Check for updates

Received: 30 June 2020 | Revised: 27 November 2020 | Accepted: 7 December 2020

DOI: 10.1002/widm.1402

ADVANCED REVIEW

WIREs DATA MINING AND KNOWLEDGE DISCOVERY

WILEY

Fuzzy rough sets and fuzzy rough neural networks for feature selection: A review

Wanting Ji[1] | Yan Pang[2] | Xiaoyun Jia[1] | Zhongwei Wang[2] | Feng Hou[1] | Baoyan Song[3] | Mingzhe Liu[4] | Ruili Wang[1]

[1]School of Natural and Computational Sciences, Massey University, Auckland, New Zealand

[2]Central South University of Forestry and Technology, Changsha, China

[3]Liaoning University, Shenyang, China

[4]Chengdu University of Technology, Chengdu, China

Correspondence
Ruili Wang, School of Natural and Computational Sciences, Massey University, Auckland, New Zealand.
Email: ruili.wang@massey.ac.nz

Funding information
China Scholarship Council; Hunan Key Laboratory of Intelligent Logistics Technology, Grant/Award Number: 2019TP1015; Liaoning Province Key R&D Plan Guidance Plan Project, Grant/Award Number: 2017231011; Social Sciences Foundation of Hunan Province, Grant/Award Number: 19YBA375

Abstract

Feature selection aims to select a feature subset from an original feature set based on a certain evaluation criterion. Since feature selection can achieve efficient feature reduction, it has become a key method for data preprocessing in many data mining tasks. Recently, many feature selection strategies have been developed since in most cases it is infeasible to obtain an optimal/reduced feature subset by using exhaustive search. Among these strategies, fuzzy rough set theory has proved to be an ideal candidate for dealing with uncertain information. This article provides a comprehensive review on the fuzzy rough set theory and two fuzzy rough set theory based feature selection methods, that is, fuzzy rough set based feature selection methods and fuzzy rough neural network based feature selection methods. We review the publications related to the fuzzy rough theory and its applications in feature selection. In addition, the challenges in the two types of feature selection methods are also discussed.

This article is categorized under:

Technologies > Machine Learning

KEYWORDS

fuzzy rough neural network, fuzzy rough set theory, fuzzy set theory, rough set theory

1 | INTRODUCTION

In the past few decades, feature selection has become an apparent need in many machine-learning tasks, such as regression, classification, and clustering. The input data of a machine learning task need to be preprocessed before processing them further for selecting useful and relevant features of the data since some features of the data may be irrelevant to or redundant for the task. Thus, feature selection (Jović, Brkić, & Bogunović, 2015; Saeys, Inza, & Larrañaga, 2007; Sheeja & Kuriakose, 2018) has been introduced, which is a process to select a feature subset (i.e., a low-dimensional feature set) from the original feature set (i.e., a high-dimensional feature set) based on a certain evaluation criterion. The objectives of feature selection are manifold, including avoiding overfitting and improving the performance of datasets. Early algorithms for feature selection include the Branch & Bound algorithm (Clausen, 1999), Focus (Almuallim & Dietterich, 1991), Relief (Kira & Rendell, 1992), Las Vegas Filter (Liu & Setiono, 1996), Correlation-based Feature Selection (Doshi, 2014), and many others.

The information in the real world is normally vague and incomplete. Since real-world information contains many subjective concepts, such as tall, good, and young, these concepts generate vagueness. Such concepts have no clear

© 2021 Wiley Periodicals LLC.

boundaries, hence are difficult to be processed using set theory. Incompleteness means that the elements in a set with the same features may lead to different outcomes. For example, a group of patients have the same symptoms but suffer from different diseases, and hence are indistinguishable. Such vague or incomplete information is incapable to be handled by set theory for feature selection.

The fuzzy rough set theory was developed by Dubois and Prade (1990), which was inspired by the relevance and complementarity between the fuzzy set theory and the rough set theory. In the fuzzy rough set theory, the relationship between any two elements in a set was described by a fuzzy relation instead of an equivalence relation. Specifically, the fuzzy rough set theory defined two operators, named fuzzy rough upper and lower approximations, to approximate a set according to fuzzy relations. In other words, the fuzzy rough set theory allowed a partial membership of an element to belong to the upper and lower approximations, and the approximate equalities between elements could be simulated by the fuzzy relations. Therefore, in the fuzzy rough set theory, the elements in a set were discernible/distinguished from each other to a certain extent, instead of being distinguishable/discernible or indistinguishable/indiscernible.

The fuzzy rough set theory and its applications are widely used to solve many problems in data mining, pattern recognition, and machine learning. Recently, fuzzy rough set theory has achieved great success in dealing with feature selection problems. In fuzzy rough set theory based feature selection methods, the fuzzy rough set theory was utilized to model the strength (or the importance) of each feature in a feature set, and then the redundant features in the feature set could be removed to obtain an optimal/reduced feature subset. This process reduced the number of features that would be processed by a decision system and achieved performance gains (Lu & Wang, 2009; Tsang, Chen, Yeung, Wang, & Lee, 2008).

This article aims to provide a comprehensive review of feature selection methods based on fuzzy rough sets or fuzzy rough neural networks. The contributions of this review include (i) reviewing the development of fuzzy rough sets and fuzzy rough neural networks for feature selection, and (ii) discussing the challenges of fuzzy rough sets and fuzzy rough neural networks for feature selection.

The rest of this review is organized as follows. Section 2 presents the background related to the fuzzy rough set theory, including the concepts in the fuzzy set theory and the rough set theory, as well as the theoretical progress of the fuzzy rough set theory. Section 3 provides a review of fuzzy rough set based feature selection methods and discusses the challenges in these methods. Section 4 provides a review of fuzzy rough neural networks based feature selection methods and discusses the challenges in these methods. Section 5 presents a conclusion of this article.

2 | BACKGROUND AND PRELIMINARIES

This section first introduces the concepts and operations of fuzzy set theory, rough set theory, and fuzzy rough set theory, and then reviews the theoretical progress of fuzzy rough set theory.

2.1 | Fuzzy set theory

If universe U is a finite and nonempty set, then the mapping from U into the unit interval [0,1] is a fuzzy set defined on U. The mapping can be written as μ_A: $U \rightarrow [0,1]$, where A denotes a fuzzy set defined on U, and $\mu_A(*)$ is a membership function which is connected with A.

For each element $x \in U$, the value of $\mu_A(x)\mu_A(x)$ presents a membership degree of the element x to A. Thus, the fuzzy cardinality of A, which is presented as $|A|$, is the sum of all membership degrees, that is, $|A| = \sum_{x \in U}\mu_A(x)$

The membership function measures the similarity degree between a fuzzy set and an element. Usually, it is designed by users' arbitrarily, such as the trapezoidal membership function (Pedrycz, 1994), triangular membership function (Pedrycz, 1994), and Gaussian membership function (Zeng, Li, Liu, Zhang, & Chen, 2015), or machine learning methods, for example, Genetic Algorithms and Artificial Neural Networks. Thus, the fuzzy set A can be presented as a collection of ordered pairs: $A = \{(x, \mu_A(x)), x \in U\}$.

Moreover, Zadeh (1965) defined several operations on fuzzy sets. Given a universe U, two fuzzy sets A and B, and a particular element x of U, then the union, intersection, and complement operations on the fuzzy sets can be presented as follows:

$$\mu_{A\cup B}(x) = \max(\mu_A(x), \mu_B(x)), \tag{1}$$

$$\mu_{A\cap B}(x) = \min(\mu_A(x), \mu_B(x)), \tag{2}$$

$$\mu_{\neg A}(x) = 1 - \mu_A(x). \tag{3}$$

2.2 | The rough set theory

In the rough set theory, equivalence relations are developed to measure the distinguishability or discernibility between any two elements in the universe U (Vluymans, D'eer, Saeys, & Cornelis, 2015). If A is a subset of U, then the lower approximation of A consists of the elements, which are definitely in A, while the upper approximation of A consists of the elements which may belong to A. Given an equivalence relation R, the upper and lower approximations of A can be mathematically presented as follows:

$$\overline{\mathrm{apr}}_R(A) = \left\{x \in U \mid [x]_R \cap A \neq \emptyset\right\} = \cup\left\{[x]_R \in U/R \mid [x]_R \cap A \neq \emptyset\right\} \tag{4}$$

$$\underline{\mathrm{apr}}_R(A) = \left\{x \in U \mid [x]_R \subseteq A\right\} = \cup\left\{[x]_R \in U/R \mid [x]_R \subseteq A\right\}, \tag{5}$$

where $[x]_R$ represents the equivalence class of the element x, and U/R represents the quotient set of equivalence classes. Moreover, in some cases, the first phase of Equations (4) and (5) can be referred to as an element-based definition, and the second phase of Equations (4) and (5) can be referred to as a granule-based definition.

The boundary region of A is constructed according to the upper and lower approximations of A. The boundary region of A contains the elements that cannot be certainly determined whether they are inside or outside of A. Formally, the boundary region of A is described by the differences between the upper and lower approximations of A: $\mathrm{bnd}_R(A) = \overline{\mathrm{apr}}_R(A) - \underline{\mathrm{apr}}_R(A)$.

If $\mathrm{bnd}_R(A) \neq \emptyset$, then A is a rough set defined on U, otherwise, there is no uncertainty in the set (or information). Figure 1 illustrates the concepts in the rough set theory. In addition, Pawlak (1982) also defined several operations on rough sets. For any rough sets A, $B \subseteq U$, the upper and lower approximations of the rough sets follow the following properties:

(i) $\underline{\mathrm{apr}}_R(A) = \neg\overline{\mathrm{apr}}_R(\neg A)$, $\overline{\mathrm{apr}}_R(A) = \neg\underline{\mathrm{apr}}_R(\neg A)$;
(ii) $\underline{\mathrm{apr}}_R(A) \subseteq A \subseteq \overline{\mathrm{apr}}_R(A)$;
(iii) $\underline{\mathrm{apr}}_R(\emptyset) = \overline{\mathrm{apr}}_R(\emptyset) = \emptyset$
(iv) $\underline{\mathrm{apr}}_R(U) = \overline{\mathrm{apr}}_R(U) = U$,
(v) $\underline{\mathrm{apr}}_R(A \cap B) = \underline{\mathrm{apr}}_R(A) \cap \underline{\mathrm{apr}}_R(B)$, $\underline{\mathrm{apr}}_R(A \cup B) \supseteq \underline{\mathrm{apr}}_R(A) \cup \underline{\mathrm{apr}}_R(B)$;
(vi) $\overline{\mathrm{apr}}_R(A \cup B) = \overline{\mathrm{apr}}_R(A) \cup \overline{\mathrm{apr}}_R(B)$, $\overline{\mathrm{apr}}_R(A \cap B) \subseteq \overline{\mathrm{apr}}_R(A) \cap \overline{\mathrm{apr}}_R(B)$;
(vii) $\underline{\mathrm{apr}}_R(A) = \underline{\mathrm{apr}}_R\left(\underline{\mathrm{apr}}_R(A)\right) = \overline{\mathrm{apr}}_R\left(\underline{\mathrm{apr}}_R(A)\right)$, $\overline{\mathrm{apr}}_R(A) = \underline{\mathrm{apr}}_R(\overline{\mathrm{apr}}_R(A)) = \overline{\mathrm{apr}}_R(\overline{\mathrm{apr}}_R(A))$.

The rough set theory is capable of dealing with the incompleteness in data. However, to obtain useful results, the discretization of any real-valued features is required (Vluymans, Fernández, Saeys, Cornelis, & Herrera, 2018). This

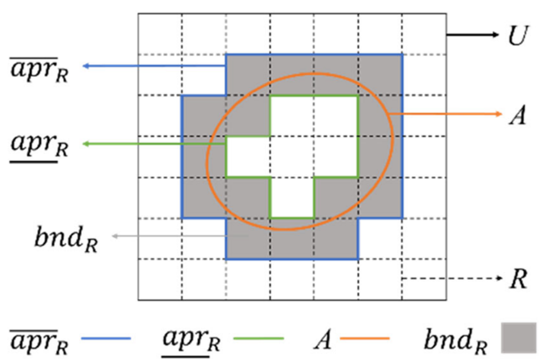

FIGURE 1 Rough approximations. Given a universe U, an equivalence relation R can be represented as the dotted line that divides U. If A is a rough set (the area inside the orange circle), then the area inside the green line is the lower approximation of A; the area inside the blue line is the upper approximation of A; the area between the blue and green lines is the boundary region of A

limits the performance/effectiveness of the rough set theory. Thus, the fuzzy rough set theory has been developed to address this problem.

2.3 | The fuzzy rough set theory

The fuzzy rough set theory is a hybrid theory, which encapsulates fuzziness and roughness into a single model. It studies the operators that approximate fuzzy sets with fuzzy relations (Vluymans et al., 2015). Thus, the discretization requirement of the rough set theory is removed by computing the similarity between instances with fuzzy relations (Ma et al., 2018; Wang et al., 2018; Wang, Ji, & Song, 2018; Zhao & Zhang, 2011).

If R is a fuzzy relation on U, and A denotes a fuzzy set defined on U, then the upper and lower approximations of A is a pair of fuzzy sets defined on U, which can be presented as:

$$\mathrm{a\bar{p}r}_R(A) = \sup\{A(y)|y \in [x]_R\}, x \in U, \tag{6}$$

$$\underline{\mathrm{apr}_R}(A) = \inf\{A(y)|y \in [x]_R\}, x \in U, \tag{7}$$

where $[x]_R$ is the equivalence class of the element x based on R. If Equation (6) is not equal to Equation (7), then A is a fuzzy rough set of U, where $\underline{\mathrm{apr}_R}(A)$ is the fuzzy rough lower approximation of A; $\mathrm{a\bar{p}r}_R(A)$ is the fuzzy rough upper approximation of A; $bnd_R(A)$ is the boundary region of A. Moreover, the operations on fuzzy rough sets follow the same properties as on rough sets.

2.4 | The theoretical progress of the fuzzy rough set theory

The fuzzy rough set theory was first developed by Dubois and Prade (1990). They defined the approximation operators according to fuzzy logical connectives. Thus, the upper and lower approximation of a subset $A \subseteq U$ can be defined by a fuzzy relation R of the Kleene–Dienes implicator and the minimum t-norm (Vluymans et al., 2015). However, this definition of fuzzy rough sets is sensitive to noisy data and perturbations. In other words, mislabeled data may lead to significantly different fuzzy approximations. Later, various fuzzy rough sets were developed.

To address the above limitation, Ziarko (1993) proposed a Variable Precision Rough Set model based on variable precision. In VPRS, if the proportion $|[x]_R \cap A|/|A|$ can exceed a certain threshold, then each $x \in U$ will belong to the upper or the lower approximation of $A \subseteq U$. However, VPRS is incapable to deal with numerical features. Thus, to address this limitation, numerous works have been proposed in the field of the fuzzy rough set theory, which can be divided into three groups: (i) creating new models, (ii) defining the approximated sets, and (iii) changing aggregation operators.

Creating new models. Vague quantifiers (such as "some" and "most") were introduced into the definition of upper and lower approximation by Vaguely Quantified Rough Set (VQRS) model (Cornelis, De Cock, & Radzikowska, 2007). The influence of mislabeled data was reduced by overseeing outliers in calculating the fuzzy upper and lower approximations to deal with class noise in the Soft Fuzzy Rough Set (SFRS) (Hu, An, & Yu, 2010; Hu, An, Yu, & Yu, 2011). Then, Variable Precision Fuzzy Rough Set (VPFRS) model was developed by Mieszkowicz-Rolka and Rolka (2004, 2008), which utilized fuzzy inclusion to calculate the fuzzy memberships of data to the upper and lower approximations. This model also can be considered as an extended version of the VPRS model. Further, a fuzzy rough set definition based on residual implication θ and its dual σ was developed by Mi and Zhang (2004). The new definition utilized fuzzy relations to produce fuzzy granulated spaces. In addition, a variable precision (θ, σ)-fuzzy rough set model was developed based on granular (θ, σ)-fuzzy rough sets by Yao, Mi, and Li (2014) by using the absolute error limit.

Defining the approximated sets. Some methods introduced a threshold for membership degrees to limit the influence of outliers. For example, Radzikowska and Kerre (2002) developed a general fuzzy rough set model, named $(\mathcal{T}, T)$-fuzzy rough sets, based on the T-equivalence relations. In their method, several fuzzy rough sets were defined using the proposed fuzzy similarity relations, and the fuzzy upper and lower approximation were defined by a triangular norm T and a fuzzy implicator $\mathcal{T}$, respectively. Further, a generalized fuzzy rough model was developed by Wang, Tsang, Zhao, Chen, and Yeung (2007), in which a β-cut was used to define the upper and lower approximation based on the similarity between two elements. Shao, Liu, and Zhang (2007) explored the rough set approximations within Formal concept analysis (FCA) in a fuzzy environment, and developed a fuzzy rough method for attribute reduction and rule acquisition. A Fuzzy Variable Precision Rough Set (FVPRS) model was developed by Zhao, Tsang, and Chen (2009), in which a controlled threshold was introduced into the knowledge representation of fuzzy rough sets and the fuzzy upper and lower approximations with variable precision were also constructed. Besides, many other methods were proposed and utilized to define the upper and lower approximations for a fuzzy rough set model, such as Robust Fuzzy Rough Sets (RFRS; Hu, Zhang, An, Zhang, & Yu, 2011) and β-Precision Fuzzy Rough Sets (β-PFRS; Riza et al., 2014).

Changing aggregation operators. Boixader, Jacas, and Recasens (2000) first studied the fuzzy upper and lower approximation operators and their relations with fuzzy rough sets. A β-PFRS model developed by Salido and Murakami (2003). They introduced β-precision aggregation triangular operators by using β-precision quasi t-norms. Yeung, Chen, Tsang, Lee, and Xizhao (2005) explored the connections among the previous fuzzy rough sets and generalized the fuzzy rough set theory utilizing arbitrary fuzzy relations. Note that, both Salido and Murakami (2003) and Yeung et al. (2005) were built on the algebraic structure of unit interval [0,1]. Later, a fuzzy rough set model with a crisp upper (or lower) approximation was developed by Hu, Xie, and Yu (2007) based on fuzzy inclusion and fuzzy granulation, where a symmetric function was introduced to calculate fuzzy similarity relations among elements, and then the similarity relations were transformed into fuzzy equivalence relations. In addition, an Ordered Weighted Average (OWA) rough set model was developed by Cornelis, Verbiest, and Jensen (2010) based on ordered weighted average operators, which respected monotonicity with regard to the used fuzzy indiscernibility relation.

In addition, a set of axioms on fuzzy rough sets was developed by Morsi and Yakout (1998) using t-norms and T-residuated implication. However, their research was limited to fuzzy T-rough sets, which were developed by fuzzy T-similarity relations. In other words, when these relations degenerate into crisp relations, they can be ordinary (crisp) equivalence relations. To solve this problem, a general framework for fuzzy rough sets was proposed by Wu, Mi, and Zhang (2003) using both constructive and axiomatic methods. Specifically, the constructive method defined a pair of upper and lower generalized approximation operators and explored the relationships between the defined fuzzy rough approximation operators and fuzzy relations. Meanwhile, the axiomatic method characterized the defined fuzzy rough approximation operators through different axiom sets.

Later, Wei, Liang, and Qian (2012) introduced several rough set models for hybrid data and investigated the relationships between these models. They also analyzed the relationships between fuzzy hybrid granules and neighborhood hybrid granules. Further, Chen, Kwong, He, and Wang (2012) proposed a geometric interpretation of membership functions, in terms of the square distances in Krein spaces, utilizing the lower approximations of fuzzy rough sets. This proved the geometric interpretation of fuzzy similarity relations in a Krein space.

Over the past few decades, various fuzzy rough models were proposed in the development of the fuzzy rough set theory. These models have successfully been applied to overcome real-world problems (Bhatt & Gopal, 2008; Fernández-Riverola, Diaz, & Corchado, 2006; Hong, Wang, Wang, & Chien, 2000; Hu, Yu, & Guo, 2010; Leung, Fischer, Wu, & Mi, 2008; Li, Shiu, & Pal, 2006; Wang, 2003), especially in feature selection (or attribute reduction) (Hassanien, 2007; Hu et al., 2007; Jensen & Shen, 2004a; Jensen & Shen, 2007). This article reviews two types of fuzzy rough set theory based feature selection methods (i.e., fuzzy rough set based feature selection methods and fuzzy rough neural network based feature selection methods). The basic idea of feature selection is to remove redundant features from a feature set on the premise of maintaining the classification ability of the set unchanged (Guo, Yi, Wang, Ye, & Zhao, 2014; Wang & Zong, 2018; Yi, Song, Guo, & Wang, 2017; Zhu, Zhang, Wang, Zheng, & Zhu, 2018). Thus, by utilizing the fuzzy rough set theory to remove redundant or irrelevant features, the dimensionality of features can be reduced, while the classification ability of the feature set can be maintained.

3 | FUZZY ROUGH SETS FOR FEATURE SELECTION

This section reviews the methods that utilized fuzzy rough sets for feature selection, and then discusses the advantages and challenges of these methods.

3.1 | Fuzzy rough set based feature selection methods

Fuzzy rough set based feature selection methods are based on the notion of the fuzzy lower approximation, so that real-valued features contained in a set can be reduced (Shen & Jensen, 2004). This process is the same as the crisp set based methods when processing nominal well-defined features.

Feature selection via fuzzy rough set models was first proposed by Kuncheva (1992), which was also used for fuzzy pattern recognition. After that, Shen and Jensen developed a series of fuzzy rough set based methods (Jensen & Shen, 2004a, 2004b, 2007; Shen & Jensen, 2004) for feature selection, which can be considered as pioneering work on fuzzy rough set based feature selection. Many studies after them were aimed at improving their work. They extended the concept of dependency functions in a crisp rough set into a fuzzy case. By using the developed dependency function, they developed a fuzzy rough set based method named QuickReduct for feature selection.

A general process of the fuzzy rough QuickReduct algorithm is presented in Algorithm 1. As shown in Algorithm 1, for a feature set $\mathbb{C}$, a general monotone measure M, which satisfied $M(\mathbb{C}) = 1$, was utilized to determine that a reduced feature set $\mathbb{C}'$ can have the same discernibility as its original feature set $\mathbb{C}$.

The QuickReduct algorithm calculated a minimal reduct without exhaustively generating all possible subsets. However, although the experimental results of QuickReduct performed well, this feature selection method lacked the mathematical foundation and theoretical analysis (Zhao et al., 2009). Thus, Jensen and Shen (2007, 2008) then proposed a fuzzy discernibility matrix based on an improved dependency function for fuzzy rough set based feature selection.

Moreover, Hu et al. (2007) proved that the computational complexity of the QuickReduct algorithm increased exponentially with the number of input variables and was multiplied by the size of patterns. Bhatt and Gopal (2005a, 2005b) also indicated that when applying the QuickReduct algorithm to some real-valued datasets, the algorithm was not convergent due to the drawback of stopping criteria. Thus, they defined new fuzzy rough set models for feature selection. Compared to the model proposed by Jensen and Shen (2004b), Bhatt and Gopal (2005a, 2005b) improved time and computational efficiency by processing on a compact domain.

Inspired by Shannon's information entropy, some research has extended the reduction search methods (Ślezak, 2002; Wang, Zhao, An, & Wu, 2005) in a crisp rough set into a fuzzy rough set. Parthaláin, Jensen, and Shen (2006) first proposed a fuzzy rough feature selection method by using fuzzy entropy. Hu and Yu (2004) extended the Shannon entropy to fuzzy rough sets to compute the information of fuzzy indiscernibility relation, and the proposed entropy was applied to redefine independence of an attribute set, reduct, relative reduct in a fuzzy rough set model. Based on this, Hu, Yu, Xie, and Liu (2006); Hu, Yu, and Xie (2006) proposed a theory related to fuzzy probabilistic approximation spaces, and an information entropy based attribute reduction method to measure attribute significance. However, the notions of attribute reduction in (Hu, Yu, & Xie, 2006) were not according to its fuzzy rough set based knowledge representation. In other words, the two parts of this method, that is, the attribute reduction part and the knowledge representation part, were independent of each other.

Later, a fuzzy entropy based heuristic method was developed by Varma, Kumari, and Kumar (2016) for ant colony optimization. For real-time intrusion detection, the proposed method could search a global optimal minimum set of network traffic features. Compared to other rough set based methods, the proposed method did not need an additional phase of data discretization as a preprocessing stage of real-valued data.

Algorithm 1 Fuzzy-rough QUICKREDUCT 1

Input: Conditional attribute set £
Output: Selected attribute set £'
(i): £' ← {};
(ii): **do**
(iii): $T \leftarrow$ £';
(iv): $best \leftarrow -1$;
(v): **for all** $x \in$ (£− £') **do**
(vi): **if** M(£' ∪ $\{x\}$) > $best$ **then**
(vii): $T \leftarrow$ £' ∪ $\{x\}$;
(viii): $best \leftarrow M$(£' ∪ $\{x\}$);
(ix): £' ← T;
(x): **until** M(£') == $best$;
(xi): **return** £'.

Jensen and Shen (2008) also proposed a fuzzy extension of the crisp discernibility matrix for feature subset selection. However, such heuristic algorithms may result in an over-reduct or under-reduct instead of a proper reduct because of their stop criteria (Chen, Hu, et al., 2011; Yang, Chen, Wang, & Wang, 2017). To find a proper reduct, the formal concepts of attributes reduction with fuzzy rough sets was introduced by Tsang et al. (2008). They also analyzed the mathematical structure of attribute reduction and developed a fuzzy rough set based attribute selection algorithm utilizing a discernibility matrix to compute all the attributes reductions.

The fuzzy rough set based feature selection methods discussed above are limited to the batch calculation. In other words, these methods process all elements in a data set at once. Such methods are usually costly and cannot be used to process big data sets (Wang et al., 2016). Therefore, incremental methods were introduced to fuzzy rough set models. These incremental methods can update knowledge (such as decision rules and approximations used for feature selection) in dynamic data sets, whose elements, features, or feature values will change over time. For example, Yang, Chen, Wang, and Wang (2017) developed a fuzzy rough set based feature selection method by investigating an incremental perspective. In their method, features can be added or deleted according to the updated relative discernibility relations between each feature and the feature set. Yang, Chen, Wang, Tsang, and Zhang (2017) also explored the incremental feature selection with fuzzy rough sets. Given a sample with multiple features, the proposed methods updated the relative discernibility relation for each feature. In other words, their methods could.

Furthermore, Liu, Cao, and Zheng (2011) proposed a concept of fuzzy similarity measure to evaluate feature dependence. The results showed that data with smaller variance had a higher measure of fuzzy similarity. Chen, Zhang, et al. (2011) proposed an attribute reduction method that used minimal elements in a discernibility matrix to compute reducts and enhanced the computational efficiency of the discernibility matrix. The relative discernibility relations of conditional attributes were defined by them, which can be used to characterize the minimal elements in the fuzzy discernibility matrix. Later, Chen and Yang (2013) proposed a feature selection method that combined fuzzy rough set with classical rough set models for decision systems. Their method could deal with heterogeneous condition features based on their discernible ability. They defined a discernibility relation that characterized the discernible ability of each feature according to the decision labels. Based on the obtained the discernibility relation, a dependence function was proposed to measure the inconsistency between decision labels and heterogeneous condition features.

Further, a fuzzy rough accelerator, named forward approximation, was developed by Qian, Wang, Cheng, Liang, and Dang (2015). The proposed accelerator can select features by combing the dimensionality reduction with the sample reduction together. It also accelerated the heuristic process of fuzzy rough feature selection, which was verified by applying the proposed accelerator to three representative heuristic fuzzy-rough feature selection algorithms.

Moreover, since the labels of multi-label data are independent to each other, the feature selection process of multi-label data usually has a high computation complexity. To deal with this problem, Qu, Rong, Deng, and Yang (2017) proposed a fuzzy-rough feature selection method for multi-label data. Specifically, they developed association rules between labels so that the combination of labels could be collapsed into a set of sub-labels. Then these relabeled data were used for fuzzy rough feature selection. Their method avoided the label overlapping phenomenon and reduced the scale of labels. However, this method did not improve the process of fuzzy rough feature selection for multi-label data based on the proposed labeling method in terms of association rules.

In addition, Dai, Hu, Wu, Qian, and Huang (2017) defined the definition of reduced maximal discernibility pairs, which can directly adopt the perspective of the object pair in fuzzy rough set models. After that, two effective attribute selection algorithms, named Reduced Maximal Discernibility Pair Selection (RMDPS) and Weighted Reduced Maximal Discernibility Pair Selection (WRMDPS), were developed according to the proposed concept. The developed algorithms only focused on partial object pairs rather than the entire object pairs in the discourse to achieve efficient attribute reduction. Selvakumar et al. (2019) proposed a fuzzy rough set based feature selection method for their adaptive intrusion detection system. The proposed feature selection method was able to select a huge number of attack data for effective prediction of attacks in wireless sensor networks. By using such a method, their system enhanced detection accuracy with reduced false positives in wireless sensor networks.

Recently, Chen, Mi, and Lin (2020) proposed a fuzzy rough feature selection method based on graph theory. They considered fuzzy rough feature selection as a process of searching the minimal transversal of a derivative hypergraph. Thus, the proposed method could avoid generating hypergraph for feature selection. This greatly reduced the time complexity of finding reducts, especially when dealing with large-scale datasets.

3.2 | Summary

The goals of existing feature selection methods based on fuzzy rough set models (Lin, Li, Wang, & Chen, 2018) can be concentrated on two aspects: (i) optimize fuzzy similarity relation (Hu, Yu, Pedrycz, & Chen, 2010; Maji & Garai, 2012; Wang, Qi, et al., 2016; Zeng et al., 2015; Zhang, Mei, Chen, & Li, 2016) and (ii) construct robust distance (Ziarko, 1993; Zhao et al., 2009; Hu, Zhang, et al., 2011; Salido and Murakami, 2003; Aggarwal, 2015; An, Hu, Pedrycz, Zhu, & Tsang, 2015; Gu, Li, & Jia, 2015; Maji, 2010). Three types of methods (i.e., discernibility matrix based methods, dependency degree based methods, and fuzzy entropy based methods) were proposed to achieve these two goals.

Discernibility matrix based methods. The discernibility matrix was originally developed in the rough set theory, which determines the distinguishable attributes/features for each pair of instances. In other words, the attributes/features for which the two instances take on different values. Since feature selection aims to obtain an attribute subset, redundant attributes can be removed from the original attribute set without reducing the discerning capability, so this discernibility matrix can be used to achieve this goal. In fuzzy rough set based feature selection methods (Chen et al., 2020; Chen & Yang, 2013; Chen, Zhang, et al., 2011; Dai et al., 2017; Jensen & Shen, 2007; Jensen & Shen, 2008; Qu et al., 2017; Selvakumar et al., 2019; Tsang et al., 2008; Varma et al., 2016; Yang, Chen, Wang, Tsang, & Zhang, 2017; Yang, Chen, Wang, & Wang, 2017), this matrix is usually customized depending on the selected fuzzy rough set model.

Dependency degree based methods. The dependency degree is a measure which represents how well a set of attributes can discern between elements. Some methods utilized a dependency degree as a criterion for feature selection to measuring how a feature set discern between elements. These methods utilized fuzzy dependency functions (Hu, Yu, & Xie, 2006; Jensen & Shen, 2004b; Jensen & Shen, 2007) or fuzzy positive regions (Chen, Hu, et al., 2011; Hu, Zhang, et al., 2011; Hu, Zhang, Chen, Pedrycz, & Yu, 2010; Hu, Zhang, Zhou, & Pedrycz, 2017; Jensen & Shen, 2002; Tsang et al., 2008; Wang, Shao, He, Qian, & Qi, 2016; Yao et al., 2014; Zhao et al., 2009) to define the dependency degree.

Fuzzy entropy based methods. The information entropy based information gain was widely used as a measure in feature selection, in which the higher the information gain, the more important the feature is. The fuzzy information entropy can be utilized for selecting a feature subset from a big data set since the uncertainty of a fuzzy binary relation can be characterized by the fuzzy information entropy (Hu et al., 2012; Hu, Yu, & Xie, 2006). Thus, some research (Hu, Yu, Xie, & Liu, 2006; Parthaláin et al., 2006; Zhang et al., 2016) introduced fuzzy entropy as a criterion for fuzzy rough set based feature selection methods.

In fact, when the fuzzy lower approximation in a fuzzy dependency function is the same as the fuzzy lower approximation in a fuzzy positive region, the reduct based on the fuzzy dependency function is equal to the reduct preserved by the positive region (Zhang, Mei, Chen, & Yang, 2018). For example, in (Zhang et al., 2016), the fuzzy positive region-preserved reduct and the fuzzy dependency function based reduct can be equivalently considered by the information entropy. In addition, some research also proposed algorithms based on mutual information (Lotfabadi, Shiratuddin, & Wong, 2013; Lotfabadi, Shiratuddin, & Wong, 2015) or fuzzy boundary region (Parthalain, Shen, & Jensen, 2009; Tsang et al., 2008) to compute the feature reducts for fuzzy rough set based feature selection.

4 | FUZZY ROUGH NEURAL NETWORKS FOR FEATURE SELECTION

The neural network is an important technique in machine learning, which usually contains an input layer, an output layer, and several hidden layers (Ma et al., 2019). Each layer in a neural network consists of several processing nodes (i.e., neurons), and the neurons in different layers are connected to each other by using weighted edges. Each neuron combines its inputs with their corresponding weights and transforms these values utilizing an activation function (e.g., sigmoid function) to obtain an output. In the training phase, by processing one training instance at a time and minimizing the prediction errors generated on the training instance, the weights in the neural network can be adaptively modified. Thus, the final output of the neural network can be considered as a complex combination of the initial inputs.

The neural network adapts to process noise data since it has strong modeling ability and outstanding fault-tolerance performance. Thus, neural network based feature selection methods (Basak & Mitra, 1999; Pal, De, & Basak, 2000; Setiono & Liu, 1997; Verikas & Bacauskiene, 2002) were well studied and applied. However, fuzzy rough neural networks for feature selection were not well investigated. In this section, we review the fuzzy rough neural network based feature selection methods over the past few years and discuss the advantages and disadvantages of these methods.

4.1 | Fuzzy rough neural network based feature selection methods

The fuzzy rough set theory has been integrated into the neural network design for many years. However, there is no accurate definition or unified structure has been proposed for fuzzy rough neural networks. Current fuzzy rough neural networks can be divided into two categories: One is to introduce rough neurons into fuzzy neural networks (Lingras, 1998; Lingras, 2001; Wen, Dan, & Lan, 2001) and the other one is based on fuzzy rough membership functions. The latter has been widely used for feature selection (Bezdek, Ehrlich, & Full, 1984; Ganivada, Dutta, et al., 2011; Sarkar & Yegnanarayana, 1998; Mao, Caiping, & Yaonan, 2010; Zhang & Wang, 2006, 2007).

The concept of rough neurons was proposed by Lingras (1998), who incorporated roughness into the design of neurons to increase the prediction accuracy. Rough neurons provide an ability to use rough patterns, in which each value is a pair of upper and lower bounds. The upper and lower bounds of the input and output values are stored by each rough neuron. After that, Lingras (2001) connected the rough neurons and fuzzy neurons in series and achieved good performance. In other words, it formalizes such a serial combination of neo-fuzzy neurons followed by rough neurons as fuzzy-rough neural networks. In addition, Wen et al. (2001) utilized rough neurons to form a sub-network of a fuzzy neural network to establish a rough fuzzy neural network. The input of the proposed sub-network was fuzzified, and then the output of the proposed sub-network was defuzzified to an accurate output.

Further, a fuzzy rough neural network for vowel classification was proposed by Sarkar and Yegnanarayana (1998) based on fuzzy rough membership functions. The proposed neural network was a three-layer feedforward network, containing an input layer, a hidden layer, and an output layer. The proposed neural network can deal with the uncertainty problem in fuzzy clustering, in which each data sample can belong to more than one cluster. However, the physical meaning of this network was not clear, and the network had limitations in adapting new data and classifying (i.e., result in low classification accuracy; Mao et al., 2010). Therefore, a four-layer backpropagation neural network was developed by Zhang and Wang (2006, 2007). The physical meaning of the proposed neural network was more explicit than the network in (Manish & Yegnanarayana, 1998) and it was suitable for feature selection.

The neural network by Zhang and Wang (2006) contains four layers: an input layer, a cluster layer, a membership layer, and an output layer. All layers in the network were fully connected. Specifically, the number of neurons in the input layer was equal to the dimensions of input features; the number of neurons in the cluster layer was equal to the number of clusters clustered by an unsupervised Fuzzy C-Means algorithm (Bezdek et al., 1984); the number of neurons in the membership layer and the output layer were equal to the number of output classes.

The mapping relationships between the inputs and outputs of each layer can be described as follows:

Input layer. Raw data is imported to the input layer without any transformation. The value of the i^{th} feature is fed to the i^{th} input neuron as the input, for each instance. The input and output neurons in the input layer can be represented as:

$$I_i^{(1)} = x_i, i = 1, 2, ..., M, \tag{8}$$

$$O_i^{(1)} = I_i^{(1)}, \tag{9}$$

where x_i is the i^{th} feature of the input $x = (x_1, x_2, ..., x_M)^T$ and M is the number of features.

Cluster layer. The fuzzy memberships of each input to clusters are calculated by the cluster layer using a Gaussian function. w_{ij}^{12} is the connection weight coming from the i^{th} neuron in the input layer to the j^{th} neuron in the cluster layer. The output of this layer represents the extent that instance x belongs to a fuzzy set A_j. The input and output neurons can be represented as:

$$I_j^{(2)} = (x_1, x_2, ..., x_M), j = 1, 2, ..., H, \tag{10}$$

$$O_j^{(2)} = \exp\left(-\left(I_j^{(2)} - c_j\right)^T \left(I_j^{(2)} - c_j\right) / 2\sigma_j^2\right), \tag{11}$$

where H is the number of clusters, and c_j and σ_j are the center and variance of the cluster, respectively. If A_j represents the j^{th} cluster, then $o_j^{(2)}$ can be expressed as $\mu_{Aj}(x)$.

Membership layer. The value of fuzzy rough membership degree is represented by the input of this layer, and a *tansig* function is used as an actuation function. The input and output neurons can be represented as:

$$I_k^{(3)} = \sum_{j=1}^{H} O_j^{(2)} \cdot w_{jk}^{23}, k = 1, 2, ..., N, \tag{12}$$

$$O_k^{(3)} = \left(2/\left(1 + \exp\left(-2I_k^{(3)}\right)\right)\right) - 1, \tag{13}$$

where N is the number of output classes, and $w_{jk}{}^{23}$ is the connection weight from the j^{th} neuron in the cluster layer to the k^{th} neuron in the membership layer. $w_{jk}{}^{23}$ can be calculated as:

$$w_{jk}^{23} = \frac{1}{H}\frac{|A_j \cap Class_k|}{|A_j|} = \frac{1}{H}\frac{\sum_{x \in Class_k} \mu_{A_j}(x)}{\sum_k \sum_{x \in Class_k} \mu_{A_j}(x)}, \tag{14}$$

where $class_k$ is the k^{th} output class.

Output layer. The output layer computes a membership degree of an input instance to each output class so that the instance can be classified into the class with the maximum membership value. The input and output neurons can be represented as:

$$I_l^{(4)} = \sum\nolimits_{k=1}^{N} O_k^{(3)} \cdot w_{kl}^{34}, l = 1, 2, ..., N, \tag{15}$$

$$O_l^{(4)} = I_l^{(4)}. \tag{16}$$

If S is the target output, then $W^{34} O^{(3)} = S$. The connection weight matrix W^{34} can be calculated as:

$$W^{34} = \left(O^{(3)}\right)^{-1} S. \tag{17}$$

The proposed neural network performed feature selection using a backward search. Initially, the network was trained by all features in a feature set and its corresponding prediction error was calculated. After that, in each iteration, the network discarded the feature with the smallest error and was retrained without the discarded feature. Compared with the resulted error in the previous iteration, if the resulted error increased too much, then the chosen feature can be included again, otherwise, the chosen feature can be permanently removed. Continue this process until no more features can be removed.

Similar to the neural network proposed by Liu et al. (2011), Zhao and Zhang (2011) proposed a fuzzy rough neural network for feature selection based on their defined fuzzy rough membership functions. Specifically, to search the optimal feature subset, the input nodes in the proposed neural network were pruned based on the descent of classification accuracy.

Moreover, based on the multilayer perceptron, a Fuzzy Rough Granular Neural Network (FRGNN) was developed by Ganivada, Dutta, et al. (2011). The proposed network integrated the fuzzy rough set theory with fuzzy neural networks, in which the input vector was described by fuzzy granules, the initial connection weights were determined by fuzzy rough set-theoretic concepts, and the target vector was defined by fuzzy class membership values and zeros. This network was first applied for classification and clustering tasks (Arunkumar & Ramakrishnan, 2018; Ganivada & Pal (2011); Ganivada, Ray, et al., 2011; Ganivada, Ray, & Pal, 2012; Qian, Liang, & Wei, 2013), in which the fuzzy rules, which was used to represent the information granules that was generated by fuzzy rough sets, were incorporated into artificial neural networks.

After that, Ganivada, Ray, and Pal (2013) defined the notions of the upper and lower approximations of a fuzzy rough set model and applied a three-layered FRGNN for feature selection based on their fuzzy rough set models. The proposed FRGNN performed well in solving unsupervised feature selection problems. In other words, the developed method can be recommended when the data has no class information, that is, class labels, the number of features is large, and the feature space is uncertain. However, it required a large number of computational times compared with other related methods for feature selection.

In addition, Gafar (2016) proposed a feature selection approach based on fuzzy rough neural networks to solve diagnostic breast cancer problems, which could remove unnecessary features from medical data. The lower and upper approximations of the input features were weighted by input synapses learnt through the training phase.

4.2 | Summary

This section reviews two types of fuzzy rough neural network based feature selection methods. One is to introduce rough neurons into fuzzy neural networks, and the other is to introduce fuzzy rough membership functions into neural networks. This article mainly focuses on the latter, which has a proper theoretical foundation and applications. Such fuzzy rough neural networks generally consist of four layers: an input layer (input data), a fuzzy layer (discrete input data through fuzzy neurons), a rule layer (each neuron represents a rule), and an output layer. Since most of the existing fuzzy rough neural networks are proposed for classification, the classification accuracy is usually used to measure the performance of the proposed fuzzy rough neural networks.

5 | CONCLUSION

This section summarizes the contributions of this article and provides the challenges of current fuzzy rough feature selection methods to be addressed in the future.

5.1 | Summary

Feature selection is a key task in machine learning, which aims to select useful and relevant features on the premise of maintaining the unchanged classification capability. The fuzzy rough set theory provides the flexibility in dealing with two distinct types of uncertainty (i.e., vagueness and incompleteness) presented in information. It can incorporate the fuzzy set theory to consider vagueness within a rough set model to process incomplete information. Thus, fuzzy rough feature selection methods are developed for data preprocessing. It can select a feature subset from the original feature set based on fuzzy rough rules for special machine learning tasks such as classification and clustering. This article critically reviews the research progresses in the fuzzy rough set theory, fuzzy rough set based feature selection methods, and fuzzy rough neural network based feature selection methods. We review the publications related to the fuzzy rough theory and its applications in the past decades.

Fuzzy rough set based methods select features from a feature set according to the notion of fuzzy lower approximation. The existing fuzzy rough set based feature selection methods select effective features by optimizing fuzzy similarity relation and constructing robust distance. These methods can be divided into three categories (i.e., discernibility matrix based methods, dependency degree based methods, and fuzzy entropy based methods). Meanwhile, fuzzy rough neural networks select features from a feature set through two ways (i.e., introducing rough neurons into fuzzy neural networks and introducing fuzzy rough membership functions into neural networks). Both two types of feature selection methods were thoroughly explored in recent years.

5.2 | Challenges

Although various methods for fuzzy rough feature selection have been proposed, the research on the fuzzy rough sets and the fuzzy rough neural networks is still immature and needs further study to meet the real-world requirements. The complexity of real-world problems demands further development of the fuzzy rough set theory and its related techniques for feature selection. This article identifies two challenges in fuzzy rough feature selection to be addressed. One of the challenges is how to deal with complex data, such as dynamic data and multi-label data, using fuzzy rough feature selection methods. Another challenge is how to combine other techniques, such as genetic algorithms, clustering algorithms, and ant colony optimization, with the fuzzy rough set theory to meet the real-world requirements and achieve the desired performance.

CONFLICT OF INTEREST

The authors have declared no conflicts of interest for this article.

AUTHOR CONTRIBUTIONS

Wanting Ji: Conceptualization; investigation; methodology; resources; validation; visualization; writing-original draft; writing-review and editing. **Yan Pang:** Data curation; formal analysis; investigation; validation; writing-review and editing. **Xiaoyun Jia:** Writing-review and editing. **Zhongwei Wang:** Data curation; formal analysis; resources; visualization; writing-review and editing. **Feng Hou:** Writing-review and editing. **Baoyan Song:** Writing-review and editing. **Mingzhe Liu:** Writing-review and editing. **Ruili Wang:** Conceptualization; investigation; project administration; resources; supervision; validation; visualization; writing-original draft; writing-review and editing.

ORCID

Ruili Wang https://orcid.org/0000-0003-2899-9816

RELATED WIREs ARTICLES

Rough clustering

REFERENCES

Aggarwal, M. (2015). Probabilistic variable precision fuzzy rough sets. *IEEE Transactions on Fuzzy Systems*, *24*(1), 29–39.
Almuallim, H., & Dietterich, T. G. (1991). *Learning with Many Irrelevant Features*, pp. 547–552. Anaheim, California: AAAI.
An, S., Hu, Q., Pedrycz, W., Zhu, P., & Tsang, E. C. (2015). Data-distribution-aware fuzzy rough set model and its application to robust classification. *IEEE Transactions on Cybernetics*, *46*(12), 3073–3085.
Arunkumar, C., & Ramakrishnan, S. (2018). Attribute selection using fuzzy rough set based customized similarity measure for lung cancer microarray gene expression data. *Future Computing and Informatics Journal*, *3*(1), 131–142.
Basak, J., & Mitra, S. (1999). Feature selection using radial basis function networks. *Neural Computing & Applications*, *8*(4), 297–302.
Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, *10*(2–3), 191–203.
Bhatt, R. B., & Gopal, M. (2005a). On fuzzy-rough sets approach to feature selection. *Pattern Recognition Letters*, *26*(7), 965–975.
Bhatt, R. B., & Gopal, M. (2005b). On the compact computational domain of fuzzy-rough sets. *Pattern Recognition Letters*, *26*(11), 1632–1640.
Bhatt, R. B., & Gopal, M. (2008). FRCT: Fuzzy-rough classification trees. *Pattern Analysis and Applications*, *11*(1), 73–88.
Boixader, D., Jacas, J., & Recasens, J. (2000). Upper and lower approximations of fuzzy sets. *International Journal of General System*, *29*(4), 555–568.
Chen, D., Hu, Q., & Yang, Y. (2011). Parameterized attribute reduction with Gaussian kernel based fuzzy rough sets. *Information Sciences*, *181*(23), 5169–5179.
Chen, D., Kwong, S., He, Q., & Wang, H. (2012). Geometrical interpretation and applications of membership functions with fuzzy rough sets. *Fuzzy Sets and Systems*, *193*, 122–135.
Chen, D., & Yang, Y. (2013). Attribute reduction for heterogeneous data based on the combination of classical and fuzzy rough set models. *IEEE Transactions on Fuzzy Systems*, *22*(5), 1325–1334.
Chen, D., Zhang, L., Zhao, S., Hu, Q., & Zhu, P. (2011). A novel algorithm for finding reducts with fuzzy rough sets. *IEEE Transactions on Fuzzy Systems*, *20*(2), 385–389.
Chen, J., Mi, J., & Lin, Y. (2020). A graph approach for fuzzy-rough feature selection. *Fuzzy Sets and Systems*, *391*, 96–116.
Clausen, J. (1999). *Branch and bound algorithms-principles and examples*, pp. 1–30. Denmark: Department of Computer Science, University of Copenhagen.
Cornelis, C., De Cock, M., & Radzikowska, A. M. (2007). *Vaguely Quantified Rough Sets*. International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-soft Computing, pp. 87–94. Berlin, Heidelberg: Springer.
Cornelis, C., Verbiest, N., & Jensen, R. (2010). *Ordered Weighted Average Based Fuzzy Rough Sets*. International Conference on Rough Sets and Knowledge Technology, pp. 78–85. Berlin, Heidelberg: Springer.
Dai, J., Hu, H., Wu, W. Z., Qian, Y., & Huang, D. (2017). Maximal-discernibility-pair-based approach to attribute reduction in fuzzy rough sets. *IEEE Transactions on Fuzzy Systems*, *26*(4), 2174–2187.
Doshi, M. (2014). Correlation based feature selection (CFS) technique to predict student performance. *International Journal of Computer Networks & Communications*, *6*(3), 197–206.
Dubois, D., & Prade, H. (1990). Rough fuzzy sets and fuzzy rough sets. *International Journal of General System*, *17*(2–3), 191–209.
Fernández-Riverola, F., Diaz, F., & Corchado, J. M. (2006). Reducing the memory size of a fuzzy case-based reasoning system applying rough set techniques. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *37*(1), 138–146.
Gafar, M. G. (2016). *Medical Diagnostic System Basing Fuzzy Rough Neural-computing for Breast Cancer*. International Conference on Advanced Intelligent Systems and Informatics, pp. 467–479. Cham: Springer.
Ganivada, A., Dutta, S., & Pal, S. K. (2011). Fuzzy rough granular neural networks, fuzzy granules, and classification. *Theoretical Computer Science*, *412*(42), 5834–5853.

Ganivada, A., & Pal, S. K. (2011). A novel fuzzy rough granular neural network for classification. *International Journal of Computational Intelligence Systems*, *4*(5), 1042–1051.

Ganivada, A., Ray, S. S., & Pal, S. K. (2011). *Fuzzy Rough Granular Self Organizing Map*. International Conference on Rough Sets and Knowledge Technology, pp. 659–668. Berlin, Heidelberg: Springer.

Ganivada, A., Ray, S. S., & Pal, S. K. (2012). Fuzzy rough granular self-organizing map and fuzzy rough entropy. *Theoretical Computer Science*, *466*, 37–63.

Ganivada, A., Ray, S. S., & Pal, S. K. (2013). Fuzzy rough sets, and a granular neural network for unsupervised feature selection. *Neural Networks*, *48*, 91–108.

Gu, X., Li, Y., & Jia, J. (2015). Feature selection for transient stability assessment based on kernelized fuzzy rough sets and memetic algorithm. *International Journal of Electrical Power & Energy Systems*, *64*, 664–670.

Guo, J., Yi, P., Wang, R., Ye, Q., & Zhao, C. (2014). Feature selection for least squares projection twin support vector machine. *Neurocomputing*, *144*, 174–183.

Hassanien, A. (2007). Fuzzy rough sets hybrid scheme for breast cancer detection. *Image and Vision Computing*, *25*(2), 172–183.

Hong, T. P., Wang, T. T., Wang, S. L., & Chien, B. C. (2000). Learning a coverage set of maximally general fuzzy rules by rough sets. *Expert Systems with Applications*, *19*(2), 97–103.

Hu, Q., An, S., & Yu, D. (2010). Soft fuzzy rough sets for robust feature evaluation and selection. *Information Sciences*, *180*(22), 4384–4400.

Hu, Q., An, S., Yu, X., & Yu, D. (2011). Robust fuzzy rough classifiers. *Fuzzy Sets and Systems*, *183*(1), 26–43.

Hu, Q., Xie, Z., & Yu, D. (2007). Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation. *Pattern Recognition*, *40*(12), 3509–3521.

Hu, Q., & Yu, D. (2004). Entropies of fuzzy indiscernibility relation and its operations. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *12*(05), 575–589.

Hu, Q., Yu, D., & Guo, M. (2010). Fuzzy preference based rough sets. *Information Sciences*, *180*(10), 2003–2022.

Hu, Q., Yu, D., Pedrycz, W., & Chen, D. (2010). Kernelized fuzzy rough sets and their applications. *IEEE Transactions on Knowledge and Data Engineering*, *23*(11), 1649–1667.

Hu, Q., Yu, D., & Xie, Z. (2006). Information-preserving hybrid data reduction based on fuzzy-rough techniques. *Pattern Recognition Letters*, *27*(5), 414–423.

Hu, Q., Yu, D., Xie, Z., & Liu, J. (2006). Fuzzy probabilistic approximation spaces and their information measures. *IEEE Transactions on Fuzzy Systems*, *14*(2), 191–201.

Hu, Q., Zhang, L., An, S., Zhang, D., & Yu, D. (2011). On robust fuzzy rough set models. *IEEE Transactions on Fuzzy Systems*, *20*(4), 636–651.

Hu, Q., Zhang, L., Chen, D., Pedrycz, W., & Yu, D. (2010). Gaussian kernel based fuzzy rough sets: Model, uncertainty measures and applications. *International Journal of Approximate Reasoning*, *51*(4), 453–471.

Hu, Q., Zhang, L., Zhou, Y., & Pedrycz, W. (2017). Large-scale multimodality attribute reduction with multi-kernel fuzzy rough sets. *IEEE Transactions on Fuzzy Systems*, *26*(1), 226–238.

Jensen, R., & Shen, Q. (2002). *Fuzzy-Rough Sets for Descriptive Dimensionality Reduction*. IEEE World Congress on Computational Intelligence. IEEE International Conference on Fuzzy Systems. FUZZ-IEEE'02. Proceedings (Cat. No. 02CH37291), pp. 29–34. Honolulu, HI: IEEE.

Jensen, R., & Shen, Q. (2004a). Semantics-preserving dimensionality reduction: Rough and fuzzy-rough-based approaches. *IEEE Transactions on Knowledge and Data Engineering*, *16*(12), 1457–1471.

Jensen, R., & Shen, Q. (2004b). Fuzzy–rough attribute reduction with application to web categorization. *Fuzzy Sets and Systems*, *141*(3), 469–485.

Jensen, R., & Shen, Q. (2007). Fuzzy-rough sets assisted attribute selection. *IEEE Transactions on Fuzzy Systems*, *15*(1), 73–89.

Jensen, R., & Shen, Q. (2008). New approaches to fuzzy-rough feature selection. *IEEE Transactions on Fuzzy Systems*, *17*(4), 824–838.

Jović, A., Brkić, K., & Bogunović, N. (2015). *A Review of Feature Selection Methods With Applications*. *MIPRO*, pp. 1200–1205. Opatija, Croatia: IEEE.

Kira, K., & Rendell, L. A. (1992). *A Practical Approach to Feature Selection*. Machine Learning Proceedings, pp. 249–256. Morgan Kaufmann.

Kuncheva, L. I. (1992). Fuzzy rough sets: Application to feature selection. *Fuzzy Sets and Systems*, *51*(2), 147–153.

Leung, Y., Fischer, M. M., Wu, W. Z., & Mi, J. S. (2008). A rough set approach for the discovery of classification rules in interval-valued information systems. *International Journal of Approximate Reasoning*, *47*(2), 233–246.

Li, Y., Shiu, S. C., & Pal, S. K. (2006). Combining feature reduction and case selection in building CBR classifiers. *IEEE Transactions on Knowledge and Data Engineering*, *18*(3), 415–429.

Lin, Y., Li, Y., Wang, C., & Chen, J. (2018). Attribute reduction for multi-label learning with fuzzy rough set. *Knowledge-Based Systems*, *152*, 51–61.

Lingras, P. (1998). Comparison of neofuzzy and rough neural networks. *Information Sciences*, *110*(3–4), 207–215.

Lingras, P. (2001). Fuzzy-rough and rough-fuzzy serial combinations in neurocomputing. *Neurocomputing*, *36*(1–4), 29–44.

Liu, C., Cao, Q., & Zheng, G. (2011). *New Definition of Dependency in Fuzzy Rough Feature Selection*. 2011 International Conference on System Science, Engineering Design and Manufacturing Informatization, pp. 271–274. IEEE.

Liu, H., & Setiono, R. (1996). *A Probabilistic Approach to Feature Selection-A Filter Solution*. pp. 319–327. Bari, Italy: ICML.

Lotfabadi, M. S., Shiratuddin, M. F., & Wong, K. W. (2013). *Feature Decreasing Methods Using Fuzzy Rough Set Based on Mutual Information*. ICIEA, pp. 1141–1146. Melbourne, VIC, Australia: IEEE.

Lotfabadi, M. S., Shiratuddin, M. F., & Wong, K. W. (2015). *Utilising Fuzzy Rough Set Based on Mutual Information Decreasing Method for Feature Reduction in an Image Retrieval System*. Innovations and Advances in Computing, Informatics, Systems Sciences, Networking and Engineering. pp. 177–184. Cham: Springer.

Lu, J., & Wang, R. (2009). An enhanced fuzzy linear regression model with more flexible spreads. *Fuzzy Sets and Systems*, *160*(17), 2505–2523.

Ma, J., Wang, R., Ji, W., Zhao, J., Zong, M., & Gilman, A. (2018). Robust multi-view continuous subspace clustering. *Pattern Recognition Letters*. https://doi.org/10.1016/j.patrec.2018.12.004.

Ma, J., Wang, R., Ji, W., Zheng, H., Zhu, E., & Yin, J. (2019). Relational recurrent neural networks for polyphonic sound event detection. *Multimedia Tools and Applications*, *78*(20), 29509–29527.

Maji, P. (2010). Fuzzy–rough supervised attribute clustering algorithm and classification of microarray data. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, *41*(1), 222–233.

Maji, P., & Garai, P. (2012). Fuzzy-rough simultaneous attribute selection and feature extraction algorithm. *IEEE Transactions on Cybernetics*, *43*(4), 1166–1177.

Mao, J., Caiping, L., & Yaonan, W. (2010). *Remote Sensing Images Classification Using Fuzzy-Rough Neural Network*. BIC-TA, pp. 761–765. Changsha, China: IEEE.

Mi, J. S., & Zhang, W. X. (2004). An axiomatic characterization of a fuzzy generalization of rough sets. *Information Sciences*, *160*(1–4), 235–249.

Mieszkowicz-Rolka, A., & Rolka, L. (2004). *Variable Precision Fuzzy Rough Sets*. Transactions on Rough Sets I, LNCS, pp. 144–160. Berlin, Heidelberg: Springer.

Mieszkowicz-Rolka, A., & Rolka, L. (2008). Fuzzy rough approximations of process data. *International Journal of Approximate Reasoning*, *49*(2), 301–315.

Morsi, N. N., & Yakout, M. M. (1998). Axiomatics for fuzzy rough sets. *Fuzzy Sets and Systems*, *100*(1–3), 327–342.

Pal, S. K., De, R. K., & Basak, J. (2000). Unsupervised feature evaluation: A neuro-fuzzy approach. *IEEE Transactions on Neural Networks*, *11*(2), 366–376.

Parthalain, N., Shen, Q., & Jensen, R. (2009). A distance measure approach to exploring the rough set boundary region for attribute reduction. *IEEE Transactions on Knowledge and Data Engineering*, *22*(3), 305–317.

Parthaláin, N. M., Jensen, R., & Shen, Q. (2006). *Fuzzy Entropy-assisted Fuzzy-rough Feature Selection*. 2006 IEEE International Conference on Fuzzy Systems, pp. 423–430. Vancouver, BC, Canada: IEEE.

Pawlak, Z. (1982). Rough sets. *International Journal of Computer Information Sciences*, *11*(5), 341–356.

Pedrycz, W. (1994). Why triangular membership functions? *Fuzzy Sets and Systems*, *64*(1), 21–30.

Qian, Y., Liang, J., & Wei, W. (2013). Consistency-preserving attribute reduction in fuzzy rough set framework. *International Journal of Machine Learning and Cybernetics*, *4*(4), 287–299.

Qian, Y., Wang, Q., Cheng, H., Liang, J., & Dang, C. (2015). Fuzzy-rough feature selection accelerator. *Fuzzy Sets and Systems*, *258*, 61–78.

Qu, Y., Rong, Y., Deng, A., & Yang, L. (2017). *Associated Multi-label Fuzzy-rough Feature Selection*. 2017 Joint 17th World Congress of International Fuzzy Systems Association and 9th International Conference on Soft Computing and Intelligent Systems, pp. 1–6. Otsu, Japan: IEEE.

Radzikowska, A. M., & Kerre, E. E. (2002). A comparative study of fuzzy rough sets. *Fuzzy Sets and Systems*, *126*(2), 137–155.

Riza, L. S., Janusz, A., Bergmeir, C., Cornelis, C., Herrera, F., Śle, D., & Benítez, J. M. (2014). Implementing algorithms of rough set theory and fuzzy rough set theory in the R package "RoughSets". *Information Sciences*, *287*, 68–89.

Saeys, Y., Inza, I., & Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics*, *23*(19), 2507–2517.

Salido, J. F., & Murakami, S. (2003). Rough set analysis of a general type of fuzzy data using transitive aggregations of fuzzy similarity relations. *Fuzzy Sets and Systems*, *139*(3), 635–660.

Sarkar, M., & Yegnanarayana, B. (1998). *Fuzzy-rough Neural Networks for Vowel Classification*. SMC'98 Conference Proceedings. 1998 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 98CH36218), pp. 4160–4165. San Diego, CA: IEEE.

Selvakumar, K., Karuppiah, M., SaiRamesh, L., Islam, S. H., Hassan, M. M., Fortino, G., & Choo, K. K. R. (2019). Intelligent temporal classification and fuzzy rough set-based feature selection algorithm for intrusion detection system in WSNs. *Information Sciences*, *497*, 77–90.

Setiono, R., & Liu, H. (1997). Neural-network feature selector. *IEEE Transactions on Neural Networks*, *8*(3), 654–662.

Shao, M. W., Liu, M., & Zhang, W. X. (2007). Set approximations in fuzzy formal concept analysis. *Fuzzy Sets and Systems*, *158*(23), 2627–2640.

Sheeja, T. K., & Kuriakose, A. S. (2018). A novel feature selection method using fuzzy rough sets. *Computers in Industry*, *97*, 111–116.

Shen, Q., & Jensen, R. (2004). Selecting informative features with fuzzy-rough sets and its application for complex systems monitoring. *Pattern Recognition*, *37*(7), 1351–1363.

Ślezak, D. (2002). Approximate entropy reducts. *Fundamenta Informatica*, *53*(3–4), 365–390.

Tsang, E. C., Chen, D., Yeung, D. S., Wang, X. Z., & Lee, J. W. (2008). Attributes reduction using fuzzy rough sets. *IEEE Transactions on Fuzzy Systems*, *16*(5), 1130–1141.

Varma, P. R. K., Kumari, V. V., & Kumar, S. S. (2016). Feature selection using relative fuzzy entropy and ant colony optimization applied to real-time intrusion detection system. *Procedia Computer Science*, *85*, 503–510.

Verikas, A., & Bacauskiene, M. (2002). Feature selection with neural networks. *Pattern Recognition Letters*, *23*(11), 1323–1335.
Vluymans, S., D'eer, L., Saeys, Y., & Cornelis, C. (2015). Applications of fuzzy rough set theory in machine learning: A survey. *Fundamenta Informaticae*, *142*(1–4), 53–86.
Vluymans, S., Fernández, A., Saeys, Y., Cornelis, C., & Herrera, F. (2018). Dynamic affinity-based classification of multi-class imbalanced data with one-versus-one decomposition: A fuzzy rough set approach. *Knowledge and Information Systems*, *56*(1), 55–84.
Wang, C., Qi, Y., Shao, M., Hu, Q., Chen, D., Qian, Y., & Lin, Y. (2016). A fitting model for feature selection with fuzzy rough sets. *IEEE Transactions on Fuzzy Systems*, *25*(4), 741–753.
Wang, C., Shao, M., He, Q., Qian, Y., & Qi, Y. (2016). Feature subset selection based on fuzzy neighborhood rough sets. *Knowledge-Based Systems*, *111*, 173–179.
Wang, G. Y., Zhao, J., An, J., & Wu, Y. (2005). A comparative study of algebra viewpoint and information viewpoint in attribute reduction. *Fundamenta Informaticae*, *68*(3), 289–301.
Wang, R., Ji, W., Liu, M., Wang, X., Weng, J., Deng, S., … Yuan, C. A. (2018). Review on mining data from multiple data sources. *Pattern Recognition Letters*, *109*, 120–128.
Wang, R., Ji, W., & Song, B. (2018). Durable relationship prediction and description using a large dynamic graph. *World Wide Web*, *21*(6), 1575–1600.
Wang, R., & Zong, M. (2018). Joint self-representation and subspace learning for unsupervised feature selection. *World Wide Web*, *21*(6), 1745–1758.
Wang, X., Tsang, E. C., Zhao, S., Chen, D., & Yeung, D. S. (2007). Learning fuzzy rules from fuzzy samples based on rough set technique. *Information Sciences*, *177*(20), 4493–4514.
Wang, Y. F. (2003). Mining stock price using fuzzy rough set system. *Expert Systems with Applications*, *24*(1), 13–23.
Wei, W., Liang, J., & Qian, Y. (2012). A comparative study of rough sets for hybrid data. *Information Sciences*, *190*, 1–16.
Wen, M. Z., Dan, H., & Lan, S. (2001). *Learning Rules Approach to R-FNN*. Proceedings Joint 9th IFSA World Congress and 20th NAFIPS International Conference (Cat. No. 01TH8569), pp. 639–644. Vancouver, BC, Canada, Canada: IEEE.
Wu, W. Z., Mi, J. S., & Zhang, W. X. (2003). Generalized fuzzy rough sets. *Information Sciences*, *151*, 263–282.
Yang, Y., Chen, D., Wang, H., Tsang, E. C., & Zhang, D. (2017). Fuzzy rough set based incremental attribute reduction from dynamic data with sample arriving. *Fuzzy Sets and Systems*, *312*, 66–86.
Yang, Y., Chen, D., Wang, H., & Wang, X. (2017). Incremental perspective for feature selection based on fuzzy rough sets. *IEEE Transactions on Fuzzy Systems*, *26*(3), 1257–1273.
Yao, Y., Mi, J., & Li, Z. (2014). A novel variable precision (θ, σ)-fuzzy rough set model based on fuzzy granules. *Fuzzy Sets and Systems*, *236*, 58–72.
Yeung, D. S., Chen, D., Tsang, E. C., Lee, J. W., & Xizhao, W. (2005). On the generalization of fuzzy rough sets. *IEEE Transactions on Fuzzy Systems*, *13*(3), 343–361.
Yi, P., Song, A., Guo, J., & Wang, R. (2017). Regularization feature selection projection twin support vector machine via exterior penalty. *Neural Computing and Applications*, *28*(1), 683–697.
Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*(3), 338–353.
Zeng, A., Li, T., Liu, D., Zhang, J., & Chen, H. (2015). A fuzzy rough set approach for incremental feature selection on hybrid information systems. *Fuzzy Sets and Systems*, *258*, 39–60.
Zhang, D., & Wang, Y. (2006). Fuzzy-rough neural network and its application to vowel recognition. *Control and Decision*, *21*(2), 221.
Zhang, D., & Wang, Y. (2007). Fuzzy-rough membership function neural network suitable for pattern classification. Journal of Hunan University: Natural. *Sciences*, *34*(5), 39–43.
Zhang, X., Mei, C., Chen, D., & Li, J. (2016). Feature selection in mixed data: A method using a novel fuzzy rough set-based information entropy. *Pattern Recognition*, *56*, 1–15.
Zhang, X., Mei, C., Chen, D., & Yang, Y. (2018). A fuzzy rough set-based feature selection method using representative instances. *Knowledge-Based Systems*, *151*, 216–229.
Zhao, J. Y., & Zhang, Z. L. (2011). *Fuzzy Rough Neural Network and Its Application to Feature Selection*. pp. 684–687. Wuhan, China: IWACI.
Zhao, S., Tsang, E. C., & Chen, D. (2009). The model of fuzzy variable precision rough sets. *IEEE Transactions on Fuzzy Systems*, *17*(2), 451–467.
Zhu, Y., Zhang, X., Wang, R., Zheng, W., & Zhu, Y. (2018). Self-representation and PCA embedding for unsupervised feature selection. *World Wide Web*, *21*(6), 1675–1688.
Ziarko, W. (1993). Variable precision rough set model. *Journal of Computer and System Sciences*, *46*(1), 39–59.

How to cite this article: Ji W, Pang Y, Jia X, et al. Fuzzy rough sets and fuzzy rough neural networks for feature selection: A review. *WIREs Data Mining Knowl Discov*. 2021;11:e1402. https://doi.org/10.1002/widm.1402