Novel fuzzy measure of similarity for fuzzy-rough feature selection

Andreja Naumoski, Miroslav Mircev
Ss. Cyril and Methodius University in Skopje, Faculty of Computers Science and Engineering, Skopje, Macedonia
{andreja.naumoski, miroslav.mircev}@finki.ukim.mk

Abstract. Many real datasets contain attributes that are not relevant to the task at hand, even more, they are not correlated with the output class. Some of the attributes may be redundant, and others may be irrelevant and noisy. The process of features selection aims to find those attributes. In order to improve the process of feature selection, in this paper we introduce a novel metric for fuzzy-rough feature selection (FRFS). The new fuzzy similarity metric is evaluated using Lukasiewicz and Kleene-Dienes fuzzy connectives for both lower and upper approximations. The experimental evaluation show that the datasets obtained with FRFS, compared with the raw datasets, achieve improvement of the ROC accuracy for some classification algorithms. Most important is that this is achieved with method that requires no information, threshold or domain information. The feature selected datasets achieve dimensionality reduction, not compromising the ROC accuracy.

Keywords: Fuzzy rough sets, feature selection, dimensionality reduction, classification.

1 Introduction

The task of knowledge discovery or pattern discovery is much easier and faster by removing attributes of the dataset that are redundant. This process in machine learning is called feature selection (FS) [1]. FS has long been a research area within machine learning, statics and pattern recognition scientific disciplines [2, 3]. Compared to the process of pre-processing and constructing new data, FS is well defined. There are several advantages of using FS; reduction the amount of data to obtain the model, improved ROC accuracy, models that are compact and easy interpretable, and allowing learning algorithms to operate faster and more effectively. Sometimes, accuracy on future classification is improved; or in other cases, the model is a more compact, easily interpreted representation of the classification task. For example, decision tree algorithms such as C4.5 [4, 5] can sometimes overfitting the training data, resulting in large trees. In many cases, removing irrelevant and redundant information can result in C4.5 producing smaller trees [6] and thus reducing overfitting. FS methods for machine learning are divided into two wide-ranging categories. The first category, known as wrappers [6, 7],
are methods that evaluate the worth of features using the learning algorithm that is ultimately being applied to the data. And the second category, known as filters [6, 7], are methods that evaluate the worth of features by using heuristics based on general characteristics of the data. Both these methods have advantages and disadvantages regarding the property of any data, mainly the correlation between the attributes and the class.

When comparing two attributes and their values, it is not easy to say that they are similar or to what extent they are similar. Two close values may differ typical as a result of noise. Rough set theory (RST) [8] is one of the most distinct and recent approaches for modelling imperfect knowledge. RST can be used for feature selection, and in the same time preserving the semantics or meaning of the features. It is possible to distinct these two attribute values, and attribute affected by noise are considered to be as different as two values of a different order of magnitude. The rough set is a set of pair of precise concepts, called lower and upper approximations that are used of classification of the objects of interest into disjoint categories. The concept of lower approximation is a description of the domain objects that are known with certainty to belong to the subset of interest. On the other hand, the upper approximation is a description of the objects that possibly belong to the subset. Until now, rough theory has been successfully used for hidden pattern discovery from data, data reduction, data significance evaluation, decision rule generation, and data-driven inference interpretation [9]. However, this does not mean that the rough theory has no flaws. In fact, by using only the dataset information, this theory assumes that the data are a true and accurate reflection of the real world. Which in many cases this is not true; handling noisy data is just one example. This method is data-driven and not threshold information is needed. Another drawback is reliance on discrete data for the successful operation. Indeed this requirement of rough set theory implies objectivity in the data that is simply not present [10]. Moreover, subjective judgment is required for establishing boundaries by the objective measurements. A more intelligent way to handle this problem is use of the fuzzy-rough sets (FRS) [11].

As we mention in [12], fuzzification aids uncertainty modelling by allowing the possibility of the membership of a value to more than one fuzzy term. A fuzzy property of many datasets arises due to a lack of sharp distinctions or boundaries in the data, while rough sets can model ambiguity resulting from a lack of information through set approximations. FRS is promising approach to combines these advantages by extending the fuzzy theory based on similarity or tolerance measures [13]. This method encapsulates the related but distinct concepts of imprecision [14] and indiscernibility (for rough sets), both of which occur as a result of uncertainty in knowledge. Since the rough sets are sets of pair of precise concepts, here the FRS considers two fuzzy sets. This is a product of extending the corresponding crisp set notions. The lower and upper approximations, are converted into fuzzy lower and upper approximations. Fuzzy lower approximations in crisp case have membership of 1, and they belong to the approximated set with absolute certainty. On other hand, element members of upper approximations may have membership in a range [0, 1], allowing greater flexibility in handling uncertainly. Using these types of sets, subjective judgments are not entirely removed, because fuzzy set membership functions still need to be defined in order fuzzy set theory
to be applied. However, the FRFS method offers a high degree of flexibility when dealing with real-valued data, enabling the vagueness and imprecision present to be modelled effectively [15].

The work presented here is in this paper, combines fuzzy-rough sets and feature selection on real-valued data using FRFS [16]. It employs the fuzzy-rough discernibility measure to examine the level of discernibility between a single feature and subsets of other features. The central idea behind the latter, is that any single feature which carries little or no further information than that subsumed by the remaining features is redundant and can therefore be eliminated [17, 18, 19]. Therefore, we proposed a novel similarity metric that combines the fuzzy information to reduce the search space of fuzzy terms. We also provide comparison of this metric with other previous used metrics [20]. The experiments are made on both regression and classification datasets.

The rest of the paper is organized as follows. Section 2 presents description of the rough and fuzzy-rough sets, and the process of FRFS reduction. In section 3, the experimental data used for test the proposed fuzzy similarity metrics and the results, while Section 4 concludes the paper and gives direction for future research.

2 Fuzzy – rough feature selection process

2.1 What is rough set?

Let’s assume that U is a nonempty set of finite objects (the universe of discourse) and A is a nonempty finite set of attributes such that a: U → V_a for every a ∈ A. V_a is the set of values that attribute a may take. Let’s now consider a dataset that consists from attributes and class indexes. A = {C ∪ D}, where C is the set of input features and D is the set of class indexes. Here, a class index d ∈ D is itself a variable d: U → {0, 1} such that for a ∈ U, d(a) = 1 if a has class d and d(a) = 0 otherwise. Now let’s assume that X ⊆ U and P ⊆ A. With any P ⊆ A there is an associated equivalence relation IND(P):

\[ IND(P) = \{(x, y) \in U^2 \mid \forall a \in P, a(x) = a(y)\} \]  

(1)

The partition of U, generated by IND(P) is denoted U/IND(P) (or U/P for simplicity) and can be calculated as follows:

\[ U / IND(P) = \mathcal{X}\{U / IND(\{a\}) \mid a \in P\} \]  

(2)

where \( \mathcal{X} \) is specifically defined as follows for sets A and B:

\[ A \mathcal{X} B = \{X \cap Y \mid X \in A, Y \in B, X \cap Y \neq \emptyset\} \]  

(3)

X can be approximated using only the information contained within P by constructing the P-lower and P-upper approximations of the classical crisp set X:

\[ PX = \{x \in U \mid [x]_P \subseteq X\} \]  

(4)

\[ \overline{PX} = \{x \in U \mid [x]_P \cap X \neq \emptyset\} \]  

(5)
The tuple \( \langle PX, \overline{PX} \rangle \) is called rough set. Let’s now assume that \( P \) and \( Q \) are set off attributes indicating equivalence relations over \( U \), so the positive, negative and boundary region are defined as:

\[
POS_p(Q) = \bigcup_{x \in U \cap Q} PX
\]

(6)

\[
NEG_p(Q) = U - \bigcup_{x \in U \cap Q} \overline{PX}
\]

(7)

\[
BND_p(Q) = \bigcup_{x \in U \cap Q} \overline{PX} - \bigcup_{x \in U \cap \overline{Q}} PX
\]

(8)

According rough set definitions [20], the positive region compromises from all the objects of \( U \) that can be classified to classes of \( U/Q \), using the information contained in attribute \( P \). In contrary the negative region, contains set of objects that cannot be classified to classes of \( U/Q \). While the boundary region (see eq. 8) is the set of objects that can possible, but not certainly, be classified in this way. Another very important issue in data analysis is discovering the dependencies between attributes. In rough set theory, dependency - \( P \) is define as: \( Q \subset A \), it is said that \( Q \) depends on \( P \) in a degree \( k \) (\( 0 \leq k \leq 1 \)), denoted \( P \Rightarrow_k Q \), if

\[
k = \gamma_p(Q) = \frac{|POS_p(Q)|}{|U|}
\]

(9)

where \(|U|\) stands for the cardinality of set \( U \). If \( k=1 \), \( Q \) totally depends on \( P \), if \( k \) is between 0 and 1, \( Q \) partially depends from \( P \), and if \( k=0 \) \( Q \) does not depend from \( P \). Because there are multiple features in one set, calculating change in dependency when a features is removed from the set, an estimate of the significance of that feature have to be obtained. The higher the change in dependency, the more significant is the feature, otherwise the feature is dispensable. The concept of reduct is defined as minimal subset \( R \) of the initial attribute set \( C \) such that for a given set of attributes \( D \), \( \gamma_R(D) = \gamma_C(D) \). \( R \) is minimal if \( \gamma_{R\setminus\{a\}}(D) \neq \gamma_R(D) \) for all \( a \in R \). Real valued dataset may have many reduct sets, and the collection of all reducts is represented by:

\[
R_{all} = \{ X | X \subseteq C, \gamma_X(D) = \gamma_{C}(D); \gamma_{X\setminus\{a\}}(D) \neq \gamma_X(D), \forall a \in X \}
\]

(10)

The intersection of all sets in \( R_{all} \) is called core. These elements are features that cannot be eliminated without introducing more contradictions to the dataset. For the task of feature selection, a reduct minimal cardinality is ideal to search for. This means locating a single element of the reduct set \( R_{min} \subseteq R_{all} \):

\[
R_{min} = \{ X | X \in R_{all}, \forall Y \in R_{all}, |X| \leq |Y| \}
\]

(11)
2.2 What is fuzzy-rough set?

The fuzzy lower and upper fuzzy approximations are redefined as:

\[
\mu_{F_i}^L(F) = \sup_{x \in U/P} \min\{\mu_F(x), \inf_{y \in U} \max\{1 - \mu_F(y), \mu_X(y)\}\}
\]

\[
\mu_{F_i}^U(F) = \sup_{x \in U/P} \min\{\mu_F(x), \sup_{y \in U} \min\{\mu_F(y), \mu_X(y)\}\}
\]

where \(F_i\) denotes a fuzzy equivalence class belonging to \(U/P\). From implementation point of view, not all \(y \in U\) need to be considered - only those where \(\mu_{F_i}(y)\) is nonzero, and \(y\) is fuzzy member of class \(F\). The tuple \(\{P_X, \bar{P}_X\}\) is called fuzzy-rough set (FRS). For FRS, the crisp positive region is defined as union of the lower approximations. Extending this notion with the extension principle [21], the membership of an object \(x \in U\), belonging to the fuzzy positive, negative and boundary regions can be defined by

\[
\mu_{POS_{R_P}}(x) = \sup_{X \in U/Q} \mu_{R_{P,X}}(x)
\]

\[
\mu_{NEG_{R_P}}(x) = \inf_{X \in U/Q} \mu_{\bar{R}_{P,X}}(x)
\]

\[
\mu_{BND_{R_P}}(x) = \mu_{R_{P,X}}(x) - \mu_{\bar{R}_{P,X}}(x)
\]

Objects like object \(x\), belong to the positive region only if their underlying corresponding class do so. So, object \(x\) will not belong to the positive region only if the corresponding class it belongs to is not a constituent of the positive region. The negative and boundary regions can be also defined in similar way. It is important to note, that fuzzy-rough negative regions is always empty when the decisions are crisp, but not empty when we consider fuzzy decisions.

Another important property of the FRS is calculating the dependency function. This function shows how much feature A or feature B is dependable from subset of features. The dependency function is define as follows:

\[
\lambda_p(Q) = \frac{\sum_{X \in U/Q} U_p(X)}{|U/Q|}
\]

The dependency of \(Q\) on \(P\) is the proportion of objects that are discernible out of the entire dataset. If the sum function returns 1, means that the object \(x\) belongs to the positive region, otherwise the function is 0. If \(Q = \{a, b\}\) or contains two fuzzy features, objects could belong to many equivalent classes, the Cartesian product of \(U/Q\) have to be consider in determining \(U/Q\). In case of huge number of feature subsets, the complexity of Cartesian product calculation becomes prohibitively high. Some optimizations to eliminate these problems are given in [16, 22], however the complexity is still too high. Further improvement of the method is presented in [23], by using different calculation for fuzzy lower approximation. A second problem is it was shown in [24] that in some situations the fuzzy lower approximation might not be a subset of the fuzzy...
upper approximation. This suggests that there is more certainty in the fuzzy upper than the fuzzy lower approximation.

\[ \mu_{R_{p}(X)}(x) = \inf_{y \in U} I(\mu_{R_{p}(X)}(x, y), \mu_{X}(y)) \]  
\[ \mu_{R_{p}(X)}(x) = \sup_{y \in U} T(\mu_{R_{p}(X)}(x, y), \mu_{X}(y)) \]

where \( I \) is a fuzzy implicator and \( T \) is a t-norm. \( R_{p} \) represent fuzzy similarity relation induced by the subset of features \( P \):

\[ \mu_{R_{p}(X)}(x, y) = \bigcap_{a \in P} \{ \mu_{Ra}(x, y) \} \]

\( \mu_{Ra}(x, y) \) is the degree to which objects \( x \) and \( y \) are similar for feature \( a \). However, here we use only lower approximation for evaluation of features subsets. In order to be useful this information, which is part of the definition of the lower and upper approximation, the subset have to have less uncertainty of objects near boundary region. Therefore, the resulting degree of dependency taking into account the boundary region is:

\[ \gamma_{p}(Q) = \frac{\sum_{x \in U} \mu_{BND_{R_{p}(x)}}(x)}{|U|}, \text{ where } U_{p}(X) = \frac{\sum_{x \in U/Q} U_{p}(x)}{|U/Q|} \]

From eq. 18, the overall uncertainty is evaluated via averaging the uncertainty of all decision concepts. A better way of measuring the uncertainty within the boundary region of a concept \( X \) is to calculate the fuzzy entropy:

\[ U_{p}(X) = \sum_{x \in U} \frac{\mu_{BND_{R_{p}(x)}}(x)}{|BND_{R_{p}(x)}(x)|} \log_{2} \frac{\mu_{BND_{R_{p}(x)}}(x)}{|BND_{R_{p}(x)}(x)|} \]

\[ \gamma_{p}(D) = \frac{\sum_{D \in R_{p}} U_{p}(D)}{\sum_{D \in R_{p}} |D|^{-1}} \]

This will be minimized when all fuzzy boundary regions are empty. Therefore, a fuzzy-rough reduct \( R \), or the process of finding FRFS can be defined as subset of features that preserves the dependency degree of the entire dataset, so that \( \gamma_{p}(D) = 0 \).

In order to calculate \( \mu_{Ra}(x, y) \) and preserve the T – transitivity [26], several fuzzy
similarity relations have been constructed, and one of them defined with eq. 24, will be taken into account:

$$\mu_{\mu_0}(x, y) = \max(0, \min(1, \frac{\beta - \alpha \ast (a(x) - a(y))}{\sigma_a}))$$  \hspace{1cm} (24)

where ($\sigma_a$)$^2$ is the variance of feature $a$. In this paper, we propose modification of this equation, in order not to be dependable from outside definition of the parameters $\alpha$ and $\beta$. The objects $x$ and $y$ depend directly from the variance of the feature $a$.

$$\mu_{\mu_0}(x, y) = \max(0, \min(1, \frac{\sigma_a}{\sigma_a} \ast |a(x) - a(y)|))$$  \hspace{1cm} (25)

Both equations 24 and 25, will be evaluated for classification task over nine benchmark datasets from [12, 27] with three classifiers: J48 [5], Naïve Bayes (NB) [28] and JRip [29].

3 Experimental setup and results

In this section, we present the results from the experimentally evaluation of the proposed metric (see eq. 25 – Metric 25) and compare the results from the method that use the eq. 24 (Metric 24). In [30], the authors recommend using the Łukasiewicz (Łuk) t-norm to produce fuzzy T-equivalence relations. Here, we test both Łuk and KD fuzzy connectives and later we evaluate them with three classification algorithms. The Łuk t-norm is defined as ($\max(x + y - 1, 0)$), while the Łuk fuzzy implicator ($\min(1 - x + y, 1)$). For both teste we have used Greedy Stepwise method as a search method. Similarly, the Kleene-Dienes (KD) t-norm is defined as ($\min(x, y)$) and while KD implicator as ($\max(x, y)$).

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Objects</th>
<th>Attributes</th>
<th>Metric 24</th>
<th>Metric 24</th>
<th>Metric 25</th>
<th>Metric 25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Credit-g</td>
<td>670</td>
<td>21</td>
<td>13</td>
<td>11</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>Dermatology</td>
<td>365</td>
<td>34</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>Ionosphere</td>
<td>351</td>
<td>34</td>
<td>33</td>
<td>11</td>
<td>23</td>
<td>32</td>
</tr>
<tr>
<td>Mushroom</td>
<td>8124</td>
<td>22</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Water 1</td>
<td>218</td>
<td>116</td>
<td>70</td>
<td>2</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>Water 2</td>
<td>218</td>
<td>116</td>
<td>73</td>
<td>3</td>
<td>13</td>
<td>6</td>
</tr>
<tr>
<td>Water 3</td>
<td>201</td>
<td>116</td>
<td>71</td>
<td>3</td>
<td>14</td>
<td>5</td>
</tr>
<tr>
<td>Water 4</td>
<td>190</td>
<td>116</td>
<td>61</td>
<td>3</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td>Water 5</td>
<td>218</td>
<td>116</td>
<td>72</td>
<td>3</td>
<td>11</td>
<td>6</td>
</tr>
</tbody>
</table>

After FRFS, the datasets are reduced according to the presented reducts. These reduced datasets are then classified using the relevant classifier. The results from the FRFS process are shown in Table 1. Table 1 results depict reduction of the features in all of the datasets, except for the datasets where KD norm with Metric 24 is applied. The Metric
25, in the case of using the Łuk operator, the method obtains lower attribute reduction, while using the KD operator, all the datasets obtain significant attribute reduction. However, this doesn’t mean that the ROC accuracy will improve either.

Three classifiers were employed for the purpose of evaluating the resulting subsets from the FRFS phase. Table 2 compares the results obtained reduced datasets with the raw dataset set. The test evaluation was obtained using 2/3 train and 1/3 test set from the original dataset. The ROC analysis is conduct to compare the performance of the obtain results.

Table 2. Experimental evaluation of both types of datasets (raw and reduced datasets). Bolded results are highest ROC value between metrics 24 and 25 using KD operator. Underlined results represent the highest ROC values between metric 24 and 25 using Łuk operator. The results market with (*), show the highest ROC values for each dataset.

<table>
<thead>
<tr>
<th>FRFS</th>
<th>Classification</th>
<th>Credit</th>
<th>Dermatology</th>
<th>Ionosphere</th>
<th>Mushroom</th>
<th>Water 1</th>
<th>Water 2</th>
<th>Water 3</th>
<th>Water 4</th>
<th>Water 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw</td>
<td>J48 0.69</td>
<td>0.94</td>
<td>0.88</td>
<td>0.85</td>
<td>0.5</td>
<td>0.53</td>
<td>0.76</td>
<td>0.49</td>
<td>0.52</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NB 0.79*</td>
<td>0.99*</td>
<td>0.94*</td>
<td>0.99*</td>
<td>0.63</td>
<td>0.62</td>
<td>0.66</td>
<td>0.43</td>
<td>0.54</td>
<td></td>
</tr>
<tr>
<td></td>
<td>JRip 0.69</td>
<td>0.89</td>
<td>0.87</td>
<td>0.99*</td>
<td>0.55</td>
<td>0.53</td>
<td>0.72</td>
<td>0.48</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>KD</td>
<td>J48 0.71</td>
<td>0.94</td>
<td>0.88</td>
<td>0.85</td>
<td>0.57</td>
<td>0.53</td>
<td>0.62</td>
<td>0.45</td>
<td>0.48</td>
<td></td>
</tr>
<tr>
<td>(24)</td>
<td>NB 0.76</td>
<td>0.99*</td>
<td>0.94*</td>
<td>0.83</td>
<td>0.63</td>
<td>0.59</td>
<td>0.68</td>
<td>0.34</td>
<td><strong>0.52</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>JRip 0.59</td>
<td>0.95</td>
<td>0.86</td>
<td>0.83</td>
<td>0.54</td>
<td>0.50</td>
<td>0.62</td>
<td>0.48</td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td>Łuk</td>
<td>J48 0.70</td>
<td>0.98</td>
<td>0.75</td>
<td>0.92</td>
<td>0.58</td>
<td>0.50</td>
<td>0.52</td>
<td>0.49</td>
<td>0.53</td>
<td></td>
</tr>
<tr>
<td>(24)</td>
<td>NB 0.74</td>
<td>0.92</td>
<td>0.87</td>
<td>0.89</td>
<td>0.55</td>
<td>0.57</td>
<td>0.54</td>
<td>0.35</td>
<td>0.56</td>
<td></td>
</tr>
<tr>
<td></td>
<td>JRip 0.54</td>
<td>0.95</td>
<td>0.75</td>
<td>0.75</td>
<td>0.57</td>
<td>0.50</td>
<td>0.59</td>
<td>0.47</td>
<td><strong>0.57</strong></td>
<td></td>
</tr>
<tr>
<td>KD</td>
<td>J48 0.71</td>
<td>0.94</td>
<td>0.84</td>
<td><strong>0.85</strong></td>
<td>0.66</td>
<td>0.57</td>
<td>0.72</td>
<td><strong>0.51</strong></td>
<td>0.45</td>
<td></td>
</tr>
<tr>
<td>(25)</td>
<td>NB 0.76</td>
<td><strong>0.99</strong></td>
<td>0.93</td>
<td>0.83</td>
<td><strong>0.73</strong></td>
<td>0.62</td>
<td>0.31</td>
<td>0.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>JRip 0.59</td>
<td>0.95</td>
<td>0.82</td>
<td>0.83</td>
<td>0.54</td>
<td>0.52</td>
<td>0.78</td>
<td>0.50</td>
<td>0.45</td>
<td></td>
</tr>
<tr>
<td>Łuk</td>
<td>J48 0.54</td>
<td>0.98</td>
<td>0.73</td>
<td>0.92</td>
<td>0.50</td>
<td>0.49</td>
<td>0.50</td>
<td>0.50</td>
<td>0.54</td>
<td></td>
</tr>
<tr>
<td>(25)</td>
<td>NB 0.68</td>
<td>0.92</td>
<td>0.91</td>
<td>0.89</td>
<td>0.59</td>
<td>0.52</td>
<td>0.56</td>
<td>0.35</td>
<td>0.57</td>
<td></td>
</tr>
<tr>
<td></td>
<td>JRip 0.53</td>
<td>0.95</td>
<td>0.66</td>
<td>0.75</td>
<td>0.55</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.54</td>
<td></td>
</tr>
</tbody>
</table>

The results shown as (*), have point out that combination of the KD operator with the proposed metric 25, have in many cases achieve better results than raw dataset in classification tasks. On the other hand, the combination of KD operator with the previous metric 24, have achieve same results in 2 out of 9 dataset, compared with the raw dataset analysis. It is interesting to note, that the combination of both metrics with the Łuk operator do not achieved any improvement despite the fact that, according Table 1, this operator obtain best results in FRFS process. Regarding the classification algorithms, the J48 and Naïve Bayes methods obtain best results compared with the JRip method among the tested datasets.
4 Conclusion

In this paper we proposed a novel fuzzy similarity metric that could be used in the process of FRFS. This metric is based on the fuzzy-rough fuzzy similarity metrics that could be applied on real measured datasets. Additionally we compare our metric with the previous similarity metric based on experimental evaluation with two aggregation operators. After feature selection, they are evaluated on three classification algorithms on real valued datasets. The results have shown that the proposed fuzzy similarity measure achieved in some cases better dimensionality reduction with KD operator and the ROC accuracy remain the same or some cases improve.

A further interesting should be done in direction of testing more new fuzzy similarity metrics to improve the method. Other methods for effective backwards search elimination method should be considered. Some problems related to the search techniques that often returns results in local optimum and it can precede do to non-optimal paths. The investigation of often more efficient search techniques may help in solving this problem and further improve the efficiency of this approach.

Acknowledgement: This work was partially financed by the Faculty of Computer Science and Engineering at the Ss. Cyril and Methodius University in Skopje.

References