On the Representation, Measurement, and Discovery of Fuzzy Associations

Didier Dubois, *Member, IEEE*, Henri Prade, and Thomas Sudkamp, *Senior Member, IEEE*

***Abstract*—The use of fuzzy sets to describe associations between data extends the types of relationships that may be represented, facilitates the interpretation of rules in linguistic terms, and avoids unnatural boundaries in the partitioning of the attribute domains. In addition, the partial membership values provide a method for incorporating the distribution of the data into the assessment of a rule. This paper investigates techniques to identify and evaluate associations in a relational database that are expressible by fuzzy if-then rules. Extensions of the classical confidence measure based on the α-cut decompositions of the fuzzy sets are proposed to incorporate the distribution of the data into the assessment of a relationship and identify robustness in an association. A rule learning strategy that discovers both the presence and the type of an association is presented.**

***Index Terms*—Data mining, fuzzy association rules, generalized implication, rule learning.**

I. Introduction

THE proliferation of large databases provides both the impetus and the need for the development of algorithmic techniques for the identification and evaluation of relationships among data. This paper considers two distinct, but closely related issues: The measurement of the degree to which data satisfy a relationship and the discovery of relationships among the data in a relational database. Data associations will be described by fuzzy rules, which extend the representational capabilities of classical association rules, facilitate the construction and interpretation of rules in natural linguistic terms, and avoid unnatural boundaries in the partitioning of the attribute domains.

A classical (crisp) association between properties A and B is frequently represented in the form of a rule $A \Rightarrow B$ indicating that an element satisfying property A also satisfies B. The standard pair of indices used to measure the validity of an association rule are the support and the confidence

$$\mathrm{Supp}(A \Rightarrow B) = \frac{|A \cap B|}{|U|} \quad \mathrm{Con}(A \Rightarrow B) = \frac{|A \cap B|}{|A|} \qquad (1)$$

where $|\,|$ denotes the cardinality and U is the set of tuples in the database [1], [2]. The support measures the extent of the simultaneous occurrence of properties A and B in elements of the database while the confidence indicates the likelihood of an element with property A also having property B.

Manuscript received June 25, 2003; revised June 1, 2004.

D. Dubois and H. Prade are with the Institut de Recherche en Informatique de Toulouse (IRIT), Université Paul Sabatier, 31062 Toulouse Cedex 4, France (e-mail: Didier.Dubois@irit.fr; Henri.Prade@irit.fr).

T. Sudkamp is with the Department of Computer Science, Wright State University, Dayton, OH 45435 USA (e-mail: tsudkamp@cs.wright.edu).

Digital Object Identifier 10.1109/TFUZZ.2004.840130

The additional representational capability afforded by fuzzy rules introduces several difficulties into the measurement of the satisfaction of a rule. These include the selection of appropriate generalizations for the support and confidence, the effect of the accumulation of small cardinalities, and the robustness of the support for an association. The partial membership values in the boundary regions of the fuzzy sets provide a method for incorporating the distribution of the data into the assessment of a rule, an issue that is frequently overlooked in the evaluation of quantitative or interval-based association rules [15], [27], [30]–[32]. The objective of this paper is to provide a unified approach to the evaluation and discovery of association rules that will be applicable to crisp as well as fuzzy associations. We begin in Section II by considering association rules in terms of multiple-valued implications, which provides a framework for comparing crisp rules, certainty rules, and gradual rules [12], the latter two being common types of relationships representable by fuzzy associations.

There are two standard methods for extending crisp measures to fuzzy sets and fuzzy rules. The first, and perhaps simplest, is to directly replace the operators in the crisp measure with appropriate fuzzy counterparts. An alternative approach is to represent a fuzzy rule as a set of crisp rules. The confidence and support for the fuzzy rule are then obtained by applying standard techniques to the associated set of crisp rules and aggregating the results. Section III compares generalizations of support and confidence obtained following these strategies. The ability of the resulting measures to discriminate between different types and distributions of data is examined in Section IV.

The generalizations of the confidence measure considered in Section III produce a scalar value by assessing either individual data elements or α-cuts and aggregating the results. Section V reverses the order of the process; an initial step summarizes the database in terms of the membership values of the tuples. The summarization preserves the information necessary for computing support and confidence measures and provides the ability to evaluate associations based on the degree of relevance and the distribution of the tuples. A combination of confidence and robustness, which avoids the anomalies presented in Section III, is proposed as the criteria for supporting an association. The paper concludes with the presentation of a rule learning algorithm that discovers both the presence and the type of an association.

In addition to support and confidence, other measures have been proposed to assess the validity of a crisp association. These are frequently based upon the independence or correlation of attributes [6], [7], [23]. The properties of fuzzy associations addressed in this paper, the accumulation of small cardinalities, robustness of support, and the distribution of examples, would be equally relevant to extensions of these measures to fuzzy sets.

II. Associations and Implication Rules

The generalization from classical to fuzzy association rules provides the ability to represent uncertainty and synergistic relationships between attributes. In this section, we review the semantics of the representation of association rules in terms of multiple-valued implication, a representation that encompasses both standard and fuzzy associations. We begin by introducing the notation that will be used throughout this paper.

The tuples of a relational database R are defined by attributes $\mathcal{A}, \mathcal{B}, \mathcal{C}, \ldots$ with domains $D_{\mathcal{A}} = \{a_1, \ldots, a_{|D_{\mathcal{A}}|}\}$, $D_{\mathcal{B}} = \{b_1, \ldots b_{|D_{\mathcal{B}}|}\}$, $D_{\mathcal{C}} = \{c_1, \ldots c_{|D_{\mathcal{C}}|}\}, \ldots$, respectively. For the purposes of this paper, it is sufficient to consider tuples consisting of two attributes $\mathcal{A}$ and $\mathcal{B}$. The tuples will be denoted $t_k = (a_{i_k}, b_{j_k})$, $k = 1, \ldots, |R|$, where a_{i_k} and b_{j_k} are the elements from domain $D_{\mathcal{A}}$ and $D_{\mathcal{B}}$ in tuple k. When the index is immaterial, a tuple will be written $t = (a, b)$.

A subset A of $D_{\mathcal{A}}$ defines a subset $R(A)$ over R; a tuple (a, b) is in $R(A)$ if, and only if, $a \in A$. When A is fuzzy, $R(A)$ is a fuzzy set over the tuples of R. The membership value of the tuple $t = (a, b)$ in $R(A)$ is $A(a)$, the degree of membership of the attribute value a in the fuzzy set A.

Using the preceding notation, the support and confidence for an association $A \Rightarrow B$ given by tuples of a relational database may be written as

$$\begin{aligned} \operatorname{supp}(A \Rightarrow B) &= |R(A) \cap R(B)| \\ \operatorname{con}(A \Rightarrow B) &= \frac{|R(A) \cap R(B)|}{|R(A)|}. \end{aligned} \tag{2}$$

Since our analysis will focus on the number of examples of an association, the normalizing term $|U|$ is omitted from the definition of support given in (1). For crisp sets, $\operatorname{supp}(A \Rightarrow B)$ is the number of tuples that support (or are examples of) the rule $A \Rightarrow B$. When A and B are fuzzy, intersection is evaluated by a T-norm [21], [26], [29]. The identification of membership in $A \cap B$ with being an example of an association $A \Rightarrow B$ yields $T(A(a), B(b))$ as the degree to which a tuple (a, b) is an example of the association. Consequently, the selection of a T-norm determines the size of the set of examples and, in turn, the confidence.

For reasons of both efficiency and the robustness of the analysis, we restrict the possible membership values of tuples in A and B to a finite scale. The scale is obtained by selecting a small set of values that serve as representative membership values for the fuzzy sets. We will let $\mathcal{L}_A = \{0 = \sigma_1 < \sigma_2 < \cdots < \sigma_r = 1\}$ and $\mathcal{L}_B = \{0 = \gamma_1 < \gamma_2 < \cdots < \gamma_s = 1\}$ be the representative values for A and B, respectively.

An association rule $A \Rightarrow B$ is generally interpreted as an implication "if X is A, then Y is B." For example, the rule "if *Age* is *Young*, then *Salary* is *Low*" associates a set of ages with a range of salaries. These sets may be defined by crisp sets, but more natural interpretations of "young" and "low" would suggest a fuzzy interpretation.

In either case, the underlying relationship can be described in terms of a multiple-valued implication. An implication operator I on [0, 1] × [0, 1] is a generalization of classical material implication that satisfies

I1) $I(x, y) \le I(x', y)$ for $x' \le x$;

I2) $I(x, y) \le I(x, y')$ for $y \le y'$;

I3) $I(1, y) = 1$ if and only if $y = 1$;

I4) $I(1, 0) = 0$, $I(0, 0) = 1$;

for $x, y \in [0, 1]$. That is, I is nonincreasing in the first variable and nondecreasing in the second. The boundary conditions together with I1) ensure that I reduces to material implication when restricted to {0, 1}. Implication is often required to satisfy additional constraints such as the exchange property $I(x, I(y, z)) = I(y, I(x, z))$, or neutrality $I(1, y) = y$, but the less restrictive set of axioms suffices for our purposes. A comprehensive survey of axioms for fuzzy implication can be found in [14].

An implication operator specifies the type of the relation between the attributes and its values may be used in the determination of the degree that tuples are examples of an association. The classical binary equivalence

$$R(A) \cap R(B) \equiv R(A) \cap (\overline{R(A)} \cup R(B)) \tag{3}$$

explicitly indicates the role of implication in the determination of examples. Evaluating $\cap$ with a T-norm, the right-hand side of (3) produces the value

$$T(A(a), I(A(a), B(b))) \tag{4}$$

for a tuple (a, b). The term $A(a)$ in the first argument of the T-norm ensures that irrelevant tuples are excluded from the set of examples. By I3), a tuple that completely satisfies the antecedent is considered to be an example of the rule to degree 1 only when the consequent is also completely satisfied. With fuzzy association rules represented as implications, the determination of the examples and the confidence is dependent upon the selection of the T-norm and the implication operator.

When the domain is restricted to the finite scales $\mathcal{L}_A$ and $\mathcal{L}_B$, an implication operator defines an implication relation over $\mathcal{L}_A \times \mathcal{L}_B$ that may be represented by a $r \times s$ matrix $\mathbf{I}$ whose i, jth entry is

$$\mathbf{I}(i, j) = I(\sigma_i, \gamma_j).$$

For a crisp association rule, 0 and 1 are the only α-levels needed in the assessment of the rule. The requirements for an implication operator produce the matrix

$$\mathbf{I}(i, j) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

The values $\mathbf{I}(0, 0)$ and $\mathbf{I}(0, 1)$ are the implication degrees for tuples that are irrelevant to the rule and consequently have no effect on the measurement of the confidence and support.

In [12], Dubois and Prade identified two types of fuzzy rules, certainty rules and gradual rules, that are distinguished intuitively by their semantics and formally by the type of implication operator that defines the relation. Certainty rules have the interpretation "the more X is A, the more certainly Y is B" and provide a natural extension of crisp association rules $A \Rightarrow B$.

Examples of implications for certainty rules include Dienes implication, $I_D(x,y) = \max\{1-x, y\}$, and Reichenbach implication, $I_R(x,y) = 1 - x + xy$, which produce the matrix

$$\begin{bmatrix} 1 & 1 & \dots & 1 & 1 \\ & & & & 1 \\ \uparrow & & \nearrow & & \vdots \\ & & & & 1 \\ 0 & & \rightarrow & & 1 \end{bmatrix}$$

where the arrows indicate that the values of the implication are nondecreasing in the designated direction. Moreover, the only entries with value 1 are those indicated in the matrix.

The semantics of certainty rules are illustrated by the relationship "the nearer to noon, the more certain the shop will be closed." The vague concept in the antecedent "nearer to noon" is well described by a fuzzy set over the domain of times. The consequent is a binary valued attribute, either the shop is open or not. The use of Dienes or Reichenbach implication indicates that a relevant tuple (a,b) completely supports the rule only if $A(a) = 1$ and $B(b) = 1$, in which case the consequent is certain.

The interpretation of gradual rules is "the more X is A, the more Y is B." Gradual rules indicate a synergistic relationship between the values of the two attributes and are based on a residuated or R-implication. R-implications are obtained from a T-norm by $I(x,y) = \sup\{z \in [0,1] \mid T(x,z) \le y\}$ and include Goguen implication

$$I_{Gn}(x,y) = \begin{cases} \min\left\{1, \frac{y}{x}\right\}, & \text{if } x > 0 \\ 1, & \text{otherwise} \end{cases}$$

generated using the product T-norm, Gödel implication

$$I_{Gl}(x,y) = \begin{cases} 1, & \text{if } x \le y \\ y, & \text{otherwise} \end{cases}$$

generated using the minimum, and Lukasiewicz implication, $I_L(x,y) = \min\{1, 1-x+y\}$ generated using $T(x,y) = \max\{0, x+y-1\}$.

The matrix of an R-implication for a continuous T-norm and $\sigma_i = \gamma_i$ has the form

$$\begin{bmatrix} 1 & 1 & 1 & \dots & 1 & 1 & \\ & 1 & 1 & \dots & 1 & 1 & \\ \uparrow & & & \ddots & \vdots & \vdots & \\ & <1 & & & 1 & 1 & \\ 0 & & \rightarrow & & & 1 & \end{bmatrix}.$$

The values below the diagonal are all less than 1 and nondecreasing in the directions of the arrows (see [21] for the proofs of these properties).

The rule "the closer the time is to 4:00 pm, the hotter the temperature" illustrates the relationship between the membership values in a gradual rule. The relationship would be formalized as a gradual rule using fuzzy sets to describe the terms "near 4:00 pm" and "hot." The rule specifies an increasingly tight set of constraints on the value of the consequent; a tuple (a,b) with $A(a) = x$ completely satisfies the restriction if $B(b) \ge x$. As $A(a)$ increases, so does the constraint on $B(b)$.

Pure gradual rules are special class of implication rules in which the relationship between the fuzzy sets A and B is specified by a nondecreasing function between the α-cuts of A and B. When restricted to finite scales $\mathcal{L}_A$ and $\mathcal{L}_B$, the implication relation defined by a nondecreasing function m is

$$\mathbf{I}_m(\sigma_i, \gamma_j) = \begin{cases} 1, & \text{if } m(\sigma_i) \le \gamma_j \\ 0, & \text{otherwise.} \end{cases}$$

The pure gradual rule relating $\mathcal{L}_A$ to $\mathcal{L}_B$ defined by m will be denoted $m \circ A \mapsto B$.

Each pure gradual rule $m \circ A \mapsto B$ is uniquely characterized by its focal set $\Lambda_m = \{(\sigma_i, \gamma_j) \mid m(\sigma_i) \le \gamma_j\}$. A family of pure gradual rules $m_i \circ A \mapsto B$, $i = 1, \dots, t$, is said to be *consonant* if either $\Lambda_{m_i} \subset \Lambda_{m_j}$ or $\Lambda_{m_j} \subset \Lambda_{m_i}$, for every pair i, j. In [11], Dubois *et al.* established the following representation of a general implication rule by a set of pure gradual rules.

Theorem 1: Let $\mathbf{I}$ be an implication relation on $\mathcal{L}_A \times \mathcal{L}_B$. There is a unique set of consonant pure gradual rules $m_k \circ A \mapsto B, i = 1, \dots, t$, and values $p_k \in (0,1]$ with $\sum p_k = 1$ such that

$$\mathbf{I}(\sigma_i, \gamma_j) = \sum_{k \mid m_k(\sigma_i) \le \gamma_j} p_k$$

for all $i = 1, \dots, r$, $j = 1, \dots, s$.

Theorem 1 provides the ability to define the support and confidence of an arbitrary implication rule using the crisp constraints provided by the associated pure gradual rules. This technique will be employed in Section III to produce a scalar-valued measure for association rules.

III. Scalar-Valued Measures

The use of fuzzy sets to describe ranges of attributes and implicative rules to express associations of data increases the types of relationships that are expressible. However, these generalizations complicate the task of determining whether data supports an association. In this section, we consider several techniques that have been proposed for measuring the support and confidence of a fuzzy association. The presentation will focus on determining the number of examples of an association in a database and exhibiting the effect of the selection of the operations on the discriminability of the resulting measure.

A common approach to extending measures from crisp sets to fuzzy sets is to replace the crisp operations in the measure with their fuzzy counterparts [8], [19], [24]. Following this strategy, support and confidence for fuzzy associations can be obtained by replacing intersection with a T-norm and cardinality with a scalar-cardinality in (1). The simplest extension of cardinality to fuzzy sets is the Σ-count, which is the sum of the membership values of the elements in a fuzzy set. The generalization from crisp sets to fuzzy sets using the Σ-count produces the confidence measure

$$\text{con}_T(A \Rightarrow B) = \frac{\sum_{k=1}^{|R|} T(A(a_{i_k}), B(b_{j_k}))}{\sum_{k=1}^{|R|} A(a_{i_k})} \tag{5}$$

where T is a T-norm and the sum is taken over the set of tuples in the database.

A consequence of employing the Σ-count is that the summation of a large number of tuples with small membership degrees makes the same contribution to the support and confidence as a small number of tuples with that are highly relevant. The accumulation of small cardinalities may produce confidence values that are contrary to an intuitive assessment of the data, as demonstrated in the following example from [25].

Example 1: In the table that follows, each row represents a set of like tuples; the first column gives the number of tuples of the type, columns two and three their membership values in A and B, and the result of the intersection using the T-norm $\min$ is in column four

Number	$A(a)$	$B(b)$	$\min(A(a), B(b))$
1	1	.01	.01
1	.01	1	.01
998	.01	.01	.01.

Measuring confidence with the Σ-count yields $\text{con}_{\min}(A \Rightarrow B) = .91$. Thus, a high measure of confidence in the association is produced, even though the only tuple that significantly satisfies A minimally satisfies B. The influence of small cardinalities on the confidence measure con_T is not limited to the T-norm $\min$. Changing the B values from .01 to 1 in final 999 tuples produces a like result regardless of the T-norm. □

Example 2: While minimum is the most common T-norm, its noncompensatory nature contributes to a loss of information when used for the intersection of fuzzy sets. This phenomena can easily be exhibited by considering tuples whose A and B values are identical. Although contrived for the example, the presence of tuples with identical pairs of attributes is not unusual since the tuples may be distinguished by other attributes

	Number	$A(a)$	$B(b)$	$\min(A(a), B(b))$
a)	3	.5	.5	.5
b)	3	.5	1	.5.

Tuples in rows a) and b) differ only in the value of the B attribute. This change does not affect the membership values of $R(A)$ or $R(A) \cap R(B)$. Consequently confidence measures using min intersection produce the same values for these sets of tuples, in this case $\text{con}_{\min} = 1$ for both a) and b). It is also clear that the confidence measures will be independent of the B attributes whenever their membership values satisfy $B(b_{j_k}) \geq A(a_{i_k})$ for all tuples t_k. □

The same type of information loss occurs with minimum when the membership values for the consequent variable are all less than the minimum antecedent membership. These examples indicate the importance of the selection of the T-norm in the measurement of support and confidence.

A. Implication-Based Measures

Since the intuitive relationship represented by an association $A \Rightarrow B$ is an IF–THEN condition, the presence of attribute A in an instance guarantees the presence of B, it may seem advantageous to base the measurement of such an association on the degree of satisfaction of an implication rather than a conjunction. This was the motivation for conviction measure proposed by Brin *et al.* [6] for crisp associations, which uses the equivalence of $A \Rightarrow B$ with $\neg(A \cap \neg B)$ and measures the degree of independence of A and $\overline{B}$. For an assessment of implication based on examples, the equivalence in (3) shows a way to determine the number of examples of $A \Rightarrow B$ using a T-norm and an implication operator I. For crisp sets, a tuple is an example if $T(A(a), I(A(a), B(b)) = 1$; the conjunction of the term $A(a)$, which is either 1 or 0, removes the irrelevant cases from consideration.

The fundamental issue for assessing the confidence of the rule $A \Rightarrow B$ when the sets A and B are fuzzy is to determine which properties of the crisp measure to preserve. The measure con_T chose to maintain the ratio of set cardinalities and used the Σ-count to incorporate intermediate membership values into the confidence assessment. The implication interpretation provides an alternative generalization of confidence to fuzzy associations. In this case, the notion of relevant instance serves as the basis for the extension to fuzzy sets. The following two measures differ in the degree to which a tuple is considered relevant to an association.

The first generalization considers a tuple (a, b) *fully* relevant to the association $A \Rightarrow B$ if $A(a) > 0$. With this interpretation, a tuple (a, b) is an example to degree $T(\delta(A(a)), I(A(a), B(b))$ where δ is the indicator function

$$\delta(x) = \begin{cases} 1, & \text{if } x > 0 \\ 0, & \text{otherwise.} \end{cases}$$

The confidence, which is the ratio of the examples to the relevant cases, becomes

$$\text{imp}_f(A \Rightarrow B) = \frac{\sum_{k=1}^{|R|} T(\delta(A(a_{i_k})), I(A(a_{i_k}), B(b_{j_k})))}{\sum_{k=1}^{|R|} \delta(A(a_{i_k}))}.$$

With this interpretation of relevance, the irrelevant cases remain the same as in the crisp case (i.e., $A(a) = 0$). The normalization factor also remains the number of positive instances.

An alternative interpretation of relevance considers a tuple (a, b) relevant only to the degree $A(a)$. In this case, both the numerator and the denominator are weighted by the membership degree of the antecedent of a relevant instance

$$\text{imp}_p(A \Rightarrow B) = \frac{\sum_{k=1}^{|R|} T(A(a_{i_k}), I(A(a_{i_k}), B(b_{j_k})))}{\sum_{k=1}^{|R|} A(a_{i_k})}. \quad (6)$$

The subscript p indicates that tuples are considered *partially* relevant based on the degree of satisfaction of the antecedent.

B. Reduction to Crisp Sets

As seen in Examples 1 and 2, the selection of the intersection operator and method of determination of the cardinality of fuzzy sets may have a significant effect on the generalization of crisp measures to fuzzy sets. A common approach used to avoid these complications is to transform the measurement of a property of fuzzy sets into that of a family of crisp sets. Such a transformation permits the use of the standard crisp set operators and avoids the subtle variations associated with the selection of a particular T-norm or implication operator. The tradeoff, however, is the need to aggregate the values produced by the crisp analyzes. In this section, we examine two methods used to transform the measurement of fuzzy associations to an assessment of crisp sets.

The α-cut decomposition of a fuzzy set is frequently used to reduce the measurement of a property of fuzzy sets into that of a nested sequence of crisp sets. In [10], Delgado *et al.* introduced a scalar cardinality of fuzzy sets based on the weighted summation of the cardinalities of its α-cuts. The motivation behind the weighting was to accentuate the contribution of elements with high membership values. Martín-Bautista *et al.* [25] proposed the use of the weighted cardinality to mitigate the impact of small cardinalities in the assessment of an association.

The confidence measure using weighted cardinality is defined as

$$\text{con}_w(A \Rightarrow B) = \sum_{i=1}^{t-1} (\alpha_i - \alpha_{i+1}) \frac{|(R(A) \cap R(B))|_{\alpha_i}}{|R(A)|_{\alpha_i}} \quad (7)$$

where $1 = \alpha_1, \ldots, \alpha_t = 0$ is an ordered listing of the union of the α-levels of $R(A) \cap R(B)$ and $R(A)$. The summation over the α-cuts places a greater emphasis on the elements with higher membership values, since a tuple with membership α_k occurs in each of the summands $k, k+1, \ldots, t-1$. Variations of this measure have been presented [3], [9] and their performance compared on a database of information from the U.S. Census Bureau.

Another method for the evaluating association rules using a reduction to crisp sets, proposed by Hüllermeier [20], is based on the decomposition of a generalized implication into a sequence of pure gradual rules described in Theorem 1. The objective is to determine the degree to which tuples are examples of an implication I based on their satisfaction of the pure gradual rules associated with I.

Let I be an implication relation on $\mathcal{L}_A$ to $\mathcal{L}_B$ with the decomposition

$$m_i \circ A \to B, \qquad p_i > 0, \text{ for } i = 1, \ldots, q$$

into pure gradual rules and probabilities as specified by Theorem 1. That is, each m_i is a nondecreasing function from $\mathcal{L}_A$ to $\mathcal{L}_B$ and $\sum p_i = 1$. Let $\text{supp}(m_i \circ A \to B)$ denote the support for the pure gradual rule $m_i \circ A \to B$ given by the tuples in the database. The support of the association $A \Rightarrow B$ can then be obtained as the weighted sum of the support of each of the gradual rules in the decomposition

$$\text{supp}(A \Rightarrow B) = \sum_{i=1}^{q} p_i \cdot \text{supp}(m_i \circ A \to B).$$

What remains is to determine $\text{supp}(m_i \circ A \to B)$. A tuple (a, b) satisfies the constraint imposed by $m_i \circ A \to B$ if, and only if, $m_i(A(a)) \leq B(b)$. However, the determination of whether a tuple is an example of an implication also incorporates the relevance of the tuple to the rule. Hüllermeier [20] proposed two methods for incorporating the relevance into support of a pure gradual rule: full and partial relevance. In the measures that follow, the subscripts f and p indicate whether a tuple with $0 < A(a) < 1$ is considered either *fully* or *partially* relevant.

In the first method, the support for the rule $m_i \circ A \to B$ by the tuple (a, b) is

$$\text{supp}_f(m_i \circ A \to B)(a, b) = \begin{cases} 1, & \text{if } A(a) > 0 \text{ and } m_i(A(a)) < B(b) \\ 0, & \text{otherwise.} \end{cases}$$

Under this interpretation, a tuple is an example if it satisfies the constraint imposed by the gradual rule and the A attribute of the tuple has a nonzero membership degree. That is, a tuple is considered fully relevant whenever for $A(a) > 0$. The support for the rule $m_i \circ A \to B$ is the sum over all the tuples in R.

In the preceding approach, a tuple was deemed either completely relevant or irrelevant. The second method assigns intermediate degrees of relevance based on the membership of a in A

$$\text{supp}_p(m_i \circ A \to B)(a, b) = \begin{cases} A(a), & \text{if } m_i(A(a)) < B(b) \\ 0, & \text{otherwise.} \end{cases}$$

Thus, a tuple that satisfies the constraint is considered to completely support the rule when $A(a) = 1$ and it gives partial support to the rule when $A(a) < 1$. As before, the support for $m_i \circ A \to B$ is the sum of the support of each tuple in the database.

The confidence may be obtained directly from the definitions of support. For consistency, the same definition of support (interpretation of relevancy) should be used in both the terms of the confidence measure. Thus

$$\text{con}_f(A \Rightarrow B) = \frac{\text{supp}_f(A \Rightarrow B)}{\text{supp}_f(A \Rightarrow \mathcal{D}_{\mathcal{B}})}$$
$$\text{con}_p(A \Rightarrow B) = \frac{\text{supp}_p(A \Rightarrow B)}{\text{supp}_p(A \Rightarrow \mathcal{D}_{\mathcal{B}})} \quad (8)$$

where $\mathcal{D}_{\mathcal{B}}$ is the entire domain of the attribute $\mathcal{B}$.

The preceding measures may also be obtained directly from the evaluation of an association that determines the degree to which a tuple (a, b) is an example of an implication using $T(A(a), I(A(a), B(b)))$. When the product is used as the T-norm in imp_f and imp_p, $\text{imp}_f = \text{con}_f$ and $\text{imp}_p = \text{con}_p$ providing a justification for the implication measures based on the satisfaction of the crisp constraints provided by the corresponding pure gradual rules.

C. Properties of Scalar-Valued Measures

The preceding sections reviewed approaches that have been proposed for generalizing the support and confidence measures to fuzzy associations. We now examine the ability of these measures to avoid the problem of the accumulation of small cardinalities and to discriminate between data. In addition to $\text{con}_{\min}$, we will consider the following measures:

- con_w weighted α-cuts;
- imp_{f_D} full relevance and Dienes implication;
- imp_{p_D} partial relevance and Dienes implication;
- imp_{f_G} full relevance and Goguen implication;
- imp_{p_G} partial relevance and Goguen implication.

The product is used for the T-norm because of the resulting equalities $\mathrm{imp}_f = \mathrm{con}_f$ and $\mathrm{imp}_p = \mathrm{con}_p$ that relate pure gradual rules and the implication interpretation of confidence. Dienes and Goguen implications were chosen to demonstrate the difference in evaluation obtained by a certainty and a gradual interpretation of an association.

Example 3: The tuples from Example 1 are used to examine the effect of the accumulation of small cardinalities on the scalar-valued confidence measures introduced in the preceding sections. With the weighted α-cuts, the contribution of the 999 tuples with membership .01 is reduced by the weighting. The result, $\mathrm{con}_w(A \Rightarrow B) = .01$, is more in accord with an intuitive assessment of the validity of the rule.

For the fully relevant implication-based measures, all the tuples are completely relevant and all but one are considered examples (to either degree .99 or 1) producing confidence values of approximately 1. The measures that reduce the degree to which a tuple is an example by the relevance $A(a)$ yield $\mathrm{imp}_{p_D}(A \Rightarrow B) = .90$ and $\mathrm{imp}_{p_G}(A \Rightarrow B) \approx .91$. In these cases the number of examples of $A \Rightarrow B$ is small, approximately 10, but the number of examples of $A \Rightarrow \mathcal{D}_B$ is also small producing the high confidence. □

Example 4: The tuples in Example 2 illustrated the inability of $\mathrm{con}_{\min}$ to discriminate between intuitively different sorts of data. Examining these tuples with the other scalar measures produces

	$\mathrm{con}_{\min}$	con_w	imp_{f_D}	imp_{p_D}	imp_{f_G}	imp_{p_G}
a)	1	.5	.5	.5	1	1
b)	1	.5	1	1	1	1

Like $\mathrm{con}_{\min}$, measures based on the analysis of α-cuts depend solely on the term with the least membership value. When $A(a) \leq B(b)$ for all tuples, the α-cuts of $A \cap B$ are completely determined by the membership values of A. The same inequality ensures that Goguen implication is 1 for all tuples, producing identical results for a) and b). However, Dienes implication discriminates between these sets of tuples. □

The previous example illustrates that different confidence values may be obtained based on whether the association is interpreted as representing a certainty or a gradual relationship. The next example further illustrates this distinction.

Example 5: Consider the fuzzy sets A and B over nonnegative integers that represent "about 5" and "about 10" defined by the membership functions

$$A(a) = \begin{cases} 1 - \frac{|5-a|}{5}, & \text{if } 0 \leq a \leq 10 \\ 0, & \text{otherwise.} \end{cases}$$
$$B(b) = \begin{cases} 1 - \frac{|10-b|}{5}, & \text{if } 5 \leq b \leq 15 \\ 0, & \text{otherwise.} \end{cases}$$

Tuples of the form $(a, a+5)$, for $0 < a \leq 5$, completely support the gradual relation "the more X is about 5, the more Y is about 10." Consider the membership values produced by the five tuples (1,6), (2,7), (3,8), (4,9), and (5,10), which produce the following membership values:

Number	$A(a)$	$B(b)$	$\min(A(a), B(b))$	I_D	I_G
1	1	1	1	1	1
1	.8	.8	.8	.8	1
1	.6	.6	.6	.6	1
1	.4	.4	.4	.6	1
1	.2	.2	.2	.8	1

The confidence values

$\mathrm{con}_{\min}$	con_w	imp_{f_D}	imp_{p_D}	imp_{f_G}	imp_{p_G}
1	1	.6	.8	1	1

indicate complete support for the gradual relationship but lesser support for the certainty of about 10 given a value that is about 5. □

The disparity between the results of the preceding examples has repercussions for both the measurement and the discovery of associations. In determining the degree to which a set of data supports an association, it is necessary to specify the type of hypothesized relationship between the variables to obtain a proper assessment. Moreover, a successful knowledge discovery strategy should not only identify the existence of a relation, but also the type of relation supported by the data.

IV. Data Distribution and Association Robustness

The use of fuzzy sets in association rules facilitates smooth transitions between subsets of attributes, provides a more natural interpretation to linguistic descriptions, but introduces the problem of small cardinalities described in the preceding section. In addition to these, the partial membership values provide the ability to incorporate information about the distribution of the data into the assessment of a rule. This, in turn, produces a more accurate summarization of the support provided by the data than can be obtained from a crisp partitioning of the domain. Moreover, the analysis of the distribution of the data provides information about the robustness of the support of a rule. We begin the examination of the effect of the distribution of data with several examples that illustrate the importance of the distribution on the measurement of rule confidence.

The relationship between age and salary will be used to demonstrate the benefits of incorporating the distribution of the data in the assessment of an association. For fuzzy rules, the distribution is determined by the membership degrees of the data in the antecedent of the rule. The terms "young" and "middle-aged" are defined by the sets of ages in Table I.

For simplicity, we will define the terms "low" and "middle-income" by the crisp sets [\$20k, \$40k) and [\$40k, \$70k), respectively. Consider the two sets of age and salary data

a) age	salary	b) age	salary
23	30k	23	30k
24	30k	24	30k
25	30k	25	30k
27	50k	23	50k
28	50k	24	50k
29	50k	25	50k

TABLE I
CRISP AND FUZZY PARTITIONS

	Young	*Middle-aged*
Crisp	[0,30)	[30,55)
Fuzzy	$\begin{cases} 1, & \text{if } a \leq 25 \\ -.2(a-30), & \text{if } 25 < a \leq 30 \\ 0, & \text{otherwise.} \end{cases}$	$\begin{cases} 0, & \text{if } a \leq 25 \\ .2(a-25), & \text{if } 25 < a \leq 30 \\ 1, & \text{if } 30 < a \leq 50 \\ .-.2(55-a), & \text{if } 50 < a \leq 55 \\ 0, & \text{otherwise.} \end{cases}$

The sets differ only in the distribution of the ages in the tuples with middle income salaries. In a), these are near the boundary between *Young* and *Middle-aged* while in b) they are distributed in the core of *Young*.

The confidence measures for the rule "if *Age* is *Young*, then *Salary* is *Low*" using the crisp definitions is .5 for both sets of tuples a) and b). The crisp partition does not differentiate between the locations of the data in the set *Young*.

All of the generalized measures produced a confidence of .5 for the tuples in b). The partial membership in the fuzzy set *Young* of the tuples with middle-income salaries in a) decreases their relevance to the assertion "if *Age* is *Young*, then *Salary* is *Low*" resulting in an increased confidence in the measures $\text{con}_{\min} = .625$, $\text{con}_w = .67$, $\text{imp}_{f_D} = .8$, $\text{imp}_{p_D} \approx .87$, and $\text{imp}_{p_G} \approx .71$, thereby distinguishing the support for the association based on the distribution of the data within the fuzzy sets.

Example 6 shows that disregarding the distribution the data can prevent the acceptance of an intuitively supportable association. A standard method of acceptance of an association is to employ a user-defined value θ as a threshold; a confidence value greater than θ indicates acceptance of the rule. In this example, .9 will be considered as the threshold of acceptance.

Example 6: The membership data

Number	Young	Low	Middle-income
7	1	1	0
1	.6	0	1
1	.4	1	0
1	.2	0	1

are used to assess the validity of the rule "if *Age* is *Young*, then *Salary* is *Low*." This data strongly reflects that intent of the rule. However the crisp partition produces a confidence of .8, which does not surpass the threshold. The partial membership in the transition between young and middle-aged reduces the impact of the middle-income salaries producing confidence values for $\text{con}_{\min}$, con_w, imp_{f_D}, imp_{p_D}, and imp_{p_G} that all exceed .9 and support the rule. □

The following two examples show that disregarding the distribution of the data may cause the acceptance of associations of dubious merit. This type of result can occur when the data supporting an association is more highly concentrated in regions of low membership in the antecedent. Intuitively the support for a rule is *robust* if it is obtained from highly relevant data or from data across all levels of relevance. The α-cut decomposition of the fuzzy set A will be used to analyze the robustness of support of a rule $A \Rightarrow B$. The high confidence value for the association in Example 1 is a result of the aggregation hiding the lack of robustness of the support. For those tuples, the α-cut confidence values are

α-level	$\text{con}(A_\alpha, B_\alpha)$
1	0
.01	.91

indicating that there is no support prior to the .01 α level.

Unfortunately, the lack of robustness may not always be as obvious as in the preceding example. Example 7 shows that the distribution of data may cause significant variations in the α-cut confidence.

Example 7: The α-level evaluation of the tuples

Number	$A(a)$	$B(b)$	$\min(A(a), B(b))$
1	1	1	1
n	.8	.7	.7
m	.7	.7	.7

produces α-cut confidence levels

α	$\text{con}(A_\alpha, B_\alpha)$
1	1
.8	$\frac{1}{(n+1)}$
.7	1

that indicate a lack of robustness in the support for the association; α-cuts .9 and .7 support the association while .8 does not. Weighting the α-cuts produces $\text{con}_w > .9$, which indicates strong support for the association regardless of the lack of robustness indicated by the α-cuts. □

Example 8: The tuples in this example further show the inability of assessment using aggregation or α-cuts to distinguish robust from nonrobust support for an association. The tuples in a) and b)

	Number	$A(a)$	$B(b)$	$\min(A(a), B(b))$
a)	n	1	1	1
	n	.6	.6	.6
	n	.1	.1	.1
b)	1	1	1	1
	$3n-1$	.1	.1	.1

both produce the α-cut confidence values

α	$\text{con}(A_\alpha, B_\alpha)$
1	1
.6	1
.1	1

.

In case a), the support for $A \Rightarrow B$ comes from tuples whose A membership values are distributed throughout the range [0,1]. This differs from b) in which only one tuple significantly supports the rule. However, neither the selection of a particular α-cut, $\text{con}_{\min}$, con_w, imp_{f_G}, nor imp_{p_G} differentiates these sets of tuples. When $n = 100$, the certainty rule interpretation based on Dienes implication produces greater support from the data in b) with little relevance, $\text{imp}_{f_D} \approx .90$ and $\text{imp}_{p_D} \approx .90$, than from the evenly distributed data in a), $\text{imp}_{f_D} \approx .83$ and $\text{imp}_{p_D} \approx .85$. This is because a tuple (a, b) with membership values $A(a) = .1$ and $B(b) = .1$ is considered to be more in agreement with (actually, less of a counterexample) of a certainty rule than one with $A(a) = .6$ and $B(b) = .6$. □

V. Data Distribution and Robustness

In the preceding section, the analysis of an association first determined the intersection of the values $A(a)$ and $B(b)$ for each tuple $t = (a, b)$ and then aggregated or normalized the results. The latter step has the concomitant loss of information about the distribution of the tuples that support the association. In [5], Bosc *et al.* proposed the reversal of this procedure. The first step consisted of summarizing the database using a fuzzy cardinality. The information in the resulting fuzzy set was then used to produce a confidence value. Moreover, the summarization permitted a more detailed analysis of relationships among the data. In this section, we will adopt this strategy and initially summarize the data in terms of its membership in A and B.

As before, let $0 = \sigma_1 < \sigma_2 < \cdots < \sigma_r = 1$ and $0 = \gamma_1 < \gamma_2 < \cdots < \gamma_s = 1$ be the representative values for A and B, respectively. A two-dimensional "membership map" $\mathbf{M}$ of the database R is constructed over the set $[1, r] \times [1, s]$ that summarizes the distribution of the tuples in A and B. Each tuple $t = (a, b)$ is associated with a pair (σ_i, γ_j) based on the proximity of $A(a)$ to σ_i and $B(b)$ to γ_j. The representative value σ_{i_t} for an element a is selected as follows:

$$\sigma_{i_t} = \begin{cases} \sigma_1, & \text{if } A(a) = 0 \\ \sigma_2, & \text{if } 0 < A(a) \le \frac{(\sigma_2+\sigma_3)}{2} \\ \sigma_i, & \text{if } \frac{(\sigma_{i-1}+\sigma_i)}{2} < A(a) \le \frac{(\sigma_i+\sigma_{i+1})}{2}, \\ & \quad \text{for } i = 2, \ldots, r-1 \\ \sigma_r, & \text{if } \frac{(\sigma_{r-1}+1)}{2} < A(a) \le 1 \end{cases}$$

That is, an element a with $A(a) > 0$ is assigned to the nearest nonzero σ. The value γ_{i_t} for b is selected in a similar manner. The entry $\mathbf{M}(i, j)$ is the number of tuples in the database that are associated with the pair (σ_i, γ_j).

This first level of processing provides two types of summarization. Organizing the tuples by their membership in the fuzzy sets A and B extracts the information required for the assessment of the association. The selection of the representative set of membership values reduces the size of the resulting map. If the fuzzy sets $R(A)$ and $R(B)$ have a small number of α-levels, the α-levels may be used as the representative values and the summarization has no resulting loss of membership information. In the examples that follow, the representative values σ_i will be the α-levels of the fuzzy sets.

The benefit of the summarization of the membership values is the ability to perform a more detailed analysis that considers the distribution of the values. The membership map provides the information needed to identify the types of tuples that contribute to the confidence measure. An α-cut evaluation suggests a straightforward method for eliminating the adverse effects of small cardinalities, augmenting the acceptance threshold with an α-cut requirement. That is, a rule is supported if the confidence of the designated α-cut exceeds θ. In [5], a threshold on the convex hull formed from the α-cut confidences was used in the determination of the acceptance of an association.

Examples 7 and 8 showed that scalar-valued confidence measures are frequently unaffected by the distribution and the relevance of the data. These examples also demonstrate that the addition of an α-cut criterion is insufficient to ensure that the support for an association is primarily from relevant data.

To distinguish a confidence value produced by tuples that are strongly relevant to a rule from one generated by tuples that barely satisfy the antecedent, it is useful to consider the distribution of the tuples. The value

$$dst_{\sigma_i} = \begin{cases} 1, & \text{if } i = 2 \\ \min\left(1, \frac{|A_{\sigma_i}|}{|\text{support}(A)|\left(1-\frac{(\sigma_{i-1}+\sigma_i)}{2}\right)}\right), & \text{if } 2 < i < r \\ \min\left(1, \frac{|A_{\sigma_r}|}{|\text{support}(A)|\left(1-\frac{\sigma_{r-1}}{2}\right)}\right), & \text{if } i = r \end{cases}$$

where $\text{support}(A) = \{(a, b) \in R \mid A(a) > 0\}$, compares the number of tuples in $|A_{\sigma_i}|$ with the expected number from a uniform distribution across σ-levels. A σ-level that contains at least as many tuples as the expectation yields $dst_{\sigma_i} = 1$. Requiring dst_{σ_i} to be greater than the threshold θ, or another threshold for tuple distribution, incorporates strength of support into the confidence assessment.

Example 9: The computation of the distribution function dst is demonstrated for a set of 100 tuples with A membership values distributed as follows:

$A(a)$	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
No.	10	15	12	7	10	14	13	8	5	6

The representative values of A are $\sigma_1 = 0$, $\sigma_2 = .25$, $\sigma_3 = .5$, $\sigma_4 = .75$, and $\sigma_5 = 1.0$.

For $i = 2, 3, 4, 5$, the interval of A membership values associated with dst_i, the number of tuples in that interval, the expected number of tuples in the interval from a random distribution 100 tuples, and the value dst_i are

i	σ_i	interval	expected	tuples	dst_i
5	1.0	$(8.75, 1]$	12.5	11	.88
4	.75	$(6.25, 1]$	37.5	32	.85
3	.50	$(3.75, 1]$	62.5	63	1.0
2	.25	$(0, 1]$	100	100	1.0

A tuple (a, b) is assigned to the nonzero σ_i that is nearest to $A(a)$. Thus, dst_5 consists of only the interval from $(\sigma_4 + \sigma_5)/2$ to 1 while dst_2 is (0, 1]. □

The addition of the distribution criterion prevents the acceptance of the association for each of the problematic cases in Examples 1, 7, and 8. The σ-level and the associated dst_σ values are given here with $m = n = 100$ in Examples 7 and 8

Ex. 1

σ	dst
1	.002
.01	1

Ex. 7

σ	dst
1	.002
.8	1
.7	1

Ex. 8 a)

σ	dst
1	1
.6	1
.1	1

Ex. 8 b)

σ	dst
1	.007
.01	1

The lack of tuples at high σ-levels in all Examples except 8 a) would preclude the associations from receiving support.

VI. Rule Discovery

Knowledge discovery [13] differs from assessing the degree of satisfaction of a rule because the discovery process generally examines all, or a large number, of possible associations to determine which are supported by the data. Thus, an important aspect of knowledge discovery is the development of efficient algorithms to consider multiple associations simultaneously (see, for example, [6], [17], [18], and [33]) and the selection of interesting associations from the resulting set [4], [16], [22], [28], [31]. In this paper, we focus on the steps required to learn whether there is an association, and if so what type, between two prescribed sets of attributes A and B. The latter question, what type, is introduced by the representation of an association by a generalized implication.

When A and B are crisp sets, the process of discovering an association and checking the validity of one are essentially identical. In this case, there is only one T-norm and implication operator or, more precisely, all T-norms and implication operators are identical when restricted to the set {0,1}. The membership map $\mathbf{M}$ is a 2×2 matrix whose entries provide the number of examples, $\mathbf{M}(2,2)$, and counterexamples, $\mathbf{M}(2,1)$, which combine to yield the number of relevant tuples, $\mathbf{M}(2,1) + \mathbf{M}(2,2)$.

The process of rule checking, and of learning, an association is determining the degree to which the membership map of the data matches the pattern of the implication matrix

$$\mathbf{I}(i,j) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

If the vast majority of the data accumulates in the positions indicated by 1 s in $\mathbf{I}$, the rule is supported.

When the attributes sets A and B are fuzzy, there are infinitely many implication operators that may be selected to assess confidence. Thus, selecting a particular implication to check introduces a bias into the learning process. To avoid the *a priori* selection of an implication, we outline a learning procedure that will construct an implication relation as it analyzes the data. The technique proposed for discovering associations will focus on determining the core of the most restrictive implication that supports an association to a specified threshold θ.

The *core* of an implication relation I over $\mathcal{L}_A \times \mathcal{L}_B$ is the set of entries in the associated matrix $\mathbf{I}$ that have value 1. The core of the implication relations produced by Dienes and R-implications are shown in Section II. There is a one-to-one correspondence between cores of implication relations on $\mathcal{L}_A \rightarrow \mathcal{L}_B$ and nondecreasing functions $g : \mathcal{L}_A \rightarrow \mathcal{L}_B$ that satisfy $g(0) = 0$ and $g(1) = 1$. Such a function g can be obtained from the core of an implication relation by

$$g(\sigma_i) = \gamma_j \text{ where } j = \min\{n \mid 1 = \mathbf{I}(\sigma_i, \gamma_n)\}.$$

Conversely, a function g defines a crisp implication relation

$$\mathbf{I}_g(i,j) = \begin{cases} 1, & \text{if } g(\sigma_i) \le \gamma_j \\ 0, & \text{otherwise.} \end{cases}$$

Using standard notation, matrix $\mathbf{I}_g \le \mathbf{I}_h$ if $\mathbf{I}_g(i,j) \le \mathbf{I}_h(i,j)$ for all $1 \le i \le r$, $1 \le j \le s$. The preceding inequality is satisfied if, and only if, $g(\sigma_i) \ge h(\sigma_i)$ for all i.

For each set of dimensions $r \times s$, there is a greatest and a least core generated by implication operators. The smallest generating function with the corresponding largest core matrix are

$$g_s(\sigma_i) = \begin{cases} 0, & \text{for } 1 \le i \le r-1 \\ 1, & \text{if } i = r \end{cases}$$

$$\mathbf{I}_{g_s} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 & 1 \\ 1 & 1 & 1 & \dots & 1 & 1 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 1 & 1 & 1 & \dots & 1 & 1 \\ 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix}.$$

This represents the least restrictive crisp implication. The implication is completely satisfied by a tuple (a,b) whenever $A(a) < 1$ and represents the generalization of material implication to the scale $\mathcal{L}_A \times \mathcal{L}_B$. $\mathbf{I}_{g_s}$ is the core of the implication generated by the drastic T-norm by residuation on the discrete scale

$$T(x,y) = \begin{cases} \min\{x,y\}, & \text{if } x = 1 \text{ or } y = 1 \\ 0, & \text{otherwise.} \end{cases}$$

The most restrictive crisp implication is that generated by Dienes implication

$$g_l(\sigma_i) = \begin{cases} 0, & \text{for } i = 1 \\ 1, & \text{otherwise.} \end{cases}$$

$$\mathbf{I}_{g_l} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 & 1 \\ 0 & 0 & 0 & \dots & 0 & 1 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix}$$

which is completely satisfied by relevant tuples only when the consequent is true.

The general form of the core of an implication is that of an "inverted staircase." The boundary and monotonicity conditions for an implication operator $\mathbf{I}$ produce a matrix of the form

$$\begin{bmatrix} 1 & 1 & \dots & 1 & 1 & \dots & 1 & 1 & \dots & 1 & 1 & \dots & 1 & 1 \\ 0 & 0 & \dots & 0 & 1 & \dots & 1 & 1 & \dots & 1 & 1 & \dots & 1 & 1 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & 1 & \dots & 1 & 1 & \dots & 1 & 1 \\ \vdots & \vdots & & \vdots & \vdots & & & & & & & & & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 0 & 1 \end{bmatrix}.$$

Learning the core consists of identifying the location of the left-most 1 in each row of the matrix $\mathbf{I}$. If the implication cannot satisfy the "inverted staircase" pattern, the association will not be supported. Subject to the monotonicity conditions, an implication matrix may have any values from [0, 1) in the noncore entries. To efficiently evaluate potential implications, a protocol must be specified for selecting these values in each row.

The learning process presented below iteratively produces values for rows $r, r-1, \dots, 2$ of $\mathbf{I}$. The entries in the rth row will be the representative values $\gamma_1, \dots, \gamma_s$. The selection of these initial values coincides with the satisfaction of $I(1, y) = y$ by the implication.

Subsequences of the values $\gamma_1, \dots, \gamma_s$ will be used in constructing the rows $r-1, r-2, \dots, 2$. The consequence of this selection of values will be discussed after the presentation of the algorithm. We begin with an example to illustrate the construction of an implication matrix from the membership map $\mathbf{M}$ of a database.

In order to determine confidence using implication, the user must select one of the two interpretations of relevance of a tuple; full or partial. The example and algorithm will use partial relevance, which specifies a tuple (a, b) to be an example of $A \Rightarrow B$ to degree $T(A(a), I(A(a), B(b)))$.

Example 10: The objective of this example is to determine if there is an association $A \Rightarrow B$ between fuzzy sets A and B and, if so, of what type based on a comparison of the number of examples and relevant cases in a database. The membership map will be constructed using the representative values $\sigma_1 = \gamma_1 = 0$, $\sigma_2 = \gamma_2 = .5$, and $\sigma_3 = \gamma_3 = 1$. Thus the membership map and the implication matrix will be 3×3 matrices. Assume that the analysis of the data in fuzzy sets A and B produces the membership map

$$\mathbf{M}_1(i,j) = \begin{bmatrix} - & - & - \\ 4 & 6 & 5 \\ 0 & 1 & 3 \end{bmatrix}.$$

The first row has been left uninstantiated since it represents tuples that are irrelevant to the association. The tuples that are recorded in the second row are those with A membership values in $(0, .75]$ and those recorded in the third row have A membership in $(.75, 1]$. As discussed in Section VI, a distribution that contains a significant proportion of highly relevant data is desired for the acceptance of an association. For this data, $dst_1 = dst_{.5} = 1$ and conditions requiring a representative data distribution are satisfied.

The initial implication matrix has the form

$$\begin{bmatrix} 1 & 1 & 1 \\ u & v & 1 \\ 0 & w & 1 \end{bmatrix}.$$

The objective of the data analysis is to determine the most restrictive implication relation, if there is one at all, for which the confidence exceeds a predetermined threshold θ. For purposes of this example, θ is .8.

The default values for row 3 are the original γ values; 0, .5, and 1. The confidence measure for the σ_3-cut using imp_p is

$$\frac{1(\mathbf{M}(3,1)\mathbf{I}(3,1) + \mathbf{M}(3,2)\mathbf{I}(3,2) + \mathbf{M}(3,3)\mathbf{I}(3,3))}{1(\mathbf{M}(3,1) + \mathbf{M}(3,2) + \mathbf{M}(3,3))} = \frac{1(0(0) + 1(.5) + 3(1))}{1(0+1+3)} = \frac{3.5}{4} = .875 \geq \theta$$

and the confidence threshold is exceeded. If the confidence requirement were not met, the process would halt and reject the association.

At this point, the implication matrix has the form

$$\begin{bmatrix} 1 & 1 & 1 \\ u & v & 1 \\ 0 & .5 & 1 \end{bmatrix}$$

and the construction of row 2 is initiated. By the monotonicity conditions of an implication matrix, the minimal values permissible in row 2 are [0, .5, 1]. This provides the first condition to be checked. The analysis is on α-cuts, so the confidence for the σ_2-cut includes elements in the rows 2 and 3. Computing the confidence associated with [0, .5, 1] in the second row produces

$$\frac{3.5 + .5(4(0) + 6(.5) + 5(1))}{4 + .5(4+6+5)} \approx .65 < \theta.$$

Since this failed, core of the matrix is increased by "shifting" the second row to the left producing [.5, 1,1]. Testing for confidence with this row produces

$$\frac{3.5 + .5(4(.5) + 6(1) + 5(1))}{4 + .5(4+6+5)} \approx .87 \geq \theta.$$

Thus, the final matrix produced is

$$\begin{bmatrix} 1 & 1 & 1 \\ .5 & 1 & 1 \\ 0 & .5 & 1 \end{bmatrix}$$

which shows that a gradual association between the fuzzy sets A and B is supported by this data.

Now, assume that the data creates the membership map

$$\mathbf{M}_2(i,j) = \begin{bmatrix} - & - & - \\ 1 & 4 & 10 \\ 0 & 1 & 3 \end{bmatrix}$$

which differs from $\mathbf{M}_1$ only in the distribution of the data in the second row. Computing the σ_2-cut confidence with row 2 of the implication matrix [0,.5, 1] produces

$$\frac{3.5 + .5(1(0) + 4(.5) + 10(1))}{4 + .5(1+4+10)} \approx .83 \geq \theta$$

and this row is accepted. The resulting implication matrix

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & .5 & 1 \\ 0 & .5 & 1 \end{bmatrix}$$

indicates a certainty type relationship between A and B. □

Following the strategy presented in the preceding example, we outline a general approach to identifying the presence and type of an association from summarized membership data. The algorithm will use the following:

$\mathbf{M}$	$r \times s$ membership map;
$\mathbf{I}$	$r \times s$ implication matrix;
θ	acceptance threshold;
psup	support from previous σ cut;
prel	relevance from previous σ cut;
c_r	confidence value in association.

The implication matrix is initialized to

$$\begin{bmatrix} 1 & 1 & 1 & \dots & 1 & 1 \\ 0 & 0 & 0 & \dots & 0 & 1 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & \gamma_2 & \gamma_3 & \dots & \gamma_{r-1} & 1 \end{bmatrix}$$

and the process begins by checking if the σ_r-cut exceeds the confidence threshold. The σ_k-cut confidence is the ratio of the support and relevance sup/rel at that level. These values are maintained separately so that they may be used in the determination of the subsequent σ-cut confidence.

1. $\text{psup} = \text{prel} = 0$ {initialization}
2. $\text{sup} = \sum_{i=1}^{s} \mathbf{I}(r,i)\mathbf{M}(r,i)$ {check rth row}
 $\text{rel} = \sum_{i=1}^{s} \mathbf{M}(r,i)$
3. If $\text{sup}/\text{rel} < \theta$, {exit: the association not supported}
4. For $k = r-1$ to 2 {loop on the remaining rows}
 4.1. $\text{psup} = \text{sup} + \text{psup}$
  $\text{prel} = \text{rel} + \text{prel}$
 4.2. For $i = 1$ to s, $I(k,i) = I(k+1,i)$
 4.3. Repeat
  4.3.1. $\text{sup} = \text{psup} + \sigma_k(\sum_{i=1}^{s} \mathbf{I}(k,i)\mathbf{M}(k,i))$
  4.3.2. $\text{rel} = \text{prel} + \sigma_k(\sum_{i=1}^{s} \mathbf{M}(k,i))$
  4.3.3. if $\text{sup}/\text{rel} < \theta$, then {shift}
   for $i = 1$ to $s-1$, $\mathbf{I}(k,i) = \mathbf{I}(k,i+1)$
  until $\text{sup}/\text{rel} \geq \theta$
4. $c_r = \text{sup}/\text{rel}$

After the satisfaction of the confidence condition for row r in step 3, the algorithm will always produce an implication relation. In the extreme case, the relation will be $\mathbf{I}_{g_s}$, the least restrictive implication. Such a result would be produced by analyzing a membership map of the form

$$\mathbf{M}_3(i,j) = \begin{bmatrix} - & - & - \\ 10 & 0 & 0 \\ 0 & 0 & 5 \end{bmatrix}.$$

It is doubtful whether associations supported by such a weak implication should be accepted. To prevent the generation of such implications, the user may specify a greatest core (or least g) that will be considered. The algorithm can be modified to terminate whenever this condition is violated. This is precisely the situation that occurs in row r, where the core is not permitted to extend to the left of column s.

The loop in step 4 produces an implication matrix $\mathbf{I}_r$ and measure of confidence c_r for the association represented by $\mathbf{I}_r$. We now show that the resulting confidence can be used to imply the presence of certain types of certainty and gradual rules as described in [12].

Theorem 2: If the core of $\mathbf{I}_r$ is $\mathbf{I}_{g_l}$, then $\theta \leq c_r \leq \text{imp}_{p_D}$.

Proof: Since the core of $\mathbf{I}_r$ is $\mathbf{I}_{g_l}$, no shifts have been required and $\mathbf{I}_r(k,j) = \gamma_j$. Combining this with $\mathbf{I}_D(k,j) = \max(1-\sigma_k, \gamma_j) \geq \gamma_j$ yields

$$c_r = \frac{\sum_{k=2}^{r} \left(\sigma_k \sum_{j=1}^{s} \mathbf{M}(k,j)\mathbf{I}_r(k,j) \right)}{\sum_{k=2}^{r} \left(\sigma_k \sum_{j=1}^{s} \mathbf{M}(k,j) \right)} \leq \frac{\sum_{k=2}^{r} \left(\sigma_k \sum_{j=1}^{s} \mathbf{M}(k,j)\mathbf{I}_D(k,j) \right)}{\sum_{k=2}^{r} \left(\sigma_k \sum_{j=1}^{s} \mathbf{M}(k,j) \right)} = \text{imp}_{p_D}.$$

■

Theorem 3: Assume that $\sigma_i = \gamma_i = (i-1)/(r-1)$. If the core of $\mathbf{I}_R$ is upper triangular, then $\theta \leq c_r = \text{imp}_{p_L}$ where imp_{p_L} is the confidence obtained using Lukasiewicz implication.

Proof: For $i > j$, Lukasiewicz implication produces

$$\mathbf{I}_L(i,j) = I_L(\sigma_i, \sigma_j) = 1 - \sigma_i + \sigma_j = 1 - \frac{(j-i)}{(r-1)}.$$

The initial vector $[0 = \sigma_1, \sigma_2, \dots, 1 = \sigma_s]$ is shifted $r-i$ times to obtain row i. Thus, the value in position $\mathbf{I}_r(i,j)$ is

$$\sigma_j + \frac{(r-i)}{(r-1)} = \frac{(j-1)}{(r-1)} + \frac{(r-i)}{(r-1)} = 1 - \frac{(j-i)}{(r-1)}$$

and the two matrices are identical. ■

The import of the preceding results is that data mining can used to hypothesize not only the presence of an association supported but also the type. If an implication matrix with minimal core is generated, the data supports a certainty rule. If a diagonal core is produced, the data supports a Lukasiewicz implication. Moreover, the generation of a core between these shows the constraints imposed on the consequent by the antecedent.

Protocols other than the "vector shift" for selecting potential partial membership values of $\mathbf{I}$ may be implemented and incorporated into step 4.3.3. A requisite property, however, is that any such strategy be efficiently computable and that the results can be related to well known implication operators.

VII. Conclusion

A fuzzy association defines a set of increasingly restrictive constraints on the consequent based upon the satisfaction of the antecedent and may be represented by a generalized implication. This paper has exhibited the dependence of the methods for assessing the support for an association on the underlying implication and upon both the relevance and distribution of the data. In addition to the satisfaction of a confidence threshold, a distribution criterion is proposed to ensure the robustness of the support for an association and mitigate the anomalies that can result from the accumulation of small cardinalities. The ability to represent different types of associations with fuzzy sets requires that data mining fuzzy associations identifies both the presence and the type of a relationship between attributes. An algorithm is presented that generates an implication matrix while analyzing the data. The core of the resulting matrix may be used to identify the type of the association.

REFERENCES

[1] R. Agrawal, T. Imielinski, and A. Swami, "Mining association rules between sets of items in large databases," in *Proc. 1993 ACM SIGMOD Int. Conf. Management of Data*, P. Buneman and S. Jajodia, Eds., Washington, DC, 26–28, 1993, pp. 207–216.

[2] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in *Proc. 20th Int. Conf. Very Large Data Bases (VLDB)*, J. B. Bocca, M. Jarke, and C. Zaniolo, Eds., Santiago, Chile, Sep. 1994, pp. 487–499.

[3] F. Berzal, I. Blanco, D. Sánchez, and M.-A. Vila, "A new framework to assess association rules," in *Advances in Intelligent Data Analysis*. Berlin, Germany: Springer-Verlag, 2001, vol. 2189, LNCS, pp. 95–104.

[4] I. Bhandari, "Attribute focusing: Data mining for the layman," IBM T. J. Watson Research Center, White Plains, NY, Tech. Rep. RC 20136, 1995.

[5] P. Bosc, D. Dubois, O. Pivert, H. Prade, and M. de Calmès, "Fuzzy summarization of data using fuzzy cardinalities," in *Proc. 9th Int. Conf. IPMU 2002*, Annecy, France, 2002, pp. 1553–1559.

[6] S. Brin, R. Motwani, J. D. Ullman, and S. Tsur, "Dynamic itemset counting and implication rules for market basket data," in *Proc. ACM SIGMOD Int. Conf. Management of Data*, J. Peckham, Ed., May 1997, pp. 255–264.

[7] K. C. C. Chan and W.-H. Au, "An effective algorithm for mining interesting quantitative association rules," in *Select. Areas Cryptogr.*, 1997, pp. 88–90.

[8] G. Chen, Q. Wei, and E. Kerre, "Fuzzy data mining: Discovery of fuzzy generalized association rules," in *Recent Issues on Fuzzy Databases*, G. Borgdona and G. Pasi, Eds. Heidelberg, Germany: Physica-Verlag, 2000, Studies in Fuzziness and Soft Computing, pp. 45–66.

[9] M. Delgado, N. Marín, D. Sánchez, and M. A. Vila, "Fuzzy association rules: General model and applications," *IEEE Trans. Fuzzy Syst.*, vol. 11, no. 2, pp. 214–225, Apr. 2003.

[10] M. Delgado, D. Sánchez, M. J. Martin-Bautista, and M. A. Vila Miranda, "A probablistic definition of a nonconvex fuzzy cardinality," *Fuzzy Sets Syst.*, no. 2, pp. 177–190, 2002.

[11] D. Dubois, E. Hullermeier, and H. Prade, "On the representation of fuzzy rules in terms of crisp rules," *Inf. Sci.*, vol. 151, pp. 301–326, 2003.

[12] D. Dubois and H. Prade, "What are fuzzy rules and how to use them," *Fuzzy Sets Syst.*, vol. 84, no. 2, pp. 169–186, 1996.

[13] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," *AI Mag.*, vol. 17, pp. 37–54, 1996.

[14] J. Fodor and R. Yager, "Fuzzy set-theoretic operators and quantifiers," in *Fundamentals of Fuzzy Sets: The Handbook of Fuzzy Sets Series*, D. Dubois and H. Prade, Eds. Boston, MA: Kluwer, 2000, The Handbook of Fuzzy Sets, pp. 125–193.

[15] T. Fukuda, Y. Morimoto, S. Morishita, and T. Tokuyama, "Mining optimized association rules for numeric data," in *Proc. 1996 ACM SIGMOD Int. Conf. Management of Data*, H. V. Jagadish and I. S. Mumick, Eds., Montreal, 1996, pp. 13–24.

[16] P. Gago and C. Bento, "A metric selection of the most promising rules," in *Proc. Principles of Data Mining and Knowledge Discovery, Second European Symp.*, vol. 1510, J. M. Zytkow and M. Quafafou, Eds., Nantes, France, Sep. 1998.

[17] J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," in *Proc. 2000 ACM SIGMOD Int. Conf. Management of Data*, W. Chen, J. Naughton, and P. A. Bernstein, Eds., 05, 2000, pp. 214–218.

[18] J. Hipp, U. Guntzer, and G. Nakhaeizadeh, "Algorithms for association rule mining-A general survey and comparison," *SIGKDD Explor.*, pp. 58–64, 2000.

[19] T.-P. Hong, C.-S. Kuo, and S.-C. Chi, "Mining association rules from quantitative data," *Intell. Data Anal.*, vol. 3, pp. 363–376, 1999.

[20] E. Hullermeier, "Fuzzy association rules semantic issues and quality measures," in *Computational Intelligence: Theory and Applications*, B. Reusch, Ed. Berlin, Germany: Springer-Verlag, 2001, LNCS, pp. 380–391.

[21] E. P. Klement, R. Mesiar, and E. Pap, *Triangular Norms*. Dordrecht, The Netherlands: Kluwer, 2000.

[22] M. Klemettinen, H. Mannila, P. Ronkainen, H. Toivonen, and A. I. Verkamo, "Finding interesting rules from large sets of discovered association rules," in *Proc. 3rd Int. Conf. Information and Knowledge Management (CIKM'94)*, N. R. Adam, B. K. Bhargava, and Y. Yesha, Eds., 1994, pp. 401–407.

[23] Y. Kodratoff, "Comparing machine learning and knowledge discovery in databases: An application to knowledge discovery in texts," in *Machine Learning and its Applications*, G. Paliouras, V. Karkaletsis, and C. D. Spyropoulos, Eds. Berlin, Germany: Springer-Verlag, 2001, vol. 2049, Lecture Notes in Computer Science, pp. 1–21.

[24] C. M. Kuok, A. W.-C. Fu, and M. H. Wong, "Mining fuzzy association rules in databases," *Proc. SIGMOD Rec.*, vol. 27, no. 1, pp. 41–46, 1998.

[25] M. J. Martin-Bautista, D. Sánchez, M. A. Vila, and H. Larsen, "Measuring effectiveness in fuzzy information retrieval," in *Flexible Query Answering in Fuzzy Information Retrieval*. Heidelberg, Germany: Physica-Verlag, 2001, pp. 396–402.

[26] R. Mesiar, "Triangular norms—An overview," in *Computational Intelligence in Theory and Practice*, B. Reusch and K.-H. Temme, Eds. Heidelberg, Germany: Phsyica-Verlag, 2001.

[27] R. J. Miller and Y. Yang, "Association rules over interval data," in *Proc. ACM SIGMOD Int. Conf. Management of Data*, Tuscon, AZ, May 1997, pp. 452–461.

[28] G. Piatetsky-Shapiro, "Discovery, analysis, and presentation of strong rules," in *Knowledge Discovery in Databases*, G. Piatetsky-Shapiro and W. Frawley, Eds. Menlo Park, CA: AAAI/MIT Press, 1991, pp. 229–248.

[29] B. Schweizer and A. Skalar, "Associative functions and statistical triangle inequalities," *Publicationes Mathmaticae*, pp. 169–186, 1961.

[30] R. Srikant and R. Agrawal, "Mining quantitative association rules in large relational tables," in *Proc. 1996 ACM SIGMOD Int. Conf. Management of Data*, H. V. Jagadish and I. S. Mumick, Eds., Montreal, 1996, pp. 1–12.

[31] ——, "Mining generalized association rules," *Future Gen. Comput. Syst.*, vol. 13, no. 2–3, pp. 161–180, 1997.

[32] J. Wijsen and R. Meersman, "On the complexity of mining quantitative association rules," *Data Mining Knowl. Disc.*, vol. 2, pp. 263–281, 1998.

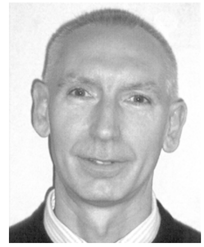
[33] M. J. Zaki, S. Parthasarathy, M. Ogihara, and W. Li, "New algorithms for fast discovery of association rules," in *Proc. 3rd Int. Conf. Knowledge Discovery and Data Mining*, D. Heckerman, H. Mannila, D. Pregibon, and R. Uthurusamy, Eds., 12–15, 1997, pp. 283–296.

Didier Dubois (M'92) received the Engineer and the Doctor Engineer degrees from the Ecole Nationale Supérieure de l'Aéronautique et de l'Espace, Toulouse, France, the "Doctorat d'Etat" degree from the University of Grenoble, France, and the "Habilitation à Dirigier des Recherches" degree from the University of Toulouse III, Toulouse, France, in 1975, 1977, 1983, and 1986, respectively.

He is a full-time Researcher at the National Center for Scientific Research (CNRS), France, where he has been Directeur de Recherche since 1990. He has coauthored two books on fuzzy sets and possibility theory. He has coedited *Readings in Fuzzy Sets for Intelligent Systems* (San Mateo, CA: Morgan and Kauffman, 1993) and *Fuzzy Information Engineering: A Guided Tour of Applications* (New York: Wiley 1997). He has also coedited the *Handbooks of Fuzzy Sets Series* (Norwell, MA: Kluwer 1998–2000). His main topics of interest are the modeling of imprecision and uncertainty, the representation of knowledge and approximate reasoning for intelligent systems, operations research, and decision analysis. He is Co-Editor-in-Chief of *Fuzzy Sets and Systems* and a Member of the editorial board of several other technical journals.

Henri Prade received the Doctor Engineer degree from ENSAE, Toulouse, France, and the "Doctorat d'Etat" and "Habilitation à Dirigier des Recherches" degrees, both from the University of Toulouse, Toulouse, France, in 1977, 1982, and 1986, respectively.

He has been a "Directeur de Recherche" at the National Center for Scientific Research (CNRS), France, and he works as a Research Advisor at the Institut de Recherche en Informatique de Toulouse (IRIT), Toulouse, France. He is a coauthor of two monographs on fuzzy sets and possibility theory. He has contributed a great number of technical papers on uncertainty modeling and applications. He is the coeditor of *Fuzzy Information Engineering: A Guided Tour of Applications* (New York: Wiley 1997) and of *Handbooks of Fuzzy Sets Series* (Norwell, MA: Kluwer 1998–2000). His current research interests are in uncertainty modeling, nonclassical logics, and approximate and plausible reasoning with applications to artificial intelligence, information systems, and operations research. He is Co-Editor-in-Chief of *Fuzzy Sets and Systems* and a Member of the editorial board of several other technical journals.



Thomas Sudkamp (M'94–SM'02) received the B.S. degree in mathematics from the University of Wisconsin, Madison, in 1974, the M.S. and Ph.D. degrees in mathematics from the University of Notre Dame, South Bend, IN, in 1976 and 1978, respectively, and the M.S. degree in computer science from Wright State University, Dayton, OH, in 1982.

He is currently a Professor in the Department of Computer Science and Engineering, Wright State University. His research interests are in the application of soft computing techniques to modeling and decision making in complex problem domains.

Dr. Sudkamp is an Associate Editor of the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS and the IEEE TRANSACTIONS ON FUZZY SYSTEMS. He has served as President of the North American Fuzzy Information Processing Society and is currently Vice President of the International Fuzzy Systems Association.