Bireducts with tolerance relations

1st M. José Benítez-Caballero
Departament of Mathematics
University of Cádiz
Cádiz, Spain
mariajose.benitez@uca.es

2nd Jesús Medina
Departament of Mathematics
University of Cádiz
Cádiz, Spain
jesus.medina@uca.es

3rd Eloísa Ramírez-Poussa
Departament of Mathematics
University of Cádiz
Cádiz, Spain
eloisa.medina@uca.es

4th Dominik Ślęzak
Institute of Informatics
University of Warsaw
Warsaw, Poland
slezak@mimuw.edu.pl

***Abstract*—Reducing the number of attributes by preventing the occurrence of incompatibilities and eliminating existing noise in the original data is an important goal in different frameworks, such as in those focused on modelling and processing incomplete information in information systems. Bireducts were introduced in Rough Set Theory (RST) as one successful solution for achieving a balance between the elimination of attributes and the characterization of objects that the remaining attributes can still distinguish. This paper considers bireducts in a general framework in which attributes induce tolerance relations over the available objects. In order to compute the new reducts and bireducts a characterization based on a general discernibility function is given.**

***Index Terms*—Attributes reduction, tolerance relations, discernibility function, information bireducts.**

I. Introduction

Two complementary approaches to treat imperfect knowledge are Fuzzy Set Theory (FST) introduced by Zadeh [11] and Rough Set Theory (RST) proposed by Pawlak [9]. In FST, the elements belong to a set considering a certain degree of truth. On the other hand, RST computes approximations of concepts from incomplete information.

One of the main goal is to reduce databases keeping the same information. To this end, the reducts, minimal subsets of attributes preserving the original information, were studied in [3], [6], [8].

In this paper, we also consider bireducts, which are an extension the notion of reduct, that is, a subset of attributes and a subset of objects that prevent the occurrence of incompatibilities and eliminating existing noise in the original data.

Throughout the paper, we work with information reducts and information bireducts, as well as with decision reducts and decision bireducts. We also take into consideration similarity and tolerance relations in order to provide a natural relationship of distance among the elements of the universe. In some cases, a tolerance relation can be more appropriate since, for instance, the transitivity constraints imposed by similarity relations may produce conflicts with user's specifications or the exclusive use of similarity relations may cause wrong modeling of vague information. The notions and results obtained considering this framework is deeply studied in [2].

Partially supported by the Spanish Science Ministry project TIN2016-76653-P.

II. (Bi)Reducts in Information System

In this section we are going to present the notions of reduct and bireduct of an information system. First of all, we recall the idea of tolerance relation.

If we consider an information system $\mathbb{A} = (U, \mathcal{A})$, a fuzzy tolerance relation family $\{E_a : V_a \times V_a \to [0,1] \mid a \in \mathcal{A}\}$ and a family of values $\Delta = \{\delta_a \in [0,1] \mid a \in \mathcal{A} \cup \{d\}\}$, we can define, for each $a \in \mathcal{A}$, the relations T_{a,δ_a} as:

$$T_{a,\delta_a} = \{(v, w) \in V_a \times V_a \mid \delta_a \leq E_a(v, w)\} \qquad (1)$$

Note that each relation T_{a,δ_a} is straightforwardly a tolerance relation. Moreover, in the general environment of an information system or decision system, each attribute can have different nature and so, different thresholds could be assumed for each attribute. For further information about tolerance relations see [7]. For some examples on how to employ tolerance relations in rough set mechanisms of attribute reduction see, e.g. [10].

Based on a family of tolerance relations $\mathcal{E} = \{R_a \subseteq V_a \times V_a \mid a \in \mathcal{A}\}$, the notion of discernibility is generalized as follows.

Definition 1: Given an information system $\mathbb{A} = (U, \mathcal{A})$, a subset $B \subseteq \mathcal{A}$ and a tolerance relation family $\mathcal{E} = \{R_a \subseteq V_a \times V_a \mid a \in \mathcal{A}\}$, we say that objects $x, y \in U$ are $\mathcal{E}_B$-*similar* if for all $a \in B$ we have

$$(a(x), a(y)) \in R_a$$

Otherwise, we say that objects $x, y \in U$ are $\mathcal{E}_B$-*discordant*, that is, if the following holds

$$\{a \in B \mid (a(x), a(y)) \notin R_a\} \neq \varnothing$$

In the following definition, we present the notion of $\mathcal{E}$-information reduct, the generalization of reduct considering tolerance relations.

Definition 2: The set $B \subseteq \mathcal{A}$ is called $\mathcal{E}$-*information reduct* if B satisfies that every pair $x, y \in U$, which is $\mathcal{E}$-discordant, is also $\mathcal{E}_B$-discordant, and B is irreducible with respect to this property, that is, there is no $C \subsetneq B$ such that all pairs $x, y \in U$ are $\mathcal{E}_C$-discordant.

Analogously, we generalize the notion of information bireduct.

Definition 3: Let $\mathbb{A} = (U, \mathcal{A})$ be an information system. The pair (X, B), where $X \subseteq U$ and $B \subseteq \mathcal{A}$, is called $\mathcal{E}$-*information bireduct* if and only if all pairs $x, y \in X$ are $\mathcal{E}_B$-discordant and, B is irreducible and X is inextensible with respect to this property.

III. (Bi)Reduct in Decision System

In this section, we will present the notions and result needed in order to study the knowledge of a decision system. We generalize the notion of decision reduct. Throughout this section, we consider a decision system $\mathbb{A} = (U, \mathcal{A} \cup d)$, that is, a set of objects, a set of attributes and a decision attribute.

Definition 4: Let $\mathbb{A} = (U, \mathcal{A} \cup d)$ a decision system. A subset $B \subseteq \mathcal{A}$ is called $\mathcal{E}$-*decision reduct* if B satisfies that every $x, y \in U$, which is $\mathcal{E}_d$-discordant and $\mathcal{E}$-discordant, is also $\mathcal{E}_B$-discordant, and B is irreducible with respect to this property.

In order to compute the reducts, we are going to use the generalization of the unidimensional discernibility function.

Definition 5: The *unidimensional* $\mathcal{E}$-*discernibility function of* $\mathbb{A}$, is defined as the following conjunctive normal form (CNF):

$$\tau_{\mathbb{A}}^{\text{uni}} = \bigwedge \Big\{ \bigvee \{a \in \mathcal{A} \mid (a(x), a(y)) \notin R_a\} \mid x, y \in U, (d(x), d(y)) \notin R_d \Big\}$$

where the elements of $\mathcal{A}$ are the propositional symbols of the language.

The following result presents a mechanism to compute the reducts, using the reduced disjunctive normal form associated with the unidimensional $\mathcal{E}$-discernibility function.

Theorem 1: An arbitrary set B, where $B \subseteq \mathcal{A}$, is a $\mathcal{E}$-decision reduct of $\mathbb{A}$ if and only if the cube $\bigwedge_{b \in B} b$ is a cube in the RDNF of $\tau_{\mathbb{A}}^{\text{uni}}$.

Also, we can define the decision bireduct considering a tolerance relation.

Definition 6: A $(\mathcal{E}, U)$-*decision bireduct* is a pair (X, B), where $X \subseteq U$ and $B \subseteq \mathcal{A}$, and satisfy that all $x \in X$ and $y \in U$, with $(d(x), d(y)) \notin R_d$, are $\mathcal{E}_B$-discordant and, B is irreducible and X is inextensible with respect to this property.

The following definition presents the conjunctive normal form with which the bidimensional $\mathcal{E}$-discernibility function is defined.

Definition 7: The conjunctive normal form

$$\tau_{\mathbb{A}}^{bi} = \bigwedge \Big\{ x \vee y \vee \bigvee \{a \in \mathcal{A} \mid (a(x), a(y)) \notin R_a\} \mid x, y \in U, x < y, \ (d(x), d(y)) \notin R_d \Big\}$$

where the elements of U and $\mathcal{A}$ are the propositional symbols of the language, is called the *bidimensional* $\mathcal{E}$-*discernibility function of* $\mathbb{A}$.

This bidimensional $\mathcal{E}$-discernibility function is used in order to characterize the computation of $\mathcal{E}$-decision bireducts.

Theorem 2: An arbitrary pair (X, B), where $X \subseteq U$ and $B \subseteq \mathcal{A}$, is a $\mathcal{E}$-decision bireduct if and only if the cube $\bigwedge_{b \in B} b \wedge \bigwedge_{x \notin X} x$ is a cube in the RDNF of $\tau_{\mathbb{A}}^{bi}$.

IV. Conclusions and Future Work

We have considered tolerance relations in order to study the reducts and bireducts in the classical environment of RST. We have generalized the classical notion of discernibility function, from which we have characterized the reducts and bireducts in these environments.

The consideration of tolerance relations within this theory provides a great flexibility in different environments and the range of possible applications increase dramatically, for example, considering fuzzy tolerance relations with thresholds.

In the future, we will consider the theory developed throughout this paper in order to provide a new reduction method in fuzzy FCA. In addition, the (bi)reduction proposed for FCA will be compared with other reduction mechanisms, which reduce the size of the concept lattice considering similarities [1], [5].

Furthermore, we will extend our approach to obtain bireducts in fuzzy environments, such as in fuzzy rough sets [3], [4] and we will apply the theory developed in both theories to practical cases.

References

[1] R. Bělohlávek, J. Dvořák, and J. Outrata. Fast factorization by similarity in formal concept analysis of data with fuzzy attributes. *Journal of Computer and System Sciences*, 73(6):1012 – 1022, 2007.

[2] M. J. Benítez-Caballero, J. Medina, E. Ramírez-Poussa, and D. Ślęzak. Bireducts with tolerance relations. *Information Sciences*, 435:26 – 39, 2018.

[3] C. Cornelis, R. Jensen, G. Hurtado, and D. Ślęzak. Attribute selection with fuzzy decision reducts. *Information Sciences*, 180:209–224, 2010.

[4] C. Cornelis, J. Medina, and N. Verbiest. Multi-adjoint fuzzy rough sets: Definition, properties and attribute selection. *International Journal of Approximate Reasoning*, 55:412–426, 2014.

[5] A. Formica. Ontology-based concept similarity in formal concept analysis. *Information Sciences*, 176(18):2624–2641, 2006.

[6] A. Janusz and D. Ślęzak. Rough set methods for attribute clustering and selection. *Applied Artificial Intelligence*, 28(3):220–242, 2014.

[7] F. Klawonn. Should fuzzy equality and similarity satisfy transitivity? comments on the paper by m. de cock and e. kerre. *Fuzzy Sets and Systems*, 133(2):175 – 180, 2003.

[8] J. Medina. Relating attribute reduction in formal, object-oriented and property-oriented concept lattices. *Computers & Mathematics with Applications*, 64(6):1992–2002, 2012.

[9] Z. Pawlak. Rough sets. *International Journal of Computer and Information Science*, 11:341–356, 1982.

[10] A. Skowron and J. Stepaniuk. Tolerance approximation spaces. *Fundamenta Informaticae*, 27(2-3):245–253, 1996.

[11] L. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.