Aberystwyth University

Ordered Weighted Average Based Fuzzy Rough Sets
Jensen, Richard; Verbiest, Nele; Cornelis, Chris

Ordered Weighted Average Based
Fuzzy Rough Sets

Chris Cornelis[*1], Nele Verbiest[1], and Richard Jensen[2]

[1] Department of Applied Mathematics and Computer Science,
Ghent University, Gent, Belgium
{Chris.Cornelis,Nele.Verbiest}@UGent.be
[2] Department of Computer Science, Aberystwyth University, Wales, UK
rkj@aber.ac.uk

Abstract. Traditionally, membership to the fuzzy-rough lower, resp. upper approximation is determined by looking only at the worst, resp. best performing object. Consequently, when applied to data analysis problems, these approximations are sensitive to noisy and/or outlying samples. In this paper, we advocate a mitigated approach, in which membership to the lower and upper approximation is determined by means of an aggregation process using ordered weighted average operators. In comparison to the previously introduced vaguely quantified rough set model, which is based on a similar rationale, our proposal has the advantage that the approximations are monotonous w.r.t. the used fuzzy indiscernibility relation. Initial experiments involving a feature selection application confirm the potential of the OWA-based model.

Key words: fuzzy rough sets, vaguely quantified rough sets, ordered weighted average, aggregation operators, noise tolerance, data analysis

1   Introduction

The operations of lower and upper approximation are at the heart of rough set theory [6] and many of its applications; based on objects' indiscernibility, they determine those objects that certainly, resp. possibly belong to a given concept. More specifically, an object belongs to the lower approximation of a concept if all objects indiscernible from it belong to the concept as well, and to its upper approximation if at least one object indiscernible from it belongs to the concept.

Fuzzy rough set theory [5] extends the approximation operators by allowing the indiscernibility relation as well as the concept itself to become fuzzy. While this generalization provides for greater flexibility, the commonly used definitions of fuzzy-rough approximations (see e.g. [7]) do not address the full potential of hybridization; in particular, the process of determining the membership degrees to the approximations still depends on a single object, as dictated by the

quantifiers $\forall$ and $\exists$ from the crisp definitions. As a consequence of this non-compensatory approach, slight changes to the data sometimes result in drastically different approximations. This, in turn, impacts the robustness of data analysis applications based on them, such as feature selection and classification.

These considerations inspired the vaguely quantified rough set (VQRS) model [1]: by replacing the crisp quantifiers $\forall$ and $\exists$ by softer versions representing *most* and *some*, it ensures that several objects contribute to an object's membership degree to the approximations. Unfortunately, this benefit goes at the expense of certain theoretically desirable properties for (fuzzy-)rough approximations (see e.g. [2] for an overview). Most importantly, as was noted in [3], the failure of monotonicity of the VQRS lower approximation w.r.t. the fuzzy indiscernibility relation hampers feature selection based on it.

In this paper, we propose an alternative fuzzy-rough hybridization based on ordered weighted average (OWA) operators [9], in which membership degrees to the approximations are computed by an aggregation process. As such, like the VQRS approach, our proposal allows for compensation; on the other hand, compared to VQRS, the OWA-based approach has a number of important benefits: 1) it is monotonous w.r.t. the fuzzy indiscernibility relation, 2) the traditional fuzzy-rough approximations can be recovered by a particular choice of the OWA weight vectors, and 3) the VQRS rationale can be maintained by introducing vague quantifiers into the OWA model, following the proposal in [10].

The remainder of the paper is organized as follows: in Section 2, we recall necessary preliminaries about fuzzy-rough hybrizidation and OWA operators, while in Section 3 we introduce OWA-based fuzzy rough sets, give some examples and examine their monotonicity characteristics. In Section 4, we apply our model to a feature selection task, comparing it to the traditional fuzzy-rough approach as well as to the VQRS model. Finally, in Section 5, we conclude and outline future work.

2   Preliminaries

2.1   Fuzzy Rough Sets

According to Pawlak [6], the lower and upper approximation of a crisp set $A \subseteq X$ w.r.t. an equivalence relation R are defined by, for y in X,

$$y \in R{\downarrow}A \text{ iff } [y]_R \subseteq A \tag{1}$$

$$y \in R{\uparrow}A \text{ iff } [y]_R \cap A \neq \emptyset \tag{2}$$

or, equivalently,

$$y \in R{\downarrow}A \text{ iff } (\forall x \in X)((x,y) \in R \Rightarrow x \in A) \tag{3}$$

$$y \in R{\uparrow}A \text{ iff } (\exists x \in X)((x,y) \in R \wedge x \in A) \tag{4}$$

When A is a fuzzy set and R is a fuzzy relation[3] in X, equations (3) and (4) can be extended using a fuzzy implication[4] $\mathcal{I}$ and a t-norm[5] $\mathcal{T}$ to

$$(R{\downarrow}A)(y) = \inf_{x \in X}(\mathcal{I}(R(x,y), A(x))) \tag{5}$$

$$(R{\uparrow}A)(y) = \sup_{x \in X}(\mathcal{T}(R(x,y), A(x))) \tag{6}$$

for $y \in X$. [7] When A and R are both crisp, (1) and (2) are recovered.

2.2   Vaguely Quantified Rough Sets

The inf and sup operations in (5) and (6) play the same role as the $\forall$ and $\exists$ quantifiers in (1) and (2). Because of this, a change to a single object can have a large impact on the approximations. This makes fuzzy rough sets equally susceptible to noisy data —which is difficult to rule out in real-life applications— as their crisp counterparts.

To make up for this shortcoming, it was proposed in [1] to replace the universal and existential quantifier by means of vague quantifiers like *most* and *some*. Mathematically, vague quantifiers are modeled by a regularly increasing fuzzy quantifier: an increasing $[0, 1] \to [0, 1]$ mapping Q that satisfies the boundary conditions $Q(0) = 0$ and $Q(1) = 1$. As an example, the following parametrized formula, for $0 \le \alpha < \beta \le 1$, and x in $[0, 1]$,

$$Q_{(\alpha,\beta)}(x) = \begin{cases} 0, & x \le \alpha \\ \frac{2(x-\alpha)^2}{(\beta-\alpha)^2}, & \alpha \le x \le \frac{\alpha+\beta}{2} \\ 1 - \frac{2(x-\beta)^2}{(\beta-\alpha)^2}, & \frac{\alpha+\beta}{2} \le x \le \beta \\ 1, & \beta \le x \end{cases} \tag{7}$$

generates a regularly increasing fuzzy quantifier. For instance, $Q_{(0.2,1)}$ and $Q_{(0,0.6)}$ may be used to model the vague quantifiers *most* and *some*. Once a couple (Q_l, Q_u) of fuzzy quantifiers is fixed, the Q_l-lower and Q_u-upper approximation of a fuzzy set A under a fuzzy relation R are defined by, for all y in X,

$$(R{\downarrow}_{Q_l}A)(y) = Q_l\left(\frac{|[y]_R \cap A|}{|[y]_R|}\right) \tag{8}$$

$$(R{\uparrow}_{Q_u}A)(y) = Q_u\left(\frac{|[y]_R \cap A|}{|[y]_R|}\right) \tag{9}$$

---

[3] Typically, it is assumed that R is at least a fuzzy tolerance relation, i.e., R is reflexive and symmetric.

[4] A fuzzy implication $\mathcal{I}$ is a $[0, 1]^2 \to [0, 1]$ mapping which is decreasing in its first argument, and increasing in its second argument, and which satisfies $\mathcal{I}(0, 0) = \mathcal{I}(0, 1) = \mathcal{I}(1, 1) = 1$ and $\mathcal{I}(1, 0) = 0$.

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

2.3   Ordered Weighted Average Aggregation

The OWA operator [9] models an aggregation process in which a sequence A of n scalar values are ordered decreasingly and then weighted according to their ordered position by a weighting vector $W = \langle w_i \rangle$, such that $w_i \in [0,1]$ and $\Sigma_i^n w_i = 1$. In particular, if c_i represents the ith largest value in A,

$$OWA_W(A) = \sum_{i=1}^{n} w_i c_i \tag{10}$$

The OWA's main strength is its flexibility, since it enables us to model a wide range of aggregation strategies. For example, the maximum, minimum and average can all be modelled by means of OWA operators:

1. Maximum: $W_{max} = \langle w_i \rangle$, where $w_1 = 1$, $w_i = 0$, $i \neq 1$
2. Minimum: $W_{min} = \langle w_i \rangle$, where $w_n = 1$, $w_i = 0$, $i \neq n$
3. Average: $W_{avg} = \langle w_i \rangle$, where $w_i = \frac{1}{n}$, $i = 1, \ldots, n$

The OWA operator can be analysed by several measures, among which the orness-degree and andness-degree that compute how similar its behaviour is to that of max, respectively min:

$$orness(W) = \frac{1}{n-1} \sum_{i=1}^{n} ((n-i) \cdot w_i) \tag{11}$$

$$andness(W) = 1 - orness(W) \tag{12}$$

Note that $orness(W_{max}) = 1$, $andness(W_{min}) = 1$ and $orness(W_{avg}) = 0.5$.

3   OWA-Based Lower and Upper Approximation

Let R be a fuzzy relation in X and A a fuzzy set in $X = \{x_1, \ldots, x_n\}$. Moreover, let $\mathcal{T}$ be a t-norm and $\mathcal{I}$ a fuzzy implication. The OWA-based lower and upper approximation of A under R with weight vectors W_l and W_u are defined as

$$(R\downarrow_{W_l} A)(y) = OWA_{W_l}\langle \mathcal{I}(R(x_i,y), A(x_i)) \rangle \tag{13}$$

$$(R\uparrow_{W_u} A)(y) = OWA_{W_u}\langle \mathcal{T}(R(x_i,y), A(x_i)) \rangle \tag{14}$$

In order to distinguish the behaviour of lower and upper approximation, we enforce the conditions $andness(W_l) > 0.5$ and $orness(W_u) > 0.5$. Note that the traditional lower and upper approximation are retrieved when $W_l = W^{min}$ and $W_u = W^{max}$. Below, we give two examples of constructing the OWA weight vectors in order to relax the traditional definitions.

Example 1. Let $m \leq n$. It is possible to define $W_l = \langle w_i^l \rangle$ and $W_u = \langle w_i^u \rangle$ as

$$w_{n+1-i}^l = \begin{cases} \frac{2^{m-i}}{2^m - 1} & i = 1, \ldots, m \\ 0 & i = m+1, \ldots, n \end{cases} \tag{15}$$

$$w_i^u = \begin{cases} \frac{2^{m-i}}{2^m - 1} & i = 1, \ldots, m \\ 0 & i = m+1, \ldots, n \end{cases} \tag{16}$$

It can be verified that $andness(W_l) > 0.5$ and $orness(W_u) > 0.5$. E.g., if $n = 8$, and $m = 3$, $W_l = \langle 0, 0, 0, 0, 0, 1/7, 2/7, 4/7 \rangle$.

Example 2. In [10], Yager proposed to simulate vague quantifiers by means of OWA operators. In particular, given regularly increasing fuzzy quantifiers Q_l and Q_u, weight vectors $W_{Q_l} = \langle w_i^l \rangle$ and $W_{Q_u} = \langle w_i^u \rangle$ can be defined as

$$w_i^l = Q_l \left(\frac{i}{n} \right) - Q_l \left(\frac{i-1}{n} \right) \tag{17}$$

$$w_i^u = Q_u \left(\frac{i}{n} \right) - Q_u \left(\frac{i-1}{n} \right) \tag{18}$$

It can be verified that if $Q_l = Q_{(\alpha,1)}$ (resp., $Q_u = Q_{(0,\beta)}$), then $andness(W_{Q_l}) > 0.5$ (resp., $orness(W_{Q_u}) > 0.5$). For instance, if $n = 8$, and $Q_l = Q_{(0.2,1)}$, $W_{Q_l} = \langle 0, 0.01, 0.09, 0.18, 0.28, 0.24, 0.14, 0.06 \rangle$.

Next, we list two important propositions which follow from the monotonicity of $\mathcal{I}$, $\mathcal{T}$, and OWA operators.

Proposition 1. *Let $A_1 \subseteq A_2$ be fuzzy sets in X, and R be a fuzzy relation in X. Then $R{\downarrow}_{W_l} A_1 \subseteq R{\downarrow}_{W_l} A_2$ and $R{\uparrow}_{W_u} A_1 \subseteq R{\uparrow}_{W_u} A_2$.*

Proposition 2. *Let $R_1 \subseteq R_2$ be fuzzy relations in X, and A be a fuzzy set in X. Then $R_1{\downarrow}_{W_l} A \supseteq R_2{\downarrow}_{W_l} A$ and $R_1{\uparrow}_{W_u} A \supseteq R_2{\uparrow}_{W_u} A$.*

Note that the second proposition does not hold for VQRS [3], which is due to the fact that R occurs both in the numerator and denominator of (8) and (9).

4   Application to Feature Selection

4.1   Fuzzy-Rough Feature Selection

The purpose of feature selection is to eliminate redundant or misleading attributes from a data set, typically with the purpose of creating faster and more accurate classifiers. For our purposes, a decision system $(X, \mathcal{A} \cup \{d\})$ consists of non-empty sets of objects $X = \{x_1, ..., x_n\}$ and conditional attributes (features) $\mathcal{A} = \{a_1, ..., a_n\}$, together with a decision attribute $d \notin \mathcal{A}$. Attributes can be either quantitative or discrete; in this paper, we assume that d is always discrete. To express the approximate equality between two objects w.r.t. a quantitative attribute a, in this paper we use the fuzzy relation R_a, defined by, for x and y in X (σ_a denotes the standard deviation of a):

$$R_a(x, y) = \max \left(\min \left(\frac{a(y) - a(x) + \sigma_a}{\sigma_a}, \frac{a(x) - a(y) + \sigma_a}{\sigma_a} \right), 0 \right) \tag{19}$$

For a discrete attribute a, $R_a(x, y) = 1$ if $a(x) = a(y)$ and $R_a(x, y) = 0$ otherwise. Given a t-norm $\mathcal{T}$, for any $B \subseteq \mathcal{A}$, the fuzzy B-indiscernibility relation can be defined by

$$R_B(x, y) = \mathcal{T}(\underbrace{R_a(x, y)}_{a \in B}) \tag{20}$$

The fuzzy B-positive region is then defined as the union of lower approximations of all decision classes w.r.t. R_B. In particular, for y in X, we have [4]

$$POS_B(y) = (R_B{\downarrow}[y]_{R_d})(y) \tag{21}$$

The predictive ability w.r.t. d of the attributes in B can be measured by the degree of dependency of d on B, defined as $\gamma_B = |POS_B|/|X|$.

The goal of fuzzy-rough feature selection is then to find decision reducts: subsets B of A that satisfy $\gamma_B = \gamma_A$ and cannot be further reduced, i.e., there exists no proper subset B' of B such that $\gamma_{B'} = \gamma_A$. In practice, minimality is often not necessary, and the corresponding subsets are called superreducts. In order to obtain a single superreduct of $(X, \mathcal{A} \cup \{d\})$, in this paper we use the following hillclimbing heuristic: starting with an empty set B, we compute $\gamma_{B \cup \{a\}}$ for every attribute a and add the attribute for which this value is highest to B. This process is repeated for the remaining attributes until $\gamma_B = \gamma_A$.

4.2   VQRS- and OWA-Based Feature Selection

Since the fuzzy positive region in (21) uses the "traditional" lower approximation (5), it is sensitive to small changes in the data. For this reason, following [3], we can replace $\downarrow$ by $\downarrow_{Q_l}$, giving rise to[6]

$$POS_B^{VQRS}(y) = (R_B \downarrow_{Q_l} [y]_{R_d})(y) \tag{22}$$

Based on this, γ_B^{VQRS} may be defined analogously to γ_B. However, when using it as an evaluation measure in the hillclimbing heuristic from Section 4.1, the non-monotonicity of $\downarrow_{Q_l}$ w.r.t. R_B may result in $\gamma_{B \cup \{a\}}$ being lower than γ_B, a counterintuitive result. This problem may be mended by replacing (22) by

$$POS_B^{OWA}(y) = (R_B \downarrow_{W_l} [y]_{R_d})(y) \tag{23}$$

and defining γ_B^{OWA} analogously as before. Because of Proposition 2, applying the hillclimbing heuristic will always result in increasing evaluation values for the consecutively selected subsets, making the OWA-based feature selection intuitively sounder.

Because of the definition of $\downarrow_{W_l}$, $\gamma_B \leq \gamma_B^{OWA}$; equality holds if $W_l = W_{min}$. In general, no order relationship exists between γ_B and γ_B^{VQRS}, or between γ_B^{VQRS} and γ_B^{OWA}. On the other hand, under certain conditions it holds that subsets obtained with the hillclimbing approach using the VQRS- and OWA-based positive region, are also decision superreducts in terms of the original definition of positive region in (21), and vice versa. For VQRS, $\gamma_B = 1 \Leftrightarrow \gamma_B^{VQRS} = 1$ if $Q_l = Q_{(\alpha, 1)}$ and $\mathcal{I}$ satisfies $x \leq y \Leftrightarrow \mathcal{I}(x, y) = 1$ [3]. The following proposition provides a sufficient and necessary condition for the OWA-based approach.

---

[6] Note that [3] also introduced an alternative definition of the VQRS-based positive region which is not considered here because of its excessive computational complexity.

Proposition 3. $W_l = \langle w_i^l \rangle$ *is an OWA weight vector such that* $w_n^l > 0$ *if and only if* $\gamma_B^{OWA} = 1 \Leftrightarrow \gamma_B = 1$.

4.3 Experimental Evaluation

In order to compare the performance of the traditional fuzzy-rough feature selection approach with those based on VQRS and OWA, we have performed an experiment on the Spambase dataset[7]. This dataset has 4601 objects, 57 conditional attributes and two decision classes.

Specifically, we apply the hillclimbing procedure from Section 4.1 using γ, γ^{VQRS} and γ^{OWA}; we use $\mathcal{I}(x,y) = \min(1, 1-x+y)$ in (5), $Q_l = Q_{(0.2,1)}$ in (8), $W_l = W_{Q_{(0.2,1)}}$ in (13) and $\mathcal{T}(x,y) = \max(0, x+y-1)$ in (20). At each step in the hillclimbing algorithm, the subsets obtained so far are evaluated by means of their classification accuracy using the Nearest Neighbour classifier IBk in the Weka toolkit [8]. As a baseline, we also consider the approach in which attributes are added in random order.
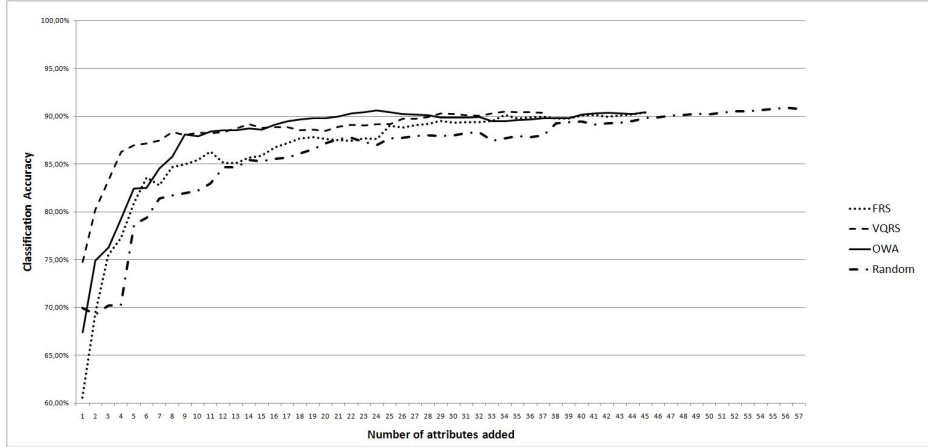
Fig. 1. Evaluation of feature selection approaches on Spambase data.

The results can be found in Figure 1. The X axis contains the number of attributes added so far, while the classification accuracy obtained with IBk for the corresponding attribute subsets can be read on the Y axis. From this figure, we can see that all fuzzy-rough approaches yield better subsets than those randomly generated, and that VQRS and OWA both outperform the fuzzy-rough approach along most of the range. Initially, the VQRS approach selects higher-quality attributes than those obtained with OWA, but roles switch after the sixteenth attribute. It is also interesting to note that OWA obtains the highest overall accuracy, viz. 90.59% using 24 attributes.

---

[7] Available from http://archive.ics.uci.edu/ml/datasets/Spambase

5   Conclusion and Future Work

In this paper, we have presented a new fuzzy-rough hybridization which relaxes the traditional property that lower and upper approximations are determined by means of a single object. As such, it provides an alternative to the previously introduced VQRS model; from a theoretical point of view, the new model is better-founded because it respects monotonicity w.r.t. the used fuzzy indiscernibility relation. As another advantage, it adheres more closely to the traditional fuzzy-rough approach, to which it converges as the andness (resp., orness) of the OWA-based lower (resp., upper) approximation nears 1. Initial experimentation has pointed out that the OWA-based model can outperform the traditional fuzzy-rough approach, and that it can compete with the VQRS approach.

More extensive experiments are needed, however, to fully investigate its performance; in particular, in order to study its noise-handling potential, we will evaluate the quality of subsets obtained on data that is deliberately contaminated with noise. Furthermore, we also plan to investigate various construction methods for the OWA weight vectors, and their optimization in function of the particular dataset used. Finally, apart from feature selection, the OWA-based model may also be applied to other data analysis tasks in which fuzzy-rough methods have been used, such as rule induction and instance selection.

References

1. Cornelis, C., De Cock, M., Radzikowska, A.M.: Vaguely quantified rough sets, Proceedings of 11th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing (RSFDGrC2007), Lecture Notes in Artificial Intelligence 4482, p. 87-94, 2007.
2. Cornelis, C., De Cock, M., Radzikowska, A.M.: Fuzzy rough sets: from theory into practice, Handbook of Granular Computing (W. Pedrycz, A. Skowron, V. Kreinovich, eds.), John Wiley and Sons, 2008, pp. 533–552.
3. Cornelis, C., Jensen, R.: A noise tolerant approach to fuzzy-rough feature selection, Proceedings of the 2008 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2008), p. 1598-1605, 2008.
4. Cornelis, C., Jensen, R., Hurtado Martín, G., Slezak, D: Attribute selection with fuzzy decision reducts, Information Sciences **180(2)** (2010) 209–224
5. Dubois, D., Prade, H.: Rough fuzzy sets and fuzzy rough sets. International Journal of General Systems **17** (1990) 91–209
6. Pawlak, Z.: Rough sets. International Journal of Computer and Information Sciences **11(5)** (1982) 341–356
7. Radzikowska, A.M., Kerre, E.E.: A comparative study of fuzzy rough sets. Fuzzy Sets and Systems **126** (2002) 137–156
8. Witten, I.H., Frank, E.: Data Mining: Practical machine learning tools and techniques, 2nd Edition, Morgan Kaufmann, San Francisco, 2005.
9. Yager, R.R: On ordered weighted averaging aggregation operators in multicriteria decision making. IEEE Transactions on Systems, Man, and Cybernetics **18** (1988) 183–190.
10. Yager, R.R.: Families of OWA operators. Fuzzy Sets and Systems **59(2)**(1993) 125–148.