# Editing Training Sets from Imbalanced Data Using Fuzzy-Rough Sets

Do Van Nguyen*, Keisuke Ogawa, Kazunori Matsumoto, and Masayuki Hashimoto

KDDI Labs, 2-1-15 Ohara, Fujimino, Saitama, 356-8502 Japan
va-nguyen@kddilabs.jp, ngdovan@gmail.com

**Abstract.** In this research, we study several instance selection methods based on rough set theory and propose an approach able to deal with inconsistency caused by noise and imbalanced data. Recent attention has focused on the significant results obtained in selecting instances from noisy data using fuzzy-rough sets. For imbalanced data, fuzzy-rough sets approach is also applied before and after using balancing methods in order to improve classification performance. In this study, we propose an approach that uses different criteria for minority and majority classes in fuzzy-rough instance selection. It thus eliminates the step of using balancing techniques employed in controversial approach. We also carry out some experiments, measure classification performance and make comparisons with other methods.

**Keywords:** Rough Set theory, fuzzy-rough sets, classification performance, instance selection, imbalanced data.

## 1 Introduction

Rough set model is a machine learning technique for acquiring knowledge from datasets. Rough set theory was first introduced by Pawlak [19,20] in the early 1980s. It is a mathematical approach designed to deal with the ambiguity, vagueness and uncertainty that exist in a database. In knowledge discovery, rough set theory constitutes a sound basis by offering mathematical tools to discover patterns hidden in data. It can, according to [25], be used for feature selection, data reduction, decision rule generation, and pattern extraction.

Recently, rough sets have been implemented for instance selection [3,12,27]. The main ideas of these approaches are extracting fuzzy memberships of positive regions and choosing the instances which have large membership degrees for training phases. However, it is believed that these criteria do not overcome all problems in relation to inconsistency issues.

One issue that may cause an inconsistency problem is imbalanced data [14]. A dataset is imbalanced when the numbers of instances in some classes are much larger than in others. Such classes are called majority classes. The classes with small cardinality are referred to as minority classes.

* Corresponding author.


R. Chbeir et al. (Eds.): AIAI 2015, IFIP AICT 458, pp. 115–129, 2015.
DOI: 10.1007/978-3-319-23868-5_9

In fact, some researchers are trying to use the advantages of rough sets to enhance classification performance when a balancing method is employed [22,23,28]. In these approaches, fuzzy-rough sets are first used to remove low quality instances. Then, a well-known balancing technique called "Synthetic Minority Oversampling Technique (SMOTE)" [4] is employed to form a candidate set. Finally fuzzy-rough sets are used again to select quality instances from the candidate sets. Obviously, such methods need considerable computation time to edit and select instances.

This research will introduce an approach that uses fuzzy-rough sets to balance and select quality instances from imbalanced datasets. In this work, different thresholds for majority and minority classes are used, thus, more items from minority classes can be selected. The experiment results show considerable improvement in classification performance and the advantage with respect to other methods in the literature.

The paper is structured as follows: Section 2 reviews the original rough set theory and an extension called fuzzy-rough set. In section 3, some rough set approaches for dealing with inconsistency are discussed along with their issues. The new algorithms and experiments are introduced in Section 4. Finally, we present our conclusions in Section 5.

## 2 Theoretical Background

### 2.1 Information Systems and Equivalence Relation

An information system is represented as a data table. Each row of this table represents an instance of an object such as people, things, etc. Information about every object is described by object attribute (feature) values.

An information system in the rough sets study is formally defined as a pair $I = (U, A)$, where $U$ is a non-empty finite set of objects called the universe and $A$ is a non-empty finite set of attributes such that $f_a : U \to V_a$ for every $a \in A$ [19,20]. The non-empty discrete value set $V_a$ is called the domain of $a$. The original rough set theory deals with complete information systems in which $\forall x \in U$, $a \in A$, $f_a(x)$ is a precise value.

Any information system taking the form $I = (U, A \cup \{d\})$ is called a decision table where $d \notin A$ is called a decision (or label) and elements of $A$ are called conditions. Let $V_d = \{d_1, ..., d_k\}$ denote the value set of the decision attribute, decision $d$ then determines a set of partitions $\{C_1, C_2, ..., C_k\}$ of universe $U$, where $C_i = \{x \in U | f_d(x) = d_i\}$, $1 \leq i \leq k$. Set $C_i$ is called the $i$-th decision class or concept on $U$. We assume that every object in $U$ has a certain decision value in $V_d$.

Formally, in complete information systems, relation $EQU_P(x, y)$, $P \subseteq A$, denotes a binary relation between objects that are equivalent in terms of values of attributes in $P$ [19]. The equivalence relation is reflexive, symmetric, and transitive. Let $E_P(x) = \{y \in U | EQU_P(y, x)\}$ be the set of all objects that are equivalent to $x$ by $P$, which is then called an equivalence class. The family of

all equivalence classes (or partitions) on $U$ based on an equivalence relation is referred to as a category and is denoted by $U/EQU_P$.

From equivalence classes, Pawlak [19,20] defined an approximation space that contains lower and upper approximations denoted by $\underline{appr}X$ and $\overline{appr}X$, respectively, of set $X \subseteq U$ as follows:

$$\underline{appr}_P X = \bigcup \{E_P(x) | x \in U, E_P(x) \subseteq X\} = \{x \in U | E_P(x) \subseteq X\}, \quad (1)$$

$$\overline{appr}_P X = \bigcup \{E_P(x) | x \in U, E_P(x) \cap X \neq \emptyset\} = \{x \in U | E_P(x) \cap X \neq \emptyset\}. \quad (2)$$

In decision table $I = (U, A \cup \{d\})$, the positive region $POS_P(X)$ of set $X$ in terms of attribute set $P \subseteq A$ is defined as:

$$POS_P(X) = \bigcup \{\underline{appr}_P E_{\{d\}}(x) | x \in X\}. \quad (3)$$

Others region definitions such as negative, boundary and rough set properties can be obtained from the original rough set theory research [19,20].

Apart from using equivalence relations to define a rough set in complete information systems, there are also numerous studies [10,13,16,17,18] that deal with incomplete or imperfect information systems in which data are not described by precise and crisp values.

### 2.2 Fuzzy Rough Set

A classical (crisp) set is normally defined as a collection of elements $x \in X$ that can be finite, countable or over countable. Each single element can either belong to or not belong to a set $X' \subseteq X$. For a fuzzy set, a characteristic function allows various degrees of membership for elements of a given set. Then a fuzzy set $\mathcal{X}$ in $X$ is a set of ordered pairs [31]:

$$\mathcal{X} = \{(x, \mu_{\mathcal{X}}(x)) | x \in X\}, \quad (4)$$

where $\mu_{\mathcal{X}}(x) \in [0, 1]$ is called the membership function or grade of membership (also the degree of compatibility or degree of truth) of $x$ in $\mathcal{X}$.

In crisp sets, for two special sets $\emptyset$ and $U$ the approximations are simply defined as $\mu_{\underline{appr}_R U}(x) = 1$ and $\mu_{\overline{appr}_R \emptyset}(x) = 0$. Based on the two equivalent definitions, lower and upper approximations may be interpreted as follows: An element $x$ belongs to the lower approximation $\underline{appr}_R X$ if all elements equivalent to $x$ belong to $X$. In other words, $x$ belongs to the lower approximation of $X$ if any element not in $X$ is not equivalent to $x$, namely, $\mu_R(x, y) = 0$. Likewise, $x$ belongs to the upper approximation of $X$ if $\mu_R(x, y) = 1$.

Now the notion of rough sets in crisp sets is extended to included fuzzy sets. Let $\mu_{\mathcal{X}}$ and $\mu_{\mathcal{R}}$ denote the membership functions of set $\mathcal{X}$ and of the set $\{(x, y) \in U \times U | \mathcal{R}(x, y)\}$, respectively. The fuzzy approximation space [21] of fuzzy set $\mathcal{X}$ on $X$ in terms of fuzzy relation $\mathcal{R}(x, y)$ can be defined as follows:

$$\mu_{\underline{appr}_{\mathcal{R}} \mathcal{X}}(x) = \inf \mathcal{I}\{\mu_{\mathcal{R}}(x, y), \mu_{\mathcal{X}}(y)\}, \quad (5)$$

$$\mu_{\overline{appr}_{\mathcal{R}} \mathcal{X}}(x) = \sup \mathcal{T}\{\mu_{\mathcal{R}}(x, y), \mu_{\mathcal{X}}(y)\}. \quad (6)$$

where $\mathcal{I}$ and $\mathcal{T}$ are fuzzy implicators [1] and triangular norms (t-norm) [2], respectively.

From the above equations, it is noticed that membership of lower and upper approximation for one instance strongly depend on only one other instance since they are defined by minimum and maximum. To soften this definition, Cornelis et. al. [5] suggest a fuzzy-rough set definition using order weight averaging (OWA) aggregation [29]. An OWA operator $F_W$ is a mapping $F : R^n \to R$ using weight vector $W =< w_1, ..., w_n >$ such that

$$F_W(a_1, \cdots, a_n) = \sum_{i=1}^{n} w_i b_i, \tag{7}$$

where $b_i$ is the $i^{th}$ largest of the values in $\{a_1, \cdots, a_n\}$.

An OWA operator is bounded by the minimum and maximum of vector $W$. It thus can soften minimum and maximum operators. Suppose that there are two vectors $W^{min} =< w_1^{min}, \cdots, w_n^{min} >$ and $W^{max} =< w_1^{max}, \cdots, w_n^{max} >$, where $w_1^{min} \leq \cdots \leq w_n^{min}$ and $w_1^{max} \geq \cdots \geq w_n^{max}$, we can define OWA fuzzy rough sets as follows:

$$\mu_{\underline{appr}_{\mathcal{R}}\mathcal{X}}(x) = F_{W^{min}}(\mathcal{I}\{\mu_{\mathcal{R}}(x, y), \mu_{\mathcal{X}}(y)\}), \tag{8}$$

$$\mu_{\overline{appr}_{\mathcal{R}}\mathcal{X}}(x) = F_{W^{max}}(\mathcal{T}\{\mu_{\mathcal{R}}(x, y), \mu_{\mathcal{X}}(y)\}). \tag{9}$$

In fact, the definition of a fuzzy relation on an attribute set depends on individual systems. One example of defining relation methods to deal with incomplete information systems is shown in [15]. Further investigation of the combination between fuzzy and rough sets can be found in [5,8,9,21,30].

## 3 Rough Selection and Problem Stated

Significant research on instance selection using rough sets model was introduced the work of Caballero et al. [3]. For this purpose, the authors calculated approximations and positive regions for each class in training sets. Instances for training phases are then selected in two ways. In the first method, they try to delete all instances in the boundary. The second employs a nearest neighbourhood algorithm to relabel instances in the boundary region. The issues associated with these approaches were raised in Jensen and Cornelis' study [12]. Another possible limitation is that they just deal with crisp sets of attributes in decision tables.

[1] A fuzzy implicator is a function $\mathcal{I} : [0, 1] \times [0, 1] \to [0, 1]$ which satisfies the following properties: $\mathcal{I}(0, 0) = 1$; $\mathcal{T}(1, a) = a$; $\mathcal{I}$ is decreasing in the first argument; $\mathcal{I}$ is increasing in the second argument.

[2] A t-norm is a function $\mathcal{T} : [0, 1] \times [0, 1] \to [0, 1]$ which satisfies the following properties: Commutativity: $T(a, b) = T(b, a)$; Monotonicity: $T(a, b) \leq T(c, d)$ if $a \leq c$ and $b \leq d$; Associativity: $T(a, T(b, c)) = T(T(a, b), c)$; The number 1 acts as an identity element: $T(a, 1) = a$

To deal with the problems identified in the study of Caballero et al., an approach called Fuzzy-Rough Instance Selection (FRIS) has been proposed [12]. The notion of this approach is using memberships of positive regions to determine which instances should be kept and which instances should be discarded for training. First, the method calculates relations among instances and then measures quality of each instance by its positive region membership. An instance will be removed if the membership degree is less than a specified threshold. For the same purpose, another approach of measuring quality of instances was introduced in [27]. Such approaches will undoubtedly refine the positive regions with quality instances.

However, one limitation to be considered is that there is only one threshold used for all classes. In some applications such as medical disease prediction, it is not uncommon to be interested in some specific groups, disease groups for example, rather than the others (healthy groups). When only one threshold is used, once noise is eliminated, valuable instances in interesting classes also disappear.

Some approaches use the advantage of fuzzy-rough to qualify instance sets formed by SMOTE [22,23,28]. Therefore, the issues above may be avoided. The steps of these methodologies can be summarized as follows:

1 Using fuzzy-rough sets to calculate the qualities of every instance and choose instances having high quality.
2 Using SMOTE to create artificial instances from the first step and add to datasets
3 Using fuzzy-rough sets to eliminate low quality instances from datasets after the second step.

There are some differences among these methods. In Ramentol et al.'s study named SMOTE-RSB$_*$ [22,23], the first step does not appear. In step 3, the algorithm removes only low quality instances in the artificial instance sets. In Verbiest's research [28], quality measurements are different for step 1 and 3. The first measures performance for imbalanced data while the last deals with balanced information. Thus, they call the algorithm FRIPS-SMOTE-FRBPS (hereinafter referred to as FRPSS in this paper). It is also claimed in [28] that FRPSS produces better training sets compared with SMOTE-RSB$_*$.

It appears that FRPSS is an excellent tool for editing training sets. However, applying three instance selections steps may require considerable computation time compared with other methods.

## 4 Multiple Thresholds Fuzzy Rough Instance Selection

As stated in [12], the Fuzzy-Rough Instance Selection (FRIS) algorithm is quite efficient as it eliminates some instances that might affect the positive region membership of the remaining objects. It thus refines the dataset by choosing quality instances for training. However, we also noted some limitations of conventional methods. An example is shown in Figure 1. If we use fuzzy lower approximation as instance qualities and compare these numbers with the threshold

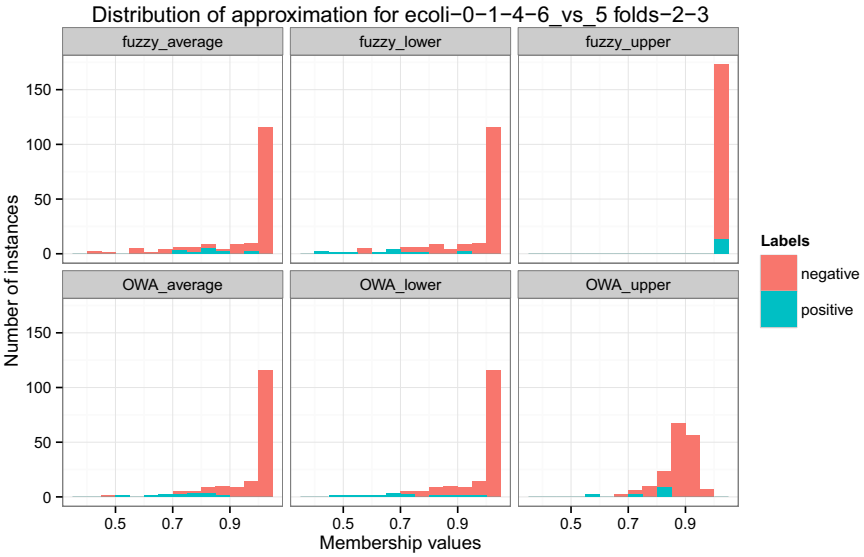


**Fig. 1.** Histogram to show distribution of approximations and average approximations for a dataset

= 0.7 (FRIS method), numerous *positive* instances (minority class) will be removed while almost all *negative* instances will be remained. The same problem occurs with FRPSS in which an average of approximations is used as a quality measurement.

Therefore, in this section, using the advantages of fuzzy rough set instance selection approach, we propose an instance selection method which uses different thresholds for different classes of decision tables. By means of this strategy, SMOTE or other balancing techniques may not be necessary.

### 4.1 MFRIS Algorithms

In decision table $I = (U, A \cup \{d\})$, let $\mathcal{R}_a(x, y)$ and $\mu_{\mathcal{R}_a}(x, y)$ denote a fuzzy relation and membership function, respectively, between objects $x, y \in U$ on attribute $a \in A$. Then the membership function of the relation on an attribute set $P \subseteq A$ is defined as:

$$\mu_{\mathcal{R}_P}(x, y) = \mathcal{T}_{a \in P} \mu_{\mathcal{R}_a}(x, y),$$

where $\mathcal{T}$ is a t-norm.

In fact, there are many ways to calculate the membership function of a fuzzy relation between two instances. It is also possible to use distance based similarity such as the normalization of Euclidean distance. It depends on the characteristics of each system.

The membership functions of approximation spaces for an object set $X \subseteq U$ on attribute $P \subseteq A$ can be defined by either equations (5) and (6) or (8) and (9).

**Data**:
$X_i, Y_j$, minority and majority classes;
$t_X$, $t_Y$, the selection thresholds for minority and majority classes.
**Result**: Decision table $(S, A \cup \{d\})$
Calculate membership degree of the relation among objects: $\mu_{\mathcal{R}}(x, y)$;
Calculate membership degree of the approximations for each class;
Calculate the quality measurement of all instances for their classes;
$S \leftarrow \emptyset$;
**for** $x \in X_i$ **do**
  **if** $\mu^{\mathcal{R}}_{\mathcal{X}_i}(x) \geq t_X$ **then**
    $S \leftarrow S \cup \{x\}$;
  **end**
**end**
**for** $x \in Y_j$ **do**
  **if** $\mu^{\mathcal{R}}_{\mathcal{Y}_i}(x) \geq t_Y$ **then**
    $S \leftarrow S \cup \{x\}$;
  **end**
**end**

**Fig. 2. MFRIS1** - The first algorithm: Choosing or eliminating instances for both majority and minority classes

In this study, as mentioned in Section 2, we must note that in the decision table $I = (U, A \cup \{d\})$ is a single label decision table such that $f_d(x)$ is a precise value. In fuzzy sets, for a single label decision table, $\mu_{C_i}(x) > 0$ if $f_d(x) = d_i$ and $\mu_{C_i}(x) = 0$, otherwise.

In selecting learning instances, we first define a function to measure the quality of an instance for each class $X$ in terms of relation $\mathcal{R}$:

$$\mu^{\mathcal{R}}_{\mathcal{X}}(x) = \alpha \mu_{\underline{appr}_{\mathcal{R}}\mathcal{X}}(x) + (1-\alpha)\mu_{\overline{appr}_{\mathcal{R}}\mathcal{X}}(x), \tag{10}$$

where $\alpha \in [0, 1]$, $\mathcal{X}$ is the fuzzy set on $X$. In the above equation, note that $x$ may or may not belong to $X$. This is where this approach differs from other approaches. When comparing $\mu^{\mathcal{R}}_{\mathcal{X}}(x)$ for $x \in X$, we can assume that $\alpha = 1$ as in the study of Jensen et al. [12] and $\alpha = 0.5$ as in the approach of Verbiest et al. [28].

Now, let $t_X$ and $t_Y$ be the selection thresholds for minority and majority classes, $\mathcal{X}_i$ and $\mathcal{Y}_j$ be the fuzzy sets on minority class $X_i$ and majority classes $Y_j$, respectively, $\mathcal{R}$ denotes the fuzzy relation on the universal. The set of instances selected for the training phase can be defined as:

$$S = \left[\cup_i \{x \in X_i | \mu^{\mathcal{R}}_{\mathcal{X}_i}(x) \geq t_X\}\right] \cup \left[\cup_j \{x \in Y_j | \mu^{\mathcal{R}}_{\mathcal{Y}_i}(x) \geq t_Y\}\right]. \tag{11}$$

Then, the algorithm to select instances can be described as shown in Figure 2.

In the first algorithm, depending on labels of instances, the quality measurement of every instance on their classes will be compared with the threshold for minority or majority thresholds. This means that an instance will be deleted

from a training set if it is low quality even it is in minority classes. The different thresholds may help us to keep more instance in minority classes if necessary.

In addition, we also propose the second algorithm to select instances. The notion of this method is keeping all instances in minority classes while removing or relabelling some instances in majority classes. To describe the algorithm, we first introduce some definitions.

Let $X, Y$ be the family of minority and majority classes respectively. Let $\mathcal{X}_i, \mathcal{Y}_j$ denote fuzzy sets on $X_i, Y_j$, respectively. Thresholds for minority and majority classes are $t_X$ and $t_Y$, respectively. The set of instances for which labels could be changed can be defined as

$$S_{Y \to X} = \bigcup_i \{x \in Y_i | \mu^{\mathcal{R}}_{\mathcal{Y}_i}(x) < t_Y \ \wedge \sup_j \{\mu^{\mathcal{R}}_{\mathcal{X}_j}(x) - t_X\} \geq 0\}. \tag{12}$$

From the above definition, there are some instances in majority classes whose labels could be changed to a minority class label. The possible minority class set can be determined as follows:

$$Z(x) = \bigcup_i \{X_i | \mu^{\mathcal{R}}_{\mathcal{X}_i}(x) \geq t_X, x \in S_{Y \to X}\}. \tag{13}$$

Then we can re-calculate the class membership functions of an instance $x \in S_{Y \to X} \cap Y_n$ as $\mu_{\mathcal{X}_m}(x) = \mu_{\mathcal{Y}_n}(x); \mu_{\mathcal{Y}_n}(x) = 0$, where $m$ is defined to satisfy: $X_m \in Z(x); \mu^{\mathcal{R}}_{\mathcal{X}_m}(x) \geq \mu^{\mathcal{R}}_{\mathcal{X}_i}(x)$, $\forall i, X_i \in Z(x)$.

Finally, the selected instances for the training phase can be defined as:

$$S = (\bigcup_i X_i) \cup (\bigcup_i \{x \in Y_i | \mu^{\mathcal{R}}_{Y_i}(x) \geq t_Y\}) \cup S_{Y \to X}. \tag{14}$$

The algorithm is then described as shown in Figure 3.

### 4.2 Experiment

Experiments were conducted on datasets from KEEL dataset repository [1] and UCI Machine Learning Repository [2] and our lifestyle-diseases dataset. The dataset lifestyle-diseases is raw data of collected health factors such as age, BMI, and waist circumference,... and the outcomes are related to lifestyle diseases. Datasets and their properties are shown in Table 1. In this table, "IR" shows imbalance rates for the number of instances between majority and minority classes.

Each dataset is divided into three parts for both minority and majority classes. For each train-test, we used two parts to form training data and the rest were used as testing instances.

Experiments were conducted using R programming[3] running on RStudio[4].

Training data are first cleaned by instance selection algorithms. For rough set based instance selection methods, we measure fuzzy tolerance relations on

[3] http://www.r-project.org/

[4] http://www.rstudio.com/

**Data**:
$X, Y$, the family set of minority and majority classes
$t_X, t_Y$, the selection thresholds for minority and majority classes
**Result**: Decision table $(S, A \cup \{d\})$
Calculate membership degree of the relation among objects: $\mu_{\mathcal{R}}(x, y)$;
Calculate membership degree of the approximations for each class;
Calculate quality measurement of all instances in the majority classes for all classes;
$S_Y \leftarrow \emptyset$;
$S_{Y \to X} \leftarrow \emptyset$;
**for** $x \in Y_i$ **do**
  **if** $\mu^{\mathcal{R}}_{Y_i}(x) \geq t_Y$) **then**
    $S_Y \leftarrow S \cup \{x\}$;
  **else**
    $Z(x) \leftarrow \bigcup_j \{X_j | \mu^{\mathcal{R}}_{X_j}(x) \geq t_X\}$;
    **if** $Z(x) \neq \emptyset$ **then**
      Find $X_m \in X$ satisfy $\mu^{\mathcal{R}}_{X_m}(x) = \max_j \{\mu^{\mathcal{R}}_{X_j}(x) | X_j \in Z(x)\}$;
      **if** *a pair occurs* **then** choose the first $X_m$;
      $\mu_{X_m}(x) \leftarrow \mu_{Y_i}(x)$;
      $\mu_{Y_i}(x) \leftarrow 0$;
      $S_{Y \to X} \leftarrow S_{Y \to X} \cup \{x\}$;
    **end**
  **end**
**end**
$S \leftarrow \bigcup_i \{X_i \in X\} \cup S_Y \cup S_{Y \to X}$;

**Fig. 3. MFRIS2** - The second algorithm: Choosing, removing or relabelling instances for majority classes

an attribute by membership function $\mu_{\mathcal{R}_a}(x, y) = 1 - \frac{|f_a(x) - f_a(y)|}{l(a)}$, where $l(a)$ is the range of value domain on $a$. In this study, we do not deal with missing values. However, it is possible to define $\mu_{\mathcal{R}_a}(x, y)$ for incomplete information by following the study in [16].

The fuzzy tolerance relation on a set of attributes can be calculated by a combination of the relation on each attribute by Lukasiewicz's t-norm $\mathcal{T} = max(x_2 + x_1 - 1, 0)$. For those which use OWA fuzzy rough sets, the OWA operators are set as follows:

$W^{min} =< w_i^{min} >= w_{n+1-i}^{min} = \frac{2^{3-i}}{2^3-1}$ for $i = 1, 2, 3$ and 0 for $i = 4, ..., n$,

$W^{max} =< w_i^{max} >= w_i^{max} = \frac{2^{3-i}}{2^3-1}$ for $i = 1, 2, 3$ and 0 for $i = 4, ..., n$,

where $n$ is number of instances in datasets.

For our approach, we also use OWA fuzzy rough set with above operators and set $\alpha = 0.5$ to calculate $\mu^{\mathcal{R}}_{\mathcal{X}}$. That is the same as instances quality measurement in FRIPS-SMOTE-FRBPS in terms of distinguishing instances quality.

**Table 1.** Datasets for experiments

| Datasets | Inst... | Feat... | IR | Datasets | Inst... | Feat... | IR |
|---|---|---|---|---|---|---|---|
| pima | 768 | 8 | 1.87 | shuttle-c0-vs-c4 | 1829 | 9 | 13.87 |
| haberman | 305 | 3 | 2.77 | page-blocks-1-3_vs_4 | 472 | 10 | 15.86 |
| liver_man | 439 | 9 | 2.78 | yeast4 | 1484 | 8 | 28.10 |
| vehicle1 | 846 | 18 | 2.90 | winequality-red-4 | 1599 | 11 | 29.17 |
| blood_transfusion | 748 | 4 | 3.20 | ozone_one_hr | 1848 | 72 | 31.42 |
| life-style-diseases | 4819 | 13 | 4.04 | yeast5 | 1484 | 8 | 32.73 |
| segment0 | 2308 | 19 | 6.02 | ecoli-0-1-3-7_vs_2-6 | 280 | 7 | 39.00 |
| page-blocks0 | 5472 | 10 | 8.79 | yeast6 | 1484 | 8 | 41.40 |
| yeast-2_vs_4 | 514 | 8 | 9.08 | mammography | 11183 | 6 | 42.01 |
| vowel0 | 988 | 13 | 9.98 | winequality-red-8_vs_6-7 | 855 | 11 | 46.50 |
| glass2 | 214 | 9 | 11.59 | winequality-white-3-9_vs_5 | 1482 | 11 | 58.28 |
| ozone_eight_hr | 1847 | 72 | 13.43 | winequality-red-3_vs_5 | 691 | 11 | 68.10 |

**Table 2.** Confusion matrix for a two-class problem

| | Positive prediction | Negative prediction |
|---|---|---|
| Positive class | True Positive (TP) | False Negative (FN) |
| Negative class | False Positive (FP) | True Negative (TN) |

Next, selected data are used for training using several classification techniques including k-Nearest Neighbours [7], Support Vector Machine [6], naive Bayes [24] and Decision Trees [26]. It should be noticed that these data classification toolboxes are available within R data mining packages.

To evaluate classification performance in different cases where classifier techniques are used as well as instance selection methods, we use the confusion matrix (Table 2) to calculate the Area Under the ROC Curver($AUC$) [11]. $AUC$ provides a single measure for comparing classification on average.

$$AUC = \frac{1}{2}\left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP}\right). \tag{15}$$

For comparison, we try to obtain the best performance for each instance selection algorithm. FRPSS suggests its own method of threshold optimization. For FRIS and MFRISs, instance measurements are quantized into a set of a limited number of elements. We then use this set as threshold candidates in order to find out if there is a significant increase in performance.

In general, from the average AUC shown in Table 3, we can see that after using the modification of rough set instances selection (both MFRIS1 and MRIS2), the classification performance increases noticeably in almost datasets. It seems that the use of multiple thresholds is more effective with the increase of IR value. Although the original fuzzy rough instance selection approach does not deal with imbalanced data, the increase in performance of MFRISs proves that it is possible to use fuzzy rough sets based instance selection with some criteria to overcome the imbalance issue.

**Table 3.** Average AUC archived by the different classifiers for each instance selection method (bold numbers show the highest)

| Datasets | NONE | FRIS | FRPSS | MFRIS1 | MFRIS2 |
|---|---|---|---|---|---|
| pima | 0.7031 | 0.7099 | 0.6967 | **0.7462** | 0.7303 |
| haberman | 0.5559 | 0.5795 | 0.6292 | **0.6565** | 0.6195 |
| liver_man | 0.5673 | 0.5903 | 0.6665 | **0.7078** | 0.6991 |
| vehicle1 | 0.6478 | 0.6545 | 0.6931 | 0.7357 | **0.7387** |
| blood_transfusion | 0.5802 | 0.6084 | 0.6290 | **0.6845** | 0.6765 |
| life-style-diseases | 0.5537 | 0.5794 | 0.5729 | **0.6055** | 0.5629 |
| segment0 | 0.8645 | 0.8769 | 0.8638 | **0.9466** | 0.9058 |
| page-blocks0 | 0.8177 | 0.8433 | 0.8650 | **0.8963** | 0.8870 |
| yeast-2_vs_4 | 0.7892 | 0.7958 | 0.7678 | **0.9102** | 0.8856 |
| vowel0 | 0.8996 | 0.8443 | 0.9482 | **0.9556** | 0.9498 |
| glass2 | 0.5305 | 0.5489 | 0.6082 | **0.7291** | 0.6918 |
| ozone_eight_hr | 0.5738 | 0.5738 | **0.7325** | 0.5738 | 0.5738 |
| shuttle-c0-vs-c4 | 0.9947 | 0.6166 | 0.9947 | **0.9987** | 0.9983 |
| page-blocks-1-3_vs_4 | 0.7996 | 0.8048 | 0.8059 | **0.9187** | 0.9046 |
| yeast4 | 0.5710 | 0.6020 | 0.8069 | **0.8416** | 0.7924 |
| winequality-red-4 | 0.5090 | 0.5094 | 0.6232 | 0.6621 | **0.6785** |
| ozone_one_hr | 0.5659 | 0.5659 | **0.7485** | 0.5659 | 0.5659 |
| yeast5 | 0.8081 | 0.8081 | 0.8736 | **0.9759** | 0.9421 |
| ecoli-0-1-3-7_vs_2-6 | 0.6487 | 0.6519 | 0.5625 | 0.7766 | **0.8593** |
| yeast6 | 0.6634 | 0.6879 | 0.8019 | **0.8996** | 0.8587 |
| mammography | 0.7559 | 0.7735 | 0.7790 | **0.8771** | 0.7587 |
| winequality-red-8_vs_6-7 | 0.5085 | 0.5102 | 0.5896 | 0.6661 | **0.7278** |
| winequality-white-3-9_vs_5 | 0.5310 | 0.5318 | 0.5715 | 0.7111 | **0.7464** |
| winequality-red-3_vs_5 | 0.5252 | 0.5128 | 0.5385 | 0.6481 | **0.8376** |

Although less consuming time may be an advantage, besides high performance archived, we do not show the time measurements in this version. This is due to the different criteria in choosing thresholds. FRISS, as stated before, has its own method to optimize thresholds. It considers the unique set of all instance measurements as the threshold candidates. This means that the algorithm has to train and test numerous times (nearly as many as the cardinality of training sets). In FRIS and MFRISs, on the other hand, we tried in limited times.

To a comparison of the two algorithms using our approach, MFRIS1 archives better performance for most datasets even though MFRIS2 still improves performance. In the second algorithm, all instances in minority classes are chosen. If such instances do not have sufficient quality, an edited training set from MFRIS2 cannot be better than that from MFRIS1.

When comparing the classifiers, in general, we can see the differences among them in terms of performance evaluation when there is no instance selection method employed (Figure 4, transparency color bars). SVM seems to be sensitive to imbalance issues, while Bayes provides better performance. This means that for a different type of dataset with respect to imbalance issues, a different choice of classifier may be required.

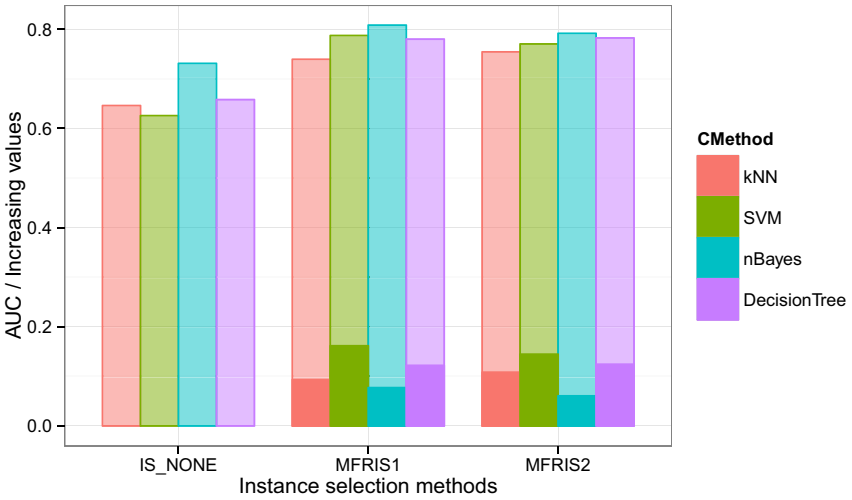


**Fig. 4.** Comparing average of AUC (on all datasets) and increase on average of AUC using MFRISs for different classifiers

**Table 4.** Comparing the thresholds for MFRIS1 that give the highest improvement of each classifier on mammography dataset (the first number is for minority)

| Training sets | kNN | SVM | nBayes | DecisionTree |
|---|---|---|---|---|
| fold2_3 | 0.5090 0.5966 | 0.5300 0.6483 | 0.5000 0.5966 | 0.5090 0.5966 |
| fold1_3 | 0.5020 0.5841 | 0.5220 0.6322 | 0.4540 0.5841 | 0.5090 0.6322 |
| fold1_2 | 0.5100 0.5911 | 0.5280 0.6517 | 0.4040 0.5911 | 0.5210 0.6517 |

**Table 5.** Comparing the thresholds for MFRIS2 that the highest improvement of each classifiers on winequality-red-8_vs_6-7 dataset (the first number is for minority)

| Training sets | kNN | SVM | nBayes | DecisionTree |
|---|---|---|---|---|
| fold2_3 | 0.2560 0.9075 | 0.2560 0.9602 | 0.1250 0.9360 | 0.2140 0.9602 |
| fold1_3 | 0.2130 0.9085 | 0.2130 0.8758 | 0.1600 0.9085 | 0.1840 0.8758 |
| fold1_2 | 0.0000 0.7619 | 0.1690 0.9575 | 0.0000 0.9082 | 0.1440 0.9082 |

Figure 4 also illustrates the difference in performance increase (dark color bars). SVM does not show a good result for imbalanced data, nevertheless, its performance improves to a greater extent than the rest. In other words, naive Bayes receives less support from instance selection methods. However, the performance of all classifiers generally improves.

Obviously, optimized thresholds are various among classifiers and among datasets. Table 4 shows that SVM needs to remove more instances by using smaller thresholds in comparing with other classifiers. In addition, Table 5

suggests that for some datasets, it is necessary to edit majority classes by discarding many instances (high majority thresholds) while moving some instances to minority classes.

## 5 Conclusion and Future Work

In this paper, we presented a modification of fuzzy-rough instance selection using multiple thresholds. Retaining the advantages of fuzzy-rough sets when selecting instances, it can measure the quality of instances for training phase in order to select valuable instances. It also deals with imbalanced data by using different thresholds between minority and majority classes.

Some experiments show that these approaches can improve performance considerably. They can also be competitive with state-of-the-art methods. Furthermore, this notion can potentially applied to modify current rough set based learning methods in order to eliminate the instance selection step.

This paper mainly proves the robustness of using multiple thresholds in fuzzy-rough sets based instance selection approach. No algorithm for selecting valuable thresholds is suggested. Therefore, in further research we intend to study a method of optimizing thresholds and will endeavour to discover a better instance quality measurement function.

## References

1. Alcala-Fdez, J., Fernandez, A., Luengo, J., Derrac, J., Garcia, S.: Keel data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework. Multiple-Valued Logic and Soft Computing 17(2–3), 255–287 (2011)
2. Bache, K., Lichman, M.U.: Machine Learning Repository (2013)
3. Caballero, Y., Bello, R., Alvarez, D., Gareia, M.M., Pizano, Y.: Improving the k-nn method: Rough set in edit training set. In: Debenham, J. (ed.) Professional Practice in Artificial Intelligence. IFIP, vol. 218, pp. 21–30. Springer, Heidelberg (2006)
4. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: Smote: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research 16, 321–357 (2002)
5. Cornelis, C., Verbiest, N., Jensen, R.: Ordered weighted average based fuzzy rough sets. In: Yu, J., Greco, S., Lingras, P., Wang, G., Skowron, A. (eds.) RSKT 2010. LNCS, vol. 6401, pp. 78–85. Springer, Heidelberg (2010)
6. Cortes, C., Vapnik, V.: Support-vector networks. Machine Learning 20(3), 273–297 (1995)
7. Cover, T., Hart, P.: Nearest neighbor pattern classification. IEEE Trans. Inf. Theor. 13(1), 21–27 (1967)
8. Dubois, D., Prade, H.: Rough fuzzy sets and fuzzy rough sets. International Journal of General Systems 17, 191–209 (1990)
9. Dubois, D., Prade, H.: Putting rough sets and fuzzy sets together. In: Slowinski, R. (ed.) Intelligent Decision Support. Theory and Decision Library, vol. 11, pp. 203–232. Springer, Netherlands (1992)

10. Grzymala-Busse, J.W., Clark, P.G., Kuehnhausen, M.: Generalized probabilistic approximations of incomplete data. International Journal of Approximate Reasoning 55(1), Part 2, 180–196 (2014). Special issue on Decision-Theoretic Rough Sets
11. Huang, J., Ling, C.: Using auc and accuracy in evaluating learning algorithms. IEEE Transactions on Knowledge and Data Engineering 17(3), 299–310 (2005)
12. Jensen, R., Cornelis, C.: Fuzzy-rough instance selection. In: 2010 IEEE International Conference on Fuzzy Systems (FUZZ), pp. 1–7, July 2010. doi:10.1109/FUZZY.2010.5584791
13. Kryszkiewicz, M.: Rough set approach to incomplete information systems. Inf. Sci. 112(1–4), 39–49 (1998)
14. Lopez, V., Fernandez, A., Garcia, S., Palade, V., Herrera, F.: An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. Information Sciences 250, 113–141 (2013)
15. Nguyen, D.V., Yamada, K., Unehara, M.: Extended tolerance relation to define a new rough set model in incomplete information systems. Advances in Fuzzy Systems, Article ID 372091 (2013)
16. Nguyen, D.V., Yamada, K., Unehara, M.: On probability of matching in probabilistiy based rough set definitions. In: IEEE-SMC2013, Manchester, The UK, pp. 449–454 (2013)
17. Nguyen, D.V., Yamada, K., Unehara, M.: Rough set approach with imperfect data based on dempster-shafer theory. Journal of Advanced Computational Intelligence and Intelligent Informatics 18(3), 280–288 (2014)
18. Nguyen, H.S.: Discretization problem for rough sets methods. In: Polkowski, L., Skowron, A. (eds.) RSCTC 1998. LNCS (LNAI), vol. 1424, pp. 545–552. Springer, Heidelberg (1998)
19. Pawlak, Z.: Rough sets. International Journal of Computer and Information Sciences 11, 341–356 (1982)
20. Pawlak, Z.: Rough Sets. Theoretical Aspects of Reasoning about Data. Kluwer Acad. (1991)
21. Radzikowska, A.M., Kerre, E.E.: A comparative study of fuzzy rough sets. Fuzzy Sets Syst. 126(2), 137–155 (2002)
22. Ramentol, E., Caballero, Y., Bello, R., Herrera, F.: SMOTE-RSB *: a hybrid preprocessing approach based on oversampling and undersampling for high imbalanced data-sets using SMOTE and rough sets theory. Knowl. Inf. Syst. 33(2), 245–265 (2011)
23. Ramentol, E., Verbiest, N., Bello, R., Caballero, Y., Cornelis, C., Herrera, F.: Smote-frst: a new resampling method using fuzzy rough set theory. In: Kahraman, C., Kerre, E., Bozbura, F.T. (eds.) World Scientific Proceedings Series on Computer Engineering and Decision Making, vol. 7, pp. 800–805. World Scientific (2012)
24. Rish, I.: An empirical study of the naive bayes classifier. Tech. rep. (2001)
25. Skowron, A., Rauszer, C.: The discernibility matrices and functions in information systems. In: Slowinski, R. (ed.) Intelligent Decision Support. Theory and Decision Library, vol. 11, pp. 331–362. Springer, Netherlands (1992)
26. Strobl, C., Malley, J., Tutz, G. An introduction to recursive partitioning: Rationale, application and characteristics of classification and regression trees, bagging and random forests (2009)
27. Verbiest, N., Cornelis, C., Herrera, F.: Frps: A fuzzy rough prototype selection method. Pattern Recognition 46(10), 2770–2782 (2013)

28. Verbiest, N., Ramentol, E., Cornelis, C., Herrera, F.: Preprocessing noisy imbalanced datasets using SMOTE enhanced with fuzzy rough prototype selection. Appl. Soft Comput. 22, 511–517 (2014)
29. Yager, R.R.: On ordered weighted averaging aggregation operators in multicriteria decisionmaking. IEEE Trans. Syst. Man Cybern. 18(1), 183–190 (1988)
30. Yao, Y.Y.: Combination of rough and fuzzy sets based on-level sets. In: Rough Sets and Data Mining: Analysis for Imprecise Data, pp. 301–321. Kluwer Academic (1997)
31. Zimmermann, H.-J.: Fuzzy Set Theory and its Applications. Springer (2001)