Fundamenta Informaticae XX (2015) 1–34
DOI 10.3233/FI-2012-0000
IOS Press

Applications of Fuzzy Rough Set Theory in Machine Learning: a Survey

Sarah Vluymans [C]
Department of Applied Mathematics, Computer Science and Statistics
Ghent University, Belgium

Inflammation Research Center
Flemish Institute for Biotechnology, Belgium

Lynn D'eer
Department of Applied Mathematics, Computer Science and Statistics
Ghent University, Belgium

Yvan Saeys
Inflammation Research Center
Flemish Institute for Biotechnology, Belgium

Department of Respiratory Medicine
Ghent University, Belgium

Chris Cornelis
Department of Computer Science and Artificial Intelligence
University of Granada, Spain

Department of Applied Mathematics, Computer Science and Statistics
Ghent University, Belgium

Abstract. Data used in machine learning applications is prone to contain both vague and incomplete information. Many authors have proposed to use fuzzy rough set theory in the development of new techniques tackling these characteristics. Fuzzy sets deal with vague data, while rough sets allow to model incomplete information. As such, the hybrid setting of the two paradigms is an ideal candidate tool to confront the separate challenges. In this paper, we present a thorough review on the use of fuzzy rough sets in machine learning applications. We recall their integration in preprocessing methods and consider learning algorithms in the supervised, unsupervised and semi-supervised domains and outline future challenges. Throughout the paper, we highlight the interaction between theoretical advances on fuzzy rough sets and practical machine learning tools that take advantage of them.

Keywords: fuzzy sets, rough sets, fuzzy rough sets, machine learning

Address for correspondence: Sarah.Vluymans@ugent.be
[C]Corresponding author

1. Introduction

Machine learning applications deal with extracting knowledge from data. Datasets obtained from real-world applications are inherently prone to contain both vague and incomplete information. An example of vagueness can be found in the evaluation of subjective concepts such as *beautiful*, *warm*, *intelligent*, *similar* and so on. Incomplete information is encountered when the available feature set describing data samples does not suffice to discern between them, that is, the dataset contains elements having the same values for all features but different values for the associated outcome. A traditional example refers to patients exhibiting the same symptoms but suffering from different diseases. This implies that no definite diagnosis can be formed based on the features (symptoms) at hand.

These two distinct sources of uncertainty are incorporated in fuzzy set theory [156] and rough set theory [114], respectively. The former models vagueness. Rather than crisp yes-or-no membership of elements to a certain set, a membership degree, drawn from the unit interval, expresses the degree to which elements belong to a concept. Rough sets on the other hand deal with incomplete information. They are based on the premise that concepts cannot be modeled exactly. Instead, concepts are approximated by two sets, a lower and upper approximation, which together form the rough set associated to the concept.

In the quest to be able to handle uncertain information, the hybridization of these two models into fuzzy rough set theory was first proposed in [43] and has been widely used and extended ever since. It involves the approximation of a vague and incomplete concept by means of two fuzzy sets, the fuzzy rough lower and upper approximation. It further allows elements to be discernible from each other to a certain extent, rather than being either discernible or not. The use of fuzzy rough set theory is encountered in a wide variety of machine learning domains. The most popular focus throughout the literature has been on attribute selection algorithms, which reduce the number of attributes describing the elements in a dataset to obtain a speed-up and possible performance gain of posterior learning algorithms. Fuzzy rough sets are used to model the strength of individual attributes and guide the search for an optimal attribute subset. In this paper, we review the history and development of these methods, considering different feature quality measures and search approaches. Apart from attribute selection, we further review the orthogonal scheme of instance selection, preprocessing datasets by removing instances instead of features. In this case, the proposed algorithms use notions from fuzzy rough set theory to assess the usefulness of each element. Both attribute and instance selection are preprocessing methods. In our review, we also consider learning algorithms themselves, constructing models based on the data at hand. We assess the supervised, unsupervised and semi-supervised domains. Our discussion on the first involves both classification and regression methods, as well as neural networks. In the unsupervised domain, we consider clustering algorithms and self organizing maps. Finally, we review an application of fuzzy rough sets in a semi-supervised self-training scheme.

Throughout this paper, we show the important interaction between theoretical studies on fuzzy rough set models and practical machine learning tools. We consider the literature in a mostly chronological order. By doing so, we offer the reader not merely an overview of proposed methods, but also an evolution of these approaches over time. The remainder of this paper is organized as follows. In Section 2, we recall rough and fuzzy sets. Their hybridizations are reviewed in Section 3. In Section 4, we consider integrations within supervised learning algorithms, including the preprocessing techniques of feature se-

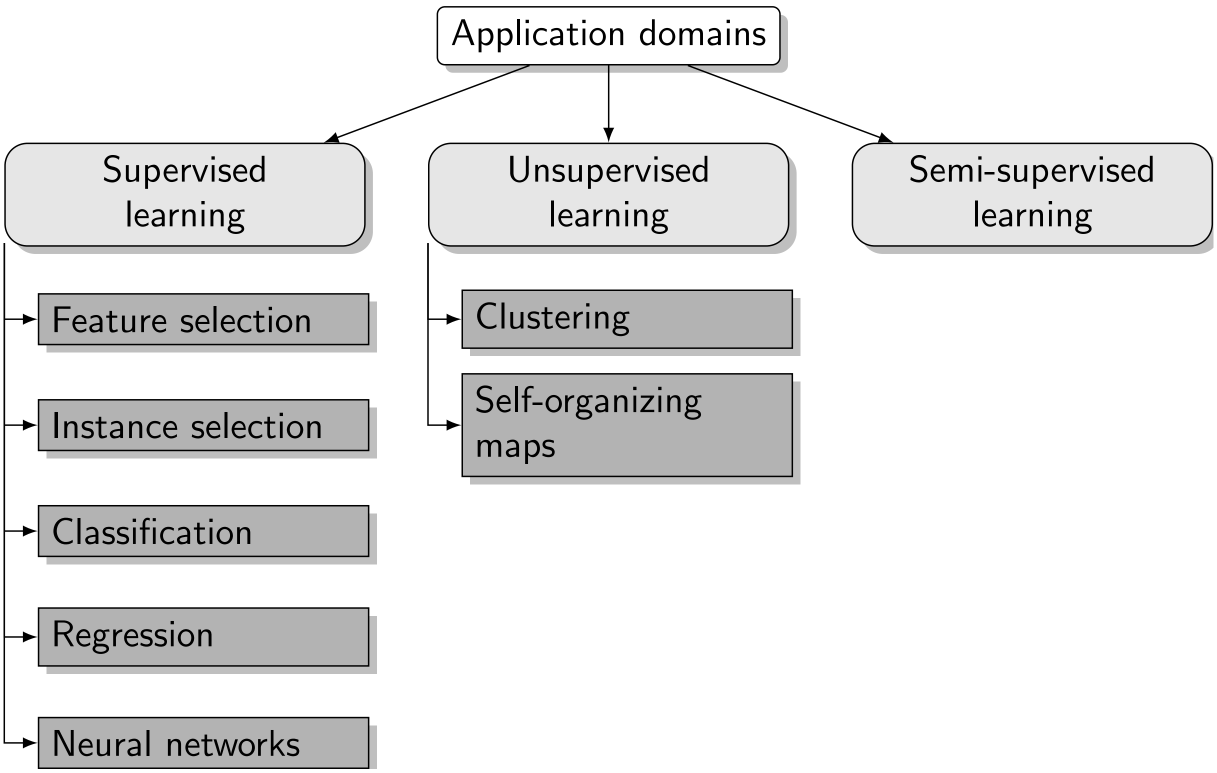

Figure 1. Overview of the application domains of fuzzy rough set theory reviewed in this paper.

lection and instance selection. We continue in Section 5 with applications in unsupervised learning. The semi-supervised domain is addressed in Section 6. An overview of these different subjects of discussion can be found in Figure 1. We note that the related domains of granular computing [5] and formal concept analysis [55] are outside the scope of this review, as we specifically focus on the integration of fuzzy rough set theory in machine learning techniques. Finally, Section 7 outlines important future challenges and we formulate our conclusion in Section 8.

2. Preliminaries

In this section, we review the most important concepts of rough and fuzzy set theory. In Section 2.1, we work with the decision system $(U, \mathcal{A} \cup \{d\})$, where U is the finite universe of discourse, $\mathcal{A}$ the set of conditional attributes and d the decision attribute, with $d \notin \mathcal{A}$. In machine learning, U is the set of data samples, $\mathcal{A}$ corresponds to the set of input variables and d is the associated outcome.

2.1. Rough Sets

Rough sets were introduced by Pawlak [114] in order to handle uncertainty due to ambiguity in information systems. Originally, an equivalence relation is used to describe the discernibility between different instances in a universe U. However, many studies replaced the equivalence relation by a general binary relation R. A survey can be found in [115, 116].

Pawlak constructed two approximations of a subset $A \subseteq U$: the lower approximation contains those elements which are certainly in the set A, while the upper approximation is assembled by the elements possibly belonging to A. Given an equivalence relation R, the lower and upper approximation of A are defined as follows:

$$\underline{\operatorname{apr}}_R(A) = \{x \in U \mid [x]_R \subseteq A\} = \bigcup\{[x]_R \in U/R \mid [x]_R \subseteq A\}, \tag{1}$$

$$\overline{\operatorname{apr}}_R(A) = \{x \in U \mid [x]_R \cap A \neq \emptyset\} = \bigcup\{[x]_R \in U/R \mid [x]_R \cap A \neq \emptyset\}, \tag{2}$$

where $[x]_R$ denotes the equivalence class of an element x and U/R denotes the quotient set of equivalence classes. Note that the first part of Equations (1) and (2) is sometimes called the element based definition and the second part is called the granule based definition [154].

Given the lower and upper approximation, the boundary region of A can be constructed. It consists of the elements of U which are not certainly inside or outside A. Formally, it is described by the difference of the upper and lower approximation: $\operatorname{bnd}_R(A) = \overline{\operatorname{apr}}_R(A) \setminus \underline{\operatorname{apr}}_R(A)$.

A prominent and wide-spread application of rough set theory can be found in data analysis [97, 98], more specifically, in feature selection and classification [136]. Many methods regarding selection and classification are based on the *positive region* of a subset of features. Given a subset $\mathcal{B} \subseteq \mathcal{A}$ of attributes, one can define the $\mathcal{B}$-indiscernibility relation $R_\mathcal{B}$ obtained by the features in $\mathcal{B}$ as

$$R_\mathcal{B} = \{(x, y) \in U \times U \mid (\forall a \in B)(a(x) = a(y))\}.$$

The positive region of $\mathcal{B}$ is the union of the lower approximations of $[y]_d$ by $R_\mathcal{B}$, where $[y]_d$ is the decision class of y given a qualitative decision attribute d [27]:

$$\operatorname{Pos}(\mathcal{B}) = \bigcup_{y \in U} \underline{\operatorname{apr}}_{R_\mathcal{B}}([y]_d).$$

Concerning the applications, two monotonicity properties of Pawlak's rough set model are important. Given two subsets of the universe, $A_1 \subseteq A_2 \subseteq U$, the approximations of A_1 will be included in the approximations of A_2: $\underline{\operatorname{apr}}_R(A_1) \subseteq \underline{\operatorname{apr}}_R(A_2)$ and $\overline{\operatorname{apr}}_R(A_1) \subseteq \overline{\operatorname{apr}}_R(A_2)$. Hence, if a decision class gets larger, so do its approximations. Moreover, given two relations $R_1 \subseteq R_2 \subseteq U \times U$, i.e., U/R_1 is a finer partition of U than U/R_2, then the lower approximation by R_1 will be larger than the lower approximation by R_2, while the upper approximation will be smaller: $\underline{\operatorname{apr}}_{R_2}(A) \subseteq \underline{\operatorname{apr}}_{R_1}(A)$ and $\overline{\operatorname{apr}}_{R_1}(A) \subseteq \overline{\operatorname{apr}}_{R_2}(A)$. That is, for a finer relation we expect the lower approximation not to decrease and the upper approximation not to increase. Hence, when more features are taken into account, the relation $R_\mathcal{B}$ will become finer and thus, the positive region will increase. Therefore, the positive region of all attributes $\operatorname{Pos}(\mathcal{A})$ yields a maximum value. This value can be used to define dependency degrees (see Section 4.1.2). A common example of a dependency degree is the proportion of $\operatorname{Pos}(\mathcal{A})$ in U [27]:

$$\gamma = \frac{|\operatorname{Pos}(\mathcal{A})|}{|U|}.$$

2.2. Fuzzy Sets

A potential drawback of rough set theory is its inability to process quantitative data and graded indiscernibility directly. Instead, a discretization step can be performed, replacing the exact attribute values by interval codes. Another solution is the use of tolerance rough sets [116], where object attribute values are considered indiscernible if they are sufficiently close. Since both approaches potentially induce information loss, an alternative is taking fuzzy set theory [156] into account.

A fuzzy set A in the universe U is a mapping from U into $[0,1]$, i.e., $A\colon U \to [0,1]$. The value $A(x)$ is called the membership degree of the element x in the fuzzy set A. The fuzzy cardinality of A, denoted by $|A|$, is the sum of all membership degrees: $\sum\limits_{x \in U} A(x)$.

Besides fuzzy sets which extend classical sets, fuzzy logical connectives can be defined to extend the Boolean connectives in a fuzzy setting. The Boolean conjunction can be extended by a conjunctor C. A conjunctor is a mapping $C\colon [0,1]\times[0,1] \to [0,1]$ which is increasing in both arguments and satisfies the border conditions $C(0,0) = C(0,1) = C(1,0) = 0$ and $C(1,1) = 1$. If $C(1,a) = a$ for all $a \in [0,1]$, then C is called border. A border conjunctor which satisfies the following two conditions is called a triangular norm of t-norm and is denoted by T:

1. $\forall a, b \in [0,1]\colon T(a,b) = T(b,a)$ (commutativity),
2. $\forall a, b, c \in [0,1]\colon T(T(a,b),c) = T(a,T(b,c))$ (associativity).

Examples of well-known t-norms are the minimum operator $T_M(a,b) = \min(a,b)$, the product t-norm $T_P(a,b) = a\cdot b$ and the Łukasiewicz t-norm $T_L(a,b) = \max(0, a+b-1)$ for $a, b \in [0,1]$.

Furthermore, the Boolean implication is extended by an implicator I. An implicator is a mapping $I\colon [0,1]\times[0,1] \to [0,1]$ which is decreasing in the first and increasing in the second argument and satisfies the border conditions $I(0,0) = I(0,1) = I(1,1) = 1$ and $I(1,0) = 0$. Examples of implicators are the Kleene-Dienes implicator $I_{KD}(a,b) = \min(1-a,b)$ and the Łukasiewicz implicator $I_L(a,b) = \min(1, a-b+1)$ for $a, b \in [0,1]$.

The final notion to extend rough sets to a fuzzy setting, is that of a fuzzy indiscernibility relation. A fuzzy relation R is a fuzzy set in $U \times U$. Given a decision system $(U, \mathcal{A} \cup \{d\})$, a subset of attributes $\mathcal{B} \subseteq \mathcal{A}$ and a t-norm T, a possible fuzzy $\mathcal{B}$-indiscernibility relation $R_{\mathcal{B}}$ based on the attributes in $\mathcal{B}$ is defined by ([27])

$$\forall x, y \in U\colon R_{\mathcal{B}}(x,y) = T(R_a(x,y)),$$

where we take the t-norm over all fuzzy relations R_a related to the attributes $a \in \mathcal{B}$. The relations R_a are discussed below. If all $a \in \mathcal{B}$ are qualitative, we obtain the crisp $\mathcal{B}$-indiscernibility relation we saw before. It is often assumed that the fuzzy relations R_a are tolerance relations, i.e., reflexive ($R_a(x,x) = 1$) and symmetric ($R_a(x,y) = R_a(y,x)$). An example of such a tolerance relation is given in [27]: let a be a quantitative attribute in $\mathcal{A} \cup \{d\}$, then the $\{a\}$-indiscernibility relation between x and y is denoted by R_a and described by

$$R_a(x,y) = \max\Big\{0, \min\Big\{\frac{a(y) - a(x) + \sigma_a}{\sigma_a}, \frac{a(x) - a(y) + \sigma_a}{\sigma_a}\Big\}\Big\},$$

where σ_a is the standard deviation of a, i.e.,

$$\sigma_a = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(a(y_i) - \bar{a})^2},$$

with $\bar{a} = \frac{1}{n}\sum_{i=1}^{n} a(y_i)$ and $|U| = n$. If a is qualitative, then $R_a(x,y) = 1$ if $a(x) = a(y)$ and $R_a(x,y) = 0$ otherwise.

Some authors assume that the $\mathcal{B}$-indiscernibility relation $R_{\mathcal{B}}$ is a fuzzy T-similarity relation. Given a t-norm T, a fuzzy relation R is called a fuzzy T-similarity relation if it is reflexive, symmetric and T-transitive, i.e.,

$$\forall x, y, z \in U : T(R(x,y), R(y,z)) \leq R(x,z).$$

Here we recall two fuzzy T-similarity relations: one is based on a distance function, while the other is based on a kernel function. Let $x, y \in U$, then the Manhattan distance between x and y given an attribute $a \in \mathcal{A}$ is defined by $d_a(x,y) = |a(x) - a(y)|$ [38]. Denote $\delta_a = \max\{d_a(x,y) \mid x, y \in U\}$, then we obtain the relation R_a by

$$R_a(x,y) = 1 - \frac{d_a(x,y)}{\delta_a}.$$

It is clear that R_a is reflexive and symmetric. Moreover, since the Manhattan distance satisfies the triangle quality, R_a is transitive for the Łukasiewicz t-norm T_L. However, R_a is not transitive for the choices of T_M and T_P.

On the other hand, let $x, y \in U$ and $a \in \mathcal{A}$, then the exponential kernel function k_a is defined by $k_a(x,y) = \exp\left(-\frac{||x-y||}{\gamma}\right)$, where $||\cdot||$ represents the Euclidean distance and $\gamma > 0$ is a parameter defining the geometric structure of the kernel function [72]. Given k_a, the fuzzy tolerance relation R_a defined by

$$R_a(x,y) = \begin{cases} \exp\left(-\frac{(a(x)-a(y))^2}{2\sigma_a^2}\right) & \sigma_a \neq 0 \\ 1 & \sigma_a = 0 \end{cases}$$

is derived. This fuzzy relation is transitive with respect to the Łukasiewicz and product t-norms T_L and T_P, but not with respect to the minimum operator T_M.

3. Fuzzy rough set models

As discussed above, hybridization of rough and fuzzy set theory allows to perform data analysis on information systems with real-valued data sets directly. The bedrock of fuzzy rough set theory is situated in the late 1980s [14, 42, 112, 149]. Fuzzy rough set theory studies operators which approximate fuzzy sets with fuzzy relations.

The seminal contribution of Dubois and Prade to fuzzy rough set theory launched the research on approximation operators based on fuzzy logical connectives [43, 44]. They defined the lower and upper approximation of a subset $A \subseteq U$ by a fuzzy relation R with respect to the Kleene-Dienes implicator and the minimum t-norm:

$$\begin{aligned}(\underline{\mathrm{apr}}_{R,I_{KD}}(A))(x) &= \inf_{y\in U} \min(1-R(y,x),A(y)),\\ (\overline{\mathrm{apr}}_{R,T_M}(A))(x) &= \sup_{y\in U} \min(R(y,x),A(y)).\end{aligned}$$

Significant work was done in [111] where both a constructive and an axiomatic approach of fuzzy rough set models was discussed, as well as in [124], where lower and upper approximation operators based on three general classes of implicators were studied. Models using general fuzzy relations were first studied in [60, 61, 147, 148]. In [107, 108], general conjunctors were considered instead of t-norms. All these different approaches using fuzzy logical connectives are encapsulated in a general implicator-conjunctor fuzzy rough set model [33, 34]: let R be a fuzzy relation, I an implicator and C a conjunctor, then the lower and upper approximation of a fuzzy set A are defined by

$$\begin{aligned}(\underline{\mathrm{apr}}_{R,I}(A))(x) &= \inf_{y\in U} I(R(y,x),A(y)),\\ (\overline{\mathrm{apr}}_{R,C}(A))(x) &= \sup_{y\in U} C(R(y,x),A(y)).\end{aligned}$$

A drawback of this model is the influence of the worst and best performing objects, as the infimum and supremum only take one element into account. Therefore, the model is sensitive to noisy data and perturbations. As this problem occurs in the crisp models as well, Ziarko proposed the Variable Precision Rough Set (VPRS) model [164]. In the VPRS model, an instance $x \in U$ belongs to the lower (or upper) approximation of $A \subseteq U$, if the proportion $\frac{|[x]_R \cap A|}{|A|}$ exceeds a certain threshold. In other words, if there are enough elements in A to which x is indiscernible.

Many authors have attempted to overcome this disadvantage in the fuzzy setting. Their proposals are thoroughly discussed in [34], both from a theoretical and a practical point of view. They can be divided into three groups, which we recall briefly:

1. The first group are frequency-based models, analogous to the VPRS model of Ziarko. The Vaguely Quantified Fuzzy Rough Set (VQFRS) model [24] uses fuzzy quantifiers. In [69, 70], a Soft Fuzzy Rough Set (SFRS) model is introduced. However, this model does not satisfy monotonicity with respect to fuzzy sets. Furthermore, the Variable Precision Fuzzy Rough Set (VPFRS) model was studied in [109, 110] and the Variable Precision Fuzzy Rough Set model based on Fuzzy Granules (FG) was proposed in [153]. Nonetheless, all these models share a defect with the VPRS model: the monotonicity with respect to fuzzy relations, i.e.,

   $$R_1 \subseteq R_2 \subseteq U \times U \Rightarrow \forall A \subseteq U\colon \underline{\mathrm{apr}}_{R_2}(A) \subseteq \underline{\mathrm{apr}}_{R_1}(A),$$

   is not guaranteed.

2. The second group adjusts the approximated set, i.e., a threshold for the membership degrees $A(x)$ is taking into account in order to limit the influence of outlying values. An example is the Fuzzy Variable Precision Rough Set (FVPRS) model [163].

3. The third group uses other aggregation operators than the infimum and supremum operator. While the β-Precision Fuzzy Rough Set (β-PFR) model [48] used β-precision quasi t-norms, the Ordered

Weighted Average (OWA) based fuzzy rough set model [28] used OWA operators. Moreover, note that certain kernel based fuzzy rough set models [72, 75, 76] can be seen as OWA based models. Improvements of the models β-PFR and OWA are presented in [32].

Although relation based fuzzy rough set models are intensively studied in literature, other hybridization approaches have also been proposed over the last decades, like covering based models [154]. Other hybridization theories include the study of classification-oriented fuzzy rough sets [77] and the axiomatic approach of fuzzy rough set models, e.g., [99, 108, 111, 118, 146, 147, 148, 155].

4. Supervised learning

Supervised learning, learning with a teacher, implies that a set of labeled elements is available in the training phase of the method. Instances are described both by a number of input attributes as well as an associated outcome. A supervised learning algorithm constructs, based on its available set of labeled instances, a model to predict the outcome of new instances based on their input feature values. In case of a categorical outcome, the learning task is called *classification*. When the outcome is a continuous variable, it corresponds to a *regression* task. Often, the available dataset is preprocessed before application of a learner, in order to gain a speed-up in learning time or an increase in the prediction ability. Fuzzy rough sets have been used in the orthogonal preprocessing stages of feature selection (Section 4.1) and instance selection (Section 4.2). They have also been applied in classification (Section 4.3) as well as regression methods (Section 4.4). Finally, we also recall their use in neural networks (Section 4.5).

4.1. Feature selection

The feature selection (FS) or attribute selection paradigm reduces the dimensionality of a problem by replacing the feature set by a subset of it (e.g. [64]). By using a subset rather than a transformation of the features (e.g. PCA [117]), their original interpretation and semantics are preserved. It has been a popular focus of efforts integrating fuzzy rough set theory in machine learning techniques.

When reducing the attribute set, an algorithm aims to sufficiently preserve the original discriminative power of the system. In rough set theory, a *discernibility matrix* determines for each pair of instances the attributes by which they can be discerned, i.e., the ones for which these two instances take on different values. This provides a list of attributes for each pair of instances. For as long as the attribute system contains at least one attribute from each of these lists, it will be able to discern exactly the same amount of instances as the original set did. An additional aim of attribute selection is to remove redundancy. To this end, its goal is to obtain an attribute set from which no attributes can be removed without reducing the discerning capability. Such a minimal set of attributes with the same discerning capability as the full system is called a *reduct*. When the set is not necessarily minimal, it is called a *superreduct*. In [113], the general notion of a μ-reduct was investigated, where μ is a monotone function evaluating the quality of attribute subsets. In a classification problem, which is the process of assigning a class label to an instance based on its attribute values, it is usually sufficient to preserve the ability to discern between classes and not necessarily between individual elements. Here, a minimal attribute set with this property is called a *decision reduct* or *relative reduct*.

The search for a decision reduct starts with the important question whether all or only one reduct is required. As noted above, in rough set theory one can define a matrix expressing discernibility between objects. Based on this matrix, the discernibility function f is constructed. This function is defined in conjunctive normal form (CNF), taking the conjunction of all disjunctions constructed from the matrix elements, which are lists of attributes. It is therefore given by

$$f = \wedge\{\vee a \,|\, a \in c_{ij}\},$$

where c_{ij} is the matrix entry on the ith row and jth column. It has been shown [132] that a feature subset is a decision reduct if and only if its elements correspond to a disjunct of f in disjunctive normal form (DNF). Although this method does determine all reducts, it is computationally very expensive as the transformation of a function in CNF to DNF is an NP-hard problem. When a single reduct, corresponding to one feature subset, is required, it is more advisable to use a fast algorithm. To this end, heuristic approaches are used to approximate a solution. They provide one near-optimal set, meaning that the outcome can be a superreduct rather than an actual reduct. A popular example is the QuickReduct algorithm, an incremental hill-climbing algorithm increasing the feature subset, initialized as an empty set, by one more feature in each iteration until no further improvement can be found.

The extension of rough to fuzzy rough discernibility implies that elements can be discerned from each other in a certain degree rather than making this a yes-or-no question. The most general study of fuzzy rough attribute selection was presented in [27], an aim to unify previous research in a single framework. A general monotone measure $\mathcal{M}$ is used to determine the extent to which a reduced feature set $\mathcal{B}$ is able to discern between classes equally well as the full feature set $\mathcal{A}$. The measure satisfies $\mathcal{M}(\mathcal{A}) = 1$. As such, a *fuzzy $\mathcal{M}$-decision reduct* to degree α, with $\alpha \in]0, 1]$, is defined as a set $\mathcal{B} \subseteq \mathcal{A}$ for which $\mathcal{M}(\mathcal{B}) \geq \alpha$ and none of its strict subsets has the same property. If the minimality property does not necessarily hold, $\mathcal{B}$ is called a *fuzzy $\mathcal{M}$-decision superreduct* to degree α.

The two approaches referenced above, either using the discernibility function or a faster heuristic, have been extended for use in fuzzy rough attribute selection. The extension of QuickReduct is presented in Algorithm 1, using the notation of [27]. Over the last two decades, several authors have looked into fuzzy rough attribute selection. We present an overview of their work in the remainder of this section. The algorithms can be divided into three groups. Figure 2 presents a visual overview. Firstly, in Section 4.1.1, we consider methods using the discernibility matrix. This matrix is custom defined depending on the chosen fuzzy rough model. In the remaining two sections, we consider algorithms which are largely related to the QuickReduct algorithm in Algorithm 1. In Section 4.1.2, proposals using the fuzzy rough dependency degree as $\mathcal{M}$ are recalled. Section 4.1.3 considers methods using the entropy measure.

4.1.1. Discernibility matrix

In this section, we review attribute selection algorithms using a discernibility matrix in their internal workings. In a first proposal [151], no explicit use of fuzzy rough sets is made, although both fuzzy and rough components are incorporated in their feature selection algorithm. The authors develop a fuzzy discernibility matrix, where matrix elements are determined for pairs of sufficiently dissimilar and

Algorithm 1 General QuickReduct

Input: Set $\mathcal{A}$ of conditional attributes, degree α
Output: Set $\mathcal{B}$ of selected attributes

$\mathcal{B} \leftarrow 0$
repeat
    $T \leftarrow \mathcal{B}$
    $best \leftarrow -1$
    for all $a \in (\mathcal{A} \setminus \mathcal{B})$ **do**
        if $\mathcal{M}(\mathcal{B} \cup a) > best$ **then**
            $T \leftarrow \mathcal{B} \cup a$
            $best \leftarrow \mathcal{M}(\mathcal{B} \cup a)$
    $\mathcal{B} \leftarrow T$
until $\mathcal{M}(\mathcal{B}) \geq \alpha$

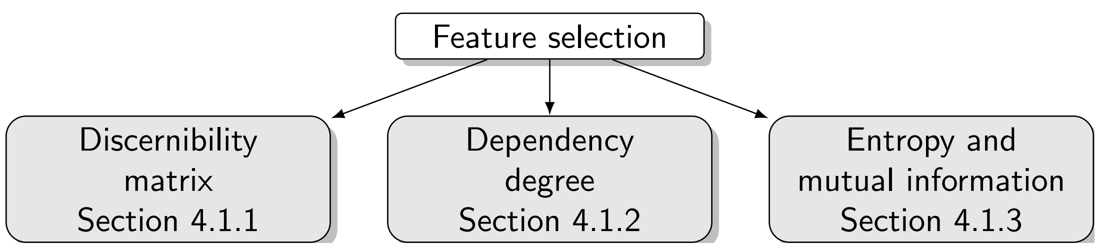

Figure 2. Overview of the different groups of FS algorithms and the sections in which they are discussed.

opposite-class instances. These matrix elements are sets of attributes, for which the values are sufficiently different between the instances. The matrix is combined with a rough attribute reduction method in a genetic algorithm.

True integrations of fuzzy rough set theory for feature selection can be found in [20], [21] and [137]. They use the traditional fuzzy rough set model from [43], but apply the Łukasiewicz t-norm T_L when measuring feature-wise similarity between instances. The similarity between two instances based on all conditional attributes is computed by taking the minimum of the feature-wise similarities. They define an *indispensable* attribute as one leading to a decrease in the positive region after its removal. If a removal of an attribute does not lead to such a decrease, it is called *dispensable*. The *core* of the attribute set is defined as the set of indispensable attributes and is shown to be the intersection of all relative reducts, that is

$$\text{core} = \cap\{R \mid R \text{ is relative reduct}\}.$$

They define a relative reduct as a subset of the full attribute set which keeps the fuzzy rough positive region of the system invariant. Placing this in the framework of [27] discussed above, they use the ratio of the fuzzy cardinality of the fuzzy rough positive region based on an attribute subset and that based on the full attribute set as the measure $\mathcal{M}$. They set α to 1. The authors develop an algorithm to construct all these relative reducts using a discernibility matrix. They define the matrix elements as empty sets, when the corresponding instances belong to the same class. For opposite-class instances x_i and x_j, the element c_{ij} in the matrix is a set of attributes a for which

$$a(x_i, x_j) + \lambda(x_i) \leq 1$$

holds, where $\lambda(x_i)$ corresponds to the membership degree of x_i to the lower approximation of its decision class and $a(\cdot, \cdot)$ represents the similarity of two instances based on the attribute a. The algorithm computes the core as the union of singleton set matrix elements. Every matrix element that is an empty set or has a non-empty overlap with the core is removed. The remaining matrix elements are aggregated in CNF, which is converted to DNF, from which it is concluded that every disjunct corresponds to a relative reduct.

In [23], the authors develop a similar approach to determine a local reduct, which is a reduct based on a subset of the decision classes. They aim to identify important conditional attributes for each decision class. They show that every attribute contained in such a local core is indispensable for at least one decision class. An attribute is again said to be indispensable if its removal results in a decrease in the fuzzy rough positive region. Based on a subset $\mathbf{K}$ of decision classes, they construct a discernibility matrix by setting the entry c_{ij} to $T(R(x_i, x_j), \lambda(x_i))$, when x_j does not belong to the decision class of x_i in $\mathbf{K}$. All other entries are empty sets. They use a t-norm T and a fuzzy T-similarity relation $R(\cdot, \cdot)$. They show that the core for $\mathbf{K}$ is the union of the matrix entries representing singletons. They present a heuristic to find one local reduct based on the discernibility matrix. After its construction, they define the core as the union of the singleton set matrix entries. The remaining entries with a non-empty intersection with the core are removed. The candidate reduct is initialized as the core. Next, an iterative procedure is performed to reduce the size of the reduct.

In [162], a process of attribute reduction preceding rule induction is developed, using a general fuzzy rough set model, less sensitive to misclassification and perturbation. The feature selection phase is based on the construction of a discernibility vector. Both an algorithm to find all decision reducts, once more applying a conversion from CNF to DNF, and a heuristic to find a near-minimal one are proposed.

Another proposal is the attribute selection algorithm of [18], that makes use of the Gaussian kernel based fuzzy rough set model [72]. They construct a discernibility matrix by setting c_{ij} to a set of conjunctions of attribute subsets $\mathbf{P}$ for opposite-class elements x_i and x_j and to an empty set otherwise. The sets $\mathbf{P}$ are determined as minimal sets of attributes based on which x_i and x_j are sufficiently dissimilar. They apply the same heuristic method as described in [23] to determine one reduct based on the discernibility matrix, keeping the fuzzy rough positive region invariant compared to that of the full system. The set of all reducts can be determined by constructing the discernibility function in CNF and converting it to DNF, concluding that every disjunct is a reduct.

The algorithm of [67] defines an individual discernibility matrix for all fuzzy decision classes. The separate matrices are joined to form the main discernibility matrix by setting its elements to the conjunction of the disjunctions of the corresponding sets for the separate decision classes. An algorithm to find all reducts is presented, with the lower approximation of all fuzzy decision classes as invariant factor. In the scheme of [27], this means $\alpha = 1$. The authors note that the proposed algorithm is time-consuming and consequently also present a heuristic alternative to find a close-to-minimal reduct, based on the discernibility matrix as well.

Finally, [22] specifically focused on the important downside of using a discernibility matrix, which is the computational complexity. For efficiency purposes, the authors proposed to only use minimal elements in the discernibility matrix. A minimal element in the discernibility matrix is defined as an entry for which there does not exist another entry that is a strict subset of it. The authors note that when supersets of other sets are included in the CNF, they will disappear during the conversion to DNF as a consequence of the absorption law, which states

$$P \wedge (P \vee Q) \equiv P.$$

By not including these supersets in the CNF, a speed-up in running time can be obtained. Additionally, if the minimal elements can be determined prior to the construction of the entire discernibility matrix, the need for this full computation is removed and the storage requirements reduced. Accordingly, the authors develop an algorithm collecting these minimal elements avoiding the construction of the full matrix. They furthermore propose a method to find a single reduct. Both algorithms have polynomial complexity. The set of minimal elements can be used to find all reducts, again by converting the corresponding discernibility function from CNF to DNF.

4.1.2. Dependency degree

As indicated in Section 2.1, the dependency degree is a measure which expresses how well a set of attributes can discern between elements, that is, what proportion of the dataset can be discerned when restricting the attributes to the current set. Its exact definition depends on the fuzzy rough set model and proposal at hand, but it makes use of the fuzzy cardinality of the positive region. The membership degree of an element to the fuzzy rough positive region expresses the extent to which instances similar to it also share its decision label. The similarity between elements is measured based on the reduced attribute set.

The first proposal of a fuzzy rough selection criterion was put forward in [94]. The authors define an approximation measure of a fuzzy set, similar to a fuzzy rough dependency degree, based on its similarity with its positive region on the one hand and the similarity with its boundary region on the other. In determining these sets, the proposal does not use a clear standalone fuzzy rough set model, but rather develops a separate one, based on a degree of inclusion of one fuzzy set into another.

A series of papers was proposed by Jensen and Shen, outlining the development of a fuzzy rough QuickReduct algorithm using the dependency degree both to guide the search and as stopping criterion. After reviewing rough and fuzzy rough attribute selection approaches in [86], the first proposals in [85] and [131] define a reduct as a minimal subset of conditional attributes that has the same dependency degree as the full attribute set. An approximation of a single decision reduct is computed by the hill-climbing paradigm in QuickReduct. Taking on a greedy approach, the resulting set is not necessarily minimal. The reduced attribute set is initialized as an empty set and consequently built up by adding the attribute resulting in the largest increase in dependency degree in each iteration. The algorithm halts when the constructed attribute set has the same dependency degree as the entire attribute set. Example applications to bookmark and website classification are provided. In [87], the authors consider the use of ant colony optimization [13] for fuzzy rough feature selection and experimentally show that the use of this meta-heuristic outperforms the previously proposed fuzzy rough QuickReduct algorithm on a two-class dataset, while the opposite holds on a three-class dataset. They also conducted an experimental metric comparison in [89] for fuzzy rough QuickReduct, showing that the fuzzy rough dependency degree is indeed appropriate to use when determining relevant features.

In the above, they defined this measure using fuzzy equivalence classes, based on the Cartesian product of fuzzy partitions of the separate attributes. This dependency degree is given as the ratio of the fuzzy cardinality of the positive region and the number of instances. The former uses the fuzzy equivalence classes to discern between elements. After comments in [9] on its convergence and in [137] on its lack of theoretical foundation, the authors formulated a response in [90] and proposed three new extensions of the feature selection method in [91]. In a first version, they use the similarity relation based on an attribute subset at hand, rather than a fuzzy partitioning of the feature space to construct fuzzy equivalence classes. Secondly, they consider an alternative definition of the dependency degree. Instead of using the fuzzy rough positive region, they turn to the fuzzy rough boundary region and use it to define a total uncertainty degree related to a feature subset. An analogous hill-climbing algorithm minimizing this total uncertainty is put forward. They note that their definition of this measure is related to that of the entropy measure discussed in Section 4.1.3. The third extension involves the construction of a fuzzy discernibility matrix. Once again, a heuristic approach is taken to build up a near-minimal feature subset satisfying as many discernibility requirements posed by the matrix to the largest degree possible. An alternative extension was developed in [39], where, building on the fuzzy rough QuickReduct algorithm, the alternative search method of harmony search [56] was applied to search for an optimal feature subset. Recently, they proposed a computationally less expensive way to calculate the dependency degree and fuzzy discernibility matrix, by precomputing the k nearest other-class neighbors for each instance in the dataset [84].

Various other authors consider using different fuzzy rough set models rather than the traditional one. In [8, 9, 10] for instance, a new fuzzy rough set model computing approximations on compact domains instead of the entire universe was proposed. The authors modified the fuzzy rough QuickReduct algorithm accordingly and were able to experimentally show a significant speed-up in computation time. Their proposal involves a tree structured search algorithm. We encounter the application of another alternative fuzzy rough set model in the work of [71], using a VPFRS model (Section 3) in their attribute reduction method. The authors define the fuzzy rough dependency degree as the ratio of the cardinality of the positive region, calculated within their proposed model, and the number of instances in the decision system. They consider the forward or backward search for a single reduct and propose to use greedy strategies. In each iteration of the forward search, the attribute leading to the largest increase in dependency degree is added to the current set. The backward search, on the other hand, sets out with the entire attribute set and removes a feature when this action does not result in a decrease of dependency degree. Similarly, VPFRS is used in [143], where attribute reduction is applied as one step in the pipeline from going from fuzzy data to a fuzzy rule set. Rather than the VPFRS model, the algorithm of [66] is based on Gaussian kernel based fuzzy rough sets [72]. Its developers determine a decision reduct as the smallest attribute subset with a positive region within a small perturbation ε of the positive region of the full decision system. In the same light, we also reference the study of [161], where, instead of resorting to an alternative model, the authors of [161] present a theoretical and experimental study on the effect of different fuzzy approximation operators and t-norms within the traditional fuzzy rough set model on the results of attribute reduction.

Several proposals illustrate the negative effect of noise on feature selection and aim to develop robust algorithms. In [26], a noise-tolerant version of fuzzy rough QuickReduct is put forward, where the search is guided by the dependency degree based on the VQFRS model of [24]. On the other hand, like [71] referenced above, the proposal of [119] uses VPFRS in their custom modification of rough QuickReduct. Their algorithm is denoted as iQuickReduct and was further extended in [120]. In the initial proposal, they use the dependency degree to assess the quality of feature subsets, defining it based on the fuzzy rough positive region in their model. The extension involves the use of different quality measure heuristics, like information gain, maximum domain cardinality or a combination of both. Another proposal considering the robustness against noise of its feature selection algorithm, can be found in [69]. The authors introduce the SFRS model (Section 3) and define the dependency degree accordingly, again setting it to the ratio of the cardinalities of the positive region, as determined by their model, and the entire instance set.

Recently, the proposal of [106] combined feature selection with feature extraction, which is the construction of new features. They use the dependency degree to evaluate feature significance. They develop an iterative process where, based on a given threshold, insignificant features are removed, significant features are used to select one out of them and the remaining features are used to construct a new one.

Example applications of fuzzy rough dimensionality reduction can for instance be found in the study of [104], presenting a unified framework for mammographic risk analysis, which combines the fuzzy rough nearest neighbor classifier (Section 4.3.2) with attribute selection. In [41], the image recognition application of driver fatigue detection is discussed. The authors use the Gaussian kernelized fuzzy rough set model. The experimental study of [101] focuses on the application of attribute selection in image retrieval, comparing fuzzy rough algorithms to other feature selection methods.

4.1.3. Entropy and mutual information

When learning decision trees, the information entropy is an often applied measure to decide on the quality of potential splits. It measures the degree of impurity in a group of observations and is in general defined as

$$\text{entropy} = \sum_{i=1}^{c} -p_i \log p_i,$$

where c is the number of classes in the dataset and p_i is the class probability of the ith class, that is, its frequency among the group of observations. The higher the value of the entropy, the more diverse or impure the observation set is. Based on the entropy, the information gain in decision tree learning is defined as the difference between entropy in a parent node and the average entropy of its children, induced by splitting on a chosen feature. The higher the information gain, the more impurity has been dispensed with by selecting this particular feature to split on, i.e., the more important this feature is. It also corresponds to the definition of mutual information, which is the difference between the regular and conditional entropy, where the latter is defined using conditional class probabilities. A modification to this measure is the gain ratio, which takes the number of branches as well as their size into account in the split selection procedure. Several authors have generalized the notion of entropy for use in fuzzy rough feature selection. We discuss them below.

The proposal of [74] uses fuzzy probabilistic approximation spaces, integrating three types of uncertainty: probability, fuzziness and roughness. The probability distribution is introduced in the definition of fuzzy cardinality of the positive region: instead of summing over the membership degrees to determine its value, i.e. assigning equal weight $\frac{1}{n}$ to all elements, elements are weighted by means of $p(\cdot)$. Within such a space, they define information quantity of a feature subset, entropy and conditional entropy and define a reduct based on the latter measure. Similarly, in [73], the computation of the strength of an attribute set to discern between elements is based on such an information measure. The authors propose two greedy heuristic algorithms, one determining a reduct and the other a relative reduct, measuring the significance or quality of features by means of the entropy (reduct) or conditional entropy (relative reduct).

In [121, 122], a more efficient approach to attribute selection is proposed, using forward approximation. It is based on an ordered list of fuzzy relations, each list element being more specific than its predecessor. The authors show that smaller similarity matrices need to be stored and that the computation of attribute significance can be greatly reduced. This significance is measured by the mutual information criterion.

The authors of [150] present an incremental algorithm computing one relative reduct, using conditional entropy to measure attribute significance. They present an example application of feature selection, corresponding to gene selection, for cancer classification. Their experimental study shows that the proposed algorithm is effective in removing noise from the data. In [31], a tumor classification application is described. The gain ratio is applied in an attribute selection method. It is used to evaluate the use of adding an attribute a to a partially constructed set $\mathcal{B}$. The gain itself is defined by means of the fuzzy equivalent of the mutual information and gain ratio by dividing out the information quantity (entropy). Using this criterion, their greedy algorithm adds the best features to the attribute subset, until no more gain is obtained.

4.2. Instance selection

Instance selection (IS) is the dual procedure of FS, discussed in the previous section. It reduces the size of the problem by removing entire instances from the dataset. In doing so, it can remove noisy elements and thereby improve the performance of posterior learning algorithms. Furthermore, a considerable reduction in the number of instances corresponds to reduced storage requirements and running times. Apart from IS, *prototype selection*, *training set selection* and *sample selection* are used to denote this set of methods as well. IS has predominantly been used as a preprocessing step for classification algorithms. Each of the methods discussed in this section was specifically designed to be executed on classification datasets, that is, we assume that all instances have an associated class label, drawn from a finite, discrete set of possibilities. Classification techniques incorporating fuzzy rough set theory themselves are discussed in Section 4.3.

4.2.1. Fuzzy rough IS methods

In [81], a first IS method using fuzzy rough set theory was developed. The authors considered three different decision criteria for the removal or inclusion of instances in the preprocessed dataset. The membership degree to the fuzzy rough positive region is used to determine how useful instances are, incorporating a granularity parameter α in the feature-wise relationship between instances. In a first version, a threshold τ is provided by the user. The algorithm removes instances for which the membership degree to the fuzzy rough positive region is strictly smaller than τ. This implies that only sufficiently typical elements are retained in the dataset. The second alternative does not require the specification of a τ value and is based on the idea that the removal of certain elements should not reduce the membership degrees of other instances to the fuzzy rough positive region. In each iteration, it removes the instance with the lowest membership degree to the fuzzy rough positive region. It halts when all instances fully belong to the fuzzy rough positive region. Finally, the third version uses backward elimination, removing instances for which this results in the largest increase in the value of the dependency degree γ. The termination criterion is a dependency degree of one. It was experimentally shown that the first and simplest version performed best.

The later proposal of [139] is a more complex IS method, using a wrapper approach to evaluate entire candidate instance subsets instead of single instances. Its specific aim is to remove noise on the data. The quality of each instance is evaluated using a granularity parameter and the fuzzy rough positive region. In this procedure, OWA-based fuzzy rough sets [28] are used. Candidate subsets are formed out of sufficiently high quality instances. The subsets themselves are evaluated by testing their performance in a classification experiment. The best performing one out of the generated set constitutes the final result of the IS method. The method was further optimized in [140], using an alternative quality measure, dispensing with the granularity parameter.

4.2.2. Combining FS and IS

The FS and IS paradigms can be used in conjunction, reducing the dataset in two ways: by selecting the most informative features and instances. In [36], the combination of evolutionary IS and fuzzy rough FS was proposed and experimentally evaluated. The framework involves iterative applications of one round of FS, to generate a feature subset, combined with the execution of a number of generations of a genetic algorithm for IS, optimizing the instance subset. An opposite approach was taken in [37], proposing the hybridization of fuzzy rough IS and evolutionary FS. Here, the IS algorithm sorts elements in increasing order of their membership degree to the positive region and outputs the subset encountered during this procedure that resulted in the highest classification accuracy of the nearest neighbor classifier. The evolution of an optimal feature subset by a genetic algorithm is alternated with the application of fuzzy rough IS to reduce the instance set. In both experimental studies, it was shown that these hybrid settings are able to significantly improve the classification performance, outperforming both the basic IS and FS models as well as other hybrid proposals. Even though applying both, the above still keep the FS and IS stages separate. In [103], the authors proposed the simultaneous application of FS and IS by using fuzzy rough bireducts.

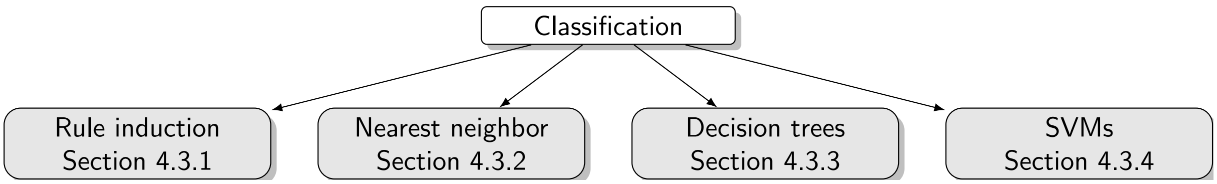

Figure 3. Overview of the four included classification algorithms, together with the sections in which they are discussed.

4.2.3. Imbalanced data

Recently, specific proposals of IS for imbalanced datasets have been put forward. In an imbalanced dataset, the instances are unevenly divided among the classes. Such class imbalance has been shown to pose problems for standard learning algorithms and several remedies have been proposed to resolve the issues related to it (e.g. [135]). A popular approach consists of preprocessing the dataset in order to create a better balance between classes. In [126], fuzzy rough set theory has been used in conjunction with the data sampling technique SMOTE [16], which attains a more balanced class distributions by artificially creating new elements of the underrepresented class. It has been noted that SMOTE can lead to the overgeneralization of the minority class (e.g. [125, 133]). The proposal of [126] aims to resolve this matter by filtering the modified dataset after application of SMOTE. In particular, it removes artificially generated or majority class instances for which the membership degree to the fuzzy rough positive region is lower than a user-specified threshold. In this way, atypical elements are not retained in the dataset. Alternatively, the authors of [142] proposed to apply an IS algorithm prior to application of SMOTE in order to clean the data and thereby ensuring that noisy elements do not take part in the construction of artificial elements. They used the algorithm of [139], although modifying it to better suit the setting of class imbalance.

4.3. Classification

In classification datasets, each instance is associated with a single class, represented by a class label. This is a value drawn from a finite and discrete set of possibilities, corresponding to the classes. A classification algorithm or classifier sets up a classification model based on the available labeled training instances. Newly presented instances are assigned a class label by processing them with the constructed model. As it can appropriately deal with the combined presence of vague and incomplete information in the training set, fuzzy rough set theory has been incorporated in several classifiers. Figure 3 presents an overview of the methods reviewed here. For each of them, we recall its original setup and list the fuzzy rough modifications that have been made to it.

4.3.1. Rule induction

Rule induction involves the construction of a number of IF-THEN rules, based on the data at hand. It provides a comprehensible classification model, as opposed to some other classification algorithms below, for which the constructed model is less human-interpretable. An example fuzzy decision rule could be

$$\text{IF } X_1 \text{ is } A_1 \text{ and } X_2 \text{ is } A_2 \text{ and} \ldots \text{and } X_m \text{ is } A_m \text{ THEN } Y \text{ is } B.$$

The rule consists of a number of antecedents A_1, A_2, ..., A_m and one consequent B, which are all fuzzy sets.

In [57], a fuzzy rough set model without using fuzzy connectives is proposed, following the argument of the authors that using the latter implies an arbitrary choice between them. Their model is solely based on the fuzzy membership degrees and forms a basis for fuzzy rule induction. They determine positive and negative relations between variables and induce lower and upper approximation decision rules. Based on them, specific lower and upper approximation operators are defined. As the rules are solely based on membership degrees and relations between variables, there is indeed no use of fuzzy connectives in the proposed model. The decision rule base is subsequently constructed by using the approximation operator.

The authors of [143] set up a complete pipeline for classification by means of fuzzy rules. As a preprocessing step, fuzzy rough attribute reduction based on a β-PFR model is performed. Next, the fuzzy rule base is constructed on the reduced dataset and the classification of new instances is performed with these rules. Similarly, in [138, 162], a separate attribute reduction step is also applied first. The authors define the consistency degree of an instance according to its membership degree to the lower approximation of its decision class. Based on this measure, the attribute set is reduced, keeping the original consistency degree invariant. In the second phase, a near-minimal rule set is induced.

In the work of [83], however, the authors noted that a clear separation between the preprocessing and rule induction phases could be suboptimal, as it can lead to the creation of too specific rules. They therefore proposed a framework hybridizing feature selection and rule induction. They present an adaptation of the fuzzy rough QuickReduct algorithm, simultaneously generating fuzzy rules covering a maximal part of the training set with a minimal amount of features in each step.

Further rule induction algorithms using fuzzy rough set theory include the algorithm of [68], which uses the fuzzy rough boundary region in its rule generation mechanism and furthermore makes use of a hierarchical structure existing in attributes to make rules more general, and the proposal of [100], which incorporates the specific aim to obtain a compact, and thereby more easily interpretable, rule set.

4.3.2. Nearest neighbor classification

The k nearest neighbor classifier [30] assigns a class label to a newly presented instance by locating its k nearest training instances, based on a given distance measure. This implies that no classification model is constructed, but rather that the training set is stored in full and evaluated entirely for each target instance. A fuzzy rough version of this classifier has been developed and improved in several stages. We recall its evolution.

The authors of [92] proposed a fuzzy version of the nearest neighbor rule (FNN). Each target instance is assigned a membership degree to each of the c classes, as opposed to a specific class label. This membership is based on the distance to its k nearest neighbors and their membership degrees to the decision classes. The latter values are determined based on the training data. A straightforward approach is the crisp one, namely assigning each instance membership degree 1 to its own class and 0 to all others.

Alternative ways are proposed as well, for instance by basing these values on distances to class centroids. In [12], FNN was modified by introducing the use of the fuzzy rough lower and upper approximation of the decision classes to classify new instances.

Similarly, in [128], the fuzzy rough nearest neighbor (FRNN) algorithm was introduced, integrating fuzzy rough set theory into the nearest neighbor classifier, supported by the claim that this leads to the enrichment of the classification semantics. The authors argue that any classification procedure inherently encounters two types of uncertainty. The first one, referred to as *fuzzy uncertainty*, is due to overlap between classes, such that elements can be considered as belonging to several classes. Secondly, they put forward the *rough uncertainty*, which concerns the fact that there may not be enough features to discern between the classes, i.e., the feature space is not sufficiently complete. Since fuzzy rough sets are the hybridization of two models dealing with these separate issues, they constitute the ideal candidate to resolve them. We note that this work makes no specific reference to the proposal of [12]. FRNN uses all elements in the training set as neighbors, instead of first determining the k nearest ones. It defines a fuzzy rough ownership function based on the two types of uncertainty mentioned above. For each class, the membership of a target instance to it is determined using all training instances, weighting them based on the distance to the target. The predicted class label is that for which this membership degree is maximal.

The main comment on this method is that it does not use the characterizing aspects of fuzzy rough sets, namely the lower and upper approximation operators. The later proposals of [80, 82] do introduce a nearest neighbor algorithm using the lower and upper approximations. Like FNN, this algorithm uses k nearest neighbors in the classification process. Based on these neighbors, lower and upper approximations of the classes are calculated. The membership degree of the newly presented instance to these fuzzy sets is used to decide on a class label. Two variants have been proposed, one using traditional fuzzy rough sets, referred to as FRNN, and one using the vaguely quantified fuzzy rough set model of [24], called VQNN. The latter is specifically put in place to deal with noisy data. The work of [123] further extends the methods in [80] by using kernel-based fuzzy rough sets.

Lastly, in [141] it is shown that the FRNN method from [82] takes into account only one instance and that VQNN does not differ from the original FNN method, when the same function is used to measure similarity between instances. They prove that FRNN will always classify a newly presented instance to the class of its nearest neighbor. They argue that fuzzy rough set theory is as such not used to its full strength, since the method basically coincides with the nearest neighbor rule, using the similarity function as distance measure. They introduce a more advanced weighting scheme for the k neighbors, using both their similarity to the target as well as their membership degree to the fuzzy rough positive region. By doing so, the classification takes the classification quality of an instance into account in its influence as a neighbor. Another modification of the method of [82] is found in the recent proposal [127]. It uses the OWA fuzzy rough set model [28] to allow for more neighbors to take part in the decision process, each being assigned a different weight. The method is specifically designed for imbalanced data: in the OWA aggregation, it uses different weight vectors for the majority and minority classes.

4.3.3. Decision trees

The classification model constructed by decision trees (e.g. [49]) is a tree, a graph without cycles. A decision tree is grown by starting with the root node, representing the entire set of instances. The construction is continued by iteratively splitting nodes. In each step, it is assessed for each node whether it will be used as a leaf of the tree or to select an attribute to induce a split. A node is used as a leaf when it is pure, meaning that it represents instances of only one class. If not, a split is constructed, leading to the generation of a number of child nodes, each representing a subset of the instances represented by their parent.

Several proposals for fuzzy rough decision trees have been developed. The integration of fuzzy rough set theory can be found in the splitting phases. A split in a decision tree is based on a single feature. To determine which feature to split on, measures like information gain or impurity reduction are traditionally used [49]. Intuitively, they allow the selection of a splitting feature such that the resulting tree is the most powerful, that is, best able to distinguish between classes. The fuzzy rough dependency degree has been used in [2, 11, 88, 157] to measure feature significance. The feature with the maximal importance is selected to induce a split. Alternatively, [45] uses a roughness measure for fuzzy partitions, of which the value is sought to be minimized by splitting on additional features.

4.3.4. Support vector machines

Support vector machines (SVMs, [29]) construct a decision hyperplane during the training phase. When the dataset is linearly separable, the hyperplane is chosen such that the boundary it induces between the classes is maximal. Obviously, the assumption of linear separability is not realistic. Therefore, soft-margin SVMs, as opposed to hard-margin ones, allow for elements to be located on the incorrect side of the hyperplane and trade off training misclassifications with separation of the classes. Furthermore, the construction of SVMs can involve the use of a kernel function, mapping the feature space to another higher-dimensional space in which the linear boundary is constructed. By doing so, a more complex boundary is obtained in the original space. Determining the optimal hyperplane is an optimization problem and is solved by using Lagrange multipliers.

In [19], the authors discuss the geometrical interpretation of a fuzzy rough similarity relations, as well as the link between them and kernel functions, as both measure similarity between objects. They reformulate the optimization problem of soft-margin SVMs by incorporating the membership degrees of instances in its constraints. Similarly, in [17], the membership degrees are used in the constraints of an extension of hard-margin SVMs.

4.4. Regression

Regression methods estimate a real-valued outcome based on the feature values of instances. A relatively smaller amount of attention has been directed to integrating fuzzy rough set theory into regression methods. As opposed to the previous section on classification, we discuss only two proposals using fuzzy rough sets for regression purposes below.

In Section 4.3.2, we recalled the FRNN and VQNN methods for classification introduced in [82]. The authors proposed alternatives to be used in regression problems as well. The algorithms are slightly modified, aggregating the outcomes of the k nearest neighbors of the target instance, using lower and upper approximation membership degrees of these neighbors to determine their weight in the aggregation.

In the recent proposal [3], a regression model using fuzzy rough set theory was developed. Their method involves a preliminary step of transforming the data into a fuzzy partition of k fuzzy sets based on the outcome. Secondly, to estimate the outcome of a test instance, its membership degree to the lower approximation of each of these k fuzzy sets is determined. For the fuzzy set F to the lower approximation of which this membership is largest, the outcome prediction is made based on the membership degrees of the instance to the lower and upper approximations of F.

4.5. Neural networks

As a final part of this section on supervised learning, we recall neural networks [65]. We discuss them separately, as they can be applied in numerous learning tasks, including classification, regression, feature selection and so on. The learning mechanism incorporated in a neural network is inspired by the biological workings of neurons, which are set up in a complex structure analyzing information using several processing layers. In machine learning, a neural network contains an input and output layer and a number of hidden layers in between. Each layer contains a number of processing nodes, which are connected among the layers with weighted edges. Each node performs an aggregation by combining its inputs with the corresponding weights and further transforming this value by means of an activation function, like the sign function. The final output is therefore a complex combination of the initial inputs. Simply put, learning a model corresponds to optimizing the connection weights. During the training phase, the weights are adaptively modified by processing the training instances one at a time and minimizing the prediction error made on them.

Fuzzy rough set theory has been integrated in the neural network design. For instance in [130], the authors develop a neural network mimicking a fuzzy rough membership function, expressing the membership degree of an instance to a fuzzy cluster. In particular, they propose a three layer neural network. The input layer has as many nodes as there are features in the dataset. The hidden layer has C nodes, with C being the number of clusters present in the training set. Lastly, the number of nodes in the output layer is put equal to the number of classes. When an input pattern is presented to the network, the corresponding output represents the class membership degree of this instance, for each class. This fuzzy rough neural network was used in vowel classification, classifying instances to the class for which the corresponding output node yielded the largest membership degree.

Similarly, in [160], a four layer feed-forward fuzzy rough neural network was set up. Their four layers correspond to separate input, clustering, membership degree and output stages. All layers are fully connected. They applied their model to feature selection. The input layer contains as many nodes as there are features in the dataset. For each instance the value of the jth feature is fed to the jth input node as input. The first hidden layer calculates the membership degree of the instance to a number of fuzzy clusters, where each fuzzy cluster is represented by one node in this hidden layer. The membership degree to fuzzy equivalent classes represented by the clusters are aggregated in the second hidden layer to compute the value of the fuzzy rough membership function. This layer contains a number of nodes equal to the number of classes in the dataset. The output layer computes a membership degree of the instance to each of the classes in the dataset. There are as many output nodes as there are classes. An instance is assigned to the class corresponding to the output node with the highest value. To perform feature selection, a backward search is applied. In the first stage, the network is trained using all features and a corresponding prediction error can be estimated. Afterward, in each iteration, the feature with the smallest variation is discarded and the network is retrained without this feature. If the resulting error has increased too much compared to the previous step, the feature is included again. In the other case, it is permanently removed. This process is continued until no more features can be deleted.

An alternative fuzzy rough set theoretic approach can be found in [129], where it is applied to determine the importance of subsets of information sources. These subsets correspond to the output of subnetworks, which are trained by dividing the input data in smaller subgroups. In [50, 51, 54] a fuzzy rough three-layer fully-connected perceptron using back-propagation to train the connection weights was proposed. The input is fed to the network in granular form, that is, three fuzzy granules *low*, *medium* and *high* are defined for each attribute and the membership degree of all instances to each of these fuzzy sets are determined. These membership degrees are used as input to the network. The number of input nodes is set equal to the number of fuzzy granules and therefore equals three times the cardinality of the attribute set. For each of the c classes in the dataset, the fuzzy rough dependency degree of each attribute with respect to this particular value of the decision attribute can be determined. These values are used as connection weights of the input and hidden nodes, where the number of hidden nodes is c. Furthermore, the connection weights between the hidden and output layers, both containing c nodes, are determined by averaging feature-wise dependency degrees. Each output node corresponds to a class, to which the membership degree is determined. The authors considered the application of their proposal for both classification [50, 51] and feature selection [54].

5. Unsupervised learning

As opposed to supervised learning, an unsupervised learner does not have the associated outcome for the training instances at its disposal. Instead, its aim is to discover patterns in the training set, solely based on the values of the input features. In Section 5.1, we discuss the use of fuzzy rough set theory in a popular clustering algorithm. In Section 5.2, we recall the self-organizing map (SOM, [65]) paradigm and review some integrations of fuzzy rough sets within it.

5.1. Fuzzy c-means clustering

In clustering methods, the aim is to construct a partition of the feature space, assigning each point of the training set to one cluster. More generally, fuzzy clustering techniques, like the fuzzy c-means method [6, 7], construct each of the c clusters as a fuzzy set, to which all training instances can have a certain membership degree. One of the goals of a clustering algorithm is to divide the dataset in groups which are as dissimilar as possible. Therefore, in [95], the fuzzy c-means algorithm was augmented with a fuzzy rough evaluation measure to evaluate between-cluster similarity of the fuzzy clusters. By means of this index, the results of several runs of the method, with varying values of c, can be compared in order to find the optimal number of clusters to be formed. On the other hand, a clustering also aims to ensure that when two instances have a high membership degree to the same fuzzy cluster, that they are highly similar. In [105], the authors looked into an application to microarray data, where the task is the clustering of genes. Genes assigned to the same cluster should be highly similar, of which the biological interpretation is that they are coregulated. Genes in separate clusters should be highly dissimilar. A fuzzy rough measure based on the dependency degree was used to measure gene similarity.

5.2. Self-organizing maps

SOMs are also used to find structure in the data solely using unlabeled information and is a form of unsupervised neural network (Section 4.5). In a SOM, the input layer is fully connected to a computational layer, which is represented by a two dimensional grid structure. Initially, all connection weights are set to small random values. For each presented training pattern, the distance to all neurons in the computational layer is computed and the pattern is assigned to the one to which this distance is the smallest. For this winning neuron, the connection weights to the input are modified, as well as for its neighboring neurons in the grid, albeit in a smaller way.

In [52, 53], the authors proposed a SOM model using fuzzy rough set theory. Similar to their work on neural networks discussed in Section 4.5, the input to the network is provided by constructing fuzzy granules based on the input data. The computational layer consists of at least c nodes, where c is the number of classes in the dataset. The initial connection weights from the input to these nodes are set using the fuzzy rough dependency degree, while the initial weights to the remaining nodes remain set to small random values.

6. Semi-supervised learning

Semi-supervised learning bridges the gap between the supervised and unsupervised settings of the previous two sections. It implies that labeled information is present for a, usually small, part of the dataset. The remaining elements have no associated outcome, but can be used to improve the generalization of the learner. To the best of our knowledge, only one specific integration of fuzzy rough set theory in a semi-supervised learning method has been proposed. In [102], the self-training approach is followed. This scheme involves the iterative retraining of a classifier. In a first stage, the initially labeled examples are used to train the model. This model is used to predict the labels of the unlabeled instances. Elements for which this prediction is most confident are assigned the estimated label. In the next phase, the enlarged labeled set is used to retrain the classifier. The process is repeated until a certain stopping criterion is satisfied. The proposal of [102] uses fuzzy rough approximation operators. In each iteration, the method computes the membership degree of the unlabeled instances to the fuzzy rough lower approximations of all decision classes, based on the labeled elements. If an unlabeled instance fully belongs to the lower approximation of some class, it is assigned the corresponding label. The process is repeated until no more unlabeled instances with a membership degree of 1 to a lower approximation are found. Any remaining unlabeled instances are labeled with the class for which the average of their membership degrees to the upper and lower approximation is maximal.

7. Future challenges

In this final section, we outline some important challenges of the use of fuzzy rough set theory within machine learning applications. By doing so, we aim to stimulate fellow researchers to consider these topics in their on-going efforts to extend and improve both the theoretical and applied frameworks.

7.1. Further extensions from rough to fuzzy rough sets

Dominance based rough sets [58, 59] have been successfully used in machine learning. They replace the equivalence relation of traditional rough sets (Section 2.1) with a dominance relation on the conditional attributes. A dominance relation is reflexive and transitive, but not symmetric. A traditional example is the 'better than'-relation. This theoretical concept has been applied to rough attribute selection [79, 152], rule induction [35] and classification [62, 93] among others. An extension of dominance based rough sets to dominance based fuzzy rough sets as in [63, 46] and an analogous extension of the corresponding rough application techniques is an inviting research path.

Substantial theoretical work has also been done in the framework of covering based rough sets (e.g. [154]). A crisp covering is a generalization of a partition: it consists of not-necessarily disjoint sets of which the union spans the entire universe. This relaxes the constraints posed by an equivalence relation, as it allows for a certain overlap between these sets. The definition of a fuzzy covering and the extension of covering based rough sets to covering based fuzzy rough sets has for instance been explored in [47, 78, 96, 145]. The extension of current fuzzy rough machine learning techniques to include fuzzy covering based approaches is certainly worth looking into.

Lastly, we note that extensive work in the literature focuses on theoretical aspects of type-2 fuzzy rough sets (e.g. [159]), intuitionistic fuzzy rough sets (e.g. [15, 25]) and interval-valued fuzzy rough sets (e.g. [134]). The theoretical developments have not been used to their full extent in applications and we find their presence in machine learning techniques somewhat lacking. These models warrant evaluation in real-world data instead of on toy-examples, as was done for instance for the attribute reduction algorithm proposed in [144] and their further practical exploration should be an important future research goal.

7.2. Multi-instance learning

Multi-instance learning [40] is a branch of supervised learning in which the nature of the dataset poses additional challenges to the learner. Traditionally, in single-instance learning, each element of a dataset is described by a number of attributes and a decision value. Every element can therefore be represented by a single vector, containing these values. In multi-instance learning, every element is a so-called bag, consisting of a number of individual instances. All instances in the bag are described by a given set of attributes, but the number of instances in each bag can be different. The bags themselves, not their instances, are associated with the decision values. An example of a bag is a set of different conformations of the same molecule [40]. A review of current multi-instance learners can be found in e.g. [1]. The challenge of multi-instance learning is to predict the decision value of a bag based on its instances and its similarity with other bags. So far, to the best of our knowledge, fuzzy rough set theory has not been used within this domain. Nevertheless, as the indiscernibility relation is an important aspect of fuzzy rough sets, its generalization to handle similarity between bags can result in an ideal tool to handle the challenges of this domain.

7.3. Fuzzy indiscernibility

As discussed in Sections 2.2 and 3, fuzzy rough set theory allows to assess a level of indiscernibility or similarity between elements, as measured by the indiscernibility relation. Currently, the definition of this relation is set up in a one-suits-all fashion, without regard to the data or application at hand. Apart from making a basic distinction between nominal and real-valued features, most approaches apply the same process of comparing data values for all features and all data sets. Nevertheless, different situations can warrant different similarity measures, depending on the importance of individual attributes, missing or noisy data, structured data types and so on. An interesting future research direction is therefore the automatic construction of appropriate similarity metrics, based on the considered dataset, application and learner.

7.4. Big data

Big data refers to problems whose size and complexity render standard machine learning algorithms unable to adequately deal with them [165]. It is a hot topic within the machine learning community, as such problems are encountered in many application fields such as bioinformatics, marketing, medicine and so on. Distributed machine learning techniques are required to handle their inherent challenges. It has been noted in the literature that the rough approximation operators do not scale well to big data (e.g. [158]). Their fuzzy rough extensions experience similar (and more complex) issues, as they require the construction of a fuzzy similarity matrix. Recently, in [4], a first distributed approach to calculate fuzzy rough lower and upper approximations was presented.

8. Conclusion

Fuzzy rough set theory offers the flexibility to deal with two types of uncertainty present in information. It incorporates fuzzy set theory, which considers vagueness, within the rough set framework, handling incomplete information. A variety of models for this integration have been proposed in the literature. In this paper, we have presented a review of integrations of fuzzy rough set theory in machine learning applications. We have observed that it has been widely used and encounter its presence in many subdomains. A large effort has in particular been put in the development of attribute selection algorithms, which reduce the attribute set of decision systems. Fuzzy rough set theory guides the search for optimal reduced attribute sets by offering measures to evaluate the strength of candidate sets. We further reviewed its use in instance selection methods, a variety of classification methods as well as regression algorithms, neural networks, clustering, self organizing maps and semi-supervised learning. We conclude that the mathematical theory of fuzzy rough sets has certainly proven its worth for the machine learning domain, although several important challenges remain relatively unexplored. Both the theoretical and applied sides keep holding promise for further development and we see new proposals arising every day.

Acknowledgments

The research of Sarah Vluymans and Lynn D'eer is funded by the Special Research Fund (BOF) of Ghent University. This work was partially supported by the Spanish Ministry of Science and Technology under the project TIN2011-28488 and the Andalusian Research Plans P11-TIC-7765 and P10-TIC-6858, and by project PYR-2014-8 of the Genil Program of CEI BioTic GRANADA.

References

[1] Amores, J.: Multiple instance classification: Review, taxonomy and comparative study, *Artificial Intelligence*, **201**, 2013, 81–105.

[2] An, S., Hu, Q.: Fuzzy Rough Decision Trees, *Proceedings of the 8th International Conference RSCTC 2012: Rough Sets and Current Trends in Computing*, Springer, 2012.

[3] An, S., Shi, H., Hu, Q., Li, X., Dang, J.: Fuzzy rough regression with application to wind speed prediction, *Information Sciences*, **282**, 2014, 388–400.

[4] Asfoor, H., Srinivasan, R., Vasudevan, G., Verbiest, N., Cornells, C., Tolentino, M., Teredesai, A., De Cock, M.: Computing fuzzy rough approximations in large scale information systems, *Big Data (Big Data), 2014 IEEE International Conference on*, IEEE, 2014.

[5] Bargiela, A., Pedrycz, W.: *Granular computing: an introduction*, Springer Science & Business Media, 2003.

[6] Bezdek, J.: *Pattern recognition with fuzzy objective function algorithms*, Kluwer Academic Publishers, 1981.

[7] Bezdek, J., Ehrlich, R., Full, W.: FCM: The fuzzy c-means clustering algorithm, *Computers & Geosciences*, **10**(2), 1984, 191–203.

[8] Bhatt, R., Gopal, M.: Improved feature selection algorithm with fuzzy-rough sets on compact computational domain, *International Journal of General Systems*, **34**(4), 2005, 485–505.

[9] Bhatt, R., Gopal, M.: On fuzzy-rough sets approach to feature selection, *Pattern recognition letters*, **26**(7), 2005, 965–975.

[10] Bhatt, R., Gopal, M.: On the compact computational domain of fuzzy-rough sets, *Pattern Recognition Letters*, **26**(11), 2005, 1632–1640.

[11] Bhatt, R., Gopal, M.: FRCT: fuzzy-rough classification trees, *Pattern analysis and applications*, **11**(1), 2008, 73–88.

[12] Bian, H., Mazlack, L.: Fuzzy-rough nearest-neighbor classification approach, *22nd International Conference of the North American Fuzzy Information Processing Society, 2003. NAFIPS 2003.*, IEEE, 2003.

[13] Bonabeau, E., Dorigo, M., Theraulaz, G.: *Swarm intelligence: from natural to artificial systems*, vol. 1, Oxford university press, 1999.

[14] del Cerro, L. F., Prade, H.: Rough sets, twofold fuzzy sets and modal logic – Fuzziness in indiscernibility and partial information, *The Mathematics of Fuzzy Systems* (A. D. Nola, A. Ventre, Eds.), Verlag TUV, Rheinland, Köln, 1986.

[15] Chakrabarty, K., Gedeon, T., Koczy, L.: Intuitionistic fuzzy rough set, *Proceedings of 4th joint conference on information sciences (JCIS), Durham, NC*, 1998.

[16] Chawla, N., Bowyer, K., Hall, L., Kegelmeyer, W.: SMOTE: synthetic minority over-sampling technique, *arXiv preprint arXiv:1106.1813*, 2011.

[17] Chen, D., He, Q., Wang, X.: FRSVMs: Fuzzy rough set based support vector machines, *Fuzzy Sets and Systems*, **161**(4), 2010, 596–607.

[18] Chen, D., Hu, Q., Yang, Y.: Parameterized attribute reduction with Gaussian kernel based fuzzy rough sets, *Information Sciences*, **181**(23), 2011, 5169–5179.

[19] Chen, D., Kwong, S., He, Q., Wang, H.: Geometrical interpretation and applications of membership functions with fuzzy rough sets, *Fuzzy Sets and Systems*, **193**, 2012, 122–135.

[20] Chen, D., Tsang, E., Zhao, S.: An approach of attributes reduction based on fuzzy T_L rough sets, *IEEE International Conference on Systems, Man and Cybernetics, 2007. ISIC.*, IEEE, 2007.

[21] Chen, D., Wang, X., Zhao, S.: Attribute reduction based on fuzzy rough sets, in: *Proceedings of the International Conference RSEISP 2007: Rough Sets and Intelligent Systems Paradigms*, Springer, 2007, 381–390.

[22] Chen, D., Zhang, L., Zhao, S., Hu, Q., Zhu, P.: A novel algorithm for finding reducts with fuzzy rough sets, *IEEE Transactions on Fuzzy Systems*, **20**(2), 2012, 385–389.

[23] Chen, D., Zhao, S.: Local reduction of decision system with fuzzy rough sets, *Fuzzy Sets and Systems*, **161**(13), 2010, 1871–1883.

[24] Cornelis, C., Cock, M. D., Radzikowska, A.: Vaguely Quantified Rough Sets, *Proceedings of 11th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing (RSFDGrC2007)*, 2007.

[25] Cornelis, C., De Cock, M., Kerre, E.: Intuitionistic fuzzy rough sets: at the crossroads of imperfect knowledge, *Expert Systems*, **20**(5), 2003, 260–270.

[26] Cornelis, C., Jensen, R.: A noise-tolerant approach to fuzzy-rough feature selection, *IEEE International Conference on Fuzzy Systems, 2008. FUZZ-IEEE 2008.(IEEE World Congress on Computational Intelligence).*, IEEE, 2008.

[27] Cornelis, C., Jensen, R., Hurtado, G., Slezak, D.: Attribute selection with fuzzy decision reducts, *Information Sciences*, **180**(16), 2010, 3006–3020.

[28] Cornelis, C., Verbiest, N., Jensen, R.: Ordered Weighted Average Based Fuzzy Rough Sets, *Proceedings of the 5th International Conference on Rough Sets and Knowledge Technology (RSKT2010)*, 2010.

[29] Cortes, C., Vapnik, V.: Support-vector networks, *Machine learning*, **20**(3), 1995, 273–297.

[30] Cover, T., Hart, P.: Nearest neighbor pattern classification, *IEEE Transactions on Information Theory*, **13**(1), 1967, 21–27.

[31] Dai, J., Xu, Q.: Attribute selection based on information gain ratio in fuzzy rough set theory with application to tumor classification, *Applied Soft Computing*, **13**(1), 2013, 211–221.

[32] D'eer, L., Verbiest, N.: Improving the Beta-Precision and OWA Based Fuzzy Rough Set Models: Definitions, Properties and Robustness Analysis, *Proceedings of the 9th International Conference on Rough Sets and Current Trends in Soft Computing (RSCTC 2014), LNAI 8536*, 2014.

[33] D'eer, L., Verbiest, N., Cornelis, C., Godo, L.: Implicator-conjunctor based models of fuzzy rough sets: definitions and properties, *Proceedings of the 14th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing (RSFDGrC 2013), LNAI 8170*, 2013.

[34] D'eer, L., Verbiest, N., Cornelis, C., Godo, L.: A Comprehensive Study of Implicator-Conjunctor-Based and Noise-Tolerant Fuzzy Rough Sets: Definitions, Properties and Robustness Analysis, *Fuzzy Sets and Systems*, 2015, In press.

[35] Dembczyński, K., Pindur, R., Susmaga, R.: Generation of exhaustive set of rules within dominance-based rough set approach, *Electronic Notes in Theoretical Computer Science*, **82**(4), 2003, 96–107.

[36] Derrac, J., Cornelis, C., García, S., Herrera, F.: Enhancing evolutionary instance selection algorithms by means of fuzzy rough set based feature selection, *Information Sciences*, **186**(1), 2012, 73–92.

[37] Derrac, J., Verbiest, N., García, S., Cornelis, C., Herrera, F.: On the use of evolutionary feature selection for improving fuzzy rough set based prototype selection, *Soft Computing*, **17**(2), 2013, 223–238.

[38] Deza, M. M., Deza, E.: *Encyclopedia of distances*, Springer, 2009.

[39] Diao, R., Shen, Q.: Feature selection with harmony search, *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, **42**(6), 2012, 1509–1523.

[40] Dietterich, T., Lathrop, R., Lozano-Pérez, T.: Solving the multiple instance problem with axis-parallel rectangles, *Artificial intelligence*, **89**(1), 1997, 31–71.

[41] Du, Y., Hu, Q., Chen, D., Ma, P.: Kernelized Fuzzy Rough Sets Based Yawn Detection for Driver Fatigue Monitoring, *Fundamenta Informaticae*, **111**(1), 2011, 65–79.

[42] Dubois, D., Prade, H.: Twofold fuzzy sets and rough sets – Some issues in knowledge representation, *Fuzzy Sets and Systems*, **23**, 1987, 3–18.

[43] Dubois, D., Prade, H.: Rough fuzzy sets and fuzzy rough sets, *International Journal of General Systems*, **17**, 1990, 191–209.

[44] Dubois, D., Prade, H.: Putting fuzzy sets and rough sets together, in: *Intelligent Decision Support - Handbook of Applications and Advances of the Rough Sets Theory* (R. Słowiński, Ed.), Kluwer Academic Publishers, Boston, 1992, 203–232.

[45] Elashiri, M., Hefny, H., Elwhab, A.: Construct Fuzzy Decision Trees Based on Roughness Measures, in: *Advances in Communication, Network, and Computing*, Springer, 2012, 199–207.

[46] Fan, T., Liau, C., Liu, D.: Dominance-based fuzzy rough set analysis of uncertain and possibilistic data tables, *International Journal of Approximate Reasoning*, **52**(9), 2011, 1283–1297.

[47] Feng, T., Zhang, S., Mi, J.: The Reduction and Fusion of Fuzzy Covering Systems based on the Evidence Theory, *International Journal of Approximate Reasoning*, **53**, 2012, 87–103.

[48] Fernández Salido, J. M., Murakami, S.: Rough set analysis of a general type of fuzzy data using transitive aggregations of fuzzy similarity relations, *Fuzzy Sets and Systems*, **139**, 2003, 635–660.

[49] Flach, P.: *Machine learning: the art and science of algorithms that make sense of data*, Cambridge University Press, 2012.

[50] Ganivada, A., Dutta, S., Pal, S.: Fuzzy rough granular neural networks, fuzzy granules, and classification, *Theoretical Computer Science*, **412**(42), 2011, 5834–5853.

[51] Ganivada, A., Pal, S.: A novel fuzzy rough granular neural network for classification, *International Journal of Computational Intelligence Systems*, **4**(5), 2011, 1042–1051.

[52] Ganivada, A., Ray, S., Pal, S.: Fuzzy rough granular self organizing map, in: *Rough Sets and Knowledge Technology*, Springer, 2011, 659–668.

[53] Ganivada, A., Ray, S., Pal, S.: Fuzzy rough granular self-organizing map and fuzzy rough entropy, *Theoretical Computer Science*, **466**, 2012, 37–63.

[54] Ganivada, A., Ray, S., Pal, S.: Fuzzy rough sets, and a granular neural network for unsupervised feature selection, *Neural Networks*, **48**, 2013, 91–108.

[55] Ganter, B., Wille, R.: *Formal concept analysis: mathematical foundations*, Springer Science & Business Media, 2012.

[56] Geem, Z., Kim, J., Loganathan, G.: A new heuristic optimization algorithm: harmony search, *Simulation*, **76**(2), 2001, 60–68.

[57] Greco, S., Inuiguchi, M., Slowinski, R.: Fuzzy rough sets and multiple-premise gradual decision rules, in: *Fuzzy Logic and Applications*, Springer, 2006, 148–163.

[58] Greco, S., Matarazzo, B., Słowinński, R.: The use of rough sets and fuzzy sets in MCDM, in: *Multicriteria Decision Making*, Springer, 1999, 397–455.

[59] Greco, S., Matarazzo, B., Słowinński, R.: Decision rule approach, in: *Multiple criteria decision analysis: state of the art surveys*, Springer, 2005, 507–555.

[60] Greco, S., Matarazzo, B., Słowinski, R.: Fuzzy similarity relation as a basis for rough approximations, in: *Rough Sets and Current Trends in Computing, Lecture Notes in Artificial Intelligence* (L. Polkowski, A. Skowron, Eds.), vol. 1442, Springer, Berlin, 1998, 283–289.

[61] Greco, S., Matarazzo, B., Słowinski, R.: Rough set processing of vague information using fuzzy similarity relations, in: *Finite Versus Infinite—Contributions to an Eternal Dilemma* (C. Calude, G. Paun, Eds.), Springer-Verlag, 2000, 149–173.

[62] Greco, S., Matarazzo, B., Słowiński, R.: Multicriteria classification by dominance-based rough set approach, *Handbook of data mining and knowledge discovery. Oxford University Press, New York*, 2002.

[63] Greco, S., Matarazzo, B., Słowiński, R.: Fuzzy set extensions of the dominance-based rough set approach, in: *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models*, Springer, 2008, 239–261.

[64] Guyon, I., Elisseeff, A.: An introduction to variable and feature selection, *The Journal of Machine Learning Research*, **3**, 2003, 1157–1182.

[65] Haykin, S.: *Neural networks: A comprehensive foundation*, Prentice Hall International, 2004.

[66] He, Q., Wu, C.: Membership evaluation and feature selection for fuzzy support vector machine based on fuzzy rough sets, *Soft Computing*, **15**(6), 2011, 1105–1114.

[67] He, Q., Wu, C., Chen, D., Zhao, S.: Fuzzy rough set based attribute reduction for information systems with fuzzy decisions, *Knowledge-Based Systems*, **24**(5), 2011, 689–696.

[68] Hong, T., Liou, Y., Wang, S.: Fuzzy rough sets with hierarchical quantitative attributes, *Expert systems with Applications*, **36**(3), 2009, 6790–6799.

[69] Hu, Q., An, S., Yu, D.: Soft fuzzy rough sets for robust feature evaluation and selection, *Information Sciences*, **180**(22), 2010, 4384–4400.

[70] Hu, Q., An, S., Yu, X., Yu, D.: Robust fuzzy rough classifiers, *Fuzzy Sets and Systems*, **183**, 2011, 26–43.

[71] Hu, Q., Xie, Z., Yu, D.: Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation, *Pattern recognition*, **40**(12), 2007, 3509–3521.

[72] Hu, Q., Yu, D., Pedrycz, W., Chen, D.: Kernelized fuzzy rough sets and their applications, *IEEE Transactions on Knowledge and Data Engineering*, **23**(11), 2011, 1649–1667.

[73] Hu, Q., Yu, D., Xie, Z.: Information-preserving hybrid data reduction based on fuzzy-rough techniques, *Pattern recognition letters*, **27**(5), 2006, 414–423.

[74] Hu, Q., Yu, D., Xie, Z., Liu, J.: Fuzzy probabilistic approximation spaces and their information measures, *IEEE Transactions on Fuzzy Systems*, **14**(2), 2006, 191–201.

[75] Hu, Q., Zhang, L., An, S., Zhang, D., Yu, D.: On robust fuzzy rough set models, *IEEE Transactions on Fuzzy Systems*, **20**, 2012, 636–651.

[76] Hu, Q., Zhang, L., Chen, D., Pedrycz, W., Yu, D.: Gaussian kernel based fuzzy rough sets: model, uncertainty measures and applications, *International Journal of Approximate Reasoning*, **51**, 2010, 453–471.

[77] Inuiguchi, M.: Two generalizations of rough sets and their fundamental properties, *Proceedings of 6th Workshop on Uncertainty Processing*, 2003.

[78] Inuiguchi, M., Tanino, T.: New fuzzy rough sets based on certainty qualification, *Rough-Neural Computing* (K. P. et al., Ed.), Springer, Berlin-Heidelberg, 2003.

[79] Inuiguchi, M., Yoshioka, Y., Kusunoki, Y.: Variable-precision dominance-based rough set approach and attribute reduction, *International Journal of Approximate Reasoning*, **50**(8), 2009, 1199–1214.

[80] Jensen, R., Cornelis, C.: A New Approach to Fuzzy-Rough Nearest Neighbour Classification, *6th International Conference on Rough Sets and Current Trends in Computing (RSCTC 2008)*, 2008.

[81] Jensen, R., Cornelis, C.: Fuzzy-rough instance selection, *2010 IEEE International Conference on Fuzzy Systems (FUZZ)*, IEEE, 2010.

[82] Jensen, R., Cornelis, C.: Fuzzy-rough nearest neighbour classification and prediction, *Theoretical Computer Science*, **412**(42), 2011, 5871–5884.

[83] Jensen, R., Cornelis, C., Shen, Q.: Hybrid fuzzy-rough rule induction and feature selection, *IEEE International Conference on Fuzzy Systems, 2009. FUZZ-IEEE 2009.*, IEEE, 2009.

[84] Jensen, R., Mac Parthaláin, N.: Nearest Neighbour-Based Fuzzy-Rough Feature Selection, *Rough Sets and Current Trends in Computing*, Springer, 2014.

[85] Jensen, R., Shen, Q.: Fuzzy–rough attribute reduction with application to web categorization, *Fuzzy sets and systems*, **141**(3), 2004, 469–485.

[86] Jensen, R., Shen, Q.: Semantics-preserving dimensionality reduction: rough and fuzzy-rough-based approaches, *IEEE Transactions on Knowledge and Data Engineering*, **16**(12), 2004, 1457–1471.

[87] Jensen, R., Shen, Q.: Fuzzy-rough data reduction with ant colony optimization, *Fuzzy Sets and Systems*, **149**(1), 2005, 5–20.

[88] Jensen, R., Shen, Q.: Fuzzy-rough feature significance for fuzzy decision trees, *Proc. 2005 UK Workshop on Computational Intelligence*, Citeseer, 2005.

[89] Jensen, R., Shen, Q.: Fuzzy-rough sets assisted attribute selection, *IEEE Transactions on Fuzzy Systems*, **15**(1), 2007, 73–89.

[90] Jensen, R., Shen, Q.: Are more features better? a response to attributes reduction using fuzzy rough sets, *IEEE Transactions on Fuzzy Systems*, **17**(6), 2009, 1456–1458.

[91] Jensen, R., Shen, Q.: New approaches to fuzzy-rough feature selection, *IEEE Transactions on Fuzzy Systems*, **17**(4), 2009, 824–838.

[92] Keller, J., Gray, M., Givens, J.: A fuzzy k-nearest neighbor algorithm, *IEEE Transactions on Systems, Man and Cybernetics*, **15**(4), 1985, 580–585.

[93] Kotłowski, W., Dembczyński, K., Greco, S., Słowiński, R.: Stochastic dominance-based rough set model for ordinal classification, *Information Sciences*, **178**(21), 2008, 4019–4037.

[94] Kuncheva, L.: Fuzzy rough sets: application to feature selection, *Fuzzy sets and Systems*, **51**(2), 1992, 147–153.

[95] Li, F., Min, F., Liu, Q.: Intra-cluster similarity index based on fuzzy rough sets for fuzzy c-means algorithm, in: *Rough Sets and Knowledge Technology*, Springer, 2008, 316–323.

[96] Li, T., Leung, Y., Zhang, W.: Generalized fuzzy rough approximation operators based on fuzzy coverings, *International Journal of Approximate Reasoning*, **48**, 2008, 836–856.

[97] Lin, T. Y., Yao, Y. Y., Zadeh, L. A., Eds.: *Data Mining, Rough Sets and Granular Computing*, vol. 95, Springer, 2002.

[98] Lingras, P. J., Yao, Y. Y.: Data mining using extensions of the rough set model, *Journal of the Association for Information Science and Technology*, **49**, 1998, 415–422.

[99] Liu, G.: Axiomatic systems for rough sets and fuzzy rough sets, *International Journal of Approximate Reasoning*, **48**, 2008, 857–867.

[100] Liu, Y., Zhou, Q., Rakus-Andersson, E., Bai, G.: A fuzzy-rough sets based compact rule induction method for classifying hybrid data, in: *Rough Sets and Knowledge Technology*, Springer, 2012, 63–70.

[101] Lotfabadi, M., Shiratuddin, M., Wong, K.: Using fuzzy rough feature selection for image retrieval system, *2013 IEEE Symposium on Computational Intelligence for Multimedia, Signal and Vision Processing (CIMSIVP)*, IEEE, 2013.

[102] Mac Parthaláin, N., Jensen, R.: Fuzzy-rough set based semi-supervised learning, *2011 IEEE International Conference on Fuzzy Systems (FUZZ)*, IEEE, 2011.

[103] Mac Parthaláin, N., Jensen, R.: Simultaneous feature and instance selection using fuzzy-rough bireducts, *2013 IEEE International Conference on Fuzzy Systems (FUZZ)*, IEEE, 2013.

[104] Mac Parthaláin, N., Jensen, R., Shen, Q., Zwiggelaar, R.: Fuzzy-rough approaches for mammographic risk analysis, *Intelligent Data Analysis*, **14**(2), 2010, 225–244.

[105] Maji, P.: Fuzzy–rough supervised attribute clustering algorithm and classification of microarray data, *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, **41**(1), 2011, 222–233.

[106] Maji, P., Garai, P.: Simultaneous Feature Selection and Extraction Using Fuzzy Rough Sets, *Proceedings of the Second International Conference on Soft Computing for Problem Solving (SocProS 2012), December 28-30, 2012*, Springer, 2014.

[107] Mi, J., Leung, Y., Zhao, H., Feng, T.: Generalized fuzzy rough sets determined by a triangular norm, *Information Sciences*, **178**, 2008, 3203–3213.

[108] Mi, J., Zhang, W.: An axiomatic characterization of a fuzzy generalization of rough sets, *Information Sciences*, **160**, 2004, 235–249.

[109] Mieszkowicz-Rolka, A., Rolka, L.: Variable precision fuzzy rough sets, in: *Transactions on Rough Sets I* (J. P. et al., Ed.), Springer-Verlag, 2004, 144–160.

[110] Mieszkowicz-Rolka, A., Rolka, L.: Fuzzy rough approximations of process data, *Internation Journal of Approximate Reasoning*, **49**, 2008, 301–315.

[111] Morsi, N., Yakout, M.: Axiomatics for fuzzy rough sets, *Fuzzy Sets and Systems*, **100**, 1998, 327–342.

[112] Nakamura, A.: Fuzzy rough sets, *Note on Multiple-Valued Logic in Japan*, **9**, 1988, 1–8.

[113] Nguyen, H. S.: Approximate boolean reasoning: foundations and applications in data mining, *Transactions on rough sets V*, Springer-Verlag, 2006.

[114] Pawlak, Z.: Rough sets, *International Journal of Computer & Information Sciences*, **11**(5), 1982, 341–356.

[115] Pawlak, Z., Skowron, A.: Rough sets: some extensions, *Information Sciences*, **177**, 2007, 28–40.

[116] Pawlak, Z., Skowron, A.: Rudiments of rough sets, *Information Sciences*, **177**, 2007, 3–27.

[117] Pearson, K.: LIII. On lines and planes of closest fit to systems of points in space, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, **2**(11), 1901, 559–572.

[118] Pei, D.: A generalized model of fuzzy rough sets, *International Journal of General Systems*, **34**(5), 2005, 603–613.

[119] Prasad, P., Rao, C.: IQuickReduct: an improvement to Quick Reduct algorithm, in: *Rough Sets, Fuzzy Sets, Data Mining and Granular Computing*, Springer, 2009, 152–159.

[120] Prasad, P., Rao, C.: Extensions to iQuickReduct, in: *Multi-disciplinary Trends in Artificial Intelligence*, Springer, 2011, 351–362.

[121] Qian, Y., Li, C., Liang, J.: An efficient fuzzy-rough attribute reduction approach, in: *Rough Sets and Knowledge Technology*, Springer, 2011, 63–70.

[122] Qian, Y., Wang, Q., Cheng, H., Liang, J., Dang, C.: Fuzzy-rough feature selection accelerator, *Fuzzy Sets and Systems*, 2014.

[123] Qu, Y., Shang, C., Shen, Q., Mac Parthaláin, N., Wu, W.: Kernel-based fuzzy-rough nearest neighbour classification, *Fuzzy Systems (FUZZ), 2011 IEEE International Conference on*, IEEE, 2011.

[124] Radzikowska, A., Kerre, E.: A Comparative Study of Fuzzy Rough Sets, *Fuzzy Sets and Systems*, **126**, 2002, 137–155.

[125] 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, *Knowledge and information systems*, **33**(2), 2012, 245–265.

[126] Ramentol, E., Verbiest, N., Bello, R., Caballero, Y., Cornelis, C., Herrera, F.: SMOTE-FRST: a new resampling method using fuzzy rough set theory, *10th International FLINS conference on uncertainty modelling in knowledge engineering and decision making*, 2012.

[127] Ramentol, E., Vluymans, S., Verbiest, N., Caballero, Y., Bello Perez, R., Cornelis, C., Herrera, F.: IFROWANN: Imbalanced Fuzzy-Rough Ordered Weighted Average Nearest Neighbor Classification, *IEEE Transactions of Fuzzy Systems*, 2014, In press.

[128] Sarkar, M.: Fuzzy-rough nearest neighbor algorithms in classification, *Fuzzy Sets and Systems*, **158**(19), 2007, 2134–2152.

[129] Sarkar, M., Yegnanarayana, B.: Application of fuzzy-rough sets in modular neural networks, *The 1998 IEEE International Joint Conference on Neural Networks Proceedings, 1998. IEEE World Congress on Computational Intelligence.*, 1, IEEE, 1998.

[130] Sarkar, M., Yegnanarayana, B.: Fuzzy-rough neural networks for vowel classification, *IEEE International Conference on Systems, Man, and Cybernetics, 1998.*, 5, IEEE, 1998.

[131] Shen, Q., Jensen, R.: Selecting informative features with fuzzy-rough sets and its application for complex systems monitoring, *Pattern recognition*, **37**(7), 2004, 1351–1363.

[132] Skowron, A., Rauszer, C.: The discernibility matrices and functions in information systems, in: *Intelligent Decision Support*, Springer, 1992, 331–362.

[133] Stefanowski, J., Wilk, S.: Selective pre-processing of imbalanced data for improving classification performance, in: *Data Warehousing and Knowledge Discovery*, Springer, 2008, 283–292.

[134] Sun, B., Gong, Z., Chen, D.: Fuzzy rough set theory for the interval-valued fuzzy information systems, *Information Sciences*, **178**(13), 2008, 2794–2815.

[135] Sun, Y., Wong, A., Kamel, M.: Classification of imbalanced data: A review, *International Journal of Pattern Recognition and Artificial Intelligence*, **23**(04), 2009, 687–719.

[136] Swiniarski, R. W., Skowron, A.: Rough set methods in feature selection and recognition, *Pattern Recognition Letters*, **24**, 2003, 833–849.

[137] Tsang, E., Chen, D., Yeung, D., Wang, X., Lee, J.: Attributes reduction using fuzzy rough sets, *IEEE Transactions on Fuzzy Systems*, **16**(5), 2008, 1130–1141.

[138] Tsang, E., Zhao, S., Lee, J.: Rule induction based on fuzzy rough sets, *2007 International Conference on Machine Learning and Cybernetics*, 5, IEEE, 2007.

[139] Verbiest, N., Cornelis, C., Herrera, F.: FRPS: A fuzzy rough prototype selection method, *Pattern Recognition*, **46**(10), 2013, 2770–2782.

[140] Verbiest, N., Cornelis, C., Herrera, F.: OWA-FRPS: A Prototype Selection Method Based on Ordered Weighted Average Fuzzy Rough Set Theory, in: *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, Springer, 2013, 180–190.

[141] Verbiest, N., Cornelis, C., Jensen, R.: Fuzzy rough positive region based nearest neighbour classification, *2012 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, IEEE, 2012.

[142] Verbiest, N., Ramentol, E., Cornelis, C., Herrera, F.: Improving SMOTE with Fuzzy Rough Prototype Selection to Detect Noise in Imbalanced Classification Data, in: *Advances in Artificial Intelligence–IBERAMIA 2012*, Springer, 2012, 169–178.

[143] Wang, X., Tsang, E., Zhao, S., Chen, D., Yeung, D.: Learning fuzzy rules from fuzzy samples based on rough set technique, *Information sciences*, **177**(20), 2007, 4493–4514.

[144] Wu, H., Wu, Y., Luo, J.: An interval type-2 fuzzy rough set model for attribute reduction, *IEEE Transactions on Fuzzy Systems*, **17**(2), 2009, 301–315.

[145] Wu, M., Han, H., Si, Y.: Properties and Axiomatization of Fuzzy Rough Sets based on Fuzzy Coverings, *Proceedings of the 2012 International Conference on Machine Learning and Cybernetics*, 2012.

[146] Wu, W., Leung, Y., Mi, J.: On characterizations of (I, T)-fuzzy rough approximation operators, *Fuzzy Sets and Systems*, **154**, 2005, 76–102.

[147] Wu, W., Mi, J., Zhang, W.: Generalized fuzzy rough sets, *Information Sciences*, **151**, 2003, 263–282.

[148] Wu, W., Zhang, W.: Constructive and axiomatic approaches of fuzzy approximation operators, *Information Sciences*, **159**, 2004, 233–254.

[149] Wygralak, M.: Rough sets and fuzzy sets – some remarks on interrelations, *Fuzzy Sets and Systems*, **29**(2), 1989, 241–243.

[150] Xu, F., Miao, D., Wei, L.: Fuzzy-rough attribute reduction via mutual information with an application to cancer classification, *Computers & Mathematics with Applications*, **57**(6), 2009, 1010–1017.

[151] Yang, M., Chen, S., Yang, X.: A Novel Approach of Rough Set-Based Attribute Reduction Using Fuzzy Discernibility Matrix., *Proceedings of the 4th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2007*, IEEE, 2007.

[152] Yang, X., Yang, J., Wu, C., Yu, D.: Dominance-based rough set approach and knowledge reductions in incomplete ordered information system, *Information Sciences*, **178**(4), 2008, 1219–1234.

[153] Yao, Y., Mi, J., Li, Z.: A novel variable precision (θ, σ)-fuzzy rough set model based on fuzzy granules, *Fuzzy Sets and Systems*, **236**, 2014, 58–72.

[154] Yao, Y., Yao, B.: Covering based Rough Sets Approximations, *Information Sciences*, **200**, 2012, 91–107.

[155] Yeung, D., Chen, D., Tsang, E., Lee, J., Xizhao, W.: On the generalization of fuzzy rough sets, *IEEE Transactions on Fuzzy Systems*, **13**(3), 2005, 343–361.

[156] Zadeh, L.: Fuzzy sets, *Information and control*, **8**(3), 1965, 338–353.

[157] Zhai, J.: Fuzzy decision tree based on fuzzy-rough technique, *Soft Computing*, **15**(6), 2011, 1087–1096.

[158] Zhang, J., Li, T., Ruan, D., G., Z., Zhao, C.: A parallel method for computing rough set approximations, *Information Sciences*, **194**, 2012, 209–223.

[159] Zhang, Z.: On characterization of generalized interval type-2 fuzzy rough sets, *Information Sciences*, **219**, 2013, 124–150.

[160] Zhao, J., Zhang, Z.: Fuzzy rough neural network and its application to feature selection, *2011 Fourth International Workshop on Advanced Computational Intelligence (IWACI)*, IEEE, 2011.

[161] Zhao, S., Tsang, E.: On fuzzy approximation operators in attribute reduction with fuzzy rough sets, *Information Sciences*, **178**(16), 2008, 3163–3176.

[162] Zhao, S., Tsang, E., Chen, D., Wang, X.: Building a rule-based classifier – a fuzzy-rough set approach, *IEEE Transactions on Knowledge and Data Engineering*, **22**(5), 2010, 624–638.

[163] Zhao, S., Tsang, E. C. C., Chen, D.: The model of fuzzy variable precision rough sets, *IEEE Transactions on Fuzzy Systems*, **17**(2), 2009, 451–467.

[164] Ziarko, W.: Variable Precision Rough Set Model, *Journal of Computer and System Sciences*, **46**, 1993, 39–59.

[165] Zikopoulos, P., Eaton, C.: *Understanding big data: Analytics for enterprise class hadoop and streaming data*, McGraw-Hill Osborne Media, 2011.