FUZZY ASSOCIATION RULES: A TWO-SIDED APPROACH

M. De Cock C. Cornelis E. E. Kerre
Dept. of Applied Mathematics and Computer Science
Ghent University, Krijgslaan 281 (S9), B-9000 Gent, Belgium
phone: +32 9 264 47 72, fax: +32 9 264 49 95
homepage: http://fuzzy.rug.ac.be
Email:{Martine.DeCock, Chris.Cornelis, Etienne.Kerre}@rug.ac.be

Abstract

We bring to the surface the fundamental two-sidedness of knowledge in the framework of association rules, until now only slumberingly present in the measures of support and confidence. We identify the set of positive as well as the set of negative examples which are not necessarily complementary. Taking this into account we introduce new quality measures comprising the existing ones. Finally, we carefully examine the generalization of our findings to fuzzy association rules.

Keyword: fuzzy association rule, support, opposition, positive example, negative example

1 INTRODUCTION

The idea behind association rules [1] is straightforward and effective, which, together with the increasing availability of large databases, probably accounts for their success story. Coming from the world of shops and customers, the underlying mechanism aims to identify frequent itemsets in market baskets, i.e. groups of products frequently bought together. This valuable information helps shop-keepers to make decisions about what to put on sale, how to place merchandise on shelfs to maximize a cross-selling effect etc. Needless to say the same mechanism can be exploited for knowledge discovery in databases in general.

Suppose we have a data table containing records described by binary attribute values. Let X be the non-empty, finite universe of these records. Each record x in X corresponds to a transaction (a market basket). For each attribute A, $A(x)$ is either 1 or 0 indicating whether or not attribute A was purchased in transaction x. An association rule is an expression of the form $A \Rightarrow B$ in which A and B are attributes (or sets of attributes), such as *bread* $\Rightarrow$ *butter*. The meaning is that when A is bought in a transaction, B is likely to be bought as well. The symbol $\Rightarrow$ does not have any further (mathematical) meaning. There exist several measures to express the quality of an association rule, such as the support (the number of transactions in which both A and B were bought) and the confidence (the percentage of transactions containing A that contain B as well). The problem of mining association rules is to generate all association rules that have support and confidence greater than user-specified thresholds.

In this respect association rule mining algorithms only look at positive examples: especially when determining the degree of support they only check how many of the transactions are in favour of the rule. However, in the set of transactions not in favour of the rule, an interesting distinction can be made between those examples that contradict the rule and those that do not carry relevant information for the rule. In ignoring this distinction, traditional association rule mining algorithms do not address the fundamental two-sidedness of knowledge.

Active exploitation of this two-sidedness can enrich information technologies significantly. A striking example is the rapid growing theory of intuitionistic fuzzy sets [2]: by complementing the membership degree (familiar from fuzzy sets) with a non-membership degree, a whole new spectrum of knowledge can be expressed. In this paper we take a similar view, namely that "not being a positive example" of a rule (i.e. not being a transaction that supports the rule) is not the same as "being a negative example" (i.e. a transaction that contradicts the rule). In Section 2 we will point out the true nature of positive and negative examples in the framework of association rule mining, and, taking them into account, we will define new measures of support and confidence comprising the commonly used measures. In Section 3 we will carefully examine the generalization of our findings to fuzzy association rules.

2 POSITIVE AND NEGATIVE EXAMPLES

For ease of notation we will use the same symbol A to denote the attribute A and the set of transactions having attribute value $A(x) = 1$, i.e. $A(x) = 1$ iff $x \in A$, and $A(x) = 0$ iff $x \notin A$. In this way, A is a subset of X. Furthermore we will only deal with simple association rules $A \Rightarrow B$ in which A and B are both attributes (and not sets of attributes; note that this is not a real limitation since we can always introduce a new attribute combining several others).

Support. The support of an association rule $A \Rightarrow B$ is usually defined as[1]

$$\text{supp}(A \Rightarrow B) = |A \cap B|$$

i.e. the number of elements belonging to both A and B. Indeed only those elements can be seen as *positive examples*, fully supporting the rule $A \Rightarrow B$. Note that exactly those elements are also the positive examples of the rule $B \Rightarrow A$, i.e. the support of $B \Rightarrow A$ is the same as that of $A \Rightarrow B$.

As soon as one identifies these "supporters" of $A \Rightarrow B$ as positive examples, the question arises what a *negative example* might look like. Note that the rule $A \Rightarrow B$ is contradicted by exactly those records belonging to A but not to B. The notion "negative example of $A \Rightarrow B$" is distinct from "negative example of $B \Rightarrow A$" as is shown in Table 1. It is also clear from this table that "negative example" differs from "non-positive example", just like "positive example" and "non-negative example" are distinct notions. The most interesting part however is that they all give rise to different measures, as shown in Table 2. Naturally,

$$\text{minsupp}(A \Rightarrow B) \leq \text{maxsupp}(A \Rightarrow B)$$

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

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	$x \notin A \vee x \in B$	$x \in A \vee x \notin B$

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

	$A \Rightarrow B$
minimum support (minsupp)	$\lvert A \cap B \rvert$
maximum opposition (maxopp)	$\lvert coA \cup coB \rvert$
minimum opposition (minopp)	$\lvert A \cap coB \rvert$
maximum support (maxsupp)	$\lvert coA \cup B \rvert$

Table 2: Overview of measures

$$\text{minopp}(A \Rightarrow B) \leq \text{maxopp}(A \Rightarrow B)$$

Since

$$\text{maxopp}(A \Rightarrow B) = |X| - \text{minsupp}(A \Rightarrow B)$$

$$\text{minopp}(A \Rightarrow B) = |X| - \text{maxsupp}(A \Rightarrow B)$$

in reality we are only dealing with two independent measures. We can for instance choose to work with minsupp and maxsupp. The measure minsupp corresponds to the symmetrical support (supp) that is traditionally studied, while maxsupp is a non-symmetrical measure taking into account all examples that do not contradict $A \Rightarrow B$. Another way of expressing the two opposition measures in terms of the support measures is

$$\text{minopp}(A \Rightarrow coB) = \text{minsupp}(A \Rightarrow B)$$

$$\text{maxopp}(A \Rightarrow coB) = \text{maxsupp}(A \Rightarrow B)$$

Confidence. If the support of an association rule $A \Rightarrow B$ exceeds a user-specified threshold, its confidence is investigated. This is usually defined as

$$\text{conf}(A \Rightarrow B) = \frac{\text{supp}(A \Rightarrow B)}{\text{supp}(A \Rightarrow X)} = \frac{\text{supp}(A \Rightarrow B)}{\text{supp}(A \Rightarrow B) + \text{supp}(A \Rightarrow co\ B)}$$

or [5]

$$\text{conf}_{\text{n}}(A \Rightarrow B) = \frac{\text{supp}(A \Rightarrow B)}{\text{supp}(A \Rightarrow co\ B)}$$

The latter can be written in terms of the newly defined measures of support and opposition:

$$\text{conf}_{\text{n}}(A \Rightarrow B) = \frac{\text{minsupp}(A \Rightarrow B)}{\text{minopp}(A \Rightarrow B)}$$

Hence the confidence of a rule is the number of positive examples of the rule divided by the number of negative examples. Using the measures of maximum support and opposition as well, we introduce two new measures of confidence, namely pessimistic and optimistic confidence:

$$\mathrm{conf_p}(A \Rightarrow B) = \frac{\mathrm{minsupp}(A \Rightarrow B)}{\mathrm{maxopp}(A \Rightarrow B)} \qquad \mathrm{conf_o}(A \Rightarrow B) = \frac{\mathrm{maxsupp}(A \Rightarrow B)}{\mathrm{minopp}(A \Rightarrow B)}$$

When determining the pessimistic confidence of a rule *bread* $\Rightarrow$ *butter* we have the following assumption in mind: if those people who did not buy bread, *would* have bought bread, they would not have bought butter as well. For the optimistic confidence measure on the other hand we assume that if those people who did not buy bread, *would* have bought bread, they would have bought butter as well. Naturally: $\mathrm{conf_p}(A \Rightarrow B) \leq \mathrm{conf_n}(A \Rightarrow B) \leq \mathrm{conf_o}(A \Rightarrow B)$. Since all these measures of confidence are defined in terms of measures of opposition and support, and furthermore the opposition measures can be described in terms of the support measures, in the remainder we will focus on minsupp and maxsupp.

3 FUZZY ASSOCIATION RULES

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 an ordinal scale. One way of dealing with a quantitative attribute is to replace it by a few other attributes that partition the range of the original one, such as *low*, *medium* and *high*. Now one can consider these new attributes as binary ones, which reduces the problem to the mining procedure described above (the generated rules are now called quantitative association rules [8]). It is however far more intuitively justifiable to allow attribute values from the interval $[0, 1]$ (instead of just $\{0, 1\}$) indicating the degree to which the record has the attribute. In this way attributes are no longer binary but fuzzy. The corresponding mining process yields fuzzy (quantitative) association rules (see e.g. [3], [6]).

To obtain such rules the measures discussed above have to be generalized in a suitable way. The power of a fuzzy set A in a finite universe X was introduced as a generalization of the classical concept of cardinality of a crisp set [4]. It is defined as

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

Fuzzy set theoretical counterparts of complement, intersection, and union are usually defined by means of a negator, a t-norm, and a t-conorm. Recall that an increasing, associative and commutative $[0, 1]^2 - [0, 1]$ 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 $[0, 1] - [0, 1]$-mapping satisfying $\mathcal{N}(0) = 1$ and $\mathcal{N}(1) = 0$. For A and B fuzzy sets in X:

$$\begin{aligned} co_{\mathcal{N}} A(x) &= \mathcal{N}(A(x)) \\ A \cap_{\mathcal{T}} B(x) &= \mathcal{T}(A(x), B(x)) \\ A \cup_{\mathcal{S}} B(x) &= \mathcal{S}(A(x), B(x)) \end{aligned}$$

Replacing the set theoretical operations in Table 2 by their fuzzy set theoretical counterparts, we obtain

$$\mathrm{minsupp}(A \Rightarrow B) = \sum_{x \in X} (A \cap_{\mathcal{T}} B)(x) \tag{1}$$

t − norm	t − conorm
$\mathcal{T}_{\mathrm{M}}(x,y) = \min(x,y)$	$\mathcal{S}_{\mathrm{M}}(x,y) = \max(x,y)$
$\mathcal{T}_{\mathrm{P}}(x,y) = xy$	$\mathcal{S}_{\mathrm{P}}(x,y) = x + y - xy$
$\mathcal{T}_{\mathrm{W}}(x,y) = \max(x+y-1,0)$	$\mathcal{S}_{\mathrm{W}}(x,y) = \min(x+y,1)$

Table 3: Well-known t-norms and t-conorms

S − implicator	Residual implicator
$\mathcal{I}_{\mathcal{S}_{\mathrm{M}}}(x,y) = \max(1-x,y)$	$\mathcal{I}_{\mathcal{T}_{\mathrm{M}}}(x,y) = \begin{cases} 1, & \text{if } x \le y \\ y, & \text{otherwise} \end{cases}$
$\mathcal{I}_{\mathcal{S}_{\mathrm{P}}}(x,y) = 1 - x + xy$	$\mathcal{I}_{\mathcal{T}_{\mathrm{P}}}(x,y) = \begin{cases} 1, & \text{if } x \le y \\ \frac{y}{x}, & \text{otherwise} \end{cases}$
$\mathcal{I}_{\mathcal{S}_{\mathrm{W}}}(x,y) = \min(1-x+y,1)$	$\mathcal{I}_{\mathcal{T}_{\mathrm{W}}}(x,y) = \min(1-x+y,1)$

Table 4: Well-known implicators

and

$$\text{maxsupp}(A \Rightarrow B) = \sum_{x \in X} (co_{\mathcal{N}} A \cup_{\mathcal{S}} B)(x) \qquad (2)$$

Formula (1) corresponds to the measure of support that is commonly used for mining fuzzy association rules. Formula (2) seems to be the most intriguing one from the semantical point of view. In the crisp case

$$\begin{aligned} x \in coA \cup B \quad &\Leftrightarrow \quad x \in coA \vee x \in B \\ &\Leftrightarrow \quad \neg(x \in A) \vee x \in B \\ &\Leftrightarrow \quad x \in A \rightarrow x \in B \end{aligned}$$

revealing that the logical connective behind the maximum support is an implication. The fuzzy logical counterpart of implication is the concept of implicator. An implicator $\mathcal{I}$ is a $[0,1]^2 - [0,1]$ 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 $\mathcal{I}(0,0) = 0$. The implicator at hand in Formula (2) 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 $[0,1]$. Another well-known kind of implicator is the residual implicator $\mathcal{I}_{\mathcal{T}}$ induced by a t-norm $\mathcal{T}$ in the following way

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

for all x and y in $[0,1]$. The question arises whether we can substitute the S-implicator in Formula (2) by a residual implicator. Tables 3 and 4 recall some well-known t-norms and t-conorms, as well as the implicators induced by them and the standard negator $\mathcal{N}_s(x) = 1-x$ for all x in $[0,1]$ (which is omitted in the notation). Table 5 shows the different contributions of several transactions x to maxsupp($A \Rightarrow B$) for all of these implicators. 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 the largest for

$A(x)$	$B(x)$	$\mathcal{I}_{\mathcal{S}_\text{M}}$	$\mathcal{I}_{\mathcal{S}_\text{P}}$	$\mathcal{I}_{\mathcal{S}_\text{W}}$	$\mathcal{I}_{\mathcal{T}_\text{M}}$	$\mathcal{I}_{\mathcal{T}_\text{P}}$	$\mathcal{I}_{\mathcal{T}_\text{W}}$
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

Table 5: Comparison of the contribution of some transactions

$\mathcal{I}_{\mathcal{T}_\text{M}}$ which completely ignores $A(x)$, and the smallest for $\mathcal{I}_{\mathcal{T}_\text{W}}$ as the implicators induced by $\mathcal{S}_\text{W}$ and $\mathcal{T}_\text{W}$ 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 indeed tend to identify x as a non-negative example, while residual implicators overlook it. 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 λ such that $\mathcal{T}(A(x), \lambda) \leq B(x)$. If $A(x) \leq B(x)$ then λ 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 λ to keep $\mathcal{T}(A(x), \lambda)$ from surpassing $B(x)$. Therefore λ tends to be low, hence x is not identified as a non-negative example.

Finally we mention that Hüllermeier [5], [6] suggested the following implication-based measure of support for a fuzzy association rule $A \Rightarrow B$:

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

It is motivated by the fact 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 (1) defined by means of $\mathcal{T}_\text{M}$). This is considered to be low since x *"does hardly violate (and hence supports) the rule"* [6]. We fully agree on the first claim (x is a non-negative example to a high degree) but not on the second (being a non-negative example does not imply being a positive example). Although the introduction of fuzzy logical implicators in the measures used for mining fuzzy association rules in itself is very meaningful, the author in [6] does not respect the fundamental difference between positive and non-negative examples, which lies in 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

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

However if $\mathcal{I}$ is the residual implicator induced by a continuous t-norm $\mathcal{T}$ then $\text{supp}_2(A \Rightarrow B) = \min(A(x), B(x))$ (see e.g. [7]) as is also noted in [6]. Hence in this case the new measure of support introduced in [5] reduces to the commonly used one. Still the author prefers these residual implicators over S-implicators, which seems to be in conflict with our findings described above. However his arguments for doing so basically come down to the fact that S-implicators detect non-negative examples overlooked by residual implicators, namely those that are not relevant to the rule. In [5], [6] this is considered to be an unwanted side effect because the author 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.

4 CONCLUSION

We refined the theory of association rules by exploiting the distinction between positive and negative examples to introduce additional quality measures that may be used in the assessment of such rules. Next, this simple but effective idea is generalized to the setting of fuzzy association rules, using appropriate classes of fuzzy connectives. In the process, it is revealed that S-implicators adhere closer to the intended semantics of positive versus negative examples than residual implicators. The obvious next step will now be to re-think algorithms for identifying (fuzzy) association rules on the basis of the newly obtained information.

References

[1] R. Agrawal, T. Imielinski, A. Swami (1993) "Mining Association Rules between Sets of Items in Large Databases", in Proceedings of ACM SIGMOD Conference on Management of Data, 207–216.

[2] K. T. Atanassov (1999) Intuitionistic Fuzzy Sets, Physica-Verlag, Heidelberg, New York.

[3] G. Chen, Q. Wei, E. E. Kerre (2000) "Fuzzy Data Mining: Discovery of Fuzzy Generalized Association Rules", in Recent Research Issues on Management of Fuzziness in Databases (G. Bordogna, G. Pasi eds.), Springer Physica-Verlag, 2000, 45–66.

[4] A. De Luca, S. Termini (1972) "A Definition of Non-probabilistic Entropy in the Setting of Fuzzy Sets", Inform. Control, 20:301–312.

[5] E. Hüllermeier (2001) "Fuzzy Association Rules: Semantic Issues and Quality Measures", in Computational Intelligence: Theory and Applications; Proceedings of the 7th Fuzzy Days Dortmund, Ed. B. Reusch; Lecture Notes in Computer Science Vol. 2206, Springer, 380–391.

[6] E. Hüllermeier (2002) "Mining Implication-Based Fuzzy Association Rules in Databases", in Proceedings of IPMU2002 (9th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems) Volume I, 101–108.

[7] V. Novák, I. Perfilieva, J. Močkoř (1999), Mathematical Principles of Fuzzy Logic, Kluwer Academic Publishers.

[8] R. Srikant, R. Agrawal (1996) "Mining Quantitative Association Rules in Large Relational Tables", in Proceedings of ACM SIGMOD Conference on Management of Data, v.25 n.2, 1–12.