

Fuzzy Rules in Data Mining: From Fuzzy Associations to Gradual Dependencies

Eyke Hüllermeier

Abstract Fuzzy rules, doubtlessly one of the most powerful tools of fuzzy logic, have not only been used successfully in established application areas like control engineering and approximate reasoning, but more recently also in the field of data mining. In this chapter, we provide a synthesis of different approaches to fuzzy association analysis, that is, the data-driven extraction of interesting patterns expressed in the form of fuzzy rules. In this regard, we highlight a specific advantage of a fuzzy in comparison to a conventional approach, namely an increased expressiveness that allows for representing patterns of interest in a more distinctive way. Therefore, we specifically focus on the modeling of a less common type of pattern, namely gradual dependencies between attributes in a data set.

1 Introduction

Tools and techniques that have been developed during the last 40 years in the field of fuzzy set theory and fuzzy logic (in the broad sense) have been applied quite successfully in a variety of application areas. A prominent example of the practical usefulness of such techniques is *fuzzy control* [22, 23]. Yet, fuzzy extensions of existing methods have also been used and developed in many other fields, including approximate reasoning, optimization, decision support, image processing, and robotics, just to name a few.

While different tools from fuzzy logic have been employed in all these fields, it is arguably fair to say that *fuzzy rule models* have received special attention. Indeed, rule-based models have always been a cornerstone of fuzzy systems and a central aspect of research in fuzzy logic. To a large extent, the popularity of rule-based models can be attributed to their potential com-

Department of Mathematics and Computer Science, University of Marburg, Germany
eyke@mathematik.uni-marburg.de

prehensibility, a distinguishing feature and key advantage in comparison to “black-box” models such as neural networks.

While aspects of knowledge representation and reasoning have dominated research in fuzzy logic for a long time, problems of *automated learning and knowledge acquisition* have more and more come to the fore in recent years [18]. There are several reasons for this development, notably the following: First, caused by the awareness of the well-known “knowledge acquisition bottleneck”, there has been an internal shift within fuzzy systems research from “modeling” to “learning”, i.e., from the knowledge-driven to the data-driven design of fuzzy systems (see e.g. [2]). Second, this trend has been further amplified by the great interest that the fields of *knowledge discovery in databases* (KDD) and its core methodological component, *data mining*, have attracted in recent years [12].

The goal of this chapter is to elaborate on the use of fuzzy rules in data mining. More specifically, using the problem of association analysis as an important and to some extent representative example, we seek to highlight one of the key advantages of fuzzy methods, namely their *increased expressiveness* for representing and discovering patterns of interest in data. To this end, we begin with a brief introduction to association analysis in Section 2. In the following sections, we shall then discuss different types of fuzzy association rules. The chapter ends with some concluding remarks in Section 6.

2 Association Analysis

Association analysis [1, 25] is a widely applied data mining technique that has been studied intensively in recent years. The goal in association analysis is to find “interesting” associations in a data set, that is, dependencies between so-called itemsets \mathbf{A} and \mathbf{B} expressed in terms of rules of the form “IF \mathbf{A} THEN \mathbf{B} ”, or $\mathbf{A} \rightarrow \mathbf{B}$ for short. To illustrate, consider the well-known example where items are products and a data record (transaction) is a shopping basket such as $\{\text{butter, milk, bread}\}$. The intended meaning of an association $\mathbf{A} \rightarrow \mathbf{B}$ is that, if \mathbf{A} is present in a transaction, then \mathbf{B} is likely to be present as well. For example, the rule $\{\text{butter, bread}\} \rightarrow \{\text{milk}\}$ suggests that people buying bread and butter are likely to purchase milk, too.

A standard problem in association analysis is to find all rules $\mathbf{A} \rightarrow \mathbf{B}$ the *support* and *confidence* of which reach user-defined thresholds minsupp and minconf , respectively. Given a database comprising N data records (transactions) $\mathbf{D} = \{x_1, \dots, x_N\}$, the support and confidence of a (candidate) rule $\mathbf{A} \rightarrow \mathbf{B}$ are defined, respectively, as follows:

$$\text{supp}(\mathbf{A} \rightarrow \mathbf{B}) = \sum_{i=1}^N \mathbf{A}(x_i) \cdot \mathbf{B}(x_i), \quad (1)$$

$$\text{conf}(\mathbf{A} \rightarrow \mathbf{B}) = \frac{\text{supp}(\mathbf{A} \rightarrow \mathbf{B})}{\sum_{i=1}^N \mathbf{A}(x_i)}, \quad (2)$$

where $\mathbf{A}(x) = 1$ if the subset of items \mathbf{A} is present in the transaction x and $\mathbf{A}(x) = 0$ otherwise. Thus, support is just the frequency of transactions containing both \mathbf{A} and \mathbf{B} , i.e., the number of positive examples of the rule $\mathbf{A} \rightarrow \mathbf{B}$, while confidence puts the number of positive examples in relation to the number of all examples (positive and negative).

More generally, the following contingency table can be taken as a point of departure for analyzing and evaluating a relationship between items or item subsets \mathbf{A} and \mathbf{B} :

	$\mathbf{B}(x) = 0$ $\mathbf{B}(x) = 1$		
$\mathbf{A}(x) = 0$	n_{00}	n_{01}	$n_{0\bullet}$
$\mathbf{A}(x) = 1$	n_{10}	n_{11}	$n_{1\bullet}$
	$n_{\bullet 0}$	$n_{\bullet 1}$	n

In this table, n_{00} denotes the number of transactions x for which $\mathbf{A}(x) = 0$ and $\mathbf{B}(x) = 0$, and the remaining entries are defined analogously. Most of the commonly used evaluation measures for association rules, including support (n_{11}/n) and confidence ($n_{11}/n_{1\bullet}$), can be expressed in terms of these numbers.

2.1 Fuzzy Items and Contingency Diagrams

In the above setting, a single item can be represented in terms of a binary (0/1-valued) attribute reflecting the presence or absence of the item, i.e., the latter is considered as a feature of a transaction. To make association analysis applicable to data sets involving numerical attributes, such attributes are typically discretized into intervals, and each interval is considered as a new binary feature. For example, the attribute `temperature` might be replaced by two binary attributes `cold` and `warm`, where `cold` = 1 (`warm` = 0) if the temperature is below 10 degrees and `warm` = 1 (`cold` = 0) otherwise.

An obvious extension is to use fuzzy sets (fuzzy partitions) instead of intervals (interval partitions), and corresponding approaches to fuzzy association analysis have been proposed by several authors [7, 8]. There are different motivations for a fuzzy approach to association rule mining. In particular, it has been emphasized that, by allowing for “soft” rather than crisp boundaries of intervals, fuzzy sets can avoid certain undesirable threshold or “boundary effects” (see e.g. [26, 19]). The latter refers to the problem that a slight variation of an interval boundary may already cause a considerable change of the

evaluation of an association rule, and therefore strongly influence the data mining result.

Here, we shall emphasize another potential advantage of fuzzy association analysis, namely an increased expressiveness that helps to represent association rules in a more distinctive way. Roughly speaking, this is due to the fact that, given a candidate rule $\mathbf{A} \rightarrow \mathbf{B}$ and a transaction x , the respective truth degrees of the rule antecedent and consequent define a tuple $(u, v) = (\mathbf{A}(x), \mathbf{B}(x)) \in [0, 1]^2$, i.e., a point in the unit square. Instead, in the non-fuzzy case, only the four corner points of this square are assumed, and the numbers of points in these corners correspond to the numbers in the above contingency table. In [16], the term *contingency diagram* was introduced for the collection of the points produced by all transactions in a database. More precisely, a contingency diagram is a two-dimensional diagram in which every transaction x defines a point $(u, v) = (\mathbf{A}(x), \mathbf{B}(x)) \in [0, 1]^2$. Thus, for every transaction x , the values on the abscissa and ordinate are given, respectively, by the degrees $u = \mathbf{A}(x)$ and $v = \mathbf{B}(x)$ to which it satisfies the antecedent and the consequent part of a candidate rule.

An example of a contingency diagram is shown in Fig. 1. This diagram has been produced for the abalone data set¹ from the UCI repository [13] (i.e., “transactions” are instances of the abalone species), modeling the fuzzy sets of high values by linearly scaling the corresponding attribute domains to the interval $[0, 1]$. This diagram may serve as a point of departure for evaluating a candidate rule of the form

$$\{\text{high length, high width}\} \rightarrow \{\text{high weight}\}, \quad (3)$$

suggesting that an abalone with high length and width tends to have a high weight. Note that, when “binarizing” the diagram by thresholding membership degrees at $1/2$, a contingency table is obtained that does not support the rule (3). In fact, since $(\mathbf{A}(x), \mathbf{B}(x)) \in (1/2, 1] \times [0, 1/2]$ for most of the observations and hence $n_{10} \gg n_{11}$, the confidence of this rule is very low.

3 Canonical Extension of Association Rules

The standard approach of association rule mining can be generalized to the fuzzy case in a rather straightforward way, namely by means of a proper generalization of the measures of support and confidence. Recall that, instead of a set-based representation of transactions in terms of itemsets, one can also adopt a feature-based representation and identify items with binary (fuzzy) attributes. Denote by \mathbb{A} the (finite) set of underlying fuzzy attributes. Thus, each transaction is represented in terms of a feature vector x , and for each $A \in \mathbb{A}$, $A(x) \in [0, 1]$ indicates the degree to which x has feature A or, say,

¹ We did not use the whole data set but only a randomly selected subset of 1500 instances.

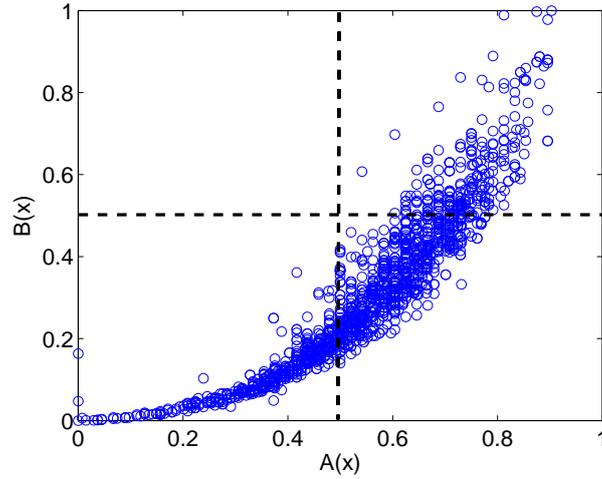


Fig. 1 Example of a contingency diagram. Each value $\mathbf{A}(x)$ on the x-axis is a conjunctive (minimum) combination of the degrees to which an abalone has high length and high width, respectively, and the corresponding value $\mathbf{B}(x)$ on the y-axis is the membership degree of the abalone in the fuzzy set of high overall weights.

to which A is present in x . Correspondingly, the degree of presence of a feature subset $\mathbf{A} = \{A_1, \dots, A_m\}$, that is, a *compound feature* considered as a conjunction of primitive features A_1, \dots, A_m , is specified as

$$\mathbf{A}(x) = \top(A_1(x), A_2(x), \dots, A_m(x)), \quad (4)$$

where \top is a t-norm serving as a generalized conjunction. The simplest way to extend evaluation measures, then, is to replace set-theoretic operations, namely Cartesian product and cardinality, by corresponding fuzzy set-theoretic operations. Modeling the Cartesian product through t-norm combination and the cardinality of a fuzzy set in terms of the σ -count, one thus obtains

$$\text{supp}(\mathbf{A} \rightarrow \mathbf{B}) = \sum_{i=1}^N \top(\mathbf{A}(x_i), \mathbf{B}(x_i)) \quad (5)$$

$$\text{conf}(\mathbf{A} \rightarrow \mathbf{B}) = \frac{\sum_{i=1}^N \top(\mathbf{A}(x_i), \mathbf{B}(x_i))}{\sum_{i=1}^N \mathbf{A}(x_i)} \quad (6)$$

Coming back to our previous example, the evaluation of the rule (3) is now better but still not very strong. Depending on the t-norm used in (6), the confidence is only around 1/2.

The choice of the t-norm \top in (5–6) is indeed an interesting question. Without going into details here, we just mention the existence of theoretical

arguments in favor of the minimum and the product operator; see [9] for a deeper discussion of this topic.

Another important point concerns computational aspects. In fact, since scalability is an issue of utmost importance in data mining, the usefulness of fuzzy extensions presupposes that fuzzy patterns can be mined without sacrificing computational efficiency. Fortunately, efficient algorithmic solutions can be assured in many cases, mainly because fuzzy extensions can usually resort to the same algorithmic principles as non-fuzzy methods. For the above extension, for instance, almost all commonly used rule mining algorithms can be adapted in a rather straightforward way.

4 Implication-based Association Rules

Fuzzy rules can be interpreted in different ways, and depending on the interpretation, different (fuzzy) logical operators are used for modeling a rule on a formal level [11]. In contrast to *conjunction-based* rules that are typically used in applications like fuzzy control, and in which the antecedent and consequent part are combined by means of a t-norm, so-called *gradual rules* combine these parts by means of a *residuated* implication operator \rightsquigarrow . The latter is derived from a t-norm \top through residuation:

$$a \rightsquigarrow b = \sup\{c \mid \top(a, c) \leq b\} \quad (7)$$

This approach to modeling a rule is in agreement with the following interpretation: “The more the antecedent part **A** is true, the more the consequent **B** is true”, for example “The higher the length and width of an abalone, the higher its weight” [24, 10].

So-called *pure* gradual rules are obtained when using the following implication operator:²

$$a \rightsquigarrow b = \begin{cases} 1 & \text{if } a \leq b \\ 0 & \text{if } a > b \end{cases} \quad (8)$$

Using this operator, a rule “The more x is in **A**, the more it is **B**” is interpreted as an ordinary constraint

$$\mathbf{A}(x) \leq \mathbf{B}(x) . \quad (9)$$

This constraint is satisfied if x has property **B** at least as much as property **A**, otherwise it is violated.

An implication-based approach to association rule mining, along with a corresponding re-interpretation of the meaning of a rule, was proposed in [15, 17]. Note, however, that (9) alone is arguably not in agreement with

² This operator is the core of all residuated implications (7).

the intuitive idea of a “positive example” of a rule. In particular, since an implication is true if its antecedent is false, a transaction x with $\mathbf{A}(x) = 0$ would fully support a rule $\mathbf{A} \rightarrow \mathbf{B}$.

As proposed in [15], a possible way out is to combine the implication $\mathbf{A}(x) \rightsquigarrow \mathbf{B}(x)$ conjunctively with the *relevance* of a transaction x for the rule, $\text{Rel}_{\mathbf{A},\mathbf{B}}(x)$, thereby expressing that x supports $\mathbf{A} \rightarrow \mathbf{B}$ if (i) it satisfies the rule in the sense of an implication, and (ii) it is a relevant or, say, non-trivial example for the rule in the sense that it satisfies the condition part. This approach suggests a support measure of the following kind:

$$\text{supp}_x(\mathbf{A} \rightarrow \mathbf{B}) = \top (\text{Rel}_{\mathbf{A},\mathbf{B}}(x), \mathbf{A}(x) \rightsquigarrow \mathbf{B}(x))$$

Regarding the definition of $\text{Rel}_{\mathbf{A},\mathbf{B}}(x)$, note that the constraint (9) is trivially satisfied only in the case $\mathbf{A}(x) = 0$. Therefore, it appears sensible to let $\text{Rel}_{\mathbf{A},\mathbf{B}}(x) = 1$ if $\mathbf{A}(x) > 0$ and $= 0$ otherwise. In combination with the implication (8), this yields the following simple support:

$$\text{supp}(\mathbf{A} \rightarrow \mathbf{B}) = \sum_{i=1}^N \begin{cases} 1 & \text{if } 0 < \mathbf{A}(x_i) \leq \mathbf{B}(x_i) \\ 0 & \text{otherwise} \end{cases}$$

Depending on the type of application, it may of course be reasonable to consider non-triviality as a gradual concept. For example, one may argue that the larger $\mathbf{A}(x)$, the more difficult it is to satisfy constraint (9), i.e., the less trivial this constraint becomes. In this case, an obvious definition of relevance (non-triviality) is $\text{Rel}_{\mathbf{A},\mathbf{B}}(x) = \mathbf{A}(x)$. In conjunction with (8), this leads to the following support measure:

$$\text{supp}(\mathbf{A} \rightarrow \mathbf{B}) = \sum_{i=1}^N \begin{cases} \mathbf{A}(x) & \text{if } \mathbf{A}(x_i) \leq \mathbf{B}(x_i) \\ 0 & \text{if } \mathbf{A}(x_i) > \mathbf{B}(x_i) \end{cases}$$

Again, this result has an intuitively appealing interpretation: A transaction x is a positive example if it satisfies the constraint $\mathbf{A}(x) \leq \mathbf{B}(x)$, 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 \mathbf{A} , i.e., to its degree of relevance.

In terms of this interpretation, our exemplary rule (3) is still not strongly supported by the data, mainly because $\mathbf{A}(x) > \mathbf{B}(x)$ for most x . However, the situation changes when modifying the rule by applying the linguistic hedge “very” to the antecedent part, formally modeled by squaring the membership function [27, 21]. The resulting rule “The more an abalone has very high length and width, the more it has high weight” is then rather well supported, and the confidence (6) increases to more than 0.8.

5 Gradual Dependencies

As mentioned above, fuzzy rules modeled by means of residuated implication operators, called *gradual rules*, are interpreted semantically as THE MORE—THE MORE relationships: “The more the antecedent \mathbf{A} is true, the more the consequent \mathbf{B} is true” [10]. This interpretation is arguable, however. In fact, since $\mathbf{A}(x) \leq \mathbf{B}(x)$ suffices to satisfy a gradual fuzzy rule in a logical sense, there is actually no real consideration of the *change* of an attribute value and, therefore, no examination of a *tendency*. For example, if $\mathbf{B}(x) \equiv 1$, then a gradual rule is always perfectly satisfied, although $\mathbf{B}(x)$ does not depend on $\mathbf{A}(x)$.

Therefore, instead of pursuing a *logical* approach using implication operators to evaluate a rule $\mathbf{A} \rightarrow \mathbf{B}$, a *statistical* alternative was proposed in [16]. This approach directly proceeds from the contingency diagram. Informally speaking, a *gradual dependency* is reflected by the relationship between the points in this diagram. In particular, a THE MORE—THE MORE relationship manifests itself in an increasing trend, i.e., an approximate functional dependency between the u - and v -values: the higher u , the higher v tends to be. In [16], it was therefore suggested to analyze contingency diagrams by means of techniques from statistical regression analysis. For example, if a linear regression line with a significantly positive slope can be fit to the data, this suggests that indeed a higher $u = \mathbf{A}(x)$ tends to come along with a higher $v = \mathbf{B}(x)$.

A qualitative, non-parametric alternative to this numerical approach was proposed in [3]. Roughly speaking, to evaluate a candidate rule $\mathbf{A} \rightarrow \mathbf{B}$, the authors count the number of pairs of points (u, v) and (u', v') in the contingency diagram for which $u < u'$ and $v < v'$. As an advantage of this approach, note that it is more flexible in the sense of not making any assumption about the type of functional dependency; as opposed to this, the regression approach implicitly assumes a linear dependency. From Fig. 1, for example, a positive trend between $u = \mathbf{A}(x)$ and $v = \mathbf{B}(x)$ is clearly visible, but this trend is not linear. On the other hand, since the actual distances between the points are ignored, there is also a disadvantage, namely a loss of information about the strength of a relationship.

The two above approaches, the numerical and the qualitative one, essentially come down to looking for two types of correlation between the u - and v -values, namely the standard Pearson correlation and the rank correlation. In order to combine the advantages of both approaches, [20] proposed to measure the strength of a dependency in terms of a *fuzzy rank correlation* measure that combines properties of both types of correlation. Indeed, as discussed in the following, this measure is able to capture the strength of a tendency while remaining flexible and free of specific model assumptions.

5.1 Fuzzy Rank Correlation

Consider $n \geq 2$ paired observations $\{(u_i, v_i)\}_{i=1}^n \subset (\mathbb{U} \times \mathbb{V})^n$ of two variables U and V , where \mathbb{U} and \mathbb{V} are two linearly ordered domains. The goal of a rank correlation measure is to measure the dependence between the two variables in terms of their tendency to increase and decrease in the same or the opposite direction. If an increase in U tends to come along with an increase in V , then the (rank) correlation is positive. The other way around, the correlation is negative if an increase in U tends to come along with a decrease in V . If there is no dependency of either kind, the correlation is (close to) 0.

Several rank correlation measures are defined in terms of the number C of *concordant*, the number D of *discordant*, and the number N of *tied* data points. For a given index pair $(i, j) \in \{1, \dots, n\}^2$, we say that (i, j) is concordant, discordant or tied depending on whether $(u_i - u_j)(v_i - v_j)$ is positive, negative or 0, respectively. A well-known example is Goodman and Kruskal's *gamma rank correlation* [14], which is defined as $\gamma = (C - D)/(C + D)$.

5.1.1 Fuzzy Equivalence and Order Relations

Bodenhofer and Klawonn [6] propose a fuzzy extension of the gamma coefficient based on concepts of fuzzy orderings and \top -equivalence relations, where \top denotes a t-norm [4].

A fuzzy relation $E : \mathbb{U} \times \mathbb{U} \rightarrow [0, 1]$ is called *fuzzy equivalence* with respect to a t-norm \top , for brevity \top -equivalence, if it is reflexive ($E(u, u) = 1$), symmetric ($E(u, u') = E(u', u)$), and \top -transitive ($\top(E(u, u'), E(u', u'')) \leq E(u, u'')$). Moreover, a fuzzy relation $L : \mathbb{U} \times \mathbb{U} \rightarrow [0, 1]$ is called *fuzzy ordering* with respect to a t-norm \top and a \top -equivalence E , for brevity \top - E -ordering, if it is E -reflexive ($E(u, u') \leq L(u, u')$), \top - E -antisymmetric ($\top(L(u, u'), L(u', u)) \leq E(u, u')$), and \top -transitive ($\top(L(u, u'), L(u', u'')) \leq L(u, u'')$). A \top - E -ordering L is called *strongly complete* if, for all $u, u' \in \mathbb{U}$, either $L(u, u') = 1$ or $L(u', u) = 1$. Finally, let R denote a strict fuzzy ordering associated with a strongly complete \top - E -ordering L ; in the case of the well-known Łukasiewicz t-norm, defined by $\top(a, b) = \max(0, a + b - 1)$, this relation can simply be taken as $R(u, u') = 1 - L(u, u')$ [5].

5.1.2 The Fuzzy Gamma Rank Correlation

Consider a set of paired data points $\{(u_i, v_i)\}_{i=1}^n \subset (\mathbb{U} \times \mathbb{V})^n$ and assume to be given two \top -equivalences $E_{\mathbb{U}}$ and $E_{\mathbb{V}}$ and two strict fuzzy order relations $R_{\mathbb{U}}$ and $R_{\mathbb{V}}$. Using these relations, the concepts of concordance and discordance of data points can be generalized as follows: Given an index pair (i, j) , the degree to which this pair is concordant, discordant, and tied is defined, respectively, as

$$\tilde{C}(i, j) = \top(R_{\mathbb{U}}(u_i, u_j), R_{\mathbb{V}}(v_i, v_j)), \quad (10)$$

$$\tilde{D}(i, j) = \top(R_{\mathbb{U}}(u_i, u_j), R_{\mathbb{V}}(v_j, v_i)), \quad (11)$$

$$\tilde{T}(i, j) = \perp(E_{\mathbb{U}}(u_i, u_j), E_{\mathbb{V}}(v_i, v_j)), \quad (12)$$

where \top is a t-norm and \perp is the dual t-conorm of \top (i.e. $\perp(a, b) = 1 - \top(1 - a, 1 - b)$). The following equality holds for all index pairs (i, j) :

$$\tilde{C}(i, j) + \tilde{C}(j, i) + \tilde{D}(i, j) + \tilde{D}(j, i) + \tilde{T}(i, j) = 1.$$

Adopting the simple σ -count principle to measure the cardinality of a fuzzy set, the number of concordant and discordant pairs can be computed, respectively, as

$$\tilde{C} = \sum_{i=1}^n \sum_{j \neq i} \tilde{C}(i, j), \quad \tilde{D} = \sum_{i=1}^n \sum_{j \neq i} \tilde{D}(i, j).$$

The *fuzzy ordering-based* gamma rank correlation measure $\tilde{\gamma}$, or simply “fuzzy gamma”, is then defined as

$$\tilde{\gamma} = \frac{\tilde{C} - \tilde{D}}{\tilde{C} + \tilde{D}}. \quad (13)$$

From the definition of $\tilde{\gamma}$, it is clear that the basic idea is to decrease the influence of “close-to-tie” pairs (u_i, v_i) and (u_j, v_j) . Such pairs, whether concordant or discordant, are turned into a partial tie, and hence are ignored to some extent. Or, stated differently, there is a smooth transition between being concordant (discordant) and being tied.

5.2 Modeling Gradual Dependencies

The idea of [20] is to evaluate a gradual dependency $\mathbf{A} \rightarrow \mathbf{B}$ in terms of two measures, namely the number of concordant pairs, \tilde{C} , and the rank correlation $\tilde{\gamma}$ as defined in (13). Comparing this approach with the classical support-confidence setting of association analysis, \tilde{C} plays the role of the support of a rule, while $\tilde{\gamma}$ corresponds to the confidence. These measures can also be nicely interpreted within the formal framework proposed in [9], in which every observation (in our case a pair of points $(\mathbf{A}(x_i), \mathbf{B}(x_i))$ and $(\mathbf{A}(x_j), \mathbf{B}(x_j))$) is considered, to a certain degree, as an *example* of a pattern, as a *counterexample*, or as being *irrelevant* for the evaluation of the pattern. In our case, these degrees are given, respectively, by the degree of concordance, the degree of discordance, and the degree to which the pair is a tie.

More formally, the support and confidence of a gradual dependency $\mathbf{A} \rightarrow \mathbf{B}$ can be defined as follows:

$$\text{supp}(\mathbf{A} \rightarrow \mathbf{B}) = \tilde{C}, \quad \text{conf}(\mathbf{A} \rightarrow \mathbf{B}) = \frac{\tilde{C} - \tilde{D}}{\tilde{C} + \tilde{D}},$$

where

$$\begin{aligned} \tilde{C} &= \sum_{x_i} \sum_{x_j} \tilde{C}(x_i, x_j) = \sum_{x_i} \sum_{x_j} \top(R(\mathbf{A}(x_i), \mathbf{A}(x_j)), R(\mathbf{B}(x_i), \mathbf{B}(x_j))), \\ \tilde{D} &= \sum_{x_i} \sum_{x_j} \tilde{D}(x_i, x_j) = \sum_{x_i} \sum_{x_j} \top(R(\mathbf{A}(x_i), \mathbf{A}(x_j)), R(\mathbf{B}(x_j), \mathbf{B}(x_i))). \end{aligned}$$

Considering the special case of the Lukasiewicz t-norm, it can be verified that $E(a, b) = [1 - |a - b|/r]_0^1$ is a \top -equivalence on \mathbb{R} and $R(a, b) = [(a - b)/r]_0^1$ is a strict fuzzy ordering, where $[\cdot]_0^1$ denotes the mapping $a \mapsto \min(1, \max(0, a))$. Note that these relations are parameterized by the value $r \in (0, 1]$. For $r \rightarrow 0$, the confidence measure converges toward the classical (non-fuzzy) rank correlation, whereas for $r = 1$, we obtain $R(a, b) = a - b$ if $a \geq b$ and $= 0$ otherwise. The degree of concordance (discordance) is then proportional to the Euclidean distances, which means that this case is very close to the numerical evaluation in terms of Pearson correlation.

Using this approach, the confidence of our exemplary rule (3) increases to more than 0.95, regardless of parameter r specifying the fuzzy ordering. This is hardly surprising, since the positive dependence of $\mathbf{B}(x)$ on $\mathbf{A}(x)$ is quite obvious from Fig. 1.

6 Concluding Remarks

In this chapter, we have reviewed several fuzzy set-based extensions of methods for association rule mining, an important problem task in data mining. Our main goal was to highlight some advantages of fuzzy extensions, notably an increased expressiveness that allows for modeling different types of patterns associated with different interpretations.

In fact, just like fuzzy rules in general, a pattern represented in terms of a fuzzy IF—THEN expression can be interpreted in different ways, and the concrete semantics is reflected by the logical operators used for modeling a dependency and the measures used for evaluating candidate patterns. In particular, we have distinguished between conjunction-based rules, which are obtained as a straightforward extension of conventional association rules, implication-based rules, in which antecedent and consequent part are combined in terms of a residuated implication operator, and fuzzy gradual dependencies, which are modeled as a kind of statistical instead of a logical relationship.

In the non-fuzzy case, where $\mathbf{A}(x), \mathbf{B}(x) \in \{0, 1\}$, the different types of rules distinguished in the fuzzy case simply coincide, since only three cases can occur. Adopting the notions of positive example, negative example and irrelevancy introduced in [9], a transaction x is a positive example supporting the rule if $\mathbf{A}(x) = \mathbf{B}(x) = 1$, a negative example if $\mathbf{A}(x) = 1$ and $\mathbf{B}(x) = 0$, and it is irrelevant if $\mathbf{A}(x) = 0$.

Finally, we like to emphasize again an important aspect of data mining that was not addressed in this chapter, namely scalable algorithms for mining (fuzzy) associations in an efficient way. In fact, a fuzzy extension would be less interesting if a gain in terms of expressiveness would come along with a loss in terms of efficiency. Fortunately, as mentioned earlier, existing algorithms for mining association rules can often be adapted in a rather straightforward way, without compromising efficiency. This is true at least in the case of conjunction-based fuzzy rules, which are indeed closest to conventional association rules. On the other hand, the development of efficient algorithms is indeed less obvious for implication-based fuzzy rules and gradual dependencies. For a closer discussion, we refer to the references given in the respective sections.

References

1. R. Agrawal and R. Srikant. Fast algorithms for mining association rules. In *Proceedings of the 20th Conference on VLDB*, pages 487–499, Santiago, Chile, 1994.
2. R. Babuska. *Fuzzy Modeling for Control*. Kluwer Academic Publishers, Boston, 1998.
3. F. Berzal, J.C. Cubero, D Sanchez, JM. Serrano, and MA. Vila. An alternative approach to discover gradual dependencies. *Int. Journal of Uncertainty, Fuzziness and Knowledge-based Systems*, 15(5):559–570, 2007.
4. U. Bodenhofer. Representations and constructions of similarity-based fuzzy orderings. *Fuzzy Sets and Systems*, 137:113–136, 2003.
5. U. Bodenhofer and M. Demirci. Strict fuzzy orderings with a given context of similarity. *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems*, 16(2):147–178, 2008.
6. U. Bodenhofer and F. Klawonn. Robust rank correlation coefficients on the basis of fuzzy orderings: Initial steps. *Mathware & Soft Computing*, 15:5–20, 2008.
7. 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.
8. M. Delgado, N. Marin, D. Sanchez, and MA. Vila. Fuzzy association rules: general model and applications. *IEEE Transactions on Fuzzy Systems*, 11(2):214–225, 2003.
9. D. Dubois, E. Hüllermeier, and H. Prade. A systematic approach to the assessment of fuzzy association rules. *Data Mining and Knowledge Discovery*, 13(2):167–192, 2006.
10. D. Dubois and H. Prade. Gradual inference rules in approximate reasoning. *Information Sciences*, 61(1,2):103–122, 1992.
11. D. Dubois and H. Prade. What are fuzzy rules and how to use them. *Fuzzy Sets and Systems*, 84:169–185, 1996.
12. UM. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. From data mining to knowledge discovery: An overview. In *Advances in Knowledge Discovery and Data Mining*, pages 1–34. MIT Press, 1996.
13. A. Frank and A. Asuncion. UCI machine learning repository, 2010.

14. L.A. Goodman and W.H. Kruskal. *Measures of Association for Cross Classifications*. Springer-Verlag, New York, 1979.
15. E. Hüllermeier. Implication-based fuzzy association rules. In *Proc. PKDD-01, 5th European Conference on Principles and Practice of Knowledge Discovery in Databases*, pages 241–252, Freiburg, Germany, 2001. Springer-Verlag.
16. E. Hüllermeier. Association rules for expressing gradual dependencies. In *Proceedings PKDD-02, 6th European Conference on Principles and Practice of Knowledge Discovery in Databases*, pages 200–211, Helsinki, Finland, 2002. Springer-Verlag.
17. E. Hüllermeier. Mining implication-based fuzzy association rules in databases. In *Proceedings IPMU-02, 9th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 101–108, Annecy, France, July 2002.
18. E. Hüllermeier. Fuzzy sets in machine learning and data mining: Status and prospects. *Fuzzy Sets and Systems*, 156(3):387–406, 2005.
19. E. Hüllermeier and Y. Yi. In defense of fuzzy association analysis. *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*, 37(4):1039–1043, 2007.
20. H.W. Koh and E. Hüllermeier. Mining gradual dependencies based on fuzzy rank correlation. In *Proceedings SMPS-2010, 5th International Conference on Soft Methods in Probability and Statistics*, Toulouse, France, 2010.
21. P.J. MacVicar-Whelan. Fuzzy sets, the concept of height, and the hedge 'very'. *IEEE Trans. Systems, Man and Cybernetics*, 8:507–511, 1978.
22. 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.
23. K. Michels, F. Klawonn, R. Kruse, and A. Nrnberger. *Fuzzy Control*. Springer-Verlag, 2006.
24. 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.
25. A. Savasere, E. Omiecinski, and S. Navathe. An efficient algorithm for mining association rules in large databases. In *VLDB-95, Proceedings of 21st International Conference on Very Large Data Bases*, pages 432–444, Zurich, September 1995.
26. T. Sudkamp. Examples, counterexamples, and measuring fuzzy associations. *Fuzzy Sets and Systems*, 149(1):57–71, 2005.
27. L.A. Zadeh. A fuzzy-set theoretic interpretation of linguistic hedges. *J. Cybernetics*, 2(3):4–32, 1972.