Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Fuzzy Sets and Systems 149 (2005) 73–85
FUZZY sets and systems
www.elsevier.com/locate/fss

# Elicitation of fuzzy association rules from positive and negative examples

M. De Cock*, C. Cornelis, E.E. Kerre

*Fuzziness and Uncertainty Modelling Research Unit, Department of Applied Mathematics and Computer Science, Ghent University, Krijgslaan 281 (S9), 9000 Gent, Belgium*

Available online 19 August 2004

**Abstract**

The aim of this paper is to provide a crystal clear insight into the true semantics of the measures of support and confidence that are used to assess rule quality in fuzzy association rule mining. To achieve this, we rely on two important pillars: the identification of transactions in a database as positive or negative examples of a given association between attributes, and the correspondence between measures of support and confidence on one hand, and measures of compatibility and inclusion on the other hand. In this way we remove the "mystery" from recently suggested quality measures for fuzzy association rules.


*Keywords:* Fuzzy association rules; Support; Confidence; Positive and negative examples; Compatibility; Inclusion

## 1. Introduction

Association rules [1] provide a convenient and effective way to identify and represent certain dependencies between attributes in a database. Originally, association rules emerged in the domain of shops and customers; the basic idea is to identify frequent itemsets in market baskets, i.e., groups of products frequently bought together, so storekeepers may use this information to decide on what to put on sale, how to place merchandize on shelves to maximize a cross-selling effect, how to advertise, etc. Evidently, the application of association rules is not limited to marketing problems: in fact they can shed light on a wide

* Corresponding author.
*E-mail addresses:* Martine.DeCock@UGent.be (M. De Cock), Chris.Cornelis@UGent.be (C. Cornelis), Etienne.Kerre@UGent.be (E.E. Kerre).
*URL:* http://fuzzy.UGent.be.

0165-0114/$ - see front matter © 2004 Published by Elsevier B.V.
doi:10.1016/j.fss.2004.07.010

range of knowledge discovery and decision making problems. Given the massive data archives maintained by most firms nowadays, it comes as no surprise that easy-to-handle and easy-to-grasp mechanisms like association rules have risen to great popularity.

Association rule mining is traditionally performed on a data table with binary attributes. Conceptually, a record $x$ in the data table represents a customer transaction, whereas the attributes represent items that may be either purchased in that transaction, or not. Therefore, for each attribute $A$, $A(x)$ is either 1 or 0 indicating whether or not item $A$ was bought in transaction $x$. An association rule is an expression of the form $A \Rightarrow B$ in which $A$ and $B$ are attributes, such as *cheese* $\Rightarrow$ *bread*. The meaning is that when $A$ is bought in a transaction, $B$ is likely to be bought as well. In an extended approach, the antecedent and the consequent of an association rule are sets of attributes. Considering this more general definition, however, would complicate the notation without providing additional benefit for the issues we want to deal with in this paper. Furthermore, since mining algorithms tend to generate too many rules, there is a trend to focus on simple association rules, i.e., those containing only one attribute in the consequent, and use them as building blocks to construct more general rules if required [8,9].

Association rules can be rated by a number of quality measures (for a recent, comprehensive overview of what is available, we refer to [24]), among which *support* and *confidence* stand out as the two essential ones. Support measures the statistical significance of a candidate rule $A \Rightarrow B$ as the fraction of transactions in which both $A$ and $B$ were bought. Confidence assesses the strength of a rule as the fraction of transactions containing $A$ that contain $B$ as well. The basic problem of mining association rules is then to generate all association rules $A \Rightarrow B$ that have support and confidence greater than user-specified thresholds.

In most real life applications, databases contain many other attribute values besides 0 and 1. Very common for instance are quantitative attributes such as *age* or *income*, taking values from a partially ordered, numerical scale, often a subset of the real numbers. One way of dealing with a quantitative attribute like *cost* is to replace it by a few other attributes that form a crisp partition of the range of the original one, such as $low = [0, 100[$, $medium = [100, 300[$ and $high = [300, +\infty[$. Now we can consider these new attributes as binary ones that have value 1 if the *cost* attribute equals a value within their range, and 0 otherwise. In this way, the problem is reduced to the mining procedure described above (the generated rules are now called quantitative association rules [23]). From an intuitive viewpoint, it makes more sense, however, to draw values from the interval $[0, 1]$ (instead of just $\{0, 1\}$), to allow records to exhibit a given attribute to a certain extent only. In this way binary attributes are replaced by fuzzy ones. The corresponding mining process yields fuzzy (quantitative) association rules (see, e.g., [4–7,9,11,13,15–17]).

In the traditional approach to association rule mining algorithms (including quantitative and fuzzy association rule mining), one merely thinks in terms of positive examples: especially when determining the degree of support, only the number of transactions in favour of the rule is accounted for. As we argued in [11], the remaining transactions can still be partitioned into those that actually violate the rule, and those which do not carry any relevant information. In other words, “not being a positive example" of a rule is not the same as “being a negative example". Realizing this provides deeper insight into the semantics of the quality measures as we will show in this paper.

On another count, it is sometimes also useful to detect negative associations (denoted $A \Rightarrow co\, B$), whose intended meaning is that transactions containing $A$ are unlikely to contain $B$ as well. As a somewhat frivolous example, we might quote *lucky-in-love* $\Rightarrow$ *co (lucky-in-games)*. Such patterns have received quite some attention lately (see, e.g., [6,21,26,28]); we will show that they can be embedded elegantly into our framework of positive and negative examples.

The goal of this paper is not to introduce yet another series of quality measures, but to shine a bright light on what has been proposed so far, with a specific focus on the quality measures of support and confidence, and to show how the pieces of this puzzle neatly fit together. Section 2 deals with the first pillar of our argument: the identification of transactions in a database as positive or negative examples of an association between attributes. Along the way we recall the basic concepts of support and confidence, initially in the framework of crisp association rules. Soon, however, we move on to the mining of fuzzy association rules as it is specifically in this setting that new and seemingly aberrant quality measures have been proposed recently, such as non-symmetrical measures of support. The second important pillar in this paper is that support and confidence measures should actually be thought of as compatibility and inclusion measures, which we discuss in Section 3. Leaning on both pillars, in Section 4 we take the mystery out of some recently proposed quality measures for fuzzy association rules by providing crystal clear insight into their true semantics.

## 2. Positive and negative examples

### 2.1. Crisp association rules

Suppose we have a non-empty data table $X$ containing records described by their values for binary attributes $A$ belonging to a set $\mathcal{A}$ . Conceptually, the attributes correspond to the items which customers may purchase, while the records represent the transactions or market baskets. For an attribute $A$ and a record $x \in X$, $A(x) = 1$ means item $A$ was purchased in transaction $x$, while $A(x) = 0$ means $A$ was not bought. In this way, $A$ can also be thought of as the set of transactions containing the item, i.e., $x \in A$ iff $A(x) = 1$, and $x \notin A$ iff $A(x) = 0$. Likewise, $co\,A$ is the set of transactions not containing the item, i.e., $x \in co\,A$ iff $A(x) = 0$, and $x \notin co\,A$ iff $A(x) = 1$.

Let $A, B \in \mathcal{A}$ . To decide whether one of $A \Rightarrow B$ or $A \Rightarrow co\,B$ is a worthwhile association rule, we can use a number of quality measures to rate these potential rules. Most commonly used are the measures of support and confidence, which are outlined below.

*Support*. The support of an association rule $A \Rightarrow B$ is usually defined as

$$supp(A \Rightarrow B) = \frac{|A \cap B|}{|X|}, \tag{1}$$

i.e., the number of elements belonging to both $A$ and $B$, scaled to a value between 0 and 1. The idea behind the definition of support is to measure the statistical significance by counting *positive examples*, i.e., transactions that explicitly support the hypothesis expressed by the association rule. It is worth noting that the positive examples of $A \Rightarrow B$ are also those of the rule $B \Rightarrow A$, i.e., support is a symmetric measure. Hence, as can be expected, it only reveals part of the global picture. This is why we also need the confidence measure, to assess the strength of a rule.

*Confidence*. Traditionally, if a rule $A \Rightarrow B$ generates a support exceeding a user-specified threshold, it is meaningful to compute its confidence, i.e., the proportion of correct applications of the rule.

$$conf(A \Rightarrow B) = \frac{|A \cap B|}{|A|}. \tag{2}$$

Table 1
The nature of transaction $x$ w.r.t. rules $A \Rightarrow B$, $B \Rightarrow A$ and $A \Rightarrow co\, B$

| $x$ | $A \Rightarrow B$ | $B \Rightarrow A$ | $A \Rightarrow co\, B$ |
|---|---|---|---|
| Positive example | $x \in A \wedge x \in B$ | $x \in A \wedge x \in B$ | $x \in A \wedge x \notin B$ |
| Non-positive example | $x \notin A \vee x \notin B$ | $x \notin A \vee x \notin B$ | $x \notin A \vee x \in B$ |
| Negative example | $x \in A \wedge x \notin B$ | $x \notin A \wedge x \in B$ | $x \in A \wedge x \in B$ |
| Non-negative example | $x \notin A \vee x \in B$ | $x \in A \vee x \notin B$ | $x \notin A \vee x \notin B$ |

Note that $|A|$ will not be 0 if we assume that the *confidence* is computed only when the *support* exceeds a certain threshold (which should be greater than 0 to be meaningful). It is easy to see that

$$conf(A \Rightarrow B) = \frac{supp(A \Rightarrow B)}{supp(A \Rightarrow B) + supp(A \Rightarrow co\, B)}. \tag{3}$$

Having identified the "*supporters*" of $A \Rightarrow B$ as positive examples, we can ask ourselves what a *negative example* of the same rule might look like. It is clear that a transaction violates the rule $A \Rightarrow B$ as soon as it contains $A$ but not $B$. As opposed to positive examples, a negative example of $A \Rightarrow B$ is no negative example of $B \Rightarrow A$, and vice versa. Also, the complement of the set of positive examples does not necessarily equal that of negative examples, just like a "non-negative example" differs from a "positive example". On the other hand, a positive example of $A \Rightarrow co\, B$ is a negative example of $A \Rightarrow B$, and vice versa. This is summarized in Table 1. Hüllermeier [15] defined an alternative measure which calculates the ratio of support for the positive rule $A \Rightarrow B$ and the negative rule $A \Rightarrow co\, B$, or equivalently, the number of positive examples of the rule divided by the number of negative examples

$$conf_n(A \Rightarrow B) = \frac{supp(A \Rightarrow B)}{supp(A \Rightarrow co\, B)}. \tag{4}$$

It was noted in [15] that (2) and (4) are equivalent in the sense that, for $\Delta \in [0, 1]$,

$$conf(A \Rightarrow B) \geqslant \Delta \iff conf_n(A \Rightarrow B) \geqslant \frac{\Delta}{1 - \Delta}. \tag{5}$$

As opposed to (1)–(3), whose values belong to [0, 1], the value of (4) can be any positive real number. This makes the choice of a meaningful threshold harder. Furthermore the mining algorithm should have a built-in test to verify the case "$supp(A \Rightarrow co\, B) = 0$" which corresponds to infinite confidence in a rule (as there are no negative examples).

It is interesting that Dubois et al. [13] also distinguish between positive and negative examples that are grouped into sets they call $S_+$ and $S_-$, respectively. Furthermore, they introduce the class of irrelevant examples $S_\pm$ as

$$S_\pm = \{x \in X \mid x \notin A\}$$

One can easily verify that our classes of non-positive and non-negative examples are obtained as unions of $S_\pm$ with the set of positive and negative examples, respectively, i.e.,

$$x \text{ is a non-positive example} \iff x \in S_- \cup S_\pm,$$
$$x \text{ is a non-negative example} \iff x \in S_+ \cup S_\pm.$$

Also, while $S_-$, $S_+$ and $S_\pm$ form a partition of $X$, this is clearly not the case for the four classes we defined. The most important reason we choose to consider them is that they all give rise to different measures.

**Definition 1.** The quality measures $\mathrm{M}_1$, $\mathrm{M}_2$, $\mathrm{M}_3$, and $\mathrm{M}_4$ of the rule $A \Rightarrow B$ are, respectively, defined as

$$\mathrm{M}_1(A \Rightarrow B) = \frac{|A \cap B|}{|X|}, \tag{6}$$

$$\mathrm{M}_2(A \Rightarrow B) = \frac{|co\, A \cup co\, B|}{|X|}, \tag{7}$$

$$\mathrm{M}_3(A \Rightarrow B) = \frac{|A \cap co\, B|}{|X|}, \tag{8}$$

$$\mathrm{M}_4(A \Rightarrow B) = \frac{|co\, A \cup B|}{|X|}, \tag{9}$$

where *co* is the set-theoretical complement. $\mathrm{M}_1$, $\mathrm{M}_2$, $\mathrm{M}_3$ and $\mathrm{M}_4$, respectively, correspond to the number of positive examples of the rule, the number of non-positive examples, the number of negative examples, and the number of non-negative examples. In [11] these measures are therefore called minimum support, maximum opposition, minimum opposition, and maximum support, respectively.

The following propositions can be easily verified:

**Proposition 1.**

$$\mathrm{M}_1(A \Rightarrow B) \leqslant \mathrm{M}_4(A \Rightarrow B), \tag{10}$$

$$\mathrm{M}_3(A \Rightarrow B) \leqslant \mathrm{M}_2(A \Rightarrow B). \tag{11}$$

**Proposition 2.**

$$\mathrm{M}_2(A \Rightarrow B) = 1 - \mathrm{M}_1(A \Rightarrow B), \tag{12}$$

$$\mathrm{M}_3(A \Rightarrow B) = 1 - \mathrm{M}_4(A \Rightarrow B). \tag{13}$$

Hence, only two measures are independent. We can for instance choose to work with $\mathrm{M}_1$ and $\mathrm{M}_4$. The measure $\mathrm{M}_1$ corresponds to the symmetrical support measure (*supp*) of formula (1), while $\mathrm{M}_4$ is a non-symmetrical measure taking into account all examples that do not violate the rule $A \Rightarrow B$. The correspondences between the new measures can also be expressed in terms of positive and negative association rules.

**Proposition 3.**

$$\mathrm{M}_3(A \Rightarrow co\, B) = \mathrm{M}_1(A \Rightarrow B), \tag{14}$$

$$\mathrm{M}_2(A \Rightarrow co\, B) = \mathrm{M}_4(A \Rightarrow B). \tag{15}$$

Note that $conf_n$ can be expressed in terms of the measures from Definition 1:

$$conf_n(A \Rightarrow B) = \frac{\mathrm{M}_1(A \Rightarrow B)}{\mathrm{M}_3(A \Rightarrow B)}. \tag{16}$$

### *2.2. Fuzzy association rules*

Recall that a fuzzy set $A$ in $X$ is an $X \rightarrow [0, 1]$ mapping. Fuzzy-set-theoretical counterparts of complementation, intersection, and union are as usual defined by means of a negator, a t-norm, and a t-conorm. Recall that an increasing, associative and commutative $[0, 1]^2 \rightarrow [0, 1]$ mapping is called a t-norm $\mathcal{T}$ if it satisfies $\mathcal{T}(x, 1) = x$ for all $x$ in [0, 1], and a t-conorm $\mathcal{S}$ if it satisfies $\mathcal{S}(x, 0) = x$ for all $x$ in [0, 1]. A negator $\mathcal{N}$ is a decreasing $[0, 1] \rightarrow [0, 1]$ mapping satisfying $\mathcal{N}(0) = 1$ and $\mathcal{N}(1) = 0$. For $A$ and $B$ fuzzy sets in $X$ and $x$ in $X$ we define

$$co_{\mathcal{N}} A(x) = \mathcal{N}(A(x)),$$
$$A \cap_{\mathcal{T}} B(x) = \mathcal{T}(A(x), B(x)),$$
$$A \cup_{\mathcal{S}} B(x) = \mathcal{S}(A(x), B(x)).$$

Let $A(x)$ be the degree to which an attribute $A$ is bought in a transaction $x$ (or in a broader context: the degree to which $x$ satisfies the attribute). This way $A$ can be thought of as a fuzzy set in the universe of transactions, and the measures discussed above have to be generalized accordingly. The cardinality of a fuzzy set in a finite universe $X$ was introduced as a generalization of the classical concept of cardinality of a crisp set [12]. It is defined by

$$|A| = \sum_{x \in X} A(x).$$

Replacing the set-theoretical operations in Definition 1 by their fuzzy-set-theoretical counterparts (defined by means of a negator $\mathcal{N}$, a t-norm $\mathcal{T}$, and a t-conorm $\mathcal{S}$), we obtain

**Definition 2.**

$$\mathrm{M}_1(A \Rightarrow B) = \frac{1}{|X|} \sum_{x \in X} (A \cap_{\mathcal{T}} B)(x) \tag{17}$$

$$\mathrm{M}_2(A \Rightarrow B) = \frac{1}{|X|} \sum_{x \in X} (co_{\mathcal{N}} A \cup_{\mathcal{S}} co_{\mathcal{N}} B)(x), \tag{18}$$

$$\mathrm{M}_3(A \Rightarrow B) = \frac{1}{|X|} \sum_{x \in X} (A \cap_{\mathcal{T}} co_{\mathcal{N}} B)(x), \tag{19}$$

$$\mathrm{M}_4(A \Rightarrow B) = \frac{1}{|X|} \sum_{x \in X} (co_{\mathcal{N}} A \cup_{\mathcal{S}} B)(x). \tag{20}$$

Since

$$\mathcal{T}(A(x), B(x)) \leqslant B(x) \leqslant \mathcal{S}(\mathcal{N}(A(x)), B(x))$$

and

$$\mathcal{T}(A(x), \mathcal{N}(B(x))) \leqslant \mathcal{N}(B(x)) \leqslant \mathcal{S}(\mathcal{N}(A(x)), \mathcal{N}(B(x)))$$

for all $x$ in $X$, Proposition 1 still remains valid when the measures are defined by (17)–(20). The natural extension of Proposition 2 linking opposition and support measures, holds when $\mathcal{N}$ is the standard negator $\mathcal{N}_s$ (defined by $\mathcal{N}_s(x) = 1 - x$ for all $x$ in [0, 1]) and $(\mathcal{T}, \mathcal{S}, \mathcal{N}_s)$ is a de Morgan triplet, i.e.,

$$\mathcal{T}(x, y) = \mathcal{N}_s(\mathcal{S}(\mathcal{N}_s(x), \mathcal{N}_s(y)))$$

for all $x$ and $y$ in [0, 1]. For a negative fuzzy association rule, we have to specify which negator $\mathcal{N}$ we are using. For simplicity, we will assume $\mathcal{N} = \mathcal{N}_s$ and continue to write $A \Rightarrow co\, B$. It is clear that equalities (14) and (15) can also easily be retained in this way.

Generalizing the confidence measures listed above to the fuzzy case, the following formulas are obtained:

$$conf(A \Rightarrow B) = \frac{\sum_{x \in X}(A \cap_\mathcal{T} B)(x)}{\sum_{x \in X} A(x)}, \tag{21}$$

$$conf_2(A \Rightarrow B) = \frac{\sum_{x \in X}(A \cap_\mathcal{T} B)(x)}{\sum_{x \in X}(A \cap_\mathcal{T} B)(x) + \sum_{x \in X}(A \cap_\mathcal{T} co_\mathcal{N} B)(x)}. \tag{22}$$

Note that the equality of $conf$ and $conf_2$ (cf. formula (3)) is not automatically transferred to the fuzzy case, since

$$\sum_{x \in X}(A \cap_\mathcal{T} B)(x) + \sum_{x \in X}(A \cap_\mathcal{T} co_\mathcal{N} B)(x) = \sum_{x \in X} A(x) \tag{23}$$

does not always hold. As a consequence, the link (5) is not always maintained either. However, as proved in [2], if $\mathcal{N} = \mathcal{N}_s$, then (23) holds iff $\mathcal{T} = \mathcal{T}_P$. As mentioned in [13], this choice of parameters is also mandatory if one wants to ensure that the sum of the respective degrees to which a transaction $x$ is a positive example (i.e., $\mathcal{T}(A(x), B(x))$), a negative example (i.e., $\mathcal{T}(A(x), \mathcal{N}_s(B(x)))$) and an irrelevant example (i.e., $\mathcal{N}_s(A(x))$) is equal to 1. In our opinion, however, dismissing a panoply of otherwise often meaningful t-norms is a very high price to pay to enforce a property that is similar in spirit to the law of the excluded middle in classical set theory. We do, however, have the following relationship between $conf$ and $conf_2$.

**Proposition 4.** *If conf and $conf_2$ are defined by a t-norm $\mathcal{T}$ such that*

- $\mathcal{T} \leqslant \mathcal{T}_P$, *then* $conf(A \Rightarrow B) \leqslant conf_2(A \Rightarrow B)$
- $\mathcal{T} \geqslant \mathcal{T}_P$, *then* $conf(A \Rightarrow B) \geqslant conf_2(A \Rightarrow B)$

## 3. Inclusion and compatibility measures

In fuzzy set theory, inclusion is, by default, defined as follows: for $A$ and $B$ fuzzy sets in $X$, $A \subseteq B$ if and only if

$$(\forall x \in X)(A(x) \leqslant B(x)),$$

i.e., $A \subseteq B$ if and only if the membership function of $A$ fits beneath the membership function of $B$. While in many theoretical and practical settings this two-valued characterization of subsethood suffices, it could be argued that the definition is overly restrictive: just as an element can belong to a fuzzy set to varying degrees, so we may also want to talk about a fuzzy set being "more or less" a subset of another one. Many researchers [3,10,14,18,19,22,27] have tried to capture this intuition by proposing concrete operators *Inc* that take a couple of fuzzy sets $(A, B)$ as their input and return a value $Inc(A, B)$ in [0, 1] indicating the degree of subsethood of *A* to *B*.

Typically, to define fuzzy subsethood one takes a definition of classical set inclusion and tries to extend ("fuzzify") it to apply to fuzzy sets. Below we quote three distinct, but essentially equivalent, [1] definitions of the inclusion of *A* into *B*, where *A* and *B* are crisp subsets of *X*:

$$A \subseteq B \iff (\forall x \in X)(x \in A \Rightarrow x \in B), \tag{24}$$

$$\iff A = \emptyset \text{ or } \frac{|\, A \cap B \,|}{|\, A \,|} = 1, \tag{25}$$

$$\iff \frac{|\, co\, A \cup B \,|}{|\, X \,|} = 1. \tag{26}$$

While (24) is stated in strictly logical terms, the other two are based on counting the elements of a set, i.e., on cardinality, and have a probabilistic (i.e., frequentist) touch about them. It is therefore not surprising that their respective generalizations to fuzzy set theory cease to be equivalent. We might roughly state that adepts of the different crisp definitions have put fuzzy subsethood on two separate tracks, one logic-based, the other frequency-based. One situation where this distinction comes to light is when one tries to mould fuzzy inclusion measures into axiomatic characterizations by listing desirable properties for them.

For instance, formula (24) can be generalized to fuzzy sets by replacing the implication by an implicator. Recall that an implicator $\mathcal{I}$ is a $[0, 1]^2 \to [0, 1]$ mapping such that $\mathcal{I}(x, .)$ is increasing and $\mathcal{I}(., x)$ is decreasing, and $\mathcal{I}(1, x) = x$ for all *x* in [0, 1], and $\mathcal{I}(0, 0) = 1$. An inclusion measure satisfying desirable axioms is then given by

$$Inc_1(A, B) = \inf_{x \in X} \mathcal{I}(A(x), B(x))$$

However, this approach has certain disadvantages in applications. Indeed, if two fuzzy sets *A* and *B* are equal everywhere, except in the point *x* for which $A(x) = 1$ and $B(x) = 0$, then $Inc_1(A, B) = 0$. One can think of very concrete instances in which this indeed makes no sense. Imagine for instance that we are to evaluate to what extent the young people in a company are also rich. Testing subsethood of the fuzzy set of young workers into that of rich workers should then be based on the relative fraction (i.e., the *frequency*) of good earners among the youngsters, and not on whether there exists or does not exist one poor, young employee. This observation has led researchers to consider extensions to definition (25) of crisp subsethood. If *A* and *B* are fuzzy sets, then one can define the subsethood of *A* into *B* as

$$Inc_2(A, B) = \frac{|A \cap_{\mathcal{T}} B|}{|A|}$$

if $A \neq \emptyset$, and 1 otherwise (see, e.g., [19]).

In formula (26) the presence of implication is also very clear. For propositions *p* and *q* in binary logic, $p \Rightarrow q$ has the same truth value as $\neg p \vee q$. The counterpart in fuzzy logic is the so-called S-implicator

[1] Arguably, (24) is more general since it can also deal with infinite sets.

Table 2
Well-known t-norms and t-conorms

| t-norm | t-conorm |
|---|---|
| $\mathcal{T}_M(x, y) = \min(x, y)$ | $\mathcal{S}_M(x, y) = \max(x, y)$ |
| $\mathcal{T}_P(x, y) = xy$ | $\mathcal{S}_P(x, y) = x + y - xy$ |
| $\mathcal{T}_W(x, y) = \max(x + y - 1, 0)$ | $\mathcal{S}_W(x, y) = \min(x + y, 1)$ |

Table 3
Well-known implicators

| S-implicator | Residual implicator |
|---|---|
| $\mathcal{I}_{\mathcal{S}_M}(x, y) = \max(1 - x, y)$ | $\mathcal{I}_{\mathcal{T}_M}(x, y) = \begin{cases} 1 & \text{if } x \leqslant y \\ y & \text{otherwise} \end{cases}$ |
| $\mathcal{I}_{\mathcal{S}_P}(x, y) = 1 - x + xy$ | $\mathcal{I}_{\mathcal{T}_P}(x, y) = \begin{cases} 1 & \text{if } x \leqslant y \\ \frac{y}{x} & \text{otherwise} \end{cases}$ |
| $\mathcal{I}_{\mathcal{S}_W}(x, y) = \min(1 - x + y, 1)$ | $\mathcal{I}_{\mathcal{S}_W}(x, y) = \min(1 - x + y, 1)$ |

induced by $\mathcal{S}$ and $\mathcal{N}$, defined by

$$\mathcal{I}_{\mathcal{S},\mathcal{N}}(x, y) = \mathcal{S}(\mathcal{N}(x), y)$$

for all $x$ and $y$ in [0, 1]. Generalizing formula (26) hence gives rise to a softened version of $Inc_1$ in which the supremum is replaced by taking the average over all elements of $X$:

$$Inc_3(A, B) = \frac{1}{|X|} \sum_{x \in X} \mathcal{I}_{\mathcal{S},\mathcal{N}}(A(x), B(x))$$

Another well-studied class of implicators are the residual implicators $\mathcal{I}_{\mathcal{T}}$, induced by a t-norm $\mathcal{T}$ in the following way:

$$\mathcal{I}_{\mathcal{T}}(x, y) = \sup\{\lambda | \lambda \in [0, 1] \text{ and } \mathcal{T}(x, \lambda) \leqslant y\}$$

for all $x$ and $y$ in [0, 1]. Tables 2 and 3 recall some well-known t-norms and t-conorms, as well as the implicators induced by them and the standard negator (which is omitted in the notation).

Another important kind of comparison measures for fuzzy sets, the so-called compatibility measures, assess their degree of overlap (see, e.g., [25]). The so-called simple matching coefficient

$$Com_1(A, B) = \frac{|A \cap_{\mathcal{T}} B|}{|X|} = \frac{1}{|X|} \sum_{x \in X} \mathcal{T}(A(x), B(x))$$

is the average degree to which the fuzzy sets $A$ and $B$ together span the universe $X$. It is in a sense a softened version of

$$Com_2(A, B) = \sup_{x \in X} \mathcal{T}(A(x), B(x))$$

Table 4
Comparison of the contribution of some transactions to $\mathrm{M}_4(A \Rightarrow B)$

| $A(x)$ | $B(x)$ | $\mathcal{I}_{\mathcal{S}_\mathrm{M}}$ | $\mathcal{I}_{\mathcal{S}_\mathrm{P}}$ | $\mathcal{I}_{\mathcal{S}_\mathrm{W}}$ | $\mathcal{I}_{\mathcal{T}_\mathrm{M}}$ | $\mathcal{I}_{\mathcal{T}_\mathrm{P}}$ | $\mathcal{I}_{\mathcal{T}_\mathrm{W}}$ |
|---|---|---|---|---|---|---|---|
| 0.1 | 0.2 | 0.9 | 0.92 | 1 | 1 | 1 | 1 |
| 0.2 | 0.1 | 0.8 | 0.82 | 0.9 | 0.1 | 0.5 | 0.9 |
| 0.6 | 0.8 | 0.8 | 0.88 | 1 | 1 | 1 | 1 |
| 0.8 | 0.6 | 0.6 | 0.68 | 0.8 | 0.6 | 0.75 | 0.8 |
| 0.5 | 0.5 | 0.5 | 0.75 | 1 | 1 | 1 | 1 |
| 0.2 | 0.8 | 0.8 | 0.96 | 1 | 1 | 1 | 1 |
| 0.8 | 0.2 | 0.2 | 0.36 | 0.4 | 0.2 | 0.25 | 0.4 |

which is the height of the $\mathcal{T}$-intersection of fuzzy sets $A$ and $B$. Compatibility measures are symmetrical but in general not reflexive.

## 4. Into the semantics of quality measures

Forgetting about the special variants of the confidence measure (namely $conf_2$ and $conf_n$) for a moment, to assess the quality of a fuzzy association rule, so far we have the measure of confidence

$$conf(A \Rightarrow B) = \frac{|A \cap_\mathcal{T} B|}{|A|}.$$

Note that it coincides with $Inc_2$. Furthermore we have the “independent" measures $\mathrm{M}_1$ and $\mathrm{M}_4$ from Definition 2, with $\mathrm{M}_1$ corresponding to the measure of support that is commonly used for mining fuzzy association rules, i.e.,

$$supp(A \Rightarrow B) = \frac{|A \cap_\mathcal{T} B|}{|X|} \quad \text{and} \quad \mathrm{M}_4(A \Rightarrow B) = \frac{|co\, A \cup_\mathcal{S} B|}{|X|}.$$

Let us first consider these last two measures in terms of positive and negative examples: $supp(A \Rightarrow B)$ is the number of positive examples of the candidate rule $A \Rightarrow B$ and coincides with the compatibility measure $Com_1$, while $\mathrm{M}_4(A \Rightarrow B)$ corresponds to the number of non-negative examples and coincides with the inclusion measure $Inc_3$ for $\mathcal{I}$ an S-implicator. The question arises whether we can substitute the S-implicator in $\mathrm{M}_4$ (or $Inc_3$) by, e.g., a residual implicator. Table 4 shows the different contributions of several transactions $x$ to $\mathrm{M}_4(A \Rightarrow B)$ for the implicators of Table 3. As explained in Section 2, this contribution corresponds to the degree to which $x$ is a non-negative example. In most of the cases the S-implicators (on the left) and the residual implicators (on the right) behave rather similar. A striking difference, however, appears in the second example. It is caused by the low value of $A(x)$ which is taken into account much more by the S-implicators than by the residual implicators. The difference is greatest for $\mathcal{I}_{\mathcal{T}_\mathrm{M}}$ which completely ignores $A(x)$, and smallest for $\mathcal{I}_{\mathcal{T}_\mathrm{W}}$ as the implicators induced by $\mathcal{S}_\mathrm{W}$ and $\mathcal{T}_\mathrm{W}$ coincide.

An example can be called non-negative if it does not contradict the rule; so either if it is in favour of the rule, or if it does not say anything about the rule. The latter situation arises when $A(x)$ is small. In this case S-implicators tend to always identify $x$ as a non-negative example, while some residual implicators overlook it for low $B(x)$ values. Indeed if $A(x)$ is low, then $\mathcal{N}(A(x))$ tends to be high and hence so does $\mathcal{I}_{\mathcal{S},\mathcal{N}}(A(x), B(x)) = \mathcal{S}(\mathcal{N}(A(x)), B(x))$. If on the other hand we use a residual implicator $\mathcal{I}_\mathcal{T}$, referring to the definition, we are basically looking for the largest $\lambda$ such that $\mathcal{T}(A(x), \lambda) \leqslant B(x)$. If $A(x) \leqslant B(x)$ then $\lambda$ will be 1 and the transaction is identified as a non-negative example. However if $A(x)$ is low but $B(x)$ is even lower, we are in a way relying on $\lambda$ to keep $\mathcal{T}(A(x), \lambda)$ from surpassing $B(x)$. Therefore $\lambda$ tends to be low (especially if $\mathcal{T}$ is large), hence $x$ is not identified as a non-negative example.

In [15], Hüllermeier suggests the following implication-based measure of support for a fuzzy association rule $A \Rightarrow B$:

$$supp_1(A \Rightarrow B) = \sum_{x \in X} \mathcal{I}(A(x), B(x)),$$

where $\mathcal{I}$ is an implicator. Note that by dividing it by $|X|$ we obtain a formula similar to $Inc_3$. The rationale behind it is that a transaction $x$ with $A(x) = 0.6$ and $B(x) = 0.4$, only contributes to degree 0.4 to the commonly used support (which is our formula (17) defined by means of $\mathcal{T}_M$). This is considered to be low since, in the words of [17] "*x does hardly violate* (*and hence supports*) *the rule*". We fully agree on the first claim ($x$ is a non-negative example to a high degree) but not on the second one (being a non-negative example does not imply being a positive example). Although the introduction of implicators in the measures used for mining fuzzy association rules in itself can be meaningful, [17] does not respect the fundamental difference between positive and non-negative examples, which becomes evident when examining those transactions that do not really tell us something about the rule (i.e., that have a low membership degree in $A$). To deal with this problem of *"trivial support"*, Hüllermeier suggests to extend the measure of support to

$$supp_2(A \Rightarrow B) = \sum_{x \in X} \mathcal{T}(A(x), \mathcal{I}(A(x), B(x))).$$

However, if $\mathcal{I}$ is the residual implicator induced by a continuous t-norm $\mathcal{T}$ then $supp_2(A \Rightarrow B) = \sum_{x \in X} \min(A(x), B(x))$ (see, e.g., [20]) as is also noted in [17]. Therefore, in this case the new measure of support introduced in [15] reduces to the commonly used one, and hence does not offer anything new. For this reason we do not fully agree with the claim of [13] that whereas the traditional support measure (i.e. *supp* or $M_1$) is in line with the conjunction-based approach to modelling fuzzy rules, the above-defined measure $supp_2$ follows the tradition of implication-based fuzzy rules.

On another count, in [15–17] a clear preference of residual implicators over S-implicators is expressed, which seems to be in conflict with our findings described above. However, the arguments raised in favour of residual implicators in those papers, basically amount to the fact that S-implicators detect non-negative examples overlooked by most residual implicators, namely those that are not relevant to the rule. This is considered to be an unwanted side effect because Hüllermeier is exclusively trying to identify positive examples. As soon as one realizes that not the positive but the negative examples (and hence also the non-negative examples) can be revealed by means of an implicator, the preference of S-implicators over residual implicators becomes very natural.

Within the literature on fuzzy association rules there exists another view on the use of $Inc_3$ as well. Chen et al. [9] call this measure "degree of implication" and use it to replace the traditional confidence measure. This should not come as a great surprise, since their reliance on $Inc_3$ yields just another way of expressing the subsethood of $A$ into $B$.

## 5. Concluding remarks

Throughout the literature the following quality measures are prominent:

| | | | |
|---|---|---|---|
| (1) | $supp(A \Rightarrow B)$ | $\frac{1}{\lvert X\rvert} \sum_{x\in X} \mathcal{T}(A(x), B(x))$ | $Com_1$ |
| (2) | $M_4(A \Rightarrow B)$ | $\frac{1}{\lvert X\rvert} \sum_{x\in X} \mathcal{I}_{\mathcal{S},\mathcal{N}}(A(x), B(x))$ | $Inc_3$ |
| (3) | $conf(A \Rightarrow B)$ | $\frac{1}{\lvert A\rvert} \sum_{x\in X} \mathcal{T}(A(x), B(x))$ | $Inc_2$ |

The first and the third measure are generally accepted as measures of support and confidence, respectively. They assess the significance and the strength of a fuzzy association rule. They coincide with a compatibility measure and an inclusion measure from fuzzy set theory. The second measure is perceived by some as an alternative support measure, and by others as an alternative measure to assess the strength (of implication) of a rule. We prefer the latter view since this second measure is non-symmetrical and coincides with another inclusion measure.

Furthermore, when counting non-negative examples it is more meaningful to use an S-implicator than a residual implicator in measure (2). Since association rule mining is concerned with finding frequent patterns in databases it seems more natural to use cardinality based rather than logical compatibility and inclusion measures, which explains why $Inc_1$ and $Com_2$ are not met in literature on fuzzy association rules.

## Acknowledgements

The results presented above were obtained within the bilateral China-Flanders Cooperation Project 174B0201. Martine De Cock and Chris Cornelis would like to thank the Fund for Scientific Research Flanders - FWO for funding the research reported in this paper.

## References

[1] R. Agrawal, T. Imielinski, A. Swami, Mining association rules between sets of items in large databases, in: Proc. ACM SIGMOD Conf. Management of Data,1993, pp. 207–216.

[2] C. Alsina, On a family of connectives for fuzzy sets, Fuzzy Sets and Systems 16 (3) (1985) 231–235.

[3] W. Bandler, L. Kohout, Fuzzy power sets and fuzzy implication operators, Fuzzy Sets and Systems 4 (1980) 13–30.

[4] P Bosc, D. Dubois, O. Pivert, H. Prade, On fuzzy association rules based on fuzzy cardinalities, in: Proc. IEEE Internat. Fuzzy Systems Conf.,2001, pp. 461–464.

[5] P Bosc, O. Pivert, On some fuzzy extensions of association rules, in: Proc. Joint 9th IFSA Worl Congress and 20th NAFIPS Internat. Conf.,2001, pp. 1104–1109.

[6] S. Brin, R. Motwani, C. Silverstein, Beyond market baskets: generalizing association rules to correlations, in: Proc. ACM SIGMOD on Management of Data,1997, pp. 265–276.

[7] G. Chen, Q. Wei, E.E. Kerre, Fuzzy data mining: discovery of fuzzy generalized association rules, in: G. Bordogna, G. Pasi (Eds.), Recent Research Issues on Management of Fuzziness in Databases, Springer Physica-Verlag, Berlin, 2000, pp. 45–66.

[8] G. Chen, Q. Wei, D. Liu, G. Wets, Simple association rules (SAR) and the SAR-based rule discovery, Comput. Ind. Eng. 43 (2002) 721–733.

[9] G. Chen, P. Yan, E.E. Kerre, Computationally efficient mining for fuzzy implication-based association rules in quantitative databases, Internat. J. Gen. Systems, 33 (2–3) (2004) 163–182.
[10] C. Cornelis, C. Van Der Donck, E.E. Kerre, Sinha–Dougherty approach to the fuzzification of set inclusion revisited, Fuzzy Sets and Systems 134 (2) (2003) 283–295.
[11] M. De Cock, C. Cornelis, E.E. Kerre, Fuzzy association rules: a two-sided approach, in: Proc. Internat. Conf. Fuzzy Information Processing-Theories and Applications,2003, pp. 385–390.
[12] A. De Luca, S. Termini, A definition of non-probabilistic entropy in the setting of fuzzy sets, Inform. Control 20 (1972) 301–312.
[13] D. Dubois, E. Hüllermeier, H. Prade, A note on quality measures for fuzzy association rules, Fuzzy Sets and Systems; in: T. Bilgiç, B. De Baets, O. Kaynak (Eds.), Proc. 10th Internat. Fuzzy Systems Association World Congress, Lecture Notes in Computer Science, vol. 2715, Springer, Berlin, 2003, pp. 346–353.
[14] J. Fodor, R. Yager, Fuzzy set theoretic operators and quantifiers, in: D. Dubois, H. Prade (Eds.), Fundamentals of Fuzzy Sets, Kluwer, Boston, MA, 2000, pp. 125–193.
[15] E. Hüllermeier, Fuzzy association rules: semantic issues and quality measures, in: B. Reusch (Ed.), Computational Intelligence: Theory and Applications; Proc. 7th Fuzzy Days Dortmund, Lecture Notes in Computer Science, vol. 2206, Springer, Berlin, 2001, pp. 380–391.
[16] E. Hüllermeier, Implication-based fuzzy association rules, in: Proc. Fifth European Conf. on Principles and Practice of Knowledge Discovery in Databases,2001, pp. 241–252.
[17] E. Hüllermeier, Mining implication-based fuzzy association rules in databases, Proc. Ninth Internat. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems, vol. I, 2002, pp. 101–108.
[18] L. Kitainik, Fuzzy Decision Procedures with Binary Relations, Kluwer, Dordrecht, The Netherlands, 1993.
[19] B. Kosko, Fuzzy entropy and conditioning, Inform. Sci. 40 (1986) 165–174.
[20] V. Novák, I. Perfilieva, J. Mocˇkorˇ, Mathematical Principles of Fuzzy Logic, Kluwer Academic Publishers, Dordrecht, 1999.
[21] A. Savasere, E. Omiecinski, S. Navathe, Mining for strong negative associations in a large database of customer transactions, in: Proc. 1998 Internat. Conf. on Data Engineering,1998, pp. 494–502.
[22] D. Sinha, E.R. Dougherty, Fuzzification of set inclusion: theory and applications, Fuzzy Sets and Systems 55 (1) (1993) 15–42.
[23] R. Srikant, R. Agrawal, Mining quantitative association rules in large relational tables, Proc. ACM SIGMOD Conf. on Management of Data 25 (2) (1996) 1–12.
[24] P.N. Tan, V. Kumar, J. Srivastava, Selecting the Right Interestingness Measure for Association Patterns, in: Proc. ACM SIGKDD Internat. Conf. Knowledge Discovery and Data Mining, Edmonton,2002, pp. 32–41.
[25] E. Tsiporkova, H.-J. Zimmermann, Aggregation of compatibility and equality: a new class of similarity measures for fuzzy sets, in: Proc. 7th Internat. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems,1998, pp. 1769–1776.
[26] X. Wu, C. Zhang, S. Zhang, Mining Both Positive and Negative Association Rules, in: Proc. 19th Internat. Conf. on Machine Learning (ICML-2002),2002, pp. 658–665.
[27] V. Young, Fuzzy subsethood, Fuzzy Sets and Systems 77 (3) (1996) 371–384.
[28] C. Zhang, S. Zhang, Association rule mining-models and algorithms, Lecture Notes in Aritificial Intelligence, vol. 2037, Springer, Berlin, 2002.