

Hybridization of Fuzzy Sets and Rough Sets: Achievements and Opportunities

Chris Cornelis

Computational Web Intelligence

Department of Applied Mathematics, Computer Science and Statistics

Ghent University

Krijgslaan 281 (S9), B-9000 Gent, Belgium

Email: Chris.Cornelis@UGent.be

Abstract—Fuzzy rough sets are the fruit of an intense and long-lasting collaboration effort between fuzzy set theory and rough set theory. Seminal research on the hybridization originated in the late 1980’s, and has inspired generations of researchers from around the globe to address both theoretical and practical challenges. In this paper, we gauge the state-of-the-art in this domain and identify opportunities for further development. In particular, we highlight the potential of fuzzy quantifiers in creating new robust fuzzy rough models, we advocate closer integration with granular computing as a stepping stone for designing rule induction algorithms, and we contemplate the role of fuzzy rough sets vis-à-vis explainable artificial intelligence.

I. INTRODUCTION

FUZZY ROUGH SETS emerge as a combination of fuzzy sets (Zadeh [62], 1965) and rough sets (Pawlak [43], 1982): while the former model vague information by recognizing that membership to certain concepts, or logical truth of certain propositions, is a matter of degree, the latter handle potentially inconsistent information by providing a lower and upper approximation of a concept, using the equivalence classes of an indiscernibility relation as building blocks. Both frameworks can be integrated from at least three different perspectives:

- 1) Concepts may be fuzzy rather than exact, allowing that objects belong to it to varying degrees. For example, in a data set containing information about hotels, one may be interested to characterize the concept of hotels considered as “expensive”, an inherently vague predicate. Then, each hotel’s membership to the fuzzy concept “expensive hotel” will be expressed using a value between 0 (not belonging to the concept at all, i.e., not expensive) and 1 (fully meeting the concept’s membership conditions, i.e., definitely expensive).
- 2) The indiscernibility relation, expressing that objects may or may not be distinguished from each other, may be gradual rather than strict; this reflects the intuitive idea that some objects are more similar to each other than others, and therefore they should be related to a higher degree (again a value between 0 and 1).

- 3) The condition for belonging to the lower and upper approximation may be expressed using *fuzzy quantifiers*. For example, classically an object is a member of the lower approximation of a concept if *all* objects indiscernible from it also belong to the concept. Here, instead of the traditional universal quantifier (\forall), one may use a fuzzy quantifier like “most”. The purpose of such a relaxation is to introduce a measure of tolerance towards inconsistency into the approximations, making them more robust.

The first two principles were already present early on in the work of Fariñas del Cerro and Prade [17], and Dubois and Prade [16], while the third one was first explored in 2007 by the theory of vaguely quantified rough sets [5] and recently revived by the introduction of Choquet-based fuzzy rough sets [49]. In this paper, we want to highlight the potential of fuzzy quantifiers in designing robust and interpretable fuzzy rough set models, which are in particular relevant for applications in data analysis [52]. The latter are becoming more and more widespread, and include, amongst others, feature selection [8], [13], [39], instance selection [29], [50], [51], instance-based classification [25], [30], [32], [33], cognitive networks [37], imbalanced classification [47], [54], multi-instance classification [53] and multi-label classification [35], [55].

Given that many of the aforementioned applications are instance-based or greedy approaches, whereas most successful applications of the rough set paradigm are rule based systems (see e.g. [23], [24]), it may be argued that the full potential of the hybrid theory has not yet been tapped (some notable exceptions are [22], [65], [66]), taking into account also the vast body of existing research on fuzzy rule based systems [2], [18], [26]. An important key to this logical next step lies in the field of granular computing [4], [59], an information processing paradigm centered on the segmentation of complex information into smaller pieces called information granules. Both rough sets and fuzzy sets relate to granular computing; it is well-known that the lower and upper approximation can be represented as unions of simple sets or granules induced from data [60], while Zadeh [64] identified fuzziness as a key part of the granulation in human cognition, and with

This work was supported by the Odysseus programme (grant number G0H9118N) of the Research Foundation – Flanders (FWO).

the help of fuzzy granules, a fuzzy rule based system may be set up. An important advantage of fuzzy rules, and of fuzzy logic in general, is that they allow for explanations using linguistic expressions. This property can be utilized for the development of interpretable machine learning algorithms, a research direction which is currently attracting a lot of researchers' attention [36].

The remainder of this paper is structured as follows: in Section II, we recall important preliminaries from both rough set and fuzzy set theory, while in Section III, we outline the main steps and results of the hybridization process. In Section IV, we pay attention to robust fuzzy rough set models, which play an important role for practical applications of the theory. Finally, in Section V, we offer an informal discussion of ongoing challenges and new opportunities for the hybrid theory.

II. PRELIMINARIES: FUZZY SETS AND ROUGH SETS

A. Rough sets

We first recall Pawlak's definition [43] of rough sets, which is also called the Indiscernibility-based Rough Set Approach (IRSA).

Definition 2.1: Let U be a set of objects and E an equivalence relation expressing indiscernibility, i.e., E is

- reflexive: $(u, u) \in E$,
- symmetric: if $(u, v) \in E$ then $(v, u) \in E$,
- transitive: if $(u, v) \in E$ and $(v, w) \in E$, then also $(u, w) \in E$.

For any $A \subseteq U$, the *lower and upper approximation* of A are defined as:

$$\underline{\text{apr}}_E(A) = \{u \in U : [u]_E \subseteq A\} \quad (1)$$

$$\overline{\text{apr}}_E(A) = \{u \in U : [u]_E \cap A \neq \emptyset\} \quad (2)$$

The couple $(\underline{\text{apr}}_E(A), \overline{\text{apr}}_E(A))$ is called the *rough set* of A . The equations (1) and (2) can be expressed equivalently using logical operators: for $u \in U$,

$$u \in \underline{\text{apr}}_E(A) \Leftrightarrow (\forall v \in U)((v, u) \in E \Rightarrow v \in A) \quad (3)$$

$$u \in \overline{\text{apr}}_E(A) \Leftrightarrow (\exists v \in U)((v, u) \in E \wedge v \in A) \quad (4)$$

It is also easily verified that $\underline{\text{apr}}_E(A) \subseteq A \subseteq \overline{\text{apr}}_E(A)$, which justifies the terms "lower and upper approximation". Moreover, the rough set approximations satisfy various other properties, for example set monotonicity:

$$A \subseteq A' \Rightarrow \underline{\text{apr}}_E(A) \subseteq \underline{\text{apr}}_E(A') \wedge \overline{\text{apr}}_E(A) \subseteq \overline{\text{apr}}_E(A') \quad (5)$$

which expresses that if a concept becomes larger, its approximations naturally should not decrease. On the other hand, relation monotonicity:

$$E \subseteq E' \Rightarrow \underline{\text{apr}}_E(A) \supseteq \underline{\text{apr}}_{E'}(A) \wedge \overline{\text{apr}}_E(A) \subseteq \overline{\text{apr}}_{E'}(A) \quad (6)$$

states that when equivalence classes become larger (more objects are indiscernible from each other), the lower approximation gets smaller, while more objects populate the upper approximation.

In case $\underline{\text{apr}}_E(A) = \overline{\text{apr}}_E(A)$, we call A an *exact set*. An equivalent way of expressing that A is exact is

$$A = \bigcup_{u \in A} [u]_E \quad (7)$$

In other words, A can be seen as a union of basic building blocks or granules, which correspond to equivalence classes of E . We call (7) the *granular representation* of A , and A is also called a *granularly representable set*. The following proposition highlights the special role of the lower and upper approximation as specific exact sets.

Proposition 2.2: For $A \subseteq U$, the greatest granularly representable set that is included in A is equal to $\underline{\text{apr}}_E(A)$, while the smallest granularly representable set that includes A is equal to $\overline{\text{apr}}_E(A)$.

On the other hand, the granular representation is also closely connected with the notion of consistency.

Proposition 2.3: Set $A \subseteq U$ is granularly representable if and only if it satisfies the consistency property, i.e., iff

$$(\forall u, v \in U)((v, u) \in E \wedge u \in A \Rightarrow v \in A) \quad (8)$$

Consistency expresses that if two objects are indiscernible and one of them belongs to a given concept A , the second object should necessarily also be part of the concept. This property is desirable in classification problems, where the goal is to establish meaningful patterns that allow to decide the membership of unseen objects to given decision classes. In this context, objects are also called instances and are characterized by their values for a number of attributes from a set \mathcal{A} . The domain of every attribute $a \in \mathcal{A}$ consists of a finite number of nominal values, and every instance $u \in U$ takes one of those values denoted with $a(u)$. Then, the equivalence relation E is constructed as

$$(u, v) \in E \Leftrightarrow (\forall a \in \mathcal{A})(a(u) = a(v)) \quad (9)$$

The granular representation of rough sets is in particular very useful from the perspective of rule induction. The problem of rule induction for classification tasks amounts to generating a set of rules which relate descriptions of objects by subsets of attributes with particular decision classes. Basic granules, from which rough sets are composed, can be interpreted as human readable "if..., then..." rules, and can be used to construct a rule based inference system as a prediction model. Well-known examples of rule induction algorithms are LEM2 [23] and MODLEM [24].

Pawlak's theory has been generalized in various different ways. For example, dropping the symmetry requirement from E leads to the Preorder-based Rough Set Approach (PRSA, [40]), which contains as a special case the Dominance-based Rough Set Approach (DRSA, [21]). In the latter, the domain of attributes now contains ordinal values, and the indiscernibility relation is replaced by a dominance relation. For clarity and brevity of the exposition, in the remainder of this paper we will focus on the indiscernibility-based approach, although many of the presented results remain valid for more general settings.

B. Fuzzy sets

Given a universal set U , Zadeh defined a fuzzy set A in U simply as a mapping from U to the unit interval $[0, 1]$, where $A(u)$ is called the membership degree of object u to A . It expresses to what extent u satisfies the vague property expressed by the fuzzy set A . For example, if U is a set of hotels from a given area, we may evaluate their expensiveness, based on their quoted nightly rate for a double room, as a fuzzy set A in U . Clearly, the assignment of membership degrees is both subjective and context-dependent, as it would depend for example on the budget of the person making the booking, and the area where the search is performed. However, an intuitive constraint in this case would be that the higher the quoted rate, the larger the membership degree should be. In this example, as in most practical applications of fuzzy set theory, there is an underlying numerical scale (a subset of the real numbers) on which the evaluation is made, and the ordering on that scale constrains the assignment of membership degrees.

In a similar vein, a binary fuzzy relation R in U is defined as a fuzzy set in U^2 , i.e., for any two objects u and v in U , $R(u, v)$ expresses the degree to which they relate. Fuzzy relations may be used for example to generalize the equivalence relation E from Section II-A, to establish to what extent two objects are *similar* (as opposed to a black-or-white assessment whether they are indiscernible or not). Such a fuzzy relation R should be at least reflexive and symmetric, i.e.,

$$R(u, u) = 1 \quad (10)$$

$$R(u, v) = R(v, u) \quad (11)$$

should hold for any u and v in U . In order to accommodate for the transitivity property, we first need an extension of the classical conjunction operator \wedge .

Definition 2.4: A triangular norm, or shortly t-norm, is a mapping $T : [0, 1]^2 \rightarrow [0, 1]$ that is commutative, associative, increasing in both arguments, and that satisfies the boundary condition $T(1, x) = x$ for all $x \in [0, 1]$.

Well-known representatives of the class of t-norms include the minimum, the product, and the Łukasiewicz t-norm defined by $T_{\mathbb{L}}(x, y) = \max(0, x + y - 1)$ for x, y in $[0, 1]$. The choice for a particular t-norm depends on the particular properties that one is interested in; for a comprehensive overview, we refer to [31].

Using a t-norm, we may now impose a kind of transitivity on fuzzy relations, and therefore extend the notion of an equivalence relation.

Definition 2.5: Let T be a t-norm. A fuzzy relation R in U that is reflexive, symmetric and satisfies

$$T(R(u, v), R(v, w)) \leq R(u, w) \quad (12)$$

for any u, v, w in U is called a fuzzy T -equivalence relation. Apart from logical conjunction, we will also require an extension of the boolean implication operator \Rightarrow .

Definition 2.6: An implicator is a mapping $I : [0, 1]^2 \rightarrow [0, 1]$ that is decreasing in its first argument and increasing

in its second one, and that satisfies the boundary conditions $I(0, 0) = I(0, 1) = I(1, 1) = 1$ and $I(1, 0) = 0$.

There exist numerous ways to construct implicators. Again, a detailed overview is out of the scope of this paper, and may be found in e.g. [3]. A popular approach is to associate implicators to t-norms by means of *residuation*, leading to the following definition of residuated implicators, or shortly R-implicators.

Definition 2.7: The R-implicator I_T associated to a t-norm T is defined by, for x, y in $[0, 1]$:

$$I_T(x, y) = \sup\{z \in [0, 1] \mid T(x, z) \leq y\} \quad (13)$$

As an example, the R-implicator associated to the Łukasiewicz t-norm can be obtained as $I_{T_{\mathbb{L}}}(x, y) = \min(1, 1 - x + y)$.

Finally, we recall that subsethood for fuzzy sets is defined as follows [62]: for fuzzy sets A and B in U ,

$$A \subseteq B \Leftrightarrow (\forall u \in U)(A(u) \leq B(u)) \quad (14)$$

III. HYBRIDIZATION: GENERAL FUZZY ROUGH SET MODEL

The equations (3) and (4) for determining membership to the classical lower and upper approximations can be “fuzzified” by making use of fuzzy logical connectives. This leads to the following definition [48], [6], [11].

Definition 3.1: Let A be a fuzzy set in U , R a fuzzy relation in U , I an implicator and T a t-norm. The lower and upper approximation of A are defined as, for $u \in U$,

$$\underline{\text{apr}}_R(A)(u) = \inf_{v \in U} I(R(v, u), A(v)) \quad (15)$$

$$\overline{\text{apr}}_R(A)(u) = \sup_{v \in U} T(R(v, u), A(v)) \quad (16)$$

The couple $(\underline{\text{apr}}_R(A), \overline{\text{apr}}_R(A))$ is called the *fuzzy rough set* of A . If $\underline{\text{apr}}_R(A) = A = \overline{\text{apr}}_R(A)$, A is called an exact fuzzy set.

Take note how these definitions extend their classical counterparts:

- 1) Object u belongs to the lower approximation of A to the extent that *for all* objects v , *if* v is related to u by R , *then* v should belong to A .
- 2) Object u belongs to the upper approximation of A to the extent that *there exists* an object v , such that *if* v is related to u by R , *and* v belongs to A .

In other words, the inf and sup operators naturally represent the \forall and \exists quantifier from Eq. (3) and (4), respectively, while the implicator I and t-norm T fulfil the role of the logical implication \Rightarrow and conjunction \wedge . When A is a classical, non-fuzzy set and R is a crisp equivalence relation, we again obtain Pawlak’s model. Depending on the specific choice of fuzzy connectives I and T , and the fuzzy relation R , some properties of this original model may or may not be preserved (see [12] for more details).

An important question is whether exact fuzzy sets possess a granular representation analogous to Eq. (7), as it would allow the above approximations to be used for generating fuzzy rules in a similar way as is done with crisp granules. Degang et al. [14] were the first to address this issue by formalizing the

notions of a fuzzy granule and granular representability of fuzzy sets.

Definition 3.2: Let R be a fuzzy T -equivalence relation in U for a given t-norm T and $\lambda \in [0, 1]$. The fuzzy granule corresponding to R , λ and T is the fuzzy set R_λ in U , defined by

$$R_\lambda(u) = \{(v, T(R(v, u), \lambda)); v \in U\} \quad (17)$$

We call a fuzzy set A in U granularly representable if

$$A = \bigcup \{R_{A(u)}(u); u \in U\} \quad (18)$$

In other words, A is granularly representable if it is the union of fuzzy granules $R_\lambda(u)$, where $\lambda = A(u)$ for each object u . The following proposition reveals that for particular choices of I and T , exact fuzzy sets indeed correspond to granularly representable ones, and vice versa.

Proposition 3.3: Let A be a fuzzy set in U , T a left-continuous t-norm and I its R-implicator. Then A is exact if and only if it is granularly representable.

Along the same lines, we can also generalize Proposition 2.2 and 2.3.

Proposition 3.4: For a fuzzy set A in U and a fuzzy T -equivalence relation, the greatest granularly representable fuzzy set that is included in A is equal to $\underline{\text{apr}}_R(A)$, while the smallest granularly representable set that includes A is equal to $\overline{\text{apr}}_R(A)$.

Proposition 3.5: Let R be a fuzzy T -equivalence relation. Fuzzy set A in U is granularly representable if and only if it satisfies the fuzzy consistency property, i.e., iff

$$(\forall u, v \in U)(T(R(v, u), A(u)) \leq A(v))$$

Note how fuzzy consistency provides us with a softened version of Eq. (8): the more similar u and v are, and the higher u 's membership to A , the more v should also belong to A .

IV. ROBUST FUZZY ROUGH SETS

The model of fuzzy rough sets described in the previous section offers considerable strength and flexibility, and lends itself very well for handling datasets with real-valued attributes, where fuzzy T -equivalence relations can be constructed by taking into account the distance between individual instances' attribute values. However, it may still be too rigid when applied in practical problems of data analysis, due to the occurrence of outliers. By the latter, we mean instances that do not follow the general data distribution, and which may negatively impact the quality of the fuzzy-rough approximations. In extreme cases, the lower approximation of a concept may be empty, while its upper approximation may contain all instances fully.

The root of the problem lies in the use of the inf and sup operators which, as we explained, correspond to the \forall and \exists quantifier, respectively. Because of this, an instance u will be fully excluded from $\underline{\text{apr}}_R(A)$ as soon as there exists another instance v such that $R(v, u) = 1$ and $A(v) = 0$, while on the other hand u will fully belong to $\overline{\text{apr}}_R(A)$ when an object

v can be found such that $R(v, u) = 1$ and $A(v) = 1$. This will occur independently of the choice of the implicator I and the t-norm T . While this effect may be mitigated by a thoughtful choice of the fuzzy relation R , it cannot be ruled out altogether as (partial) inconsistencies are commonplace in real applications.

In classical rough set theory, researchers also faced this problem, leading to probabilistic approaches like Ziarko's Variable Precision Rough Set (VPRS) model [67]. The latter relaxes Eq. (3) and (4) into

$$u \in \underline{\text{apr}}_E^p(A) \Leftrightarrow \frac{|[u]_E \cap A|}{|[u]_E|} \geq p \quad (19)$$

$$u \in \overline{\text{apr}}_E^q(A) \Leftrightarrow \frac{|[u]_E \cap A|}{|[u]_E|} > q \quad (20)$$

where $1 \geq p > q \geq 0$ are parameters of the model. In other words, an object belongs to the VPRS lower approximation if at least a fraction p of its equivalence class belongs to A , while it belongs to the upper approximation if more than a fraction q of $[u]_E$ is inside A . The model assumes that U is finite (which is not a problem considering that its application is in data analysis), and that $p > q$, to ensure that $\underline{\text{apr}}_E^p(A) \subseteq \overline{\text{apr}}_E^q(A)$. When $p = 1$ and $q = 0$, we recover Pawlak's original equations (3) and (4). In general, probabilistic rough set approaches have been exploited successfully for classification purposes, most notably within Yao's framework of three-way decisions [61].

Ziarko's VPRS model served as an inspiration source for different robust fuzzy rough set proposals. One of them, the Vaguely Quantified Rough Set (VQRS) model [5] softens the membership criterion for an object u to belong to the lower approximation of A into "most elements of $[u]_E$ are inside A ". Similarly, u belongs to the VQRS upper approximation of A to the extent that "at least some elements in $[u]_E$ belong to A ". To formalize this idea, the inherently fuzzy quantifiers "most" and "at least some" are modeled as specific fuzzy sets in the unit interval [63]:

Definition 4.1: A fuzzy set Q in $[0, 1]$ is called a regular increasing monotone (RIM) quantifier if Q is non-decreasing, $Q(0) = 0$ and $Q(1) = 1$.

The class of RIM quantifiers includes as special cases the existential and the universal quantifier:

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

Examples of RIM quantifiers that also take on values from the interior of the unit interval can be obtained using the following parametrized formula [5], 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 to represent the fuzzy quantifiers "at least some" and "most" from natural

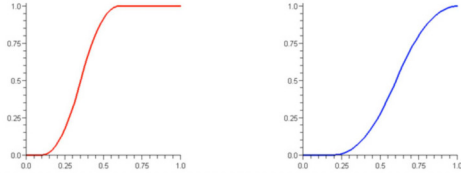


Fig. 1. The RIM quantifiers $Q_{(0.1,0.6)}$ (left) and $Q_{(0.2,1)}$ (right)

language. They are depicted in Figure 1. In general, assuming RIM quantifiers Q_1 and Q_2 such that $Q_1 \subseteq Q_2$, we may define the VQRS approximations of A : for $u \in U$,

$$\underline{\text{apr}}_E^{Q_1}(A)(u) = Q_1\left(\frac{|[u]_E \cap A|}{|[u]_E|}\right) \quad (21)$$

$$\overline{\text{apr}}_E^{Q_2}(A)(u) = Q_2\left(\frac{|[u]_E \cap A|}{|[u]_E|}\right) \quad (22)$$

The above definition has the peculiarity that although both the set to be approximated and the equivalence relation are non-fuzzy, the resulting approximations may well be fuzzy. The reasoning behind this is that for example the membership degree in Eq. (21) evaluates the degree of fulfilment of the condition “ Q_1 elements of $[u]_E$ are in A ”. Note that if $Q_1 = Q_\forall$ and $Q_2 = Q_\exists$, we again arrive at Pawlak’s lower and upper approximation, while the VPRS equations (19) and (20) are also special cases of (19) and (20) using the RIM quantifiers

$$Q_1(x) = \begin{cases} 0, & x \leq p \\ 1, & x \geq p \end{cases} \quad Q_2(x) = \begin{cases} 0, & x < q \\ 1, & x > q \end{cases}$$

The VQRS equations may be further generalized to a fuzzy set A and a fuzzy relation R ; for details, we refer to [5]. Despite its intuitive appeal, the VQRS model has an important shortcoming which it shares with the VPRS model: it does not satisfy relation monotonicity, Eq. (6). This is in particular problematic in applications where the (fuzzy) indiscernibility relation is iteratively refined by adding more information (additional attributes). For example, in [7] it is shown how this affects the operation of the greedy QuickReduct feature selection algorithm.

A solution to this problem can be found by revisiting the equations (15) and (16) and replacing the inf and sup operators by less extreme ones. First note that for finite universes, inf and sup correspond to min and max, respectively. So, the lower approximation (15) is determined solely by the smallest one among all $I(R(v, u), A(u))$ values, and the single largest value $T(R(v, u), A(u))$ will set the upper approximation. A more balanced evaluation is offered by using ordered weighted average (OWA) operators [57]: given an input vector of $n \geq 1$ real values $\langle a_1, \dots, a_n \rangle$ and a weight vector $W = \langle w_1, \dots, w_n \rangle$ such that each $w_i \in [0, 1]$ and $\sum_{i=1}^n w_i = 1$, we first order the input values a_i from large to small obtaining $\langle c_1, \dots, c_n \rangle$ and then compute

$$OWA_W \langle a_1, \dots, a_n \rangle = \sum_{i=1}^n w_i c_i \quad (23)$$

This leads to the definitions of the OWA-based lower and upper approximation [9]:

$$\underline{\text{apr}}_R^{W_1}(A)(u) = OWA_{W_1} \underbrace{I(R(v, u), A(u))}_{v \in U} \quad (24)$$

$$\overline{\text{apr}}_R^{W_2}(A)(u) = OWA_{W_2} \underbrace{T(R(v, u), A(u))}_{v \in U} \quad (25)$$

Using $W_1 = \langle 0, \dots, 0, 1 \rangle$ and $W_2 = \langle 1, 0, \dots, 0 \rangle$, we obtain the original Eq. (15) and (16). Because of the monotonicity properties of T , I and the OWA operator, relation monotonicity is guaranteed, and moreover it always holds that

$$\underline{\text{apr}}_R(A) \subseteq \underline{\text{apr}}_R^{W_1}(A) \text{ and } \overline{\text{apr}}_R^{W_2}(A) \subseteq \overline{\text{apr}}_R(A) \quad (26)$$

In other words, OWA-based fuzzy rough sets indeed relax the original definitions, enlarging the lower approximation and restricting the upper one. In [56], different weighting schemes were discussed and evaluated experimentally.

Recently, in [41] it was shown that for certain choices of the t-norm T (including the product and Łukasiewicz t-norm, but not minimum), the OWA-based lower and upper approximations of any fuzzy set A are exact sets, i.e.

$$\underline{\text{apr}}_R(\underline{\text{apr}}_R^{W_1}(A)) = \overline{\text{apr}}_R(\underline{\text{apr}}_R^{W_1}(A)) = \underline{\text{apr}}_R^{W_1}(A) \quad (27)$$

$$\underline{\text{apr}}_R(\overline{\text{apr}}_R^{W_2}(A)) = \overline{\text{apr}}_R(\overline{\text{apr}}_R^{W_2}(A)) = \overline{\text{apr}}_R^{W_2}(A) \quad (28)$$

This means in particular that these approximations possess a granular representation, and can be used as a basis for fuzzy rule induction algorithms, as discussed in the next section.

V. DISCUSSION: CHALLENGES AND OPPORTUNITIES FOR FUZZY ROUGH SETS

Even though the VQRS model, considering its violation of the relation monotonicity property, has been mostly abandoned in favour of the OWA-based approach, its interpretation of membership to the fuzzy-rough approximations in terms of fuzzy quantifiers expressing “most” and “at least some” is arguably more intuitive and transparent than the one using the rather less compact representation of OWA weight vectors.

Yet this does not mean that OWA fuzzy rough sets are isolated from vague quantification. In fact, as explained in [49], from any OWA weight vector $W = \langle w_1, \dots, w_n \rangle$, a corresponding RIM quantifier Q can be derived by setting

$$Q(x) = \sum_{i \leq xn} w_i \quad (29)$$

and, vice versa, for every RIM quantifier Q and $n \geq 1$, the associated OWA weights w_i ($i = 1, \dots, n$) are determined by

$$w_i = Q\left(\frac{i}{n}\right) - Q\left(\frac{i-1}{n}\right) \quad (30)$$

As such, the weight vectors W_1 and W_2 featured in Eq. (24) and (25) are mutually interchangeable with their VQRS counterparts Q_1 and Q_2 . The resulting membership degrees, however, carry a different meaning; for example, $\underline{\text{apr}}_R^{W_1}(A)(u)$ should be understood as the truth value of the statement “for

most objects in U , it holds that if they are indiscernible from u , then they also belong to A "; in other words, the quantification also takes into account objects completely unrelated to u (i.e., fully discernible from u), while for the computation of the membership to the VQRS lower approximation, these objects are excluded.

A more serious limitation of the OWA fuzzy rough set model lies in the fact that it treats all objects symmetrically during the aggregation process, i.e., an individual object's impact is determined merely by its fulfillment of a specific logical formula. In practice, this means that we hypothesize that "a limited amount" of objects are outliers, and that their effect will be cancelled out by the chosen weighting scheme. Suppose however that we have specific knowledge that some specific objects are in fact certainly outliers, e.g., based on an outlier score that was calculated for them separately. Then, a more natural way to evaluate whether an object u belongs to the lower approximation of concept A is by checking if all objects indiscernible from u , *except* perhaps those which are considered outliers, belong to A .

In order to accommodate the above and other related use case scenarios, new definitions were proposed for the fuzzy-rough lower and upper approximation in [49] recently. They are based on the Choquet integral, a generalization of the classical Lebesgue integral to non-additive measures which has become popular as an aggregation function in decision making [20]. It was shown how the resulting Choquet-based fuzzy rough sets (CFRS) contain the OWA-based model as a special case, while inheriting some of its desirable properties, including set and relation monotonicity. At the same time, they also maintain the intuitive interpretation in terms of vague quantification. Considering that most of the existing fuzzy rough set models still rely on "traditional" approaches to fuzzy quantifiers such as those proposed by Zadeh [63] and Yager [58], which were shown to suffer from some serious conceptual flaws [19], this opens up exciting research opportunities involving more recent developments (see e.g. [15] for an overview).

An important unresolved question about the new CFRS model is whether it still conforms to the granular structure that the OWA-based model from Eq. (24) and (25), and the traditional fuzzy rough set model from Eq. (15) and (16) exhibit, in other words: whether its approximations are exact fuzzy sets in the sense of Eq. (18). While technical in nature, if this question can be answered positively, it opens the doors to fuzzy rule induction methods based on these fuzzy-rough approximations. Indeed, the fuzzy granules corresponding to the approximations can be used inside the antecedent part of fuzzy rules, and an unseen test object's membership in them may be interpreted as the firing strength of the corresponding rule.

As a concrete example, let us consider rule-based classification. In this case, U is partitioned into a number of decision classes (concepts). Let C be one such decision class, and denote its lower approximation, computed according to one of the models discussed in this paper, by $\underline{apr}(C)$. Then, if

$\underline{apr}(C)$ is granularly representable, by Eq. (18) a corresponding decision rule will be generated for every training object u , such that for a given test object v , the firing strength of this rule is obtained as

$$T(R(v, u), \underline{apr}(C)(u)) \quad (31)$$

in other words, as a conjunction between $R(v, u)$, the observed similarity between u and v , and $\underline{apr}(C)(u)$, the membership of the training object u to the lower approximation.

Decision rules based on the lower approximation are usually called "certain" rules, while we refer to those based on the upper approximation as "possible" rules, distinguishing their relative strength. More generally, fuzzy decision rules can be derived from any granularly representable fuzzy set associated to decision classes, for instance from the so-called granular fuzzy-rough approximation introduced in [42], which is defined as the closest granularly representable fuzzy set (w.r.t. a certain loss function) to a given concept, and which is obtained as the result of a linear programming problem.

In practice, however, generating one rule per training object is not a viable approach, and an important challenge is therefore to design proper rule induction algorithm that can at the same time reduce the number of rules, as well as maximize the number of objects that each rule covers. Such a strategy was already pursued in [28], where it was combined with fuzzy rough set guided feature selection, and various other attempts (see e.g. [22], [38], [65], [66]) have also been made to integrate fuzzy rough sets and rule induction; yet, a convincing proposal of a "fuzzy LEM" classification algorithm is still missing and could represent a breakthrough in this domain, not in the least from the perspective of interpretable machine learning.

Indeed, the generation of compact fuzzy rules benefits the human understanding of classification algorithms based on them, as rule-based models are some of the most interpretable models, and they closely resemble human cognition [1]. In [2], various criteria were distinguished for interpretability at different levels of a fuzzy rule-based system, including linguistic variables and fuzzy granules. An integration of these criteria with fuzzy-rough rule induction is therefore at hand to develop a coherent and compact model of interpretable granular computing. Apart from the granules themselves, an important role should again be reserved for fuzzy quantifiers, as they are useful to summarize knowledge in a concise, linguistic way.

ACKNOWLEDGMENTS

The author would like to thank Marko Palanetić and Adnan Theerens for insightful discussions that helped shape this paper.

REFERENCES

- [1] V. Aleven, Rule-Based Cognitive Modeling for Intelligent Tutoring Systems, Springer, pp. 33–62, 2010.
- [2] J. Alonso, C. Castiello, L. Magdalena, C. Mencar, Explainable fuzzy systems: paving the way from interpretable fuzzy systems to explainable AI Systems, Springer, 2021.
- [3] M. Baczyński, B. Jayaram, Fuzzy implications, Springer, 2008.

- [4] A. Bargiela, W. Pedrycz, The roots of granular computing, in: 2006 IEEE International Conference on Granular Computing, pp. 806–809, 2006.
- [5] C. Cornelis, M. De Cock, A. Radzikowska, Vaguely quantified rough sets, in: Proceedings of 11th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing (RSFDGrC2007), Lecture Notes in Artificial Intelligence 4482, 2007, pp. 87–94.
- [6] C. Cornelis, M. De Cock, A.M. Radzikowska, Fuzzy rough sets: from theory into practice, in: Handbook of Granular Computing (W. Pedrycz, A. Skowron, V. Kreinovich, eds.), Wiley, 2008, pp. 533–552.
- [7] C. Cornelis, R. Jensen, A noise-tolerant approach to fuzzy-rough feature selection, in: Proc. 2008 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2008), 2008, pp. 1598–1605.
- [8] C. Cornelis, R. Jensen, G. Hurtado Martín, and D. Ślęzak, Attribute selection with fuzzy decision reducts, *Information Sciences*, **180**(2), 2010, pp. 209–224.
- [9] C. Cornelis, N. Verbiest, and R. Jensen, Ordered weighted average based fuzzy rough sets, in: Proc. 5th International Conference on Rough Sets and Knowledge Technology (RSKT 2010), 2010, pp. 78–85.
- [10] M. De Cock, C. Cornelis, E.E. Kerre, Fuzzy rough sets: beyond the obvious, in: Proc. 2004 IEEE Int. Conf. on Fuzzy Systems, FUZZ-IEEE'04, Volume 1, 2004, pp. 103–108.
- [11] M. De Cock, C. Cornelis, E.E. Kerre, Fuzzy rough sets: the forgotten step, *IEEE Transactions on Fuzzy Systems* **15**(1), 2007, pp. 121–130.
- [12] L. D'eer, N. Verbiest, C. Cornelis, L. Godo, A comprehensive study of implicator-conjunctive based and noise-tolerant fuzzy rough sets: definitions, properties and robustness analysis, *Fuzzy Sets and Systems* **275**, 2015, pp. 1–38.
- [13] C. Degang, Z. Suyun, Local reduction of decision system with fuzzy rough sets, *Fuzzy Sets and Systems* **161**(13), 2010, pp. 1871–1883. Chen Degang, Zhao Suyun,
- [14] C. Degang, Y. Yongping, W. Hui, Granular computing based on fuzzy similarity relations, *Soft Computing* **15**(6), 2011, pp. 1161–1172.
- [15] M. Delgado, M. D. Ruiz, D. Sánchez, M. A. Vila, Fuzzy association rules: general model and applications, *IEEE Transactions on Fuzzy Systems* **11**(2), 2003, pp. 214–225.
- [16] M. Delgado, M. D. Ruiz, D. Sánchez, M. A. Vila, Fuzzy quantification: a state of the art, *Fuzzy Sets and Systems* **242**, 2014, pp. 1–30.
- [17] D. Dubois and H. Prade, Rough fuzzy sets and fuzzy rough sets, *International Journal of General Systems* **17**, 1990, pp. 91–209.
- [18] L. Fariñas del Cerro, H. Prade, Rough sets, twofold fuzzy sets and modal logic—Fuzziness in indiscernibility and partial information, In: A. Di Nola, A.G.S. Ventre, Eds., *The Mathematics of Fuzzy Systems*, Verlag TUV Rheinland, Köln, 1986, pp. 103–120.
- [19] A. Fernández, V. Lopez, M.J. del Jesus, F. Herrera, Revisiting evolutionary fuzzy systems: taxonomy, applications, new trends and challenges, *Knowledge-Based Systems* **80**, 2015, pp. 109–121.
- [20] I. Glöckner, *Fuzzy quantifiers: a computational theory*, *Studies in Fuzziness and Soft Computing* 193, Springer, 2008.
- [21] M. Grabisch, C. Labreuche, A decade of application of the Choquet and Sugeno integrals in multi-criteria decision aid, *Ann. Oper. Res.* **175** (1), 2010, pp. 247–286.
- [22] S. Greco, B. Matarazzo, and R. Słowiński, Rough sets theory for multicriteria decision analysis, *European journal of operational research* **129**(1), 2001, pp. 1–47.
- [23] S. Greco, M. Inuiguchi, R. Slowinski, Fuzzy rough sets and multiple-premise gradual decision rules, *International Journal of Approximate Reasoning* **41**, 2005, 179–211.
- [24] J.W. Grzymala-Busse, LERS—a system for learning from examples based on rough sets, in: *Intelligent decision support*, Springer, 1992, pp. 3–18.
- [25] J.W. Grzymala-Busse, J. Stefanowski, Three discretization methods for rule induction, *International Journal of Intelligent Systems* **16**(1), 2001, pp. 29–38.
- [26] Q. Hu, S. An, X. Yu, D. Yu, Robust fuzzy rough classifiers, *Fuzzy Sets and Systems* **183**(1), 2011, pp. 26–43.
- [27] J. Hühn, E. Hüllermeier, FURIA: an algorithm for unordered fuzzy rule induction, *Data Mining and Knowledge Discovery* **19**(3), 2009, pp. 293–319.
- [28] M. Inuiguchi, W.Z. Wu, C. Cornelis, N. Verbiest, Fuzzy-rough hybridization, in: *Springer Handbook of Computational Intelligence*, 2015, pp. 425–451.
- [29] R. Jensen, C. Cornelis, Q. Shen, Hybrid fuzzy-rough rule induction and feature selection, in: Proc. 2009 IEEE International Conference on Fuzzy Systems, 2009, pp. 1151–1156.
- [30] R. Jensen, C. Cornelis, Fuzzy-rough instance selection, in: Proc. 19th International Conference on Fuzzy Systems (FUZZ-IEEE 2010), 2010, pp. 1776–1782.
- [31] R. Jensen and C. Cornelis, Fuzzy-rough nearest neighbour classification, *Transactions on rough sets*, vol. XIII, 2011, pp. 56–72.
- [32] E.P. Klement, R. Mesiar, E. Pap, *Triangular norms*, Springer, 2000.
- [33] O.U. Lenz, D. Peralta, C. Cornelis, Scalable approximate FRNN-OWA classification, *IEEE Transactions on Fuzzy Systems* **28**(5), 2020, pp. 929–938.
- [34] O.U. Lenz, D. Peralta, C. Cornelis, fuzzy-rough-learn 0.1: A Python library for machine learning with fuzzy rough sets, in: Proc. International Joint Conference on Rough Sets, 2020, pp. 491–499.
- [35] M.J. Lesot, G. Moysse, B. Bouchon-Meunier, Interpretability of fuzzy linguistic summaries, *Fuzzy Sets and Systems* **292**, 2016, pp. 307–317.
- [36] Y. Lin, Y. Li, C. Wang, J. Chen, Attribute reduction for multi-label learning with fuzzy rough set, *Knowledge-based systems* **52**, 2018, pp. 51–61.
- [37] C. Molnar, *Interpretable machine learning*, Lulu.com, 2020.
- [38] G. Nápoles, C. Mosquera, R. Falcon, I. Grau, R. Bello, K. Vanhoof, Fuzzy-Rough Cognitive Networks, *Neural Networks* **97**, 2018, pp. 19–27.
- [39] A. Naumoski, G. Mirceva, K. Mitreski, Novel t-norm for fuzzy-rough rule induction algorithm and its influence, in: *ICT Innovations 2021. Digital Transformation, Communications in Computer and Information Science*, vol. 1521. Springer, 2021, pp. 115–125.
- [40] P. Ni, S. Zhao, X. Wang, H. Chen, C. Li, E.C.C. Tsang, Incremental feature selection based on fuzzy rough sets, *Information Sciences* **536**, 2020, pp. 185–204.
- [41] M. Palangetić, C. Cornelis, S. Greco, R. Słowiński, Fuzzy extensions of the dominance-based rough set approach, *International Journal of Approximate Reasoning* **129**, 2021, pp. 1–19.
- [42] M. Palangetić, C. Cornelis, S. Greco, R. Słowiński, Granular representation of OWA-based fuzzy rough sets, *Fuzzy Sets and Systems*, in press.
- [43] M. Palangetić, C. Cornelis, S. Greco, R. Słowiński, A novel machine learning approach to data inconsistency with respect to a fuzzy relation, arXiv:2111.13447 [cs.AI].
- [44] Z. Pawlak, Rough sets, *International Journal of Computer and Information Sciences* **11**(5), 1982, pp. 341–356.
- [45] Z. Pawlak, A. Skowron, Rudiments of rough sets, *Information Sciences* **177**(1), 2007, pp. 3–27.
- [46] Z. Pawlak, A. Skowron, Rough sets: some extensions. *Information Sciences* **177**(1), 2007, pp. 28–40.
- [47] Z. Pawlak, A. Skowron, Rough sets and boolean reasoning, *Information Sciences* **177**(1), 2007, pp. 41–73.
- [48] E. Ramentol, S. Vluymans, N. Verbiest, Y. Caballero, R. Bello, C. Cornelis, F. Herrera, IFROWANN: imbalanced fuzzy-rough ordered weighted average nearest neighbor classification, *IEEE Transactions on Fuzzy Systems* **23**(5), 2015, pp. 1622–1637.
- [49] A.M. Radzikowska, E.E. Kerre, A comparative study of fuzzy rough sets, *Fuzzy Sets and Systems* **126**, 2002, pp. 137–156.
- [50] A. Theerens, O.U. Lenz, C. Cornelis, Choquet-based fuzzy rough sets, *International Journal of Approximate Reasoning*, in press.
- [51] N. Verbiest, C. Cornelis, F. Herrera, FRPS: a fuzzy rough prototype selection method, *Pattern Recognition* **46**(10), 2013, pp. 2770–2782.
- [52] N. Verbiest, E. Ramentol, C. Cornelis, F. Herrera, Preprocessing Noisy Imbalanced Datasets using SMOTE enhanced with Fuzzy Rough Prototype Selection **22**, 2014, pp. 511–517.
- [53] S. Vluymans, L. D'eer, Y. Saeys, C. Cornelis, Applications of fuzzy rough set theory in machine learning: a survey, *Fundamenta Informaticae* **142**(1–4), 2015, pp. 53–86.
- [54] S. Vluymans, D. Sánchez Tarragó, Y. Saeys, C. Cornelis, F. Herrera, Fuzzy rough classifiers for class imbalanced multi-instance data, *Pattern Recognition* **53**, 2016, pp. 36–45.
- [55] S. Vluymans, A. Fernández, Y. Saeys, C. Cornelis, F. Herrera, Dynamic affinity-based classification of multi-class imbalanced data with one-vs-one decomposition: a fuzzy rough approach, *Knowledge and Information Systems* **6**(1), 2018, pp. 55–84.
- [56] S. Vluymans, C. Cornelis, F. Herrera, Y. Saeys, Multi-label classification using a fuzzy rough neighborhood consensus, *Information Sciences* **433–434**, 2018, pp. 96–114.
- [57] S. Vluymans, N. Mac Parthalaín, C. Cornelis, Y. Saeys, Weight selection strategies for ordered weighted average based fuzzy rough sets, *Information Sciences* **501**, 2019, pp. 155–171.

- [58] R.R. Yager, On ordered weighted averaging aggregation operators in multicriteria decision making, *IEEE Transactions on systems, Man, and Cybernetics* **18**(1), 1988, pp. 183–190.
- [59] R.R. Yager, Quantifier guided aggregation using OWA operators, *International Journal of Intelligent Systems* **11**(1), 1996, pp. 49–73.
- [60] J.T. Yao, A.V. Vasilakos, W. Pedrycz, Granular computing: perspectives and challenges, *IEEE Transactions on Cybernetics* **43**(6), 2013, pp.1977–1989.
- [61] Y. Yao, *Granular computing using neighborhood systems*, Advances in soft computing, Springer, 1999, pp. 539–553.
- [62] Y. Yao, Three-way decisions with probabilistic rough sets, *Information Sciences* **180**(3), 2010, pp. 341–353.
- [63] L.A. Zadeh, Fuzzy sets, *Information and Control* **8**, 1965, pp. 338–353.
- [64] L.A. Zadeh, A computational approach to fuzzy quantifiers in natural languages, in: *Computational linguistics*, Elsevier, 1983, pp. 149–184.
- [65] L.A. Zadeh, Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, *Fuzzy sets and systems* **90**(2), 1997, pp. 111–127.
- [66] S. Zhao, E.C.C. Tsang, D. Chen, X. Wang, Building a rule-based classifier—a fuzzy-rough set approach, *IEEE Transactions on Knowledge and Data Engineering* **22**(5), 2009, pp. 624–638.
- [67] S. Zhao, Z. Dai, X. Wang, P. Ni, H. Chen, C. Li, An accelerator for rule induction in fuzzy rough theory, *IEEE Transactions on Fuzzy Systems* **29**(12), 2021, pp. 3635–3649.
- [68] W. Ziarko, Variable precision rough set model, *Journal of computer and system sciences* **46**(1), 1993, pp. 39–59.