood granule and Definition 4, combine with the rough decision \( \text{RD} \) to calculate the conditional entropy \( E_c(\text{RD} \mid A) \) of decision attribute set \( D \) with respect to \( A \)

Step 5. \( \forall a \in A \), calculate \( \text{SIG}(a, A, D) \), and select the attribute \( a \) that maximizes the value of \( \text{SIG}(a, A, D) \). Then, \( \text{red} = \text{red} \cup \{a\} \)

Step 6. \( \forall a \in A \), calculate \( E_c(\text{RD} \mid \text{red}) \), if \( E_c(\text{RD} \mid \text{red}) - E_c(\text{RD} \mid A) \leq \beta \), execute Step 9, otherwise execute Step 7

Step 7. Let \( B = A - \text{red} \), \( \forall r \in B \), calculate \( \text{SIG}(r, \text{red}, D) \)

Step 8. Select the attribute \( r \) that maximizes the value of \( \text{SIG}(r, \text{red}, D) \), then \( B = B - \{r\}, \text{red} = \text{red} \cup \{r\} \), execute Step 6

Step 9. Return \( \text{red} \)

**Algorithm 1:** Feature genes selection based on the rough uncertainty metric.

### Table 1: Description of data sets.

| Data sets | Sample | Attributes | Classes |
|-----------|--------|------------|---------|
| WPBC      | 198    | 32         | 2       |
| WDBC      | 569    | 30         | 2       |
| Colon     | 62     | 2000       | 2       |
| Heart-Cle | 303    | 13         | 5       |

**Figure 1:** Accuracy varying with parameters \( \alpha \) and \( \beta \) (WPBC).

**Figure 2:** Accuracy varying with parameters \( \alpha \) and \( \beta \) (WDBC).

**Figure 3:** Accuracy varying with parameters \( \alpha \) and \( \beta \) (Colon).

**Figure 4:** Accuracy varying with parameters \( \alpha \) and \( \beta \) (Heart-Cle).

### Table 2: Suitable parameters and selected feature genes of different data sets.

| Data sets | Values of \( \alpha \) | Values of \( \beta \) | Selected feature genes |
|-----------|------------------------|-----------------------|------------------------|
| WPBC      | 0.1                    | 0.25                  | 1, 32, 24              |
| WDBC      | 0.1                    | 0.1                   | 28, 22, 1, 25, 7       |
| Colon     | 0.1                    | 0.3                   | 1895, 267, 897, 1990   |
| Heart-Cle | 0.35                   | 0.35                  | 9, 12, 11, 7, 13       |
on the Colon data set. On the remaining data sets, the classification accuracy obtained by the RUM algorithm is the highest, especially on the WPBC data set, which is about 7-8 percent higher than the FNRS algorithm.

Comparing the classification accuracy on a classifier alone, the persuasive power may be insufficient. In this paper, the classifier 3NN is added to further verify the advantages of the proposed RUM algorithm and is shown in Table 5. First, the average classification accuracy obtained by the RUM algorithm is the highest. Apparently, the classification accuracy of the RUM algorithm on the WDBC data set is slightly lower than that of the original data and the FNRS algorithm, but this is not enough to show that the RUM algorithm is not good. It can be seen from Table 3 that the number of feature genes obtained by the RUM algorithm on the WDBC data set is the least, even less than one-third of the FNRS algorithm and the FISEN algorithm, which is only one-sixth of the original data. In the meantime, the classification accuracy achieved by the RUM algorithm is only less than 1 percent lower than the original data and FNRS algorithm, and higher than the FISEN algorithm. In addition, on the other three data sets, data selected by the RUM algorithm has the highest classification accuracy, especially on the WPBC and Colon data sets, which is about 10 percent higher than other algorithms.

From the above concept, the advantage of the rough uncertainty metric model in the feature genes selection has been well verified by comparing the number of selected feature genes and analyzing the classification accuracy on the two classifiers.

5. Conclusions and Future Works

A novel model is established in this paper. In this model, the fuzzy neighborhood granule of the sample are constructed by combining the fuzzy concept with the neighborhood concept, and the decision equivalence class is further accurately expressed as rough decision by using the fuzzy similarity relation between samples. The original information between the data is reserved as perfectly as possible during the data characterization phase. Then fuzzy neighborhood granule and rough decision are introduced into conditional entropy, and a metric method is proposed to evaluate the significance of feature genes. Based on the nature of this model, which is proved by the four theorems in the paper, a feature gene selection algorithm is designed. Finally, the proposed algorithm is compared with the existing similar algorithms on the common data sets. The experimental results show that the proposed algorithm can obtain a relatively small subset of feature genes and achieve better classification results, which verify the validity of the proposed model. However, this model still has something inadequate: the selected parameters based on the corresponding single data set are not generalized, requiring further study and improvement in the future work. In the next step, our study will focus on the parameter selection problem and find out how to adaptively set the appropriate generalized parameters.

Data Availability

Four tumor microarray data sets used to support the findings of this study have been deposited in the public data sources, which include UCI Machine Learning Repository and http://datam.i2r.a-star.edu.sg/datasets/krrbd/.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. 61772176, 61370169, and 61402153), the Plan for Scientific Innovation Talent of Henan Province (No. 184100510003), the Science and Technology Department, Henan Province (Nos. 182102210362 and 162102210261), the Young Scholar Program of Henan Province (No. 2017GGJS041), and the Key Scientific and Technological Project of Xinxiang City of China (No. CXGG17002).

References

[1] J. C. Xu, T. Li, L. Sun, and Y. H. Li, “Feature gene selection based on SNR and neighborhood rough set,” Journal of Data Acquisition and Processing, vol. 30, no. 5, pp. 973–981, 2015.
[2] M. Kohl, “Analysis of microarray data,” Journal of Computational Biology, vol. 1, no. 1, pp. 207–208, 2012.
[3] J. C. Xu, T. H. Xu, L. Sun, and J. Y. Ren, “Feature selection for cancer classification based on neighborhood rough set and
particle swarm optimization,” Journal of Chinese Computer Systems, vol. 35, no. 11, pp. 2528–2532, 2014.

[4] K. H. Chen, K. J. Wang, K. M. Wang, and M. A. Angelia, “Applying particle swarm optimization-based decision tree classifier for cancer classification on gene expression data,” Applied Soft Computing, vol. 24, pp. 773–780, 2014.

[5] J. C. Xu, S. Feng, and H. Y. Mu, "Tumor feature gene selection based on SNR and random forest," Journal of Henan Normal University (Natural Science Edition), vol. 45, no. 2, pp. 87–92, 2017.

[6] T. H. Xu, Y. Y. Ma, and J. C. Xu, "Efficient gene selection technique based on maximum neighborhood mutual information and particle swarm optimization," Journal of Chinese Computer Systems, vol. 37, no. 8, pp. 1775–1779, 2016.

[7] Z. Pawlak, "Rough sets," International Journal of Computer and Information Sciences, vol. 11, no. 5, pp. 341–356, 1982.

[8] G. Lang, Q. Li, and T. Yang, "An incremental approach to attribute reduction of dynamic set-valued information systems," International Journal of Machine Learning and Cybernetics, vol. 5, no. 5, pp. 775–788, 2014.

[9] J. Y. Liang, F. Wang, C. Y. Dang, and Y. H. Qian, "A group incremental approach to feature selection applying rough set technique," IEEE Transaction on Knowledge and Data Engineering, vol. 26, no. 2, pp. 294–304, 2014.

[10] Q. Hu, D. Yu, J. Liu, and C. Wu, "Neighborhood rough set based heterogeneous feature subset selection," Information Sciences, vol. 178, no. 18, pp. 3577–3594, 2008.

[11] H. Chen, T. Li, Y. Cai, C. Luo, and H. Fujita, "Parallel attribute reduction in dominance-based neighborhood rough set," Information Sciences, vol. 373, pp. 351–368, 2016.

[12] C. Z. Wang, Q. H. Hu, X. Z. Wang, D. G. Chen, Y. H. Qian, and Z. Dong, "Feature selection based on neighborhood discrimination index," IEEE Transactions on Neural Networks and Learning Systems, vol. 29, no. 7, pp. 2986–2999, 2018.

[13] J. Xu, H. Mu, Y. Wang, and F. Huang, "Feature genes selection using supervised locally linear embedding and correlation coefficient for microarray classification," Computational and Mathematical Methods in Medicine, vol. 2018, Article ID 5490513, 11 pages, 2018.

[14] D. Dubois and H. Prade, "Rough fuzzy sets and fuzzy rough sets", International Journal of General Systems, vol. 17, no. 2-3, pp. 191–209, 2007.

[15] R. Jensen and Q. Shen, "Fuzzy-rough attribute reduction with application to web categorization," Fuzzy Sets and Systems, vol. 141, no. 3, pp. 469–485, 2004.

[16] D. Chen and Y. Yang, "Attribute reduction for heterogeneous data based on the combination of classical and fuzzy rough set models," IEEE Transactions on Fuzzy Systems, vol. 22, no. 5, pp. 1325–1334, 2014.

[17] Q. Hu, D. Yu, W. Pedrycz, and D. Chen, "Kernelized fuzzy rough sets and their applications," IEEE Transactions on Knowledge and Data Engineering, vol. 23, no. 11, pp. 1649–1667, 2011.

[18] A. Mieszkowicz-Rolka and L. Rolka, Variable Precision Fuzzy Rough Sets, Springer, Berlin, Germany, 2004.

[19] C. Wang, M. Shao, Q. He, Y. Qian, and Y. Qi, "Feature subset selection based on fuzzy neighborhood rough sets," Knowledge-Based Systems, vol. 111, pp. 173–179, 2016.

[20] C. Wang, Y. Qi, M. Shao et al., "A fitting model for feature selection with fuzzy rough sets," IEEE Transactions on Fuzzy Systems, vol. 25, no. 4, pp. 741–753, 2017.

[21] Y. Qian, J. Liang, W. Pedrycz, and C. Dang, "Positive approximation: an accelerator for attribute reduction in rough set theory," Artificial Intelligence, vol. 174, no. 9-10, pp. 597–618, 2010.

[22] Z. Meng and Z. Shi, "A fast approach to attribute reduction in incomplete decision systems with tolerance relation-based rough sets," Information Sciences, vol. 179, no. 16, pp. 2774–2793, 2009.

[23] D. Yao, X. Liu, X. Zhang, and C. Wang, "Type-2 fuzzy cross-entropy and entropy measures and their applications," Journal of Intelligent and Fuzzy Systems, vol. 30, no. 4, pp. 2169–2180, 2016.

[24] J. Dai, W. Wang, and Q. Xu, "An uncertainty measure for incomplete decision tables and its applications," IEEE Transactions on Cybernetics, vol. 43, no. 4, pp. 1277–1289, 2013.

[25] S. Yao, J. Wang, F. Xu, and J. Chen, "Uncertainty measure and attribute reduction in inconsistent neighborhood rough set," Journal of Chinese Computer Systems, vol. 39, no. 4, pp. 700–706, 2018.

[26] D. Małyszko and J. Stepaniuk, "Adaptive multilevel rough entropy evolutionary thresholding," Information Sciences, vol. 180, no. 7, pp. 1138–1158, 2010.

[27] L. Sun, J. Xu, and Y. Tian, "Feature selection using rough entropy-based uncertainty measures in incomplete decision systems," Knowledge-Based Systems, vol. 36, pp. 206–216, 2012.

[28] N. Zhang and N. B. Fan, "Heuristic attribute reduction based on neighborhood approximate conditional entropy," Application Research of Computers, vol. 35, no. 5, 2018.

[29] Q. H. Hu, D. R. Yu, and Z. X. Xie, "Numerical attribute reduction based on neighborhood granulation and rough approximation," Journal of Software, vol. 19, no. 3, pp. 640–649, 2008.

[30] L. Q. Xie, J. S. Lei, and F. F. Xu, "Vibrant fault diagnosis for hydro-turbine generating unit based on improved neighborhood rough sets and PNN," Journal of Shanghai University of Electric Power, vol. 32, no. 2, pp. 181–187, 2016.

[31] E. Rashidi and H. Nezamabadi-pour, "Feature subset selection using improved binary gravitational search algorithm," Journal of Intelligent and Fuzzy Systems, vol. 26, no. 3, pp. 693–696, 2014.

[32] Q. Hu, D. Yu, and Z. Xie, "Information-preserving hybrid data reduction based on fuzzy-rough techniques," Pattern Recognition Letters, vol. 27, no. 5, pp. 414–423, 2006.
