

A Systematic Approach to the Assessment of Fuzzy Association Rules

Didier Dubois
IRIT-UPS
Toulouse, France

Eyke Hüllermeier
Department of Computer Science, LS1
University of Dortmund, 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 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.

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 $\mathcal{A} \rightarrow \mathcal{B}$ 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 $\{\text{paper, envelopes}\} \rightarrow \{\text{stamps}\}$ 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 has been motivated by numerous authors (see [5, 7] for recent overviews). By allowing for “soft” rather than crisp boundaries of intervals, fuzzy sets can avoid certain undesirable threshold effects [35]. 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 [14]. Especially, it assumes an understanding of the semantics of a fuzzy rule [20]. 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 and elaborated in Section 4. The paper concludes with a summary and an outlook on future work in Section 5.

2. Association Rules

2.1. Binary Association Rules

Let $\mathcal{D} = \{x_1, x_2 \dots x_n\}$ be a set of objects and $\mathcal{R} = \{A_1, A_2 \dots A_m\}$ a set of binary attributes. Each attribute is interpreted as a property (feature). We define $A_i(x_j) = 1$ if the i -th property applies to the object x_j and $A_i(x_j) = 0$ otherwise. Each property is thus encoded by a binary attribute with domain $\{0, 1\}$, and the data we proceed from can be thought of as a binary relation of dimension $n \times m$.

As mentioned above, the attributes A_i are often referred to as *items* in the context of association rule mining. Moreover, the set $\{A_i \mid A_i(x_j) = 1\} \subseteq \mathcal{R}$ of features that apply to an object x_j 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. To illustrate, let an object again be a purchase, and let attributes correspond to different products. Then, $A_i(x_j) = 1$ means that the i -th product is contained in the purchase x_j .

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

$$\mathcal{A}(x_j) = 1 \Leftrightarrow \forall A_i \in \mathcal{A} : A_i(x_j) = 1, \quad (1)$$

i.e., \mathcal{A} is considered as a logical conjunction of individual attributes. An association rule is a directed association $\mathcal{A} \rightarrow \mathcal{B}$ with $\mathcal{A}, \mathcal{B} \subseteq \mathcal{R}$ and $\mathcal{A} \cap \mathcal{B} = \emptyset$. The intended meaning of such a rule is that $\mathcal{A}(x) = 1$ usually entails $\mathcal{B}(x) = 1$.

In order to find “interesting” association rules in a database (binary relation), a potential rule $\mathcal{A} \rightarrow \mathcal{B}$ 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, where $\mathcal{D}_{\mathcal{A}}$ denotes the set of objects $x \in \mathcal{D}$ such that $\mathcal{A}(x) = 1$: A measure of *support* defines the number of objects that satisfy both \mathcal{A} and \mathcal{B} , i.e.,

$$\text{supp}(\mathcal{A} \rightarrow \mathcal{B}) = \text{supp}(\mathcal{A} \cup \mathcal{B}) \stackrel{\text{df}}{=} |\mathcal{D}_{\mathcal{A} \cup \mathcal{B}}|, \quad (2)$$

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 $\text{Pr}(\mathcal{B} \mid \mathcal{A})$ that the consequent is true given that

the antecedent holds:

$$\text{conf}(\mathcal{A} \rightarrow \mathcal{B}) \stackrel{\text{df}}{=} \frac{|\mathcal{D}_{\mathcal{A} \cup \mathcal{B}}|}{|\mathcal{D}_{\mathcal{A}}|}. \quad (3)$$

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

$$\text{sign}(\mathcal{A} \rightarrow \mathcal{B}) \stackrel{\text{df}}{=} \frac{|\mathcal{D}_{\mathcal{A} \cup \mathcal{B}}|}{|\mathcal{D}_{\mathcal{A}}|} - \frac{|\mathcal{D}_{\mathcal{B}}|}{|\mathcal{D}|},$$

suggesting that $\mathcal{A} \rightarrow \mathcal{B}$ is interesting only if the occurrence of \mathcal{A} does indeed have a positive influence on the occurrence of \mathcal{B} (i.e. $\text{Pr}(\mathcal{B} \mid \mathcal{A}) > \text{Pr}(\mathcal{B})$).

As can be seen, the support measure plays a central role. In fact, other measures can generally be derived from the support. For example, the confidence of an association $\mathcal{A} \rightarrow \mathcal{B}$ 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 A_i whose domain $\text{dom}(A_i)$ is a subset of the real numbers rather than $\{0, 1\}$. One possibility to deal with a numerical attribute A_i is to replace it by several binary attributes $A_i^1, A_i^2 \dots A_i^k$ [29, 34]. Each A_i^j is identified with the property that $A_i(x)$ lies in a subset of $\text{dom}(A_i)$, 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 $\text{dom}(A_i)$), i.e., we shall employ the same symbol for both of them.

The following property should of course be satisfied:

$$\text{dom}(A_i) \subseteq \bigcup_{j=1}^k A_i^j,$$

i.e., the binary properties should cover the domain of the original attribute A_i . Note that the A_i^j 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

$$\text{IF } A_i(x) \in A_i^j \text{ THEN } A_p(x) \in A_p^q, \quad (4)$$

which can be written equivalently as $A_i^j(x) \rightarrow A_p^q(x)$. Note that (4) does make sense only if the attributes in the antecedent and consequent are different ($i \neq p$), i.e., the binary attributes A_i^j 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 clustering techniques can be applied which create intervals and rules at the same time.

3. Fuzzy Association Rules

Fuzzy sets formalize the idea of *graded membership*, i.e. the idea that an element belongs “more or less” to a set. 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 [36]. 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, $F(x)$ denotes the degree of membership of the element u in the fuzzy set F .

In the context of association rule mining, the fuzzy approach allows for a “soft” partitioning of numerical attributes, that is, the partitioning of a numerical domain into fuzzy sets (intervals) rather than ordinary sets (intervals). Thus, the domain $\text{dom}(A_j)$ of the numerical attribute is described by means of a set $F_j^1 \dots F_j^n$ of fuzzy properties, and each of them can be viewed as a $[0, 1]$ -valued attribute A_j^i of objects such that $A_j^i(x) = F_j^i(A_j(x))$. 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 $\mathcal{A} \rightarrow \mathcal{B}$, where \mathcal{A} and \mathcal{B} are, respectively, sets of fuzzy attributes A_i and B_j . To illustrate, consider a rule suggesting that experienced managers have high income:

$$\{\text{manager, experienced}\} \rightarrow \{\text{high_income}\}$$

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 $\text{dom}(A_i)$ (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_i can be combined into one multi-dimensional attribute \mathcal{A} by means of a t-norm \otimes [24, 33]:

$$\mathcal{A}(x) \stackrel{\text{df}}{=} \bigotimes_{A_i \in \mathcal{A}} A_i(x).$$

This is a direct extension of definition (1) to the fuzzy case. A 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 $0 \leq \alpha \leq 1$. 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 $(\alpha, \beta) \mapsto \max(\alpha + \beta - 1, 0)$.

3.1. Quality Measures for Fuzzy Associations

Several generalizations of the quality measures (2) and (3) have been proposed in literature [14]. The standard approach is to replace set-theoretic operations, namely Cartesian product and cardinality, by corresponding fuzzy set-theoretic operations (e.g. [3]). Modeling the Cartesian product by a t-norm and the cardinality of a fuzzy set by the sum of the values of its membership function (the so-called sigma-count) [9], this gives

$$\text{supp}(\mathcal{A} \rightarrow \mathcal{B}) \stackrel{\text{df}}{=} \sum_{x \in \mathcal{D}} \mathcal{A}(x) \otimes \mathcal{B}(x), \quad (5)$$

$$\text{conf}(\mathcal{A} \rightarrow \mathcal{B}) \stackrel{\text{df}}{=} \frac{\sum_{x \in \mathcal{D}} \mathcal{A}(x) \otimes \mathcal{B}(x)}{\sum_{x \in \mathcal{D}} \mathcal{A}(x)}. \quad (6)$$

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

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

$$\text{supp}_x(\mathcal{A} \rightarrow \mathcal{B}) \stackrel{\text{df}}{=} \mathcal{A}(x) \otimes \mathcal{B}(x). \quad (7)$$

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

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 the corresponding conditional probability. This observation [27] has led to a vivid debate in philosophical circles since the late seventies [19] 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. [11]). In fact, an (association) rule $\mathcal{A} \rightarrow \mathcal{B}$ 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:

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

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 [6] in 1936. It is also the basis of DE FINETTI’s approach to conditional probability. Indeed, it is obvious that the probability $\Pr(\mathcal{B} \mid \mathcal{A})$ is entirely defined by $\Pr(\mathcal{A} \cap \mathcal{B})$ and $\Pr(\mathcal{A} \cap \bar{\mathcal{B}})$. (As an aside, note that *events* should formally be distinguished from attributes and sets of attributes. An event can again be associated with a subset \mathcal{A} of the attribute set \mathcal{R} , namely those A_i with $A_i(x) = 1$ for a given x . However, the logical conjunction of two events \mathcal{A} and \mathcal{B} , expressing that $A_i(x) = 1$ for the attributes in \mathcal{A} and the attributes in \mathcal{B} , corresponds to a set-theoretical *intersection* $\mathcal{A} \cap \mathcal{B}$. Thus, an intersection of events has the same meaning as the *union* $\mathcal{A} \cup \mathcal{B}$ of the corresponding attribute sets in the context of association rules, for example in rules such as $\mathcal{A} \cup \mathcal{B} \rightarrow \mathcal{C}$.)

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 $(\mathcal{A} \cap \mathcal{B}, \mathcal{A} \cap \bar{\mathcal{B}})$. This definition has several consequences. First, it justifies the claim made by DE FINETTI that a conditional probability $\Pr(\mathcal{A} \cap \mathcal{B})$ is the probability of a particular entity which can be called a *conditional event*, denoted $\mathcal{B} \mid \mathcal{A}$. 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, $\{x \mid \neg\mathcal{A}(x) \vee \mathcal{B}(x)\}$, 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 agree-

ment 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 $\{x \mid \mathcal{A}(x) \wedge \mathcal{B}(x)\}$ 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.

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:

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

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:

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

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

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

where \rightsquigarrow is a generalized implication operator, i.e., 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.)

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

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

for all potential objects x [32]. This leads us to the **admissible operators 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 three parts: strict preference, indifference, and incompatibility [15].

Before investigating this problem in more detail, let us anticipate the reproach 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 generalized logical disjunction (t-conorm), i.e., an operator $[0, 1] \times [0, 1] \rightarrow [0, 1]$ which is associative, commutative, monotone increasing (in both places) and such that $\alpha \oplus 1 = 1$ and $\alpha \oplus 0 = \alpha$ for all $0 \leq \alpha \leq 1$ [24, 33]. In fact, 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 fuzzy 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 (such as e.g. the product). 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.

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

$$\alpha \rightsquigarrow \beta = (1 - \alpha) + (\alpha \otimes \beta) \quad (12)$$

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 [2] 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

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

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

Here are some examples of standard conjunctions \otimes together with induced implications:

\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: $\max(1 - \alpha, \beta)$

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, a simple counterexample is the (weakly) drastic product [24], defined by $\alpha \otimes \beta = \min(\alpha, \beta)$ if $\max(\alpha, \beta) = 1$ and 0 otherwise, for which (12) becomes

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

Besides, there are even continuous t-norms that violate the above monotonicity condition. For instance, consider the HAMACHER family [18] of t-norms:

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

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. [25] for definitions and references).

Note that the monotonicity condition

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

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

$$(\alpha \leq \alpha') \Rightarrow (\alpha' \otimes \beta) - (\alpha \otimes \beta) \leq \alpha' - \alpha. \quad (14)$$

As a consequence, we find that a t-norm \otimes is admissible in (12) if it is a so-called *copula* [30, 33]. In fact, the following result is stated as a theorem in [33]: 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 generator¹ is convex. For many parameterized families of t-norms [24], 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 [33] 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 [16], 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 .

A further interesting result concerns the possibility of combining admissible t-norms into new admissible t-norms (see [30]).

Proposition 3: The ordinal sum² of copulas is again a copula.

Since each element of the family

$$\otimes_\gamma : (\alpha, \beta) \mapsto \frac{\alpha\beta}{\max(\alpha, \beta, \gamma)}, \quad 0 < \gamma \leq 1 \quad (15)$$

of t-norms, introduced by DUBOIS and PRADE [8], is an ordinal sum of the minimum and the product, we obtain that each t-norm (15) is admissible in the sense of (14).

¹ A definition of an additive generator can be found in the appendix.

² A definition of an ordinal sum is given in the appendix.

4.3. Two Special Cases

In this section, we shall consider two special cases (refinements) of the admissible operators problem that will further reduce the class of solutions. In fact, in both cases a unique solution will be obtained.

4.3.1. Fully Confident Rules 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 $\mathcal{A} \rightarrow \mathcal{A}$ with identical premise and conclusion part.³ 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 $\text{conf}(\mathcal{A} \rightarrow \mathcal{A}) < 1$. As this might strike as odd, one might think of postulating $\text{conf}(\mathcal{A} \rightarrow \mathcal{A}) = 1$ as an axiom. More generally, it seems sensible to require the following property to hold:

$$(\forall x \in \mathcal{D} : \mathcal{A}(x) \leq \mathcal{B}(x)) \Rightarrow (\text{conf}(\mathcal{A} \rightarrow \mathcal{B}) = 1). \quad (16)$$

Now, it is obvious that this property holds iff

$$\mathcal{A}(x) \leq \mathcal{B}(x) \Rightarrow \mathcal{S}_-(x) = 0 \quad (17)$$

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 $\otimes = \min$.

4.3.2. Strong Implication Operators A second specialization of the admissible operators problem is obtained by assuming a particular type of implication in (10), namely a strong implication operator. The latter is of the form $\alpha \rightsquigarrow \beta \stackrel{\text{df}}{=} n(\alpha) \oplus \beta$, 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

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

³ 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.

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):

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

The admissible operators 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

$$(\alpha \otimes \beta) + \alpha \otimes (1 - \beta) \equiv \alpha. \quad (19)$$

Interestingly enough, from ALSINA's results in [1] 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

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

4.4. 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 [12]. 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:

$$\alpha \rightsquigarrow \beta \stackrel{\text{df}}{=} \sup\{\gamma \mid \alpha \otimes \gamma \leq \beta\}. \quad (20)$$

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” [31, 10], for example “The larger an object, the heavier it is”.

So-called *pure* gradual rules are obtained when using the implication operator⁴

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

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

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

⁴ 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.

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 $\mathcal{A}(x) \wedge \mathcal{B}(x)$.⁵ 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 of a rule $\mathcal{A} \rightarrow \mathcal{B}$.

As proposed in [21], 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, $\text{Rel}_{\mathcal{A},\mathcal{B}}(x)$, thereby expressing that x supports $\mathcal{A} \rightarrow \mathcal{B}$ 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 it satisfies the condition part (non-triviality).

This approach suggests a support measure of the following kind:

$$\text{supp}_x(\mathcal{A} \rightarrow \mathcal{B}) = \text{Rel}_{\mathcal{A},\mathcal{B}}(x) \otimes (\mathcal{A}(x) \rightsquigarrow \mathcal{B}(x)).$$

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

	$\mathcal{A}(x) = 0$	$0 < \mathcal{A}(x) \leq \mathcal{B}(x)$	$\mathcal{A}(x) > \mathcal{B}(x)$
$\mathcal{S}_+(x)$	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 $\text{Rel}_{\mathcal{A},\mathcal{B}}(x) = \mathcal{A}(x)$.

⁵ This conjunctive approach to modeling fuzzy rules has first been proposed by MAMDANI and ASSILIAN [28] in the context of fuzzy control and is now widely known as Mamdani rules.

This approach suggests the following specification of a fuzzy partition:

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

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 $\{(\alpha, \alpha) \mid 0 \leq \alpha \leq 1\}$): 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

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

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.⁶ 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 yet another result that emphasizes the particular role of the support measure (5) with $\otimes = \min$.

4.5. Summary of Results

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

1. $\text{supp}_x(\mathcal{A} \rightarrow \mathcal{B}) = \min\{\mathcal{A}(x), \mathcal{B}(x)\}$
2. $\text{supp}_x(\mathcal{A} \rightarrow \mathcal{B}) = \mathcal{A}(x) \cdot \mathcal{B}(x)$
3. $\text{supp}_x(\mathcal{A} \rightarrow \mathcal{B}) = \mathcal{A}(x) \cdot (\mathcal{A}(x) \rightsquigarrow \mathcal{B}(x))$

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

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

5. 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

⁶ Using the definition (20) of a residuated implication, this equality is easy to prove.

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. 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 application, especially in connection with implication-based fuzzy rules (gradual rules). 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.

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 operator (cf. Section 4.4) has been proposed in [22].

Before concluding the paper, let us point to a problem that concerns the derivation of frequency information in the fuzzy case where attributes can apply to statistical entities in a partial way. To exemplify the problem, consider the two 10×10 arrays shown in Fig. 1. Both 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 $\text{supp}(\text{light}) = 75$ in this case. As opposed to this, the property `light` does not seem to be well supported in the second case (right), where actually none of the fields is really light. Still, if $\text{supp}(\text{light})$ is computed by the sum of membership degrees $\mathcal{S}_+(x)$, the support is again 75.

This counter-intuitive result is due to the fact that, in the fuzzy case, several small membership degrees can compensate for a few large degrees [14]. Admittedly, the second

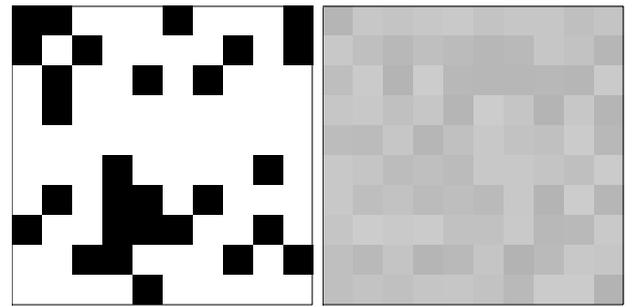


Figure 1. Objects (statistical entities) are the fields of a 10×10 array. The property of interest is the attribute `light` which is in direct correspondence with the level of grey.

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* [4, 13]. 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

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

The idea of this measure, which is directly related to a measure of fuzziness introduced in [23], 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. In our example, the clarity degree (25) is 1 for the first case in our example whereas it is only $\approx .5$ for the second 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 in a topic of ongoing work.

References

- [1] C. Alsina. On a family of connectives for fuzzy sets. *Fuzzy Sets and Systems*, 16:231–235, 1985.
- [2] C. Alsina, E. Trillas, and L. Valverde. On some logical connective for fuzzy sets theory. *J. Math. Anal. Appl.*
- [3] W.-H. Au and K. Chan. An effective algorithm for discovering fuzzy rules in relational databases. In *Proceedings IEEE World Congress on Computational Intelligence*, pages 1314–1319, 1998.
- [4] P. Bosc, D. Dubois, O. Pivert, and H. Prade. On fuzzy association rules based on fuzzy cardinalities. In *Proceedings FUZZ-IEEE-2001, 10th IEEE International Conference on Fuzzy Systems*, Melbourne, Australia, 2001.
- [5] 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.
- [6] B. DeFinetti. La logique des probabilités. *Int. Cong. de Philosophie Scientifique*, 5:1–9, 1936.
- [7] M. Delgado, N. Marin, D. Sanchez, and M. Vila. Fuzzy association rules: general model and applications. *IEEE Transactions on Fuzzy Systems*, 11(2):214–225, 2003.
- [8] D. Dubois and H. Prade. New results about properties and semantics of fuzzy-set-theoretic operators. In P. Wang and S. Chang, editors, *Fuzzy Sets: Theory and Applications to Policy Analysis and Information Systems*, pages 59–75. Plenum Publ., 1980.
- [9] D. Dubois and H. Prade. Fuzzy cardinality and the modeling of imprecise quantification. *Fuzzy Sets and Systems*, 16:199–230, 1985.
- [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. Conditional objects as non-monotonic consequence relationships. *IEEE Transactions on Systems, Man and Cybernetics*, 24(12):1724–1739, 1994.
- [12] D. Dubois and H. Prade. What are fuzzy rules and how to use them. *Fuzzy Sets and Systems*, 84:169–185, 1996.
- [13] 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.
- [14] 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, Istanbul, 2003.
- [15] J. Fodor and M. Roubens. *Fuzzy Preference Modelling and Multicriteria Decision Support*. Kluwer, 1994.
- [16] M. Frank. On the simultaneous associativity of $f(x, y)$ and $x + y - f(x, y)$. *Aeq. Math.*, 19:194–226, 1979.
- [17] I. Goodman, H. Nguyen, and E. Walker. *Conditional Inference and Logic for Intelligent Systems*. North-Holland, Amsterdam, 1991.
- [18] H. Hamacher. *Über logische Aggregationen nichtbinär explizierter Entscheidungskriterien; Ein axiomatischer Beitrag zur normativen Entscheidungstheorie*. R.G. Fischer Verlag, 1978.
- [19] W. Harper, R. Stalnaker, and G. Pearce, editors. *IFS*. D. Reidel, Dordrecht, The Netherlands, 1981.
- [20] E. Hüllermeier. Fuzzy association rules: Semantic issues and quality measures. In B. Reusch, editor, *Proceedings of the International Conference on Computational Intelligence – 7th Fuzzy Days*, number 2206 in LNCS, pages 380–391, Dortmund, Germany, 2001. Springer-Verlag.
- [21] E. Hüllermeier. Implication-based fuzzy association rules. In L. D. Raedt and A. Siebes, editors, *Proceedings PKDD-01, 5th European Conference on Principles and Practice of Knowledge Discovery in Databases*, number 2168 in LNAI, pages 241–252, Freiburg, Germany, September 2001. Springer-Verlag.
- [22] E. Hüllermeier and J. Beringer. Mining implication-based fuzzy association rules in databases. In B. Bouchon-Meunier, L. Foulloy, and R. Yager, editors, *Intelligent Systems for Information Processing: From Representation to Applications*. Elsevier, 2003.
- [23] A. Kaufmann. *Introduction to the theory of fuzzy subsets*. Academic Press, New York, 1975.
- [24] E. Klement, R. Mesiar, and E. Pap. *Triangular Norms*. Kluwer Academic Publishers, 2002.
- [25] G. Klir and T. Folger. *Fuzzy Sets, Uncertainty, and Information*. Prentice Hall, 1988.
- [26] S. Kraus, D. Lehmann, and M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44:167–207, 1990.
- [27] D. Lewis. Probabilities of conditionals and conditional probabilities. *Journal of Philosophical Logic*, 3, 1973.
- [28] 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.
- [29] R. 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.
- [30] R. Nelsen. *An Introduction to Copulas*. Springer-Verlag, 1999.
- [31] 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.
- [32] E. Ruspini. A new approach to clustering. *Information Control*, 15:22–32, 1969.
- [33] B. Schweizer and A. Sklar. *Probabilistic Metric Spaces*. North-Holland, New York, 1983.
- [34] 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.
- [35] T. Sudkamp. Examples, counterexamples, and measuring fuzzy associations. *Fuzzy Sets and Systems*. To appear.
- [36] L. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.