IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 16, 2020, accepted July 5, 2020, date of publication July 20, 2020, date of current version July 30, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3010314

Multi-Label Attribute Reduction Based on Variable Precision Fuzzy Neighborhood Rough Set

PANPAN CHEN[1,2], MENGLEI LIN[1,2], AND JINGHUA LIU[2,3]

[1]School of Mathematics and Statistics, Minnan Normal University, Zhangzhou 363000, China
[2]Institute of Meteorological Big Data-Digital Fujian, Minnan Normal University, Zhangzhou 363000, China
[3]College of Computer Science and Technology, Huaqiao University, Xiamen 361000, China

Corresponding authors: Menglei Lin (menglei36@126.com) and Jinghua Liu (zzliujinghua@163.com)

This work was supported in part by the National Youth Science Foundation of China under Grant 61603173, in part by the Natural Science Foundation of Fujian Province under Grant 2018J01422, and in part by the Institute of Meteorological Big Data-Digital Fujian and the Fujian Key Laboratory of Data Science and Statistics, Minnan Normal University, under Grant 42010701.

ABSTRACT Multi-label attribute reduction as a common dimensionality reduction technique has obtained widely research in recent years. Most existing multi-label attribute reduction methods adopt discretization to deal with mixed data and have strict requirements on the condition of sample classification. However, the process of discretization may lead to information loss, moreover, strict conditions will increase the possibility of a sample classified into a wrong class. Based on this, we construct a multi-label attribute reduction method based on variable precision fuzzy neighborhood rough set. The main motivation is that the variable precision fuzzy neighborhood rough set can process multiple types of data without discretization and tolerate noisy data. Specifically, we first use the parameterized fuzzy neighborhood granule to define the fuzzy decision and decision class of each sample under different labels. Then, the fuzzy decision and decision classes under different labels are fused into the entire multi-label learning space. Finally, a multi-label attribute reduction algorithm is designed according to the defined maximum attribute significance criterion. Our experiments are conducted on a series of multi-label datasets, and the experimental results verify that the proposed algorithm achieves better classification performance than other state-of-the-art comparison algorithms.

INDEX TERMS Attribute reduction, multi-label learning, rough set model, fuzzy neighborhood, variable precision.

I. INTRODUCTION

Multi-label learning is ubiquitous in multiple real-world applications, such as image automatic annotation [32], [40], text categorization [12], [29], and functional genomics [4], [44]. For instance, an image may have multiple semantics, including lakes, forest, and landscape; a document may belong to economy, politics, and sport.

As we know, multi-label data in reality always contains a large amount of attributes [7], [11], [37], [38], [41], [50], and some attributes may be irrelevant and/or redundant, which will severely interfere with the classification performance of multi-label classifier. Attribute reduction, as an important means for data pre-processing, can effectively solve the curse of dimensionality. Generally, there are three main methods in multi-label attribute reduction, that is embedded, wrapper, and filter [26]–[28]. The embedded method combines a specific learning algorithm to conduct attribute reduction in the training process, such as Correlated Multi-Label Feature Selection (CMLFS) [8], Multi-Label Feature Selection (MLFS) [8], and Correlated Label Rank SVM (CLaRank SVM) [8]. The wrapper method uses a predetermined multi-label classifier to evaluate attribute subsets and representative algorithms including feature selection for Multi-Label Naive Bayes (MLNB) [47], Binary Relevance Random Forest (BRRF) [9], and Random Forest Label Power-set (RFLP) [9]. The filter method chooses attributes by using attribute evaluation criteria, including dependency [16], [45], [46], distance metric [24], [25], information metric [13], [14], [17], [39], and rough set [18], [20], [30], [42].

Rough set theory, proposed by Pawlak [23], has been proved to be an effective attribute reduction tool due to

The associate editor coordinating the review of this manuscript and approving it for publication was Zhan Bu.

its understandability and high approximate ability. Attribute reduction based on rough set can be roughly classified as the attribute reduction algorithm based on the positive region [15], the attribute reduction algorithm based on discernibility matrix and its improvement [49], and the attribute reduction algorithm based on heuristic information entropy [22], [35]. However, the classical rough set theory has one main defect, that is, this model is defined on the stringent equivalence relation, so it can not work effectively on the hybrid attributes in classification learning. Therefore, many extensive models of rough set have been proposed [6], [10], [21], [34], [36], which provide an important theoretical tool for attribute reduction. For instance, Lin [19] proposed a hypothesis that samples with the similar attribute values should be classified into the same class or neighborhood class. Based on this hypothesis, Hu *et al.* [10] extended the equivalent relation into neighborhood relation, and then presented a neighborhood rough set (NRS). Unfortunately, this model be unable to characterize the samples' fuzziness in fuzzy conditions. Then, Dubois and Prade [3] used fuzzy similarity relation to construct a fuzzy rough set model (FRS), which combines rough set and fuzzy set. But fuzzy rough set is sensitive to noise. These noisy data will affect the computation of fuzzy lower approximation and astrict their practical utilization. To solve the above problems, Wang *et al.* [33] defined the fuzzy decision and decision class of samples by combining the neighborhood rough set and fuzzy rough set, and then constructed a fuzzy neighborhood rough set model (FNRS), which overcomes the shortcoming of fuzzy rough set model. Meanwhile, the fuzzy neighborhood rough set model can better describe the sample decisions by using fuzzy information granules, thereby reducing the classification error rate. Due to the fuzzy neighborhood rough set is too strict to tolerate the noisy data, they also proposed a variable precision fuzzy neighborhood rough set to overcome this drawback. These models are more effective to deal with the uncertainty of numerical data and noise data, but they cannot be used to multi-label learning directly. Different from single-label learning, each sample may be related to multiple labels in multi-label learning, and the label space is a closely related whole. In case of the integrity of label space, it is a severe challenge to construct fuzzy decision and decision class.

Inspired by this, we generalize variable precision fuzzy neighborhood rough set in single-label learning to fit multi-label learning, and introduce a novel multi-label attribute reduction model based on variable precision fuzzy neighborhood rough set. The proposed method uses fuzzy relations to estimate the similarity among samples under different labels, and evaluates the quality of attributes on multi-label data. Meanwhile, this method can reduce the interference of noise data and avoid the incorrect classification of samples caused by the strictness of reduction model. In order to exactly calculate the lower and upper approximation operators of multi-label variable precision neighborhood rough set, we first take a large function and small function to define the fuzzy decision of a sample in a multi-label fuzzy neighborhood decision system, and define the decision classes in the whole label space. Then, we employ the threshold to construct the multi-label approximation space. In addition, we present an optimization objective function, which can be use to evaluate the quality of the candidate attributes. Finally, extensive experimental results display that the proposed multi-label attribute reduction algorithm can reduce the dimension of attribute space, and highly improve the classification performance of a multi-label classifier. The major contributions of this study are summarized below:

- A multi-label variable precision fuzzy neighborhood rough set model is proposed, which can process high-dimensional multi-label data and solve the restriction problem of data type.
- The proposed model considers the integrity of label space.
- FNRS-ML achieves highly classification performance than some other state-of-the-art multi-label attribute reduction algorithms in our experiments.

The remaining of this paper is formed as follows. In Section II, we introduce multi-label learning and variable precision fuzzy neighborhood rough set. In Section III, we redefine fuzzy decision and decision classes for multi-label learning in variable precision fuzzy neighborhood rough and give our multi-label attribute reduction algorithm. Then, we show the experimental results on several multi-label datasets in Section IV, and at last Section V is our conclusions.

II. PRELIMINARIES

A. MULTI-LABEL LEARNING

Suppose $X = R^{n\times d}$ be the input space of n samples, $x_i = [x_{i1}, x_{i2}, \ldots, x_{id}] \in X$ is a d-dimensional attribute vector, and $L = \{l_1, l_2, \ldots l_m\}$ is a set of m labels. Each sample x_i is related to a subset of L, and the subset can be represented as a m dimensional vector $y_i = [y_i^1, y_i^2, \ldots, y_i^m]$, where $y_i^j = 1$ only if x_i is labeled with label l_j, otherwise $y_i^j = 0$.

Currently, several popular evaluation metrics are designed from different evaluation perspectives in multi-label learning. In our experiment, we choose Average precision, Ranking Loss, Hamming Loss, and One-error as evaluation metrics [43], [47]. Let $T = \{(x_i, Y_i)|1 \le i \le N\}$ is a testing set, where $Y_i \subseteq L$ is a correct label set and $Y_i' \subseteq L$ is expressed as a binary predicted label vector. In the following, the four evaluation metrics are introduced.

(1) **Average Precision** evaluates the average probability of predictive labels ordered before a true relevant label $y \in Y_i$ that do belong to Y_i.

$$avgPre(f) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{|Y_i|} \cdot \sum_{y\in Y_i}\frac{|\{y' \in Y_i : rank(x_i, y') \le rank(x_i, y)\}|}{rank(x_i, y)}. \quad (1)$$

(2) **Ranking Loss** evaluates the average probability of label pairs that are not sorted correctly for one sample,

which is expressed as

$$rLoss(f) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{|Y_i||\overline{Y_i}|}|\{(y_1, y_2)|f(x_i, y_1) \geq f(x_i, y_2), (y_1, y_2) \in Y_i \times \overline{Y_i}\}|, \quad (2)$$

where $\overline{Y_i}$ indicates the complementary set of Y_i.

(3) **Hamming Loss** evaluates the times of misclassified sample-label pairs.

$$hLoss(h) = \frac{1}{N}\sum_{i=1}^{N}\frac{|Y_i' \oplus y_i|}{M}, \quad (3)$$

where $\oplus$ denotes the XOR operation.

(4) **One-error** evaluates the proportion of test samples whose first ranked label is not in a set of relevant labels.

$$oneError(f) = \frac{1}{N}\sum_{i=1}^{N}[\![[\arg\max_{y\in L} f(x_i, y)] \notin Y_i']\!], \quad (4)$$

where for any predicate π holds, then $[\![\pi]\!]$ equals 1; Otherwise 0.

For these evaluation metrics, Hamming Loss concerns with the label set prediction, and the other three evaluation metrics more focus on the label ranking. Note that for Average Precision, bigger value indicates the better the method's performance, and the best value is 1. While for Ranking Loss, Hamming Loss, and One-error, smallervaluedemonstrates the better the method's performance, and the optimal value is 0.

B. VARIABLE PRECISION FUZZY NEIGHBORHOOD ROUGH SET

Given a fuzzy neighborhood decision system, which denoted by $NDT = \langle U, C, D\rangle$. $U = \{x_1, x_2, \ldots, x_n\}$ is a nonempty set of samples, C is an attribute set of real-valued to characterize samples, D is a decision attribute. Suppose the universe be divided into r equivalence classes by D, which is expressed as $U/D = \{D_1, D_2, \ldots, D_r\}$.

Definition 1 [33]: Assume that $NDT = \langle U, C, D\rangle$, $B \subseteq C$, and B can induce a fuzzy binary relation R_B, then $R_B(x, y)$ is a fuzzy similarity relation if:

(1) *Reflectivity* : $R_B(x, x) = 1, \forall x \in U$;

(2) *Symmetry* : $R_B(x, y) = R_B(y, x), \forall x, y \in U$.

Suppose R_a is a fuzzy similarity relation for any $a \in B$, then we can express $R_B = \bigcap_{a\in B} R_a$. The fuzzy neighborhood granule for any $x \in U$ is defined as

$$[x]_B^{\delta}(y) = \begin{cases} 0, & R_B(x, y) < \delta; \\ R_B(x, y), & R_B(x, y) \geq \delta. \end{cases} \quad (5)$$

The parameter δ is the fuzzy neighborhood radius of samples, and it satisfies $0 < \delta \leq 1$. Obviously, the membership degree of fuzzy neighborhood granule is determined by the two factors δ and B.

Definition 2 [33]: Given a fuzzy neighborhood decision system $NDT = \langle U, C, D\rangle$, $U/D = \{D_1, D_2, \ldots, D_r\}$, and R_C is a fuzzy similarity relation generated by C. For $\forall x \in U$, the fuzzy decision of sample x is defined as

$$\tilde{D}_i(x) = \frac{|[x]_C \cap D_i|}{|[x]_C|}, \quad i = 1, 2 \ldots r, \quad (6)$$

where $\tilde{D}_i$ denotes the membership degree of each sample x to D_i, and it is a fuzzy set. $\{\tilde{D}_1, \tilde{D}_2, \ldots, \tilde{D}_r\}$ is the set of fuzzy decision of samples, which is generated by decision attribute D.

Definition 3 [33]: Assume A and B are two fuzzy sets, the inclusion degree between the two fuzzy sets can be defined as

$$I(A, B) = \frac{|A \subseteq B|}{|U|}, \quad (7)$$

where $I(A, B)$ represents the inclusion degree of A in B, and $|A \subseteq B|$ indicates the number of samples whose membership degrees to the fuzzy set A are not greater than those to the fuzzy set B.

Definition 4 [33]: Given a fuzzy neighborhood decision system $NDT = \langle U, C, D\rangle$, $B \subseteq C$, $U/D = \{D_1, D_2, \ldots, D_r\}$, R_B is the fuzzy similarity relation on U induced by B, $\{\tilde{D}_1, \tilde{D}_2, \ldots, \tilde{D}_r\}$ are the fuzzy decisions induced by D, the variable precision fuzzy neighborhood approximation space of decision attribute D with respect to B is defined as

$$\underline{R}_B^{\delta,\alpha}(D) = \{\underline{R}_B^{\delta,\alpha}(\tilde{D}_1), \underline{R}_B^{\delta,\alpha}(\tilde{D}_2) \ldots \underline{R}_B^{\delta,\alpha}(\tilde{D}_r)\}, \quad (8)$$

$$\overline{R}_B^{\delta,\beta}(D) = \{\overline{R}_B^{\delta,\beta}(\tilde{D}_1), \overline{R}_B^{\delta,\beta}(\tilde{D}_2) \ldots \overline{R}_B^{\delta,\beta}(\tilde{D}_r)\}, \quad (9)$$

where

$$\underline{R}_B^{\delta,\alpha}(\tilde{D}_i) = \{x_i \in D_i | I([x_i]_B^{\delta}, \tilde{D}_i) \geq \alpha\}, \ 0.5 \leq \alpha \leq 1, \quad (10)$$

$$\overline{R}_B^{\delta,\beta}(\tilde{D}_i) = \{x_i \in D_i | I([x_i]_B^{\delta}, \tilde{D}_i) > \beta\}, \ 0 \leq \beta < 0.5. \quad (11)$$

The lower approximation of the decision system is defined as the union of the lower approximation of each decision class, expressed as, $\underline{R}_B^{\delta,\alpha}(D) = \cup_{i=1}^{r}\underline{R}_B^{\delta,\alpha}(\tilde{D}_i)$. Similarly, $\overline{R}_B^{\delta,\beta}(D) = \cup_{i=1}^{r}\overline{R}_B^{\delta,\beta}(\tilde{D}_i)$. $\underline{R}_B^{\delta,\alpha}(D)$ is also called the variable precision fuzzy positive region, denoted by $POS_B^{\delta,\alpha}(D)$. Concretely, $POS_B^{\delta,\alpha}(D)$ is a set of samples that satisfy the threshold of inclusion degree and consistent with classification.

Definition 5 [33]: Given a fuzzy neighborhood decision system $NDT = \langle U, C, D\rangle$, $B \subseteq C$, the variable precision fuzzy dependency of D to conditional attribute set B is defined as

$$\gamma_B^{\delta,\alpha}(D) = \frac{|POS_B^{\delta,\alpha}(D)|}{|U|}, \quad (12)$$

$\gamma_B^{\delta,\alpha}(D)$ reflects the proportion of samples in the universe that can be classified accurately. Apparently, $0 \leq \gamma_B^{\delta,\alpha}(D) \leq 1$. The larger the positive region is, the stronger the ability of condition attributes B to describe D is.

Lemma 1 [33]: For a fuzzy neighborhood decision system $NDT = \langle U, C, D\rangle$, with two parameters of δ and α, if $B_1 \subseteq B_2 \subseteq C$, then we have

(1) $POS_{B_1}^{\delta,\alpha}(D) \subseteq POS_{B_2}^{\delta,\alpha}(D)$;

(2) $\gamma_{B_1}^{\delta,\alpha}(D) \leq \gamma_{B_2}^{\delta,\alpha}(D)$.

Definition 6 [33]: Given $NDT = \langle U, C, D\rangle$ and a fuzzy neighborhood radius δ, $B \subseteq C$, if attribute subset B satisfies:

(1) $\gamma_B^{\delta,\alpha}(D) = \gamma_C^{\delta,\alpha}(D)$;

(2) $\forall a \in B, \gamma_{B-a}^{\delta,\alpha}(D) < \gamma_B^{\delta,\alpha}(D)$.

Specially, we can say that B is a variable precision fuzzy reduct.

III. VARIABLE PRECISION FUZZY NEIGHBORHOOD ROUGH SET IN MULTI-LABEL LEARNING

A. FUZZY DECISION AND DECISION CLASSES

Variable precision fuzzy neighborhood rough set model, proposed by Wang *et al.* [33], can describe the fuzziness of samples under fuzzy background, and reduce the influence of noisy data on the classification model. Although the model can effective to deal with the uncertainty of numerical and noise data. However, it only be used to single-label learning. Recently, many multi-label attribute reduction methods based on fuzzy rough set have been presented and discussed, but their common characteristic is that these algorithms handle multi-label data by transforming multi-label data set into single-label data set, which ignore the integrality of label space and loses some important information. Moreover, the calculation of fuzzy decision and decision class for multi-label learning are somewhat complicated as each object is related to multiple class labels simultaneously. Therefore, considering the whole information of label space, we present a multi-label variable precision fuzzy neighborhood rough set method by using the large and small function to fuse the fuzzy decision and decision class of samples in multi-label learning.

Definition 7: Given a non-empty set of samples $U = \{x_1, x_2, \ldots, x_n\}$, an attribute set is C, label set is $L = \{l_1, l_2, \ldots, l_m\}$, C can induce a fuzzy similarity relation R, and we then call $MNDT =< U, R, L >$ as a multi-label fuzzy neighborhood decision system, which usually recorded as $MNDT =< U, C \cup L >$.

Definition 8: The fuzzy similarity degree r_{ij} between multi-label samples x_i and x_j relative to an attribute a can be calculated by

$$r_{ij} = \begin{cases} 1 - |x_i - x_j|, & |x_i - x_j| \le 1 - \delta, \\ 0, & |x_i - x_j| > 1 - \delta. \end{cases}$$

Here, when $r_{ij} = r_{ji}$ and $0 \le r_{ij} < 1$, the matrix $R_a^{\delta} = (r_{ij})_{n\times n}$ is a fuzzy similarity relation.

Definition 9: Given $MNDT =< U, C \cup L >$, $L = \{l_1, l_2, \ldots, l_m\}$ is label set, $\{L_1^{l_j}, L_2^{l_j}, \ldots, L_r^{l_j}\}$ is a set of mutually exclusive decision classes generated by l_j, and R_C is a fuzzy similarity relation. For $\forall x \in U$, the fuzzy decision of sample x in decision system can be defined as

$$\widetilde{L}_i(x) = \vee_{j=1}^{m} \frac{|[x]_C \cap L_i^{l_j}|}{|[x]_C|}, \quad i = 1. \tag{13}$$

$$\widetilde{L}_i(x) = \wedge_{j=1}^{m} \frac{|[x]_C \cap L_i^{l_j}|}{|[x]_C|}, \quad i = 2, 3, \ldots r. \tag{14}$$

where, $[x]_C$ is obtained by R_C, R_C is a fuzzy similarity matrix, and the values of r_{ij} in the matrix is calculated by Definition 3.2. For multi-label learning, the membership degree of each sample under different labels to each decision class is different, namely, the fuzzy decision of each sample is also different under each label. Therefore, we take

TABLE 1. Example of multi-label data.

U	a_1	a_2	a_3	a_4	l_1	l_2	l_3
x_1	0.14	0.08	0.23	0.60	-1	1	-1
x_2	0.40	0.35	0.17	0.85	1	-1	1
x_3	0.56	0.36	0.15	0.79	1	-1	-1
x_4	0.18	0.24	0.25	0.44	-1	-1	-1
x_5	0.36	0.15	0.10	0.13	-1	-1	-1

Eqs. (13)-(14) to fuse the fuzzy decision of samples under each label in the decision system. Considering the integrity of the label space, we need to redefine the decision classes of the decision system.

Definition 10: Given $MNDT =< U, C \cup L >$, $L = \{l_1, l_2, \ldots, l_m\}$ is label set, l_j generates the decision classes $\{L_1^{l_j}, L_2^{l_j}, \ldots, L_r^{l_j}\}$, then the decision classes in multi-label decision system are obtained:

$$L_i = \bigcup_{j=1}^{m} L_i^{l_j} - L_{i-1} - L_{i-2} - \cdots - L_1, \ i = 1, 2, \ldots, r. \tag{15}$$

The right side of (15) expresses the difference operation of the set from left to right according to the subtraction operation rule, where the symbol of minus represents the relative complement set between two sets. Such as, $A - B$ denotes a collection of all elements belonging to set A but not to B, that is, the set of elements in A that are left after removing all the elements that belong to set B. Meanwhile, L_i is a partition of the universe. We describe the process by Example 3.1 to show the definition of fuzzy decision and decision classes more clearly and specifically.

Example 1: Given a multi-label data in TABLE 1. There are three labels l_1, l_2 and l_3, $B = \{a_1, a_2, a_3, a_4\}$, and B induces the fuzzy similarity relation R_B. Here $\delta = 0.4$, we can obtain

$$R_B = \begin{bmatrix} 1 & 0.55 & 0.45 & 0.77 & 0.46 \\ 0.55 & 1 & 0.83 & 0.51 & 0 \\ 0.45 & 0.83 & 1 & 0.46 & 0 \\ 0.77 & 0.51 & 0.46 & 1 & 0.60 \\ 0.46 & 0 & 0 & 0.60 & 1 \end{bmatrix}$$

Accordingly, the fuzzy decision and decision classes under the labels l_1, l_2 and l_3 are as follows:

$$L_1^{l_1} = \{x_2, x_3\}, \quad L_2^{l_1} = \{x_1, x_4, x_5\}.$$
$$\widetilde{L}_1^{l_1} = \frac{0.31}{x_1} + \frac{0.63}{x_2} + \frac{0.67}{x_3} + \frac{0.30}{x_4} + \frac{0}{x_5},$$
$$\widetilde{L}_2^{l_1} = \frac{0.69}{x_1} + \frac{0.37}{x_2} + \frac{0.33}{x_3} + \frac{0.70}{x_4} + \frac{1}{x_5}.$$
$$L_1^{l_2} = \{x_1\}, \quad L_2^{l_2} = \{x_2, x_3, x_4, x_5\}.$$
$$\widetilde{L}_1^{l_2} = \frac{0.31}{x_1} + \frac{0.19}{x_2} + \frac{0.17}{x_3} + \frac{0.23}{x_4} + \frac{0.22}{x_5},$$
$$\widetilde{L}_2^{l_2} = \frac{0.69}{x_1} + \frac{0.81}{x_2} + \frac{0.83}{x_3} + \frac{0.77}{x_4} + \frac{0.78}{x_5}.$$
$$L_1^{l_3} = \{x_2\}, \quad L_2^{l_3} = \{x_1, x_3, x_4, x_5\}.$$
$$\widetilde{L}_1^{l_3} = \frac{0.17}{x_1} + \frac{0.35}{x_2} + \frac{0.30}{x_3} + \frac{0.15}{x_4} + \frac{0}{x_5},$$
$$\widetilde{L}_2^{l_3} = \frac{0.83}{x_1} + \frac{0.65}{x_2} + \frac{0.70}{x_3} + \frac{0.85}{x_4} + \frac{1}{x_5}.$$

Then, that's from (13)-(14), the fuzzy decision of the entire label space can be given as follows:

$$\widetilde{L}_1 = \widetilde{L}_1^{l_1} \vee \widetilde{L}_1^{l_2} \vee \widetilde{L}_1^{l_3} = \frac{0.31}{x_1} + \frac{0.63}{x_2} + \frac{0.67}{x_3} + \frac{0.30}{x_4} + \frac{0.22}{x_5},$$
$$\widetilde{L}_2 = \widetilde{L}_2^{l_1} \wedge \widetilde{L}_2^{l_2} \wedge \widetilde{L}_2^{l_3} = \frac{0.69}{x_1} + \frac{0.37}{x_2} + \frac{0.33}{x_3} + \frac{0.70}{x_4} + \frac{0.78}{x_5}.$$

By above knowable, $\forall x \in U$, $\widetilde{L}_1(x) + \widetilde{L}_2(x) = 1$, then, $\{\widetilde{L}_1, \widetilde{L}_2\}$ is a fuzzy partition on U.

From the matrix, we can get the decision classes of multi-label decision system:

$$L_1 = \bigcup_{j=1}^{3} L_1^{l_j} = L_1^{l_1} \cup L_1^{l_2} \cup L_1^{l_3} = \{x_1, x_2, x_3\},$$
$$L_2 = \bigcup_{j=1}^{3} L_2^{l_j} - L_1 = L_2^{l_1} \cup L_2^{l_2} \cup L_2^{l_3} - L_1 = \{x_4, x_5\}.$$

B. MULTI-LABEL VARIABLE PRECISION FUZZY NEIGHBORHOOD ROUGH SET MODEL

According to the above definition of fuzzy decision and decision classes, we can construct the multi-label variable precision fuzzy neighborhood rough set model, and we define its approximation space as follows.

Definition 11: Given a multi–label fuzzy neighborhood decision system $MNDT =< U, C \cup L >$, the label set is $L = \{l_1, l_2, \ldots, l_m\}$, and the radius of fuzzy neighborhood is δ, $B \subseteq C$, R_B is a fuzzy similarity relation on U induced by B. Then the lower and upper approximations are defined as

$$\underline{R}_B^{\delta,\alpha}(L) = \{\underline{R}_B^{\delta,\alpha}(\tilde{L}_1), \underline{R}_B^{\delta,\alpha}(\tilde{L}_2) \ldots \underline{R}_B^{\delta,\alpha}(\tilde{L}_r)\}, \quad (16)$$
$$\overline{R}_B^{\delta,\beta}(L) = \{\overline{R}_B^{\delta,\beta}(\tilde{L}_1), \overline{R}_B^{\delta,\beta}(\tilde{L}_2) \ldots \overline{R}_B^{\delta,\beta}(\tilde{L}_r)\}, \quad (17)$$

where

$$\underline{R}_B^{\delta,\alpha}(\tilde{L}_i) = \{x_i \in L_i | I([x_i]_B^{\delta}, \tilde{L}_i) \geq \alpha\}, \ 0.5 \leq \alpha \leq 1, \quad (18)$$
$$\overline{R}_B^{\delta,\beta}(\tilde{L}_i) = \{x_i \in L_i | I([x_i]_B^{\delta}, \tilde{L}_i) > \beta\}, \ 0 \leq \beta < 0.5. \quad (19)$$

According to Definition 11, we can define the corresponding multi-label fuzzy positive region.

$$POS_B^{\delta,\alpha}(L) = \underline{R}_B^{\delta,\alpha}(L) = \bigcup_{i=1}^{r} \underline{R}_B^{\delta,\alpha}(\tilde{L}_i). \quad (20)$$

$POS_B^{\delta,\alpha}(L)$ is the subset of samples whose fuzzy decision satisfying inclusion degree threshold. The greater the positive region is, the greater the characterizing power of attribute has. For classification learning, we always try to find a minimal attributes subset which makes the classification algorithm have the highest performance. Based on (20), the dependency function is computed as

$$\gamma_B^{\delta,\alpha}(L) = \frac{|POS_B^{\delta,\alpha}(L)|}{|U|}. \quad (21)$$

Eq. (21) indicates that the dependency degree of L with respect to B is a positive real number, that is, $\gamma_B^{\delta,\alpha}(L) \in [0, 1]$, and $\gamma_B^{\delta,\alpha}(L)$ characterizes the approximation power of B to L. If $\gamma_B^{\delta,\alpha}(L) = 1$, label set L is completely dependent on B; Otherwise, label set L depends on $\gamma_B^{\delta,\alpha}(L)$, that is, partial samples can be classified consistently.

Theorem 1: Given $MNDT =< U, C \cup L >$, for two parameters of δ and α, if $B_1 \subseteq B_2 \subseteq C$, then

(1) $POS_{B_1}^{\delta,\alpha}(L) \subseteq POS_{B_2}^{\delta,\alpha}(L)$;

(2) $\gamma_{B_1}^{\delta,\alpha}(L) \leq \gamma_{B_2}^{\delta,\alpha}(L)$.

Proof: Since $B_1 \subseteq B_2$, according to definition 1, we have $\forall x, y \in U$, $R_{B_2}(x, y) \leq R_{B_1}(x, y)$. As the definition of the fuzzy neighborhood granule, we have $\forall x \in U$, $[x]_{B_2}^{\delta} \subseteq [x]_{B_1}^{\delta}$. Based on Eq.(18), $\underline{R}_{B_1}^{\delta,\alpha}(\tilde{L}_i) \subseteq \underline{R}_{B_2}^{\delta,\alpha}(\tilde{L}_i)$ holds. Then, $POS_{B_1}^{\delta,\alpha}(L) \subseteq POS_{B_2}^{\delta,\alpha}(L)$. Obviously, according to Theorem 3.1 (1) and Eq. (21), formula (2) holds. ■

Theorem 1 indicates that the multi-label fuzzy positive region and fuzzy dependency function are consistent with the monotonicity of attributes subset size. This property is particularly important for constructing a forward reduction algorithm since it can ensure that adding any candidate attribute to an existing attribute subset that does not reduce the fuzzy dependency of a new subset. Then, the definition of variable precision multi-label fuzzy reduct can be given.

Definition 12: Given $MNDT =< U, C \cup L >$ and the neighborhood radius δ, $B \subseteq C$, we can say attribute subset B is variable precision multi-label fuzzy reduct if

(1) $\gamma_B^{\delta,\alpha}(L) = \gamma_C^{\delta,\alpha}(L)$;

(2) $\forall a \in B, \gamma_{B-a}^{\delta,\alpha}(L) < \gamma_B^{\delta,\alpha}(L)$.

The definition indicates that the reduction set has the same or similar approximate ability as the whole attribute set. From definition 12, we can present the following definition of attribute significance.

Definition 13: Given $MNDT =< U, C \cup L >$, the label set is $L = \{l_1, l_2, \ldots, l_m\}$, for any $a \in C - B$, the significance of attribute a with respect to L can be given by

$$sig(a, B, L) = \gamma_{B \cup a}^{\delta,\alpha}(L) - \gamma_B^{\delta,\alpha}(L). \quad (22)$$

Eq. (22) shows the importance of an attribute a with respect to conditional attributes B, and it can be measured by the dependency change caused by adding a to B. If an attribute a is added to attributes subset B, the dependency of decision attribute on B does not change, that is, $sig(a, B, L) = 0$, it means that a can be removed from B, (i.e., attribute a is redundant).

C. MULTI-LABEL ATTRIBUTE REDUCTION ALGORITHM

According to formulas (13)-(22), we can construct a heuristic algorithm for multi-label attribute reduction, which is called multi-label attribute reduction algorithm based on variable precision fuzzy neighborhood rough set. The greedy search strategy is adopted in the algorithm of FNRS-ML, which starts with an empty set and one attribute is added to each step to maximize the discrimination ability until the degree of dependency is invariant.

In Algorithm 1, the algorithm terminates when any remaining attributes are added to the existing attribute set without increasing the multi-label fuzzy dependency $\gamma_{red \cup a_i}^{\delta}(L)$. The time complexity for calculating a fuzzy similarity relation

Algorithm 1 Multi-Label Attributes Reduction Algorithm Based on Variable Precision Fuzzy Neighborhood Rough Set(FNRS-ML)

Input: U: sample set; C: attribute set; L: label space; δ: the neighborhood radius; α: the parameter of inclusion degree.
Output: one reduct.

$red \longleftarrow \emptyset$
for each $a_i \in C - red$ **do**
  compute the fuzzy similarity relation $R^{\delta}_{red \cup a_i}$;
  for each $l \in L$ **do**
    compute the fuzzy decision $\tilde{L}_i$ and decision classes $L_1, L_2, \ldots, L_r$;
  end for
  compute $\gamma^{\delta,\alpha}_{red \cup a_i}(L)$;
  compute $sig(a_i, B, L) = \gamma^{\delta,\alpha}_{red \cup a_i}(L) - \gamma^{\delta,\alpha}_{red}(L)$;
end for
find attribute a_k with maximum values $sig(a_k, B, L)$;
if $sig(a_k, B, L) > 0$ **then**
  $red \leftarrow red \cup a_k$;
else
  return red;
end if
return red;

is $O(\frac{1}{2}|U|^2)$, and the worst time complexity for searching a reduct is $O(|C|^2)$. Overall, the computational complexity of Algorithm 1 is $O(\frac{1}{2}|U|^2 \times |C| \times |L| + |C|^2)$.

IV. EXPERIMENTS

A. DATASETS

We select nine benchmark multi-label datasets as our testbeds to test the proposed method FNRS-ML, and all datasets are from different application fields [1], [31]. Among these datasets, Arts, Recreation, Society, Entertainment and Science are extensively used to Web page text categorization. Cal500 is a content-based music annotation and retrieval research data, which consists of 500 Western popular music tracks. Flags is a image processing and classification data, including 194 image instances and 7 labels. Emotions is a benchmark for music, containing 593 music objects and each of which belongs to at least one of 6 labels. Yeast is applied to predict gene functional classes, which contains 2417 instances where each instance indicates a yeast gene and 14 feasible labels. TABLE 2 shows the detailed presentation information of these datasets.

B. EXPERIMENT SETTINGS

In this subsection, we compare the experimental results with six advanced multi-label attribute reduction methods, including MDDMspc [45], MDDMproj [45], RF-ML [25], PMU [13], MLNB [47], and RMFRS [21]. For MDDMspc and MDDMproj, μ is set as 0.5 as recommended in the literature [45]. For PMU, we use an equal-width strategy to discretize continuous attributes to two bins, while categorical attributes maintained unchanged, as suggested in [13]. For MLNB, the threshold parameters smooth and ratio are set as 1 and 0.3 [47]. For RMFRS, the weight parameter β and μ are set as 0.9 and 0.1 [21]. For our proposed FNRS-ML, there are two important thresholds δ and α. The parameter δ is regarded as the fuzzy neighborhood radius, and α is the threshold of inclusion degree. In this paper, we set $\delta = 0.1$, the threshold α is set to 0.7, 0.8, 0.85, and 0.95 for the datasets from different fields containing Web page categorization, Gene, Image processing, and Music, respectively. The inclusion degree of upper approximation satisfies $0 \leq \beta < 0.5$. Meanwhile, ML-KNN (K = 10) [48] is used to evaluate the classification performance of all multi-label attribute reduction methods. Ultimately, we choose *Average Precision*, *Ranking Loss*, *Hamming Loss* and *One-error* to estimate the predictive performance of these multi-label attribute reduction methods. As we know, the four criteria make the performance comparison from different aspects, and normally a few of algorithms are superior to other algorithms on all these criteria.

TABLE 2. Description of multi-label datasets.

Datasets	Instances	Attributes	Labels	Training	Test
Cal500	502	68	174	251	251
Flags	194	19	7	129	65
Emotions	593	72	6	391	202
Yeast	2417	103	14	1499	918
Arts	5000	462	26	2000	3000
Recreation	5000	606	22	2000	3000
Society	5000	636	27	2000	3000
Entertainment	5000	640	21	2000	3000
Science	5000	743	40	2000	3000

C. EXPERIMENTAL RESULTS

1) EVALUATION OF PREDICTIVE PERFORMANCE

We compare FNRS-ML with MDDMspc, MDDMproj, RF-ML, PMU, MLNB, and RMFRS in terms of predictive classification performance to testify its effectiveness. In which, MDDMspc, MDDMproj, RF-ML, PMU, and RMFRS can obtain the attribute ranking, while MLNB and FNRS-ML get the attribute subset directly. In this algorithm, we choose the same number of attributes with the quantity determined by FNRS-ML as the final attribute subset for the comparability of the results. TABLES 3-6 list the experimental results of seven attribute reduction algorithms on four evaluation criteria. For given evaluation criteria, " $\downarrow$" means that "the smaller the better" and " $\uparrow$" means that "the larger the better". In these results, bold font indicates the best classification performance for each dataset, and italic represents the average classification performance of every algorithm.

As the experimental results are exhibited in TABLES 3-6, it can be found that: (1) FNRS-ML achieves superior performance with Average Precision on seven datasets. (2) For Ranking Loss and Hamming Loss, FNRS-ML outperforms all comparing algorithms at least on six datasets, and there is not great difference between FNRS-ML and the optimum value on the other datasets. (3) For One-error, FNRS-ML obtains the best classification performance on all datasets

TABLE 3. Performance comparison of different algorithms on Average precision (↑).

Datasets	MDDMspc	MDDMproj	RF-ML	PMU	MLNB	RMFRS	FNRS-ML
Cal500	0.4715	0.4715	0.4756	0.4737	0.4776	0.4793	**0.4844**
Flags	0.7828	0.7828	0.8081	0.7779	0.7762	0.7808	**0.8091**
Emotions	0.7588	0.7637	0.7652	0.7574	0.7529	**0.7808**	0.7788
Yeast	0.7125	0.7101	0.7209	0.7185	**0.7355**	0.7229	0.7300
Arts	0.4937	0.4691	0.4917	0.4909	0.4991	0.5017	**0.5129**
Recreation	0.4497	0.4592	0.4507	0.4434	0.4790	0.4776	**0.4866**
Society	0.5550	0.5523	0.5781	0.5763	0.5894	0.5777	**0.5914**
Entertainment	0.5425	0.5542	0.5533	0.5687	0.5575	0.5567	**0.5791**
Science	0.4270	0.4222	0.4542	0.4492	0.4613	0.4880	**0.4884**
Average	*0.5771*	*0.5761*	*0.5886*	*0.5840*	*0.5921*	*0.5962*	***0.6067***

TABLE 4. Performance comparison of different algorithms on Ranking Loss (↓).

Datasets	MDDMspc	MDDMproj	RF-ML	PMU	MLNB	RMFRS	FNRS-ML
Cal500	0.1951	0.1952	0.1901	0.1902	0.1913	0.1916	**0.1888**
Flags	0.2579	0.2579	0.2277	0.2503	0.2518	**0.1729**	0.2297
Emotions	0.2027	0.2014	0.2005	0.2110	0.2055	0.2211	**0.1770**
Yeast	0.2052	0.2093	0.1984	0.2045	**0.1871**	0.1991	0.1943
Arts	0.1549	0.1651	0.1528	0.1567	0.1542	0.1486	**0.1464**
Recreation	0.1912	0.1909	0.1865	0.1925	0.1879	0.1849	**0.1827**
Society	0.1573	0.1612	0.1483	0.1500	0.1456	**0.1405**	0.1439
Entertainment	0.1289	0.1255	0.1240	0.1224	0.1254	0.1279	**0.1202**
Science	0.1470	0.1483	0.1378	0.1382	0.1364	0.1298	**0.1289**
Average	*0.1822*	*0.1839*	*0.1740*	*0.1795*	*0.1761*	*0.1685*	***0.1680***

TABLE 5. Performance comparison of different algorithms on Hamming Loss (↓).

Datasets	MDDMspc	MDDMproj	RF-ML	PMU	MLNB	RMFRS	FNRS-ML
Cal500	0.1409	0.1409	0.1424	0.1413	0.1426	0.1401	**0.1394**
Flags	0.6857	0.6857	0.3121	0.3253	0.7099	0.3020	**0.2967**
Emotions	0.2277	0.2409	0.2410	0.2343	0.2451	**0.2137**	0.2170
Yeast	0.2309	0.2278	0.2302	0.2287	0.2080	0.2245	**0.2011**
Arts	0.0618	0.0632	0.0615	0.0613	0.0612	0.0606	**0.0604**
Recreation	0.0632	0.0620	0.0632	0.0639	0.0611	0.0610	**0.0608**
Society	0.0601	0.0598	0.0573	0.0568	**0.0559**	0.0565	0.0562
Entertainment	0.0630	0.0632	**0.0611**	0.0615	0.0614	0.0617	0.0616
Science	0.0354	0.0353	0.0350	0.0352	0.0346	0.0343	**0.0341**
Average	*0.1743*	*0.1753*	*0.1338*	*0.1343*	*0.1755*	0.1283	***0.1253***

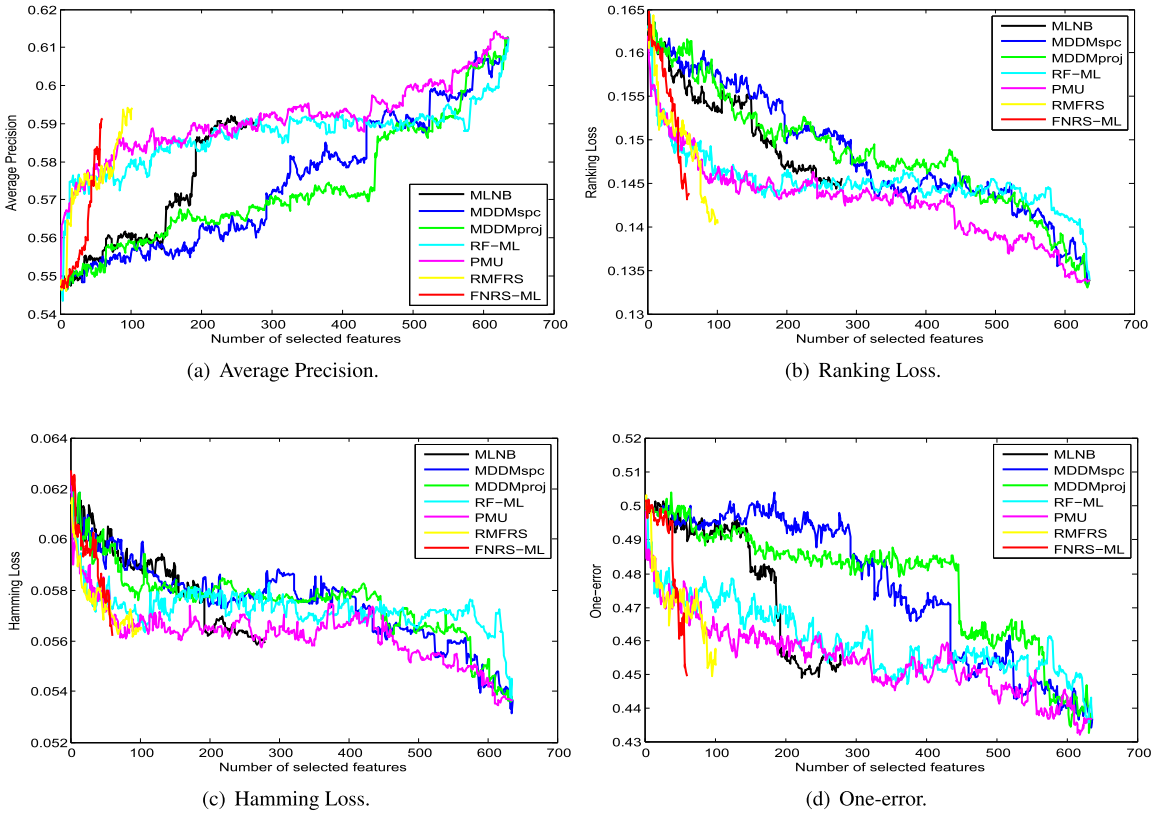
TABLE 6. Performance comparison of different algorithms on one-error (↓).

Datasets	MDDMspc	MDDMproj	RF-ML	PMU	MLNB	RMFRS	FNRS-ML
Cal500	0.1673	0.1673	0.1195	0.1116	0.1434	0.1355	**0.1076**
Flags	0.2154	0.2154	**0.2000**	0.2615	0.2769	0.3317	**0.2000**
Emotions	0.3515	0.3317	0.3218	0.3317	0.3762	0.3317	**0.3119**
Yeast	0.2527	0.2538	0.2462	0.2549	0.2560	0.2495	**0.2440**
Arts	0.6610	0.6907	0.6553	0.6477	0.6433	**0.6193**	0.6277
Recreation	0.7137	0.7017	0.7183	0.7183	0.6643	0.6720	**0.6607**
Society	0.4957	0.4980	0.4707	0.4743	0.4540	0.4530	**0.4497**
Entertainment	0.6170	0.6100	0.6017	0.5750	0.5930	0.5923	**0.5707**
Science	0.7163	0.7223	0.6767	0.6887	0.6713	0.6373	**0.6343**
Average	*0.4656*	*0.4657*	*0.4456*	*0.4515*	*0.4532*	*0.4469*	***0.4230***

except for Arts on RMFRS. (4) In addition, FNRS-ML acquires better than these comparison algorithms over all evaluation metrics with respect to average classification performance. In short, the experimental results of performance comparison demonstrate that FNRS-ML tends to work better than other baselines.

To sufficiently testify the authenticity and reliability of the predictive performance of those seven multi-label attribute reduction algorithms, we carry out several experiments to illustrate the change tendency of predictive classification performance with different number of selected attributes. In this study, we select three multi-label datasets, i.e., Society, Entertainment, and Science to show the change tendency on four evaluation metrics as shown in FIGS. 1-3. In which, the horizontal axis denotes the number of selected attributes, and the vertical axis represents the predictive classification performance of different metrics. Additionally, seven lines in each figure stand for seven comparison algorithms.

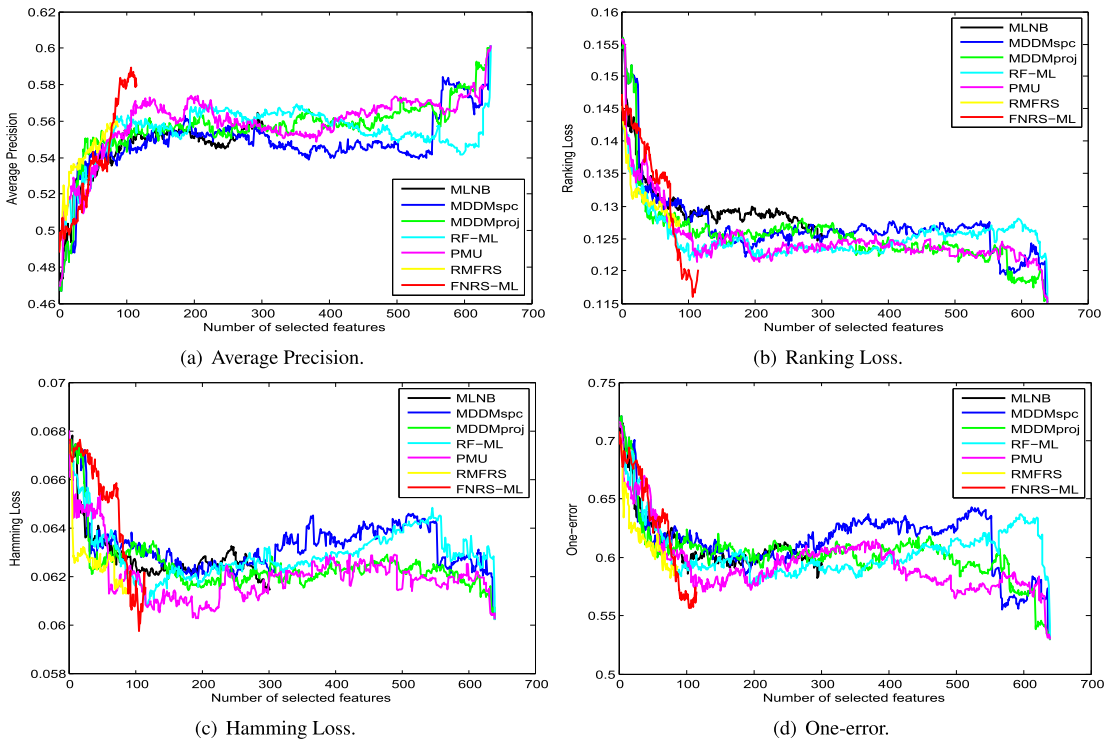
From FIGS. 1-3, we can observe that: (1) The lines in each figure are not monotonous. (2) Regardless of the trend of the curve, FNRS-ML can get the best classification performance as the number of selected attributes increases. It should be



(a) Average Precision. (b) Ranking Loss.

(c) Hamming Loss. (d) One-error.

FIGURE 1. Society.



(a) Average Precision. (b) Ranking Loss.

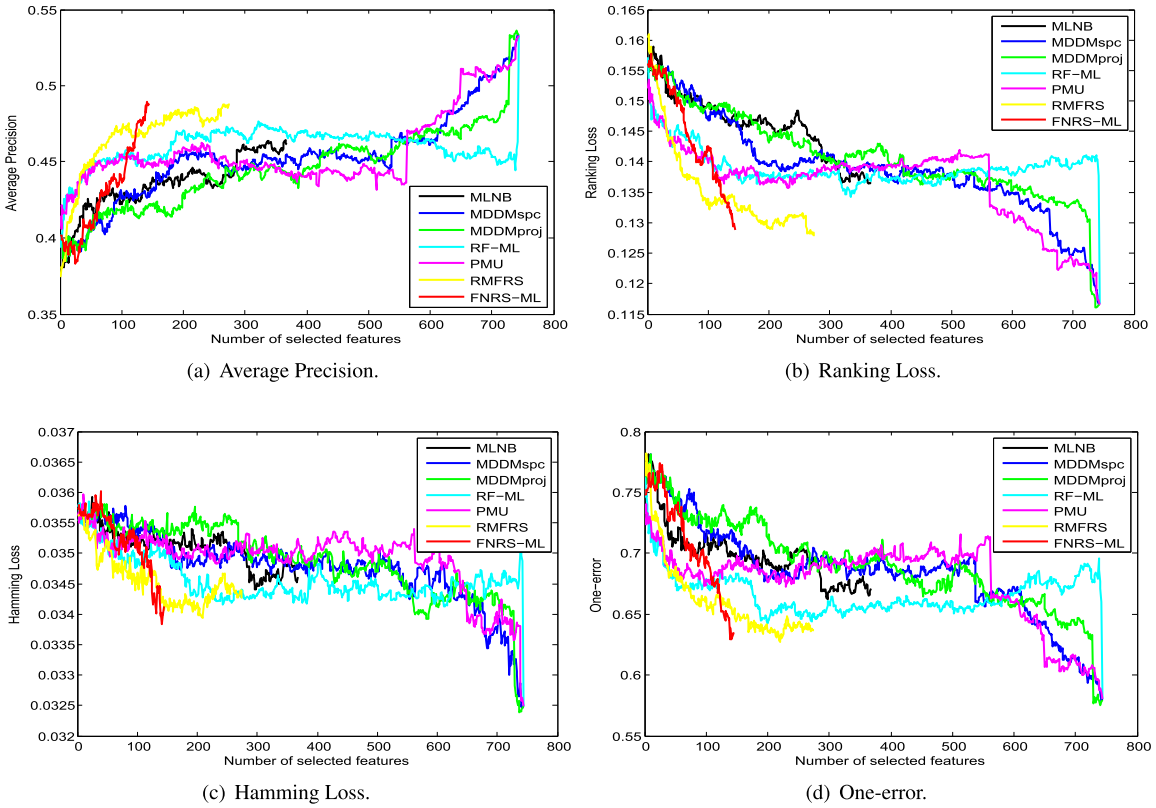
(c) Hamming Loss. (d) One-error.

FIGURE 2. Entertainment.

noted that FNRS-ML will achieve better classification performance with a certain number of attributes, and it accords with the actual situation.

2) STATISTICAL TEST

To systematically discuss the comparative algorithms in statistical sense, we carry out a nonparametric Friedman test



(a) Average Precision. (b) Ranking Loss.

(c) Hamming Loss. (d) One-error.

FIGURE 3. Science.

TABLE 7. Sum up the Friedman statistics F_F ($k = 7, N = 9$) on the basis of different metrics and the critical value.

Evaluation measure	F_F	Critical value($\alpha = 0.10$)
Average Precision	13.5806	1.93
Ranking loss	11.5538	
Hamming Loss	7.3910	
One-error	8.0218	

[5] and use it to compare multiple datasets. Suppose there are k comparison algorithms and N datasets. The average rank of the j-th algorithm among all multi-label datasets is expressed as $R_i = \frac{1}{N}\sum_{i=1}^{N} r_i^j$, where r_{ij} be the rank of the j-th algorithm on i-th dataset. The Friedman statistic F_F follows the F-distribution with $(k-1)$ and $(k-1)(N-1)$ degrees of freedom under the null-hypothesis, and it is defined as:

$$F_F = \frac{(N-1)\chi_F^2}{N(k-1)-\chi_F^2}, \tag{23}$$

where

$$\chi_F^2 = \frac{12N}{k(k+1)}\left(\sum_{i=1}^{k} R_i^2 - \frac{k(k+1)^2}{4}\right).$$

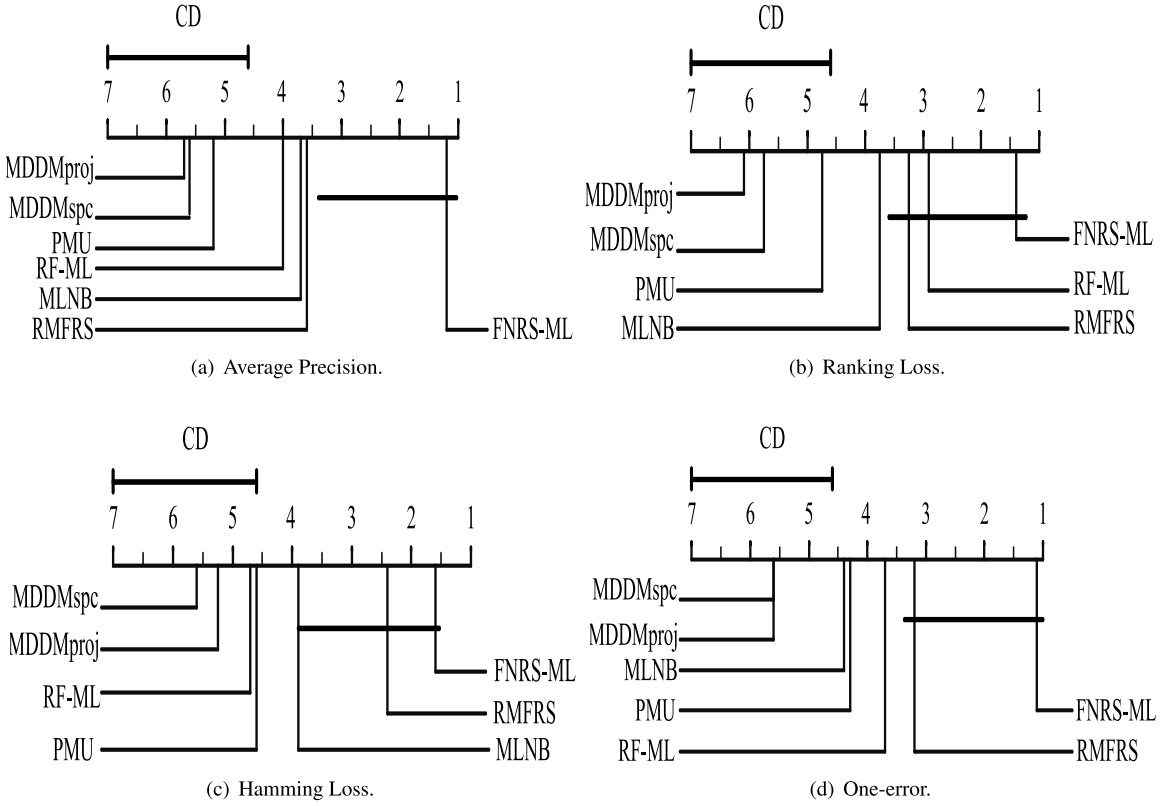
The Friedman statistic F_F and the corresponding critical value are summed up in TABLE 7. Moreover, the null hypothesis is that all algorithms are executed equally, is explicitly rejected at the significance level $\alpha = 0.10$. As we mainly aim at the performance difference between the proposed FNRS-ML algorithm with the other seven comparison algorithms, we use Bonferroni-Dunn test [2] to set FNRS-ML as the control algorithm. It is considered that there is a significant difference if the average rank between the control algorithm and one comparison algorithms is greater than the critical value (CD):

$$CD_\alpha = q_\alpha\sqrt{\frac{k(k+1)}{6N}}. \tag{24}$$

Here, for the Bonferroni-Dunn test, we have $q_\alpha = 2.394$ at significance level $\alpha = 0.10$, and thus CD $= 2.4379$ ($k = 7, N = 9$).

To display the relative performance of FNRS-ML and other comparison algorithms more intuitively, FIG. 4 illustrates the CD diagrams on different evaluation metrics. In every subfigure, the average rank of each comparison algorithm is drawn along the axis from high to low, and the rightmost algorithm is considered as the best one. The thick lines of connection between the control algorithm and the comparison algorithm indicates that the two algorithms have not apparent difference.

According to the results shown in FIG. 4, we can come to conclusion: (1) FNRS-ML is significantly better than MDDMspc, MDDMproj, and PMU over all evaluation metrics. (2) FNRS-ML performs comparable performance against MLNB in Average Precision, Hamming Loss, and One-error. (3) FNRS-ML remarkably outperforms other four algorithms in terms of Ranking Loss and Hamming Loss. To summarize, FNRS-ML achieves highly competitive performance comparing to the other several state-of-the-art multi-label attribute reduction algorithms.



(a) Average Precision. (b) Ranking Loss.

(c) Hamming Loss. (d) One-error.

FIGURE 4. Performance comparison on terms of four evaluation criteria.

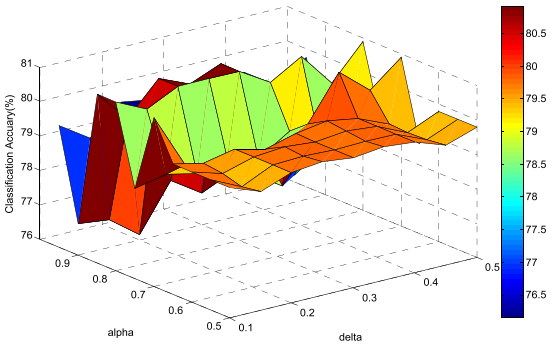


FIGURE 5. The classification accuracy varying with δ and α of Flags.

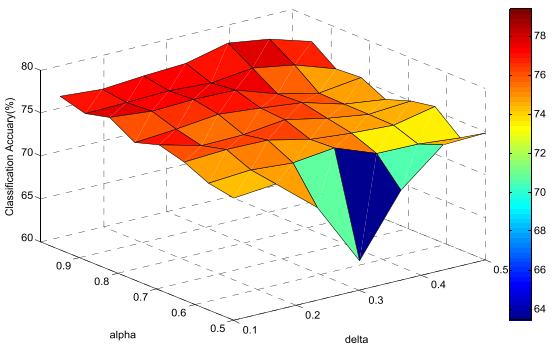


FIGURE 6. The classification accuracy varying with δ and α of Emotions.

3) THE INFLUENCE OF PARAMETER α AND δ

To observe the influence of parameter α and δ on classification performance, we display the figures about classification accuracy varying with parameter α and δ of the datasets in FIGS. 5-6. From FIGS. 5-6, different color areas represent different classification accuracy. We can see that the two datasets obtain higher accuracy in a large area, especially when δ is 0.1. That is, when δ is 0.1, the α of these two datasets is 0.85 and 0.95 respectively, the classification performance reaches the highest value. Thus, FNRS-ML method is of stability and feasibility.

V. CONCLUSION AND FUTURE WORKS

In this paper, we have proposed a multi-label attribute reduction algorithm based on variable precision fuzzy neighborhood rough set. The algorithm effectively removes redundant attributes, and solves the restriction problem of data type in general multi-label attribute reduction. Meanwhile, the variable precision model averts the influence of the strictness of fuzzy neighborhood rough set on the calculation of upper and lower approximations. The experiment has shown that, FNRS-ML obtains highly classification performance comparing with other state-of-the-art multi-label algorithms. In the future, our works will research on how to design other multi-label attribute reduction methods based on variable precision model by considering the label correlations and how to automatically set the neighborhood size for each dataset.

REFERENCES

[1] K. Bache and M. Lichman, "UCI machine learning repository," School Inf. Comput. Sci., Univ. Califirnia, Irvine, CA, USA, 2013. [Online]. Available: http://archive.ics.uci.edu/ml

[2] O. J. Dunn, "Multiple comparisons among means," *J. Amer. Stat. Assoc.*, vol. 56, no. 293, pp. 52–64, Mar. 1961.
[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] A. Elisseeff and J. Weston, "A kernel method for multi-labelled classifcation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2002, pp. 681–687.
[5] M. Friedman, "A comparison of alternative tests of significance for the problem of *m* rankings," *Ann. Math. Statist.*, vol. 11, no. 1, pp. 86–92, Mar. 1940.
[6] X. Fan, W. Zhao, C. Wang, and Y. Huang, "Attribute reduction based on max-decision neighborhood rough set model," *Knowl.-Based Syst.*, vol. 151, pp. 16–23, Jul. 2018.
[7] W. Gao and Z.-H. Zhou, "On the consistency of multi-label learning," *Artif. Intell.*, vols. 199–200, pp. 22–44, Apr. 2013.
[8] Q. Gu, Z. Li, and J. Han, "Correlated multi-label feature selection," in *Proc. 20th ACM Int. Conf. Inf. Knowl. Manage.*, 2011, pp. 1087–1096.
[9] O. Gharroudi, H. Elghazel, and A. Aussem, "A comparison of multi-label feature selection methods using the random forest paradigm," in *Proc. Can. Conf. Artif. Intell.* Cham, Switzerland: Springer, 2014, pp. 95–106.
[10] Q. Hu, D. Yu, J. Liu, and C. Wu, "Neighborhood rough set based heterogeneous feature subset selection," *Inf. Sci.*, vol. 178, no. 18, pp. 3577–3594, Sep. 2008.
[11] H. Han, M. Huang, Y. Zhang, X. Yang, and W. Feng, "Multi-label learning with label specific features using correlation information," *IEEE Access*, vol. 7, pp. 11474–11484, 2019.
[12] J.-Y. Jiang, S.-C. Tsai, and S.-J. Lee, "FSKNN: Multi-label text categorization based on fuzzy similarity and k nearest neighbors," *Expert Syst. Appl.*, vol. 39, no. 3, pp. 2813–2821, Feb. 2012.
[13] J. Lee and D.-W. Kim, "Feature selection for multi-label classification using multivariate mutual information," *Pattern Recognit. Lett.*, vol. 34, no. 3, pp. 349–357, Feb. 2013.
[14] J. Lee and D.-W. Kim, "Memetic feature selection algorithm for multi-label classification," *Inf. Sci.*, vol. 293, pp. 80–96, Feb. 2015.
[15] L. Yong, H. Wenliang, J. Yunliang, and Z. Zhiyong, "Quick attribute reduct algorithm for neighborhood rough set model," *Inf. Sci.*, vol. 271, pp. 65–81, Jul. 2014.
[16] Y. Lin, Q. Hu, J. Liu, and J. Duan, "Multi-label feature selection based on max-dependency and min-redundancy," *Neurocomputing*, vol. 168, pp. 92–103, Nov. 2015.
[17] Y. Lin, Q. Hu, J. Liu, J. Li, and X. Wu, "Streaming feature selection for multilabel learning based on fuzzy mutual information," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 6, pp. 1491–1507, Dec. 2017.
[18] Y. Lin, Y. Li, C. Wang, and J. Chen, "Attribute reduction for multi-label learning with fuzzy rough set," *Knowl.-Based Syst.*, vol. 152, no. 15, pp. 51–61, Jul. 2018.
[19] T. Y. Lin, "Granular computing on binary relations I: Data mining and neighborhood systems," *Rough Sets Knowl. Discovery*, vol. 1, pp. 107–121, Jul. 1998.
[20] H. Li, D. Li, Y. Zhai, S. Wang, and J. Zhang, "A novel attribute reduction approach for multi-label data based on rough set theory," *Inf. Sci.*, vols. 367–368, pp. 827–847, Nov. 2016.
[21] Y. Li, Y. Lin, J. Liu, W. Weng, Z. Shi, and S. Wu, "Feature selection for multi-label learning based on kernelized fuzzy rough sets," *Neurocomputing*, vol. 318, pp. 271–286, Aug. 2018.
[22] D.-Q. Miao and G.-R. Hu, "A heuristic algorithm for reduction of knowledge," *J. Comput. Res. Develop.*, vol. 36, no. 6, pp. 681–684, 1999.
[23] Z. Pawlak, "Rough sets," *Int. J. Comput. Inf. Sci.*, vol. 11, no. 5, pp. 341–356, Oct. 1982.
[24] O. Reyes, C. Morell, and S. Ventura, "Scalable extensions of the ReliefF algorithm for weighting and selecting features on the multi-label learning context," *Neurocomputing*, vol. 161, no. 5, pp. 168–182, Aug. 2015.
[25] N. Spolaôr, E. A. Cherman, M. C. Monard, and H. D. Lee, "ReliefF for multi-label feature selection," in *Proc. Brazilian Conf. Intell. Syst.*, Oct. 2013, pp. 6–11.
[26] N. Spolaôr, M. C. Monard, G. Tsoumakas, and H. Lee, "Label construction for multi-label feature selection," in *Proc. Brazilian Conf. Intell. Syst. (BRACIS)*, Oct. 2014, pp. 247–252.
[27] N. Spolaôr, E. A. Cherman, M. C. Monard, and H. D. Lee, "A comparison of multi-label feature selection methods using the problem transformation approach," *Electron. Notes Theor. Comput. Sci.*, vol. 292, pp. 135–151, Mar. 2013.
[28] N. Spolaôr, E. A. Cherman, M. C. Monard, and H. Lee, "Filter approach feature selection methods to support multi-label learning based on relieff and information gain," in *Proc. Brazilian Symp. Artif. Intell.*, 2012, pp. 72–81.
[29] R. E. Schapire and Y. Singer, "BoosTexter: A boosting-based system for text categorization," *Mach. Learn.*, vol. 39, nos. 2–3, pp. 135–168, May 2000.
[30] L. Sun, X. Zhang, J. Xu, and S. Zhang, "An attribute reduction method using neighborhood entropy measures in neighborhood rough sets," *Entropy*, vol. 21, no. 2, p. 155, Feb. 2019.
[31] G. Tsoumakas, E. Spyromitros-Xioufis, J. Vilcek, and I. Vlahavas, "Mulan: A java library for multi-label learning," *J. Mach. Learn. Res.*, vol. 12, pp. 2411–2414, Jun. 2011.
[32] B. Wu, S. Lyu, B.-G. Hu, and Q. Ji, "Multi-label learning with missing labels for image annotation and facial action unit recognition," *Pattern Recognit.*, vol. 48, no. 7, pp. 2279–2289, Jul. 2015.
[33] C. Wang, M. Shao, Q. He, Y. Qian, and Y. Qi, "Feature subset selection based on fuzzy neighborhood rough sets," *Knowl.-Based Syst.*, vol. 111, pp. 173–179, Aug. 2016.
[34] C. Wang, Y. Qi, M. Shao, Q. Hu, D. Chen, Y. Qian, and Y. Lin, "A fitting model for feature selection with fuzzy rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 4, pp. 741–753, Aug. 2017.
[35] C. Wang, Y. Lin, and J. Liu, "Feature selection for multi-label learning with missing labels," *Int. J. Speech Technol.*, vol. 49, no. 8, pp. 3027–3042, Aug. 2019.
[36] C. Wang, Y. Shi, X. Fan, and M. Shao, "Attribute reduction based on k-nearest neighborhood rough sets," *Int. J. Approx. Reasoning*, vol. 106, pp. 18–31, Mar. 2019.
[37] C. Wang, Y. Huang, M. Shao, and X. Fan, "Fuzzy rough set-based attribute reduction using distance measures," *Knowl.-Based Syst.*, vol. 164, pp. 205–212, Jan. 2019.
[38] C. Wang, Y. Wang, M. Shao, Y. Qian, and D. Chen, "Fuzzy rough attribute reduction for categorical data," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 5, pp. 818–830, May 2020, doi: 10.1109/TFUZZ.2019.2949765.
[39] C. Z. Wang, Y. Huang, M. W. Shao, Q. H. Hu, and D. G. Chen, "Feature selection based on neighborhood self-Information," *IEEE Trans. Cybern.*, early access, Jul. 9, 2019, doi: 10.1109/TCYB.2019.2923430.
[40] Y. Yu, W. Pedrycz, and D. Miao, "Neighborhood rough sets based multi-label classification for automatic image annotation," *Int. J. Approx. Reasoning*, vol. 54, no. 9, pp. 1373–1387, Nov. 2013.
[41] Y. Yu, J. Wang, Q. Tan, L. Jia, and G. Yu, "Semi-supervised multi-label dimensionality reduction based on dependence maximization," *IEEE Access*, vol. 5, pp. 21927–21940, Oct. 2017.
[42] Y. Yang, D. Chen, and H. Wang, "Active sample selection based incremental algorithm for attribute reduction with rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 4, pp. 825–838, Aug. 2017.
[43] M.-L. Zhang and Z.-H. Zhou, "A review on multi-label learning algorithms," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 8, pp. 1819–1837, Aug. 2014.
[44] M.-L. Zhang and Z.-H. Zhou, "Multilabel neural networks with applications to functional genomics and text categorization," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 10, pp. 1338–1351, Oct. 2006.
[45] Y. Zhang and Z.-H. Zhou, "Multilabel dimensionality reduction via dependence maximization," *ACM Trans. Knowl. Discovery Data*, vol. 4, no. 3, pp. 1–21, Oct. 2010.
[46] L. J. Zhang, H. Q. Hu, J. Duan, and X. X. Wang, "Multi-label feature selection with fuzzy rough sets," in *Proc. Int. Conf. Rough Sets Knowl. Technol.* Cham, Switzerland: Springer, 2014, pp. 121–128.
[47] M.-L. Zhang, J. M. Peña, and V. Robles, "Feature selection for multi-label naive Bayes classification," *Inf. Sci.*, vol. 179, no. 19, pp. 3218–3229, Sep. 2009.
[48] M.-L. Zhang and Z.-H. Zhou, "ML-KNN: A lazy learning approach to multi-label learning," *Pattern Recognit.*, vol. 40, no. 7, pp. 2038–2048, Jul. 2007.
[49] X. Zhang, W. Xu, and W. Zhang, "A novel algorithm of matrix computation for lower approximation reduction in IOIS," in *Proc. 7th Int. Conf. Fuzzy Syst. Knowl. Discovery*, vol. 4, Aug. 2010, pp. 1876–1880.
[50] P. Zhu, Q. Xu, Q. Hu, C. Zhang, and H. Zhao, "Multi-label feature selection with missing labels," *Pattern Recognit.*, vol. 74, pp. 488–502, Feb. 2018.

PANPAN CHEN is currently pursuing the M.S. degree with the School of Mathematics and Statistics, Minnan Normal University. Her research interests include information processing and data mining.

MENGLEI LIN received the M.S. degree in mathematics from Xiamen University. He currently is a Professor with the School of Mathematics and Statistics, Minnan Normal University. His research interests include data mining and fuzzy set theory and its application.

JINGHUA LIU received the Ph.D. degree in systems engineering from the Department of Automation, Xiamen University, Xiamen, China, in 2020. She is currently a Lecturer with the College of Computer Science and Technology, Huaqiao University, Xiamen. Her current research interests include data mining, machine learning, and granular computing.