Available online at www.sciencedirect.com

ScienceDirect

Fuzzy Sets and Systems 401 (2020) 113–132

FUZZY sets and systems

www.elsevier.com/locate/fss

Representation by levels: An alternative to fuzzy sets for fuzzy data mining

Carlos Molina [a], M. Dolores Ruiz [b,*], José M. Serrano [a]

[a] *Dept. of Computer Science, University of Jaén, Spain*
[b] *CITIC-UGR, Dept. Computer Science and A.I., University of Granada, Spain*

Received 13 March 2019; received in revised form 4 September 2019; accepted 28 November 2019
Available online 10 December 2019

Abstract

In this paper we describe and discuss the main contributions of the representation by levels approach to fuzzy data mining. Representation by levels is an alternative representation of fuzziness in information and data, which is complementary to fuzzy sets in the sense that it provides tools and algebraic structures beyond the capabilities of fuzzy set theories, based on t-norms, t-conorms and fuzzy negations. Our approach allows to extend any crisp mining technique to the fuzzy case in a simple way, keeping all the properties of the crisp technique. We illustrate our discussion with examples and existing approaches based on representation by levels to fuzzy association rules and the related issues of mining exception/anomalous rules and mining fuzzy bag databases.

Keywords: Representation by levels; Fuzzy data mining; Assessment measures; Fuzzy association rules; Mining bag databases

1. Introduction

The potential contributions of fuzzy set theory and extensions in Knowledge Discovery in Databases (KDD) and machine learning (ML) are widely recognized [41,33,32]. Fuzzy set theory has the advantage to produce models that are more comprehensible, less complex, and more robust, being especially useful for representing “vague” patterns and modeling and processing various forms of uncertain and incomplete information [32].

Fuzzy set theory is a way to mathematically represent concepts and objects affected by fuzziness, and to operate and reason with them. In the last years, a different mathematical theory, called *representation by levels* (RL for short) [44, 46], has been proposed for the same purpose, offering important capabilities that cannot be provided by fuzzy set theory. There are several proposals in the literature which go in a similar direction, with some differences [24,35] that will be highlighted in next section. Additionally, the reader can find in the literature, several models that have been proposed in the ambit of fuzzy quantification in order to fulfill the desirable properties of Fuzzy Quantification Mechanisms. In [26] the author defined three-valued cuts, or γ-cuts, to define models of fuzzy quantification fulfilling

* Corresponding author.
E-mail addresses: carlosmo@ujaen.es (C. Molina), mdruiz@decsai.ugr.es (M.D. Ruiz), jschica@ujaen.es (J.M. Serrano).

https://doi.org/10.1016/j.fss.2019.11.014
0165-0114/

the symmetry properties involving the negation operator. In [20] the authors proposed a probabilistic model based on crisp representatives of a fuzzy set, which is then applied to define a Quantifier Fuzzification Mechanism, fulfilling properties such as correct generalization of crisp expressions and induced operators.

It is important to remark that RLs are not fuzzy sets in general, and the operations are completely different. In particular, RLs with ordinary set operations have a Boolean structure, something that cannot be achieved with fuzzy set theories on the basis of t-norms, t-conorms, and fuzzy negations [21,22]. In addition, the representation of numbers affected by fuzziness using RLs has nothing to do with fuzzy numbers (that are in fact fuzzy intervals of numbers [23]), and keep all the arithmetic properties and the same algebraic structure that the corresponding crisp numbers. Hence, the proposal is not another point of view for considering fuzzy set theory, but a completely different framework, and the use of RLs in fuzzy data mining offer several features and properties that cannot be provided by fuzzy sets.

We do not claim that data mining techniques based on fuzzy sets are not correct. We just show that RLs contribute with certain properties that can be requested in particular fuzzy mining contexts, and cannot be provided by fuzzy set theory. Hence, RLs add to the existing fuzzy mining toolbox some capabilities that fuzzy sets cannot, that can be useful to the extent that they are necessary in a specific application. For instance, as a consequence of the Boolean structure of RLs, the calculations of frequencies and probability measures keep all the properties of the crisp case, something not existing in fuzzy set theory. We shall see examples for such of these properties, and some of their potential contributions to fuzzy data mining. In addition, since RLs are able to represent and operate with fuzziness and datasets, they can be used in general for providing the advantages described in [41,33,32].

In this paper we explain in detail the main advantages that RLs can offer to fuzzy association rule mining when requested. We introduce the RL approach to data mining, which gives a general methodology to extend a data mining definition or technique to the fuzzy case, using RLs for representation and operations. In this approach, fuzzy sets play an important role as the way to get inputs and generate meaningful outputs; however, the intermediate operations are not performed by fuzzy sets, and hence the results are not those of fuzzy sets in general. We illustrate our discussion with several examples and existing techniques in the literature that support these contributions in the setting of fuzzy association rule mining.

The paper is organized as follows: in Section 2 we introduce the basic notions of RLs and general contributions to data mining. Then, in Section 3, we describe different approaches for fuzzy data mining based on RLs in the literature, and the particular contributions to such problems. Specifically, fuzzy association rules are discussed in Section 4. Finally, section 5 concludes the paper.

2. Representation by levels

A set is affected by fuzziness when its boundaries are fuzzy in the sense that, in addition to elements that are or are not in the set, there are other elements for which membership is not crisp, but gradual.

As it is well known, fuzzy sets are a very good mathematical model for representing sets affected by fuzziness. Fuzzy set theory has been shown to be very useful for solving problems in which fuzziness appear, with an overwhelming amount of theoretical and practical results in the literature.

Representation by levels [44,46] is *a different mathematical model for representing fuzziness*, akin to the notion of gradual set by Dubois and Prade [24]. The main difference between RLs and the latter is that, whilst in [24] gradual sets are presented as a new contribution *within fuzzy set theory*, in our view, presented in [44,46], RLs are *neither fuzzy sets nor just an alternative representation of fuzzy sets*. RLs are intended for representing and operating with fuzziness in a different way, with different semantics. Of course, since both fuzzy sets and RLs are intended to mathematically model the same kind of objects (sets affected by fuzziness), it is possible to find relations between them, specially from the point of view of the representation structure (more specifically, between the representation of fuzzy sets via alpha-cuts, and RLs). However, there is no one-to-one correspondence between fuzzy sets and RLs, and operations and their properties are completely different.

RLs approach is also different to the $X - \mu$ approach presented in [35]. In this model the authors define a kind of "inverse" function of the fuzzy membership function as $\mu{*}_{\geq} : [0, 1] \longrightarrow \mathcal{P}(X)$ where $\mu *_{\geq} (\alpha) = \{x \in X : f(x) \geq \alpha\}$. The definition is similar to that of RLs but in the continuous range of [0,1]. However when the authors applied their approach for data mining purposes (see for instance [36], the assessment measures are computed using the discrete set of known membership values which have at least one appearance in the dataset. For very large datasets this may slow the efficiency of the algorithm due to the large number of different values that can appear in the dataset. On the contrary, for RLs the set Λ is set in advance and there is no necessity of aggregating new levels during the execution of the algorithm over a dataset.

Let us remark that we do not claim that RLs are better, subsume, or outperform fuzzy sets as a model for representing and operating with fuzziness in general. RLs just enlarge our available toolbox of mathematical models for dealing with fuzziness, incorporating abilities and properties that fuzzy set theory cannot provide. On its turn, RLs have some drawbacks that fuzzy sets do not have. These are mostly related to understandability of the representation for humans, as well as storage and computational issues, for which nevertheless some solutions are available [44,46]. As we shall see later, it is possible to use fuzzy sets and RLs in systems at the same time, every one providing their best capabilities.

In the following we provide a brief introduction to the basic aspects of RLs.

2.1. Notation

The basic idea of RLs is that fuzzy concepts defined on a set of objects X may be described by an assignment of crisp sets to levels represented by values in $(0, 1]$, where level 1 corresponds to the most restrictive view of the concept, containing objects for which we are completely sure they satisfy the concept. Value 0 means *no restriction at all*, i.e., at this level every object in X satisfies the concept (and every other concept), and hence 0 is not used in RLs since it does not make a difference. The semantics of intermediate levels is based on the distance to 1 and 0, i.e., level 0.5 is halfway between being totally strict and no strict at all.

In practical applications we assume that RLs are defined using a finite set of levels $\Lambda = \{\alpha_1, \ldots, \alpha_m\}$ satisfying $1 = \alpha_1 > \alpha_2 > \cdots > \alpha_m > \alpha_{m+1} = 0$, $m \geq 1$, since we deal with finite computers and a finite set of objects.[1] An RL of a fuzzy concept on X is defined as follows:

Definition 1 *([46]).* An RL is represented by a pair (Λ, ρ) where $\Lambda \subset (0, 1]$ is a finite set of levels and ρ is a function

$$\rho : \Lambda \to \mathcal{P}(X) \tag{1}$$

As a convention, a concept and its RL will be denoted in the same way, i.e., A is a concept that will be represented by an RL $A = (\Lambda_A, \rho_A)$.

Definition 2 *([46]).* The set of crisp representatives Ω_A of an RL (Λ_A, ρ_A) is

$$\Omega_A = \{\rho_A(\alpha) \mid \alpha \in \Lambda_A\} \tag{2}$$

From the point of view of the structure of representation, the collection of significant alpha-cuts of a fuzzy set with a finite amount of different membership degrees may be viewed as a particular case of RL. As an example, the RL for a fuzzy set A is the pair (Λ_A, ρ_A), where Λ_A is obtained using equation (3),

$$\Lambda_A = \{A(x) \mid x \in support(A)\} \cup \{1\} \tag{3}$$

and $\rho_A(\alpha) = A_\alpha \ \forall \alpha \in \Lambda_A$, with A_α being the alpha-cut of level α of the fuzzy set A.

However, in RLs crisp sets are not required to be nested with respect to levels in general, and hence not every RL corresponds to the alpha-cut representation of a fuzzy set.

Though defined through a finite set of levels, the function ρ is defined for the whole interval $(0, 1]$, allowing to operate with RLs defined on different sets of levels:

Definition 3 *([46]).* Let (Λ, ρ) be an RL with $\Lambda = \{\alpha_1, \ldots, \alpha_m\}$ satisfying $1 = \alpha_1 > \alpha_2 > \cdots > \alpha_m > \alpha_{m+1} = 0$. Let $\alpha \in (0, 1]$ and $\alpha_i, \alpha_{i+1} \in \Lambda$ be such that $\alpha_i \geq \alpha > \alpha_{i+1}$. Then

$$\rho(\alpha) = \rho(\alpha_i) \tag{4}$$

[1] However, RLs are always defined for all levels in the complete interval (0,1], finite representations being extended as we shall see later.

Definition 4 *([46]).* Let (Λ, ρ) and (Λ', ρ') be two RLs on X. We say that both representations are *equivalent*, noted $(\Lambda, \rho) \equiv (\Lambda', \rho')$, iff $\forall \alpha \in (0, 1]$

$$\rho(\alpha) = \rho'(\alpha).$$

2.2. Fuzzy summary of an RL

For every RL, it is possible to obtain a fuzzy set measuring the *amount of membership* of every object to the concept represented by the RL. This fuzzy set is just a summary of the membership of each element with a particular semantics, and neither a proper representation of the RL nor an equivalent view; in particular, many different RLs can yield the same memberships for objects in X, that is, the same fuzzy summary.

Definition 5 *([46]).* Let (Λ, ρ) be an RL. We define the associated probability distribution $\omega : \Lambda \to [0, 1]$ as

$$\omega(\alpha_i) = \alpha_i - \alpha_{i+1}. \tag{5}$$

On this basis, a probability assignment $m : \Omega \to [0, 1]$ can be associated to the crisp representatives of an RL:

$$m(Y) = \sum_{\alpha_i \mid Y=\rho(\alpha_i)} \alpha_i - \alpha_{i+1} \quad \forall Y \in \Omega. \tag{6}$$

Note that the concept of probability of fuzzy sets has been dealt in many occasions in the ambit of Fuzzy Set Theory. An example is the proposal made in [3] to define the mass assignment function of a normalized fuzzy set. In our case, we have followed the natural definition of discrete mass probability function using the sigma-count.

Now, given an RL A, the associated fuzzy summary $\nu_A : X \to [0, 1]$ is

$$\nu_A(x) = \sum_{Y \in \Omega_A \mid x \in Y} m_A(Y) \tag{7}$$

The assignment of probabilities to crisp representatives in Eq. (6) can be seen as a random set. However, it is not a representation of uncertainty neither about the value of a variable in a dataset to be analyzed, nor about the correct representation of the RL (a representation, given by an assignment of crisp representatives to levels, which is assumed to be certain and precise, as it is also the case of fuzzy sets). On the contrary, $m(Y)$ represents the probability that the crisp representative corresponding to a level α chosen at random in $(0, 1]$ is Y. The value $m(Y)$ can be seen just as a measure of the *intensity* or *extent* to which Y is a valid representative of the RL, and it is useful in order to provide a more easily understandable view of the concept represented by the RL to users, that can be expressed either using $m(Y)$ or, even better, through the fuzzy summary of Eq. (7). From that point of view, the information expressed by this probability assignment, and its potential use, is different from the application of random sets to data analysis that can be found in the literature, since it is not giving information neither about the unknown value of a variable, nor about the uncertainty in the value of a set-valued variable.

The fuzzy summary of an RL is useful for providing a (partial) view of the information contained in an RL in the form of a measure of membership of each individual element, since fuzzy sets are much easier to understand than RLs for humans, specially when the number of levels is high. The membership value of an element is simply the probability that the object appears in the representative of a certain level α chosen at random in $(0, 1]$, again measuring the intensity of the presence of the element in the RL. Note that this interpretation is also valid for fuzzy sets.

Let μ_A be a fuzzy set representing concept A and let (Λ_A, ρ_A) be the RL given by the alpha-cut representation of μ_A. Then it is $\nu_A(x) = \mu_A(x)$. However, this partial correspondence between RLs and fuzzy sets is broken by operations. In general, given two fuzzy sets μ_A and μ_B and their corresponding RLs (Λ_A, ρ_A) and (Λ_B, ρ_B), and given an operation $\star$, the fuzzy summary of $A \star B$ is not necessarily the same as $\mu_A \star \mu_B$ for any fuzzy realization of $\star$.

Due to this fuzzy summary operation, fuzzy sets and RLs can be used jointly in a system, providing each one their best capabilities. In the RL-system view for building computational systems dealing with fuzziness [46], fuzzy sets are used as inputs and outputs in the system, since they are more intuitive than RLs for humans. However, the internal computations are performed using RLs. This is possible since, as we have seen, fuzzy sets and RLs can be related (though not biunivocally) under certain assumptions: every fuzzy set can be interpreted as an RL comprised of its set of alpha-cuts, under the assumption that the crisp representatives in each level are nested (which in RLs is not always the case in general). In addition, for every RL it is possible to obtain a fuzzy set summarizing the membership information of each object to the RL. RL-systems are a good alternative when computations with sets affected by fuzziness are required to satisfy properties that cannot be provided by fuzzy set operations, but can be provided by the corresponding operations with RLs. Hence, in general, the fuzzy set obtained after a operation $\star$ in an RL-system is different from that provided by any fuzzy set theory implementation of $\star$ for the same operations, as we shall see in the next subsection.

2.3. Operations

Crisp operations defined on concepts are extended to RLs by operating in each level independently. In general, let $f : \mathcal{P}(X)^n \to \mathcal{P}(X)$ be a crisp operation and let $(P_1, \ldots, P_n)$ be fuzzy concepts defined on X with associated RLs $(\Lambda_{P_i}, \rho_{P_i})$. Then,

$$f(P_1, \ldots, P_n) = (\Lambda_{f(P_1,\ldots,P_n)}, \rho_{f(P_1,\ldots,P_n)}) \tag{8}$$

where

$$\Lambda_{f(P_1,\ldots,P_n)} = \bigcup_{1 \le i \le n} \Lambda_{P_i} \tag{9}$$

and, $\forall \alpha \in \Lambda_{f(P_1,\ldots,P_n)}$,

$$\rho_{f(P_1,\ldots,P_n)}(\alpha) = f(\rho_{P_1}(\alpha), \ldots, \rho_{P_n}(\alpha)) \tag{10}$$

In particular, conjunction and disjunction are extended as follows:

Definition 6 *([46])*. Let P, Q be fuzzy concepts represented by (Λ_P, ρ_P), (Λ_Q, ρ_Q). Then, $P \wedge Q$ and $P \vee Q$ are fuzzy concepts represented by $(\Lambda_{P\wedge Q}, \rho_{P\wedge Q})$ and $(\Lambda_{P\vee Q}, \rho_{P\vee Q})$, respectively, where

$$\Lambda_{P\wedge Q} = \Lambda_{P\vee Q} = \Lambda_P \cup \Lambda_Q \tag{11}$$

with, $\forall \alpha \in (0, 1]$,

$$\rho_{P\wedge Q}(\alpha) = \rho_P(\alpha) \cap \rho_Q(\alpha) \tag{12}$$

and

$$\rho_{P\vee Q}(\alpha) = \rho_P(\alpha) \cup \rho_Q(\alpha) \tag{13}$$

Definition 7 *([46])*. Let P be a fuzzy concept represented by (Λ_P, ρ_P). Then, $\neg P$ is a fuzzy concept represented by $(\Lambda_{\neg P}, \rho_{\neg P})$, where

$$\Lambda_{\neg P} = \Lambda_P \tag{14}$$

and, $\forall \alpha \in (0, 1]$,

$$\rho_{\neg P}(\alpha) = \overline{\rho_P(\alpha)} \tag{15}$$

where $\overline{Y}$ is the usual set complement of a crisp set Y.

Example 1 *([46])*. In order to illustrate the differences between operations with fuzzy sets and operations with RLs, let us consider an universe $X = \{x_1, \ldots, x_5\}$ and the fuzzy sets

$$A = 1/x_1 + 0.8/x_2 + 0.5/x_3 + 0.4/x_5$$

$$B = 0.9/x_1 + 0.6/x_3 + 0.5/x_4$$

Table 1
Negation of A.

α	$\rho_A(\alpha)$	$\rho_{\neg A}(\alpha)$
1	$\{x_1\}$	$\{x_2, x_3, x_4, x_5\}$
0.8	$\{x_1, x_2\}$	$\{x_3, x_4, x_5\}$
0.5	$\{x_1, x_2, x_3\}$	$\{x_4, x_5\}$
0.4	$\{x_1, x_2, x_3, x_5\}$	$\{x_4\}$

Table 2
Negation of B.

α	$\rho_B(\alpha)$	$\rho_{\neg B}(\alpha)$
1	$\emptyset$	X
0.9	$\{x_1\}$	$\{x_2, x_3, x_4, x_5\}$
0.6	$\{x_1, x_3\}$	$\{x_2, x_4, x_5\}$
0.5	$\{x_1, x_3, x_4\}$	$\{x_2, x_5\}$

Table 3
Several concepts derived from the fuzzy sets A and B (I).

α	$\rho_A(\alpha)$	$\rho_{\neg A}(\alpha)$	$\rho_B(\alpha)$	$\rho_{\neg B}(\alpha)$	$\rho_{A\wedge\neg B}(\alpha)$	$\rho_{B\wedge\neg A}(\alpha)$
1	$\{x_1\}$	$\{x_2, x_3, x_4, x_5\}$	$\emptyset$	X	$\{x_1\}$	$\emptyset$
0.9	$\{x_1\}$	$\{x_2, x_3, x_4, x_5\}$	$\{x_1\}$	$\{x_2, x_3, x_4, x_5\}$	$\emptyset$	$\emptyset$
0.8	$\{x_1, x_2\}$	$\{x_3, x_4, x_5\}$	$\{x_1\}$	$\{x_2, x_3, x_4, x_5\}$	$\{x_2\}$	$\emptyset$
0.6	$\{x_1, x_2, x_3\}$	$\{x_4, x_5\}$	$\{x_1, x_3\}$	$\{x_2, x_4, x_5\}$	$\{x_2\}$	$\emptyset$
0.5	$\{x_1, x_2, x_3\}$	$\{x_4, x_5\}$	$\{x_1, x_3, x_4\}$	$\{x_2, x_5\}$	$\{x_2\}$	$\{x_4\}$
0.4	$\{x_1, x_2, x_3, x_5\}$	$\{x_4\}$	$\{x_1, x_3, x_4\}$	$\{x_2, x_5\}$	$\{x_2, x_5\}$	$\{x_4\}$

Table 4
Several concepts derived from concepts A and B (II).

α	$\rho_{A\wedge B}(\alpha)$	$\rho_{A\vee B}(\alpha)$	$\rho_{\neg(A\wedge B)}(\alpha)$	$\rho_{\neg(A\vee B)}(\alpha)$	$\rho_{\neg A\vee\neg B}(\alpha)$
1	$\emptyset$	$\{x_1\}$	X	$\{x_2, x_3, x_4, x_5\}$	X
0.9	$\{x_1\}$	$\{x_1\}$	$\{x_2, x_3, x_4, x_5\}$	$\{x_2, x_3, x_4, x_5\}$	$\{x_2, x_3, x_4, x_5\}$
0.8	$\{x_1\}$	$\{x_1, x_2\}$	$\{x_2, x_3, x_4, x_5\}$	$\{x_3, x_4, x_5\}$	$\{x_2, x_3, x_4, x_5\}$
0.6	$\{x_1, x_3\}$	$\{x_1, x_2, x_3\}$	$\{x_2, x_4, x_5\}$	$\{x_4, x_5\}$	$\{x_2, x_4, x_5\}$
0.5	$\{x_1, x_3\}$	$\{x_1, x_2, x_3, x_4\}$	$\{x_2, x_4, x_5\}$	$\{x_5\}$	$\{x_2, x_4, x_5\}$
0.4	$\{x_1, x_3\}$	X	$\{x_2, x_4, x_5\}$	$\emptyset$	$\{x_2, x_4, x_5\}$

Table 1 shows the RLs for A and $\neg A$; Table 2 shows the RLs corresponding to B and $\neg B$. Note that the set of levels employed in each case is different.

It is immediate that the representation of the negation of a fuzzy set (e.g. $\neg A$) is not the set of α-cuts of the complement of the corresponding fuzzy set (e.g. the fuzzy set $\overline{A}(x) = 1 - A(x)$). This is also true for B in Table 2. In fact, the representations of $\neg A$ and $\neg B$ are not fuzzy sets.

Tables 3 and 4 show the representation of several concepts derived from A and B by different operations and combination of operations.

When one is used to consider fuzzy sets and their corresponding alpha-cuts, the fact that the crisp sets are not nested seems strange. But the idea of operations with RLs is very clear. Consider for instance A and its negation in Table 3. If we assume that at level 1 the representation of A is $\{x_1\}$ (since by the semantics of this level we are being maximally strict in our requirements for an element to be in A, and only x_1 satisfies such requirements in this level), then it is natural to consider that, in the same level and under the same requirements, the set of elements that are not in A is $X \backslash \{x_1\} = \{x_2, x_3, x_4, x_5\}$.

It is important to remark that with this extension of set operations, RLs form a Boolean algebra, whilst at the same time an element can appear in both the representation of a concept and its negation. On the contrary, no standard fuzzy set theory (FST), i.e., no triple (t-norm, t-conorm, fuzzy negation) is a Boolean algebra [21,22].

Hence, in an RL-system, given a fuzzy set A defined on X as input, performing RL-operations, and providing the fuzzy summary as output, it is always $A \cap A = A$, $A \cap \overline{A} = \emptyset$, and $A \cup \overline{A} = X$, something that cannot be provided by a system based on a specific FST. This is an example of the capabilities that RLs add to our toolbox for dealing with fuzziness, to be used particularly when these properties are necessary or convenient in our specific application. As we shall see, these properties have important consequences in data mining applications.

In addition, the problem of managing fuzzy information is transformed into a collection of independent problems of crisp information management, that can be solved by using ordinary crisp approaches. Hence, any crisp procedure may (in principle) be extended to the case of fuzzy information in a unique, straightforward way, different in general from the possibilities that FSTs can offer (hence complementing FSTs and enlarging our tools for managing fuzziness), and preserving at the same time all the ordinary properties of the crisp case.

Again, this feature of RLs offers interesting advantages in fuzzy data mining.

2.4. RL-numbers

RL-numbers [45] (corresponding to gradual numbers as defined in [24]) are assignments of numbers to levels. The representation by levels of a fuzzy real quantity (RL-real number for short [46]) is a pair (Λ, ρ) in which Λ is a set of levels and $\rho : \Lambda \to \mathbb{R}$. This idea can also be extended to other kind of numbers like integers, complex numbers, etc. Note that for RL-numbers, Ω is a finite set of numbers.

Following the general idea of RLs, operations on RL-numbers are performed by levels:

Definition 8. Let $f : \mathbb{R}^n \to \mathbb{R}$ and let $R_1 \dots R_n$ be RL-real numbers. Then $f(R_1, \dots, R_n)$ is an RL-real number with

$$\Lambda_{f(R_1,\dots,R_n)} = \bigcup_{1 \le i \le n} \Lambda_{R_i} \tag{16}$$

and, $\forall \alpha \in \Lambda_{f(R_1,\dots,R_n)}$

$$\mathcal{R}_{f(R_1,\dots,R_n)}(\alpha) = f(\mathcal{R}_{R_1}(\alpha), \dots, \mathcal{R}_{R_n}(\alpha)) \tag{17}$$

It is important to remark that we consider that operations not defined for certain combinations of real values are not defined for RL-numbers that satisfy that combination in at least one restriction level.

RL-numbers satisfy some important properties:

- RL-numbers are not fuzzy numbers, the latter being represented in each level by an interval, whilst RL-numbers have a single number in each level.
- Fuzziness of RL-numbers (corresponding to the entropy calculated from the probability distribution on $\mathbb{R}$ as given by Eq. (6) does not necessarily increase with operations, and may even decrease. In fact, the operation with RL-numbers not corresponding to crisp numbers, can yield a crisp number.
- Crisp operations are extended to RL-real numbers directly and uniquely operating by levels.
- All the properties of arithmetic operations are kept. The algebraic structure of RL-real numbers is exactly that of crisp real numbers.

As we shall see, RL-real numbers can be used in fuzzy data mining for representing quality measures for fuzzy patterns, fuzzy parameters, etc.

A summary of the information contained in an RL-number R can be provided by the probability distribution m_x for every real value $x \in \Omega_R$. But we can also use an scalar value as a way to summarize the information in an RL-number R as follows:

$$c_R = \sum_{x \in \Omega_R} x \cdot m_x \tag{18}$$

were m_x is calculated as in Eq. (6). The value c_R can be seen as a center of gravity of m_x with respect to the real line, with a semantics of expected value. We shall see some examples later.

Table 5
Membership degrees to "large", "¬ large", and "blue" for ten balls in an urn. Negation has been calculated using the standard negation.

	a	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*	*j*
large	1	0.7	0.5	1	0.8	0	0.8	0.7	1	0.5
blue	0.8	1	1	1	0.2	0.2	0.5	1	0	0.7
¬ large	0	0.3	0.5	0	0.2	1	0.2	0.3	0	0.5

3. Some contributions to fuzzy data mining

As pointed out in the previous section, RLs provide some properties that fuzzy sets cannot. When such properties are necessary, RLs become a suitable alternative to fuzzy sets in fuzzy data mining. In this section we explain two of the possible contributions of RLs to fuzzy data mining. Note that this section is not intended to be exhaustive with respect of the advantages that RLs may provide in fuzzy data mining, but to explain some of the most general and emphasizing those concerning the assessment measures employed during the process. More detailed contributions to fuzzy association rules is shown in section 4.

3.1. Frequencies and probabilities of fuzzy events

One of the most important aspects in data mining is providing appropriate quality measures for the obtained patterns. In many cases, such measures are based on calculating frequencies in datasets. For instance, when mining for association rules, the support of items is the percentage of transactions in which the item appear, that is, the frequency of appearance of the item in the transaction. In many other occasions, these frequencies are used for estimating or measuring probabilities in clustering, classification, etc.

When introducing fuzziness in data mining, we have to face the problem of calculating frequencies/probabilities of fuzzy events. For instance, consider the case of an urn X containing ten balls of different sizes, painted in different shades of blue. Notions as "large" or "blue" or, more generally, those related to size and color, are prototypical and well-known examples of fuzzy properties; hence, every ball has a membership degree to both concepts. The probability that a ball drawn at random from the urn is "large" (resp. "blue") is given by the frequency of "large" (resp. "blue") in the set of balls. "Drawing a large ball" and "drawing a blue ball" are examples of fuzzy events.

The usual way to calculate the probabilities of such fuzzy events in fuzzy data mining is by using the sigma-count to measure the cardinality of the fuzzy set of "large" (resp. "blue") balls, and dividing it by the number of balls. Let us consider for instance the dataset in Table 5, where balls in the set X are identified by letters from a to j.

The probabilities calculated using the sigma-count are:

$$p(large) = \frac{1}{10} \sum_{x \in X} large(x) = 0.7 \tag{19}$$

$$p(blue) = \frac{1}{10} \sum_{x \in X} blue(x) = 0.64 \tag{20}$$

On this basis, in the literature the conditional probabilities are obtained as in the crisp case. For the fuzzy event "draw a large blue ball", the point is that we need to compute the intersection of the sets of large and blue balls. If we use the minimum for such computation, we get:

$$p(large \wedge blue) = \frac{1}{10} \sum_{x \in X} (large \wedge blue)(x) = 0.49 \tag{21}$$

and hence, the probability that a blue ball is large is

$$p(large|blue) = \frac{p(large \wedge blue)}{p(blue)} \approx 0.765 \tag{22}$$

Of course, the result would be different if t-norms other than the minimum are employed.

Table 6
RL-representation of several fuzzy events. L=large, B=blue. Expressions of the type "a–e" mean events from a to e.

	L	B	$\neg L$	$L \wedge B$	$\neg L \wedge B$	$\neg L \wedge L$	$\neg L \vee L$
1	$\{a, d, i\}$	$\{b$–$d, h\}$	$\{b, c, e$–$h, j\}$	$\{d\}$	$\{b, c, h\}$	$\emptyset$	X
0.8	$\{a, d, e, g, i\}$	$\{a$–$d, h\}$	$\{b, c, f, h, j\}$	$\{a, d\}$	$\{b, c, h\}$	$\emptyset$	X
0.7	$\{a, b, d, e, g$–$i\}$	$\{a$–$d, h, j\}$	$\{c, f, j\}$	$\{a, b, d, h\}$	$\{c, j\}$	$\emptyset$	X
0.5	$\{a$–e, g–$j\}$	$\{a$–$d, g, h, j\}$	$\{f\}$	$\{a$–$d, g, h, j\}$	$\emptyset$	$\emptyset$	X
0.2	$\{a$–e, g–$j\}$	$\{a$–$h, j\}$	$\{f\}$	$\{a$–$e, g, h, j\}$	$\{f\}$	$\emptyset$	X

Table 7
RL-numbers representing probabilities for several fuzzy events in Table 6 and some associated conditionals, L=large, B=blue. The last row contains the expected value c for each RL-number.

Level	$p(L)$	$p(B)$	$p(\neg L)$	$p(L \wedge B)$	$p(\neg L \wedge B)$	$p(L\|B)$	$p(\neg L\|B)$
1	0.3	0.4	0.7	0.1	0.3	0.25	0.75
0.8	0.5	0.5	0.5	0.2	0.3	0.4	0.6
0.7	0.7	0.6	0.3	0.4	0.2	2/3	1/3
0.5	0.9	0.7	0.1	0.7	0	1	0
0.2	0.9	0.9	0.1	0.8	0.1	8/9	1/9
c	0.7	0.64	0.3	0.49	0.15	0.39+14/45	0.21+ 4/45

Despite the t-norm, negation, and cardinality employed, there are properties that cannot hold simultaneously with this scheme. In order to illustrate this claim, let us consider that in a data mining application the following properties, which hold in the crisp case, are needed simultaneously:

1. $p(A|A) = 1$
2. $p(A|C) + p(\neg A|C) = 1$

These properties cannot be simultaneously satisfied using a fuzzy set theory, since $p(A|A) = 1$ requires $A \cap A = A$, that can be only achieved using the minimum as t-norm, and $p(A|C) + p(\neg A|C) = 1$ requires $(A \cap C) \cap (\neg A \cap C) = \emptyset$, which cannot be achieved with the minimum in general (in particular when C is the whole referential universe, the requirement is $A \cap \neg A = \emptyset$). In our example, using the minimum it is $p(large|large) = p(blue|blue) = 1$ but it is $p(\neg large \wedge blue) = 0.22$ and

$$p(\neg large|blue) = \frac{p(\neg large \wedge blue)}{p(blue)} \approx 0.343 \tag{23}$$

and hence $p(large|blue) + p(\neg large|blue) \approx 1.1$.

On the contrary, these (and every other property of crisp probabilities) are kept if we use RLs. By assuming that "large" and "blue" can be represented by RLs with nested representatives, the corresponding RLs and some operations are shown in Table 6. The corresponding probabilities, in the form of RL-numbers, are in Table 7.

Let us remark that:

- Since $L \wedge L = L$, it is $p(L \wedge L) = p(L)$. It is also $L \wedge \neg L = \emptyset$ and $L \vee \neg L = X$, hence $p(L \wedge \neg L) = 0$ and $p(L \vee \neg L) = 1$. These properties do not hold simultaneously for any fuzzy set theory.
- All the ordinary properties of probability hold for the RL-numbers representing the probability of a fuzzy event, since both set and arithmetic operations are performed in each level independently. In particular, $p(L|B) + p(\neg L|B) = 1$ in every level and hence it is the crisp value 1, as requested.
- When operations with probabilities are to be performed, following the idea of RL-systems explained in the previous section, they are to be performed by levels, until final results are obtained in each level.
- Otherwise, if the final result to provide as output are the probabilities themselves, or if we want to have a global view of them as a partial result, we may use either m or c as a summary. The latter is shown as the final row in Table 7. As shown in [12], the expected value of the cardinality of levels for a fuzzy set is always equal to the sigma-count. Since in every level we have represented the cardinality divided by 10, the expected value of each column equals the sigma-count divided by 10. For example, for the first column it is

$$c_{p(L)} = (1-0.8)\cdot 0.3 + (0.8-0.7)\cdot 0.5 + (0.7-0.5)\cdot 0.7 + \\ + (0.5-0.2)\cdot 0.9 + (0.2-0)\cdot 0.9 = 0.7$$

Hence, the values c in Table 7 for L and B are equal to the probabilities obtained by using the sigma-count. This is also the case for $\neg L$ and $L \wedge B$ since the property holds as well for the RLs obtained by a single negation of a fuzzy set, and for the intersection of fuzzy sets performed via the minimum, which coincides with the summary of the intersection by levels of the corresponding RLs [46]. However, this is not the case in general for other combinations of operators, since RLs have structure of Boolean algebra, whilst fuzzy sets do not. As a consequence, the result of set operations with RLs does not correspond to a fuzzy set in general; particularly, crisp representatives are not necessarily nested after operations. This is the case for $\neg L \wedge B$, and hence the expected value of the RL-number $p(\neg L \wedge B)$ (0.15) is different from the probability obtained by using the sigma-count from the fuzzy sets (0.22). Note also that when for two RL-numbers it is $R_1 + R_2 = 1$, then it is easy to show that $c_{R_1} + c_{R_2} = 1$, and hence $c_{p(L|B)} + c_{p(\neg L|B)} = 1$, as it can be seen in Table 7. However, note that in general, the calculation of expected values is not functionally expressible, but requires to perform the set operations on the corresponding events, as in the case of ordinary probabilities.

We have employed an example based on a couple of properties of probability that are derived from Boolean properties for illustrating a potential contribution of using RLs in data mining. However, similar examples can be put forward related to many other combination of operations, and even individual derived properties of Boolean algebras that do not hold for any fuzzy set theory [1]. The same happens to important problems related to frequencies in a fuzzy environment, like fuzzy quantification [19].

Again, let us remark that we are not claiming that fuzzy data mining techniques developed so far on the basis of fuzzy sets, sigma-counts, usual fuzzy quantification, etc., are wrong. We only claim that they are right to the extent that the properties they cannot provide are not required for the results to be valid in the specific problem where they are being applied. On the other hand, in those cases in which those properties are necessary for the results to be valid, RLs offer a suitable alternative, as we have seen.

3.2. Extension of existing crisp algorithms

An important feature of using RLs is that a fuzzy mining problem is represented via a collection of independent crisp mining problems, corresponding to each level. Hence, any algorithm for crisp data mining can be immediately applied to solve the fuzzy mining problem in a single and direct way: applying the algorithm to each level independently. As a consequence, well-known and tested crisp algorithms can be employed, guaranteeing the quality of the results in each level. That is, the quality of the results using RLs comes from the quality of the existing crisp algorithms and the fact that all their properties are kept.

Another interesting feature is that there is no necessity to determine how to extend operations and measures to the fuzzy case, or to study the different ways in which this can be done. This is useful for addressing fuzzy data mining problems quickly using existing crisp solutions.

The result of a data mining process using RLs is a collection of RL-patterns, that is, an assignment of sets of patterns to levels. For example, in the case of association rule mining, as we shall see later, a set of crisp association rules is obtained in each level. As another example, if we intend to learn a classifier using fuzzy features of objects in a dataset (like the "large" and "blue" properties in our previous example), we can learn a classifier in every level using any of the very good and widely employed techniques available. The result will be an RL-classifier, that is, an assignment of classifiers to levels.

Let us remark that the fact that we have to solve a collection of problems does not increase significantly the computational complexity of the data mining procedure. First, only a finite and relatively small amount of levels is enough. In our experiments for the techniques that we shall explain in detail in the next sections, using 20 equidistributed levels gave very good results. We have employed no more than 100 levels, which is far beyond our capability for distinguishing membership in practice. But an important advantage of RLs that alleviates time complexity is that, since computations in every level are independent, the mining procedures for every level can be computed completely in parallel. Hence, the complexity of the fuzzy mining problem is the same as its crisp counterpart.

Depending on the mining task, the final results can be employed in different ways. For instance, consider the case of an RL-classifier. Given an object with (possibly fuzzy) features, we can obtain an RL-representation of the object, where features will hold or not in every level and then, by applying the classifiers in each level independently to the corresponding features in that level, we can obtain a class in each level. This information can be summarized by using m in Eq. (6), so that we have a fuzzy class as result. This is reasonable: as the information employed is fuzzy, the result is also fuzzy. However, note that the result can be crisp as well despite the input information being fuzzy, particularly when classifiers in all levels yield the same class as result.

Summarizing, the RL general approach to fuzzy data mining consists of the following steps:

1. Represent the input fuzzy information using RLs with a fixed collection of relevant levels. If fuzzy sets are given as input, use the corresponding alpha-cut representation. Each level is a crisp version of the problem.
2. Solve the crisp version of the problem in each level by applying a suitable crisp mining algorithm in each level independently.
3. The fuzzy result is represented by an assignment of crisp results to levels. If the results are to be used as input to some procedure, proceed by levels independently. If they are to be provided to an end user, or we want to obtain a summary of the information, combine the information in different levels.

This approach, with small variations according to the specific problem, allows us to extend easily crisp algorithms to the fuzzy case, and to keep properties of the crisp case when necessary, particularly in calculating and operating with frequencies and probabilities. In addition to these general contributions, there are particular features that are provided when using RLs for specific fuzzy mining tasks, making again RLs an interesting alternative to fuzzy sets when such features are convenient or necessary. We shall see some illustrative examples in the following sections.

4. Fuzzy association rules

The natural way to consider the use of fuzziness in the association rule mining (ARM) problem is taking the elements of the ARM framework and studying possible and sensible fuzzy extensions of them [37]. There are many possible ways to introduce fuzziness in items, transactions, rules and measures, which are called *fuzzy frameworks* in [37]. Among them, the framework based on *fuzzy transactions* is particularly interesting, since the classical view of fuzzy association rules as fuzzy rules of the kind used in fuzzy control is just a particular case of such framework [37], though the framework is much more general, as described in [11]. In this section we describe some contributions of RLs for mining fuzzy associations within this framework emphasizing how the quality assessment measures are extended using RLs.

4.1. Mining fuzzy association rules on fuzzy transactions using RLs

As introduced in [11], let $I = \{i_1, \dots, i_m\}$ be a finite set of items. A *fuzzy transaction* is defined as a non empty fuzzy subset $\tau \subseteq I$. An item $i \in I$ is then satisfied in τ to a certain degree in the unit interval, i.e. $\tau(i) \in [0, 1]$. By extension, let A be an itemset of I, i.e. a subset of items in a fuzzy transaction τ. The membership degree of $A \subseteq I$ to the fuzzy transaction τ is defined as $\tau(A) = \min_{i \in A} \tau(i)$. In particular, a crisp transaction is a special case of fuzzy transaction where every item in the transaction has membership degree equal to 1 when it is satisfied in the transaction, and 0 otherwise. Then, a *fuzzy association rule*, $A \Rightarrow B$ is satisfied in $\tilde{D}$ if and only if $\tau(A) \leq \tau(B)$ for all $\tau \in \tilde{D}$.

The usual view of fuzzy association rules based on fuzzy partitions of the domains of quantitative variables is a particular case of the general model in [11]. Items correspond to pairs (attribute,label), for every label of a fuzzy partition of the domain of the attribute. Fuzzy transactions correspond to tuples/records. The membership of the item (X, L_i^X) in the transaction corresponding to the tuple t is given by $L_i^X(t[X])$, where $t[X]$ is the value of the attribute X in the tuple t. However, the general model covers many other scenarios, as shown in [11,37].

4.1.1. The RL approach

In [18], the RL approach is employed in order to mine association rules in a set of fuzzy transactions $\tilde{D}$ on the basis of crisp association mining algorithms, by following the same technique that we have described in previous sections:

Table 8
Set of fuzzy transactions $\tilde{D}$.

	i_1	i_2	i_3	i_4	i_5
t_1	1	0.2	1	0.8	0.9
t_2	1	1	0.8	0	0
t_3	0.4	0.1	0.6	0.6	0
t_4	0.6	0	0.4	0.4	0.5
t_5	0.4	0.1	0	0	0
t_6	0	1	0	0	0

Table 9
RLs of the fuzzy sets of transactions for several itemsets.

Level	$\rho_{\{i_1\}}$	$\rho_{\{i_3\}}$	$\rho_{\{i_4\}}$	$\rho_{\{i_1,i_3\}}$	$\rho_{\{i_1,i_3,i_4\}}$
1	$\{t_1,t_2\}$	$\{t_1\}$	$\emptyset$	$\{t_1\}$	$\emptyset$
0.8	$\{t_1,t_2\}$	$\{t_1,t_2\}$	$\{t_1\}$	$\{t_1,t_2\}$	$\{t_1\}$
0.6	$\{t_1,t_2,t_4\}$	$\{t_1,t_2,t_3\}$	$\{t_1,t_3\}$	$\{t_1,t_2\}$	$\{t_1\}$
0.4	$\{t_1,t_2,t_3,t_4,t_5\}$	$\{t_1,t_2,t_3,t_4\}$	$\{t_1,t_2,t_3\}$	$\{t_1,t_2,t_3,t_4\}$	$\{t_1,t_2,t_3\}$

- First, the fuzzy set of transactions that contain each item is represented by levels, each level containing the crisp set of transactions to which the item pertains with degree greater or equal to the level. This way, we can see items $i \in I$, and more generally an itemset $A \subset I$, as an RL (Λ_A, ρ_A) defined over the fuzzy database $\tilde{D}$, where Λ_A is a predefined set of levels and $\rho_A(\alpha) = \{\tau \in \tilde{D} : \tau(A) \geq \alpha\}$ for all $\alpha \in \Lambda_A$. By abuse of notation, we will employ A and B to refer and name their associated RL-sets $A = (\Lambda_A, \rho_A)$, $B = (\Lambda_B, \rho_B)$.
- Second, a crisp association rule mining algorithm is employed for mining rules in each level in parallel. Any existing algorithm can be used in this step, depending on the preferences of the user and the characteristics of the dataset. In the same way, any of the existing quality measures for association rules can be employed, without the necessity to take into account how to extend them to the fuzzy case, or which of the possible extensions to consider. This way, extending any crisp algorithm and/or measure to the fuzzy case is straightforward.
- Finally, the result of the previous procedure is an assignment of sets of association rules to each level. If rules are to be used as input for some other procedure, we proceed by levels (for instance, for building a classifier based on them, as in [6,29], obtaining a classifier in each level). If the rules are to be provided to the user, as a final result or just to have a summary of the information, we can provide a fuzzy set of association rules as result. The membership of each rule to the result is calculated using Eq. (7), meaning the degree to which the rule is a *strong* rule. Note that the classical notion of strong rule is crisp, given the thresholds. This fuzzy notion of strong rule is not related to using fuzzy thresholds for support and confidence, but derives from the fact that the items are in transactions to a certain degree.

Let us illustrate these ideas with an example:

Example 2. Let $I = \{i_1, i_2, \ldots, i_5\}$ be a set of items and $\tilde{D}$ be the set of fuzzy transactions in Table 8, showing the membership of each item in I to each of the six transactions. Table 9 shows the corresponding RLs for several itemsets.

Table 10 shows the support and confidence by levels of several rules, where support of itemsets in each level is calculated from Table 9, and confidence of the rule $A \Rightarrow B$ is calculated as $conf(A \Rightarrow B) = supp(A \cup B)/supp(A)$ as usual.[2]

When applied to each level independently, algorithms for crisp association rule mining yield the set of strong rules in each level. Strong rules are those having support and confidence above certain user-defined thresholds, hence such set depends on the thresholds employed. Table 11 shows which rules from Table 10 are strong in each level. This information can be summarized as a fuzzy subset of rules using Eq. (7); the resulting fuzzy sets are shown in the last

[2] During calculation of the confidence, indeterminations of the form "$\frac{0}{0}$" may appear when $|\rho_A(\alpha_i)| = 0$ for some level α_i. In such levels we take $conf(A \Rightarrow B) = 1$ because there are no exceptions to the rule. Note that this is not a problem since in such case it is $supp(A \Rightarrow B) = 0$ and the rule is not important.

Table 10
Support and confidence by levels of some rules in the fuzzy transactions of Table 8.

Level	$i_1 \Rightarrow i_3$		$i_3 \Rightarrow i_1$		$i_1 i_3 \Rightarrow i_4$	
	supp	conf	supp	conf	supp	conf
1	1/6	1/2	1/6	1	0	0
0.8	1/3	1	1/3	1	1/6	1/2
0.6	1/3	2/3	1/3	2/3	1/6	1/3
0.4	2/3	4/5	2/3	1	1/2	3/4

Table 11
Strong rules by levels among those in Table 10 for different support/confidence thresholds. The last row contains the summary of each column as a fuzzy subset of rules.

Level	Minimum thresholds		
	supp 1/6, conf 2/3	supp 1/6, conf 1/2	supp 1/3, conf 1
1	$i_3 \Rightarrow i_1$	$i_1 \Rightarrow i_3, i_3 \Rightarrow i_1$	None
0.8	$i_1 \Rightarrow i_3, i_3 \Rightarrow i_1$	$i_1 \Rightarrow i_3, i_3 \Rightarrow i_1$ $i_1 i_3 \Rightarrow i_4$	$i_1 \Rightarrow i_3, i_3 \Rightarrow i_1$
0.6	$i_1 \Rightarrow i_3, i_3 \Rightarrow i_1$	$i_1 \Rightarrow i_3, i_3 \Rightarrow i_1$	None
0.4	$i_1 \Rightarrow i_3, i_3 \Rightarrow i_1$ $i_1 i_3 \Rightarrow i_4$	$i_1 \Rightarrow i_3, i_3 \Rightarrow i_1$ $i_1 i_3 \Rightarrow i_4$	$i_3 \Rightarrow i_1$
Summary	$0.8/(i_1 \Rightarrow i_3)+$ $+1/(i_3 \Rightarrow i_1)+$ $+0.4/(i_1 i_3 \Rightarrow i_4)$	$1/(i_1 \Rightarrow i_3)+$ $+1/(i_3 \Rightarrow i_1)+$ $+0.6/(i_1 i_3 \Rightarrow i_4)$	$0.2/(i_1 \Rightarrow i_3)+$ $+0.6/(i_3 \Rightarrow i_1)$

Table 12
Summaries of RL-numbers representing support and confidence in Table 10.

$i_1 \Rightarrow i_3$		$i_3 \Rightarrow i_1$		$i_1 i_3 \Rightarrow i_4$	
supp	conf	supp	conf	supp	conf
$13/30$ ≈ 0.43	$103/150$ ≈ 0.68	$13/30$ ≈ 0.43	$28/30$ ≈ 0.93	$8/30$ ≈ 0.26	$14/30$ ≈ 0.46

row of Table 10. For instance, the membership of the rule $i_3 \Rightarrow i_1$ to the fuzzy set of strong rules in the last column is 0.6 since this rule appears in levels 0.8 and 0.4 only, hence its membership is calculated as $(0.8-0.6)+(0.4-0)=0.6$. Note that we have only studied three rules in this example, but many other rules can be extracted from the fuzzy transactions in Table 8. These rules would also appear in different levels (and also, with a certain membership, in the final fuzzy set of strong rules) in Table 10.

In addition to their membership to the fuzzy set of strong rules, we can give additional information about the rules in the form of summaries of quality measures, calculated as in Eq. (18). However let us remark that, as we have just seen, we are not using such summaries for determining whether a rule is strong or not. Table 12 shows the corresponding summaries for the RL-numbers (columns in Table 10) measuring support and confidence of the rules in our example. Let us point out that, when RLs representing itemsets are nested (hence representing fuzzy sets), the value of these summaries coincide with the measures proposed in [11] on the basis of fuzzy quantification methods. As in the case of probabilities and conditional probabilities we discussed in section 3.1, the values also coincide with those calculated using the sigma-count in the case of support, but are different in general for confidence.

4.1.2. Contributions of the RL approach

In the ordinary approaches to fuzzy association rule mining, a single scalar value is used for assessing support and confidence of fuzzy rules. Then, a crisp set of strong rules are obtained as those that exceed the user-defined minimum thresholds. On the contrary, the RL approach provides a fuzzy subset of strong rules.

The membership of rules to the final result is very useful in order to distinguish rules that seem to be similar, but represent different situations. For instance, consider the following two situations:

Table 13
Solving two different fuzzy rule mining cases using RLs with minsupp=0.05.

Level	First case		Second case	
	$supp(i_1 \Rightarrow i_2)$	$conf(i_1 \Rightarrow i_2)$	$supp(i_1 \Rightarrow i_2)$	$conf(i_1 \Rightarrow i_2)$
1	0	1	0.05	1
0.05	1	1	0.05	1
Summary	0.05	1	0.05	1
Result	$0.05/(i_1 \Rightarrow i_2)$		$1/(i_1 \Rightarrow i_2)$	

1. A dataset with 1000 fuzzy transactions and items i_1 and i_2 such that for every fuzzy transaction τ it is $\tau(i_1) = \tau(i_2) = 0.05$.
2. A dataset with 1000 fuzzy transactions and items i_1 and i_2 such that for 50 fuzzy transactions it is $\tau(i_1) = \tau(i_2) = 1$, whilst for the other 950 transactions it is $\tau(i_1) = \tau(i_2) = 0$.

In the ordinary approach to fuzzy association rules using sigma-count and minimum, we have in both cases the same solution

$$supp(i_1 \Rightarrow i_2) = 0.05 = supp(i_1) = supp(i_2)$$
$$conf(i_1 \Rightarrow i_2) = 1$$

If the minimum support threshold is less or equal than 0.05, then this rule will be found to be strong in both cases, and provided to the user in the set of strong rules obtained from the dataset. However, the situations are very different. In the first case, the membership of i_1 and i_2 is very low, so it is hard to say that the rule is supported by the dataset. In the second case, the rule is absolutely crisp and there is no doubt about its meaning and the existence of transactions supporting the rule.

When this distinction is unimportant for the user, there is no problem in applying the classical approach. When on the contrary it is relevant to distinguish both cases, RLs can help. Table 13 shows the solution of both cases by levels. Let us remark that $supp(i_1) = supp(i_2) = supp(i_1 \Rightarrow i_2)$ in all levels, and that in the case of indetermination of confidence we assign value 1 as there are no exceptions, as explained in the previous section.

We can see that when using RLs it is easy to distinguish both cases: in the first case, the rule has a membership of 0.05 to the set of strong rules, whilst in the second case, the membership is 1. This is concordant with data in each case: in the first case, the rule holds to the extent that we can accept i_1 and i_2 as items appearing in the transaction, which is 0.05. In the second case, items are in the transactions to degree 1, hence the data completely supports the rule. Note also that the summary of support and confidence of the rule is the same for both cases, and it is the same as that obtained by using sigma-counts in ordinary fuzzy association rule mining (something that is not the usual case, as we have seen before). However, in the RL approach we do not use these summaries for determining the set of strong rules. As we have pointed out, another important difference in this example is that using the RL approach we are able to provide a different membership degree of the rule to the fuzzy set of strong rules in each of the two cases, providing very important information that allows to distinguish both cases.

As a final remark, the result for the first case using sigma-count can be different if other t-norms are used. In this example, if we use the product for instance, the support of the rule drops to 0.0025, inviting to not consider the rule. However, in such case it is also $conf(i_1 \Rightarrow i_2) = conf(i_1 \Rightarrow i_1) = 1/2$, different from the expected value of 1. Hence, this solution is only valid if we accept the possibility $conf(i \Rightarrow i) < 1$ as valid for our problem.

In the following section we describe the use of RLs for mining fuzzy rules with absent items, and additional contributions of the RL approach for that problem.

4.2. *Extending quantifier-based quality assessment*

RLs have been employed in [18] for extending existing accuracy measures from the crisp case to that of association rules in fuzzy transactions in an easy, direct and unique way. The foundation of the work in [18] comes from a formalization [15] that is based in two basic notions:

- four fold contingency tables (*4ft* for short) representing the frequencies of appearance of the itemsets involved in the association rule and their negations, and
- $4ft$-*quantifiers*,[3] which represent quality measures used in the evaluation of association rules.

For a certain rule $A \Rightarrow B$, and every considered level α_i, a four fold table is defined, noted by $\mathcal{M}_{\alpha_i} = 4ft(A, B, \tilde{D}, \alpha_i)$, containing the cardinalities in level α_i associated to the crisp partition of the support of $\tilde{D}$ generated by them:

$\mathcal{M}_{\alpha_i}$	B	$\neg B$
A	a_i	b_i
$\neg A$	c_i	d_i

where a_i, b_i, c_i and d_i are non negative integers such that $a_i = |\rho_{A\wedge B}(\alpha_i)|$, $b_i = |\rho_{A\wedge\neg B}(\alpha_i)|$ and analogously with c_i and d_i.

$4ft$-quantifiers define quality measures by combining the values a_i, b_i, c_i and d_i. For instance, in level α_i, the support and confidence are particular cases of $4ft$-quantifiers defined as follows:

$$supp(A \Rightarrow B) = \frac{a_i}{a_i + b_i + c_i + d_i} \tag{24}$$

$$conf(A \Rightarrow B) = \frac{a_i}{a_i + b_i} \tag{25}$$

Many other $4ft$-quantifiers define alternative measures. This formalization offers a good framework to study the properties of these measures. The obtained measures using RLs keep the main properties that we already have when mining crisp association rules, while keeping the ordinary boolean properties at the same time [18]. Particularly, in every level it is guaranteed $a_i + b_i + c_i + d_i = |\tilde{D}|$, a property that cannot be guaranteed with fuzzy sets when a, b, c and d are calculated using sigma-counts since, similarly to other examples in previous sections, negation of items and conjunctions are involved. On the basis of the information provided by $\mathcal{M}_{\alpha_i}$ we are able to generalize every interest measure ($4ft$-quantifier) from the crisp to the fuzzy case. If necessary, a scalar summarizing the information can be provided using Eq. (18), as we have seen with other examples.

4.3. Further contributions of RLs to fuzzy association rule mining and related problems

As we mentioned before, in this work it is not our intention to be exhaustive about the potential contributions of RLs in fuzzy data mining, but to illustrate such potential with a number of significant examples.

In the particular case of mining fuzzy association rules and related problems, RLs have provided additional properties and advantages, that can be found in the literature. We summarize some of them in this section.

4.3.1. Fuzzy association rules with absent items

The problem of mining for association rules involving also the absence of items has been studied by several authors [9,48,49,51,52,2]. In this problem, a so-called negated item is added for every item in the dataset, representing the absence of the item in a transaction.

There have been also some studies on fuzzy extensions of this problem, that is, mining fuzzy rules involving negated items [27,28]. In the fuzzy case, one of the main problems discussed in the literature is that, due to the properties of fuzzy negation, some properties that are important for developing efficient algorithms are lost; particularly, closely related to our discussion in section 3.1, properties like $p(a \wedge a) = p(a)$ and $p(a \wedge b) + p(a \wedge \neg b) = p(a)$ cannot hold simultaneously using fuzzy set theories in general. As a consequence, for instance, the downward closure property necessary in the apriori-like algorithms is lost. Hence, in these algorithms it is necessary to increase the searching space including all the combinations involving negated items, since $p(a \wedge \neg b)$ cannot be obtained from $p(a \wedge b)$ and $p(b)$, with the resultant increment in time and space complexity of the mining process. This is again a consequence of the lack of Boolean structure of fuzzy sets theories.

[3] The notion of quantifier in this context is different from that of fuzzy quantification.

Table 14
RLs of the fuzzy sets of transactions in Table 8 for several itemsets with items and negated items.

Level	$\rho_{\{i_1\}}$	$\rho_{\{i_3\}}$	$\rho_{\{\neg i_1\}}$	$\rho_{\{i_1,i_3\}}$	$\rho_{\{\neg i_1,i_3\}}$
1	$\{t_1,t_2\}$	$\{t_1\}$	$\{t_3,t_4,t_5,t_6\}$	$\{t_1\}$	$\emptyset$
0.8	$\{t_1,t_2\}$	$\{t_1,t_2\}$	$\{t_3,t_4,t_5,t_6\}$	$\{t_1,t_2\}$	$\emptyset$
0.6	$\{t_1,t_2,t_4\}$	$\{t_1,t_2,t_3\}$	$\{t_3,t_5,t_6\}$	$\{t_1,t_2\}$	$\{t_3\}$
0.4	$\{t_1,t_2,t_3,t_4,t_5\}$	$\{t_1,t_2,t_3,t_4\}$	$\{t_6\}$	$\{t_1,t_2,t_3,t_4\}$	$\emptyset$

Table 15
Support of items and rules, and confidence of rules involving items i and $\neg i$ when $t(i)=0.5$ for every fuzzy transaction t.

Level	$supp(i)$	$supp(\neg i)$	$supp(i \Rightarrow i)$	$conf(i \Rightarrow i)$	$supp(i \Rightarrow \neg i)$	$conf(i \Rightarrow \neg i)$
1	0	1	0	1	0	1
0.5	1	0	1	1	0	0
Summary	0.5	0.5	0.5	1	0	0.5

In [18], the RL approach has been employed for providing a solution to the problem of mining fuzzy rules with negated items. As it has been shown before, RLs have a Boolean structure, and hence all the ordinary properties of probabilities are kept, so RLs offer suitable properties for contributing to solve the problem. The approach is similar to the RL approach presented in the previous section, with the following particularities:

- In the first step, negated items are also considered. Negated items are added to the set of items in the problem. The crisp set of transactions where the negated item $\neg i$ appears in a certain level is computed as the complement of the set of transactions where the item i appears in the same level. Table 14 shows an example for the set of fuzzy transactions in Table 8.
- Any algorithm for mining crisp association rules with negated items can be applied in each level independently and in parallel. Since in each level we have a crisp problem and all the properties of set and arithmetic operations are kept, the heuristics employed for increasing the efficiency of the search are valid in each level. In the proposal presented in [18], a BitSet approach for the representation of the presence of items in transactions at each level is employed to accelerate the calculations.
- Results by levels and the corresponding summary as a fuzzy set of strong rules are obtained as in the last section.

An important contribution of RLs to this problem is the possibility to employ the existing efficient algorithms. However, some other properties are provided by using RLs that cannot be provided by fuzzy sets, so again the RL approach offer a suitable alternative when these properties are considered to be necessary.

For instance, consider that we want to keep the property that $Conf(A \Rightarrow A)=1$ using sigma-count. Then, the minimum must be employed as t-norm. Now, consider a set of fuzzy transactions and an item i such that $t(i)=0.5$ for every transaction. Then it is

$$supp(i \Rightarrow i) = 0.5 = supp(i \Rightarrow \neg i)$$

$$conf(i \Rightarrow i) = 1 = conf(i \Rightarrow \neg i)$$

and both $i \Rightarrow i$ and $i \Rightarrow \neg i$ are strong rules for any minimum support threshold below 0.5. In general, using fuzzy sets, every transaction where $t(i) \le 0.5$ supports the rule $i \Rightarrow \neg i$ and contributes to increase its confidence.

This situation can be avoided if necessary by using RLs. Table 15 shows the corresponding computations for this example.

Note that the rule $i \Rightarrow i$ is strong for all levels in $(0, 0.5]$ despite the minimum support and confidence thresholds, so the fuzzy set of strong rules will be $0.5/(i \Rightarrow i)$ independently of the thresholds employed. This is reasonable since $t(i)=0.5$ for every transaction, and hence we do not have complete support to the rule. This information is independent from the summary of support and confidence, which are 0.5 and 1, respectively. This result is consistent with our discussion in the previous section. Now, regarding the rule $i \Rightarrow \neg i$, we can see that it is not strong in any level for any minimum threshold above 0, since in every level it is $supp(i \Rightarrow \neg i)=0$. Hence, this rule will never appear in the resulting fuzzy set of strong rules, as desired. The fact that $conf(i \Rightarrow \neg i)=0.5$ is a consequence of our treatment of the $0/0$ case which, as we pointed out before, is not problematic since when the support of the rule is 0, it is never strong.

In addition to this particular example, let us remark that:

- The rule $i \Rightarrow \neg i$ has always support 0 in every level for any collection of fuzzy transactions. Hence, this rule is not strong in any level for any dataset, and never appears in the support of the fuzzy set of strong rules.
- When it is $t(i) = 1$ for a percentage of transactions greater or equal to the minimum support, then $i \Rightarrow i$ has support and confidence 1 in every level, and appears in the final fuzzy set of strong rules with degree 1.

4.3.2. Exception/anomalies

Exception rules were first introduced by Suzuki in [50] as a pair of rules noted by (csr,exc) where csr stands for *common sense rule*, and exc represents the exception rule associated to the csr.

$$A \Rightarrow B \quad (csr)$$
$$A \wedge E \Rightarrow \neg B \quad (exc)$$

For instance, if A represents young students and B graduated in five years, we may find the following exception rule

"IF young student, THEN (s)he graduates in five years
EXCEPT when (s)he has children.(E)"

This example shows how the presence of E changes the consequent of the common sense rule $A \Rightarrow B$, in this case not graduating in five years ($\neg B$).

A similar idea was proposed by Berzal et al. in [5], where they mined anomalous rules representing, in this case, the deviation from the common sense rule (i.e. the deviation from the usual behavior).

$$X \Rightarrow Y \quad (csr)$$
$$X \wedge \neg Y \Rightarrow A \quad (anom)$$

For instance, if X represents that a vertebrate animal can fly and Y the class of birds, we can find the following anomalous rule:

"IF a vertebrate animal flies (X)
THEN it is a bird (Y) (usually);
OR the vertebrate is a mammal (A) (unusually)".

In this example, a vertebrate animal that flies it is usually a bird but unusually a mammal (A), for instance a bat. In other words: when X, then we have either Y (usually) or A (unusually).

When mining exception and anomalous rules from fuzzy transactional datasets we have to pay special attention to the treatment of the negation like in the case of absent items. Particularly, as we have seen before, using fuzzy sets we can find that both $A \Rightarrow B$ and $A \Rightarrow \neg B$ are strong rules. In such case, since the *exc* rule is not expected to have a large support (otherwise the *csr* rule would not be very confident), and its confidence is lower or equal to the confidence of $A \wedge E \Rightarrow B$, it is likely that almost every itemset E seems to be an exception to the rule. That is, we have the risk of obtaining a large amount of misleading exception rules. Something similar can happen to anomalous rules. Both problems can be avoided by using the RL approach.

Besides this, the efficiency of an algorithm to discover exception and anomalous association rules is extremely important because we have to first extract the common sense rules and after that to look for those items that satisfy the exception or the anomalous rule assumption. To this aim, in [43] a parallel algorithm was proposed employing the RL approach and the BitSet representation of items for accelerating the logical computations involving both conjunction and negation between items.

As in the case of absent items, when mining the exception and anomalous rules, keeping the structure of Boolean Algebra in set operations can be very helpful in order to compute by levels the conjunction of items involving negation. This enables a more efficient implementation since some of the computations can be reformulated using some properties of Boolean Algebras, allowing to prune the search space, and crisp algorithms can be employed in parallel in each level, as shown in [43,16].

4.3.3. Bags and bag mining

Bags, also called multisets, are set-like structures where elements can appear more than once. This is often the case in transactions; for instance, in market baskets it is usual to find more than one unit of an item (bread, milk, etc.).

This information, which is usually discarded in market basket analysis, can be useful for extracting knowledge. We can discover different types of association rules relating not only the presence of items together, but also the quantities appearing in the transactions. Existing works in this field consider different kinds of rules [30]

- Simple rules of the form "$i_1 = value_1 \Rightarrow i_2 = value_2$" relating that items i_1 and i_2 appear together with the indicated quantities ($value_j$).
- General rules of the form "$i_1 = value_1 \Rightarrow i_2 \geq value_2$" offering also information about the minimum quantity of i_2 to be related to the quantity of i_1.
- Semantic rules like "large quantity of juice ⇒ small quantity of coca-cola" which use semantic terms for describing volume quantities.

In the last example, "large" and "small" are fuzzy concepts, and hence fuzziness appear in the analysis of bag datasets. In [14] an approach is presented for mining fuzzy rules representing useful patterns involving the quantities of items in a bag dataset. Appropriate linguistic terms are defined for quantities for different items, and the bag dataset is transformed into a fuzzy transactional one. In [17] the authors proposed to use RLs to manage the fuzziness by levels yielding to the new concept of RL-bag, based on the previous approach to fuzzy bags introduced in [13]. Besides the fulfillment of boolean properties and keeping the ordinary properties of crisp bags, using level-based representation for bags is a good tool to generalize measures involving bags like similarity or distance measures that are often used in problems like decision making or clustering [42,8,39].

RL bags can be employed for mining bag datasets with fuzzy quantities since, by using the RL approach, we have crisp bags in each level and, once more, existing algorithms for mining associations in bag databases can be employed for solving the fuzzy version of the problem.

5. Conclusions

The representations by levels of fuzzy concepts offer several advantages for fuzzy data mining: crisp mining procedures can be applied directly by operating in each level independently; numbers affected by fuzziness can be defined, which are not fuzzy numbers in the usual sense; and for both concepts (represented by sets) and numbers affected by fuzziness, all properties of the arithmetic and set operations of their crisp counterparts are kept. To the best of our knowledge, this has no correspondence in fuzzy number and fuzzy set theories.

The potential contributions of RLs with respect to using only fuzzy sets in fuzzy data mining are interesting *to the extent that such properties are needed or convenient in the corresponding data mining applications*. In no way we mean that conventional fuzzy data mining is not useful or must be replaced by using RLs. We just claim and show that RLs i) are a suitable alternative to fuzzy sets for representing fuzziness, and ii) offer very important properties and possibilities that fuzzy sets cannot, providing solutions when such properties are required. They are simply an alternative to consider, in the same way that the different t-norms in fuzzy set theory are alternatives to consider when modeling a fuzzy system, and we cannot say that a certain t-norm is better than another in general. We have illustrated the potential contributions of RLs with several examples and also with descriptions of techniques in the literature that have successfully employed the RL approach in fuzzy data mining and particularly in fuzzy association rule mining.

Neither our description about the potential contributions of RLs in fuzzy data mining, nor our overview of existing fuzzy data mining techniques in the literature based on RLs, are intended to be exhaustive, but illustrative. For instance, RLs have been applied also in the setting of rough sets [10] or in fuzzy clustering [25]. In fuzzy association rule mining, many possibilities are currently present regarding the different fuzzy frameworks for mining associations proposed in [37]. Other related approaches have been applied in this field, particularly the $X - \mu$ approach [38,34,35].

The RL approach opens an enormous research field full of possibilities. Among our ideas for future work in this area we can mention to apply the RL approach to fuzzy approximate dependencies [4] on the basis of the crisp version proposed in [47], and also to gradual dependencies [31,40] on the basis of the crisp approach in [7].

Acknowledgements

This work was conceived as an homage to Prof. Miguel Delgado and refers to one of his main research lines in the last two decades which combines Fuzzy Logic, and in particular RLs, with Data Mining techniques. The work has been partially supported by the Spanish Ministry of Science, Innovation and Universities and the European Regional Development Fund (Fondo Europeo de Desarrollo Regional - FEDER) under projects TIN2014-58227-P, PGC2018-096156-B-I00 *Recuperación y Descripción de Imágenes mediante Lenguaje Natural usando Técnicas de Aprendizaje Profundo y Computación Flexible*, and TIN2015-64776-C3-1-R.

References

[1] C. Alsina, E. Trillas, On Iterative Boolean-Like Laws of Fuzzy Sets, in: EUSFLAT 2005, 2005, pp. 389–394.
[2] M.L. Antonie, O.R. Zaïane, Mining positive and negative association rules: an approach for confined rules, in: European PKDD Conference, 2004, pp. 27–38.
[3] J.F. Baldwin, J. Lawry, T.P. Martin, A mass assignement theory of the probability of fuzzy events, Fuzzy Sets and Systems 83 (1996) 353–367.
[4] F. Berzal, I. Blanco, D. Sánchez, J.M. Serrano, M.A. Vila, A definition for fuzzy approximate dependencies, Fuzzy Sets and Systems 149 (1) (2005) 105–129.
[5] F. Berzal, J.C. Cubero, N. Marín, M. Gámez, Anomalous association rules, in: IEEE International Conference on Data Mining, 2004.
[6] F. Berzal, J.C. Cubero, D. Sánchez, J.M. Serrano ART, A hybrid classification model, Mach. Learn. 54 (1) (2004) 67–92.
[7] F. Berzal, J.C. Cubero, D. Sánchez, M.A. Vila, J.M. Serrano, An alternative approach to discover gradual dependencies, Int. J. Uncertain. Fuzziness Knowl.-Based Syst. 15 (05) (2007) 559–570.
[8] R. Biswas, An application of Yager's bag theory in multicriteria based decision making problems, Int. J. Intell. Syst. 14 (1999) 1231–1238.
[9] S. Brin, R. Motwani, J.D. Ullman, S. Tsur, Dynamic itemset counting and implication rules for market basket data, SIGMOD Rec. 26 (2) (1997) 255–264.
[10] L. D'eer, C. Cornelis, D. Sánchez, Fuzzy covering based rough sets revisited, in: Proceedings IFSA-EUSFLAT 2015, 2015.
[11] M. Delgado, N. Marín, D. Sánchez, M.A. Vila, Fuzzy association rules: general model and applications, IEEE Trans. Fuzzy Syst. 11 (2) (2003) 214–225.
[12] M. Delgado, D. Sánchez, M.J. Martín-Bautista, M.A. Vila, A probabilistic definition of a nonconvex fuzzy cardinality, Fuzzy Sets and Systems 126 (2002) 177–190.
[13] M. Delgado, M.J. Martín-Bautista, D. Sánchez, M.A. Vila, An extended characterization of fuzzy bags, Int. J. Intell. Syst. 24 (2009) 706–721.
[14] M. Delgado, M.D. Ruiz, D. Sánchez, Pattern extraction from bag databases, Int. J. Uncertain. Fuzziness Knowl.-Based Syst. 16 (4) (2008) 475–494.
[15] M. Delgado, M.D. Ruiz, D. Sánchez, Studying interest measures for association rules through a logical model, Int. J. Uncertain. Fuzziness Knowl.-Based Syst. 18 (1) (2010) 87–106.
[16] M. Delgado, M.D. Ruiz, D. Sánchez, New approaches for discovering exception and anomalous rules, Int. J. Uncertain. Fuzziness Knowl.-Based Syst. 19 (2) (2011) 361–399.
[17] M. Delgado, M.D. Ruiz, D. Sánchez, RL-bags: a conceptual, level-based approach to fuzzy bags, Fuzzy Sets and Systems 208 (2012) 111–128.
[18] M. Delgado, M.D. Ruiz, D. Sánchez, J.M. Serrano, A formal model for mining fuzzy rules using the RL representation theory, Inf. Sci. 181 (23) (2011) 5194–5213.
[19] M. Delgado, M.D. Ruiz, D. Sánchez, M.A. Vila, Fuzzy quantification: a state of the art, Fuzzy Sets and Systems 242 (2014) 1–30.
[20] F. Díaz-Hermida, D.E. Losada, A. Bugarín, S. Barro, A probabilistic quantifier fuzzification mechanism: the model and its evaluation for information retrieval, IEEE Trans. Fuzzy Syst. 13 (5) (2005) 688–700.
[21] D. Dubois, H. Prade, New results about properties and semantics of fuzzy set-theoretic operators, in: P.P. Wang, S.K. Chang (Eds.), Fuzzy Sets: Theory and Applications to Policy Analysis and Information Systems, Plenum Publ., 1980, pp. 59–75.
[22] D. Dubois, H. Prade, An introduction to possibility and fuzzy logics, in: P. Smets, et al. (Eds.), Non-Standard Logics for Automated Reasoning, Academic Press, 1988, pp. 742–755.
[23] D. Dubois, H. Prade, Fuzzy intervals versus fuzzy numbers: Is there a missing concept in fuzzy set theory?, in: Linz Seminar 2005 Abstracts, 2005, pp. 45–46.
[24] D. Dubois, H. Prade, Gradual elements in a fuzzy set, Soft Comput. 12 (2008) 165–175.
[25] D. Dubois, D. Sánchez, Fuzzy clustering based on coverings, in: Towards Advanced Data Analysis by Combining Soft Computing and Statistics, Springer Berlin Heidelberg, Berlin–Heidelberg, 2013, pp. 319–330.
[26] I. Glöckner, Models defined in terms of three-valued cuts and fuzzy-median aggregation, in: Fuzzy Quantifiers, in: Studies in Fuzziness and Soft Computing, vol. 193, Springer-Verlag Berlin Heidelberg, Berlin–Heidelberg, 2006, pp. 179–220.

[27] J. Han, M. Beheshti, Discovering both positive and negative fuzzy association rules in large transaction databases, J. Adv. Comput. Intell. Intell. Inform. 10 (3) (2006) 287–294.
[28] J. Han, M. Beheshti, Mining fuzzy association rules: interestingness measure and algorithm, in: IEEE International Conference on Granular Computing, 2006, pp. 659–662.
[29] R. Hernández-León, J.A. Carrasco-Ochoa, J.F. Martínez-Trinidad, J. Hernández-Palancar, CAR-NF: a classifier based on specific rules with high netconf, Intell. Data Anal. 16 (1) (2012) 49–68.
[30] P.Y. Hsu, Y.L. Chen, C.C. Ling, Algorithms for mining association rules in bag databases, Inf. Sci. 166 (2004) 31–47.
[31] E. Hüllermeier, Association rules for expressing gradual dependencies, in: Principles of Data Mining and Knowledge Discovery, Springer, 2002, pp. 200–211.
[32] E. Hüllermeier, Fuzzy sets in machine learning and data mining, Appl. Soft Comput. 11 (2) (2011) 1493–1505.
[33] R. Kruse, C. Borgelt, D. Nauck, Problems and prospects in fuzzy data analysis, in: Benham Azvine, DetlefD Nauck, Nader Azarmi (Eds.), Intelligent Systems and Soft Computing, in: Lecture Notes in Computer Science, vol. 1804, Springer, Berlin–Heidelberg, 2000, pp. 95–109.
[34] D.J. Lewis, T.P. Martin, X-μ fuzzy association rule method, in: 13th UK Workshop on Computational Intelligence, UKCI 2013, Guildford, United Kingdom, September 9–11, 2013, pp. 144–150.
[35] D.J. Lewis, T.P. Martin, The x-μ approach: in theory and practice, in: Information Processing and Management of Uncertainty in Knowledge-Based Systems – 15th International Conference, Proceedings, Part III, IPMU 2014, Montpellier, France, July 15–19, 2014, pp. 406–415.
[36] D.J. Lewis, T.P. Martin, The X-μ fuzzy association rule method, in: 13th UK Workshop on Computational Intelligence, Proceedings, UKCI 2013, Guildford, UK, September 9–11, 2013, pp. 144–150.
[37] N. Marín, M.D. Ruiz, D. Sánchez, Fuzzy frameworks for mining data associations: fuzzy association rules and beyond, Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 6 (2) (2016) 50–69.
[38] T.P. Martin, B. Azvine, The x-mu approach: fuzzy quantities, fuzzy arithmetic and fuzzy association rules, in: IEEE Symposium on Foundations of Computational Intelligence, FOCI 2013, 2013, pp. 24–29.
[39] S. Miyamoto, Information clustering based on fuzzy multisets, Inf. Process. Manag. 39 (2003) 195–213.
[40] C. Molina, J.M. Serrano, D. Sánchez, M.A. Vila, Mining gradual dependencies with variation strength, Mathw. Soft Comput. (15) (2008) 75–93.
[41] W. Pedrycz, Fuzzy set technology in knowledge discovery, Fuzzy Sets and Systems 98 (1998) 279–290.
[42] A. Rebaï, BBTOPSIS: a bag based technique for order preference by similarity to ideal solution, Fuzzy Sets and Systems 60 (1993) 143–162.
[43] M.D. Ruiz, D. Sánchez, M. Delgado, M.J. Martin-Bautista, Discovering fuzzy exception and anomalous rules, IEEE Trans. Fuzzy Syst. 24 (4) (2016) 930–944.
[44] D. Sánchez, M. Delgado, M.A. Vila, A restriction level approach to the representation of imprecise properties, in: Int. Conference on Information Processing and Management of Uncertainty, Málaga, Spain, 2008, pp. 153–159.
[45] D. Sánchez, M. Delgado, M.A. Vila, RL-numbers: an alternative to fuzzy numbers for the representation of imprecise quantities, in: IEEE International Conference on Fuzzy Systems, 2008, pp. 2058–2065.
[46] D. Sánchez, M. Delgado, M.A. Vila, J. Chamorro-Martínez, On a non-nested level-based representation of fuzziness, Fuzzy Sets and Systems 192 (2012) 159–175.
[47] D. Sánchez, J.M. Serrano, I. Blanco, M.J. Martín-Bautista, Using association rules to mine for strong approximate dependencies, Data Min. Knowl. Discov. 16 (3) (2008) 313–348.
[48] A. Savasere, E. Omiecinski, S. Navathe, An efficient algorithm for mining association rules in large databases, in: Proceedings of the 21st Conference on Very Large Databases, Zürich, Switzerland, 1995, pp. 432–444.
[49] A. Savasere, E. Omiecinski, S. Navathe, Mining for strong negative associations in a large database of customer transactions, in: ICDE, IEEE Computer Society, 1998, pp. 494–502.
[50] E. Suzuki, Autonomous discovery of reliable exception rules, in: Proceedings of KDD-97, 1997, pp. 259–262.
[51] X. Wu, C. Zhang, S. Zhang, Mining both positive and negative association rules, in: Proceedings of ICML, 2002, pp. 658–665.
[52] X. Yuan, B. Buckles, Z. Yuan, J. Zhang, Mining negative association rules, in: Proceedings of ISCC, 2002, pp. 623–629.