A Systematic Approach to the Assessment of Fuzzy Association Rules*

Didier Dubois
IRIT-UPS, Toulouse, France

Eyke Hüllermeier
Faculty of Computer Science
University of Magdeburg, Germany

Henri Prade
IRIT-UPS, Toulouse, France

Abstract

In order to allow for the analysis of data sets including numerical attributes, several generalizations of association rule mining based on fuzzy sets have been proposed in the literature. While the formal specification of fuzzy associations is more or less straightforward, the assessment of such rules by means of appropriate quality measures is less obvious. Particularly, it assumes an understanding of the semantic meaning of a fuzzy rule. This aspect has been ignored by most existing proposals, which must therefore be considered as ad-hoc to some extent. In this paper, we develop a systematic approach to the assessment of fuzzy association rules. To this end, we proceed from the idea of partitioning the data stored in a database into examples of a given rule, counterexamples, and irrelevant data. Evaluation measures are

*This article is a revised and extended version of a paper presented at the 10th International Fuzzy Systems Association World Congress, Istambul, 2003 [25].

then derived from the cardinalities of the corresponding subsets. The problem of finding a proper partition has a rather obvious solution for standard association rules but becomes less trivial in the fuzzy case. Our results not only provide a sound justification for commonly used measures but also suggest a means for constructing meaningful alternatives.

Keywords: association rules, fuzzy sets, quality measures, fuzzy partition.

1 Introduction

Association rules provide a means for representing dependencies between attribute values of objects (data records) stored in a database. Typically, an association involves two sets of binary attributes (features), $\mathcal{A}$ and $\mathcal{B}$. Then, the intended meaning of a rule $\mathcal{A} \rightharpoonup \mathcal{B}$ is that an object having all the features in $\mathcal{A}$ is likely to have all the features in $\mathcal{B}$ as well.

Association rules of such type are often employed in the context of market-basket analysis, where an object is a purchase and features are associated with products or items. In this context, the association $\{\texttt{paper}, \texttt{envelopes}\} \rightharpoonup \{\texttt{stamps}\}$ suggests, for example, that a purchase containing paper and envelopes is likely to contain stamps as well.

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 numerical or ordered categorical attributes. In *quantitative association rules*, attribute values are specified by means of subsets, typically in the form of 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 [72] – has been motivated by numerous authors (see [15, 22] for recent overviews). By allowing for "soft" rather than crisp boundaries of intervals, fuzzy sets can avoid certain undesirable threshold effects [85].

Furthermore, fuzzy association rules are very appealing from a knowledge representational point of view: The very idea of fuzzy sets is to act as an interface between a numerical 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 receive considerable incomes."

As can be seen, moving from set-based (interval-based) to fuzzy associations is formally accomplished by replacing sets (intervals) by fuzzy sets (fuzzy intervals). While the formal specification of fuzzy associations is hence more or less straightforward, the *evaluation* of fuzzy associations through appropriate quality measures, notably the well-known support and confidence measures, is more intricate [32]. Especially, it assumes an understanding of the semantics of a fuzzy rule [50]. In this respect, many existing proposals can be considered ad-hoc to some extent.

In this paper, we suggest a formal framework for the systematic derivation of quality measures which is based on the classification of stored data into examples of a rule, counterexamples of that rule, and irrelevant cases. In the fuzzy case, this means defining a corresponding fuzzy partition of the set of examples into *positive*, *negative*, and *irrelevant* examples of a rule. Among other contributions, this framework sheds light on and provides a sound justification of several measures that have been proposed in literature.

The remainder of the paper is organized as follows: 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 and further elaborated in Section 5. The paper concludes with a summary and an outlook on future work in Section 6.

2 Association Rules

2.1 Binary Association Rules

Let $\mathcal{D} = \{x_1, x_2 \ldots x_n\}$ be a set of objects and $\mathcal{R} = \{A_1, A_2 \ldots A_m\}$ a set of features or properties. Each property can be considered as a logical predicate or, alternatively, a binary attribute with domain $\{0, 1\}$. Correspondingly, $A_\imath(x_\jmath)$ is true (false) or $A_\imath(x_\jmath) = 1$ $(= 0)$ if the $\imath$-th property applies (does not apply) to the object $x_\jmath$.

As mentioned above, the $A_\imath$ are often referred to as *items* in the context of association rule mining. Moreover, the set $\{A_\imath \,|\, A_\imath(x_\jmath)\} \subseteq \mathcal{R}$ of features that apply to an object $x_\jmath$ is often called a *transaction*. In other words, a transaction can either be considered as a subset of $\mathcal{R}$ or as an m-dimensional binary vector $(t_\jmath^1 \ldots t_\jmath^m)$, where $t_\jmath^\imath = 1$ if $A_\imath(x_\jmath)$ and $= 0$ otherwise. To illustrate, let an object again be a purchase, and let features correspond to different products. Then, $t_\jmath^\imath = 1$ means that the $\imath$-th product is contained in the purchase $x_\jmath$ and $t_\jmath^\imath = 0$ that it is not contained.

For a subset $\mathcal{A} \subseteq \mathcal{R}$ of features, let

$$\mathcal{A}(x_\jmath) \overset{\text{df}}{\Leftrightarrow} \forall A_\imath \in \mathcal{A} \,:\, A_\imath(x_\jmath), \tag{1}$$

i.e., $\mathcal{A}$ is a new predicate (binary attribute) that corresponds to the logical conjunction of individual properties. An association rule is a directed association $\mathcal{A} \rightharpoonup \mathcal{B}$ with $\mathcal{A}, \mathcal{B} \subseteq \mathcal{R}$ and $\mathcal{A} \cap \mathcal{B} = \emptyset$. The intended meaning of such a rule is that $\mathcal{A}(x)$ usually entails $\mathcal{B}(x)$.

In order to find "interesting" association rules in a database (binary relation), a potential rule $\mathcal{A} \rightharpoonup \mathcal{B}$ 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 objects that satisfy both $\mathcal{A}$ and $\mathcal{B}$, i.e.,

$$\mathsf{supp}(\mathcal{A} \rightharpoonup \mathcal{B}) \overset{\text{df}}{=} \big|\{x \in \mathcal{D} \,|\, \mathcal{A}(x) \wedge \mathcal{B}(x)\}\big|\,, \tag{2}$$

where $|\cdot|$ denotes cardinality. Support can also be defined by the proportion rather than the absolute number of objects, in which case (2) is divided by

$n = |\mathcal{D}|$. The *confidence* is the proportion of correct applications of the rule and can be thought of as (an estimation of) the conditional probability $\mathsf{Prob}(\mathcal{B} \,|\, \mathcal{A})$ that the consequent is true given that the antecedent holds:

$$\mathsf{conf}(\mathcal{A} \rightharpoonup \mathcal{B}) \stackrel{\mathrm{df}}{=} \frac{\big|\{x \in \mathcal{D} \,|\, \mathcal{A}(x) \wedge \mathcal{B}(x)\}\big|}{\big|\{x \in \mathcal{D} \,|\, \mathcal{A}(x)\}\big|}. \tag{3}$$

Further reasonable measures can be considered such as, e.g., the deviation (significance, lift)

$$\mathsf{sign}(\mathcal{A} \rightharpoonup \mathcal{B}) \stackrel{\mathrm{df}}{=} \mathsf{conf}(\mathcal{A} \rightharpoonup \mathcal{B}) - \frac{\big|\{x \in \mathcal{D} \,|\, \mathcal{B}(x)\}\big|}{|\mathcal{D}|},$$

suggesting that $\mathcal{A} \rightharpoonup \mathcal{B}$ is interesting only if the occurrence of $\mathcal{A}$ does indeed have a positive influence on the occurrence of $\mathcal{B}$ (i.e. $\mathsf{Prob}(\mathcal{B} \,|\, \mathcal{A}) > \mathsf{Prob}(\mathcal{B})$).

As can be seen, the support measure plays a central role. In fact, other measures can generally be derived from the support [86, 69]. For example, the confidence of an association $\mathcal{A} \rightharpoonup \mathcal{B}$ is the support of that association divided by the support of its antecedent, $\mathcal{A}$.

2.2 Quantitative Association Rules

In addition to binary attributes, a database (relation) usually contains numerical attributes, i.e., attributes $\mathcal{A}_\imath$ whose domain $\mathrm{dom}(A_\imath)$ is a subset of the real numbers rather than $\{0, 1\}$. One possibility to deal with a numerical attribute $A_\imath$ is to replace it by several binary attributes $A_\imath^1, A_\imath^2 \ldots A_\imath^k$ [67, 80]. Each $A_\imath^\jmath$ is identified with the property that $A_\imath(x)$ lies in a subset of $\mathrm{dom}(A_\imath)$, i.e., it is just the indicator function of that subset. For the sake of simplicity, we shall not distinguish between the binary attribute itself and its associated property (a subset of $\mathrm{dom}(A_\imath)$), i.e., we shall employ the same symbol for both of them.

The following property should of course be satisfied:

$$\mathrm{dom}(A_\imath) \subseteq \bigcup_{\jmath=1}^{k} A_\imath^\jmath,$$

i.e., the binary properties should cover the domain of the original attribute $A_\imath$. Note that the $A_\imath^\jmath$ are usually defined as intervals. In other words, quantitative association rules, understood as association rules involving numerical attributes, are usually interval-based rules of the form

$$\text{IF} \;\; A_\imath(x) \in A_\imath^\jmath \;\; \text{THEN} \;\; A_p(x) \in A_p^q, \tag{4}$$

which can be written equivalently as $A_\imath^\jmath(x) \rightharpoonup A_p^q(x)$. Note that (4) does make sense only if the attributes in the antecedent and consequent are different ($\imath \neq p$), i.e., the binary attributes $A_\imath^\jmath$ and A_p^q are related to different numerical attributes. Moreover, (4) can of course be generalized to the case where the antecedent and consequent part consist of several attributes.

By transforming numerical into binary attributes, not only the rating but also the mining of associations can be reduced to the standard setting. Still, finding a useful transformation (discretization) 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 standard discretization methods [24], clustering techniques can be applied which create intervals and rules at the same time [60, 94].[1]

2.3 Mining Algorithms for Association Rules

Apart from the formal problem specification, an important issue is the efficient mining of associations, that is algorithms for the extraction of all interesting rules from a given database. Even though the focus of this paper is on evaluation measures rather than mining algorithms, we briefly address this issue since the latter is clearly not independent of the former. A particularly relevant question, for example, is whether or not the "fuzzification" of a standard evaluation measure will preclude the use of an existing mining approach, i.e., whether or not that approach can still be used in order to extract those rules which are interesting according to that measure.

As databases are in general very large, mining algorithms should be scalable.

[1] Some authors reserve the term "quantitative association rule" for the case where partitions are not predefined.

The mining of association rules heavily exploits the structure of patterns which presents itself in the form of a generalization/specialization relation. Several efficient algorithms have been devised so far [2, 70, 76]. Typically, such algorithms perform by generating a set of candidate rules from selected itemsets which are then filtered according to several quality criteria. For example, the well-known APRIORI algorithm [2] generates rules from so-called *frequent itemsets*: One subset of the itemset becomes the premise of the rule and the complement becomes the conclusion. Due to definition (2), the support of any rule derived from a frequent itemset equals the support of the itemset itself. Thus, the problem of finding sufficiently supported rules reduces to the problem of finding frequent (= sufficiently supported) itemsets, which constitutes the main part of the APRIORI algorithm.

Alternative techniques have been developed to avoid the costly process of candidate generation and testing. The data mining method FP-growth (frequent pattern growth) is introduced in [47]. This method uses an extended prefix-tree (FP-tree) structure to store the database in a compressed form. FP-growth adopts a divide-and-conquer approach to decompose both the mining tasks and the databases. [18] describes how to compute *partial support counts* in one pass over the database and how to store them in an enumeration tree, a so-called P-tree. A related data structure, called T-tree (a compressed set enumeration tree), has been proposed in [19]. Several new algorithms combining different features (database format, the decomposition technique, and the search procedure) have been introduced in [97]. See [49, 38] for a comparison of different mining algorithms and [39] for a report on the performance of different frequent itemset mining implementations on selected real-world and artificial databases.

2.4 Further Topics in Association Rule Mining

A crucial problem in association rule mining concerns the often huge number of frequent itemsets and interesting rules that can be found in a dataset. In this connection, one idea is to focus on so-called *closed and maximal* frequent itemsets [98, 71] which are typically (but not necessarily[2]) by orders of magnitude fewer than all frequent itemsets (a frequent itemset is maximal if it is no proper subset of any other frequent itemset; an itemset is closed if no proper superset is contained in every transaction in which this set is contained). Still, it is guaranteed that all frequent itemsets can be generated from these itemsets. Hence, algorithms mining closed and maximal frequent itemsets are often more effective. Efficient algorithms for mining closed itemsets have recently been described in [64, 88, 96].

Constraint-based mining of association rules aims at reducing the number of rules presented to the user by incorporating constraints [73]. This way, uninteresting rules should be filtered out. Constraints can be realized, e.g., in the form of metarules [56, 37] or templates [81]. In this connection, it is also worth mentioning techniques for presenting the association rules found in a database in a more compact and intelligible way [84, 82].

Association rule mining has also been generalized in other directions. For example, alternative interest measures (besides support and confidence) have been proposed such as, e.g., *collective strength* [1] or *share frequency* [9]. So-called *multilevel association rules* involve items at different levels of abstraction [46]. For instance, the item *hardware* is an abstraction of the items *computer*, *monitor* and *printer*. An important topic in multilevel association rule mining concerns the avoidance of redundancy, that is redundant association rules [79]. The traditional model of association rule mining has also been adapted to handle *weighted* association rule mining, i.e., problems where each item is allowed to have a weight [13, 87]. Here, the goal is to steer the mining focus to interesting relationships involving items with significant weights.

[2] In fact, there are of course cases where the number of frequent itemsets will hardly be reduced. In sparse domains, for example, most frequent itemsets are already closed.

3 Fuzzy Association Rules

3.1 Background on Fuzzy Sets

A fuzzy subset F of a reference set U is identified by a so-called *membership function*, which is a generalization of the characteristic function of an ordinary subset [95]. For each element $u \in U$, this function specifies the degree of membership of u in the fuzzy set. Usually, membership degrees are taken from the unit interval $[0,1]$, i.e., a membership function is a mapping $U \rightarrow [0,1]$. 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, $F(u)$ denotes the degree of membership of the element u in the fuzzy set F. Note that an ordinary set F can be considered as a "degenerate" fuzzy set with membership degrees $F(u) = \mathbb{I}_F(u) \in \{0,1\}$.

Fuzzy sets formalize the idea of *graded membership*, i.e., the idea that an element belongs "more or less" to a set. A fuzzy set can have "non-sharp" boundaries. Consider the set of tall people as an example. Is it reasonable to say that (in a certain context) 180 cm is tall and 179 cm is not tall? In fact, any sharp boundary of the set of tall people will appear rather arbitrary. Modeling the concept "tall" as a fuzzy set F, it becomes possible to express, for example, that a height of 190 cm is completely in accordance with this concept ($F(190) = 1$), 175 cm is "more or less" tall ($F(175) = 0.5$, say), and 160 cm is clearly not tall ($F(160) = 0$).

As can be seen, fuzzy sets can provide a reasonable interpretation of linguistic expressions such as "tall people" or "high income" (in a given context). This way, they act as a smooth interface between a quantitative, numerical level and a qualitative level where knowledge is expressed in terms of natural language. In data mining, fuzzy sets thus allow for expressing patterns found at the quantitative level in terms of natural language.

Apart from that, the use of fuzzy partitions of the domains of quantitative attributes can avoid some undesirable threshold effects which are usually produced by crisp (non-fuzzy) partitions. Such effects are well-known, for instance, from *histograms* in statistics: A slight variation of the boundary points of the intervals can have a considerable effect on the histogram induced by a number of observations (it may even lead to qualitative changes, i.e., changes of the shape of the histogram) [83]. Likewise, the variation of a partition can strongly influence the evaluation of association rules [59].

To operate with fuzzy sets in a formal way, fuzzy set theory offers generalized set-theoretical resp. logical connectives and operators (as in the classical case, there is a close correspondence between set-theory and logic). In the following, we recall some basic operators that will be used in later parts of the paper.

A so-called t-norm $\otimes$ is a generalized logical *conjunction*, i.e., an operator $[0,1] \times [0,1] \rightarrow [0,1]$ which is associative, commutative, monotone increasing (in both places) and which satisfies the boundary conditions $\alpha \otimes 0 = 0$ and $\alpha \otimes 1 = \alpha$ for all $0 \leq \alpha \leq 1$ [55, 77]. Well-known examples of t-norms include the minimum $(\alpha, \beta) \mapsto \min(\alpha, \beta)$, the product $(\alpha, \beta) \mapsto \alpha\beta$, and the Lukasiewicz t-norm $(\alpha, \beta) \mapsto \max(\alpha + \beta - 1, 0)$.

A t-norm is used for defining the *intersection* of fuzzy sets $F, G : U \rightarrow [0,1]$ as follows: $(F \cap G)(u) \stackrel{\mathrm{df}}{=} F(u) \otimes G(u)$ for all $u \in U$. In a quite similar way, the *Cartesian product* of fuzzy sets $F : X \rightarrow [0,1]$ and $G : Y \rightarrow [0,1]$ is defined: $(F \times G)(x, y) \stackrel{\mathrm{df}}{=} F(x) \otimes G(y)$ for all $(x, y) \in X \times Y$.

The logical *disjunction* is generalized by a so-called t-conorm $\oplus$, an operator $[0,1] \times [0,1] \rightarrow [0,1]$ which is associative, commutative, monotone increasing (in both places) and such that $\alpha \otimes 1 = 1$ and $\alpha \otimes 0 = \alpha$ for all $0 \leq \alpha \leq 1$. Well-known examples of t-conorms include the maximum $(\alpha, \beta) \mapsto \max(\alpha, \beta)$, the algebraic sum $(\alpha, \beta) \mapsto \alpha + \beta - \alpha\beta$, and the Lukasiewicz t-conorm $(\alpha, \beta) \mapsto \min(\alpha + \beta, 1)$. A t-conorm can be used for defining the union of fuzzy sets: $(F \cup G)(u) \stackrel{\mathrm{df}}{=} F(u) \oplus G(u)$ for all u.

A generalized *implication* $\rightsquigarrow$ is an operator $[0,1] \times [0,1] \rightarrow [0,1]$ that is monotone decreasing in the first and monotone increasing in the second argument and that satisfies the boundary conditions $\alpha \rightsquigarrow 1 = 1$, $0 \rightsquigarrow \beta = 1$, $1 \rightsquigarrow \beta = \beta$. (Apart from that, additional properties are sometimes required.) Implication operators of that kind such as, e.g., the Lukasiewicz implication $(\alpha, \beta) \mapsto \min(1 - \alpha + \beta, 1)$, are especially important in connection with the modeling of fuzzy rules.

A repertoire of *negation* operators $n(\cdot)$ is also available, even though in practice one commonly employs the simple mapping $\alpha \mapsto 1 - \alpha$.

The *cardinality* of a fuzzy set F is usually defined in terms of the so-called σ-count, that is, by the sum of the values of its membership function [27]: $|F| \stackrel{\text{df}}{=} \sum_{u \in U} F(u)$.

3.2 Fuzzy Associations

In the context of association rule mining, the comments above motivate a "soft" partitioning of numerical attributes, that is, the partitioning of a numerical domain into fuzzy sets (fuzzy intervals) rather than ordinary sets (intervals). Thus, the domain $\mathrm{dom}(A_\jmath)$ of the numerical attribute is described by means of a set $F_\jmath^1 \ldots F_\jmath^n$ of fuzzy properties, and each of them can be viewed as a $[0,1]$-valued attribute $A_\jmath^\imath$ of objects such that $A_\jmath^\imath(x) = F_\jmath^\imath(A_\jmath(x))$. Such $[0,1]$-valued attributes can be called *fuzzy attributes* and are actually fuzzy subsets (events) of the set of objects $\mathcal{D}$.

A fuzzy association rule is then understood as a rule of the form $\mathcal{A} \rightharpoonup \mathcal{B}$, where $\mathcal{A}$ and $\mathcal{B}$ are, respectively, sets of fuzzy attributes $A_\imath$ and $B_\jmath$. To illustrate, consider a rule suggesting that experienced managers have high income:

$$\{\,\texttt{manager}, \texttt{experienced}\,\} \rightharpoonup \{\,\texttt{high\_income}\,\}$$

Here, the attributes `experienced` and `high_income` are reasonably modeled as fuzzy attributes.

As an aside, we note that one should be careful with this type of modeling via fuzzy attributes. In fact, it assumes that one disposes of a single scale (the unit interval) for all attributes. Membership functions defined on the domains of numerical attributes define a rescaling of these domains through the fuzzy properties. While this is reasonable for attributes with an underlying numerical domain $\mathrm{dom}(A_i)$ (such as, e.g., height and the related fuzzy property "tall" or income and the related property "high"), it is much less obvious for attributes with complex domains or attributes the underlying dimensions of which are ill-defined. For instance, it might be quite tricky to measure the property "experienced" on the unit interval if it does not only depend on the length of time a manager has worked in his job, but also on other dimensions such as, e.g., size of the company, number and type of projects, etc.

Note that several (one-dimensional) fuzzy attributes $A_\imath$ can be combined into one multi-dimensional attribute $\mathcal{A}$ by means of a t-norm $\otimes$:

$$\mathcal{A}(x) \stackrel{\mathrm{df}}{=} \bigotimes_{A_\imath \in \mathcal{A}} A_\imath(x).$$

This is a direct extension of definition (1) to the fuzzy case.

3.3 Quality Measures for Fuzzy Associations

Several generalizations of the quality measures (2) and (3) have been proposed in literature [32]. The standard approach is to replace set-theoretic operations, namely Cartesian product and cardinality, by corresponding fuzzy set-theoretic operations. Modeling the Cartesian product and the cardinality of a fuzzy set as defined in Section 3.1, one thus obtains

$$\mathsf{supp}(\mathcal{A} \rightharpoonup \mathcal{B}) \stackrel{\mathrm{df}}{=} \sum_{x \in \mathcal{D}} \mathcal{A}(x) \otimes \mathcal{B}(x), \tag{5}$$

$$\mathsf{conf}(\mathcal{A} \rightharpoonup \mathcal{B}) \stackrel{\mathrm{df}}{=} \frac{\sum_{x \in \mathcal{D}} \mathcal{A}(x) \otimes \mathcal{B}(x)}{\sum_{x \in \mathcal{D}} \mathcal{A}(x)}. \tag{6}$$

The most common choice for the t-norm $\otimes$ is the minimum, yet the product has also been applied.

Note that the support of $\mathcal{A} \rightharpoonup \mathcal{B}$ corresponds to the sum of the *individual supports* provided by the objects $x \in \mathcal{D}$ if the individual support is defined as

$$\mathsf{supp}_x(\mathcal{A} \rightharpoonup \mathcal{B}) \stackrel{\mathrm{df}}{=} \mathcal{A}(x) \otimes \mathcal{B}(x). \tag{7}$$

According to (7), an object x supports the rule $\mathcal{A} \rightharpoonup \mathcal{B}$ if it satisfies both, the antecedent $\mathcal{A}$ and the consequent $\mathcal{B}$.

3.4 Related Work

Even though the majority of contributions to fuzzy association analysis is based on the canonical extensions as outlined above (e.g. [5, 21, 41]), it is worth to mention some alternative proposals. Firstly, the use of implication operators instead of generalized conjunctions for modeling association rules has been motivated by some authors [12, 16, 51, 33], a point that we shall come back to in later sections. Secondly, interestingness indices other than the standard support and confidence measures have been discussed in several publications [44, 22, 32]. In [59], a *significance factor* and a kind of *certainty factor* have been proposed as a generalization of support and confidence, respectively. Moreover, so-called measures of *adjusted difference* and *weight of evidence* have been suggested in [5].

There are also some approaches in which the interpretation of fuzzy associations is quite different from the common understanding. In [92], for example, a fuzzy itemset is not a crisp set of fuzzy items (fuzzy sets) but rather a fuzzy set of crisp items. That is, a degree $\alpha \in [0, 1]$ is associated with each item, reflecting its relative importance in the itemset (importance can refer to different aspects such as, e.g., the frequency of the item in the itemset). Using the standard extension of the set-theoretic inclusion relation ($\mathcal{A} \subseteq \mathcal{B}$ iff $\mathcal{A}(x) \leq \mathcal{B}(x)$ for all x), the support and confidence measures for itemsets resp. association rules can be derived in the usual way.

Since the quality and efficiency of association analysis is strongly influenced by the fuzzy partitions of the numerical attributes involved, the problem of determining such partitions constitutes one of the most important preprocessing steps. Thus, it is hardly astonishing that much research effort has been devoted to this problem. For example, the effect of normalizing fuzzy membership degrees (in order to guarantee a partition of unity) has been investigated in [43]. One of the earliest automated methods for finding fuzzy sets in association rule mining has been proposed in [36]. Like many other approaches, this method makes use of clustering techniques in order to find fuzzy partitions that are in line with the structure of the data [17]. Apart from common clustering techniques such as k-means, approaches based on genetic algorithms have recently been developed [54]. An extension of the equi-depth partitioning algorithm [80], which allows for combining crisp values, intervals and fuzzy sets in the antecedent and consequent part of association rules, has been proposed in [99]. As a disadvantage of purely data-driven approaches to partitioning let us mention that they cannot guarantee the linguistic interpretability of the resulting fuzzy sets. In principle, this problem can be avoided by using predefined fuzzy partitions [6]. Needless to say, however, specifying all fuzzy partitions by hand might be a tedious job for the user of a data mining system, all the more if the dataset under consideration comprises a large number of attributes.

Apart from key questions concerning the preprocessing of numerical attributes and the fuzzy-logical modeling of association rules, several other extensions addressing diverse aspects of fuzzy association analysis can be found in the literature. For example, why should the frequency of a *fuzzy* pattern in a database be expressed in terms of a *scalar* cardinality? This question is as legitimate as the question why the cardinality of a fuzzy set should be a precise number [27]. An indeed, at the end of the paper we will point to a potential disadvantage of the σ-count (5). In [11, 31], the authors proposed to measure the support of a fuzzy association rule in terms of a *fuzzy set-valued cardinality.* On the one hand, a fuzzy cardinality of that kind comprises more information about the (statistical) occurrence of a fuzzy pattern. On the other hand, while scanning a database, the simple counting procedure (addition of membership degrees) sufficient for the standard support (5) has to be replaced by a more complex updating procedure.

Fuzzy association rules with *weighted* items are considered in [78]. Here, a degree of importance is assigned to each item and, correspondingly, weighted versions of the (fuzzy) support and confidence measures are proposed. An extension along the same line has been developed in [42]. In [8], the problem of mining changes in association rules is addressed: Given collections of association rules, which have been mined (in an evolving database) for different time periods, the authors seek to discover systematic changes in these rule sets (and hence of the underlying database). To this end, they induce so-called fuzzy meta-rules on the basis of a fuzzy decision tree. The problem of mining rules in fuzzy taxonomies, i.e., hierarchically structured sets of items, has been considered in [14]. Fuzzy taxonomies reflect partial belongings among items on different levels (e.g., a tomato can be regarded as both a fruit and a vegetable). The same paper also addresses the incorporation of *linguistic hedges* in fuzzy association rules.

Efficient algorithms supporting the mining process, i.e., the extraction of interesting associations from a database, have received less attention in the fuzzy community. This might be explained to some extent by the fact that, for fuzzy extensions of association analysis, standard algorithms can often be used or at least adapted in a relatively straightforward way. Still, some contributions have also been made in this field. For instance, a method for the parallel mining of fuzzy association rules was proposed in [90]. Besides, some authors have worked on methods for the reduction and the intelligent structuring of (fuzzy) association rules using the theoretical framework of formal concept analysis [10, 93].

Finally, a notable number of interesting applications of fuzzy association analysis have been realized. Just to mention some examples, fuzzy association rules have been used for medical data mining [23], for intrusion detection [65], for web access case adaptation [89], and for mining in bank-account databases [7].

4 Fuzzy Partitions of Examples

4.1 Rules and Conditional Objects

In the tradition of expert systems, a "rule" is understood as a production rule and associated with a modus-ponens-like deduction process. Thus, it is a kind of inference rule, even though it does not have a clear mathematical status. In more recent probabilistic expert systems, rules are encoded as conditional probabilities in a belief network. Even though this view of a weighted rule is mathematically sound, it is at odds with the logical tradition, since the probability of a material implication describing a rule clearly differs from

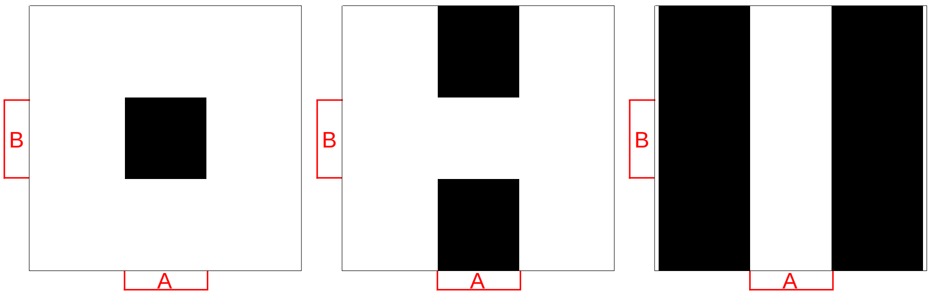

Figure 1: Illustration of the partition into positive (black region in the left picture), negative (middle) and irrelevant examples (right) induced by an association rule.

the corresponding conditional probability. This observation [61] has led to a vivid debate in philosophical circles since the late seventies [48] without fully settling the case.

The important point to notice is that a rule is not a two-valued entity but a three-valued one (see e.g. [29]). In fact, an (association) rule $\mathcal{A} \rightharpoonup \mathcal{B}$ partitions the data $\mathcal{D}$ into three types of objects, namely *positive examples* $\mathcal{S}_+$ that verify the rule, *negative examples* $\mathcal{S}_-$ that falsify the rule, and *irrelevant examples* $\mathcal{S}_\pm$ that do not provide any information about the rule:

$$\begin{aligned} \mathcal{S}_+ &\stackrel{\text{df}}{=} \{\, x \in \mathcal{D} \,|\, \mathcal{A}(x) \wedge \mathcal{B}(x) \,\} \\ \mathcal{S}_- &\stackrel{\text{df}}{=} \{\, x \in \mathcal{D} \,|\, \neg(\mathcal{A}(x) \Rightarrow \mathcal{B}(x)) \,\} \\ \mathcal{S}_\pm &\stackrel{\text{df}}{=} \{\, x \in \mathcal{D} \,|\, \neg\mathcal{A}(x) \,\} \end{aligned} \tag{8}$$

Fig. 1 illustrates this partition for an interval-based association rule with one-dimensional attributes.

Each of the three cases should be encoded by means of a different truth-value. The first and the second case correspond respectively to the usual truth-values "true" and "false" for the rule. The last case corresponds to a third truth-value, however. Depending on the context, it can be interpreted as, e.g., unknown, undetermined, or, as suggested above, as irrelevant. This idea of a rule as a "tri-event" actually goes back to De Finetti [20] in 1936. It is also the basis of De Finetti's approach to conditional probability.

Indeed, it is obvious that the probability $\mathsf{Prob}(\mathcal{B} \,|\, \mathcal{A})$ is entirely defined by $\mathsf{Prob}(\mathcal{A} \cap \mathcal{B})$ and $\mathsf{Prob}(\mathcal{A} \cap \overline{\mathcal{B}})$. (In the framework of association analysis, an *event* corresponds to a transaction, i.e., the subset $\mathcal{A} \subseteq \mathcal{R}$ of attributes with $A_\imath(x) = 1$ for a given x. The logical conjunction of two events $\mathcal{A}$ and $\mathcal{B}$, expressing that $A_\imath(x) = 1$ for the attributes in $\mathcal{A}$ *and* the attributes in $\mathcal{B}$, corresponds to an intersection of sets of objects $\mathcal{A} \cap \mathcal{B}$. This again clarifies the connection between the confidence of an association $\mathcal{A} \rightharpoonup \mathcal{B}$ and the conditional probability $\mathsf{Prob}(\mathcal{B} \,|\, \mathcal{A}) = \mathsf{Prob}(\mathcal{A} \cap \mathcal{B})/\mathsf{Prob}(\mathcal{A})$.)

This framework for modeling a rule suggests a mathematical model in which a rule is formalized as a *pair* of disjoint sets representing its examples and counter-examples, namely $(\mathcal{A} \cap \mathcal{B},\, \mathcal{A} \cap \overline{\mathcal{B}})$. This definition has several consequences. First, it justifies the claim made by De Finetti that a conditional probability $\mathsf{Prob}(\mathcal{A} \,|\, \mathcal{B})$ is the probability of a particular entity which can be called a *conditional event*, denoted $(\mathcal{B} \,|\, \mathcal{A})$. Second, it shows that the material implication does not fully capture the intended meaning of an "if-then" rule. It is obvious that the set of objects for which the material implication is true, $\{x \,|\, \neg\mathcal{A}(x) \vee \mathcal{B}(x)\}$, is the complement of the set of counter-examples of a rule. Thus, the usual logical view does not single out the examples of the rule, only its counter-examples. This is clearly in agreement with the fact that propositions in classical logic represent negative information in the sense of stating what is impossible (and combining in a conjunctive way what is left possibly true by different pieces of information). Still, the set of examples of a rule is $\{x \,|\, \mathcal{A}(x) \wedge \mathcal{B}(x)\}$ and clearly represents positive information. Thus, the three-valued representation of an "if-then" rule strongly suggests that a rule contains both positive and negative information. This also explains why two indices, support and confidence, are necessary to evaluate the quality of an association rule. In fact, the primitive quality indices of an association rule are the proportion of its examples and the proportion of its counter-examples.

As an aside, let us mention that this three-valued representation also tolerates non-monotonicity. It is intuitively satisfying to consider a rule R_1 = "If $\mathcal{A}$ then $\mathcal{B}$" to entail a rule R_2 = "If $\mathcal{C}$ then $\mathcal{D}$" if R_1 has more examples and less counter-examples than R_2 (in the sense of set inclusion). This can be formally written as

$$(\mathcal{B} \mid \mathcal{A}) \mid\!\sim (\mathcal{D} \mid \mathcal{C}) \Leftrightarrow \mathcal{A} \cap \mathcal{B} \models \mathcal{C} \cap \mathcal{D},\ \mathcal{C} \cap \overline{\mathcal{D}} \models \mathcal{A} \cap \overline{\mathcal{B}}.$$

The second condition (on exceptions) of this entailment corresponds to the classical inference between material conditionals. The non-monotonicity of this entailment is patent if we notice that the conditional object $(\mathcal{B} \mid \mathcal{A})$ does not entail $(\mathcal{B} \mid \mathcal{A} \cap \mathcal{C})$ generally, since the latter actually has less examples than the former. Indeed, it has been shown in [29] that the three-valued semantics of rules provides a representation for the calculus of conditional assertions of Kraus, Lehmann and Magidor [58], which is the basis of the rationality postulate-based approach to nonmotonic reasoning. The above entailment is also in agreement with probability theory since it was proved that, if $\mathcal{A} \cap \mathcal{B} \neq \emptyset$ and $\mathcal{C}$ is not a subset of $\mathcal{D}$, then, $(\mathcal{B} \mid \mathcal{A}) \mid\!\sim (\mathcal{D} \mid \mathcal{C})$ if and only if $\mathsf{Prob}(\mathcal{B}|\mathcal{A}) \leq \mathsf{Prob}(\mathcal{C}|\mathcal{D})$ for all probability measures $\mathsf{Prob}(\cdot)$ such that $\mathsf{Prob}(\mathcal{A}) \neq 0$ and $\mathsf{Prob}(\mathcal{C}) \neq 0$ on the underlying space [40].

4.2 Fuzzy Partitions

The key idea of our approach is to provide a sound basis for the assessment of fuzzy association rules by generalizing the aforementioned classification of data into *positive*, *negative*, and *irrelevant* examples of a rule. In fact, the most important quality measures for association rules (support and confidence) are naturally expressed in terms of the cardinalities 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:

$$\begin{aligned} \mathsf{supp}(\mathcal{A} \rightharpoonup \mathcal{B}) &\stackrel{\text{df}}{=} |\mathcal{S}_+|, \\ \mathsf{conf}(\mathcal{A} \rightharpoonup \mathcal{B}) &\stackrel{\text{df}}{=} |\mathcal{S}_+| \cdot \big(|\mathcal{S}_+| + |\mathcal{S}_-| \big)^{-1}. \end{aligned}$$

The basic question in connection with fuzzy association rules now concerns the generalization of the partition (8). Clearly, if $\mathcal{A}$ and $\mathcal{B}$ 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, an object x can be a positive (negative) example to some degree, and may also be irrelevant to some extent. We denote by $\mathcal{S}_+(x)$ the degree of membership of x in the fuzzy set $\mathcal{S}_+$ of positive examples and employ the same notation for $\mathcal{S}_-$ and $\mathcal{S}_\pm$.

The bipolar view of a rule as a conditional object suggests the following (logical) specification of positive, negative, and irrelevant examples:

$$\begin{aligned} x \in \mathcal{S}_+ &\overset{\text{df}}{\Leftrightarrow} \quad \mathcal{A}(x) \ \wedge \ \mathcal{B}(x) \\ x \in \mathcal{S}_- &\overset{\text{df}}{\Leftrightarrow} \neg(\mathcal{A}(x) \Rightarrow \mathcal{B}(x)) \\ x \in \mathcal{S}_\pm &\overset{\text{df}}{\Leftrightarrow} \neg\, \mathcal{A}(x) \end{aligned} \tag{9}$$

where $\Rightarrow$ denotes the material implication. In the fuzzy case, (9) translates into

$$\begin{aligned} \mathcal{S}_+(x) &\overset{\text{df}}{=} \mathcal{A}(x) \otimes \mathcal{B}(x) \\ \mathcal{S}_-(x) &\overset{\text{df}}{=} 1 - \big(\, \mathcal{A}(x) \rightsquigarrow \mathcal{B}(x) \,\big) \\ \mathcal{S}_\pm(x) &\overset{\text{df}}{=} 1 - \mathcal{A}(x) \end{aligned} \tag{10}$$

where $\rightsquigarrow$ is a generalized implication operator.

Moreover, a proper fuzzy partition into positive, negative, and irrelevant examples should satisfy

$$\mathcal{S}_+(x) + \mathcal{S}_-(x) + \mathcal{S}_\pm(x) = 1 \tag{11}$$

for all potential objects x [75]. This leads us to the **admissible operator problem:** Which generalized conjunctions (t-norms) $\otimes$ and generalized implications $\rightsquigarrow$ do satisfy (11) with $\mathcal{S}_+$, $\mathcal{S}_-$, and $\mathcal{S}_\pm$ given by (10)?

As an aside, let us note 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 four parts: strict preference (in both directions), indifference, and incompatibility [34].

Before investigating this problem in more detail, let us anticipate the critique that it might have been stated in an overly restrictive manner. First, it is true that fuzzy logic offers negation operators $n(\cdot)$ more general than the mapping $\alpha \mapsto 1 - \alpha$ as employed in (10). However, apart from the fact that this operator is the standard choice in most applications, it does also have desirable theoretical properties. For example, (up to an isomorphic transformation) it is the only involutive operator, i.e., the only $n(\cdot)$ such that $n(n(\alpha)) \equiv \alpha$. Second, one might think of replacing the addition of membership degrees in (11) by a disjunctive combination:

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

where $\oplus$ is a t-conorm. As noted above, t-conorms are commonly used as operations for set-theoretic union. One should realize, however, that in the context of data mining we are dealing with *frequency* information. From this point of view, (11) appears reasonable since it guarantees $|\mathcal{S}_+|+|\mathcal{S}_-|+|\mathcal{S}_\pm| = |\mathcal{D}|$, i.e., the sum of positive, negative, and irrelevant examples corresponds to the overall number of objects in the database (when using the standard cardinality). Moreover, note that (11) would be equivalent to

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

in the case where $\oplus$ is a strictly increasing t-conorm. Needless to say, this is a questionable property since it means that each object is either a completely positive or completely negative or completely irrelevant example.

4.3 Solution to the Fuzzy Partition Problem

Now, let us come back to the admissible operator problem as stated above. First, note that (11) in conjunction with (10) implies

$$\alpha \rightsquigarrow \beta \; = \; (1-\alpha) + (\alpha \otimes \beta) \tag{12}$$

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, (12) defines a form of the so-called QL-implication [4] with t-conorm $(\alpha, \beta) \mapsto \min(1, \alpha + \beta)$ as a disjunction (and $\alpha \mapsto 1-\alpha$ as a negation). A QL-implication is derived from a negation $n(\cdot)$, a t-conorm $\oplus$, and a t-norm $\otimes$ as follows: $\alpha \rightsquigarrow \beta = n(\alpha) \oplus (\alpha \otimes \beta)$. Thus, noting that $0 \leq (1-\alpha) + (\alpha \otimes \beta) \leq 1$ always holds since $\alpha \otimes \beta \leq \alpha$ for any t-norm $\otimes$, we obtain

$$\begin{aligned} \alpha \rightsquigarrow \beta \; &= \; (1-\alpha) + (\alpha \otimes \beta) \\ &= \; \min((1-\alpha) + (\alpha \otimes \beta), 1) \\ &= \; n(\alpha) \oplus (\alpha \otimes \beta) \end{aligned}$$

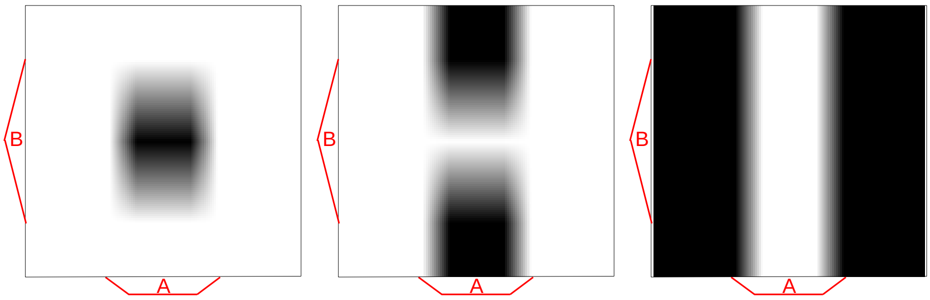

Figure 2: Illustration of the partition into positive (left), negative (middle) and irrelevant examples (right) in the fuzzy case ($\otimes$ = product).

with $n(\cdot)$ and $\oplus$ as above.

Here are some examples of standard conjunctions $\otimes$ together with induced implications:[3]

$\otimes$	$\rightsquigarrow$	
$\min(\alpha, \beta)$	Lukasiewicz:	$\min(1, 1 - \alpha + \beta)$
$\alpha\beta$	Reichenbach:	$1 - \alpha(1 - \beta)$
$\max(\alpha + \beta - 1, 0)$	Kleene-Dienes:	$\max(1 - \alpha, \beta)$

Note that $\mathcal{S}_-(x) = 1 - \mathcal{B}(x)$ appears rather natural in the case $\mathcal{A}(x) = 1$. Depending on the actual truth degree of the premise, $\alpha = \mathcal{A}(x)$, this basic evaluation is modified by the above implication operators in different ways, namely by shifting (Lukasiewicz), scaling (Reichenbach), or bounding (Kleene-Dienes); see Fig. 3.

The general question concerning the operators $\otimes$ and $\rightsquigarrow$ that can be chosen in (10) can be stated as follows: For which t-norms $\otimes$ does (12) 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, (12) is obviously increasing in β. Thus, as a major point it remains to guarantee that (12) is monotonically decreasing in α.

First of all, let us show that indeed not all t-norms are admissible, i.e., there are t-norms $\otimes$ for which (12) is not monotone decreasing in α. In fact,

[3] See Fig. 2 for an illustration of the product t-norm.

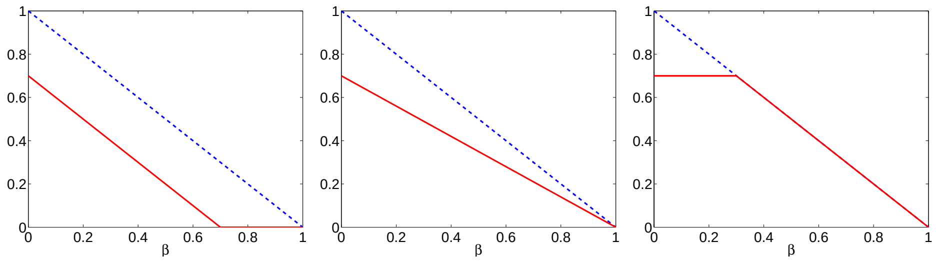

Figure 3: Modification of $\mathcal{S}_{-}(x) = 1 - \beta$ (dashed line) by shifting (Lukasiewicz, left), scaling (Reichenbach, middle), or bounding (Kleene-Dienes, right) in the case $\alpha = 0.7$.

a simple counter-example is the (weakly) drastic product [55], defined by $\alpha \otimes \beta = \min(\alpha, \beta)$ if $\max(\alpha, \beta) = 1$ and 0 otherwise, for which (12) becomes

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

Besides, there are even continuous t-norms that violate the above monotonicity condition. For instance, consider the HAMACHER family [45] of t-norms:

$$\alpha \otimes_\gamma \beta \; = \; \frac{\alpha\beta}{\gamma + (1-\gamma)(\alpha + \beta - \alpha\beta)}, \tag{13}$$

where γ is a non-negative parameter. With $\gamma = 10$, (12) 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 (see e.g. [57] for definitions and references).

Note that the monotonicity condition

$$(\alpha \le \alpha') \;\Rightarrow\; 1 - \alpha + (\alpha \otimes \beta) \;\ge\; 1 - \alpha' + (\alpha' \otimes \beta)$$

is equivalent to the following Lipschitz-condition on $\otimes$ which hence characterizes admissible operators:

$$(\alpha \le \alpha') \;\Rightarrow\; (\alpha' \otimes \beta) - (\alpha \otimes \beta) \;\le\; \alpha' - \alpha. \tag{14}$$

As a consequence, we find that a t-norm $\otimes$ is admissible in (12) if it is a so-called *copula* [68, 77]. In fact, the following result is stated as a theorem in [77]: A t-norm $\otimes$ is a copula iff (14) holds. A related result concerns continuous Archimedean t-norms in particular and shows that such t-norms are admissible in the sense of (14) if and only if their additive generator[4] is convex. For many parameterized families of t-norms [55], the latter result makes it easy to check whether or not a parameter is admissible. For instance, $\gamma \leq 1$ is necessary (and sufficient) for the HAMACHER family (13) to satisfy (14).

The following results provide further insight into the class of admissible t-norms (see [77] for proofs).

Proposition 1: The Lukasiewicz t-norm $\otimes_L : (\alpha, \beta) \mapsto \max(\alpha + \beta - 1, 0)$ is the smallest copula (i.e., the smallest t-norm admissible in the sense of (14)).

Proposition 2: For the family of Frank t-norms [35], parameterized through $\rho > 0$ according to

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

the operator $1 - \alpha + \alpha \otimes_\rho \beta$ is always monotone decreasing in α.

This proposition immediately follows from the fact that $(\alpha \otimes \beta) + (\alpha \oplus \beta) = \alpha + \beta$ holds for all Frank t-norms. Thus, substituting $\alpha \otimes \beta$ in (12) by $\alpha + \beta - (\alpha \oplus \beta)$, the right-hand side becomes $1 + \beta - (\alpha \oplus \beta)$, which is obviously monotone decreasing in α.

A further interesting result concerns the possibility of combining admissible t-norms into new admissible t-norms (see [68]).

Proposition 3: The ordinal sum[5] of copulas is again a copula.

[4] A definition of an additive generator can be found in the appendix.

[5] A definition of an ordinal sum is given in the appendix.

Since each element of the family

$$\otimes_\gamma : (\alpha, \beta) \mapsto \frac{\alpha\beta}{\max(\alpha, \beta, \gamma)}, \qquad 0 < \gamma \le 1 \tag{15}$$

of t-norms, introduced by Dubois and Prade [26], is an ordinal sum of the minimum and the product, we obtain that each t-norm (15) is admissible in the sense of (14).

5 Particular Types of Fuzzy Associations

In this section, we shall consider two special cases (refinements) of the admissible operator problem that will further reduce the class of solutions. In fact, in both cases a unique solution will be obtained. Moreover, in Section 5.3 we will reconsider the idea of *gradual* fuzzy rules in the context of association analysis.

5.1 Self-Implication

The first refinement concerns the demand for a property (axiom) which is motivated by the following observation: Consider a tautology in the form of a rule $\mathcal{A} \rightharpoonup \mathcal{A}$ with identical premise and conclusion part.[6] According to the solution that we have obtained above, such a rule is usually not fully confident, that is, it is thoroughly possible to have $\mathsf{conf}(\mathcal{A} \rightharpoonup \mathcal{A}) < 1$. As this might strike as odd, one might think of postulating $\mathsf{conf}(\mathcal{A} \rightharpoonup \mathcal{A}) = 1$ as an axiom. This postulate is closely related to the *self-implication* property of implication operators, namely $\alpha \rightsquigarrow \alpha = 1$ for all $\alpha \in [0, 1]$. More generally, it seems sensible to require the following property to hold:

$$(\forall\, x \in \mathcal{D} \,:\, \mathcal{A}(x) \le \mathcal{B}(x)) \;\Rightarrow\; (\mathsf{conf}(\mathcal{A} \rightharpoonup \mathcal{B}) = 1). \tag{16}$$

Now, it is obvious that this property holds iff

$$\mathcal{A}(x) \le \mathcal{B}(x) \;\Rightarrow\; \mathcal{S}_-(x) = 0 \tag{17}$$

[6]Strictly speaking, such a rule is of course forbidden if an attribute is not allowed to be part of the antecedent and consequent at the same time.

and, therefore, $\mathcal{S}_+(x) = 1 - \mathcal{S}_\pm(x) = \mathcal{A}(x) = \min\{\mathcal{A}(x), \mathcal{B}(x)\}$. In other words, requiring (16) to hold leads to the unique solution $\otimes = \min$, i.e., the minimum is the only admissible operator.

In connection with (17) it is interesting to note that the t-norm $\otimes$ can indeed be used to control the "punishment" of a pattern, as expressed by $\mathcal{S}_-(x)$. Using (12), $\mathcal{S}_-(x)$ can be written in the following form:

$$\mathcal{S}_-(x) = \mathcal{A}(x) - \mathcal{A}(x) \otimes \mathcal{B}(x).$$

As can be seen, the larger $\otimes$, the smaller $\mathcal{S}_-(x)$. Therefore, the smallest degree possible, $\mathcal{S}_-(x) = 0$, is obtained for the largest t-norm $\otimes = \min$.

5.2 Strong Implication Operators

A second specialization of the admissible operator problem is obtained by assuming a particular type of implication in (10), namely a strong implication operator. The latter is of the form $\alpha \rightsquigarrow \beta \stackrel{\mathrm{df}}{=} n(\alpha) \oplus \beta$, where $n(\cdot)$ is a strong negation. This definition is obviously derived from the logical equivalence $A \Rightarrow B \equiv \neg A \vee B$.

If $n(\cdot)$ is the standard negation and $\oplus$ the t-conorm associated with the t-norm $\otimes$ (i.e., $\alpha \oplus \beta = n(n(\alpha) \otimes n(\beta))$, then

$$1 - (\alpha \rightsquigarrow \beta) = 1 - (n(\alpha) \oplus \beta) = \alpha \otimes (1 - \beta).$$

Thus, the expression for $\mathcal{S}_-(x)$ can be simplified to $\mathcal{A}(x) \otimes (1 - \mathcal{B}(x))$, and we obtain the following special case of (10):

$$\begin{aligned} \mathcal{S}_+(x) &\stackrel{\mathrm{df}}{=} \mathcal{A}(x) \otimes \mathcal{B}(x) \\ \mathcal{S}_-(x) &\stackrel{\mathrm{df}}{=} \mathcal{A}(x) \otimes (1 - \mathcal{B}(x)) \\ \mathcal{S}_\pm(x) &\stackrel{\mathrm{df}}{=} 1 - \mathcal{A}(x) \end{aligned} \tag{18}$$

The admissible operator problem is now to find a t-norm $\otimes$ such that (11) holds with $\mathcal{S}_+$, $\mathcal{S}_-$, and $\mathcal{S}_\pm$ given by (18). This is equivalent to finding $\otimes$ such that

$$(\alpha \otimes \beta) + \alpha \otimes (1 - \beta) \equiv \alpha. \tag{19}$$

Interestingly enough, from ALSINA's results in [3] it follows that the only t-norm solving this problem is the product. In fact, in his paper ALSINA even considers a problem more general than (19), seeking solutions $(\otimes, \oplus, n)$ to the functional equation

$$(\alpha \otimes \beta) \oplus (\alpha \otimes n(\beta)) \equiv \alpha.$$

5.3 The Case of Gradual Rules

Fuzzy rules of the form "If P then C", where P and C are fuzzy propositions, play an important role not only in data mining (association analysis) but also in other fields such as e.g. automated control and approximate reasoning. Rules of this type can be interpreted in different ways [30]. Depending on the interpretation, different (fuzzy) logical operators are used for modeling a rule at a formal level.

A special type of fuzzy rule, referred to as *gradual rules*, combines the premise part P and the conclusion part C of a rule by means of a *residuated* implication operator $\rightsquigarrow$. The latter is derived from a t-norm $\otimes$ through residuation:

$$\alpha \rightsquigarrow \beta \stackrel{\text{df}}{=} \sup\{\, \gamma \,|\, \alpha \otimes \gamma \leq \beta \,\}. \tag{20}$$

This approach to modeling a rule is in agreement with the following interpretation of a gradual rule: "The more the premise P is true, the more the conclusion C is true" [74, 28], for example "The larger an object, the heavier it is".

So-called *pure* gradual rules are obtained when using the implication operator[7]

$$\alpha \rightsquigarrow \beta = \begin{cases} 1 & \text{if } \alpha \leq \beta \\ 0 & \text{if } \alpha > \beta \end{cases} \tag{21}$$

A rule "The more x is $\mathcal{A}$, the more x is $\mathcal{B}$" can then be interpreted as an ordinary constraint

$$\mathcal{A}(x) \leq \mathcal{B}(x) \tag{22}$$

[7]This operator is the core of all residuated implications (20).

This constraint is satisfied if x has property $\mathcal{B}$ at least as much as property $\mathcal{A}$, otherwise it is violated. (The operator (20) satisfies the property of self-implication, cf. Section 5.1.)

Having this constraint-based interpretation in mind, one might argue that the specification of positive examples in (10) is not fully in line with the semantics of a gradual rule. In fact, the meaning of the above constraint is obviously not captured by requiring an object x to satisfy both the condition and the conclusion part, as suggested by a conjunctive combination $\mathcal{A}(x) \wedge \mathcal{B}(x)$.[8] Indeed, requiring $\mathcal{A}$ and $\mathcal{B}$ to hold is clearly different from requiring $\mathcal{B}$ to hold at least as much as $\mathcal{A}$. However, simply replacing the conjunction in the definition of $\mathcal{S}_+(x)$ by an implication is questionable. For example, since an implication is true if its antecedent is false, an object x with $\mathcal{A}(x) = 0$ would fully support a rule $\mathcal{A} \rightharpoonup \mathcal{B}$.

As proposed in [51], a possible way out is to combine the implication $\mathcal{A}(x) \rightsquigarrow \mathcal{B}(x)$ conjunctively with the *relevance* of an object x for the rule, $\mathrm{Rel}_{\mathcal{A},\mathcal{B}}(x)$, thereby expressing that x supports $\mathcal{A} \rightharpoonup \mathcal{B}$ if

- it satisfies the rule in the sense of an implication (correctness), and
- it is a relevant (non-trivial) example for the rule in the sense that is satisfies the condition part (non-triviality).

This approach suggests a support measure of the following kind:

$$\mathsf{supp}_x(\mathcal{A} \rightharpoonup \mathcal{B}) \;=\; \mathrm{Rel}_{\mathcal{A},\mathcal{B}}(x) \otimes \big(\mathcal{A}(x) \rightsquigarrow \mathcal{B}(x) \big)\,.$$

Regarding the definition of $\mathrm{Rel}_{\mathcal{A},\mathcal{B}}(x)$, note that the constraint (22) is trivially satisfied only in the case $\mathcal{A}(x) = 0$. Therefore, it appears sensible to let $\mathrm{Rel}_{\mathcal{A},\mathcal{B}}(x) = 1$ if $\mathcal{A}(x) > 0$ and $= 0$ otherwise. In combination with the implication (21), this yields the following simple (non-fuzzy) partition:

[8] This conjunctive approach to modeling fuzzy rules has first been proposed by Mamdani and Assilian [66] in the context of fuzzy control and is now widely known as Mamdani rules.

	$\mathcal{A}(x) = 0$	$0 < \mathcal{A}(x) \leq \mathcal{B}(x)$	$\mathcal{A}(x) > \mathcal{B}(x)$
$\mathcal{S}_+(x)$	0	1	0
$\mathcal{S}_-(x)$	0	0	1
$\mathcal{S}_\pm(x)$	1	0	0

Depending on the type of application it might of course be reasonable to consider relevance resp. non-triviality as a gradual concept. For example, one might argue that the larger $\mathcal{A}(x)$, the more difficult it is to satisfy constraint (22), i.e., the less trivial this constraint becomes. In this case, an obvious definition of relevance (non-triviality) is $\mathrm{Rel}_{\mathcal{A},\mathcal{B}}(x) = \mathcal{A}(x)$. This approach suggests the following specification of a fuzzy partition:

$$
\begin{aligned}
\mathcal{S}_+(x) &\stackrel{\text{df}}{=} \mathcal{A}(x) \otimes \big(\mathcal{A}(x) \rightsquigarrow \mathcal{B}(x)\big) \\
\mathcal{S}_-(x) &\stackrel{\text{df}}{=} \mathcal{A}(x) \otimes \big(1 - (\mathcal{A}(x) \rightsquigarrow \mathcal{B}(x))\big) \\
\mathcal{S}_\pm(x) &\stackrel{\text{df}}{=} 1 - \mathcal{A}(x)
\end{aligned}
\tag{23}
$$

With regard to the admissible operators $\otimes$ and $\rightsquigarrow$, i.e., those operators satisfying (11) with $\mathcal{S}_+$, $\mathcal{S}_-$, and $\mathcal{S}_\pm$ given by (23), we can again refer to ALSINA's result. That is, the t-norm $\otimes$ is necessarily the product. Apart from that, however, any implication operator can in principle be chosen.

To illustrate, the following partition is obtained when using (21), regardless of the t-norm $\otimes$:

	$\mathcal{A}(x) \leq \mathcal{B}(x)$	$\mathcal{A}(x) > \mathcal{B}(x)$
$\mathcal{S}_+(x)$	$\mathcal{A}(x)$	0
$\mathcal{S}_-(x)$	0	$\mathcal{A}(x)$
$\mathcal{S}_\pm(x)$	$1 - \mathcal{A}(x)$	$1 - \mathcal{A}(x)$

Again, this result has an intuitively appealing interpretation: An object x is a positive example (and not a negative example, i.e., a negative example to degree 0) if it satisfies the constraint $\mathcal{A}(x) \leq \mathcal{B}(x)$, otherwise it is a negative example. The degree to which x is a positive resp. negative example corresponds to the degree to which it satisfies the antecedent $\mathcal{A}$, i.e., to its degree of relevance.

Of course, a questionable property of the above measure is its discontinuity (on the diagonal $\{(\alpha,\alpha) \,|\, 0 \leq \alpha \leq 1\}$): A slight variation of $\mathcal{B}(x)$ can have an extreme influence on the degree to which an object is a positive resp. negative example. This problem is obviously caused by the discontinuous implication operator (21) and the associated "hard" constraint (22).

A weakening of this constraint can be obtained by using other (larger) residuated implication operators such as, e.g., the GOGUEN implication

$$\alpha \rightsquigarrow \beta = \begin{cases} 1 & \text{if } \alpha \leq \beta \\ \beta/\alpha & \text{if } \alpha > \beta \end{cases} \tag{24}$$

In this case, a rule can be violated in a partial way. Or, stated differently, a rule can be satisfied to some extent even if $\mathcal{A}(x) > \mathcal{B}(x)$. When using (24) in conjunction with the product t-norm, the following partition is obtained:

	$\mathcal{A}(x) \leq \mathcal{B}(x)$	$\mathcal{A}(x) > \mathcal{B}(x)$
$\mathcal{S}_+(x)$	$\mathcal{A}(x)$	$\mathcal{B}(x)$
$\mathcal{S}_-(x)$	0	$\mathcal{A}(x) - \mathcal{B}(x)$
$\mathcal{S}_\pm(x)$	$1 - \mathcal{A}(x)$	$1 - \mathcal{A}(x)$

Now, x is to some extent a positive example even if it violates the constraint $\mathcal{A}(x) \leq \mathcal{B}(x)$. In fact, the degree to which it is a negative example now depends on "how much" the constraint is violated, as expressed by the difference $\mathcal{A}(x) - \mathcal{B}(x)$. Interestingly enough, this result exactly corresponds to the result that is obtained when using $\otimes = \min$ in the general approach (10). And indeed, it was already noticed that min is the only t-norm that – in connection with (10) – guarantees property (16), which in turn is in agreement with the concept of a gradual rule.

More generally, $\alpha \otimes (\alpha \rightsquigarrow \beta) = \min(\alpha, \beta)$ if $\otimes$ is a continuous t-norm and $\rightsquigarrow$ is the implication derived from that t-norm through residuation.[9] In other words, when using a continuous t-norm $\otimes$ together with the associated residuated implication, then $\mathcal{S}_+(x) = \min(\mathcal{A}(x), \mathcal{B}(x))$. Thus, we have obtained

[9] Using the definition (20) of a residuated implication,this equality is easy to prove.

yet another result that emphasizes the particular role of the support measure (5) with $\otimes = \min$.

5.4 Summary of Results

In summary, the results that we have obtained in this section suggest three reasonable support measures of the form $\sum_{x \in \mathcal{D}} \mathsf{supp}_x(\mathcal{A} \rightharpoonup \mathcal{B})$ for fuzzy associations, where

$$\begin{aligned}
1.\quad \mathsf{supp}_x(\mathcal{A} \rightharpoonup \mathcal{B}) &= \min\{\mathcal{A}(x), \mathcal{B}(x)\} \\
2.\quad \mathsf{supp}_x(\mathcal{A} \rightharpoonup \mathcal{B}) &= \mathcal{A}(x) \cdot \mathcal{B}(x) \\
3.\quad \mathsf{supp}_x(\mathcal{A} \rightharpoonup \mathcal{B}) &= \mathcal{A}(x) \cdot (\mathcal{A}(x) \rightsquigarrow \mathcal{B}(x))
\end{aligned}$$

The first measure guarantees the confidence condition (16) to hold. The second measure is obtained when defining the counterexamples of a rule in a particular (though natural) way, namely in the form (18). The third measure is in agreement with a gradual rule interpretation of a fuzzy association.

As an aside, note that since the degree of relevance of an example is $\mathcal{A}(x)$ in all of the above cases, the confidence measure of a fuzzy association is still given by

$$\mathsf{conf}(\mathcal{A} \rightharpoonup \mathcal{B}) = \frac{\sum_{x \in \mathcal{D}} \mathsf{supp}_x(\mathcal{A} \rightharpoonup \mathcal{B})}{\sum_{x \in \mathcal{D}} \mathcal{A}(x)}$$

6 Concluding Remarks

Several fuzzy extensions of association rule mining have already been proposed in literature. Despite of this great interest in fuzzy associations, the question of how to evaluate fuzzy patterns in a proper way has mostly been approached in an ad-hoc manner.

In this paper, we have proposed a formal framework for constructing evaluation measures in a systematic way. Our approach is based on the idea of partitioning the data into positive, negative, and irrelevant examples of a rule, and to derive measures from the cardinalities of these sets. This approach is rather general and can in principle be applied to any type of fuzzy pattern.

The possibility to specify positive and negative examples in a logical way appears to be especially useful in connection with association rules, as it allows one to reflect the semantics of a rule in a more or less direct way. Indeed, our approach has shown that different types of fuzzy rules call for different evaluation measures, depending on their semantic interpretation. Interestingly enough, the differences between various types of rules, such as conjunction-based and implication-based rules, becomes obvious only in the fuzzy case. It remains invisible, however, in the non-fuzzy case, where different interpretations formally coincide. This might explain to some extent that semantic issues have not received much attention in association rule mining so far.

On the one hand, our results provide a sound justification of the commonly used support and confidence measures (5–6) and, in this connection, point out the particular role of the product and minimum t-norms. On the other hand, alternative measures might be meaningful for certain applications, especially in connection with implication-based fuzzy rules (gradual rules). The best choice of an operator does of course depend on the type of application and cannot be answered in general. Anyway, an interesting question that deserves consideration concerns the difference between operators regarding the data mining results. Our current impression is that changing the t-norm in the evaluation measures (5–6) does not change the results (set of interesting association rules) dramatically. There is a stronger difference, however, between the conjunction-based and the implication-based approach. Studying these issues on an experimental basis is an important topic of future work.

Since algorithmic aspects of rule mining have not been addressed in this paper, let us mention that the standard methods based on the Apriori principle can easily be extended to the fuzzy case. If the support measure is expressed in terms of a conjunction, this extension is indeed straightforward and has already been implemented by several authors. An extension for the case where the logical expression of positive examples involves an implication

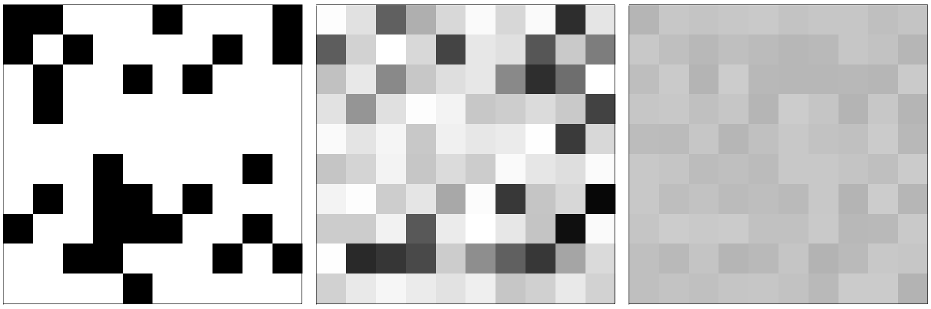

Figure 4: Objects (statistical entities) are the fields of a 10×10 array. The property of interest is the attribute `light` which is in direct correspondence with the level of grey.

operator (cf. Section 5.3) has been proposed in [52].

Before concluding the paper, let us point to a caveat that concerns the derivation of frequency information such as, e.g., frequency-based evaluation measures, in the fuzzy case where attributes can apply to statistical entities in a partial way. To exemplify the problem, consider the three 10×10 arrays shown in Fig. 4. Each of these arrays consists of 100 fields (the statistical entities) that are light to a certain degree (as represented by the corresponding level of grey). Suppose that we are interested in the frequency of light fields: What is the support of the attribute `light`?

Intuitively, `light` seems to be well supported in the first case (left picture) where fields are either completely black or completely white. Indeed, 75 out of 100 fields are white, i.e., the support is $\texttt{supp}(\texttt{light}) = 75$ in this case. As opposed to this, the property `light` does not seem to be well supported in the third case (right), where actually none of the fields is really light. Still, if $\texttt{supp}(\texttt{light})$ is computed by the sum of membership degrees $\mathcal{S}_+(x)$, the support is again 75. In fact, the same support is obtained for the second case as well.

This counter-intuitive result is due to the fact that, when using the σ-count for computing the cardinality in the fuzzy case, several small membership degrees can compensate for a few large degrees [32]. Admittedly, the third case above is an extreme example. Still, it clearly reveals a weakness of the standard aggregation of membership degrees.

There are different possibilities for coping with the above problem. One idea is to replace the scalar cardinality (σ-count) of a fuzzy set by a fuzzy set-valued cardinality. As noted in Section 3.4, the latter comprises more information about the (statistical) distribution of membership degrees. Still, as a disadvantage of this approach let us note that it makes the specification of support thresholds more difficult, since such thresholds would no longer be scalar. Moreover, the computation of measures that are derived from the support requires arithmetic operations for fuzzy cardinalities.

An alternative approach is to complement the assessment of fuzzy associations by a measure of "clarity" which might be defined as

$$\frac{1}{n}\sum_{x} |1 - 2\mathcal{S}_{+}(x)|. \tag{25}$$

The idea of this measure, which is directly related to a measure of fuzziness introduced in [53], is to compare membership degrees to the midpoint 1/2 which is considered to be the "least clear" situation. Of course, this measure could be replaced by any other measure of fuzziness or fuzzy entropy [63] which aims at quantifying a lack of distinction between a fuzzy set and its complement (e.g. [91]). In our example, the clarity degree (25) is 1 for the first case in our example whereas it is only ≈ 0.5 for the third case. Using such an additional measure for fuzzy associations seems to be natural: Since fuzzy rules are more complex resp. flexible objects than non-fuzzy rules, it is hardly surprising that their proper evaluation requires taking further criteria into account. Elaborating on these ideas in more detail is a topic of ongoing work.

A Additive Generator of a T-Norm

An additive generator of a t-norm $\otimes$ is a mapping $f : [0,1] \to [0,\infty]$ such that

- f is continuous and monotone decreasing,

- $f(1) = 0$,
- $\alpha \otimes \beta = f^{(-1)}(f(\alpha) + f(\beta))$ for all $0 \leq \alpha, \beta \leq 1$,

where $f^{(-1)}(\cdot)$ denotes the pseudo-inverse of $f(\cdot)$, i.e.,

$$f^{(-1)}(x) \;=\; \begin{cases} y & \text{if } 0 \leq x \leq f(0) \text{ and } y = f(x) \\ 0 & \text{if } f(0) < x \leq \infty \end{cases}$$

for all $x \geq 0$. It can be shown that an additive generator does exist for each Archimedian t-norm, and that this generator is unique up to a positive multiplicative constant [62]. (A t-norm $\otimes$ is Archimedian if $\alpha \otimes \alpha < \alpha$ for all $0 < \alpha < 1$.)

B Ordinal Sums

Suppose n t-norms $\otimes^1 \ldots \otimes^n$ to be given. Moreover, let $u_1 \ldots u_n$ and $v_1 \ldots v_n$ be numbers such that $0 \leq u_1 < v_1 < u_2 < v_2 < \ldots < u_n < v_n \leq 1$. The ordinal sum $\otimes$ of $\otimes^1 \ldots \otimes^n$ is given by

$$\alpha \otimes \beta \stackrel{\mathrm{df}}{=} \begin{cases} u_\imath + (v_\imath - u_\imath)\left(\frac{\alpha - u_\imath}{v_\imath - u_\imath} \otimes^\imath \frac{\beta - u_\imath}{v_\imath - u_\imath}\right) & \text{if } u_\imath < \alpha, \beta < v_\imath \\ \min(\alpha, \beta) & \text{otherwise} \end{cases},$$

for all $0 \leq \alpha, \beta \leq 1$.

As an example, the t-norm (15) is obtained for $n = 1$, $\otimes^1 =$ product, $u_1 = 0$, $v_1 = \gamma$.

References

[1] CC. Aggarwal and PS. Yu. A new framework for itemset generation. In *PODS-98, Symposium on Principles of Database Systems*, pages 18–24, Seattle, WA, USA, 1998.

[2] R. Agrawal and R. Srikant. Fast algorithms for mining association rules. In *Proceedings of the 20th Conference on* VLDB, pages 487–499, Santiago, Chile, 1994.

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

[4] C. Alsina, E. Trillas, and L. Valverde. On some logical connective for fuzzy sets theory. *J. Math. Anal. Appl.*, 93:15–26, 1983.

[5] Wai-Ho Au and KCC. Chan. An effective algorithm for discovering fuzzy rules in relational databases. In *Proceedings* IEEE *World Congress on Computational Intelligence*, pages 1314–1319, Piscataway, NJ, 1998.

[6] Wai-Ho Au and KCC. Chan. FARM: A data mining system for discovering fuzzy association rules. In *Proceedings* FUZZ–IEEE–99, pages 1217–1222, Seoul, Korea, 1999.

[7] Wai-Ho Au and KCC. Chan. Mining fuzzy association rules in a bank-account database. *IEEE Transactions on Fuzzy Systems*, 11(2):238–248, 2003.

[8] Wai-Ho Au and KCC. Chan. Mining changes in association rules: a fuzzy approach. *Fuzzy Sets and Systems*, 149(1):87–104, 2005.

[9] B. Barber and HJ. Hamilton. Extracting share frequent itemsets with infrequent subsets. *Data Mining and Knowledge Discovery*, 7:153–185, 2003.

[10] R. Belohlavek. Fuzzy Galois connections. *Mathematical Logic Journal*, 45(4):497–504, 1999.

[11] P. Bosc, D. Dubois, O. Pivert, and H. Prade. On fuzzy association rules based on fuzzy cardinalities. In *Proceedings* FUZZ-IEEE-2001*, 10th* IEEE *International Conference on Fuzzy Systems*, Melbourne, Australia, 2001.

[12] P. Bosc and O. Pivert. On some fuzzy extensions of association rules. In *Proc. IFSA/NAFIPS-2001*, Vancouver, Canada, 2001.

[13] CH. Cai, A. Fu, CH. Cheng, and WW. Kwong. Mining association rules with weighted items. In *IDEAS–98, International Database Engineering and Application Symposium*, pages 68–77, Cardiff, Wales, 1998.

[14] G. Chen and Q. Wei. Fuzzy association rules and the extended mining algorithms. *Information Sciences*, 147(1–4):201–228, 2002.

[15] G. Chen, Q. Wei, E. Kerre, and G. Wets. Overview of fuzzy associations mining. In *Proc. ISIS–2003, 4th International Symposium on Advanced Intelligent Systems*. Jeju, Korea, September 2003.

[16] 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.

[17] BC. Chien, ZL. Lin, and TP. Hong. An efficient clustering algorithm for mining fuzzy quantitative association rules. In *Proc. 9th Int. Fuzzy Systems Association World Congress*, pages 1306–1311, Vancouver, 2001.

[18] F. Coenen, G. Goulbourne, and P. Leng. Tree structures for mining association rules. *Data Mining and Knowledge Discovery*, 8:25–51, 2004.

[19] F. Coenen, P. Leng, and S. Ahmed. Data structures for association rule mining: T-trees and P-trees. *IEEE Transactions on Knowledge and Data Engineering*, 16(6):774–778, 2004.

[20] B. DeFinetti. La logique des probabilités. *Int. Cong. de Philosophie Scientifique*, 5:1–9, 1936.

[21] JM. DeGraaf, WA. Kosters, and JJW. Witteman. Interesting fuzzy association rules in quantitative databases. In *Proc. 5th European Conference on Principles and Practice of Knowledge Discovery in Databases, PKDD–01*, pages 140–151, Freiburg, Germany, 2001.

[22] M. Delgado, N. Marin, D. Sanchez, and MA. Vila. Fuzzy association rules: general model and applications. *IEEE Transactions on Fuzzy Systems*, 11(2):214–225, 2003.

[23] 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.

[24] J. Dougherty, R. Kohavi, and M. Sahami. Supervised and unsupervised discretization of continuous features. In A. Prieditis and S. Russell, editors, *Machine Learning: Proceedings of the 12th International Conference*, pages 194–202. Morgan Kaufmann, 1995.

[25] D. Dubois, E. Hüllermeier, and H. Prade. A note on quality measures for fuzzy association rules. In *Proc.* IFSA–03*, 10th International Fuzzy Systems Association World Congress*, number 2715 in LNAI, pages 677–648, Istambul, 2003. Springer-Verlag.

[26] D. Dubois and H. Prade. New results about properties and semantics of fuzzy-set-theoretic operators. In PP. Wang and SK. Chang, editors, *Fuzzy Sets: Theory and Applications to Policy Analysis and Information Systems*, pages 59–75. Plenum Publ., 1980.

[27] D. Dubois and H. Prade. Fuzzy cardinality and the modeling of imprecise quantification. *Fuzzy Sets and Systems*, 16:199–230, 1985.

[28] D. Dubois and H. Prade. Gradual inference rules in approximate reasoning. *Information Sciences*, 61(1,2):103–122, 1992.

[29] D. Dubois and H. Prade. Conditional objects as non-monotonic consequence relationships. *IEEE Transactions on Systems, Man and Cybernetics*, 24(12):1724–1739, 1994.

[30] D. Dubois and H. Prade. What are fuzzy rules and how to use them. *Fuzzy Sets and Systems*, 84:169–185, 1996.

[31] D. Dubois and H. Prade. Fuzzy sets in data summaries – outline of a new approach. In *Proceedings* IPMU-2000*, International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 1035–1040, Madrid, Spain, 2000.

[32] D. Dubois, H. Prade, and T. Sudkamp. A discussion of indices for the evaluation of fuzzy associations in relational databases. In *Proceedings* IFSA–03*, 10th International Fuzzy Systems Association World Congress*, pages 111–118, Istambul, 2003.

[33] D. Dubois, H. Prade, and T. Sudkamp. On the representation, measurement, and discovery of fuzzy associations. *IEEE Transactions on Fuzzy Systems*, 13(2):250–262, 2005.

[34] J. Fodor and M. Roubens. *Fuzzy Preference Modelling and Multicriteria Decision Support*. Kluwer, 1994.

[35] M.J. Frank. On the simultaneous associativity of $f(x,y)$ and $x+y-f(x,y)$. *Aeq. Math.*, 19:194–226, 1979.

[36] 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 International Symposium on Intelligent Data Engineering and Learning*, pages 263–268, Hong Kong, 1998.

[37] Yo Fu and J. Han. Meta-rule-guided mining of association rules in relational databases. In *KDOOD/TDOOD*, pages 39–46, Singapore, 1995.

[38] B. Goethals and MJ. Zaki. Advances in frequent itemset mining implementations. In B. Goethals and MJ. Zaki, editors, *Proc. IEEE ICDM Workshop on Frequent Itemset Mining Implementations*, Melbourne, Florida, USA, 2003.

[39] B. Goethals and MJ. Zaki. Advances in frequent itemset mining implementations: Report on FIMI'03. *SIGKDD Explorations*, 6(1):109–117, 2004.

[40] IR. Goodman, HT. Nguyen, and EA. Walker. *Conditional Inference and Logic for Intelligent Systems*. North-Holland, Amsterdam, 1991.

[41] A. Gyenesei. A fuzzy approach for mining quantitative association rules. Technical Report 336, Turku Centre for Computer Science, 2000.

[42] A. Gyenesei. Mining weighted association rules for fuzzy quantitative items. Technical Report 346, Turku Centre for Computer Science, 2000.

[43] A. Gyenesei. A fuzzy approach for mining quantitative association rules. *Acta Cybernetica*, 15:305–320, 2001.

[44] A. Gyenesei. Interestingness measures for fuzzy association rules. In *Proc. 5th European Conference on Principles and Practice of Knowledge Discovery in Databases, PKDD–01*, pages 152–164, Freiburg, Germany, 2001.

[45] H. Hamacher. *Über logische Aggregationen nichtbinär explizierter Entscheidungskriterien; Ein axiomatischer Beitrag zur normativen Entscheidungstheorie (About logical aggregation of non-binary explicated decision criteria; an axiomatic contribution to normative decision theory).* R.G. Fischer Verlag, 1978.

[46] J. Han and Y. Fu. Discovery of multi-level association rules from large databases. In *Proceedings of the International Conference on Very Large Databases*, pages 420–431, Zurich, Switzerland, 1995.

[47] J. Han, J. Pei, Y. Yin, and R. Mao. Mining frequent patterns without candidate generation. *Data Mining and Knowledge Discovery*, 8:53–87, 2004.

[48] WL. Harper, R. Stalnaker, and G. Pearce, editors. *IFS.* D. Reidel, Dordrecht, The Netherlands, 1981.

[49] J. Hipp, U. Güntzer, and G. Nakhaeizadeh. Algorithms for association rule mining – a general survey and comparison. *Newsletter of the Special Interest Group on Knowledge Discovery and Data Mining*, 2(1):58–64, 2000.

[50] E. Hüllermeier. Fuzzy association rules: Semantic issues and quality measures. In *Proc. International Conference on Computational Intelligence – 7th Fuzzy Days*, number 2206 in LNCS, pages 380–391, Dortmund, Germany, 2001. Springer-Verlag.

[51] E. Hüllermeier. Implication-based fuzzy association rules. In *Proc. 5th European Conference on Principles and Practice of Knowledge Discovery in Databases, PKDD–01*, pages 241–252, Freiburg, Germany, 2001. Springer-Verlag.

[52] 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.

[53] A. Kaufmann. *Introduction to the theory of fuzzy subsets*. Academic Press, New York, 1975.

[54] M. Kaya and R. Alhajj. Genetic algorithm based framework for mining fuzzy association rules. *Fuzzy Sets and Systems*, 152(3):587–601, 2005.

[55] EP. Klement, R. Mesiar, and E. Pap. *Triangular Norms*. Kluwer Academic Publishers, 2002.

[56] M. Klemettinen, H. Mannila, P. Ronkainen, H. Toivonen, and A. Verkamo. Finding interesting rules from large sets of discovered association rules. In *Third International Conference on Information and Knowledge Management, CIKM–94*, pages 401–407, 1994.

[57] G.J. Klir and T.A. Folger. *Fuzzy Sets, Uncertainty, and Information*. Prentice Hall, 1988.

[58] S. Kraus, D. Lehmann, and M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44:167–207, 1990.

[59] C. Man Kuok, A. Fu, and M. Hon Wong. Mining fuzzy association rules in databases. *SIGMOD Record*, 27:41–46, 1998.

[60] B. Lent, A. Swami, and J. Widom. Clustering association rules. In *Proceedings* ICDE–97, Birmingham, UK, 1997.

[61] D. Lewis. Probabilities of conditionals and conditional probabilities. *Journal of Philosophical Logic*, 3, 1973.

[62] CH. Ling. Representations of associative functions. *Publications Mathematicae Debrecen*, 12:189–212, 1965.

[63] A. De Luca and S. Termini. A definition of non-probabilistic entropy in the setting of fuzzy sets theory. *Information and Control*, 24:301–312, 1972.

[64] C. Lucchesse, S. Orlando, and R. Perego. DCI-Closed: a fast and memory efficient algorithm to mine frequent closed itemsets. In B. Goethals and MJ. Zaki, editors, *Proc. IEEE ICDM Workshop on Frequent Itemset Mining Implementations*, Brighton, UK, 2004.

[65] J. Luo and S. Bridges. Mining fuzzy association rules and fuzzy frequency episodes for intrusion detection. *International Journal of Intelligent Systems*, 15(8):687–703, 2000.

[66] E. Mamdani and S. Assilian. An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies*, 7:1–13, 1975.

[67] R.J. Miller and Y. Yang. Association rules over interval data. In *Proceedings of the* ACM SIGMOD *International Conference on Management of Data*, pages 452–461, 1997.

[68] RB. Nelsen. *An Introduction to Copulas*. Springer-Verlag, 1999.

[69] M. Ohsaki, S. Kitaguchi, K. Okamoto, H. Yokoi, and T. Yamaguchi. Evaluation of rule interestingness measures with a clinical dataset on hepatitis. pages 362–373, Pisa, Italy, 2004. Springer-Verlag.

[70] J.S. Park, M.S. Chen, and P.S. Yu. An efficient hash-based algorithm for mining association rules. In *Proceedings* ACM SIGMOD *International Conference on Management of Data*, pages 175–186, 1995.

[71] N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal. Discovering frequent closed itemsets for association rules. In *Proc. 7th International Conference on Database Theory*, number 1540 in LNCS, pages 398–416. Springer, 1999.

[72] W. Pedrycz. Data mining and fuzzy modeling. In *Proc. of the Biennial Conference of the NAFIPS*, pages 263–267, Berkeley, CA, 1996.

[73] J. Pei, J. Han, and LVS. Lakshmanan. Mining frequent itemsets with convertible constraints. In *Proc. 17th International Conference on Data Engineering*, pages 433–442, Heidelberg, Germany, 2001.

[74] H. Prade. Raisonner avec des règles d'inférence graduelle - Une approche basée sur les ensembles flous. *Revue d'Intelligence Artificielle*, 2(2):29–44, 1988.

[75] E.H. Ruspini. A new approach to clustering. *Information Control*, 15:22–32, 1969.

[76] A. Savasere, E. Omiecinski, and S. Navathe. An efficient algorithm for mining association rules in large databases. In VLDB–95*, Proceedings of 21th International Conference on Very Large Data Bases*, pages 11–15, Zurich, September 1995.

[77] B. Schweizer and A. Sklar. *Probabilistic Metric Spaces.* North-Holland, New York, 1983.

[78] J. Shu-Yue, E. Tsang, D. Yengg, and S. Daming. Mining fuzzy association rules with weighted items. In *Proc. IEEE Int. Conference on Systems, Man, and Cybernetics*, pages 1906–1911, Nashville, TN, 2000.

[79] R. Srikant and R. Agrawal. Mining generalized association rules. In *Proceedings of the International Conference on Very Large Databases*, pages 407–419, Zurich, Switzerland, 1995.

[80] 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.

[81] R. Srikant, Q. Vu, and R. Agrawal. Mining association rules with item constraints. In *Proc. of the 3rd Int'l Conference on Knowledge Discovery in Databases and Data Mining*, pages 67–73, Newport Beach, California, 1997.

[82] M. Steinbach, PN. Tan, and V. Kumar. Support envelopes: A technique for exploring the structure of association patterns. In *Proc. 10th ACM SIGKDD Conference*, Seattle, WA, 2004.

[83] O. Strauss, F. Comby, and MJ. Aldon. Rough histograms for robust statistics. In *ICPR–2000, 15th International Conference on Pattern Recognition*, pages 2684–2687. Barcelona, September 2000.

[84] G. Stumme, R. Taouil, Y. Bastide, N. Pasquier, and L. Lakhal. Intelligent structuring and reducing of association rules with formal concept analysis. In *Proc. 24th German Conference on Artificial Intelligence*, volume 2174 of *LNCS*. Springer-Verlag, 2001.

[85] T. Sudkamp. Examples, counterexamples, and measuring fuzzy associations. *Fuzzy Sets and Systems*, 149(1), 2005.

[86] P. Tan, V. Kumar, and J. Srivastava. Selecting the right interestingness measure for association patterns. In *Proc. 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 32–41. ACM Press, 2002.

[87] F. Tao, F. Murtagh, and M. Farid. Weighted association rule mining using weighted support and significance framework. In *Proc. 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 661–666, Washington DC, 2003.

[88] T. Uno, M. Kiyomi, and H. Arimura. LCM ver. 2: Efficient mining algorithms for frequent/closed/maximal itemsets. In B. Goethals and MJ. Zaki, editors, *Proc. IEEE ICDM Workshop on Frequent Itemset Mining Implementations*, Brighton, UK, 2004.

[89] C. Wong, S. Shiu, and SK. Pal. Mining fuzzy association rules for web access case adaptation. In S. Shiu, SK. Pal, and TS. Dillon, editors, *ICCBR-Workshop on Soft Computing in Case-Based Reasoning*, Vancouver, 2001.

[90] B. Xu, J. Lu, Y. Zhang, L. Xu, H. Chen, and H. Yang. Parallel algorithm for mining fuzzy association rules. In *International Conference on Cyberworlds*, Singapore, 2003.

[91] RR. Yager. On the measure of fuzziness and negation-part I: Membership in the unit interval. *Int. Journal of General Systems*, 5:221–229, 1979.

[92] S. Ben Yahia and A. Jaoua. A top-down approach for mining fuzzy association rules. In *Proc. 8th Int. Conf. Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 952–959, Madrid, 2000.

[93] S. Ben Yahia and A. Jaoua. Discovering knowledge from fuzzy concept lattice. In A. Kandel, M. Last, and H. Bunke, editors, *Data mining and Computational Intelligence*, volume 68 of *Studies in Fuzziness and Soft Computing*, pages 167–190. Physica-Verlag, 2001.

[94] 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.

[95] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[96] M. Zaki and CJ. Hsiao. Efficient algorithms for mining closed itemsets and their lattice structure. *IEEE Transactions on Knowledge and Data Engineering*, 17(4):462–478, 2005.

[97] MJ. Zaki. Scalable algorithms for association mining. *IEEE Transactions on Knowledge and Data Engineering*, 12(3):372–390, 2000.

[98] MJ. Zaki, S. Parthasarathy, M. Ogihara, and W. Li. New algorithms for fast discovery of association rules. Technical Report 651, Computer Science Department, University of Rochester, Rochester, NY 14627, July 1997.

[99] W. Zhang. Mining fuzzy quantitative association rules. In *Proceedings of the 11th* IEEE *International Conference on Tools with Artificial Intelligence*, Chicago, Illinois, 1999.