

# Implication-Based Fuzzy Association Rules

Eyke Hüllermeier

Statistics and Decision Theory  
University of Paderborn, Germany  
eyke@upb.de

**Abstract.** Fuzzy association rules provide a data mining tool which is especially interesting from a knowledge-representational point of view since fuzzy attribute values allow for expressing rules in terms of natural language. In this paper, we show that fuzzy associations can be interpreted in different ways and that the interpretation has a strong influence on their assessment and, hence, on the process of rule mining. We motivate the use of multiple-valued implication operators in order to model fuzzy association rules and propose quality measures suitable for this type of rule. Moreover, we introduce a semantic model of fuzzy association rules which suggests to consider them as a convex combination of simple association rules. This model provides a sound theoretical basis and gives an explicit meaning to fuzzy associations. Particularly, the aforementioned quality measures can be justified within this framework.

## 1 Introduction

Association rules, syntactically written  $A \Rightarrow B$ , provide a means for representing dependencies between attributes in databases. Typically,  $A$  and  $B$  denote sets of binary attributes, also called features or items. The intended meaning of a (binary) rule  $A \Rightarrow B$  is that a transaction (a data record stored in the database) that contains the set of items  $A$  is likely to contain the items  $B$  as well. Several efficient algorithms for mining association rules in large databases have been devised [1, 19, 21]. Typically, such algorithms perform by generating a set of candidate rules from selected itemsets which are then filtered according to several quality criteria.

Generally, a database does not only contain binary attributes but also attributes with values ranging on (completely) ordered scales, e.g. cardinal or ordinal attributes. This has motivated a corresponding generalization of (binary) association rules. Typically, a *quantitative association rule* specifies attribute values by means of intervals, as e.g. in the simple rule “Employees at the age of 30 to 40 have incomes between \$50,000 and \$70,000.”

This paper investigates *fuzzy association rules*, which are basically obtained by replacing intervals in quantitative rules by fuzzy sets (intervals). The use of fuzzy sets in connection with association rules – as with data mining in general [20] – has recently been motivated by several authors (e.g. [2, 3, 5–8, 13, 15, 23, 25]). Among other aspects, fuzzy sets avoid an arbitrary determination of crisp

boundaries for intervals. Furthermore, fuzzy associations are very interesting from a knowledge representational point of view: The very idea of fuzzy sets is to act as an interface between a numeric scale and a symbolic scale which is usually composed of linguistic terms. Thus, the rules discovered in a database might be presented in a linguistic and hence comprehensible and user-friendly way. Example: “Middle-aged employees dispose of considerable incomes.”

Even though fuzzy association rules have already been considered by some authors, the investigation of their semantics in the context of data mining has not received much attention as yet. This is somewhat surprising since a clear semantics is a necessary prerequisite, not only for the interpretation, but also for the rating and, hence, for the mining of fuzzy association rules.

The semantics of fuzzy associations and their assessment by means of adequate quality measures constitute the main topics of the paper. By way of background, Section 2 reviews the aforementioned types of association rules. In Section 3, we discuss quality measures for fuzzy associations. In this connection, it is shown that a generalization of (quantitative) association rules can proceed from different perspectives, which in turn suggest different types of measures. We especially motivate the use of multiple-valued implication operators in order to model fuzzy association rules. In Section 4, we introduce a semantic model of fuzzy associations which considers them as convex combinations of simple association rules. This model clarifies the meaning and provides a sound theoretical basis of fuzzy association rules. Particularly, the aforementioned quality measures can be justified within this framework.

## 2 Association Rules

### 2.1 Binary Association Rules

Consider an association rule of the form  $A \Rightarrow B$ , where  $A$  and  $B$  denote subsets of an underlying set  $\mathcal{A}$  of *items* (which can be considered as binary attributes). As already said above, the intended meaning of  $A \Rightarrow B$  is that a *transaction*  $T \subset \mathcal{A}$  which contains the items in  $A$  is likely to contain the items in  $B$  as well.

In order to find “interesting” associations in a database  $D$ , a potential rule  $A \Rightarrow B$  is generally rated according to several criteria, none of which should fall below a certain (user-defined) threshold. In common use are the following measures ( $D_X \doteq \{T \in D \mid X \subset T\}$  denotes the transactions in the database  $D$  which contain the items  $X \subset \mathcal{A}$ , and  $|D_X|$  is its cardinality):

- A measure of *support* defines the absolute number or the proportion of transactions in  $D$  containing  $A \cup B$ :

$$\text{supp}(A \Rightarrow B) \doteq |D_{A \cup B}| \quad \text{or} \quad \text{supp}(A \Rightarrow B) \doteq \frac{|D_{A \cup B}|}{|D|}. \quad (1)$$

- The *confidence* is the proportion of correct applications of the rule:

$$\text{conf}(A \Rightarrow B) \doteq \frac{|D_{A \cup B}|}{|D_A|}. \quad (2)$$

- A rule  $A \Rightarrow B$  should be interesting in the sense that it provides new information. That is, the occurrence of  $A$  should indeed have a positive influence on the occurrence of  $B$ . A common measure of the *interest* of a rule is

$$\text{int}(A \Rightarrow B) \doteq \frac{|D_{A \cup B}|}{|D_A|} - \frac{|D_B|}{|D|}. \quad (3)$$

This measure can be seen as an estimation of  $\Pr(B | A) - \Pr(B)$ , that is the increase in probability of  $B$  caused by the occurrence of  $A$ .

## 2.2 Quantitative Association Rules

In the above setting, a transaction  $T$  can be seen as a sequence  $(x_1, \dots, x_m)$  of values of binary variables  $X_i$  with domain  $\mathfrak{D}_{X_i} = \{0, 1\}$ , where  $x_i = T[X_i] = 1$  if the  $i$ th item  $X_i$  is contained in  $T$  and  $x_i = 0$  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  $\mathfrak{D}_X, \mathfrak{D}_Y \subset \mathfrak{R}$ . A quantitative association rule involving the variables  $X$  and  $Y$  is then of the following form:

$$A \Rightarrow B : \text{ If } X \in A = [x_1, x_2] \text{ then } Y \in B = [y_1, y_2], \quad (4)$$

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 vectors 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  $(x, y) = (T[X], T[Y])$ , i.e. as a projection of the original database.

Note that the quality measures from Section 2.1 are applicable in the quantitative case as well:<sup>1</sup>

$$\text{supp}(A \Rightarrow B) = |\{(x, y) \in D \mid x \in A \wedge y \in B\}|, \quad (5)$$

$$\text{conf}(A \Rightarrow B) = \frac{|\{(x, y) \in D \mid x \in A \wedge y \in B\}|}{|\{(x, y) \in D \mid x \in A\}|}. \quad (6)$$

In fact, each interval  $[x_1, x_2]$  does again define a binary attribute  $X_{x_1, x_2} = \mathbb{I}_{[x_1, x_2]}$ . Thus, not only the rating but also the mining of quantitative rules can be reduced to the mining of binary association rules, by simply transforming the numerical data into binary data [18, 22]. Still, finding a useful transformation (binarization) of the data is a non-trivial problem by itself which affects both, the efficiency of subsequently applied mining algorithms and the potential quality of discovered rules. Apart from data transformation methods, clustering techniques can be applied which create intervals and rules at the same time [16, 24].

<sup>1</sup> Subsequently we focus on *support* and *confidence* measures. The results can be transferred to other measures such as *interest* in a straightforward way.

### 2.3 Fuzzy Association Rules

Replacing the sets (intervals)  $A$  and  $B$  in (4) by fuzzy sets (intervals) leads to fuzzy (quantitative) association rules. Thus, a fuzzy association rule is understood as a rule of the form  $A \Rightarrow B$ , 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. We shall use the same notation for ordinary sets and fuzzy sets. Moreover, we shall not distinguish between a fuzzy set and its membership function, that is,  $A(x)$  denotes the degree of membership of the element  $x$  in the fuzzy set  $A$ . Note that an ordinary set  $A$  can be considered as a “degenerate” fuzzy set with membership degrees  $A(x) = \mathbb{I}_A(x) \in \{0, 1\}$ .

## 3 Quality Measures for Fuzzy Association Rules

The standard approach to generalizing the quality measures for fuzzy association rules is to replace set-theoretic by fuzzy set-theoretic operations. The Cartesian product  $A \times B$  of two fuzzy sets  $A$  and  $B$  is usually defined by the membership function  $(x, y) \mapsto \min\{A(x), B(y)\}$ . Moreover, the cardinality of a finite fuzzy set is simply the sum of its membership degrees [17]. Thus, (5) and (6) can be generalized as follows:

$$\text{supp}(A \Rightarrow B) \doteq \sum_{(x,y) \in D} \min\{A(x), B(y)\}, \quad (7)$$

$$\text{conf}(A \Rightarrow B) \doteq \frac{\sum_{(x,y) \in D} \min\{A(x), B(y)\}}{\sum_{(x,y) \in D} A(x)}. \quad (8)$$

Note that the support of  $A \Rightarrow B$  corresponds to the sum of the *individual supports*, provided by tuples  $(x, y) \in D$ :<sup>2</sup>

$$\text{supp}_{[x,y]}(A \Rightarrow B) = \min\{A(x), B(y)\}. \quad (9)$$

According to (9),  $(x, y)$  supports  $A \Rightarrow B$  if both,  $x \in A$  and  $y \in B$ .

### 3.1 Support

The fact that the antecedent  $A$  and the consequent  $B$  play symmetrical roles in (9) might appear strange. Indeed, a more logic-oriented approach to modeling a fuzzy rule “If  $X$  is  $A$  then  $Y$  is  $B$ ” would use a generalized implication operator  $\rightsquigarrow$ , i.e. a mapping  $[0, 1] \times [0, 1] \rightarrow [0, 1]$  which generalizes the classical material implication (particularly,  $\rightsquigarrow$  is non-increasing in the first and non-decreasing in the second argument). Thus, individual support can be defined as

$$\text{supp}_{[x,y]}(A \Rightarrow B) \doteq A(x) \rightsquigarrow B(y) \quad (10)$$

<sup>2</sup> See [12] for an alternative approach where the frequency of a fuzzy item is measured by a *fuzzy cardinality*, i.e. by a fuzzy (rather than by a crisp) number.

and, hence, the overall (now *asymmetric*) support as

$$\text{supp}(A \Rightarrow B) \doteq \sum_{(x,y) \in D} A(x) \rightsquigarrow B(y). \tag{11}$$

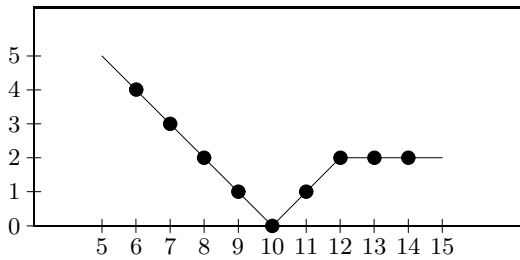
In order to realize the difference between (9) and (10), consider a simple rule “ $A \Rightarrow B$ : If  $X$  is approximately 10 then  $Y$  is almost 0” defined by two fuzzy subsets of the non-negative integers,

$$A : x \mapsto \begin{cases} 1 - \frac{|10-x|}{5} & \text{if } 6 \leq x \leq 14 \\ 0 & \text{otherwise} \end{cases}, \quad B : y \mapsto \begin{cases} 1 - \frac{y}{5} & \text{if } 0 \leq y \leq 4 \\ 0 & \text{otherwise} \end{cases}.$$

To which degree does the tuple  $(x, y) = (8, 3)$  support the above rule? According to (9), the support is  $2/5$ , namely the minimum of the membership of 8 in  $A$  and the membership of 3 in  $B$ . According to (10) with  $\rightsquigarrow$  the Goguen implication

$$\alpha \rightsquigarrow \beta \doteq \begin{cases} 1 & \text{if } \alpha = 0 \\ \min\{1, \beta/\alpha\} & \text{if } \alpha > 0 \end{cases}$$

the individual support is larger, namely  $2/3$ . In fact,  $(x, y) = (8, 3)$  does hardly violate (and hence supports) the rule in the sense of (10): It is true that  $y = 3$  does not fully satisfy the conclusion part of the rule; however, since  $x = 8$  does not fully meet the condition part either, it is actually not expected to do so.



**Fig. 1.** A simple functional relation between two variables that can be described by means of a gradual fuzzy rule.

As can be seen, the definition of adequate quality measures for fuzzy association rules depends strongly on the interpretation of the rule.<sup>3</sup> For further illustration, consider the nine observations shown in Figure 1.<sup>4</sup> In the sense of (10), each of these observations does fully support the rule “If  $X$  is approximately 10 then  $Y$  is almost 0.” In fact, this rule is actually interpreted as “The closer  $X$

<sup>3</sup> Particularly, the strategy of first using (7) and (8) to find interesting fuzzy associations and then interpreting these rules as *implications* appears questionable [4].

<sup>4</sup> Needless to say, this is a somewhat artificial example not at all typical of data mining.

is to 10, the closer  $Y$  is to 0” or, more precisely, “The closer  $X$  is to 10, the more it is guaranteed that  $Y$  is close to 0.” Therefore, (11) yields  $\text{supp}(A \Rightarrow B) = 9$ . As opposed to this, the overall support is only 5 according to (7), since the individual support  $\min\{A(x), B(y)\}$  which comes from a point  $(x, y)$  is bounded by  $B(y)$ , the closeness of  $y$  to 0. For instance, the support through  $(6, 4)$  is only  $1/5$  rather than 1. Note that, in the sense of (11), the support of the rule  $A \Rightarrow B$  is larger than the support of  $B \Rightarrow A$ ! This reflects the fact that the closeness of  $x$  to 10 is less guaranteed by the closeness of  $y$  to 0 than vice versa. For instance,  $y = 2$  is “rather” close to 0, whereas  $x = 14$  is only “more or less” close to 10, hence  $\text{supp}_{[2,14]}(B \Rightarrow A) = 1/3 < 1 = \text{supp}_{[14,2]}(A \Rightarrow B)$ .

The above example shows that the implication-based approach should be preferred whenever an association rule is thought of as expressing a gradual “the more ... the more ...”-relation between variables and all the more if this relation is not symmetric. For instance, the rule “Young people have low income” might actually be understood as “The younger a person, the lower the income,” and this rule might well be distinguished from its inversion “The lower the income, the younger the person.”

Concerning the adequacy of (11) as a support measure for association rules, two points deserve mentioning. Firstly, the concept of *support* in the context of association rules is actually intended as *non-trivial support*. Yet, the truth degree of a (generalized) implication  $\alpha \rightsquigarrow \beta$  is 1 whenever  $\alpha = 0$ : From a logical point of view, a false premise entails any conclusion. That is, the rule  $A \Rightarrow B$  would also be supported by those points  $(x, y)$  for which  $x \notin A$ . In order to avoid this effect, (10) can be modified as follows:

$$\text{supp}_{[x,y]}(A \Rightarrow B) = \begin{cases} A(x) \rightsquigarrow B(y) & \text{if } A(x) > 0 \\ 0 & \text{if } A(x) = 0 \end{cases} \quad (12)$$

According to (12), a point  $(x, y)$  supports a rule if both, it satisfies the rule (from a logical point of view) and it is non-trivial. Here, non-triviality is considered as a binary concept. However, it can also be quantified as a gradual property, namely as the degree to which  $x$  is in  $A$ . Combining satisfaction and non-triviality by means of a generalized logical conjunction  $\top$ , a so-called t-norm, then yields  $\text{supp}_{[x,y]}(A \Rightarrow B) = \top(A(x), A(x) \rightsquigarrow B(y))$ . For example, by using the product operator as a special t-norm we obtain

$$\text{supp}_{[x,y]}(A \Rightarrow B) = A(x) \cdot (A(x) \rightsquigarrow B(y)). \quad (13)$$

Note that (9), (12), and (13) are identical in the case where  $A$  and  $B$  are intervals, that is where  $A(x), B(y) \in \{0, 1\}$ .

The second point concerns the choice of the implication operator  $\rightsquigarrow$ . In fact, different types of implication operators exist which support different interpretations of a fuzzy rule [11]. The gradual “the more ... the more...”-interpretation discussed above is supported by so-called R(esiduated)-implications. An implication  $\rightsquigarrow$  of this type can be derived from a t-norm  $\top$  by residuation (hence the name):

$$\alpha \rightsquigarrow \beta \doteq \sup\{\gamma \in [0, 1] \mid \top(\alpha, \gamma) \leq \beta\}.$$

A second important class is given by so-called S(trong)-implications, which are defined by  $\alpha \rightsquigarrow \beta \doteq n(\alpha) \oplus \beta$ , where  $\oplus$  is a t-conorm (a generalized disjunction) and  $n$  a strong (hence the name) negation. For example, taking  $n(\cdot) = 1 - (\cdot)$  and  $\oplus = \max$ , one obtains the Kleene-Dienes implication  $\alpha \rightsquigarrow \beta = \max\{1 - \alpha, \beta\}$ .

S-implications support a different type of fuzzy rule, often referred to as *certainty rules*. Basically, they attach a level of uncertainty to the conclusion part of the rule, in correspondence with the truth degree of the condition part. These rules, however, appear less reasonable in the context of association rules. This can be exemplified by the Kleene-Dienes implication. For this operator, the truth degree is *lower-bounded* by  $1 - A(x)$ . Thus, (13) entails  $\text{supp}_{[x,y]}(A \Rightarrow B) \geq \min\{A(x), 1 - A(x)\}$  regardless of the value  $B(y)$ . For example, if  $A(x) = 1/2$ , then  $\text{supp}_{[x,y]}(A \Rightarrow B) = 1/2$ , no matter whether  $B(y)$  is  $1/2$ ,  $1/4$ , or even 0. This contrasts with R-implications, for which  $\beta = 0$  generally implies  $\alpha \rightsquigarrow \beta = 0$ .

### 3.2 Confidence

A measure of confidence of a fuzzy association rule can be derived from a corresponding measure of support. Indeed, in the non-fuzzy case, the confidence of  $A \Rightarrow B$  is nothing else than the support of  $A \Rightarrow B$  over the support of  $A$ , that is, the support of  $A \Rightarrow \mathfrak{D}_Y$ . Interestingly enough, however, the minimal confidence condition  $\text{conf}(A \Rightarrow B) \geq \Delta$  (where  $\Delta$  is a user-specified threshold) can be interpreted in different ways, which in turn suggest different generalizations.

According to the aforementioned interpretation which relates the support of  $A \Rightarrow B$  to the support of  $A$ , one obtains the generalized confidence measure

$$\text{conf}(A \Rightarrow B) \doteq \frac{\sum_{(x,y) \in D} \text{supp}_{[x,y]}(A \Rightarrow B)}{\sum_{(x,y) \in D} \text{supp}_{[x,y]}(A \Rightarrow \mathfrak{D}_Y)}. \tag{14}$$

Note that  $A(x) \rightsquigarrow \mathfrak{D}_Y(y) = A(x) \rightsquigarrow 1 = 1$  for all  $(x, y)$ . Thus, the denominator in (14) simplifies to  $\sum_{(x,y) \in D} \mathbb{I}_{(0,1]}(A(x))$  for (12) and  $\sum_{(x,y) \in D} A(x)$  for (13).

A second possibility is to relate the support of  $A \Rightarrow B$  to the support of  $A \Rightarrow \neg B$ . In this case, the minimal confidence condition means that the rule  $A \Rightarrow B$  should be supported much better than  $A \Rightarrow \neg B$ :

$$\text{conf}(A \Rightarrow B) = \frac{\sum_{(x,y) \in D} \text{supp}_{[x,y]}(A \Rightarrow B)}{\sum_{(x,y) \in D} \text{supp}_{[x,y]}(A \Rightarrow \neg B)}. \tag{15}$$

Note that  $A(x) = \text{supp}_{[x,y]}(A \Rightarrow \mathfrak{D}_Y) = \text{supp}_{[x,y]}(A \Rightarrow B) + \text{supp}_{[x,y]}(A \Rightarrow \neg B)$  for all  $(x, y)$  in the non-fuzzy case, which means that (14) and (15) are equivalent in the sense that one criterion can mimic the other one by adapting its threshold:

$$\frac{\text{supp}(A \Rightarrow B)}{\text{supp}(A \Rightarrow \mathfrak{D}_Y)} \geq \Delta \Leftrightarrow \frac{\text{supp}(A \Rightarrow B)}{\text{supp}(A \Rightarrow \neg B)} \geq \frac{\Delta}{1 - \Delta}.$$

## 4 Semantic Interpretation of Fuzzy Association Rules

In this section, we propose a semantic model of implication-based fuzzy rules which can directly be applied to fuzzy association rules. The idea is to represent a fuzzy rule as a collection of crisp (implication-based) rules. According to this model, a fuzzy association rule can be considered as a convex combination of non-fuzzy association rules. In this connection, we shall also justify the support measures (12) and (13).

### 4.1 Pure Gradual Rules

Consider two variables  $X$  and  $Y$  ranging on domains  $\mathfrak{D}_X$  and  $\mathfrak{D}_Y$ , respectively. Moreover, let  $A$  and  $B$  denote fuzzy subsets of  $\mathfrak{D}_X$  and  $\mathfrak{D}_Y$ . For the sake of simplicity, we assume the range of  $A$  and  $B$  to be a finite subset  $\mathcal{L} \subset [0, 1]$ . That is, membership degrees  $A(x)$  and  $B(y)$  are elements of  $\mathcal{L} = \{\lambda_1, \dots, \lambda_n\}$ , where  $0 = \lambda_1 < \lambda_2 < \dots < \lambda_n = 1$ .

A special type of fuzzy rule, called *pure gradual rule* [10], is obtained for the Rescher-Gaines implication

$$\alpha \rightarrow \beta \doteq \begin{cases} 1 & \text{if } \alpha \leq \beta \\ 0 & \text{if } \alpha > \beta \end{cases} \quad (16)$$

A pure gradual rule does actually induce a *crisp* relation of admissible tuples  $(x, y)$ . In fact, the fuzzy rule  $A \Rightarrow B$ , modeled by the implication (16), is equivalent to the following class of non-fuzzy constraints:

$$X \in A_\lambda \Rightarrow Y \in B_\lambda \quad (\lambda \in \mathcal{L}) \quad (17)$$

where  $A_\lambda = \{x \mid A(x) \geq \lambda\}$  is the  $\lambda$ -cut of  $A$ . Now, in some situations one might wish to modify the constraints (17), that is to weaken or to strengthen a conclusion  $Y \in B_\lambda$  drawn from the condition  $X \in A_\lambda$ . This leads to a collection

$$X \in A_\lambda \Rightarrow Y \in B_{m(\lambda)} \quad (\lambda \in \mathcal{L}) \quad (18)$$

of (non-fuzzy) constraints, where  $m$  is a mapping  $\mathcal{L} \rightarrow \mathcal{L}$ . These constraints can be written compactly in terms of membership functions as  $m(A(X)) \leq B(Y)$ , and correspond to the rule  $A \Rightarrow B$  modeled by the *modified* Rescher-Gaines implication  $\rightarrow_m$  with  $\alpha \rightarrow_m \beta = 1$  if  $m(\alpha) \leq \beta$  and 0 otherwise. Given two fuzzy sets  $A$  and  $B$ , we can thus associate a gradual rule  $A \rightarrow_m B$  (which is short for:  $A \Rightarrow B$  modeled by the implication  $\rightarrow_m$ ) with each function  $m : \mathcal{L} \rightarrow \mathcal{L}$ . Note that  $m$  should be non-decreasing: If the premise  $X \in A_\lambda$  entails the conclusion  $Y \in B_{m(\lambda)}$ , then a more restrictive premise  $X \in A_{\lambda'}$  ( $\lambda < \lambda'$ ) justifies this conclusion all the more, that is  $m(\lambda) \leq m(\lambda')$ . Thus, the scale  $\mathcal{L}$  gives rise to the following class of gradual rules:

$$\mathcal{G} = \mathcal{G}_{A,B} = \{A \rightarrow_m B \mid m : \mathcal{L} \rightarrow \mathcal{L} \text{ is non-decreasing}\}. \quad (19)$$



### 4.2 Other Implication-Based Rules

The tuples  $(x, y)$  that satisfy a pure gradual rule  $A \rightarrow_m B$  define an ordinary relation  $\pi_m \subset \mathfrak{D}_X \times \mathfrak{D}_Y$ :

$$\pi_m(x, y) = \begin{cases} 1 & \text{if } m(A(x)) \leq B(y) \\ 0 & \text{if } m(A(x)) > B(y) \end{cases} \quad (20)$$

More generally, an implication operator  $\rightsquigarrow$  induces a *fuzzy* relation  $\pi_{\rightsquigarrow}$ , where  $\pi_{\rightsquigarrow}(x, y) = A(x) \rightsquigarrow B(y)$  is the degree of admissibility of  $(x, y)$ . Subsequently, we assume a multiple-valued implication  $\rightsquigarrow$  to be non-increasing in the first and non-decreasing in the second argument, and to satisfy the identity property:  $\lambda \rightsquigarrow 1 = 1$  for all  $\lambda \in \mathcal{L}$ .

### 4.3 Randomized Gradual Rules

We are now going to establish a helpful relationship between an implication-based rule  $A \rightsquigarrow B$ , i.e. the rule  $A \Rightarrow B$  modeled by the implication operator  $\rightsquigarrow$ , and the class (19) of pure gradual rules  $A \rightarrow_m B$  associated with  $A$  and  $B$ .

**Definition 1 (randomized rule).** *A randomized rule associated with a conditional statement “If  $X$  is  $A$  then  $Y$  is  $B$ ” is a tuple  $(\mathcal{G}, p)$ , where  $\mathcal{G} = \mathcal{G}_{A,B}$  is the (finite) set of pure gradual rules (19) and  $p$  is a probability distribution on  $\mathcal{G}$ . Each rule  $A \rightarrow_m B$  is identified by the corresponding function  $m : \mathcal{L} \rightarrow \mathcal{L}$ . Moreover,  $p_m = p(A \rightarrow_m B)$  is interpreted as the probability (or, more generally, the weight) of the rule  $A \rightarrow_m B$ .*

Recall that each pure gradual rule  $A \rightarrow_m B$  induces an admissible set (20) of tuples  $(x, y)$ . Therefore, a randomized rule  $(\mathcal{G}, p)$  gives rise to a *random set* over  $\mathfrak{D}_X \times \mathfrak{D}_Y$  and, hence, induces the following fuzzy relation:

$$\pi_{(\mathcal{G},p)} = \sum_{m \in \mathcal{G}} p_m \cdot \pi_m. \quad (21)$$

Moreover, (21) is completely determined by the following implication operator associated with  $(\mathcal{G}, p)$ :

$$\lambda_i \stackrel{(\mathcal{G},p)}{\rightsquigarrow} \lambda_j = \sum_{m \in \mathcal{G} : m(\lambda_i) \leq \lambda_j} p_m. \quad (22)$$

Namely,  $\pi_{(\mathcal{G},p)}(x, y) = A(x) \stackrel{(\mathcal{G},p)}{\rightsquigarrow} B(y)$  for all  $(x, y) \in \mathfrak{D}_X \times \mathfrak{D}_Y$ .

Recall that a pure gradual rule corresponds to a collection of simple, non-fuzzy constraints and, hence, disposes of a very simple semantics. Since a random rule is a convex combination of pure gradual rules, it can also be interpreted in a very simple way. This lets the representation of a general implication-based fuzzy rule in terms of a random rule seem appealing. Concerning this representation, we have proved the following existence and uniqueness results [9]: For

each implication operator  $\rightsquigarrow$ , a probability  $p$  exists such that the rule  $A \rightsquigarrow B$  is equivalent to the randomized rule  $(\mathcal{G}, p)$  in the sense that  $\pi_{\rightsquigarrow} = \pi_{(\mathcal{G}, p)}$ . That is, the rule  $A \rightsquigarrow B$  and the randomized rule  $(\mathcal{G}, p)$  induce the same admissibility relation on  $\mathfrak{D}_X \times \mathfrak{D}_Y$ . Moreover, the probability  $p$  is guaranteed to be unique if the implication operator  $\rightsquigarrow$  does not have a certain (strict) monotonicity property.

**Theorem 1.** *For each fuzzy rule  $A \Rightarrow B$  formalized by means of an implication operator  $\rightsquigarrow$  an equivalent random rule  $(\mathcal{G}, p)$  exists. Moreover, the representation in terms of  $(\mathcal{G}, p)$  is unique if the condition  $\neg(\gamma_{kj} < \gamma_{lj} < \gamma_{il}) \vee \neg(\gamma_{kj} < \gamma_{kl} < \gamma_{il})$  holds for all  $1 \leq i < k \leq n$  and  $1 \leq j < l \leq l$ , where  $\gamma_{ij} = \lambda_i \rightsquigarrow \lambda_j$ .*

#### 4.4 Application to Association Rules

In Section 2, we have proposed to consider a fuzzy association  $A \Rightarrow B$  as an implication-based (gradual) fuzzy rule. Referring to the above interpretation, a rule  $A \Rightarrow B$  can hence be seen as a convex combination of simple or *pure gradual* association rules

$$A \rightarrow_m B \quad (m \in \mathcal{G}), \tag{23}$$

weighted by the probability degrees  $p_m$ . Each of these gradual association rules in turn corresponds to a collection

$$A_\lambda \Rightarrow B_{m(\lambda)} \quad (\lambda \in \mathcal{L}) \tag{24}$$

of ordinary association rules. In fact, if the level-cuts  $A_\lambda$  and  $B_{m(\lambda)}$  of the fuzzy sets  $A$  and  $B$  are intervals, which holds true for commonly used membership functions, then (24) reduces to a class of interval-based association rules.

The interpretation as a randomized rule assigns an association rule a concrete meaning and might hence be helpful in connection with the acquisition (mining) and interpretation of such rules. Apart from this, it provides a basis for justifying quality measures for fuzzy association rules. In fact, proceeding from the convex combination of rules (23) it is obvious to define the support of  $A \Rightarrow B$  as the convex combination of the supports of the rules  $A \rightarrow_m B$ :

$$\text{supp}(A \Rightarrow B) = \sum_{m \in \mathcal{G}} p_m \cdot \text{supp}(A \rightarrow_m B). \tag{25}$$

Thus, it remains to define the (non-trivial) support of a pure gradual association rule  $A \rightarrow_m B$ , that is of a collection of ordinary association rules (24). To which degree does a point  $(x, y)$  support this class of constraints? One possibility is to say that  $(x, y)$  supports  $A \rightarrow_m B$  if it satisfies *all* individual constraints, and, furthermore, at least one of these constraints is non-trivial:

$$\text{supp}_{[x,y]}(A \rightarrow_m B) = \begin{cases} 1 & \text{if } A(x) > 0 \wedge m(A(x)) \leq B(y) \\ 0 & \text{otherwise} \end{cases}. \tag{26}$$

A second possibility is to define  $\text{supp}_{[x,y]}(A \rightarrow_m B)$  as the sum of weights of those individual constraints which are indeed non-trivially satisfied:

$$\text{supp}_{[x,y]}(A \rightarrow_m B) = \begin{cases} A(x) & \text{if } m(A(x)) \leq B(y) \\ 0 & \text{otherwise} \end{cases}. \quad (27)$$

It is readily verified that (26), in conjunction with (25), yields the support measure (12), and that (27) in place of (26) implies (13). This result provides a sound basis for these measures of (individual) support and, hence, for further quality measures derived from them.

## 5 Concluding Remarks

We have proposed an implication-based approach to fuzzy association rules as well as a semantic model which suggests of consider such rules as a convex combination of simple, non-fuzzy association rules. Thus, a fuzzy association can be seen as a compact representation of a class of simple rules. This model clarifies the meaning and provides a sound basis of fuzzy association rules.

The paper has mainly focused on theoretical foundations of fuzzy association rules. An important aspect of ongoing research is the practical realization of the results, that is the development of rule mining procedures. Our current implementation (not presented here due to space limitations, see [14]) is an extension of the APRIORI algorithm [1] which is able to cope with fuzzy attribute values and asymmetric support measures. This algorithm takes advantage of the fact that the support (13) of  $A \Rightarrow B$  is *lower-bounded* by the support of the premise  $A$ :  $\text{supp}_{[x,y]}(A \Rightarrow B) = A(x) \cdot (A(x) \rightsquigarrow B(y)) \leq A(x)$ . Consequently, the premise  $A$  of a minimally supported rule  $A \Rightarrow B$  must be a frequent itemset or, put in a different way, the frequent itemsets (which can be found by APRIORI) constitute a superset of the condition parts of minimally supported association rules. Furthermore, the algorithm makes use of a monotonicity property for implications which is similar to the monotonicity property of frequent itemsets employed by APRIORI:  $\text{supp}(A \Rightarrow B) \leq \text{supp}(A \Rightarrow B')$  for all  $B' \subset B$ . Thus, if  $A \Rightarrow B$  satisfies the minimum support condition, the same condition holds for each rule  $A \Rightarrow B'$  with  $B' \subset B$ . This provides the basis for filtering candidate rules (obtained by combining frequent itemsets  $A$  with conclusions  $B$ ) in an efficient way.

## References

1. R. Agrawal and S. Srikant. Fast algorithms for mining association rules. In *Proceedings of the 20th Conference on VLDB*, Santiago, Chile, 1994.
2. 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.
3. Wai-Ho Au and K.C.C. Chan. FARM: A data mining system for discovering fuzzy association rules. In *Proceedings FUZZ-IEEE-99*, pages 1217–1222, 1999.

4. M. Bernadet. Basis of a fuzzy knowledge discovery system. In D.A. Zighed, J. Komorowski, and J. Zytkow, editors, *Principles of Data Mining and Knowledge Discovery*, pages 24–33. Springer-Verlag, 2000.
5. K.C.C. Chan and Wai-Ho Au. Mining fuzzy association rules. In *Proceedings CIKM-97, 6th Int. Conf. on Inform. and Knowl. Management*, pages 10–14, 1997.
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. G. Chen, G. Wets, and K. Vanhoof. Representation and discovery of fuzzy association rules. Technical Report 00/01, ITEO Research Papers, 2000.
8. 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.
9. D. Dubois, E. Hüllermeier, and H. Prade. Toward the representation of implicative fuzzy rules in terms of crisp rules. In *Proc. IFSA/NAFIPS-2001*, Vancouver, Canada, July 2001. To appear.
10. D. Dubois and H. Prade. Gradual inference rules in approximate reasoning. *Information Sciences*, 61(1,2):103–122, 1992.
11. D. Dubois and H. Prade. What are fuzzy rules and how to use them. *Fuzzy Sets and Systems*, 84:169–185, 1996.
12. 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.
13. A. Fu, M.H. Wong, S.C. Sze, W.C. Wong, W.L. Wong, and W.K. Yu. Finding fuzzy sets for the mining of fuzzy association rules for numerical attributes. In *IDEAL-98, 1st Int. Symp. on Intell. Data Engineering and Learning*, pages 263–268, 1998.
14. E. Hüllermeier. Mining implication-based fuzzy association rules. Submitted.
15. C. Man Kuok, A. Fu, and M. Hon Wong. Mining fuzzy association rules in databases. *SIGMOD Record*, 27:41–46, 1998.
16. B. Lent, A. Swami, and J. Widom. Clustering association rules. In *Proceedings ICDE-97*, Birmingham, UK, 1997.
17. A. De Luca and S. Termini. Entropy of L-fuzzy sets. *Information and Control*, 24:55–73, 1974.
18. R.J. Miller and Y. Yang. Association rules over interval data. In *Proc. ACM SIGMOD Int. Conf. on Management of Data*, pages 452–461, 1997.
19. J.S. Park, M.S. Chen, and P.S. Yu. An efficient hash-based algorithm for mining association rules. In *ACM SIGMOD Int. Conf. on Management of Data*, 1995.
20. W. Pedrycz. Data mining and fuzzy modeling. In *Proc. of the Biennial Conference of the NAFIPS*, pages 263–267, Berkeley, CA, 1996.
21. A. Savasere, E. Omiecinski, and S. Navathe. An efficient algorithm for mining association rules in large databases. In *Proceedings VLDB-95*, Zurich, 1995.
22. R. Srikant and R. Agrawal. Mining quantitative association rules in large relational tables. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pages 1–12, 1996.
23. S. Ben Yahia and A. Jaoua. Mining linguistic summaries of databases using Lukasiewicz implication fuzzy functional dependency. In *Proceedings FUZZ-IEEE-99*, pages 1246–1250, 1999.
24. Y. Yang and M. Singhal. Fuzzy functional dependencies and fuzzy association rules. In *Data Warehousing and Knowledge Discovery, Proceedings DAWAK-99*, pages 229–240. Springer-Verlag, 1999.
25. W. Zhang. Mining fuzzy quantitative association rules. In *Proc. 11th IEEE Int. Conf. on Tools with Artificial Intelligence*, Chicago, Illinois, 1999.