Vaguely Quantified Rough Sets

Chris Cornelis[1], Martine De Cock[1], and Anna Maria Radzikowska[2]

[1] Computational Web Intelligence, Department of Mathematics and Computer Science, Ghent University, Krijgslaan 281 (S9), B-9000 Gent, Belgium
{Chris.Cornelis,Martine.DeCock}@UGent.be

[2] Faculty of Mathematics and Information Science, Warsaw University of Technology, Plac Politechniki 1, 00-661 Warsaw, Poland
annrad@mini.pw.edu.pl

Abstract. The hybridization of rough sets and fuzzy sets has focused on creating an end product that extends both contributing computing paradigms in a conservative way. As a result, the hybrid theory inherits their respective strengths, but also exhibits some weaknesses. In particular, although they allow for gradual membership, fuzzy rough sets are still abrupt in a sense that adding or omitting a single element may drastically alter the outcome of the approximations. In this paper, we revisit the hybridization process by introducing vague quantifiers like "some" or "most" into the definition of upper and lower approximation. The resulting vaguely quantified rough set (VQRS) model is closely related to Ziarko's variable precision rough set (VPRS) model.

Keywords: vague quantifiers, fuzzy sets, rough sets, VPRS model

1 Introduction

In rough set theory, an object belongs to the upper approximation of a set as soon as it is related to *one* of the elements in the set, while the lower approximation only retains those objects related to *all* the elements in the set. This is due to the use of an existential quantifier in the definition of upper approximation, and of a universal quantifier for the lower approximation. In applications that use real-life data (which is usually noisy to some extent, and hence prone to classification errors and inconsistency), the definition of upper approximation might be too loose (easily resulting in very large sets), while the definition of lower approximation might be too strict (easily resulting in the empty set). A similar phenomenon can be observed at the level of fuzzy rough set theory, where the $\exists$ and $\forall$ quantifiers are replaced by the sup and inf operations (see e.g. [1, 6]), which prove just as susceptible to noise as their crisp counterparts.

In his variable precision rough set (VPRS) model, Ziarko [8, 9] introduced thresholds to deal with these problems in the crisp case. In general, given $0 \leq l < u \leq 1$, an element y is added to the lower approximation of a set A if at least $100 * u$ percent of the elements related to y are in A. Likewise, y belongs to the upper approximation of A if more than $100 * l$ percent of the elements related to y

are in A. This can be interpreted as a generalization of the rough set model using crisp quantifiers *at least* $100 * u$ *percent* and *more than* $100 * l$ *percent* to replace the universal quantifier (which corresponds to "at least 100 percent") and the existential quantifier (which corresponds to "more than 0 percent"). Also, some attempts have been made to pursue this approach within the fuzzy rough set model, having in common that they still rely on the use of crisp thresholds l and u (see e.g. [2, 4]).

In this paper, we go one step further by introducing vague quantifiers like *most* and *some* into the model. In this way, an element y belongs to the lower approximation of A if most of the elements related to y are included in A. Likewise, an element belongs to the upper approximation of A if some of the elements related to y are included in A. Mathematically, we model vague quantifiers in terms of Zadeh's notion of fuzzy quantifiers [7]. As such, the new model inherits both the flexibility of VPRSs for dealing with classification errors (by relaxing the membership *conditions* for the lower approximation, and tightening those for the upper approximation) and that of fuzzy sets for expressing partial constraint satisfaction (by distinguishing different *levels* of membership to the upper/lower approximation). Moreover, we illustrate that the model can be used in a meaningful way, regardless of whether the relation R and the set A to be approximated are crisp or fuzzy. In each case, the outcome of the approximations will be a pair of fuzzy sets delineating A in a flexible way.

The remainder of this paper is structured as follows. In Section 2, we review basic notions of classical rough sets and VPRSs, while Section 3 introduces vaguely quantified rough sets in the crisp case, and illustrates their relevance in the context of information retrieval. In Section 4, we lift the VQRS paradigm to the level of fuzzy rough set theory, distinguish it from related work that combines VPRSs with fuzzy sets and detail an experiment on a benchmark dataset to show the performance of the proposed extension vis-à-vis the classical approach in a rough data analysis problem. Finally, in Section 5, we conclude.

2 Variable Precision Rough Sets

Recall that the traditional upper and lower approximation [5] of a set A in the approximation space (X, R) are defined by

$$y \in R{\uparrow}A \text{ iff } A \cap Ry \neq \emptyset \tag{1}$$

$$y \in R{\downarrow}A \text{ iff } Ry \subseteq A \tag{2}$$

in which Ry is used to denote the equivalence class (also called R-foreset) of y. Furthermore, the rough membership function R_A of A is defined by

$$R_A(y) = \frac{|Ry \cap A|}{|Ry|} \tag{3}$$

$R_A(y)$ quantifies the degree of inclusion of Ry into A, and can be interpreted as the conditional probability that y belongs to A, given knowledge about the equivalence class Ry that y belongs to. One can easily verify that

$$y \in R{\uparrow}A \text{ iff } R_A(y) > 0 \tag{4}$$

$$y \in R{\downarrow}A \text{ iff } R_A(y) = 1 \tag{5}$$

In other words, y is added to the upper approximation as soon as Ry overlaps with A, while even a small inclusion error of Ry in A results in the rejection of the whole class from the lower approximation.

Example 1. Consider a document collection $D = \{d_1, \ldots, d_{20}\}$ in which the documents are arranged according to topic into four categories : $D_1 = \{d_1, \ldots, d_5\}$, $D_2 = \{d_6, \ldots, d_{10}\}$, $D_3 = \{d_{11}, \ldots, d_{15}\}$ and $D_4 = \{d_{15}, \ldots, d_{20}\}$. Hence, the categorization defines an equivalence relation R on X. Suppose now that a user launches a query, and that the relevant documents turn out to be (automatically determined) the set $A = \{d_2, \ldots, d_{12}\}$. This suggests that the information retrieval system simply might have missed d_1 since all other documents from D_1 are in A. Furthermore, the fact that only d_{11} and d_{12} are retrieved from D_3 might indicate that these documents are less relevant to the query than the documents of D_2, which all belong to A. Pawlak's original rough set approach does not allow to reflect these nuances, since $R{\downarrow}A = D_2$ and $R{\uparrow}A = D_1 \cup D_2 \cup D_3$, treating D_1 and D_3 in the same way.

Since in real life, data may be affected by classification errors caused by humans or noise, Ziarko [9] relaxes the constraints in (4) and (5) to obtain the following parameterized definitions:

$$y \in R{\uparrow}_l A \text{ iff } R_A(y) > l \tag{6}$$

$$y \in R{\downarrow}_u A \text{ iff } R_A(y) \geq u \tag{7}$$

Formulas (4)–(7) can also be read in terms of quantifiers, i.e.

$$y \in R{\uparrow}A \text{ iff } (\exists x \in X)((x, y) \in R \wedge x \in A) \tag{8}$$

$$y \in R{\downarrow}A \text{ iff } (\forall x \in X)((x, y) \in R \Rightarrow x \in A) \tag{9}$$

$$y \in R{\uparrow}_l A \text{ iff more than } 100 * l\% \text{ elements of } Ry \text{ are in } A \tag{10}$$

$$y \in R{\downarrow}_u A \text{ iff at least } 100 * u\% \text{ elements of } Ry \text{ are in } A \tag{11}$$

Note that the quantifiers used above are all crisp: the existential quantifier $\exists$, the universal quantifier $\forall$, as well as two threshold quantifiers $> 100 * l\%$ and $\geq 100 * u\%$. As such, although the VPRS model warrants a measure of tolerance towards problematic elements, it still treats them in a black-or-white fashion: depending on the specific choice of l and u, an element either fully belongs, or does not belong to the upper or lower approximation.

Example 2. Let us return to the document retrieval problem from Example 1. Ziarko's model offers more flexibility to distinguish the roles of D_1 and D_3, but the choice of the thresholds is crucial. In a symmetric VPRS model, l is chosen equal to $1 - u$ [9]. For $u = 0.8$ we obtain $R{\downarrow}_{.8}A = D_1 \cup D_2$ and $R{\uparrow}_{.2}A = D_1 \cup D_2 \cup D_3$. For $u = 0.9$, however, we obtain the same results as in Example 1.

3 Vaguely Quantified Rough Sets

The VPRS definitions for upper and lower approximation from the previous section can be softened by introducing vague quantifiers, to express that y belongs to the upper approximation of A to the extent that *some* elements of Ry are in A, and y belongs the lower approximation of A to the extent that *most* elements of Ry are in A. In this approach, it is implicitly assumed that the approximations are fuzzy sets, i.e., mappings from X to $[0,1]$, that evaluate to what degree the associated condition is fulfilled.

To model the quantifiers appropriately, we use Zadeh's concept of a fuzzy quantifier [7], i.e. a $[0,1] \to [0,1]$ mapping Q. Q is called regularly increasing if it is increasing and it satisfies the boundary conditions $Q(0) = 0$ and $Q(1) = 1$.

Example 3. Possible choices for Q are the existential and the universal quantifier

$$Q_{\exists}(x) = \begin{cases} 0, & x = 0 \\ 1, & x > 0 \end{cases} \qquad Q_{\forall}(x) = \begin{cases} 0, & x < 1 \\ 1, & x = 1 \end{cases}$$

for x in $[0,1]$, that will lead us to (4) and (5); or the quantifiers

$$Q_{>l}(x) = \begin{cases} 0, & x \leq l \\ 1, & x > l \end{cases} \qquad Q_{\geq u}(x) = \begin{cases} 0, & x < u \\ 1, & x \geq u \end{cases}$$

for x in $[0,1]$, that will lead us to (6) and (7).

Example 4. The quantifiers in Example 3 are crisp, in the sense that the outcome is either 0 or 1. An example of a fuzzy quantifier taking on also intermediate values is the following parametrized formula, for $0 \leq \alpha < \beta \leq 1$, and x in $[0,1]$,

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

For example, $Q_{(0.1,0.6)}$ and $Q_{(0.2,1)}$ could be used respectively to reflect the vague quantifiers *some* and *most* from natural language.

Given sets A_1 and A_2 in X and a fuzzy quantifier Q, Zadeh [7] computes the truth value of the statement "Q A_1's are also A_2's" by the formula

$$Q\left(\frac{|A_1 \cap A_2|}{|A_1|}\right) \tag{12}$$

Once we have fixed a couple (Q_l, Q_u) of fuzzy quantifiers, we can formally define the Q_l-upper and Q_u-lower approximation of A by

$$R{\uparrow}_{Q_l}A(y) = Q_l\left(\frac{|Ry \cap A|}{|Ry|}\right) = Q_l(R_A(y)) \tag{13}$$

$$R{\downarrow}_{Q_u}A(y) = Q_u\left(\frac{|Ry \cap A|}{|Ry|}\right) = Q_u(R_A(y)) \tag{14}$$

for all y in X. It is straightforward to verify that $R{\uparrow}_{Q_\exists}A = R{\uparrow}A$ and $R{\downarrow}_{Q_\forall}A = R{\downarrow}A$, and that $R{\uparrow}_{Q_{>l}}A = R{\uparrow}_l A$ and $R{\downarrow}_{Q_{\geq u}}A = R{\downarrow}_u A$. Moreover, if $Q_u \subseteq Q_l$, i.e., $Q_u(x) \leq Q_l(x)$ for all x in $[0,1]$, then $R{\downarrow}_{Q_u}A \subseteq R{\uparrow}_{Q_l}A$.

Example 5. Let us return once more to the document retrieval problem discussed in Example 1 and 2. In our VQRS model with fuzzy quantifiers $Q_u = Q_{(0.2,1)}$ and $Q_l = Q_{(0.1,0.6)}$ the lower approximation $R{\downarrow}_{Q_u}A$ equals

$$\{(x_6,1),\ldots,(x_{10},1),(x_1,0.875),\ldots,(x_5,0.875),(x_{11},0.125),\ldots,(x_{15},0.125)\}$$

In this weighted list a document ranks higher if most of the elements in its topic category are in A. The gradations reflect the different roles of the categories in a desirable way. For example, category D_3 is not excluded but its documents are presented only at the bottom of the list. A similar phenomenon occurs with the upper approximation $R{\uparrow}_{Q_l}A = \{(x_1,1),\ldots,(x_{10},1),(x_{11},0.68),\ldots,(x_{15},0.68)\}$.

4 Vaguely Quantified Fuzzy Rough Sets

As the definition of vaguely quantified rough sets brings together ideas from fuzzy sets and rough sets, it is instructive to examine their relationship to, and combine them with existing work on fuzzy-rough hybridization. Throughout this section, we assume that $\mathcal{T}$ is a triangular norm (t-norm for short), i.e., any increasing, commutative and associative $[0,1]^2 \to [0,1]$ mapping satisfying $\mathcal{T}(1,x) = x$, for all x in $[0,1]$, and that $\mathcal{I}$ is an implicator, i.e. any $[0,1]^2 \to [0,1]$-mapping $\mathcal{I}$ that is decreasing in its first, and increasing in its second component and that satisfies $\mathcal{I}(0,0) = 1, \mathcal{I}(1,x) = x$, for all x in $[0,1]$. We also assume that the upper and lower approximation of a fuzzy set A in X under a fuzzy relation R in X are defined by [6]

$$R{\uparrow}A(y) = \sup_{x\in X} \mathcal{T}(R(x,y),A(x)) \quad (15)$$

$$R{\downarrow}A(y) = \inf_{x\in X} \mathcal{I}(R(x,y),A(x)) \quad (16)$$

for y in X. Note how these formulas paraphrase the definitions (8) and (9) which hold in the crisp case. In particular, the sup and inf operations play the same role as the $\exists$ and $\forall$ quantifiers, and as such a change in a single element can still have a large impact on (15) and (16).

This observation has inspired some researchers to propose altered definitions of fuzzy-rough approximations in the spirit of the VPRS model. For example, Mieszkowicz-Rolka and Rolka [4] used the concept of a fuzzy inclusion set (based on an implicator) and the notion of α-inclusion error (based on α-level sets), while Fernández-Salido and Murakami [2] defined new approximations based on the so-called β-precision quasi minimum $\min_\beta$ and maximum $\max_\beta$ (aggregation operators dependent on a parameter β in $[0,1]$). A serious drawback of these models is that they still rely on crisp thresholds l and u like Ziarko's model, which requires a fairly complex and not wholly intuitive mathematical apparatus.

The VQRS approach, on the other hand, lends itself to a much smoother and more elegant fuzzification. In fact, formulas (13) and (14) can simply be maintained in the fuzzy case, i.e., for y in X we have

$$R{\uparrow}_{Q_l}A(y) = Q_l\left(\frac{|Ry \cap A|}{|Ry|}\right) \tag{17}$$

$$R{\downarrow}_{Q_u}A(y) = Q_u\left(\frac{|Ry \cap A|}{|Ry|}\right) \tag{18}$$

with the conventions that the R-foreset Ry is defined by $Ry(x) = R(x, y)$ for x in X, the intersection $A \cap B$ of two fuzzy sets A and B in X is defined by $(A \cap B)(x) = \min(A(x), B(x))$ and the cardinality $|A|$ of a fuzzy set A in X is defined by $\sum\limits_{x \in X} A(x)$.

It is interesting that no implicator appears inside the VQRS lower approximation (18), as opposed to (16). In fact, $|Ry \cap A|/|Ry|$ and $\inf\limits_{x \in X} \mathcal{I}(R(x, y), A(x))$ are considered in fuzzy set literature as two alternatives, to compute the inclusion degree of Ry into A, the former set- or frequency-based and the latter logic-based (see e.g. [3]).

To demonstrate that the VQRS construct offers a worthwhile alternative to the traditional "logic"-based operations of fuzzy rough set theory in the context of rough data analysis, we ran an experiment on the `housing` benchmark dataset[3]. This dataset concerns housing prices in suburbs of Boston; it has 506 instances, 13 conditional attributes (12 continuous, one binary) and a continuous class attribute called MEDV (median value of owner-occupied homes in $1000s).

The setup of our experiment is as follows. Based on the distribution of the data, we defined a fuzzy partition on the universe of MEDV, containing three fuzzy classes *low*, *medium* and *high* in the range [0,50] as shown in Figure 1a. We also defined a fuzzy relation R in the universe X of instances expressing indistinguishability between instances x_1 and x_2 based on the conditional attributes:

$$R(x_1, x_2) = \min_{i=1}^{13} \max\left(0, \min\left(1, 1.2 - \alpha\frac{|c_i(x_1) - c_i(x_2)|}{l(c_i)}\right)\right) \tag{19}$$

in which c_i denotes the i^{th} conditional attribute, $l(c_i)$ is its range, and α is a parameter ≥ 1.2 that determines the granularity of R (the higher α, the finer-grained the R-foresets).

We divided the instances into 11 folds for cross validation: in each step, we selected one fold as test set and used the remaining folds as training set X' to compute the lower approximation of each decision class. For traditional fuzzy rough sets, we used formula (16), with three popular implicators $\mathcal{I}_L$ (Lukasiewicz), $\mathcal{I}_{KD}$ (Kleene-Dienes), and $\mathcal{I}_G$ (Gödel) defined in Table 1. For the VQRS model, we used formula (18), with a fixed quantifier $Q_u = Q_{(0.2,1)}$ (shown in Figure 1b). We then predicted the membership of each test instance y to each class A

[3] available at http://www.ics.uci.edu/∼mlearn/MLRepository.html

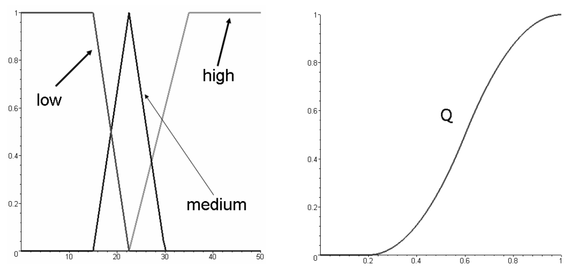

Fig. 1. a). Fuzzy partition of class attribute b) Fuzzy quantifier $Q_{(0.2,1)}$ for "most"

as the extent to which there exists a similar training instance x belonging to the previously learned lower approximation C of A:

$$\sup_{x \in X'} \mathcal{T}(R(x,y), C(x)) \tag{20}$$

In this formula, $\mathcal{T}$ is a t-norm; in our experiments, we used $\mathcal{T}_M$ (minimum) and $\mathcal{T}_L$ (Lukasiewicz), which are also shown in Table 1.[4]

Table 1. Implicators and t-norms used in the experiment.

$\mathcal{I}_L(x,y) = \min(1-x+y, 1)$	$\mathcal{T}_L(x,y) = \max(0, x+y-1)$
$\mathcal{I}_{KD}(x,y) = \max(1-x, y)$	$\mathcal{T}_M(x,y) = \min(x,y)$
$\mathcal{I}_G(x,y) = \begin{cases} 1, \text{ if } x \leq y \\ y, \text{ otherwise} \end{cases}$	

The average absolute error between the predicted and the actual membership values of the test instances was used as a metric for comparing the approaches. Also, we let α in (19) range from 2 to 8. From the results in Table 2, we observe that all approaches perform better for increasing values of α. This corresponds to the idea that a finer-grained relation allows for better approximation. However, for a too fine-grained relation, the average errors start increasing again, indicating an overfit of the model.

Comparing $\mathcal{I}_L$-$\mathcal{T}_L$ with VQRS-$\mathcal{T}_L$, we notice that in both cases the smallest error is obtained for $\alpha = 4$. The corresponding relation is still relatively coarse-grained. We observe that our VQRS-$\mathcal{T}_L$ approach is least hampered by this: it in fact achieves the lowest average error of all approaches displayed in the table. The approaches with $\mathcal{T}_M$ score worse in general, but again the smallest error is obtained with our VQRS-$\mathcal{T}_M$ model.

[4] $\mathcal{I}_{KD}$ and $\mathcal{I}_G$ are, respectively, the S-implicator and R-implicator of $\mathcal{T}_M$, while the S- and R-implicator of $\mathcal{T}_L$ coincide in $\mathcal{I}_L$.

Table 2. Experimental results for 11-fold cross-validation

α	$\mathcal{I}_L$-$\mathcal{T}_L$	$\mathcal{I}_{KD}$-$\mathcal{T}_M$	$\mathcal{I}_G$-$\mathcal{T}_M$	VQRS-$\mathcal{T}_L$	VQRS-$\mathcal{T}_M$
2	0.276	0.320	0.321	0.257	0298
3	0.264	0.301	0.315	0.238	0.280
4	**0.258**	0.288	0.299	**0.236**	0.265
5	0.263	0.268	0.274	0.246	**0.256**
6	0.272	**0.264**	**0.270**	0.261	0.258
7	0.282	0.269	0.271	0.274	0.266
8	0.291	0.280	0.280	0.286	0.279

5 Conclusion

In the VQRS model introduced in this paper, an element y belongs to the lower approximation of a set A to the extent that *most* elements related to y are in A. Likewise, y belongs to the upper approximation to the extent that *some* elements related to y are in A. The use of vague quantifiers "most" and "some", as opposed to the traditionally used crisp quantifiers "all" and "at least one" makes the model more robust in the presence of classification errors. Experimental results on the `housing` dataset show that VQRS consistently outperforms the classical approach.

Acknowledgment. Chris Cornelis would like to thank the Research Foundation—Flanders for funding his research.

References

1. Dubois, D., Prade, H.: Rough fuzzy sets and fuzzy rough sets. International Journal of General Systems **17** (1990) 91–209
2. Fernández Salido, J.M., Murakami, S.: Rough set analysis of a general type of fuzzy data using transitive aggregations of fuzzy similarity relations. Fuzzy Sets and Systems **139(3)** (2003) 635–660
3. Fodor, J., Yager, R.: Fuzzy set theoretic operators and quantifiers. Fundamentals of Fuzzy Sets (D. Dubois, H. Prade, eds.), Kluwer, Boston, Mass. (2000) 125–193
4. Mieszkowicz-Rolka, A., Rolka, L.: Variable precision fuzzy rough sets. Transactions on Rough Sets I. Lecture Notes in Computer Science **3100** (2004) 144–160
5. Pawlak, Z.: Rough sets. International Journal of Computer and Information Sciences **11(5)** (1982) 341–356
6. Radzikowska, A.M., Kerre, E.E.: A comparative study of fuzzy rough sets. Fuzzy Sets and Systems **126** (2002) 137–156
7. Zadeh, L.A.: A computational approach to fuzzy quantifiers in natural languages. Computers and Mathematics with Applications **9** (1983) 149–184
8. Ziarko, W.: Variable precision rough set model. Journal of Computer and System Sciences **46** (1993) 39-59
9. Ziarko, W.: Set approximation quality measures in the variable precision rough set model. Soft Computing Systems: Design, Management and Applications (A. Abraham, J. Ruiz-del-Solar, M. Koppen, eds.), IOS Press, (2002) 442–452