Abstract: The paper aims to improve the multi-label classification performance using the feature reduction technique. According to the determination of the dependency among features based on fuzzy rough relation, features with the highest dependency score will be retained in the reduction set. The set is subsequently applied to enhance the performance of the multi-label classifier. We investigate the effectiveness of the proposed model against the baseline via time complexity.

Keywords: Fuzzy rough relation, label-specific feature, feature reduction set.
removing any feature in the original feature space. However, the selection of features for reduction is randomly selected. Although FRS-LIFT improves the performance of multi-label learning via reducing redundant label-specific feature dimensionalities, its computational complexity is high. FRS-SS-LIFT that is the multi-label learning approach to reduce the label-specific feature by sample selection. Thus, the time and memory consumption of FRS-SS-LIFT is lower than that of FRS-LIFT. But both the two approaches do not take full account of the correlations between different labels. Moreover, the feature selection approaches to obtain the optimal reduction set is randomized completely. Recently, Thi-Ngan Nguyen et al [9] propose a semi-supervised multi-label classification algorithm MULTICS to exploit specific features of label set. The algorithm MULTICS use the functions TEST which is called recursively to identify components from labeled and unlabeled sets, but it does not concern with the feature reduction. Daniel et al [10] propose the new data dimensionality reduction approach using the Forest Optimization Algorithm (FOA) to obtain domain knowledge from feature weighting.

In this paper, we focus on studying the fuzzy rough relationships and contribute in two aspects. Firstly, we determine the fuzzy rough relation to calculate the approximate dependency between samples on each feature. Then, we select the purpose-based feature with the greatest dependency to give into the optimal reduction set. Secondly, we propose a new algorithm to improve the LIFT [7] which has processed the increased feature dimensionalities by reducing the feature space. We calculate the degree of the membership function for each element $x$ in universe $\mathcal{X}$ and improve a new systematic reduction via a review per feature which has the highest dependency before classification. In fact, we based on the greatest dependency on each feature to select the most dominant features into the feature reduction set. Thereby, it may help to reduce set using a given threshold.

The remaining parts of this paper are organised as follows: The next section introduces the multi-label training method, LIFT method, the fuzzy rough relationship, FRS-LIFT method and the factors related to feature reduction. Section 3 introduces about the label-specific feature reduction. Section 4 presents our proposed algorithm. Finally, several conclusions and plans to develop in the future are drawn in Section 5.

2. Related work

2.1. Multi-Label training

Multi-label training is stated as follows [11]:

Let $\mathcal{X} = \mathbb{R}^d$ be a sample space and $\mathcal{L}$ be a finite set of $q$ labels $\mathcal{L} = \{l_1, l_2, \ldots, l_q\}$.

Let $\mathcal{T} = [(x_i, Y_i)]_{i = 1, 2, \ldots, n}$ be multi-label training set with $n$ samples where each $x_i \in \mathcal{X}$ is a $d$-dimensional feature vector, $x_i = [x_i^1, x_i^2, \ldots, x_i^d]$ and $Y_i \subseteq \mathcal{L}$ is the set of labels associated with $x_i$. The desired purpose is that the training system will create a real-valued function $f: \mathcal{X} \times P(\mathcal{L}) \to \mathbb{R}$; where $P(\mathcal{L})$ is a power set of $\mathcal{L}$. $P(\mathcal{L}) = 2^\mathcal{L} / \emptyset$ is the set of the non-empty label sets $Y_i$ that connect to $x_i$.

The problem of multi-label classification is also shown in the context of semi-supervised multi-label learning model [3] as follows:

Let $D$ be the set of documents in a considered domain. Let $L = \{l_1, \ldots, l_q\}$ be the set of labels.

Let $\overline{D}$ and $\overline{D}^u$ be the collections of labeled and unlabeled documents, correspondingly. For each $d$ in $\overline{D}$, $\text{label}(d)$ denotes the set of labels assigned to $d$. The task is to derive a multi-label classification function $f: \overline{D} \to 2^\mathcal{L}$, i.e., given a new unlabeled document $d \in \overline{D}$, the function identifies a set of relevant labels $f(d) \subseteq \mathcal{L}$.  

2.2. Approach to LIFT

As can be seen in [7], in order to train a multi-label learning model successfully, approach to LIFT perform three steps. The first step is to create label-specific features for each label $l_k \in \mathcal{L}$ which is done by dividing the training set $\mathcal{T}$ into two sample sets:
\[ P_k = \{ x_i \mid (x_i, Y_i) \in T, l_k \in Y_i \}; \]
\[ N_k = \{ x_i \mid (x_i, Y_i) \in T, l_k \notin Y_i \}; \]  
(\( P_k \) and \( N_k \) are called two positive and negative training sample sets for each label \( l_k \), respectively.)

Finally, the last step is to define the predicted label set for \( x \in \mathcal{X} \) sample:
\[ Y = \{ l_k \mid f(\varphi_k(x), l_k) > 0, \ 1 \leq k \leq q \}. \]

2.3. Approach to fuzzy rough relation

In the following, we remind some basic definitions [3, 7, 12] which use throughout this paper.

Let a nonempty universe \( \mathcal{X} \), \( R \) is a similarity relation on \( \mathcal{X} \) where every \( x \in \mathcal{X} \), \( [x]_R \) stands for the similarity class of \( R \) defined by \( x \), i.e. \( [x]_R = \{ y \in \mathcal{X} : (x, y) \in R \} \).

Given \( A \) be the set of condition features, \( B \) be the set of decision feature and \( F \) be a fuzzy set on \( \mathcal{X} \) \( (F: \mathcal{X} \rightarrow [0,1]) \). A fuzzy rough set is the pair of lower and upper approximations of \( F \) in \( \mathcal{X} \) on a fuzzy relation \( R \).

**Definition 1**. Let \( \mathcal{X} \) be a nonempty universe and \( \alpha \) is a feature, \( \alpha \in A \). The fuzzy similarity relation between two patterns \( x \) and \( y \) on the feature \( \alpha \) is determined:
\[ R_\alpha(x, y) = 1 - \frac{|a(x) - a(y)|}{\max_{i=1:n} a(x_i) - \min_{i=1:n} a(x_i)} \]  
(5)

**Definition 2**. Let \( \mathcal{X} \) be a nonempty universe and \( B \) is a feature reduction set, \( B \subseteq A \). The fuzzy similarity relation among all samples in \( \mathcal{X} \) on the reductant \( B \) is determined as follows \( \forall x, y \in \mathcal{X} \):
\[ R_B(x, y) = \min_{\alpha \in B} \{ R_\alpha(x, y) \} \]
\[ = \min_{\alpha \in B} \left\{ 1 - \frac{|a(x) - a(y)|}{\max_{i=1:n} a(x_i) - \min_{i=1:n} a(x_i)} \right\} \]  
(6)

The relationship \( R_B(x, y) \) is the fuzzy similarity relation that satisfies to be reflexive, symmetrical and transitive [2, 13].

Determining the approximations of each fuzzy similarity relation with the corresponding decision set \( D_k \) in the label \( l_k \), respectively.
\[ R_B D(x) = \inf_{y \in \mathcal{X}} \max_{y \in \mathcal{X}} (1 - R(x, y), F(y)); \]
\[ \bar{R}_B D(x) = \sup_{y \in \mathcal{X}} \min_{y \in \mathcal{X}} (R(x, y), F(y)); \]  
(7)

Thus, there may be the method to determine the approximation of \( B \) for \( D_k \) as follows in Eq. (8):

Figure 1. The flowchart of LIFT\(_i\) Classification Model.

Subsequently, the \textit{k-means} clustering is performed to split in \( P_k \), \( N_k \) into discrete clusters with the clustering centers are respectively \( \{ p_1^k, p_2^k, ..., p_{m_k}^k \} \) and \( \{ n_1^k, n_2^k, ..., n_{m_k}^k \} \), in which:
\[ m_k^+ = m_k^- = m_k \left[ |r \cdot \min \{ |P_k|, |N_k| \} \right] \]  
(2)
where \( m_k^+ \) and \( m_k^- \) are the cluster numbers divided in \( P_k, N_k \) respectively; and \( r \) is the ratio parameter controlling the number of given clusters.

Creating the label-specific feature space \( \text{LIFT}_i \) with \( 2m_k \) dimension bases using an appropriable metric to compute distance between samples.
\[ \varphi_k: \mathcal{X} \rightarrow \text{LIFT}_k \]  
(3)
\[ \varphi_k(x_i) = [d(x_i, p_1^k), ..., d(x_i, p_{m_k}^k), \]
\[ d(x_i, n_1^k), ..., d(x_i, n_{m_k}^k)] \]

The second step is to build a family of \( q \) classification models \( \text{LIFT}_i \) \( (1 \leq k \leq q) \) \( \{ f_1, f_2, ..., f_q \} \) respectively for \( l_k \in \mathcal{L} \) labels. In which, a binary training set is created in the form of:
\[ \mathcal{B}_k = \{ (\varphi_k(x_i), \omega(Y_i, l_k)) \mid (x_i, Y_i) \in T \} \]  
(4)
where, \( \omega(Y_i, l_k) = 1, \) if \( l_k \in Y_i, \)
\[ \omega(Y_i, l_k) = -1, \) if \( l_k \notin Y_i \)

We initialize the classification model for each label based on \( \mathcal{B}_k \) as follows:
\[ f_k: \text{LIFT}_k \rightarrow \mathbb{R} \]
\[
R_B D(x) = \inf_{y \in \mathcal{Y}} \max_{x \in \mathcal{X}} \left( 1 - \min_{a \in \mathcal{B}} \left( 1 - \frac{|a(x) - a(y)|}{\max_{i=1}^{n} a(x_i) - \min_{i=1}^{n} a(x_i)} \right) \right) \quad F(y)
\]

(8)

The fuzzy set \( F \) actually affect to the values of the approximations in Eq. (8). The common approach is using Zadeh’s extension principle to determine an appropriate fuzzy set on the given universe \( \mathcal{X} \) [12].

**Definition 3.** Let \( \mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \ldots \times \mathcal{X}_m \) be a nonempty universe and the fuzzy set \( F = F_1 \times F_2 \times \ldots \times F_m \) on the universe \( \mathcal{X} \) with the membership function \( \mu_F(x) = \min\{\mu_{F_1}(x_1), \mu_{F_2}(x_2), \ldots , \mu_{F_m}(x_m)\} \) where \( x = (x_1, x_2, \ldots, x_m) \), \( \mu_{F_i} \) membership function of the fuzzy set \( F_i \) on the universe \( \mathcal{X}_i \), \( i = 1, 2, \ldots, m \).

The mapping \( f: \mathcal{X} \rightarrow \mathcal{Y} \) is determined for the new fuzzy set \( B \) on the universe \( \mathcal{Y} \) with the membership function \( \mu_B(x) \) as follows:

\[
\mu_B(x) = \begin{cases} 
\sup\{\mu_F(x)\} & \text{if } f^{-1}(y) \neq \emptyset \\
0 & \text{if } f^{-1}(y) = \emptyset 
\end{cases}
\]

(9)

where \( f^{-1}(y) = \{x \in \mathcal{X} : f(x) = y\} \).

**Definition 4.** [2, 14]: Let \( R \) be a fuzzy similarity relation on the universe \( \mathcal{X} \) and \( D_k \) is a decision set, \( D_k \subseteq D \) The approximate cardinality represents the dependency of the feature set \( B \) on \( D_k \) in the form is computed as follows:

\[
\gamma(B, D) = \frac{\sum_{x \in \mathcal{X}} P_O S_{B}(D)}{|X|}
\]

(10)

In which, \( |X| \) denotes the cardinality of the set. And \( P_O S_{B}(D) = \bigcup_{x \in \mathcal{X}/D} R_B D(x) \), where \( P_O S_{B}(D) \) is the definite area of the partition \( \mathcal{X}/D \) with \( B \). In fact, \( 0 \leq \gamma(B, D_k) \leq 1 \), its meaning is to represent the proportion of all elements of \( \mathcal{X} \) which can be uniquely classified \( \mathcal{X}/D \) using features \( B \). Moreover, the dependency \( \gamma(B, D_k) \) is always defined on the fuzzy equivalence approximation values of all finite samples.

\( B \) is the best reduced feature set in \( A \) if \( B \) satisfied simultaneously:

\[
\forall B \subseteq A, \gamma(A, D_k) > \gamma(B, D_k) \quad \text{and} \quad \forall B' \subseteq B, \gamma(B', D_k) < \gamma(B, D_k)
\]

(11)

Using threshold \( \varepsilon \) without restrictions [8], \( B \) is the reduction of the set \( A \) if satisfied:

\[
(i) \quad \gamma(A, D) - \gamma(B, D) \leq \varepsilon \\
(ii) \quad \forall C \subseteq B, \gamma(A, D) - \gamma(C, D) > \varepsilon
\]

(12)

The threshold parameter \( \varepsilon \) performs a role in controlling the change of the approximation quality to loosen the limitations of reduction. The purpose of using \( \varepsilon \) is to reduce redundant information as much as possible [13].

2.4. An FRS-LIFT multi-label learning approach

FRS-LIFT is a multi-label learning approach with label-specific feature reduction based on fuzzy rough set [13]. To define the membership functions of the fuzzy lower and upper approximations, Xu et al firstly use a fuzzy set \( F \) followed [1]. Next, they carry out calculating the approximation quality to review the significance of specific dimension using the forward greedy search strategy. They select the most significant features until no more deterministic rules generating with the increasing of features. There are two determined coefficients to identify the significance of a considered feature in the predictable reduction set \( B \) in which: \( \forall a_i \in B, B \subseteq A \):

\[
\text{Sig}_{\text{in}}(a_i, B, D) = \gamma(B, D) - \gamma(B - \{a_i\}, D)
\]

(13)

\[
\text{Sig}_{\text{out}}(a_i, B, D) = \gamma(B + \{a_i\}, D) - \gamma(B, D)
\]

(14)

where \( \text{Sig}_{\text{in}}(a_i, B, D) \) means the significance of \( a_i \) in \( B \) relative to \( D \), and \( \text{Sig}_{\text{out}}(a_i, B, D) \) measures the change of approximate quality when \( a_i \) is chosen into \( B \).

This algorithm improves the performance of multi-label learning using reducing redundant label-specific feature dimensionalities.

However, its computational complexity is high. FRS-SS-LIFT is also be limited time and memory consumption.

3. The label-specific feature reduction for classification model

3.1. Problem formulation

According to LIFT [7], the label-specific space has an expanded dimension 2 times greater
than the number of created clusters. In which, the sample space contains:
\[
A = \{a_1, a_2, \ldots, a_{2m_k}\} = \{p_k^1, p_k^2, \ldots, p_k^{m_k}, n_k^1, n_k^2, \ldots, n_k^{m_k}\}
\]
be the feature sets in \(\mathcal{X}\).

\[\forall x_i \in \mathcal{X}, i = 1, n \text{ be the feature vector, } x_i = [x_i^1, \ldots, x_i^{2m_k}], \text{ each } x_i^j \text{ be a distance } d(x_i, p_k^j).\]

\[D_k = [d_k^1, d_k^2, \ldots, d_k^n] \text{ be the decided classification},\]

\[d_k^i = 1 \text{ if } x_i \in l_k; d_k^i = 0 \text{ if } x_i \notin l_k.\]

Thus, when we have the multi-label training set \(\mathcal{T}\) and the necessary input parameters, the obtained result is a predicted label set \(\mathcal{Y}\) for any sample \(x\). In order to be able to have an effective set \(\mathcal{Y}\), it is necessary to solve the label-specific feature reduction [8]. Therefore, our main goal is to build a classification model that represents the mapping form: \(\mathcal{T}: \mathcal{X} \rightarrow \text{FRR-MLL}_k\)

This proposed task is to build the feature reduction space \(\text{FRR-MLL}_k\) based on the properties of the fuzzy rough relation to satisfy:

- Selecting a better fuzzy set for determining the degree of the membership function of approximates.

- The feature \(a_i\) which has the highest dependency \(\gamma(A_i, D_k)\) is chosen into the reduced feature set \(B\) in this space \((B \subseteq A)\) on \(D_k\). This work is performed if \(B\) satisfy Eq. 11 and \(\gamma(A, D) - \gamma(B, D)\) obtains the greatest value with the threshold parameter \(\epsilon \in [0, 0.1]\).

3.2. Reducing the feature set for multi-label classification

In this subsection, we propose the reductive feature set \(B\) be satisfied simultaneously: The dependency of the feature which is added into reduction set \(B\) on the partition \(X/D\), \(\gamma(a_i, D)\) is the greatest one.

The dependency difference between the initial feature in the set \(A\) with \(D_k\) and the dependency between the reduced feature set \(B\) with \(D_k\) must be within the given threshold \(\epsilon (\epsilon \in [0, 0.1])\), et., \(\gamma(A, D_k) - \gamma(B, D_k) \leq \epsilon;\)

We focus on selecting the proposed feature into the reduction set \(B\) and conducted experimentally on many datasets:

- The feature that has the greatest dependency and was determined from the fuzzy approximations on the samples, is first selected to be included in the set \(B\).

- Next, other features are considered to be included in the reduction set \(B\) if guaranteed using threshold \(\epsilon\) without restrictions [13] i.e. \(B\) is the reduction of the set \(A\) if satisfied Eq. (11).

We note that finding a good fuzzy set is more meaningful for identification between elements. It directly affects the result of the membership function of approximates. In fact, searching a great fuzzy set to model concepts can be challenging and subjective, but it is more significant than making an artificial crisp distinction between elements [5]. Here, we temporarily based on the cardinality of a fuzzy set \(F\) by determining the sum of the membership values of all elements in \(\mathcal{X}\) to \(F\).

For examples: Given the set \(\mathcal{X}\) by the under table and the dependency parameter \(\epsilon = 0.1\), we respectively determine the fuzzy equivalence relationship \(R_k(x, y)\) and the lower approximation of the features with \(D_k\) before calculating the dependencies \(\gamma(A_i, D_k)\) and \(\gamma(a_i, D_k)\):

|  | \(a_1\) | \(a_2\) | \(a_3\) | \(a_4\) | \(d_k\) |
|---|---|---|---|---|---|
| \(x_3\) | 3.3 | 2.0 | 3.0 | 4.2 | 1.0 |
| \(x_2\) | 1.1 | 3.8 | 1.7 | 2.3 | 1.0 |
| \(x_3\) | 2.0 | 4.7 | 2.1 | 2.5 | 0.0 |
| \(x_4\) | 2.9 | 4.2 | 2.9 | 1.8 | 0.0 |
| \(x_5\) | 1.9 | 2.5 | 1.7 | 2.9 | 0.0 |
| \(x_6\) | 2.4 | 1.7 | 2.3 | 3.1 | 1.0 |
| \(x_7\) | 2.5 | 3.9 | 2.3 | 1.6 | 0.0 |

First, we choose the feature \(a_4\) and add it to the set \(B\). Next, we select the feature \(a_1\) and add it to the set \(B\). Calculate \(\gamma(B, D) = 0.15\), we obtained: \(\gamma(A, D) - \gamma(B, D) = \epsilon\):

\[\gamma(A, D_k) = 0.25,\]

\[\gamma(a_1, D_k) = 0.092\]

\[\gamma(a_2, D_k) = 0.07\]

\[\gamma(a_3, D_k) = 0\]

\[\gamma(a_4, D_k) = 0.094\]
4. The proposed algorithms

4.1. The specific feature reduction algorithm

Finding the optimal reductive set from the given set A is seen as the significant phase. It is necessary to decide the classification efficiency. So, we propose a new method FRR_RED to search an optimal set.

Algorithm 1: FRR-RED algorithm

**Input:** The finite set of n samples $X$; The set of condition features $A$; The set of decision $D$; The threshold $\varepsilon$ for controlling the change of approximate quality.

$X = \{x_1, ..., x_n\}$,
$A = \{a_1, ..., a_{2+m}\}$, $D = \{d_1, ..., d_n\}$

**Output:** Feature reduction $B$.

Method:

1. For each $a_i \in A$
   1. Compute $\gamma(A, D), \gamma_l = \gamma(a_i, D) \forall a_i \in A$ according to Eq. (5);
   2. Create $B = \{\}$; $\gamma(B, D) = 0$;
   3. If $(\gamma(A, D) - \gamma(B, D) > \varepsilon)$ then
   5. Compute $\gamma_{\text{max}}$ for $\forall a_i \in A \text{ and } \forall a_i \notin B$
   6. If $\gamma_{\text{max}}$ then $B = B \cup \{a_j\}$
   7. Compute $\gamma(B, D)$ by Eq. (10);
   8. End if
   9. End if
   10. End for

From step 4 to step 11, selecting the features that have the highest dependency to put into the reductive set $B$ and this is implemented continuously until satisfy Eq. (11). This proposed method which hopefully finds the optimal reductive set is different to the previous approach because this selecting process is not random.

4.2. Approach to FRR-MLL for multi-label classification with FRR-RED

Improving the FRS-LIFT algorithm [8], we apply the above FRR-LIFT algorithm to step 5, details as follows:

Algorithm 2: FRR-MLL algorithm

**Input:** The multi-label training set $\mathcal{T}$, The ratio parameter $r$ for controlling the number of clusters; The threshold $\varepsilon$ for controlling the change of approximate quality; The unseen sample $x'$.

**Output:** The predicted label set $Y'$.

Method:

1. For $k = 1$ to $q$ do
2. Form the set of positive samples $\mathcal{P}_k$ and the set of negative samples $\mathcal{N}_k$ based on $\mathcal{T}$ according to Eq. (1);
3. Perform $k$-means clustering on $\mathcal{P}_k$ and $\mathcal{N}_k$, each with $m_k$ clusters as defined in Eq. (2);
4. $\forall (x_i, y_i) \in \mathcal{T}$, create the mapping $\varphi_k(x_i)$ according to Eq. (3), form the original label-specific feature space $\mathcal{LIFT}_k$ for label $l_k$;
5. Perform find decision reduct $B$ such as FRR-RED:
6. With $B$, form the dimension-reduced label-specific feature space $\mathcal{FRR-MLL}_k$ for label $l_k$ (etc., mapping $\varphi'_k(x_i)$);
7. End for
8. For $k = 1$ to $q$ do
9. Construct the binary training set $\mathcal{T}_k^*$ in $\varphi_k(x_i)$ according to Eq. (4);
10. Induce the classification model $f_k$: $\mathcal{FRR-MLL}_k \rightarrow \mathbb{R}$ by invoking any binary learner on $\mathcal{T}_k^*$;
11. End for
12. The predicted label set:
13. $Y = \{l_k | f_k(\varphi'_k(x_i)) > 0, 1 \leq k \leq q\}$

The FRR-MLL algorithm is performed to create the $\mathcal{FRR-LIFT}_k$ space, then reduce the label-specific feature based on selecting the maximum dependency of the features. The dataset on the reductive feature set is trained in the next step. Finally, build the classification model $\mathcal{FRR-MLL}_k$ and make the label prediction set $Y$ for the element $x'$.

We calculate the time complexity of FRR-LIFT and compare to FRS-LIFT. The result shows that the proposed algorithm is better.
The time complexity of FRS-LIFT [12] as following:
\[ O(m_k(t_1|P_k| + t_2|N_k|) + 2m_k|\mathcal{T}| + 2t_3|\mathcal{T}| + 4m_k^2|\mathcal{T}|^2) \]

And the time complexity of FRR-LIFT is shown below:
\[ O(m_k(t_1|P_k| + t_2|N_k|) + 2m_k|\mathcal{T}| + 4|\mathcal{T}|m_k) \]
where \( t_1, t_2, t_3 \) are the iterations of k-means on \( P_k, N_k \) and \( |\mathcal{T}| \), respectively.

Table 1 shows the detailed computing steps of FRS-LIFT and FRR-LIFT. Basically, the time complexity is the same, but the only difference is in reducing feature step. With the proposed algorithm, we prioritize selecting the features with the highest dependency in order to satisfy the conditions of Eq. (12). On the other hand, while reducing, we determine to calculate the approximations of the samples on partition \( X/D_k \). This work decreases some computing steps, thus, the time complexity of FRR-LIFT is more optimal than FRS-LIFT’s.

| Order | Steps | The time complexity of FRR-LIFT | The time complexity of FRS-LIFT |
|-------|-------|--------------------------------|--------------------------------|
| 1     | Clustering on \( P_k \) and \( N_k \) using k-means | \( O\left(m_k(t_1|P_k| + t_2|N_k|)\right) \) | \( O\left(m_k(t_1|P_k| + t_2|N_k|)\right) \) |
| 2     | Creating the label-specific feature space \( \text{LIFT}_k \) | \( O(2m_k|\mathcal{T}|) \) | \( O(2m_k|\mathcal{T}|) \) |
| 3     | Selecting samples on the label-specific feature space \( \text{LIFT}_k \) | \( 2t_3|\mathcal{T}| \) | \( 2t_3|\mathcal{T}| \) |
| 4     | Reducing features using the fuzzy rough relationship | \( O(4|\mathcal{T}|m_k) \) | \( O(4|\mathcal{T}|m_k) \) |
| 5     | Total time complexity | \( O\left(m_k(t_1|P_k| + t_2|N_k|) + 2m_k|\mathcal{T}| + 2t_3|\mathcal{T}| + 4|\mathcal{T}|m_k\right) \) | \( O\left(m_k(t_1|P_k| + t_2|N_k|) + 2m_k|\mathcal{T}| + 2t_3|\mathcal{T}| + 4m_k^2|\mathcal{T}|^2\right) \) |

5. Conclusion

The paper proposed the algorithm for reducing the set of features. Finding the most significant features can determine the new reduction set rapidly, because we do not have to calculate all most features if the reduction set satisfy all conditions to be verified. In the future, we continue to conduct experiments on real databases to evaluate the efficiency of the proposed algorithms and improve the fuzzy set \( F \) which is the set of the membership functions on \( X \).

References

[1] Richard Jensen, Chris Cornelis, Fuzzy-Rough Nearest Neighbor Classification and Prediction. Proceedings of the 6th International Conference on Rough Sets and Current Trends in Computing, 2011, pp. 310-319.
[2] Y.H. Qian, Q. Wang, H.H. Cheng, J.Y. Liang, C.Y. Dang, Fuzzy-Rough feature selection accelerator, Fuzzy Syst Syst. 258 (2014) 61-78.
[3] Quang-Thuy Ha, Thi-Ngân Pham, Van-Quang Nguyen, Minh-Chau Nguyen, Thanh-Huyen Pham, Tri-Thanh Nguyen, A New Text Semi-supervised Multi-label Learning Model Based on Using the Label-Feature Relations, International Conference on Computational Collective Intelligence, LNAI 11055, Springer, 2018, pp. 403-413.
[4] Daniel Kostrzewa, Robert Brzeski, The data Dimensionality Reduction and Feature Weighting in the Classification Process Using Forest Optimization Algorithm, ACIDS, 2019, pp. 97-108.
[5] Nele Verbiest, Fuzzy Rough and Evolutionary Approaches to Instance Selection, PhD Thesis, Ghent University, 2014.
[6] Y. Yu, W. Pedrycz, D.Q. Miao, Multi-label classification by exploiting label correlations, Expert syst. Appl. 41 (2014) 2989-3004.
[7] M.L. Zhang, LIFT: Multi-label learning with label-specific features, IEEE Trans, Pattern Anal, Mach, Intell 37 (2015) 107-120.

[8] Suping Xu, Xibei Yang, Hualong Yu, Dong-Jun Yu, Jingyu Yang, Eric CC Tsang, Multi-label learning with label-specific feature reduction, Knowledge-Based Systems 104 (2016) 52-61. https://doi.org/10.1080/24751839.2017.1364925.

[9] Thi-Ngan Pham, Van-Quang Nguyen, Van-Hien Tran, Tri-Thanh Nguyen, Quang-Thuy Ha, A Semi-supervised multi-label classification framework with feature reduction and enrichment, Journal of Information and Telecommunication 1(4) (2017) 305-318.

[10] M. Ghaemi, M.R. Feizi-Derakhsh, Feature selection using forest optimization algorithm, Pattern Recognition 60 (2016) 121-129.

[11] M.L. Zhang, Z.H. Zhou, ML-KNN: A lazy learning approach to multi-label learning, Pattern Recognition 40 (2007) 2038-2048.

[12] M.Z. Ahmad, M.K. Hasan, A New Approach for Computing Zadeh's Extension Principle, MATHEMATIKA. 26(1) (2010) 71-81.

[13] Richard Jensen, Neil Mac Parthaláin and Qiang Shen. Fuzzy-rough data mining (using the Weka data mining suite), A Tutorial, IEEE WCCI 2014, Beijing, China, July 6, 2014.

[14] D. Dubois, H. Prade. Rough fuzzy sets and fuzzy rough sets, Int. J. Gen. Syst. 17 (1990) 191-209.