Improving the β-Precision and OWA Based Fuzzy Rough Set Models: Definitions, Properties and Robustness Analysis

Lynn D'eer and Nele Verbiest

Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Krijgslaan 281 (S9), B-9000 Gent, Belgium
{Lynn.Deer,Nele.Verbiest}@UGent.be

Abstract. Since the early 1990s, many authors have studied fuzzy rough set models and their application in machine learning and data reduction. In this work, we adjust the β-precision and the ordered weighted average based fuzzy rough set models in such a way that the number of theoretical properties increases. Furthermore, we evaluate the robustness of the new models a-β-PREC and a-OWA to noisy data and compare them to a general implicator-conjunctor-based fuzzy rough set model.

Keywords: fuzzy sets, rough sets, hybridization, lower and upper approximation, implication, conjunction, beta-precision, OWA, robustness.

1 Introduction

Rough set theory (Pawlak [1], 1982) characterizes uncertainty due to incomplete information, by dividing a set of objects according to their indiscernibility towards each other, modeled by an equivalence relation. In particular, the lower and upper approximation of a set are constructed. The former includes the objects certainly belonging to the set, while the latter excludes the objects certainly not belonging to the set.

Furthermore, fuzzy set theory (Zadeh [2], 1965) extends classical or crisp sets in the sense that intermediary membership degrees, mostly between 0 and 1, can be obtained. This theory is used when dealing with gradual information or vague concepts.

Hybridization of both theories has its origin in the early 1990s, when Dubois and Prade [3] presented the first fuzzy rough set model. From then on, research on fuzzy rough set models grows, mainly due to its proven application in machine learning and, in particular, in feature selection.

Many fuzzy rough set models are intuitively constructed by substituting the Boolean implication and conjunction in Pawlak's model by fuzzy implicators and conjunctors, as well as the universal and existential quantifier by the infimum and supremum operator. In addition, approximations by general fuzzy relations are studied, instead of considering fuzzy equivalence relations. All these studies can be covered by a general implicator-conjunctor-based fuzzy rough set model (IC-model), discussed in [4].

C. Cornelis et al. (eds.): RSCTC 2014, LNAI 8536, pp. 23–34, 2014.

However, the main disadvantage of this model is its use of the infimum and supremum operator. Both operators are very sensitive to noise in the data and/or outlying samples. To overcome this problem, authors have studied robust fuzzy rough set models. We focus on the following two models: the β-precision fuzzy rough set model introduced by Fernández-Salido and Murakami in 2003 [5,6] and the ordered weighted average (OWA) based fuzzy rough set model introduced by Cornelis et al. in 2010 [7]. Both models use aggregation operators instead of the inf- and sup-operator and preliminary work showed that they have interesting theoretical and practical assets. Unfortunately, they do not satisfy the inclusion property, which is required if we want the approximations to be on both sides of the set to be approximated, and is important for feature selection [8].

In this work, we overcome this drawback by adjusting the two models. Inspiration for this adjustment was given by Inuiguchi [9]. We present the adapted β-precision and OWA based fuzzy rough set models, shortly called the a-β-PREC- and a-OWA-model. Moreover, we discuss their properties and analyze their robustness in comparison with the IC-model and the original robust models.

The remainder of this paper is as follows: in Section 2, we briefly recall the IC-model. In Section 3, we recall the β-precision and OWA based fuzzy rough set models and propose adaptations. Furthermore, we discuss which properties of the IC-model are maintained by the new models. In Section 4, we compare the five models with respect to their robustness to noisy data. Finally, we conclude and state future work in Section 5.

2 The IC-Model

Consider a *fuzzy approximation space*, i.e., a couple (U, R) consisting of a non-empty set U and a binary fuzzy relation R in U, and a general format for approximation operators using implicators[1] and conjunctors[2]:

Definition 1. *[4] Let (U, R) be a fuzzy approximation space, A a fuzzy set in U, $\mathcal{I}$ an implicator and $\mathcal{C}$ a conjunctor. The $(\mathcal{I}, \mathcal{C})$-*fuzzy rough approximation *of A by R is the pair of fuzzy sets $(R{\downarrow}_{\mathcal{I}}A, R{\uparrow}_{\mathcal{C}}A)$ defined by, for $x \in U$,*

$$(R{\downarrow}_{\mathcal{I}}A)(x) = \inf_{y \in U} \mathcal{I}(R(y,x), A(y)),$$
$$(R{\uparrow}_{\mathcal{C}}A)(x) = \sup_{y \in U} \mathcal{C}(R(y,x), A(y)).$$

A pair (A_1, A_2) of fuzzy sets in U is called a fuzzy rough set *in (U, R) if there exists a fuzzy set A in U such that $A_1 = R{\downarrow}_{\mathcal{I}}A$ and $A_2 = R{\uparrow}_{\mathcal{C}}A$.*

[1] An *implicator* $\mathcal{I}$ is a mapping $\mathcal{I}\colon [0,1]^2 \to [0,1]$ satisfying $\mathcal{I}(1,0) = 0$, $\mathcal{I}(1,1) = \mathcal{I}(0,1) = \mathcal{I}(0,0) = 1$ which is decreasing in the first and increasing in the second argument.

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

In Table 1, the extensions of the classical rough set properties to a fuzzy approximation space are shown; (U, R), (U, R_1) and (U, R_2) are fuzzy approximation spaces, A, B and $\hat{\alpha}$[3] are fuzzy sets in U, $\mathcal{I}$ is an implicator, $\mathcal{C}$ a conjunctor, $\mathcal{N}$ an involutive negator[4] and R' the inverse relation of R, defined by, for $x, y \in U$, $R'(y, x) = R(x, y)$.

The following proposition summarizes under which conditions all properties in Table 1 hold.

Proposition 1. *[4] Let $\mathcal{C}$ be a left-continuous t-norm $\mathcal{T}$ and $\mathcal{I}$ its R-implicator, i.e., for x, y in $[0,1]$, $\mathcal{I}_{\mathcal{T}}(x,y) = \sup\{\gamma \in [0,1] \mid \mathcal{T}(x,\gamma) \leq y\}$. If the fuzzy relation R is reflexive, i.e., for x in U, $R(x,x) = 1$, and $\mathcal{T}$-transitive, i.e., for x, y, z in U, $\mathcal{T}(R(x,y), R(y,z)) \leq R(x,z)$, then all properties in Table 1 hold.*

Table 1. Properties in a fuzzy approximation space

(D) Duality	$R{\downarrow_{\mathcal{I}}}A = (R{\uparrow_{\mathcal{C}}}(A^{\mathcal{N}}))^{\mathcal{N}}$
(A) Adjointness	$A \subseteq R{\uparrow_{\mathcal{C}}}B \Leftrightarrow R'{\downarrow_{\mathcal{I}}}A \subseteq B$
(INC) Inclusion	$R{\downarrow_{\mathcal{I}}}A \subseteq A \subseteq R{\uparrow_{\mathcal{C}}}A$
(SM) Set monotonicity	$A \subseteq B \Rightarrow \begin{cases} R{\downarrow_{\mathcal{I}}}A \subseteq R{\downarrow_{\mathcal{I}}}B \\ R{\uparrow_{\mathcal{C}}}A \subseteq R{\uparrow_{\mathcal{C}}}B \end{cases}$
(RM) Relation monotonicity	$R_1 \subseteq R_2 \Rightarrow \begin{cases} R_2{\downarrow_{\mathcal{I}}}A \subseteq R_1{\downarrow_{\mathcal{I}}}A \\ R_1{\uparrow_{\mathcal{C}}}A \subseteq R_2{\uparrow_{\mathcal{C}}}A \end{cases}$
(IU) Intersection and Union	$R{\downarrow_{\mathcal{I}}}(A \cap B) = R{\downarrow_{\mathcal{I}}}A \cap R{\downarrow_{\mathcal{I}}}B$ $R{\uparrow_{\mathcal{C}}}(A \cup B) = R{\uparrow_{\mathcal{C}}}A \cup R{\uparrow_{\mathcal{C}}}B$
(ID) Idempotence	$R{\downarrow_{\mathcal{I}}}(R{\downarrow_{\mathcal{I}}}A) = R{\downarrow_{\mathcal{I}}}A$ $R{\uparrow_{\mathcal{C}}}(R{\uparrow_{\mathcal{C}}}A) = R{\uparrow_{\mathcal{C}}}A$
(CS) Constant sets	$R{\downarrow_{\mathcal{I}}}\hat{\alpha} = \hat{\alpha}$ $R{\uparrow_{\mathcal{C}}}\hat{\alpha} = \hat{\alpha}$

3 Robust Fuzzy Rough Set Models

A disadvantage of the IC-model is that the infimum and supremum operator are very sensitive to noise in the data: a small error in the data set can change the outcome of the model drastically. To overcome this problem, robust fuzzy rough set models are studied.

In literature, many robust fuzzy rough set models are defined. A first group of robust models is based on frequency: these models only take a subset of U into account when computing the lower and upper approximation [10,11,12]. Furthermore, there are robust models that use vague quantifiers to compute the

[3] $\forall \alpha \in [0,1]$, we denote with $\hat{\alpha}$ the constant α-set in U, i.e., $\forall x \in U\colon \hat{\alpha}(x) = \alpha$.

[4] A *negator* $\mathcal{N}$ is a decreasing mapping $\mathcal{N}\colon [0,1] \to [0,1]$ which satisfies $\mathcal{N}(0) = 1$ and $\mathcal{N}(1) = 0$. It is *involutive* if for all $x \in [0,1]$, $\mathcal{N}(\mathcal{N}(x)) = x$. The standard negator $\mathcal{N}_S$ is defined by, for x in $[0,1]$, $\mathcal{N}_S(x) = 1 - x$.

approximation operators [13] or that modify the fuzzy set A which is approximated [14].

In this work, we focus on robust models that replace the inf- and sup-operator by aggregation operators. It is known that they are highly robust against noise and that they satisfy the properties (SM) and (RM), which are important in feature selection where the goal is to find a minimal subset of features [15]. More specifically, we concentrate on the β-precision (β-PREC, [5,6]) and OWA based (OWA, [7]) fuzzy rough set models. We adapt both models in such a way that the inclusion property is guaranteed, and hence the original idea of Pawlak. In addition, we compare the original and adapted models to the IC-model from both theoretical and practical view.

In the remainder of this article, we assume the universe U to be finite. This is not a limitation for practical purposes, since data sets in real applications are always finite.

3.1 The Original β-Precision and OWA Based Fuzzy Rough Set Models

First, we recall the fuzzy rough set model based on β-precision quasi-t-norms and quasi-t-conorms:

Definition 2. *[5,6] Given a t-norm $\mathcal{T}$, a t-conorm*[5] *$\mathcal{S}$, $\beta \in [0,1]$ and $n \in \mathbb{N} \setminus \{0,1\}$, the corresponding* β-precision quasi-t-norm $\mathcal{T}_\beta$ *and* β-precision quasi-t-conorm $\mathcal{S}_\beta$ *of order n are $[0,1]^n \to [0,1]$ mappings such that for all $\boldsymbol{x} = (x_1, \ldots, x_n)$ in $[0,1]^n$,*

$$\mathcal{T}_\beta(\boldsymbol{x}) = \mathcal{T}(y_1, \ldots, y_{n-m}),$$
$$\mathcal{S}_\beta(\boldsymbol{x}) = \mathcal{S}(z_1, \ldots, z_{n-p}),$$

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

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

Note that for $\beta = 1$ the original t-norm $\mathcal{T}$ and t-conorm $\mathcal{S}$ are retrieved.

The β-precision fuzzy rough set model (shortly, β-PREC-model) is defined as follows:

[5] A *t-conorm* $\mathcal{S}$ is a mapping $\mathcal{S}\colon [0,1]^2 \to [0,1]$ that is increasing in both arguments, commutative, associative and satisfies for x in $[0,1]$, $\mathcal{S}(x,0) = x$.

Definition 3. *[6] Let $\mathcal{T}$ be a t-norm, $\mathcal{S}$ a t-conorm and $\beta \in [0,1]$. Given an implicator $\mathcal{I}$ and a conjunctor $\mathcal{C}$, the* β-precision fuzzy rough approximation *of A by R is the pair of fuzzy sets $(R{\downarrow}_{\mathcal{I},\mathcal{T}_\beta}A, R{\uparrow}_{\mathcal{C},\mathcal{S}_\beta}A)$, defined by, for $x \in U$:*

$$(R{\downarrow}_{\mathcal{I},\mathcal{T}_\beta}A)(x) = \underset{y\in U}{\mathcal{T}_\beta} \langle \mathcal{I}(R(y,x), A(y))\rangle,$$
$$(R{\uparrow}_{\mathcal{C},\mathcal{S}_\beta}A)(x) = \underset{y\in U}{\mathcal{S}_\beta} \langle \mathcal{C}(R(y,x), A(y))\rangle.$$

For $\mathcal{T} = \min$ and $\mathcal{S} = \max$ the following hold: if $\beta = 1$, the IC-model is obtained, and if $\beta < 1$, the approximation operators of the β-PREC-model satisfy $R{\downarrow}_{\mathcal{I}}A \subseteq R{\downarrow}_{\mathcal{I},\mathcal{T}_\beta}A$ and $R{\uparrow}_{\mathcal{T},\mathcal{S}_\beta}A \subseteq R{\uparrow}_{\mathcal{T}}A$.

The following properties hold for the β-PREC-model:

Proposition 2. *Let $\mathcal{T}$ be a t-norm and $\mathcal{S}$ its $\mathcal{N}_S$-dual t-conorm, i.e., for x, y in $[0,1]$, $\mathcal{S}(x,y) = 1 - \mathcal{T}(1-x, 1-y)$. Let $\beta \in [0,1]$. If the pair $(\mathcal{I},\mathcal{C})$ consists of an implicator $\mathcal{I}$ and the conjunctor induced by $\mathcal{I}$ and $\mathcal{N}_S$, i.e., for x, y in $[0,1]$, $\mathcal{C}(x,y) = 1 - \mathcal{I}(x, 1-y)$, then (D) w.r.t. $\mathcal{N}_S$ holds for the β-PREC-model.*

Proposition 3. *The β-PREC-model satisfies (SM) and (RM).*

Secondly, we recall the fuzzy rough set model based on OWA operators:

Definition 4. *[16] Given a sequence D of n scalar values and a weight vector $W = \langle w_1, \ldots, w_n\rangle$ of length n, such that for all $i \in \{1,\ldots,n\}$, $w_i \in [0,1]$, and $\sum_{i=1}^{n} w_i = 1$ (an* OWA weight vector of length *n). 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* $\text{OWA}_W(D) = \sum_{i=1}^{n} w_i d_{\sigma(i)}$.

The OWA operator allows to consider a wide variety of aggregation strategies. For instance, the maximum and minimum are represented by the weight vectors $W_{\max} = \langle 1, 0, \ldots, 0\rangle$ and $W_{\min} = \langle 0, \ldots, 0, 1\rangle$, respectively. To measure how similar an OWA operator is to the maximum and minimum, the orness and andness degree are used. Let W be an OWA weight vector of length n, the *orness* and *andness degree* of W are defined by

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

The OWA based fuzzy rough set model (shortly OWA-model) is defined as follows:

Definition 5. *[7] Given an implicator $\mathcal{I}$, a conjunctor $\mathcal{C}$ and OWA weight vectors W_1 and W_2 of length n, with $n = |U|$, and such that* $\text{andness}(W_1) > 0.5$

and orness$(W_2) > 0.5$. *The* (W_1, W_2)-fuzzy rough approximation *of* A *by* R *is the pair of fuzzy sets* $(R{\downarrow}_{\mathcal{I},W_1}A, R{\uparrow}_{\mathcal{C},W_2}A)$ *defined by, for* $x \in U$*:*

$$(R{\downarrow}_{\mathcal{I},W_1}A)(x) = \underset{y \in U}{\mathrm{OWA}_{W_1}} \langle \mathcal{I}(R(y,x), A(y)) \rangle,$$
$$(R{\uparrow}_{\mathcal{C},W_2}A)(x) = \underset{y \in U}{\mathrm{OWA}_{W_2}} \langle \mathcal{C}(R(y,x), A(y)) \rangle.$$

By varying the OWA weight vectors, different fuzzy rough set models can be obtained. For the weight vectors $W_1 = W_{\min}$ and $W_2 = W_{\max}$, 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 $R{\downarrow}_{\mathcal{I}}A \subseteq R{\downarrow}_{\mathcal{I},W_1}A$ and $R{\uparrow}_{\mathcal{C},W_2}A \subseteq R{\uparrow}_{\mathcal{C}}A$.

The following properties hold for the OWA-model:

Proposition 4. *Let* W_1 *be a weight vector such that* andness$(W_1) > 0.5$ *and let* W_2 *be of the same length* n *such that* $(W_2)_i = (W_1)_{n-i+1}$ *for* $i \in \{1, \ldots, n\}$. *Let* $(\mathcal{I}, \mathcal{C})$ *be a pair consisting of an implicator* $\mathcal{I}$ *and the conjunctor induced by* $\mathcal{I}$ *and* $\mathcal{N}_S$*, then (D) w.r.t.* $\mathcal{N}_S$ *holds for the OWA-model.*

Proposition 5. *The OWA-model satisfies (SM) and (RM).*

3.2 The Adapted β-Precision and OWA Based Fuzzy Rough Set Models

A drawback of both models is that they do not satisfy the inclusion property. It means that the lower approximation is not necessarily contained in the approximated set. This is something we want to avoid in feature selection, where the goal is to find a smaller set of features. However, by adjusting the models, we can force the inclusion property to hold.

We begin with adapting the β-precision fuzzy rough set model:

Definition 6. *Let* $\mathcal{T}$ *be a t-norm,* $\mathcal{S}$ *a t-conorm and* $\beta \in [0, 1]$. *Given an implicator* $\mathcal{I}$ *and a conjunctor* $\mathcal{C}$*, the* adapted β-precision fuzzy rough approximation *of* A *by* R *is the pair of fuzzy sets* $(R{\downarrow}^a_{\mathcal{I},\mathcal{T}_\beta}A, R{\uparrow}^a_{\mathcal{C},\mathcal{S}_\beta}A)$*, defined by, for* $x \in U$*:*

$$(R{\downarrow}^a_{\mathcal{I},\mathcal{T}_\beta}A)(x) = \min\{A(x), \underset{y \in U}{\mathcal{T}_\beta} \langle \mathcal{I}(R(y,x), A(y)) \rangle\},$$
$$(R{\uparrow}^a_{\mathcal{C},\mathcal{S}_\beta}A)(x) = \max\{A(x), \underset{y \in U}{\mathcal{S}_\beta} \langle \mathcal{C}(R(y,x), A(y)) \rangle\}.$$

We refer to this model as the a-β-PREC-model.

In the a-β-PREC-model, (D), (SM) and (RM) still hold and moreover, the properties (INC) and (CS) for $\alpha = 0$ and $\alpha = 1$ hold.

Proposition 6. *Let* $\mathcal{T}$ *be a t-norm and* $\mathcal{S}$ *its* $\mathcal{N}_S$*-dual t-conorm. Let* $\beta \in [0, 1]$. *If the pair* $(\mathcal{I}, \mathcal{C})$ *consists of an implicator* $\mathcal{I}$ *and the conjunctor induced by* $\mathcal{I}$ *and* $\mathcal{N}_S$*, then (D) w.r.t.* $\mathcal{N}_S$ *holds for the a-*β*-PREC-model.*

Proposition 7. *The a-β-PREC-model satisfies (INC), (SM), (RM) and (CS) for $\alpha = 0$ and $\alpha = 1$.*

The a-β-PREC-model does not satisfy (A), (IU), (ID) and (CS) for $\alpha \in]0,1[$.

In a similar way, we adjust the OWA based model:

Definition 7. *Given an implicator $\mathcal{I}$, a conjunctor $\mathcal{C}$ 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* adapted (W_1, W_2)-fuzzy rough approximation *of A by R is the pair of fuzzy sets $(R{\downarrow}^a_{\mathcal{I},W_1} A, R{\uparrow}^a_{\mathcal{C},W_2} A)$ defined by, for $x \in U$:*

$$(R{\downarrow}^a_{\mathcal{I},W_1} A)(x) = \min\{A(x), \underset{y\in U}{\mathrm{OWA}_{W_1}} \langle \mathcal{I}(R(y,x), A(y))\rangle\},$$
$$(R{\uparrow}^a_{\mathcal{C},W_2} A)(x) = \max\{A(x), \underset{y\in U}{\mathrm{OWA}_{W_2}} \langle \mathcal{C}(R(y,x), A(y))\rangle\}.$$

We refer to this model as the a-OWA-model.

The a-OWA-model still satisfies (D), (SM) and (RM) and additionally, it satisfies (INC) and (CS) for all α in $[0,1]$.

Proposition 8. *Let W_1 be a weight vector such that* andness$(W_1) > 0.5$ *and let W_2 be of the same length n such that $(W_2)_i = (W_1)_{n-i+1}$ for $i \in \{1, \ldots, n\}$. Let $(\mathcal{I}, \mathcal{C})$ be a pair consisting of an implicator $\mathcal{I}$ and the conjunctor induced by $\mathcal{I}$ and $\mathcal{N}_S$, then (D) w.r.t. $\mathcal{N}_S$ holds for the a-OWA-model.*

Proposition 9. *The a-OWA-model satisfies (INC), (SM), (RM) and (CS) for $\alpha = 0$ and $\alpha = 1$. If the implicator $\mathcal{I}$ and the conjunctor $\mathcal{C}$ satisfy $\mathcal{I}(1, x) = x$ and $\mathcal{C}(1, x) = x$ for all x in $[0,1]$, then the a-OWA-model satisfies (CS) for all α in $[0,1]$.*

The a-OWA-model does not satisfy (A), (IU) and (ID).

To end, we compare the two adapted models to the IC-model in Table 2. 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 ✰. Note that the property (UE) stands for the (CS)-property with $\alpha \in \{0, 1\}$. In both cases we see that the adapted models satisfy more properties than the original ones, but the IC-model remains the best model from theoretical point of view.

4 Analysis of Robustness

Besides a theoretical comparison of the five models, we evaluate their robustness to noisy data. For this, we use five real-valued data sets from the KEEL data set repository[6]: 'Diabetes' ($|U| = 43$, $|\mathcal{A}| = 2$), 'Ele-1' ($|U| = 495$, $|\mathcal{A}| = 2$), 'AutoMPG6' ($|U| = 392$, $|\mathcal{A}| = 5$), 'MachineCPU' ($|U| = 209$, $|\mathcal{A}| = 6$) and 'Baseball' ($|U| = 337$, $|\mathcal{A}| = 16$). Each data set can be considered as a decision system $(U, \mathcal{A} \cup \{d\})$, where U is the finite set of instances, $\mathcal{A}$ is the set

[6] www.keel.es

Table 2. Overview of properties for the different fuzzy rough set models

Property	IC	β-PREC	a-β-PREC	OWA	a-OWA
(D)	☆	☆	☆	☆	☆
(A)	☆	✗	✗	✗	✗
(INC)	☆	✗	✓	✗	✓
(SM)	✓	✓	✓	✓	✓
(RM)	✓	✓	✓	✓	✓
(IU)	✓	✗	✗	✗	✗
(ID)	☆	✗	✗	✗	✗
(CS)	☆	✗	✗	✗	☆
(UE)	☆	✗	✓	✗	✓

of features (conditional attributes) and $d \notin \mathcal{A}$ the decision attribute. We only consider regression problems, so the decision attributes in the five data sets are continuous.

In many data mining tasks based on fuzzy rough set theory [17,18], the *positive region* is used, defined by, for x in U, $\text{POS}(x) = \sup\limits_{y \in U}(R{\downarrow}R_d y)(x)$. In this paper, $\downarrow$ is as in one of the definitions in Section 3, R is the indiscernibilitiy relation defined by

$$\forall x, y \in U : R(x,y) = \frac{1}{|\mathcal{A}|} \cdot \left(\sum_{a \in \mathcal{A}} 1 - \frac{|a(x) - a(y)|}{\text{range}(a)} \right)$$

and $R_d y$ is the indiscernibility class of y by R_d defined by

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

Here, the range of an attribute $a \in \mathcal{A} \cup \{d\}$ is given by the difference between the maximum and the minimum value of a.

If the positive region based on a certain fuzzy rough model does not change drastically when small errors in the data occur, we call the model robust. These errors can occur both in the features values (attribute noise) and in the decision attribute values (class noise). To evaluate the robustness of the fuzzy rough set models discussed in this paper, we compare the values of the fuzzy rough positive region calculated based on the original dataset and based on the same dataset with artificial noise added.

Given a decision system $(U, \mathcal{A} \cup \{d\})$ and a certain noise level n, we define the altered decision system $(U, \mathcal{A}^n \cup \{d\})$ with artificial attribute noise as the decision system where each attribute value $a(x)$ has an $n\%$ chance of being altered to an attribute value in the range of a. For instance, if a takes values in the interval $[0, 10]$ and if the noise level is 10%, for each instance $x \in U$ there is a 10 percent chance that the value of $a(x)$ in the altered decision system is not equal to $a(x)$ but is a random value in the interval $[0, 10]$. We add artificial class noise in a similar manner. The altered decision system $(U, \mathcal{A} \cup \{d^n\})$ is the decision system where each decision value $d(x)$ has an $n\%$ chance of being altered to a random value in the range of d.

We denote by $\text{POS}_n^a(x)$ the fuzzy rough positive region of $x \in U$ based on the decision system $(U, \mathcal{A}^n \cup \{d\})$, where n percent artificial attribute noise is added. The value $\text{POS}_n^d(x)$ refers to the fuzzy rough positive region of $x \in U$ based on the decision system $(U, \mathcal{A} \cup \{d^n\})$, where n percent artificial class noise is added.

As we are interested in how much the value of the fuzzy rough positive region based on the altered decision systems with artificial noise deviates from the original values, we define the following error measures:

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

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

These measures reflect the average deviation of the fuzzy rough positive region when the decision system has $n\%$ attribute or class noise.

In Algorithm 1 we outline the experiment that we carry out for each dataset, represented by a decision system, and each fuzzy rough set model. We consider 30 noise levels, and calculate the average errors over 10 runs associated with this noise level and dataset in lines 3 to 10. Note that the processes in line 6 and 7 are stochastic, and therefore this procedure is repeated 10 times. As a result, for each dataset and fuzzy rough model, we have 60 error values, namely the average error related to attribute noise and the average error related to class noise for each noise level n in $1, \ldots, 30$.

Algorithm 1. Procedure carried out in our experimental evaluation to assess the robustness of a given fuzzy rough set model on a dataset

1: **Input:** Dataset represented by a decision system $(U, \mathcal{A} \cup \{d\})$, fuzzy rough set model
2: **for** $n = 1, \ldots, 30$ **do**
3:     $\text{av-error}_n^a \leftarrow 0$
4:     $\text{av-error}_n^d \leftarrow 0$
5:     **for** $i = 1, \ldots, 10$ **do**
6:         $\text{av-error}_n^a \leftarrow \text{av-error}_n^a + \text{error}_n^a$
7:         $\text{av-error}_n^d \leftarrow \text{av-error}_n^d + \text{error}_n^d$
8:     **end for**
9:     $\text{av-error}_n^a \leftarrow \text{av-error}_n^a/10$
10:     $\text{av-error}_n^d \leftarrow \text{av-error}_n^d/10$
11:     **Output:** av-error_n^a and av-error_n^d
12: **end for**

The parameters used in the different fuzzy rough set models are as follows: in all the models we use the Łukasiewicz implicator $\mathcal{I}(a, b) = \min(1, 1 - a + b)$ for

$a, b \in [0, 1]$. In the β-PREC- and a-β-PREC-model, $\mathcal{T} = \min$ and $\beta = 0.96$ are used. In the OWA- and a-OWA-model we use the weight vector W defined by

$$w_i = \frac{1}{\sum_{i=1}^{n} w_i} \cdot \frac{1}{n - i + 1}.$$

The results are shown in Figures 1 and 2. For both attribute and class noise, we see that the IC-model is the most sensitive model. We observe that the adapted models are more or less equally robust as their respective original models. The practical benefits of the β-PREC- and OWA-model are not lost due to the adaptations: both the a-β-PREC- and a-OWA-model perform well in the robustness analysis. However, it is not possible to decide which robust model performs best. We note that the robustness of the β-PREC- and OWA-models come with an extra computational cost. Assuming that the values of the similarity relation are known, the complexity of the IC-model is $\mathcal{O}(|U|)$, whereas the complexity of the β-PREC- and OWA-models is $\mathcal{O}(|U| \log(|U|))$ due to the sorting operations required by these models. The adapted β-PREC- and OWA-models have the same complexity as their respective original models, which means that the extra theoretical properties do not come with a higher complexity.

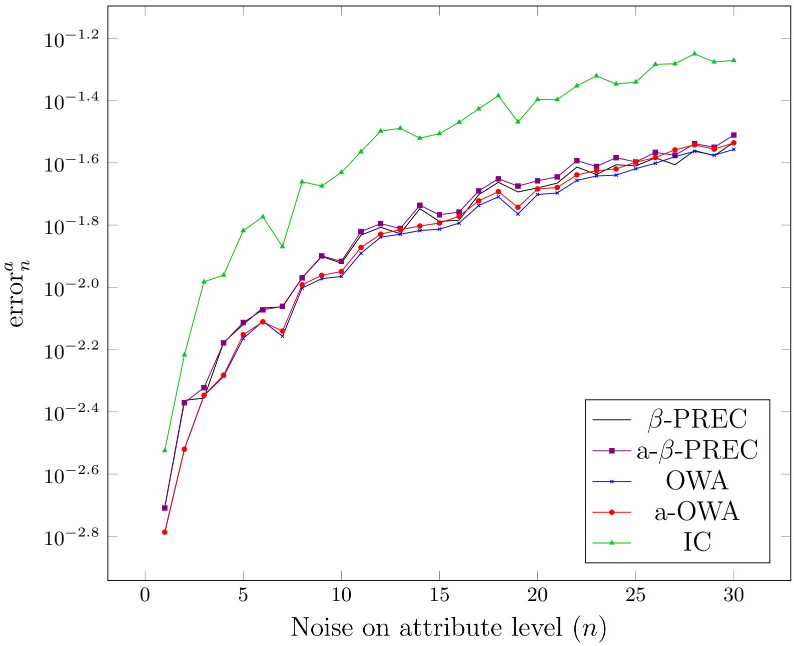

Fig. 1. Average error over the five data sets for attribute noise

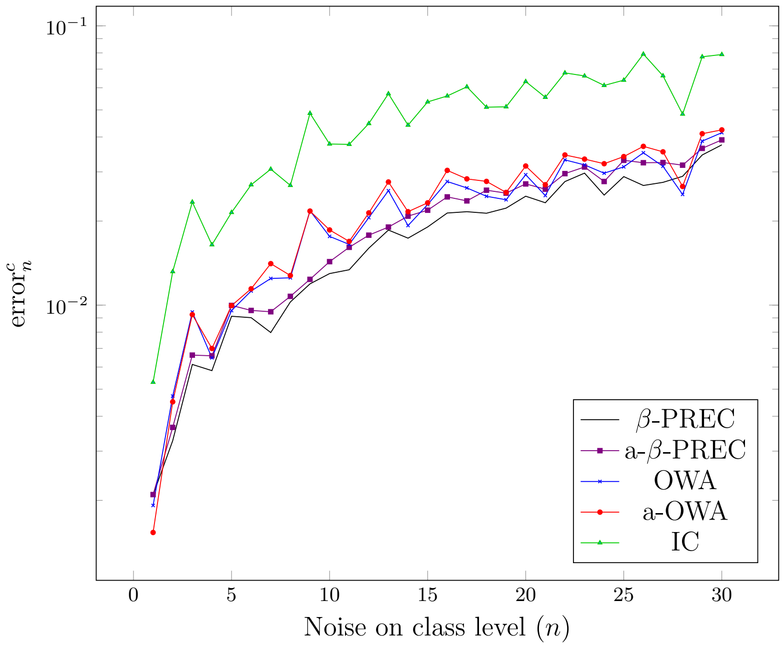

Fig. 2. Average error over the five data sets for class noise

5 Conclusion and Future Work

In this work, we adapted two state-of-the-art robust fuzzy rough set models such that the inclusion property is guaranteed, in order to obtain the required theoretical properties for using the models in feature selection. Furthermore, we compared the robust models to a general implicator-conjunctor-based fuzzy rough set model from a practical point of view. The benefits of the original models are not lost due to the proposed adaptation.

Future work consists of studying a formal framework for data reduction techniques based on fuzzy rough set models, and in particular, for the implicator-conjunctor-based model and the two adapted robust fuzzy rough set models.

Acknowledgment. Lynn D'eer has been supported by the Ghent University Special Research Fund.

References

1. Pawlak, Z.: Rough sets. International Journal of Computer and Information Sciences 11(5), 341–356 (1982)
2. Zadeh, L.: Fuzzy sets. Information and Control 8, 338–353 (1965)

3. Dubois, D., Prade, H.: Rough fuzzy sets and fuzzy rough sets. International Journal of General Systems 17, 191–209 (1990)
4. D'eer, L., Verbiest, N., Cornelis, C., Godo, L.: Implicator-conjunctor based models of fuzzy rough sets: definitions and properties. In: Ciucci, D., Inuiguchi, M., Yao, Y., Ślęzak, D., Wang, G. (eds.) RSFDGrC 2013. LNCS, vol. 8170, pp. 169–179. Springer, Heidelberg (2013)
5. Fernández Salido, J., Murakami, S.: On β-precision aggregation. Fuzzy Sets and Systems 139, 547–558 (2003)
6. Fernández Salido, J., Murakami, S.: 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)
7. Cornelis, C., Verbiest, N., Jensen, R.: Ordered weighted average based fuzzy rough sets. In: Yu, J., Greco, S., Lingras, P., Wang, G., Skowron, A. (eds.) RSKT 2010. LNCS, vol. 6401, pp. 78–85. Springer, Heidelberg (2010)
8. Cornelis, C., Jensen, R., Hurtado Martín, G., Slezak, D.: Attribute selection with fuzzy decision reducts. Information Sciences 180(2), 209–224 (2010)
9. Inuiguchi, M.: Classification- versus approximation-oriented fuzzy rough sets. In: Bouchon-Meunier, B., Coletti, G., Yager, R. (eds.) Proceedings of Information Processing and Management of Uncertainty in Knowledge-Based Systems. CD-ROM (2004)
10. Mieszkowicz-Rolka, A., Rolka, L.: Fuzzy rough approximations of process data. International Journal of Approximate Reasoning 49, 301–315 (2008)
11. Hu, Q., An, S., Yu, X., Yu, D.: Robust fuzzy rough classifiers. Fuzzy Sets and Systems 183, 26–43 (2011)
12. Yao, Y., Mi, J., Li, Z.: A novel variable precision (θ, σ)-fuzzy rough set model based on fuzzy granules. Fuzzy Sets and Systems 236, 58–72 (2014)
13. Cornelis, C., De Cock, M., Radzikowska, A.: Vaguely quantified rough sets. In: An, A., Stefanowski, J., Ramanna, S., Butz, C.J., Pedrycz, W., Wang, G. (eds.) RSFDGrC 2007. LNCS (LNAI), vol. 4482, pp. 87–94. Springer, Heidelberg (2007)
14. Zhao, S., Tsang, E., Chen, D.: The model of fuzzy variable precision rough sets. IEEE 17(2), 451–467 (2009)
15. D'eer, L., Verbiest, N., Cornelis, C., Godo, L.: Modelos de conjuntos rugosos difusos tolerantes al ruido: definiciones y propiedades. In: Proceedings of XVII Congreso Español sobre Tecnologías y Lógica Fuzzy (ESTYLF 2014), pp. 27–32 (2014)
16. Yager, R.: On ordered weighted averaging aggregation operators in multicriteria decision making. IEEE Transactions on Systems, Man and Cybernetics 18, 183–190 (1988)
17. Verbiest, N., Cornelis, C., Herrera, F.: FRPS: a fuzzy rough prototype selection method. Pattern Recognition 46(10), 2770–2782 (2013)
18. Verbiest, N., Cornelis, C., Jensen, R.: Quality, frequency and similarity based fuzzy nearest neighbor classification. In: Proceedings of the IEEE International Conference on Fuzzy Systems (2013)