A Note on Quality Measures for Fuzzy Association Rules

Didier Dubois[1], Eyke Hüllermeier[2], and Henri Prade[1]

[1] Institut de Recherche en Informatique de Toulouse, France
[2] Informatics Institute, Marburg University, Germany

Abstract. Several approaches generalizing association rules to fuzzy association rules have been proposed so far. While the formal specification of fuzzy associations is more or less straightforward, the evaluation of such rules by means of appropriate quality measures assumes an understanding of the semantic meaning of a fuzzy rule. In this respect, most existing proposals can be considered ad-hoc to some extent. In this paper, we suggest a theoretical basis of fuzzy association rules by generalizing the classification of the data stored in a database into positive, negative, and irrelevant examples of a rule.

1 Introduction

Association rules provide a means for representing dependencies between attributes in databases. Typically, an association involves two sets, A and B, of so-called items (binary features). Then, the intended meaning of a (binary) rule $A \rightharpoonup B$ is that a data record stored in the database that contains the set of items A is likely to contain the items B as well. Example: If a data record is a purchase, the association $\{\text{paper}, \text{envelopes}\} \rightharpoonup \{\text{stamps}\}$ suggests that a purchase containing paper and envelopes is likely to contain stamps as well. Several efficient algorithms for mining association rules in large databases have already been devised [1, 18].

A generalization of binary association rules is motivated by the fact that a database is usually not restricted to binary attributes but also contains attributes with values ranging on (completely) ordered scales, such as cardinal or ordinal attributes. In *quantitative association rules*, attribute values are specified by means of subsets which are typically intervals. Example: "Employees at the age of 30 to 40 have incomes between $50,000 and $70,000".

The use of fuzzy sets in connection with association rules – as with data mining in general [17] – has recently been motivated by several authors (e.g. [3, 6, 7, 16]). Moving from set-based (interval-based) to fuzzy associations is formally accomplished by replacing sets (intervals) by fuzzy sets (fuzzy intervals). Still, the evaluation of fuzzy associations through appropriate quality measures, notably the well-known support and confidence measures, is more intricate. Especially, it assumes an understanding of the semantics of a fuzzy rule. In this respect, many existing proposals can be considered ad-hoc to some extent. Here, we suggest a theoretical justification of existing measures by generalizing the classification of stored data into positive, negative, and irrelevant examples of a rule.

By way of background, Section 2 reviews classical association rules, and Section 3 gives a brief overview of existing approaches to fuzzy associations. The idea of basing the support and confidence of a fuzzy association on a fuzzy partition of examples is presented in Section 4.

2 Association Rules

Consider a set $\mathcal{A} = \{a_1, \ldots, a_m\}$ of items, and let a transaction (data record) be a subset $T \subseteq \mathcal{A}$. Let $D_X \doteq \{T \in D \,|\, X \subseteq T\}$ denote the transactions in the database D that contain the items $X \subseteq \mathcal{A}$; the cardinality of this set is $|D_X| = \mathrm{card}(D_X)$. In order to find "interesting" association rules in a database D, a potential rule $A \rightharpoonup B$ is generally rated according to several criteria. For each criterion an appropriate measure is defined, and none of these measures must fall below a certain (user-defined) threshold. In common use are the following measures: A measure of *support* defines the number of transactions in D that contain both A and B:

$$\mathsf{supp}(A \rightharpoonup B) \doteq |D_{A \cup B}|. \tag{1}$$

Support can also be defined by the proportion rather than the absolute number of transactions, in which case (1) is divided by $|D|$. The *confidence* is the proportion of correct applications of the rule:

$$\mathsf{conf}(A \rightharpoonup B) \doteq \frac{|D_{A \cup B}|}{|D_A|}. \tag{2}$$

Further reasonable measures can be considered such as, e.g., the deviation (significance) $\mathsf{int}(A \rightharpoonup B) \doteq |D_{A \cup B}| \cdot |D_A|^{-1} - |D_B| \cdot |D|^{-1}$, expressing that $A \rightharpoonup B$ is interesting only if the occurrence of A does indeed have a positive influence on the occurrence of B. As can be seen, the support measure plays a central role. In fact, all other measures can generally be derived from the support. For example, the confidence of an association $A \rightharpoonup B$ is the support of that association divided by the support of its antecedent, A.

Rather than looking at a transaction T as a subset of items, it can also be seen as a sequence $(x_1, \ldots, x_m)$ of values of binary variables $X_\imath$ with domain $\mathfrak{D}_{X_\imath} = \{0, 1\}$, where $x_\imath = 1$ if the $\imath$th item, $a_\imath$, is contained in T and $x_\imath = 0$ otherwise.

Now, let X and Y be quantitative attributes (such as age or income) with completely ordered domains $\mathfrak{D}_X$ and $\mathfrak{D}_Y$, respectively. Without loss of generality we can assume that $\mathfrak{D}_X, \mathfrak{D}_Y \subseteq \mathfrak{R}$, where $\mathfrak{R}$ denotes the set of real numbers. Let x_T and y_T denote, respectively, the values that X and Y take for the transaction T. A quantitative, interval-based association rule $A \rightharpoonup B$ involving the variables X and Y is then of the following form:

$$\text{If } X \in A = [x_1, x_2] \text{ then } Y \in B = [y_1, y_2], \tag{3}$$

where $x_1, x_2 \in \mathfrak{D}_X$ and $y_1, y_2 \in \mathfrak{D}_Y$. This approach can simply be generalized to the case where X and Y are multi-dimensional variables and, hence, A and B hyper-rectangles rather than intervals. Subsequently, we proceed from fixed variables X and Y, and consider the database D as a collection of data points $(x, y) = (x_T, y_T)$, i.e. as a projection of the original database to $\mathfrak{D}_X \times \mathfrak{D}_Y$.

Note that the above quality measures are applicable in the quantitative case as well:

$$\mathsf{supp}(A \rightharpoonup B) = \text{card}\left(\{(x,y) \in D \,|\, x \in A \wedge y \in B\}\right),$$

$$\mathsf{conf}(A \rightharpoonup B) = \frac{\text{card}\left(\{(x,y) \in D \,|\, x \in A \wedge y \in B\}\right)}{\text{card}\left(\{(x,y) \in D \,|\, x \in A\}\right)}.$$

In fact, each interval $A = [x_1, x_2]$ does again define a binary attribute $X_A(x)$ defined by $X_A(x) = 1$ if $x \in A$ and 0 otherwise. In other words, each quantitative attribute X is replaced by k binary attributes $X_{A_\imath}$ such that $\mathfrak{D}_X \subseteq \bigcup_{\imath=1}^{k} A_\imath$.

3 Fuzzy Association Rules

Replacing the sets (intervals) A and B in (3) by fuzzy sets (intervals) leads to fuzzy (quantitative) association rules. Thus, a fuzzy association rule is understood as a rule of the form $A \rightharpoonup B$, where A and B are now fuzzy subsets rather than crisp subsets of the domains $\mathfrak{D}_X$ and $\mathfrak{D}_Y$ of variables X and Y, respectively. In other words, a variable X is now replaced by a number of fuzzy attributes rather than by a number of binary attributes.

The standard approach to generalizing the quality measures for fuzzy association rules is to replace set-theoretic operations, namely Cartesian product and cardinality, by corresponding fuzzy set-theoretic operations:

$$\mathsf{supp}(A \rightharpoonup B) \doteq \sum_{(x,y) \in D} A(x) \otimes B(y), \tag{4}$$

$$\mathsf{conf}(A \rightharpoonup B) \doteq \frac{\sum_{(x,y) \in D} A(x) \otimes B(y)}{\sum_{(x,y) \in D} A(x)}, \tag{5}$$

where $\otimes$ is a t-norm; the usual choice is $\otimes = \min$. Note that the support of $A \rightharpoonup B$ can be expressed by the sum of the *individual supports*, provided by tuples $(x, y) \in D$:

$$\mathsf{supp}_{[x,y]}(A \rightharpoonup B) = A(x) \otimes B(y). \tag{6}$$

According to (6), the tuple (x, y) supports $A \rightharpoonup B$ if both, $x \in A$ and $y \in B$.

Let us mention the possibility of measuring the frequency (support) of a fuzzy itemset $A \cup B$ by a *fuzzy cardinality*, i.e. a fuzzy number, rather than by a single number [4, 9].

The support measure (6) is obviously in line with the conjunction-based approach to modeling fuzzy rules, well-known from Mamdani-like fuzzy control systems. Taking into account the asymmetric nature of a rule, the use of implication-based fuzzy rules and, hence, of implication operators in place of conjunctions for the modeling of associations has been proposed by some authors [5, 6, 13]. For example, the following type of measure was suggested in [13]:

$$\mathsf{supp}_{[x,y]}(A \rightharpoonup B) = A(x) \otimes \big(A(x) \rightsquigarrow B(y)\big)\,. \tag{7}$$

As one advantage of taking the implicative nature of a rule into account, note that (7) avoids the following questionable property of (4) and (5): Suppose that attribute A is *perfectly associated* with attribute B, which means that $A(x) = B(y)$ for all tuples $(x, y) \in D$. Thus, one may find it natural that $A \rightharpoonup B$ has full confidence. Yet, since $\alpha \otimes \alpha < \alpha$ if $\otimes$ is not idempotent (i.e. $\otimes \neq \min$), (5) usually yields $\mathsf{conf}(A \rightharpoonup B) < 1$ [14]. (Note that (7) is equivalent to (6) with minimum t-norm if $\otimes$ is continuous and $\rightsquigarrow$ is the R-implication induced by $\otimes$.)

4 Fuzzy Partitions of Examples

The key idea of the approach as outlined in this section is to provide a sound basis of fuzzy association rules by generalizing the classification of data into *positive*, *negative*, and *irrelevant* examples of a rule. In fact, an set-based association rule $A \rightharpoonup B$ partitions the database into three types of transactions, namely *positive examples* $\mathcal{S}_+$ that verify the rule, *negative examples* $\mathcal{S}_-$ that falsify the rule, and *irrelevant examples* $\mathcal{S}_\pm$:

$$\mathcal{S}_+ \doteq \{(x,y) \,|\, x \in A \wedge y \in B\} \tag{8}$$

$$\mathcal{S}_- \doteq \{(x,y) \,|\, x \in A \wedge y \notin B\} \tag{9}$$

$$\mathcal{S}_\pm \doteq \{(x,y) \,|\, x \notin A\} \tag{10}$$

The most important quality measures for association rules (support and confidence) are expressed in a natural way in terms of the cardinality of the above sets. Namely, the support is the number of positive examples, and the confidence is the number of positive over the number of relevant examples:

$$\mathsf{supp}(A \rightharpoonup B) \doteq |\mathcal{S}_+|, \quad \mathsf{conf}(A \rightharpoonup B) \doteq |\mathcal{S}_+| \cdot \big(\,|\mathcal{S}_+| + |\mathcal{S}_-|\,\big)^{-1}$$

The basic question in connection with fuzzy association rules now concerns the generalization of the partition (8–10). Clearly, if A and B are fuzzy sets rather than ordinary sets, then $\mathcal{S}_+$, $\mathcal{S}_-$, and $\mathcal{S}_\pm$ will be fuzzy sets as well. In other words, a point (x, y) can be a positive (negative) example to some degree, and may also be irrelevant to some extent. We denote by $\mathcal{S}_+(x, y)$ the degree of membership of the point (x, y) in the fuzzy set $\mathcal{S}_+$ of positive examples and employ the same notation for $\mathcal{S}_-$ and $\mathcal{S}_\pm$.

There are different ways to proceed since the logical specification of positive and negative examples is not unique. In fact, the logical specification of irrelevant examples via $(x, y) \in \mathcal{S}_\pm \Leftrightarrow \neg\,(x \in A)$ is actually clear, but there are different options to characterize $\mathcal{S}_+$ and $\mathcal{S}_-$. A straightforward possibility is of course

$$\begin{array}{lcl} (x,y) \in \mathcal{S}_+ & \doteq & (x \in A) \;\wedge\; (y \in B), \\ (x,y) \in \mathcal{S}_- & \doteq & (x \in A) \;\wedge\; \neg(y \in B). \end{array} \tag{11}$$

Still, a viable alternative could be

$$\begin{array}{lcl} (x,y) \in \mathcal{S}_+ & \doteq & \quad (x \in A) \ \wedge \ (y \in B), \\ (x,y) \in \mathcal{S}_- & \doteq & \neg((x \in A) \Rightarrow (y \in B)), \end{array} \tag{12}$$

where $\Rightarrow$ is the standard logical (material) implication. Moreover, referring to (7), one could think of

$$\begin{array}{lcl} (x,y) \in \mathcal{S}_+ & \doteq & (x \in A) \ \wedge \quad ((x \in A) \Rightarrow (y \in B)), \\ (x,y) \in \mathcal{S}_- & \doteq & (x \in A) \ \wedge \ \neg((x \in A) \Rightarrow (y \in B)). \end{array} \tag{13}$$

When taking (11) and the standard negation $\alpha \mapsto 1 - \alpha$ as a point of departure, our problem can be specified as follows: Find a generalized conjunction (t-norm) $\otimes$ such that

$$\mathcal{S}_+(x,y) + \mathcal{S}_-(x,y) + \mathcal{S}_\pm(x,y) = 1 \tag{14}$$

holds for all $(x,y) \in \mathfrak{D}_X \times \mathfrak{D}_Y$, where

$$\begin{array}{l} \mathcal{S}_+(x,y) \doteq A(x) \otimes B(y) \\ \mathcal{S}_-(x,y) \doteq A(x) \otimes (1 - B(y)) \\ \mathcal{S}_\pm(x,y) \doteq 1 - A(x) \end{array} \tag{15}$$

From Alsina's results in [2] it follows that the only t-norm solving this problem is the product. In fact, Alsina even solves a somewhat more general problem, seeking solutions $(\otimes, \oplus, n)$ to the functional equation

$$(\alpha \otimes \beta) \oplus (\alpha \otimes n(\beta)) = \alpha$$

for all $0 \leq \alpha, \beta \leq 1$, where $\oplus$ is a t-conorm and $n(\cdot)$ a negation. However, assuming addition (that is the Lukasiewicz t-conorm) as a generalized conjunction is clearly reasonable in our context of data mining, where we are basically interested in generalizing frequency information. Particularly, (14) guarantees that $|\mathcal{S}_+| + |\mathcal{S}_-| + |\mathcal{S}_\pm| = |D|$, which is clearly a reasonable property.

It should be noted that questions of similar type have also been studied, e.g., in fuzzy preference modeling, where the problem is to decompose a weak (valued) preference relation into three parts: strict preference, indifference, and incompatibility [10].

When taking (12) rather than (11) as a point of departure, the problem is to find a generalized conjunction (t-norm) $\otimes$ and a generalized implication operator $\rightsquigarrow$ such that (14) holds with

$$\begin{array}{l} \mathcal{S}_+(x,y) \doteq A(x) \otimes B(y) \\ \mathcal{S}_-(x,y) \doteq 1 - \big(A(x) \rightsquigarrow B(y) \big) \\ \mathcal{S}_\pm(x,y) \doteq 1 - A(x) \end{array} \tag{16}$$

Note that (14) in conjunction with (16) implies

$$\alpha \rightsquigarrow \beta \ = \ (1 - \alpha) + (\alpha \otimes \beta) \tag{17}$$

for all $0 \leq \alpha, \beta \leq 1$ and, hence, suggests a definition of the implication $\rightsquigarrow$ in terms of the conjunction $\otimes$. In fact, (17) defines the QL-implication with t-conorm $(\alpha, \beta) \mapsto \min\{1, \alpha + \beta\}$ as a disjunction ($0 \leq (1-\alpha) + (\alpha \otimes \beta) \leq 1$ always holds since $\alpha \otimes \beta \leq \alpha$ for any t-norm $\otimes$). Here are some examples of standard conjunctions $\otimes$ together with induced implications:

$\otimes$	$\rightsquigarrow$
$\min\{\alpha, \beta\}$	$\min\{1, 1-\alpha+\beta\}$
$\alpha\beta$	$1-\alpha(1-\beta)$
$\max\{\alpha+\beta-1, 0\}$	$\max\{1-\alpha, \beta\}$

The question concerning the operators $\otimes$ and $\rightsquigarrow$ that can be chosen in (16) can be stated as follows: For which t-norms $\otimes$ does (17) define a proper implication operator? Note that the boundary conditions $\alpha \rightsquigarrow 1 = 1$ and $0 \rightsquigarrow \beta = 1$ do hold for all $0 \leq \alpha, \beta \leq 1$. Apart from that, (17) is obviously increasing in β. Thus, as a major point it remains to guarantee comonotonicity in α. (Of course, apart from that further properties of $\rightsquigarrow$ might be required.)

First of all, let us show that indeed not all t-norms are admissible, i.e. there are t-norms $\otimes$ for which (17) is not monotone decreasing in α. In fact, a simple counter-example is the (weakly) drastic product ($\alpha \otimes \beta \doteq \min\{\alpha, \beta\}$ if $\max\{\alpha, \beta\} = 1$ and 0 otherwise), for which (17) becomes

$$\alpha \rightsquigarrow \beta = \begin{cases} 1 & \text{if} \quad \beta = 1 \\ \beta & \text{if} \quad \alpha = 1 \\ 1-\alpha & \text{if} \quad \alpha < 1 \end{cases}.$$

Besides, there are even continuous t-norms that violate the above monotonicity condition. For instance, consider the Hamacher family [12] of t-norms:

$$\alpha \otimes_\gamma \beta \doteq \frac{\alpha\beta}{\gamma + (1-\gamma)(\alpha + \beta - \alpha\beta)}, \tag{18}$$

where γ is a non-negative parameter. With $\gamma = 10$, (17) yields $0.9 \rightsquigarrow 0.5 \approx 0.41 < 0.5 = 1 \rightsquigarrow 0.5$. Similar counter-examples can also be constructed for the families of t-norms introduced by Yager, Schweizer-Sklar, and Dombi [15].

Note that the comonotonicity condition

$$(\alpha \leq \alpha') \;\Rightarrow\; 1 - \alpha + (\alpha \otimes \beta) \;\geq\; 1 - \alpha' + (\alpha' \otimes \beta)$$

is equivalent to

$$(\alpha \leq \alpha') \;\Rightarrow\; (\alpha' \otimes \beta) - (\alpha \otimes \beta) \;\leq\; \alpha' - \alpha. \tag{19}$$

Thus, it follows that a t-norm $\otimes$ is admissible in (17) if it is a so-called *copula*. In fact, the following result is stated as a theorem in [19]: A t-norm $\otimes$ is a copula iff (19) holds. A related result concerns continuous Archimedean t-norms in particular and shows that such t-norms are admissible if and only if their additive generator is convex.

For many parameterized families of t-norms, the latter result makes it easy to check whether or not a parameter is admissible. For instance, $\gamma \leq 1$ is necessary for the Hamacher family (18).

As a direct consequence of the above results one can prove

Proposition 1: The Lukasiewicz t-norm $\otimes_L : (\alpha, \beta) \mapsto \max\{\alpha + \beta - 1, 0\}$ is the smallest t-norm admissible in (17).

Proposition 2: For the family of Frank t-norms [11], parameterized through $\rho > 0$ according to

$$\otimes_\rho : (\alpha, \beta) \mapsto \begin{cases} \min(\alpha, \beta) & \text{if } \rho = 0 \\ \alpha\beta & \text{if } \rho = 1 \\ \max\{0, 1 - \alpha + \beta\} & \text{if } \rho = \infty \\ \ln_\rho \left(1 + \frac{(\rho^\alpha - 1)(\rho^\beta - 1)}{\rho - 1}\right) & \text{otherwise} \end{cases},$$

(17) is always monotone decreasing in α.

A further interesting result concerns the possibility of combining admissible t-norms into new admissible t-norms.

Proposition 3: The ordinal sum of admissible t-norms is again admissible.

Corollary 4: Each element of the family of t-norms

$$\otimes_\gamma : (\alpha, \beta) \mapsto \frac{\alpha\beta}{\max\{\alpha, \beta, \gamma\}}, \qquad 0 < \gamma \leq 1 \tag{20}$$

introduced by Dubois and Prade [8], is admissible in (17).

Finally, reconsider model (13). In conjunction with (14), we obtain

$$(\alpha \otimes \beta) + (\alpha \otimes \neg\beta) = \alpha$$

with $\alpha = A(x)$ and $\beta = A(x) \rightsquigarrow B(y)$. Thus, we can again refer to Alsina's result, showing that $\otimes$ should be the product. Apart from that, any implication operator can be used. It should be noted, however, that some implications are unacceptable when imposing further requirements. For example, the reasonable property that $\mathcal{S}_+(x, y)$ is upper-bounded by $B(y)$, especially $\mathcal{S}_+(x, y) = 0$ if $B(y) = 0$, is not satisfied by all operators.

5 Concluding Remarks

The approach outlined in this paper justifies the use of certain fuzzy logical operators in connection with different types of support measures for fuzzy itemsets. Particularly, the t-norm generalizing the logical conjunction should be either the product (and not, as usually, the minimum!) or a so-called copula, depending on how positive and negative negative examples are specified.

Since the membership of a tuple (x, y) in the fuzzy set of *relevant* examples (the complement of $\mathcal{S}_\pm$) is $A(x)$ in any case, our approach also justifies the standard confidence measure (5).

Our results might appear not fully satisfactory since they still permit a rather large class of support measures. Restricting this class further by assuming additional properties is hence an important topic of ongoing research.

References

1. R. Agrawal and R. Srikant. Fast algorithms for mining association rules. In *Proceedings of the 20th Conference on* VLDB, Santiago, Chile, 1994.
2. C. Alsina. On a family of connectives for fuzzy sets. *Fuzzy Sets and Systems*, 16:231–235, 1985.
3. Wai-Ho Au and K.C.C. Chan. An effective algorithm for discovering fuzzy rules in relational databases. In *Proceedings* IEEE *World Congress on Computational Intelligence*, pages 1314 –1319, 1998.
4. P. Bosc, D. Dubois, O. Pivert, and H. Prade. On fuzzy association rules based on fuzzy cardinalities. In *Proc. IEEE Int. Fuzzy Systems Conference*, Melbourne, 2001.
5. P. Bosc and O. Pivert. On some fuzzy extensions of association rules. In *Proc. IFSA/NAFIPS-2001*, Vancouver, Canada, 2001.
6. G. Chen, Q. Wei, and E.E. Kerre. Fuzzy data mining: Discovery of fuzzy generalized association rules. In G. Bordogna and G. Pasi, editors, *Recent Issues on Fuzzy Databases*. Springer-Verlag, 2000.
7. M. Delgado, D. Sanchez, and M.A. Vila. Acquisition of fuzzy association rules from medical data. In S. Barro and R. Marin, editors, *Fuzzy Logic in Medicine*. Physica Verlag, 2000.
8. D. Dubois and H. Prade. *Fuzzy Sets and Systems: Theory and Applications*. Academic Press, New York, 1980.
9. D. Dubois and H. Prade. Fuzzy sets in data summaries – outline of a new approach. In *Proceedings* IPMU-2000, pages 1035–1040, Madrid, Spain, 2000.
10. J. Fodor and M. Roubens. *Fuzzy Preference Modelling and Multicriteria Decision Support*. Kluwer, 1994.
11. M.J. Frank. On the simulataneous associativity of $f(x, y)$ and $x + y - f(x, y)$. *Aeq. Math.*, 19:194–226, 1979.
12. H. Hamacher. *Über logische Aggregationen nichtbinär explizierter Entscheidungskriterien; Ein axiomatischer Beitrag zur normativen Entscheidungstheorie*. R.G. Fischer Verlag, 1978.
13. E. Hüllermeier. Implication-based fuzzy association rules. In *Proceedings* PKDD–01, pages 241–252, Freiburg, Germany, September 2001.
14. E. Hüllermeier and J. Beringer. Mining implication-based fuzzy association rules in databases. In B. Bouchon-Meunier, L. Foulloy, and R.R. Yager, editors, *Intelligent Systems for Information Processing: From Representation to Applications*. Elsevier, 2003. To appear.
15. G.J. Klir and B. Yuan. *Fuzzy Sets and Fuzzy Logic – Theory ad Applications*. Prentice Hall, 1995.
16. C. Man Kuok, A. Fu, and M. Hon Wong. Mining fuzzy association rules in databases. *SIGMOD Record*, 27:41–46, 1998.
17. W. Pedrycz. Data mining and fuzzy modeling. In *Proc. of the Biennial Conference of the NAFIPS*, pages 263–267, Berkeley, CA, 1996.
18. A. Savasere, E. Omiecinski, and S. Navathe. An efficient algorithm for mining association rules in large databases. In VLDB–95, Zurich, 1995.
19. B. Schweizer and A. Sklar. *Probabilistic Metric Spaces*. North Holland, 1983.