Data-Distribution-Aware Fuzzy Rough Set Model and its Application to Robust Classification

Shuang An, Qinghua Hu, *Senior Member, IEEE*, Witold Pedrycz, *Fellow, IEEE*, Pengfei Zhu, and Eric C. C. Tsang

***Abstract*—Fuzzy rough sets (FRSs) are considered to be a powerful model for analyzing uncertainty in data. This model encapsulates two types of uncertainty: 1) fuzziness coming from the vagueness in human concept formation and 2) roughness rooted in the granulation coming with human cognition. The rough set theory has been widely applied to feature selection, attribute reduction, and classification. However, it is reported that the classical FRS model is sensitive to noisy information. To address this problem, several robust models have been developed in recent years. Nevertheless, these models do not consider a statistical distribution of data, which is an important type of uncertainty. Data distribution serves as crucial information for designing an optimal classification or regression model. Thus, we propose a data-distribution-aware FRS model that considers distribution information and incorporates it in computing lower and upper fuzzy approximations. The proposed model considers not only the similarity between samples, but also the probability density of classes. In order to demonstrate the effectiveness of the proposed model, we design a new sample evaluation index for prototype-based classification based on the model, and a prototype selection algorithm is developed using this index. Furthermore, a robust classification algorithm is constructed with prototype covering and nearest neighbor classification. Experimental results confirm the robustness and effectiveness of the proposed model.**

***Index Terms*—Data distribution, fuzzy rough covering, fuzzy rough sets (FRSs), prototype selection, robust classification.**

Manuscript received June 18, 2015; revised October 11, 2015; accepted October 19, 2015. Date of publication November 12, 2015; date of current version November 15, 2016. This work was supported in part by the National Program on Key Basic Research Project under Grant 2013CB329304, in part by the National Natural Science Foundation of China under Grant 61432011, Grant 61222210, Grant 61202259, and Grant 61071179, in part by the Natural Science Foundation of Hebei Province of China under Grant F2013501052, in part by the China Postdoctoral Science Foundation under Grant 2013M530874, and in part by the Macau Science and Technology Development Fund under Grant 100/2013/A3. This paper was recommended by Associate Editor Y. Yang. *Corresponding author: Qinghua Hu.*

S. An is with Northeastern University, Shenyang, China, and also with Tianjin University, Tianjin 150001, China (e-mail: anshuang_001@163.com).

Q. Hu and P. Zhu are with the School of Computer Science and Technology, Tianjin University, Tianjin 150001, China, and also with the Tianjin Key Laboratory of Cognitive Computing and application, Tianjin 150001, China (e-mail: huqinghua@tju.edu.cn).

W. Pedrycz is with the Department of Electrical and Computer Engineering, Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia, and also with the Systems Research Institute, Polish Academy of Sciences, Warsaw 00-901, Poland (e-mail: wpedrycz@ualberta.ca).

E. C. C. Tsang is with the Faculty of Information Technology, Macau University of Science and Technology, Macau 519020, China (e-mail: cctsang@must.edu.mo).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2015.2496425

I. Introduction

ROUGH set theory has attracted considerable attention owing to its ability to characterize uncertainty in imperfect data [1], [2]. The classical model has been extended to fuzzy rough sets (FRSs) to handle uncertainty in existing information [3]. Various applications of FRSs were successfully developed these years, including feature evaluation, attribute reduction [4], [5], prototype selection, classification, and regression tasks [6]–[9].

Unfortunately, it has been reported that rough sets are sensitive to noisy samples; a single noisy sample may exert large impact on the computation of the lower and upper approximations of a set [10]–[13]. Such sensitivity severely limits the practical applications of rough sets. In recent years, researchers try to alleviate the effects of noise and develop robust rough set models. Initially, Yao *et al.* [12] proposed a decision-theoretic rough set model to enhance the robustness of Pawlak's rough sets, and this model was successfully applied to attribute reduction. Later, Slezak and Ziarko [14] introduced variable precision into Pawlak's rough sets by allowing for misclassification aspect, which was considered as noise. Further, a Bayesian rough set model was proposed by defining set approximations on the basis of prior probability [14]. Recently, Qian *et al.* [15] proposed a multigranulation decision-theoretic rough set model that is robust to noise by applying multigranulation theory to rough sets.

Some noise-tolerant models and algorithms have been developed for fuzzy data [16]. Salido and Murakami [17] presented a robust model by applying the concept of variable precision rough sets to FRSs (VPRSs-FRSs). Further, a variable precision FRS (VPFRS) model was proposed as a general robust FRS framework [10]. In this model, the membership grades of a sample to the lower and upper approximations are computed with fuzzy inclusion. Later, Cornelis *et al.* [18] presented the vaguely quantified rough set (VQRS) model and used it for robust feature selection. The ordered weighted average FRS theory was proposed inspired by VQRS [19]. In 2009, fuzzy VPRSs (FVPRSs) were developed to enhance the robustness of FRSs by increasing the lower approximation memberships (LAMs) of boundary points [13]. Cornelis *et al.* [20] and Verbiest *et al.* [21] constructed a model of FRSs based on ordered weighted average operators, and this model was applied to prototype selection. In addition, Hu *et al.* [22] proposed soft FRSs (SFRSs) by achieving a tradeoff between the number of ignored samples and an increase in the LAM. Further, in 2012, Hu *et al.* [23] and An *et al.* [24] proposed

robust statistics and soft minimum enclosing ball (SMEB)-based robust FRS models. In the same year, Ma and Sun [25] presented a probabilistic rough set model over two universes. In 2014, Sun and Ma [26] introduced soft sets into FRSs to present SFRSs for robust decision making. In the same year, Yao *et al.* [27] proposed a novel robust variable precision (θ, σ)-FRS model based on fuzzy granules.

The essential concept underlying existing robust FRS models is to ignore so-called noisy samples when computing lower and upper approximations. These models adopt different approaches to ignore noisy samples. For instance, VPRS-FRS and VPFRS overlook some boundary samples according to the variable precision theory. SFRS finds noisy samples by considering a tradeoff between the number of ignored samples and an increase in the LAM [22]. A common feature of the ignored samples in the above-mentioned models is that they all have low probability density (PD) values. Zhao *et al.* [13] used this information to develop the FVPRS model, which sets a threshold for PD values. When computing the lower and upper approximations, if the PD value of a sample is lower than the threshold, the sample is considered as a noisy one. Under this approach, different thresholds can be set for different datasets. It is known that different mechanisms for detecting noise result in different distributions [28]. Therefore, data distribution should be considered when detecting noisy samples. However, current FRS models are data-distribution-blind, i.e., these models do not consider the data distribution when computing the lower and upper approximations.

In 2006, a theory of fuzzy probabilistic approximation spaces was proposed by introducing probability into fuzzy approximation spaces [29]. Under this approach, a probability is assigned to each sample, and the probabilities are used to compute the fuzzy cardinalities instead of the lower and upper approximations. Following this approach, we propose a robust FRS model by considering data distribution when computing the lower and upper approximations. Accordingly, we refer to this model as the probabilistic FRS (PFRS) model. When computing the LAM of a sample x to a class A, we establish a tradeoff between the similarity of x and $y \notin A$ and the PD values $y \notin A$. Similarly, when computing the upper approximation membership of a sample x to a class A, we establish a tradeoff between the similarity of x and $y \in A$ and the PD values $y \in A$. In this way, the same boundary sample points are determined as noise and ignored when computing lower and upper approximations. Using the PFRS model, we design a prototype selection algorithm (PSA) and a robust classification algorithm (RCA), both of which are based on fuzzy rough covering theory. First, we propose a sample evaluation measure by combining PFRS and sample density values. With the new measure, the samples with larger evaluation values are classified correctly with higher probability. Then, the PSA is designed using the proposed sample evaluation measure based on fuzzy rough covering theory. Finally, using the PSA, the RCA is also developed on the basis of fuzzy rough covering. In addition, some experiments are conducted to test the effectiveness of the PSA and RCA.

In summary, the contributions of this paper are threefold. First, we propose a data-distribution-aware FRS model. In this model, we calculate the lower and upper approximations by making a tradeoff between similarity and PD values. Second, we develop a novel prototype evaluation index with the proposed model, and construct an RCA based on the prototype selection approach. Finally, we conduct extensive experiments to verify the effectiveness of the proposed model and algorithms.

This paper is organized as follows. Section II introduces and discusses the classical FRS model as well as several existing robust FRS models; in addition, this section formulates the problem statement based on the disadvantages of these models. Section III introduces the proposed PFRS model and analyzes its properties. Section IV describes the application of the proposed model to prototype selection and classification modeling. Section V describes a set of experiments conducted to test the effectiveness of PFRS, PSA, and RCA. Finally, Section VI summarizes our findings and concludes this paper.

II. Preliminaries and Problem Description

We here introduce the preliminaries and describe the problem associated with existing robust FRS models.

Let U be a nonempty set with a finite number of objects, and R be a fuzzy binary relation on U. R is a fuzzy equivalence relation if $R(x, x) = 1$; $R(x, y) = R(y, x)$, and $R(x, y) \geq \text{supmin}_{z \in U}\{R(x, z), R(z, y)\}$. The fuzzy equivalence class $[x]_R$ associated with x and R is a fuzzy set on U, where $[x]_R(y) = R(x, y)$ for all $y \in U$. $F(y)$ is the membership of y belonging to fuzzy set F. T is a triangular norm (t-norm) and S is a triangular conorm (t-conorm). $S(x, y) = \max(x, y)$ is the standard max operator. For a t-conorm S, σ is defined as $\sigma(a, b) = \inf\{c \in [0, 1] : S(a, c) > b\}, a, b \in [0, 1]$ [30], [31]. An involutive negator N is a decreasing mapping $[0, 1] \to [0, 1]$ satisfying $N(0) = 1$, $N(1) = 0$, and $N(N(x)) = x$. The standard negator is defined as $N_S(x) = 1 - x$ [32]. Further, $\vartheta(a, b) = \sup\{c \in [0, 1] : T(a, c) \leq b\}, a, b \in [0, 1]$ is called a R-implicator. If T is lower semicontinuous, then ϑ is called the T-residuation implication [31].

Definition 1: Let U be a nonempty universe, R be a fuzzy equivalence relation on U, and $F(U)$ be the fuzzy power set of U. Given a fuzzy set $F \in F(U)$, the lower and upper approximations of F are defined as [3]

$$\begin{aligned} \underline{R}F(x) &= \inf_{y \in U} \max\{1 - R(x, y), F(y)\} \\ \overline{R}F(x) &= \sup_{y \in U} \min\{R(x, y), F(y)\}. \end{aligned} \tag{1}$$

The lower approximation indicates the certainty that a sample belongs to a class, and the upper approximation indicates the possibility that a sample belongs to a class. The predictive ability of a sample is proportional to its LAM. Subsequently, this model was generalized using other fuzzy operators [30], [31], [33], [34]

$$\begin{aligned} \overline{R_T}F(x) &= \sup_{u \in U} T\big(R(x, u), F(u)\big) \\ \underline{R_S}F(x) &= \inf_{u \in U} S(N(R(x, u), F(u)) \\ \overline{R_\sigma}F(x) &= \sup_{u \in U} \sigma\big(N(R(x, u)), F(u)\big) \\ \underline{R_\vartheta}F(x) &= \inf_{u \in U} \vartheta\big(R(x, u), F(u)\big). \end{aligned} \tag{2}$$

FRSs have been successfully applied to attribute reduction, and rule extraction [6], [7]. However, it is reported that they are sensitive to noise, which limits their practical applications. As mentioned in [23] and [24], the lower and upper approximations of FRSs are sensitive to label noise because FRSs are computed on the basis of the statistics of minimum and maximum, which are sensitive to outliers.

Some robust FRS models have been proposed these years, including β-precision FRS model [17], VPFRS model [10], VQRS [18], FVPRSs [13], SFRSs [22], ordered weighted average FRS theory [19], and robust statistics-based FRS models [23]. Due to the page limitation, we just briefly review a typical model: β-precision FRSs.

In [17], β-precision aggregation was introduced into FRSs to present a robust model called the β-precision FRS model. It is defined by the following four operators:

$$\begin{aligned}
\underline{R_S}F(x) &= T_{\beta_{y\in U}} S\Big(N\big(R(x,y)\big), F(y)\Big)\\
\overline{R_T}F(x) &= S_{\beta_{y\in U}} T\big(R(x,y), F(y)\big)\\
\underline{R_\vartheta}F(x) &= T_{\beta_{y\in U}} \vartheta\big(R(x,y), F(y)\big)\\
\overline{R_\sigma}F(x) &= S_{\beta_{y\in U}} \sigma\Big(N\big(R(x,y)\big), F(y)\Big).
\end{aligned} \tag{3}$$

Here, T_β and S_β are β-precision quasi-T-norms and β-precision quasi-T-conorms [17]. $T_\beta(x_1, x_2, \ldots, x_N) = T(x_1, x_2, \ldots, x_n), \beta \in [0, 0.5)$, where $x_1 \geq x_2 \geq \cdots \geq x_N$, and $n = \max_k\{k \in [0, 1, \ldots, N] | k \leq \sum_{i=1}^{N} x_i(1-\beta)\}$. $S_\beta(x_1, x_2, \ldots, x_N) = S(x_1, x_2, \ldots, x_n), \beta \in [0, 1]$, where $x_1 \geq x_2 \geq \cdots \geq x_N$, and $n = \max_k\{k \in [0, 1, \ldots, N] | k \leq \sum_{i=1}^{N}(1-x_i)(1-\beta)\}$.

Let $T(x, y) = \min(x, y)$ and $S(x, y) = \max(x, y)$, the lower and upper approximations of β-PFRS, degenerate to the following form:

$$\begin{aligned}
\underline{R}_\beta F(x) &= \min_{\beta\ y\in U} \max\big(1 - R(x,y), F(y)\big)\\
\overline{R}_\beta F(x) &= \max_{\beta\ y\in U} \min\big(R(x,y), F(y)\big).
\end{aligned} \tag{4}$$

Given a decision class d, the LAM of $x \in d$ reads as follows:

$$\begin{aligned}
\underline{R}_\beta d(x) &= \min_{\beta\ y\in U-d} \max\big(1 - R(x,y), d(y)\big)\\
&= \min_{\beta}(g_1, \ldots g_v) = \min(g_1, \ldots, g_u)
\end{aligned} \tag{5}$$

where $g_1 \geq g_2 \geq \cdots \geq g_v$, $|U - d| = v$, $g_i = 1 - R(x, y_i)$ $(y_i \in U - d)$, and $u = \max_k\{k \in [0, 1, 2, \ldots, v] : k \leq \sum_i^v g_i(1-\beta)\}$. And the number of samples neglected in computing lower approximation of β-PFRS is $v - u$. β is the direct parameter used for determining the number of ignored samples.

The robustness analysis shows that $\underline{R}_\beta d(x)$ of β-PFRS achieves robustness to noise by overlooking the some nearest neighbors of x from different classes.

In fact, the robust FRS models can be categorized into two classes. In the first approach, samples located around the classification boundary that are considered as noisy samples are ignored. Models that follow this approach include β-PFRS, VPFRS, SFRS, k-trimmed FRS, and SMEB-FRS. These robust models differ only in terms of the method used to determine the samples that should be ignored. The second approach uses robust approximation operators. Models that follow this approach include VQRS, FVPRS, k-mean FRS, ordered weighted average-FRS (OWA-FRS), and k-median FRS.

As mentioned above, the first approach adopts different measures to ignore samples that are considered as noisy samples. By analyzing existing robust FRS models, we can conclude that the ignored samples have low probability densities. If the PD of a sample is high, the possibility that the sample belongs to noise is small. In addition, the PD of a sample is closely related to its influence on the lower approximation of FRS.

The above analysis shows that the probability densities of samples are closely related to the determination of noise. In addition, collected samples are usually contaminated by noise, which submits to various distributions in practice. Thus, several robust FRS models exhibit poor performance or are ineffective in various applications. This is because the critical parameters used for noise determination in different FRS models are set on the basis of experience. For datasets with various distributions, these parameter values are usually adjusted using experimental results. For example, for datasets that submit to heavy-tailed distributions, the parameter k in k-trimmed FRS should be larger than that for datasets with Gaussian distributions. Thus, by using the distribution information, a robust model can adaptively select the values of k for different datasets. In conclusion, it is important to consider distribution information of datasets when building a robust FRS framework.

III. Robust Fuzzy Rough Set Model Considering Data Distributions

In this section, we first present a robust FRS model considering data distribution. Then, we discuss some properties of the proposed model.

A. Data-Distribution-Aware Fuzzy Rough Sets

For FRS, the LAM $\underline{R}F(x)$ of x to its own class equals the dissimilarity between x and the nearest neighbor y in other classes. If y is a mislabeled sample, $\underline{R}F(x)$ is reduced considerably. Further, if y is a noisy sample with a small error, $\underline{R}A(x)$ is reduced slightly. Thus, two types of noise influence the lower approximation of FRS. First, we introduce a measure for evaluating the influence of noisy samples on the lower approximation of FRS during classification.

Definition 2: Let U be a nonempty universe, and let R be a fuzzy equivalence relation on U. D is the decision attribute. Sample x belongs to decision class d. $p(y)$ is the probability of $\{y | y \notin d\}$. The influence of noisy sample y on the LAM of x to d is defined as

$$IF(y_x) = \big(1 - P(y)\big) \cdot R(x, y). \tag{6}$$

The similarity between x and y can be computed with Gaussian kernel function.

It is shown that the influence of noisy sample $y \notin d$ on $\underline{R}d(x)$ is related to the similarity $R(x, y)$ between x and y, and the

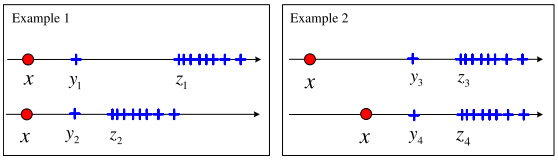


Fig. 1. Influence of $P(y)$ and $R(x, y)$ on $\underline{R}d(x)$. (a) $R(x, y_1) = R(x, y_2)$. (b) $P(y_3) = P(y_4)$.

density at y. Suppose that the noisy sample y is the nearest neighbor of x in $U - d$. The smaller the value of $P(y)$ is, the farther is y from $U - d$. Accordingly, y has a considerable influence on the LAM of x to d. In addition, the higher the similarity $R(x, y)$, the greater is the influence of y on $\underline{R}d(x)$. Fig. 1 illustrates the above situation.

Considering Example 1, assume that y_1 and y_2 are two noisy samples. Since $R(x, y_1) = R(x, y_2)$ and $P(y_1) < P(y_2)$, $IF(y_1) > IF(y_2)$. If y_1 and y_2 are used to compute $\underline{R}d(x)$, the relative error values of $\underline{R}d(x)$ are

$$\text{err}_1 = \frac{\big(1 - R(x, z_1)\big) - \big(1 - R(x, y_1)\big)}{1 - R(x, y_1)} \tag{7}$$

$$\text{err}_2 = \frac{\big(1 - R(x, z_2)\big) - \big(1 - R(x, y_2)\big)}{1 - R(x, y_2)}. \tag{8}$$

Obviously, $\text{err}_1 > \text{err}_2$, which implies that the lower the PD value of the sample is, the larger the IF of the sample is. In the case of Example 2, since $R(x, y_3) < R(x, y_4)$ and $P(y_3) = P(y_4)$, $IF(y_3) < IF(y_4)$. The relative errors of $\underline{R}d(x)$ are

$$\text{err}_3 = \frac{\big(1 - R(x, z_3)\big) - \big(1 - R(x, y_3)\big)}{1 - R(x, y_3)} \tag{9}$$

$$\text{err}_4 = \frac{\big(1 - R(x, z_4)\big) - \big(1 - R(x, y_4)\big)}{1 - R(x, y_4)}. \tag{10}$$

Clearly, $\text{err}_3 < \text{err}_4$. It means that as y_4 is more similar to x, and it has a greater influence on $\underline{R}d(x)$ than does y_3.

The above analysis shows that if the nearest neighbor of x is a noisy sample in other classes, is more similar to x, and has a lower density, then its influence on the computed value of $\underline{R}d(x)$ is greater. Thus, both PD and similarity of a sample influence the computation of the lower approximation. Now, we present a way of determining noisy samples by looking at the underlying probability.

Definition 3: Let U be a nonempty universe, R be a fuzzy equivalence relation on U, and $F(U)$ be the fuzzy power set of U. The minimal probability of a normal sample for x is defined as

$$P_{\min}(x) = \inf_{y \in U} S\Big(N\big(P(y) \cdot R(x, y)\big), P(y)\Big) \tag{11}$$

where $P(y)$ is the PD value at point y. Given a fuzzy set $A \in F(U)$, $\forall y \in U$, if $A(y) \geq P_{\min}$, y is defined as a normal sample; otherwise, y is a noisy sample.

Now, we present a robust FRS model which considers data distribution.

Definition 4: Let U be a nonempty universe; R, a fuzzy equivalence relation on U; and $F(U)$, the fuzzy power set of U, and $A \in F(U)$. The probabilistic lower and upper approximations of A are defined as

$$\begin{aligned}
\overline{PR}_T A(x) &= \sup_{y \in U} T\big(R_P(x, y), A(y)\big) \\
\underline{PR}_S A(x) &= \inf_{y \in U} S\Big(N\big(R_P(x, y)\big), A(y)\Big) \\
\overline{PR}_\sigma A(x) &= \sup_{y \in U} \sigma\Big(N\big(R_P(x, y)\big), A(y)\Big) \\
\underline{PR}_\vartheta A(x) &= \inf_{y \in U} \vartheta\big(R_P(x, y), A(y)\big)
\end{aligned} \tag{12}$$

where

$$R_P(x, y) = \begin{cases} R(x, y), & P(y) \geq P_{\min}(x) \\ 0, & P(y) < P_{\min}(x). \end{cases} \tag{13}$$

The above definitions (12) can be simplified as

$$\begin{aligned}
\overline{PR}_T A(x) &= \sup_{P(y) \geq P_{\min}} T\big(R(x, y), A(y)\big) \\
\underline{PR}_S A(x) &= \inf_{P(y) \geq P_{\min}} S\Big(N\big(R(x, y)\big), A(y)\Big) \\
\overline{PR}_\sigma A(x) &= \sup_{P(y) \geq P_{\min}} \sigma\Big(N\big(R(x, y)\big), A(y)\Big) \\
\underline{PR}_\vartheta A(x) &= \inf_{P(y) \geq P_{\min}} \vartheta\big(R(x, y), A(y)\big).
\end{aligned} \tag{14}$$

It means that the proposed model ignores the samples that have lower PD values than $P_{\min}$ when computing the lower and upper approximation memberships.

Now, we present the above definition for a classification problem. Let $S(a, b) = S_M(a, b) = \max(a, b)$. The lower approximation $\underline{PR}_S A(x)$ can be simplified as follows:

$$\underline{PR}_S A(x) = \inf_{\substack{y \notin A \\ P(y) \geq P_{\min}(x)}} \big(1 - R(x, y)\big). \tag{15}$$

Similarly, let $T(a, b) = T_M(a, b) = \min(a, b)$, then the T-upper approximation is simplified as

$$\overline{PR}_T A(x) = \sup_{\substack{y \in A \\ p(y) \geq p_{\min}(x)}} \big(R(x, y)\big). \tag{16}$$

We can see that the LAM of x to A is the minimal dissimilarity between x and its nearest neighbor $y \notin A$ that satisfies $P(y) \geq P_{\min}$. And the upper approximation membership of x to A is the maximal similarity between x and its nearest neighbor $y \in A$ that satisfies $P(y) \geq P_{\min}$. Thus, the new lower and upper approximation operators achieve robustness to noise by ignoring samples that have lower PD values.

In Definition 4, $P_{\min}$ plays an important role in the development of the PFRS model. Next, we demonstrate how to determine $P_{\min}$ in classification problems. When computing the LAM $\underline{PR}_S A(x)$ ($\underline{PR}_\vartheta A(x)$), $P_{\min}(x)$ is calculated as

$$P_{\min}(x) = \inf_{y \notin A} S\big(N\big(P(y) \cdot R(x, y)\big), P(y)\big). \tag{17}$$

Let $S(a, b) = \max(a, b)$, the above expression is reduced to

$$P_{\min}(x) = \inf_{y \notin A} \max\big(1 - P(y) \cdot R(x, y), P(y)\big) \tag{18}$$

where P is the probability distribution of $U - A$. Note that (17) returns the minimal PD value of the normal samples in $U - A$, which is used to select a normal sample y for computing

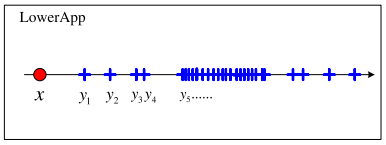


Fig. 2. Artificial dataset.

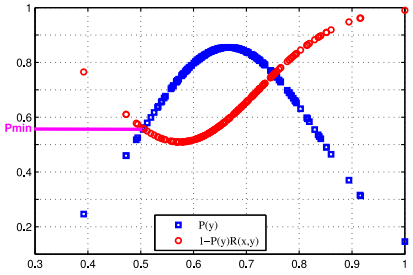


Fig. 3. Determining $P_{\min}$ with Gaussian noise distribution.

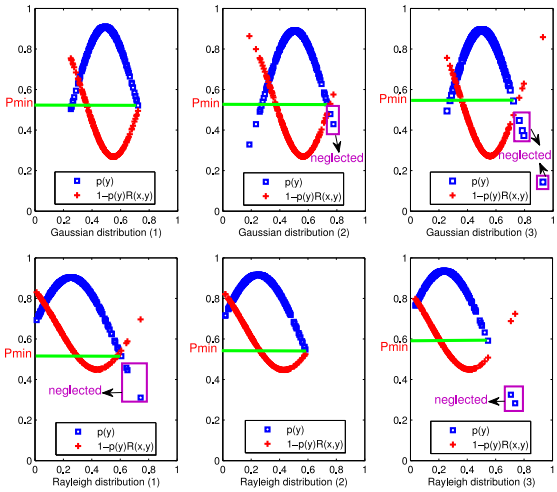


Fig. 4. Numbers of samples neglected with symmetrical and dissymmetrical distributions.

$\underline{PR_S}A(x)$. For computing the LAMs of different samples, the returned values of $P_{\min}$ are different. In practice, the probability distribution of data is usually unknown, and $P(\cdot)$ can be estimated as

$$P(x) = \frac{1}{n}\sum_{i=1}^{n} G(x, y_i) = \frac{1}{n}\sum_{i=1}^{n} \exp\left(-\frac{\| x - y_i \|^2}{2\sigma^2}\right). \quad (19)$$

Now, we present an example to illustrate the computation of $P_{\min}$. As shown in Fig. 2, $x \in A$, $y_i \notin A$; when computing $\underline{PR_S}A(x)$, it is important to determine $P_{\min}$. Suppose that the probability distribution of $\{y_i\}$ is a Gaussian distribution. Sort the elements of $\{y_i\}$ in descending order based on the similarity between x and y_i, and obtain an ordered sample set $\{y_i'\}$. Thus, $R(x, y)(y \in \{y_i'\})$ is a monotone minus function, and $P(y)(y \in \{y_i'\})$ first increases and then decreases. There is a minimum $1-P(y_k')\cdot R(x, y_k')$ of $1-P(y)\cdot R(x, y)$. When y_k' is a normal sample, it can be used to compute $\underline{PR_S}A(x)$. However, y_k' may have higher density. Therefore, (17) addresses this problem by selecting the maximum between $1-P(y_k')\cdot R(x, y_k')$ and $P(y_k')$. In Fig. 3, $1 - P(y) \cdot R(x, y)$ does have a minimum. From the figure, we can easily find $P_{\min}$, and there are four samples with lower PD values than $P_{\min}$. Thus, the fifth nearest neighbor of x is determined to be a normal sample, and y_5 is used to compute the LAM of x to A, that is

$$\underline{PR_S}A(x) = \inf_{\substack{y\notin A \\ p(y)\geq p_{\min}}} \big(1 - R(x, y)\big) = 1 - R(x, y_5). \quad (20)$$

The above analysis shows that when computing $\underline{PR_S}A(x)$, it is preferable that both $R(x, y)$ and $P(y)$ be large, where $y \notin A$. For determining $P_{\min}(x)$, a tradeoff must be achieved between $R(x, y)$ and $P(y)$. Similarly, when computing $\overline{PR_T}A(x)$, it is preferable that both $R(x, y)$ and $P(y)$ be large, where $y \in A$. PFRS achieves robustness to datasets following different distributions by introducing PD distribution into FRS model.

B. Property Analysis

Yao and Wong [35] proposed probabilistic approximations with respect to two cases. For $0 \leq \beta < \alpha \leq 1$, the lower and upper probabilistic approximations of $A \subseteq U$ are, respectively, defined as

$$\begin{aligned} \overline{R}A(x) &= \cup\{[x]|Pr(A|[x]) > \beta\} \\ \underline{R}A(x) &= \cup\{[x]|Pr(A|[x]) \geq \alpha\}. \end{aligned} \quad (21)$$

Pawlak and Skowron [36] suggested using a rough membership function $\mu_A(x) = Pr(A|[x])$ to redefine the two approximations. And rough sets can be represented as

$$\begin{aligned} \overline{R}A(x) &= \{x \in U|Pr(A|[x]) > 0\} \\ \underline{R}A(x) &= \{x \in U|Pr(A|[x]) = 1\}. \end{aligned} \quad (22)$$

In [37], the (α, β)-lower approximation and upper approximation with $\mu_A(x)$ are defined, respectively, as

$$\begin{aligned} \overline{R}_{(\alpha,\beta)}A(x) &= \{x \in U|Pr(A|[x]) > \beta\} \\ \underline{R}_{(\alpha,\beta)}A(x) &= \{x \in U|Pr(A|[x]) \geq \alpha\}. \end{aligned} \quad (23)$$

Ma and Sun [25] defined the probabilistic rough set over two universes.

Definitely, there is an obvious difference between PFRS and above models in addressing noise. The mentioned probabilistic rough set models address noise via controlling the values of the rough membership function $\mu_A(x)$, and they are not related to data distributions. And our model richly utilizes data distribution information in handling noise, which makes PFRS be robust to datasets with different distributions. This conclusion is illustrated as follows.

In practice, sampled data usually follow different distributions. Here, suppose that $\{y_i\}$ follows two kinds of distributions, namely, Gaussian distribution (symmetrical distribution), and Rayleigh distribution (asymmetrical distribution). Fig. 4 shows the determined $P_{\min}$ values with three datasets following the same Gaussian distribution and three datasets following the same Rayleigh distribution. In the figure, x equals 0.85. $P(y)$ is the PD at y. It is shown that there are, respectively, 0, 2, 4, 3, 0, and 2 samples with lower PD than determined values of $P_{\min}$. With our definition of FRS, there are, respectively, 0, 2, 4, 3, 0, and 2 samples neglected in computing the LAM of x to a class A.

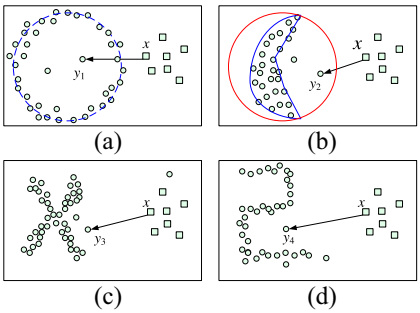


Fig. 5. Some special data distributions. (a) Circle distribution. (b) Moon distribution. (c) Connected graph (1). (d) Connected graph (2).

Furthermore, if the data satisfy a uniform distribution, $P(y)(y \in \{y_i\})$ is equivalent. Obviously, $P_{\min}(x) = P(y)$. Thus, no sample should be overlooked in computing LAMs. On occasion, datasets follow special or irregular distributions, such as circle distribution, moon distribution, and connected graphs shown in Fig. 5. PFRS is robust to noise with lower PD in other classes by introducing probability distribution of datasets, which guarantees that the samples used to compute lower approximation of PFRS have higher densities than $P_{\min}$. This results illustrate our PFRS model is neatly adapt to datasets with different distributions, which is the key difference from the existing robust FRS models.

In β-PFRS and VPFRS models, the number of samples neglected is determined by the values of β. With the six datasets used in Fig. 4, given a value of β, the number of samples neglected are the same as each other. Some experiments are operated on the above six cases. Particularly, if datasets follow an uniform distribution, there should not be any samples neglected. However, β-PFRS still neglects some samples in computing the lower approximation.

For SFRS model, it is to find a tradeoff between LAMs and number of samples neglected, which is similar to PFRS on anti-noise. In this case, the SFRS model is not appropriate to some special distributions, such as circle and connected graph distributions [Fig. 5(a), (c), and (d)]. Some lower density samples may be selected to compute the lower approximation of SFRS.

In k-trimmed FRS model, the parameter k plays an important role for anti-noise. As shown in Fig. 4, the parameter k should have different values for six datasets. As to uniform distribution, k should equal zero. However, the value of k is set with subjectivity or via experiments. There is no relationship between values of k and data distributions. With datasets following different distributions, it is very difficult to select suitable values of k. This will cause k-trimmed FRS exhibiting unstable performance on different datasets.

1) SMEB-FRS Model: It is easily shown that SMEB-FRS is only adapt to datasets following spherical distribution. For example, given a dataset submitting half moon distribution, the LAM of x to its own class with SMEB-FRS is sensitive to noise, which is shown in Fig. 5(b). This effectively illustrates that SMEB-FRS model is not adapted to nonconvex classification problems. PFRS can easily address this kind of problems by considering the probability distribution.

For the standard min operator $T_M(x, y) = \min\{x, y\}$ (t-conorm operator $S_M(x, y) = \max\{x, y\}$), $T_L(x, y) = \max\{0, x+y-1\}$ (t-conorm operator $S_L(x, y) = \min\{1, x+y\}$) and standard negator $N(x) = 1 - x$, some properties of PFRS are discussed. The related conclusions are easily extended if other fuzzy operators are used.

Proposition 1: $\forall\ A \in F(U)$, the following statements hold:

$$\begin{aligned} &1)\ \underline{PR}_S A = N\big(\overline{PR}_T(N(A))\big), \overline{PR}_T A = N\Big(\underline{PR}_S\big(N(A)\big)\Big) \\ &2)\ \underline{PR}_\vartheta A = N\big(\overline{PR}_\sigma\big(N(A)\big)\big), \overline{PR}_\sigma A = N\big(\underline{PR}_\vartheta\big(N(A)\big)\big). \end{aligned} \tag{24}$$

Proof: 1) $\forall\ x \in U$

$$\begin{aligned} N\Big(\overline{PR}_T\big(N(A(x))\big)\Big) &= N\Big(\sup_{y\in U}\Big(T\Big(R_P(x,y), N\big(A(y)\big)\Big)\Big)\Big) \\ &= \inf_{y\in U}\Big(N\Big(T\Big(R_P(x,y), N\big(A(y)\big)\Big)\Big)\Big) \\ &= \inf_{y\in U}\Big(S\Big(N\big(R_P(x,y)\big), A(y)\Big)\Big) = \underline{PR}_S A. \end{aligned} \tag{25}$$

Similarly, $\overline{PR}_T A = N(\underline{PR}_S(N(A)))$, $\underline{PR}_\vartheta A = N(\overline{PR}_\sigma(N(A)))$, and $\overline{PR}_\sigma A = N(\underline{PR}_\vartheta(N(A)))$ holds. ■

Proposition 2: $\forall\ A \in F(U)$, if x is a normal sample, the following statements hold:

$$1)\ \underline{PR}_S A \subseteq A \subseteq \overline{PR}_T A.\ 2)\ \underline{PR}_\vartheta A \subseteq A \subseteq \overline{PR}_\sigma A. \tag{26}$$

Proof: $\forall\ x \in U$

$$\begin{aligned} \underline{PR}_S A(x) &= \inf_{y\in U} S\Big(N\big(R_P(x,y)\big), A(y)\Big) \\ &\le S\Big(N\big(R_p(x,x)\big), A(x)\Big). \end{aligned} \tag{27}$$

As x is a normal sample, x can be used to compute $\underline{PR}_S A(x)$

$$\begin{aligned} S(N(R_p(x,x)), A(x)) &= S\Big(N\big(R(x,x)\big), A(x)\Big) \\ &= S\big(0, A(x)\big) = A(x). \end{aligned} \tag{28}$$

Thus, $\underline{PR}_S A \subseteq A$

$$\overline{PR}_T A(x) = \sup_{y\in U} T\big(R_P(x,y), A(y)\big) \ge T\big(R_p(x,x), A(x)\big). \tag{29}$$

As x is a normal sample, x can be used to compute $\overline{PR}_T A(x)$

$$T\big(R(x,x), A(x)\big) = T\big(1, A(x)\big) = A(x). \tag{30}$$

Thus, $A \subseteq \overline{PR}_T A$. Accordingly, $\underline{PR}_S A \subseteq A \subseteq \overline{PR}_T A$ holds.

Similarly, $\underline{PR}_\vartheta A \subseteq A \subseteq \overline{PR}_\sigma A$ holds. ■

Proposition 3: Suppose $A \subset U$ is a set, then the following statements hold:

$$\begin{aligned} &1)\ \underline{PR}_S A \supseteq \underline{R}_S A, \overline{PR}_T A \subseteq \overline{R}_T A. \\ &2)\ \underline{PR}_\vartheta A \supseteq \underline{R}_\vartheta A, \overline{PR}_\sigma A \subseteq \overline{R}_\sigma A. \end{aligned} \tag{31}$$

Proof: $\forall x \in U$, $\underline{PR}_S A(x) = \inf_{y\notin A} S(1 - R_P(x,y), A(y)) \ge \inf_{y\notin A} S(1 - R(x,y), A(y)) = \underline{R}_S A(x)$. Thus $\underline{PR}_S A \supseteq \underline{R}_S A$.

$\forall x \in U$, $\overline{PR}_T A(x) = \sup_{y\in A} T(R_P(x,y), A(y)) \le \inf_{y\in A} T(R(x,y), A(y)) = \overline{R}_T A(x)$. Thus $\overline{PR}_T A \subseteq \overline{R}_T A$.

Similarly, $\underline{PR}_\vartheta A \supseteq \underline{R}_\vartheta A$, and $\overline{PR}_\sigma A \subseteq \overline{R}_\sigma A$ holds. ■

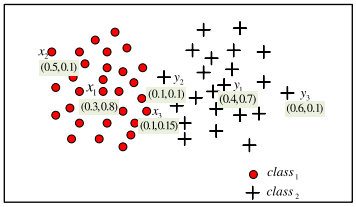


Fig. 6. Artificial datasets.

IV. Prototype Selection and Classification

A. Prototype Evaluation

A sample can be regarded as a prototype if it has a large LAM to its class and high PD. We perform prototype evaluation for each sample by combining the lower approximation of FRSs and the PD value of the sample. The evaluation measure is defined as

$$q_x = \underline{R}_S A(x) \cdot P(x). \tag{32}$$

Note that $\underline{R}_S A(x)$ can be replaced with $\underline{R}_\vartheta A(x)$. A large q_x implies that x is a good prototype. Furthermore, the sphere of influence of each prototype is defined as $r_x = \underline{R}_S A(x)$. As classical FRSs are sensitive to noise, we use $\underline{PR}_S A(x)$ to compute the lower approximation of samples.

The lower approximation of FRS can be used to measure the certainty of samples to a class. And the PD can be used to measure the reliability of samples. By incorporating the above two measures, it is ensured that the selected prototypes with the new measure are inner points which can express richly the character of a class.

Now, an example is used to demonstrate the effectiveness of (32). Fig. 6 shows the distribution of a dataset with two classes, where $y_1, y_2, y_3 \in \text{class}_1$ and $x_1, x_2, x_3 \in \text{class}_2$. The LAM and PD of $x_i (i = 1, 2, 3)$ and $y_j (j = 1, 2, 3)$ are (0.3, 0.8), (0.5, 0.1), (0.1, 0.15), (0.4, 0.7),(0.1, 0.1), and (0.6, 0.1), respectively. Here, we take the product of LAM and PD as the criterion for a sample to qualify as a prototype. If the product is large, the sample is suitable as a prototype. Thus, the evaluation values of x_i are $q_{x_1} = 0.24$, $q_{x_2} = 0.05$, and $q_{x_3} = 0.015$, respectively, which implies that x_1 is the most suitable sample as a prototype. Further, x_2 is a better prototype than x_3. The spheres of influence are $r_{x_1} = 0.3$, $r_{x_2} = 0.5$, and $r_{x_3} = 0.1$, respectively. Similarly, the evaluation values of y_j are $q_{y_1} = 0.28$, $q_{y_2} = 0.01$, and $q_{y_3} = 0.06$, respectively. Clearly, y_1 is the most suitable sample as a prototype, followed by y_3. The spheres of influence are $r_{y_1} = 0.4$, $r_{y_2} = 0.1$, and $r_{y_3} = 0.6$, respectively.

B. Fuzzy Rough Prototype Selection

A PSA is constructed using the proposed prototype evaluation measure and sphere of influence. The algorithm is presented in Table I.

The PSA is divided into two phases. The first phase evaluates the quality of each sample as a prototype using $q_s = \underline{PR}_S A(s) \cdot P(s), s \in S$. The second phase selects the prototypes with fuzzy rough covering theory. First, the most competent prototype is selected, where r_s is the sphere of influence of the prototype. Then, the samples ss that are covered by the sphere of influence of the selected prototype s are deleted from S.

TABLE I
Fuzzy Rough PSA

Input	Training set S.
2	$PS = \emptyset$, $R = \emptyset$;
3	Evaluate the quality of each sample as a prototype q_s;
4	Compute LAM of each sample s, denoted by r_s;
5	Estimate PD of each sample s, denoted by p_s;
6	Compute $q_s = r_s \times p_s$;
7	While $S \neq \emptyset$
8	Select the sample $s \in S$ with the largest value of q_s;
9	$PS = PS \bigcup \{s\}$, $S = S - \{s\}$, $R = R \bigcup \{r_s\}$;
10	Compute the dissimilarity dd between s and samples $ss \in S$;
11	$S = S - ss$ where $dd < r_s$;
12	End
Output	Prototype set PS and the sphere of influence R.

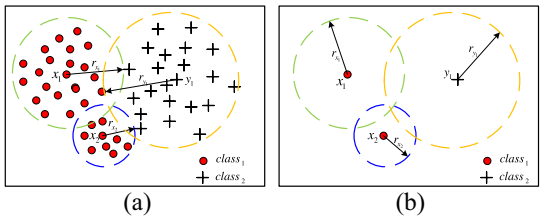


Fig. 7. Prototype selection with fuzzy rough covering. (a) Raw data set. (b) Prototype set.

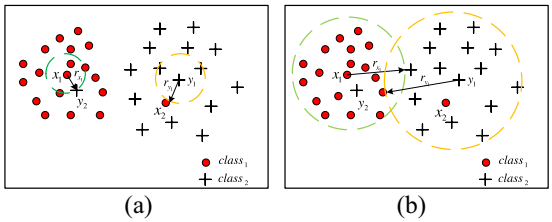


Fig. 8. Prototype selection with (a) FRS and (b) PFRS.

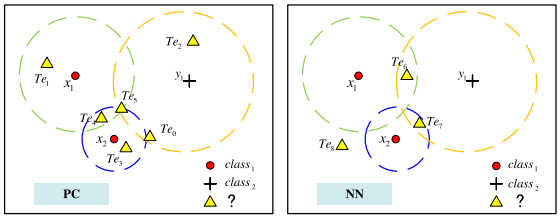
As shown in Fig. 7(a), let us suppose that x_1 is the selected prototype, and its sphere of influence is r_{x_1}. Then, x_1 is added to the prototype set PS, and the samples for which the dissimilarity with x_1 is less than r_{x_1} are removed from the training set S. Similarly, y_1 is added to the PS, and the samples for which the dissimilarity with y_1 is less than r_{y_1} are deleted from the training set S. Nevertheless, some samples in class_1 are not covered by the sphere of influence of x_1. In this case, another prototype is selected among the remaining samples of class_1. If x_2 is the best prototype candidate, it is added to the prototype set and deleted from class_1. The PSA will repeat this process until class_1 becomes empty. The same principle applies to class_2. Accordingly, the PSA outputs the PS and the set R of spheres of influence of each prototype. Fig. 7(b) shows the selected prototypes and their spheres of influence.

In practice, data are usually corrupted by noise. When FRSs are used to evaluate prototypes, noisy samples are selected as prototypes. Thus, the selected prototype set cannot be used for classification. Here, we evaluate prototypes using the proposed PFRSs, and noisy samples are not selected as prototypes. Fig. 8 illustrates the above-mentioned problem.

The figure shows a dataset with two noisy samples, x_2 and y_2. Suppose that x_1 is selected as a prototype. Its sphere

TABLE II
PFRS-BASED PROTOTYPE COVERING CLASSIFICATION ALGORITHM

Input	Test set TS with m samples; Prototype set PS with k prototypes from C classes; Set of spheres of influence, R.
2	$L = \emptyset$;
3	for $l = 1$ to m
4	Compute the dissimilarity dr_{p_i} between $Te_l \in TS$ and each prototype $p_i \in PS$;
5	if $(dr_{p_i} \le r_{p_i})$ or $(dr_{p_i} \le r_{p_i}$ and $dr_{p_j} \le r_{p_j}$ and $\cdots$ and $dr_{p_k} <= r_{p_k})$, where $r_* \in R$;
6	$TL_{Te_l} = TL_{p_i}$ or $TL_{Te_l} = majority\{TL_{p_i}, TL_{p_j}, ..., TL_{p_k}\}$;
7	else
8	$TL_{Te_l} = TL_{p_t}$, where $dr_{p_t} = \min\{dr_{p_i} \mid p_i \in P\}$;
9	end
10	$L(l) = TL_{Te_l}$;
11	end
Output	Label set L of test samples.



(a) (b)

Fig. 9. Classification with prototypes. (a) Prototype cover classification. (b) Nearest neighbor classification.

TABLE III
SUMMARY OF DATASETS

Datasets	Samples	Features	Continuous	Symbolic	Classes
glass	214	9	9	0	6
haberman	360	3	3	0	2
ICU	200	20	4	16	2
ionosphere	351	34	32	2	2
iris	150	4	4	0	3
rice	104	5	5	0	2
sonar	208	60	60	0	2
segmentation	200	19	16	3	7
thyroid	215	2	5	0	3
WDBC	569	30	30	0	2
wine	178	13	13	0	3
WPBC	198	32	32	0	2
DLBCL	88	4026	4026	0	6

of influence is r_{x_1}, and the number of covered samples is one [Fig. 8(a)]. Moreover, x_2 and y_2 cannot be covered by any sphere of influence. Finally, x_2 and y_2 are selected as prototypes. If we use PFRS to select prototypes, the sphere of influence of x_1 is expanded [Fig. 8(b)], and all the samples in class_1 are covered. Similarly, all the samples in class_2 are covered by the sphere of influence of y_2. Thus, the noisy samples x_2 and y_2 are not selected as prototypes. Therefore, prototype selection based on PFRS is more robust to noise than that based on FRS.

C. Prototype Covering for Robust Classification

Using the selected prototypes, we propose an RCA based on fuzzy rough covering theory and the nearest neighbor rule, which is presented in Table II. The input consists of the test sample set TS, prototype set PS, and set of spheres of influence, R. Further, the output of the classification algorithm consists of the label set L of test samples.

Given a test sample Te_l, the algorithm first computes the dissimilarity dr_{p_i} between Te_l and $p_i (p_i \in \text{PS})$. Then, it determines whether Te_l is covered by the sphere of influence of a prototype.

1) If Te_l is covered only by the sphere of a certain prototype $p_i (p_i \in \text{PS})$, i.e., $dr_{p_i} \le r_{p_i}$, then Te_l is marked with the label TL_{p_i} of the prototype p_i. Here, dr_{p_i} denotes the dissimilarity between Te_l and prototype p_i. In Fig. 9, x_1 and x_2 are two prototypes selected from class_1, and y_1 is the prototype selected from class_2. Further, $Te_0, Te_1, \ldots, Te_8$ are test samples. In Fig. 9(a), the labels of Te_1–Te_3 can be predicted using the above method.
2) If Te_l is covered by the spheres of influence of several prototypes $p_i, p_j, \ldots, p_k$, then Te_l is classified into the class to which the maximum prototypes belong. For instance, Te_4–Te_7 are all covered by the spheres of influence of multiple prototypes. In this case, if the $n (n = 1, \ldots, C)$ largest number of prototypes in each class are equal(for instance, Te_6 and Te_7), we apply the NN rule to classify a test sample. Thus, the test sample is classified into the class TL_{p_t} to which the nearest prototype p_t belongs, i.e., $dr_{p_t} = \min\{dr_{p_i} \mid p_i \in \text{PS}\}$.
3) If a test sample is not covered by the sphere of influence of any prototype, its label is also determined by the NN rule. For example, in Fig. 9(b), Te_8 is labeled class_1 by the NN rule. This usually occurs in the case of boundary points.

V. EXPERIMENTAL ANALYSIS

In this section, we describe some experiments conducted to test the robustness of the proposed PFRS set model using 13 datasets from the UCI database [38]. The datasets are summarized in Table III.

Now, we test the robustness of the PSA on a binary artificial dataset containing 400 samples. Here, the samples in each class agree to a Gaussian distribution. First, the artificial dataset is contaminated into five datasets having noise levels of 2%, 4%, 6%, 8%, and 10%. Here, noise is some mislabeled samples. In experiments, we let $\beta \in (0, 0.05)$, $k \in \{3, 4, 5\}$.

Fig. 10 shows a comparison of the selected prototypes using Table I with the FRS, k-mean FRS, VPFRS, and PFRS models. The selected prototypes were compared using raw

TABLE IV
PERFORMANCE COMPARISON OF DIFFERENT CLASSIFIERS

	FRS-PSRC	FRPS-NNC			
		FRPS-1	FRPS-2	FRPS-3	FRPS-4
glass	67.3	66.4	62.7	66.9	62.7
haberman	72.6	73.1	72.6	73.9	72.2
ICU	92.6	88.9	94.1	93.1	94.1
ionosphere	92.2	89.5	86.4	89.0	86.7
iris	97.3	96.0	95.3	95.3	95.3
rice	85.1	83.2	78.2	83.9	76.2
sonar	85.2	86.1	84.6	86.6	84.1
segmentation	91.9	85.2	86.7	85.7	87.6
thyroid	96.3	95.8	94.8	96.7	94.3
WDBC	96.1	96.5	95.6	96.7	96.0
wine	98.3	95.4	94.9	93.7	95.4
WPBC	79.8	73.2	72.7	73.2	71.1
DLBCL	97.3	96.3	96.3	96.3	96.3
AVE.	88.6	86.6	85.8	87.0	85.5

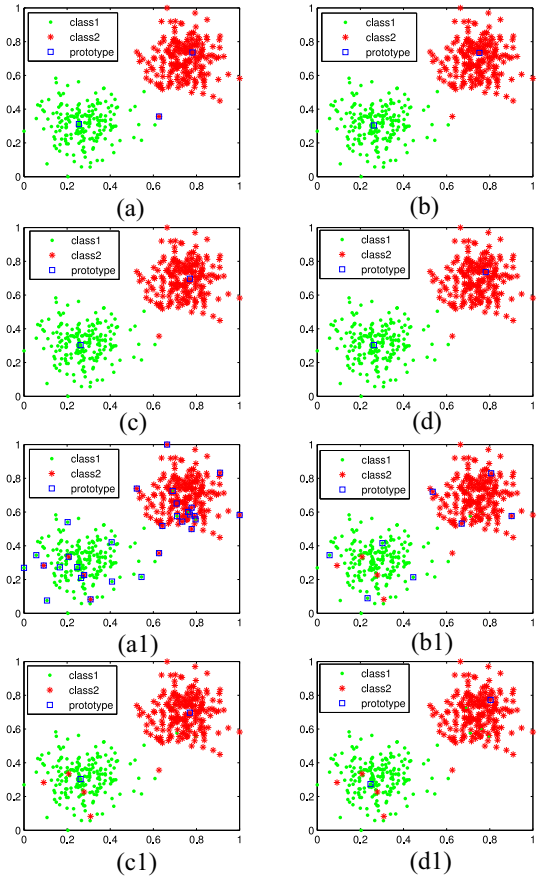


Fig. 10. Prototype selection with different FRS models. (a) FRS, noiselevel = 0%. (b) k-mean FRS, noiselevel = 0%. (c) VPFRS, noiselevel = 0%. (d) PFRS, noiselevel = 0%. (a1) FRS, noiselevel = 2%. (b1) k-mean FRS, noiselevel = 2%. (c1) VPFRS, noiselevel = 2%. (d1) PFRS, noiselevel = 2%.

artificial data and noisy data with 2% noise. In the case of the raw artificial dataset, three prototypes were obtained using Table I with FRS. It is easy to see that a border sample in class_2 was selected as a prototype, which is not what we want [Fig. 10(a)]. Further, two prototypes were obtained with k-mean FRS, VPFRS, and PFRS [Fig. 10(b)–(d)]. When the noise level of the dataset was 2%, the number of selected prototypes was 30 with FRS [Fig. 10(a1)]. It is shown that eight mislabeled samples were selected as prototypes with FRS, which implies that Table I is sensitive to noise using FRS. On the other hand, the number of prototypes obtained with k-mean FRS, VPFRS, and PFRS are, respectively, 8, 2, and 2 [Fig. 10(b1)–(d1)]. And there are none noisy samples selected as prototypes. Thus, the PSA based on k-mean FRS, VPFRS, and PFRS are more robust to noise than that based on FRS.

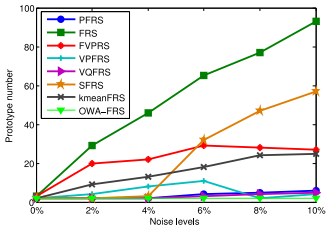


Fig. 11. Numbers of prototypes selected using different FRS models.

Fig. 11 shows the prototypes selected from the six datasets (an artificial dataset and five noisy datasets) using the PSA. It can be seen that the number of selected prototypes increases rapidly with the noise level in the case of FRS, whereas it increases gradually in the case of PFRS and VQFRS. Noise does not have any influence on the number of selected prototypes in the case of OWA-FRS. The figure shows that prototype selection based on PFRS, FVPRS, VPFRS, VQFRS, SFRS, k-mean-FRS, and OWA-FRS is more robust to noise than that based on FRS.

Some experiments were conducted to test the prototype-based RCA. First, we compare our classification algorithm with FRPS based nearest neighbor classifier (FRPS-NNC) [21]. FRPS-NNC method first computes the minimum granularity for each samples with FRS theory, and selects the samples with smaller granularity than a threshold as prototypes. And it adopts NN classification to predict the labels of unknown samples with prototypes. The classification accuracies of 13 datasets with FRS-prototype selection for robust classification (PSRC) and FRPS-NNC methods are shown in Table IV, where FRPS-1–FRPS-4 are four methods for computing minimum granularity. It is shown that our method performs best performance on glass, ionosphere, iris, rice, segmentation, wine, wisconsin prognostic breast cancer, and diffuse large B-cell lymphoma datasets.

Further, we compared our classification algorithm with several state-of-the-art algorithms such as multilayer perceptron (MLP) [39], linear support vector machine (LSVM) [40], Bayes net [41], C4.5 [42], and k-nearest neighbors

TABLE V
PERFORMANCE COMPARISON OF DIFFERENT CLASSIFIERS ON RAW DATASETS AND NOISY DATASETS

	NoiseLevel	PFRS-PSRC	FRS-PSRC	MLP	LSVM	BayesNet	C4.5	KNN
glass	0%	69.5	67.3	67.3	61.3	74.7	67.3	66.3
	5%	64.5	66.9	64.5	59.0	71.6	66.8	65.8
	10%	63.3	64.1	64.4	56.6	64.5	57.5	64.0
haberman	0%	73.3	72.6	72.9	73.9	72.9	71.9	70.5
	5%	73.2	72.2	72.2	73.5	70.2	70.6	69.5
	10%	72.5	72.2	68.3	73.4	68.3	66.3	68.3
ICU	0%	94.1	94.1	90.0	92.5	91.5	91.5	90.5
	5%	94.1	94.1	91.0	91.9	93.0	88.5	93.1
	10%	94.1	94.1	90.0	92.0	92.0	83.0	92.3
ionosphere	0%	92.4	92.9	91.2	87.6	89.5	86.6	86.4
	5%	90.1	89.6	88.3	86.9	88.3	82.8	85.5
	10%	86.7	85.5	81.6	86.2	84.4	82.4	83.8
iris	0%	97.3	98.0	97.3	96.7	92.7	96.0	95.3
	5%	97.3	95.0	96.0	96.3	93.3	93.1	95.3
	10%	96.0	95.2	98.6	95.3	97.7	89.1	93.3
rice	0%	88.8	86.0	84.6	78.2	76.9	81.0	84.0
	5%	86.5	83.6	87.3	78.0	81.9	75.5	81.9
	10%	83.1	80.6	75.7	77.2	80.3	75.4	80.6
sonar	0%	87.0	87.0	81.7	77.9	79.3	76.4	88.0
	5%	87.0	86.6	80.0	76.9	79.0	78.7	86.6
	10%	83.2	82.8	80.0	75.2	73.4	74.6	85.6
segmentation	0%	89.1	91.0	88.1	89.1	85.7	89.0	88.1
	5%	88.0	90.0	88.6	89.1	87.7	84.2	88.1
	10%	88.1	86.2	86.7	88.1	77.2	83.8	88.5
thyroid	0%	96.3	95.3	96.7	89.8	94.4	92.0	95.3
	5%	94.1	91.0	96.2	88.4	94.0	90.7	95.3
	10%	92.1	86.5	94.7	86.7	94.0	88.4	94.8
WDBC	0%	96.9	96.5	96.5	97.7	95.1	94.4	97.7
	5%	96.1	94.2	95.1	97.1	95.0	91.5	97.0
	10%	95.6	91.8	91.2	96.5	94.4	89.1	95.1
wine	0%	97.8	98.3	96.2	98.9	98.3	91.6	98.3
	5%	97.5	95.4	97.9	98.4	97.5	90.1	97.2
	10%	96.2	92.8	91.3	97.7	94.4	86.3	96.7
WPBC	0%	80.4	80.4	72.7	77.4	74.7	73.2	72.7
	5%	79.3	79.3	73.0	77.1	77.0	70.2	71.7
	10%	78.9	78.9	70.0	77.2	66.0	69.1	71.5
DLBCL	0%	97.3	97.3	94.3	97.3	97.7	90.9	97.7
	5%	96.3	95.3	93.8	95.7	96.3	87.1	97.3
	10%	95.3	93.6	85.6	94.9	92.0	80.5	97.3
AVE.		87.9	87.0	85.4	85.2	85.3	82.0	86.3

(KNNs) [43]. The classification accuracies of 13 datasets with different classifiers are listed in Table V. The experimental results contain two parts, i.e., classification accuracies of raw datasets and that of noisy datasets. Here, noise level equals 0% means this is a raw dataset. Noise level equals 5% (or 10%) means there are 5% (or 10%) samples mislabeled artificially. With the total average accuracies, it is concluded both PFRS-PSRC and FRS-PSRC conduct higher classification accuracies than other algorithms, and PFRS-PSRC conducts better performance than FRS-PSRC.

Considering the sensitivity of FRS, we replaced FRS with FVPRS, VPFRS, VQFRS, SFRS, k-mean FRS, and OWA-FRS for RCAs, and these algorithms are denoted by FVPRS based prototype selection for robust classification, VPFRS-PSRC, VQFRS-PSRC, SFRS-PSRC, k-mean FRS-PSRC, and OWA-FRS-PSRC, respectively. The experimental results are listed in Table VI. With the average classification accuracies, it is shown PFRS-PSRC, FRS-PSRC, KNN, VPFRS-PSRC, SFRS-PSRC, k-mean FRS-PSRC, and OWA-FRS-PSRC perform better classification performance than other algorithms.

Furthermore, Fig. 12 shows the average classification accuracies on raw datasets and noisy datasets with Tables V and VI. The numbers $1, 2, \ldots, 13$ are 13 algorithms, i.e., PFRS-PSRC, FRS-PSRC, MLP, LSVM, BayesNet, C4.5, KNN, FVPRS-PSRC, VPFRS-PSRC, VQFRS-PSRC, SFRS-PSRC, k-mean FRS-PSRC, and OWA-FRS-PSRC. With Fig. 12(a), we obtain that PFRS-PSRC, FRS-PSRC, SFRS-PSRC, and OWA-FRS-PSRC produce higher classification accuracies.

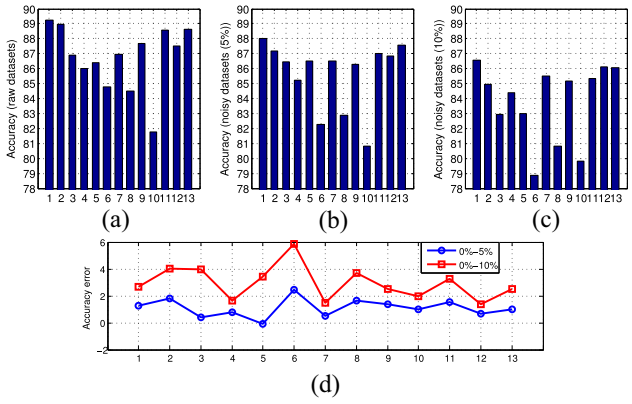


Fig. 12. Average classification accuracies on raw datasets and noisy datasets. (a)–(d) Algorithm index.

With Fig. 12(b) and (c), it is shown that all algorithms are affected by noise, and the classification accuracies descend along with noise increasing. Furthermore, Fig. 12(d) describes the accuracy error between the average accuracies of raw datasets and that of noisy datasets. The figure shows noise produces larger influence on FRS-PSRC than that on other algorithms except C4.5, and all FRS models mentioned are more robust than classical FRS model. Moreover, we also conclude PFRS-PSRC performs the best average classification performance both on raw datasets and noisy datasets.

Furthermore, a statistical test, namely, the Wilcoxon signed-rank test [44], was used to analyze the experimental results.

TABLE VI
ROBUSTNESS COMPARISON OF DIFFERENT FRS MODELS

	NoiseLevel	FVPRS-PSRC	VPFRS-PSRC	VQFRS-PSRC	SFRS-PSRC	k-meanFRS-PSRC	OWA-FRS-PSRC
glass	0%	41.6	68.7	55.0	67.1	66.2	67.2
	5%	41.0	66.9	47.4	66.7	64.4	66.2
	10%	36.8	63.9	44.3	64.4	64.1	64.9
haberman	0%	70.6	71.6	74.5	73.2	71.9	74.9
	5%	66.9	71.6	74.5	73.2	71.6	72.9
	10%	62.4	71.5	73.9	72.3	71.6	72.0
ICU	0%	92.6	94.1	94.1	94.1	94.1	94.6
	5%	91.0	93.6	94.1	94.1	94.1	94.1
	10%	82.3	93.1	93.1	94.1	93.6	94.1
ionosphere	0%	91.6	91.0	83.8	91.5	91.2	91.2
	5%	87.8	89.4	80.7	87.3	90.4	90.4
	10%	87.8	88.7	79.8	85.2	89.3	89.0
iris	0%	96.7	96.0	96.0	98.0	96.0	96.0
	5%	96.7	96.7	96.0	95.3	96.0	96.0
	10%	96.0	96.7	95.8	95.3	96.0	95.3
rice	0%	85.2	83.7	77.3	84.8	84.0	82.4
	5%	85.2	82.9	76.2	84.3	83.9	82.4
	10%	84.8	82.7	76.0	82.2	82.2	80.0
sonar	0%	73.5	83.2	77.4	86.1	84.7	86.6
	5%	72.3	80.3	77.0	79.5	82.8	83.7
	10%	70.4	78.4	74.6	73.6	79.9	79.8
segmentation	0%	82.9	86.2	83.6	89.1	84.8	91.0
	5%	82.6	85.5	82.3	88.3	83.3	90.7
	10%	82.3	85.3	80.1	86.2	82.6	89.8
thyroid	0%	95.4	94.4	87.9	95.8	97.2	95.8
	5%	92.5	94.3	87.6	95.3	96.3	93.9
	10%	91.5	93.8	86.6	93.5	94.7	93.3
WDBC	0%	96.0	96.7	96.7	96.3	96.5	96.7
	5%	95.6	96.2	96.5	95.9	96.4	96.6
	10%	94.9	96.1	96.0	94.9	96.4	96.2
wine	0%	95.5	97.2	92.6	97.2	97.2	98.3
	5%	92.1	96.7	93.9	96.9	96.3	96.9
	10%	90.2	94.6	94.3	95.7	96.2	95.0
WPBC	0%	79.9	79.8	78.3	79.3	78.4	79.9
	5%	78.3	73.3	77.8	78.8	78.3	79.3
	10%	78.3	71.2	77.3	76.8	78.3	77.8
DLBCL	0%	97.3	97.3	66.0	99.0	95.3	97.3
	5%	95.6	94.7	66.3	95.7	95.1	95.6
	10%	93.0	91.2	66.0	95.0	94.6	91.8
AVE.		82.7	86.4	80.8	87.0	86.8	87.4

TABLE VII
WILCOXON SIGNED-RANK TEST RESULTS FOR TABLES V AND VI, $T(38, 0.01) = 2.712$

Algorithms	FRS-PSRC	MLP	LSVM	BayesNet	C4.5	KNN
Test results T	3.7609	3.5655	2.9933	3.8865	5.2541	1.5420
Algorithms	FVPRS-PSRC	VPFRS-PSRC	VQFRS-PSRC	SFRS-PSRC	k-meanFRS-PSRC	OWA-FRS-PSRC
Test results T	4.7238	3.3422	4.5284	3.6911	3.1189	2.8817

The comparison results between PFRS-PSRC and other classification algorithms listed in Tables V and VI are shown in Table VII. Using this method, the test results satisfying $T > T(38, 0.01) = 2.712$ imply that there is a significant difference with significance level $\alpha = 0.01$ between the PFRS-PSRC algorithm and a certain algorithm. It is shown that there are significance differences between PFRS-PSRC and other algorithms except KNN. The above analysis shows PFRS-PSRC performs better classification performance than other algorithm except KNN.

VI. CONCLUSION

In this paper, we proposed a novel robust FRS model: PFRS, which considers data distribution as important information in computing lower and upper approximations and a prototype selection method is proposed, a prototype-based RCA is designed. The main conclusions of this paper can be summarized as follows.

1) We showed that PFRS is robust to noise because it ignores samples that are identified as noise by achieving a tradeoff between similarity and class PD. Thus, the distribution information of samples is considered when developing the PFRS model.
2) We designed a PSA and an RCA based on PFRS.
3) Extensive experiments were conducted to test the robustness of the proposed model and algorithms. The robustness of PFRS was tested and compared with other models. The experimental results showed that PFRS is robust to noise and the proposed classification algorithm is more effective than some state-of-the-art algorithms.

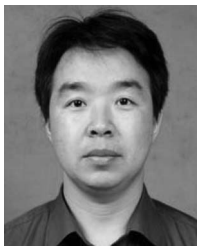
REFERENCES

[1] Z. Pawlak, "Rough sets," *Int. J. Comput. Inf. Sci.*, vol. 11, no. 5, pp. 341–356, 1982.
[2] J. H. Dai, W. T. Wang, and Q. Xu, "An uncertainty measure for incomplete decision tables and its applications," *IEEE Trans. Cybern.*, vol. 43, no. 4, pp. 1277–1289, Aug. 2013.
[3] D. Dubois and H. Prade, "Rough fuzzy sets and fuzzy rough sets," *Int. J. Gen. Syst.*, vol. 17, nos. 2–3, pp. 191–209, 1990.
[4] D. G. Chen, L. Zhang, S. Y. Zhao, Q. Hu, and P. Zhu, "A novel algorithm for finding reducts with fuzzy rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 2, pp. 385–389, Apr. 2012.
[5] P. Maji and P. Garai, "Fuzzy-rough simultaneous attribute selection and feature extraction algorithm," *IEEE Trans. Cybern.*, vol. 43, no. 4, pp. 1166–1177, Aug. 2013.

[6] S. An, H. Shi, Q. H. Hu, X. Q. Li, and J. W. Dang, "Fuzzy rough regression with application to wind speed prediction," *Inf. Sci.*, vol. 282, pp. 388–400, Oct. 2014.
[7] S. Y. Zhao, E. C. C. Tsang, D. G. Chen, and X. Z. Wang, "Building a rule-based classifier by using fuzzy rough set technique," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 5, pp. 624–638, May 2010.
[8] R. Diao, F. Chao, T. Peng, N. Snooke, and Q. Shen, "Feature selection inspired classifier ensemble reduction," *IEEE Trans. Cybern.*, vol. 44, no. 8, pp. 1259–1268, Aug. 2014.
[9] J. T. Yao, A. V. Vasilakos, and W. Pedrycz, "Granular computing: Perspectives and challenges," *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1977–1989, Dec. 2013.
[10] A. Mieszkowicz-Rolka and L. Rolka, "Variable precision fuzzy rough sets," in *Transactions on Rough Sets I* (LNCS 3100). Berlin, Germany: Springer, pp. 144–160, 2004.
[11] P. Maji and P. Garai, "IT2 fuzzy-rough sets and max relevance-max significance criterion for attribute selection," *IEEE Trans. Cybern.*, vol. 45, no. 8, pp. 1657–1668, Aug. 2015.
[12] Y. Y. Yao, S. K. M. Wong, and P. Lingras, "A decision-theoretic rough set model," in *Methodologies for Intelligent Systems*. New York, NY, USA: North-Holland, 1990, pp. 17–24.
[13] S. Y. Zhao, E. C. C. Tsang, and D. G. Chen, "The model of fuzzy variable precision rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 2, pp. 451–467, Apr. 2009.
[14] D. Slezak and W. Ziarko, *Variable Precision Bayesian Rough Set Model* (LNCS 2639). Berlin, Germany: Springer, pp. 312–315, 2003.
[15] Y. H. Qian, H. Zhang, Y. L. Sang, and J. Y. Liang, "Multigranulation decision-theoretic rough sets," *Int. J. Approx. Reason.*, vol. 55, no. 1, pp. 225–237, Jan. 2014.
[16] N. Manwani and P. S. Sastry, "Noise tolerance under risk minimization," *IEEE Trans. Cybern.*, vol. 43, no. 3, pp. 1146–1151, Jun. 2013.
[17] J. M. F. Salido and S. Murakami, "Rough set analysis of a general type of fuzzy data using transitive aggregations of fuzzy similarity relations," *Fuzzy Sets Syst.*, vol. 139, no. 3, pp. 635–660, 2003.
[18] C. Cornelis, M. De Cock, and A. M. Radzikowska, *Vaguely Quantified Rough Sets* (LNAI 4482). Berlin, Germany: Springer, 2007, pp. 87–94.
[19] N. Verbiest, C. Cornelis, and F. Herrera, "OWA-FRPS: A prototype selection method based on ordered weighted average fuzzy rough set theory," in *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*. Berlin, Germany: Springer, pp. 180–190, 2013.
[20] C. Cornelis, N. Verbiest, and R. Jensen, "Ordered weighted average based fuzzy rough sets," in *Proc. 5th Int. Conf. Rough Sets Knowl. Technol.*, Beijing, China, 2010, pp. 78–85.
[21] N. Verbiest, C. Cornelis, and F. Herrera, "FRPS: A fuzzy rough prototype selection method," *Pattern Recognit.*, vol. 46, no. 10, pp. 2770–2782, 2013.
[22] Q. H. Hu, S. An, and D. R. Yu, "Soft fuzzy rough sets for robust feature evaluation and selection," *Inf. Sci.*, vol. 180, no. 22, pp. 4384–4400, 2010.
[23] Q. H. Hu, L. Zhang, S. An, D. Zhang, and D. R. Yu, "On robust fuzzy rough set models," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 4, pp. 636–651, Aug. 2012.
[24] S. An, Q. H. Hu, D. R. Yu, and J. F. Liu, "Soft minimum-enclosing-ball based robust fuzzy rough sets," *Fundam. Informat.*, vol. 115, nos. 2–3, pp. 189–202, 2012.
[25] W. M. Ma and B. Z. Sun, "Probabilistic rough set over two universes and rough entropy," *Int. J. Approx. Reason.*, vol. 53, no. 4, pp. 608–619, 2012.
[26] B. Z. Sun and W. M. Ma, "Soft fuzzy rough sets and its application in decision making," *Artif. Intell. Rev.*, vol. 41, no. 1, pp. 67–80, 2014.
[27] Y. Q. Yao, J. S. Mi, and Z. J. Li, "A novel variable precision (θ, σ)-fuzzy rough set model based on fuzzy granules," *Fuzzy Sets Syst.*, vol. 236, pp. 58–72, Feb. 2014.
[28] Q. H. Hu, S. G. Zhang, Z. X. Xie, J. S. Mi, and J. Wang, "Noise model based v-support vector regression with its application to short-term wind speed forecasting," *Neural Netw.*, vol. 57, pp. 1–11, Sep. 2014.
[29] Q. H. Hu, D. R. Yu, Z. X. Xie, and J. Liu, "Fuzzy probabilistic approximation spaces and their information measures," *IEEE Trans. Fuzzy Syst.*, vol. 14, no. 2, pp. 191–201, Apr. 2006.
[30] J.-S. Mi and W.-X. Zhang, "An axiomatic characterization of a fuzzy generalization of rough sets," *Inf. Sci.*, vol. 160, nos. 1–4, pp. 235–249, 2004.
[31] D. S. Yueng, D. G. Chen, E. C. C. Tsang, J. W. T. Lee, and X. Wang, "On the generalization of fuzzy rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 3, pp. 343–361, Jun. 2005.
[32] G. J. Klir and B. Yuan, *Fuzzy Logic: Theory and Applications*. Upper Saddle River, NJ, USA: Prentice-Hall, 1995.
[33] N. N. Morsi and M. M. Yakout, "Axiomatics for fuzzy rough sets," *Fuzzy Sets Syst.*, vol. 100, nos. 1–3, pp. 327–342, 1998.
[34] W.-Z. Wu and W.-X. Zhang, "Constructive and axiomatic approaches of fuzzy approximation operators," *Inf. Sci.*, vol. 159, nos. 3–4, pp. 233–254, 2004.
[35] Y. Y. Yao and S. K. M. Wong, "A decision theoretic framework for approximating concepts," *Int. J. Man-Mach. Stud.*, vol. 37, no. 6, pp. 793–809, 1992.
[36] Z. Pawlak and A. Skowron, "Rough membership functions," in *Advances in the Dempster-Shafer Theory of Evidence*. New York, NY, USA: Wiley, pp. 251–271, 1994.
[37] Y. Y. Yao, "Two semantic issues in a probabilistic rough set model," *Fundam. Informat.*, vol. 108, nos. 3–4, pp. 249–265, 2011.
[38] M. Lichman, *UCI Machine Learning Repository*, School Inf. Comput. Sci., Univ. California, Irvine, CA, USA, 2013. [Online]. Available: http://archive.ics.uci.edu/ml
[39] G. J. Gibson and C. F. N. Cowan, "On the decision regions of multilayer perceptrons," *Proc. IEEE*, vol. 78, no. 10, pp. 1590–1594, Oct. 1990.
[40] O. Chapelle, P. Haffner, and V. N. Vapnik, "Support vector machines for histogram-based image classification," *IEEE Trans. Neural Netw.*, vol. 10, no. 5, pp. 1055–1064, Sep. 1999.
[41] J. Li *et al.*, "Bayes net classifiers for prediction of renal graft status and survival period," *Int. J. Med. Med. Sci.*, vol. 1, no. 4, pp. 215–221, 2010.
[42] J. R. Quinlan, *C4.5: Programs for Machine Learning*, vol. 16. San Francisco, CA, USA: Morgan Kaufmann, 1993, pp. 235–240.
[43] D. W. Aha, D. Kibler, and M. K. Albert, "Instance-based learning algorithms," *Mach. Learn.*, vol. 6, no. 1, pp. 37–66, 1991.
[44] F. Wilcoxon, *Some Rapid Approximate Statistical Procedures*. New York, NY, USA: Amer. Cyanamid, 1945, p. 13.

Shuang An received the B.S. degree from Shenyang Normal University, Shenyang, China, in 2005, the M.S. degree from Northeastern University, Shenyang, in 2008, and the Ph.D. degree from the Harbin Institute of Technology, Harbin, China, in 2011.

She is currently with Northeastern University at Qinhuangdao, Qinhuangdao, China. Her current research interests include robust machine learning and pattern recognition.



Qinghua Hu (SM'13) received the B.S., M.S., and Ph.D. degrees from the Harbin Institute of Technology, Harbin, China, in 1999, 2002, and 2008, respectively.

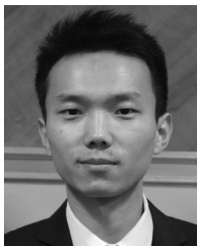
He was a Post-Doctoral Fellow with the Department of Computing, Hong Kong Polytechnic University, Hong Kong, from 2009 to 2011. He is currently a Full Professor and the Vice Dean of the School of Computer Science and Technology, Tianjin University, Tianjin, China. He has authored over 100 journal and conference papers in the areas of granular computing-based machine learning, reasoning with uncertainty, pattern recognition, and fault diagnosis. His current research interests include rough sets, granular computing, and data mining for classification and regression.

Prof. Hu was the Program Committee Co-Chair of the International Conference on Rough Sets and Current Trends in Computing in 2010, the Chinese Rough Set and Soft Computing Society in 2012 and 2014, and the International Conference on Rough Sets and Knowledge Technology and the International Conference on Machine Learning and Cybernetics in 2014, and the General Co-Chair of IJCRS 2015. He serves as a Referee for a great number of journals and conferences.

Witold Pedrycz (F'99) received the M.Sc., Ph.D., and the D.Sci. degrees from the Silesian University of Technology, Gliwice, Poland.

He is a Professor and the Canada Research Chair of Computational Intelligence with the Department of Electrical and Computer Engineering, Faculty of Engineering, King Abdulaziz University, Jeddah, Saudi Arabia. He is also with the Systems Research Institute of the Polish Academy of Sciences, Warsaw, Poland. He holds an appointment of special professorship with the School of Computer Science, University of Nottingham, Nottingham, U.K. His current research interests include computational intelligence, fuzzy modeling and granular computing, knowledge discovery and data mining, fuzzy control, pattern recognition, knowledge-based neural networks, relational computing, and software engineering. He has published numerous papers in the above areas. He has also authored 15 research monographs covering various aspects of computational intelligence, data mining, and software engineering.

Dr. Pedrycz was a recipient of the prestigious Norbert Wiener Award from the IEEE Systems, Man, and Cybernetics Society in 2007, the IEEE Canada Computer Engineering Medal 2008, the Cajastur Prize for Soft Computing from the European Centre for Soft Computing for "pioneering and multifaceted contributions to granular computing" in 2009, the Killam Prize in 2013, and the Fuzzy Pioneer Award 2013 from the IEEE Computational Intelligence Society. He was elected as a Foreign Member of the Polish Academy of Sciences in 2009 and a fellow of the Royal Society of Canada in 2012. He has been a member of numerous program committees of the IEEE conferences in the area of fuzzy sets and neurocomputing.



Pengfei Zhu received the B.S. and M.S. degrees from the Harbin Institute of Technology, Harbin, China, in 2009 and 2011, respectively, and the Ph.D. degree from Hong Kong Polytechnic University, Hong Kong, in 2014.

He is currently an Associate Professor with the School of Computer Science and Technology, Tianjin University, Tianjin, China. His current research interests include machine learning and computer vision.

Eric C. C. Tsang received the B.S. degree in computer studies from the City University of Hong Kong, Hong Kong, in 1990, and the Ph.D. degree in computing from Hong Kong Polytechnic University, Hong Kong, in 1996.

He is an Associate Professor with the Faculty of Information Technology, Macau University of Science and Technology, Macau, China. His current research interests include fuzzy systems, rough sets, fuzzy rough sets, and multiple classifier systems.