A Clear View on Quality Measures for Fuzzy Association Rules

Martine De Cock, Chris Cornelis, and Etienne E. Kerre

Fuzziness and Uncertainty Modelling Research Unit
Department of Applied Mathematics and Computer Science
Ghent University, Krijgslaan 281 (S9), 9000 Gent, Belgium
tel: +32 9 2644742, fax: +32 9 2644995, http://fuzzy.UGent.be
{Martine.DeCock, Chris.Cornelis, Etienne.Kerre}@UGent.be

1 Introduction

In today's information–driven economy, companies may benefit a lot from suitable knowledge management. Although knowledge management is not just a technology–based concept but rather a business practice in general, the possible and even indispensable support of IT–tools in this context is obvious. Because of the large data repositories many firms maintain nowadays, an important role is played by data mining techniques that dig up useful knowledge from these large data volumes. Among them, 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. Storekeepers may use this information to decide on how to place merchandise on shelves to maximize a cross–selling effect, how to advertise, what to put on sale (for instance lowering the price of product A to attract customers, meanwhile increasing the price of product B that is frequently bought together with A), ... Evidently, the application of association rules can shed light on a wide range of decision making and marketing problems going beyond the scope of straightforward storekeeping.

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 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 [9]. 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 association rules (see e.g. [2, 4, 5, 6, 7]).

Association rules can be rated by a number of quality measures, among which *support* and *confidence* stand out as the two essential ones. Support measures the statistical significance of a candidate rule $A \Rightarrow B$, whereas confidence assesses its strength. 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. These measures can be generalized for fuzzy association rules in several ways.

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. 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 respectively (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 clear insight into their true semantics.

2 Positive and Negative Examples

2.1 Crisp Association Rules

Let X be a non–empty data table containing records described by their values for binary attributes A belonging to a set $\mathcal{A}$. 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 coA is the set of transactions not containing the item, i.e. $x \in coA$ iff $A(x) = 0$, and $x \notin coA$ iff $A(x) = 1$. Let $A, B \in \mathcal{A}$. The support of an association rule $A \Rightarrow B$ is usually defined as

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

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. 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.

$$\text{conf}(A \Rightarrow B) = |A \cap B|/|A| \tag{2}$$

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).

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". This is summarized in Table 1 (see also [4]). It is interesting that Dubois et al. [5]

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

x	$A \Rightarrow B$	$B \Rightarrow A$
positive example	$x \in A \wedge x \in B$	$x \in A \wedge x \in B$
non-positive example	$x \notin A \vee x \notin B$	$x \notin A \vee x \notin B$
negative example	$x \in A \wedge x \notin B$	$x \notin A \wedge x \in B$
non-negative example	$x \notin A \vee x \in B$	$x \in A \vee x \notin B$

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. 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 M_4 of the rule $A \Rightarrow B$ are respectively defined as

$$\mathrm{M}_1(A \Rightarrow B) = |A \cap B|/|X| \qquad \mathrm{M}_3(A \Rightarrow B) = |A \cap coB|/|X|$$

$$\mathrm{M}_2(A \Rightarrow B) = |coA \cup coB|/|X| \qquad \mathrm{M}_4(A \Rightarrow B) = |coA \cup B|/|X|$$

It can be easily verified that

$$\mathrm{M}_2(A \Rightarrow B) = 1 - \mathrm{M}_1(A \Rightarrow B) \quad \text{and} \quad \mathrm{M}_3(A \Rightarrow B) = 1 - \mathrm{M}_4(A \Rightarrow B) \quad (3)$$

Hence, only two measures are independent. We can for instance choose to work with M_1 and M_4. The measure M_1 corresponds to the symmetrical support measure (supp) of Formula (1), while M_4 is a non–symmetrical measure taking into account all examples that do not violate the rule $A \Rightarrow B$.

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 defined, as usual, 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 we define $co_{\mathcal{N}}A(x) = \mathcal{N}(A(x))$, $A \cap_{\mathcal{T}} B(x) = \mathcal{T}(A(x), B(x))$, and $A \cup_{\mathcal{S}} B(x) = \mathcal{S}(A(x), B(x))$ for all x in 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 is defined as usual as the sum of the individual membership degrees. 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. The quality measures $\mathrm{M}_1, \mathrm{M}_2, \mathrm{M}_3$, and M_4 of the rule $A \Rightarrow B$ are respectively defined as

$$\mathrm{M}_1(A \Rightarrow B) = \frac{1}{|X|} \sum_{x \in X} (A \cap_{\mathcal{T}} B)(x) \qquad \mathrm{M}_3(A \Rightarrow B) = \frac{1}{|X|} \sum_{x \in X} (A \cap_{\mathcal{T}} co_{\mathcal{N}}B)(x)$$

$$\mathrm{M}_2(A \Rightarrow B) = \frac{1}{|X|} \sum_{x \in X} (co_{\mathcal{N}}A \cup_{\mathcal{S}} co_{\mathcal{N}}B)(x) \qquad \mathrm{M}_4(A \Rightarrow B) = \frac{1}{|X|} \sum_{x \in X} (co_{\mathcal{N}}A \cup_{\mathcal{S}} B)(x)$$

The natural extension of Formula (3) 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 $[0,1]$. Generalizing the confidence measure listed above to the fuzzy case, the following formula is obtained:

$$\text{conf}(A \Rightarrow B) = \frac{\sum\limits_{x \in X} (A \cap_{\mathcal{T}} B)(x)}{\sum\limits_{x \in X} A(x)} \tag{4}$$

3 Inclusion and Compatibility of Fuzzy Sets

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:

$$A \subseteq B \iff (\forall x \in X)(x \in A \Rightarrow x \in B), \tag{5}$$

$$\iff A = \emptyset \text{ or } \frac{\mid A \cap B \mid}{\mid A \mid} = 1, \tag{6}$$

$$\iff \frac{\mid coA \cup B \mid}{\mid X \mid} = 1 \tag{7}$$

While (5) 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) flavour. It is therefore not surprising that their respective generalizations to fuzzy set theory cease to be equivalent and give rise to cardinality–based and logical inclusion measures, respectively [3]. For instance, formula (5) can be generalized to fuzzy sets by replacing the two–valued implication by a $[0,1]$–valued 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 properties 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

[1] Arguably, (5) is more general since it can also deal with infinite sets.

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 (6) 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.

In formula (7) 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 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 (7) 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: $\mathcal{I}_{\mathcal{T}}(x,y) = sup\{\lambda | \lambda \in [0,1]$ and $\mathcal{T}(x,\lambda) \leq y\}$ for all x and y in $[0,1]$.

Another important kind of comparison measures for fuzzy sets, the so–called compatibility measures, assess their degree of overlap (see e.g. [10]). 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 a softened version of

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

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 A Clear View on the Semantics of the Measures

Throughout the literature on fuzzy association rules, the quality measures listed in Table 2 are prominent. 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 (Com_1) and an inclusion measure (Inc_2) from fuzzy set theory.

Table 2. Quality measures for fuzzy association rules

(1)	$M_1(A \Rightarrow B)$ or $\text{supp}(A \Rightarrow B)$	$\frac{1}{\|X\|} \sum_{x \in X} \mathcal{T}(A(x), B(x))$	Com_1
(2)	$M_4(A \Rightarrow B)$	$\frac{1}{\|X\|} \sum_{x \in X} \mathcal{I}_{\mathcal{S},\mathcal{N}}(A(x), B(x))$	Inc_3
(3)	$\text{conf}(A \Rightarrow B)$	$\frac{1}{\|A\|} \sum_{x \in X} \mathcal{T}(A(x), B(x))$	Inc_2

The second measure $M_4(A \Rightarrow B)$ corresponds to the number of non–negative examples of the rule, and coincides with the inclusion measure Inc_3 for $\mathcal{I}$ an S–implicator. In [4] we tackled the question whether we can substitute the S–implicator in M_4 by a residual implicator, and concluded that such a replacement is not desirable. This can be roughly explained as follows: 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 correctly as a non–negative example, while some residual implicators overlook it for low $B(x)$ values. In [6], Hüllermeier suggests the following implication–based measure of support for a fuzzy association rule $A \Rightarrow B$:

$$\text{supp}_1(A \Rightarrow B) = \sum_{x \in X} \mathcal{I}(A(x), B(x))$$

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 (2) defined by means of $\mathcal{T} = \min$). This is considered to be low since, in the words of [7] *"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). Indeed the fundamental difference between positive and non–negative examples does not seem to be respected in [7], 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

$$\text{supp}_2(A \Rightarrow B) = \sum_{x \in X} \mathcal{T}(A(x), \mathcal{I}(A(x), B(x)))$$

Furthermore he is in favour of using residual implicators over S–implicators, which seems to be in conflict with our findings. In [4] we go into this in detail. However if $\mathcal{I}$ is the residual implicator induced by a continuous t-norm $\mathcal{T}$ then $\text{supp}_2(A \Rightarrow B) = \sum_{x \in X} \min(A(x), B(x))$ (see e.g. [8]) as is also noted in [7]. Therefore in this case the new measure of support introduced in [6] reduces to the commonly used one, and hence does not offer anything new. For this reason we disagree with the claim of [5] 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. Within the literature on fuzzy association rules there exists another view on the use of Inc_3 as well. Chen et al. [2] 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. For this reason we also prefer to view the non–symmetrical measure M_4 as a confidence measure. Finally, 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 in this paper 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. Agrawal R, Imielinski T, Swami A (1993) Mining Association Rules between Sets of Items in Large Databases. Proceedings of ACM SIGMOD conference on Management of Data, 207–216
2. Chen G, Yan P, Kerre E E (2004) Computationally Efficient Mining for Fuzzy Implication–based Association Rules in Quantitative Databases. To appear in: International Journal of General Systems
3. Cornelis C, Kerre E E (2003) Inclusion Measures in Intuitionistic Fuzzy Set Theory. Lecture Notes in Computer Science 2711:345–356
4. De Cock M, Cornelis C, Kerre E E (2003) Fuzzy Association Rules: a Two-Sided Approach. Proceedings of FIP2003, 385–390
5. Dubois D, Hüllermeier E, Prade H (2003) A Note on Quality Measures for Fuzzy Association Rules. Lecture Notes in Computer Science 2715:346–353
6. Hüllermeier E (2001) Fuzzy Association Rules: Semantic Issues and Quality Measures. Lecture Notes in Computer Science 2206:380–391
7. Hüllermeier E (2002) Mining Implication-Based Fuzzy Association Rules in Databases. Proceedings of IPMU2002, Vol. I:101–108
8. Novák V, Perfilieva I, Močkoř J (1999) Mathematical Principles of Fuzzy Logic. Kluwer
9. Srikant R, Agrawal R (1996) Mining Quantitative Association Rules in Large Relational Tables. Proceedings of ACM SIGMOD Conference on Management of Data, v.25 n.2:1–12
10. Tsiporkova E, Zimmermann H-J (1998) Aggregation of Compatibility and Equality: a New Class of Similarity Measures for Fuzzy Sets. Proceedings of IPMU1998, 1769–1776