Article

Combining Fuzzy C-Means Clustering with Fuzzy Rough Feature Selection

Ruonan Zhao, Lize Gu * and Xiaoning Zhu

Institute of Cyberspace Security Beijing University of Posts and Telecommunications, Beijing 100876, China; zrn@bupt.edu.cn (R.Z.); xiaoning158@bupt.edu.cn (X.Z.)

* Correspondence: glzisc@bupt.edu.cn

Received: 17 January 2019; Accepted: 12 February 2019; Published: 16 February 2019

Abstract: With the rapid development of the network, data fusion becomes an important research hotspot. Large amounts of data need to be preprocessed in data fusion; in practice, the features of datasets can be filtered to reduce the amount of data. The feature selection based on fuzzy rough sets can process a large number of continuous and discrete data to reduce the data dimension, making the selected feature subset highly correlated with the classification but less dependent on other features. In this paper, a new method of fuzzy rough feature selection is proposed which combines the membership function determination method of fuzzy c-means clustering and fuzzy equivalence to the original selection. Different from the existing research, our method takes full advantage of knowledge about the dataset itself and the differences between datasets, which makes the features selected have a higher correlation with the classification, improves the classification accuracy, and reduces the data dimension. Experimental results on the UCI machine learning repository datasets confirmed the performance and effectiveness of our method. Compared to the existing method, smaller subsets of features and an average of 1% higher classification accuracies were achieved.

Keywords: feature selection; fuzzy set; rough set; membership function

1. Introduction

We live in an age of data explosion—dealing with the growth of datasets usually requires much time and expense if we use the existing computers and algorithms. We want the dataset to contain more and more features to increase the likelihood of distinguishing different categories. Unfortunately, it may not be right. A higher-dimensional dataset increases the possibility of discovering incompletely valid false patterns. An effective way of resolving this problem is to select some of the most relevant and informative features from the dataset and eliminate redundant or irrelevant features. Unlike other dimensional reduction methods, feature selection retains the original meaning of features. This method can effectively reduce the size of a dataset without influencing the information expressed by the data, thus reducing cost and saving time.

Researchers have carried out different definitions of feature selection. Ideally, feature selection is to find the minimum feature subset which is necessary and sufficient to identify the target [1]. The definition of feature selection from the angle of improving the prediction accuracy is a process which can increase the classification accuracy or reduce the characteristic dimension without lowering the classification accuracy [2]. The basic method of feature selection is to generate a feature subset (search algorithm) and then evaluate the subset (evaluation criteria). The selection algorithm and evaluation criteria are two important parts of feature selection; an excellent search algorithm can speed up the search of features to find the optimal solution. The normal search algorithms contain global optimization, random search, and heuristic search. Evaluation criteria are defined as to evaluate

the feature subset which is selected with some evaluation criterions. The evaluation decides directly the output of algorithm and performance of a classification model. An excellent evaluation criterion ensures the chosen subset contains a large amount of information and tiny redundancy. Evaluation functions can be divided into filter (evaluation function is independent of classifier), wrapper (the error probability of classifier is used as evaluation function), and embedding (a mixture of the first two). Common feature selection methods include Relief (relevant features), LVW (Las Vegas wrapper), LARS (least angle regression), and attribute reduction of rough set.

Fuzzy rough set is the extension of rough set. Two important applications of rough set are rule induction and attribute reduction (the meaning is the same as feature selection, but in rough set theory it is often called attribute reduction). However, rough set feature selection can only reduce the dataset of discrete data, and to overcome the drawback, people combined fuzzy set to rough set. Fuzzy rough feature selection (FRFS) provides a method which can effectively reduce the datasets of real value or discrete noise data (or both) without users' information. In addition, the technology can be applied to data with a continuous or nominal decision attribute, thus it can be used in the regression and classification dataset, and only requires additional information for each feature of fuzzy partition which can be automatically derived from the data. An important problem of FRFS is the determination of the similarity relations between objects. The existing methods are not flexible nor pertinent. In this paper, a method combines the membership function generation method based on fuzzy c-means clustering and fuzzy equivalence is proposed. The improving FRFS can automatically generate a membership function based on the knowledge of the dataset itself and complete feature selection.

This paper is structured as follows: Section 2 is literature review. The theoretical background is given in Section 3, introducing the rough set attribute reduction, membership function, and fuzzy rough feature selection. In Section 4, the improved fuzzy rough feature selection is presented, and an experiment is provided in Section 5. The paper is concluded in Section 6.

2. Literature Review

The rough set theory is a frame represented by Zdzislaw Pawlak in 1989, which can construct concept approximation with incomplete information. The available information contains a set of examples of concept and the relationship to each other, such as indiscernibility, set approximation, reducts, and dependency [3,4]. Rough set as a method of soft computing receives more and more attention. Nowadays, rough set is still a research hotspot in the field of artificial intelligence. Hu and Yao proposed structured rough set approximation in complete and incomplete information systems to serve as a basis of three-way decisions with rough set [5]. To deal with an incomplete information system, a more generalized approach that considered potential candidates was presented [6].

Rule induction and feature selection are two important applications of rough set. Every component of the model of rule induction is introduced in detail in [7]. In the literature [8,9], rule induction is carried out for the absence of feature values in the information system. In the literature [10,11], the researchers used the result of attribute reduction to classify datasets with neural networks; the testing result indicated that with less study time the misclassification does not increase significantly, and they declared that the attribute reduction of rough set has the possibility of practical application. Because the attribute reduction of rough set is an NP-hard problem, many pieces of research pay attention to the acceleration algorithm [12–14]. Recently, two quick feature selection algorithms based on the neighbor inconsistent pair were presented which can reduce the time consumed in finding a reduct [15].

Fuzzy sets were introduced independently by Lotfi A. Zadeh and Dieter Klaua in 1965 as an extension of the classical notion of set [16]. Because both rough set and fuzzy set are used to deal with uncertain data, so many scholars compared the two methods and make great contributions [17,18]. Dubois and Prade first combined fuzzy set and rough set together [19], hereafter the research centering on fuzzy rough set appear one after another [20–25], and in the meantime, the accelerating algorithm came out, such as feature selection based on ant colony optimization [26] and information

entropy [27]. In recent years, the feature selection algorithm based on a new definition of fuzzy rough set approximations based on the divergence measured of fuzzy sets is proposed and its properties were explored [28]. Another interest is the accelerator of fuzzy rough feature selection, a method based on sample reduction and dimensionality reduction was proposed [29].

3. Theoretical Background

Rough set theory is presented by Zdzislaw Pawlak in 1981 [30,31]. A basic notion of rough set is the concept of lower and upper approximations. A vague concept is an approximation of precise concepts. Objects belonging to the same category have the same attributes that are indistinguishable.

3.1. Rough Set Attribute Reduction

The central notion of RSAR (rough set attribute reduction) is indiscernibility. Assume there is an information system $I = (U, A)$, where U is a non-empty finite subset of objects (the universe), A is a non-empty finite set of attributes such as for each $a \in A, U \rightarrow V_a$, where V_a is the value set for attribute a . $A = \{C \cup D\}$ where C is the set of condition attributes and D is the set of decision attributes.

Definition 1 (Indiscernibility). For any $P \subseteq A$, there is an associated equivalence relation $IND(P) : IND(P) = \{(x, y) \in U^2 | \forall a \in P, a(x) = a(y)\}$, if $(x, y) \in IND(P)$, then x and y are indiscernible from attributes of P . The equivalence classes of the P -indiscernibility relation are denoted as $[x]_P$.

Definition 2 (Lower Approximation). The lower approximation of P is defined as $\underline{P}X = \{x | [x]_P \subseteq X\}, X \subseteq U$.

Definition 3 (Positive Region). P and Q are equivalence relations over U , the positive region is defined as $POS_P(Q) = \bigcup_{X \in U/Q} \underline{P}X$. The positive region contains the objects of U which can be divided into U/Q class, using the knowledge of attribute P .

Definition 4 (Dependency). An important work of data analysis is finding the dependency of attributes. If all values of a set of attributes Q depend uniquely on another set of attributes P , or there is a functional dependency between values of Q and P , then Q depends totally on P , denoted $P \Rightarrow Q$. Dependency can be defined as: for $P, Q \subseteq A$, Q depends on P with degree k ($0 \leq k \leq 1$), defined $P \Rightarrow_k Q, k = \gamma_P(Q) = \frac{|POS_P(Q)|}{|U|}$, if $k = 1$, Q depends totally on P . If $k < 1$, Q depends partially on P with degree k . If $k = 0$, Q does not depend on P .

A basic idea is to calculate the dependencies of all the possible subsets of C , any subset with $\gamma(D) = 1$ is a reduct. The smallest subset is the minimum reduct. However, this idea is impossible for large datasets. Algorithm 1 is called QUICKREDUCT Algorithm [20]; we do not need to calculate all the possible subsets. Start with an empty set and add attributes to the set in proper order to obtain the maximum increase of dependency until we get the maximum possible number (usually equal to 1). We need to pay attention to that the algorithm may not get the minimum set of reduct necessarily every time. In the worst situation, the complexity reaches $n!$ for attribute dimensionality of n .

Algorithm 1: RSAR QUICKREDUCT

Input: $R \leftarrow \{\text{empty set}\}$, $C \leftarrow \{\text{set of all features}\}$,
Output: R
1: **do**
2: $T \leftarrow R$
3: $\forall x \in (C - R)$
4: **if** $\gamma_{R \cup \{x\}}(D) > \gamma_T(D)$
5: $T \leftarrow R \cup \{x\}$
6: $R \leftarrow T$
7: **until** $\gamma_R(D) = \gamma_C(D)$
8: **return** R

3.2. Membership Function

American cybernetician L.A. Zadeh created fuzzy sets theory with a groundbreaking paper in 1965 [16]. Fuzzy set is the extension of classical set, which contains more general and various mathematical concepts, forming a new fuzzy mathematics discipline.

Assuming X is the universe, A is a subset of X and can be represented by characteristic functions such as map $A : X \rightarrow \{0, 1\}$, $\forall x \in X$, $A(x) = \begin{cases} 0, & x \notin A \\ 1, & x \in A \end{cases}$. For fuzzy subset A of X , any $x \in X$, x is not absolutely belonging to A nor absolutely not belonging to A . The degree to which x belongs to A can be represented by the value in $[0, 1]$.

Definition 5 (Membership Function). Assuming X is the universe, the map $A : X \rightarrow [0, 1]$ is called a fuzzy subset of X , F set for short. Map A is called membership function of F set about A , $A(x)$ is called degree of membership of x about A .

The generation of the fuzzy membership function is fundamentally important. It is important to find a proper membership function. A basic method of constructing membership functions is the reference function; there are some commonly used membership functions, choosing proper parameters to get the membership functions which we need, such as triangular membership functions:

$$f(x, a, b, c) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ \frac{c-x}{c-b}, & b < x \leq c \\ 0, & x > c \end{cases} \quad (1)$$

Trapezoidal membership function:

$$f(x, a, b, c, d) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x < b \\ 1, & b \leq x < c \\ \frac{d-x}{d-c}, & c \leq x \leq d \\ 0, & x > d \end{cases} \quad (2)$$

Gaussian membership function:

$$f(x, \sigma, c) = \exp\left(-\frac{(x-c)^2}{2\sigma^2}\right). \quad (3)$$

Membership function can be generated from available data. Many methods can be used, such as the histogram method, the transformation of the probability distribution to the probability distribution,

clustering, and neural network [32–36]. We need an effective membership function generation mechanism to make full use of fuzzy theory and it must have the following advantages [37]:

1. Accuracy. Membership function should reflect the knowledge contained in the data accurately.
2. Flexibility. This method can provide a broad family of membership function.
3. Computability. The method should be feasible in calculation so that it has practical application value. The literature [38] emphasizes the importance of easy optimization and adjustment for membership functions.
4. Ease of use. Once the membership function is determined, for any given x , the corresponding $A(x)$ can be found easily.

In this paper, we use clustering to find membership functions. The fuzzy c-mean clustering (FCM) method is used to generate the fuzzy membership function during clustering. During the clustering process of FCM, the fuzzy membership function is generated.

3.3. Fuzzy Rough Feature Selection

The RSAR we mentioned only applies on the discrete datasets, but in real life, datasets usually contain real values and noises. Using fuzzy set theory, we can deal with this complex situation. Fuzzy mathematics is a mathematical theory in which studies utilize fuzzy phenomena. The core of fuzzy mathematics is the fuzzy set, which is different from the classical set. It has no definite elements and can only be mastered by membership function.

The intersection, union, and complement operations of fuzzy sets are similar to those of classical sets, but in some cases, the general and operator may fail. The selection of the and operator should be analyzed in detail. T norm and s norm can be thought of as generalized operations, but they are not actually. This question remains uncertain.

Definition 6 (t-norm). A triangular norm or shortly *t*-norm better reflects the nature of the logic operator and. *T*-norm is a binary function T on $[0, 1]$, which satisfies the exchange law, associative law, monotonicity, and the boundary condition. That is to say $T : [0, 1] \times [0, 1] \rightarrow [0, 1], \forall x, y, z \in [0, 1]$:

$$T(x, y) = T(y, x)$$

$$T(x, T(y, z)) = T(T(x, y), z)$$

$$y \leq z, T(x, y) \leq T(x, z)$$

$$T(x, 1) = x.$$

There are some frequently used *t*-norms $\otimes$:

The standard min operator $x \otimes y = \min\{x, y\}$

The algebraic product $x \otimes y = xy$

Lukasiewicz *t*-norm $x \otimes y = \max\{x + y - 1, 0\}$.

Definition 7 (s-norm). *S*-norm is also called a triangular conorm and shortly *t*-conorm. *S*-norm is a binary function S on $[0, 1]$, which satisfies the exchange law, associative law, monotonicity, and the boundary condition. That is to say $S : [0, 1] \times [0, 1] \rightarrow [0, 1], \forall x, y, z \in [0, 1]$:

$$S(x, y) = S(y, x)$$

$$S(x, S(y, z)) = S(S(x, y), z)$$

$$y \leq z, S(x, y) \leq S(x, z)$$

$$S(x, 0) = x.$$

Three well known *s*-norms $\oplus$ are:

The standard max operator $x \oplus y = \max\{x, y\}$

The probabilistic sum $x \oplus y = x + y - xy$

The bounded sum $x \oplus y = \min\{x + y, 1\}$.

The fuzzy rule can be expressed as “if A then B ”, shortly $A \rightarrow B$. A and B are fuzzy sets, the true degrees are expressed as $A(x)$, $B(y)$. The true degree of $A \rightarrow B$ is expressed as $A(x) \rightarrow B(y)$; the degree of the proposition depends on the true degree of the former and the latter.

Definition 8 (Fuzzy Implicator). Find a binary function $R : [0, 1] \times [0, 1] \rightarrow [0, 1]$ to represent the fuzzy rules properly. Meet left monotone (or right monotone)

$$R(1, 0) = 0, R(1, 1) = R(0, 1) = R(0, 0) = 1.$$

The frequently used implicators are:

Mamdani implicator: $x \rightarrow y = x \wedge y$

Zadeh implicator: $x \rightarrow y = (1 - x) \vee (x \wedge y)$

Kleene–Dienes implicator: $x \rightarrow y = (1 - x) \vee (x \wedge y)$

Lukasiewicz implicator: $x \rightarrow y = \min\{1, 1 - x + y\}$.

Fuzzy upper approximation and lower approximation are defined with the fuzzy division of input [39] or with the following definition:

$$\mu_{R_P X}(x) = \inf_{y \in U} I(\mu_{R_P}(x, y), \mu_X(y)), \quad (4)$$

where I is the fuzzy implicator and R_P represents the fuzzy similar relation in the subset of feature set P .

$$\mu_{R_P}(x, y) = \cap_{a \in P} \{\mu_{R_a}(x, y)\}, \quad (5)$$

where $\mu_{R_a}(x, y)$ represents the degree of similarity of x and y about feature a .

$$\mu_{R_a}(x, y) = \max\left(\min\left(\frac{a(y) - (a(x) - \sigma_a)}{a(x) - (a(x) - \sigma_a)}, \frac{a(x) + \sigma_a - a(y)}{a(x) + \sigma_a - a(x)}\right), 0\right) \quad (6)$$

The fuzzy positive region and dependency are defined as in rough set [40].

$$\mu_{POS_{R_P}(Q)}(x) = \sup_{x \in U/Q} \mu_{R_P X}(x) \quad (7)$$

$$\gamma'_P(Q) = \frac{\sum_{x \in U} \mu_{POS_{R_P}(Q)}(x)}{|U|} \quad (8)$$

An example with the dataset in Table 1 follows. There are six objects, features (condition attribute) a, b and c , label (decision attribute) q . Assuming $C = \{a, b, c\}$, $D = \{q\}$, $IND(q) = \{\{2, 4, 5\}, \{1, 3, 6\}\}$, using Equation (6), we can obtain the fuzzy similarity matrixes as follows:

$$U_{R_a}(x, y) = \begin{pmatrix} 1.0 & 1.0 & 0.987 & 0.013 & 0.006 & 0.006 \\ 1.0 & 1.0 & 0.987 & 0.013 & 0.006 & 0.006 \\ 0.987 & 0.987 & 1.0 & 0.026 & 0.019 & 0.019 \\ 0.013 & 0.013 & 0.026 & 1.0 & 0.993 & 0.993 \\ 0.006 & 0.006 & 0.019 & 0.993 & 1.0 & 1.0 \\ 0.006 & 0.006 & 0.019 & 0.993 & 1.0 & 1.0 \end{pmatrix}$$

$$R_b(x, y) = \begin{pmatrix} 1.0 & 0.0 & 0.568 & 1.0 & 1.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.137 \\ 0.568 & 0.0 & 1.0 & 0.568 & 0.568 & 0.0 \\ 1.0 & 0.0 & 0.568 & 1.0 & 1.0 & 0.0 \\ 1.0 & 0.0 & 0.568 & 1.0 & 1.0 & 0.0 \\ 0.0 & 0.137 & 0.0 & 0.0 & 0.0 & 1.0 \end{pmatrix}$$

$$R_c(x, y) = \begin{pmatrix} 1.0 & 0.0 & 0.036 & 0.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.036 & 0.518 & 0.518 & 0.518 \\ 0.036 & 0.036 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.518 & 0.0 & 1.0 & 1.0 & 1.0 \\ 0.0 & 0.518 & 0.0 & 1.0 & 1.0 & 1.0 \\ 0.0 & 0.518 & 0.0 & 1.0 & 1.0 & 1.0 \end{pmatrix}$$

$$\mu_{\underline{R}_a\{1,3,6\}}(x) = \inf_{y \in U} I(\mu_{R_a}(x, y), \mu_{\{1,3,6\}}(y))$$

$$\mu_{\underline{R}_a\{1,3,6\}}(3) = \inf_{y \in U} I(\mu_{R_a}(3, y), \mu_{\{1,3,6\}}(y)) = \inf\{I(0.699, 1), I(0.699, 0), I(1, 1), I(0, 0), I(0, 0), I(0, 0)\} = 0.301.$$

Table 1. Dataset with continuous data.

Object	a	b	c	q
1	−0.4	−0.3	−0.5	1
2	−0.4	0.2	−0.1	2
3	−0.3	−0.4	−0.3	1
4	0.3	−0.3	0	2
5	0.2	−0.3	0	2
6	0.2	0	0	1

Others are the same as object 3.

$$\mu_{\underline{R}_a\{1,3,6\}}(1) = 0.0$$

$$\mu_{\underline{R}_a\{1,3,6\}}(2) = 0.0$$

$$\mu_{\underline{R}_a\{1,3,6\}}(4) = 0.0$$

$$\mu_{\underline{R}_a\{1,3,6\}}(5) = 0.0$$

$$\mu_{\underline{R}_a\{1,3,6\}}(6) = 0.0$$

For $\{2, 4, 5\}$

$$\mu_{\underline{R}_a\{2,4,5\}}(1) = 0.0$$

$$\mu_{\underline{R}_a\{2,4,5\}}(2) = 0.0$$

$$\mu_{\underline{R}_a\{2,4,5\}}(3) = 0.0$$

$$\mu_{\underline{R}_a\{2,4,5\}}(4) = 0.301$$

$$\mu_{\underline{R}_a\{2,4,5\}}(5) = 0.0$$

$$\mu_{\underline{R}_a\{2,4,5\}}(6) = 0.0.$$

Thus, the positive regions of every object are:

$$\mu_{POS_{R_a}(Q)}(1) = 0.0$$

$$\begin{aligned}
\mu_{POS_{R_a}(Q)}(2) &= 0.0 \\
\mu_{POS_{R_a}(Q)}(3) &= 0.301 \\
\mu_{POS_{R_a}(Q)}(4) &= 0.301 \\
\mu_{POS_{R_a}(Q)}(5) &= 0.0 \\
\mu_{POS_{R_a}(Q)}(6) &= 0.0.
\end{aligned}$$

The dependency of feature a :

$$\gamma'_{\{a\}}(Q) = \frac{\sum_{x \in U} \mu_{POS_{R_a}(Q)}(x)}{|U|} = \frac{0.602}{6} = 0.1003.$$

Using the same algorithm, we can calculate the $\gamma'_{\{b\}}(Q) = 0.3597$, $\gamma'_{\{c\}}(Q) = 0.4078$, in the first circulation we choose feature c . Then, in the second circulation we obtain $\gamma'_{\{a,c\}}(Q) = 0.5501$, $\gamma'_{\{b,c\}}(Q) = 1.0$, and in the end, we choose the feature subset $\{b, c\}$.

4. New Method

In the existing fuzzy rough feature selection algorithm, there are two methods for choosing the fuzzy set, one is given a fuzzy set while inputting data [39]. The other is definite with fuzzy similarity relations and a fuzzy implicator [40,41]. Both methods have their own drawbacks. The first one complicates the algorithm, so we need to add some knowledge to the feature selection, which is departing from our original intension. The second one has some problems in the definition of fuzzy similarity relations. The common definitions of relation at present are:

$$\mu_{R_a}(x, y) = 1 - \frac{|a(x) - a(y)|}{|a_{max} - a_{min}|}, \quad (9)$$

$$\mu_{R_a}(x, y) = \exp\left(-\frac{(a(x) - a(y))^2}{2\sigma_a^2}\right), \quad (10)$$

$$\mu_{R_a}(x, y) = \max\left(\min\left(\frac{(a(y) - (a(x) - \sigma_a))}{(a(x) - (a(x) - \sigma_a))}, \frac{((a(x) + \sigma_a) - a(y))}{((a(x) + \sigma_a) - a(x))}\right), 0\right), \quad (11)$$

where σ_a^2 is the variance of feature a . As we can see, what is said above defines the fuzzy similarity relations of all the datasets with a single equation but ignores the difference between the datasets. Generating a fuzzy set automatically is extremely urgent. A dataset is the universe of fuzzy sets that contains many fuzzy sets. We can abstract a fuzzy set and fuzzy similarity relations from a dataset and make it different between datasets, so the algorithm model has better generalization ability.

4.1. Reduction

The fuzzy c-means clustering algorithm (FCM) has a wide range application and is more successful in numerous fuzzy clustering algorithms. It obtains the membership degree of every sample point to the class center through the optimization of objective function [42].

Objective function is represented by the Euclidean distance of clustering center and sample point. Solving every clustering center to the minimum of the value function of the non-similarity index. The vague generalization is:

$$J(U, c_1, c_2, \dots, c_c) = \sum_{i=1}^c J_i = \sum_{i=1}^c \sum_j^n u_{ij}^m d_{ij}^2 \quad (12)$$

where u_{ij} is between 0 and 1, c_i is the clustering center of the fuzzy set, $d_{ij} = \|c_i - x_j\|$ is the Euclidean distance of i th clustering center j th sample point. m is the weighted index number. To construct the Lagrangian multiplier of the constraint formula, derivate all the input parameters to make Equation (12) reach the minimum:

$$c_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{j=1}^n u_{ij}^m} \quad (13)$$

$$\mu_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{kj}} \right)^{2/(m-1)}}. \quad (14)$$

The whole processors of the clustering algorithm are as in Algorithm 2. The output of FCM are centers c and membership matrix U . U contains the degree of every object belonging to each center.

Algorithm 2: Fuzzy C-Means (FCM)

Input: Matrix of membership degree U initialized with random numbers between 0 and 1, and satisfied the normalization, accuracy of convergence ϵ , maximum number of iterations T and weighted index m .

Output: U

```

1:  $t = 1$ 
2: do
3:    $c_i, i = 1, 2, \dots, c$  with Equation (13)
4:    $U = [u_{ij}]$  with Equation (14).
5:   Update  $c_i$  and  $U$ 
6:    $t \leftarrow t + 1$ 
7: Until  $t = T$  or  $\|U_t - U_{t-1}\| \leq \epsilon$ 
8: Return  $U$ 

```

The definition of the lower approximation of fuzzy rough set is:

$$\mu_{R_P X}(x) = \inf_{y \in U} I(\mu_{R_P}(x, y), \mu_X(y)) \quad (15)$$

where I is the fuzzy implicator. $\mu_{R_P}(x, y)$ represents the similarity relation between x and y in the whole subsets of feature set P . In order to contain only one similarity relation, we take the intersection of relations in all subsets of P , where the intersection is calculated with the t-norm.

$$\mu_{R_P}(x, y) = \bigcap_{a \in P} \{\mu_{R_a}(x, y)\} \quad (16)$$

where $\mu_{R_a}(x, y)$ represents the similarity degree of x and y about feature a .

Fuzzy clustering on every feature with Algorithm 2 to get the membership degree of every object to the feature:

$$\mu_{R_a}(x) = U_{FCM(a)}[x]. \quad (17)$$

Because equivalence relations are used to model equality, fuzzy equivalence relations are commonly considered to represent approximate equality or similarity [41,43]. We use the fuzzy equivalence relation R in the literature [44]:

$$R(x, y) = \frac{a - a|x - y| + b \min(x, y)}{a - (a - 1)|x - y| + b \min(x, y)}. \quad (18)$$

Figure 1 shows that with different values of a and b , the equation has a different form. The value of a defines the basic shape of the function. Additionally, the value of b defines the size of the opening on one side. As we use the function to describe the similarity of two objects, so we choose b equal to 0 to make sure that the function is in balance of the two objects even if they are great values or

small values. When a goes to 0, the function performs a crisp relation that when two objects are equal the function output is 1, otherwise it is 0. Thus, we chose a small value on a , because after the FCM algorithm, the difference between objects was small.

Figure 1. Fuzzy equivalence relation with different a and b . (a) $a = 0.1$ and $b = 0$; (b) $a = 0.01$ and $b = 0$; (c) $a = 0.1$ and $b = 0.1$; (d) $a = 0.1$ and $b = 0.1$.

Algorithm 3: Fuzzy C-Means Fuzzy Rough Feature Select (C-FRFS)

Input: the set of all conditional attributes C , the set of all decision attributes D and the set of all objects U . $R \leftarrow \{\text{empty set}\}$

Output: R

```

1: For each  $x, y$  in  $U$ :
2:    $\mu_x = FCM(x), \mu_y = FCM(y)$ 
3:   Calculate  $E(x, y)$ 
4:    $\gamma'_{prev} = 0, \gamma'_{best} = 0$ 
5: do
6:    $T \leftarrow R$ 
7:    $\gamma'_{prev} = \gamma'_{best}$ 
8:   if  $\gamma_{R \cup \{a\}}(D) > \gamma_T(D), \forall a \in (C - R)$ ,
9:      $T \leftarrow R \cup \{a\}$ 
10:   $R \leftarrow T$ 
11:   $\gamma'_{best} \leftarrow \gamma_R(D)$ 
12: until  $(\gamma'_{best} - \gamma'_{prev}) \times |U| < 1$ 
13: return  $R$ 

```

According to the clustering membership and Equation (18), we can get the fuzzy similarity relation of two objects.

$$\mu_{R_a}(x, y) = \frac{a - a \left| U_{FCM(a)}[x] - U_{FCM(a)}[y] \right|}{a - (a - 1) \left| U_{FCM(a)}[x] - U_{FCM(a)}[y] \right|} \quad (19)$$

The definitions of positive region and dependency are the same as we mentioned above [40].

$$\mu_{POS_{R_P}(Q)}(x) = \sup_{x \in U/Q} \mu_{R_P X}(x) \quad (20)$$

$$\gamma'_P(Q) = \frac{\sum_{x \in U} \mu_{POS_{R_P}(Q)}(x)}{|U|} \quad (21)$$

The steps of the algorithm are as in Algorithm 3, and we simply called it C-FRFS, which means fuzzy rough feature selection based on clustering. We apply the fuzzy c-means clustering on every object in the universe C . For every two objects x and y , use the fuzzy equivalence relation Equation (19) to describe the fuzzy similarity relation. Then, according to Equations (15), (20), and (21), we can obtain the dependency degree γ of every feature in C . Start with an empty set R , and each time select a feature which has the greatest increase in the dependency degree. The algorithm stops when adding a feature cannot result in classifying at least one object.

4.2. Example

We still use the example in Table 1, and with the FCM we calculate the membership matrix as in Table 2. Assuming $C = \{a, b, c\}$, $D = \{q\}$, $IND(q) = \{\{2, 4, 5\}, \{1, 3, 6\}\}$, with Equation (19), we can get the fuzzy similarity matrix:

$$U_{R_a}(x, y) = \begin{pmatrix} 1 & 0.0707 & 0.5893 & 1 & 0.1196 & 0.5893 \\ 0.0707 & 1 & 0.0744 & 0.0707 & 0.1492 & 0.0744 \\ 0.5893 & 0.0744 & 1 & 0.5893 & 0.1306 & 1 \\ 1 & 0.0707 & 0.5893 & 1 & 0.1196 & 0.5893 \\ 0.1196 & 0.1492 & 0.1306 & 0.1196 & 1 & 0.1306 \\ 0.5893 & 0.0744 & 1 & 0.5893 & 0.1306 & 1 \end{pmatrix}$$

$$U_{R_b}(x, y) = \begin{pmatrix} 1 & 0.0490 & 0.0094 & 0.0343 & 1 & 1 \\ 0.0490 & 1 & 0.0120 & 0.1066 & 0.0490 & 0.0490 \\ 0.0094 & 0.0120 & 1 & 0.0135 & 0.0094 & 0.0094 \\ 0.0343 & 0.1066 & 0.0135 & 1 & 0.0343 & 0.0343 \\ 1 & 0.0490 & 0.0094 & 0.0343 & 1 & 1 \\ 1 & 0.0490 & 0.0094 & 0.0343 & 1 & 1 \end{pmatrix}$$

$$U_{R_c}(x, y) = \begin{pmatrix} 1 & 0.0085 & 0.0306 & 0.0433 & 0.0306 & 0.0306 \\ 0.0085 & 1 & 0.0064 & 0.0110 & 0.0064 & 0.0064 \\ 0.0306 & 0.0064 & 1 & 0.0178 & 1 & 1 \\ 0.0433 & 0.0110 & 0.0178 & 1 & 0.0178 & 0.0178 \\ 0.0306 & 0.0064 & 1 & 0.0178 & 1 & 1 \\ 0.0306 & 0.0064 & 1 & 0.0178 & 1 & 1 \end{pmatrix}$$

Table 2. Membership matrix of dataset in Table 1.

Object	a	b	c
1	0.9973	0.9967	0.9657
2	0.9843	0.9777	0.8617
3	0.9966	0.9014	0.9965
4	0.9973	0.9693	0.9441
5	0.9900	0.9967	0.9965
6	0.9966	0.9967	0.9965

For every element in $IND(q)$, calculate the lower approximation of every object. We give an example with $\{1, 3, 6\}$ in $IND(q)$ and object 2:

$$\mu_{\underline{R}_a\{1,3,6\}}(x) = \inf_{y \in U} I(\mu_{R_a}(x, y), \mu_{\{1,3,6\}}(y)).$$

$$\begin{aligned} \mu_{\underline{R}_a\{1,3,6\}}(2) &= \inf_{y \in U} I(\mu_{R_a}(2, y), \mu_{\{1,3,6\}}(y)) \\ &= \inf\{I(0.0707, 1), I(1, 0), I(0.0744, 1), I(0.0707, 0), I(0.1492, 0), I(0.0744, 1)\} \\ &= 0.8508 \end{aligned}$$

Others in the same way:

$$\mu_{\underline{R}_a\{1,3,6\}}(1) = 0.0$$

$$\mu_{\underline{R}_a\{1,3,6\}}(3) = 0.0$$

$$\mu_{\underline{R}_a\{1,3,6\}}(4) = 0.0$$

$$\mu_{\underline{R}_a\{1,3,6\}}(5) = 0.0$$

$$\mu_{\underline{R}_a\{1,3,6\}}(6) = 0.0.$$

For $\{2, 4, 5\}$

$$\mu_{\underline{R}_a\{2,4,5\}}(1) = 0.0$$

$$\mu_{\underline{R}_a\{2,4,5\}}(2) = 0.0$$

$$\mu_{\underline{R}_a\{2,4,5\}}(3) = 0.0$$

$$\mu_{\underline{R}_a\{2,4,5\}}(4) = 0.0$$

$$\mu_{\underline{R}_a\{2,4,5\}}(5) = 0.8508$$

$$\mu_{\underline{R}_a\{2,4,5\}}(6) = 0.0.$$

Thus, the positive region of every object:

$$\mu_{POS_{R_a}(Q)}(1) = 0.0$$

$$\mu_{POS_{R_a}(Q)}(2) = 0.8508$$

$$\mu_{POS_{R_a}(Q)}(3) = 0.0$$

$$\mu_{POS_{R_a}(Q)}(4) = 0.0$$

$$\mu_{POS_{R_a}(Q)}(5) = 0.8508$$

$$\mu_{POS_{R_a}(Q)}(6) = 0.0.$$

The dependency of feature a :

$$\gamma'_{\{a\}}(Q) = \frac{\sum_{x \in U} \mu_{POS_{R_P}(Q)}(x)}{|\mathbb{U}|} = \frac{1.7016}{6} = 0.2836.$$

In the same way, we can obtain $\gamma'_{\{b\}}(Q) = 0.4624$, $\gamma'_{\{c\}}(Q) = 0.4837$.

Thus, in the first circulation, we choose feature c . Then, in the second circulation, we obtain $\gamma'_{\{a,c\}}(Q) = 0.6305$, $\gamma'_{\{b,c\}}(Q) = 0.9831$, and in the end we choose feature b and c as our result of feature selection.

5. Experiments

In this part, we used nine classifiers to classify nine datasets from UCI machine learning repository [45], to verify and evaluate our feature selection algorithm. The algorithm uses fuzzy clustering to generate fuzzy sets, which are defined by Equation (17). After feature selection, the datasets are reduced. These reduced datasets are classified using related classifiers (unreduced datasets do not use feature selection steps).

The T-FRFS is the short of threshold fuzzy rough feature select which was represented in [46] and FRFS is mentioned in [40]. C-FRFS is our proposed method mentioned in this paper. All the three methods were written in MATLAB 2017b which runs on a computer with following characteristics:

OS: Microsoft Windows 10

CPU: Intel Core i5-8400 CPU 2.80GHz

RAM: 8GB.

We use these two proposed successors to compare with our new method and this indicated that we have made some progress on fuzzy rough feature selection. Nine datasets as depicted in Table 3, were employed to evaluate each methods' performance. The results of all methods, as well as unreduced datasets in terms of the number of selected features, are also demonstrated in Table 3. As shown, our new method C-FRFS always gets the smallest reduced subset and then the T-FRFS. The FRFS does not perform well and sometimes cannot select features.

Table 3. Information of nine datasets and reduct size of three feature selection methods. C-FRFS: c-means fuzzy rough feature select. T-FRFS: threshold fuzzy rough feature select.

No.	Dataset	Objects	Features	Classes	Reduct Size		
					C-FRFS	T-FRFS	FRFS
1	Glass	214	9	7	4	8	9
2	Iris	150	4	3	3	3	4
3	Cleveland	303	13	5	4	6	8
4	Wine	178	13	3	4	4	5
5	User	431	5	4	3	4	5
6	Vertebral	310	6	4	4	5	6
7	Sonar	208	60	2	2	4	5
8	Yeast	1484	8	10	6	7	8
9	Hillvalley	606	100	2	24	96	100

Figure 2 shows the running time of three methods. In Figure 2a, our new method is always slower than the other two methods. However, in the Hillvalley dataset, our new method performs better, which means it is more suitable for high-dimensional data than the other two methods. The reason is that our method obtains a smaller reduct, so the program ends earlier and running time is less.

To test the performance of the new method, we applied three methods on different sizes of datasets. With the number of objects from 200 to 5000, the time consumption increased exponentially. However, as shown in Figure 3, the three methods have the same tendency and are almost coincident,

which demonstrates that clustering nearly not influences the performance of our new method. In the same case, our new method can get the best consequence but has no impact on efficiency.

(a)

(b)

Figure 2. Running time of the three methods for dealing with different datasets in Table 3.

Figure 3. Running time of three methods for dealing with different sizes of datasets.

Nine classifiers of different categories such as Bayes Net, Naïve Bayes, RBFNetwork, Jrip, PART, BFTree, FT, J48, and NBTree in Weka were selected to classify resulting subsets of features by each FS method [47]. To obtain more accurate results, we used 10-fold cross validation in the modeling of each classifier method. The results are presented in Table 4 where the average classification accuracies of nine classifiers for each dataset worked with three selection methods along with unreduced datasets shown in each column, and the last column indicated the mean of nine classification accuracies. The highest mean of classification accuracies was gained in expense of employing all features, which means that feature selection methods are not always successful in increasing classification accuracy, but in decreasing model complexity by sacrificing inconsiderable accuracy. As Table 4 shows, for four datasets, the original model gained the highest classification accuracy, three for both FRFS and our new method.

Table 4. Resulting classification accuracies. Unred.: unreduced.

Datasets	Glass	Iris	Cleveland	Wine	User	Vertebral	Sonar	Yeast	Hillvalley	Mean
C-FRFS	63.66	94.52	55.01	95.51	91.01	80.90	72.44	56.36	50.59	73.33
T-FRFS	64.85	94.52	54.46	94.82	90.99	71.11	72.49	56.55	50.84	72.29
FRFS	66.46	94.82	54.61	95.38	90.65	81.11	70.04	56.99	51.03	73.45
Unred.	66.46	94.82	54.42	95.13	90.65	81.11	76.07	56.99	51.03	74.08

Since both the classification accuracies and the number of selected features are important, divisions of classification accuracies, as shown in Table 4, by the number of selected features, as shown in Table 3, were considered as a measure to compare the results [48]. Therefore, a method which has a higher classification accuracy and smaller set of features is regarded as a better method. The results are shown in Table 5. The lowest value was from the original datasets, and our new method reaches the highest value in each dataset. Sometimes, the T-FRFS also reaches the highest value.

Table 5. Division of accuracies and number of selected features.

Datasets	Glass	Iris	Cleveland	Wine	User	Vertebral	Sonar	Yeast	Hillvalley	Mean
C-FRFS	15.92	31.51	13.75	23.88	30.34	20.23	36.22	9.39	2.11	24.55
T-FRFS	8.11	31.51	9.08	23.71	22.75	14.22	18.12	8.08	0.53	18.21
FRFS	7.38	23.71	6.83	19.08	18.13	13.52	14.01	7.12	0.51	14.66
Unred.	7.38	23.71	4.19	7.32	18.13	13.52	1.27	7.12	0.51	10.79

A statistical test with the Friedman test was used to compare the methods. Table 6 shows the average ranks of methods which correspond to Table 5. The chi-square with 3 freedom was 26.1 and P-value was 0, which means there are statistical difference between methods.

Table 6. Average ranks of methods.

Method	Rank
C-FRFS	1.0556
T-FRFS	1.9444
FRFS	3.3333
Unred.	3.6667

6. Discussion and Conclusions

A new method named C-FRFS based on fuzzy rough feature selection has been presented in this paper. The first development based on the fuzzy c-means clustering uses a new kind of membership function which automatically forms the membership matrix only with the knowledge contained in the dataset itself. The second development employs the setting up of the fuzzy relation—different to methods used in many articles which use a single equation—we have combined fuzzy equivalence mentioned in other research and the membership function generated through the above way to set

up the fuzzy relation. An example is given to illustrate how reduction may be achieved. Note that no user-defined thresholds are required for the method, although a choice must be made regarding fuzzy connections and the priori coefficient of fuzzy equivalence. Experimental results over nine datasets taken from UCI showed the applicability and effectiveness of our proposed new method. In order to measure the performance of every method, we considered the number of selected features, classification accuracies, and the division of classification accuracies by the number of selected features. We compared the results operated on different methods, such as T-FRFS, FRFS, unreduced, and our new method C-FRFS. The comparisons indicated that our method was outperformed. We also compared the time taking of three methods on different sizes of datasets. The results demonstrated that our new method did not influence the efficiency remarkably. That was to say, we improved the accuracy and validity of feature selection with nearly no time loss.

Further research in this area will include a more in-depth experimental investigation of the prior coefficient with its evaluation and a more efficient fuzzy clustering method to raise up the accuracies of our feature selection method.

Author Contributions: Conceptualization, R.Z. and L.G.; Data curation, X.Z.; Funding acquisition, L.G.; Investigation, R.Z.; Methodology, R.Z. and X.Z.; Project administration, L.G. and X.Z.; Resources, L.G.; Software, R.Z.; Supervision, L.G.; Validation, R.Z., L.G., and X.Z.; Visualization, R.Z.; Writing—original draft, R.Z.

Funding: This research was funded by the National Key R&D Program of China (Grant No. 2017YFB080301). The APC was funded by L.G.

Acknowledgments: The authors would like to thank the reviewers for their helpful advice.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kira, K.; Rendell, L.A. The feature selection problem: Traditional methods and a new algorithm. In Proceedings of the Tenth National Conference on Artificial Intelligence, San Jose, CA, USA, 12–16 July 1992; pp. 129–134.
2. John, G.H.; Kohavi, R.; Pfleger, K. Irrelevant features and the subset selection problem. In Proceedings of the Eleventh International Conference, New Brunswick, NJ, USA, 10–13 July 1994; pp. 121–129.
3. Komorowski, J.; Pawlak, Z.; Polkowski, L.; Skowron, A. Rough sets: A tutorial. In *Rough Fuzzy Hybridization: A New Trend in Decision-Making*; Springer: New York, NY, USA, 1999; pp. 3–98.
4. Yao, Y. Probabilistic approaches to rough sets. *Expert Syst.* **2003**, *20*, 287–297. [[CrossRef](#)]
5. Hu, M.; Yao, Y. Structured approximations as a basis for three-way decisions in rough set theory. *Knowl.-Based Syst.* **2019**, *165*, 92–109. [[CrossRef](#)]
6. Cao, T.; Yamada, K.; Unehara, M.; Suzuki, I.; Nguyen, D.V. Rough Set Model in Incomplete Decision Systems. *J. Adv. Comput. Intell. Intell. Inform.* **2017**, *21*, 1221–1231. [[CrossRef](#)]
7. Grzymala-Busse, J.W. LERS—A system for learning from examples based on rough sets. In *Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory*; Słowiński, R., Ed.; Springer: Dordrecht, The Netherlands, 1992; pp. 3–18.
8. Grzymala-Busse, J.W. Data with missing attribute values: Generalization of indiscernibility relation and rule induction. In *Transactions on Rough Sets I*; Springer: Berlin/Heidelberg, Germany, 2004; pp. 78–95.
9. Kryszkiewicz, M. Rough set approach to incomplete information systems. *Inf. Sci.* **1998**, *112*, 39–49. [[CrossRef](#)]
10. Jelonek, J.; Krawiec, K.; Słowiński, R. Rough set reduction of attributes and their domains for neural networks. *Comput. Intell.* **1995**, *11*, 339–347. [[CrossRef](#)]
11. Swiniarski, R.W.; Skowron, A. Rough set methods in feature selection and recognition. *Pattern Recognit. Lett.* **2003**, *24*, 833–849. [[CrossRef](#)]
12. Ke, L.; Feng, Z.; Ren, Z. An efficient ant colony optimization approach to attribute reduction in rough set theory. *Pattern Recognit. Lett.* **2008**, *29*, 1351–1357. [[CrossRef](#)]
13. Qian, Y.; Liang, J.; Pedrycz, W.; Dang, C. Positive approximation: An accelerator for attribute reduction in rough set theory. *Artif. Intell.* **2010**, *174*, 597–618. [[CrossRef](#)]

14. Wang, X.; Yang, J.; Teng, X.; Xia, W.; Jensen, R. Feature selection based on rough sets and particle swarm optimization. *Pattern Recognit. Lett.* **2007**, *28*, 459–471. [\[CrossRef\]](#)
15. Dai, J.; Hu, Q.; Hu, H.; Huang, D. Neighbor Inconsistent Pair Selection for Attribute Reduction by Rough Set Approach. *IEEE Trans. Fuzzy Syst.* **2018**, *26*, 937–950. [\[CrossRef\]](#)
16. Zadeh, L.A. Fuzzy sets. *Inf. Control* **1965**, *8*, 338–353. [\[CrossRef\]](#)
17. Pawlak, Z. Rough sets and fuzzy sets. *Fuzzy Sets Syst.* **1985**, *17*, 99–102. [\[CrossRef\]](#)
18. Yao, Y. A comparative study of fuzzy sets and rough sets. *Inf. Sci.* **1998**, *109*, 227–242. [\[CrossRef\]](#)
19. Dubois, D.; Prade, H. Rough fuzzy sets and fuzzy rough sets. *Int. J. Gen. Syst.* **1990**, *17*, 191–209. [\[CrossRef\]](#)
20. Shen, Q.; Chouchoulas, A. A modular approach to generating fuzzy rules with reduced attributes for the monitoring of complex systems. *Eng. Appl. Artif. Intell.* **2000**, *13*, 263–278. [\[CrossRef\]](#)
21. Jensen, R.; Shen, Q. Semantics-preserving dimensionality reduction: Rough and fuzzy-rough-based approaches. *IEEE Trans. Knowl. Data Eng.* **2004**, *16*, 1457–1471. [\[CrossRef\]](#)
22. Wu, W.-Z.; Mi, J.-S.; Zhang, W.-X. Generalized fuzzy rough sets. *Inf. Sci.* **2003**, *151*, 263–282. [\[CrossRef\]](#)
23. Bhatt, R.B.; Gopal, M. On fuzzy-rough sets approach to feature selection. *Pattern Recognit. Lett.* **2005**, *26*, 965–975. [\[CrossRef\]](#)
24. Tsang, E.C.; Chen, D.; Yeung, D.S.; Wang, X.-Z.; Lee, J.W. Attributes reduction using fuzzy rough sets. *IEEE Trans. Fuzzy Syst.* **2008**, *16*, 1130–1141. [\[CrossRef\]](#)
25. Jensen, R.; Shen, Q. Fuzzy-rough sets assisted attribute selection. *IEEE Trans. Fuzzy Syst.* **2007**, *15*, 73–89. [\[CrossRef\]](#)
26. Jensen, R.; Shen, Q. Fuzzy-rough data reduction with ant colony optimization. *Fuzzy Sets Syst.* **2005**, *149*, 5–20. [\[CrossRef\]](#)
27. Mac Parthaláin, N.; Jensen, R.; Shen, Q. Fuzzy entropy-assisted fuzzy-rough feature selection. In Proceedings of the 2006 IEEE International Conference on Fuzzy Systems, Vancouver, BC, Canada, 16–21 July 2006; pp. 423–430.
28. Sheeja, T.K.; Kuriakose, A.S. A novel feature selection method using fuzzy rough sets. *Comput. Ind.* **2018**, *97*, 111–116. [\[CrossRef\]](#)
29. Qian, Y.; Wang, Q.; Cheng, H.; Liang, J.; Dang, C. Fuzzy-rough feature selection accelerator. *Fuzzy Sets Syst.* **2015**, *258*, 61–78. [\[CrossRef\]](#)
30. Pawlak, Z. *Rough Sets: Theoretical Aspects of Reasoning about Data*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1992; p. 248.
31. Pawlak, Z. Rough sets. *Int. J. Comput. Inf. Sci.* **1982**, *11*, 341–356. [\[CrossRef\]](#)
32. Yang, C.-C.; Bose, N. Generating fuzzy membership function with self-organizing feature map. *Pattern Recognit. Lett.* **2006**, *27*, 356–365. [\[CrossRef\]](#)
33. Nieradka, G.; Butkiewicz, B. A method for automatic membership function estimation based on fuzzy measures. In *International Fuzzy Systems Association World Congress*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 451–460.
34. De Oliveira, J.V. Semantic constraints for membership function optimization. *IEEE Trans. Syst. Man Cybern. Part A Syst. Hum.* **1999**, *29*, 128–138. [\[CrossRef\]](#)
35. Jiang, X.; Yi, Z.; Lv, J.C. Fuzzy SVM with a new fuzzy membership function. *Neural Comput. Appl.* **2006**, *15*, 268–276. [\[CrossRef\]](#)
36. Cheng, H.-D.; Chen, J.-R. Automatically determine the membership function based on the maximum entropy principle. *Inf. Sci.* **1997**, *96*, 163–182. [\[CrossRef\]](#)
37. Medaglia, A.L.; Fang, S.-C.; Nuttle, H.L.; Wilson, J.R. An efficient and flexible mechanism for constructing membership functions. *Eur. J. Oper. Res.* **2002**, *139*, 84–95. [\[CrossRef\]](#)
38. Medasani, S.; Kim, J.; Krishnapuram, R. An overview of membership function generation techniques for pattern recognition. *Int. J. Approx. Reason.* **1998**, *19*, 391–417. [\[CrossRef\]](#)
39. Jensen, R.; Shen, Q. Fuzzy-rough attribute reduction with application to web categorization. *Fuzzy Sets Syst.* **2004**, *141*, 469–485. [\[CrossRef\]](#)
40. Jensen, R.; Shen, Q. New approaches to fuzzy-rough feature selection. *IEEE Trans. Fuzzy Syst.* **2009**, *17*, 824. [\[CrossRef\]](#)
41. Radzikowska, A.M.; Kerre, E.E. A comparative study of fuzzy rough sets. *Fuzzy Sets Syst.* **2002**, *126*, 137–155. [\[CrossRef\]](#)

42. Bezdek, J.C.; Ehrlich, R.; Full, W. FCM: The fuzzy c-means clustering algorithm. *Comput. Geosci.* **1984**, *10*, 191–203. [[CrossRef](#)]
43. De Cock, M.; Cornelis, C.; Kerre, E.E. Fuzzy rough sets: The forgotten step. *IEEE Trans. Fuzzy Syst.* **2007**, *15*, 121–130. [[CrossRef](#)]
44. Li, Y.; Qin, K.; He, X. Some new approaches to constructing similarity measures. *Fuzzy Sets Syst.* **2014**, *234*, 46–60. [[CrossRef](#)]
45. Dua, D.; Karra Taniskidou, E. *UCI Machine Learning Repository*; University of California, Irvine, School of Information and Computer Sciences: Irvine, CA, USA, 2017.
46. Anaraki, J.R.; Eftekhari, M. Improving fuzzy-rough quick reduct for feature selection. In Proceedings of the 2011 19th Iranian Conference on Electrical Engineering (ICEE), Tehran, Iran, 17–19 May 2011; pp. 1–6.
47. Parsania, V.; Bhalodiya, N.; Jani, N. Applying Naïve bayes, BayesNet, PART, JRip and OneR Algorithms on Hypothyroid Database for Comparative Analysis. *Int. J. Darshan Inst. Eng. Res. Emerg. Technol.* **2013**, *3*, 60–64.
48. Anaraki, J.R.; Eftekhari, M.; Ahn, C.W. Novel improvements on the fuzzy-rough quickreduct algorithm. *IEICE Trans. Inf. Syst.* **2015**, *98*, 453–456. [[CrossRef](#)]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).