Available online at www.sciencedirect.com
ScienceDirect
Fuzzy Sets and Systems 440 (2022) 112–130
FUZZY sets and systems
www.elsevier.com/locate/fss

Granular representation of OWA-based fuzzy rough sets

Marko Palangetić [a,*], Chris Cornelis [a], Salvatore Greco [b,c], Roman Słowiński [d,e]

[a] *Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Ghent, Belgium*
[b] *Department of Economics and Business, University of Catania, Catania, Italy*
[c] *Portsmouth Business School, Centre of Operations Research and Logistics (CORL), University of Portsmouth, Portsmouth, United Kingdom*
[d] *Institute of Computing Science, Poznań University of Technology, Poznań, Poland*
[e] *Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland*

Received 6 April 2020; received in revised form 28 February 2021; accepted 20 April 2021
Available online 26 April 2021

Abstract

Granular representations of crisp and fuzzy sets play an important role in rule induction algorithms based on rough set theory. In particular, arbitrary fuzzy sets can be approximated using unions of simple fuzzy sets called granules. These granules, in turn, have a straightforward interpretation in terms of human-readable fuzzy "*if..., then...*" rules. In this paper, we are considering a fuzzy rough set model based on ordered weighted average (OWA) aggregation over considered values. We show that this robust extension of the classical fuzzy rough set model, which has been applied successfully in various machine learning tasks, also allows for a granular representation. In particular, we prove that when approximations are defined using a directionally convex t-norm and its residual implicator, the OWA-based lower and upper approximations are definable as unions of fuzzy granules. This result has practical implications for rule induction from such fuzzy rough approximations.

Keywords: Fuzzy rough sets; Granular computing; Ordered weighted average; Rule induction

1. Introduction

Rough sets were introduced by Pawlak [1] to deal with inconsistencies within information tables where objects are described by a set of attributes. Pawlak's approach produces two sets, called lower and upper approximation, which represent elements being, respectively, necessarily consistent (lower approximation), and possibly consistent (upper approximation) with knowledge contained in the information table. The original theory was designed to deal with nominal attributes, and relies on an equivalence relation, expressing indiscernibility between elements. Greco et al. [2] extended the original theory with their Dominance-based Rough Set Approach (DRSA) allowing attributes to have ordinal value sets, and replacing the indiscernibility relation with a dominance relation. To distinguish between both approaches, Pawlak's original theory is also called the Indiscernibility-based Rough Set Approach (IRSA).

* Corresponding author.
E-mail addresses: marko.palangetic@ugent.be (M. Palangetić), chris.cornelis@ugent.be (C. Cornelis), salgreco@unict.it (S. Greco), roman.slowinski@cs.put.poznan.pl (R. Słowiński).

https://doi.org/10.1016/j.fss.2021.04.018
0165-0114/

On the other hand, fuzzy set theory [3] studies the gradual truth of logical statements, and is used extensively in modeling imprecise and vague information. The combination of fuzzy sets and IRSA was first proposed by Dubois and Prade [4], allowing to approximate fuzzy sets using a fuzzy indiscernibility relation. A similar extension of DRSA to fuzzy set theory was proposed by Greco et al. [5].

It is well-known that the classical definitions of fuzzy rough sets in both the indiscernibility and dominance case are vulnerable to noise, in a similar way as their crisp counterparts: small fluctuations in data may cause huge changes in membership values of the approximations. For this reason, various robust versions of the fuzzy rough approximations were proposed [6–10]. In this paper, we focus on the Ordered Weighted Average (OWA) approach, which was shown to the be the most noise tolerant among a variety of robust fuzzy rough set models in [11].

Both IRSA and DRSA are very useful from the point of view of granular computing, as they possess a so-called granular representation; indeed, lower and upper approximations can be represented as unions of simple sets or granules, induced from the data [12]. In contrast to crisp sets, the granular properties of fuzzy rough sets do not stem directly from the proposed definitions. Degang et al. [13] were the first to show that fuzzy IRSA has indeed a granular representation which means that fuzzy rough approximations can be represented as a union of simple fuzzy sets or fuzzy granules. Later, Yao et al. [14] pointed out that the symmetry of the fuzzy relation is not essential for the granular representation, and hence it can be extended to fuzzy DRSA as well.

The granular representation of rough sets and fuzzy 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 description of objects by subsets of attributes with particular decision classes. Basic granules, from which rough sets and fuzzy 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 the LEM2 algorithm [15] for IRSA, and the DomLEM algorithm [16] for DRSA. Similarly, the granularity of fuzzy rough sets has also been used for rule induction. In this case, we obtain a fuzzy inference system, with flexible fuzzy rules instead of strict crisp rules [17,18]. The main advantage of fuzzy rules is that they can model complex shapes of data, and still keep an intuitive interpretation of those shapes.

In this paper, we distinguish concepts of rough approximations and granular approximations. While rough approximations are constructed based on the assumption that two elements should relate identically (or similarly) on condition and decision attributes in data, granular approximations are constructed based on the representability of sets by means of granules. We prove that these two concepts are equivalent in IRSA and DRSA, as well as in fuzzy IRSA and fuzzy DRSA, under certain conditions on the fuzzy connectives. Moreover, as our main contribution, we show that under certain conditions on the logical connectives, OWA-based fuzzy rough approximations also possess such a granular representation. As a consequence, this robust extension of fuzzy rough sets naturally induces a set of associated fuzzy rules.

The remainder of this paper is structured as follows. In Section 2, we recall some required preliminaries about fuzzy sets and rough sets, and unify the definitions of IRSA and DRSA for practical purposes into the Preorder-based Rough Set Approach (PRSA). Section 3 introduces the notion of a granularly representable set, and investigates its relationship with definable sets and rough approximations. In this way, we provide a new view on granularity of sets in general, and on the relationship between granularity and rough approximations. In Section 4, we define granularly representable fuzzy sets and provide analogous propositions as for the crisp case. Section 5 deals with the granularity of OWA-based approximations, while Section 6 goes deeper into the topic of characterizing convex t-norms which are crucial for the representation. Section 7 contains our conclusion and outlines a future work.

2. Preliminaries

2.1. Fuzzy logic connectives and fuzzy relations

In this subsection, the definitions and terminology are based on [19].

Recall that a *t-norm* $T : [0, 1]^2 \to [0, 1]$ is a binary operator which is commutative, associative, non-decreasing in both parameters, and for which holds that $\forall x \in [0, 1]$, $T(x, 1) = x$. Since a t-norm is associative, we may extend it unambiguously to a $[0, 1]^n \to [0, 1]$ mapping for any $n > 2$.

We say that a t-norm has an idempotent element $x \in [0, 1]$ if $T(x, x) = x$. 0 and 1 are called trivial idempotent elements. Also, we call a t-norm Archimedean if

Table 1
Some common t-norms.

Name	Definition		
Minimum	$T_M(x, y)$	=	$\min(x, y)$
Product	$T_P(x, y)$	=	xy
Łukasiewicz	$T_L(x, y)$	=	$\max(0, x + y - 1)$
Drastic	$T_D(x, y)$	=	$\begin{cases} \min(x, y) & \text{if } \max(x, y) = 1 \\ 0 & \text{otherwise} \end{cases}$
Nilpotent minimum	$T_{nM}(x, y)$	=	$\begin{cases} \min(x, y) & \text{if } x + y > 1 \\ 0 & \text{otherwise} \end{cases}$

Table 2
Residual implicators of the t-norms from Table 1.

Name	Definition		
Gödel	$I_{T_M}(x, y)$	=	$\begin{cases} 1 & \text{if } x \le y \\ y & \text{otherwise} \end{cases}$
Goguen	$I_{T_P}(x, y)$	=	$\begin{cases} 1 & \text{if } x \le y \\ \frac{y}{x} & \text{otherwise} \end{cases}$
Łukasiewicz	$I_{T_L}(x, y)$	=	$\min(1, 1 - x + y)$
Drastic	$I_{T_D}(x, y)$	=	$\begin{cases} y & \text{if } x = 1 \\ 1 & \text{otherwise} \end{cases}$
Nilpotent minimum	$I_{T_{nM}}(x, y)$	=	$\begin{cases} 1 & \text{if } x \le y \\ \max(1 - x, y) & \text{otherwise} \end{cases}$

$$(\forall (x, y) \in (0, 1)^2)(\exists n \ge 2)\left(T(\underbrace{x, \ldots, x}_{n \text{ times}}) < y \right).$$

Some commonly used t-norms are listed in Table 1. T_P, T_L and T_D are Archimedean, while T_M and T_{nM} are not. It is well-known [19] that a continuous t-norm is Archimedean if and only if it has only trivial idempotent elements.

Related to the notion of t-norm is that of a *copula*. A (bivariate) copula C is a $[0, 1]^2 \to [0, 1]$ mapping which satisfies the boundary conditions $\forall x,\ C(0, x) = C(x, 0) = 0,\ C(1, x) = C(x, 1) = x$, and the 2-increasingness property: $C(x, y) + C(x', y') \ge C(x', y) + C(x, y')$ for all $x \ge x'$ and $y \ge y'$.

Some t-norms are copulas, while others are not: for example, T_M, T_P and T_L are copulas, while T_D and T_{nM} are not. Vice versa, there also exist copulas which are not t-norms.

An *implicator* (or *fuzzy implication*) $I : [0, 1]^2 \to [0, 1]$ is a binary operator which is non-increasing in the first component, non-decreasing in the second one and for which holds that $I(1, 0) = 0$, $I(0, 0) = I(0, 1) = I(1, 1) = 1$.

The residuation property holds for a t-norm T and an implicator I if $T(x, y) \le z \Leftrightarrow x \le I(y, z)$. It is well-known that the residuation property holds if and only if T is left-continuous and I is defined as the residual implicator of T, that is

$$I(x, y) = \sup\{\lambda \in [0, 1] : T(x, \lambda) \le y\}.$$

The proposition below recalls some useful properties that hold in this case.

Proposition 2.1. *Let T be a left-continuous t-norm, I its residual implicator, $x, y, z \in [0, 1]$. It holds that:*

a. $T(x, I(x, y)) \le y$,
b. $I(T(x, y), z) = I(x, I(y, z))$.

Table 2 contains the residual implicators of the t-norms from Table 1. Note that all of them, except for I_{T_D}, satisfy the residuation property.

Given a non-empty set U, a fuzzy relation $\widetilde{R}$ on U is a mapping $\widetilde{R} : U \times U \to [0, 1]$ which indicates how much two elements from U are related. Some relevant properties of fuzzy relations include the following:

- $\widetilde{R}$ is reflexive if $\forall u \in U,\ \widetilde{R}(u,u) = 1$.
- $\widetilde{R}$ is symmetric if $\forall u, v \in U,\ \widetilde{R}(u,v) = \widetilde{R}(v,u)$.
- $\widetilde{R}$ is T-transitive w.r.t. t-norm T if $\forall u, v, w \in U$ it holds that $T(\widetilde{R}(u,v), \widetilde{R}(v,w)) \le \widetilde{R}(u,w)$.

A reflexive and T-transitive fuzzy relation is called a fuzzy T-preorder or fuzzy T-dominance relation. If it is also symmetric, it is called a fuzzy T-equivalence relation.

2.2. Rough sets

We first recall Pawlak's definition of IRSA [1]. Let U be the set of objects and E an equivalence relation on U, which is also called indiscernibility relation. With $[u]_E$ we denote an equivalence class of E containing element u. Lower and upper approximations of set A are defined as:

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

In DRSA [2], approximations are based on a dominance relation D which is a preorder, i.e., a reflexive and transitive binary relation on U. The sets which are approximated in DRSA are so-called upward and downward unions of objects. Without going into the detail of their specific construction (for this, we refer, e.g., to [2]), let $A \subseteq U$ be some upward union. Its complement $A^c = U \setminus A$ will then be a downward union by construction. By $D^+(u)$, we denote the set of elements v for which it holds that $(v,u) \in D$, while by $D^-(u)$ we denote the set of elements v for which $(u,v) \in D$. The DRSA approximations of A and A^c are then given by:

$$\underline{\text{apr}}_D(A) = \{u \in U : D^+(u) \subseteq A\}, \quad \overline{\text{apr}}_D(A) = \{u \in U : D^-(u) \cap A \neq \emptyset\},$$
$$\underline{\text{apr}}_D(A^c) = \{u \in U : D^-(u) \subseteq U\}, \quad \overline{\text{apr}}_D(A^c) = \{u \in U : D^+(u) \cap A \neq \emptyset\}.$$

As we can see, if D is a symmetric relation then $D^+(u) = D^-(u)$ and the approximations are reduced to the IRSA definition. So, we may conclude that DRSA is a generalization of IRSA. As mentioned, DRSA is only applied to upward or downward unions, and this specification is purely motivated by the practical applications of DRSA. As it does not affect any theoretical property of the DRSA approximations, for further use we will introduce the Preorder-based Rough Set Approach (PRSA) in which DRSA is applied to a general set instead of an upward or downward union.

The question might be raised whether PRSA should use the approximations of A or those of A^c from the DRSA definitions. However, we may see that they are in fact equivalent: the approximations of A^c may be obtained from those of A by replacing relation D with its inverse relation D^{-1}. Therefore, let R be a preorder relation and let $R^+(u) = \{v \in U, (v,u) \in R\}$ and $R^-(u) = \{v \in U, uRv\}$. The lower and upper PRSA approximations of set $A \subseteq U$ are defined as:

$$\underline{\text{apr}}_R(A) = \{u \in U : R^+(u) \subseteq A\}, \quad \overline{\text{apr}}_R(A) = \{u \in U : R^-(u) \cap A \neq \emptyset\}.$$

We list some important properties of PRSA which will be used later. All the proofs may be found in [20] and in its references.

- (**inclusion**): $\underline{\text{apr}}_R(A) \subseteq A \subseteq \overline{\text{apr}}_R(A)$,
- (**exact approximation**): $\underline{\text{apr}}_R(A) = A \Leftrightarrow A = \overline{\text{apr}}_R(A)$,
- (**idempotence**) $\underline{\text{apr}}_R(\underline{\text{apr}}_R(A)) = \underline{\text{apr}}_R(A)$, $\overline{\text{apr}}_R(\overline{\text{apr}}_R(A)) = \overline{\text{apr}}_R(A)$.

2.3. Fuzzy rough sets

There exist multiple definitions of fuzzy rough sets. The initial one given by Dubois and Prade [4] evolved into many versions, and an overview was presented in [11]. The main differences among them refer to the following two aspects:

- the fuzzy relation used,

- type of fuzzy set connectives.

Depending on the selection of relations and connectives, different fuzzy rough approximations exhibit different properties. Therefore, we first need to determine which fuzzy relation and which type of fuzzy connectives will be suitable for our task. For the question of which fuzzy relation to use, we follow the intuition of the relations used in the crisp case (equivalence for IRSA, and dominance for DRSA). As mentioned, the granules obtained from the IRSA and DRSA approximations have nice interpretations in the form of the human-readable "*if..., then...*" rules. We want to transfer that interpretability property to the fuzzy case too. So, we will focus only on fuzzy IRSA and fuzzy DRSA. On the other hand, the connectives which are used in the definitions are a t-norm T and an implicator I. As mentioned in [11], all relevant properties of rough sets can be maintained when T is a left-continuous t-norm and I is its residual implicator, so we will keep that assumption throughout this paper.

Given a fuzzy T-equivalence relation $\widetilde{E}$ on U, the lower and upper approximations of a fuzzy set A in U in fuzzy IRSA are defined as:

$$\underline{\text{apr}}_{\widetilde{E}}^{\inf,I}(A)(u) = \inf(I(\widetilde{E}(u,v), A(v)); v \in U),$$
$$\overline{\text{apr}}_{\widetilde{E}}^{\sup,T}(A)(u) = \sup(T(\widetilde{E}(u,v), A(v)); v \in U).$$

In fuzzy DRSA, we are approximating fuzzy upward and downward unions. Definitions and examples how to construct such sets in practice appear in [20], but are not important for the current exposition. Let $A \in \mathcal{F}(U)$ be a fuzzy upward union, then the construction ensures that the fuzzy set complement coA is a fuzzy downward union. Let $\widetilde{D}$ be a fuzzy T-dominance relation. The approximations of A and coA are given as:

$$\underline{\text{apr}}_{\widetilde{D}}^{\inf,I}(A)(u) = \inf(I(\widetilde{D}(v,u), A(v)); v \in U),$$
$$\overline{\text{apr}}_{\widetilde{D}}^{\sup,T}(A)(u) = \sup(T(\widetilde{D}(u,v), A(v)); v \in U),$$
$$\underline{\text{apr}}_{\widetilde{D}}^{\inf,I}(coA)(u) = \inf(I(\widetilde{D}(u,v), (coA)(v)); v \in U),$$
$$\overline{\text{apr}}_{\widetilde{D}}^{\sup,T}(coA)(u) = \sup(T(\widetilde{D}(v,u), (coA)(v)); v \in U).$$

Adding symmetry to $\widetilde{D}$ reduces both fuzzy DRSA definitions to the fuzzy IRSA definition, in the same way as in the crisp case. Hence, fuzzy DRSA is a generalization of fuzzy IRSA provided it is applied to a general fuzzy set instead of an upward or downward union. We will refer to this general case as fuzzy PRSA. As before, the definitions using A and coA are equivalent: indeed, the second definition may be obtained from the first one using the inverse fuzzy relation $\widetilde{D}^{-1}$, defined as $\widetilde{D}^{-1}(u,v) = \widetilde{D}(v,u)$. Hence, let $\widetilde{R}$ be a fuzzy T-preorder. The lower and upper PRSA approximations of a fuzzy set $A \in \mathcal{F}(U)$ are defined as:

$$\underline{\text{apr}}_{\widetilde{R}}^{\inf,I}(A)(u) = \inf(I(\widetilde{R}(v,u), A(v)); v \in U),$$
$$\overline{\text{apr}}_{\widetilde{R}}^{\sup,T}(A)(u) = \sup(T(\widetilde{R}(u,v), A(v)); v \in U).$$

As before, we list some important properties of the fuzzy PRSA, some of which will be needed later. All proofs may be found in [11,20] and in its references.

- (**inclusion**) $\forall u \in U$: $\underline{\text{apr}}_{\widetilde{R}}^{\inf,I}(A) \subseteq A \subseteq \overline{\text{apr}}_{\widetilde{R}}^{\sup,T}(A)$.
- (**exact approximation**) $\underline{\text{apr}}_{\widetilde{R}}^{\inf,I}(A) = A \Leftrightarrow \overline{\text{apr}}_{\widetilde{R}}^{\sup,T}(A) = A$.
- (**idempotence**) It holds that $\underline{\text{apr}}_{\widetilde{R}}^{\inf,I}(\underline{\text{apr}}_{\widetilde{R}}^{\inf,I}(A)) = \underline{\text{apr}}_{\widetilde{R}}^{\inf,I}(A)$, $\overline{\text{apr}}_{\widetilde{R}}^{\sup,T}(\overline{\text{apr}}_{\widetilde{R}}^{\sup,T}(A)) = \overline{\text{apr}}_{\widetilde{R}}^{\sup,T}(A)$.

2.4. OWA-based fuzzy rough sets

In the definitions of fuzzy PRSA, we used operators inf and sup to aggregate different values. Such an approach suffers from a lack of robustness since inf and sup values may be achieved in some outlying points, which may significantly affect the approximations. Because of that, a softer version of inf and sup operators was introduced in terms of Ordered Weighted Average (OWA) operators [21]. We recall the definition of OWA.

Definition 2.1. [22] The OWA aggregation of set V of n real numbers with weight vector $W = (w_1, w_2, ..., w_n)$, where $w_i \in [0, 1]$ and $\Sigma_{i=1}^{n} w_i = 1$, is given by

$$\mathrm{OWA}_W(V) = \sum_{i=1}^{n} w_i v_{(i)},$$

where $v_{(i)}$ is the i-th largest element in the set V.

Different weights are used for softer versions of inf and sup. When sup is replaced, the larger values are among the initial elements of vector W while for the replacement of inf, the larger values are at the end of the same vector. By W_L we denote a weight vector replacing inf, while W_U refers to a weight vector which replaces sup. Following that, we have the definition of OWA-based fuzzy PRSA.

$$\underline{\mathrm{apr}}_{\widetilde{R}}^{W_L,I}(A)(u) = \mathrm{OWA}_{W_L}(\{I(\widetilde{R}(v,u), A(v)); v \in U\}),$$
$$\overline{\mathrm{apr}}_{\widetilde{R}}^{W_U,T}(A)(u) = \mathrm{OWA}_{W_U}(\{T(\widetilde{R}(u,v), A(v)); v \in U\}).$$

Properties of such defined OWA-based fuzzy rough approximations may be found in [11,20]. Our goal in this paper will be to prove that these approximations may be represented as a fuzzy union of simple fuzzy sets, under the condition of convexity for the involved fuzzy logic connectives.

3. Granular view of PRSA

Granular properties of IRSA have been discussed in [23], while a similar analysis was carried out for DRSA in [24]. More recently, the granular representation of DRSA was also studied from the perspective of covering-based rough sets in [25]; in particular, the notion of a definable set was introduced as a union of elementary sets or granules: equivalence classes $[u]_E$ in the case of IRSA, and sets $D^-(u)$ and $D^+(u)$ in the case of DRSA.

In this section, we investigate the granular representation of PRSA from a new viewpoint of granular approximations. We introduce the notion of a granularly representable set: a set which can be disintegrated into building blocks that are interpreted as human readable rules. Let U be the set of objects and let $A \subseteq U$. Let R be a preorder relation on U and $R^+(u) = \{v \in U; (v,u) \in R\}$. We say that set A is granularly representable w.r.t. relation R if

$$A = \bigcup_{u \in A} R^+(u).$$

The blocks $R^+(u)$ may be interpreted as indiscernibility rules in case of IRSA, or monotonic rules in the case of DRSA. Optimality of the rules in the sense of a minimal number of blocks covering A is not guaranteed, and while there exist ways to reduce the number of building blocks of A, this falls outside the scope of the present paper. Here, we focus on the link between granular representability and rough approximations.

Proposition 3.1. *Set A is granularly representable if and only if $\underline{apr}_R(A) = A = \overline{apr}_R(A)$.*

Proof. For the right side of the equivalence it is enough to prove or assume that $\underline{\mathrm{apr}}_R(A) = A$ since it holds that $\underline{\mathrm{apr}}_R(A) = A \Leftrightarrow A = \overline{\mathrm{apr}}_R(A)$ due to the exact approximation property.

($\Rightarrow$) Assume that A is granulary representable. For $u \in A$ we have that also $R^+(u) \subseteq A$ which leads to $u \in \underline{\mathrm{apr}}_R(A)$. Hence $A \subseteq \underline{\mathrm{apr}}_R(A)$. Combining that with the inclusion property, we have that $\underline{\mathrm{apr}}_R(A) = A$.

($\Leftarrow$) Assume that $\underline{\mathrm{apr}}_R(A) = A$. We have that

$$u \in A \Rightarrow u \in \underline{\mathrm{apr}}_R(A) \Rightarrow R^+(u) \subseteq A.$$

So we have that $\bigcup_{u \in A} R^+(u) \subseteq A$. On the other hand, from the reflexivity of R it holds that $A \subseteq \bigcup_{u \in A} R^+(u)$. Therefore, A is granularly representable. □

Corollary 3.1. *$\underline{apr}_R(A)$ and $\overline{apr}_R(A)$ are granularly representable sets.*

Proof. This follows from the idempotence property of lower and upper approximation. □

Corollary 3.2. *We may write the rough approximations in the granular form:*

$$\underline{apr}_R(A) = \bigcup\{R^+(u) : u \in U, R^+(u) \subseteq A\}, \quad \overline{apr}_R(A) = \bigcup\{R^+(u) : u \in A\}.$$

Proof. We have that:

$$\underline{\text{apr}}_R(A) = \bigcup\{R^+(u), u \in \underline{\text{apr}}_R(A)\} = \bigcup\{R^+(u) : u \in U, R^+(u) \subseteq A\},$$

since $u \in \underline{\text{apr}}_R(A) \Leftrightarrow u \in U \wedge R^+(u) \subseteq A$. For the upper approximation, from the definition we have that $\overline{\text{apr}}_R(A) = \bigcup\{R^+(u), u \in \overline{\text{apr}}_R(A)\}$. From the inclusion property we know that $A \subseteq \overline{\text{apr}}_R(A)$ so we may conclude that $\bigcup\{R^+(u), u \in A\} \subseteq \bigcup\{R^+(u), u \in \overline{\text{apr}}_R(A)\}$. For the opposite direction we have the following:

$$\begin{aligned} v \in \bigcup\{R^+(u), u \in \overline{\text{apr}}_R(A)\} &\Leftrightarrow \exists u \in \overline{\text{apr}}_R(A), v \in R^+(u) \\ &\Leftrightarrow \exists u \in U, R^-(u) \cap A \neq \emptyset, v \in R^+(u) \\ &\Leftrightarrow \exists w \in A, w \in R^-(u) \wedge v \in R^+(u) \\ &\Leftrightarrow \exists w \in A, u \in R^+(w) \wedge v \in R^+(u) \\ &\Rightarrow \exists w \in A, v \in R^+(w) \\ &\Leftrightarrow v \in \bigcup\{R^+(u), u \in A\}. \end{aligned}$$

where for the implication we use the transitivity of R. So, we conclude that also $\bigcup\{R^+(u), u \in \overline{\text{apr}}_R(A)\} \subseteq \bigcup\{R^+(u), u \in A\}$ which gives us the desired result. □

Corollary 3.3.

$$R^+(u) \subseteq A \Leftrightarrow R^+(u) \subseteq \underline{apr}_R(A).$$

Proof. The (⇐) part is obvious because of the inclusion property. (⇒) is a consequence of the definition of the granular representation and Corollary 3.2. □

Proposition 3.2. *Let $A \subseteq U$ and R a preorder on U. The largest granularly representable set contained in A is $\underline{apr}_R(A)$, while the smallest granularly representable set containing A is $\overline{apr}_R(A)$.*

Proof. Let B be some granularly representable set containing A. We have that

$$\overline{\text{apr}}_R(A) = \bigcup\{R^+(u) : u \in A\} \subseteq \bigcup\{R^+(u) : u \in B\} = B.$$

Since $\overline{\text{apr}}_R(A)$ is contained in every granularly representable set containing A, $\overline{\text{apr}}_R(A)$ is the smallest such set since it also contains A by the inclusion property. On the other hand, let C be a granularly representable set contained in A. We have that:

$$u \in C \Rightarrow R^+(u) \subseteq C \Rightarrow R^+(u) \subseteq A \Rightarrow u \in \underline{\text{apr}}_R(A).$$

So we conclude that $C \subseteq \underline{\text{apr}}_R(A)$. Since $\underline{\text{apr}}_R(A)$ contains every granularly representable set contained in A, $\underline{\text{apr}}_R(A)$ is the largest such set since it is also contained in A by the inclusion property. □

In conclusion, we saw that for every set A and preorder R there is a granular enclosing in the form of rough approximations. They represent families of building blocks which are necessarily (lower approximations) or possibly (upper approximation) contained in A. This may be further translated into possible and certain rules induced from $\overline{\text{apr}}_R(A)$ and $\underline{\text{apr}}_R(A)$, respectively, as done using LEM2 (for IRSA) or DomLEM (for DRSA) algorithms.

4. Granular representation of fuzzy PRSA

In this section, we extend the granular representation from the previous section to fuzzy sets and relate it to the fuzzy rough set definitions. Let $\widetilde{R}$ be a fuzzy T-preorder relation. We replace $R^+(u)$ from above with fuzzy set $\widetilde{R}^+(u)$ where the membership degree of $v \in U$ is given by $\widetilde{R}(u, v)$. Granular properties of fuzzy rough approximations were first introduced in [13], where the authors defined a parameterized family of fuzzy granules:

$$\widetilde{R}^+_\lambda(u) = \{(v, T(\widetilde{R}(v, u), \lambda)); v \in U\}, \tag{1}$$

where λ is a real parameter from $[0, 1]$. In the original work, $\widetilde{R}$ was also symmetric, but later on it was noticed that symmetry does not contribute to the granular properties of fuzzy rough approximations. We observe that Eq. (1) is not the only possible way to define fuzzy granules. An alternative was proposed in [26], using implicators and coimplicators. However, in order to extend the granular representation introduced in the previous section, we will focus on the original formula (1).

The idea that a set A is granularly representable if it is the union of building blocks $R^+(u)$ with $u \in A$ can be fuzzified by putting $\lambda = A(u)$ in Eq. (1). In particular, we call $A \in \mathcal{F}(U)$ granulary representable if

$$A = \bigcup\{\widetilde{R}^+_{A(u)}(u); u \in U\}.$$

We first prove two simple lemmas necessary for future proofs.

Lemma 4.1. *For $\lambda_1 \leq \lambda_2$ and for $u \in U$ we have that:*

$$\widetilde{R}^+_{\lambda_1}(u) \subseteq \widetilde{R}^+_{\lambda_2}(u).$$

Proof. Obvious from the monotonicity of a t-norm. □

Lemma 4.2.

$$\widetilde{R}^+_\lambda(u) \subseteq A \Leftrightarrow \lambda \leq \underline{apr}^{\inf,T}_{\widetilde{R}}(A)(u).$$

Proof. We use the residuation property:

$$\widetilde{R}^+_\lambda(u) \subseteq A \Leftrightarrow \forall v \in U,\ T(\widetilde{R}(v, u), \lambda) \leq A(v) \Leftrightarrow \forall v \in U, \lambda \leq I(\widetilde{R}(v, u), A(v))$$
$$\Leftrightarrow \lambda \leq \inf_{v \in U} I(\widetilde{R}(v, u), A(v)) \Leftrightarrow \lambda \leq \underline{\mathrm{apr}}^{\inf,T}_{\widetilde{R}}(A)(u). \quad \square$$

Next, we prove the main result about the granular representability of fuzzy sets.

Proposition 4.1. *Fuzzy set A is granularly representable w.r.t. relation $\widetilde{R}$ if and only it is definable, i.e., $\underline{apr}^{\inf,I}_{\widetilde{R}}(A) = A = \overline{apr}^{\sup,T}_{\widetilde{R}}(A)$.*

Proof. As before, for the right side of equivalence it is enough to prove or assume that $\overline{\mathrm{apr}}^{\sup,T}_{\widetilde{R}}(A) = A$ since $\underline{\mathrm{apr}}^{\inf,I}_{\widetilde{R}}(A) = A \Leftrightarrow A = \overline{\mathrm{apr}}^{\sup,T}_{\widetilde{R}}(A)$ due to the exact approximation property.

($\Rightarrow$) Assume that A is granularly representable. For $v \in U$, we have that

$$A(v) = \sup(T(\widetilde{R}(v, u), A(u)); u \in U) = \overline{\mathrm{apr}}^{\sup,T}_{\widetilde{R}}(A)(v).$$

($\Leftarrow$) Assume that $\overline{\mathrm{apr}}^{\sup,T}_{\widetilde{R}}(A) = A$. Then by the same reasoning, we find that A is granularly representable. □

Corollary 4.1. *$\underline{apr}^{\inf,I}_{\widetilde{R}}(A)$ and $\overline{apr}^{\sup,T}_{\widetilde{R}}(A)$ are granularly representable.*

Proof. This follows from the idempotence of lower and upper approximations. □

Corollary 4.2. *We may write the fuzzy rough approximations definitions in the granular form:*

$$\underline{apr}_{\widetilde{R}}^{\inf,I}(A) = \bigcup\{\widetilde{R}_\lambda^+(u);\, \widetilde{R}_\lambda^+(u) \subseteq A\}, \quad \overline{apr}_{\widetilde{R}}^{\sup,T}(A) = \bigcup\{\widetilde{R}_{A(u)}(u)\}.$$

Proof. For the lower approximation we have that:

$$\begin{aligned}\bigcup\{\widetilde{R}_\lambda^+(u);\, \widetilde{R}_\lambda^+(u) \subseteq A\} &= \bigcup\{\widetilde{R}_\lambda^+(u);\, \lambda \le \underline{\text{apr}}_{\widetilde{R}}^{\inf,I}(A)(u)\} \\ &= \bigcup\{\widetilde{R}_{\underline{\text{apr}}_{\widetilde{R}}^{\inf,I}(A)(u)}(u)\} = \underline{\text{apr}}_{\widetilde{R}}^{\inf,I}(A).\end{aligned}$$

The first equality holds because of Lemma 4.2 while the second one because of Proposition 4.1. For the upper approximation, it follows directly from the definitions:

$$\bigcup\{\widetilde{R}_{A(u)}(u)\} = \sup(T(\widetilde{R}(u,v), A(u));\, u \in U) = \overline{\text{apr}}_{\widetilde{R}}^{\sup,T}(A). \quad \square$$

Corollary 4.3.

$$\widetilde{R}_\lambda^+(u) \subseteq A \Leftrightarrow \widetilde{R}_\lambda^+(u) \subseteq \underline{apr}_{\widetilde{R}}^{\inf,I}(A).$$

Proof. The (⇐) part is obvious since the lower approximation is a subset of the approximated set. (⇒) is a consequence of the definition of the granular representation and Corollary 4.2. □

Proposition 4.2. *Let $A \in \mathcal{F}(U)$ and $\widetilde{R}$ a fuzzy T-preorder. The largest fuzzy granularly representable set contained in A is $\underline{apr}_{\widetilde{R}}^{\inf,I}(A)$, while the smallest granularly representable set containing A is $\overline{apr}_{\widetilde{R}}^{\sup,T}(A)$.*

Proof. Assume that there is granularly representable set B containing A. We have that:

$$\begin{aligned}\overline{\text{apr}}_{\widetilde{R}}^{\sup,T}(A)(u) &= \sup(T(\widetilde{R}(u,v), A(v));\, v \in U) \\ &\le \sup(T(\widetilde{R}(u,v), B(v));\, v \in U) = B(u).\end{aligned}$$

Hence $\overline{\text{apr}}_{\widetilde{R}}^{\sup,T}(A) \subseteq B$. Since $\overline{\text{apr}}_{\widetilde{R}}^{\sup,T}(A)$ is contained in every fuzzy granularly representable set containing A, $\overline{\text{apr}}_{\widetilde{R}}^{\sup,T}(A)$ is the smallest such set since it also contains A by the inclusion property.

On the other hand, assume that C is a fuzzy granularly representable set contained in A. We have that

$$\begin{aligned}\underline{\text{apr}}_{\widetilde{R}}^{\inf,I}(A)(u) &= \inf(I(\widetilde{R}(v,u), A(v)), v \in U) \ge \inf(I(\widetilde{R}(v,u), C(v)), v \in U) \\ &= \underline{\text{apr}}_{\widetilde{R}}^{\inf,I}(C)(u) = C(u).\end{aligned}$$

Since $\underline{\text{apr}}_{\widetilde{R}}^{\inf,I}(A)$ contains every fuzzy granularly representable set contained in A, $\underline{\text{apr}}_{\widetilde{R}}^{\inf,I}(A)(u)$ is the largest such set since it is also contained in A by the inclusion property. □

As we saw before, for any fuzzy set A, there is a fuzzy granular enclosing composed of fuzzy rough approximations. With that we obtain families of fuzzy building blocks which may be interpreted as certain and possible fuzzy rules. Concrete examples of fuzzy rough rule induction may be found in [17,18].

5. Granules and their interpretation

As we mentioned before, granules are important from the perspective of rule induction. We keep granules simple, such that one granule corresponds to one rule. Since a granularly representable set is a union of granules, it can be seen as a union of rules, so it is fully readable by a human. With granules in PRSA and fuzzy PRSA we are able to identify four types of rules: two types for the crisp case (IRSA and DRSA) and two types for the fuzzy case (fuzzy IRSA and fuzzy DRSA).

We first discuss IRSA granules and rules. Assume we are given an information table as a 4-tuple $< U, Q \cup \{d\}, V, f >$ where $U = \{u_1, \ldots, u_n\}$ is a finite set of objects or alternatives, $Q = \{q_1, \ldots, q_m\}$ is a finite set of condition attributes, d is a decision attribute, $V = \cup_{q \in Q\cup\{d\}} V_q$, where V_q is a domain of attribute $q \in Q \cup \{d\}$ and

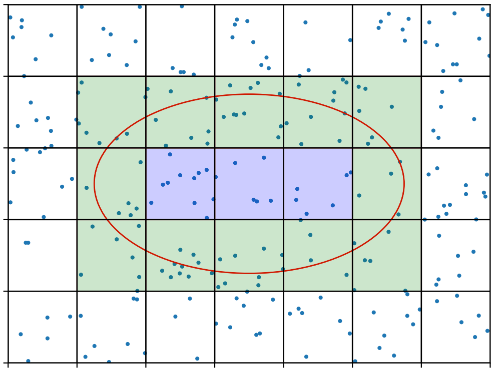

Fig. 1. Crisp approximations with equivalence relation. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

$f : U \times Q \cup \{d\} \to V$ is an information function such that $f(u,q) \in V_q$ for each $u \in U$ and $q \in Q$. Since we are dealing with classification problems, we require that $|V_d|$ is finite. We assume for the moment that V_q contains only categorical values without quantitative meaning (categories like: cat, dog, house ...). We are able to construct an equivalence relation E among objects based on identity as $(u,v) \in E$ if $f(u,q) = f(v,q),\ \forall q \in Q$. Two objects relate if they are identically evaluated on all condition attributes.

In Fig. 1, we show an example of binary classification ($|V_d| = 2$) where 250 objects (points) are separated by the elliptical curve into interior and exterior classes. The equivalence classes in the set of objects are represented by the squares in the figure (35 equivalence classes). The lower approximation of the interior class is marked with light blue color, while its upper approximation is the union of light green and light blue squares. The approximations are granularly representable sets so they are equal to the union of the equivalence classes of their objects. We notice that we can choose one granule per representative element for each equivalence class. Therefore, the interior can be represented as a union of three classes as we can see in the figure. Each such granule can be seen as a rule. Since equivalence classes consist of objects with equal values on all attributes, the rules have the following form:

IF $\text{att}_1 = \text{val}_1$ AND ... AND $\text{att}_m = \text{val}_m$ THEN decision is dec.

Here “att”, “val” and “dec” are abbreviations for “attribute”, “value” and “decision”. The lower approximation generates certain rules, while the upper approximation generates possible rules.

We continue with rules obtained from DRSA. We now assume that data are ordinal, i.e., there exists a total order $\geq_q$ on domain V_q of every attribute $q \in Q$ and on d. From this, we may induce a dominance relation D defined as $(u,v) \in D \Leftrightarrow u \geq_q v,\ \forall q \in Q$. We denote $D^+(u) = \{v \in U : (v,u) \in D\}$, which will play the role of DRSA granules.

Using DRSA, we can approximate upward and downward unions, which are sets of objects having at least, resp. at most, a particular value of the decision attribute. By the granular representation, their lower and upper approximations are unions of granules. Using the simple property that $(u,v) \in D \implies D^+(v) \subseteq D^+(u)$, we can eliminate redundant granules (those contained in a bigger granule) and reduce the number of granules covering lower and upper approximations. The rules which can be obtained in this case are called monotonic rules which have the following form for upward unions:

IF $\text{att}_1 \geq \text{val}_1$ AND ... AND $\text{att}_m \geq \text{val}_m$ THEN decision is at least dec.

Here $\text{val}_1, \ldots, \text{val}_m$ are obtained from the attribute values of the generating object of that particular granule. Analogously, rules with opposite direction ($\leq$) can be constructed for downward unions.

Next, we assume that our data are numerical. This data type contains measurable information collected in number form. It does not contain numerical codes of categories which are of nominal type.

Such numerical data have ordinal properties that can be captured by the DRSA approach, but additionally one can also measure similarity between instances. Assume $\widetilde{E}_q$ is a fuzzy T-equivalence relation on U for attribute q. To construct the relation over all attributes we aggregate particular values: $\widetilde{E}(u,v) = T_{q \in Q}\widetilde{E}_q(u,v)$. We may use some other aggregation operator here, but the use of the t-norm is motivated by the fact that T-transitivity is preserved in this way. In Fig. 2, we show the widely used triangular similarity, defined by:

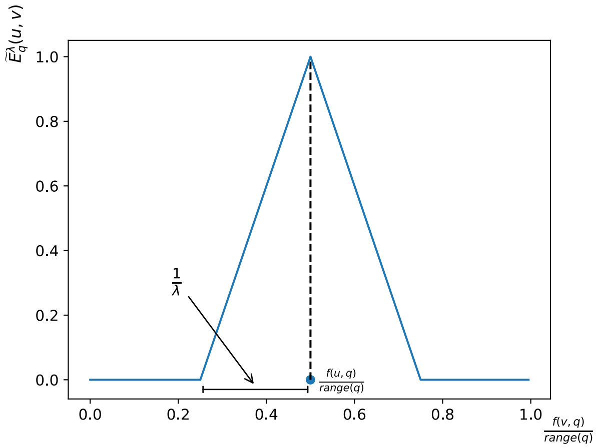

Fig. 2. Triangular similarity.

$$\widetilde{E}_q^\lambda(u, v) = \max\left(1 - \lambda\frac{|f(u, q) - f(v, q)|}{range(q)}, 0\right).$$

where $range(q)$ is the range of the attribute q and $\lambda > 0$ is a shrinking parameter.

In Fig. 3, we show an example of granularity of the lower approximation. In this example triangular similarity, Łukasiewicz t-norm and its R-implicator are used. The top-left sub-figure shows a fuzzy set denoted by the blue line together with its lower approximation, denoted by the green line. The top right sub-figure contains examples of a few granules that can be extracted from the lower approximation.They are represented by red triangles with points on their top. We displayed only seven granules, but in reality every object generates its own granule. We may see that some granules are included in others (small triangles inside the bigger ones), so we may safely remove them since they do not contribute to the granular representation of the lower approximation. In the bottom-left subfigure, we see the same example where redundant granules are removed. Sometimes, we want to obtain an even smaller number of granules in order to reduce the number of rules. For example, we may impose the condition that every object which belongs to the lower approximation to degree at least 0.5, is covered by some granule with degree at least 0.5. In the bottom-right image, we show the reduced set of granules which satisfies this condition.

If we assume that we use IRSA on crisp decision classes, we can induce the rules of the form:

IF $\text{att}_1 \sim \text{val}_1$ AND ... AND $\text{att}_m \sim \text{val}_m$ THEN decision is dec,

where $\sim$ stands for expression “is similar to”. Here, as before, $\text{val}_1, \ldots, \text{val}_m$ are obtained from the object which generates the granule.

The fourth type of rules corresponds to fuzzy DRSA. Again, we assume that we have numerical data and in this case we also take into account their ordinal properties. A fuzzy T-dominance relation may be constructed from a fuzzy T-equivalence relation on a particular attribute as:

$$\widetilde{D}_q(x, y) = \begin{cases} 1 & \text{if } x \geq y \\ \widetilde{E}_q(x, y) & \text{otherwise.} \end{cases}$$

It is easy to check that such fuzzy relation is indeed reflexive and T-transitive. It induces two regions: the region of strict dominance and the region of similarity. Hence, the interpretation of rules which correspond to the granules obtained from fuzzy dominance relation is “greater or similar” (“lower or similar” for the opposite direction). If we assume that we use fuzzy DRSA on crisp upward unions, then the induced rules are of the form:

IF $\text{att}_1 \gtrsim \text{val}_1$ AND ... AND $\text{att}_m \gtrsim \text{val}_m$ THEN decision is at least dec.

Here, $\gtrsim$ stands for the expression “is greater or similar”, and $\text{val}_1, \ldots, \text{val}_m$ are obtained from the object which generates the granule.

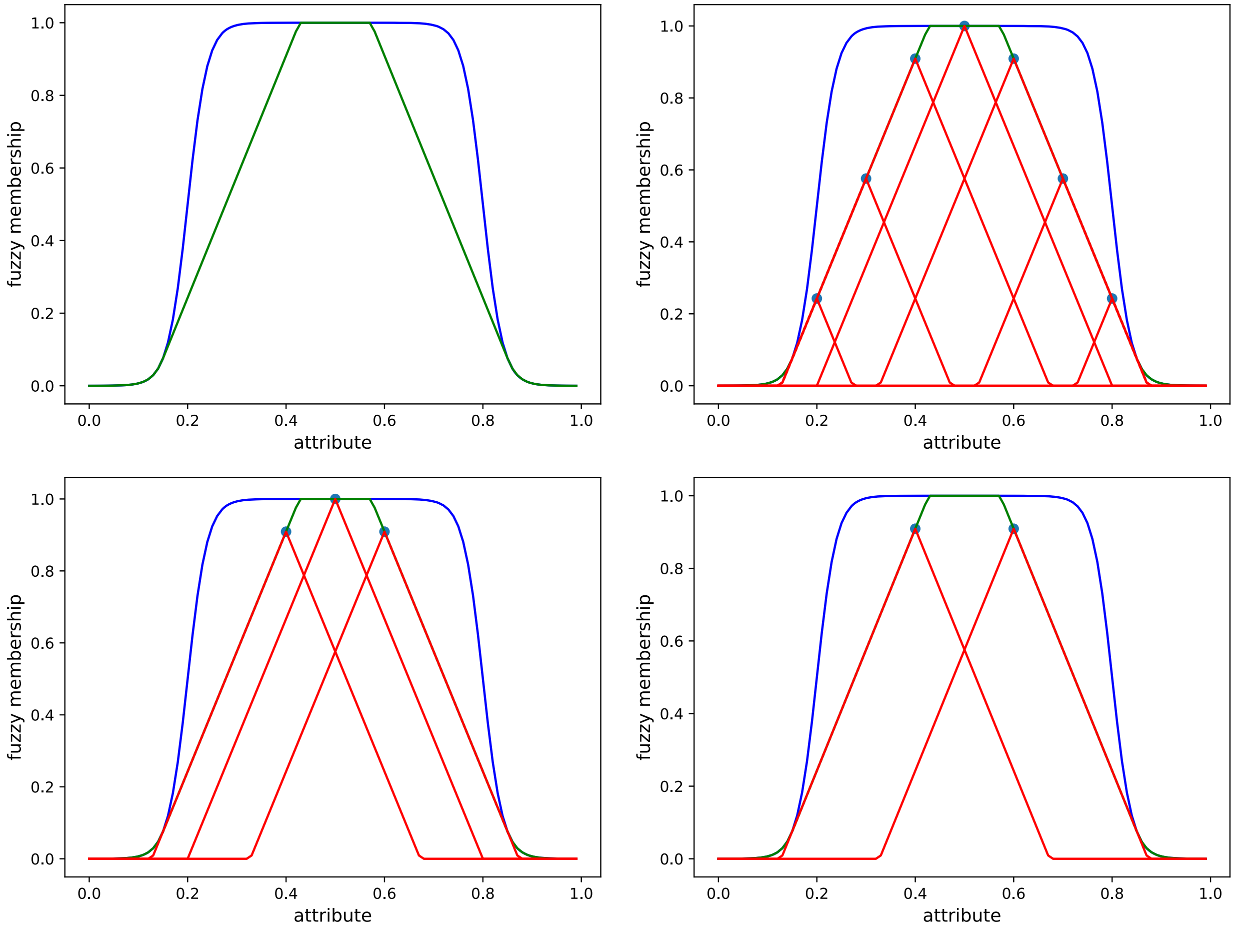

Fig. 3. Example of lower approximation and its granules.

6. Granular representation of OWA based approximations

In practice, data collected for real machine learning problems may be represented as unknown values plus noise. If the amount of noise is negligible, we can use the standard fuzzy PRSA approach to calculate rough and granular approximations. In the opposite case, we require noise-tolerant methods. As already mentioned in the introduction, the OWA-based approach was identified as the most robust known fuzzy rough approach. OWA-based PRSA will yield different lower and upper approximations. Since the purpose of the approach is to reduce the influence of the noise as much as possible, we may interpret these new approximations as (standard) PRSA approximations of some unknown "real" fuzzy set. As such, we would like to be able to treat them as granular approximations too. In this section, we will prove that under specific conditions, OWA-based fuzzy rough approximations are granularly representable fuzzy sets.

From Proposition 4.1, we already know that a fuzzy set has a granular representation if and only if it is equal to its standard fuzzy rough approximations. Therefore, we should find out under what conditions it holds that:

$$\underline{\mathrm{apr}}_{\widetilde{R}}^{\inf,I}(\underline{\mathrm{apr}}_{\widetilde{R}}^{W_L,I}(A)) = \underline{\mathrm{apr}}_{\widetilde{R}}^{W_L,I}(A), \quad \overline{\mathrm{apr}}_{\widetilde{R}}^{\sup,T}(\overline{\mathrm{apr}}_{\widetilde{R}}^{W_U,T}(A)) = \overline{\mathrm{apr}}_{\widetilde{R}}^{W_U,T}(A).$$

To this aim, we recall some definitions and properties about convexity for binary fuzzy logic connectives.

Definition 6.1. [27] We say that a binary operator $H : [0,1] \times [0,1] \to [0,1]$ is convex (concave) if for all $x_1, x_2, y_1, y_2 \in [0,1]$ and $w_1, w_2 \in [0,1]$ such that $w_1 + w_2 = 1$ holds that:

$$H(w_1x_1 + w_2x_2, w_1y_1 + w_2y_2) \le (\ge)\ w_1H(x_1, y_1) + w_2H(x_2, y_2).$$

Definition 6.2. [27] We say that a binary operator $H : [0,1] \times [0,1] \to [0,1]$ is midpoint convex (concave) if for all $x_1, x_2, y_1, y_2 \in [0,1]$ holds that:

$$H\left(\frac{x_1}{2} + \frac{x_2}{2}, \frac{y_1}{2} + \frac{y_2}{2}\right) \leq (\geq) \ \frac{H(x_1, y_1)}{2} + \frac{H(x_2, y_2)}{2}.$$

Proposition 6.1. *[27] A continuous midpoint convex (concave) t-norm is convex (concave).*

Definition 6.3. [28] We say that a binary operator $H : [0,1] \times [0,1] \to [0,1]$ is directionally convex or D-convex (directionally concave or D-concave) if it is a convex (concave) function in both of its arguments, i.e., for all $x_1, x_2, y \in [0,1]$ and $w_1, w_2 \in [0,1]$ such that $w_1 + w_2 = 1$ holds that:

$$H(w_1x_1 + w_2x_2, y) \leq (\geq) \ w_1H(x_1, y) + w_2H(x_2, y) \quad \text{and}$$
$$H(y, w_1x_1 + w_2x_2) \leq (\geq) \ w_1H(y, x_1) + w_2H(y, x_2).$$

This definition expresses that the partial mappings of H are convex (concave) functions. We prove a simple proposition:

Proposition 6.2. *Every convex (concave) operator is also D-convex (D-concave).*

Proof. Just take $x_1 = x_2 = x$ or $y_1 = y_2 = y$ in the previous definitions. □

The reverse implication is not necessarily satisfied. Now, we formulate and prove the following important result.

Proposition 6.3. *Let T be a convex left-continuous t-norm and let I be its R-implicator. Then I is concave.*

Proof. Assume we are given $x_1, x_2, y_1, y_2, w_1, w_2 \in [0,1]$ such that $w_1 + w_2 = 1$. We have to prove that

$$w_1I(x_1, y_1) + w_2I(x_2, y_2) \leq I(w_1x_1 + w_2x_2, w_1y_1 + w_2y_2).$$

Using the residuation property, we can express this condition as

$$T(w_1I(x_1, y_1) + w_2I(x_2, y_2), w_1x_1 + w_2x_2) \leq w_1y_1 + w_2y_2.$$

By the convexity of T we have that:

$$T(w_1I(x_1, y_1) + w_2I(x_2, y_2), w_1x_1 + w_2x_2) \leq w_1T(x_1, I(x_1, y_1)) + w_2T(x_2, I(x_2, y_2)).$$

Using Proposition 2.1a., we have that

$$T(x_1, I(x_1, y_1)) \leq y_1 \text{ and } T(x_2, I(x_2, y_2)) \leq y_2.$$

which completes the proof. □

Proposition 6.4. *Let T be a D-convex left-continuous t-norm and I its R-implicator. Then I is concave in its second argument.*

Proof. Similarly as for Proposition 6.3. □

We recall the following well-known inequality from calculus which will be needed further on.

Proposition 6.5 *(Jensen's inequality). [29] Let $f : \mathbb{R} \to \mathbb{R}$ be a convex (concave) function. Let $x_1, \ldots, x_n$ be real numbers and $w_1, \ldots, w_n$ real weights which sum up to 1. Then we have*

$$f(w_1x_1 + \cdots + w_nx_n) \leq (\geq) \ w_1f(x_1) + \cdots + w_nf(x_n).$$

Before we proceed to the main theorem, we provide the following easily verified algebraic result.

Proposition 6.6. *Let $\{a_{i,j}, i \in \{1, \dots, n\}, j \in \{1, \dots, m\}\}$ be a matrix. Then we have that*

$$\inf_i \sum_j a_{i,j} \geq \sum_j \inf_i a_{i,j}, \quad \sup_i \sum_j a_{i,j} \leq \sum_j \sup_i a_{i,j}.$$

Theorem 6.1. *Let T be a D-convex left-continuous t-norm and I its R-implicator. Then for every $A \in \mathcal{F}(U)$ it holds that*

$$\underline{apr}_{\widetilde{R}}^{\inf,I}(\underline{apr}_{\widetilde{R}}^{W_L,I}(A)) = \underline{apr}_{\widetilde{R}}^{W_L,I}(A), \quad \overline{apr}_{\widetilde{R}}^{\sup,T}(\overline{apr}_{\widetilde{R}}^{W_U,T}(A)) = \overline{apr}_{\widetilde{R}}^{W_U,T}(A).$$

Proof. Observe that using T-transitivity of $\widetilde{R}$ we find

$$\forall w \in U, \widetilde{R}(v,u) \geq T(\widetilde{R}(v,w), \widetilde{R}(w,u)) \implies \widetilde{R}(v,u) \geq \sup_{w \in U} T(\widetilde{R}(v,w), \widetilde{R}(w,u)).$$

First, we provide the proof for the lower approximation. From the inclusion property we know that

$$\underline{\text{apr}}_{\widetilde{R}}^{\inf,I}(\underline{\text{apr}}_{\widetilde{R}}^{W_L,I}(A)) \subseteq \underline{\text{apr}}_{\widetilde{R}}^{W_L,I}(A).$$

We proceed to prove the opposite inequality. Due to Proposition 6.4, I is a concave function in its second argument, to which we can apply Jensen's inequality. We find:

$$\begin{aligned}
\underline{\text{apr}}_{\widetilde{R}}^{W_L,I}(A)(u) &= \sum_{v \in U} w_v I(\widetilde{R}(v,u), A(v)) \\
&\leq \sum_{v \in U} w_v I\left(\sup_{w \in U} T(\widetilde{R}(v,w), \widetilde{R}(w,u)), A(v)\right) \\
&\leq \sum_{v \in U} w_v \inf_{w \in U} I\left(T(\widetilde{R}(v,w), \widetilde{R}(w,u)), A(v)\right) \\
&\leq \inf_{w \in U} \sum_{v \in U} w_v I\left(T(\widetilde{R}(v,w), \widetilde{R}(w,u)), A(v)\right) \\
&= \inf_{w \in U} \sum_{v \in U} w_v I\left(\widetilde{R}(w,u), I(\widetilde{R}(v,w), A(v))\right) \\
&\leq \inf_{w \in U} I\left(\widetilde{R}(w,u), \sum_{v \in U} w_v I(\widetilde{R}(v,w), A(v))\right) \\
&= \inf_{w \in U} I\left(\widetilde{R}(w,u), \underline{\text{apr}}_{\widetilde{R}}^{W_L,I}(A)(w)\right) \\
&= \underline{\text{apr}}_{\widetilde{R}}^{\inf,I}(\underline{\text{apr}}_{\widetilde{R}}^{W_L,I}(A))(u).
\end{aligned}$$

The first inequality follows from T-transitivity of $\widetilde{R}$ and the monotonicity of I. The latter is also used in the second inequality. The third one holds because of Proposition 6.6, while the equality in the fourth step is due to Proposition 2.1b.

Next, we provide the proof for the upper approximation. From the inclusion property we have that:

$$\underline{\text{apr}}_{\widetilde{R}}^{\sup,T}(\underline{\text{apr}}_{\widetilde{R}}^{W_U,T}(A)) \supseteq \underline{\text{apr}}_{\widetilde{R}}^{W_U,T}(A).$$

We proceed to prove the opposite inequality, using similar arguments:

$$\begin{aligned}
\underline{\text{apr}}_{\widetilde{R}}^{W_U,T}(A)(u) &= \sum_{v \in U} w_v T(\widetilde{R}(u,v), A(v)) \\
&\geq \sum_{v \in U} w_v T\left(\sup_{w \in U} T(\widetilde{R}(u,w), \widetilde{R}(w,v)), A(v)\right) \\
&= \sum_{v \in U} w_v \sup_{w \in U} T(T(\widetilde{R}(u,w), \widetilde{R}(w,v)), A(v))
\end{aligned}$$

$$
\begin{aligned}
&\geq \sup_{w\in U} \sum_{v\in U} w_v T(T(\widetilde{R}(u,w), \widetilde{R}(w,v)), A(v)) \\
&= \sup_{w\in U} \sum_{v\in U} w_v T(\widetilde{R}(u,w), T(\widetilde{R}(w,v), A(v))) \\
&\geq \sup_{w\in U} T\left(\widetilde{R}(u,w), \sum_{v\in U} w_v T(\widetilde{R}(w,v), A(v))\right) \\
&= \sup_{w\in U} T\left(\widetilde{R}(u,w), \underline{\mathrm{apr}}_{\widetilde{R}}^{W_U,T}(A)(w)\right) \\
&= \underline{\mathrm{apr}}_{\widetilde{R}}^{\sup,T}(\underline{\mathrm{apr}}_{\widetilde{R}}^{W_U,T}(A))(u). \quad \square
\end{aligned}
$$

Therefore, using Corollaries 4.2 and 4.3 we have the following granular representation of the OWA-based approximations.

$$
\begin{aligned}
\underline{\mathrm{apr}}_{\widetilde{R}}^{W_L,I}(A) &= \bigcup\{\widetilde{R}_\lambda(u):\ \widetilde{R}_\lambda(u) \subseteq \underline{\mathrm{apr}}_{\widetilde{R}}^{W_L,I}(A)\}, \\
\overline{\mathrm{apr}}_{\widetilde{R}}^{W_U,T}(A) &= \bigcup\{\widetilde{R}_\lambda(u):\ \widetilde{R}_\lambda(u) \subseteq \overline{\mathrm{apr}}_{\widetilde{R}}^{W_U,T}(A)\}.
\end{aligned}
$$

With this result, we can conclude that OWA-based approximations are not only robust fuzzy rough approximations, but also robust granular approximations. The question remains which connectives preserve the granularity property, or in other words, which left-continuous t-norms are also D-convex.

7. Partial characterization of D-convex t-norms

Convexity is a crucial property for the granularity of the OWA-based operators. The general characterization of convex t-norms is still an open problem, but D-convex t-norms with some additional characteristics may be well characterized. The results in this section are mainly an adaptation of the existing work on characterizing convex copulas [30].

Assume that we have a continuous D-convex t-norm T. For a t-norm it is known that it is continuous as a function of two variables, if and only if its partial mappings are continuous [19]. From basic calculus we know that convex functions are continuous on the interior of the domain, which is in this case the interval $(0, 1)$ [31]. However, we may have a discontinuity at the points 0 and 1. An example of a discontinuous D-convex t-norm is the drastic t-norm T_D.

In this section, we want to characterize left-continuous D-convex t-norms. The following proposition shows that such t-norms are necessarily continuous.

Proposition 7.1. *Every left-continuous D-convex t-norm is continuous.*

Proof. As we noted before, a D-convex t-norm can only have discontinuities in 0 or 1. Moreover, a left-continuous t-norm cannot have a discontinuity in 1. Hence, the only possibility is that it is discontinuous in 0. However, we will prove that the partial mappings of any t-norm are right-continuous in 0.

Let $c \in [0, 1]$ be a constant. For every $\epsilon > 0$, we need to find $\delta > 0$ such that $x - 0 < \delta \implies T(x, c) - T(0, c) < \epsilon \Leftrightarrow x < \delta \implies T(x, c) < \epsilon$. Taking $\delta = \epsilon/2$ we have that

$$
x < \epsilon/2 \implies T(x,c) \leq \min(x,c) \leq \min(\epsilon/2, c) \leq \epsilon/2 < \epsilon
$$

which is true. From this, we conclude there is no discontinuity in 0, i.e., T is continuous in $[0, 1]$. $\square$

We proceed with the characterization. First, we show that T cannot have any non-trivial idempotent element. Assume that it has an idempotent point $z \in (0, 1)$. From [19], we may then conclude that $T(x, z) = \min(x, z)$ for all $x \in (z, 1]$. However, it is easy to see that the function $f(x) = \min(x, c)$ is not convex for any constant $c \in (0, 1]$, so T is not a D-convex t-norm. Because of that, we have a contradiction, and T cannot have idempotent points. In particular, the minimum t-norm T_M is not convex.

Under the assumption of continuity, T does not have idempotent points if and only if it is Archimedean [19]. Since T is a continuous Archimedean t-norm, it has a unique representation [19]:

$$T(x, y) = f^{-1}(\min(f(x) + f(y), f(0))),$$

where f is a decreasing generator, i.e., $f : [0, 1] \to \mathbb{R}^+$ is a strictly decreasing continuous $[0, 1] \to [0, +\infty]$ mapping for which $f(1) = 0$.

In [30], necessary and sufficient conditions for D-convexity of Archimedean copulas are derived. Here we repeat the proof, adapting it for t-norms.

Theorem 7.1. *Let T be a continuous Archimedean t-norm with a twice differentiable generator f. Then T is D-convex if and only if $1/f'$ is a convex function.*

Proof. We have a representation of T from above as:

$$T(x, y) = f^{-1}(\min(f(x) + f(y), f(0))).$$

Since f is twice differentiable, T is D-convex if and only if $T_{xx}(x, y) \geq 0$ and $T_{yy}(x, y) \geq 0$ for all x, y such that $f(x) + f(y) \leq f(0)$, where T_{xx} is the second partial derivative for the first component, while T_{yy} is the second partial derivative for the second one. Due to the symmetry of T, it suffices to show that $T_{xx} \geq 0$. We find:

$$T_{xx}(x, y) = \frac{f''(x)(f'(f^{-1}(f(x) + f(y)))) - f'(x)^2 f''(f^{-1}(f(x) + f(y)))}{(f'(f^{-1}(f(x) + f(y))))^3}.$$

It holds that $f'(x) < 0$ since f is a strictly decreasing function, so the condition that $T_{xx}(x, y) \geq 0$ is equivalent to

$$\frac{f''(x)}{f'(x)^2} \leq \frac{f''(f^{-1}(f(x) + f(y)))}{f'(f^{-1}(f(x) + f(y)))^2}.$$

We introduce a new variable $u = f^{-1}(f(x) + f(y))$. From the definition we conclude that $u = f^{-1}(f(x) + f(y)) \leq f^{-1}(f(x)) = x$ due to the fact that f^{-1} is also a strictly decreasing function. Now the condition above becomes

$$\frac{f''(x)}{f'(x)^2} \leq \frac{f''(u)}{f'(u)^2}.$$

We have that $\frac{f''(x)}{f'(x)^2} = -(\frac{1}{f'(x)})'$, so the above condition may be rewritten as

$$\left(\frac{1}{f'(x)}\right)' \geq \left(\frac{1}{f'(u)}\right)'. \tag{2}$$

Note that for a fixed x, u can take any value smaller or equal to x. Indeed, from the condition that $f(x) + f(y) \leq f(0)$, it follows that y takes values from the interval $[f^{-1}(f(0) - f(x)), 1]$. Using this as a domain for y, we have that the function $u(y) = f^{-1}(f(x) + f(y))$ is a bijective mapping $[f^{-1}(f(0) - f(x)), 1]$ to $[0, x]$. So for every $u \leq x$, we can choose some y to obtain u.

Since u may take all the values smaller or equal to x, we have that the condition (2) states that $\left(\frac{1}{f'(x)}\right)'$ is an non-decreasing function. This is further equivalent to $\left(\frac{1}{f'(x)}\right)'' \geq 0$ which means that $\frac{1}{f'(x)}$ is a convex function. □

Example 7.1. We present a way to construct a generator satisfying the conditions of the previous theorem. The construction is also inspired by [30] but adapted here to t-norms. Let $g : [0, 1] \to [0, \infty]$ be a convex function with $g(1) > 0$. Then the generator can be constructed as:

$$f(x) = \int_x^1 \frac{1}{g(u)} du.$$

By the positivity of g, we ensure that f is a decreasing function, while its convexity ensures that $\frac{1}{f'}$ is a convex function.

To illustrate that our adaptation of the work from [30] brings new knowledge, we need to show that there exists a D-convex t-norm which is not a copula. The following example confirms this.

Example 7.2. In Example 7.1 take $g(u) = 2 - u$. We have the following:

$$f(x) = \int_x^1 \frac{1}{2-u} du = -\log(2-1) + \log(2-x) = \log(2-x),$$

while $f^{-1}(x) = 2 - e^x$. Using such generator, we construct the associated t-norm:

$$\begin{aligned} T(x, y) &= 2 - e^{\min(\log(2-x)+\log(2-y), \log(2))} = 2 - e^{\log(\min((2-x)(2-y),2))} \\ &= 2 - \min((2-x)(2-y), 2) = \max(2(x+y-1) - xy, 0). \end{aligned}$$

If we take values $x = 0.5$, $y = 0.9$, $x' = 0.4$, $y' = 0.8$, we can see that the 2-increasingness property does not hold, i.e.

$$T(x, y) + T(x', y') < T(x', y) + T(x, y').$$

which means that T is not a copula.

Furthermore, we can easily check, with the same values, that T is not midpoint convex, which is equivalent to stating that it is not convex.

Example 7.3. We check the D-convexity of the left-continuous t-norms from Table 1.

- Łukasiewicz t-norm $T_L(x, y) = \max(x + y - 1, 0)$ is D-convex since its partial mappings are a composition of a linear function and max, which are both convex. It was proven in [32] that T_L is even convex.
- Product t-norm $T_P(x, y) = xy$ is D-convex because its partial mappings are linear functions.
- From the above exposition, we already know that the minimum t-norm $T_M(x, y) = \min(x, y)$ is not D-convex.
- The nilpotent minimum t-norm

$$T_{nM}(x, y) = \begin{cases} \min(x, y) & \text{if } x + y > 1, \\ 0 & \text{otherwise.} \end{cases}$$

  is not D-convex because its partial mappings have discontinuities in the interior (0, 1) of its domain.

8. Conclusion and future work

In this paper, we have studied the granular representability of crisp and fuzzy sets w.r.t. a (fuzzy) preorder relation. We introduced the notion of a granularly representable (fuzzy) set as a union of simple granules, where granules represent the fuzzy equivalence or dominance classes of individual objects. As our main contribution, we proved that OWA-based fuzzy rough approximations are granularly representable sets when we use D-convex left-continuous t-norms and their residual implicators for calculating the approximations. From that perspective, we may conclude that OWA-based fuzzy rough approximations are also granular approximations. Finally, we characterized continuous convex t-norms and we presented a method how to construct them.

Our future work will explore the direction of rule induction. We have seen that granules may be interpreted as fuzzy rules and we want to investigate if such rules can lead to better classification models. Some of the challenges to tackle include:

- Finding a proper induction algorithm to reduce the number of covering rules.
- Reducing the length of individual rules: a rule does not have to include all attributes in its condition part. For this task we may benefit from rough and fuzzy rough reducts.

- Inducing different sets of covering rules as basic classifiers and using them in ensemble classifiers.
- Merging different rules in order to get less rules which would cover larger portions of objects.
- Evaluating how the OWA-based approach may improve the accuracy of classification.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Marko Palangetić and Chris Cornelis would like to thank Odysseus project from Flanders Research Foundation (FWO), grant G0H9118N for funding their research. Salvatore Greco wishes to acknowledge the support of the Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) - PRIN 1576 2017, project "Multiple Criteria Decision Analysis and Multiple Criteria Decision Theory", grant 2017CY2NCA. The research of Roman Słowiński was supported by the Polish Ministry of Education and Science, grant no. 0311/SBAD/0709.

References

[1] Z. Pawlak, Rough sets, Int. J. Comput. Inf. Sci. 11 (5) (1982) 341–356.
[2] S. Greco, B. Matarazzo, R. Slowinski, Rough sets theory for multicriteria decision analysis, Eur. J. Oper. Res. 129 (1) (2001) 1–47.
[3] L. Zadeh, Fuzzy sets, Inf. Control 8 (3) (1965) 338–353.
[4] D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, Int. J. Gen. Syst. 17 (2–3) (1990) 191–209.
[5] S. Greco, B. Matarazzo, R. Slowinski, Fuzzy extension of the rough set approach to multicriteria and multiattribute sorting, in: Preferences and Decisions Under Incomplete Knowledge, Springer, 2000, pp. 131–151.
[6] T.-F. Fan, C.-J. Liau, D.-R. Liu, Variable consistency and variable precision models for dominance-based fuzzy rough set analysis of possibilistic information systems, Int. J. Gen. Syst. 42 (6) (2013) 659–686.
[7] A. Mieszkowicz-Rolka, L. Rolka, Variable precision fuzzy rough sets, in: Transactions on Rough Sets I, Springer, 2004, pp. 144–160.
[8] C. Cornelis, M. De Cock, 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.
[9] C. Cornelis, N. Verbiest, R. Jensen, Ordered weighted average based fuzzy rough sets, in: Proceedings of the 5th International Conference on Rough Sets and Knowledge Technology, RSKT 2010, 2010, pp. 78–85.
[10] M. Palangetić, C. Cornelis, S. Greco, R. Słowiński, Extension of the fuzzy dominance-based rough set approach using ordered weighted average operators, in: 2019 Conference of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology, EUSFLAT 2019, Atlantis Press, 2019.
[11] L. D'eer, N. Verbiest, C. Cornelis, L. Godo, A comprehensive study of implicator–conjunctor-based and noise-tolerant fuzzy rough sets: definitions, properties and robustness analysis, Fuzzy Sets Syst. 275 (2015) 1–38.
[12] Y. Yao, Rough sets, neighborhood systems and granular computing, Engineering Solutions for the Next Millennium. 1999 IEEE Canadian Conference on Electrical and Computer Engineering (Cat. No. 99TH8411), vol. 3, IEEE, 1999, pp. 1553–1558.
[13] C. Degang, Y. Yongping, W. Hui, Granular computing based on fuzzy similarity relations, Soft Comput. 15 (6) (2011) 1161–1172.
[14] Y. Yao, J. Mi, Z. Li, A novel variable precision (θ, σ)-fuzzy rough set model based on fuzzy granules, Fuzzy Sets Syst. 236 (2014) 58–72.
[15] J.W. Grzymala-Busse, Lers-a system for learning from examples based on rough sets, in: Intelligent Decision Support, Springer, 1992, pp. 3–18.
[16] S. Greco, B. Matarazzo, R. Slowinski, J. Stefanowski, An algorithm for induction of decision rules consistent with the dominance principle, in: International Conference on Rough Sets and Current Trends in Computing, Springer, 2000, pp. 304–313.
[17] S. Zhao, E.C. Tsang, D. Chen, X. Wang, Building a rule-based classifier—a fuzzy-rough set approach, IEEE Trans. Knowl. Data Eng. 22 (5) (2009) 624–638.
[18] R. Jensen, C. Cornelis, Q. Shen, Hybrid fuzzy-rough rule induction and feature selection, in: 2009 IEEE International Conference on Fuzzy Systems, IEEE, 2009, pp. 1151–1156.
[19] E.P. Klement, R. Mesiar, E. Pap, Triangular Norms, vol. 8, Springer Science & Business Media, 2013.
[20] M. Palangetić, C. Cornelis, S. Greco, R. Słowiński, Fuzzy extensions of the dominance-based rough set approach, Int. J. Approx. Reason. 129 (2021) 1–19.
[21] C. Cornelis, N. Verbiest, R. Jensen, Ordered weighted average based fuzzy rough sets, in: International Conference on Rough Sets and Knowledge Technology, Springer, 2010, pp. 78–85.
[22] R.R. Yager, On ordered weighted averaging aggregation operators in multicriteria decisionmaking, IEEE Trans. Syst. Man Cybern. 18 (1) (1988) 183–190.
[23] Y. Yao, Information granulation and rough set approximation, Int. J. Intell. Syst. 16 (1) (2001) 87–104.
[24] S. Greco, B. Matarazzo, R. Słowiński, Dominance-based rough set approach as a proper way of handling graduality in rough set theory, in: Transactions on Rough Sets VII, Springer, 2007, pp. 36–52.

[25] L. D'eer, C. Cornelis, Y. Yao, A semantical approach to rough sets and dominance-based rough sets, in: Proceedings of 16th International Conference on Information Processing and Management of Uncertainty (IPMU2016), Part II, in: CCIS, vol. 611, 2016, pp. 23–35.
[26] B.W. Fang, B.Q. Hu, Granular fuzzy rough sets based on fuzzy implicators and coimplicators, Fuzzy Sets Syst. 359 (2019) 112–139.
[27] C. Alsina, B. Schweizer, M.J. Frank, Associative Functions: Triangular Norms and Copulas, World Scientific, 2006.
[28] J. Matoušek, On directional convexity, Discrete Comput. Geom. 25 (3) (2001) 389–403.
[29] J. Jensen, Sur les fonctions convexes et les inégalités entre les valeurs moyennes, Acta Math. 30 (1) (1906) 175–193.
[30] E.P. Klement, M. Manzi, R. Mesiar, et al., Ultramodularity and copulas, Rocky Mt. J. Math. 44 (1) (2014) 189–202.
[31] W. Rudin, et al., Principles of Mathematical Analysis, vol. 3, McGraw-Hill, New York, 1964.
[32] C. Alsina, M. Tomás, Smooth convex t-norms do not exist, Proc. Am. Math. Soc. (1988) 317–320.