A Comprehensive Study of Implicator-Conjunctor-Based and Noise-Tolerant Fuzzy Rough Sets: Definitions, Properties and Robustness Analysis

Lynn Deer[a,*], Nele Verbiest[a], Chris Cornelis[a,b], Lluís Godo[c]

[a]*Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Belgium.*
[b]*Department of Computer Science and Artificial Intelligence, Research Center on Information and Communications Technology (CITIC-UGR), University of Granada, Spain.*
[c]*Artificial Intelligence Research Institute (IIIA), Spanish National Research Council (CSIC), Bellaterra, Spain.*

Abstract

Both rough and fuzzy set theory offer interesting tools for dealing with imperfect data: while the former allows us to work with uncertain and incomplete information, the latter provides a formal setting for vague concepts. The two theories are highly compatible, and since the late 1980s many researchers have studied their hybridization. In this paper, we critically evaluate most relevant fuzzy rough set models proposed in the literature. To this end, we establish a formally correct and unified mathematical framework for them. Both implicator-conjunctor-based definitions and noise-tolerant models are studied. We evaluate these models on two different fronts: firstly, we discuss which properties of the original rough set model can be maintained and secondly, we examine how robust they are against both class and attribute noise. By highlighting the benefits and drawbacks of the different fuzzy rough set models, this study appears a necessary first step to propose and develop new models in future research.

Keywords: fuzzy sets, rough sets, hybridization, lower and upper approximation, noise-tolerance

1. Introduction

Rough set theory was originally proposed by Pawlak [47] in 1982 to deal with uncertainty due to incompleteness and indiscernibility. The basic idea of rough set theory is that it provides a lower and upper approximation of a concept with respect to a binary indiscernibility relation. The lower approximation contains all the elements of the universe certainly belonging to the concept, while the upper approximation contains the elements possibly belonging to the concept. In the original model of Pawlak, an equivalence relation is used to model indiscernibility. Yet, many authors have generalized Pawlak's model by using binary non-equivalence relations (see e.g. [48, 49] for a survey).

*Corresponding author
Email addresses: Lynn.Deer@UGent.be (Lynn Deer), Nele.Verbiest@UGent.be (Nele Verbiest), Chris.Cornelis@decsai.ugr.es (Chris Cornelis), godo@iiia.csic.es (Lluís Godo)

Applications of rough set theory are widespread and are especially prominent in data analysis [33, 35] and more specific in feature selection and classification [53]. However, since the traditional rough set is designed to process qualitative (discrete) data, it faces important limitations when dealing with real-valued data sets [31]. Fuzzy set theory proposed in 1965 by Zadeh [68] is very useful to overcome these limitations, as it can deal effectively with vague concepts and graded indiscernibility.

It was recognized early that both theories are complementary, rather than competitive. To that end, rough set theory has been extended in two ways [14]. Rough fuzzy set theory discusses the approximation of a fuzzy set by a crisp relation. If moreover the indiscernibility relation to distinguish different objects is fuzzy as well, fuzzy rough set theory is considered. Since every crisp relation can be seen as a special case of a fuzzy relation, all results obtained in fuzzy rough set theory also hold for rough fuzzy set theory.

The vestiges of fuzzy rough set theory date back to the late 1980s, and originate from work by Fariñas del Cerro and Prade [12], Dubois and Prade [13], Nakamura [45] and Wygralak [62]. From 1990 onwards, research on the hybridization between rough sets and fuzzy sets flourished. The inspiration to combine rough and fuzzy set theory was found in different mathematical fields. For instance, Lin [34] studied fuzzy rough sets using generalized topological spaces (Frechet spaces) and Nanda and Majumdar [46] discussed fuzzy rough sets based on an algebraic approach. Moreover, Thiele [54] examined the relationship with fuzzy modal logic. Later on, Yao [66] and Liu [39] used level sets to combine fuzzy and rough set theory.

This work focuses on fuzzy rough set models using fuzzy relations and fuzzy logical connectives. The seminal papers of Dubois and Prade [14, 15] are probably the most important in the evolution of these fuzzy rough set models, since they influenced numerous authors who used different fuzzy logical connectives and fuzzy relations. Essential work was done by Morsi and Yakout [44] who studied both constructive and axiomatic approaches and by Radzikowska and Kerre [51] who defined fuzzy rough sets based on three general classes of fuzzy implicators: S-, R- and QL-implicators. However, despite generalizing the fuzzy connectives, they still used fuzzy similarity relations. A first attempt to use reflexive fuzzy relations instead of fuzzy similarity relations was done by Greco et al. [22, 23]. Thereafter, Wu et al. [60, 61] were the first to consider general fuzzy relations. Besides generalizing the fuzzy relation, Mi et al. [40, 41] considered conjunctors instead of t-norms. Furthermore, Yeung et al. [67] discussed two pairs of dual approximation operators from both a constructive and an axiomatic point of view. Hu et al. [26, 28] for their part studied fuzzy relations based on kernel functions. In this work, we consider all these different proposals within a general Implicator-Conjunctor (IC) based fuzzy rough set model that encapsulates all of them, as discussed in Section 3.1.

However, the aforementioned models only consider the worst and best performing objects to determine the fuzzy rough lower and upper approximations respectively. Consequently, these approximations are sensitive to noisy and/or outlying samples. This, in turn, impacts the robustness of data analysis applications based on them, such as attribute selection and classification. To mitigate this problem in the crisp case, Ziarko [70] proposed the Variable Precision Rough Set (VPRS) model in 1993. This model also served as a starting point for the design of several noise-tolerant fuzzy rough set approaches, such as [5, 7, 17, 18, 24, 25, 42, 43, 65, 69], which will be discussed in detail in Section 4.

In this paper, we critically evaluate most relevant fuzzy rough set models proposed in the literature. To this end, we establish a formally correct and unified mathematical framework for them. A structured and critical analysis of the current research of constructive methods for fuzzy rough set models is presented. Note that we do not consider axiomatic approaches

(see e.g. [36, 37, 41, 44, 50, 59–61, 67]). We review the definitions of noise-tolerant models, generalizing them where appropriate and in some cases applying modifications to correct errors in the original proposal. Where applicable, we also establish relationships between these models and the corresponding IC based definitions, as well as Pawlak's and Ziarko's crisp approaches. Furthermore, we examine which theoretical properties of traditional rough sets and IC fuzzy rough sets can still be maintained for the noise-tolerant models; indeed, similarly as for Ziarko's VPRS model, providing mechanisms for making the approximations less strict usually involves sacrificing some desirable properties. Finally, we evaluate whether the considered approaches really live up to the claim of being more "robust" approximations, by performing a stability analysis on four real datasets, and comparing them to the IC model. This will allow us to obtain a comprehensive overview of the benefits and the drawbacks of the robust fuzzy rough set models, in order to acquire the expertise for future research opportunities.

The remainder of this article is structured as follows: in Section 2, we summarize preliminary definitions concerning fuzzy logical connectives, fuzzy sets and relations, and rough set theory. In Section 3, we introduce the general IC based fuzzy rough set model, and review specific cases of it studied in literature. We also establish sufficient conditions under which its approximations satisfy the same theoretical properties as Pawlak's model. In Section 4, we review the seven most important noise-tolerant fuzzy rough set models introduced in the literature: Mieszkowicz-Rolka and Rolka's Variable Precision Fuzzy Rough Set model (VPFRS [42, 43]), Cornelis et al.'s Vaguely Quantified Fuzzy Rough Set model (VQFRS, [5]), Hu et al.'s Soft Fuzzy Rough Set model (SFRS, [24, 25]), Yao et al.'s Variable Precision Fuzzy Rough Set model based on Fuzzy Granules (FG, [65]), Zhao et al.'s Fuzzy Variable Precision Rough Set model (FVPRS, [69]), Fernández-Salido and Murakami's β-Precision Fuzzy Rough Set model (β-PREC, [17, 18]) and Cornelis et al.'s Ordered Weighted Average based Fuzzy Rough Set model (OWA, [7]). In Section 5, we test the robustness of these proposals and the IC model in the presence of class and attribute noise, and in Section 6, we conclude and state future work. In the Appendix, we give counterexamples for the properties that are not maintained in the different noise-tolerant models. Finally, we mention that a small part of the results in this paper (pertaining only to the IC model) has been published in a conference paper [10].

2. Preliminaries

In this section, we present basic notions related to fuzzy logical connectives, fuzzy sets and relations, and the lower and upper approximation operators in rough set theory.

2.1. Fuzzy Logical Connectives

We start by recalling some important fuzzy logical connectives. First, we discuss conjunctors, disjunctors and negators that are fuzzy extensions of the Boolean conjunctor, disjunctor and negation.

- A *conjunctor* is a mapping $\mathscr{C}\colon [0,1]^2 \to [0,1]$ that is increasing in both arguments and that satisfies the boundary conditions $\mathscr{C}(0,0) = \mathscr{C}(0,1) = \mathscr{C}(1,0) = 0$ and $\mathscr{C}(1,1) = 1$. It is called a *border conjunctor* if for all $x \in [0,1]$ it holds that $\mathscr{C}(1,x) = x$. A commutative and associative border conjunctor is called a *t-norm* and is denoted by $\mathscr{T}$.

Example 1. *For $x,y \in [0,1]$, the minimum t-norm or minimum operator $\mathscr{T}_M$ is defined by $\mathscr{T}_M(x,y) = \min(x,y)$, the product t-norm $\mathscr{T}_P$ is defined by $\mathscr{T}_P(x,y) = x \cdot y$ and the Łukasiewicz t-norm $\mathscr{T}_L$ is defined by $\mathscr{T}_L(x,y) = \max(0, x+y-1)$.*

- A *disjunctor* is a mapping $\mathscr{D}\colon [0,1]^2 \to [0,1]$ that is increasing in both arguments and that satisfies the boundary conditions $\mathscr{D}(1,1) = \mathscr{D}(0,1) = \mathscr{D}(1,0) = 1$ and $\mathscr{D}(0,0) = 0$. It is called a *border disjunctor* if for all $x \in [0,1]$ it holds that $\mathscr{D}(0,x) = x$. A commutative and associative border disjunctor is called a *t-conorm* and is denoted by $\mathscr{S}$.

  Example 2. *For $x,y \in [0,1]$, the maximum t-conorm or maximum operator $\mathscr{S}_M$ is defined by $\mathscr{S}_M(x,y) = \max(x,y)$, the probabilistic sum $\mathscr{S}_P$ is defined by $\mathscr{S}_P(x,y) = x+y-x\cdot y$ and the Łukasiewicz t-conorm $\mathscr{S}_L$ is defined by $\mathscr{S}_L(x,y) = \min(1, x+y)$.*

- A *negator* is a decreasing mapping $\mathscr{N}\colon [0,1] \to [0,1]$ that satisfies $\mathscr{N}(0) = 1$ and $\mathscr{N}(1) = 0$. A negator is called *involutive* if for all $x \in [0,1]$ it holds that $\mathscr{N}(\mathscr{N}(x)) = x$.

  Example 3. *For $x \in [0,1]$, the standard negator $\mathscr{N}_S$ is defined by $\mathscr{N}_S(x) = 1 - x$.*

When we use an involutive negator, we can define a dual couple of a conjunctor and a disjunctor with respect to the chosen negator. Given an involutive negator $\mathscr{N}$, a conjunctor $\mathscr{C}$ and a disjunctor $\mathscr{D}$, the *$\mathscr{N}$-dual of $\mathscr{C}$* is a disjunctor $\mathscr{D}_{\mathscr{C},\mathscr{N}}$ defined by

$$\forall x,y \in [0,1] : \mathscr{D}_{\mathscr{C},\mathscr{N}}(x,y) = \mathscr{N}(\mathscr{C}(\mathscr{N}(x), \mathscr{N}(y)))$$

and the *$\mathscr{N}$-dual of $\mathscr{D}$* is a conjunctor $\mathscr{C}_{\mathscr{D},\mathscr{N}}$ defined by

$$\forall x,y \in [0,1] : \mathscr{C}_{\mathscr{D},\mathscr{N}}(x,y) = \mathscr{N}(\mathscr{D}(\mathscr{N}(x), \mathscr{N}(y))).$$

It can be verified that the $\mathscr{N}$-dual of a t-norm is a t-conorm and vice versa.

Besides conjunctors, disjunctors and negators, we recall the notion of implicators. They extend the Boolean implication to the fuzzy setting.

An *implicator* is a mapping $\mathscr{I}\colon [0,1]^2 \to [0,1]$ that is decreasing in the first and increasing in the second argument and that satisfies the boundary conditions $\mathscr{I}(0,0) = \mathscr{I}(0,1) = \mathscr{I}(1,1) = 1$ and $\mathscr{I}(1,0) = 0$. It is called a *border implicator* if for all $x \in [0,1]$ it holds that $\mathscr{I}(1,x) = x$. It satisfies the *weak confinement principle* if $\forall x,y \in [0,1] : x \leq y \Rightarrow \mathscr{I}(x,y) = 1$. An implicator that satisfies the weak confinement principle is called a *WCP-implicator* in this paper.

Let $\mathscr{I}$ be an implicator. The *induced negator* of $\mathscr{I}$ is the negator $\mathscr{N}_{\mathscr{I}}$ defined by

$$\forall x \in [0,1] : \mathscr{N}_{\mathscr{I}}(x) = \mathscr{I}(x,0).$$

There are two important classes of implicators: S-implicators based on disjunctors and negators and R-implicators based on border conjunctors.

- Let $\mathscr{D}$ be a disjunctor and $\mathscr{N}$ a negator. The *S-implicator* $\mathscr{I}_{\mathscr{D},\mathscr{N}}$ based on $\mathscr{D}$ and $\mathscr{N}$ is defined by
$$\forall x,y \in [0,1] : \mathscr{I}_{\mathscr{D},\mathscr{N}}(x,y) = \mathscr{D}(\mathscr{N}(x), y).$$

- Let $\mathscr{C}$ be a border conjunctor. The *R-implicator* $\mathscr{I}_{\mathscr{C}}$ based on $\mathscr{C}$ is defined by

$$\forall x,y \in [0,1] : \mathscr{I}_{\mathscr{C}}(x,y) = \sup\{z \in [0,1] \mid \mathscr{C}(x,z) \leq y\}.$$

Both S- and R-implicators are border implicators. R-implicators are also WCP-implicators. A special class of R-implicators are IMTL-implicators, where IMTL stands for 'Involutive Monoidal T-norm based Logic' [16, 21]: these R-implicators are based on a left-continuous t-norm and have an involutive induced negator. A left-continuous t-norm of which the R-implicator is an IMTL-implicator, is called an IMTL-t-norm.

Example 4. *The Kleene-Dienes (KD) implicator $\mathscr{I}_{KD}$ is an S-implicator defined by, for $x,y \in [0,1]$, $\mathscr{I}_{KD}(x,y) = \max(1-x,y)$. The Łukasiewicz implicator $\mathscr{I}_L$ is both an S-implicator and R-implicator defined by, for $x,y \in [0,1]$, $\mathscr{I}_L(x,y) = \min(1,1-x+y)$.*

Finally, we recall that we can construct conjunctors based on implicators and involutive negators. Given an involutive negator $\mathscr{N}$ and an implicator $\mathscr{I}$, the *induced conjunctor of $\mathscr{I}$ and $\mathscr{N}$* is the conjunctor $\mathscr{C}_{\mathscr{I},\mathscr{N}}$ defined by

$$\forall x,y \in [0,1] : \mathscr{C}_{\mathscr{I},\mathscr{N}}(x,y) = \mathscr{N}(\mathscr{I}(x,\mathscr{N}(y))),$$

which is not necessarily a t-norm.

2.2. Fuzzy Sets and Relations

In this section, we recall basic notions about fuzzy sets and fuzzy relations [68].

- A *fuzzy set* A in a non-empty universe U is a mapping $A\colon U \to [0,1]$. The collection of all fuzzy sets in U is denoted by $\mathscr{F}(U)$. If U is finite, the *cardinality* of A is defined by

$$|A| = \sum_{x \in U} A(x).$$

The *support* of A is the crisp set $\mathrm{supp}(A) = \{x \in U \mid A(x) > 0\}$. Given $\alpha \in [0,1]$, the *α-level set* A_α of A in U is a crisp set in U such that $x \in A_\alpha$ if and only if $A(x) \geq \alpha$.

- Given $\alpha \in [0,1]$, the *constant (fuzzy) set* $\hat{\alpha}$ in U is defined by

$$\forall x \in U\colon \hat{\alpha}(x) = \alpha.$$

In the crisp case, the only constant sets in U are $\hat{0} = \emptyset$ and $\hat{1} = U$.

- Let $A,B \in \mathscr{F}(U)$. Given a negator $\mathscr{N}$, a conjunctor $\mathscr{C}$ and a disjunctor $\mathscr{D}$, the *$\mathscr{N}$-complement $A^{\mathscr{N}}$ of A* is given by

$$\forall x \in U : A^{\mathscr{N}}(x) = \mathscr{N}(A(x)).$$

The *$\mathscr{C}$-intersection of A and B* is the fuzzy set $A \cap_{\mathscr{C}} B$ defined by

$$\forall x \in U : (A \cap_{\mathscr{C}} B)(x) = \mathscr{C}(A(x),B(x)).$$

The *$\mathscr{D}$-union of A and B* is the fuzzy set $A \cup_{\mathscr{D}} B$ given by

$$\forall x \in U : (A \cup_{\mathscr{D}} B)(x) = \mathscr{D}(A(x),B(x)).$$

For $\mathscr{C} = \mathscr{T}_M$ and $\mathscr{D} = \mathscr{S}_M$, we shortly denote $A \cap B$ and $A \cup B$.

- A binary fuzzy relation R in U is a fuzzy set in $U \times U$. We define its inverse fuzzy relation R' by $R'(x,y) = R(y,x)$ for $x,y \in U$. The relation R is called *reflexive* if for all x in U it holds that $R(x,x) = 1$ and it is called *serial* if for all $x \in U$: $\sup_{y \in U} R(x,y) = 1$. We call the relation *symmetric* if for all $x,y \in U$, $R(x,y) = R(y,x)$. It is obvious that for a symmetric binary fuzzy relation R it holds that $R = R'$. Given a t-norm $\mathscr{T}$, the relation R is called *$\mathscr{T}$-transitive* if for all x,y and z in U it holds that

$$\mathscr{T}(R(x,y),R(y,z)) \leq R(x,z).$$

  Furthermore, the relation R is called *$\mathscr{T}$-euclidean* if for all x,y and z in U it holds that

$$\mathscr{T}(R(y,x),R(y,z)) \leq R(x,z).$$

  Every symmetric and $\mathscr{T}$-transitive relation is $\mathscr{T}$-euclidean. If a relation R is reflexive, symmetric and $\mathscr{T}$-transitive, it is called a *(fuzzy) $\mathscr{T}$-similarity relation*. When $\mathscr{T} = \mathscr{T}_M$, we also call it a *similarity relation*. Since the minimum operator is the largest t-norm, a similarity relation is $\mathscr{T}$-transitive for every t-norm $\mathscr{T}$.

- Let R be a binary fuzzy relation and $x \in U$, then the *R-foreset of x* is the fuzzy set Rx such that for every $y \in U$, $Rx(y) = R(y,x)$. When R is a crisp binary relation, the R-foreset of x is the crisp set $\{y \in U \mid (y,x) \in R\}$.

2.3. Rough Sets

In this section, we recall the definition of the lower and upper approximation in a Pawlak approximation space [47].

A *classic* or *Pawlak approximation space* is a couple (U,R) consisting of a non-empty universe U and an equivalence relation R in U. The *rough approximation* of a crisp set A in U by R is the pair of sets $(\underline{\text{apr}}_R(A), \overline{\text{apr}}_R(A))$ defined by, for $x \in U$,

$$x \in \underline{\text{apr}}_R(A) \quad \Leftrightarrow \quad (\forall y \in U)((y,x) \in R \Rightarrow y \in A), \tag{1}$$

$$x \in \overline{\text{apr}}_R(A) \quad \Leftrightarrow \quad (\exists y \in U)((y,x) \in R \wedge y \in A). \tag{2}$$

A pair (A_1,A_2) of sets in U is called a *rough set* in (U,R) if there is a set A in U such that $A_1 = \underline{\text{apr}}_R(A)$ and $A_2 = \overline{\text{apr}}_R(A)$. A_1 is called the *lower approximation* of A and A_2 is its *upper approximation*. A set A is *definable* when $\underline{\text{apr}}_R(A) = A = \overline{\text{apr}}_R(A)$ holds.

We list the most important properties of the rough approximation in a Pawlak approximation space in Table 1. The properties (D), (INC), (SM), (IU), (ID), (LU) and (UE) were already stated in [47].

The idea of property (A) comes from the concept of a Galois connection [30, 56]: let U_1 and U_2 be two universes and $f\colon U_1 \to U_2$ and $g\colon U_2 \to U_1$ two mappings between these universes. We call (f,g) a *Galois connection* in (U_1,U_2) if for all $x \in U_1$ and $y \in U_2$ it holds that

$$f(x) \leq_2 y \Leftrightarrow x \leq_1 g(y),$$

where $\leq_i$ is an order relation in U_i. It is clear that if f and g are the upper and lower approximation operator respectively, then they form a Galois connection in $(\mathscr{F}(U),\mathscr{F}(U))$ with $\leq_1 = \leq_2 = \subseteq$. In the special case that $B = A$ in Table 1, the properties (A) and (INC) imply that

$$\underline{\text{apr}}_R(A) = A \Leftrightarrow A = \overline{\text{apr}}_R(A).$$

Finally, we add the (RM)-property since it is relevant in many applications of rough sets; in particular, in the context of attribute selection in information systems [52, 53], R represents an indiscernibility equivalence relation between objects based on their attribute values. When attributes are omitted from the information system, the granularity imposed by the equivalence relation becomes coarser and it is then desirable that the lower approximation does not increase, and the upper approximation does not decrease.

Property	Abbreviation	Definition
Duality	(D)	$\underline{\mathrm{apr}}_R(A) = (\overline{\mathrm{apr}}_R(A)^{\mathrm{c}})^{\mathrm{c}}$
Adjointness	(A)	$\overline{\mathrm{apr}}_R(A) \subseteq B \Leftrightarrow A \subseteq \underline{\mathrm{apr}}_R(B)$
Inclusion	(INC)	$\underline{\mathrm{apr}}_R(A) \subseteq A \subseteq \overline{\mathrm{apr}}_R(A)$
Set Monotonicity	(SM)	$A \subseteq B \Rightarrow \begin{cases} \underline{\mathrm{apr}}_R(A) \subseteq \underline{\mathrm{apr}}_R(B) \\ \overline{\mathrm{apr}}_R(A) \subseteq \overline{\mathrm{apr}}_R(B) \end{cases}$
Relation Monotonicity	(RM)	$R_1 \subseteq R_2 \Rightarrow \begin{cases} \underline{\mathrm{apr}}_{R_1}(A) \supseteq \underline{\mathrm{apr}}_{R_2}(A) \\ \overline{\mathrm{apr}}_{R_1}(A) \subseteq \overline{\mathrm{apr}}_{R_2}(A) \end{cases}$
Intersection and Union	(IU)	$\underline{\mathrm{apr}}_R(A \cap B) = \underline{\mathrm{apr}}_R(A) \cap \underline{\mathrm{apr}}_R(B)$ $\overline{\mathrm{apr}}_R(A \cup B) = \overline{\mathrm{apr}}_R(A) \cup \overline{\mathrm{apr}}_R(B)$
Idempotence	(ID)	$\underline{\mathrm{apr}}_R(\underline{\mathrm{apr}}_R(A)) = \underline{\mathrm{apr}}_R(A)$ $\overline{\mathrm{apr}}_R(\overline{\mathrm{apr}}_R(A)) = \overline{\mathrm{apr}}_R(A)$
Interaction Lower and Upper	(LU)	$\overline{\mathrm{apr}}_R(\underline{\mathrm{apr}}_R(A)) = \underline{\mathrm{apr}}_R(A)$ $\underline{\mathrm{apr}}_R(\overline{\mathrm{apr}}_R(A)) = \overline{\mathrm{apr}}_R(A)$
Universe and Empty Set	(UE)	$\underline{\mathrm{apr}}_R(U) = U = \overline{\mathrm{apr}}_R(U)$ $\underline{\mathrm{apr}}_R(\emptyset) = \emptyset = \overline{\mathrm{apr}}_R(\emptyset)$

Table 1: Properties of the lower and upper approximation in a Pawlak approximation space

A disadvantage of Pawlak's model is that the lower and upper approximation operators are fairly rigid: for instance, an element belongs to the lower approximation of a set A only if its entire equivalence class is included in A. This makes the model less suitable for practical applications in data analysis, where some tolerance to noisy data is required. To mend this shortcoming, Ziarko and Katzberg [32, 70] designed the Variable Precision Rough Set (VPRS) model, only defined for finite universes U. They use the notion of the *rough membership function* R_A of A, given by

$$\forall x \in U : R_A(x) = \frac{|[x]_R \cap A|}{|[x]_R|},$$

where R is an equivalence relation.

Now let $0 \leq l < u \leq 1$. The *VPRS lower approximation* $\underline{\mathrm{apr}}^u_R A$ of A is defined by

$$\forall x \in U : x \in \underline{\mathrm{apr}}^u_R A \Leftrightarrow R_A(x) \geq u$$

and the *VPRS upper approximation* $\overline{\text{apr}}^l_R A$ of A by

$$\forall x \in U : x \in \overline{\text{apr}}^l_R A \Leftrightarrow R_A(x) > l.$$

It is clear that when $l = 0$ and $u = 1$, Pawlak's model is retrieved. When $0 < l < u < 1$, the VPRS model satisfies (SM) and (UE) and it satisfies (D) if $0 \leq l < 0.5$ and $u = 1 - l$.

Pawlak's original rough set model has been generalized in various other ways. In the remainder of this paper, we study the case where A is a fuzzy set and R is a fuzzy relation.

3. Implicator-Conjunctor-Based Fuzzy Rough Set Model

Equations (1) and (2) of the lower and upper approximation operators in a Pawlak approximation space can be generalized in many ways. In this section, we consider a general fuzzy rough set model based on a conjunctor, an implicator and a general binary fuzzy relation that encapsulates many existing definitions of fuzzy rough sets. We also study under which conditions the properties of Pawlak's model listed in Table 1 can be maintained by the extension.

3.1. Definition

For the remainder of the article, we consider a *fuzzy approximation space* (U,R). The couple (U,R) consists of a non-empty universe U and a binary fuzzy relation R on U. We consider the following general format for the approximation operators.

Definition 1. *Let (U,R) be a fuzzy approximation space, A a fuzzy set in U, $\mathscr{I}$ an implicator and $\mathscr{C}$ a conjunctor. The* $(\mathscr{I},\mathscr{C})$-fuzzy rough approximation *of A by R is the pair of fuzzy sets $(\underline{\text{apr}}^{\mathscr{I}}_R(A), \overline{\text{apr}}^{\mathscr{C}}_R(A))$ defined by, for $x \in U$:*

$$(\underline{\text{apr}}^{\mathscr{I}}_R(A))(x) = \inf_{y \in U} \mathscr{I}(R(y,x), A(y)), \tag{3}$$

$$(\overline{\text{apr}}^{\mathscr{C}}_R(A))(x) = \sup_{y \in U} \mathscr{C}(R(y,x), A(y)). \tag{4}$$

A pair (A_1,A_2) of fuzzy sets in U is called a fuzzy rough set *in (U,R) if there is a fuzzy set A in U such that $A_1 = \underline{\text{apr}}^{\mathscr{I}}_R(A)$ and $A_2 = \overline{\text{apr}}^{\mathscr{C}}_R(A)$. A fuzzy set A is called* definable *if $\underline{\text{apr}}^{\mathscr{I}}_R(A) = A = \overline{\text{apr}}^{\mathscr{C}}_R(A)$.*

In the remainder of this work, we will refer to this model as the implicator-conjunctor-based (fuzzy rough set) model or shortly, IC model.

When the fuzzy set we want to approximate is crisp and the relation R is a crisp equivalence relation, this definition coincides with the lower and upper approximation operators in a Pawlak approximation space:

$$\begin{aligned}
(\underline{\text{apr}}^{\mathscr{I}}_R(A))(x) = 1 &\Leftrightarrow \forall y \in U : \mathscr{I}(R(y,x), A(y)) = 1 \\
&\Leftrightarrow \forall y \in U : R(y,x) = 1 \Rightarrow A(y) = 1, \\
(\overline{\text{apr}}^{\mathscr{C}}_R(A))(x) = 1 &\Leftrightarrow \exists y \in U : \mathscr{C}(R(y,x), A(y)) = 1 \\
&\Leftrightarrow \exists y \in U : R(y,x) = 1 \wedge A(y) = 1.
\end{aligned}$$

Reference	Conjunctor	Implicator	Fuzzy relation
[14, 15] Dubois & Prade, 1990	min	KD-impl.	$\mathscr{T}_M$-similarity
[44] Morsi & Yakout, 1998	left-cont. t-norm	R-impl.	$\mathscr{T}$-similarity
[22, 23] Greco et al., 1998	t-norm	S-impl.	reflexive
[4] Boixader et al., 2000	cont. t-norm	R-impl.	$\mathscr{T}$-similarity
[51] Radzikowska & Kerre, 2002	t-norm	border impl.	$\mathscr{T}_M$-similarity
[60, 61] Wu et al., 2003	min	S-impl.	general
[41] Mi & Zhang, 2004	conjunctor	R-impl.	general
[50] Pei, 2005	min	S-impl.	general
[59] Wu et al., 2005	cont. t-norm	implicator	general
[67] Yeung et al., 2005	left-cont. t-norm	S-impl.	general
[67] Yeung et al., 2005	conjunctor	R-impl.	general
[38] Liu, 2006	t-norm	R-impl.	general
[9] De Cock et al., 2007	t-norm	border impl.	general
[36] Liu, 2008	min	S-impl.	general
[40] Mi et al., 2008	cont. t-norm	S-impl.	general
[26, 28] Hu et al., 2010	left-cont. t-norm	S-impl.	$\mathscr{T}_{\cos}$-similarity
[26, 28] Hu et al., 2010	conjunctor	R-impl.	$\mathscr{T}_{\cos}$-similarity

Table 2: Overview of implicator-conjunctor-based fuzzy rough set model in literature

Definition 1 covers many fuzzy rough set models that have been proposed in literature, and which emerge by choosing a specific conjunctor $\mathscr{C}$, implicator $\mathscr{I}$ and binary fuzzy relation R. We list these models in Table 2. Dubois and Prade [14] used the Kleene-Dienes implicator and the minimum operator to replace the Boolean implication and conjunction respectively. Some authors [28, 44, 67] require lower semicontinuity of $\mathscr{T}$ instead of left-continuity, but by a result from [20] these two notions are equivalent for t-norms. Also, some papers [38, 40, 41, 59–61] consider fuzzy relations in $U \times W$, with both U and W non-empty, finite universes, but here we restrict ourselves to the case where $U = W$. As can be seen from Table 2, Greco et al. [22, 23] were the first to consider reflexive binary fuzzy relations and Wu et al. [60, 61] were the first to consider general binary fuzzy relations. Also note that Greco et al. used arbitrary t-norms and t-conorms instead of the infimum and supremum operators. Mi and Zhang [41] initiated the use of conjunctors that are not necessarily t-norms. Also note that the t-norm $\mathscr{T}_{\cos}$ used in [28] is defined, for x, y in $[0,1]$, by $\mathscr{T}_{\cos}(x,y) = \max(xy - \sqrt{(1-x^2)(1-y^2)}, 0)$. Its use is inspired by the fact that some commonly used kernel functions in machine learning are in fact $\mathscr{T}_{\cos}$-similarity relations.

A variant of the IC model was proposed in [29] by Inuiguchi: the lower approximation $\underline{\mathrm{apr}}^{I,\mathscr{I}}_{R^*}(A)$ of A by a fuzzy relation R^* in U is given by, for x in U,

$$(\underline{\mathrm{apr}}^{I,\mathscr{I}}_{R^*}(A))(x) = \min(A(x), \inf_{y \in U} \mathscr{I}(R^*(y,x), A(y))),$$

while the upper approximation $\overline{\mathrm{apr}}^{I,\mathscr{C}}_{R^*}(A)$ of A by R^* is given by, for x in U,

$$(\overline{\mathrm{apr}}^{I,\mathscr{C}}_{R^*}(A))(x) = \max(A(x), \sup_{y \in U} \mathscr{C}(R^*(x,y), A(y))).$$

Note that this model can be seen as a special case of the IC model if in Definition 1 a border implicator $\mathscr{I}$ and a border conjunctor $\mathscr{C}$ are chosen and if the relation R defined by

$$\forall x,y \in U: R(x,y) = \max(R^*(x,y), Id(x,y)),$$

is symmetric, where Id is defined by

$$\forall x,y \in U: Id(x,y) = \begin{cases} 1 & x = y, \\ 0 & \text{otherwise.} \end{cases}$$

3.2. Properties

In this section, we consider the properties of the IC model. In particular, we discuss which properties of Pawlak's rough set model are always maintained, and which conditions need to be imposed in order for the remaining ones to hold.

In Table 3, we list the adaptation of the properties in Table 1 to a fuzzy approximation space, and provide references to the papers where those properties have been considered. Since most of these properties have been studied extensively in literature in one form or another, and the proofs for the IC model emerge by small modifications of existing results, we do not provide detailed proofs in this section.

Note that in generalizing the adjointness condition to a fuzzy approximation space, we have replaced R in the right hand side of the equivalence by its inverse fuzzy relation R'. Clearly, if R is symmetric, this modification is redundant. Also note that the constant set property (CS) emerges by extending (UE) to every constant fuzzy set $\hat{\alpha}$ for $\alpha \in [0,1]$. If (CS) is satisfied, then a priori (UE) also holds.

In the following, we assume that (U,R), (U,R_1), (U,R_2) are fuzzy approximation spaces, $\mathscr{I}$ an implicator, $\mathscr{C}$ a conjunctor and $\mathscr{N}$ an involutive negator.

Proposition 1. *If $\mathscr{C}$ is the induced conjunctor of $\mathscr{I}$ and $\mathscr{N}$, i.e., $\mathscr{C} = \mathscr{C}_{\mathscr{I},\mathscr{N}}$, then (D) holds.*

Corollary 1. *Let $\mathscr{D}$ be the $\mathscr{N}$-dual disjunctor of $\mathscr{C}$. If the pair $(\mathscr{I},\mathscr{C})$ consists of the S-implicator $\mathscr{I}_{\mathscr{D},\mathscr{N}}$ and the conjunctor $\mathscr{C}$, then (D) holds.*

Corollary 2. *Let $\mathscr{T}$ be an IMTL-t-norm such that $\mathscr{N}_{\mathscr{I}_{\mathscr{T}}} = \mathscr{N}$. If the pair $(\mathscr{I},\mathscr{C})$ consists of the R-implicator $\mathscr{I}_{\mathscr{T}}$ and the t-norm $\mathscr{T}$, then (D) holds.*

To see Corollary 2, note that $\mathscr{C}_{\mathscr{I}_{\mathscr{T}},\mathscr{N}} = \mathscr{T}$ holds (see e.g. [56]): for x,y in $[0,1]$, $\mathscr{C}_{\mathscr{I}_{\mathscr{T}},\mathscr{N}}(x,y) = \mathscr{N}(\mathscr{I}_{\mathscr{T}}(x,\mathscr{N}(y))) = \mathscr{N}(\mathscr{I}_{\mathscr{T}}(x,\mathscr{I}_{\mathscr{T}}(y,0))) = \mathscr{N}(\mathscr{I}_{\mathscr{T}}(\mathscr{T}(x,y),0)) = \mathscr{N}(\mathscr{N}(\mathscr{T}(x,y))) = \mathscr{T}(x,y)$.

Proposition 2. *If the pair $(\mathscr{I},\mathscr{C})$ consists of the R-implicator $\mathscr{I}_{\mathscr{T}}$ and the left-continuous t-norm $\mathscr{T}$, then (A) holds.*

Note that Proposition 2 does not hold for S-implicators:

Example 5. *Let $U = \{x,y\}$, R the similarity relation with $R(x,y) = 0.5$ and A the fuzzy set such that $A(x) = 0.3$ and $A(y) = 0.4$. Let $\mathscr{T}$ be the minimum t-norm and $\mathscr{I}$ the Kleene-Dienes implicator. Then it holds that $A = \underline{\mathrm{apr}}_R^{\mathscr{I}}(A)$, but since $(\overline{\mathrm{apr}}_R^{\mathscr{T}}(A))(x) = 0.4$, it does not hold that $\overline{\mathrm{apr}}_R^{\mathscr{T}}(A) \subseteq A$.*

Property	Definition	References
(D)	$\underline{\mathrm{apr}}_R^{\mathscr{I}}(A) = (\overline{\mathrm{apr}}_R^{\mathscr{C}}(A^{\mathscr{N}}))^{\mathscr{N}}$	[4, 29, 36, 40, 41, 50, 51, 59–61, 67]
(A)	$\overline{\mathrm{apr}}_R^{\mathscr{C}}(A) \subseteq B \Leftrightarrow A \subseteq \underline{\mathrm{apr}}_{R'}^{\mathscr{I}}(B)$	[26, 44], for $B = A$
(INC)	$\underline{\mathrm{apr}}_R^{\mathscr{I}}(A) \subseteq A \subseteq \overline{\mathrm{apr}}_R^{\mathscr{C}}(A)$	[4, 9, 26, 28, 29, 36, 40, 41, 44, 51] [59–61, 67]
(SM)	$A \subseteq B \Rightarrow \begin{cases} \underline{\mathrm{apr}}_R^{\mathscr{I}}(A) \subseteq \underline{\mathrm{apr}}_R^{\mathscr{I}}(B) \\ \overline{\mathrm{apr}}_R^{\mathscr{C}}(A) \subseteq \overline{\mathrm{apr}}_R^{\mathscr{C}}(B) \end{cases}$	[4, 9, 29, 36, 40, 44, 50, 51, 59–61]
(RM)	$R_1 \subseteq R_2 \Rightarrow \begin{cases} \underline{\mathrm{apr}}_{R_1}^{\mathscr{I}}(A) \supseteq \underline{\mathrm{apr}}_{R_2}^{\mathscr{I}}(A) \\ \overline{\mathrm{apr}}_{R_1}^{\mathscr{C}}(A) \subseteq \overline{\mathrm{apr}}_{R_2}^{\mathscr{C}}(A) \end{cases}$	[28, 50]
(IU)	$\underline{\mathrm{apr}}_R^{\mathscr{I}}(A \cap B) = \underline{\mathrm{apr}}_R^{\mathscr{I}}(A) \cap \underline{\mathrm{apr}}_R^{\mathscr{I}}(B)$ $\overline{\mathrm{apr}}_R^{\mathscr{C}}(A \cup B) = \overline{\mathrm{apr}}_R^{\mathscr{C}}(A) \cup \overline{\mathrm{apr}}_R^{\mathscr{C}}(B)$	[4, 26, 28, 29, 36, 38, 40, 41, 44, 50] [51, 59–61, 67]
(ID)	$\underline{\mathrm{apr}}_R^{\mathscr{I}}(\underline{\mathrm{apr}}_R^{\mathscr{I}}(A)) = \underline{\mathrm{apr}}_R^{\mathscr{I}}(A)$ $\overline{\mathrm{apr}}_R^{\mathscr{C}}(\overline{\mathrm{apr}}_R^{\mathscr{C}}(A)) = \overline{\mathrm{apr}}_R^{\mathscr{C}}(A)$	[4, 26, 28, 29, 36, 40, 41, 44, 50, 51] [59–61, 67]
(LU)	$\overline{\mathrm{apr}}_R^{\mathscr{C}}(\underline{\mathrm{apr}}_R^{\mathscr{I}}(A)) = \underline{\mathrm{apr}}_R^{\mathscr{I}}(A)$ $\underline{\mathrm{apr}}_R^{\mathscr{I}}(\overline{\mathrm{apr}}_R^{\mathscr{C}}(A)) = \overline{\mathrm{apr}}_R^{\mathscr{C}}(A)$	[29, 44, 50, 51, 59–61, 67]
(CS)	$\underline{\mathrm{apr}}_R^{\mathscr{I}}(\hat{\alpha}) = \hat{\alpha}$ $\overline{\mathrm{apr}}_R^{\mathscr{C}}(\hat{\alpha}) = \hat{\alpha}$	[4, 38, 40, 44, 51, 59–61]
(UE)	$\underline{\mathrm{apr}}_R^{\mathscr{I}}(U) = U = \overline{\mathrm{apr}}_R^{\mathscr{C}}(U)$ $\underline{\mathrm{apr}}_R^{\mathscr{I}}(\emptyset) = \emptyset = \overline{\mathrm{apr}}_R^{\mathscr{C}}(\emptyset)$	[29, 36, 50]

Table 3: Properties of the lower and upper approximation in a fuzzy approximation space

Proposition 3. *If R is reflexive, $\mathscr{I}$ is a border implicator and $\mathscr{C}$ is a border conjunctor, then (INC) holds.*

Corollary 3. *Let $\mathscr{T}$ be a t-norm and $\mathscr{S}$ its $\mathscr{N}$-dual t-conorm. If R is reflexive, and $(\mathscr{I},\mathscr{C}) = (\mathscr{I}_{\mathscr{S},\mathscr{N}},\mathscr{T})$ or $(\mathscr{I},\mathscr{C}) = (\mathscr{I}_{\mathscr{T}},\mathscr{T})$, then (INC) holds.*

Proposition 4. *The properties (SM), (RM) and (IU) always hold.*

Proposition 5. *If R is a $\mathscr{T}$-transitive relation, where $\mathscr{T}$ is a left-continuous t-norm and the pair $(\mathscr{I},\mathscr{C})$ consists of the R-implicator $\mathscr{I}_{\mathscr{T}}$ and the t-norm $\mathscr{T}$, or $(\mathscr{I},\mathscr{C})$ consists of the S-implicator $\mathscr{I}_{\mathscr{S},\mathscr{N}}$ based on the $\mathscr{N}$-dual t-conorm of $\mathscr{T}$ with respect to an involutive negator $\mathscr{N}$ and the t-norm $\mathscr{T}$, then (ID) holds.*

Proposition 6. *Let $\mathscr{T}$ be a left-continuous t-norm. If R is a reflexive and $\mathscr{T}$-euclidean relation, or if R is a $\mathscr{T}$-similarity relation, and the pair $(\mathscr{I},\mathscr{C})$ consists of the R-implicator $\mathscr{I}_{\mathscr{T}}$ and the t-norm $\mathscr{T}$, then (LU) holds.*

Proposition 6 does not hold for S-implicators.

Example 6. *Let $U = \{x,y\}$, R the similarity relation with $R(x,y) = 0.5$ and A the fuzzy set such that $A(x) = 0.2$ and $A(y) = 0.6$. Let $\mathscr{T}$ be the minimum t-norm and $\mathscr{I}$ the Kleene-Dienes implicator. Then it holds that $(\underline{\mathrm{apr}}_R^{\mathscr{I}}(A))(x) = 0.2$ and $(\underline{\mathrm{apr}}_R^{\mathscr{I}}(A))(y) = 0.5$, but since*

$(\overline{\text{apr}}_R^{\mathscr{T}}(\underline{\text{apr}}_R^{\mathscr{I}}(A)))(x) = 0.5$*, it does not hold that* $\overline{\text{apr}}_R^{\mathscr{T}}(\underline{\text{apr}}_R^{\mathscr{I}}(A)) = \underline{\text{apr}}_R^{\mathscr{I}}(A)$*. Moreover, by duality, it does not hold that* $\underline{\text{apr}}_R^{\mathscr{I}}(\overline{\text{apr}}_R^{\mathscr{T}}(A)) = \overline{\text{apr}}_R^{\mathscr{T}}(A)$.

Proposition 7. *If R is reflexive,* $\mathscr{I}$ *a border implicator and* $\mathscr{C}$ *a border conjunctor, then (CS) and (UE) hold.*

Corollary 4. *Let* $\mathscr{T}$ *be a t-norm and* $\mathscr{S}$ *its* $\mathscr{N}$*-dual t-conorm. If R is reflexive, and* $(\mathscr{I},\mathscr{C}) = (\mathscr{I}_{\mathscr{S},\mathscr{N}},\mathscr{T})$ *or* $(\mathscr{I},\mathscr{C}) = (\mathscr{I}_{\mathscr{T}},\mathscr{T})$*, then (CS) and (UE) hold.*

Summarizing, in order to satisfy all properties in Table 3, $\mathscr{C}$ should be an IMTL-t-norm $\mathscr{T}$ and $\mathscr{I}$ its R-implicator, while *R* needs to be at least reflexive, $\mathscr{T}$-transitive and $\mathscr{T}$-euclidean or symmetric:

Proposition 8. *Let* $\mathscr{C}$ *be an IMTL-t-norm* $\mathscr{T}$ *and* $\mathscr{I}$ *its R-implicator, and R a reflexive and* $\mathscr{T}$*-transitive relation. If R is either* $\mathscr{T}$*-euclidean or symmetric, then all the properties listed in Table 3 hold.*

It is interesting to observe that Dubois and Prade's original fuzzy rough set model [14] satisfies all properties except (A) and (LU).

4. Robust Fuzzy Rough Set Models

The fuzzy rough set models encapsulated by Definition 1 allow for a lot of flexibility in terms of the choice of logical connectives and the fuzzy relation. However, the use of the inf- and sup-operators in formulas (3) and (4) limits their practical use, in a similar way as the $\forall$- and $\exists$-quantifiers restrict the application potential of Pawlak's original rough set model.

The core of the problem is that the result of the approximations is determined by a single best (sup) or worst (inf) element. This can be a disadvantage in a data analysis context, since data samples may be erroneous. Such noisy data can perturb the approximations and therefore weaken the machine learning algorithms that invoke them [27].

To address this problem, many authors have defined robust alternatives for the lower and upper approximation operators in fuzzy rough set theory, in a similar way as the VPRS model provides a noise-tolerant alternative for Pawlak's rough set definition. In this section and the next one, we critically examine existing robust fuzzy rough set models, from both theoretical and practical point of view.

Different alternatives for robust fuzzy rough set models have been constructed. There are models which are frequency-based, analogous to the VPRS model of Ziarko [5, 24, 25, 42, 43, 65]. Another model adjusts the set which is approximated [69]. Moreover, some models use other aggregation operators than the infimum and supremum operators [7, 18]. To the best of our knowledge, the seven models we discuss here are the most widely used robust fuzzy rough set models [27, 65].

In this section we discuss for each considered model the specific criteria it uses for harnessing the approximations against noise. We generalize, correct or simplify its definition, without harming the original ideas it is based on. We reveal relationships that exist between the models, and also evaluate which properties of Pawlak's rough set model can be maintained; similarly as for VPRS, making the models more flexible towards noise typically involves sacrificing some of the desirable properties they satisfy. In such cases, we will construct explicit counterexamples, which are gathered in the Appendix. In Section 5, we will evaluate the claim

that the noise-tolerant models are more robust than the IC model from a practical perspective, by examining how stable their approximations are when the data are contaminated by noise.

Throughout this section, unless specified otherwise, we assume that A is a fuzzy set in a fuzzy approximation space (U,R), with U finite and R an arbitrary binary fuzzy relation in U. The restriction to finite universes is not a real limitation, since the number of data samples (elements of U) in a real information system is always finite.

4.1. Noise-Tolerant Models based on Frequency

We start with discussing frequency-based fuzzy rough set models. Models of this type are the Variable Precision Fuzzy Rough Set model of Mieskowicz-Rolka and Rolka [42, 43], the Vaguely Quantified Fuzzy Rough Set model of Cornelis et al. [5], the Soft Fuzzy Rough Set model of Hu et al. [24, 25] and the Variable Precision Fuzzy Rough Set model based on Fuzzy Granules of Yao et al. [65]. The idea behind this type of robust models is that only a subset of the R-foreset of an object x is taken into account when computing the lower and upper approximation in x, which is similar to the Variable Precision Rough Set model of Ziarko [70].

We start with the model of Mieskowicz-Rolka and Rolka.

4.1.1. Variable Precision Fuzzy Rough Set Model

The Variable Precision Fuzzy Rough Set (VPFRS) model was proposed by Mieszkowicz-Rolka and Rolka [42] in 2004, and later appeared in slightly revised form in [43]. They intended to design a robust fuzzy rough set model that covers both the seminal fuzzy rough set approach of Dubois and Prade [14, 15] and the VPRS model of Ziarko [32, 70]. We start by recalling the definition of the model as it was introduced by Mieszkowicz-Rolka and Rolka, and then show a simpler expression for computing the approximations.

Definition 2. *[43] Given a WCP-implicator $\mathscr{I}$, a t-norm $\mathscr{T}$ and fuzzy sets A,B in U. The* implication-based inclusion set $\mathrm{Incl}_{\mathscr{I}}(A,B)$ *of A in B is defined by*[1]

$$\forall x \in U : \mathrm{Incl}_{\mathscr{I}}(A,B)(x) = \mathscr{I}(A(x),B(x))$$

and the t-norm-based inclusion set[2] $\mathrm{Incl}_{\mathscr{T}}(A,B)$ *of A in B is defined by*

$$\forall x \in U : \mathrm{Incl}_{\mathscr{T}}(A,B)(x) = \mathscr{T}(A(x),B(x)).$$

Definition 3. *[43] Given a WCP-implicator $\mathscr{I}$, a t-norm $\mathscr{T}$, a non-empty fuzzy set A and a fuzzy set B in U, and $\alpha \in [0,1]$. The* lower α-inclusion error $\underline{e}_\alpha$ *of A in B is given by*

$$\underline{e}_\alpha(A,B) = 1 - \frac{|A \cap (\mathrm{Incl}_{\mathscr{I}}(A,B))_\alpha|}{|A|}$$

[1] In [43], the authors defined $\mathrm{Incl}_{\mathscr{I}}$ as follows:

$$\mathrm{Incl}_{\mathscr{I}}(A,B)(x) = \begin{cases} \mathscr{I}(A(x),B(x)) & A(x) > 0 \\ 0 & A(x) = 0. \end{cases}$$

However, the special treatment of the case $A(x) = 0$ is not necessary to properly define the VPFRS model, hence our simplified definition of $\mathrm{Incl}_{\mathscr{I}}$.

[2] Note that the name 'inclusion set" is slightly misleading because this definition is not actually related to fuzzy inclusion. However, we adopt the terminology of [43].

and the upper α-inclusion error $\bar{e}_\alpha$ *of A in B is given by*

$$\bar{e}_\alpha(A,B) = 1 - \frac{|A \cap ((\mathrm{Incl}_{\mathscr{T}}(A,B))^{\mathscr{N}_S})_\alpha|}{|A|}.$$

For the empty set $\emptyset$*, we define the inclusion errors by*[3]

$$\underline{e}_\alpha(\emptyset,B) = \bar{e}_\alpha(\emptyset,B) = 0.$$

Below is the definition of the VPFRS model. Mieszkowicz-Rolka and Rolka worked with a $\mathscr{T}$-similarity relation, which is generalized here to an arbitrary binary fuzzy relation.

Definition 4. *[43] Given a WCP-implicator* $\mathscr{I}$*, a t-norm* $\mathscr{T}$ *and* $0 \leq l < u \leq 1$*. The* (u,l)-fuzzy rough approximation *of A by R is the pair of fuzzy sets* $(\underline{\mathrm{apr}}_R^{\mathscr{I},u}(A), \overline{\mathrm{apr}}_R^{\mathscr{T},l}(A))$ *defined by, for x in U,*

$$\begin{aligned}
(\underline{\mathrm{apr}}_R^{\mathscr{I},u}(A))(x) &= \inf_{y \in S_{x,u}} \mathrm{Incl}_{\mathscr{I}}(Rx,A)(y), \\
(\overline{\mathrm{apr}}_R^{\mathscr{T},l}(A))(x) &= \sup_{y \in S_{x,l}} \mathrm{Incl}_{\mathscr{T}}(Rx,A)(y),
\end{aligned}$$

with

$$\begin{aligned}
S_{x,u} &= \mathrm{supp}(Rx) \cap \left((\mathrm{Incl}_{\mathscr{I}}(Rx,A))_{\alpha_{x,u}}\right) \\
&= \{y \in U \mid R(y,x) > 0 \text{ and } (\mathrm{Incl}_{\mathscr{I}}(Rx,A))(y) \geq \alpha_{x,u}\}, \\
\alpha_{x,u} &= \sup\{\alpha \in [0,1] \mid \underline{e}_\alpha(Rx,A) \leq 1-u\} \\
S_{x,l} &= \mathrm{supp}(Rx) \cap \left(((\mathrm{Incl}_{\mathscr{T}}(Rx,A))^{\mathscr{N}_S})_{\alpha_{x,l}}\right) \\
&= \{y \in U \mid R(y,x) > 0 \text{ and } (\mathrm{Incl}_{\mathscr{T}}(Rx,A))(y) \leq 1-\alpha_{x,l}\}, \\
\alpha_{x,l} &= \sup\{\alpha \in [0,1] \mid \bar{e}_\alpha(Rx,A) \leq l\}.
\end{aligned}$$

Remark 1. Note that when $Rx = \emptyset$, $S_{x,u}$ and $S_{x,l}$ are empty and $\alpha_{x,u} = \alpha_{x,l} = 1$. In this case, we obtain that $(\underline{\mathrm{apr}}_R^{\mathscr{I},u}(A))(x) = 1$ and $(\overline{\mathrm{apr}}_R^{\mathscr{T},l}(A))(x) = 0$.

The VPFRS model excludes certain elements from the inf and sup computations. In this sense, $\underline{\mathrm{apr}}_R^{\mathscr{I}}(A) \subseteq \underline{\mathrm{apr}}_R^{\mathscr{I},u}(A)$ and $\overline{\mathrm{apr}}_R^{\mathscr{T},l}(A) \subseteq \overline{\mathrm{apr}}_R^{\mathscr{T}}(A)$ always hold. In order to get more insight in this element exclusion process, we show that the approximations of Definition 4 can be simplified. To do this, we first obtain a simpler expression for $\alpha_{x,u}$ and $\alpha_{x,l}$.

Lemma 1. *Given a WCP-implicator* $\mathscr{I}$*, a t-norm* $\mathscr{T}$ *and* $0 \leq l < u \leq 1$*. For* $x \in U$ *such that* $Rx \neq \emptyset$*, it holds that*

$$\begin{aligned}
\alpha_{x,u} &= \max\left\{\alpha \in \{\mathscr{I}(R(y,x),A(y)) \mid y \in U\} \mid \frac{|Rx \cap (\mathrm{Incl}_{\mathscr{I}}(Rx,A))_\alpha|}{|Rx|} \geq u\right\}, \\
\alpha_{x,l} &= \max\left\{\alpha \in \{1-\mathscr{T}(R(y,x),A(y)) \mid y \in U\} \mid \frac{|Rx \cap ((\mathrm{Incl}_{\mathscr{T}}(Rx,A))^{\mathscr{N}_S})_\alpha|}{|Rx|} \geq 1-l\right\}.
\end{aligned}$$

[3] In [43], Mieszkowicz-Rolka and Rolka only defined the inclusion errors for a non-empty fuzzy set A. We extend their definition to include the empty set, in order to allow the use of a general binary fuzzy relation R in Definition 4.

Proof. Take x in U such that Rx is not empty. Since U is finite, we can number and rename its elements such that $U = \{z_1, \ldots, z_n\}$ and

$$\mathrm{Incl}_{\mathscr{I}}(Rx,A)(z_1) \geq \mathrm{Incl}_{\mathscr{I}}(Rx,A)(z_2) \geq \ldots \geq \mathrm{Incl}_{\mathscr{I}}(Rx,A)(z_n).$$

Define $F_x\colon [0,1] \to [0,1]$ such that, for $\alpha \in [0,1]$,

$$F_x(\alpha) = \frac{|Rx \cap (\mathrm{Incl}_{\mathscr{I}}(Rx,A))_\alpha|}{|Rx|}.$$

Then F_x is a decreasing mapping in α. We have the following expression for $\alpha_{x,u}$:

$$\alpha_{x,u} = \sup\{\alpha \in [0,1] \mid F_x(\alpha) \geq u\}.$$

We prove that we only need to consider the values $\mathrm{Incl}_{\mathscr{I}}(Rx,A)(y)$ to compute $\alpha_{x,u}$, that is,

$$\sup\{\alpha \in [0,1] \mid F_x(\alpha) \geq u\} = \sup\{\alpha \in \{\mathscr{I}(R(y,x),A(y)) \mid y \in U\} \mid F_x(\alpha) \geq u\}.$$

Take $\alpha^* \in [0,1]$ such that $\alpha^* \notin \{\mathscr{I}(R(y,x),A(y)) \mid y \in U\}$. We prove that this α^* does not influence the supremum. First, assume that α^* is such that

$$\mathrm{Incl}_{\mathscr{I}}(Rx,A)(z_i) > \alpha^*$$

for $i \in \{1,\ldots,n\}$, then $F_x(\alpha^*) = F_x(\mathrm{Incl}_{\mathscr{I}}(Rx,A)(z_i))$. Since $\mathrm{Incl}_{\mathscr{I}}(Rx,A)(z_i) > \alpha^*$, we do not need to take α^* into account when we compute the supremum of $\{\alpha \in [0,1] \mid F_x(\alpha) \geq u\}$. On the other hand, if $\alpha^* > \mathrm{Incl}_{\mathscr{I}}(Rx,A)(z_1)$, then $(\mathrm{Incl}_{\mathscr{I}}(Rx,A))_{\alpha^*} = \emptyset$, hence $F_x(\alpha^*) = 0$. Since we assume $u > 0$, it holds that

$$\alpha^* \notin \{\alpha \in [0,1] \mid F_x(\alpha) \geq u\}.$$

In both cases we conclude that α^* will not influence the supremum. Hence, we obtain that

$$\begin{aligned} \alpha_{x,u} &= \sup\{\alpha \in [0,1] \mid F_x(\alpha) \geq u\} \\ &= \sup\{\alpha \in \{\mathscr{I}(R(y,x),A(y)) \mid y \in U\} \mid F_x(\alpha) \geq u\} \\ &= \max\{\alpha \in \{\mathscr{I}(R(y,x),A(y)) \mid y \in U\} \mid F_x(\alpha) \geq u\}, \end{aligned}$$

since U is finite.

The proof for $\alpha_{x,l}$ is analogous. □

Proposition 9. *Given a WCP-implicator $\mathscr{I}$, a t-norm $\mathscr{T}$ and $0 \leq l < u \leq 1$. It holds that, for x in U,*

$$\begin{aligned} (\underline{\mathrm{apr}}_R^{\mathscr{I},u}(A))(x) &= \alpha_{x,u}, \\ (\overline{\mathrm{apr}}_R^{\mathscr{T},l}(A))(x) &= 1 - \alpha_{x,l}. \end{aligned}$$

Proof. By definition of $S_{x,u}$, it holds that

$$(\underline{\mathrm{apr}}_R^{\mathscr{I},u}(A))(x) = \inf_{y \in S_{x,u}} \mathscr{I}(R(y,x),A(y)) \geq \alpha_{x,u}.$$

By Lemma 1, let y^* be the element in U such that $\alpha_{x,u} = \mathscr{I}(R(y^*,x),A(y^*))$.

- If $R(y^*,x) > 0$, then $y^* \in S_{x,u}$ and hence $(\underline{\text{apr}}_R^{\mathscr{I},u}(A))(x) \leq \mathscr{I}(R(y^*,x),A(y^*)) = \alpha_{x,u}$.
- If $R(y^*,x) = 0$, then it holds that $\alpha_{x,u} = \mathscr{I}(0,A(y^*)) = 1$ and thus, $(\underline{\text{apr}}_R^{\mathscr{I},u}(A))(x) = 1$.

In both cases, we conclude that $(\underline{\text{apr}}_R^{\mathscr{I},u}(A))(x) = \alpha_{x,u}$.

In an analogous way, by definition of $S_{x,l}$, we know that $(\overline{\text{apr}}_R^{\mathscr{T},l}(A))(x) \leq 1 - \alpha_{x,l}$. By Lemma 1, let z^* be the element in U such that $\alpha_{x,l} = 1 - \mathscr{T}(R(z^*,x),A(z^*))$.

- If $R(z^*,x) > 0$, then $z^* \in S_{x,l}$ and hence $(\overline{\text{apr}}_R^{\mathscr{T},l}A)(x) \geq \mathscr{T}(R(z^*,x),A(z^*)) = 1 - \alpha_{x,l}$.
- If $R(z^*,x) = 0$, then it holds that $\alpha_{x,l} = 1 - \mathscr{T}(0,A(z^*)) = 1$ and thus, $(\overline{\text{apr}}_R^{\mathscr{T},l}(A))(x) = 0$.

Hence, we conclude that $(\overline{\text{apr}}_R^{\mathscr{T},l}(A))(x) = 1 - \alpha_{x,l}$. □

The above results give more insight into how the VPFRS model operates. The fuzzy set $\big(Rx \cap (\text{Incl}_{\mathscr{I}}(Rx,A))_{\text{Incl}_{\mathscr{I}}(Rx,A)(y)}\big)$ can be seen as the R-foreset of x with those elements z excluded, for which $\mathscr{I}(R(z,x),A(z)) < \mathscr{I}(R(y,x),A(y))$. If the cardinality of this restricted R-foreset is at least a fraction u of that of the entire R-foreset, and there is no smaller such fuzzy set satisfying this condition, then the VPFRS lower approximation equals $\mathscr{I}(R(y,x),A(y))$. An analogous interpretation can be given to the VPFRS upper approximation.

In the following propositions, we examine the conditions under which the VPFRS model coincides with the IC model and the VPRS model, respectively.

First, when $u = 1$ and $l = 0$, we derive a specific case of the IC model.

Proposition 10. *Given a WCP-implicator $\mathscr{I}$ and a t-norm $\mathscr{T}$. It holds that, for x in U,*

$$\begin{aligned}(\underline{\text{apr}}_R^{\mathscr{I},1}(A))(x) &= \inf_{y\in U} \mathscr{I}(R(y,x),A(y)),\\ (\overline{\text{apr}}_R^{\mathscr{T},0}(A))(x) &= \sup_{y\in U} \mathscr{T}(R(y,x),A(y)).\end{aligned}$$

Proof. Note that when Rx is empty, both equalities hold trivially.

Now take $x \in U$ such that Rx is not empty. With $u = 1$, we obtain that

$$\alpha_{x,1} = \sup\{\alpha \in [0,1] \mid \forall y \in U : R(y,x) > 0 \Rightarrow \mathscr{I}(R(y,x),A(y)) \geq \alpha\}.$$

If $R(y,x) = 0$, then $\mathscr{I}(R(y,x),A(y)) = 1 \geq \alpha$ for any $\alpha \in [0,1]$, so we obtain that

$$\alpha_{x,1} = \inf_{y\in U} \mathscr{I}(R(y,x),A(y)).$$

On the other hand, with $l = 0$, we derive that

$$\alpha_{x,0} = \sup\{\alpha \in [0,1] \mid \forall y \in U : R(y,x) > 0 \Rightarrow 1 - \mathscr{T}(R(y,x),A(y)) \geq \alpha\}.$$

If $R(y,x) = 0$, then $1 - \mathscr{T}(R(y,x),A(y)) = 1 \geq \alpha$ for any $\alpha \in [0,1]$, so it holds that

$$\begin{aligned}\alpha_{x,0} &= \inf_{y\in U} (1 - \mathscr{T}(R(y,x),A(y)))\\ &= 1 - \sup_{y\in U} \mathscr{T}(R(y,x),A(y)).\end{aligned}$$

The proof now follows from Proposition 9. □

Furthermore, when A is a crisp set and R is a crisp equivalence relation, the VPFRS model coincides with the VPRS model of Ziarko.

Proposition 11. *Given a WCP-implicator $\mathscr{I}$, a t-norm $\mathscr{T}$ and $0 \leq l < u \leq 1$. Let A be a crisp subset of U and R a crisp equivalence relation on U. For x in U it holds that*

$$\begin{aligned}(\underline{\mathrm{apr}}_R^{\mathscr{I},u}(A))(x) &= (\underline{\mathrm{apr}}_R^{u}(A))(x),\\ (\overline{\mathrm{apr}}_R^{\mathscr{T},l}(A))(x) &= (\overline{\mathrm{apr}}_R^{l}(A))(x).\end{aligned}$$

Proof. Take x in U arbitrary. Since A and R are crisp, $\alpha_{x,u}$ and $\alpha_{x,l}$ are either 1 or 0.

First, note that, for $y \in U$,

$$y \in [x]_R \cap (\mathrm{Incl}_{\mathscr{I}}([x]_R, A))_1 \Leftrightarrow R(y,x) = 1 \text{ and } A(y) = 1 \Leftrightarrow y \in [x]_R \cap A.$$

For $\alpha_{x,u}$ we derive that

$$\begin{aligned}\alpha_{x,u} = 1 &\Leftrightarrow \frac{|[x]_R \cap (\mathrm{Incl}_{\mathscr{I}}([x]_R, A))_1|}{|[x]_R|} \geq u\\ &\Leftrightarrow \frac{|[x]_R \cap A|}{|[x]_R|} \geq u\\ &\Leftrightarrow (\underline{\mathrm{apr}}_R^{u}(A))(x) = 1.\end{aligned}$$

On the other hand, since

$$y \in [x]_R \cap ((\mathrm{Incl}_{\mathscr{T}}([x]_R, A))^{\mathscr{N}_S})_1 \Leftrightarrow R(y,x) = 1 \text{ and } A(y) = 0 \Leftrightarrow y \in [x]_R \cap A^{\mathrm{c}},$$

we obtain for $\alpha_{x,l}$ that

$$\begin{aligned}\alpha_{x,l} = 1 &\Leftrightarrow \frac{|[x]_R \cap ((\mathrm{Incl}_{\mathscr{T}}([x]_R, A))^{\mathscr{N}_S})_1|}{|[x]_R|} \geq 1 - l\\ &\Leftrightarrow \frac{|[x]_R \cap A^{\mathrm{c}}|}{|[x]_R|} \geq 1 - l\\ &\Leftrightarrow \frac{|[x]_R \cap A|}{|[x]_R|} \leq l\\ &\Leftrightarrow (\overline{\mathrm{apr}}_R^{l}(A))(x) = 0.\end{aligned}$$

The proof now follows from Proposition 9. □

Finally, we prove that (SM) holds for the VPFRS model. The other properties do not hold (see Appendix, Example 9).

Proposition 12. *The VPFRS model satisfies (SM).*

Proof. Let A, B be fuzzy sets in U such that $A \subseteq B$. Due to Proposition 9, we have to prove that $\alpha_{x,u}^{A} \leq \alpha_{x,u}^{B}$ and $\alpha_{x,l}^{A} \geq \alpha_{x,l}^{B}$ for every x in U.

Take $x \in U$ arbitrary. We reorder the elements of U such that $U = \{z_1, \ldots, z_n\}$ and

$$\mathscr{I}(R(z_1,x), A(z_1)) \geq \mathscr{I}(R(z_2,x), A(z_2)) \geq \ldots \geq \mathscr{I}(R(z_n,x), A(z_n)).$$

By Lemma 1, there exists an $m \in \{1,\ldots,n\}$ such that $\alpha^A_{x,u} = \mathscr{I}(R(z_m,x),A(z_m))$. Now, for every $i \in \{1,\ldots,n\}$ it holds that $\mathscr{I}(R(z_i,x),B(z_i)) \geq \mathscr{I}(R(z_i,x),A(z_i))$ and thus for all $i \in \{1,\ldots,m\}$,

$$\mathscr{I}(R(z_i,x),B(z_i)) \geq \alpha^A_{x,u}.$$

By definition of $\alpha^A_{x,u}$, it holds that

$$\sum_{\substack{z\in U:\\ \mathscr{I}(R(z,x),A(z))\geq\alpha^A_{x,u}}} R(z,x) \geq u \cdot \sum_{z\in U} R(z,x)$$

and thus it also holds that

$$\sum_{\substack{z\in U:\\ \mathscr{I}(R(z,x),B(z))\geq\alpha^A_{x,u}}} R(z,x) \geq u \cdot \sum_{z\in U} R(z,x).$$

In other words,

$$\alpha^A_{x,u} \in \{\alpha \in [0,1] \mid \underline{e}_\alpha(Rx,B) \leq 1-u\}$$

Hence, $\alpha^A_{x,u} \leq \alpha^B_{x,u}$.

In an analogous way, we obtain that

$$\alpha^B_{x,l} \in \{\alpha \in [0,1] \mid \bar{e}_\alpha(Rx,A) \leq l\},$$

and thus, $\alpha^B_{x,l} \leq \alpha^A_{x,l}$. □

We continue with Cornelis et al.'s Vaguely Quantified Fuzzy Rough Set model.

4.1.2. Vaguely Quantified Fuzzy Rough Set Model

In 2007, Cornelis et al. [5] introduced the Vaguely Quantified Fuzzy Rough Set (VQFRS) model. In contrast to the other fuzzy rough set approaches, they did not make use of implicators and conjunctors, but they worked with fuzzy quantifiers to extend Ziarko's VPRS model.

Definition 5. *[5] A* regularly increasing fuzzy quantifier *is an increasing mapping* $Q\colon [0,1] \to [0,1]$ *that satisfies* $Q(0) = 0$ *and* $Q(1) = 1$.

The VQFRS model is then defined based on two such regularly increasing fuzzy quantifiers.

Definition 6. *[5] Given a couple* (Q_u, Q_l) *of regularly increasing fuzzy quantifiers, the* (Q_u, Q_l)-vaguely quantified fuzzy rough approximation *of A by R is the pair of fuzzy sets* $(\underline{\mathrm{apr}}^{Q_u}_R(A), \overline{\mathrm{apr}}^{Q_l}_R(A))$ *defined by, for* $x \in U$*:*

$$(\underline{\mathrm{apr}}^{Q_u}_R(A))(x) = \begin{cases} Q_u\left(\frac{|Rx\cap A|}{|Rx|}\right) & Rx \neq \emptyset \\ 1 & Rx = \emptyset, \end{cases}$$

$$(\overline{\mathrm{apr}}^{Q_l}_R(A))(x) = \begin{cases} Q_l\left(\frac{|Rx\cap A|}{|Rx|}\right) & Rx \neq \emptyset \\ 1 & Rx = \emptyset. \end{cases}$$

In [5], the interpretation of this model is as follows: x belongs to the lower approximation if "most" of the elements related to x belong to A and x belongs to the upper approximation if "at least some" elements related to x belong to A; the linguistic quantifiers "most" and "at least some" generalize the crisp $\forall$- and $\exists$-quantifiers, and are modeled by means of the fuzzy quantifiers Q_u and Q_l, respectively. In [5], as a specific example, the authors put $Q_u = Q_{(0.2,1)}$ and $Q_l = Q_{(0.1,0.6)}$, where $Q_{(\alpha,\beta)}$ with $\alpha, \beta \in [0,1]$ is defined by

$$\forall x \in [0,1]\colon Q_{(\alpha,\beta)}(x) = \begin{cases} 0 & x \leq \alpha \\ \frac{2(x-\alpha)^2}{(\beta-\alpha)^2} & \alpha \leq x \leq \frac{\alpha+\beta}{2} \\ 1 - \frac{2(x-\beta)^2}{(\beta-\alpha)^2} & \frac{\alpha+\beta}{2} \leq x \leq \beta \\ 1 & \beta \leq x. \end{cases}$$

As pointed out in [5], the VQFRS model has Pawlak's model and Ziarko's VPRS model as specific cases when A is a crisp set and R is a crisp equivalence relation. In the former case, $Q_u = Q_\forall$ and $Q_l = Q_\exists$, where

$$\forall x \in [0,1]\colon Q_\forall(x) = \begin{cases} 0, & x < 1 \\ 1, & x = 1, \end{cases}$$

$$\forall x \in [0,1]\colon Q_\exists(x) = \begin{cases} 0, & x = 0 \\ 1, & x > 0. \end{cases}$$

In the case of the VPRS model, $Q_u = Q_{\geq u}$ and $Q_l = Q_{>l}$, where

$$\forall x \in [0,1]\colon Q_{\geq u}(x) = \begin{cases} 0, & x < u \\ 1, & x \geq u, \end{cases}$$

$$\forall x \in [0,1]\colon Q_{>l}(x) = \begin{cases} 0, & x \leq l \\ 1, & x > l. \end{cases}$$

There is no connection between the VQFRS model and the IC model, i.e., we cannot find fuzzy quantifiers Q_u and Q_l such that $\forall A \in \mathscr{F}(U)$: $\underline{\text{apr}}_R^{Q_u}(A) = \underline{\text{apr}}_R^{\mathscr{I}}(A)$ and $\overline{\text{apr}}_R^{Q_l}(A) = \overline{\text{apr}}_R^{\mathscr{C}}(A)$.

It can be verified that for the VQFRS model, (SM) holds and (UE) holds if R is serial. Other properties do not hold (see Appendix, Example 10).

Proposition 13. *The VQFRS model satisfies (SM).*

Proof. This follows from the monotonicity of regularly increasing fuzzy quantifiers. □

Proposition 14. *If R is a serial fuzzy relation, (UE) holds.*

Proof. For every x in U, Rx is not empty and $Q_u(0) = Q_l(0) = 0$ and $Q_u(1) = Q_l(1) = 1$. □

Next, we discuss Hu et al.'s Soft Fuzzy Rough Set model.

4.1.3. Soft Fuzzy Rough Set Model

Inspired by soft margin Support Vector Machines [8], Hu et al. [24] proposed in 2010 the Soft Fuzzy Rough Set (SFRS) model as a new robust fuzzy rough set model.

An important ingredient of the model is the so-called *soft distance* between an element x and a crisp set A of U, defined as follows in [25]:

$$SD(x,A) = \arg_{d(x,y)} \max_{y\in A}\{d(x,y) - \delta m_{x,y}\},$$

where d is a distance funtion, $\delta > 0$ is a penalty factor and

$$m_{x,y} = |\{z \in U \mid d(x,z) < d(x,y)\}|.$$

However, we may encounter a problem with this definition due to the use of the function $\arg_{d(x,y)}$: when the value of the maximum is reached for different values of y, it is not clear which y should generate the soft distance. The following example illustrates this.

Example 7. *Let* $U = \{x, y_1, y_2, y_3\}$, $A = \{y_1, y_2, y_3\}$, $\delta = 0.1$ *and*

$$d(x,x) = 0, d(x,y_1) = 0.2, d(x,y_2) = 0.3, d(x,y_3) = 0.4.$$

Because $d(x,y_1) - \delta m_{x,y_1} = d(x,y_2) - \delta m_{x,y_2} = d(x,y_3) - \delta m_{x,y_3} = 0.1$, $SD(x,A)$ *could be either 0.2, 0.3 or 0.4.*

Based on the soft distance, the authors define the SFRS model with a distance function d determined by $d(x,y) = 1 - R(y,x)$ for all $x, y \in U$. However, since the use of the arg function leads to ambiguity as illustrated above, we introduce a slightly adapted definition of the model.

Definition 7. *Given an implicator* $\mathscr{I}$*, a conjunctor* $\mathscr{C}$ *and a penalty factor* $\delta > 0$*. The* soft fuzzy rough approximation *of A by R is the pair of fuzzy sets* $(\underline{\text{apr}}_R^{\mathscr{I},\delta}(A), \overline{\text{apr}}_R^{\mathscr{C},\delta}(A))$ *defined by, for x in U,*

$$\begin{aligned}
(\underline{\text{apr}}_R^{\mathscr{I},\delta}(A))(x) &= \mathscr{N}_{\mathscr{I}}\left(\min_{y\in\Omega_x} R(y,x)\right),\\
(\overline{\text{apr}}_R^{\mathscr{C},\delta}(A))(x) &= \min_{y\in\Pi_x} R(y,x),
\end{aligned}$$

with

$$\begin{aligned}
\Omega_x &= \big\{z \in U \mid A(z) \le \mu_x \text{ and}\\
&\quad (\forall z' \in U)(A(z') \le \mu_x \Rightarrow R(z',x) + \delta m_{x,z'} \ge R(z,x) + \delta m_{x,z})\big\},\\
\mu_x &= \max\{A(z) \mid z \in U \text{ and } \mathscr{I}(R(z,x),A(z)) = \sigma_x\},\\
\sigma_x &= \inf_{z\in U} \mathscr{I}(R(z,x),A(z)),\\
m_{x,z} &= |\{w \in U \mid A(w) \le \mu_x \text{ and } R(w,x) > R(z,x)\}|
\end{aligned}$$

and

$$\begin{aligned}
\Pi_x &= \big\{z \in U \mid A(z) \ge \nu_x \text{ and}\\
&\quad (\forall z' \in U)(A(z') \ge \nu_x \Rightarrow R(z',x) + \delta n_{x,z'} \ge R(z,x) + \delta n_{x,z})\big\},\\
\nu_x &= \min\{A(z) \mid z \in U \text{ and } \mathscr{C}(R(z,x),A(z)) = \tau_x\},\\
\tau_x &= \sup_{z\in U} \mathscr{C}(R(z,x),A(z)),\\
n_{x,z} &= |\{w \in U \mid A(w) \ge \nu_x \text{ and } R(w,x) > R(z,x)\}|.
\end{aligned}$$

We briefly explain the intuition behind the lower approximation; the explanation for the upper approximation is then analogous. First, we identify those elements z in U for which the value of the infimum considered in the IC lower approximation, $\sigma_x = (\underline{\mathrm{apr}}_R^{\mathscr{I}}(A))(x)$, is reached. As there may be several of them, we consider the one that has the highest membership to A; we denote this membership degree by μ_x. To obtain Ω_x, we look for z in U such that $A(z) \leq \mu_x$ and $R(z,x) + \delta m_{x,z}$ is minimal, where $m_{x,z}$ counts the number of elements w in U such that $A(w) \leq \mu_x$ and $R(w,x) > R(z,x)$. Among all the $y \in \Omega_x$, we choose the one that has the smallest value of $R(y,x)$. Finally, to compute the lower approximation in x, we take the negation of that value $R(y,x)$.

By tuning the penalty factor δ, we may allow for more or less noise tolerance. If δ is sufficiently large, the result will be determined by the largest value of $R(z,x)$ (or equivalently, the smallest value of $m_{x,z}$) among the considered z, since other elements will not satisfy the minimality condition. For small values of δ, more noise tolerance is allowed. In this case, the largest values of $R(z,x)$ are overlooked, and consequently the membership degree to the lower approximation gets larger.

Next, we discuss the properties of the SFRS model. When $\mathscr{N}_{\mathscr{I}}$ is involutive, (D) holds.

Proposition 15. *Let $\mathscr{I}$ be an implicator and $\mathscr{C}$ the conjunctor induced by $\mathscr{I}$ and $\mathscr{N}_{\mathscr{I}}$. If $\mathscr{N}_{\mathscr{I}}$ is involutive, the SFRS model satisfies (D) with respect to $\mathscr{N}_{\mathscr{I}}$.*

Proof. Let $x \in U$. Note that since $\mathscr{N}_{\mathscr{I}}$ is involutive, it is continuous (see [2], Corollary 1.4.6). We compute $(\overline{\mathrm{apr}}_R^{\mathscr{C},\delta}(A^{\mathscr{N}_{\mathscr{I}}}))^{\mathscr{N}_{\mathscr{I}}}(x)$:

$$\begin{aligned}
\tau_x^{A^{\mathscr{N}_{\mathscr{I}}}} &= \sup_{z\in U} \mathscr{C}(R(z,x), \mathscr{N}_{\mathscr{I}}(A(z))) \\
&= \sup_{z\in U} \mathscr{N}_{\mathscr{I}}(\mathscr{I}(R(z,x), A(z))) \\
&= \mathscr{N}_{\mathscr{I}}\left(\inf_{z\in U} \mathscr{I}(R(z,x), A(z))\right) \\
&= \mathscr{N}_{\mathscr{I}}(\sigma_x^A), \\
\nu_x^{A^{\mathscr{N}_{\mathscr{I}}}} &= \min\left\{\mathscr{N}_{\mathscr{I}}(A(z)) \mid z \in U \text{ and } \mathscr{C}(R(z,x), \mathscr{N}_{\mathscr{I}}(A(z))) = \tau_x^{A^{\mathscr{N}_{\mathscr{I}}}}\right\} \\
&= \min\left\{\mathscr{N}_{\mathscr{I}}(A(z)) \mid z \in U \text{ and } \mathscr{N}_{\mathscr{I}}(\mathscr{I}(R(z,x), A(z))) = \mathscr{N}_{\mathscr{I}}(\sigma_x^A)\right\} \\
&= \mathscr{N}_{\mathscr{I}}\left(\max\{A(z) \mid z \in U \text{ and } \mathscr{I}(R(z,x), A(z)) = \sigma_x^A\}\right) \\
&= \mathscr{N}_{\mathscr{I}}(\mu_x^A).
\end{aligned}$$

Furthermore, we have for every $z \in U$ that

$$\begin{aligned}
n_{x,z}^{A^{\mathscr{N}_{\mathscr{I}}}} &= |\{w \in U \mid \mathscr{N}_{\mathscr{I}}(A(w)) \geq \nu_x^{A^{\mathscr{N}_{\mathscr{I}}}} \wedge R(w,x) > R(z,x)\}| \\
&= |\{w \in U \mid \mathscr{N}_{\mathscr{I}}(A(w)) \geq \mathscr{N}_{\mathscr{I}}(\mu_x^A) \wedge R(w,x) > R(z,x)\}| \\
&= |\{w \in U \mid A(w) \leq \mu_x^A \wedge R(w,x) > R(z,x)\}| \\
&= m_{x,z}^A.
\end{aligned}$$

Hence,

$$\begin{aligned}
\Pi_x^{A^{\mathscr{N}_\mathscr{I}}} &= \big\{z \in U \mid \mathscr{N}_\mathscr{I}(A(z)) \geq \nu_x^{A^{\mathscr{N}_\mathscr{I}}} \text{ and} \\
&\quad (\forall z' \in U)(\mathscr{N}_\mathscr{I}(A(z')) \geq \nu_x^{A^{\mathscr{N}_\mathscr{I}}} \Rightarrow R(z',x) + \delta n_{x,z'}^{A^{\mathscr{N}_\mathscr{I}}} \geq R(z,x) + \delta n_{x,z}^{A^{\mathscr{N}_\mathscr{I}}})\big\}, \\
&= \big\{z \in U \mid \mathscr{N}_\mathscr{I}(A(z)) \geq \mathscr{N}_\mathscr{I}(\mu_x^A) \text{ and} \\
&\quad (\forall z' \in U)(\mathscr{N}_\mathscr{I}(A(z')) \geq \mathscr{N}_\mathscr{I}(\mu_x^A) \Rightarrow R(z',x) + \delta m_{x,z'}^A \geq R(z,x) + \delta m_{x,z}^A)\big\}, \\
&= \big\{z \in U \mid A(z) \leq \mu_x^A \text{ and} \\
&\quad (\forall z' \in U)(A(z') \leq \mu_x^A \Rightarrow R(z',x) + \delta m_{x,z'}^A \geq R(z,x) + \delta m_{x,z}^A)\big\}, \\
&= \Omega_x^A.
\end{aligned}$$

We conclude that

$$\begin{aligned}
(\overline{\text{apr}}_R^{\mathscr{C},\delta}(A^{\mathscr{N}_\mathscr{I}}))^{\mathscr{N}_\mathscr{I}}(x) &= \mathscr{N}_\mathscr{I}\left(\min_{y \in \Pi_x^{A^{\mathscr{N}_\mathscr{I}}}} R(y,x)\right) \\
&= \mathscr{N}_\mathscr{I}\left(\min_{y \in \Omega_x^A} R(y,x)\right) \\
&= (\underline{\text{apr}}_R^{\mathscr{I},\delta}(A))(x).
\end{aligned}$$

□

In particular, Proposition 15 holds if the pair $(\mathscr{I},\mathscr{C})$ consists of an S-implicator based on a t-conorm and its $\mathscr{N}_\mathscr{I}$-dual t-norm, or if it consists of an IMTL-t-norm and its R-implicator.

Other properties do not hold for this model (see Appendix, Examples 11, 12, 13, 14 and 15). In [24], the authors claimed that (IU) holds for the Kleene-Dienes implicator, the minimum operator and a similarity relation. However, below we show that this claim is false.

Example 8. *Let $U = \{x,y,z\}$ and A,B fuzzy sets in U such that $A(x) = 0.4$, $A(y) = 0.8$, $A(z) = 0.2$ and $B(x) = 0.8$, $B(y) = 0.5$, $B(z) = 0.8$. Let $\mathscr{I}$ be the Kleene-Dienes implicator, $\mathscr{C}$ the minimum operator, R the similarity relation such that $R(x,y) = 0.8$, $R(x,z) = 0.4$, $R(y,z) = 0.4$ and $\delta = 0.1$.*

We obtain the following results:

$$\begin{aligned}
(\underline{\text{apr}}_R^{\mathscr{I},\delta}(A))(x) &= 1 - R(z,x) = 0.6, \\
(\underline{\text{apr}}_R^{\mathscr{I},\delta}(B))(x) &= 1 - R(y,x) = 0.2, \\
(\underline{\text{apr}}_R^{\mathscr{I},\delta}(A \cap B))(x) &= 1 - R(z,x) = 0.6.
\end{aligned}$$

We conclude that

$$\underline{\text{apr}}_R^{\mathscr{I},\delta}(A) \cap \underline{\text{apr}}_R^{\mathscr{I},\delta}(B) \neq \underline{\text{apr}}_R^{\mathscr{I},\delta}(A \cap B).$$

For the upper approximations we derive that

$$\begin{aligned}
(\overline{\text{apr}}_R^{\mathscr{C},\delta}(A))(x) &= R(y,x) = 0.8, \\
(\overline{\text{apr}}_R^{\mathscr{C},\delta}(B))(x) &= R(z,x) = 0.4, \\
(\overline{\text{apr}}_R^{\mathscr{C},\delta}(A \cup B))(x) &= R(z,x) = 0.4.
\end{aligned}$$

We conclude that

$$\overline{\text{apr}}_R^{\mathscr{C},\delta}(A) \cup \overline{\text{apr}}_R^{\mathscr{C},\delta}(B) \neq \overline{\text{apr}}_R^{\mathscr{C},\delta}(A \cup B).$$

To end this subsection, we study Yao et al.'s Variable Precision Fuzzy Rough Set model based on Fuzzy Granules.

4.1.4. Variable Precision Fuzzy Rough Set Model based on Fuzzy Granules

The last model we discuss based on frequency is proposed in 2014 by Yao et al. [65] and is a granule-based fuzzy rough set model instead of an element-based one. This means that the lower and upper approximation operators are defined by means of a union, resp. intersection, of certain elementary fuzzy sets which are called fuzzy granules. In this approach, called FG model henceforth, the authors work with a left-continuous t-norm $\mathscr{T}$, its $\mathscr{N}_S$-dual t-conorm $\mathscr{S}$ and a $\mathscr{T}$-similarity relation R.

The relationship between fuzzy granules and fuzzy rough set theory was discussed for the first time by Degang et al. in [11]. In traditional set theory, a set is determined by the elements it contains. In fuzzy set theory, a fuzzy point plays a similar role: a non-empty fuzzy set is the union of certain fuzzy points. Fuzzy granules can then be constructed around these fuzzy points.

Definition 8. *[11] Let $x \in U$, $\lambda \in (0,1]$. A* fuzzy point *is a fuzzy set x_λ in U defined by*

$$\forall y \in U : x_\lambda(y) = \begin{cases} \lambda & y = x \\ 0 & y \neq x. \end{cases}$$

Definition 9. *[11] Let $x \in U$, $\lambda \in [0,1]$ and R a binary fuzzy relation. Let $\mathscr{T}$ be a left-continuous t-norm and $\mathscr{S}$ its $\mathscr{N}_S$-dual t-conorm. The* fuzzy information granules *$[x_\lambda]_R^{\mathscr{T}}$ and $[x_\lambda]_R^{\mathscr{S}}$ are defined by:*

$$\begin{aligned} \forall y \in U : [x_\lambda]_R^{\mathscr{T}}(y) &= \mathscr{T}(R(y,x), \lambda), \\ \forall y \in U : [x_\lambda]_R^{\mathscr{S}}(y) &= \mathscr{S}(1 - R(y,x), 1 - \lambda) \\ &= 1 - \mathscr{T}(R(y,x), \lambda). \end{aligned}$$

Degang et al. proved that the IC approximation operators $\underline{\text{apr}}_R^{\mathscr{I}}(A)$ and $\overline{\text{apr}}_R^{\mathscr{C}}(A)$ can be expressed using fuzzy granules for $(\mathscr{I}, \mathscr{C}) = (\mathscr{I}_{\mathscr{T}}, \mathscr{T})$. Note that R needs to be $\mathscr{T}$-transitive for this proposition to hold.

Proposition 16. *[11, 65] Let $\mathscr{T}$ be a left-continuous t-norm, $\mathscr{S}$ its $\mathscr{N}_S$-dual t-conorm and R a fuzzy $\mathscr{T}$-similarity relation. Then it holds that*

$$\begin{aligned} \underline{\text{apr}}_R^{\mathscr{I}_{\mathscr{T}}}(A) &= \bigcup \left\{ [x_\lambda]_R^{\mathscr{T}} \mid x \in U, \lambda \in [0,1], [x_\lambda]_R^{\mathscr{T}} \subseteq A \right\}, \\ \overline{\text{apr}}_R^{\mathscr{T}}(A) &= \bigcap \left\{ [x_\lambda]_R^{\mathscr{S}} \mid x \in U, \lambda \in [0,1], A \subseteq [x_\lambda]_R^{\mathscr{S}} \right\}. \end{aligned}$$

Now, instead of considering the fuzzy granules $[x_\lambda]_R^{\mathscr{T}}$ and $[x_\lambda]_R^{\mathscr{S}}$ such that for all $y \in U$ $[x_\lambda]_R^{\mathscr{T}}(y) \leq A(y)$, resp. $A(y) \leq [x_\lambda]_R^{\mathscr{S}}(y)$, Yao et al. consider the fuzzy granules such that for "many" elements y these inequalities hold. This is controlled by a parameter $\gamma \leq 1$ that is recommended to be chosen very close to 1. Although in [65] the relation R is a $\mathscr{T}$-similarity relation, we consider the model for arbitrary binary fuzzy relations.

Definition 10. *[65] Let $\mathscr{T}$ be a left-continuous t-norm, $\mathscr{S}$ its $\mathscr{N}_S$-dual t-conorm and R a binary fuzzy relation. Let $\gamma \in [0,1]$. The* fuzzy rough approximation of A by R based on fuzzy granules *is the pair of fuzzy sets $(\underline{\mathrm{apr}}_R^{\mathscr{T},\gamma}(A), \overline{\mathrm{apr}}_R^{\mathscr{S},\gamma}(A))$ defined by, for $x \in U$,*

$$\begin{aligned}
(\underline{\mathrm{apr}}_R^{\mathscr{T},\gamma}(A))(x) &= \sup\left\{ [z_\lambda]_R^{\mathscr{T}}(x) \mid z \in U, \lambda \in [0,1], \frac{|\{y \in U \mid [z_\lambda]_R^{\mathscr{T}}(y) \leq A(y)\}|}{|U|} \geq \gamma \right\}, \\
(\overline{\mathrm{apr}}_R^{\mathscr{S},\gamma}(A))(x) &= \inf\left\{ [z_\lambda]_R^{\mathscr{S}}(x) \mid z \in U, \lambda \in [0,1], \frac{|\{y \in U \mid A(y) \leq [z_\lambda]_R^{\mathscr{S}}(y)\}|}{|U|} \geq \gamma \right\}.
\end{aligned}$$

If $\gamma = 1$ and R is a $\mathscr{T}$-similarity relation, this model coincides with the IC model for the couple $(\mathscr{I}_{\mathscr{T}}, \mathscr{T})$ as shown in Proposition 16. For general $\gamma \in [0,1]$, $\underline{\mathrm{apr}}_R^{\mathscr{I}_{\mathscr{T}}}(A) \subseteq \underline{\mathrm{apr}}_R^{\mathscr{T},\gamma}(A)$ and $\overline{\mathrm{apr}}_R^{\mathscr{S},\gamma}(A) \subseteq \overline{\mathrm{apr}}_R^{\mathscr{T}}(A)$.

In [65], another characterisation of the FG model was proposed. It is proven that for a $\mathscr{T}$-similarity relation R, the supremum, resp. the infimum, is reached in x, when computing the lower, resp. the upper approximation in x.

Proposition 17. *[65] Let $\mathscr{T}$ be a left-continuous t-norm, $\mathscr{S}$ its $\mathscr{N}_S$-dual t-conorm and R a fuzzy $\mathscr{T}$-similarity relation. Let $\gamma \in [0,1]$. It holds, for $x \in U$, that*

$$\begin{aligned}
(\underline{\mathrm{apr}}_R^{\mathscr{T},\gamma}(A))(x) &= \sup\left\{ [x_\lambda]_R^{\mathscr{T}}(x) \mid \lambda \in [0,1], \frac{|\{y \in U \mid [x_\lambda]_R^{\mathscr{T}}(y) \leq A(y)\}|}{|U|} \geq \gamma \right\}, \\
(\overline{\mathrm{apr}}_R^{\mathscr{S},\gamma}(A))(x) &= \inf\left\{ [x_\lambda]_R^{\mathscr{S}}(x) \mid \lambda \in [0,1], \frac{|\{y \in U \mid A(y) \leq [x_\lambda]_R^{\mathscr{S}}(y)\}|}{|U|} \geq \gamma \right\}.
\end{aligned}$$

Note that it suffices that R is reflexive and $\mathscr{T}$-transitive for Proposition 17 to hold.

We continue with discussing the properties of this model. As was proven in [65], (D) and (SM) hold.

Proposition 18. *[65] The FG model satisfies (D) with respect to the standard negator.*

Proposition 19. *[65] The FG model satisfies (SM).*

Although it was not mentioned in [65], (RM) holds under conditions:

Proposition 20. *Let $\mathscr{T}$ be a left-continuous t-norm and $\mathscr{S}$ its $\mathscr{N}_S$-dual t-conorm. If R_1 and R_2 are reflexive and $\mathscr{T}$-transitive fuzzy relations, then the FG model satisfies (RM).*

Proof. Assume $R_1 \subseteq R_2$ and $\gamma \in [0,1]$. We introduce the notation λ_x^i for $x \in U$ and $i = 1,2$:

$$\lambda_x^i = \sup\left\{ \lambda \in [0,1] \mid \frac{|\{y \in U \mid [x_\lambda]_{R_i}^{\mathscr{T}}(y) \leq A(y)\}|}{|U|} \geq \gamma \right\}.$$

Due to Proposition 17 and the left-continuity of $\mathscr{T}$, we have for $x \in U$ and $i = 1,2$:

$$\begin{aligned}
(\underline{\mathrm{apr}}_{R_i}^{\mathscr{T},\gamma}(A))(x) &= \sup\left\{ [x_\lambda]_{R_i}^{\mathscr{T}}(x) \mid \lambda \in [0,1], \frac{|\{y \in U \mid [x_\lambda]_{R_i}^{\mathscr{T}}(y) \leq A(y)\}|}{|U|} \geq \gamma \right\} \\
&= [x_{\lambda_x^i}]_{R_i}^{\mathscr{T}}(x) \\
&= \mathscr{T}(R_i(x,x), \lambda_x^i) \\
&= \lambda_x^i.
\end{aligned}$$

Now, since $R_1 \subseteq R_2$, it holds for all $x \in U$ and $\lambda \in [0,1]$ that

$$[x_\lambda]^{\mathscr{T}}_{R_1} \subseteq [x_\lambda]^{\mathscr{T}}_{R_2}.$$

Hence $\lambda^2_x \leq \lambda^1_x$ for all $x \in U$ and thus,

$$\underline{\text{apr}}^{\mathscr{T},\gamma}_{R_2}(A) \subseteq \underline{\text{apr}}^{\mathscr{T},\gamma}_{R_1}(A).$$

By duality, it also holds that

$$\overline{\text{apr}}^{\mathscr{S},\gamma}_{R_1}(A) \subseteq \overline{\text{apr}}^{\mathscr{S},\gamma}_{R_2}(A).$$

Hence, the FG model satisfies (RM). □

Other properties do not hold for the FG model (see Appendix, Examples 16, 17 and 18).

Next, we discuss the Fuzzy Variable Precision Rough Set model proposed by Zhao et al., where a level of uncertainty is introduced in the approximated set.

4.2. Fuzzy Variable Precision Rough Set Model

Another model designed to make approximation operators more robust, is the Fuzzy Variable Precision Rough Set (FVPRS) model, proposed by Zhao et al. [69] in 2009. It introduces a level α of uncertainty into the IC model. Below, we recall the definition of the approximation operators in this model[4].

Definition 11. *[69] Given an implicator $\mathscr{I}$, a conjunctor $\mathscr{C}$, an involutive negator $\mathscr{N}$, a disjunctor $\mathscr{D}$, its $\mathscr{N}$-dual conjunctor $\mathscr{C}_{\mathscr{D},\mathscr{N}}$ and $\alpha \in [0,1)$. The* α-variable precision fuzzy rough approximation *of A by R is the pair of fuzzy sets $(\underline{\text{apr}}^{\mathscr{I},\alpha}_R(A), \overline{\text{apr}}^{\mathscr{C},\alpha}_R(A))$ defined by, for $x \in U$:*

$$\begin{aligned}(\underline{\text{apr}}^{\mathscr{I},\alpha}_R(A))(x) &= \inf_{y\in U} \mathscr{I}(R(y,x), \mathscr{D}(\alpha, A(y))),\\ (\overline{\text{apr}}^{\mathscr{C},\alpha}_R(A))(x) &= \sup_{y\in U} \mathscr{C}(R(y,x), \mathscr{C}_{\mathscr{D},\mathscr{N}}(\mathscr{N}(\alpha), A(y))).\end{aligned}$$

In this way, elements with very small membership degrees to A are smoothed with an uncertainty level α to limit their impact on the lower approximation, while the opposite happens for the upper approximation. For this reason, it is clear that $\underline{\text{apr}}^{\mathscr{I}}_R(A) \subseteq \underline{\text{apr}}^{\mathscr{I},\alpha}_R(A)$ and $\overline{\text{apr}}^{\mathscr{C},\alpha}_R(A) \subseteq \overline{\text{apr}}^{\mathscr{C}}_R(A)$ always hold.

In general, α will be chosen close to 0. Note that if α is equal to 0 and $\mathscr{D}$ is a border disjunctor, we derive the IC model as a special case of the FVPRS model.

On the other hand, we can also interpret the FVPRS model as a specific instance of the IC model. Indeed, if we define the fuzzy sets B_1 and B_2 in U by, for $x \in U$,

$$\begin{aligned}B_1(x) &= \mathscr{D}(\alpha, A(x)),\\ B_2(x) &= \mathscr{C}_{\mathscr{D},\mathscr{N}}(\mathscr{N}(\alpha), A(x)),\end{aligned}$$

then it is clear that $\underline{\text{apr}}^{\mathscr{I},\alpha}_R(A) = \underline{\text{apr}}^{\mathscr{I}}_R(B_1)$ and $\overline{\text{apr}}^{\mathscr{C},\alpha}_R(A) = \overline{\text{apr}}^{\mathscr{C}}_R(B_2)$.

Based on this relationship, it is now easy to see that (SM), (RM) and (IU) always hold for the FVPRS model, and that (D) and (ID) hold under the same conditions as for the IC model.

[4]We note that in [69] the following logical connectives are used: for a lower semicontinuous t-norm $\mathscr{T}$ and a involutive negator $\mathscr{N}$, $(\mathscr{I},\mathscr{C})$ is either $(\mathscr{I}_{\mathscr{T}}, \mathscr{C}_{\mathscr{I},\mathscr{N}})$ or $(\mathscr{I}_{\mathscr{S}}, \mathscr{T})$ with $\mathscr{S}$ the $\mathscr{N}$-dual t-conorm of $\mathscr{T}$. For the disjunctor $\mathscr{D}$, the maximum operator is used.

Proposition 21. *The FVPRS model satisfies (SM), (RM) and (IU).*

Proposition 22. *If $\mathscr{C}$ is the induced conjunctor of $\mathscr{I}$ and $\mathscr{N}$, i.e., $\mathscr{C} = \mathscr{C}_{\mathscr{I},\mathscr{N}}$, then the FVPRS model satisfies (D).*

We remark that a restricted version of Propositions 21 and 22 was proven in [69]. Note also that we do not have to assume that $\mathscr{C}_{\mathscr{D},\mathscr{N}}$ and $\mathscr{C}_{\mathscr{I},\mathscr{N}}$ are the same conjunctors.

Proposition 23. *If R is a $\mathscr{T}$-transitive relation, where $\mathscr{T}$ is a left-continuous t-norm and the pair $(\mathscr{I},\mathscr{C})$ consists of the t-norm $\mathscr{T}$ and its R-implicator, or $(\mathscr{I},\mathscr{C})$ consists of the S-implicator based on the $\mathscr{N}$-dual t-conorm of $\mathscr{T}$ with respect to an involutive negator $\mathscr{N}$ and the t-norm $\mathscr{T}$, then (ID) holds.*

(A) holds under certain conditions, which are stricter than for the IC model.

Proposition 24. *Let $(\mathscr{I},\mathscr{C})$ be a pair consisting of an IMTL-t-norm $\mathscr{T}$ and its R-implicator $\mathscr{I}$ and let $\mathscr{D} = \mathscr{S}$ be the $\mathscr{N}_{\mathscr{I}}$-dual of $\mathscr{T}$, then (A) holds.*

Proof. For the sake of notation, we write $\mathscr{I}$ instead of $\mathscr{I}_{\mathscr{T}}$. First note that (see e.g. [56]):

$$\forall a,b \in [0,1]\colon \mathscr{I}(\mathscr{N}_{\mathscr{I}}(a),b) = \mathscr{N}_{\mathscr{I}}(\mathscr{T}(\mathscr{N}_{\mathscr{I}}(a),\mathscr{N}_{\mathscr{I}}(b))) = \mathscr{S}(a,b).$$

Take $x,y \in U$ and let A and B be fuzzy sets in U. Then it holds for all $x,y \in U$ that

$$\begin{aligned}
\mathscr{T}(R(y,x),\mathscr{T}(\mathscr{N}_{\mathscr{I}}(\alpha),A(y))) \leq B(x) \quad &\Leftrightarrow \quad \mathscr{T}(\mathscr{T}(R(y,x),\mathscr{N}_{\mathscr{I}}(\alpha)),A(y)) \leq B(x) \\
&\Leftrightarrow \quad A(y) \leq \mathscr{I}(\mathscr{T}(R(y,x),\mathscr{N}_{\mathscr{I}}(\alpha)),B(x)) \\
&\Leftrightarrow \quad A(y) \leq \mathscr{I}(R(y,x),\mathscr{I}(\mathscr{N}_{\mathscr{I}}(\alpha),B(x))) \\
&\Leftrightarrow \quad A(y) \leq \mathscr{I}(R'(x,y),\mathscr{S}(\alpha,B(x))),
\end{aligned}$$

where we first use the associativity of $\mathscr{T}$, then the residuation condition between $\mathscr{T}$ and $\mathscr{I}$ and in the last step a property of $\mathscr{T}$ and $\mathscr{I}$ that is due to the fact that $\mathscr{T}$ is an IMTL-t-norm (see [56], Proposition 13). In other words, $\overline{\mathrm{apr}}_R^{\mathscr{T},\alpha}(A) \subseteq B \Leftrightarrow A \subseteq \underline{\mathrm{apr}}_R^{\mathscr{I}_{\mathscr{T}},\alpha}(B)$. □

The properties (INC), (LU), (CS) and (UE) do not hold for the FVPRS model (see Appendix, Examples 19 and 20).

In the last subsection, we discuss robust fuzzy rough set models which make use of other aggregation operators then the infimum and supremum operators.

4.3. Noise-Tolerant Models based on Aggregation Operators

A drawback of the IC model is its use of the infimum and supremum operators. The lower, respectively upper, approximation is fully determined by the worst, respectively best, value. To overcome this problem, fuzzy rough set models based on other aggregation operators have been defined. Models of this type are the β-Precision Fuzzy Rough Set model of Fernández-Salido and Murakami [18] and the Ordered Weighted Average based Fuzzy Rough Set model proposed by Cornelis et al. [7]. We start with the former one.

4.3.1. β-Precision Fuzzy Rough Set Model

The oldest noise-tolerant fuzzy rough set model is due to Fernández-Salido and Murakami [18], who proposed the β-Precision Fuzzy Rough Set (β-PREC) model in 1993. They tackle the noise problem by replacing the infimum and supremum operators by less strict aggregation operators, which are recalled first.

Definition 12. *[17, 18] Given a t-norm $\mathscr{T}$, a t-conorm $\mathscr{S}$, $\beta \in [0,1]$ and $n \in \mathbb{N} \setminus \{0,1\}$, the corresponding β-*precision quasi-t-norm *$\mathscr{T}_\beta$ and β-*precision quasi-t-conorm *$\mathscr{S}_\beta$ of order n are $[0,1]^n \to [0,1]$ mappings such that for all $\boldsymbol{x} = (x_1,\ldots,x_n)$ in $[0,1]^n$,*

$$\begin{aligned} \mathscr{T}_\beta(\boldsymbol{x}) &= \mathscr{T}(y_1,\ldots,y_{n-m}), \\ \mathscr{S}_\beta(\boldsymbol{x}) &= \mathscr{S}(z_1,\ldots,z_{n-p}), \end{aligned}$$

where y_i is the i^{th} greatest element of $\boldsymbol{x}$ and z_i is the i^{th} smallest element of $\boldsymbol{x}$, and

$$\begin{aligned} m &= \max\left\{ i \in \{0,\ldots,n\} \mid i \leq (1-\beta)\sum_{j=1}^{n} x_j \right\}, \\ p &= \max\left\{ i \in \{0,\ldots,n\} \mid i \leq (1-\beta)\sum_{j=1}^{n} (1-x_j) \right\}. \end{aligned}$$

When $\beta = 1$, we obtain the original t-norm $\mathscr{T}$ and t-conorm $\mathscr{S}$. The β-PREC model is then defined as follows:

Definition 13. *[18] Let $\mathscr{T}$ be a t-norm, $\mathscr{S}$ a t-conorm and $\beta \in [0,1]$. Given an implicator $\mathscr{I}$ and a conjunctor $\mathscr{C}$, the* β-precision fuzzy rough approximation *of A by R is the pair of fuzzy sets $(\underline{\text{apr}}_R^{\mathscr{I},\mathscr{T}_\beta}(A), \overline{\text{apr}}_R^{\mathscr{C},\mathscr{S}_\beta}(A))$, defined by, for $x \in U$:*

$$\begin{aligned} (\underline{\text{apr}}_R^{\mathscr{I},\mathscr{T}_\beta}(A))(x) &= \mathop{\mathscr{T}_\beta}_{y\in U} \langle \mathscr{I}(R(y,x),A(y)) \rangle, \\ (\overline{\text{apr}}_R^{\mathscr{C},\mathscr{S}_\beta}(A))(x) &= \mathop{\mathscr{S}_\beta}_{y\in U} \langle \mathscr{C}(R(y,x),A(y)) \rangle. \end{aligned}$$

Controlled by β, the smallest elements are omitted in the calculation of the lower approximation. Analogously, the largest elements will not influence the upper approximation. Therefore, when $\mathscr{T} = \mathscr{T}_M$ and $\mathscr{S} = \mathscr{S}_M$, the approximation operators of the β-PREC model satisfy $\underline{\text{apr}}_R^{\mathscr{I}}(A) \subseteq \underline{\text{apr}}_R^{\mathscr{I},\mathscr{T}_\beta}(A)$ and $\overline{\text{apr}}_R^{\mathscr{T},\mathscr{S}_\beta}(A) \subseteq \overline{\text{apr}}_R^{\mathscr{T}}(A)$. On the other hand, the use of t-norms and t-conorms other than the minimum and maximum operator allows for more interaction among the arguments to be aggregated. When $\beta = 0$, all elements smaller than the average $\mathscr{I}(R(y,x),A(y))$ value (resp., higher than the average $\mathscr{C}(R(y,x),A(y))$ value) are ignored. If $\beta = 1$, $\mathscr{T} = \mathscr{T}_M$ and $\mathscr{S} = \mathscr{S}_M$, we derive the IC model. Fernández-Salido and Murakami recommended to choose β very close to 1.

Below, we evaluate the properties of the β-PREC model.

Proposition 25. *Let $\mathscr{T}$ be a t-norm and and $\mathscr{S}$ its $\mathscr{N}_S$-dual t-conorm. Let $\beta \in [0,1]$. If the pair $(\mathscr{I},\mathscr{C})$ consists of an implicator $\mathscr{I}$ and the conjunctor induced by $\mathscr{I}$ and $\mathscr{N}_S$, i.e., $\mathscr{C} = \mathscr{C}_{\mathscr{I},\mathscr{N}_S}$, then (D) holds.*

Proof. We will prove that $\mathscr{T}_\beta$ and $\mathscr{S}_\beta$ are dual with respect to the standard negator $\mathscr{N}_S$. Let us take $(x_1,\ldots,x_n) \in [0,1]^n$ and σ the permutation on $\{1,\ldots,n\}$ such that $x_{\sigma(i)}$ is the i^{th} biggest element of $(x_1,\ldots,x_n)$. Define m as:

$$m = \max\left\{ i \in \{0,\ldots,n\} \mid i \leq (1-\beta)\cdot \sum_{j=1}^{n} x_j \right\}.$$

Now, since

$$m \leq (1-\beta)\cdot \sum_{j=1}^{n} x_j \Leftrightarrow m \leq (1-\beta)\cdot \sum_{j=1}^{n} 1-(1-x_j),$$

we omit m values to calculate $\mathscr{T}_\beta$ and we omit m values to calculate $\mathscr{S}_\beta$. Hence,

$$\begin{aligned}
\mathscr{N}_S(\mathscr{T}_\beta(x_1,\ldots,x_n)) &= 1-\left(\mathscr{T}(x_{\sigma(1)},\ldots,x_{\sigma(n-m)})\right)\\
&= \mathscr{S}(1-x_{\sigma(1)},\ldots,1-x_{\sigma(n-m)})\\
&= \mathscr{S}_\beta(1-x_1,\ldots,1-x_n)\\
&= \mathscr{S}_\beta(\mathscr{N}_S(x_1),\ldots,\mathscr{N}_S(x_n)).
\end{aligned}$$

The rest of the proof now follows from Proposition 1. □

In particular, Proposition 25 holds if the pair $(\mathscr{I},\mathscr{C})$ consists of an S-implicator $\mathscr{I}_{\mathscr{S},\mathscr{N}_S}$ and the $\mathscr{N}_S$-dual t-norm of $\mathscr{S}$, or if it consists of an IMTL-t-norm $\mathscr{T}$ and its R-implicator, assuming that the induced negator of the R-implicator is the standard negator, i.e., $\mathscr{N}_{\mathscr{I}_\mathscr{T}} = \mathscr{N}_S$.

We note that duality only holds with respect to the standard negator $\mathscr{N}_S$. Indeed, if $\mathscr{S}$ is the $\mathscr{N}$-dual t-conorm of a t-norm $\mathscr{T}$ for an involutive negator $\mathscr{N} \neq \mathscr{N}_S$, then $\mathscr{N}(\mathscr{T}_\beta(x_1,\ldots,x_n))$ is not necessarily equal to $\mathscr{S}_\beta(\mathscr{N}(x_1),\ldots,\mathscr{N}(x_n))$.

Proposition 26. *The β-PREC model satisfies (SM) and (RM).*

Proof. This follows from the monotonicity of conjunctors, implicators and β-precision t-norms and t-conorms. □

None of the other properties holds for this model (see Appendix: Examples 21, 22 and 23).

To end this subsection, we discuss Cornelis et al.'s Ordered Weighted Average based Fuzzy Rough Set model.

4.3.2. Ordered Weighted Average based Fuzzy Rough Set Model

In 2010, Cornelis et al. [7] constructed a fuzzy rough set model in which they replaced the infimum and supremum operators of the IC model by Ordered Weighted Average (OWA) aggregation operators, which are recalled first:

Definition 14. *[63] Given a sequence D of n scalar values and a weight vector $W = \langle w_1,\ldots,w_n\rangle$ of length n (henceforth called an OWA weight vector of length n), such that for all $i \in \{1,\ldots,n\}$, $w_i \in [0,1]$, and $\sum_{i=1}^{n} w_i = 1$. Let σ be the permutation on $\{1,\ldots,n\}$ such that $d_{\sigma(i)}$ is the i^{th} largest value of D. The* OWA operator acting on *D yields the value*

$$\mathrm{OWA}_W(D) \quad = \quad \sum_{i=1}^{n} w_i d_{\sigma(i)}.$$

The OWA operator allows to consider a wide variety of aggregation strategies. For instance, the weight vectors $\langle 0,\ldots,0,1\rangle$ and $\langle 1,0,\ldots,0\rangle$ represent the minimum and the maximum, respectively. Using other OWA weight vectors, it is possible to soften the lower and upper approximation such that not only the smallest, resp. largest element contribute to the result. In order to assure that the obtained operators still behave sufficiently like the minimum and the maximum, the orness and andness degree of the weight vector are used.

Definition 15. *[63] Let W be an OWA weight vector of length n. The* orness *and* andness degree *of W are defined by*

$$\begin{aligned} \text{orness}(W) &= \frac{1}{n-1}\sum_{i=1}^{n}((n-i)\cdot w_i), \\ \text{andness}(W) &= 1-\text{orness}(W). \end{aligned}$$

Definition 16. *[7] Given an implicator $\mathscr{I}$, a conjunctor $\mathscr{C}$ and OWA weight vectors W_1 and W_2 of length n, with $n=|U|$ and such that* andness$(W_1)>0.5$ *and* orness$(W_2)>0.5$*, the* (W_1,W_2)-fuzzy rough approximation *of A by R is the pair of fuzzy sets $(\underline{\text{apr}}_R^{\mathscr{I},W_1}(A),\overline{\text{apr}}_R^{\mathscr{C},W_2}(A))$ defined by, for $x\in U$:*

$$\begin{aligned} (\underline{\text{apr}}_R^{\mathscr{I},W_1}(A))(x) &= \underset{y\in U}{\text{OWA}_{W_1}}\ \langle \mathscr{I}(R(y,x),A(y))\rangle, \\ (\overline{\text{apr}}_R^{\mathscr{C},W_2}(A))(x) &= \underset{y\in U}{\text{OWA}_{W_2}}\ \langle \mathscr{C}(R(y,x),A(y))\rangle. \end{aligned}$$

By varying the OWA weight vectors, different fuzzy rough set models can be maintained. Clearly, for the weight vectors $W_1=\langle 0,\ldots,0,1\rangle$ and $W_2=\langle 1,0,\ldots,0\rangle$, we obtain the IC model. If other OWA weight vectors are used, more weight will be given to higher, resp. lower values, so it always holds that $\underline{\text{apr}}_R^{\mathscr{I}}(A)\subseteq\underline{\text{apr}}_R^{\mathscr{I},W_1}(A)$ and $\overline{\text{apr}}_R^{\mathscr{C},W_2}(A)\subseteq\overline{\text{apr}}_R^{\mathscr{C}}(A)$.

We note that the OWA model also has the robust nearest neighbor fuzzy rough approximation operators, proposed in 2012 by Hu et al. [27], as special cases. For instance, with the *k-trimmed minimum* operator, the authors consider a variation on the IC lower approximation, in which the k smallest implication values $\mathscr{I}(R(y,x),A(y))$ are omitted in the infimum computation; clearly, this approach can be modeled by using the OWA weight vector in which the element on position $n-k$ is equal to 1 and the remaining values are 0. In Table 4, we show for each robust nearest neighbor fuzzy rough approximation proposed in [27] its corresponding OWA weight vector. Note that the first three correspond to lower approximations, and the last three to upper approximations.

Next, we consider the properties of the OWA model. Duality holds under conditions on the weight vectors W_1 and W_2. Let W_1 be a weight vector such that andness$(W_1)>0.5$, then it holds that

$$1-\frac{n}{n-1}\cdot 1+\frac{1}{n-1}\sum_{i=1}^{n} i\cdot(W_1)_i>0.5$$

and thus

$$\frac{1}{n-1}\sum_{i=1}^{n} i\cdot(W_1)_i>\frac{n}{n-1}-0.5.$$

Operator	OWA weight vector
k-trimmed minimum	$w_i = \begin{cases} 1 & \text{if } i = n-k \\ 0 & \text{otherwise} \end{cases}$
k-mean minimum	$w_i = \begin{cases} \frac{1}{k} & \text{if } i > n-k \\ 0 & \text{otherwise} \end{cases}$
k-median minimum	$w_i = \begin{cases} 1 & \text{if } k \text{ odd, } i = n-\frac{k-1}{2} \\ \frac{1}{2} & \text{if } k \text{ even, } i = n-\frac{k}{2} \text{ or } i = n-\frac{k-2}{2} \\ 0 & \text{otherwise} \end{cases}$
k-trimmed maximum	$w_i = \begin{cases} 1 & \text{if } i = k+1 \\ 0 & \text{otherwise} \end{cases}$
k-mean maximum	$w_i = \begin{cases} \frac{1}{k} & \text{if } i < k+1 \\ 0 & \text{otherwise} \end{cases}$
k-median maximum	$w_i = \begin{cases} 1 & \text{if } k \text{ odd, } i = \frac{k+1}{2} \\ \frac{1}{2} & \text{if } k \text{ even, } i = \frac{k}{2} \text{ or } i = \frac{k}{2}+1 \\ 0 & \text{otherwise} \end{cases}$

Table 4: Correspondence between robust nearest neighbor fuzzy rough approximation operators and OWA weight vectors

If we define W_2 as $(W_2)_i = (W_1)_{n-i+1}$ for $i \in \{1,\ldots,n\}$, then it holds that

$$\begin{aligned}
\text{orness}(W_2) &= \frac{1}{n-1}\sum_{i=1}^{n}(n-i).(W_2)_i \\
&= \frac{1}{n-1}\sum_{i=1}^{n}(n-i).(W_1)_{n-i+1} \\
&= \frac{1}{n-1}\sum_{j=1}^{n}(n-(n-j+1)).(W_1)_j \\
&= \frac{1}{n-1}\sum_{j=1}^{n}(j-1).(W_1)_j \\
&= \frac{1}{n-1}\sum_{j=1}^{n}j.(W_1)_j - \frac{1}{n-1} \\
&> \frac{n}{n-1} - 0.5 - \frac{1}{n-1} \\
&= 1-0.5 \\
&= 0.5.
\end{aligned}$$

Hence, if $\text{andness}(W_1) > 0.5$, then $\text{orness}(W_2) > 0.5$. We can now prove the duality property.

Proposition 27. *Let W_1 be a weight vector such that* $\text{andness}(W_1) > 0.5$. *If the pair $(\mathscr{I},\mathscr{C})$ consists of an implicator $\mathscr{I}$ and the conjunctor induced by $\mathscr{I}$ and $\mathscr{N}_S$, i.e., $\mathscr{C} = \mathscr{C}_{\mathscr{I},\mathscr{N}_S}$, and W_2 is the weight vector defined by $(W_2)_i = (W_1)_{n-i+1}$ for $i \in \{1,\ldots,n\}$ then (D) holds with respect to $\mathscr{N}_S$.*

Proof. Let $x \in U$. Since U is finite, we can rename the elements of U such that $U = \{z_1,\ldots,z_n\}$

and

$$\mathscr{I}(R(z_1,x),A(z_1)) \leq \ldots \leq \mathscr{I}(R(z_n,x),A(z_n)).$$

As $\mathscr{C} = \mathscr{C}_{\mathscr{I},\mathscr{N}_S}$, it holds for every $z_i \in U$ that

$$1 - \mathscr{C}(R(z_i,x), 1 - A(z_i)) = \mathscr{I}(R(z_i,x),A(z_i)).$$

Hence,

$$\mathscr{C}(R(z_1,x), 1 - A(z_1)) \geq \ldots \geq \mathscr{C}(R(z_n,x), 1 - A(z_n)).$$

We now compute $(\overline{\text{apr}}_R^{\mathscr{C},W_2}(A^{\mathscr{N}_S}))^{\mathscr{N}_S}(x)$:

$$\begin{aligned}
(\overline{\text{apr}}_R^{\mathscr{C},W_2}(A^{\mathscr{N}_S}))^{\mathscr{N}_S}(x) &= 1 - \sum_{i=1}^{n} (W_2)_i \cdot \mathscr{C}(R(z_i,x), 1 - A(z_i)) \\
&= 1 - \sum_{i=1}^{n} (W_1)_{n-i+1} \cdot (1 - \mathscr{I}(R(z_i,x),A(z_i))) \\
&= 1 - \left(1 - \sum_{i=1}^{n} (W_1)_{n-i+1} \cdot \mathscr{I}(R(z_i,x),A(z_i))\right) \\
&= (W_1)_n \cdot \mathscr{I}(R(z_1,x),A(z_1)) + \ldots + (W_1)_1 \cdot \mathscr{I}(R(z_n,x),A(z_n)) \\
&= (\underline{\text{apr}}_R^{\mathscr{I},W_1}(A))(x),
\end{aligned}$$

where the last step holds, since $\mathscr{I}(R(z_1,x),A(z_1)) \leq \ldots \leq \mathscr{I}(R(z_n,x),A(z_n))$. □

In particular, Proposition 27 holds if the pair $(\mathscr{I},\mathscr{C})$ consists of an S-implicator $\mathscr{I}_{\mathscr{S},\mathscr{N}_S}$ and the $\mathscr{N}_S$-dual t-norm of $\mathscr{S}$, or if it consists of an IMTL-t-norm $\mathscr{T}$ and its R-implicator, assuming that the induced negator of the R-implicator is the standard negator, i.e., $\mathscr{N}_{\mathscr{I}_\mathscr{T}} = \mathscr{N}_S$.

Set and relation monotonicity also hold for this model, as was stated in [7].

Proposition 28. *[7] The OWA model satisfies (SM) and (RM).*

None of the remaining properties is satisfied (see Appendix, Examples 24, 25 and 26).

4.4. Summary

We summarize the properties of the IC model discussed in Section 3 and the robust fuzzy rough set models discussed in Section 4 in Table 5. If a property holds, we denote this with ✓; if a property does not hold, we indicate this by ✗ and if a property holds under certain conditions, we write ☆.

The most important conclusion seems to be that none of the noise-tolerant models is able to retain all the properties of Pawlak's original model and its implicator-conjunctor-based extension, even when R is a similarity relation. FVPRS is the model that satisfies most theoretical properties; incidentally, this is also the model that adheres most closely to the IC model.

Note that none of the noise-tolerant models satisfies (INC), a defect they share with the crisp VPRS model. This is quite remarkable, since intuitively one would expect the lower and upper approximations to be positioned on either side of the fuzzy set they are supposed to approximate.

Furthermore, the properties (SM) and (RM) are important from the application perspective. The fact that SFRS does not satisfy (SM), not even for crisp subsets of U (see Appendix, Example 13), makes its practical application very problematic; for instance, in a classification problem,

Property	IC	VPFRS	VQFRS	SFRS	FG	FVPRS	β-PREC	OWA
(D)	☆	✗	✗	☆	☆	☆	☆	☆
(A)	☆	✗	✗	✗	✗	☆	✗	✗
(INC)	☆	✗	✗	✗	✗	✗	✗	✗
(SM)	✓	✓	✓	✗	✓	✓	✓	✓
(RM)	✓	✗	✗	✗	☆	✓	✓	✓
(IU)	✓	✗	✗	✗	✗	✓	✗	✗
(ID)	☆	✗	✗	✗	✗	☆	✗	✗
(LU)	☆	✗	✗	✗	✗	✗	✗	✗
(CS)	☆	✗	✗	✗	✗	✗	✗	✗
(UE)	☆	✗	☆	✗	✗	✗	✗	✗

Table 5: Evaluation of properties for different fuzzy rough set models

one would expect that when a decision class gets larger, so do its approximations. On the other hand, the violation of (RM) has its own consequences for applications, like attribute selection, that consider different levels of granulation of data: indeed, when the granulation of the data imposed by the fuzzy relations R becomes finer, we expect the lower approximation not to shrink. Some attribute selection methods even rely on (RM), e.g., the QuickReduct algorithm [6] assumes that the lower approximation of classes increases when attributes are added, since adding attributes means that the membership values of the fuzzy relation increase. If the (RM) property is not satisfied, the QuickReduct algorithm could get stuck in an infinite loop. In this sense, the fuzzy rough set models of FG, FVPRS, β-PREC and OWA are prefered over the other ones.

The other seven properties we have discussed are less important for practical applications, and are mainly interesting from a theoretical perspective. If a pair of operators satisfies (D), then there is an interaction between the lower and upper approximation operator: one can be defined by the other. The property (A) comes from the concept of a Galois connection. However, in many data analysis applications only the lower approximation is used and thus, the properties (D) and (A) are less important from a practical point of view. The property (IU) is very intuitive, since it states that the membership degrees of an element to the intersection $A \cap B$ and the union $A \cup B$ are fully determined by its membership degrees to A and B. The properties (ID) and (LU) express that approximating once gives immediately all the information. Furthermore, the property (CS) expresses that no extra information is gained by approximating a constant set. Moreover, the property (UE) states that the universe contains all information, and that the empty set contains none.

In the following section, we empirically analyze the robustness of the different models considered in this paper.

5. Robustness of Fuzzy Rough Set Models: Experimental Evaluation

To evaluate the robustness of the previously discussed fuzzy rough set models, we set up an experiment involving four real-world data sets. Each data set can be considered as a decision system $(U, \mathscr{A} \cup \{d\})$ where U is the finite set of instances that are described by the attributes in $\mathscr{A}$ (also called conditional attributes) and one decision attribute $d \notin \mathscr{A}$. The value of an instance $x \in U$ for a certain attribute $a \in \mathscr{A} \cup \{d\}$ is denoted by $a(x)$.

A common task in machine learning is to predict the value of the decision attribute of an instance, given the other attributes of that instance and previously labeled training data. To this end, many fuzzy rough set applications of machine learning (see e.g. [6, 57]), use the so-called positive region, defined as follows:

$$\forall x \in U : \mathrm{POS}(x) = \sup_{y \in U}(\underline{\mathrm{apr}}_R(R_d y))(x)$$

In this formula, $\underline{\mathrm{apr}}_R$ refers to one of the lower approximation operators considered in the previous sections, and R and R_d are fuzzy relations in U that represent approximate equality between instances based on the conditional attributes, and on the decision attribute, respectively. They are assumed to be (at least) reflexive and symmetric. In this paper, we use the following definition for R:

$$\forall x, y \in U : R(x,y) = \frac{\sum_{a \in \mathscr{A}} 1 - \frac{|a(x)-a(y)|}{\mathrm{range}(a)}}{|\mathscr{A}|}$$

where $\mathrm{range}(a)$ equals the difference between the maximum value of the attribute a and its minimum value. The fuzzy relation R_d is defined as follows when d is a categorical (discrete) decision attribute:

$$\forall x, y \in U : R_d(x,y) = \begin{cases} 1 & d(x) = d(y) \\ 0 & \text{otherwise} \end{cases}$$

and as follows for a continuous decision attribute:

$$\forall x, y \in U : R_d(x,y) = 1 - \frac{|d(x) - d(y)|}{\mathrm{range}(d)}.$$

For a robust fuzzy rough set model, we would like that the positive region does not change drastically when small changes in the data occur. This should hold both when the conditional attributes are affected by noise (attribute noise), as well as when the decision attribute contains errors (class noise). Given a certain noise level $n \in [0, 100]$, we define the altered decision system $(U, \mathscr{A}^n \cup \{d\})$ as the decision system where each attribute a has a $n\%$ chance of having their values altered to other attribute values in the range of a. To that end, the attribute values of all instances in the decision system are considered separately. For each attribute and each instance, a random number $r \in [0,1]$ is generated. If this number r is lower than $n\%$, the attribute value of the instance is changed to a random value in the range of a. This means that in the asymptotic case of an infinite amount of data, $n\%$ of the attribute values is altered. Analogously, we can define an altered decision system $(U, \mathscr{A} \cup \{d^n\})$ where each decision value has an $n\%$ chance of being altered to a value in the range of d. That is, for each $x \in U$, we generate a random number $r \in [0,1]$ and if this value is lower than $n\%$, the value $d(x)$ is altered to a random value in the range of d.

In order to evaluate the robustness of the fuzzy rough models, we therefore carry out the following procedure for each data set, modeled as a decision system $(U, \mathscr{A} \cup \{d\})$:

1. Calculate the positive regions $\mathrm{POS}(x)$ of all instances $x \in U$ using the specified fuzzy rough set model.
2. For noise levels $n = 1, 2, \ldots, 30$, calculate the positive regions $\mathrm{POS}^a_n(x)$ of all instances, now on the altered decision system $(U, \mathscr{A}^n \cup \{d\})$ where each attribute value $a(x)$ has an $n\%$ chance of being altered to an attribute value in the range of a.

	# instances	# features
Appendicitis	106	7
Iris	150	4
Diabetes	43	2
CPU	209	6

Table 6: Data sets used in the experimental evaluation

3. For noise levels $n = 1, 2, \ldots, 30$, calculate the positive regions $\text{POS}_n^d(x)$ of all instances, now on the altered decision system $(U, \mathscr{A} \cup \{d^n\})$ where each decision value $d(x)$ has an $n\%$ chance of being altered to an attribute value in the range of d.
4. For each noise level, calculate the average distances between the original positive regions and the altered positive regions:

$$\text{error}_n^a = \frac{\sum_{x \in U} |\text{POS}(x) - \text{POS}_n^a(x)|}{|U|},$$

$$\text{error}_n^d = \frac{\sum_{x \in U} |\text{POS}(x) - \text{POS}_n^d(x)|}{|U|}.$$

5. Repeat steps 1 through 4 ten times and report the average errors over ten runs.

These errors express to what extent the fuzzy rough set model changes when a certain level of attribute or class noise is imposed on the data.

We carry out this procedure for four data sets from the KEEL[5] data set repository; their properties are listed in Table 6. Note that the first two data sets, 'Appendicitis' and 'Iris'[6], have a categorical decision attribute (classification problems), while the last two, 'Diabetes' and 'CPU', have a continuous decision attribute (regression problems).

The specific parameters that we used for the different fuzzy rough set models are described in Table 7. We followed the parameter suggestions of the authors for their models in the corresponding papers. In some cases, multiple parameter settings were suggested, in which case we tested all of them and report the results of the models with the most stable parameter setting, highlighted in bold in Table 7. For the IC model, we use the Łukasiewicz implicator to calculate the lower approximation, and we maintain this choice for the other models to make the evaluation implicator independent.

In Figures 1 and 2, we show the results of the experiment. In each case, the X-axis shows the noise level, while the corresponding error can be seen on the Y-axis. Therefore, the less the increase of a curve, the better the corresponding model performs. It can be seen that the OWA model always outperforms the IC model and is in most cases the most robust model, both for attribute and class noise. The VQFRS model also performs well for most of the data sets, but it is mostly outperformed by the OWA model. The VPFRS model performs remarkably bad for

[5] `www.keel.es`

[6] Also available at the UCI Machine Learning Repository [1], `http://archive.ics.uci.edu/ml`

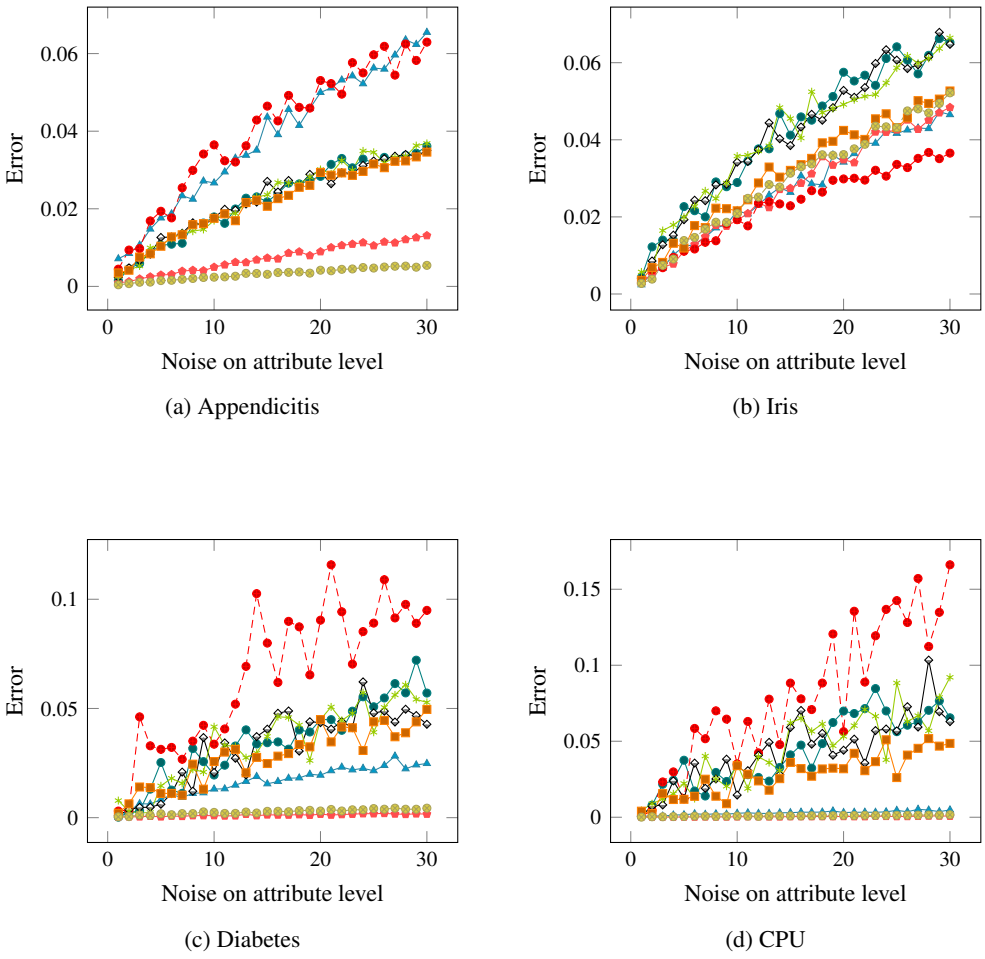

Figure 1: Robustness of the Fuzzy Rough Set models to Attribute Noise

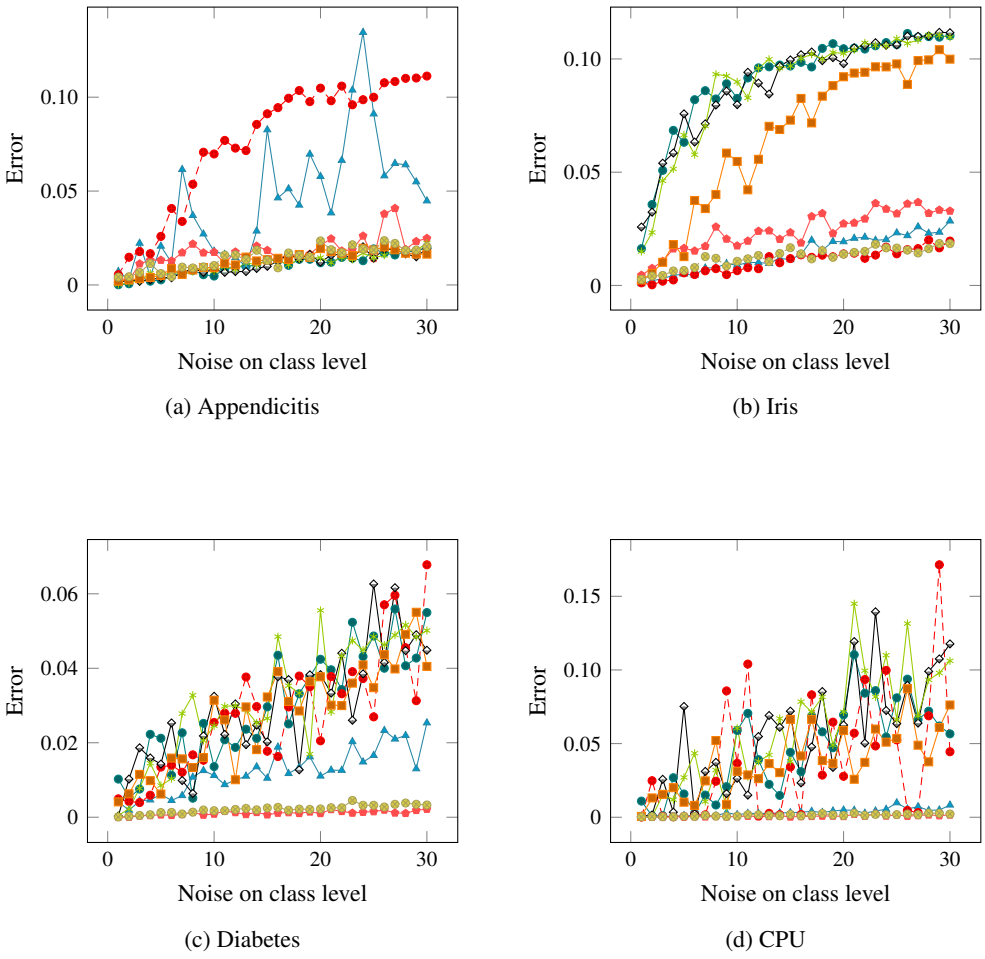

Figure 2: Robustness of the Fuzzy Rough Set models to Class Noise

Model	Parameters
IC	-
VPFRS	$u: 0.80, \mathbf{0.85}, 0.90$
VQFRS	$Q_{(0.2,1)}$
SFRS	$\delta: 0.10, \mathbf{0.15}, 0.20$
FG	$\gamma: \mathbf{0.80}, 0.85, 0.90$
FVPRS	$\alpha: 0.10, \mathbf{0.15}, 0.20$
β-PREC	$\beta: \mathbf{0.97}, 0.98, 0.99$
OWA	Aggregation: minimum operator Exponential weights, **Yager weights** [64] based on $Q_{(0.2,1)}$

Table 7: Parameter settings for the fuzzy rough set models

the Appendicitis data set, where it is even less robust than the IC model. On the other hand, the VPFRS model is one of the most robust models for the Iris and CPU data sets. For all data sets, the FVPRS model and the FG model are more or less equally robust as the IC model. The β-PREC model is more robust than the IC model but is mostly outperformed by other models such as the OWA model or VQFRS. Finally, the SFRS model performs better than the OWA model for the Iris data set, but performs badly for the other data sets. Overall, the OWA model is the most robust against attribute and class noise.

6. Conclusion and Future Work

In this article, we have established a unified framework for fuzzy rough set models. We have compared the approximation operators corresponding to a general implicator-conjunctor (IC) fuzzy rough set model with those used in a number of existing noise-tolerant fuzzy rough set approaches. We have adapted the definition of the Variable Precision Fuzzy Rough Set model [42, 43] in order to obtain a more general model and we have stated a simplified version. Moreover, we have adjusted the Soft Fuzzy Rough Set model [24] such that the idea of the model is maintained, but the problem concerning the soft distance is avoided. Additionally, we have rejected the claim stated in [24] that the Soft Fuzzy Rough Set model satisfies property (IU) for the Kleene-Dienes implicator, the minimum t-norm and a similarity relation.

In order to obtain a critical comparison of all the models, we have utilized a unified notation. Our comparison focused both on desirable theoretical properties which we would like a model to inherit from Pawlak's original rough set definition, as well as on its robustness when data are contaminated with different types of noise. In addition, we have provided proofs and counterexamples for the properties which were not yet discussed by the authors of the noise-tolerant models. Furthermore, we have performed experiments on four real-valued data sets to evaluate to what extent the IC model and the seven noise-tolerant models are sensitive to attribute and class noise.

Our findings indicate that from a theoretical perspective, none of the noise-tolerant models satisfies as many properties as the IC model. When we look at the properties that matter most for practical applications, namely set and relation monotonicity, we conclude that the only models satisfying them are Yao et al.'s Variable Precision Fuzzy Rough Set model based on Fuzzy

Granules (FG, [65]), Zhao et al.'s Fuzzy Variable Precision Rough Set model (FVPRS, [69]), Fernández-Salido and Murakami's β-Precision Fuzzy Rough Set model (β-PREC, [17, 18]) and Cornelis et al.'s Ordered Weighted Average based Fuzzy Rough Set model (OWA, [7]). Among these four, from a practical point of view, the OWA model stands out as the most interesting one, due to its stable behavior with respect to both attribute and class noise, improving the implicator-conjunctor-based model for each of the problems we considered. However, none of the four models satisfies the inclusion property.

An important asset of the OWA approach is its association of weights with ordered positions of elements rather than with the elements themselves; this provides a flexible way to weight data samples differently according to their relevance for different tasks. Clearly, the success of this approach depends on the choice of appropriate OWA weights. While simple rules of thumb have been used for this purpose so far (see e.g. [7, 58]), a more systematic study of their generation is mandatory, e.g., by using learning methods as proposed in [3, 19, 55]. This study is left for future research.

Furthermore, the experiments in this article aim to illustrate the sensitivity of the considered fuzzy rough set models to attribute and class noise. We plan to examine different fuzzy rough set based algorithms and to study their effect on practical applications such as feature selection.

In addition, we continue to search for new robust models that strike the right balance between theoretical attractiveness and practical applicability.

Acknowledgements

Lynn D'eer has been supported by the Ghent University Special Research Fund, Chris Cornelis 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 and Lluis Godo has been partially supported by the Spanish MINECO project EdeTRI TIN2012-39348-C02-01. The authors would like to thank the anonymous referees for their valuable comments and suggestions, in particular the suggestion by one referee to consider the (LU) property.

Appendix A.

In this Appendix, we have gathered counterexamples for those properties that are not valid for the noise-tolerant fuzzy rough set models in this paper, i.e., for all the ✗ marks in Table 5. In all of our examples, we work fuzzy $\mathscr{T}$-similarity relations R, where the t-norm $\mathscr{T}$ will be specified in each example.

Appendix A.1. Noise-Tolerant Models based on Frequency

Appendix A.1.1. Variable Precision Fuzzy Rough Set Model

Since the VPRS model of Ziarko is a special case of the VPFRS model, the properties (D), (A), (INC), (RM), (IU), (ID) and (LU) do not hold. We give a counterexample for (UE) and (CS).

Example 9. *Let $U = \{x, y, z\}$ and R the $\mathscr{T}_L$-similarity relation such that $R(x,y) = 0.7$, $R(x,z) = 0.8$, $R(y,z) = 0.8$. Let $\mathscr{I} = \mathscr{I}_L$ and $\mathscr{T} = \mathscr{T}_L$. Let $l = 0.4$ and $u = 0.6$. Computing the lower*

approximation of $\emptyset = \hat{0}$ and the upper approximation of $U = \hat{1}$ in y gives us the following results:

$$\begin{aligned}(\underline{\mathrm{apr}}_R^{\mathscr{I},u}(\hat{0}))(y) &= 0.2,\\ (\overline{\mathrm{apr}}_R^{\mathscr{T},l}(\hat{1}))(y) &= 0.9.\end{aligned}$$

Hence, (UE) and (CS) do not hold for the VPFRS model.

Appendix A.1.2. Vaguely Quantified Fuzzy Rough Set Model

Since the VPRS model of Ziarko is a special case of the VQFRS model, the properties (D), (A), (INC), (RM), (IU), (ID) and (LU) do not hold. We give a counterexample for (CS).

Example 10. *Let $U = \{x,y,z\}$, $\alpha = 0.1$ and R the similarity relation such that $R(x,y) = R(x,z) = R(y,z) = 1$. Let $Q_u = Q_{(0.2,1)}$ and $Q_l = Q_{(0.1,0.6)}$. We compute the lower and upper approximation of $\hat{\alpha}$ in x. We obtain the following results:*

$$\begin{aligned}(\underline{\mathrm{apr}}_R^{Q_u}(\hat{\alpha}))(x) &= 0,\\ (\overline{\mathrm{apr}}_R^{Q_l}(\hat{\alpha}))(x) &= 0.\end{aligned}$$

Hence, (CS) does not hold for the VQFRS model.

Appendix A.1.3. Soft Fuzzy Rough Set Model

We give counterexamples for (A), for (UE), (CS), (INC) and (LU), for (SM), for (RM) and for (ID). Note that the property (SM) not even holds for crisp subsets A and B. A counterexample for (IU) was already given in Example 8.

Example 11. *Let $U = \{x,y,z\}$ and A, B fuzzy sets in U such that $A(x) = 0.3$, $A(y) = 0.5$, $A(z) = 0.5$ and $B(x) = 0.1$, $B(y) = 0.4$, $B(z) = 0.1$. Let R be the $\mathscr{T}_L$-similarity relation such that $R(x,y) = 0.3$, $R(x,z) = 0.5$, $R(y,z) = 0.5$. Let $\mathscr{I} = \mathscr{I}_L$, $\mathscr{C} = \mathscr{T}_L$ and $\delta = 0.1$. The lower approximation of B is*

$$(\underline{\mathrm{apr}}_R^{\mathscr{I},\delta}(B))(x) = 0.5, (\underline{\mathrm{apr}}_R^{\mathscr{I},\delta}(B))(y) = 0.7, (\underline{\mathrm{apr}}_R^{\mathscr{I},\delta}(B))(z) = 0.5,$$

so it holds that $A \subseteq \underline{\mathrm{apr}}_R^{\mathscr{I},\delta}(B)$. On the other hand, since $(\overline{\mathrm{apr}}_R^{\mathscr{C},\delta}(A))(x) = 0.3$, it does not hold that $\overline{\mathrm{apr}}_R^{\mathscr{I},\delta}(A) \subseteq B$. Hence, (A) does not hold for the SFRS model.

Example 12. *Let $U = \{x,y\}$ and R the similarity relation such that $R(x,y) = 0.1$. Let $\mathscr{I} = \mathscr{I}_L$, $\mathscr{C} = \mathscr{T}_L$ and $\delta = 0.1$. We compute the lower approximation of $\emptyset = \hat{0}$ and the upper approximation of $U = \hat{1}$ in x:*

$$\begin{aligned}(\underline{\mathrm{apr}}_R^{\mathscr{I},\delta}(\hat{0}))(x) &= 0.9,\\ (\overline{\mathrm{apr}}_R^{\mathscr{C},\delta}(\hat{1}))(x) &= 0.1.\end{aligned}$$

Hence, (UE), (CS) and (INC) do not hold for the SFRS model. Moreover, $(\underline{\mathrm{apr}}_R^{\mathscr{I},\delta}(\hat{0}))(y) = 0.9$. This leads to the following result:

$$(\overline{\mathrm{apr}}_R^{\mathscr{C},\delta}(\underline{\mathrm{apr}}_R^{\mathscr{I},\delta}(\hat{0}))(x) = 0.1.$$

Hence, the SFRS model does not satisfy (LU).

Example 13. *Let $U = \{x,y\}$ and A, B crisp subsets of U such that $A = \{x\}$ and $B = \{x,y\}$, i.e., $A \subseteq B$. Let R be the similarity relation such that $R(x,y) = 0.9$. Let $\mathscr{I} = \mathscr{I}_L$ and $\delta = 0.2$. We compute the lower approximation of A and B in x. We have the following results:*

$$\begin{aligned}(\underline{\mathrm{apr}}_R^{\mathscr{I},\delta}(A))(x) &= 0.1,\\ (\underline{\mathrm{apr}}_R^{\mathscr{I},\delta}(B))(x) &= 0.\end{aligned}$$

A similar counterexample can be constructed to illustrate that $\overline{\mathrm{apr}}_R^{\mathscr{C},\delta}(A) \subseteq \overline{\mathrm{apr}}_R^{\mathscr{C},\delta}(B)$ not always holds. Hence, (SM) does not hold for the SFRS model.

Example 14. *Let $U = \{x,y,z\}$ and A the fuzzy set such that $A(x) = 0.6$, $A(y) = 0.6$, $A(z) = 0.2$. Let R_1 and R_2 be similarity relations such that $R_1(x,y) = 0.4$, $R_1(x,z) = 0.3$, $R_1(y,z) = 0.3$ and $R_2(x,y) = 0.9$, $R_2(x,z) = 0.3$, $R_2(y,z) = 0.3$, i.e., $R_1 \subseteq R_2$. Let $\mathscr{I} = \mathscr{I}_L$ and $\delta = 0.2$. We compute the lower approximation of A in x with respect to R_1 and R_2:*

$$\begin{aligned}(\underline{\mathrm{apr}}_{R_1}^{\mathscr{I},\delta}(A))(x) &= 0.6,\\ (\underline{\mathrm{apr}}_{R_2}^{\mathscr{I},\delta}(A))(x) &= 0.7.\end{aligned}$$

A similar counterexample can be constructed to illustrate that $\overline{\mathrm{apr}}_{R_1}^{\mathscr{C},\delta}(A) \subseteq \overline{\mathrm{apr}}_{R_2}^{\mathscr{C},\delta}(A)$ not always holds. Hence, (RM) does not hold for the SFRS model.

Example 15. *Let $U = \{x,y,z\}$ and A the fuzzy set such that $A(x) = 0$, $A(y) = 0.2$, $A(z) = 0.3$. Let R be the $\mathscr{T}_L$-similarity relation such that $R(x,y) = 0.5$, $R(x,z) = 0.8$, $R(y,z) = 0.6$. Let $\mathscr{I} = \mathscr{I}_L$ and $\delta = 0.1$. We compute the lower approximation of A:*

$$(\underline{\mathrm{apr}}_R^{\mathscr{I},\delta}(A))(x) = 0, (\underline{\mathrm{apr}}_R^{\mathscr{I},\delta}(A))(y) = 0.5, (\underline{\mathrm{apr}}_R^{\mathscr{I},\delta}(A))(z) = 0.2.$$

Now, since $(\underline{\mathrm{apr}}_R^{\mathscr{I},\delta}(\underline{\mathrm{apr}}_R^{\mathscr{I},\delta}(A)))(y) = 0.4$, it does not hold that $\underline{\mathrm{apr}}_R^{\mathscr{I},\delta}(\underline{\mathrm{apr}}_R^{\mathscr{I},\delta}(A)) = \underline{\mathrm{apr}}_R^{\mathscr{I},\delta}(A)$. A similar counterexample can be constructed to illustrate that $\overline{\mathrm{apr}}_R^{\mathscr{C},\delta}(\overline{\mathrm{apr}}_R^{\mathscr{C},\delta}(A)) = \overline{\mathrm{apr}}_R^{\mathscr{C},\delta}(A)$ not always holds. Hence, (ID) does not hold for the SFRS model.

Appendix A.1.4. A Variable Precison Fuzzy Rough Set Model based on Fuzzy Granules

We give counterexamples for (A), for (UE), (CS), (INC), (ID) and (LU) and for (IU).

Example 16. *Let $U = \{x,y,z\}$ and A, B fuzzy sets such that $A(x) = 0.5$, $A(y) = 0.8$, $A(z) = 0.2$ and $B(x) = 1$, $B(y) = 0.3$, $B(z) = 0.3$. Let R be the similarity relation such that $R(x,y) = 0.3$, $R(x,z) = 0.3$, $R(y,z) = 1$. Let $\mathscr{T} = \mathscr{T}_L$, $\mathscr{S} = \mathscr{S}_L$ and $\gamma = 0.6$. We compute the upper approximation of A:*

$$(\overline{\mathrm{apr}}_R^{\mathscr{S},\gamma}(A))(x) = 0.1, (\overline{\mathrm{apr}}_R^{\mathscr{S},\gamma}(A))(y) = 0.2, (\overline{\mathrm{apr}}_R^{\mathscr{S},\gamma}(A))(z) = 0.2,$$

i.e., $\overline{\mathrm{apr}}_R^{\mathscr{S},\gamma}(A) \subseteq B$. However, since $(\underline{\mathrm{apr}}_R^{\mathscr{T},\gamma}(B))(y) = 0.3$, it does not hold that $A \subseteq \underline{\mathrm{apr}}_R^{\mathscr{T},\gamma}(B)$. Hence, (A) does not hold for the FG model.

Example 17. *Let $U = \{y_1,\ldots,y_{100}\}$ and R the similarity relation such that for all $y_i, y_j \in U$ with $i \neq j$ it holds that $R(y_i,y_j) = 0.5$. Let $\mathscr{T} = \mathscr{T}_L$, $\mathscr{S} = \mathscr{S}_L$ and $\gamma = 0.95$. The lower approximation of $\emptyset = \hat{0}$ in y_1 is*

$$(\underline{\mathrm{apr}}_R^{\mathscr{T},\gamma}(\hat{0}))(y_1) = 0.5.$$

The upper approximation of $U = \hat{1}$ in y_1 is

$$(\overline{\text{apr}}_R^{\mathscr{S},\gamma}(\hat{1}))(y_1) = 0.5.$$

Hence, (UE), (CS) and (INC) do not hold for the FG model. Furthermore, it holds, for every $y_i \in U$, that

$$(\underline{\text{apr}}_R^{\mathscr{T},\gamma}(\hat{0}))(y_i) = 0.5 = (\overline{\text{apr}}_R^{\mathscr{S},\gamma}(\hat{1}))(y_i).$$

This leads to the following results:

$$(\overline{\text{apr}}_R^{\mathscr{S},\gamma}(\overline{\text{apr}}_R^{\mathscr{S},\gamma}(\hat{1})))(y_1) = 0$$

and

$$(\overline{\text{apr}}_R^{\mathscr{S},\gamma}(\underline{\text{apr}}_R^{\mathscr{T},\gamma}(\hat{0})))(y_1) = 1.$$

Hence, (ID) and (LU) do not hold for the FG model.

Example 18. *Let $U = \{x,y,z\}$ and A, B fuzzy sets such that $A(x) = 0.6$, $A(y) = 0.6$, $A(z) = 0.2$ and $B(x) = 0.6$, $B(y) = 0$, $B(z) = 0.7$. Let R be the similarity relation such that $R(x,y) = 0.4$, $R(x,z) = 0.4$, $R(y,z) = 0.5$. Let $\mathscr{T} = \mathscr{T}_L$, $\mathscr{S} = \mathscr{S}_L$ and $\gamma = 0.6$. We compute the lower approximation of A, B and $A \cap B$ in z:*

$$\begin{aligned}(\underline{\text{apr}}_R^{\mathscr{T},\gamma}(A))(z) &= 1,\\ (\underline{\text{apr}}_R^{\mathscr{T},\gamma}(B))(z) &= 0.7,\\ (\underline{\text{apr}}_R^{\mathscr{T},\gamma}(A \cap B))(z) &= 0.5.\end{aligned}$$

A similar counterexample can be constructed to prove that the upper approximation of the union of two fuzzy sets is not necessarily equal to the union of the upper approximations of the two fuzzy sets. Hence, (IU) does not hold for the FG model.

Appendix A.2. Fuzzy Variable Precision Rough Set Model

We give a counterexample for (UE), (CS) and (INC) and for (LU).

Example 19. *Let $U = \{x,y,z\}$ and R the similarity relation such that $R(x,y) = R(x,z) = R(y,z) = 1$. Let $\mathscr{I} = \mathscr{I}_L$, $\mathscr{C} = \mathscr{T}_L$, $\mathscr{N} = \mathscr{N}_S$, $\mathscr{D} = \mathscr{S}_M$ and $\alpha = 0.1$. We compute the lower approximation of $\emptyset = \hat{0}$ and the upper approximation of $U = \hat{1}$ in x:*

$$\begin{aligned}(\underline{\text{apr}}_R^{\mathscr{I},\alpha}(\hat{0}))(x) &= 0.1,\\ (\overline{\text{apr}}_R^{\mathscr{C},\alpha}(\hat{1}))(x) &= 0.9.\end{aligned}$$

Hence, (UE), (CS) and (INC) do not hold for the FVPRS model.

Example 20. *Let $U = \{x,y,z\}$ and R the similarity relation such that $R(x,y) = R(x,z) = R(y,z) = 1$. Let $\mathscr{I} = \mathscr{I}_L$, $\mathscr{C} = \mathscr{T}_L$, $\mathscr{N} = \mathscr{N}_S$, $\mathscr{D} = \mathscr{S}_M$ and $\alpha = 0.1$. The upper approximation of $\emptyset$ is $\emptyset$:*

$$\overline{\text{apr}}_R^{\mathscr{C},\alpha}(\hat{0}) = \hat{0}.$$

However, from Example 19, we obtain that

$$(\underline{\text{apr}}_R^{\mathscr{I},\alpha}(\overline{\text{apr}}_R^{\mathscr{C},\alpha}(\hat{0})))(x) = (\underline{\text{apr}}_R^{\mathscr{I},\alpha}(\hat{0}))(x) = 0.1.$$

Hence, (LU) does not hold for the FVPRS model.

Appendix A.3. Noise-Tolerant Models based on Aggregation Operators

Appendix A.3.1. β-Precision Fuzzy Rough Set Model

We give counterexamples for (A), for (UE), (CS), (INC), (ID) and (LU) and for (IU).

Example 21. *Let $U = \{x,y,z\}$, A and B fuzzy sets such that $A(x) = 0.5$, $A(y) = 0.8$, $A(z) = 0.3$ and $B(x) = 0.5$, $B(y) = 0.9$, $B(z) = 0.7$. Let $\mathscr{I} = \mathscr{I}_L$, $\mathscr{C} = \mathscr{T}_L$, $\mathscr{T} = \mathscr{T}_L$, $\mathscr{S} = \mathscr{S}_L$ and $\beta = 0.9$. Let R be the $\mathscr{T}_L$-similarity relation such that $R(x,y) = 0.3$, $R(x,z) = 0.6$ and $R(y,z) = 0.7$. We compute the lower approximation of B:*

$$(\underline{\mathrm{apr}}_R^{\mathscr{I},\mathscr{T}_\beta}(B))(x) = 0.5, (\underline{\mathrm{apr}}_R^{\mathscr{I},\mathscr{T}_\beta}(B))(y) = 0.9, (\underline{\mathrm{apr}}_R^{\mathscr{I},\mathscr{T}_\beta}(B))(z) = 0.6.$$

We conclude that $A \subseteq \underline{\mathrm{apr}}_R^{\mathscr{I},\mathscr{T}_\beta}(B)$, but since $(\overline{\mathrm{apr}}_R^{\mathscr{C},\mathscr{S}_\beta}(A))(x) = 0.6$, $\overline{\mathrm{apr}}_R^{\mathscr{C},\mathscr{S}_\beta}(A) \not\subseteq B$. Hence, (A) does not hold for the β-PREC model.

Example 22. *Let $U = \{y_1,\ldots,y_{100}\}$ and R the similarity relation such that for all $y_i, y_j \in U$ with $i \neq j$ it holds that $R(y_i,y_j) = 0.5$. Let $\mathscr{I} = \mathscr{I}_L$, $\mathscr{C} = \mathscr{T}_P$, $\mathscr{T} = \mathscr{T}_M$, $\mathscr{S} = \mathscr{S}_M$ and $\beta = 0.95$. The lower approximation of $\emptyset = \hat{0}$ in y_1 is*

$$(\underline{\mathrm{apr}}_R^{\mathscr{I},\mathscr{T}_\beta}(\hat{0}))(y_1) = 0.5.$$

The upper approximation of $U = \hat{1}$ in y_1 is

$$(\overline{\mathrm{apr}}_R^{\mathscr{C},\mathscr{S}_\beta}(\hat{1}))(y_1) = 0.5.$$

Hence, (UE), (CS) and (INC) do not hold for the β-PREC model. Furthermore, it holds, for every $y_i \in U$, that

$$(\underline{\mathrm{apr}}_R^{\mathscr{I},\mathscr{T}_\beta}(\hat{0}))(y_i) = 0.5 = (\overline{\mathrm{apr}}_R^{\mathscr{C},\mathscr{S}_\beta}(\hat{1}))(y_i).$$

This leads to the following results:

$$(\underline{\mathrm{apr}}_R^{\mathscr{I},\mathscr{T}_\beta}(\underline{\mathrm{apr}}_R^{\mathscr{I},\mathscr{T}_\beta}(\hat{0})))(y_1) = 1$$

and

$$(\overline{\mathrm{apr}}_R^{\mathscr{C},\mathscr{S}_\beta}(\underline{\mathrm{apr}}_R^{\mathscr{I},\mathscr{T}_\beta}(\hat{0})))(y_1) = 0.25.$$

Hence, (ID) and (LU) do not hold for the β-PREC model.

Example 23. *Let $U = \{y_0,\ldots,y_{10}\}$ and A, B fuzzy sets such that*

$$\begin{gathered} A(y_0) = 1, B(y_0) = 0, \\ A(y_i) = 1, B(y_i) = \tfrac{i}{10},\ i \text{ even, not } 0, \\ A(y_i) = \tfrac{i}{10}, B(y_i) = 1,\ i \text{ odd, not } 0. \end{gathered}$$

Let $\mathscr{I} = \mathscr{I}_{KD}$, $\mathscr{C} = \mathscr{T}_M$, $\mathscr{T} = \mathscr{T}_M$ and $\mathscr{S} = \mathscr{S}_M$. Let R be the similarity relation such that for all $y_i, y_j \in U$ it holds that $R(y_i,y_j) = 1$ and $\beta = 0.8$. We compute the lower approximation of A, B and $A \cap B$ in y_0. We have the following results:

$$\begin{aligned} (\underline{\mathrm{apr}}_R^{\mathscr{I},\mathscr{T}_\beta}(A))(y_0) &= 0.3, \\ (\underline{\mathrm{apr}}_R^{\mathscr{I},\mathscr{T}_\beta}(B))(y_0) &= 0.2, \\ (\underline{\mathrm{apr}}_R^{\mathscr{I},\mathscr{T}_\beta}(A \cap B))(y_0) &= 0.1. \end{aligned}$$

A similar counterexample can be constructed to prove that the upper approximation of the union of two fuzzy sets is not necessarily equal to the union of the upper approximations of the two fuzzy sets. Hence, (IU) does not hold for the β-PREC model.

Appendix A.3.2. Ordered Weighted Average based Fuzzy Rough Set Model

We give counterexamples for (A), for (UE), (CS), (INC), (ID) and (LU) and for (IU).

Example 24. *Let $U = \{x,y,z\}$, and A, B fuzzy sets such that $A(x) = 0.7$, $A(y) = 0$, $A(z) = 0.1$ and $B(x) = 0.8$, $B(y) = 0.3$, $B(z) = 0.7$. Let R be the $\mathscr{T}_L$-similarity relation such that $R(x,y) = 0.8$, $R(x,z) = 0.3$, $R(y,z) = 0.2$. Let $\mathscr{I} = \mathscr{I}_L$, $\mathscr{C} = \mathscr{T}_L$, $W_1 = \langle \frac{1}{6}, \frac{2}{6}, \frac{3}{6} \rangle$ and $W_2 = \langle \frac{3}{6}, \frac{2}{6}, \frac{1}{6} \rangle$. The upper approximation of A is:*

$$(\overline{\text{apr}}_R^{\mathscr{C},W_2}(A))(x) = 0.35, (\overline{\text{apr}}_R^{\mathscr{C},W_2}(A))(y) = 0.25, (\overline{\text{apr}}_R^{\mathscr{C},W_2}(A))(z) = 0.05,$$

i.e., $\overline{\text{apr}}_R^{\mathscr{C},W_2}(A) \subseteq B$. On the other hand, since $(\underline{\text{apr}}_R^{\mathscr{I},W_1}(B))(x) = \frac{41}{60} = 0.6833...$, it does not hold that $A \subseteq \underline{\text{apr}}_R^{\mathscr{I},W_1}(B)$. Hence, (A) does not hold for the OWA model.

Example 25. *Let $U = \{x,y\}$ and R the similarity relation such that $R(x,y) = 0.5$. Let $\mathscr{I} = \mathscr{I}_L$, $\mathscr{C} = \mathscr{T}_L$, $W_1 = \langle \frac{1}{3}, \frac{2}{3} \rangle$ and $W_2 = \langle \frac{2}{3}, \frac{1}{3} \rangle$. We compute the lower approximation of $\emptyset = \hat{0}$ and the upper approximation of $U = \hat{1}$ in x:*

$$\begin{aligned} (\underline{\text{apr}}_R^{\mathscr{I},W_1}(\hat{0}))(x) &= \frac{1}{6}, \\ (\overline{\text{apr}}_R^{\mathscr{C},W_2}(\hat{1}))(x) &= \frac{5}{6}. \end{aligned}$$

Hence, (UE), (CS) and (INC) do not hold for the OWA model.

Furthermore, it also holds that $(\underline{\text{apr}}_R^{\mathscr{I},W_1}(\hat{0}))(y) = \frac{1}{6}$ and $(\overline{\text{apr}}_R^{\mathscr{C},W_2}(\hat{1}))(y) = \frac{5}{6}$. This leads to the following results:

$$\begin{aligned} (\underline{\text{apr}}_R^{\mathscr{I},W_1}(\underline{\text{apr}}_R^{\mathscr{I},W_1}(\hat{0})))(x) &= \frac{1}{3}, \\ (\overline{\text{apr}}_R^{\mathscr{C},W_2}(\underline{\text{apr}}_R^{\mathscr{I},W_1}(\hat{0})))(x) &= \frac{1}{9}. \end{aligned}$$

Hence, (ID) and (LU) do not hold for the OWA model.

Example 26. *Let $U = \{y_0, \ldots, y_{10}\}$ and A, B fuzzy sets such that*

$$\begin{gathered} A(y_0) = 1, B(y_0) = 0, \\ A(y_i) = 1, B(y_i) = \tfrac{i}{10},\ i \text{ even, not } 0, \\ A(y_i) = \tfrac{i}{10}, B(y_i) = 1,\ i \text{ odd, not } 0. \end{gathered}$$

Let R be the similarity relation such that for all $y_i, y_j \in U$ it holds that $R(y_i, y_j) = 1$. Let $\mathscr{I} = \mathscr{I}_L$, $\mathscr{C} = \mathscr{T}_L$ and let W_1 and W_2 be as follows:

$$\begin{aligned} W_1 &= \langle 0.10, 0.09, 0.08, 0.07, 0.06, 0.05, 0.04, 0.03, 0.02, 0.01, 0.45 \rangle, \\ W_2 &= \langle 0.45, 0.10, 0.09, 0.08, 0.07, 0.06, 0.05, 0.04, 0.03, 0.02, 0.01 \rangle. \end{aligned}$$

This leads to the following results:

$$\begin{aligned}
(\underline{\mathrm{apr}}_R^{\mathscr{I},W_1}(A))(y_0) &= 0.565,\\
(\underline{\mathrm{apr}}_R^{\mathscr{I},W_1}(B))(y_0) &= 0.51,\\
(\underline{\mathrm{apr}}_R^{\mathscr{I},W_1}(A\cap B))(y_0) &= 0.385,
\end{aligned}$$

$$\begin{aligned}
(\overline{\mathrm{apr}}_R^{\mathscr{C},W_2}(A))(y_0) &= 0.945,\\
(\overline{\mathrm{apr}}_R^{\mathscr{C},W_2}(B))(y_0) &= 0.93,\\
(\overline{\mathrm{apr}}_R^{\mathscr{C},W_2}(A\cup B))(y_0) &= 1.
\end{aligned}$$

Hence, (IU) does not hold for the OWA model.

[1] K. Bache and M. Lichman. UCI Machine Learning Repository, 2013.

[2] M. Baczyński and B. Jayaram. *Fuzzy Implications*. Springer-Verlag Berlin, 2008.

[3] G. Beliakov. Learning weights in the generalized OWA operators. *Fuzzy Optimization and Decision Making*, 4(2):119–130, 2005.

[4] D. Boixader, J. Jacas, and J. Recasens. Upper and lower approximations of fuzzy sets. *International Journal of General Systems*, 29(4):555–568, 2000.

[5] C. Cornelis, M. De Cock, and A.M. Radzikowska. Vaguely quantified rough sets. In *Proceedings of 11th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing (RSFDGrC2007)*, pages 87–94, 2007.

[6] C. Cornelis, R. Jensen, G. Hurtado Martín, and D. Ślęzak. Attribute selection with fuzzy decision reducts. *Information Sciences*, 180(2):209–224, 2010.

[7] C. Cornelis, N. Verbiest, and R. Jensen. Ordered weighted average based fuzzy rough sets. In *Proceedings of the 5th International Conference on Rough Sets and Knowledge Technology (RSKT2010)*, pages 78–85, 2010.

[8] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20:273–297, 1995.

[9] M. De Cock, C. Cornelis, and E.E. Kerre. Fuzzy rough sets: the forgotten step. *IEEE Transactions on Fuzzy Systems*, 15(1):121–130, 2007.

[10] L. D'eer, N. Verbiest, C. Cornelis, and L. Godo. Implicator-conjunctor based models of fuzzy rough sets: definitions and properties. In *Proceedings of the 14th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing (RSFDGrC 2013), LNAI 8170*, pages 169–179, 2013.

[11] C. Degang, Y. Yongping, and W. Hui. Granular computing based on fuzzy similarity relations. *Soft Computing*, 15:1161–1172, 2011.

[12] L. Fariñas del Cerro and H. Prade. Rough sets, twofold fuzzy sets and modal logic – Fuzziness in indiscernibility and partial information. In A. Di Nola and A.G.S. Ventre, editors, *The Mathematics of Fuzzy Systems*, pages 103–120. Verlag TUV, Rheinland, Köln, 1986.

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

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

[15] D. Dubois and H. Prade. Putting fuzzy sets and rough sets together. In R. Słowiński, editor, *Intelligent Decision Support - Handbook of Applications and Advances of the Rough Sets Theory*, pages 203–232. Kluwer Academic Publishers, Boston, 1992.

[16] F. Esteva and L. Godo. Monoidal t-norm based logic: towards a logic for left-continuous t-norms. *Fuzzy Sets and Systems*, 124:271–288, 2001.

[17] J. M. Fernández Salido and S. Murakami. On β-precision aggregation. *Fuzzy Sets and Systems*, 139:547–558, 2003.

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

[19] D. Filev and R. R. Yager. On the issue of obtaining OWA operator weights. *Fuzzy Sets and Systems*, 94:157–169, 1998.

[20] J. Fodor. Left-continuous t-norms in fuzzy logic: an overview. *Journal of applied sciences at Budapest Tech Hungary*, 1(2), 2004.

[21] S. Gottwald and S. Jenei. A new axiomatization for involutive monoidal t-norm-based logic. *Fuzzy Sets and Systems*, 124:303–307, 2001.
[22] S. Greco, B. Matarazzo, and R. Słowinski. Fuzzy similarity relation as a basis for rough approximations. In L. Polkowski and A. Skowron, editors, *Rough Sets and Current Trends in Computing, Lecture Notes in Artificial Intelligence*, volume 1442, pages 283–289. Springer, Berlin, 1998.
[23] S. Greco, B. Matarazzo, and R. Słowinski. Rough set processing of vague information using fuzzy similarity relations. In C.S. Calude and G. Paun, editors, *Finite Versus Infinite—Contributions to an Eternal Dilemma*, pages 149–173. Springer-Verlag, 2000.
[24] Q. Hu, S. An, and D. Yu. Soft fuzzy rough sets for robust feature evaluation and selection. *Information Sciences*, 180:4384–4400, 2010.
[25] Q. Hu, S. An, X. Yu, and D. Yu. Robust fuzzy rough classifiers. *Fuzzy Sets and Systems*, 183:26–43, 2011.
[26] Q. Hu, D. Yu, W. Pedrycz, and D. Chen. Kernelized fuzzy rough sets and their applications. *IEEE Transactions on Knowledge and Data Engineering*, 23(11):1649–1667, 2011.
[27] Q. Hu, L. Zhang, S. An, D. Zhang, and D. Yu. On robust fuzzy rough set models. *IEEE Transactions on Fuzzy Systems*, 20:636–651, 2012.
[28] Q. Hu, L. Zhang, D. Chen, W. Pedrycz, and D. Yu. Gaussian kernel based fuzzy rough sets: model, uncertainty measures and applications. *International Journal of Approximate Reasoning*, 51:453–471, 2010.
[29] M. Inuiguchi. Classification- versus approximation-oriented fuzzy rough sets. In B. Bouchon-Meunier, G. Coletti, and R.R. Yager, editors, *Proceedings of Information Processing and Management of Uncertainty in Knowledge-Based Systems*, 2004, CD-rom.
[30] J. Järvinen. Lattice theory for rough sets. *Transactions on Rough Sets VI LNCS 4374*, pages 400–498, 2007.
[31] R. Jensen and Q. Shen. Fuzzy-rough attributes reduction with application to web categorization. *Fuzzy Sets and Systems*, 141:469–485, 2004.
[32] J.D. Katzberg and W. Ziarko. Variable precision extension of rough sets. *Fundamenta Informaticae*, 27(2-3):155–168, 1996.
[33] T. Y. Lin, Y. Y. Yao, and L. A. Zadeh, editors. *Data Mining, Rough Sets and Granular Computing*, volume 95. Springer, 2002.
[34] T.Y. Lin. Topological and fuzzy rough sets. In R. Słowiński, editor, *Intelligent Decision Support - Handbook of Applications and Advances of Rough Sets Theory*, pages 287–304. Kluwer Academic Publishers, Boston, 1992.
[35] P. J. Lingras and Y. Y. Yao. Data mining using extensions of the rough set model. *Journal of the Association for Information Science and Technology*, 49:415–422, 1998.
[36] G. Liu. Axiomatic systems for rough sets and fuzzy rough sets. *International Journal of Approximate Reasoning*, 48:857–867, 2008.
[37] G. Liu. Using one axiom to characterize rough sets and fuzzy rough sets. *Information Sciences*, 223:285–296, 2013.
[38] M. Liu, D.G. Chen, C. Wu, and H.X. Li. Fuzzy reasoning based on a new fuzzy rough set and its application to scheduling problems. *Computers and Mathematics with Applications*, 51(9-10):1507–1518, 2006.
[39] W.N. Liu, J.T. Yao, and Y.Y. Yao. Rough approximations under level fuzzy sets. In *Rough Sets and Current Trends in Computing (RSCTC2004)*, pages 78–83, 2004.
[40] J.S. Mi, Y. Leung, H.Y. Zhao, and T. Feng. Generalized fuzzy rough sets determined by a triangular norm. *Information Sciences*, 178:3203–3213, 2008.
[41] J.S. Mi and W.X. Zhang. An axiomatic characterization of a fuzzy generalization of rough sets. *Information Sciences*, 160:235–249, 2004.
[42] A. Mieszkowicz-Rolka and L. Rolka. Variable precision fuzzy rough sets. In J.F. Peters et al., editor, *Transactions on Rough Sets I*, pages 144–160. Springer-Verlag, 2004.
[43] A. Mieszkowicz-Rolka and L. Rolka. Fuzzy rough approximations of process data. *Internation Journal of Approximate Reasoning*, 49:301–315, 2008.
[44] N.N. Morsi and M.M. Yakout. Axiomatics for fuzzy rough sets. *Fuzzy Sets and Systems*, 100:327–342, 1998.
[45] A. Nakamura. Fuzzy rough sets. *Note on Multiple-Valued Logic in Japan*, 9:1–8, 1988.
[46] S. Nanda and S. Majumdar. Fuzzy rough sets. *Fuzzy Sets and Systems*, 45:157–160, 1992.
[47] Z. Pawlak. Rough sets. *International journal of computer and information sciences*, 11(5):341–356, 1982.
[48] Z. Pawlak and A. Skowron. Rough sets: some extensions. *Information Sciences*, 177:28–40, 2007.
[49] Z. Pawlak and A. Skowron. Rudiments of rough sets. *Information Sciences*, 177:3–27, 2007.
[50] D. Pei. A generalized model of fuzzy rough sets. *International Journal of General Systems*, 34(5):603–613, 2005.
[51] A.M. Radzikowska and E.E. Kerre. A comparative study of fuzzy rough sets. *Fuzzy Sets and Systems*, 126:137–155, 2002.
[52] A. Skowron and C. Rauszer. The discernibility matrices and functions in information systems. In R. Słowiński, editor, *Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory*, pages 331–362. Kluwer Academic Publishers, the Netherlands, 1992.

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

[54] H. Thiele. Fuzzy rough sets versus rough fuzzy sets – an interpretation and a comparative study using concepts of modal logic. Technical report, University of Dortmund, 1998. ISSN 1433-3325.

[55] V. Torra. On the learning of weights in some aggregation operators: the weighted mean and OWA operators. *Mathware & Soft Computing*, 6(2–3), 1999.

[56] E. Turunen. *Mathematics Behind Fuzzy Logic*. Advances in Soft Computing. Physica-Verlag, Heidelberg, 1999.

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

[58] N. Verbiest, C. Cornelis, and F. Herrera. OWA-FRPS: A prototype selection method based on ordered weighted average fuzzy rough set theory. In *Proceedings of the 14th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing (RSFDGrC 2013), LNAI 8170*, pages 180–190, 2013.

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

[60] W.Z. Wu, J.S. Mi, and W.X. Zhang. Generalized fuzzy rough sets. *Information Sciences*, 151:263–282, 2003.

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

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

[63] R.R. Yager. On ordered weighted averaging aggregation operators in multicriteria decision making. *IEEE Transactions on Systems, Man and Cybernetics*, 18:183–190, 1988.

[64] R.R. Yager. Families of owa operators. *Fuzzy Sets and Systems*, 59(2):125–148, 1993.

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

[66] Y. Y. Yao. Combination of rough and fuzzy sets based on α-level sets. In T. Y. Lin and N. Cercone, editors, *Rough Sets and Data Mining: Analysis for Imprecise Data*, pages 301–321. Kluwer Academic Publishers, 1997.

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

[68] L. A. Zadeh. Fuzzy sets. *Information and control*, 8:338–353, 1965.

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

[70] W. Ziarko. Variable precision rough set model. *Journal of Computer and System Sciences*, 46:39–59, 1993.