International Journal of Education, Modern Management, Applied Science & Social Science (IJEMMASSS)
ISSN : 2581-9925, Impact Factor: 6.340, Volume 03, No. 03(III), July - September, 2021, pp.98-110

ANALYZING AND MEASURING PERFORMANCE FOR FUZZY-ROUGH SET FEATURES SELECTION

Rahul Kumar*
Dr. Suyakant Pathak**

ABSTRACT

Our Environment is like 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. The huge measure of research currently being carried out in fuzzy and unpleasant sets is illustrative of this. Numerous profound connections have been set up, and recent investigations have concluded regarding the complementary idea of the two techniques. Hence, it is attractive to expand and hybridize the fundamental concepts to manage extra aspects of information imperfection. Such improvements offer a serious level of adaptability and give powerful arrangements and advanced instruments for information investigation. Fuzzy-unpleasant set-based feature (FS) selection has been demonstrated to be exceptionally valuable at reducing information dimensionality however has a few issues that render it ineffective for enormous datasets. In this paper the creator examines estimating the different performance of wide scope of fuzzy-unpleasant based feature selection. Creator's additionally compares the consequence of this fuzzy-harsh based feature selection. With the quick improvement of the organization, information combination becomes a significant research area of interest. A lot of information should be preprocessed in information combination; in practice, the features of datasets can be separated to reduce the measure of information. The feature selection dependent on fuzzy harsh sets can process an enormous number of continuous and discrete information to reduce the information measurement, making the selected feature subset profoundly correlated with the classification however less reliant upon different features. We compare strategy for fuzzy unpleasant feature selection is proposed which combines the participation function assurance technique for fuzzy c-means clustering and fuzzy equivalence to the first selection. Clarified strategy exploits information about the dataset itself and the differences between datasets, which makes the features selected have a higher correlation with the classification, further develops the classification accuracy, and reduces the information measurement.

Keywords: *Data Explosion, Growth of Datasets, Information Investigation, Classification Accuracy.*

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 [1].

6 Research Scholar, Department of CSE, Dr. K.N. Modi University, Newai, Rajasthan, India.
** Associate Professor, Department of CSE, Dr .K.N. Modi University, Newai, Rajasthan, India.

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 [3].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 [4]. 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 feature selection (FRFS) provides a means by which discrete or real-valued noisy data (or a mixture of both)can be effectively reduced without the need for user-supplied information. Additionally, this technique can be applied to data with continuous or nominal decision attributes, and as such can be applied to regression as well as classification datasets. The only additional information required is in the form of fuzzy partitions for each feature that can be automatically derived from the data. However, there are several problems with the approach from theoretical and practical viewpoints that motivate further developments in this area. This paper proposes three new methods for FRFS that address these problems and provide robust strategies for dimensionality reduction. In particular, the notion of the fuzzy discernibility matrix is proposed to compute reductions [2].

Literature Review

The rough set theory is a casing addressed by Zdzislaw Pawlak in 1989, which can construct concept guess with incomplete data. The accessible data contains a set of instances of concept and the relationship to each other, such as indiscernibility, set guess, redact, and dependency [5, 6]. Rough set as a strategy for delicate computing receives increasingly more consideration. These days, rough set is as yet a research hotpot in the field of artificial intelligence. Hu and Yao proposed structured rough set guess in complete and incomplete data frameworks to fill in as a premise of three-way decisions with rough set [7]. To manage an incomplete data framework, a more summed up approach that considered potential candidates was introduced [8]. Rule induction and feature selection are two significant applications of rough set.

Each component of the model of rule induction is introduced exhaustively in [9]. In the writing [10, 11], rule induction is carried out for the absence of feature esteems in the data framework. In the writing [12,13], the researchers utilized the consequence of property reduction to classify datasets with neural organizations; the testing result indicated that with less review time the misclassification doesn't increase significantly, and they declared that the quality reduction of rough set has the chance of practical application. Because the property reduction of rough set is a NP-difficult issue, many pieces of research focus on the acceleration algorithm [14–16]. Recently, two quick feature selection algorithms dependent on the neighbor inconsistent pair were introduced which can reduce the time consumed in discovering a reduct [17].

Fuzzy sets were introduced freely by Lotfi A. Zadeh and Dieter Klaua in 1965 as an augmentation of the classical thought of set [18]. Because both rough set and fuzzy set are utilized to manage uncertain information, such countless scholars compared the two techniques and make extraordinary contributions [19, 20]. Dubois and Prade previously combined fuzzy set and rough set together [21], in the future the research centering on fuzzy rough set seem in a steady progression [22–27], and meanwhile, the accelerating algorithm came out, such as feature selection dependent on subterranean insect colony improvement [28] and data entropy [29]. In recent years, the feature selection algorithm dependent on another meaning of fuzzy rough set approximations dependent on the divergence estimated of fuzzy sets is proposed and its properties were investigated [30]. Another interest is the accelerator of fuzzy rough feature selection, a strategy dependent on example reduction and dimensionality reduction was proposed [31].

Theoretical Background

Rough set attribute reduction (RSAR) [32] gives a channel based device by which information might be extracted from a space briefly: holding the data content while reducing the measure of information included. The primary benefit that rough set examination has is that it requires no extra boundaries to work other than the provided information [33]. It works by utilizing the granularity structure of the information as it were. This is a significant difference when compared with Dempster–Shafer theory [34] and fuzzy set theory that require likelihood tasks and enrollment esteems, respectively. Notwithstanding, this doesn't imply that no model suppositions are made. In fact, by utilizing just the given data, the theory expects that the information is a valid and accurate reflection of this present reality (which may not be the situation). The numerical and other contextual aspects of the information are overlooked which might appear to be a significant exclusion however downplays model suspicions.

- **Fuzzy-Rough Feature Selection**

The RSAR process described beforehand can just work effectively with datasets containing discrete qualities. Moreover, there is no chance of dealing with loud information. As most datasets contain genuine esteemed attributes, it is necessary to play out a discretization step in advance.

- **Fuzzy Equivalence Classes**: In the same way that crisp equivalence classes are central to rough sets, fuzzy equivalence classes are central to the fuzzy-rough set approach [36], [39], [40]. For typical applications, this means that the decision esteems and the conditional qualities may all be fuzzy. The concept of crisp equivalence classes can be reached out by the inclusion of a fuzzy closeness connection S on the universe, which decides the degree to which two components are comparable in S. The standard properties of reflexivity($\mu_S(x, x) = 1$), symmetry ($\mu_S(x, y) = \mu_S(y, x)$), and T-transitivity ($\mu_S(x, z) \geq \mu_S(x, y) \quad _T\mu_S(y, z)$) hold.

The group of ordinary fuzzy sets produced by a fuzzy apportioning f the universe of discourse can assume the part of fuzzy equivalence classes [36]. Consider the crisp parceling of a vast expanse of discourse U by the attributes in Q: U/Q = {{1, 3, 6}, {2, 4, 5}}. This contains two equivalence classes ({1, 3, 6} and {2, 4, 5}) that can be considered as declined fuzzy sets, with those components having a place with the class having a participation of one, zero in any case. For the five star, for instance, the objects 2, 4, and 5 have an enrollment of nothing. Stretching out this to the case of fuzzy equivalence classes is clear.

- **Fuzzy-Rough Sets:** There have been two primary lines of thought in the hybridization of fuzzy and rough sets, the constructive approach and the axiomatic approach. An overall structure for the investigation of fuzzy-rough sets from both of these perspectives is introduced in [41]. For the constructive approach, summed up lower and upper approximations are characterized dependent on fuzzy relations. At first, these were fuzzy comparability/equivalence relations [36] yet have since been reached out to discretionary fuzzy relations [41]. The axiomatic approach is fundamentally for the investigation of the mathematical properties of fuzzy-rough sets [42]. Here, different classes of fuzzy-rough approximation operators are characterized by various sets of sayings that ensure the existence of kinds of fuzzy relations producing similar operators.

A unique definition for fuzzy P-lower and P-upper approximations was given as follows [36]: A unique definition for fuzzy P-lower and P-upper approximations was given as follows [36]:

$$\mu_{\underline{P}X}(F_i) = \inf_x \max\{1 - \mu_{F_i}(x), \mu_X(x)\} \quad \forall i$$

$$\mu_{\overline{P}X}(F_i) = \sup_x \min\{\mu_{F_i}(x), \mu_X(x)\} \quad \forall i$$

Where Fi is a fuzzy equivalence class, and X is the (fuzzy) concept to be approximated. The tuple<¯P X,_P X >is called a fuzzy-rough set. These definitions wander a little from the crisp upper and lower approximations, as the enrollments of individual objects to the approximations are not explicitly accessible. Subsequently, the fuzzy lower and upper approximations are reclassified as [43]

$$\mu_{\underline{P}X}(x) = \sup_{F \subset \mathbb{U}/P} \min(\mu_F(x), \inf_{y \in \mathbb{U}} \max\{1 - \mu_F(y), \mu_X(y)\})$$

$$\mu_{\overline{P}X}(x) = \sup_{F \in \mathbb{U}/P} \min(\mu_F(x), \sup_{y \in \mathbb{U}} \min\{\mu_F(y), \mu_X(y)\}).$$

- The specific utilization of min and max operators in the aforementioned definitions is extended in [46], where a wide group of fuzzy-rough sets is constructed, where each part is addressed by a particular implicator and t-standard. The properties of three notable implicators (S-, R-and QL-implicators) are researched. Further examinations in this space can be found in [47], [39], [48], and [41]

- **Fuzzy-Rough Reduction Process:** Fuzzy-rough set-put together FS works with respect to the thought of the fuzzy lower approximation to empower reduction of datasets containing genuine esteemed attributes. As will be shown, the process becomes identical to the crisp approach when managing ostensible obvious attributes. The crisp positive area in conventional rough set theory is characterized as the association of the lower approximations. By the expansion principle [49], the enrollment of an object x ∈ U having a place with the fuzzy positive district can be characterized by

$$\mu_{\mathrm{POS}_P(Q)}(x) = \sup_{X \in \mathbb{U}/Q} \mu_{\underline{P}X}(x).$$

Object x won't have a place with the positive locale provided that the equivalence class it has a place with is anything but a constituent of the positive district. This is comparable to the crisp form where objects have a place with the positive locale provided that their hidden equivalence class does as such. Utilizing the meaning of the fuzzy positive district, the fuzzy rough dependency function can be characterized as follows[2]:

$$\gamma'_P(Q) = \frac{|\mu_{\mathrm{POS}_P(Q)}(x)|}{|\mathbb{U}|} = \frac{\sum_{x \in \mathbb{U}} \mu_{\mathrm{POS}_P(Q)}(x)}{|\mathbb{U}|}$$

On the off chance that the fuzzy-rough reduction process is to be valuable, it should have the option to manage different attributes, discovering the dependency between different subsets of the first attribute set. For instance, it very well might be necessary to have the option to decide the level of dependency of the decision attribute(s) with respect to P = {a, b}. In the crisp case, U/P contains sets of objects gathered that are indiscernible according to the two attributes an and b. In the fuzzy case, objects might have a place with numerous equivalence classes, and consequently, the Cartesian product of U/IND({a}) and U/IND({b}) should be considered in deciding U/P[2]. In general[2]

$$\mathbb{U}/P = \otimes\{\mathbb{U}/\mathrm{IND}(\{a\})\,|a \in P\}$$

Where,

$$A \otimes B = \{X \cap Y |\, X \in A, Y \in B, X \cap Y \neq \emptyset\}$$

if P = {a, b}, U/IND({a}) = {N_a, Z_a} and U/IND({b}) = {N_b, Z_b}, then

U/P = {$N_a \cap N_b$, $N_a \cap Z_b$, $Z_a \cap N_b$, $Z_a \cap Z_b$}.

The extent to which an object belongs to such an equivalence class is therefore calculated by using the conjunction of

FRQuickReduct($\mathbb{C}$,$\mathbb{D}$).
$\mathbb{C}$, the set of all conditional attributes;
$\mathbb{D}$, the set of decision attributes.

$R \leftarrow \{\}$; $\gamma'_{best} = 0$; $\gamma'_{prev} = 0$
do
  $T \leftarrow R$
  $\gamma'_{prev} = \gamma'_{best}$
  foreach $x \in (\mathbb{C} - R)$
    if $\gamma'_{R\cup\{x\}}(\mathbb{D}) > \gamma'_T(\mathbb{D})$
      $T \leftarrow R \cup \{x\}$
      $\gamma'_{best} = \gamma'_T(\mathbb{D})$
  $R \leftarrow T$
until $\gamma'_{best} == \gamma'_{prev}$
return R

Fig. 1. Fuzzy-rough QUICKREDUCT algorithm.[2]

constituent fuzzy equivalence classes, say F_i ,i = 1, 2, . . . , n

$$\mu_{F_1 \cap \ldots \cap F_n}(x) = \min(\mu_{F_1}(x), \mu_{F_2}(x), \ldots, \mu_{F_n}(x))$$

- **Fuzzy-Rough QUICKREDUCT:** A problem might emerge when this approach is compared to the crisp approach. In conventional RSAR, a reduct is characterized as a subsetRof the attributes that have a similar data content as the full attribute set A. As far as the dependency function, this means that the qualities γ(R) and γ (A) are identical and equivalent to 1 if the dataset is consistent. Nonetheless, in the fuzzy-rough approach, this isn't necessarily the case as the uncertainty encountered when objects have a place with numerous fuzzy equivalence classes' outcomes in a reduced absolute dependency [2].

Note that with the fuzzy-rough QUICKREDUCT algorithm, fora dimensionality of n, (n2 + n)/2, assessments of the dependency function might be performed for the most pessimistic scenario dataset. However, as FRFS is utilized for dimensionality reduction earlier to any inclusion of the framework that will utilize those attributes belonging to the resultant reduct, this activity has no negative impact upon the run-time efficiency of the framework [2].

New Fuzzy Rough FS

This section presents three new techniques for fuzzy-rough FS, based on fuzzy similarity relations.

- **Fuzzy Lower Approximation-Based FS**

The previous method for FRFS used a fuzzy partitioning of the input space in order to determine fuzzy equivalence classes. Alternative definitions for the fuzzy lower and upper approximation scan be found in [46], where a T-transitive fuzzy similarity relation is used to approximate a fuzzy concept X

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

$$\mu_{\overline{R_P}X}(x) = \sup_{y\in\mathbb{U}} T(\mu_{R_P}(x,y), \mu_X(y)).$$

Here, I is a fuzzy implicator and T a t-norm. R_P is the fuzzy similarity relation induced by the subset of features P

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

where $\mu_{R_a}(x, y)$ is the degree to which objects x and y are similar for feature a. Many fuzzy similarity relations can be constructed for this purpose, for example

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

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

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

Where σ^2 is the variance of feature a. As these relations do not necessarily display T-transitivity, the fuzzy transitive closure must be computed for each attribute [47]. The combination of feature relations in (23) has been shown to preserve T-transitivity [51].

- Reduction: In a similar way to the original FRFS approach, the fuzzy positive region can be defined as

$$\mu_{\mathrm{POS}_{R_P}(Q)}(x) = \sup_{X\in\mathbb{U}/Q} \mu_{\underline{R_P}X}(x)$$

The resulting degree of dependency is A fuzzy-rough reduct R can be defined as a subset of features that preserves the dependency degree of the entire dataset, i.e., $\gamma'_R(D) = \gamma'_C(D)$. Based on this, a new fuzzy-rough QUICKREDUCT algorithm can be constructed that operates in the same way but uses (28) to gauge subset quality. A proof of the monotonicity of the dependency function can be found in the Appendix. Core features may be determined by considering the change in dependency of the full set of conditional features when individual attributes are removed

$$Core(\mathbb{C}) = \{a \in \mathbb{C} | \gamma'_{\mathbb{C}-\{a\}}(Q) < \gamma'_{\mathbb{C}}(Q)\}.$$

- **Fuzzy Boundary Region-Based FS**

Most approaches to crisp rough set FS and all approaches to fuzzy-rough FS use only the lower approximation for the evaluation of feature subsets. The lower approximation contains information regarding the extent of certainty of object membership to a given concept. However, the upper approximation contains information regarding the degree of uncertainty of objects, and hence, this information can be used to discriminate between subsets. For example, two subsets may result in the same lower approximation, but one subset may produce a smaller upper approximation. This subset will be more useful as there is less uncertainty concerning objects within the boundary region (the difference between upper and lower approximations). The fuzzy-rough boundary region for a fuzzy concept X may thus be defined as

$$\mu_{\mathrm{BND}_{R_P}(X)}(x) = \mu_{\overline{R_P}X}(x) - \mu_{\underline{R_P}X}(x)$$

The fuzzy-rough negative region for all decision concepts can be defined as follows:

$$\mu_{\mathrm{NEG}_{R_P}}(x) = N\Big(\sup_{X \in \mathbb{U}/Q} \mu_{\overline{R_P}X}(x)\Big).$$

Reduction: As the search for an optimal subset progresses, the object memberships to the boundary region for each concept diminish until a minimum is achieved. For crisp rough set FS, the boundary region will be zero for each concept when a reduct is found. This may not necessarily be the case for fuzzy-rough FS due to the additional uncertainty involved. The uncertainty for a concept X using features in P can be calculated as follows:

$$U_P(X) = \frac{\sum_{x \in \mathbb{U}} \mu_{\mathrm{BND}_{R_P}(X)}(x)}{|\mathbb{U}|}.$$

This is the average extent to which objects belong to the fuzzy boundary region for the concept X. The total uncertainty degree for all concepts, given a feature subset P, is defined as

$$\lambda_P(Q) = \frac{\sum_{X \in \mathbb{U}/Q} U_P(X)}{|\mathbb{U}/Q|}.$$

This is related to the conditional entropy measure that considers a combination of conditional probabilities H(Q|P) in order to gauge the uncertainty present using features in P. In the crisp case, the minimization of this measure can be used to discover reducts: if the entropy for a feature subset P is zero, then the subset is a reduct [43]. Again, a QUICKREDUCT-style algorithm can be constructed for locating fuzzy-rough reducts based on this measure. Instead of maximizing the dependency degree, the task of the algorithm is to minimize the total uncertainty degree. When this reaches the minimum for the dataset, a fuzzy-rough reduct has been found.

- **Fuzzy Discernibility Matrix-Based FS**

As mentioned previously, there are two main branches of research in crisp rough set-based FS: those based on the dependency degree and those based on discernibility matrices. The developments given earlier are solely concerned with the extension of the dependency degree to the fuzzy-rough case.

  - **Fuzzy Discernibility:** The approach presented here extends the crisp discernibility matrix by employing fuzzy clauses. Each entry in the fuzzy discernibility matrix is a fuzzy set to which every feature belongs to a certain degree. The extent to which a feature a belongs to the fuzzy clause C_{ij} is determined by the fuzzy discernibility measure [2]

$$\mu_{C_{ij}}(a) = N(\mu_{R_a}(i,j))$$

where N denotes fuzzy negation, and μ_{Ra} (i, j) is the fuzzy similarity of objects i and j; hence, μ_{Cij} (a) is a measure of the fuzzy discernibility. For the crisp case, if μ_{Cij} (a) = 1, then the two objects are distinct for this feature; if μ_{Cij} (a) = 0, the two objects are identical. For fuzzy cases, where μ_{Cij} (a) ∈ (0, 1), the objects are partly discernible. (The choice of fuzzy similarity relation must be identical to that of the fuzzy-rough dependency degree approach to find corresponding reducts.) Each entry in the fuzzy indiscernibility matrix is then a set of attributes and their corresponding memberships [2]

$$C_{ij} = \{a_x \mid a \in \mathbb{C}, x = N(\mu_{R_a}(i,j))\} \quad i,j = 1,\ldots,|\mathbb{U}|.$$

For example, an entry C_{ij} in the fuzzy discernibility matrix might be

$$C_{ij} : \{a_{0.4}, b_{0.8}, c_{0.2}, d_{0.0}\}.$$

This denotes that $\mu_{C_{ij}}(a) = 0.4$, $\mu_{C_{ij}}(b) = 0.8$, etc. In crisp discernibility matrices, these values are either 0 or 1 as the underlying relation is an equivalence relation. The example clause can be viewed as indicating the value of each feature—the extent to which the feature discriminates between the two objects i and j. The core of the dataset is defined as [2]

$$Core(\mathbb{C}) = \{a \in \mathbb{C} \mid \exists C_{ij}, \mu_{C_{ij}}(a) > 0, \\ \forall f \in \{\mathbb{C} - a\} \mu_{C_{ij}}(f) = 0\}.$$

- **Fuzzy Discernibility Function:** As with the crisp approach, the entries in the matrix can be used to construct the fuzzy discernibility function [2]

$$f_D(a_1^*, \ldots, a_m^*) = \wedge\{\vee\, C_{ij}^* | 1 \leq j < i \leq |\mathbb{U}|\}$$

where $C^*_{ij} = \{a^*_x | a_x \in C_{ij}\}$. The function returns values in [0, 1], which can be seen to be a measure of the extent to which the function is satisfied for a given assignment of truth values to variables. To discover reducts from the fuzzy discernibility function, the task is to find the minimal assignment of the value 1 to the variables such that the formula is maximally satisfied. By setting all variables to 1, the maximal value for the function can be obtained as this provides the most discernibility between objects.

- **Decision-Relative Fuzzy Discernibility Matrix:** As withthe crisp discernibility matrix, for a decision system, the decision feature must be taken into account for achieving reductions; only those clauses with different decision values are included in the crisp discernibility matrix. For the fuzzy version, this is encodedas[2]

$$f_D(a_1^*, \ldots, a_m^*) = \{\wedge\{\{\vee\, C_{ij}^*\} \leftarrow q_{N(\mu_{R_q}(i,j))}\} | \\ 1 \leq j < i \leq |\mathbb{U}|\}$$

for decision feature q, where $\leftarrow$ denotes fuzzy implication. This construction allows the extent to which decision values differ to affect the overall satisfiability of the clause. If $\mu_{C_{ij}}(q) = 1$, then this clause provides maximum discernibility (i.e., the two objects are maximally different according to the fuzzy similarity measure). When the decision is crisp and crisp equivalence is used, $\mu_{C_{ij}}(q)$ becomes 0 or 1.

- **Reduction:** For the purposes of finding reducts, use of thefuzzy intersection of all clauses in the fuzzy discernibility function may not provide enough information to evaluate subsets. Here, it may be more informative to consider the individual satisfaction of each clause for a given set of features. The degree of satisfaction of a clause C_{ij} for a subset of features P is definedas[2]

$$\mathrm{SAT}_P(C_{ij}) = \bigcup_{a \in P} \{\mu_{C_{ij}}(a)\}.$$

Returning to the example, if the subset P = {a, c} is chosen, the resulting degree of satisfaction of the clause is

$$\mathrm{SAT}_P(C_{ij}) = \{0.4 \vee 0.2\} = 0.6$$

using the Łukasiewicz t-conorm min(1, x + y). For the decision-relative fuzzy indiscernibility matrix, the decision feature q must also be taken into account [2]

$$\mathrm{SAT}_{P,q}(C_{ij}) - \mathrm{SAT}_P(C_{ij}) \leftarrow \mu_{C_{ij}}(q).$$

For the example clause, if the corresponding decision values are crisp and are different, the degree of satisfaction of the clause is

$$\mathrm{SAT}_{P,q}(C_{ij}) = \mathrm{SAT}_P(C_{ij}) \leftarrow 1 \\ = 0.6 \leftarrow 1 \\ = 0.6.$$

For a subset P, the total satisfiability of all clauses can be calculated as

$$\mathrm{SAT}(P) = \frac{\sum_{i,j \in \mathbb{U}, i \neq j} \mathrm{SAT}_{P,q}(C_{ij})}{\sum_{i,j \in \mathbb{U}, i \neq j} \mathrm{SAT}_{\mathbb{C},q}(C_{ij})}$$

where C is the full set of conditional attributes, and hence, the denominator is a normalizing factor. If this value reaches 1 for a subset P, then the subset is a fuzzy-rough reduct. A proof of the monotonicity of the function SAT(P) can be found in the Appendix.

The complexity of the algorithm is the same as that of FRFS in that $O((n^2 + n)/2)$ calculations of the evaluation function (SAT(P)) are performed in the worst case. Additionally, this approach requires the construction of the fuzzy discernibility matrix, which has a complexity of $O(a * o^2)$ for a dataset containing a attributes and o objects[2].

- **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 [56]. The other is definite with fuzzy similarity relations and a fuzzy implicator [2, 57]. 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:

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

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

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

where σ^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.

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 [58].

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:

$$\mathrm{J}(\mathrm{U}, c_1, c_2, \ldots, c_c) = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c} \sum_{j}^{n} u_{ij}^m d_{ij}^2$$

where u_{ij} is between 0 and 1, ci is the clustering center of the fuzzy set, d_{ij} = ||ci - xj|| is the Euclidean distance of ith clustering center jth 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:

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

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

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 ", maximum number of iterations T and weighted index m. Output: U

t = 1
do
    $c_i, i = 1, 2, \ldots, c$ with Equation (13)
    $U = [u_{ij}]$ with Equation (14).
    Update c_i and U
    t←t + 1
Until $t = T$ *or* $\|U_t - U_{t-1}\| \leq \varepsilon$
Return U

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

$$\mu_{\underline{R_P}X}(x) = inf_{y \in U} I\left(\mu_{R_P}(x, y), \mu_X(y)\right)$$

where I is the fuzzy implicator. m_{RP} (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.

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

where m_{Ra} (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:

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

Because equivalence relations are used to model equality, fuzzy equivalence relations are commonly considered to represent approximate equality or similarity [57, 59]. We use the fuzzy equivalence relation R in the literature [60]:

$$R(x, y) = \frac{a - a|x - y| + bmin(x, y)}{a - (a - 1)|x - y| + bmin(x, y)}.$$

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 ‹ {empty set}

Output: R

For each x, y in U:
    $\mu_x = FCM(x)$, $\mu_y = FCM(y)$
    Calculate $E(x, y)$
    $\gamma'_{prev} - 0, \gamma'_{best} - 0$
do
    $T \leftarrow R$
    $\gamma'_{prev} - \gamma'_{best}$
    if $\gamma_{R \cup \{a\}}(D) > \gamma_T(D), \forall a \in (C - R)$,
        $T \leftarrow R \cup \{a\}$
    $R \leftarrow T$
    $\gamma'_{best} \leftarrow \gamma_R(D)$
until $\left(\gamma'_{best} - \gamma'_{prev}\right) \times |U| < 1$
return R

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

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

The definitions of positive region and dependency are the same as we mentioned above [2].

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

$$\gamma'_P(Q) - \frac{\sum_{x \subset U} \mu_{POS_{R_P}(Q)}(x)}{|\mathbb{U}|}$$

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 g 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.

Experimentation

This section presents the initial experimental evaluation of the selection methods for the task of pattern classification over nine benchmark datasets from [61] and [38] with two classifiers.

- **Experimental Setup**

FRFS uses a precategorization step that generates associated fuzzy sets for a dataset. For the new fuzzy-rough methods, the Łukasiewicz fuzzy connectives are used, with fuzzy similarity defined in (26). After FS, the datasets are reduced according to the various discussed techniques

We used five reducts techniques derived from fuzzy rough set theory in order to analyzing performance. Seven datasets as shown in table 1, were employed to evaluate performance of each methods. The results of all methods, as well as unreduced datasets in terms of the number of selected features, are also demonstrated in Table 1.C-FRFS always gets the smallest reduced subset andFRFS does not perform well and sometimes cannot select features.

Dataset	Objects	Features	Reduct Size					
			FRFS	B-FRFS	L-FRFS	FDM	C-FRFS	T-FRFS
Cleveland	297	14	11	9	9	9	4	6
Glass	214	10	9	9	10	9	4	8
Wine	178	14	10	6	6	6	4	4
Iris	150	4	4	3	3	3	3	3
User	431	5	5	4	4	4	3	4
Sonar	208	60	5	4	4	3	2	4
Yeast	1484	8	8	7	7	6	6	7

So, all the techniques based upon fuzzy rough set theory are enough capable to produce effective reducts. Among all of these, C-FRFS produce quiet better reducts among the others.

Conclusion

This paper has presented six new techniques for FRFS based on the use of fuzzy *T*-transitive similarity relations that alleviate problems encountered with FRFS. The first development, based on fuzzy lower approximations, uses the similarity relations to construct approximations of decision concepts and evaluates these through a new measure of feature dependency. The next techniques employs the information in the fuzzy boundary region to guide the FS search process. When this is minimized, a fuzzy-rough reduct has been obtained. Another extends the concept of the discernibility matrix to the fuzzy case, allowing features to belong to entries to a certain degree. An example FS algorithm is given to illustrate how reductions may be achieved. Other techniques, based upon the threshold based fuzzy rough set features selection that also produce effective reducts. The most effective techniques, based upon C-Means fuzzy rough set theory produce smallest reducts among all of the mentioned.

References

1. Zhao, Ruonan, LizeGu, and Xiaoning Zhu. "Combining fuzzy C-means clustering with fuzzy rough feature selection." *Applied Sciences* 9.4 (2019): 679.
2. Jensen, Richard, and Qiang Shen. "New approaches to fuzzy-rough feature selection." IEEE Transactions on fuzzy systems 17.4 (2008): 824-838.
3. 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.
4. 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.
5. 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.
6. Yao, Y. Probabilistic approaches to rough sets. Expert Syst. 2003, 20, 287–297. [CrossRef]
7. 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]
8. 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]
9. 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; Słowi ´ nski, R., Ed.; Springer: Dordrecht, The Netherlands, 1992; pp. 3–18.
10. 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.
11. Kryszkiewicz, M. Rough set approach to incomplete information systems. Inf. Sci. 1998, 112, 39–49. [CrossRef]
12. Jelonek, J.; Krawiec, K.; Slowi ´ nski, R. Rough set reduction of attributes and their domains for neural networks. Comput. Intell. 1995, 11, 339–347. [CrossRef]
13. Swiniarski, R.W.; Skowron, A. Rough set methods in feature selection and recognition. Pattern Recognit. Lett. 2003, 24, 833–849. [CrossRef]
14. 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]
15. 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] Appl. Sci. 2019, 9, 679 17 of 18
16. 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]
17. 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]
18. Zadeh, L.A. Fuzzy sets. Inf. Control 1965, 8, 338–353. [CrossRef]
19. Pawlak, Z. Rough sets and fuzzy sets. Fuzzy Sets Syst. 1985, 17, 99–102. [CrossRef]
20. Yao, Y. A comparative study of fuzzy sets and rough sets. Inf. Sci. 1998, 109, 227–242. [CrossRef]
21. Dubois, D.; Prade, H. Rough fuzzy sets and fuzzy rough sets. Int. J. Gen. Syst. 1990, 17, 191–209. [CrossRef]
22. 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]
23. Jensen, R.; Shen, Q. Semantics-preserving dimensionality reduction: Rough and fuzzy-rough-based approaches. IEEE Trans. Knowl. Data Eng. 2004, 16, 1457–1471. [CrossRef]

24. Wu,W.-Z.; Mi, J.-S.; Zhang, W.-X. Generalized fuzzy rough sets. Inf. Sci. 2003, 151, 263–282. [CrossRef]
25. Bhatt, R.B.; Gopal, M. On fuzzy-rough sets approach to feature selection. Pattern Recognit. Lett. 2005, 26, 965–975. [CrossRef]
26. 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]
27. Jensen, R.; Shen, Q. Fuzzy-rough sets assisted attribute selection. IEEE Trans. Fuzzy Syst. 2007, 15, 73–89. [CrossRef]
28. Jensen, R.; Shen, Q. Fuzzy-rough data reduction with ant colony optimization. Fuzzy Sets Syst. 2005, 149, 5–20. [CrossRef]
29. 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.
30. Sheeja, T.K.; Kuriakose, A.S. A novel feature selection method using fuzzy rough sets. Comput. Ind. 2018, 97, 111–116. [CrossRef]
31. Qian, Y.;Wang, Q.; Cheng, H.; Liang, J.; Dang, C. Fuzzy-rough feature selection accelerator. Fuzzy Sets Syst 2015, 258, 61–78. [CrossRef]
32. Chouchoulas and Q. Shen, “Rough set-aided keyword reduction for text categorisation,” Appl. Artif. Intell., vol. 15, no. 9, pp. 843–873, 2001
33. D¨untsch and G. Gediga, Rough Set Data Analysis: A Road to Non- Invasive Knowledge Discovery. Bangor, ME: Methodos, 2000
34. G. Shafer, AMathematical Theory of Evidence. Princeton, NJ: Princeton Univ. Press, 1976.
35. Q. Shen and A. Chouchoulas, “A fuzzy-rough approach for generating classification rules,” Pattern Recognit., vol. 35, no. 11, pp. 341–354, 2002.
36. D. Dubois and H. Prade, “Putting rough sets and fuzzy sets together,” Intell. Decis. Support, pp. 203–232, 1992.
37. Rough-Fuzzy Hybridization: A New Trend in Decision Making, S. K. Pal and A. Skowron, Eds. New York: Springer-Verlag, 1999.
38. R. Jensen and Q. Shen, “Fuzzy-rough sets assisted attribute selection,” IEEE Trans. Fuzzy Syst., vol. 15, no. 1, pp. 73–89, Feb. 2007.
39. H. Thiele, “Fuzzy rough sets versus rough fuzzy sets—An interpretation and a comparative study using concepts of modal logics,” Tech. Rep. CI-30/98, Univ. Dortmund, Dortmund, Germany, 1998.
40. Y. Y. Yao, “A comparative study of fuzzy sets and rough sets,” Inf. Sci., vol. 109, pp. 21–47, 1998.
41. D. S. Yeung, D. Chen, E. C. C. Tsang, J. W. T. Lee, and W. Xizhao, “On the generalization of fuzzy rough sets,” IEEE Trans. Fuzzy Syst., vol. 13, no. 3, pp. 343–361, Jun. 2005.
42. W. Z. Wu and W. X. Zhang, “Constructive and axiomatic approaches of fuzzy approximation operators,” Inf. Sci., vol. 159, no. 3–4, pp. 233–254, 2004.
43. R. Jensen and Q. Shen, “Semantics-preserving dimensionality reduction: Rough and fuzzy-rough based approaches,” IEEE Trans. Knowl. Data Eng., vol. 16, no. 12, pp. 1457–1471, Dec. 2004.
44. R. Jensen and Q. Shen, “Fuzzy-rough attribute reduction with application to Web categorization,” Fuzzy Sets Syst., vol. 141, no. 3, pp. 469–485, 2004.
45. Y. Y. Yao, “Combination of rough and fuzzy sets based on α-level sets,” in Rough Sets and Data Mining: Analysis of Imprecise Data, T. Y. Lin and N. Cereone, Eds. Norwell, MA: Kluwer, 1997, pp. 301–321.
46. M. Radzikowska and E. E. Kerre, “A comparative study of fuzzy rough sets,” Fuzzy Sets Syst., vol. 126, no. 2, pp. 137–155, 2002.
47. M.DeCock, C. Cornelis, and E. E.Kerre, “Fuzzy rough sets: The forgotten step,” IEEE Trans. Fuzzy Syst., vol. 15, no. 1, pp. 121–130, Feb. 2007.

48. W. Z. Wu, Y. Leung, and J. S. Mi, “On characterizations of (I,T)-fuzzy rough approximation operators,” Fuzzy Sets Syst., vol. 154, no. 1, pp. 76– 102, 2005.

49. L. A. Zadeh, “The concept of a linguistic variable and its application to approximate reasoning—I,” Inf. Sci., vol. 8, pp. 199–249, 1975.

50. Y. Y. Yao, Y. Zhao, and J. Wang, “On reduct construction algorithms,” in Proc. 1st Int. Conf. Rough Sets Knowl. Technol., 2006, pp. 297– 304.

51. M. Wallace, Y. Avrithis, and S. Kollias, “Computationally efficient sup-t transitive closure for sparse fuzzy binary relations,” Fuzzy Sets Syst., vol. 157, no. 3, pp. 341–372, 2006.

52. Y.Y.Yao, “Decision-theoretic rough setmodels,” in Proc. Int.Conf. Rough Sets Knowl. Technol., 2007 (Lecture Notes in Artificial Intelligence), 2007, vol. 4481, pp. 1–12.

53. G. C. Y. Tsang, D. Chen, E. C. C. Tsang, J. W. T. Lee, and D. S. Yeung, “On attributes reduction with fuzzy rough sets,” in Proc. 2005 IEEE Int. Conf. Syst., Man, Cybern., Oct. 2005, vol. 3, pp. 2775–2780.

54. X. Z.Wang, Y. Ha, and D. Chen, “On the reduction of fuzzy rough sets,” in Proc. 2005 Int. Conf. Mach. Learn. Cybern., 2005, vol. 5, pp. 3174–3178.

55. Øhrn, “Discernibility and rough sets in medicine: Tools and applications,” Dept. Comput. Inf. Sci., Norwegian Univ. Sci. Technol., Trondheim, Norway, Rep., vol. 239, 1999.

56. Jensen, R.; Shen, Q. Fuzzy–rough attribute reduction with application to web categorization. Fuzzy Sets Syst. 2004, 141, 469–485. [CrossRef]

57. Radzikowska, A.M.; Kerre, E.E. A comparative study of fuzzy rough sets. Fuzzy Sets Syst. 2002, 126, 137–155. [CrossRef]

58. Bezdek, J.C.; Ehrlich, R.; Full,W. FCM: The fuzzy c-means clustering algorithm. Comput. Geosci. 1984, 10, 191–203. [CrossRef]

59. De Cock, M.; Cornelis, C.; Kerre, E.E. Fuzzy rough sets: The forgotten step. IEEE Trans. Fuzzy Syst. 2007, 15, 121–130. [CrossRef]

60. Li, Y.; Qin, K.; He, X. Some new approaches to constructing similarity measures. Fuzzy Sets Syst. 2014, 234, 46–60. [CrossRef]

61. C. L. Blake and C. J. Merz. (1998) UCI Repository of machine learning databases, University of California, Irvine. Available: http://www.ics.uci.edu/~mlearn/.

❖◆❖