Proceedings of the of the 17th Conference on Computer Science and Intelligence Systems pp. 269–278
DOI: 10.15439/2022F231
ISSN 2300-5963 ACSIS, Vol. 30

Fuzzy Quantifier-Based Fuzzy Rough Sets

Adnan Theerens*, Chris Cornelis*
*Ghent University, Ghent, Belgium
Computational Web Intelligence
Dept. of Applied Mathematics, Computer Science and Statistics
Email: {adnan.theerens, chris.cornelis}@ugent.be

***Abstract*—In this paper we apply vague quantification to fuzzy rough sets to introduce *fuzzy quantifier-based fuzzy rough sets* (FQFRS), an intuitive generalization of fuzzy rough sets. We show how several existing models fit in this generalization as well as how it inspires novel models that may improve these existing models. In addition, we introduce several new binary quantification models. Finally, we introduce an adaptation of FQFRS that allows seamless integration of outlier detection algorithms to enhance the robustness of the applications based on FQFRS.**

I. Introduction

FUZZY quantification is an important part of fuzzy logic that models quantified sentences such as "Most Dutch people are tall" and "Nearly half of the S&P 500 stocks are down 10%". Quantifiers are an effective tool to describe the quantity of elements that satisfy a certain condition. This is especially true if the condition is of a vague nature, as for example in the quantified sentence "Most Dutch people are tall", since the quantity of elements satisfying a fuzzy condition (being tall) is hard to assess. The two most studied types of quantifiers are unary and binary quantifiers, unary quantifiers being of the form "Q_1 elements are A" (e.g. "Some people are tall") and binary quantifiers being of the form "Q_2 A's are B's" (e.g. "Most Dutch people are tall"). The first evaluation method for fuzzy quantified statements was introduced by Zadeh [1]. His idea was to define a cardinality measure for fuzzy sets to evaluate the quantity of elements satisfying a condition. The problem with this approach is that the cardinality measure is cumulative, implying that a situation involving two people with a degree of tallness of 0.5 is regarded the same as one with one tall person (tallness 1) and one short person (tallness 0). An improved evaluation method was proposed by Yager [2], which is based on the Ordered Weighted Averaging (OWA) operator. This method is semantically more reasonable for unary quantifiers but still lacks soundness for binary quantifiers. To resolve these issues, Glöckner [3] developed a general framework for fuzzy quantification. In this framework, fuzzy quantifiers are fully determined by how they act on classical (i.e. non-fuzzy) sets and by the choice of a quantifier fuzzification mechanism (QFM). A QFM thus reduces the evaluation of any quantified statement to the evaluation of quantified statements with crisp arguments.

Rough set theory, introduced by Pawlak [4], provides a lower and upper approximation of a concept with respect to the indiscernibility relation between objects. The lower and upper approximation contain all objects that are certainly, respectively possibly part of the concept. That is to say, an element is a member of the lower approximation of a concept if every element indiscernible from it belongs to the concept; and an element is a member of the upper approximation of the concept if there exists an element indiscernible from it that belongs to the concept. Rough set theory was first extended to fuzzy rough set theory by Dubois and Prade [5], where both the concept and the indiscernibility relation can be fuzzy. Fuzzy rough set theory has been used successfully for classification and other machine learning purposes, such as feature and instance selection [6], but due to the fact that the approximations in classical fuzzy rough sets are determined using the minimum and maximum operators, these approximations (and the applications based on them) are sensitive to noisy and outlying samples. To mitigate this problem, many noise-tolerant versions of fuzzy rough sets (FRS) have been proposed, such as Vaguely Quantified FRS (VQFRS) [7], β-Precision FRS [8], [9], Variable Precision FRS [10], Variable Precision (θ, σ)-FRS [11], Soft Fuzzy Rough Sets [12], Automatic Noisy Sample Detection FRS [13], Data-Distribution-Aware FRS [14], Probability Granular Distance based FRS [15], Ordered Weighted Averaging (OWA) based FRS (OWAFRS) [16] and Choquet-based FRS (CFRS) [17]. VQFRS and, as noted in [17], OWAFRS and CFRS are fuzzy rough set models based on vague quantification. In this paper, we introduce a generalization of fuzzy rough sets, called fuzzy quantifier-based fuzzy rough sets (FQFRS), that takes the idea behind VQFRS and CFRS one step further. It does this by using binary and unary quantification models to determine the lower and upper approximation of a concept, respectively. Furthermore, we explain how to adapt FQFRS to use normalized outlier scores [18] to boost the robustness of the lower and upper approximations in fuzzy rough sets.

This paper is structured as follows: in Section II, we recall the required prerequisites for (Choquet-based) fuzzy rough sets and vague quantification. Section III discusses different binary quantification models and introduces several new ones. In Section IV, fuzzy quantifier-based fuzzy rough sets (FQFRS) and confidence-based FQFRS are introduced and their relation with existing models is discussed as well as the possible benefits they may have. Sections V and VI conclude this paper and describe opportunities for future research.

II. Preliminaries

A. Fuzzy logic

In this subsection, we recall the necessary notions of fuzzy set and fuzzy logical connectives. We start with the definition of a fuzzy set and a fuzzy relation.

Definition II.1. *[19] A* fuzzy set *or* membership function *A on X is a function from X to the unit interval, i.e. $A : X \to [0,1]$. The value $A(x)$ of an element $x \in X$ is called the* degree of membership *of x in the fuzzy set A. The set of all fuzzy sets on X is denoted as $\widetilde{\mathcal{P}}(X)$.*

Definition II.2. *A* fuzzy relation *R on X is an element of $\widetilde{\mathcal{P}}(X \times X)$. A fuzzy relation R is called* reflexive *if $R(x,x) = 1$ for every $x \in X$. For an element $y \in X$ and a fuzzy relation $R \in \widetilde{\mathcal{P}}(X \times X)$, we define the R-foreset of y as the fuzzy set $Ry(x) := R(x,y)$.*

We will also make use of conjunctors, implicators and negators which extend their Boolean counterparts to the fuzzy setting.

Definition II.3.

- *A function $\mathcal{C} : [0,1]^2 \to [0,1]$ is called a* conjunctor *if it is increasing in both arguments and satisfies $\mathcal{C}(0,0) = \mathcal{C}(0,1) = 0$ and $\mathcal{C}(1,x) = x$ for all $x \in [0,1]$. A commutative and associative conjunctor $\mathcal{T}$ is called a* t-norm. *We will use the following notation $x \wedge_{\mathcal{C}} y := \mathcal{C}(x,y)$.*
- *A function $\mathcal{S} : [0,1]^2 \to [0,1]$ is called a* t-conorm *if it is non-decreasing in both arguments, commutative, associative, and satisfies $\mathcal{S}(0,x) = x$ for all $x \in [0,1]$. We will use the following notation $x \vee_{\mathcal{S}} y := \mathcal{S}(x,y)$.*
- *A function $\mathcal{I} : [0,1]^2 \to [0,1]$ is called an* implicator *if $\mathcal{I}(0,0) = \mathcal{I}(0,1) = \mathcal{I}(1,1) = 1$, $\mathcal{I}(1,0) = 0$ and for all x_1, x_2, y_1, y_2 in $[0,1]$ the following holds:*
  1) $x_1 \leq x_2 \Rightarrow \mathcal{I}(x_1,y_1) \geq \mathcal{I}(x_2,y_1)$ *(non-increasing in the first argument),*
  2) $y_1 \leq y_2 \Rightarrow \mathcal{I}(x_1,y_1) \leq \mathcal{I}(x_1,y_2)$ *(non-decreasing in the second argument),*

  We will use the following notation $x \to_{\mathcal{I}} y := \mathcal{I}(x,y)$.
- *A function $\mathcal{N} : [0,1] \to [0,1]$ is called a* negator *if it is non-increasing and satisfies $\mathcal{N}(0) = 1$ and $\mathcal{N}(1) = 0$. A negator is called a strong negator if it is an involution.*
- *Suppose $\mathcal{S}$ is a t-conorm and $\mathcal{N}$ is a negator. The mapping*
  $$\mathcal{I}(x,y) = \mathcal{N}(x) \vee_{\mathcal{S}} y, \ \ \forall x,y \in [0,1],$$
  is called the S-implicator induced by $\mathcal{S}$ and $\mathcal{N}$.

Example II.1. *The Kleene-Dienes implicator is defined as $\mathcal{I}_{KD}(x,y) := \max(1-x,y)$. It is the S-implicator induced by the standard negator $\neg(x) := 1-x$ and the standard t-conorm $x \vee y = \max(x,y)$.*

Since t-norms are required to be associative, they can be extended naturally to a function $[0,1]^n \to [0,1]$ for any natural number $n \geq 2$.

Definition II.4. *The notation $A \subseteq B$ for two fuzzy sets A and B, expresses that $A(x) \leq B(x)$ for all $x \in X$. The fuzzy set $A \cap B \in \widetilde{\mathcal{P}}(X)$ is defined by $(A \cap B)(x) = \min(A(x), B(x))$. We denote Zadeh's Sigma count as $|A| := \sum_{x \in X} A(x)$ for every fuzzy set $A \in \widetilde{\mathcal{P}}(X)$, it is a conservative extension of classical set cardinality to fuzzy sets.*

Definition II.5. *Given a negator $\mathcal{N}$, conjunctor $\mathcal{C}$, t-conorm $\mathcal{S}$, implicator $\mathcal{I}$, and two fuzzy sets $A, B \in \widetilde{\mathcal{P}}(X)$, we define the following:*

$$(\neg_{\mathcal{N}} A)(x) = \mathcal{N}(A(x)),$$
$$(A \cap_{\mathcal{C}} B)(x) := A(x) \wedge_{\mathcal{C}} B(x),$$
$$(A \cup_{\mathcal{S}} B)(x) := A(x) \vee_{\mathcal{S}} B(x),$$
$$(A \to_{\mathcal{I}} B)(x) := A(x) \to_{\mathcal{I}} B(x),$$

for all $x \in X$.

B. OWA-based fuzzy rough sets

A downside to the classical definition of lower and upper approximation in fuzzy rough set theory is their lack of robustness. The value of the membership of an element in the lower and upper approximation is fully determined by a single element because of the minimum and maximum operators in the definition. To solve this undesirable behaviour, many alternative definitions of fuzzy rough sets were introduced. One of these is OWA-based fuzzy rough sets [16], which has been shown to have an excellent trade-off between performance (robustness) and theoretical properties [20]. The Ordered Weighted Average [21] is an aggregation operator that is defined as follows:

Definition II.6 (OWA). *Let $X = \{x_1, x_2, \ldots, x_n\}$, $f : X \to \mathbb{R}$ and $\mathbf{w} = (w_1, w_2, \ldots, w_n)$ be a weighting vector, i.e. $\mathbf{w} \in [0,1]^n$ and $\sum_{i=1}^n w_i = 1$, then the* ordered weighted average *of f with respect to $\mathbf{w}$ is defined as*

$$OWA_{\mathbf{w}}(f) := \sum_{i=1}^{n} f(x_{\sigma(i)}) w_i,$$

where σ is a permutation of $\{1,2,\ldots,n\}$ such that

$$f(x_{\sigma(1)}) \geq f(x_{\sigma(2)}) \geq \cdots \geq f(x_{\sigma(n)}).$$

Example II.2. *The maximum, mean and minimum operators can all be seen as OWA-operators with weight vectors $(1,0,\ldots,0,0)$, $\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right)$ and $(0,0,\ldots,0,1)$ respectively.*

In OWA-based fuzzy rough sets, OWA operators replace the minimum and maximum in the lower and upper approximations in classical fuzzy rough sets. To not deviate too strongly from the original definitions, some requirements may be enforced on the weight vectors of the OWA-operators used [16]. In particular, the authors required that the OWA-operator for the lower approximation is a soft minimum and for the upper approximation a soft maximum.

Definition II.7. *The* orness *and* andness *of a weight vector* $\mathbf{w} = (w_i)_{i=1}^n$ *are defined as*

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

If $\text{orness}(\mathbf{w}) < 0.5$*, then* $OWA_{\mathbf{w}}$ *is called a* soft minimum. *If* $\text{orness}(\mathbf{w}) > 0.5$*,* $OWA_{\mathbf{w}}$ *is called a* soft maximum.

As can be seen from Equation (1), the orness indicates how much weight is given to the largest elements. The orness tells us how "close" the OWA-operator is to the maximum. Using this definition OWA-based fuzzy rough sets are then defined as:

Definition II.8. *[16] Given* $R \in \widetilde{\mathcal{P}}(X \times X)$*, weight vectors* $\mathbf{w}_l$ *and* $\mathbf{w}_u$ *with* $\text{orness}(\mathbf{w}_l) < 0.5$ *and* $\text{orness}(\mathbf{w}_u) > 0.5$ *and* $A \in \widetilde{\mathcal{P}}(X)$*, the* OWA lower *and* upper approximation *of* A *w.r.t.* R*,* $\mathbf{w}_l$ *and* $\mathbf{w}_u$ *are given by:*

$$(\underline{apr}_{R,\mathbf{w}_l}A)(x) = OWA_{\mathbf{w}_l}\left(\mathcal{I}(R(x,y), A(y))\right), \quad (2)$$
$$(\overline{apr}_{R,\mathbf{w}_u}A)(x) = OWA_{\mathbf{w}_u}\left(\mathcal{C}(R(x,y), A(y))\right), \quad (3)$$

where $\mathcal{I}$ *is an implicator,* $\mathcal{C}$ *a conjunctor and* $\mathcal{I}(R(x,y), A(y))$ *and* $\mathcal{C}(R(x,y), A(y))$ *are seen as functions in* y*.*

C. The Choquet integral

The Choquet integral induces a large class of aggregation functions, namely the class of all comonotone linear aggregation functions [22]. Since we will view the Choquet integral as an aggregation operator, we will restrict ourselves to measures (and Choquet integrals) on finite sets. For the general setting, we refer the reader to e.g. [23].

Definition II.9. *Let* X *be a finite set. A function* $\mu : \mathcal{P}(X) \to [0,1]$ *is called a* monotone measure *if:*

- $\mu(\emptyset) = 0$ *and* $\mu(X) = 1$*,*
- $(\forall A, B \in (\mathcal{P}(X))(A \subseteq B \implies \mu(A) \leq \mu(B))$*.*

A monotone measure is called:

- additive *if* $\mu(A \cup B) = \mu(A) + \mu(B)$ *when* A *and* B *are disjoint,*
- symmetric *if* $\mu(A) = \mu(B)$ *when* $|A| = |B|$*.*

Definition II.10. *[23] Let* μ *be a monotone measure on* X *and* $f : X \to \mathbb{R}$ *a real-valued function. The* Choquet integral *of* f *with respect to the measure* μ *is defined as:*

$$\int f\,d\mu = \sum_{i=1}^{n} \mu(A_i^*) \cdot \left[f(x_i^*) - f(x_{i-1}^*)\right],$$

where $(x_1^*, x_2^*, \ldots, x_n^*)$ *is a permutation of* $X = (x_1, x_2, \ldots, x_n)$ *such that*

$$f(x_1^*) \leq f(x_2^*) \leq \cdots \leq f(x_n^*),$$

$A_i^* := \{x_i^*, \ldots, x_n^*\}$ *and* $f(x_0^*) := 0$*.*

The class of aggregation operators induced by the Choquet integral contains the weighted mean and the OWA operator. In fact, the weighted mean and OWA operator are the Choquet integrals with respect to additive and symmetric measures, respectively.

Proposition II.11. *[22] The Choquet integral with respect to an additive measure* μ *is the weighted mean* $M_{\mathbf{w}}$ *with weight vector* $\mathbf{w} = (w_i)_{i=1}^n = (\mu(\{x_i\}))_{i=1}^n$*. Conversely, the weighted mean* $M_{\mathbf{v}}$*, with weight vector* $\mathbf{v} = (v_i)_{i=1}^n$ *is a Choquet integral with respect to the uniquely defined additive measure* μ *for which* $(\mu(\{x_i\}))_{i=1}^n = (v_i)_{i=1}^n$*.*

Proposition II.12. *[22] The Choquet integral with respect to a symmetric measure* μ *is the OWA operator with weight vector* $\mathbf{w} = (w_i)_{i=1}^n = (\mu(A_i) - \mu(A_{i-1}))_{i=1}^n$*, where* A_i *denotes any subset with cardinality* i*. Conversely, the OWA operator with weight vector* $\mathbf{v} = (v_i)_{i=1}^n$ *is a Choquet integral with respect to the symmetric measure* μ *defined as*

$$(\forall A \subseteq X)(\mu(A) := \sum_{i=1}^{|A|} v_i).$$

D. Glöckner's framework for fuzzy quantification

Glöckner's framework for fuzzy quantification deals with defining vague quantifiers in two steps. The first step is the specification of the vague quantifier on crisp sets, i.e. to specify the "underlying" *semi-fuzzy quantifier*. The second step is to extend this description to fuzzy arguments, i.e. applying a *quantifier fuzzification mechanism*.

Definition II.13. *[3] An* n*-ary semi-fuzzy quantifier on* $X \neq \emptyset$ *is a mapping* $Q : (\mathcal{P}(X))^n \to [0,1]$*. An* n*-ary fuzzy quantifier on* $X \neq \emptyset$ *is a mapping* $\widetilde{Q} : \left(\widetilde{\mathcal{P}}(X)\right)^n \to [0,1]$*.*

Definition II.14. *[3] A quantifier fuzzification mechanism (QFM)* $\mathcal{F}$ *assigns to each semi-fuzzy quantifier* $Q : (\mathcal{P}(X))^n \to [0,1]$ *a corresponding fuzzy quantifier* $\mathcal{F}(Q) : (\mathcal{P}(X))^n \to [0,1]$ *of the same arity* $n \in \mathbb{N}$ *and on the same universe* X*.*

Glöckner defined an axiomatic framework for plausible models of fuzzy quantification which he called the Determiner Fuzzification Scheme (DFS) axioms. Since introducing DFS would take up too much space we refer the reader to chapter three, four, and five of [3].

We now take a look at Zadeh's and Yager's traditional approaches, where they describe fuzzy quantifiers using fuzzy sets of the unit interval.

Definition II.15. *[1] A fuzzy set* $\Lambda \in \widetilde{\mathcal{P}}([0,1])$ *is called a* regular increasing monotone (RIM) quantifier *if* Λ *is a non-decreasing function such that* $\Lambda(0) = 0$ *and* $\Lambda(1) = 1$*.*

Example II.3. *The following RIM quantifiers represent the quantifiers "more than* $100 * k\%$*" and "at least* $100 * k\%$*":*

$$\Lambda_{>k}(p) = \begin{cases} 1 & \textit{if } p > k \\ 0 & \textit{elsewhere} \end{cases} \qquad \Lambda_{\geq k}(p) = \begin{cases} 1 & \textit{if } p \geq k \\ 0 & \textit{elsewhere} \end{cases}.$$

These RIM quantifiers also include (a representation of) the universal and existential quantifier, $\Lambda_\forall := \Lambda_{>0}$ *and* $\Lambda_\exists :=$

$\Lambda_{\geq 1}$. *Linguistic quantifiers such as "most" and "some" can be modelled using Zadeh's S-function* ($0 \leq \alpha < \beta \leq 1$) *:*

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

for example, we could use $\Lambda_{(0.3,0.9)}$ *and* $\Lambda_{(0.1,0.4)}$ *to model "most" and "some", respectively.*

In Zadeh's model, unary sentences of the form "Λ X's are A's" and binary sentences of the form "Λ A's are B's", where Λ is a RIM quantifier and $A, B \in \widetilde{\mathcal{P}}(X)$, are evaluated as

$$\widetilde{Z}_\Lambda(A) = \Lambda\left(\frac{|A|}{|X|}\right), \tag{4}$$

$$\widetilde{Z}^2_\Lambda(A, B) = \Lambda\left(\frac{|A \cap B|}{|A|}\right), \tag{5}$$

respectively, while in Yager's OWA model, the unary sentence "Λ X's are A's" is evaluated as

$$\widetilde{Y}_\Lambda(A) := OWA_{\mathbf{w}^\Lambda}(A), \tag{6}$$

where

$$w_i^\Lambda := \Lambda\left(\frac{i}{n}\right) - \Lambda\left(\frac{i-1}{n}\right). \tag{7}$$

For the binary sentence "Λ A's are B's", there is no definite evaluation, although there are two that are most common in literature (cf. [24], [25]). The first one evaluates the sentence as

$$\widetilde{Y}^{\mathcal{I}}_\Lambda(A, B) := \widetilde{Y}_\Lambda(\mathcal{I}(A, B)) = OWA_{\mathbf{w}^\Lambda}(\mathcal{I}(A, B)), \tag{8}$$

where $\mathcal{I}$ is an implicator, while the second one evaluates it as:

$$\widetilde{Y}^2_\Lambda(A, B) := OWA_{\mathbf{v}}(\mathcal{I}(A, B)), \tag{9}$$

where

$$v_i := \Lambda\left(\frac{\sum_{j=1}^{i} A(x_j^*)}{|A|}\right) - \Lambda\left(\frac{\sum_{j=1}^{i-1} A(x_j^*)}{|A|}\right), \tag{10}$$

with $A(x_i^*)$ being the ith smallest $A(x)$ for $x \in X$ and $\sum_{j=1}^{0} A(x_j^*) = \sum_{x\in\emptyset} x = 0$. Note that both $\widetilde{Z}_\Lambda$ and $\widetilde{Y}_\Lambda$ extend the semi-fuzzy quantifier

$$Q_\Lambda(A) := \Lambda\left(\frac{|A|}{|X|}\right). \tag{11}$$

E. Choquet-based fuzzy rough sets

Choquet-based fuzzy rough sets (CFRS) [17] have been introduced by noting that by Proposition II.12, we can rewrite OWAFRS as follows:

$$(\underline{\text{apr}}_{R,\mu_l} A)(y) = \int \mathcal{I}\left(R(x, y), A(x)\right) \, \mathrm{d}\mu_l(x),$$

$$(\overline{\text{apr}}_{R,\mu_u} A)(y) = \int \mathcal{C}(R(x, y), A(x)) \, \mathrm{d}\mu_u(x),$$

where μ_l and μ_u are two symmetric measures. Allowing non-symmetric measures gives us the definition of CFRS:

Definition II.16. *[17] Given* $R \in \widetilde{\mathcal{P}}(X \times X)$*, monotone measures* μ_l *and* μ_u *on* X *and* $A \in \widetilde{\mathcal{P}}(X)$*, then the* Choquet lower *and* upper approximation *of* A *w.r.t.* R*,* μ_l *and* μ_u *are given by:*

$$(\underline{apr}_{R,\mu_l} A)(y) = \int \mathcal{I}(R(x, y), A(x)) \, d\mu_l(x) \tag{12}$$

$$(\overline{apr}_{R,\mu_u} A)(y) = \int \mathcal{C}(R(x, y), A(x)) \, d\mu_u(x), \tag{13}$$

where $\mathcal{I}$ *is an implicator and* $\mathcal{C}$ *is a conjunctor.*

Example II.4. *Suppose we have a crisp set* O *containing all the instances that are outliers, unreliable or inaccurate, then a useful pair of quantifiers could be "for all except (maybe) elements of* O*" and "there exists an element in* $X \setminus O$*". These quantifiers can be modelled by the partial minimum and maximum, which in turn are Choquet-integral operators with respect to non-symmetric measures (cf. [17]).*

Using these non-symmetric measures in Equation (12) and (13), we get that the degree of membership of an element y to the lower approximation is equal to the truth value of the proposition "All trustworthy elements that are indiscernible to y are in A". An analogous interpretation holds for the upper approximation.

As we will see in Subsection II-F, it is possible to extend this approach of the previous example to fuzzy sets O and quantifiers representing "most of the trustworthy objects". The following examples show how such fuzzy sets O can be constructed in practice.

Example II.5. *Suppose we have a decision system* $(X, \mathcal{A} \cup \{d\})$ *where* d *is a categorical attribute. Then we can define* $O(x)$ *as the normalized outlier score [18] of* x *(obtained from a certain outlier detection algorithm) when compared to other elements of* $[x]_d$ *(based on the conditional attributes). An outlier score measures the degree to which a data point differs from other observations, and normalization transforms this score in such a way that it can be interpreted as a degree of outlierness.*

Example II.6. *Suppose* X *consists of patients from several different hospitals,* A *is the subset of patients that have a disease and* R *is a similarity relation between patients based on a set of symptoms. Then a confidence score* c_i *can be attached to each hospital* i *based on the accuracy of the tests performed to trace the disease (and the symptoms). The membership degree of a patient* x *of hospital* i *to* O *can then be defined as* $O(x) = 1 - c_i$*.*

F. Examples of non-symmetric measures

As described in the previous subsection we can accommodate non-symmetry by introducing a fuzzy set O in X that represents the lack of confidence. The value $O(x)$ could, for example, be seen as an outlier score in $[0, 1]$ (Example II.5) or it could represent the unreliability or inaccuracy of the observation (Example II.6). We now recall several non-symmetric measures that were introduced in [17] using the fuzzy set O.

1) Fuzzy removal: the first option that was proposed to define a non-symmetric measure using O is as follows:

$$\mu_{\forall x \in X \setminus O}(S) = \begin{cases} 1 & \text{if } S = X \\ 0 & \text{if } S = \emptyset \\ \underbrace{\mathcal{T}(O(x))}_{x \in X \setminus S} & \text{elsewhere} \end{cases}, \tag{14}$$

where $\mathcal{T}$ is a t-norm (e.g. minimum) and $S \in \mathcal{P}(X)$. This measure is called the *fuzzy removal* measure, since in the case O is crisp, the Choquet integral with respect to this measure is equal to the partial minimum. The vague quantifier interpretation of the fuzzy removal measure could thus be "for all except (maybe) elements of O".

2) Weighted ordered weighted average: the second idea for a non-symmetric measure was:

$$\mu_\Lambda^{\neg O}(S) := \Lambda\left(\frac{|\neg O \cap S|}{|\neg O|}\right) = \Lambda\left(\sum_{x_i \in S} p_i\right), \tag{15}$$

where Λ is a RIM quantifier and $\mathbf{p}$ a weight vector describing the confidence, reliability, accuracy or non-outlierness of each observation:

$$p_i = \frac{1 - O(x_i)}{n - \sum_{j=1}^n O(x_j)}. \tag{16}$$

The measure in Equation (15) corresponds to the Weighted Ordered Weighted Averaging (WOWA) operator [26], [27], which is a generalization of the OWA and the weighted mean. The RIM quantifier Λ determines the OWA part of the WOWA and the weight vector $\mathbf{p}$ the weighted mean part. The WOWA operator is also equivalent with Yager's importance weighted quantifier guided aggregation [2]. These measures could be interpreted as quantifiers of the form "Λ of the trustworthy/reliable objects".

III. Binary quantification models

We now take a deeper look at binary quantifiers (i.e. 2-ary fuzzy quantifiers), and in particular, proportional quantifiers such as "Most A's are B's". To focus our attention we will make use of Zadeh's approach of using RIM-quantifiers to model these quantifiers in their bare form.

A. QFM-based binary quantification models

To define binary fuzzy quantifiers using QFM's, we first need semi-fuzzy quantifiers. The following definition proposes the two most viable options for semi-fuzzy quantifiers that model "Λ A's are B's", with Λ a RIM-quantifier.

Definition III.1. *Given a RIM-quantifier Λ, we define the following semi-fuzzy quantifiers:*

$$Q_\Lambda^2(A, B) := \Lambda\left(\frac{|A \cap B|}{|A|}\right),$$
$$Q_\Lambda^{\rightarrow}(A, B) = \Lambda\left(\frac{|A \rightarrow B|}{|X|}\right) := \Lambda\left(\frac{|\neg A| + |A \cap B|}{|X|}\right),$$

for crisp sets $A, B \in \mathcal{P}(X)$.

The first one is the most intuitive definition, but the second one is (as we will see) used a lot in practice, perhaps due to its simplicity as we shall see later in this section (Corollary III.6). The semantical difference between Q_Λ^2 and $Q_\Lambda^{\rightarrow}$ is that of "Most A's are B's" and of "For most X's, if they are in A, they are in B". This is a very subtle difference and in day to day life both mean the same. In the first one only elements of A matter, while for the second one all elements matter. The following example demonstrates how important the difference is.

Example III.1. *Let us look at the difference between "Most Belgian people are not Belgian" and "For most people in the world, if they are Belgian, they are not Belgian". Most people would say both sentences are plainly wrong. Let X denote the set of all people and $B \in \mathcal{P}(X)$ the subset of Belgian people, if we evaluate the first sentence using Q_Λ and the second one using $Q_\Lambda^{\rightarrow}$, we get the following:*

$$Q_\Lambda^2(B, \neg B) = \Lambda\left(\frac{|\emptyset|}{|B|}\right) = 0,$$
$$Q_\Lambda^{\rightarrow}(B, \neg B) = \Lambda\left(\frac{|\neg B|}{|X|}\right) \approx 1,$$

because the percentage of Belgians in the world is minuscule. So the second one is still correct, since for most people it holds that if they are Belgian, then they are not Belgian, since they are simply not from Belgium.

If $\mathcal{F}$ is a QFM, we can use $\mathcal{F}(Q_\Lambda^2)$ and $\mathcal{F}(Q_\Lambda^{\rightarrow})$ to evaluate sentences of the form "Λ A's are B's" for $A, B \in \widetilde{\mathcal{P}}(X)$. We now take a look at the differences between the two. The first difference is the monotonicity. In the second argument both Q_Λ and $Q_\Lambda^{\rightarrow}$ are non-decreasing, hence so are $\mathcal{F}(Q_\Lambda)$ and $\mathcal{F}(Q_\Lambda^{\rightarrow})$ for a DFS $\mathcal{F}$ (argument monotonicity [3]). The difference between the two is in the first argument; let us add an element a to A and suppose Λ is a strictly increasing RIM-quantifier, if $a \in B$, then $Q_\Lambda^2(A, B)$ will strictly increase, while $Q_\Lambda^{\rightarrow}(A, B)$ stays unchanged, if $a \notin B$, then $Q_\Lambda^2(A, B)$ and $Q_\Lambda^{\rightarrow}(A, B)$ will strictly decrease. So in summary, $Q_\Lambda^{\rightarrow}$ is non-increasing in the first argument (hence $\mathcal{F}(Q_\Lambda^{\rightarrow})$ is as well), while Q_Λ^2 is not monotone in the first argument.

Proposition III.2. *We have the following inequality:*

$$Q_\Lambda^{\rightarrow}(A, B) \geq Q_\Lambda^2(A, B),$$

for every $A, B \in \mathcal{P}(X)$.

Proof.

$$\Lambda\left(\frac{|\neg A| + |A \cap B|}{|X|}\right) \geq \Lambda\left(\frac{|A \cap B|}{|A|}\right)$$
$$\iff \frac{|\neg A| + |A \cap B|}{|\neg A| + |A|} \geq \frac{|A \cap B|}{|A|}$$
$$\iff (|\neg A| + |A \cap B|) * |A| \geq |A \cap B| * (|\neg A| + |A|)$$
$$\iff |\neg A| * |A| \geq |A \cap B| * |\neg A|$$
$$\iff |A| \geq |A \cap B|$$

□

Corollary III.3. *We have the following inequality for every DFS $\mathcal{F}$:*

$$\mathcal{F}(Q_\Lambda^{\rightarrow})(A,B) \geq \mathcal{F}(Q_\Lambda^2)(A,B),$$

for every $A,B \in \widetilde{\mathcal{P}}(X)$.

Proof. Every DFS satisfies quantifier monotonicity [3]. □

We will now show that evaluating the binary quantifier $\mathcal{F}(Q_\Lambda^{\rightarrow})$ for fuzzy sets A,B and a DFS $\mathcal{F}$ simply amounts to evaluating the fuzzy set $A \tilde{\rightarrow} B$ using the unary quantifier $\mathcal{F}(Q_\Lambda)$, where $\tilde{\rightarrow}$ is the implicator induced by the DFS $\mathcal{F}$ (cf. [3]).

Definition III.4. *Let $\widetilde{Q} : (\widetilde{\mathcal{P}}(X))^n \to [0,1]$ be a fuzzy quantifier, then the fuzzy quantifier $\widetilde{Q}\tilde{\rightarrow} : (\widetilde{\mathcal{P}}(X))^{n+1} \to [0,1]$ is defined as:*

$$\widetilde{Q}\tilde{\rightarrow}(A_1,\ldots,A_{n+1}) := \widetilde{Q}(A_1,\ldots,A_{n-1},(A_n \tilde{\rightarrow} A_{n+1})).$$

For a semi-fuzzy quantifier Q the semi-fuzzy quantifier $Q\rightarrow$ is defined analogously.

Proposition III.5. *For every semi-fuzzy quantifier Q and DFS $\mathcal{F}$ we have:*

$$\mathcal{F}(Q\rightarrow) = \mathcal{F}(Q)\tilde{\rightarrow}.$$

Proof. This follows from

$$Q\rightarrow = Q\neg\cap\neg$$

and the fact that a DFS is compatible with internal meets and internal negations [3]. □

Corollary III.6. *Let $\mathcal{F}$ be a DFS and Q_Λ the unary quantifier from Equation* (11)*, then:*

$$\mathcal{F}\left(Q_\Lambda^{\rightarrow}\right)(A,B) = \mathcal{F}(Q_\Lambda)(A \tilde{\rightarrow} B),$$

for every $A,B \in \widetilde{\mathcal{P}}(X)$.

Applying this to one of the most used QFM's, Glöckner's $\mathcal{F}_{owa}$ [3], we can write one of Yager's binary quantification models as a QFM-based model:

Corollary III.7.

$$\mathcal{F}_{owa}(Q_\Lambda^{\rightarrow})(A,B) = \int \mathcal{I}_{KD}(A,B)\, d\mu_\Lambda = \widetilde{Y}_\Lambda^{\mathcal{I}_{KD}}(A,B)$$

Proof. Follows from the fact that $\mathcal{F}_{owa}$ is a standard DFS (thus the induced implicator is $\mathcal{I}_{KD}$) [3] and

$$\mathcal{F}_{owa}(Q)(A) = \int A\, \mathrm{d}Q,$$

for every $A \in \widetilde{\mathcal{P}}(X)$ and every semi-fuzzy quantifier Q that is also a monotone measure [3]. □

Remark III.2. *We can apply the exact same reasoning as in the previous corollary to the standard DFS $\mathcal{M}_{CX}$ [3], to obtain*

$$\mathcal{M}_{CX}(Q_\Lambda^{\rightarrow})(A,B) = (S)\int \mathcal{I}_{KD}(A,B)\, d\mu_\Lambda,$$

where $(S)\int$ denotes the Sugeno integral [23].

B. Integral-based binary quantification models

We start off with rewriting Zadeh's and Yager's models using the Choquet integral to get a unifying view of these models.

Proposition III.8. *Let Λ be a RIM-quantifier, $A,B \in \widetilde{\mathcal{P}}(X)$ and $\mathcal{I}$ an implicator. We can rewrite Zadeh's and Yager's evaluation models as follows:*

$$\widetilde{Z}_\Lambda(A) = \Lambda\left(\int A\, d\mu_{\boldsymbol{id}}\right)$$
$$\widetilde{Z}_\Lambda^2(A,B) = \Lambda\left(\frac{\int A\cap B\, d\mu_{\boldsymbol{id}}}{\int A\, d\mu_{\boldsymbol{id}}}\right)$$
$$\widetilde{Y}_\Lambda(A) = \int A\, d\mu_\Lambda$$
$$\widetilde{Y}_\Lambda^{\mathcal{I}}(A,B) = \int \mathcal{I}(A,B)\, d\mu_\Lambda$$
$$\widetilde{Y}_\Lambda^2(A,B) = \int \mathcal{I}(A,B)\, d\mu'_\Lambda,$$

where

$$\mu_\Lambda(S) := \Lambda\left(\frac{|S|}{|X|}\right)$$
$$\mu'_\Lambda(S) := \Lambda\left(\frac{\sum_{j=1}^{|S|} A(x_j^*)}{|A|}\right),$$

with $A(x_i^*)$ being the ith smallest $A(x)$ for $x \in X$ and $S \in \mathcal{P}(X)$.

Proof. Follows from Propositions II.11 and II.12. □

So both models can be written using integrals, but whereas Zadeh's model first integrates and then applies the RIM quantifier, Yager's model already incorporates the RIM quantifier in the measure used for integration.

Looking at Yager's model $\widetilde{Y}_\Lambda^{\mathcal{I}}$ we can see that an element not in A contributes as much to the truth value as an element that is in A and in B, therefore the model has the same issues as mentioned in Example III.1. Quantifiers based on the semi-fuzzy quantifier Q_Λ^2 and a DFS do not suffer from this issue, but are computationally more complex. Therefore we now introduce a new binary quantification model that does an extra weighting on elements of A to compensate for the issues of $\widetilde{Y}_\Lambda^{\mathcal{I}}$:

Definition III.9. *Let Λ be a RIM-quantifier, $\mathcal{I}$ an implicator and $A,B \in \widetilde{\mathcal{P}}(X)$. We define the fuzzy quantifier $\widetilde{W}_\Lambda^{\mathcal{I}} : \widetilde{\mathcal{P}}(X) \to [0,1]$ as:*

$$\widetilde{W}_\Lambda^{\mathcal{I}}(A,B) := \int \mathcal{I}(A,B)\, d\mu_\Lambda^A, \;\; \mu_\Lambda^A(S) := \Lambda\left(\frac{|S\cap A|}{|A|}\right),$$

where $S \in \mathcal{P}(X)$.

Remark III.3. *By replacing the Choquet integral in the previous definition with the Sugeno integral, or even general pan-integrals (for more information about these integrals see [23]), we obtain other novel quantifiers that are worth studying. Because Glöckner's model $\mathcal{M}_{CX}$ corresponds to the Sugeno* *integral (Remark III.2), this can give us a good compromise (between computational simplicity and semantical soundness) for the $\mathcal{M}_{CX}(Q^2_\Lambda)$ model.*

This quantifier indeed resembles $\widetilde{Y}^{\mathcal{I}}_\Lambda(A,B)$ but with an extra weighting on A. The quantifier $\widetilde{Y}^2_\Lambda$ also does this but in a less intuitive and elegant way, by doing an extra ordering of the elements. Comparing the two, we see that the new definition uses a weighted ordered weighted averaging (WOWA) with Λ for the OWA part and $p_i = A(x_i)/|A|$ for the weighted mean part, while the quantifier $\widetilde{Y}^2_\Lambda$ uses an OWA operator with weights given by Equation (10).
The following proposition describes how all of these fuzzy quantifiers act on crisp sets, i.e., what their underlying semi-fuzzy quantifiers are.

Proposition III.10. *The following equalities hold:*

$$\begin{aligned}\widetilde{W}^{\mathcal{I}}_\Lambda(A,B) &= Q_\Lambda(A,B),\\ \widetilde{Y}^2_\Lambda(A,B) &= Q_\Lambda(A,B),\\ \widetilde{Z}^2_\Lambda(A,B) &= Q_\Lambda(A,B),\\ \widetilde{Y}^{\mathcal{I}}_\Lambda(A,B) &= Q^{\rightarrow}_\Lambda(A,B),\end{aligned}$$

for all crisp sets $A, B \in \mathcal{P}(X)$.

Proof. We will only prove the first and second equality, the rest are analogous or trivial. Let $A, B \in \mathcal{P}(X)$ be two crisp sets, then:

$$\begin{aligned}\mathcal{U}\left(\widetilde{W}^{\mathcal{I}}_\Lambda\right)(A,B) &= \int \mathcal{I}(A,B)\,\mathrm{d}\mu^A_\Lambda = \mu^A_\Lambda(\mathcal{I}(A,B))\\ &= \mu^A_\Lambda(\neg A \cup B)\\ &= \Lambda\left(\frac{|A\cap B|}{|A|}\right) = Q_\Lambda(A,B),\end{aligned}$$

which proves the first equality. For the second equality:

$$\begin{aligned}\mathcal{U}\left(\widetilde{Y}^2_\Lambda\right)(A,B) &= \int \mathcal{I}(A,B)\,\mathrm{d}\mu'_\Lambda = \mu'_\Lambda(\mathcal{I}(A,B))\\ &= \mu'_\Lambda(\neg A \cup B)\\ &= \mu'_\Lambda(\neg A \cup (A\cap B))\end{aligned}$$

But for crisp sets A the measure μ'_Λ reduces to:

$$\mu'_\Lambda(S) = \begin{cases} 0 & \text{if } |S| \leq |\neg A| \\ \frac{|S|-|\neg A|}{|A|} & \text{if } |S| > |\neg A| \end{cases},$$

from which we get the desired

$$\begin{aligned}\mathcal{U}\left(\widetilde{Y}^2_\Lambda\right)(A,B) &= \mu'_\Lambda(\neg A \cup (A\cap B))\\ &= \Lambda\left(\frac{|\neg A \cup (A\cap B)| - |\neg A|}{|A|}\right)\\ &= \Lambda\left(\frac{|A\cap B|}{|A|}\right).\end{aligned}$$

□

The previous proposition thus shows that both $\widetilde{Y}^2_\Lambda$ and $\widetilde{W}^{\mathcal{I}}_\Lambda$ apply the correct weighting on A such that for crisp arguments the quantifiers act intuitively.

IV. Fuzzy quantifier-based fuzzy rough sets

Let us take another look at OWAFRS and rewrite its approximations as follows:

$$\begin{aligned}(\underline{\mathrm{apr}}_{R,\mu_l}A)(y) &= \int \mathcal{I}\left(R(x,y),A(x)\right)\,\mathrm{d}\mu_l(x),\\ &= \widetilde{Y}^{\mathcal{I}}_\Lambda(Ry,A) \quad (17)\\ (\overline{\mathrm{apr}}_{R,\mu_u}A)(y) &= \int \mathcal{C}(R(x,y),A(x))\,\mathrm{d}\mu_u(x)\\ &= \widetilde{Y}_\Upsilon(Ry \cap_{\mathcal{C}} A), \quad (18)\end{aligned}$$

where μ_l and μ_u are symmetric measures, and Λ and Υ are their corresponding RIM-quantifiers (cf. Section 3.4 in [17]). Thus, the lower and upper approximations of OWAFRS are evaluated by evaluating vaguely quantified propositions using Yager's quantification model ($\widetilde{Y}^{\rightarrow}_\Lambda$ and $\widetilde{Y}_\Lambda$). We now introduce fuzzy quantifier-based fuzzy rough sets (FQFRS) by allowing general (binary for lower approximation and unary for upper approximation) quantification models.

Definition IV.1 (($\widetilde{Q}_l, \widetilde{Q}_u$)-fuzzy rough set)**.** *Given a reflexive fuzzy relation $R \in \mathcal{F}(X\times X)$, fuzzy quantifiers $\widetilde{Q}_l : (\widetilde{\mathcal{P}}(X))^2 \to [0,1]$ and $\widetilde{Q}_u : \widetilde{\mathcal{P}}(X) \to [0,1]$, and $A \in \mathcal{F}(X)$, then the* lower *and* upper approximation *of A w.r.t. R are given by:*

$$\begin{aligned}(\underline{apr}_{R,\widetilde{Q}_l}A)(y) &= \widetilde{Q}_l(Ry,A),\\ (\overline{apr}_{R,\widetilde{Q}_u}A)(y) &= \widetilde{Q}_u(Ry\cap_{\mathcal{C}} A),\end{aligned}$$

where $\mathcal{C}$ is a conjunctor.

Suppose $\widetilde{Q}_l$ and $\widetilde{Q}_u$ represent the (linguistic) quantifiers "most" and "some", respectively. Then the degree of membership of an element y to the lower approximation of A is equal to the truth value of the statement "Most elements similar to y are in A". The degree of membership of y to the upper approximation is equal to the truth value of the statement "Some elements are similar to y and are in A".

A. Examples of FQFRS models

1) $(\widetilde{Z}^2_\Lambda, \widetilde{Z}_\Upsilon)$-FRS: let us take a look at the model derived from the most simple quantification model, the one from Zadeh. Let Λ and Υ be two RIM-quantifiers, then the lower and upper approximation for $(\widetilde{Z}^2_\Lambda, \widetilde{Z}_\Upsilon)$-fuzzy rough sets are defined as:

$$\begin{aligned}(\underline{apr}_{R,\Lambda}A)(y) &:= \widetilde{Z}^2_\Lambda(Ry,A) = \Lambda\left(\frac{|Ry\cap A|}{|Ry|}\right),\\ (\overline{apr}_{R,\Upsilon}A)(y) &:= \widetilde{Z}_\Upsilon(Ry\cap A) = \Upsilon\left(\frac{|Ry\cap A|}{|X|}\right).\end{aligned}$$

This closely resembles the Vaguely Quantified Fuzzy Rough Sets (VQFRS) model [7], which uses the following lower and upper approximations:

$$\begin{aligned}(\underline{apr}^{\mathrm{VQFRS}}_{R,\Lambda}A)(y) &:= \Lambda\left(\frac{|Ry\cap A|}{|Ry|}\right) = \widetilde{Z}^2_\Lambda(Ry,A),\\ (\overline{apr}^{\mathrm{VQFRS}}_{R,\Upsilon}A)(y) &:= \Upsilon\left(\frac{|Ry\cap A|}{|Ry|}\right) = \widetilde{Z}^2_\Upsilon(Ry,A).\end{aligned}$$

For both models the lower approximations are identical, but whereas for VQFRS the lower and upper approximation only differ in their used RIM-quantifier, $(\widetilde{Z}^2_\Lambda, \widetilde{Z}_\Upsilon)$-FRS evaluates the upper approximations using Zadeh's unary quantifier $\widetilde{Z}_\Upsilon$. Comparing the upper approximations of these two models, we can see that VQFRS will always be larger ($|X| \geq |Ry|$ and Υ is a RIM-quantifier). In some cases the upper approximation of VQFRS might be too large. For example, as soon as an element does not have a similar element ($Ry = \emptyset$) it is in the upper approximation of any concept A, even if there are many elements that are not that similar to y (e.g. $(Ry)(x) = 0.01$) but are in A (and not many elements similar to y) it will be in the upper approximation (cf. next example). Thus if one wants to discard the outlying elements from the upper approximation, this is problematic. This happens to a lesser extent with $(\widetilde{Z}^2_\Lambda, \widetilde{Z}_\Upsilon)$-FRS, but it is still susceptible to it due to the accumulative nature of the Σ-count, as the following example shows.

Example IV.1. *Suppose there are 10 elements in A with a similarity of 0.1 to $y \notin A$ and the rest of the elements are not similar to y at all ($|Ry \cap A| = 1$ and $|Ry| = 2$), then the upper approximation would always be 1 in the VQFRS approach:*

$$(\overline{apr}^{\textit{VQFRS}}_{R,\Upsilon} A)(y) = \widetilde{Z}^2_\Upsilon(Ry, A) = \Upsilon\,(0.5) := 1,$$

since Υ should represent "some". In $(\widetilde{Z}^2_\Lambda, \widetilde{Z}_\Upsilon)$-FRS we get a less extreme result:

$$(\overline{apr}_{R,\Upsilon} A)(y) = \widetilde{Z}_\Upsilon(Ry \cap A) = \Upsilon\left(\frac{1}{|X|}\right).$$

So in conclusion, with VQFRS the outliers will always belong more to the upper approximation than with $(\widetilde{Z}^2_\Lambda, \widetilde{Z}_\Upsilon)$-FRS. Lastly we note that using the existential quantifier (i.e. $\Upsilon = \Lambda_\exists$) the two upper approximations are equivalent ($|Ry| \geq 1$ since R is reflexive).

2) $(\widetilde{Y}^{\rightarrow}_\Lambda, \widetilde{Y}_\Upsilon)$-FRS: since Yager's model is generally accepted as a better model compared to Zadeh's, we now take a look at $(\widetilde{Y}^{\rightarrow}_\Lambda, \widetilde{Y}_\Upsilon)$-fuzzy rough sets. As shown in Equations (17) and (18), $(\widetilde{Y}^{\rightarrow}_\Lambda, \widetilde{Y}_\Upsilon)$-FRS corresponds to OWAFRS which is preferred over VQFRS [20]. So this justifies the improvement from a fuzzy quantifier perspective why OWAFRS are better than VQFRS. To justify even more why this is a good model we know from Corollary III.7 that when using the Kleene-Dienes implicator, OWAFRS are equal to $(\mathcal{F}_{owa}(Q^{\rightarrow}_\Lambda), \mathcal{F}_{owa}(Q_\Upsilon))$-FRS. So OWAFRS use a DFS mechanism, which is known to be semantically sound.

3) Other FQFRS: a problem with OWAFRS from a fuzzy quantifier perspective is that it makes use of the semi-fuzzy quantifier $Q^{\rightarrow}_\Lambda$, which is not that intuitive. Therefore $(\mathcal{F}_{owa}(Q^2_\Lambda), \mathcal{F}_{owa}(Q_\Upsilon))$-FRS are preferable, since they make more sense on crisp sets (semi-fuzzy quantifiers), and are thus better suited for explainability when used for classification e.g. "Most people similar to y are not able to pay off their mortgage". Instead of $\mathcal{F}_{owa}$ other DFS could also be used like the M_{CX} model [3], which is known to be a standard DFS with many good properties, or the $\mathcal{F}^{\mathcal{A}}$ a non-standard DFS [28]. For a compromise between semantical soundness and computational efficiency, one could also use the quantifier $\widetilde{Y}^2_\Lambda$, $\widetilde{W}^2_\Lambda$ or $\widetilde{W}^2_\Lambda$ using another integral (Remark III.3).

Example IV.2. *Evaluating "Λ Ry are A" using $Q^{\rightarrow}_\Lambda$ gives us:*

$$\Lambda\left(\frac{|Ry \rightarrow A|}{|X|}\right) = \Lambda\left(\frac{|\neg Ry| + |Ry \cap A|}{|X|}\right).$$

Thus the smaller the cardinality of Ry, the more true the statement is. This is not really what one would expect, because this causes the lower approximation to be large. Let $y \notin A$ be an instance that is an outlier (Ry only contains y), then the membership of y to the lower approximation of A is always very high (regardless of A). Using Q^2_Λ instead would not result in this problem.

So this example suggests that using quantifiers with Q^2_Λ as underlying semi-fuzzy quantifier might be preferable.

4) Choquet-based fuzzy rough sets: Choquet-based fuzzy rough sets correspond to $(C^{\mathcal{I}}_{\mu_l}, C_{\mu_u})$-FRS with the fuzzy quantifiers $C^{\mathcal{I}}_{\mu_l}, C_{\mu_u}$ being defined as

$$C^{\mathcal{I}}_{\mu_l}(A, B) := \int \mathcal{I}(A, B)\, \mathrm{d}\mu_l,$$
$$C_{\mu_u}(A) := \int A\, \mathrm{d}\mu_u,$$

for every implicator $\mathcal{I}$ and monotone measures μ_l and μ_u. When the measures μ_u and μ_l are symmetric these quantifiers are quantitative (i.e. each element is regarded as the same) and reduce to Yager's quantifiers like mentioned before. Note that all quantifiers mentioned in this section up until this point are quantitative. The interesting part of CFRS, when compared to OWAFRS, was that it allowed non-symmetric measures. We will now take a look at this from a FQFRS perspective for the non-symmetric measures discussed in Section II-F.

- *Fuzzy removal measure Equation* (14)*:*
  If $\mu_l = \mu_\forall$, then the quantifier $\mathcal{C}^{\mathcal{I}}_{\mu_l}$ represents "for all except (maybe) elements of O", thus it can be seen as the quantifier $\widetilde{Y}^{\mathcal{I}}_{\Lambda_\forall}$ but not regarding elements of O.
- *WOWA measure Equation* (15):
  If $\mu_l = \mu_\forall$, then the quantifier $\mathcal{C}^{\mathcal{I}}_{\mu_l}$ represents "Λ elements except (maybe) elements of O", thus it can be seen as the quantifier $\widetilde{Y}^{\mathcal{I}}_\Lambda$ but not regarding elements of O. Do note that for $\Lambda = \Lambda_\forall$ this quantifier represents the same quantifier as the fuzzy removal measure, but the evaluation is different.

B. Confidence-based FQFRS

In the previous section we have seen that we are able to seamlessly incorporate outlier information in FQFRS by making use of non-quantitative quantifiers (an implicit way). But it is also possible to do this using quantitative quantifiers (a more explicit way):

Definition IV.2. *Given a reflexive fuzzy relation $R \in \mathcal{F}(X \times X)$, $O \in \mathcal{F}(X)$, fuzzy quantifiers $\widetilde{Q}_l : \left(\widetilde{\mathcal{P}}(X)\right)^3 \rightarrow [0,1]$ and*

$\widetilde{Q}_u : \left(\widetilde{\mathcal{P}}(X)\right)^2 \to [0,1]$ *and* $A \in \mathcal{F}(X)$*, then the* lower *and* upper approximation *of* A *w.r.t.* R*,* Q_l *and* Q_u *are given by:*

$$(\underline{apr}_{R,\widetilde{Q}_l} A)(y) = \widetilde{Q}_l\left(O, Ry, A\right),$$
$$(\overline{apr}_{R,\widetilde{Q}_u} A)(y) = \widetilde{Q}_u\left(O, Ry \cap_{\mathcal{C}} A\right).$$

For example, let CS be a fuzzy set describing the accuracy/-confidence of the instances in X, $\widetilde{Q}_l$ a quantifier modelling "Most elements except maybe non confident elements" and $\widetilde{Q}_u$ a quantifier modelling "Some confident elements", then an element y is "in" the lower approximation if most accurate/confident elements indiscernible from y are in A, and an element y is "in" the upper approximation if there are some accurate/confident elements that are indiscernible to y and are "in" A.

Example IV.3. *Using a binary quantifier* $\widetilde{Q}_l$ *(e.g. representing "Most") and a unary quantifier* $\widetilde{Q}_u$ *(e.g. representing "Some") we can do this as follows:*

$$(\underline{apr}_{R,\widetilde{Q}_l} A)(y) = \widetilde{Q}_l\left((\tilde{\neg} O) \cap_{\mathcal{C}} Ry, A\right),$$
$$(\overline{apr}_{R,\widetilde{Q}_u} A)(y) = \widetilde{Q}_u\left((\tilde{\neg} O) \cap_{\mathcal{C}} Ry \cap_{\mathcal{C}} A\right).$$

V. Conclusion

We have introduced *fuzzy quantifier-based fuzzy rough sets* (FQFRS), a general definition of fuzzy rough sets based on fuzzy quantifiers. FQFRS allows to position existing models and compare them on the basis of their associated fuzzy quantifiers. In addition, this general model can lead to improved models in terms of performance and interpretability. Currently, there are only a few models that make explicit use of quantifiers, but these can be improved by using more semantically sound evaluation models. Furthermore, we have introduced novel binary quantification models based on integrals, that might give us a good compromise between computational efficiency and semantical soundness. Finally, we have introduced confidence-based FQFRS that are able to perform active outlier/noise reduction, i.e., taking into account outlier information (e.g., obtained by an outlier detection algorithm), in a more explicit and general way compared to CFRS.

VI. Future Work

From a theoretical perspective, it is interesting to find out how the properties of the used quantifiers translate to properties of the corresponding fuzzy rough sets, and vice versa. Also, it is possible to study the new fuzzy quantifiers including the quantifiers corresponding with the fuzzy removal and WOWA measures, and comparing them with existing quantifiers. Lastly we plan to perform an experimental study of the performance of the different fuzzy quantifier-based fuzzy rough sets by testing it in fuzzy rough set based classifiers similar to those considered by Lenz et al [29].

Acknowledgment

The research reported in this paper was conducted with the financial support of the Odysseus programme of the Research Foundation – Flanders (FWO). The grant number is G0H9118N.

References

[1] L. A. Zadeh, "A computational approach to fuzzy quantifiers in natural languages," in *Computational linguistics*. Elsevier, 1983, pp. 149–184.

[2] R. R. Yager, "Quantifier guided aggregation using owa operators," *International Journal of Intelligent Systems*, vol. 11, no. 1, pp. 49–73, 1996.

[3] I. Glöckner, *Fuzzy quantifiers: a computational theory*. Springer, 2008, vol. 193.

[4] Z. Pawlak, "Rough sets," *International journal of computer & information sciences*, vol. 11, no. 5, pp. 341–356, 1982.

[5] D. Dubois and H. Prade, "Rough fuzzy sets and fuzzy rough sets," *International Journal of General System*, vol. 17, no. 2-3, pp. 191–209, 1990.

[6] S. Vluymans, L. D'eer, Y. Saeys, and C. Cornelis, "Applications of fuzzy rough set theory in machine learning: a survey," *Fundamenta Informaticae*, vol. 142, no. 1-4, pp. 53–86, 2015.

[7] C. Cornelis, M. De Cock, and A. M. Radzikowska, "Vaguely quantified rough sets," in *International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing*. Springer, 2007, pp. 87–94.

[8] J. Fernández Salido and S. Murakami, "On β-precision aggregation," *Fuzzy Sets and Systems*, vol. 139, no. 3, pp. 547–558, 2003. doi: https://doi.org/10.1016/S0165-0114(03)00003-4. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165011403000034

[9] ——, "Rough set analysis of a general type of fuzzy data using transitive aggregations of fuzzy similarity relations," *Fuzzy Sets and Systems*, vol. 139, no. 3, pp. 635–660, 2003. doi: https://doi.org/10.1016/S0165-0114(03)00124-6. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165011403001246

[10] A. Mieszkowicz-Rolka and L. Rolka, "Variable precision fuzzy rough sets," in *Transactions on Rough Sets I*. Springer, 2004, pp. 144–160.

[11] 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*, vol. 236, pp. 58–72, 2014. doi: https://doi.org/10.1016/j.fss.2013.06.012 Theme: Algebraic Aspects of Fuzzy Sets. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165011413002753

[12] Q. Hu, S. An, and D. Yu, "Soft fuzzy rough sets for robust feature evaluation and selection," *Information Sciences*, vol. 180, no. 22, pp. 4384–4400, 2010. doi: https://doi.org/10.1016/j.ins.2010.07.010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025510003282

[13] A. Hadrani, K. Guennoun, R. Saadane, and M. Wahbi, "Fuzzy rough sets: Survey and proposal of an enhanced knowledge representation model based on automatic noisy sample detection," *Cognitive Systems Research*, vol. 64, pp. 37–56, 2020. doi: https://doi.org/10.1016/j.cogsys.2020.05.001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1389041720300255

[14] S. An, Q. Hu, W. Pedrycz, P. Zhu, and E. C. C. Tsang, "Data-distribution-aware fuzzy rough set model and its application to robust classification," *IEEE Transactions on Cybernetics*, vol. 46, no. 12, pp. 3073–3085, 2016. doi: 10.1109/TCYB.2015.2496425

[15] S. An, Q. Hu, and C. Wang, "Probability granular distance-based fuzzy rough set model," *Applied Soft Computing*, vol. 102, p. 107064, 2021. doi: https://doi.org/10.1016/j.asoc.2020.107064. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494620310024

[16] C. Cornelis, N. Verbiest, and R. Jensen, "Ordered weighted average based fuzzy rough sets," in *International Conference on Rough Sets and Knowledge Technology*. Springer, 2010, pp. 78–85.

[17] A. Theerens, O. U. Lenz, and C. Cornelis, "Choquet-based fuzzy rough sets," *International Journal of Approximate Reasoning*, 2022. doi: 10.1016/j.ijar.2022.04.006

[18] H.-P. Kriegel, P. Kroger, E. Schubert, and A. Zimek, "Interpreting and unifying outlier scores," in *Proceedings of the 2011 SIAM International Conference on Data Mining*. SIAM, 2011, pp. 13–24.

[19] L. A. Zadeh, "Fuzzy sets," *Information and Control*, 1965.

[20] L. D'eer, N. Verbiest, C. Cornelis, and L. Godo, "A comprehensive study of implicator–conjunctor-based and noise-tolerant fuzzy rough sets: definitions, properties and robustness analysis," *Fuzzy Sets and Systems*, vol. 275, pp. 1–38, 2015.

[21] R. R. Yager, "On ordered weighted averaging aggregation operators in multicriteria decisionmaking," *IEEE Transactions on systems, Man, and Cybernetics*, vol. 18, no. 1, pp. 183–190, 1988.

[22] G. Beliakov, A. Pradera, T. Calvo *et al.*, *Aggregation functions: A guide for practitioners*. Springer, 2007, vol. 221.
[23] Z. Wang and G. J. Klir, *Generalized measure theory*. Springer Science & Business Media, 2010, vol. 25.
[24] A. Cascallar-Fuentes, A. Ramos-Soto, and A. Bugarín-Diz, "An experimental study on the behaviour of fuzzy quantification models," in *ECAI 2020*. IOS Press, 2020, pp. 267–274.
[25] M. Delgado, M. D. Ruiz, D. Sánchez, and M. A. Vila, "Fuzzy quantification: a state of the art," *Fuzzy Sets and Systems*, vol. 242, pp. 1–30, 2014.
[26] V. Torra, "The weighted owa operator," *International Journal of Intelligent Systems*, vol. 12, no. 2, pp. 153–166, 1997.
[27] ——, "On some relationships between the wowa operator and the Choquet integral," in *Proceedings of the IPMU 1998 Conference, Paris, France*. Citeseer, 1998, pp. 818–824.
[28] F. Diaz-Hermida, D. Losada, A. Bugarin, and S. Barro, "A probabilistic quantifier fuzzification mechanism: The model and its evaluation for information retrieval," *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 5, pp. 688–700, 2005. doi: 10.1109/TFUZZ.2005.856557
[29] O. U. Lenz, D. Peralta, and C. Cornelis, "Scalable approximate frnn-owa classification," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 5, pp. 929–938, 2019. doi: 10.1109/TFUZZ.2019.2949769