



Multi-class granular approximation by means of disjoint and adjacent fuzzy granules

Marko Palangetic^{a,*}, Chris Cornelis^a, Salvatore Greco^b, Roman Słowiński^{c,d}

^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 Institute of Computing Science, Poznań University of Technology, Poznań, Poland

^d Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

ARTICLE INFO

Keywords:

Granular computing
Fuzzy sets
Machine learning

ABSTRACT

In granular computing, fuzzy sets can be approximated by granularly representable sets that are as close as possible to the original fuzzy set w.r.t. a given closeness measure. Such sets are called granular approximations. In this article, we introduce the concepts of disjoint and adjacent granules and we examine how the new definitions affect the granular approximations. First, we show that the new concepts are important for binary classification problems since they help to keep decision regions separated (disjoint granules) and at the same time to cover as much as possible of the attribute space (adjacent granules). Later, we consider granular approximations for multi-class classification problems leading to the definition of a multi-class granular approximation. Finally, we show how to efficiently calculate multi-class granular approximations for Łukasiewicz fuzzy connectives. We also provide graphical illustrations for a better understanding of the introduced concepts.

1. Introduction

Granular computing is a paradigm in information processing which includes a segmentation of complex information into smaller pieces called *information granules* [1–3]. An information granule (or just a *granule*) is a collection of instances that can be interpreted jointly. For example, an image of a human body can be disentangled into certain body parts that have precise meanings. Also, those parts can be later segmented into even smaller meaningful parts, etc. The previous example also shows the hierarchical nature of granulation, i.e., the definition of granules depends on the level of detail that we want to capture. Granules are usually constructed based on a common association (indiscernibility, similarity, functionality, proximity, coherency etc.) of instances [4,5].

Fuzzy logic and fuzzy set theory are used to model partial truth of logical expressions [6]. In other words, the expression is not only true or false, but it possesses a degree of truth represented by a value from interval [0, 1]. Value 0 stands for a completely false statement, while value 1 stands for a completely true statement. With the help of fuzzy logic, one can introduce the concept of a fuzzy granule [7] where every instance has a degree of membership to a certain granule. Fuzzy granules are useful when it is hard

* Corresponding author.

E-mail addresses: marko.palangetic@ugent.be (M. Palangetic), 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.2023.108765>

Received 14 February 2022; Received in revised form 21 October 2023; Accepted 23 October 2023

Available online 31 October 2023

0165-0114/© 2023 Elsevier B.V. All rights reserved.

to determine sharp boundaries of pieces obtained from a disentanglement of a complex object. In such case, soft boundaries are expressed using fuzzy sets.

Lotfi Zadeh identified granulation as one of three basic concepts in underlying human cognition [8], the other two being organization and causation. While organization represents the integration of parts into a whole, granulation refers to the opposite process. With fuzziness as a key part of the granulation in human cognition, humans are able to make reasonable decisions in a world that is characterized with partial knowledge, partial certainty, partial truth and imprecision in general.

In this article, granules are identified in *information tables* based on the concept of data consistency, following our previous work in [9,10]. Assume we have a prediction problem where we want to assign a decision label to a given instance described by condition attributes. In this setting, we say that two instances are consistent w.r.t. a given relation, if their relation on the condition attributes implies the same type of relation on the decision attribute. The relations that we consider here are (fuzzy) indiscernibility and (fuzzy) dominance. Based on that, an instance is consistent in a dataset if it is consistent with all other instances. For a consistent instance (w.r.t. a given relation), a granule is formed as a conjunction of two concepts:

- the set of instances that relate to the given consistent instance, and
- the association of the consistent instance to a particular decision.

Due to consistency, the instances that relate (w.r.t. a given relation like e.g. indiscernibility or dominance) to a given consistent instance will be associated to the same decision or to a decision that relates to the decision of the consistent instance. In a classification problem, we have decision classes and association of the consistent instance refers to the membership of the instance to a decision class. In a regression problem, the association refers to the numerical value that the consistent instance takes in the decision attribute.

In practice, due to perturbation in data caused by incomplete knowledge or by random effects that occur during data generation, datasets contain instances that are not consistent. To make a dataset consistent, different approaches have been applied. The best-known symbolic approach to this issue is the rough set approach (or the indiscernibility-based rough set approach (IRSA)) and its generalizations like the dominance-based rough set approach (DRSA) and fuzzy rough sets (fuzzy IRSA and fuzzy DRSA) [11–13,9]. The disentanglement of crisp rough sets into granules was discussed in [5] and of fuzzy rough sets in [4,14,15]. The IRSA and DRSA were integrated into the preorder-based rough set approach (PRSA), and their fuzzy counterparts into the fuzzy PRSA. A more comprehensive overview is provided in [9].

On the other hand, removing inconsistencies using machine learning w.r.t. a crisp preorder relation is covered in [16] and w.r.t. a fuzzy preorder relation in [10]. The result of the former approach is called *monotone approximation* (due to the monotonicity properties of the granules), while the result of the latter approach is called *granular approximation*. These machine learning approaches include an optimization procedure that removes inconsistencies at the least possible cost (w.r.t. a certain loss function).

In this article, we extend the definition of granular approximations to the multi-class context. We introduce concepts of disjoint and adjacent fuzzy granules and we discuss how these concepts relate to the formerly introduced granular approximations. They are important in classification problems since they help us to keep decision regions separated (disjoint granules) while covering as much as possible of an attribute space (adjacent granules). Then, we formulate an optimization procedure in order to extend granular approximations to the multi-class classification problem leading to the definition of *multi-class granular approximations*. Such approximation is a union of granules constructed in the way described above; it is a fuzzy set which membership degrees are conjunctions of two fuzzy values: an evaluation of fuzzy relation and an association value. These association values, as discussed in [10], can be interpreted as the degree up to which an instance belongs to a certain decision class.

In some classification problems, application of “the degree up to which an instance belongs to a class” may not be possible, since a decision class may be defined in a strictly discrete way, i.e., an instance belongs to a decision class or not. Examples of the case where consideration of membership degrees is more appropriate are recommender systems. A degree of preference for a certain product by a user can be modeled using values between 0 and 1. It is also possible that collected data contain only binary preferences (e.g., likes and dislikes) while the underlying degree of preference is hidden and can be estimated using machine learning techniques.

It is also important to mention that granules play a crucial role in developing rule induction algorithms based on rough sets. Namely, as discussed in the previous work of the same authors [9,10], the granules can be interpreted as decision rules and granular properties of rough sets have been used to develop rule induction algorithms [17–19]. Using the results of this paper, we will be able to achieve the following:

- The disjoint granules will enable to precisely associate a particular rule to a decision class (without overlapping).
- The adjacent granules will allow rules to cover as much as space as possible (when all other parameters are fixed).
- Multi-class granular approximations will help us to develop rule induction algorithms for multi-class classification problems.

The remainder of this paper is structured as follows. In Section 2, we recall some useful preliminaries about fuzzy logic theory, fuzzy rough and granularly representable sets, and granular approximations. Section 3 deals with the concept of disjoint granules and adjacent granules. It provides definitions of the concepts together with an analysis of how these definitions pertain to the granular approximations introduced in the previous section. Section 4 explains how the concepts from the previous sections can be applied in binary and multi-class classification problems, and it introduces the definition of a multi-class granular approximation. Section 5 presents how to efficiently calculate multi-class granular approximations and provides a graphical illustration of how the granules look in practice. Section 6 contains the conclusion and outlines future work.

Table 1
Some common t -norms and their R-implicators.

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

2. Preliminaries

2.1. Fuzzy logic connectives

In this subsection, the definitions and terminology are based on [20]. Recall that t -norm $T : [0, 1]^2 \rightarrow [0, 1]$ is a binary operator which is commutative, associative, non-decreasing in both arguments, and $\forall x \in [0, 1], T(x, 1) = x$. Since a t -norm is associative, we may extend it unambiguously to a $[0, 1]^n \rightarrow [0, 1]$ mapping for any $n > 2$. Some commonly used t -norms are listed in the left-hand side of Table 1.

An *implicator* (or *fuzzy implication*) $I : [0, 1]^2 \rightarrow [0, 1]$ is a binary operator which is non-increasing in the first component, non-decreasing in the second one and such that $I(1, 0) = 0$ and $I(0, 0) = I(0, 1) = I(1, 1) = 1$. The residuation property holds for a t -norm T and implicator I if for all $x, y, z \in [0, 1]$, it holds that

$$T(x, y) \leq z \Leftrightarrow x \leq I(y, z).$$

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 (R-implicator) of T , that is

$$I_T(x, y) = \sup\{\beta \in [0, 1] : T(x, \beta) \leq y\}.$$

The right-hand side of Table 1 shows the residual implicators of the corresponding t -norms. Note that all of them, except I_{T_D} , satisfy the residuation property.

If T is a continuous t -norm, it is *divisible*, i.e., for all $x, y \in [0, 1]$, it holds that

$$\min(x, y) = T(x, I_T(x, y)) = T(y, I_T(y, x)).$$

A *negator* (or *fuzzy negation*) $N : [0, 1] \rightarrow [0, 1]$ is a unary and non-increasing operator for which it holds that $N(0) = 1$ and $N(1) = 0$. A negator is involutive if $N(N(x)) = x$ for all $x \in [0, 1]$. The standard negator, defined as

$$N_s(x) = 1 - x,$$

is involutive.

For implicator I , we define the negator induced by I as $N(x) = I(x, 0)$. We will call triplet (T, I, N) , obtained as previously explained, a residual triplet. For a residual triplet we have that the following properties hold for all $x, y, z \in [0, 1]$:

- $T(x, y) \leq x$ and $T(x, y) \leq y$, (1a)

- $I(x, y) \geq y$, (1b)

- $T(x, I(x, y)) \leq y$, (1c)

- $x \leq y \Leftrightarrow I(x, y) = 1$, (ordering property) (1d)

- $T(x, I(y, z)) \leq I(I(x, y), z)$, (1e)

- $I(T(x, y), z) = I(x, I(y, z))$, (1f)

- $T(x, N(y)) \leq N(I(x, y))$ (consequence of (1e) when $z = 0$), (1g)

- $N(T(x, y)) = I(x, N(y))$ (consequence of (1f) when $z = 0$). (1h)

Two fuzzy binary operators B^1 and B^2 are isomorphic if there exists a bijection $\varphi : [0, 1] \rightarrow [0, 1]$ such that $B^1(x, y) = \varphi^{-1}(B^2(\varphi(x), \varphi(y)))$ while unary operators V^1 and V^2 are isomorphic if $V^1(x) = \varphi^{-1}(V^2(\varphi(x)))$. Moreover, we write $B^1 \equiv B_\varphi^2$ and $V^1 \equiv V_\varphi^2$.

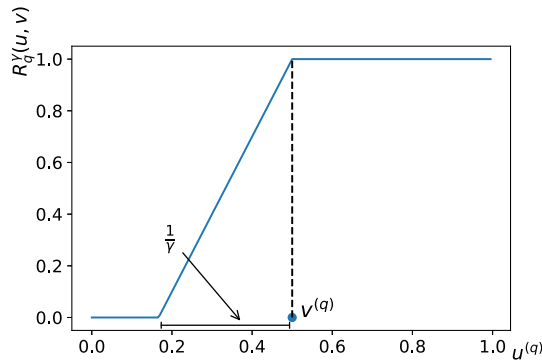


Fig. 1. Illustration of the T -preorder relation on criterion q for pair of instances (u, v) .

If residual triplet (T, I, N) is generated by t -norm T , then the residual triplet generated by T_φ is $(T_\varphi, I_\varphi, N_\varphi)$.

A t -norm for which the induced negator of its R-implicator is involutive is called an IMTL t -norm. In Table 1, T_L and T_{nM} are IMTL t -norms where the corresponding induced negator is N_s . A residual triplet (T, I, N) that is generated with an IMTL t -norm is called an IMTL triplet. If (T, I, N) is an IMTL triplet, then $(T_\varphi, I_\varphi, N_\varphi)$ is also an IMTL triplet.

For an IMTL triplet, we have that the following property holds for all $x, y \in [0, 1]$:

- $I(N(x), N(y)) = I(y, x)$ (2a)

- $T(x, N(y)) = N(I(x, y))$ (2b)

A continuous t -norm is IMTL if and only if it is isomorphic to the Łukasiewicz t -norm. Such t -norm and the corresponding R-implicator I_T are *strongly max-definable*, i.e., for all $x, y \in [0, 1]$, it holds that

$$\max(x, y) = I_T(I_T(x, y), y) = I_T(I_T(y, x), x). \tag{3}$$

A residual triplet generated by a t -norm isomorphic to T_L , $T_{L,\varphi}$ is denoted by $(T_{L,\varphi}, I_{L,\varphi}, N_{L,\varphi})$. Note that $N_L \equiv N_s$.

2.2. Fuzzy sets and fuzzy relations

Given a non-empty universe set U , a fuzzy set A on U is an ordered pair (U, m_A) , where $m_A : U \rightarrow [0, 1]$ is a membership function that indicates how much an element from U is contained in A . Instead of $m_A(u)$, the membership degree is often written as $A(u)$. If the image of m_A is $\{0, 1\}$ then A is a crisp (ordinary) set. For negator N , the fuzzy complement coA is defined as $coA(u) = N(A(u))$ for $u \in U$. If A is crisp then coA reduces to the standard complement. For $\alpha \in (0, 1]$, the α -level set of fuzzy set A is a crisp set defined as $A_\alpha = \{u \in U; A(u) \geq \alpha\}$.

A fuzzy relation \tilde{R} on U is a fuzzy set on $U \times U$, i.e., a mapping $\tilde{R} : U \times U \rightarrow [0, 1]$ which indicates how much two elements from U are related. Some relevant properties of fuzzy relations include:

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

A reflexive and T -transitive fuzzy relation is called a T -preorder relation while a symmetric T -preorder is a T -equivalence relation.

To illustrate some of these fuzzy relations, we assume that instances from U are described with a finite set of numerical attributes Q . For attribute $q \in Q$. Let $u^{(q)}$ and $v^{(q)}$ be the evaluations of instances u and v on attribute q . An example of a T_L -preorder relation (expressing dominance) on attribute q , given in [21], is

$$\tilde{R}_q^\gamma(u, v) = \max\left(\min\left(1 - \gamma \frac{v^{(q)} - u^{(q)}}{range(q)}, 1\right), 0\right), \tag{4}$$

where γ is a positive parameter and $range(q)$ is the difference between the maximal and minimal value on q . The illustration of such relation is given in Fig. 1. In the figure, we have attribute q with range equal to 1. The value of $v^{(q)}$ is 0.5, while the value of $u^{(q)}$ goes from 0 to 1. For every $u^{(q)}$ and for $\gamma = 3$, the value of $\tilde{R}_q^\gamma(u, v)$ is calculated and depicted. An example of a T_L -equivalence relation (expressing indiscernibility) on the same attribute is

$$\tilde{R}_q^\gamma(u, v) = \max\left(1 - \gamma \frac{|u^{(q)} - v^{(q)}|}{range(q)}, 0\right), \tag{5}$$

while the illustration of such relation is given in Fig. 2. The description of the figure is same as for Fig. 1. In both cases, the relation over all attributes from Q is defined as $\tilde{R}(u, v) = \min_{q \in Q} \tilde{R}_q^\gamma(u, v)$.

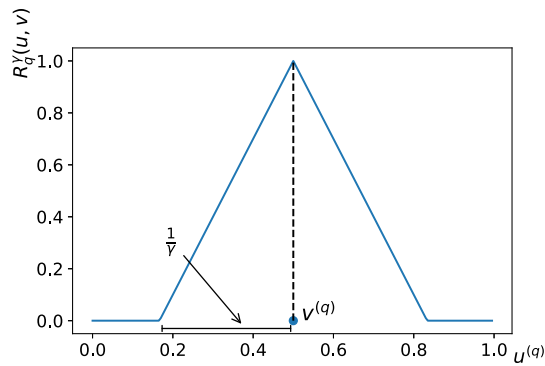


Fig. 2. Illustration of the T -equivalence relation on criterion q for pair of instances (u, v) .

2.3. Fuzzy rough and granularly representable sets

This subsection is based on [9]. Let U be a finite set of instances, A a fuzzy set on U and \tilde{R} a T -preorder relation on U . The fuzzy PRSA lower and upper approximations of A are fuzzy sets for which the membership function is defined as:

$$\begin{aligned} \underline{\text{apr}}_R^{\text{min},I}(A)(u) &= \min\{I(\tilde{R}(v, u), A(v)); v \in U\} \\ \overline{\text{apr}}_R^{\text{max},T}(A)(u) &= \max\{T(\tilde{R}(u, v), A(v)); v \in U\}. \end{aligned} \tag{6}$$

For an IMTL t -norm and the corresponding implicator I and negator N , we have the well known duality property.

$$\begin{aligned} N(\underline{\text{apr}}_R^{\text{min},I}(A)(u)) &= \overline{\text{apr}}_R^{\text{max},T}(coA)(u) \\ N(\overline{\text{apr}}_R^{\text{max},T}(A)(u)) &= \underline{\text{apr}}_R^{\text{min},I}(coA)(u) \end{aligned} \tag{7}$$

A fuzzy granule with respect to fuzzy relation \tilde{R} , parameter $\lambda \in [0, 1]$ and instance u is defined as a parametric fuzzy set

$$\tilde{R}_\lambda^+(u) = \{(v, T(\tilde{R}(v, u), \lambda)); v \in U\}, \tag{8}$$

while the granule with respect to the inverse fuzzy relation \tilde{R}^{-1} is

$$\tilde{R}_\lambda^-(u) = \{(v, T(\tilde{R}(u, v), \lambda)); v \in U\}. \tag{9}$$

Here, parameter λ describes the association of instance u to a particular decision. For example, in classification problems, it represents the membership degree of u to a particular decision class. A fuzzy set A is granularly representable (GR) w.r.t. relation \tilde{R} if

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

where the union is defined using max operator. Some equivalent forms to define granular representability are such that for all $u, v \in U$:

$$T(\tilde{R}(v, u), A(u)) \leq A(v) \Leftrightarrow \tilde{R}(v, u) \leq I(A(u), A(v)) \tag{10}$$

Proposition 2.1. [10] If fuzzy set A is granularly representable w.r.t. T -preorder relation \tilde{R} , then coA is granularly representable w.r.t. \tilde{R}^{-1} .

Proposition 2.2. [9] It holds that $\underline{\text{apr}}_R^{\text{min},I}(A)$ is the largest GR set contained in A , while $\overline{\text{apr}}_R^{\text{max},T}(A)$ is the smallest GR set containing A .

2.4. Granular approximations

This subsection is based on [10]. A granular approximation is a granularly representable set that is as close as possible to the observed fuzzy set (set that is approximated) with respect to the given closeness criterion. The closeness is measured by a loss function $L : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}^+$. For a given loss function L , fuzzy set A , relation \tilde{R} and residual triplet (T, I, N) , the granular approximation \hat{A} is obtained as a result of the following optimization problem:

$$\begin{aligned} \text{minimize} \quad & \sum_{u \in U} L(A(u), \hat{A}(u)) \\ \text{subject to} \quad & T(\tilde{R}(u, v), \hat{A}(v)) \leq \hat{A}(u), \quad u, v \in U \\ & 0 \leq \hat{A}(u) \leq 1, \quad u \in U. \end{aligned} \tag{11}$$

The objective function in (11) ensures that the resulting fuzzy set \hat{A} is as close as possible to the given fuzzy set A (w.r.t. loss function L) while the constraints of (11) guarantee that \hat{A} is granularly representable. We recall two well-known loss functions; p -quantile loss:

$$L_p(y, \hat{y}) = (y - \hat{y})(p - \mathbf{1}_{y-\hat{y}<0}) = \begin{cases} p|y - \hat{y}| & \text{if } y - \hat{y} > 0, \\ (1 - p)|y - \hat{y}| & \text{otherwise,} \end{cases} \tag{12}$$

where $p \in [0, 1]$ is a probability parameter, and squared error loss:

$$L_{SEL}(y, \hat{y}) = (y - \hat{y})^2. \tag{13}$$

The p -quantile loss for $p = \frac{1}{2}$ is called absolute error loss:

$$L_{AEL}(y, \hat{y}) = |y - \hat{y}| \tag{14}$$

It was shown that optimization problem (11) can be efficiently solved if T is isomorphic to T_L ($T = T_{L,\varphi}$) and for L being the scaled p -quantile loss: $L_{p,\varphi} = L_p(\varphi(y), \varphi(\hat{y}))$ or the scaled squared error loss: $L_{SEL,\varphi} = L_{SEL}(\varphi(y), \varphi(\hat{y}))$.

Definition 2.1. Loss function L is *symmetric* if $L(y, \hat{y}) = L(\hat{y}, y)$.

It is easy to verify that $L_{SEL,\varphi}$ and $L_{AEL,\varphi}$ are symmetric loss functions, while $L_{p,\varphi}$ for $p \neq \frac{1}{2}$ is not. However, it can be observed that $L_{p,\varphi}(y, \hat{y}) = L_{1-p,\varphi}(\hat{y}, y)$.

Definition 2.2. We say that loss function L is of \vee -type if for any real number a , it holds that

- $L(a, a) = 0$,
- functions $L(x, a)$ and $L(a, x)$ are increasing for $x > a$ and
- functions $L(x, a)$ and $L(a, x)$ are decreasing for $x < a$.

The previous definition says that the loss is greater if x is more distant from a . It is easy to verify that the squared error loss and p -quantile loss for $p \in (0, 1)$ are of \vee -type. The p -quantile loss for $p \in \{0, 1\}$ is not of \vee -type since $L_0(a, x) = 0$ for $x < a$ and $L_1(a, x) = 0$ for $x > a$.

Definition 2.3. A loss function $L : [0, 1] \times [0, 1] \rightarrow \mathbb{R}^+$ is *N -duality preserving* if $L(y, \hat{y}) = L(N(\hat{y}), N(y))$ for N from the residual triplet (T, I, N) .

In [10], it was shown that both $L_{p,\varphi}$ and $L_{SEL,\varphi}$ are N -duality preserving for IMTL triplet $(T_{L,\varphi}, I_{L,\varphi}, N_{L,\varphi})$.

3. Disjoint and adjacent granules

3.1. Definitions and basic properties

Note that from now on we assume that \tilde{R} is a T -preorder relation for a residual triplet (T, I, N) .

Definition 3.1. Two fuzzy sets A and B , defined on universe U , are called T -disjoint if

$$T(A(u), B(u)) = 0 \text{ for every } u \in U.$$

For the fuzzy granules, we have the following property:

Proposition 3.1. Let $u, v \in U$. Two fuzzy granules $\tilde{R}_{\lambda_1}^+(u)$ and $\tilde{R}_{\lambda_2}^-(v)$ are T -disjoint if and only if

$$T(\lambda_1, \lambda_2) \leq N(\tilde{R}(v, u)). \tag{15}$$

Proof. The statement that two granules are T -disjoint is equivalent to:

$$\begin{aligned} & \max_{w \in U} T(\tilde{R}(w, u), \lambda_1), T(\tilde{R}(v, w), \lambda_2) = 0 \\ \Leftrightarrow & \max_{w \in U} T(T(\tilde{R}(v, w), \tilde{R}(w, u)), T(\lambda_1, \lambda_2)) = 0 \\ \Leftrightarrow & T\left(\max_{w \in U} T(\tilde{R}(v, w), \tilde{R}(w, u)), T(\lambda_1, \lambda_2)\right) = 0 \\ \Leftrightarrow & T(\tilde{R}(v, u), T(\lambda_1, \lambda_2)) = 0 \end{aligned}$$

$$\begin{aligned} &\Leftrightarrow T(\lambda_1, \lambda_2) \leq I(\tilde{R}(v, u), 0) \\ &\Leftrightarrow T(\lambda_1, \lambda_2) \leq N(\tilde{R}(v, u)). \end{aligned}$$

The first equivalence holds because of the commutativity and associativity of T . The second one holds because T is left-continuous. The third one is a consequence of the T -transitivity of \tilde{R} while the fourth equivalence follows from the residuation property. \square

Please note that the T -disjointness is characterised by the above proposition only for granules of opposite types, i.e., granules w.r.t. relations \tilde{R} and \tilde{R}^{-1} respectively.

Proposition 3.2. *Let (T, I, N) be an IMTL triplet. Then, fuzzy set A is granularly representable w.r.t. \tilde{R} if and only if the granules from A (w.r.t. \tilde{R}) and coA (w.r.t. \tilde{R}^{-1}) are disjoint.*

Proof. We have the following equivalences.

$$\begin{aligned} A(u) \geq T(\tilde{R}(u, v), A(v)) &\Leftrightarrow \tilde{R}(u, v) \leq I(A(v), A(u)) \\ &\Leftrightarrow N(\tilde{R}(u, v)) \geq N(I(A(v), A(u))) \\ &\Leftrightarrow N(\tilde{R}(u, v)) \geq T(A(v), N(A(u))). \end{aligned}$$

The last equivalence holds because of expression (2b) (after negator N is applied on both sides), while the second one follows from the fact that N is a non-increasing function. The equivalences state that the granular representability of A is equivalent to the T -disjointness condition of granules $\tilde{R}_{A(v)}^+(v)$ from A and $\tilde{R}_{coA(u)}^-(u)$ from coA , as formulated in Proposition 3.1. \square

Corollary 3.1. *Let (T, I, N) be an IMTL triplet. The granules from $\underline{apr}_R^{\min, I}(A)$ and $\overline{apr}_R^{\max, T}(coA)$ are disjoint (analogously, the granules from $\underline{apr}_R^{\min, I}(coA)$ and $\overline{apr}_R^{\max, T}(A)$ are disjoint too).*

Proof. The result holds from the duality property of the lower and upper approximations (7). \square

Next, we examine a pair of granules $\tilde{R}_{\lambda_1}^+$ and $\tilde{R}_{\lambda_2}^-$ that are not only disjoint, but are *adjacent* to each other in the following sense. If their parameters are λ_1 and λ_2 , then adding any ϵ to either λ_1 or λ_2 will cause the granules to overlap. For fixed λ_1 , the largest λ_2 for which the granules are still disjoint is:

$$\lambda_2^{\max} = \sup\{\lambda; T(\lambda_1, \lambda) \leq N(\tilde{R}(v, u))\} = I(\lambda_1, N(\tilde{R}(v, u))).$$

Obviously,

$$T(\lambda_1, \lambda_2^{\max}) = T(\lambda_1, I(\lambda_1, N(\tilde{R}(v, u)))) \leq N(\tilde{R}(v, u))$$

due to the modus ponens property (1c). The previous discussion is formalized in the following definition.

Definition 3.2. Granule $\tilde{R}_{\lambda_2}^-(v)$ is adjacent to granule $\tilde{R}_{\lambda_1}^+(u)$ if

$$\lambda_1 = I(\lambda_2, N(\tilde{R}(v, u))),$$

while $\tilde{R}_{\lambda_1}^+(v)$ is adjacent to $\tilde{R}_{\lambda_2}^-(u)$ if

$$\lambda_2 = I(\lambda_1, N(\tilde{R}(u, v))).$$

We call such defined relationship among granules the adjacency relation.

Proposition 3.3. *Every granule is adjacent to all granules with parameter 1, under the assumption that they are disjoint.*

Proof. If $\lambda_1 = 1$, from the T -disjointness property we have:

$$T(1, \lambda_2) \leq N(\tilde{R}(v, u)) \Leftrightarrow 1 \leq I(\lambda_2, N(\tilde{R}(v, u))) \Rightarrow 1 = I(\lambda_2, N(\tilde{R}(v, u))). \quad \square$$

From the proof of Proposition 3.3, we may conclude that granule $\tilde{R}_{\lambda_1}^+(u)$ for $\lambda_1 = 1$ is adjacent to $\tilde{R}_{\lambda_2}^-(v)$ if and only if $\lambda_2 = N(\tilde{R}(v, u))$.

The previous reasoning also reveals that the adjacency relation is not necessarily symmetric.

Proposition 3.4. *For parameters λ_1 and λ_2 that are smaller than 1 and for continuous t -norm T from the IMTL triplet (T, I, N) , we have that the adjacency relation is symmetric. In other words, if $\lambda_1 = I(\lambda_2, N(\tilde{R}(v, u)))$, then also $\lambda_2 = I(\lambda_1, N(\tilde{R}(v, u)))$.*

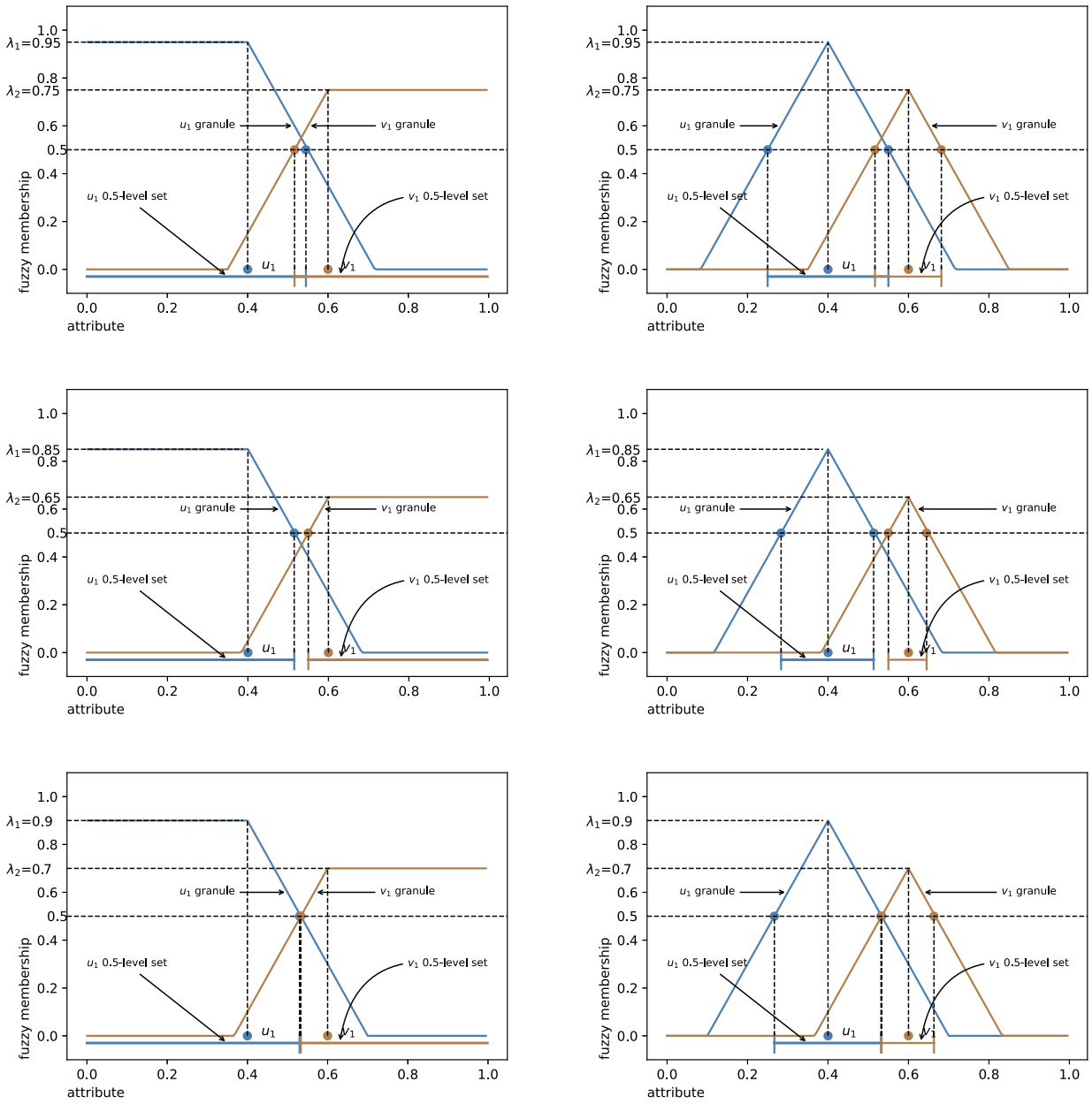


Fig. 3. Granules in one dimension.

Proof. Ordering property (1d) implies that if $\lambda_1 < 1$, then also $\lambda_2 > N(\tilde{R}(v, u))$. Using the strong max-definability (3), we have that

$$\lambda_2 = I(I(\lambda_2, N(\tilde{R}(v, u))), N(\tilde{R}(v, u))) = I(\lambda_1, N(\tilde{R}(v, u))). \quad \square$$

Example 3.1. Figs. 3 and 4 illustrate different relationships between granules, in one and two dimensions respectively. In Fig. 3, instances are represented using one condition attribute q whose range is 1. There are two instances u_1 and v_1 with respective attribute values 0.4 and 0.6. Their granules $R_{\lambda_1}^-(u_1)$ and $R_{\lambda_2}^-(v_1)$ are formed based on a T -preorder relation (left side of the figure) and T -equivalence relation (right side of the figure), parameter value $\gamma = 3$ and the Łukasiewicz t -norm. We vary parameters λ_1 and λ_2 in order to represent different relationship among two granules. We depict the fuzzy granules together with their 0.5-level sets. In the upper two images, the values of parameters are $\lambda_1 = 0.95$ and $\lambda_2 = 0.75$ which leads to overlapping granules (i.e., they are not T -disjoint). In the two images in the middle, the values of parameters are $\lambda_1 = 0.85$ and $\lambda_2 = 0.65$ which leads to T -disjoint granules, while in the lower two images, the values of parameters are $\lambda_1 = 0.9$ and $\lambda_2 = 0.7$ which leads to adjacent granules (here, the adjacency relation is symmetric). It is easy to verify that in this case, the 0.5-level sets follow the relation between granules, i.e.,

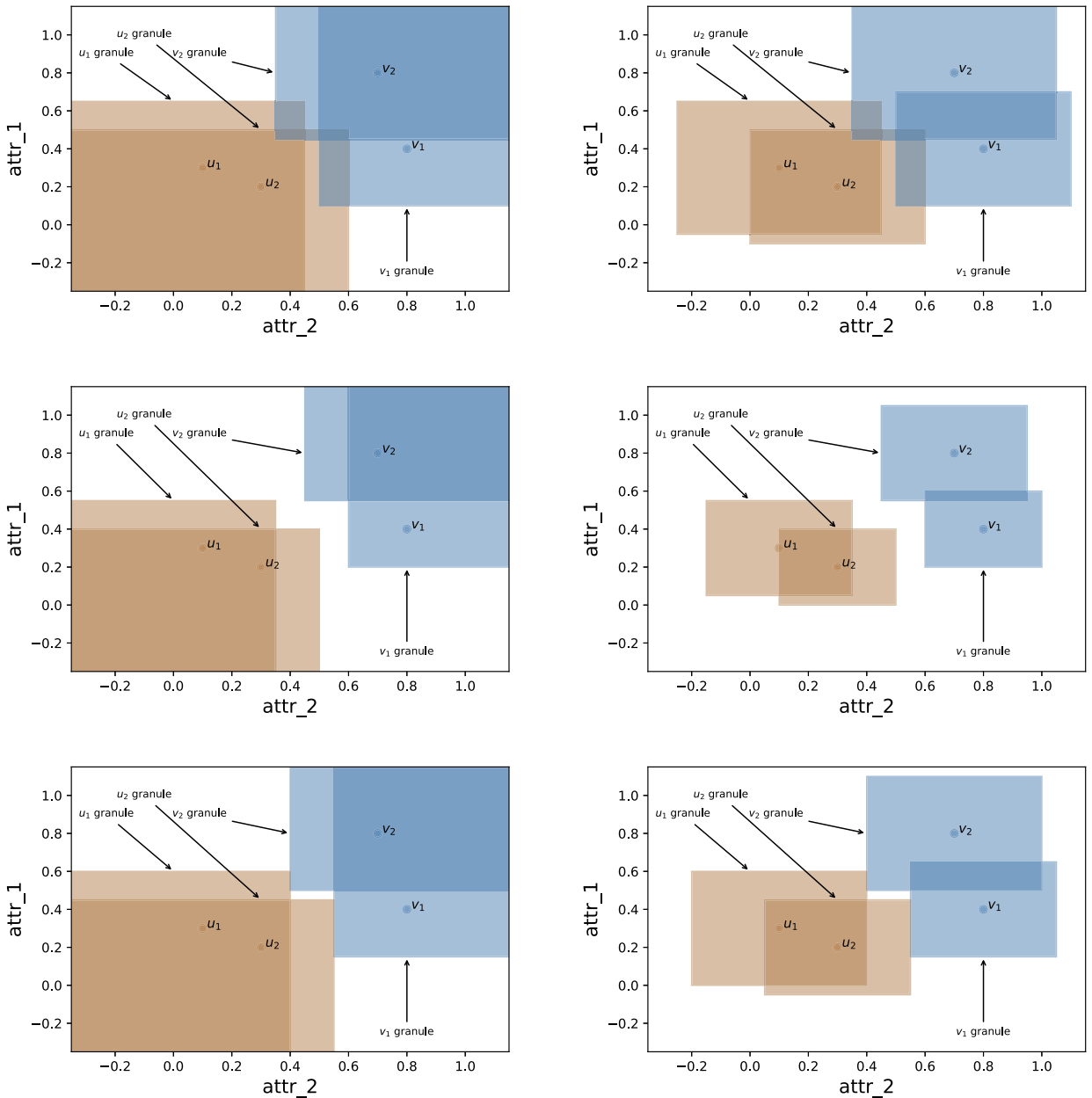


Fig. 4. Granules in two dimensions.

if the granules overlap, then the level sets overlap, if the granules are T -disjoint, then the level sets are disjoint and if the granules are adjacent, then the level sets have one common point.

The use of 0.5-level sets is in particular useful to visualize the granules in the case of two dimensions. In Fig. 4, we have four instances from two classes, described by two condition attributes; u_1 and u_2 are from one class and v_1 and v_2 are from the other one. We illustrate the relationship of granules from different classes. The granules are formed using a T -preorder relation (left side of the figure) and T -equivalence relation (right side of the figure). While the granules in one dimension had triangular (T -equivalence) or “half-triangular” shape (T -preorder), in two dimensions they have pyramidal (T -equivalence) or “half-pyramidal” (T -preorder) shape. However, for the purpose of the visualisation in 2 dimensions of the 3-dimensional granules, we use 0.5-level sets of the granules. The level sets for the T -preorder relation take the form of quarter-plane (left images) and the rectangular form (right images) for the T -equivalence relation. Again, we can distinguish granules from different classes that overlap (upper two images), that are disjoint (two images in the middle) and that are adjacent (lower two images). In this case, the level sets of adjacent granules share an edge.

In Proposition 3.2 we showed that granules from A and coA are T -disjoint for granularly representable set A . Now we examine in which cases some of them are also adjacent. If $\tilde{R}_{coA(v)}^-(v)$ is adjacent to $\tilde{R}_{A(u)}^+(u)$ for $u, v \in U$, we have that

$$A(u) = I(N(A(v)), N(\tilde{R}(v, u))) \Leftrightarrow A(u) = I(\tilde{R}(v, u), A(v)), \tag{16}$$

where the equivalence holds because of (2a). If $\tilde{R}_{A(v)}^+(v)$ is adjacent to $\tilde{R}_{coA(u)}^-(u)$ then

$$\begin{aligned} N(A(u)) = I(A(v), N(\tilde{R}(u, v))) &\Leftrightarrow A(u) = N(I(A(v), N(\tilde{R}(u, v)))) \\ &\Leftrightarrow A(u) = T(\tilde{R}(u, v), A(v)), \end{aligned} \tag{17}$$

where the second equivalence holds because of (1h).

3.2. Application to granular approximations

The next lemma, theorem, and corollary investigate the adjacency relationship of granules from the granular approximations, i.e., from the solutions of optimization problem (11).

Lemma 3.1. *Let loss function L be of \vee -type and let \hat{A} be a solution of optimization problem (11). If $\hat{A}(u) > A(u)$, then it holds that*

$$\hat{A}(u) = \max\{T(\tilde{R}(u, v), \hat{A}(v)); v \in U, v \neq u\},$$

while if $\hat{A}(u) < A(u)$, then it holds that

$$\hat{A}(u) = \min\{I(\tilde{R}(v, u), \hat{A}(v)); v \in U, v \neq u\}.$$

Proof. Let $\alpha_u = \max\{T(\tilde{R}(u, v), \hat{A}(v)); v \in U, v \neq u\}$ for some $u, \hat{A}(u) > A(u)$. If the first condition of the theorem is not satisfied, then from the granular representability it holds that $\alpha_u < \hat{A}(u)$. Replacing $\hat{A}(u)$ with $\max(\alpha_u, A(u))$ leads to a solution that is also granularly representable (easily verifiable) and that ensures a smaller value of the objective function since the loss function L is of \vee -type. That is a contradiction.

Now, let $\alpha_u = \min\{I(\tilde{R}(v, u), \hat{A}(v)); v \in U, v \neq u\}$ for some $u, \hat{A}(u) < A(u)$. If the second condition of the theorem is not satisfied, then from the granular representability it holds that $\alpha_u > \hat{A}(u)$. Replacing $\hat{A}(u)$ with $\min(\alpha_u, A(u))$ leads to a solution that is also granularly representable (easily verifiable) and that ensures a smaller value of the objective function since the loss function L is of \vee -type. That is a contradiction. \square

Theorem 3.1. *Let loss function L be of \vee -type and let \hat{A} be a solution of optimization problem (11). We define three sets:*

- $U^- = \{u \in U; \hat{A}(u) < A(u)\},$
- $U^0 = \{u \in U; \hat{A}(u) = A(u)\},$
- $U^+ = \{u \in U; \hat{A}(u) > A(u)\}.$

It holds that

$$\hat{A}(u) = \max\{T(\tilde{R}(u, v), \hat{A}(v)); v \in U^- \cup U^0\}, \tag{18}$$

for $u \in U^+$ and

$$\hat{A}(u) = \min\{I(\tilde{R}(v, u), \hat{A}(v)); v \in U^+ \cup U^0\}, \tag{19}$$

for $u \in U^-$.

Proof. Condition (18) can be reformulated as

$$\forall u \in U^+, \exists v \in U^- \cup U_0; \hat{A}(u) = T(\tilde{R}(u, v), \hat{A}(v)).$$

Let $U_1^+ \subseteq U^+$ be a set of instances which satisfy the previous condition and let $U_2^+ = U^+ - U_1^+$. Let $u^* \in U_2^+$ be an instance with the largest $\hat{A}(u)$. From Lemma 3.1, there is $v \in U, v \neq u^*$ such that $\hat{A}(u^*) = T(\tilde{R}(u^*, v), \hat{A}(v))$.

By the assumption $u^* \in U_2^+$, we have that $v \in U^+$. If $v \in U_1^+$, then there is $w \in U^- \cup U^0$ such that $\hat{A}(v) = T(\tilde{R}(v, w), \hat{A}(w))$. We have that

$$\begin{aligned} \hat{A}(u^*) &= T(\tilde{R}(u^*, v), \hat{A}(v)) \\ &= T(\tilde{R}(u^*, v), T(\tilde{R}(v, w), \hat{A}(w))) \\ &= T(T(\tilde{R}(u^*, v), \tilde{R}(v, w)), \hat{A}(w)) \\ &\leq T(\tilde{R}(u^*, w), \hat{A}(w)). \end{aligned}$$

The last inequality holds because of the T -transitivity of \tilde{R} and the monotonicity of T . The opposite inequality holds from the granular representability which leads to the conclusion that $\hat{A}(u^*) = T(\tilde{R}(u^*, w), \hat{A}(w))$ which contradicts the assumption that $u^* \in U_2^+$. Hence, $v \in U_2^+$.

From $\hat{A}(u^*) = T(\tilde{R}(u^*, v), \hat{A}(v))$, it holds $\hat{A}(v) \geq \hat{A}(u^*)$ due to (1a). Since $A(u^*)$ is the largest in U_2^+ by the assumption, then $\hat{A}(u^*) = \hat{A}(v)$. Denote with $U_3^+ \subseteq U_2^+$ instances from U_2^+ for which the membership degree in \hat{A} is $\hat{A}(u^*)$. Every pair of instances from U_3^+ satisfies (10) since they have the same membership value in \hat{A} . Due to maximality of $\hat{A}(u)$ it holds that for $u \in U_3^+$ and for $v \in U - U_3^+$, it holds that $\hat{A}(u) > T(\tilde{R}(u, v), \hat{A}(v))$. Denote

$$\beta^+ = \max(\max\{A(u); u \in U_3^+\}, \max\{\max\{T(\tilde{R}(u, v), \hat{A}(v)); v \in U - U_3^+\}; u \in U_3^+\}).$$

From the assumptions above, it holds that $\beta^+ < \hat{A}(u^*)$ which implies $\beta^+ < \hat{A}(u)$ for $u \in U_3^+$. Now, let \hat{A}^* be a fuzzy set where values $\hat{A}(u)$ for $u \in U_3^+$ are replaced with β^+ . We observe that \hat{A}^* is granularly representable since $\hat{A}^*(u)$ are pairwise equal for $u \in U_3^+$ and also for every $u \in U_3^+$ and for every $v \in U - U_3^+$ it holds that $T(\tilde{R}(u, v), \hat{A}^*(v)) \leq \hat{A}^*(u)$ by the definition of β^+ . Next, we observe that the objective value with \hat{A}^* is smaller than with \hat{A} because $A(u) < \hat{A}^*(u) < \hat{A}(u)$ for $u \in U_3^+$ and due to the fact that L is of \vee -type.

Therefore, we obtained a feasible solution with a smaller objective function, which contradicts the optimality of \hat{A} . This contradiction implies that U_2^+ must be empty which is equivalent to (18).

On the other hand, condition (19) can be reformulated as

$$\forall u \in U^-, \exists v \in U^+ \cup U_0; \hat{A}(u) = I(\tilde{R}(v, u), \hat{A}(v)).$$

Let $U_1^- \subseteq U^-$ be a set of instances which satisfy the previous condition and let $U_2^- = U^- - U_1^-$. Let $u^* \in U_2^-$ be an instance with the smallest $\hat{A}(u)$. From Lemma 3.1, there is $v \in U, v \neq u^*$ such that $\hat{A}(u^*) = I(\tilde{R}(v, u), \hat{A}(v))$.

By the assumption $u^* \in U_2^-$, it holds that $v \in U^-$. Assume that $v \in U_1^-$. Then, there is $w \in U^+ \cap U^0$ such that $\hat{A}(v) = I(\tilde{R}(w, v), \hat{A}(w))$. We have that

$$\begin{aligned} A(u^*) &= I(\tilde{R}(v, u^*), \hat{A}(v)) \\ &= I(\tilde{R}(v, u^*), I(\tilde{R}(w, v), \hat{A}(w))) \\ &= I(T(\tilde{R}(v, u^*), \tilde{R}(w, v)), \hat{A}(w)) \\ &\geq I(\tilde{R}(w, u^*), \hat{A}(w)). \end{aligned}$$

The last equality holds because of (1f). The last inequality holds because of the T -transitivity of \tilde{R} and the fact that I is non-increasing in its first argument. The opposite inequality holds from the granular representability which leads to the conclusion that $A(u^*) = I(\tilde{R}(w, u^*), \hat{A}(w))$ which contradicts the assumption that $u^* \in U_2^-$. Hence, $v \in U_2^-$.

From $\hat{A}(u^*) = I(\tilde{R}(v, u^*), \hat{A}(v))$, it holds that $\hat{A}(v) \leq A(u^*)$ due to (1b). Since $\hat{A}(u^*)$ is the smallest by the assumption, then $\hat{A}(u^*) = \hat{A}(v)$. Denote with $U_3^- \subseteq U_2^-$ instances from U_2^- that have value $\hat{A}(u^*)$. Every pair of instances from U_3^- satisfy (10) since they have the same membership degree in \hat{A} . For every $u \in U_3^-$ and for every $v \in U - U_3^-$ it holds that $\hat{A}(u) < I(\tilde{R}(v, u), \hat{A}(v))$. Denote

$$\beta^- = \min(\min\{A(u); u \in U_3^-\}, \min\{\min\{I(\tilde{R}(v, u), \hat{A}(v)); v \in U - U_3^-\}; u \in U_3^-\}).$$

From the above assumption, it holds that $\beta^- > \hat{A}(u^*)$, which implies $\beta^- > \hat{A}(u)$ for $u \in U_3^-$. Now, let \hat{A}^{**} be a fuzzy set where values $\hat{A}(u)$ for $u \in U_3^-$ are replaced with β^- . We observe that \hat{A}^{**} is granularly representable since $\hat{A}^{**}(u)$ are pairwise equal for $u \in U_3^-$ and also for every $u \in U_3^-$ and for every $v \in U - U_3^-$ it holds that $I(\tilde{R}(v, u), \hat{A}^{**}(v)) \leq \hat{A}^{**}(u)$ by the definition of β^- . Next, we observe that the objective value with \hat{A}^{**} is smaller than with \hat{A} because $A(u) > \hat{A}^{**}(u) > \hat{A}(u)$ for $u \in U_3^-$ and due to the fact that L is of \vee -type.

Therefore, we obtained a feasible solution with a smaller objective function, which contradicts the optimality of \hat{A} . This contradiction implies that U_2^- must be empty, which is equivalent to (19). \square

Corollary 3.2. Let loss function L be of \vee -type and let \hat{A} be a solution of optimization problem (11) defined w.r.t. an IMTL triplet (T, I, N) . Let U^-, U^0, U^+ be defined as in Theorem 3.1. Then, the following holds.

- For all $u \in U^+$, there is $v \in U^- \cup U_0$ such that $R_{\hat{A}(v)}^+(v)$ is adjacent to $R_{co\hat{A}(u)}^-(u)$.
- For all $u \in U^-$, there is $v \in U^+ \cup U_0$ such that $R_{co\hat{A}(v)}^-(v)$ is adjacent to $R_{\hat{A}(u)}^+(u)$.

Proof. The corollary is a direct consequence of Theorem 3.1 and equations (16) and (17). \square

Note that Theorem 3.1 does not require for residual triplet (T, I, N) to be an IMTL triplet, hence it can lead to more general results that are not related to the granular adjacency relationships.

4. Case of a classification problem

First, we consider a binary classification problem, i.e., we distinguish two classes in U : A and coA which are now crisp (ordinary) sets. Notations A and coA will be also used for the fuzzy sets that encode the corresponding decision class, i.e., $A(u) = 1$ if $u \in A$ while $A(u) = 0$ if $u \in coA$.

Proposition 4.1. *Let loss function L be of \vee -type and let \hat{A} be a granular approximation of a crisp set A w.r.t. an IMTL triplet (T, I, N) . Then, the following holds.*

- For all $u \in A$, there is $v \in coA$ such that $R_{co\hat{A}(v)}^-$ is adjacent to $R_{\hat{A}(u)}^+$.
- For all $u \in coA$, there is $v \in A$ such that $R_{\hat{A}(v)}^+$ is adjacent to $R_{co\hat{A}(u)}^-$.

Proof. Let U^+, U^0, U^- be the sets defined in Theorem 3.1. Obviously, it holds that $U^- \subseteq A$ and $U^+ \subseteq coA$. We prove the first part of the proposition, while the second part holds by analogy.

Let $u \in A$. If $u \in A - U^- \Leftrightarrow \hat{A}(u) = 1$, then from Proposition 3.3 we have that for all $v \in coA$, $R_{co\hat{A}(v)}^-$ is adjacent to $R_{\hat{A}(u)}^+$. If $u \in U^-$, then from Corollary 3.2, there is $v \in U^0 \cup U^+$ such that $R_{co\hat{A}(v)}^-$ is adjacent to $R_{\hat{A}(u)}^+$. If $v \in U^+$, then also $v \in coA$ since $U^+ \subseteq coA$. If $v \in U^0$, then either $\hat{A}(v) = 0$ or $\hat{A}(v) = 1$. If $\hat{A}(v) = 1$ then the fact that $R_{co\hat{A}(v)}^-$ is adjacent to $R_{\hat{A}(u)}^+$ is equivalent to

$$\hat{A}(u) = I(N(\hat{A}(v)), N(\tilde{R}(v, u))) = I(0, N(\tilde{R}(v, u))) = 1,$$

which contradicts the assumption that $u \in U^-$. The last equality holds because of the ordering property (1d). Therefore, it holds that $\hat{A}(v) = 0$ which implies $v \in coA$. This proves the first part of the proposition. \square

For a solution \hat{A} of optimization problem (11), we have that $\hat{A}(u)$ for $u \in U$ represents the degree up to which u belongs to decision class A , while $co\hat{A}(u)$ represents the degree up to which u belongs to decision class coA . We may be interested to calculate only the degrees $\hat{A}(u)$ for $u \in A$ and degrees $co\hat{A}(u)$ for $u \in coA$.

Proposition 4.2. *Denote $\beta_u = \hat{A}(u)$ for $u \in A$ and $\beta_u = N(\hat{A}(u))$ for $u \in coA$. Let L be of \vee -type and N -duality preserving and symmetric. Then, in the classification case, problem (11) is equivalent to*

$$\begin{aligned} & \text{minimize} && \sum_{u \in U} L(1, \beta_u) \\ & \text{subject to} && T(\beta_u, \beta_v) \leq N(\tilde{R}(v, u)), \quad u \in A, v \in coA \\ & && 0 \leq \beta_u \leq 1, \quad u \in U. \end{aligned} \tag{20}$$

Proof. With the new notation and for L being N -duality preserving, the objective function of (11) becomes:

$$\sum_{u \in A} L(1, \beta_u) + \sum_{u \in coA} L(0, N(\beta_u)) = \sum_{u \in A} L(1, \beta_u) + \sum_{u \in coA} L(\beta_u, 1) = \sum_{u \in U} L(1, \beta_u).$$

The granularity constraints from (11) are now divided into 3 groups:

- Granularity constraints for pairs of instances $u, v \in A$:

$$\beta_u \geq T(\tilde{R}(u, v), \beta_v).$$

- Granularity constraints for pairs of instances $u, v \in coA$:

$$N(\beta_u) \geq T(\tilde{R}(u, v), N(\beta_v)) \Leftrightarrow \beta_v \geq T(\tilde{R}(v, u), \beta_u).$$

- Granularity constraints for pairs of instances $u \in A, v \in coA$. In that case, the granularity condition is expressed using T -disjointness (according to Proposition 3.2) as:

$$T(\beta_u, \beta_v) \leq N(\tilde{R}(v, u)).$$

The goal is to show that the first two groups of constraints are redundant. We first prove that the adjacency from Proposition 4.1 still holds in problem (20), i.e., for every $u \in A$, there is $v \in coA$ such that $\beta_u = I(\beta_v, N(\tilde{R}(v, u)))$ and that for all $v \in coA$, there is $u \in A$ such that $\beta_v = I(\beta_u, N(\tilde{R}(v, u)))$. Using the residuation property, we have that

$$T(\beta_u, \beta_v) \leq N(\tilde{R}(v, u)) \Leftrightarrow \beta_u \leq I(\beta_v, N(\tilde{R}(v, u))).$$

If for some u and for all v it holds that $\beta_u < I(\beta_v, N(\tilde{R}(v, u)))$, then there is $\epsilon > 0$ such that replacing β_u with $\beta_u + \epsilon$ leads to a smaller objective function since the loss function is of \vee -type. This leads to a contradiction with the assumption that β is optimal. Again, using the residuation property we have

$$T(\beta_u, \beta_v) \leq N(\tilde{R}(v, u)) \Leftrightarrow \beta_v \leq I(\beta_u, N(\tilde{R}(v, u))).$$

Using the same arguments as above, we get the second equality.

Next, we prove that the granularity criteria for β_u and β_v for $u, v \in A$ are satisfied. Let $w \in coA$ such that $\beta_u = I(\beta_w, N(\tilde{R}(w, u)))$. From the constraints, it holds that $\beta_v \leq I(\beta_w, N(\tilde{R}(w, v))) \Leftrightarrow \beta_w \leq I(\beta_v, N(\tilde{R}(w, v)))$. Then, we have that

$$\begin{aligned} \beta_u &= I(\beta_w, N(\tilde{R}(w, u))) \\ &\geq I(I(\beta_v, N(\tilde{R}(w, v))), N(\tilde{R}(w, u))) \\ &\geq T(\beta_v, I(N(\tilde{R}(w, v)), N(\tilde{R}(w, u)))) \\ &= T(\beta_v, I(\tilde{R}(w, u), \tilde{R}(w, v))) \\ &\geq T(\beta_v, \tilde{R}(u, v)), \end{aligned}$$

which is exactly the granularity condition for β_u and β_v . Now, let $u, v \in coA$ and let $w \in A$ be such that $\beta_u = I(\beta_w, N(\tilde{R}(u, w)))$. From the constraints, it holds that $\beta_w \leq I(\beta_v, N(\tilde{R}(v, w)))$. Using a similar reasoning as above, we conclude that the granularity condition is also satisfied for β_u and β_v when $u, v \in coA$. Since the granularity constraints for pairs of instances from the same class are a consequence of the T -disjointness constraints, they can be omitted in the optimization problem. \square

From now on, we assume that $\tilde{R}(u, v)$ is also a symmetric relation, i.e., it is a T -equivalence. In such case, the granules in A and coA are of the same type. We now consider crisp equivalence relation S on U defined as $S(u, v) = 1$ if u and v are from the same decision class, and $S(u, v) = 0$ otherwise. If u and v are from different decision classes then $I(\tilde{R}(u, v), S(u, v)) = N(\tilde{R}(u, v))$, while $I(\tilde{R}(u, v), S(u, v)) = 1$ otherwise. With relation S , the T -disjointness constraints from (20) may be reformulated as

$$T(\beta_u, \beta_v) \leq I(\tilde{R}(u, v), S(u, v)), \quad u, v \in U. \tag{21}$$

Here, we need to note that S , as an equivalence relation, can distinguish among more than two decision classes. In other words, S can be used to model multi-class classification problems. Bearing this in mind, a multi-class extension of problem (20) can be formulated as:

$$\begin{aligned} &\text{minimize} \quad \sum_{u \in U} L(1, \beta_u) \\ &\text{subject to} \quad T(\beta_u, \beta_v) \leq I(\tilde{R}(u, v), S(u, v)), \quad u, v \in U \\ &\quad \quad \quad 0 \leq \beta_u \leq 1, \quad u \in U. \end{aligned} \tag{22}$$

We name the result of problem (22) as *multi-class granular approximation*.

We need to stress that while the binary classification problem (20) with a T -preorder relation is suitable for the binary monotone classification problems, i.e., classification problems where there exists a monotone relationship between condition attributes and a decision attribute, the problem (22) with a T -equivalence relation is suitable for ordinary classification problems i.e., problems where such monotone relationship cannot be inferred.

5. Calculation

In this section, we use the notation: $M(u, v) = I(\tilde{R}(u, v), S(u, v))$. We start with an important property.

Proposition 5.1. *Problem (22) has a feasible solution.*

Proof. We construct a feasible solution. Let u_1, \dots, u_n be an ordering of instances from U . We apply the following procedure.

- 1) β_{u_1} is a random value from $[0, 1]$.
- 2) For $1 < i \leq n$, $\beta_{u_i} = \min\{I(\beta_{u_j}, M(u_j, u_i)); j < i\}$.

The adjacency property is obvious from the construction. We have to prove the granularity property. Let u_i and u_k be two instances for which $k < i$. Since $\beta_{u_i} = \min_{j < i} I(\beta_{u_j}, M(u_j, u_i))$, it holds that $\beta_{u_i} \leq I(\beta_{u_k}, M(u_i, u_k))$. From the residuation property, this is equivalent to $T(\beta_{u_i}, \beta_{u_k}) \leq M(u_i, u_k)$. \square

For different IMTL fuzzy connectives and for different loss functions L , problem (22) may take forms that cannot be efficiently solved in practice. However, we will consider the problem for L being a scaled form of absolute error loss or squared error loss (loss functions (14) and (13)), and T being isomorphic to the Łukasiewicz t -norm.

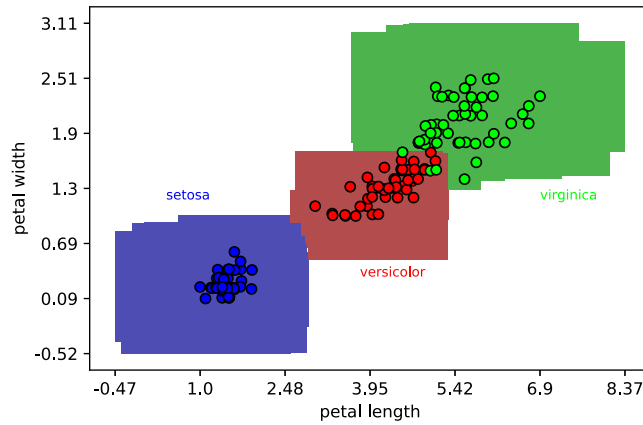


Fig. 5. An example of the multi-class granular approximation on iris dataset constructed with relation (5). (For interpretation of the colours in the figure(s), the reader is referred to the web version of this article.)

For such fuzzy connectives, the constraints of (20) are expressed as

$$\begin{aligned} \varphi^{-1}(\max(\varphi(\beta_u) + \varphi(\beta_v) - 1, 0)) &\leq M(u, v) \\ \Leftrightarrow \max(\varphi(\beta_u) + \varphi(\beta_v) - 1, 0) &\leq \varphi(M(u, v)) \\ \Leftrightarrow \varphi(\beta_u) + \varphi(\beta_v) &\leq 1 + \varphi(M(u, v)), \end{aligned}$$

for isomorphism φ and for $u, v \in U$. We introduce new variables $\forall u \in U, \alpha_u = \varphi(\beta_u)$ and $\forall u, v \in U, M_\varphi(u, v) = \varphi(M(u, v))$. With the new notations, the previous constraints may be expressed as

$$\alpha_u + \alpha_v \leq 1 + M_\varphi(u, v).$$

For the scaled absolute error loss $L_{AEL, \varphi}$, the objective function becomes

$$\sum_{u \in U} |\varphi(1) - \varphi(\beta_u)| = |U| - \sum_{u \in U} \alpha_u,$$

which leads to the optimization problem

$$\begin{aligned} &\text{maximize} \quad \sum_{u \in U} \alpha_u \\ &\text{subject to} \quad \alpha_u + \alpha_v \leq 1 + M_\varphi(u, v), \quad u, v \in U \\ &\quad \quad \quad 0 \leq \alpha_u \leq 1, \quad u \in U. \end{aligned} \tag{23}$$

Optimization problem (23) can be solved efficiently using linear programming techniques like the simplex method [22].

For the scaled squared error loss $L_{SEL, \varphi}$, the objective function becomes

$$\sum_{u \in U} (\varphi(1) - \varphi(\beta_u))^2 = \sum_{u \in U} (1 - \alpha_u)^2,$$

which leads to the optimization problem

$$\begin{aligned} &\text{maximize} \quad \sum_{u \in U} (1 - \alpha_u)^2 \\ &\text{subject to} \quad \alpha_u + \alpha_v \leq 1 + M_\varphi(u, v), \quad u, v \in U \\ &\quad \quad \quad 0 \leq \alpha_u \leq 1, \quad u \in U. \end{aligned} \tag{24}$$

Optimization problem (24) can be solved efficiently using quadratic programming techniques like the simplex method variation for quadratic programming [23].

Example 5.1. We may see another example in Fig. 5, where the instances come from the well-known iris dataset with, in this case, two features (petal length and petal width) and three classes (setosa, versicolor and virginica). The multi-class granular approximation is calculated by solving problem (24) for \tilde{R} being triangular similarity (5), and the granules are depicted using the obtained solution. On this figure, we can observe how granules look on the larger scale (in this case 150 instances).

We provide another example with more complex shapes of granules.

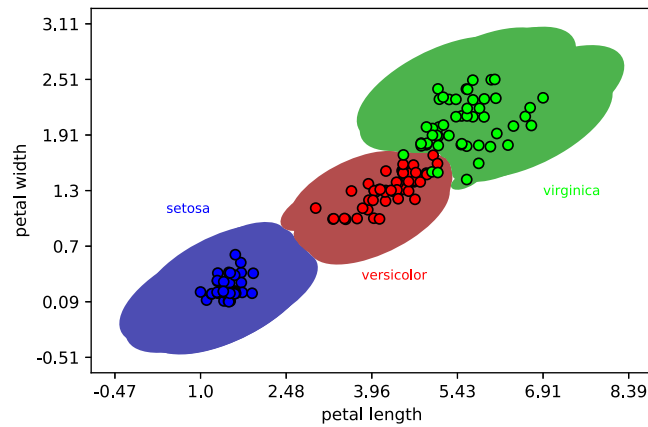


Fig. 6. An example of the multi-class granular approximation on iris dataset constructed with relation (25).

Example 5.2. Consider a family of fuzzy relations defined as

$$\tilde{R}(u, v) = \max \left(1 - \frac{d(u, v)}{a}, 0 \right), \tag{25}$$

where d is a metric (or distance function) on U and a is a positive real value (a parameter). It is easy to verify that such fuzzy relations are T_L -equivalences. More on the relationship between metrics and T -equality relations can be found in [24]. We now consider the Mahalanobis distance defined as [25]:

$$d(u, v)_{\Sigma} = \sqrt{(\mathbf{u} - \mathbf{v})^T \Sigma (\mathbf{u} - \mathbf{v})}, \tag{26}$$

where \mathbf{u} is a numerical vector representing condition attributes of instance u while Σ is a symmetric and positive-definite matrix. If Σ is an identity matrix, then $d(u, v)_{\Sigma}$ is equal to the Euclidean distance.

It is also easy to verify that the shape of level sets of granules, used to represent them in 2 dimensions, is, in the case of family (25), equal to the shape of equidistant points from the origin w.r.t. metric d . In the case of the Mahalanobis distance, the shape of granules will be elliptical. The axis of such ellipses is controlled by the eigenvalues of Σ while the rotation is controlled by the eigenvectors of Σ .

In Fig. 6, we present an example of granules from the multi-class granular approximation calculated by solving (24) and by using fuzzy relation (25) with d being the Mahalanobis distance. The approximation is calculated on the iris dataset with two attributes and three decision classes as described above. The granules have the elliptical shape where the ratio of width and height of the ellipses is 2 : 1. The rotation angle in this case is 45°.

In Figs. 5 and 6, we can observe that some green points are depicted without their granules and are completely surrounded by the granules of red points. This basically means that the multi-class granular approximation values of these green points are smaller than 0.5 (hence, the granules cannot be drawn), and that the red granules are covering those green points. Therefore, it is suitable to change the labels of those green points into red. We can conclude that the learning, characterized by optimization problems (23) and (24), can be applied in classification problems and lead to an optimal relabeling of instances based on the loss function L .

6. Conclusion and future work

We have introduced the concepts of disjoint and adjacent fuzzy granules and discussed their connection with the concept of granular approximation introduced before. Based on disjoint and adjacent granules, a granular approximation concept was applied to the multi-class classification problem leading to the definition of a multi-class granular approximation. At the end, we explained how to calculate it efficiently in practice for the Lukasiewicz t -norm and other fuzzy connectives that it generates, using linear and quadratic programming methods.

We consider the following future research directions.

- We discussed in Section 5 that newly obtained multi-class granular approximations can be used for prediction purposes. The main focus of the future work will be to explore that more in depth i.e., to investigate the performance of optimization procedures (23) and (24) in classification problems. The goal will be to select classification datasets, set up a benchmark and compare the performance of the approaches (23) and (24) with other classification approaches.
- As mentioned in the introduction, the concept of granules in this case is closely related to that of decision rules. Therefore, one aim will be to explore how multi-class granular approximations can be used to develop rule induction algorithms.
- In order to define a multi-class granular approximation, we used only symmetric loss functions compared to the original granular approximation that considers also non-symmetric loss functions (e.g., p -quantile loss) [10]. It would be worth investigating if

application of a non-symmetric loss can help to control “preference” toward some decision class (which is important in, e.g., imbalanced classification problems) and how such loss functions would perform in classification tasks.

- In this article, S is a crisp equivalence relation. However, there is an option to extend it to a fuzzy relation in order to handle regression tasks. In such a scenario, the understanding of granules and other properties is different than for S being crisp. Hence, more analysis is required in the case of S being a fuzzy relation.
- It would be worth to explore if the extension as for (22) can be obtained for a T -preorder relation instead of a T -equivalence relation. Such extension could be then applied in monotone classification problems.

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.

Data availability

No data was used for the research described in the article.

Acknowledgements

Marko Palangetić and Chris Cornelis would like to thank Odysseus project from Flanders Research Foundation (FWO), grant no. 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. Roman Słowiński is acknowledging the support of grant 0311/SBAD/0700.

References

- [1] J.T. Yao, A.V. Vasilakos, W. Pedrycz, Granular computing: perspectives and challenges, *IEEE Trans. Cybern.* 43 (6) (2013) 1977–1989.
- [2] W. Pedrycz, Allocation of information granularity in optimization and decision-making models: towards building the foundations of granular computing, *Eur. J. Oper. Res.* 232 (1) (2014) 137–145.
- [3] A. Bargiela, W. Pedrycz, The roots of granular computing, in: 2006 IEEE International Conference on Granular Computing, IEEE, 2006, pp. 806–809.
- [4] C. Degang, Y. Yongping, W. Hui, Granular computing based on fuzzy similarity relations, *Soft Comput.* 15 (6) (2011) 1161–1172.
- [5] Y. Yao, Granular computing using neighborhood systems, in: *Advances in Soft Computing*, Springer, 1999, pp. 539–553.
- [6] L. Zadeh, Fuzzy sets, *Inf. Control* 8 (3) (1965) 338–353.
- [7] L.A. Zadeh, Fuzzy sets and information granularity, in: *Advances in Fuzzy Set Theory and Applications*, vol. 11, 1979, pp. 3–18.
- [8] L.A. Zadeh, Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, *Fuzzy Sets Syst.* 90 (2) (1997) 111–127.
- [9] M. Palangetić, C. Cornelis, S. Greco, R. Słowiński, Granular representation of OWA-based fuzzy rough sets, *Fuzzy Sets Syst.* 440 (2022) 112–130.
- [10] M. Palangetić, C. Cornelis, S. Greco, R. Słowiński, Granular approximations: a novel statistical learning approach for handling data inconsistency with respect to a fuzzy relation, *Inf. Sci.* 629 (2023) 249–275.
- [11] Z. Pawlak, Rough sets, *Int. J. Comput. Inf. Sci.* 11 (5) (1982) 341–356.
- [12] S. Greco, B. Matarazzo, R. Słowiński, A new rough set approach to evaluation of bankruptcy risk, in: *Operational Tools in the Management of Financial Risks*, Springer, 1998, pp. 121–136.
- [13] D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, *Int. J. Gen. Syst.* 17 (2–3) (1990) 191–209.
- [14] C.Y. Wang, B.Q. Hu, Granular variable precision fuzzy rough sets with general fuzzy relations, *Fuzzy Sets Syst.* 275 (2015) 39–57.
- [15] B.W. Fang, B.Q. Hu, Granular fuzzy rough sets based on fuzzy implicators and coimplicators, *Fuzzy Sets Syst.* 359 (2019) 112–139.
- [16] W. Kotłowski, R. Słowiński, Statistical approach to ordinal classification with monotonicity constraints, in: *Preference Learning ECML/PKDD 2008 Workshop*, 2008.
- [17] J.W. Grzymala-Busse, Lers—a system for learning from examples based on rough sets, in: *Intelligent Decision Support*, Springer, 1992, pp. 3–18.
- [18] S. Greco, B. Matarazzo, R. Słowiński, 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.
- [19] J.W. Grzymala-Busse, J. Stefanowski, Three discretization methods for rule induction, *Int. J. Intell. Syst.* 16 (1) (2001) 29–38.
- [20] E.P. Klement, R. Mesiar, E. Pap, *Triangular Norms*, vol. 8, Springer Science & Business Media, 2013.
- [21] 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.
- [22] P. Wolfe, The simplex method for quadratic programming, *Econometrica* (1959) 382–398.
- [23] S.I. Gass, *Linear Programming: Methods and Applications*, Courier Corporation, 2003.
- [24] B. De Baets, R. Mesiar, Metrics and t-equalities, *J. Math. Anal. Appl.* 267 (2) (2002) 531–547.
- [25] P.C. Mahalanobis, On the generalised distance in statistics, *Proc. Natl. Inst. Sci. India* (1936) 49–55.