UNIVERSITY OF OSTRAVA
FACULTY OF SCIENCE
DEPARTMENT OF MATHEMATICS

FUZZY ASSOCIATION ANALYSIS IN DATA MINING

Ph.D. THESIS

AUTHOR: Pavel Rusnok
SUPERVISOR: prof. Ing. Vilém Novák, DrSc.

2018

OSTRAVSKÁ UNIVERZITA
PŘÍRODOVĚDECKÁ FAKULTA
KATEDRA MATEMATIKY

FUZZY ASOCIAČNÍ ANALÝZA V DOLOVÁNÍ DAT

DOKTORSKÁ DISERTAČNÍ PRÁCE

AUTOR: Pavel Rusnok
VEDOUCÍ PRÁCE: prof. Ing. Vilém Novák, DrSc.

2018

Já, níže podepsaný student, tímto čestně prohlašuji, že text mnou odevzdané závěrečné práce v písemné podobě i na CD nosiči je totožný s textem závěrečné práce vloženým v databázi DIPL2.

Prohlašuji, že předložená práce je mým původním autorským dílem, které jsem vypracoval samostatně. Veškerou literaturu a další zdroje, z nichž jsem při zpracování čerpal, v práci řádně cituji a jsou uvedeny v seznamu použité literatury.

Ostrava ....

(podpis)

Beru na vědomí, že tato doktorská disertační práce je majetkem Ostravské univerzity (autorský zákon Č. 121/2000 Sb., §60 odst. 1), bez jejího souhlasu nesmí být nic z obsahu práce publikováno.

Souhlasím s prezenčním zpřístupněním své práce v Univerzitní knihovně Ostravské univerzity.

Ostrava ....

(podpis)

Acknowledgements

I want to express my gratitude to my supervisor, prof. Ing. Vilém Novák, DrSc. for his support. In every meeting, he gave me valuable comments and encouragement, which pushed me to the completion of this thesis. I would like to also acknowledge Jiří Kupka, Martin Štěpnička, Michal Burda, Petr Hurtík and Nicolas Madrid for their inspiring co-operation.

I would like to thank all of my colleagues from the Institute for Research and Applications of Fuzzy Modeling, which I have not mentioned explicitly for many inspiring and encouraging conversations at the institute halls.

Compiègne, September 2018 Pavel Rusnok

Summary

There are many methods proposed for mining association rules, IF-THEN rules, and linguistic IF-THEN rules. However, many of these methods are proposed rather ad-hoc and lack theoretical and mathematical justification. The aim of this thesis is to define a framework for rules mining based on a logic that includes the semantics of evaluative expressions. We define such logic in Chapter 4. We propose a generalization of a four-fold table known from classical statistics to the case when attributes in a data set are fuzzy. We define the semantics of fuzzy quantifiers as mappings from the generalized four-fold tables into unit interval and investigate various classes of such quantifiers. In Chapter 5, we continue with a study of properties of fuzzy four-fold tables and several quality measures based on them and relations between the choice of used t-norms and chosen quality measures. We present a new theoretical approach to fuzzy association analysis and also discuss the best choice of a t-norm used for mining linguistic associations.

In Chapter 6, we present a method which allows to automatically create an ensemble of regression models and compare this method to standard approaches. The ensemble is created using mined linguistic rule bases which are further used by Perception-based Logical Deduction. As a possible side effect, we can obtain a description of the evaluative process.

In Chapter 7, we investigate fuzzy IF-THEN rule bases reduction methods. We define a probabilistic data coverage measure for evaluating the quality of a set of IF-THEN rules. We also propose a rule reduction procedure based on this measure and compare it theoretically and experimentally in the context of Perception-based Logical Deduction.

Keywords: fuzzy association analysis, linguistic IF-THEN rules, fuzzy quantifiers, four-fold table, quality measures, linguistic description, Perception-based Logical Deduction, regression, ensemble of models, data coverage.

Anotace

Existuje mnoho metod navržených pro dolování asociačních pravidel, IF-THEN pravidel a jazykových IF-THEN pravidel. Nicméně, mnoho z těchto metod je navrženo poněkud ad-hoc a chybí jim teoretické a matematické odůvodnění. Cílem této práce je definovat rámec pro dolování pravidel založený na logice, která zahrnuje sémantiku evaluačních výrazů. Tuto logiku definujeme v kapitole 4. Navrhujeme zobecnění čtyřpolní tabulky známé z klasické statistiky pro případy, kdy atributy v datové sadě jsou fuzzy. Definujeme sémantiku fuzzy kvantifikátorů jako zobrazení ze zobecněných čtyřpolních tabulek do intervalu $[0, 1]$ a zkoumáme různé třídy takových kvantifikátorů. V kapitole 5 pokračujeme ve studiu vlastností fuzzy čtyřpolních tabulek a několika měr kvality založených na nich a vztahů mezi výběrem použitých t-norem a vybranými měrami kvality. Představujeme nový teoretický přístup k fuzzy asociační analýze a také diskutujeme nejlepší volbu t-normy pro dolování jazykových asociací.

V kapitole 6 představujeme metodu, která umožňuje automaticky vytvořit ensemble regresních modelů a porovnáváme tuto metodu se standardními přístupy. Ensemble je vytvářen dolováním jazykových popisů, které jsou dále využívány skrze Perception-based logical deduction[1]. Jako možný vedlejší efekt můžeme získat popis hodnotícího procesu.

V kapitole 7 zkoumáme redukční metody pro báze fuzzy IF-THEN pravidel. Definujeme pravděpodobnostní míru pokrytí dat pro vyhodnocení kvality souboru IF-THEN pravidel. Navrhujeme také postup redukce založený na této míře a porovnáme jej teoreticky a experimentálně v kontextu Perception-based Logical Deduction.

Klíčová slova: fuzzy asociační analýza, jazykové IF-THEN pravidla, fuzzy kvantifikátory, čtyřpolní tabulky, míry kvality, jazykový popis, Perception-based logical deduction, regrese, ensemble modelů, pokrytí dat.

[1] Anglický název, který nemá alternativu v češtině, ale dal by se volně přeložit jako “Logická dedukce na základě percepce vnímání."

Table of Contents

List of Figures

Chapter 1

Introduction

In the late 60's, the first parts of GUHA (general unary hypotheses automaton) theory were formulated, see [20]. The research culminated with the publication of a book [21] under the name Mechanizing Hypothesis Formation. The research still continued and the newest results were summarized in [55] or [22]. GUHA (or its part called Observational Calculi), based on the year in which the research started, may be considered as a first data mining method invented.

Part of the Observational Calculi is so-called *association rules mining*. It belongs to an exploratory part of data mining called pattern recognition. Association rules are formulas of the form $\varphi \rightharpoonup \psi$, where φ is called antecedent and ψ consequent. The rule $\varphi \rightharpoonup \psi$ denotes some kind of implicative relationship between φ and ψ, but even more general $\varphi \rightleftharpoons \psi$ (non-implicative) associations might be mined from data. The semantics of such relationships depend on quality (interestingness) measures [18] that are used to evaluate $\varphi \rightleftharpoons \psi$. Association rules were rediscovered by Agrawal in the 90's [1] and an algorithm under the name Apriori was developed for mining association rules [2]. An algorithmic process of searching such rules in data is called *Association analysis*.

Many versions of fuzzy association rules have been proposed. For a summary of various approaches and their comparison, see [53]. A more theoretically founded approach to fuzzy association rules mining was proposed in [15].

Yager in 1982 in [63] (even before Agrawal), proposed a notion of *linguistic summary* independently on fuzzy associations. Linguistic summaries were further developed by Kacprzyk [28] and are very closely related to fuzzy association rules. There is a significant overlap between these two research directions. For example, *support* in association rules is in fact *degree of focus* in linguistic summaries [62].

However, there is an alternative approach for describing a relationship between entities in a natural language such as *linguistic descriptions*, which are sets of fuzzy IF-THEN rules with linguistic semantics [43]. The linguistic descriptions can be comprised of *evaluative linguistic expressions*, which are basic building blocks of natural language expressions and were mathematically formalized for instance in [43] based on ideas of Lakoff [36]. Evaluative expressions were developed inside of Fuzzy Natural Logic [40, 48].

Mining of linguistic IF-THEN rules that consists of evaluative linguistic expressions was proposed in [49]. The mining was using classical GUHA approach even though the linguistic IF-THEN rules consists of statements that are true for the data they are describing in a degree and not strictly 0 or 1. Thus the need to suggest some more appropriate approach was needed. This is precisely done in Chapters 4, where we define a logical calculus of fuzzy association rules, whose models we investigate in Chapter 5.

A crisp association analysis was successfully applied to the task of classification (see [23, Chapter 6.8] and references therein). Similarly, fuzzy association analysis was used for various regression or prediction type of problems (see [59, 60, 32, 56]). Fuzzy association mining was used to mine IF-THEN rules for an implicative fuzzy inference mechanism called Perception-based Logical Deduction (PbLD). Definition of PbLD and other parts of Fuzzy natural logic are recalled in Section 3.3. Various tasks were solved with this approach as peak flood prediction in [12], general non-linear regression in [32, 33], prediction of time series in [59, 60], and multivariate time series modeling in [56].

In Chapter 6, we investigate further the possibility of using PbLD and linguistic descriptions in regression tasks. We introduce an ensemble modeling approach that uses linguistic description to predict weights of particular models used in an ensemble. We provide several results related to our proposed approach on various datasets from the UCI machine learning repository [3].

Sometimes, too many rules are mined and the resulting linguistic description in PbLD is not interpretable nor readable in a reasonable time. To solve this issue we propose a notion of *probabilistic coverage* of data in Chapter 7. We also propose a new algorithm based on the probabilistic coverage and we compare our approach with previous works in this area. Again we provide several empirical testings on data sets from the UCI machine learning repository [3].

Chapter 2

The author's contribution

My contributions to the thesis can be summarized as follows:

- Chapter 4: Definition of the Logical Calculus of fuzzy association rules that incorporates the theory of the trichotomous evaluative expressions.
- Chapter 5: The study of the relationship between quality measures of associational analysis and a fuzzy four-fold table. All definitions, properties and results that were proven in this chapter.
- Chapter 6: Application of linguistic rules mining into the area of ensemble learning. The definition of two ensemble approaches for combining regression models. Experimental study of their quality and applicability on various data matrices from the UCI machine learning repository [3].
- Chapter 7: Definition of a probabilistic coverage measure of a whole set of linguistic rules. Study of its properties and relationship to a previously defined data coverage. Proposal of a new algorithm for a rule base reduction, based on the probabilistic coverage. Experimental testing of the proposed algorithm on data matrices from the UCI machine learning repository [3].

List of author's publications

Journal articles

- Kupka, J. a Rusnok, P. Regression analysis based on linguistic associations and Perception-based Logical Deduction. *EXPERT SYSTEMS WITH APPLICATIONS*, 67, 2017, pp. 107–114.
- Burda, M., Rusnok, P. a Štěpnička, M. Mining Linguistic Associations for Emergent Flood Prediction Adjustment. Advances in Fuzzy Systems. 2013, pp. 1–10.
- Rusnok, P. a Madrid, N. A Top-K retrieval algorithm based on a decomposition of ranking functions. *Information Sciences*, (accepted).
- Kupka, J. a Rusnok, P. Fuzzy four-fold tables: their properties and use in fuzzy association analysis. *International Journal of Approximate Reasoning*, (submitted).

Conference articles

- Kupka, J. and Rusnok, P. Regression Ensemble with Linguistic Descriptions. In International Conference Series on Soft Methods in Probability and Statistics, Springer, 2018, pp. 141–148
- Rusnok, P. and Burda, M. Global quality measures for fuzzy association rule bases. In: Proceedings of the Conference of the European Society for Fuzzy Logic and Technology 2017-09-11 Varšava. Cham: Springer, 2017, pp. 268–276.
- Kupka, J. and Rusnok, P. Fuzzy quantifiers on fuzzy four-fold tables. In: Proceedings of the 20th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty 2017-09-17 Pardubice. Pardubice: University of Ostrava, 2017, pp. 82–87.
- Madrid, N., Števuliáková, P. and Rusnok, P. Residuated F-transforms extensions for data analysis. In: THE 12TH INTERNATIONAL FLINS CONFERENCE: Uncertainty Modelling in Knowledge Engineering and Decision Making 2016-08-24 Roubaix. Singapore: World Scientific, 2016, pp. 313–318.

- Rusnok, P. Probabilistic Coverage of Linguistic IF-THEN Rules. In: 2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2016): IEEE International Fuzzy Systems Conference Proceedings 2016-07-24 Vancouver. Vancouver, CANADA: Institute of Electrical and Electronics Engineers Inc., 2016, pp. 798–803.
- Rusnok, P. and Madrid, N. M. Verification of Top-k algorithm for a family of non-monotonic ranking function. In: The 2015 IEEE International Conference on Systems, Man, and Cybernetics: 2015 IEEE International Conference on Systems, Man, and Cybernetics (SMC2015) 2015-10-09 Hongkong. Kowloon: IEEE, 2015. IEEE, 2015, pp. 2643–2648.
- Rusnok, P. and Kupka, J. The Role of a T-norm and Partitioning in Fuzzy Association Analysis. In: Strengthening Links between Data Analysis and Soft Computing. Springer-Verlag, 2014, pp. 283–291.
- Kupka, J. and Rusnok, P. On regression methods based on linguistic descriptions. In: 2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2015): IEEE International Fuzzy Systems Conference Proceedings 2015-08-02 Istanbul. Istanbul: IEEE, 2015. IEEE, 2015, pp. 1-7.
- Rusnok, P. Linguistic Descriptions As a Modeling Tool For Multivariate Time Series. In: IFSA-EUSFLAT 2015: Proceedings of the 2015 Conference of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology 2015-06-30 Gijon, Spain. Atlantis Press, 2015, pp. 981–986.

Chapter 3

Preliminaries

We list in this chapter the basic mathematical symbols used throughout the thesis. We also recall basic notions from fuzzy set theory, theory evaluative linguistic expressions and we describe Perception-based Logical Deduction. The definitions in this chapter are from [50, 48] or are slightly syntactically adjusted for the purposes of this thesis.

3.1 Basic mathematical symbols

Definition 1 *An* n-tuple *is an ordered sequence of real numbers, if not stated otherwise.*

Definition 2 *Cardinality of a set is denoted as* $|.|$.

Definition 3 *We denote the set of real numbers as* $\mathbb{R}$ *and the n-tuple of non-negative real numbers is denoted by* $(\mathbb{R}_0^+)^n$. *The set of natural numbers by* $\mathbb{N}$ *and the subset* $\{1, \ldots, k\}$ *of natural numbers as* $\hat{k}$.

3.2 Basic concepts of fuzzy set theory

The research in fuzzy set theory was initiated in 1965 by the paper with the title *Fuzzy Sets* [64] by Lotfi Zadeh. We recall here some basic notions from fuzzy set theory that are used throughout the thesis.

Definition 4 *By a* fuzzy set A *on a universe* U *we mean a map* $A : U \to [0, 1]$*. We denote the fact that* A *is a fuzzy set on* U *as* $A \underset{\sim}{\subset} U$*.*

Definition 5 *A* support *of a fuzzy set* A *is the set of all elements of* U*, whose membership degree to* A *is greater than zero, i.e.,*

$$\mathrm{Supp}(A) = \{u | u \in U,\ A(u) > 0\}.$$

A kernel *of a fuzzy set* A *is set of elements of* U *that fully belong to* A*, i.e.,*

$$\mathrm{Ker}(A) = \{u | u \in U,\ A(u) = 1\}.$$

We say that fuzzy set A *is* normal *if* $\mathrm{Ker}(A) = \emptyset$*.*

A fuzzy set $A \underset{\sim}{\subset} U \subseteq \mathbb{R}$ *is* convex*, if for all elements* $u, v \in U$ *and any* $0 \le \lambda \le 1$ *it holds that*

$$\min\{A(u), A(v)\} \le A(\lambda u + (1 - \lambda)v).$$

Definition 6 *A* t-norm *is a mapping* $\otimes : [0, 1]^2 \to [0, 1]$*, that is commutative, associative, monotone and* 1 *is an identity element:* $\otimes(a, 1) = a$ *for all* $a \in [0, 1]$*. Instead of* $\otimes(a, b)$ *we will alternatively write* $a \otimes b$*.*

Example 7 *There are three main t-norms (Minimum, Product, and Łukasiewicz) that are very often used in theory and applications*

$$\otimes_m(a, b) := \min\{a, b\}, \quad (3.1)$$

$$\otimes_p(a, b) := a \cdot b, \quad (3.2)$$

$$\otimes(a, b) := \max\{a + b - 1, 0\}. \quad (3.3)$$

We recall the definitions of a partial ordering of t-norms and strictness of t-norms. We use the same symbol for the ordering of reals and t-norms even though it is a partial ordering in the latter case. It will be clear from the context which we mean.

Definition 8 *We say that* $\otimes_1 \le \otimes_2$ *if and only if* $\otimes_1(a, b) \le \otimes_2(a, b)$ *for all* $a, b \in [0, 1]$*.*

Definition 9 *A t-norm* $\otimes$ *is called* strict *if it is continuous and strictly monotone, i.e.* $x \otimes z < y \otimes z$ *holds for all* $x, y, z \in (0, 1)$ *and* $x < y$*.*

Definition 10 *A* t-conorm *is a mapping* $\oplus : [0,1]^2 \to [0,1]$*, that is commutative, associative, monotone and* 0 *is an identity element:* $\otimes(a,0) = a$ *for all* $a \in [0,1]$*.*

Example 11 *There are three main t-conorms (Maximum, Product, and Łukasiewicz t-conorm) that are very often used in theory and applications*

$$\oplus_m(a,b) \;:=\; \max\{a,b\}, \tag{3.4}$$

$$\oplus_p(a,b) \;:=\; a + b - a \cdot b, \tag{3.5}$$

$$\oplus(a,b) \;:=\; \min\{1, a+b\}. \tag{3.6}$$

Definition 12 *An operation* $\to$ *is a residuation operation (or residuum) with respect to t-norm* $\otimes$ *if*

$$a \otimes b \leq c \text{ iff } a \leq b \to c.$$

The operation $\to$ is also called as *residuated implication.* There are again three main examples of residuated implications with respect to three basic t-norms mentioned in Example 7.

Example 13 *If* $a \leq b \in [0,1]$ *then* $a \to b = 1$ *for all residuated implications. If* $a > b$ *then the residuated implications derived from Minimum, Product, and Łukasiewicz t-norms are as follows*

$$\to_m (a,b) \;:=\; b, \tag{3.7}$$

$$\to_p (a,b) \;:=\; b/a, \tag{3.8}$$

$$\to (a,b) \;:=\; 1 - a + b. \tag{3.9}$$

The only residuated continues implication, up to isomorphisms, is the Łukasiewicz implication. Therefore, it is argued in [50] that it is the most reasonable choice also for applications.

Definition 14 *A* negation *is a non-increasing mapping* $\neg : [0,1] \to [0,1]$*, such that* $\neg 0 = 1$ *and* $\neg 1 = 0$*. If not mentioned otherwise, we will always assume the involutive negation* $\neg a = 1 - a$*.*

More general class of implications are fuzzy implications and they will be used in construction theorems of Subsection 5.3.2.

Definition 15 *A* fuzzy implication *is a binary operation* $I : [0,1]^2 \to [0,1]$ *for which* $I(0,0) = I(1,1) = 1$*,* $I(1,0) = 0$ *and* $x' \leq x$*,* $y' \geq y$ *implies* $I(x',y') \geq I(x,y)$*.*

Finally, we recall the notions of *fuzzy covering* and *fuzzy partition.*

Definition 16 *Let $A_1, \ldots, A_n \subsetsim U$ be fuzzy sets. We say that fuzzy sets $A_1, \ldots, A_n$ form a* fuzzy covering *of U if*

$$U \subseteq \mathrm{Supp}\left(\bigcup_{i=1}^{n} A_i\right).$$

We say that fuzzy sets $A_1, \ldots, A_n$ form a fuzzy partition *if they are normal, convex,*

$$U = \mathrm{Supp}\left(\bigcup_{i=1}^{n} A_i\right),$$

and it holds that

$$A_i(u) + A_j(u) \leq 1$$

for all $u \in U$.

3.3 Fuzzy natural logic

In this section, we present parts of *Fuzzy Natural Logic* (FNL), which is a continuation of work developed in [40] under the name fuzzy logic in a broader sense. FNL is a set of mathematical theories that extend fuzzy logic in a narrow sense (see [50, 44]). The parts of FNL presented here are the theory of evaluative linguistic expressions, the theory of fuzzy/linguistic IF-THEN rules, and logical inference based on them.

3.3.1 Evaluative linguistic expressions

We recall some basic notions from the theory of trichotomous evaluative linguistic expressions (for short *evaluative expressions*), which was developed in the framework of the fuzzy type theory in [43]. Evaluative expressions are a mathematical formalization of expressions in natural language used in decision making processes and various kinds of evaluation. (e. g. *significantly different*, *very small*, etc.). L. A. Zadeh proposed to model their semantics mathematically in [66] and later in [67]. We present here a simplified definition of evaluative expressions, that is suitable for our applications in data mining in subsequent chapters. For more elaborated definitions and philosophy behind these definitions please refer to [45] or [48].

Hedge	Abbreviation
extremely	Ex
significantly	Si
very	Ve

Table 3.1: Linguistic hedges with narrowing effect

Hedge	Abbreviation
rather	Ra
more or less	ML
roughly	Ro
very roughly	VR

Table 3.2: Linguistic hedges with widening effect

Evaluative expressions[1] have the following syntax:

$$\langle\text{linguistic hedge}\rangle\langle\text{atomic evaluative expression}\rangle, \tag{3.10}$$

where *atomic evaluative expression* is one of the *canonical* adjectives *small*, *medium*, *big* (Sm, Me, Bi for short). Vast majority of adjectives (trichotomous evaluative adjectives) can be expressed in equivalent trichotomy (e. g. "thin", "medium", "thick"; "young", "middle aged", "old" etc.). They can be substituted by atomic evaluative expressions when the context is provided. We denote the set of all evaluative expressions as $\mathcal{EE}$.

Furthermore, *linguistic hedges* are adverbs used to make the meaning of evaluative expressions more or less specific. They can have *narrowing* or *widening* effect. See Tables 3.1 and 3.2 for examples of such hedges. In the following, the evaluative expressions will be denoted by script letters $\mathcal{A}$, $\mathcal{B}$, etc. The evaluative expressions are used to characterize linguistically values of variables.. We call such descriptions *evaluative linguistic predications* (in short *evaluative predications*) that have a form:

$$X \text{ is } \mathcal{A}. \tag{3.11}$$

Examples of evaluative predications are "temperature is very high", "gas consumption is medium", etc.

To model the meaning of evaluative predications, we have to be aware that the meaning of atomic evaluative expressions changes in various *contexts*.

[1] We use here only a part of the evaluative linguistic expressions called pure evaluative expressions. For more details see [48]

Definition 17 *Let $v_L, v_M, v_R \in \mathbb{R}$ be numbers such that $v_L < v_M < v_R$. Then the* context *is a strictly increasing bijection*

$$w : [0,1] \to [v_L, v_M] \cup [v_M, v_R], \tag{3.12}$$

where $w(0) = v_L$, $w(0.5) = v_M$, $w(1) = v_R$,.

We may alternatively model the context as a triple of real numbers $w = \langle v_L, v_M, v_R \rangle$, where $[v_L, v_M]$ (resp. $[v_M, v_R]$) is a range of small (resp. big) numbers, v_M is the most typical medium value, and v_L (resp. v_R) is the smallest (resp. largest) thinkable value .

If we want to speak about the values of a variable that stores the number of inhabitants of various cities, then in the case of cities in China the context would be, e.g. $\langle 5000, 50000, 2\cdot 10^7 \rangle$, but in the context of central Europe $\langle 500, 10000, 2\cdot 10^6 \rangle$. See that the middle value in the context does not have to be the average of the values on the left and right but an arbitrary value between them.

The context $\langle 0, 0.5, 1 \rangle$ is called *standard*.

Definition 18 *The* set of all contexts *for evaluative expressions is*

$$W = \{w \mid w \textit{ is a context according to Definition } 17\}. \tag{3.13}$$

Definition 19 *Let w be a context from (3.12), then the* extended inverse *of w is a function $w^{(-1)} : \mathbb{R} \to [0,1]$ defined as*

$$w^{(-1)}(x) = \begin{cases} w^{-1}(x), & \textit{if } x \in [v_L, v_R], \\ 0, & \textit{if } x < v_L, \\ 1, & \textit{if } x > v_R. \end{cases} \tag{3.14}$$

Originally in [43], the semantics of evaluative predications was defined via *intension* and *extension*, e.g. *Small* is assigned a function from a set of intervals to a set of fuzzy sets, because the meaning of *Small* values changes from context to context, i.e. Small towns in China have more inhabitants than Small towns in Europe. This model of linguistic semantics was motivated by natural language research done by linguists. The following definition is due to Carnap [14].

Definition 20 *Let W be a set of contexts and $\mathcal{A}$ be an evaluative expression. The* intension *of $\mathcal{A}$ is a function*

$$\text{Int}(\mathcal{A}) : W \to \mathcal{F}(\mathbb{R}), \tag{3.15}$$

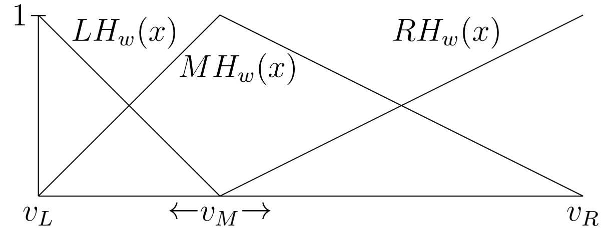

Figure 3.1: Functions representing the three horizons in general context $w = \langle v_L, v_M, v_R \rangle$.

which assigns to any context $w \in W$ a fuzzy set of $\mathbb{R}$. This fuzzy set is called extension of the evaluative expression $\mathcal{A}$ in the context w, i.e.,

$$\mathrm{Ext}_w(\mathcal{A}) = \mathrm{Int}(\mathcal{A})(w) \subsetneqq \mathbb{R}. \tag{3.16}$$

Not all functions from W to $\mathcal{F}(\mathbb{R})$ are suitable as intensions, i.e., they do not capture well the semantics of evaluative expressions. In the following we show en example of parametrized family of intensions. First, we will return to the notion of context and its related notion of *horizon*.

The values of a variable characterized by an evaluative expression are ordered in the context w from v_L to v_R, but it is impossible to stipulate the first small (resp. big) value. However, the small values run from v_L to a certain horizon of small values and after this point, the values are definitely not small. This reasoning applies also to big and medium values, wherein the case of medium values we have to consider horizons on both sides of the most typical value v_M. We now introduce a simple mathematical model of a horizon.

Definition 21 *Let w be a context. Then the horizons are defined as follows*

$$LH_w(x) = \tfrac{v_M - x}{v_M - v_L} \text{ for } x \in [v_L, v_M], \text{ else } LH_w(x) = 0,$$
$$RH_w(x) = \tfrac{x - v_M}{v_R - v_M} \text{ for } x \in [v_M, v_R], \text{ else } RH_w(x) = 0,$$
$$MH_w(x) = \neg LH_w(x) \otimes_m \neg RH_w(x).$$

When context w is standard context $\langle 0, 0.5, 1 \rangle$ then we drop the index w, i.e. the horizons are denoted LH, RH, and MH. An example of horizon functions is in Figure 3.1.

Various mathematical models were suggested by many authors [65, 39, 6] Zadeh, in [65] proposed for a linguistic hedge *very* a function x^2. In general, he proposed to compute extensions of more complicated linguistic expressions from a simpler one. However, G. Lakoff in [36] pointed out that the kernel of membership functions is supposed to be made wider or narrower depending on the type of linguistic hedge and also the slope of the function is supposed to be made steeper or less steep.

This leads us to a mathematical model of hedges as a class of non-decreasing functions $\nu_{a,b,c} : [0,1] \to [0,1]$, with parameters $a, b, c \in [0,1]$, such that $a < b < c$ and the functions fulfill following three conditions

1. $\nu_{a,b,c}(x) = 0$ for all $x \leq a$,
2. $\nu_{a,b,c}(b) = b$,
3. $\nu_{a,b,c}(x) = 1$ for all $x \geq c$.

An example of parametrized functions for modeling linguistic hedges are defined by the following formula

$$\nu_{a,b,c}(x) = \begin{cases} 1, & c \leq x, \\ 1 - \frac{(c-x)^2}{(c-b)(c-a)}, & b \leq x < c, \\ \frac{(x-a)^2}{(b-a)(c-a)}, & a \leq x < b, \\ 0, & x < a. \end{cases}$$

$\langle$empty$\rangle$ hedge is also assigned one of the $\nu_{a,b,c}$ functions. There are infinitely many possible assignments. Note an example of eight $\nu_{a,b,c}$ functions applied to LH_w function representing atomic adjective *small* in Figure 3.2.

Now we are ready to finish the semantics of evaluative predications. First, we define the semantics of evaluative expressions, which are abstract and it is enough to consider only standard context $w = \langle 0, 0.5, 1 \rangle$. This means that their extensions may be identified with their intensions.

$$\begin{aligned} \text{Int}(\langle\text{linguistic hedge}\rangle Sm)(x) &= \nu_{a,b,c}(LH(x)), \\ \text{Int}(\langle\text{linguistic hedge}\rangle Me)(x) &= \nu_{a,b,c}(MH(x)), \\ \text{Int}(\langle\text{linguistic hedge}\rangle Bi)(x) &= \nu_{a,b,c}(RH(x)), \end{aligned}$$

where $x \in w = \langle 0, 0.5, 1 \rangle$ and $\nu_{a,b,c}$ is a function assigned to linguistic hedge.

Recall, that evaluative predication has the form X is $\mathcal{A}$, where $\mathcal{A}$ is an evaluative expression of the form (3.10). We take into consideration what was mentioned above

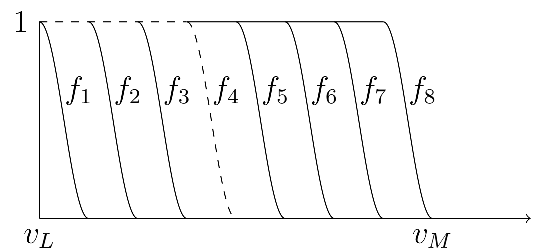

Figure 3.2: All linguistic hedges applied to atomic adjective small. Fuzzy sets $f_1, \ldots, f_8$ capture the meaning of expressions: extremely small, significantly small, very small, small, rather small, more or less small, roughly small, very roughly small in a given context w.

and the fact that a context w is a bijection, then the extensions of X is $\mathcal{A}$ based on definition (20) has the following form

$$\mathrm{Ext}_w(X \text{ is } \langle\text{linguistic hedge}\rangle Sm)(x) = \nu_{a,b,c}(LH(w^{(-1)}(x))), \tag{3.17}$$

$$\mathrm{Ext}_w(X \text{ is } \langle\text{linguistic hedge}\rangle Me)(x) = \nu_{a,b,c}(MH(w^{(-1)}(x))), \tag{3.18}$$

$$\mathrm{Ext}_w(X \text{ is } \langle\text{linguistic hedge}\rangle Bi)(x) = \nu_{a,b,c}(RH(w^{(-1)}(x))), \tag{3.19}$$

where $w \in W$, $x \in w$ and $w^{(-1)}$ is extended inverse according to Definition (19).

In Section 3.3.3, we define an inference mechanism that is relying on ordering of evaluative expressions and for that purposes we define a partial order called *specificity ordering* as follows

$$\begin{aligned} &ExAdj \preceq_s SiAdj \preceq_s VeAdj \preceq_s \langle\text{empty}\rangle\, Adj \preceq_s \\ &\preceq_s RaAdj \preceq_s MLAdj \preceq_s RoAdj \preceq_s VRAdj, \end{aligned} \tag{3.20}$$

where $Adj \in \{Sm, Me, Bi\}$. Linguistic expressions with different atomic adjectives are incomparable (e.g. $ExSm \not\preceq_s Me$).

In the following, we use symbol $\mathcal{A}$ interchangeably for evaluative expressions and fuzzy sets representing evaluative predications X is $\mathcal{A}$. It will be obvious what we mean from the context.

3.3.2 Linguistic descriptions

The evaluative predications can be used in conditional statements of the form of *fuzzy / linguistic IF-THEN rules* (in short IF-THEN rules). They are of the form

$$\mathcal{R} := \textbf{IF } X \text{ is } \mathcal{A} \textbf{ THEN } Y \text{ is } \mathcal{C}, \tag{3.21}$$

when describing relations between two variables. The expression X is $\mathcal{A}$ is called *antecedent* and Y is $\mathcal{C}$ is called *consequent*. Compound antecedents of more variables are also possible:

$$\textbf{IF } (X_1 \text{ is } \mathcal{A}_1 \textbf{ AND } X_2 \text{ is } \mathcal{A}_2 \ldots) \textbf{ THEN } Y \text{ is } \mathcal{C}. \tag{3.22}$$

Below, we describe a method to mine rules of this form. Conjunctions are computed with the minimum t-norm $\wedge$.

For simplicity, the rule $\mathcal{R}$ of the form (3.22) can be written as $\mathcal{A} \rightharpoonup \mathcal{C}$, where

$$\mathcal{A} := \mathcal{A}_1 \wedge \mathcal{A}_2 \wedge \ldots . \tag{3.23}$$

A set of IF-THEN rules is called *linguistic description* and is denoted as $LD = \{\mathcal{R}_1, \ldots, \mathcal{R}_m\}$, where

$$\begin{aligned} \mathcal{R}_1 &:= \textbf{IF } X_1 \text{ is } \mathcal{A}_1 \textbf{ AND } X_2 \text{ is } \mathcal{B}_1 \ldots \textbf{ THEN } Y \text{ is } \mathcal{C}_1, \\ &\vdots \\ \mathcal{R}_m &:= \textbf{IF } X_1 \text{ is } \mathcal{A}_m \textbf{ AND } X_2 \text{ is } \mathcal{B}_m \ldots \textbf{ THEN } Y \text{ is } \mathcal{C}_m. \end{aligned} \tag{3.24}$$

Every linguistic description can be rewritten as a text consisting of conditional statements that describe behavior of some system. In our case, it describes non-linear relationships between variables in a given data set.

Linguistic descriptions are provided by the experts or can be mined from data sets, for which we provide logical calculus in the next chapter.

3.3.3 PbLD inference

In this section, we describe an implicative inference mechanism called Perception-based Logical Deduction (PbLD). For more detailed description and its variants, see [42, 17], and [32].

At the begining of PbLD we are given *input* u.

PbLD based on input u and linguistic description LD deduces a fuzzy set as an output whose defuzzified value is a regression/prediction of the attribute Y appearing on the right side of IF-THEN rules in LD.

Based on input u, the rules in LD are ordered by $\leq_u$ ordering: For two rules $\mathcal{R}_1 := (\mathcal{A}_1 \rightharpoonup \mathcal{C}_1)$, $\mathcal{R}_2 := (\mathcal{A}_2 \rightharpoonup \mathcal{C}_2)$ the ordering is defined as follows:

$$\begin{aligned} &\mathcal{R}_1 \leq_u \mathcal{R}_2 \text{ if} \\ (i)\ &\mathcal{A}_1(u) > \mathcal{A}_2(u), \text{ or} \\ (ii)\ \mathcal{A}_1(u) = \mathcal{A}_2(u) \text{ and } &\mathcal{A}_1 \preceq_s \mathcal{A}_2, \end{aligned} \tag{3.25}$$

where $\preceq_s$ is specificity ordering defined in (3.20). For more complex rules (3.22), where the conjunctions within rules are computed with the help of the minimum t-norm $\otimes_m = \wedge$ (see (3.23)), the definition is analogous.

In the first step of PbLD inference, a set $T^{LD}(u) \subseteq LD$ is chosen. The set consists only of rules with antecedents evaluated to a positive number, namely

$$T^{LD}(u) = \{\mathcal{R} \in LD \,|\, \mathcal{A}(u) > 0 \text{ for } \mathcal{R} := (\mathcal{A} \rightharpoonup \mathcal{C})\}\,.$$

Then, based on the set $T^{LD}(u)$, we form a set called *local perception* $P^{LD}(u)$, containing only the antecedents of the most specific rules:

$$P^{LD}(u) = \left\{\mathcal{A} \in \mathcal{R} \,|\, \nexists \mathcal{R}' \in T^{LD}(u), \mathcal{R}' \leq_u \mathcal{R}\right\}. \tag{3.26}$$

As the final step in PbLD, the *rule of inference* is applied. The rule is given as follows

$$r_{inf}(u) := \frac{P^{LD}(u), LD}{C(u)},$$

where $C(u)$ is a finite set of conclusions (fuzzy sets). For each $\mathcal{A} \in P^{LD}(u)$ from the rule $\mathcal{R} := (\mathcal{A} \rightharpoonup \mathcal{C})$, we have a fuzzy set $C_{\mathcal{R}} \in C(u)$ defined for $x \in Y$ as

$$(C_{\mathcal{R}})(x) = \mathcal{A}(u) \rightarrow_L \mathcal{C}(x).$$

In the case that $C(u)$ contains more than one fuzzy set, we aggregate them by $\wedge$ to get the *final fuzzy set* F. To the resulting fuzzy set we have to apply a *defuzzification*, which is an operation that assigns to nonempty fuzzy set an element from its support.

We mention here three examples of defuzzification operations, which are used in PbLD inference. Assume, that $\{u_j^{max} \,|\, j = 1, \dots, r_{max}\}$ are all elements of the support of a fuzzy set A whose membership degree $A(u_j^{max})$ is maximal, then

- *Mean of maxima* (MOM) operation is defined as

$$\mathrm{MOM}(A) = \frac{1}{r_{max}} \sum_{j=1}^{r_{max}} u_j^{max},$$

- *First of maxima* (FOM) operation is defined as

$$\mathrm{FOM}(A) = \min\{u_j^{max} \,|\, j = 1, \dots, r_{max}\},$$

- *Last of maxima* (LOM) operation is defined as

$$\mathrm{LOM}(A) = \max\{u_j^{max} \,|\, j = 1, \dots, r_{max}\}.$$

The three operations are used in combination in PbLD. The defuzzification operation DEE applied on final fuzzy set F is defined as

$$\mathrm{DEE}(A) = \begin{cases} \mathrm{LOM}(A) & \text{if } A \text{ is non-increasing} \\ \mathrm{MOM}(A) & \text{if } A \text{ is increasing and decreasing} \\ \mathrm{FOM}(A) & \text{if } A \text{ is non-decreasing.} \end{cases}$$

Through defuzzification, we obtain the predicted/regressed value.

PbLD inference was implemented in R language [52] inside lfl: Linguistic Fuzzy Logic package [9, 11] and also as a stand-alone application called LFL Controller [46].

There are also other variants of PbLD developed in [47, 42, 58], and [32].

Chapter 4

Logical Calculus of fuzzy association rules

In this chapter, we define the logical calculus of fuzzy associations, which allow us to formalize statements like

$$\begin{gathered}\text{“If people have high income and are very young then} \\ \text{they very often spend high amounts of money on the entertainment."}\end{gathered} \tag{4.1}$$

This kind of speculations and vague hypotheses are very often formulated in natural language, and we often dispute them in various social occasions. Is it possible to somehow quantify or evaluate such statements? Is it possible to do it based on data in a sound mathematical formalism? We try to answer both questions positively in the remainder of this chapter.

4.1 Syntax

The theory of linguistic expressions (see Section 3.3.1) is suitable to at least partly formalize statement (4.1).

Example 22 *Possible partial formalization of statement (4.1) is as follows*

$$\begin{gathered}\textit{IF Income} \text{ is } \textit{Big AND Age} \text{ is } \textit{Very Small} \\ \textit{THEN VERY OFTEN MoneyOnEnterteinment} \text{ is } \textit{Big}.\end{gathered} \tag{4.2}$$

However, it remains to formalize the words in up-case formatting, for which we need greater expressiveness. We now define the language of fuzzy association rules that will be suitable to formalize the statement in equation (4.1) fully.

Definition 23 *The* language $\mathcal{L}$ *of fuzzy association rules with evaluative expressions* $\mathcal{EE}$ *is a set, which consists of*

- *non-empty finite set of basic attributes* $X_1, \ldots, X_m$*,*
- *non-empty set of evaluative expressions* $\mathcal{EE}$*,*
- *finite set of conjunctions* $\&_1, \ldots, \&_c$*,*
- *finite set of fuzzy quantifiers* $\rightleftharpoons_1 \ldots \rightleftharpoons_q$*.*

We are going to extend the definition of *evaluative predications* from previous chapter. The basic building blocks are evaluative expressions combined with logical connectives of language $\mathcal{L}$.

Definition 24 *The* evaluative predications *of a language* $\mathcal{L}$ *of fuzzy association rules with evaluative expressions* $\mathcal{EE}$ *are defined in the following steps*

1 If X *is a basic attribute and* $\mathcal{A} \in \mathcal{EE}$ *is an evaluative expression then* X is $\mathcal{A}$ *is an evaluative predication.*

2 If EP_1 *and* EP_2 *are (basic) evaluative predications then* $EP_1 \& EP_2$ *is an evaluative predication, where* & *is an arbitrary conjunction from language* $\mathcal{L}$*.*

Finally, we define syntactically the notion of fuzzy/linguistic association.

Definition 25 *Let* EP_1 *and* EP_2 *be evaluative predications and* $\rightleftharpoons$ *be a fuzzy quantifier of the language* $\mathcal{L}$ *with evaluative expressions* $\mathcal{EE}$*. Then* $EP_1 \rightleftharpoons EP_2$ *is a* fuzzy (linguistic) association*.*

Example 26 *Now we can finish the formalization of the statement in equation (4.1) which is an example of fuzzy (linguistic) association.*

$$Income \text{ is } Big \;\&\; Age \text{ is } Very\ Small \rightleftharpoons MoneyOnEnterteinment \text{ is } Big. \tag{4.3}$$

Obviously, the conjunction symbol & is the syntactical representation of the conjunction"and" from natural language, and "IF ... THEN VERY OFTEN" is represented by a quantifier symbol $\rightleftharpoons$*.*

It might be surprising that the whole statement “IF ... THEN VERY OFTEN” is represented by a singular quantifier symbol $\rightleftharpoons$ that looks symmetric. In the original statement in equation (4.1) we can see certain causality direction from left to right as spending money cannot make somebody younger. But we can make similar statements that are implicative in both directions, e.g. “High activity in right-wing political parties is related to the involvement in the business sector”. This latter example of a statement in natural language is more symmetrical than (4.1), and different quantifier would be suitable than in the former case. Symmetry and directionality lead us to the semantics and various classes of functions that would give meaning to symbol $\rightleftharpoons$.

4.2 Semantics

Now we define the semantics of logical calculus of fuzzy association rules. After defining the semantics, we will be able to evaluate the fuzzy associations $EP_1 \rightleftharpoons EP_2$, i.e. how much true they are. First, we formalize the intuitively known notion of data.

Definition 27 *The* data matrix *is a structure $\mathcal{DM} = \langle D, F \rangle$, where D is a non-empty finite set and F is a non-empty finite set of unary functions from D to $\mathbb{R}$. The elements of D are called* rows *(or* objects*) of $\mathcal{DM}$ and the elements of F are called columns (or variables).*

We assume that the number of attributes in the language is the same as the number of unary functions in F. It makes sense because we want to use the syntax of the language from the previous section to form statements that are true about the data matrices of the form $\mathcal{DM}$.

Definition 28 *The model of logical calculus of fuzzy association rules is a structure $\mathcal{M} = \langle \mathcal{DM}, W_m, \bigotimes, \bigotimes_E, Q \rangle$, where $\mathcal{DM}$ is a data matrix from Definition 27, $W_m = \{w_1, ..., w_m\}$. is a set contexts, where m is the cardinality of the set of attributes in language $\mathcal{L}$ (i.e. each attribute is assigned a context), $\bigotimes = \otimes_1, \dots, \otimes_c$ is a set of t-norms, where c is the number of conjunctions in the language $\mathcal{L}$, $\bigotimes_E$ is a set of t-norms used in fuzzy four-fold table, and Q is a set of functions from non-negative real-valued 4-tuples to real unit interval, i.e. $(\mathbb{R}_0^+)^4 \to [0, 1]$.*

We interpret the attributes present in the language $\mathcal{L}$ by data matrices.

Definition 29 *Let X be an attribute from the language $\mathcal{L}$ of logical calculus of fuzzy association rules. Then the interpretation of X by model $\mathcal{M}$ is the function $f \in F$, i.e. $\mathcal{M}(X) = (f(o_1), \ldots, f(o_n))$, where $o_j \in D$ are objects.*

The following definition is an extension of the semantics of evaluative predications from previous chapter using data matrices and models defined in this chapter.

Definition 30 *Let $\mathcal{A}$ be an evaluative expression of the form*

$$\langle \text{linguistic hedge} \rangle Adj,$$

where $Adj \in \{Sm, Me, Bi\}$. The interpretation of an evaluative predication X is $\mathcal{A}$ in a model $\mathcal{M}$ denoted $\mathcal{M}(X \text{ is } \mathcal{A})$ is a fuzzy set defined for every object $o \in D$ as

$$\mathcal{M}(X \text{ is } \mathcal{A})(o) = \begin{cases} \text{Ext}_w(X \text{ is } \langle \text{linguistic hedge} \rangle Sm)(o) = \nu_{a,b,c}(LH(w^{(-1)}(f(o)))), \\ \text{Ext}_w(X \text{ is } \langle \text{linguistic hedge} \rangle Me)(o) = \nu_{a,b,c}(MH(w^{(-1)}(f(o)))), \\ \text{Ext}_w(X \text{ is } \langle \text{linguistic hedge} \rangle Bi)(o) = \nu_{a,b,c}(RH(w^{(-1)}(f(o)))), \end{cases} \tag{4.4}$$

where $f \in F$ is a function assigned to X and $w^{(-1)}$ is an extended inverse of context w assigned to X.

When EP_1 and EP_2 are evaluative predications then the interpretation of a compound evaluative predication $EP_1 \& EP_2$ is a fuzzy set defined for every object $o \in D$ as

$$\mathcal{M}(EP_1 \& EP_2)(o) = \mathcal{M}(EP_1)(o) \otimes \mathcal{M}(EP_1)(o),$$

where $\otimes$ is a t-norm assigned to a conjunction &.

Originally in [43], the semantics of evaluative predications were defined via *intension* and *extension*, e.g. *Small* was assigned a function from a set of intervals to a set of fuzzy sets, because the meaning of *Small* values changes from context to context, i.e. Small towns in China have more inhabitants than Small towns in Europe. This model of linguistic semantics was motivated by natural language research done by linguists.

If we want to be closer to the linguistic semantics of natural language than the extensions of evaluative expressions should be partially overlapping in kernels as in Figure 4.1, because something that is very small is small too. On the other hand, it is usual to define fuzzy sets as in Figure 4.2. In this case, the linguistic expressions are rather labels of prototypical examples. More on this issue can be found in [45].

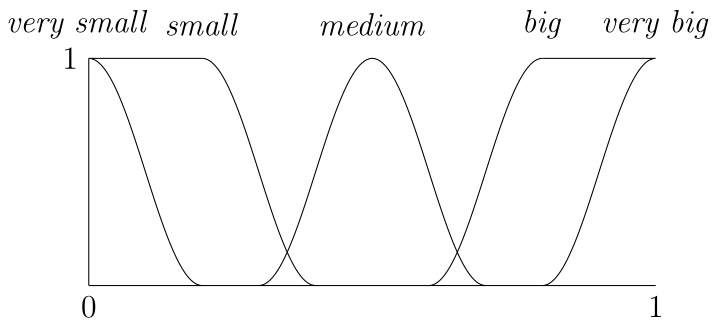

Figure 4.1: Example of extensions of evaluative expressions on standard context.

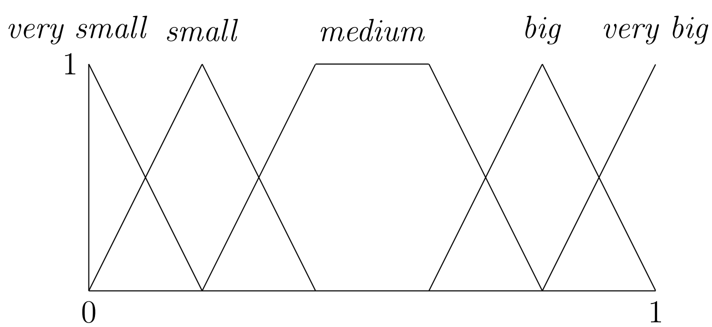

Figure 4.2: Example of linguistic type with non-overlapping kernels.

For simplicity, we denote interpretations of evaluative predications in model $\mathcal{M}$ simply by φ and ψ. Now it remains to assign meaning to the fuzzy associations $EP_1 \rightleftharpoons_i EP_2$. We will use the two remaining constituents of the structure $\langle \mathcal{DM}, W_m, \bigotimes, \bigotimes_E, Q \rangle$. Using $\bigotimes_E$ we define a notion of a fuzzy four-fold table.

Definition 31 *For any two interpretations of evaluative predications φ, ψ, and t-norms $\otimes_a, \otimes_b, \otimes_c, \otimes_d \in \bigotimes_E$ a* generalized fuzzy four-fold table $E(\varphi, \psi, \otimes_a, \otimes_b, \otimes_c, \otimes_d, \neg)$ *is an element of $(\mathbb{R}_0^+)^4$ and is expressed in the following way,*

$$E := \begin{array}{c|c|c} & \psi & \neg\psi \\ \hline \varphi & a & b \\ \hline \neg\varphi & c & d \end{array}, \tag{4.5}$$

where

$$\begin{aligned} a = &\ \textstyle\sum_{o_i \in D} \varphi(o_i) \otimes_a \psi(o_i), \\ b = &\ \textstyle\sum_{o_i \in D} \varphi(o_i) \otimes_b \neg\psi(o_i), \\ c = &\ \textstyle\sum_{o_i \in D} \neg\varphi(o_i) \otimes_c \psi(o_i), \\ d = &\ \textstyle\sum_{o_i \in D} \neg\varphi(o_i) \otimes_d \neg\psi(o_i). \end{aligned}$$

In some cases we simply write $E(\varphi, \psi, \otimes, \neg)$ provided only one t-norm is used (i.e. $\bigotimes_E = \{\otimes\}$, and $\otimes = \otimes_a = \otimes_b = \otimes_c = \otimes_d$) for aggregation of relevant truth degrees.

Observe the difference between the sets of t-norms $\bigotimes$ and $\bigotimes_E$. The former is used to give meaning to the compound evaluative predications and the latter is used to produce generalized fuzzy four-fold tables.

Example 32 *Let us consider a trivial data set from Table 4.1 consisting of two objects and two variables. Let φ be an interpretation of an evaluative predication* Age is Rather Big *and ψ be an interpretation of an evaluative predication* Hight is Very Big. *Their truth degrees for every object are calculated in the second table. The fuzzy four-fold table $E(\varphi, \psi, \otimes_m, \neg)$ with the minimum t-norm $\otimes_m$ and the involutive negation is calculated next to it.*

Observe that the sum $a+b+c+d = 2.6$ is not equal to 2, as it is in the classical case.

Observe also that the t-norms used in the interpretation of compound evaluative predications may be different from the t-norms used in the calculation of fuzzy four-fold table. We make this distinction intentionally. We will argue later that the usage of t-norms for compound evaluative predications may be arbitrary, but in the case of a fuzzy four-fold table, it should be restricted.

	Age	*Height (cm)*
o_1	*18*	*210*
o_2	*34*	*150*

	φ	ψ
o_1	*0.3*	*1*
o_2	*0.6*	*0.3*

$E(\varphi, \psi, \otimes_m, \neg) :=$

	ψ	$\neg\psi$
φ	0.6	0.6
$\neg\varphi$	1	0.4

Table 4.1: Example of a data matrix, intepretations of evaluative predications, and a fuzzy four-fold table.

Definition 33 Fuzzy quantifier function q *associated with fuzzy quantifier* $\rightleftharpoons$ *is a map defined on the set of the fuzzy four-fold tables, i.e.* $q : (\mathbb{R}_0^+)^4 \to [0,1]$, $(a,b,c,d) \mapsto e$.

To simplify the notation, we occasionally write fuzzy quantifier functions as maps of fewer variables, when they do not depend on all variables.

In the sequel we will often use the term *fuzzy quantifier* also for the *fuzzy quantifier function* and it will be clear from the context whether we mean the function or the syntactical symbol.

Definition 34 *The interpretation of a fuzzy association* $EP_1 \rightleftharpoons EP_2$ *in* $\mathcal{M}$ *is*

$$\mathcal{M}(EP_1 \rightleftharpoons EP_2) = q(a,b,c,d),$$

where q *is a fuzzy quantifier function associated with the fuzzy quantifier* $\rightleftharpoons$ *and* $(a,b,c,d) \in (\mathbb{R}_0^+)^4$ *is a fuzzy four-fold table* E *of the evaluative predications' interpretations* φ *and* ψ.

Fuzzy association $EP_1 \rightleftharpoons EP_2$ *is true in a degree* $d \in [0,1]$ *in* $\mathcal{M}$ *if* $\mathcal{M}(EP_1 \rightleftharpoons EP_2) = d$.

In this place, we finish the definition of the semantics of Logical calculus of fuzzy association rules. We extended the semantics of evaluative linguistic expressions. Now, for a given data matrix, we are able to compute a truth degree of an evaluative linguistic expression about our data variables. Furthermore, we can mine linguistic descriptions (see Section 3.3.2) from a given data matrix.

Various generalized quantifiers were proposed [41, 16]. Statements like (4.1) and natural language in general motivated the research of linguistic quantifiers in [41]. Another approach to fuzzy quantifiers determined by fuzzy measures was proposed in [16].

However, in this thesis, we continue the work done by Hájek and Havránek in [20, 21, 55], where general quantifiers were proposed as mappings defined on statistical four-fold tables of boolean attributes. We generalized their approach and combined it with Novák's theory of evaluative linguistic expressions [43].

In practical applications, part of the model $\mathcal{M} = \langle \mathcal{DM}, W_m, \bigotimes, \bigotimes_E, Q \rangle$ is given, e.g., $\mathcal{DM}$ is taken as a set of measurements of some physical process that was repeatedly observed and W_m consists of the minimas and maximas of columns in $\mathcal{DM}$. We want to model the physical process, whose measurements are in $\mathcal{DM}$, with linguistic descriptions and use PbLD. In PbLD, the compound antecedents are formed using minimum t-norm and so $\bigotimes = \{\otimes_m\}$. It remains to discuss $\bigotimes_E$ and Q. And we do it in the next chapter.

Chapter 5

Fuzzy quantifiers and quality measures

We are going to study models $\langle \mathcal{DM}, W_m, \bigotimes, \bigotimes_E, Q\rangle$, and as hinted at the end of the previous chapter we concentrate our investigation on t-norms in the set $\bigotimes_E$ and functions in Q, which give meaning to the quantifiers in the language of Logical calculus of association rules. The results of this chapter are also part of an article [31], that was submitted at the time of submission of this thesis.

Instead of explicitly stating, that φ and ψ are interpretations of some evaluative predications, we will simply say they are *fuzzy attributes*.

5.1 Classes of quantifiers for mining IF-THEN rules

We define various classes of fuzzy quantifiers[1]. There are many classes of classic (non-fuzzy) quantifiers, for a survey of them we refer to [54].

Definition 35 *A fuzzy quantifier q is*

(i) implicational *if and only if it satisfies a* monotonicity condition*: $a' \geq a, b' \leq b$ implies*

$$q(a', b') \geq q(a, b). \tag{5.1}$$

(ii) double implicational *if and only if $a' \geq a$, $b' \leq b$ and $c' \leq c$ implies*

$$q(a', b', c') \geq q(a, b, c).$$

[1] In the whole chapter we use the notion of fuzzy quantifier instead of fuzzy quantifier function for brevity.

(iii) an equivalence *if and only if* $a' \geq a$*,* $b' \leq b$*,* $c' \leq c$ *and* $d' \geq d$ *implies*

$$q(a', b', c', d') \geq q(a, b, c, d).$$

(iv) ratio-implicational *if and only if* $a'b \geq ab'$ *implies*

$$q(a', b') \geq q(a, b).$$

A fuzzy quantifier q is A fuzzy quantifier q is A fuzzy quantifier q is

It follows easily from these definitions that every equivalence quantifier is double implicational and every double implicational quantifier is implicational as well. Moreover, every ratio-implicational quantifier is also an implicational one.

Example 36 *Here we present examples of each class of fuzzy quantifiers introduced above. A quantifier*

$$q(a, b) = \frac{a}{a + b} \tag{5.2}$$

is implicational and known from association rule mining as a confidence measure. *A quantifier*

$$q(a, b, c) = \frac{a}{a + b + c}$$

is double implicational and is known as Jaccard measure. *A quantifier*

$$q(a, b, c, d) = \frac{a + d}{a + b + c + d}$$

is an equivalence. And finally, a quantifier

$$q(a, b) = \frac{a}{a + \theta \cdot b}, \qquad \text{where } \theta > 0,$$

is ratio-implicational.

5.2 Fuzzy confirmation measures and involutively dual t-norms

In this section we show the relationship of generalized fuzzy four-fold tables to the known quality measures of fuzzy association rules and point out their relationship to *involutively dual t-norms* defined in this section.

5.2.1 Confidence and support measures

We will show that our definition is more general and allows us to study the quality measures of fuzzy association rules from different points of views. We first recall the definitions of two very frequently used quality measures. The following definitions can be found e.g. in [8, 15, 19]. Let us consider a data set D consisting of n objects. The first measure is a *support measure* which is defined for a fuzzy attribute φ by

$$\mathrm{supp}(\varphi) = \frac{\sum_{o_i \in D} \varphi(o_i)}{n}, \tag{5.3}$$

and for a fuzzy association rule $\varphi \rightharpoonup \psi$ by

$$\mathrm{supp}_{\otimes}(\varphi \rightharpoonup \psi) = \frac{\sum_{o_i \in D} \varphi(o_i) \otimes \psi(o_i)}{n}. \tag{5.4}$$

In the case of the support of a fuzzy association rule, we use the symbol $\otimes$ of a t-norm in the subscript to denote which t-norm is used. Sometimes the support measure is defined without the division by n, for example in [15]. In that case, the more convenient notion could be a support count (in analogy to classical case [23]).

A *confidence measure* of a given rule is defined by

$$\mathrm{conf}_{\otimes}(\varphi \rightharpoonup \psi) = \frac{\mathrm{supp}_{\otimes}(\varphi \rightharpoonup \psi)}{\mathrm{supp}(\varphi)} = \frac{\sum_{o_i \in D} \varphi(o_i) \otimes \psi(o_i)}{\sum_{o_i \in D} \varphi(o_i)}. \tag{5.5}$$

There are several ways how to define support and confidence measures. Another option is to use a fuzzy four-fold table $E(\varphi, \psi, \otimes_a, \otimes_b, \otimes_c, \otimes_d, \neg)$ (see Definition 31) and suitable fuzzy quantifier function. A support of φ is defined by a fuzzy quantifier function $\mathrm{supp}_E(\varphi)$ by

$$\mathrm{supp}_E(\varphi) = \frac{a+b}{a+b+c+d}, \tag{5.6}$$

and a support of fuzzy association rule $\varphi \rightharpoonup \psi$ is then defined in the following way

$$\mathrm{supp}_E(\varphi \rightharpoonup \psi) = \frac{a}{a+b+c+d}, \tag{5.7}$$

where a, b, c, and d are calculated accordingly to (4.5).

In Section 5.1 we showed an example of an implicational fuzzy quantifier (5.2). We claimed that it is known as the confidence measure of an association rule $\varphi \rightharpoonup \psi$

$$\mathrm{conf}_E(\varphi \rightharpoonup \psi) = \frac{a}{a+b}. \tag{5.8}$$

In the crisp (non-fuzzy) case equations (5.3)–(5.5) coincide with equations (5.6)–(5.8), but they need not coincide in the fuzzy case and may lead to different types of notions.

Example 37 *We continue with Example 32 of two fuzzy attributes φ and ψ. Their fuzzy four-fold table $E(\varphi, \psi, \otimes_m, \neg) :=$ is in Table 4.1. Then the support of the association $\varphi \rightharpoonup \psi$ is*

$$\mathrm{supp}_E(\varphi \rightharpoonup \psi) = \frac{0.6}{0.6+0.6+1+0.4} \doteq 0.23, \quad \mathrm{conf}_E(\varphi \rightharpoonup \psi) = \frac{0.6}{0.6+0.6} \doteq 0.5.$$

We obtain different results when we compute the support and confidence of the same rule with the help of (5.4)-(5.5) and the minimum t-norm $\otimes_m$. Then we obtain

$$\mathrm{supp}_{\otimes_m}(\varphi \rightharpoonup \psi) = \frac{0.3+0.3}{2} = 0.3, \quad \mathrm{conf}_{\otimes_m}(\varphi \rightharpoonup \psi) = \frac{0.3+0.3}{0.9} = 0.67.$$

Later we will see several cases, in which definitions (5.4)–(5.5) and (5.7)–(5.8) coincide. Another natural question is why we count evidence in favor of an association $\varphi \rightharpoonup \psi$ (i.e. a in four-fold table E) with one particular t-norm $\otimes_a$ and the evidence against the same association (i.e. b in E) by a different t-norm $\otimes_b$. One of the reasons is that we want to give higher or lower weight to evidence for/against a rule. Moreover, as Lemma 38 shows, this is very frequently unknowingly done in many fuzzy association applications (see [12, 15, 60, 24]), where the equation (5.5) for calculating confidence is chosen together with the minimum t-norm.

Lemma 38 *Let $\otimes_a = \otimes_m$, $\otimes_b = \otimes$ and $E(\varphi, \psi, \otimes_a, \otimes_b, \otimes_c, \otimes_d, \neg)$ be a fuzzy four-fold table, where $\otimes_c$ and $\otimes_d$ are arbitrary. Then the confidence in equation (5.5) is equal to the confidence in equation (5.8), i.e. following equality holds*

$$\mathrm{conf}_{\otimes_m}(\varphi \rightharpoonup \psi) = \mathrm{conf}_E(\varphi \rightharpoonup \psi) = \frac{a}{a+b}. \tag{5.9}$$

Proof. From Definition 31 and the assumption that $\otimes_a = \otimes_m$ we have

$$a = \textstyle\sum_{o_i \in D} \varphi(o_i) \otimes_a \psi(o_i) = \sum_{o_i \in D} \varphi(o_i) \otimes_m \psi(o_i),$$

and

$$\begin{aligned} a+b &= \textstyle\sum_{o_i \in D} \varphi(o_i) \otimes_a \psi(o_i) + \sum_{o_i \in D} \varphi(o_i) \otimes_b \neg\psi(o_i) \\ &= \textstyle\sum_{o_i \in D} \min\{\varphi(o_i), \psi(o_i)\} + \sum_{o_i \in D} \max\{\varphi(o_i) + \neg\psi(o_i) - 1, 0\} \\ &= \textstyle\sum_{o_i \in D} \min\{\varphi(o_i), \psi(o_i)\} + \max\{\varphi(o_i) + \neg\psi(o_i) - 1, 0\} \\ &= \textstyle\sum_{o_i \in D} \min\{\varphi(o_i), \psi(o_i)\} + \max\{\varphi(o_i) - \psi(o_i), 0\} \\ &= \textstyle\sum_{o_i \in D} \varphi(o_i). \end{aligned}$$

Therefore equation (5.9) holds. □

Under the assumptions of Lemma 38 we cannot expect supports in equations (5.4) and (5.7) to be equal. Their equality is strongly related to the notion of the involutively dual pairs of t-norms (see Definition 39) as it is demonstrated in the next subsection.

5.2.2 Involutively dual pair of t-norms

In this section, we introduce involutively dual pairs of t-norms — in Definition 39 — and study their relationships to fuzzy four-fold tables. This study is also motivated by the paper [15], where natural decompositions of fuzzy association rules into fuzzy partitions were studied. While in [15] each fuzzy association rule can be decomposed into its positive, negative and irrelevant parts, we consider the following natural constraint related to the definition of a fuzzy four-fold table:

$$\varphi(o_i) \otimes_a \psi(o_i) + \varphi(o_i) \otimes_b \neg\psi(o_i) + \neg\varphi(o_i) \otimes_c \psi(o_i) + \neg\varphi(o_i) \otimes_d \neg\psi(o_i) = 1 \quad (5.10)$$

Namely, if (5.10) holds then $a + b + c + d = n$, where n is the number of objects in the data matrix $\mathcal{DM}$. For instance, without (5.10) we usually obtain that supports calculated accordingly to equations (5.4) and (5.7) are different.

Definition 39 *Let $\otimes_1$ and $\otimes_2$ be two t-norms. Then they form an* involutively dual pair *of t-norms if the following holds for each $x, y \in [0, 1]$*

$$(x \otimes_1 y) + x \otimes_2 (1 - y) = x. \quad (5.11)$$

An example of such a pair of two t-norms is the minimum t-norm and the Łukasiewicz t-norm. The only t-norm that is involutively dual to itself is the product t-norm. There are even infinitely many of such pairs. For instance consider the family of Frank t-norms

$$T_p^F(x, y) = \log_p \left(1 + \frac{(p^x - 1)(p^y - 1)}{p - 1}\right), \qquad \text{where } p \in (0, 1) \cup (1, \infty).$$

The limit cases ($p \in \{0, 1, \infty\}$) are defined by

$$T_0^F = \otimes_m, \quad T_1^F = \otimes_p, \quad T_\infty^F = \otimes.$$

It was proven in [38] that for Frank t-norms the following holds

$$T_p^F(x, y) + T_{1/p}^F(x, 1 - y) = x.$$

In other words, there are infinitely many pairs of t-norms fulfilling Definition 39. And for such pairs, as the following lemma states, we always get equality (5.12).

Theorem 40 *Let φ, ψ be fuzzy attributes and $\otimes_1$, $\otimes_2$ be an involutively dual pair of t-norms. Let the generalized fuzzy four-fold table $E(\varphi, \psi, \otimes_a, \otimes_b, \otimes_c, \otimes_d, \neg)$ be such that arbitrary two t-norms from $\{\otimes_a, \otimes_b, \otimes_c, \otimes_d\}$ are equal to $\otimes_1$ and the remaining two are equal to $\otimes_2$. Then*

$$a + b + c + d = n. \quad (5.12)$$

Proof. Assume that $\otimes_a = \otimes_b = \otimes_1$ and $\otimes_c = \otimes_d = \otimes_2$ holds. Then the following calculation

$$\begin{aligned}
a+b+c+d &= \textstyle\sum_{o_i \in D} \varphi(o_i) \otimes_1 \psi(o_i) + \sum_{o_i \in D} \varphi(o_i) \otimes_1 \neg\psi(o_i) \\
&+ \textstyle\sum_{o_i \in D} \neg\varphi(o_i) \otimes_2 \psi(o_i) + \sum_{o_i \in D} \neg\varphi(o_i) \otimes_2 \neg\psi(o_i) = \\
&= \textstyle\sum_{o_i \in D} \psi(o_i) + \sum_{o_i \in D} \neg\psi(o_i) \\
&= \textstyle\sum_{o_i \in D} 1 = n
\end{aligned}$$

easily proves (5.12). Similarly for $\otimes_a = \otimes_c = \otimes_1$ and $\otimes_b = \otimes_d = \otimes_2$ we obtain

$$\begin{aligned}
a+b+c+d &= \textstyle\sum_{o_i \in D} \varphi(o_i) \otimes_1 \psi(o_i) + \sum_{o_i \in D} \varphi(o_i) \otimes_2 \neg\psi(o_i) \\
&+ \textstyle\sum_{o_i \in D} \neg\varphi(o_i) \otimes_1 \psi(o_i) + \sum_{o_i \in D} \neg\varphi(o_i) \otimes_2 \neg\psi(o_i) = \\
&= \textstyle\sum_{o_i \in D} \varphi(o_i) + \sum_{o_i \in D} \neg\varphi(o_i) \\
&= \textstyle\sum_{o_i \in D} 1 = n.
\end{aligned}$$

Analogously we deal with the remaining combination. □

The following example shows that not every t-norm must have its involutively dual counterpart.

Example 41 *Consider the drastic t-norm*

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

Assume by (5.11) that $T_{D'}(x,y) = x - T_D(x, 1-y)$ *defines an involutively dual t-norm to the drastic t-norm* T_D*, then it is easy to see that the resulting mapping is not a t-norm, because it is not commutative. Indeed, for arbitrary* $a, b \in (0,1)$

$$T_{D'}(a,b) = a - T_D(a, 1-b) = a - 0 = a,$$

and

$$T_{D'}(b,a) = b - T_D(b, 1-a) = b - 0 = b.$$

This finishes the example.

To finish our discussion on involutively dual pairs of t-norms we mention the result from [29] that the family of Frank t-norms characterizes the whole class of involutively dual pairs of continuous t-norms.

Theorem 42 ([29]) *Let* $\otimes_1$, $\otimes_2$ *be two continuous t-norms. Then they form an involutively dual pair of t-norms if and only if there exists a number* $\lambda \in [0, \infty]$ *and an increasing bijection* $f : [0, 1] \to [0, 1]$ *such that, for all* $x, y \in [0, 1]$, $f(x) + f(1 - x) = 1$ *and*

$$\otimes_1(x, y) = f^{-1}(T^F_\lambda(f(x), f(y))) \text{ and } \otimes_2(x, y) = f^{-1}(T^F_{1/\lambda}(f(x), f(y))).$$

Later we use the following facts in our proofs.

Remark 43 *Note again that* $\otimes_p$ *is an involutively dual t-norm to itself. Thus, for any fuzzy four-fold table* $E(\varphi, \psi, \otimes, \neg)$ *we have (5.12) if* $\otimes = \otimes_p$. *If* $\otimes \leq \otimes_p$ *(see Definition 8) we have* $a+b+c+d \leq n$. *We get the opposite inequality* $a+b+c+d \geq n$ *if* $\otimes \geq \otimes_p$.

Now we prove two generalized versions of Lemma 38 using the notion of involutively dual pairs of t-norms.

Theorem 44 *Let the t-norms* $\otimes_a$ *and* $\otimes_b$ *be involutively dual to each other and* $\otimes_c$ *and* $\otimes_d$ *be arbitrary. Let* $E(\varphi, \psi, \otimes_a, \otimes_b, \otimes_c, \otimes_d, \neg)$ *be a fuzzy four-fold table. Then (5.5)=(5.8), i.e. the following equality holds*

$$\text{conf}_{\otimes_a}(\varphi \rightharpoonup \psi) = \text{conf}_E(\varphi \rightharpoonup \psi) = \frac{a}{a+b}. \tag{5.13}$$

Proof. From Definition 31 we have $a = \sum_{o_i \in D} \varphi(o_i) \otimes_a \psi(o_i)$ and $b = \sum_{o_i \in D} \varphi(o_i) \otimes_b \psi(o_i)$. Because the t-norms $\otimes_a$ and $\otimes_b$ are involutively dual to each other we get

$$\begin{aligned} a + b &= \textstyle\sum_{o_i \in D} \varphi(o_i) \otimes_a \psi(o_i) + \sum_{o_i \in D} \varphi(o_i) \otimes_b \neg\psi(o_i) \\ &= \textstyle\sum_{o_i \in D} \varphi(o_i) \otimes_a \psi(o_i) + \varphi(o_i) \otimes_b (1 - \psi(o_i)) \\ &= \textstyle\sum_{o_i \in D} \varphi(o_i). \end{aligned}$$

Therefore equation (5.13) holds. □

By adding assumptions also on $\otimes_c$ and $\otimes_d$ to be also an involutively dual pair, we obtain equality for different support measures as well.

Theorem 45 *Let the assumptions of Theorem 44 be satisfied. Furthermore, let* $\otimes_c$ *and* $\otimes_d$ *be an involutively dual pair of t-norms. Then (5.3)=(5.6) and (5.4)=(5.7).*

Proof. Analogously to the proof of Theorem 44 we have $a + b = \sum_{o_i \in D} \varphi(o_i)$ and $c + d = \sum_{o_i \in D} \neg\varphi(o_i)$. Consequently, $a + b + c + d = n$. □

5.2.3 Fully true association rules

In this subsection, we examine the so-called *fully true* rules, i.e. rules whose confidence is equal to 1. In this case, we obtain a sort of a transitivity property, which enriches other studies of dependencies among mined fuzzy association rules, see e.g. [35].

Lemma 46 *Let φ, ψ, and χ be fuzzy attributes and $E(\varphi, \psi, \otimes_a, \otimes_b, \otimes_c, \otimes_d, \neg)$, $E'(\psi, \chi, \otimes'_a, \otimes'_b, \otimes'_c, \otimes'_d, \neg)$, and $E''(\varphi, \chi, \otimes''_a, \otimes''_b, \otimes''_c, \otimes''_d, \neg)$ be fuzzy four-fold tables, and $\otimes_b$ and $\otimes'_b$ be strict t-norms. Then*

$$\mathrm{conf}_E(\varphi \rightharpoonup \psi) = \frac{a}{a+b} = \mathrm{conf}_E(\psi \rightharpoonup \chi) = \frac{a'}{a'+b'} = 1$$

implies

$$\mathrm{conf}_E(\varphi \rightharpoonup \chi) = \frac{a''}{a''+b''} = 1.$$

Proof. It follows from $\frac{a}{a+b} = \frac{a'}{a'+b'} = 1$ that $b = b' = 0$ and hence

$$\varphi(o_i) \otimes_b \neg\psi(o_i) = \psi(o_i) \otimes'_b \neg\chi(o_i) = 0 \qquad (5.14)$$

holds for all objects in the data set. Assume, on the contrary that $\mathrm{conf}_E(\varphi \rightharpoonup \chi) < 1$. Then $\varphi(o^*) \otimes''_b \neg\chi(o^*) > 0$ and $\varphi(o^*) > 0$ holds for some object o^*. Consequently, using strictness of $\otimes_b$ and $\otimes'_b$ we get $\psi(o^*) = 1$ and $\chi(o^*) = 1$, i.e., $\neg\chi(o^*) = 0$, because (5.14) holds for all objects in the data set. This contradicts $\varphi(o^*) \otimes''_b \neg\chi(o^*) > 0$. □

Observe that the t-norms $\otimes_a$ and $\otimes_b$ used in E, E', and E'' for calculation of a, b, a', b', a'', b'' may be chosen arbitrarily in each possible case and the proof will go through, because we are using general properties (i.e. axioms) of t-norms. Similarly, also the negations in fuzzy four-fold tables can be chosen arbitrarily, because we only use boundary conditions from the definition of negation.

It was shown in [15] (see also [35]) that the condition $\mathrm{conf}_{\otimes_m}(\varphi \rightharpoonup \psi) = 1$ of fully true rules implies $\varphi(o) \leq \psi(o)$ for all objects in the data set $\mathcal{D}$. In Theorem 47, we show necessary and sufficient conditions for fully true rules. The result is a generalization of [15].

Theorem 47 *Let $\varphi \rightharpoonup \psi$ be a fuzzy association, $E(\varphi, \psi, \otimes_a, \otimes_b, \otimes_c, \otimes_d, \neg)$ be its fuzzy four-fold table and $\otimes_b$ be the Łukasiewicz t-norm $\otimes$. Then*

$$\mathrm{conf}_E(\varphi \rightharpoonup \psi) = 1 \textit{ iff } \varphi(o_i) \leq \psi(o_i)$$

holds for all $o_i \in \mathcal{D}$.

Proof. As above, the equality $\mathrm{conf}_E(\varphi \rightharpoonup \psi) = \frac{a}{a+b} = 1$ is equivalent to $b = 0$, which means that

$$\varphi(o_i) \otimes_b \neg\psi(o_i) = \max\{0, \varphi(o_i) + (1 - \psi(o_i)) - 1\} = 0$$

for all $o_i \in \mathcal{D}$. The last expression is clearly equivalent to $\varphi(o_i) \leq \psi(o_i)$ for all $o_i \in \mathcal{D}$. $\square$

The role of the Łukasiewicz t-norm in the last theorem is crucial. Observe that the sum $\varphi(o_i) \otimes \neg\psi(o_i) = 0$ if and only if $\neg(\varphi(o_i) \otimes \neg\psi(o_i)) = 1$. But $\neg(\varphi(o_i) \otimes \neg\psi(o_i))$ can be rewritten as $\neg\varphi(o_i) \oplus_l \psi(o_i)$, which is the definition of S-implication. It is known that, in the class of continuous t-norms, S-implication is residuated (equal to 1 when $\varphi(o_i) \leq \psi(o_i)$) only in the case of the Łukasiewicz t-norm.

Remark 48 *In the class of discontinuous t-norms, we may use a nilpotent minimum $\otimes_0$ defined by*

$$a \otimes_0 b = \begin{cases} \min(a, b), & \text{if } a + b > 1, \\ 0, & \text{otherwise.} \end{cases}$$

Indeed, $\varphi(o_i) \otimes_0 \neg\psi(o_i) = 0$ is equivalent to $\varphi(o_i) + 1 - \psi(o_i) \leq 1$, which is equivalent to $\varphi(o_i) \leq \psi(o_i)$. Moreover, there is a family of left-continuous t-norms called nilpotent ordinal sums *(see [27]) whose limit cases are the Łukasiewicz t-norm and the nilpotent minimum, and for which Theorem 47 holds as well. But we are convinced that considering discontinuous t-norms is worthless for practical purposes.*

We realize that considering just one t-norm (in the case of continuous t-norms), in Theorem 47, is very restrictive. Instead of the Łukasiewicz t-norm we consider a class T of t-norms for which $x, y \in (0, 1)$ implies $x \otimes y > 0$. In this case we get only the implication from the left to the right.

Lemma 49 *Let $\varphi \rightharpoonup \psi$ be a fuzzy association, $E(\varphi, \psi, \otimes_a, \otimes_b, \otimes_c, \otimes_d, \neg)$ be its fuzzy four-fold table and $\otimes_b \in T$. Then $\mathrm{conf}_E(\varphi \rightharpoonup \psi) = 1$ implies $\varphi(o_i) \leq \psi(o_i)$ for all $o_i \in \mathcal{D}$.*

Proof. As above, $\mathrm{conf}_E(\varphi \rightharpoonup \psi) = \frac{a}{a+b} = 1$ is equivalent to $\varphi(o_i) \otimes_b \neg\psi(o_i) = 0$. This, by definition of the class T, implies either $\varphi(o_i) = 0$ or $\psi(o_i) = 1$ for every object $o_i \in \mathcal{D}$. $\square$

The converse implication ($\varphi(o_i) \leq \psi(o_i)$ for all $o_i \in \mathcal{D}$ implies $\mathrm{conf}_E(\varphi \rightharpoonup \psi) = 1$) is natural and much more interesting.

Lemma 50 *Let $\varphi \rightharpoonup \psi$ be a fuzzy association, $E(\varphi, \psi, \otimes_a, \otimes_b, \otimes_c, \otimes_d, \neg)$ be its fuzzy four-fold table and let $x \otimes_b y = 0$ for all $x, y \in [0,1]$ such that $x + y \leq 1$. Then $\varphi(o_i) \leq \psi(o_i)$ for all $o_i \in \mathcal{D}$ implies $\mathrm{conf}_E(\varphi \rightharpoonup \psi) = 1$.*

Proof. If $\varphi(o_i) \leq \psi(o_i)$ then $\varphi(o_i) + 1 - \psi(o_i) \leq 1$ for all $o_i \in \mathcal{D}$. From assumption about $\otimes_b$ we get $\varphi(o_i) \otimes_b \neg\psi(o_i) = 0$ for all $o_i \in \mathcal{D}$, which means that $\mathrm{conf}_E(\varphi \rightharpoonup \psi) = 1$. □

5.2.4 Lift, conviction, and leverage confirmation measures

We turn our attention now towards three other confirmation measures which are also relatively frequently used besides the support and confidence measures. We discuss their usefulness for different choices of t-norms in fuzzy four-fold tables, and we conclude our discussion by our recommendation of the most appropriate choice.

For a given t-norm $\otimes$, the three confirmation measures we study are the *lift*, *leverage*, and *conviction* measures and are defined as follows

$$\mathrm{lift}_{\otimes}(\varphi \rightharpoonup \psi) = \frac{\mathrm{supp}_{\otimes}(\varphi \rightharpoonup \psi)}{\mathrm{supp}(\varphi) \cdot \mathrm{supp}(\psi)} = n \cdot \frac{\sum_{o \in D} \varphi(o) \otimes \psi(o)}{\sum_{o \in D} \varphi(o) \cdot \sum_{o \in D} \psi(o)}, \tag{5.15}$$

$$\mathrm{lever}_{\otimes}(\varphi \rightharpoonup \psi) = \mathrm{supp}_{\otimes}(\varphi \rightharpoonup \psi) - \mathrm{supp}(\varphi) \cdot \mathrm{supp}(\psi), \tag{5.16}$$

$$\mathrm{conv}_{\otimes}(\varphi \rightharpoonup \psi) = \frac{\mathrm{supp}(\varphi) \cdot \mathrm{supp}(\neg\psi)}{\mathrm{supp}_{\otimes}(\varphi \rightharpoonup \neg\psi)}. \tag{5.17}$$

Justification and motivation for their use as well as their classical (crisp) versions can be found in [8, 10, 7, 51]. For stochastically independent variables we would expect the lift and conviction measures to have values close to 1 and the leverage close to 0, respectively. For positively dependent variables we expect the lift measure to be above 1 and on the other hand, it could be below 1 for negatively dependent variables. Analogous behavior is expected in the case of the leverage and conviction measures. However, this kind of behavior is observed only in the case of the product t-norm. But the measures (5.15)–(5.17) are often used with the help of other t-norms, and this may lead to inappropriate use of those measures.

The problem was recently discussed by Burda in [8], who prepared a short demonstration of this phenomenon in which two vectors of numbers were uniformly generated and their lift, leverage, and conviction were computed. For demonstration purposes, we have repeated the experiment from [8] with analogical results — they are mentioned in Table 5.1, where you can see that indeed the expected behavior of the quality measures is observed only for the product t-norm.

	$\text{lift}_\otimes$	$\text{lever}_\otimes$	$\text{conv}_\otimes$
Łukasiewicz t-norm	0.678	-0.081	1.503
Product t-norm	1.007	0.002	1.007
Minimum t-norm	1.340	0.086	0.759

Table 5.1: Lift, leverage and conviction computed with different t-norms on a stochastically independent data set.

To solve the above issue with lift, leverage, and conviction quality measures, we suggest using analogical definitions based on a fuzzy four-fold table with the restriction that only one kind of t-norm is used. In the following part of this subsection, we propose new definitions of the three quality measures based on fuzzy four-fold table E. We study properties of these definitions, and at the end of this subsection, we repeat the experiment from [8] for our new definitions to demonstrate the change of behavior.

Definition 51 *Let $E(\varphi, \psi, \otimes_a, \otimes_b, \otimes_c, \otimes_d, \neg)$ be a fuzzy four-fold table. Then the* lift, leverage, *and* conviction *quality measures of a rule $\varphi \rightharpoonup \psi$ are defined by*

$$\text{lift}_E(\varphi \rightharpoonup \psi) = \frac{a \cdot (a+b+c+d)}{(a+b) \cdot (a+c)}, \tag{5.18}$$

$$\text{lever}_E(\varphi \rightharpoonup \psi) = \frac{a}{a+b+c+d} - \frac{(a+b) \cdot (a+c)}{(a+b+c+d)^2}, \tag{5.19}$$

$$\text{conv}_E(\varphi \rightharpoonup \psi) = \frac{(a+b) \cdot (b+d)}{b \cdot (a+b+c+d)}. \tag{5.20}$$

In the rest of this subsection, we study the simplest case when fuzzy four-fold table $E(\varphi, \psi, \otimes, \neg)$ is constructed with just one t-norm, i.e. $\otimes = \otimes_a = \otimes_b = \otimes_c = \otimes_d$. The lift measure based on fuzzy four-fold tables has the following properties.

Theorem 52 *Let $E(\varphi, \psi, \otimes, \neg)$ be a fuzzy four-fold table of fuzzy attributes φ and ψ, then the following inequality holds*

$$\frac{\text{supp}_E(\varphi \rightharpoonup \psi)}{\text{supp}_E(\varphi) \otimes_m \text{supp}_E(\psi)} \leq \text{lift}_E(\varphi \rightharpoonup \psi) \leq \frac{\text{supp}_E(\varphi \rightharpoonup \psi)}{\text{supp}_E(\varphi) \otimes \text{supp}_E(\psi)}. \tag{5.21}$$

Moreover,

- *if* $\otimes \leq \otimes_p$ *then*

$$\text{lift}_E(\varphi \rightharpoonup \psi) \geq \frac{a+b+c+d}{n} \cdot \frac{\text{supp}_\otimes(\varphi \rightharpoonup \psi)}{\text{supp}(\varphi)\cdot\text{supp}(\psi)}. \tag{5.22}$$

- *If* $\otimes \geq \otimes_p$ *then*

$$\text{lift}_E(\varphi \rightharpoonup \psi) \leq \frac{a+b+c+d}{n} \cdot \frac{\text{supp}_\otimes(\varphi \rightharpoonup \psi)}{\text{supp}(\varphi)\cdot\text{supp}(\psi)}.$$

- *If* $\otimes = \otimes_p$ *then*

$$\text{lift}_E(\varphi \rightharpoonup \psi) = \frac{\text{supp}_\otimes(\varphi \rightharpoonup \psi)}{\text{supp}(\varphi)\cdot\text{supp}(\psi)}.$$

Proof. From Definition 51 and equations (5.6) and (5.7) we get

$$\begin{aligned}
\text{lift}_E(\varphi \rightharpoonup \psi) &= \frac{a\cdot(a+b+c+d)}{(a+b)\cdot(a+c)} \cdot \frac{a+b+c+d}{a+b+c+d} \\
&= \frac{a}{a+b+c+d} \cdot \frac{1}{\frac{a+b}{a+b+c+d}} \cdot \frac{1}{\frac{a+c}{a+b+c+d}} \\
&= \frac{\text{supp}_E(\varphi \rightharpoonup \psi)}{\text{supp}_E(\varphi)\cdot\text{supp}_E(\psi)}
\end{aligned}$$

and then simply by the ordering of t-norms we have

$$\frac{\text{supp}_E(\varphi \rightharpoonup \psi)}{\text{supp}_E(\varphi)\otimes_m\text{supp}_E(\psi)} \leq \frac{\text{supp}_E(\varphi \rightharpoonup \psi)}{\text{supp}_E(\varphi)\cdot\text{supp}_E(\psi)} \leq \frac{\text{supp}_E(\varphi \rightharpoonup \psi)}{\text{supp}_E(\varphi)\otimes\text{supp}_E(\psi)}.$$

This proves (5.21).

To prove (5.22) we assume $\otimes \leq \otimes_p$ for now. Then we have

$$(a+b) = \left(\sum_{o_i\in D}\varphi(o_i)\otimes\psi(o_i) + \sum_{o_i\in D}\varphi(o_i)\otimes\neg\psi(o_i)\right) \leq \sum_{o_i\in D}\varphi(o_i) \tag{5.23}$$

and

$$(a+c) = \left(\sum_{o_i\in D}\varphi(o_i)\otimes\psi(o_i) + \sum_{o_i\in D}\neg\varphi(o_i)\otimes\psi(o_i)\right) \leq \sum_{o_i\in D}\psi(o_i). \tag{5.24}$$

Consequently,

$$\begin{aligned}
\text{lift}_E(\varphi \rightharpoonup \psi) &= \frac{a\cdot(a+b+c+d)}{(a+b)\cdot(a+c)} \\
&= \frac{(a+b+c+d)\cdot\sum_{o_i\in D}\varphi(o_i)\otimes\psi(o_i)}{(a+b)\cdot(a+c)} \\
&\geq \frac{(a+b+c+d)\cdot\sum_{o_i\in D}\varphi(o_i)\otimes\psi(o_i)}{\sum_{o_i\in D}\varphi(o_i)\cdot\sum_{o_i\in D}\psi(o_i)} \\
&= \frac{(a+b+c+d)}{n}\cdot\frac{\text{supp}_\otimes(\varphi \rightharpoonup \psi)}{\text{supp}(\varphi)\cdot\text{supp}(\psi)}.
\end{aligned}$$

This proves (5.22). The proof of the last two expressions is analogous — see also Remark 43. ☐

The subsequent two theorems show that similar relationships can also be proved for the leverage and conviction measures.

Theorem 53 *Let φ and ψ be fuzzy attributes and $E(\varphi, \psi, \otimes, \neg)$ be a fuzzy four-fold table.*

- *If $\otimes \leq \otimes_p$ then*

$$\text{lever}_E(\varphi \rightharpoonup \psi) \geq \text{supp}_\otimes(\varphi \rightharpoonup \psi) - \left(\frac{n}{a+b+c+d}\right)^2 \cdot \text{supp}(\varphi) \cdot \text{supp}(\psi). \tag{5.25}$$

- *If $\otimes \geq \otimes_p$ then*

$$\text{lever}_E(\varphi \rightharpoonup \psi) \leq \text{supp}_\otimes(\varphi \rightharpoonup \psi) - \left(\frac{n}{a+b+c+d}\right)^2 \cdot \text{supp}(\varphi) \cdot \text{supp}(\psi).$$

- *If $\otimes = \otimes_p$ then*

$$\text{lever}_E(\varphi \rightharpoonup \psi) = \text{supp}_\otimes(\varphi \rightharpoonup \psi) - \text{supp}(\varphi) \cdot \text{supp}(\psi).$$

Proof. First, from $\otimes \leq \otimes_p$ we get $a+b+c+d < n$ and therefore

$$\frac{a}{a+b+c+d} \geq \text{supp}_\otimes(\varphi \rightharpoonup \psi).$$

Also as in the proof of Theorem 52 (see (5.23) and (5.24)) we have

$$(a+b) \leq \sum_{o_i \in D} \varphi(o_i) \qquad \text{and} \qquad (a+c) \leq \sum_{o_i \in D} \psi(o_i).$$

Now a few simple operations applied to (5.19) give

$$\begin{aligned}
\text{lift}_E(\varphi \rightharpoonup \psi) &= \frac{a}{a+b+c+d} - \frac{(a+b)\cdot(a+c)}{(a+b+c+d)^2} \\
&= \frac{a}{a+b+c+d} - \frac{(a+b)\cdot(a+c)}{(a+b+c+d)^2} \cdot \frac{n^2}{n^2} \\
&\geq \text{supp}_\otimes(\varphi \rightharpoonup \psi) - \frac{(a+b)\cdot(a+c)}{(a+b+c+d)^2} \cdot \frac{n^2}{n^2} \\
&\geq \text{supp}_\otimes(\varphi \rightharpoonup \psi) - \frac{\sum_{o_i \in D} \varphi(o_i) \cdot \sum_{o_i \in D} \psi(o_i)}{(a+b+c+d)^2} \cdot \frac{n^2}{n^2} \\
&= \text{supp}_\otimes(\varphi \rightharpoonup \psi) - \left(\frac{n}{a+b+c+d}\right)^2 \cdot \text{supp}(\varphi) \cdot \text{supp}(\psi).
\end{aligned}$$

This finishes the proof of the first expression (5.25). The second expression can be proved analogously. The last expression of this theorem is a consequence of Remark 43. □

Theorem 54 *Let φ and ψ be fuzzy attributes and $E(\varphi, \psi, \otimes, \neg)$ be a fuzzy four-fold table.*

- *If $\otimes \leq \otimes_p$ then*

$$\mathrm{conv}_E(\varphi \rightharpoonup \psi) \leq \frac{n}{a+b+c+d} \cdot \frac{\mathrm{supp}(\varphi) \cdot \mathrm{supp}(\neg\psi)}{\mathrm{supp}_{\otimes}(\varphi \rightharpoonup \neg\psi)}. \tag{5.26}$$

- *If $\otimes \geq \otimes_p$ then*

$$\mathrm{conv}_E(\varphi \rightharpoonup \psi) \geq \frac{n}{a+b+c+d} \cdot \frac{\mathrm{supp}(\varphi) \cdot \mathrm{supp}(\neg\psi)}{\mathrm{supp}_{\otimes}(\varphi \rightharpoonup \neg\psi)}.$$

- *If $\otimes = \otimes_p$ then*

$$\mathrm{conv}_E(\varphi \rightharpoonup \psi) = \frac{\mathrm{supp}(\varphi) \cdot \mathrm{supp}(\neg\psi)}{\mathrm{supp}_{\otimes}(\varphi \rightharpoonup \neg\psi)}.$$

Proof. Again, as in the proof of Theorem 53, for the case when $\otimes \leq \otimes_p$, we can obtain the following

$$\begin{aligned}
\mathrm{conv}_E(\varphi \rightharpoonup \psi) &= \frac{(a+b)\cdot(b+d)}{b\cdot(a+b+c+d)} \\
&\leq \frac{\sum_{o_i \in D} \varphi(o_i) \cdot \sum_{o_i \in D} \neg\psi(o_i)}{b\cdot(a+b+c+d)} \\
&= \frac{\sum_{o_i \in D} \varphi(o_i) \cdot \sum_{o_i \in D} \neg\psi(o_i)}{b\cdot(a+b+c+d)} \cdot \frac{n^2}{n^2} \\
&= \frac{n}{a+b+c+d} \cdot \frac{\mathrm{supp}(\varphi) \cdot \mathrm{supp}(\neg\psi)}{\mathrm{supp}_{\otimes}(\varphi \rightharpoonup \neg\psi)}.
\end{aligned}$$

This proves (5.26). The remaining two expressions can be proven analogously (see again Remark 43). □

One can see that the measures lift_E, lever_E, and conv_E coincide with their counterparts in equations (5.15)–(5.17) only if the t-norm $\otimes$ used in a fuzzy four-fold table $E(\varphi, \psi, \otimes, \neg)$ is the product t-norm.

Furthermore, we have repeated the experiment from the beginning of this section, and the results are in accordance with the expectations this time, see Table 5.2. This

	lift_E	lever_E	conv_E
Łukasiewicz t-norm	1.014	0.003	1.013
Product t-norm	1.007	0.002	1.007
Minimum t-norm	1.007	0.002	1.007

Table 5.2: The lift, leverage and conviction measures based on a fuzzy four-fold table E computed with different t-norms on a stochastically independent data set.

improved behavior of our lift, leverage, and conviction quality measures is related to the fact that we aggregate a, b, c, d in a fuzzy four-fold table E with the same t-norm, which is analogous to what was discussed in Section 5.2.1.

In the light of our theoretical and experimental investigations, we propose that the most appropriate t-norm for above-mentioned quality measures of fuzzy associational rules is the product t-norm. In this case, we aggregate the evidence for and against an association in the same way, and the sum of degrees in fuzzy four-fold table is equal to the number of rows in our data set. Only for this t-norm, we can obtain reasonable and expected behavior.

However, the situation can be different if more than one t-norm is used in the definition of the fuzzy four-fold table $E(\varphi, \psi, \otimes_a, \otimes_b, \otimes_c, \otimes_d, \neg)$. This topic and correct definitions of lift, conviction and leverage measures for fully general fuzzy four-fold tables are left for future work.

5.3 Constructions of new fuzzy quantifiers

Results mentioned in this section are highly motivated by the results obtained by Ivánek in [25, 26]. He proved seven construction theorems for quantifiers defined on classical four-fold tables. With the help of his ideas in the proofs, we generalize his results to fuzzy quantifiers on fuzzy four-fold tables.

5.3.1 From implicational quantifiers to equivalences

It is possible to define a double implicational quantifier from implicational one as the following lemma shows.

Lemma 55 *Let q be an implicational quantifier and let $\otimes : [0,1]^2 \to [0,1]$ be a map nondecreasing in each argument. Then the quantifier $\bar{q}$ constructed by*

$$\bar{q}(a, b, c) := \otimes(q(a, b), q(a, c))$$

is double implicational. If $\otimes$ is a t-norm then $\bar{q}$ satisfies the inequality

$$\bar{q}(a, b, c) \leq \min(q(a, b), q(a, c))$$

for all generalized four-fold tables (4.5).

Proof. This follows directly from the definitions of implicational and double implicational quantifiers and the fact that the map $\otimes$ is nondecreasing in each argument. $\square$

Furthermore, we can get an equivalence from a double implicational quantifier by the following construction.

Lemma 56 *Let q be a double implicational quantifier and $\oplus : [0, 1]^2 \to [0, 1]$ be a map nondecreasing in each argument. Then the quantifier $\tilde{q}$ constructed by*

$$\tilde{q}(a, b, c, d) := \oplus(q(a, b, c), q(d, b, c))$$

is an equivalence. Moreover, if $\oplus$ is a t-conorm, $\tilde{q}$ satisfies the inequality

$$\tilde{q}(a, b, c, d) \geq \max(q(a, b, c), q(d, b, c))$$

for all generalized four-fold tables (4.5).

Proof. Again, from definitions of an equivalence and a double implicational quantifier and the map $\oplus$ being nondecreasing in each argument we get that $\tilde{q}$ is an equivalence. $\square$

We have shown how to construct double implicational quantifiers from implicational quantifiers first and then equivalencies with the help of double implicational quantifiers. The ways of constructing implicational quantifiers themselves are shown in the following subsection.

5.3.2 Fuzzy implications and fuzzy implicational quantifiers

Below we consider *boundary conditions* for implicational quantifiers — $q(\infty, \infty) = q(0, 0) = 1$ and $q(0, \infty) = 0$. To shorten our notation, fuzzy implicational quantifiers can be seen as maps $q : (\mathbb{R}_0^+)^2 \to [0, 1]$, $(a, b) \mapsto e$, satisfying boundary and the monotonicity condition (5.1).

Remark 57 *In the original definition ([25]), every implicational quantifier is a map defined on pairs of natural numbers (i.e. on* $(\mathbb{N}\cup\{0\})^2$*) preserving the monotonicity condition. Our boundary condition (and also dealing with fuzzy attributes) requires an extension of this notion to* $(\mathbb{R}_0^+)^2$*. Boundary conditions are simple natural restrictions coming from the notion of a confidence of an association rule used in data mining.*

Theorem 58 *Let* I *be a fuzzy implication and* $\varphi_1, \varphi_0 : \mathbb{R}_0^+ \to [0,1]$ *be functions such that, for* $i = 1, 2$*,*

(i) φ_i *is nonincreasing,*

(ii) $\varphi_i(0) = 1$*, and*

(iii) $\varphi_i(\infty) = 0$*.*

Then $q_I : \bar{\mathbb{R}}^2 \to [0,1]$ *defined by* $q_I(a,b) := I(\varphi_1(a), \varphi_0(b))$ *is a fuzzy implicational quantifier that fulfills the boundary conditions.*

Proof. First, let us check the boundary conditions. Clearly by (ii) and (iii),

$$q_I(\infty,\infty) := I(\varphi_1(\infty), \varphi_0(\infty)) = I(0,0) = 1,$$
$$q_I(0,0) := I(\varphi_1(0), \varphi_0(0)) = I(1,1) = 1,$$
$$q_I(0,\infty) := I(\varphi_1(0), \varphi_0(\infty)) = I(1,0) = 0.$$

Further, by (i) and the definition of a fuzzy implication, for $a' \geq a, b' \leq b$ we get

$$q_I(a',b') = I(\varphi_1(a'), \varphi_0(b')) \geq I(\varphi_1(a), \varphi_0(b)) = q_I(a,b). \tag{5.27}$$

□

It is also possible to go the other way around and construct fuzzy implications from fuzzy implicational quantifiers. Note that fuzzy implicational quantifiers fulfill boundary conditions automatically due to the construction in the proof of Theorem 58. However, we need to impose the boundary conditions on q for the construction in the following Theorem 59 to work.

Theorem 59 *Let* q *be a fuzzy implicational quantifier fulfilling the boundary conditions and* $\psi_1, \psi_0 : [0,1] \to \mathbb{R}_0^+$ *be functions such that, for* $i = 1, 2$*,* *(i)* ψ_i *is nonincreasing,*

(ii) $\psi_i(1) = 0$*, and*

(iii) $\psi_i(0) = \infty$*.*

Then $I_q : [0, 1]^2 \to [0, 1]$ *defined by* $I_q(a, b) := q(\psi_1(a), \psi_0(b))$ *is a fuzzy implication.*

Proof. Let us check boundary conditions first. Clearly by (ii) and (iii),

$$I_q(0, 0) = q(\psi_1(0), \psi_0(0) = q(\infty, \infty) = 1,$$
$$I_q(1, 1) = q(\psi_1(1), \psi_0(1)) = q(0, 0) = 1,$$
$$I_q(1, 0) = q(\psi_1(1), \psi_0(0)) = q(0, \infty) = 0.$$

Further, by (i) and the definition of fuzzy implication, for $a' \leq a, b' \geq b$ we get

$$I_q(a', b') = q(\psi_1(a'), \psi_0(b')) \geq q(\psi_1(a), \psi_0(b)) = I_q(a, b). \tag{5.28}$$

This finishes the proof. □

Remark 60 *In Theorems above one can simply take any rotation* $\tilde{\varphi} : (\mathbb{R}_0^+)^2 \to [0, 1]^2$ *(or its inverse) for which*

$$\tilde{\varphi}(0, 0) = (1, 1),\ \ \tilde{\varphi}(\infty, 0) = (0, 1),$$

$$\tilde{\varphi}(0, \infty) = (1, 0),\ \ \tilde{\varphi}(\infty, \infty) = (0, 0),$$

and which reverses monotonicity of each argument. Special cases of such rotations are maps in Theorems 58 and 59 above.

5.3.3 Fuzzy ratio-implicational quantifiers

In this subsection, we discuss the class of ratio-implicational quantifiers and as above we introduce construction theorems for them. The proofs in this subsection are essentially similar to Ivánek's proofs in [25]. But we mention them as they are true on a wider class of quantifiers than it was originally proven. Ratio-implicational quantifiers have certain properties summarized in the following lemma.

Lemma 61 *Let* $q(a, b)$ *be a ratio-implicational fuzzy quantifier. Then following properties hold.*

(i) $q(a,b)$ *is implicational,*

(ii) *if* $a'b = ab'$ *then* $q(a',b') = q(a,b)$,

(iii) *there are numbers* m^*, $M^* \in [0,1]$ *such that*

(a) $m^* = q(0,b)$ *for all* $b > 0$,

(b) $M^* = q(a,0)$ *for all* $a > 0$,

(c) $m^* \leq q(a,b) \leq M^*$ *for all* $a, b > 0$,

(iv) *there is a non-decreasing function* f^* *defined on non-negative real numbers and* ∞ *such that*

$$q(a,b) = f^*\left(\frac{a}{b}\right).$$

Proof. The statement (i) follows from definitions, because $a' \geq a, b' \leq b$ implies $a'b \geq ab'$ in $\mathbb{R}_0^+$. The statement (ii) follows straightforwardly from definitions.

Concerning the statement (iii), the first two equalities follow from (ii). The inequalities $m^* \leq q(a,b) \leq M^*$ for all $a, b > 0$ follow from (i). To prove the statement (iv), we define f^* for all $a, b \in (0, \infty)$ as follows

$$f^*(0) = m^*, \quad f^*(\infty) = M^*, \quad f^*\left(\frac{a}{b}\right) = q(a,b).$$

The uniqueness of f^* is guaranteed by (ii), because for $\frac{a}{b} = \frac{a'}{b'}$ we get $ab' = a'b$, i.e. we satisfy the assumption of the statement (ii). The monotonicity of f^* follows from q being implicational quantifier. □

Lemma 61 allows us to prove the following two theorems which link together the ratio-implicational quantifiers, fuzzy implications and ratios $\frac{b}{a+b}$ and $\frac{a}{a+b}$.

Theorem 62 *Let* I *be a fuzzy implication then* $q_I(a,b) = I\left(\frac{b}{a+b}, \frac{a}{a+b}\right)$ *is a ratio-implicational quantifier with* $m^* = 0$ *and* $M^* = 1$.

Proof. If $a'b \geq ab'$ then

$$\frac{b'}{a'+b'} \leq \frac{b}{a+b}, \quad \frac{a'}{a'+b'} \geq \frac{a}{a+b},$$

and

$$q_I(a',b') = I\left(\frac{b'}{a'+b'}, \frac{a'}{a'+b'}\right) \geq I\left(\frac{b}{a+b}, \frac{a}{a+b}\right) = q_I(a,b). \tag{5.29}$$

Furthermore $m^* = q_I(0,b) = I(1,0) = 0$ and $M^* = q_I(a,0) = I(0,1) = 1$. □

Theorem 63 provides an interesting statement saying that every ratio-implicational quantifier, which fulfills the conditions $q(0,b) = 0$ for all $b > 0$ and $q(a,0) = 1$ for all $a > 0$, can be expressed as a fuzzy implication.

Theorem 63 *Let q be a ratio-implicational quantifier with $m^* = 0$ and $M^* = 1$. Then there exists a fuzzy implication I_q such that*

$$q(a,b) = I_q\left(\frac{b}{a+b}, \frac{a}{a+b}\right).$$

Proof. Let f^* be a monotone function from Lemma 61, associated with q. We then define

$$\begin{aligned} I_q(x,y) &= f^*\left(\sqrt{\tfrac{(1-x)y}{x(1-y)}}\right) && \text{for } x,y \in [0,1], x \neq 0, y \neq 1, \\ I_q(0,y) &= I_q(x,1) = 1 && \text{for } x,y \in [0,1]. \end{aligned}$$

Then $I_q(0,0) = I_q(1,1) = 1$ holds by definitions and also $I_q(1,0) = f^*(0) = m^* = 0$. Assume that $0 < x' \leq x$, $1 > y' \geq y$. Consequently, the following three inequalities hold

$$\frac{y'}{x'} \geq \frac{y}{x}, \qquad \frac{1-x'}{1-y'} \geq \frac{1-x}{1-y}, \qquad \frac{(1-x')y'}{x'(1-y')} \geq \frac{(1-x)y}{x(1-y)}.$$

Hence by the monotonicity of f^* and the square root we get

$$I_q(x',y') = f^*\left(\sqrt{\frac{(1-x')y'}{x'(1-y')}}\right) \geq f^*\left(\sqrt{\frac{(1-x)y}{x(1-y)}}\right) = I_q(x,y). \tag{5.30}$$

For $x' = 0$ or $y' = 1$ the inequality also holds because $I_q(x',y') = 1$ by the definition. Finally, for $b = 0$, we obtain $I_q(0,1) = 1 = M^* = q(a,0)$ and for $b > 0$

$$I_q\left(\frac{b}{a+b}, \frac{a}{a+b}\right) = f^*\left(\sqrt{\frac{\left(1-\frac{b}{a+b}\right)\frac{a}{a+b}}{\frac{b}{a+b}\left(1-\frac{a}{a+b}\right)}}\right) = f^*\left(\frac{a}{b}\right) = q(a,b). \tag{5.31}$$

This proves that $I_q\left(\frac{b}{a+b}, \frac{a}{a+b}\right) = q(a,b)$ and finishes the whole proof. □

In this chapter, we elaborated and presented a novel theoretical approach to fuzzy association analysis. We introduced fuzzy four-fold tables, new classes of fuzzy quantifiers and studied properties of the proposed notions. In the first part, we studied fuzzy four-fold tables based on pairs of involutively dual t-norms and properties of three commonly used confirmation measures. In light of the properties of fuzzy four-fold tables and fuzzy confirmation measures, see e.g. Remark 43 and

Theorems 52, 53 and 54, the product t-norm seems to be the best choice for practical purposes. From this part, one natural question has appeared — confirmation measures like lift, coverage, and leverage have been defined with the help of just one t-norm so far. However, for instance a systematic study of their generalizations for the class of involutive dual pairs of t-norms is missing.

In the second part, we have presented a few possible ways of constructing fuzzy quantifiers on fuzzy four-fold tables. Some of the constructions can be useful in post-processing of fuzzy association rules which are mined from data sets. With the help of constructions presented in this paper, we may mine double implications and equivalences in the sense of Subsection 5.1. This direction, in our opinion, indicates that this can be a way how to create a logical model of mined fuzzy associations, which would definitely simplify the mining procedure itself and postprocessing as well. One possible direction is not to consider transitivity (as in Subsection 5.2.3) but fuzzy transitivity instead.

Chapter 6

Applications in regression and prediction

Linguistic descriptions can be applied in regression tasks as was demonstrated in [33]. In this chapter, we continue in this direction, and we investigate the usability of linguistic rules for building an ensemble of regression models. We propose a procedure similar to the one in [60], but some differences due to the application domain appeared. The results and some of the formulations presented in this chapter were already published in [34].

	F_1	F_2	$\dots$	F_m	Y
o_1	f_{11}	f_{21}	$\dots$	f_{m1}	y_1
o_2	f_{12}	f_{22}	$\dots$	f_{m2}	y_2
$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$	$\vdots$
o_n	f_{1n}	f_{2n}	$\dots$	f_{mn}	y_n

Table 6.1: A table representing an initial data matrix $\mathcal{DM}$.

For the simplicity of presentation in this chapter, we now slightly modify the notation that was established in Definition 27, where a data matrix was defined. Instead of denoting i-th variable values as $f_i(o_1, \dots, o_n)$ we denote them $f_{i1}, \dots, f_{in}$ and the whole column as F_i. One of the variables will be denoted as Y, and its values as $y_1, \dots, y_n$, to distinguish it from the rest.

In general, we consider an $(m+1)$-dimensional table of $F_1, \dots, F_m, Y$ variables, which is shown in Table 6.1. Y is our target variable, i.e., we search for a model (function) M, such that $M(F_1, \dots, F_m) = Y$.

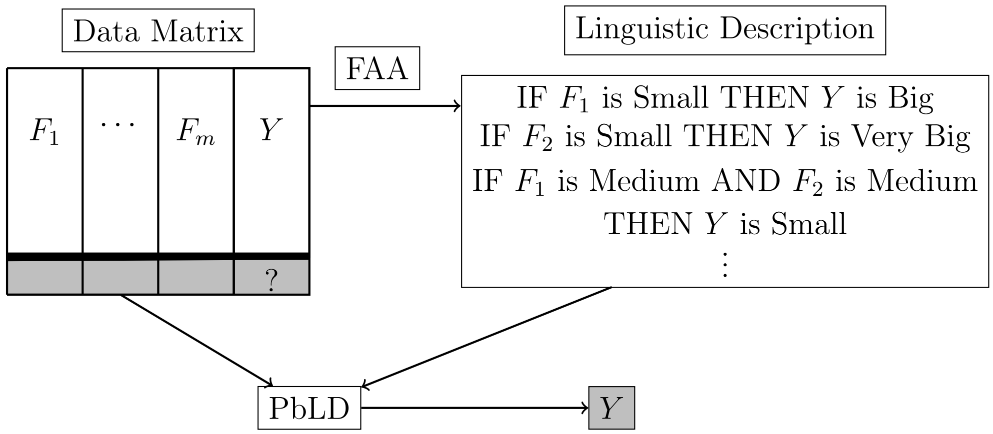

Figure 6.1: Diagram representing the process of performing regression with linguistic rules and Perception-based Logical Deduction.

6.1 Modeling with Linguistic Descriptions

Rule bases (linguistic descriptions) can be extracted from data matrices by various algorithms like OPUS [61], Apriori [2], or a bit string approach [53]. These rules are then used for modeling purposes as described below.

Assume, we are given a data matrix shown in Table 6.1. We then use one of the mentioned algorithms and suitable quality measures (studied in Chapter 5) to search for a set of rules that would describe the relationships between variables $F_1, \dots, F_m$ and the target variable Y. In this way, we obtain a linguistic description from a given data matrix. This process is usually called Fuzzy Associational Analysis (FAA).

After obtaining a new input (gray area of variables $F_1, \dots, F_m$ in a data matrix in Figure 6.1), we take mined Linguistic Description and using PbLD (as was described in Section 3.3.3) we model the unknown part of a target variable Y. The whole process, that was just described, is schematically captured in Figure 6.1.

We slightly mix semantical and syntactical notation in Figure 6.1. We will use the semantical symbol of a variable F in the syntax of linguistic descriptions (see also Equation 6.2). This simplifies the presentation and the meaning will be obvious from the context.

This approach to regression and prediction was successfully applied in many contexts, e.g in flood predictions [12], time series prediction [59] and even in ensembling time series prediction methods [60]. In the next section, we propose its application in regression ensemble modeling.

6.2 Ensemble with linguistic rules

We now use the regression technique described above to produce an ensemble of regression methods.

Let there be r regression models $M_1, \ldots, M_r$ with errors $E_1, \ldots, E_r$:

$$\begin{aligned} Y &= M_1(F_1, \ldots, F_m) + E_1, \\ Y &= M_2(F_1, \ldots, F_m) + E_2, \\ &\vdots \\ Y &= M_r(F_1, \ldots, F_m) + E_r. \end{aligned} \tag{6.1}$$

The models' values for objects $o_1, \ldots, o_n$ are stored as variables in a data matrix and we denote them with the same symbols $M_1, \ldots, M_r$. The regression models are obtained by a usual process of cross-validation [23], while the values of the best model are stored. We treat the models as new columns in our data matrix. Based on the regression models, we define weights variables as

$$W_i = \frac{1}{|Y - M_i(F_1, \ldots, F_m)|},$$

that are also added to the data matrix. Then we get an extended version of the data matrix shown in Table 6.2.

	F_1	F_2	$\ldots$	F_m	Y	M_1	M_2	$\cdots$	M_r	W_1	W_2	$\cdots$	W_r
o_1	f_{11}	f_{21}	$\ldots$	f_{m1}	y_1	m_{11}	m_{21}	$\cdots$	m_{r1}	w_{11}	w_{21}	$\cdots$	w_{r1}
o_2	f_{12}	f_{22}	$\ldots$	f_{m2}	y_2	m_{12}	m_{22}	$\cdots$	m_{r2}	w_{12}	w_{22}	$\cdots$	w_{r2}
$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
o_n	f_{1n}	f_{2n}	$\ldots$	f_{mn}	y_n	m_{1n}	m_{2n}	$\cdots$	m_{rn}	w_{1n}	w_{2n}	$\cdots$	w_{rn}

Table 6.2: A table representing an extended data matrix with regression models with their weights.

Similarly, to the construction of r regression models, we divide the data matrix into learning and testing parts. We then mine linguistic rules from the learning part of data matrix of the following form

$$\textbf{IF } (F_i \text{ is } \mathcal{A}_i \textbf{ AND } M_j \text{ is } \mathcal{A}_j \ldots) \textbf{ THEN } W_i \text{ is } \mathcal{B}_i. \tag{6.2}$$

We mine set of rules $\mathcal{R}_i$ for every weight variable W_i. We are then given testing part of data matrix $F_1, F_2, \ldots, F_m$ and regression models variables $M_1, M_2, \ldots, M_r$ and with the rules $\mathcal{R}_i$ we predict weights W_i in testing part of a data matrix using the PbLD inference method described in Section 3.3.3. From the weights and regression models we build two ensemble models *ens* and *ensMax* as, $i = 1, 2, \dots, N$,

$$ens_i = \frac{\sum_j w_{ji} \cdot m_{ji}}{\sum_j w_{ji}},$$

and

$$ensMax_i = m_{ji}, \quad \text{where } j = \arg\max_{j \in \{1,\dots,r\}} \{w_{ji}\}.$$

The first ensemble model *ens* is a weighted average, but its weights are dynamically changing based on an input and mined rules. The ensemble model *ensMax* dynamically chooses the best model M_j (i.e., with the highest weight) again based on an input and mined rules.

6.3 Experiments

For testing the proposed ensemble methods from the previous section, we have downloaded several data files from the UCI machine learning repository [3], of which some had multiple target variables. Overall, we performed twenty-nine regression tasks. The methods we used for our ensemble learning were Linear Regression (LR), Multivariate Adaptive Regression Splines (MARS), Support Vector Machines with Polynomial Kernel (SVM), k-nearest neighbor (KNN), regression trees (RT) via caret package [30]. With these methods we obtained five models $M_1, \dots, M_5$. For mining of rules and inference with the rules, we used the implementation in Linguistic Fuzzy Logic package [11].

For every data matrix, we mined five sets of rules $\mathcal{R}_1, \dots, \mathcal{R}_5$, which formed the linguistic model of weights $W_1, \dots, W_5$. We predicted the weights on testing parts of data matrices and created the ensemble models described in the previous section. We compared our ensemble methods with a simple Mean of models $M_1, \dots, M_5$ as a baseline.

We computed root mean squared errors (RMSE) on every data matrix for all five models $M_1, \dots, M_5$ and three ensemble models (Mean, *ens*, and *ensMax*). For s testing rows the RMSE is calculated as

$$\text{RMSE}(\hat{y}, y) = \sqrt{\frac{1}{s} \sum_{i=1}^{s} (\hat{y}_i - y_i)^2},$$

where $\hat{y}$ are values given by a model and y are true values of target variable Y. Based on the RMSE we calculated the ranking $(1, \dots, 8)$ of all eight models based on RMSE, i.e., the model with the lowest RMSE was assigned rank 1 and the model with the highest RMSE was assigned the lowest rank 8. The target variables were standardized and therefore the averaging of RMSE through all data matrices is reasonable.

We calculated the average ranking and average RMSE of every method through all the data matrices and results are shown in Table 6.3. Both our ensemble methods

	LR	SVM	MARS	KNN	RT	Mean	*ensMax*	*ens*
ranking	6.310	4.621	3.621	6.310	5.828	2.724	3.138	3.448
RMSE	12.151	9.626	9.637	10.373	10.476	9.341	9.247	9.249

Table 6.3: Average ranking and root mean square errors for multiple regressions.

ensMax and *ens* have the best average RMSE, 9.247 and 9.249 respectively. They are not always the best as the average ranking of the simple mean (2.724) is below the average ranking of our ensemble methods.

The models $M_1, \dots, M_5$ were quite correlated. The median of average correlations of models on all data matrices was 0.67. So we divided the data matrices into two groups. The first one was with the average correlation between models above 0.7 and the second was with the average correlation between models below 0.7. In the higher correlation scenario (see Table 6.4) our proposed ensemble method *ens* was the best with respect to other models in both ranking and RMSE. In the sce-

	LR	SVM	MARS	KNN	RT	Mean	*ensMax*	*ens*
ranking	7.600	4.333	3.200	5.667	5.600	3.400	3.267	2.933
RMSE	13.989	10.136	10.027	11.326	11.462	9.722	9.447	9.427

Table 6.4: Average ranking and root mean square errors for multiple regressions for data matrices with average correlation between regression models *above* 0.7.

nario of less correlated models the performance of our ensemble methods is slightly worse, see Table 6.5. It seems that in lower correlation scenario it is better to simply average the models.

Our ensemble approach, through mining linguistic descriptions, has an advantage, that we can study the relationship between the models (LR, SVM, MARS, KNN, and RT) more deeply. In the remaining part, we present two examples of

	LR	SVM	MARS	KNN	RT	Mean	*ensMax*	*ens*
ranking	4.929	4.929	4.071	7.000	6.071	2.000	3.000	4.000
RMSE	10.182	9.080	9.219	9.352	9.420	8.933	9.033	9.058

Table 6.5: Average ranking and root mean square errors for multiple regressions for data matrices with average correlation between regression models *below* 0.7.

the mined rules of the form (6.2) stated in natural language, which we used for prediction of model weights in our ensemble models.

A rule

IF F_1 is *small* AND MARS is *medium* THEN W_1 is *significantly small*,

states that when the first variable in a data matrix has *small* values and MARS model is modeling *medium* values then the weight W_1 (LR weight) is supposed to be *significantly small*. In other words, the Linear Regression behaves poorly when F_1 has *small* values, and Multivariate Adaptive Regression Splines have *medium* values. Another rule

IF F_1 is *medium* AND F_2 is *small* THEN W_5 is *big*,

states that if the first variable in a data matrix has *medium* values and the second variable has *small* values then the weight W_5 (RT weight) is *big*.

Beware, that such rules are not valid for all data matrices, but only for one particular data matrix. Notice that different linguistic descriptions are usually obtained for the different data matrices.

One can see all root mean squared errors for all models and all data matrices in Table 6.6. The proposed ensemble methods *ens* and *ensMax* behave in a stable manner. They are never the worst with respect to RMSE statistic and often one of them is the best method for particular data matrix, e.g., data matrix 1 and 16.

We proposed an approach to ensemble regression models with linguistic rules. We tested our proposed approach on twenty-nine different regression tasks, and we observed that our proposed method behaves very well in situations when the models in the ensemble are highly correlated, i.e. the average correlation between models is above 0.7 - see Table 6.4.

The mining of IF-THEN rules sometimes leads to vast amounts of rules. In the next chapter, we discuss how to reduce such vast rule bases so the resulting models are not overfitting and the rule bases are better interpretable.

	LR	SVM	MARS	KNN	RT	Mean	*ensMax*	*ens*
data matrix 1	10.797	7.579	7.461	8.534	7.815	7.194	7.072	7.087
data matrix 2	8.351	7.522	8.214	8.330	7.553	6.119	6.503	6.506
data matrix 3	3.163	1.086	0.808	5.682	6.493	2.758	1.008	1.018
data matrix 4	12.819	9.795	8.574	8.897	9.265	8.814	8.950	8.945
data matrix 5	13.057	8.043	7.270	9.990	8.371	7.958	7.472	7.432
data matrix 6	26.483	24.427	22.064	26.654	24.870	22.660	23.150	23.140
data matrix 7	23.084	21.754	19.653	24.736	23.069	20.352	20.662	20.534
data matrix 8	6.618	2.347	2.426	9.368	6.073	3.871	2.548	2.513
data matrix 9	7.928	5.871	1.752	2.304	4.986	3.213	1.867	1.873
data matrix 10	8.648	7.020	4.467	5.030	6.289	5.222	6.026	6.024
data matrix 11	12.482	12.732	11.806	12.702	12.088	11.878	12.233	12.252
data matrix 12	6.918	6.980	7.008	6.973	7.377	6.912	6.966	6.972
data matrix 13	8.632	8.589	8.441	8.790	8.538	8.322	8.371	8.397
data matrix 14	7.767	7.903	7.781	7.976	7.813	7.687	7.702	7.712
data matrix 15	9.940	10.301	9.710	10.131	10.127	9.731	9.995	10.017
data matrix 16	5.430	5.412	5.439	5.569	5.816	5.444	5.426	5.426
data matrix 17	14.852	15.217	14.866	15.442	14.842	14.692	14.799	14.829
data matrix 18	14.228	14.463	13.109	14.629	13.540	13.435	13.873	13.896
data matrix 19	5.711	5.746	5.699	5.927	5.837	5.701	5.716	5.717
data matrix 20	6.214	6.240	6.208	6.399	6.362	6.106	6.180	6.183
data matrix 21	5.374	5.323	5.328	5.461	5.659	5.261	5.331	5.338
data matrix 22	5.951	5.991	6.059	6.164	6.111	5.913	5.995	6.000
data matrix 23	14.977	14.962	14.271	16.299	13.904	14.105	14.041	14.028
data matrix 24	16.035	8.434	13.892	8.850	12.699	9.007	9.439	9.395
data matrix 25	19.746	9.084	14.108	10.276	9.276	9.854	9.388	9.368
data matrix 26	21.421	9.668	12.073	10.407	12.802	10.316	10.236	10.413
data matrix 27	18.575	7.799	9.463	9.232	14.370	9.902	8.415	8.375
data matrix 28	17.633	12.562	15.539	14.353	14.964	13.666	13.643	13.659
data matrix 29	19.552	16.313	15.980	15.711	16.894	14.804	15.159	15.174

Table 6.6: RMSE statistics for multiple data matrices.

Chapter 7

Reduction of rule bases

Association rule mining, in general, may lead to vast amounts of rules/associations, which leads to overfitting of the model captured in the mined rules and the loss of interpretability. To get rid of the overfitting, a method for reducing the rules based on *data coverage* measure (see below Definition 64) was proposed in [13]. However, the data coverage measure, proposed in [13], does not take into account the probability distribution of data. We therefore propose an alternative definition of data coverage called *probabilistic* (Equation 7.2). The new measure captures rather probability information instead of fuzziness. However, this allows to describe exactly how often we will encounter a situation of getting some input when no rule in the linguistic description fires. This was not possible with previous data coverage from [13]. We propose a new rule pruning method based on probabilistic data coverage and conditional firing of rules in PbLD (local perception of PbLD — see Equation 3.26). In our experiments, the rule pruning method leads to smaller and more interpretable rule bases and also to improved precision. The results described here were already published in [57].

Below, we recall the definition of data coverage from [13] and propose new probabilistic data coverage. Then we describe a reduction method based on probabilistic data coverage and in Section 7.3 we describe our experimental comparison of two reduction methods based on simple data coverage and probabilistic data coverage.

7.1 Data coverage of a rule base

In this section, we define two *measures of fuzzy rule bases*. Our aim is to measure the quality of the whole rule bases not only the quality of a particular rule that we have studied in Chapter 5.

7.1.1 Data coverage

The data coverage of a rule base LD was proposed in [13]. We recall its definition here.

Definition 64 *Let $\mathcal{DM} = \langle D, F \rangle$ be a data matrix and $LD = \{\mathcal{R}_1, \dots, \mathcal{R}_k\}$ be a rule base, where $D = \{o_1, \dots, o_n\}$ is a set of objects of $\mathcal{DM}$ and a rule $\mathcal{R} \in LD$ is of the form $\varphi \rightharpoonup \psi$, where φ and ψ are fuzzy attributes. Then the coverage of a data matrix $\mathcal{DM}$ by the rule base LD is given as*

$$\mathrm{cov}_D = \frac{1}{n} \sum_{j=1}^{n} \bigvee_{i=1}^{k} \varphi_i(o_j). \tag{7.1}$$

We recall some properties of data coverage mentioned in [13].

Proposition 65 *Let D be a set of objects of some data matrix $\mathcal{DM}$ and φ, ψ, and χ be fuzzy attributes, then the following holds:*

1. $\mathrm{cov}_D(LD) \in [0, 1]$
2. $\mathrm{cov}_D(\emptyset) = 0$
3. $LD' \subseteq LD$ *then* $\mathrm{cov}_D(LD') \leq \mathrm{cov}_D(LD)$
4. *If* $\mathcal{S}, \mathcal{T} \in LD$ *such that* $\mathcal{S} = \varphi \rightharpoonup \chi$, $\mathcal{T} = \psi \rightharpoonup \chi$ *and* $\varphi \subset \psi$, *then* $\mathrm{cov}_D(LD \setminus \{\mathcal{S}\}) = \mathrm{cov}_D(LD)$
5. *Let* $\mathcal{S} = \varphi \rightharpoonup \psi$ *then* $\mathrm{supp}_{\otimes}(\varphi) = \mathrm{cov}_D(\{\mathcal{S}\})$

All five properties are direct consequences of equation (7.1).

A mountain climbing style of the algorithm based on cov_D for reducing rule bases was proposed in [13] and was also implemented in lfl: Linguistic Fuzzy Logic package [9].

However, we disregard the underlying joint probability distribution of the variables in our data matrix, when the rules are reduced only based on $\mathrm{cov}_{\mathcal{O}}$. For that reason, we propose *probabilistic data coverage* in the next subsection.

7.1.2 Probabilistic data coverage

If we have m columns $f_1, \ldots, f_m$ in a data matrix $\mathcal{DM} = \langle D, F = \{f_1, \ldots, f_m\}\rangle$ (see Definition 27). The columns are, in general, distributed in $\mathbb{R}^m$. The rows in $\mathcal{DM}$ are sampled from an unknown joint probability distribution.

We want to estimate a probability of firing a rule from a rule base LD on a random input. In PbLD the fired rules consist of most specific rules (see equation (3.25)). Let $\iota \in I = \mathbb{R}^m$ represent an input vector from an (infinite) space $I = \mathbb{R}^m$ of all possible inputs and let $f_D(\iota)$ represents a probability density function on I. (Obviously, $D \subseteq I$.) The rule base LD splits the input space I into two parts: I^+ denoting inputs, for which there is a rule in LD that fires, and I^- denoting the rest. Thus $I = I^+ \cup I^-$ and $I^+ \cap I^- = \emptyset$. The probability that at least one rule from a rule base will fire is equal to

$$P(I^+) = \int_{I^+} f_D(\iota)\mathrm{d}\iota.$$

However, we do not know f_D and we estimate the probability $P(I^+)$ with probabilistic coverage.

Definition 66 *Let $\mathcal{DM} = \langle D, F\rangle$ be a data matrix and $LD = \{\mathcal{R}_1, \ldots, \mathcal{R}_k\}$ be a rule base, where $D = \{o_1, \ldots, o_n\}$ is a set of objects of $\mathcal{DM}$ and a rule $\mathcal{R} \in LD$ is of the form $\varphi \rightharpoonup \psi$, where φ and ψ are fuzzy attributes. Then the* probabilistic coverage *of a data matrix $\mathcal{DM}$ by the rule base LD is given as*

$$\mathrm{cov}_{\mathcal{P}}(LD) = \frac{1}{n}\#\{o \,|\, o \in D,\ \varphi_i(o) > 0,\ i \in \hat{k}\}, \tag{7.2}$$

where k is the number of rules in LD and $\#\{\ldots\}$ denotes the number of elements in a set.

Both equations (7.1) and (7.2) describe somehow the quality of a rule base LD with respect to a data matrix $\mathcal{DM}$. When comparing two rule bases LD_1 and LD_2, it can easily happen that the coverage of data by the first rule base is smaller $\mathrm{cov}_D(LD_1) < \mathrm{cov}_D(LD_2)$, but the probabilistic coverage evaluates them the other way around $\mathrm{cov}_{\mathcal{P}}(LD_1) > \mathrm{cov}_{\mathcal{P}}(LD_2)$, because the equation (7.1) does not take into account whether the point is an outlier or from the dense part of the data matrix. Thus the rule base LD_2, though having better data coverage, will fire less frequently and will more frequently assign to the input most typical value (e.g. mean). We think that this leads to worse regression at the end. Therefore, we want to optimize the rule sets with respect to the probabilistic data coverage.

We briefly mention some of the probabilistic data coverage properties and its relationships to cov_D.

Proposition 67 *Let* cov_D *(resp.* $\mathrm{cov}_{\mathcal{P}}$*) be defined as in equation (7.1) (resp. equation (7.2)). Then the following holds:*

- $\mathrm{cov}_D(LD) = 1$ *implies* $\mathrm{cov}_{\mathcal{P}}(LD) = 1$
- $\mathrm{cov}_{\mathcal{P}}(LD) = 1$ *implies* $\mathrm{cov}_D(LD) > O$

Proof. Direct consequence of Equations (7.1) and (7.2). □

As Proposition 67 states, it is theoretically possible for $\mathrm{cov}_{\mathcal{O}}$ to be close to 0, while $\mathrm{cov}_{\mathcal{P}}$ is equal to one. However, that would mean that the fuzzy sets defined on the variables are of very unusual "flat shapes" and will not occur in practical applications.

On the other hand, $\mathrm{cov}_{\mathcal{P}}$ has almost all the properties mentioned in Proposition 65.

Proposition 68 *Let* $\mathrm{cov}_{\mathcal{P}}$ *be defined as in equation (7.2) and* φ *be a fuzzy attribute then first four properties for* cov_D *from Proposition 65 hold also for* $\mathrm{cov}_{\mathcal{P}}$ *and the fifth property has the form*

5' Let $\mathcal{S} = \varphi \rightharpoonup \psi$ *then* $\frac{1}{n}\#\{o|o \in D,\ \varphi(o) > 0\} = \mathrm{cov}_{\mathcal{P}}(\{\mathcal{S}\})$

Proof. Simply follows from Equation (7.2). □

7.2 Rules reduction method

The condition for the firing of a rule in implicative expert systems and Mamdani style expert systems after receiving input u is $\varphi(u) > 0$. See [4, 5] for implicative systems and [37] for Mamdani fuzzy expert system. But in the case of PbLD, the situation is slightly more complicated as the partial ordering of the antecedents plays an important role in rules selection. Therefore, we propose the following procedure for pruning the knowledge base of a system.

Let LD be a knowledge base we want to reduce and LD' its reduced version, then the *reduction ratio* for data coverage is defined as

$$\frac{\mathrm{cov}_D(LD')}{\mathrm{cov}_D(LD)}.$$

Similarly for the probabilistic data coverage.

At the beginning of rule pruning, we decide on the reduction ratio we want to achieve. Then we sample a set of objects from the training data. Then for each sampled object o_i we infer $P^{LD}(o_i)$. We create a new knowledge base LD' from the rules that fired at least once. LD' is already significantly reduced but still with $\mathrm{cov}_D(LD') = 1$ (on the sampled data). We count how many times each rule from the knowledge base was fired and then, based on the relative frequencies, we further discard rules from LD' till the desired reduction ratio is reached.

There are two strategies applicable in the final discarding of rules and that is ascending and descending, i.e. discarding first the least probable rules to be fired or the most likely rules. If there are too many rules mined, it often happens that a higher amount of rules fires at once and thus it is possible to discard rules with the highest probabilistic data coverage. On the other hand, when there are fewer rules, the strategy of discarding first the rules with the lowest probabilistic data coverage leads to a better solution.

The pruning procedure just described is written in a more formal way as a pseudocode in Algorithm 1.

7.3 Experiments

For an experimental comparison of the reduction method based on $\mathrm{cov}_{\mathcal{O}}$ and published in [13] and our method proposed in Section 7.2, we used five different data sets from UCI Machine Learning Repository [3]. We randomly divided the data sets into test data and learning data with ratio 1:9. We mined the learning data for fuzzy associations with thresholds set as $\mathrm{supp}_t = 0.2$ and $\mathrm{conf}_t = 0.9$. Thus, we got a knowledge base LD which we used in PbLD and we calculated the resulting root

Algorithm 1 Reduction of a rule base driven by the rule base probabilistic data coverage

Inputs: n sampled set of rows/objects $\mathcal{O}$, rule base LD with m rules, probability threshold p
Output: reduced rule base LD'

$LD' \leftarrow \emptyset$
fired[] $\leftarrow \overbrace{[0, \ldots, 0]}^{m}$
for each sampled row/object $o \in \mathcal{O}$ **do**
    infer $P^{LD}(o)$
    for each rule $\mathcal{R}_i \in LD$ **do**
        if $\mathcal{R}_i \in P^{LD}(o)$ **then**
            fired[i] $\leftarrow$ fired[i]+1
        end if
    end for
end for
$LD' \leftarrow \cup_{o \in \mathcal{O}} P^{LD}(o)$
while $\mathrm{cov}_{\mathcal{P}}(LD') > p$ **do**
    $LD' \leftarrow LD' \setminus \mathcal{R}_i$,
     (ascending/descending according to fired[])
end while

mean square error from the output r given by PbLD and test data y:

$$\mathrm{RMSE}(r, y) = \sqrt{\frac{1}{s}\sum_{i=1}^{s}(r_i - y_i)^2}$$

The numbers of the mined rules for each data set are in column FAA in rows denoted *number* in Table 7.1. Similarly, root mean squared errors for every set of rules and each data set are written in rows denoted *error* in Table 7.1.

We used the same knowledge base LD in the reduction methods based on both coverages. We repeated the reductions with ratios 1, 0.99, 0.95 and 0.9. The resulting eight numbers of rules and their respective errors when used in PbLD are also in Table 7.1 for each data set we used.

The reductions with the best(smallest) root mean square errors are highlighted for both cov_D and $\mathrm{cov}_{\mathcal{P}}$. Note that only comparisons on the line make sense as the regressed variables have different ranges in each data set. Reduction with $\mathrm{cov}_{\mathcal{P}}$ always leads to the best or almost the best regression error.

In the case of automobile and yacht data, a very interesting phenomenon occurs

Dataset		FAA	cov_D				$\mathrm{cov}_{\mathcal{P}}$			
			0.9	0.95	0.99	1	0.9	0.95	0.99	1
auto-mpg	number	196	9	11	**18**	24	3	6	15	**32**
	error	4.78	5.04	5	**4.83**	4.83	8.42	8.42	6.09	**4.78**
automobile	number	2577	32	49	87	**140**	**139**	207	203	533
	error	8209	10688	10913	8450	**8431**	**5507**	5593	5614	8123
housing	number	2883	20	32	58	**110**	17	63	114	**481**
	error	8.06	8.41	8.38	8.13	**8.02**	10.57	9.5	11.05	**8.06**
airfoil	number	172	**6**	10	13	25	**22**	22	31	33
	error	8.02	**8.02**	8.02	8.02	8.02	**8.02**	8.02	8.02	8.02
yacht	number	110	15	16	**19**	20	3	7	8	**8**
	error	8.94	14.4	12.99	**12.41**	12.41	17.84	10.47	10.36	**10.36**

Table 7.1: Results of reductions of rules based on cov_D and $\mathrm{cov}_{\mathcal{P}}$

as the $\mathrm{cov}_{\mathcal{P}}$ leads to a significantly better regression with a smaller amount of rules than the $\mathrm{cov}_{\mathcal{O}}$. Furthermore, in the case of automobile data, the method based on $\mathrm{cov}_{\mathcal{P}}$ was able to reduce the rules in such a way as to obtain a smaller regression error than the original rules mined by fuzzy association analysis. This is a very important result as it clearly shows that both cov_D and $\mathrm{cov}_{\mathcal{P}}$ capture very different aspects of knowledge bases with respect to the data from which they were mined.

Obviously, both methods based on cov_D and $\mathrm{cov}_{\mathcal{P}}$ suffer from local optima, but they are somehow complementary and useful to be used in tandem.

Based on the probabilistic measure and also on the conditional firing of rules in PbLD, a new method for optimizing the knowledge base of PbLD was created. The method works against overfitting of the model produced with the fuzzy associational analysis.

Chapter 8

Conclusion

The main goal of the thesis was to join the Fuzzy natural logic, which consists besides other theories of theory of evaluative expression and Perception-based Logical Deduction, together with Observational Calculus. This was done in Chapter 4 by definition of the logical calculus of fuzzy association rules.

We investigated thoroughly the models of the logical calculus of fuzzy association rules, i.e. a structure $\mathcal{M} = \langle \mathcal{DM}, W_m, \bigotimes, \bigotimes_E, Q \rangle$. We concentrated our investigation on subparts of the models, i.e. $\bigotimes_E$ and Q. We showed many relations between generalized fuzzy four-fold table $E(\varphi, \psi, \otimes_a, \otimes_b, \otimes_c, \otimes_d, \neg)$ and various quality measures used by fuzzy associational practitioners. From our theoretical investigation, we derived also practical advice for the choice of t-norms in $\bigotimes$ and $\bigotimes_E$.

In Chapter 6, we proposed a new regression ensemble technique based on Perception-based Logical Deduction and fuzzy association analysis. We showed how is it possible to use fuzzy association mining as a technique for combining multiple regression models into a single ensemble model. The advantage of this technique is that the ensemble model consists of linguistics descriptions that describe in natural language relationships between the models and data itself.

We empirically tested our approach on multiple data sets and found out that our ensemble model achieves higher precision also in the case when the models in the ensemble are highly correlated.

Fuzzy associational analysis very often generates too many rules. As a result, the linguistic descriptions in PbLD might be too long and the whole system loses interpretability. We proposed a probabilistic coverage measure in Chapter 7 and based on it we proposed an algorithm for reducing the number of rules in linguistic descriptions. We tested our algorithm on multiple data sets and compared it with previously proposed measure and its algorithm. Our algorithm is able to find smaller linguistic descriptions while increasing the precision of the regression of the whole model.

This thesis contributed to both theoretical and applied research in fuzzy associational analysis.

Bibliography

[1] R. Agrawal, T. Imieliński, A. Swami, Mining association rules between sets of items in large databases, Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data (1993) 207–216.

[2] R. Agrawal, R. Srikant, Fast algorithms for mining association rules in large databases, in: Proceedings of the 20th International Conference on Very Large Data Bases, VLDB '94, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1994, pp. 487–499.

[3] K. Bache, M. Lichman, UCI machine learning repository (2013).
URL `http://archive.ics.uci.edu/ml`

[4] J. Baldwin, N. Guild, Modelling controllers using fuzzy relations, Kybernetes 9 (3) (1980) 223–229.

[5] U. Bodenhofer, M. Daňková, M. Štěpnička, V. Novák, A plea for the usefulness of the deductive interpretation of fuzzy rules in engineering applications, in: 2007 IEEE International Fuzzy Systems Conference, 2007, pp. 1–6.

[6] B. Bouchon-Meunier, Y. Jia, Linguistic modifiers and imprecise categories, International Journal of Intelligent Systems 7 (1) (1992) 25–36.

[7] A. L. Buczak, C. M. Gifford, Fuzzy association rule mining for community crime pattern discovery, in: ACM SIGKDD Workshop on Intelligence and Security Informatics, ACM, 2010, p. 2.

[8] M. Burda, Interest measures for fuzzy association rules based on expectations of independence, Advances in Fuzzy Systems 2014 (2014) 2.

[9] M. Burda, lfl: Linguistic Fuzzy Logic (R package on CRAN), r package version 1.2 (2015).

[10] M. Burda, Lift measure for fuzzy association rules, in: P. Grzegorzewski, M. Gagolewski, O. Hryniewicz, M. Á. Gil (Eds.), Strengthening Links Between Data Analysis and Soft Computing, Springer International Publishing, 2015, pp. 249–260.
URL http://dx.doi.org/10.1007/978-3-319-10765-3_30

[11] M. Burda, Linguistic fuzzy logic in R, in: IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), IEEE, Istanbul, 2015, pp. 1–7.

[12] M. Burda, P. Rusnok, M. Štěpnička, Mining linguistic associations for emergent flood prediction adjustment, Advances in Fuzzy Systems (2013) 4.

[13] M. Burda, M. Štěpnička, Reduction of fuzzy rule bases driven by the coverage of training data, in: Proc. 16th World Congress of the International Fuzzy Systems Association and 9th Conference of the European Society for Fuzzy Logic and Technology, IFSA-EUSFLAT, 2015, pp. 463–470.

[14] R. Carnap, Meaning and necessity: a study in semantics and modal logic, University of Chicago Press, 1988.

[15] D. Dubois, E. Hüllermeier, H. Prade, A systematic approach to the assessment of fuzzy association rules, Data Mining and Knowledge Discovery 13 (2) (2006) 167–192.

[16] A. Dvořák, M. Holčapek, Type $\langle 1, 1\rangle$ fuzzy quantifiers determined by fuzzy measures on residuated lattices. part i. basic definitions and examples, Fuzzy Sets and Systems 242 (2014) 31–55.

[17] A. Dvořák, M. Štěpnička, L. Štěpničková, On redundancies in systems of fuzzy/linguistic if–then rules under perception-based logical deduction inference, Fuzzy Sets and Systems.

[18] L. Geng, H. J. Hamilton, Interestingness measures for data mining: A survey, ACM Computing Surveys (CSUR) 38 (3) (2006) 9.

[19] D. H. Glass, Fuzzy confirmation measures, Fuzzy Sets and Systems 159 (4) (2008) 475–490.

[20] P. Hájek, I. Havel, M. Chytil, The guha method of automatic hypotheses determination, Computing 1 (4) (1966) 293–308.

[21] P. Hájek, T. Havránek, Mechanizing Hypothesis Formation (Mathematical Foundations for a General Theory), Springer-Verlag, 1978.

[22] P. Hájek, M. Holeňa, J. Rauch, The guha method and its meaning for data mining, Journal of Computer and System Sciences 76 (1) (2010) 34–48.

[23] J. Han, Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2005.

[24] E. Hullermeier, Y. Yi, In defense of fuzzy association analysis, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 37 (4) (2007) 1039–1043.

[25] J. Ivánek, Construction of implicational quantifiers from fuzzy implications, Fuzzy sets and systems 151 (2) (2005) 381–391.

[26] J. Ivánek, Combining implicational quantifiers for equivalence ones by fuzzy connectives, International journal of intelligent systems 21 (3) (2006) 325–334.

[27] S. Jenei, New family of triangular norms via contrapositive symmetrization of residuated implications, Fuzzy Sets and Systems 110 (2) (2000) 157–174.

[28] J. Kacprzyk, R. R. Yager, S. Zadrożny, A fuzzy logic based approach to linguistic summaries of databases, International Journal of Applied Mathematics and Computer Science 10 (4) (2000) 813–834.

[29] E. P. Klement, R. Mesiar, E. Pap, Invariant copulas, Kybernetika 38 (3) (2002) 275–286.

[30] M. Kuhn, J. Wing, S. Weston, A. Williams, C. Keefer, A. Engelhardt, T. Cooper, Z. Mayer, B. Kenkel, the R Core Team, M. Benesty, R. Lescarbeau, A. Ziem, L. Scrucca, Y. Tang, C. Candan, T. Hunt., caret: Classification and Regression Training, r package version 6.0-77 (2017).
URL `https://CRAN.R-project.org/package=caret`

[31] J. Kupka, P. Rusnok, Fuzzy four-fold tables: their properties and use in fuzzy association analysis, Int. J. Approx. Reasoning (submitted).

[32] J. Kupka, P. Rusnok, On regression methods based on linguistic descriptions, in: IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), IEEE, 2015, pp. 1–7.

[33] J. Kupka, P. Rusnok, Regression analysis based on linguistic associations and perception-based logical deduction, Expert Systems with Applications 67 (2017) 107–114.

[34] J. Kupka, P. Rusnok, Regression ensemble with linguistic descriptions, in: International Conference Series on Soft Methods in Probability and Statistics, Springer, 2018, pp. 141–148.

[35] J. Kupka, I. Tomanová, Dependencies among attributes given by fuzzy confirmation measures, Expert Systems with Applications 39 (9) (2012) 7591–7599.

[36] G. Lakoff, Hedges: a study in meaning criteria and the logic of fuzzy concepts, Journal of Philosophical Logic 2 (1973) 458–508.

[37] E. H. Mamdani, S. Assilian, An experiment in linguistic synthesis with a fuzzy logic controller, International journal of man-machine studies 7 (1) (1975) 1–13.

[38] M. Navara, P. Pták, Uncertainty and dependence in classical and quantum logic—the role of triangular norms, in: Language, Quantum, Music, Springer, 1999, pp. 249–261.

[39] V. Novák, The Alternative Mathematical Model of Linguistic Semantics and Pragmatics (1992), Plenum, New York.[L, Z], 1992.

[40] V. Novák, Towards formalized integrated theory of fuzzy logic, in: Fuzzy Logic and Its Applications to Engineering, Information Sciences, and Intelligent Systems, Springer, 1995, pp. 353–363.

[41] V. Novák, Antonyms and linguistic quantifiers in fuzzy logic, Fuzzy Sets and Systems 124 (3) (2001) 335–351.

[42] V. Novák, Perception-based logical deduction, in: Computational Intelligence, Theory and Applications, Springer, 2005, pp. 237–250.

[43] V. Novák, A comprehensive theory of trichotomous evaluative linguistic expressions, Fuzzy Sets and Systems 159 (22) (2008) 2939–2969.

[44] V. Novák, Reasoning about mathematical fuzzy logic and its future, Fuzzy Sets and Systems 192 (2012) 25–44.

[45] V. Novák, Evaluative linguistic expressions vs. fuzzy categories, Fuzzy Sets and Systems 281 (2015) 73–87.

[46] V. Novák, V. Pavliska, R. Valášek, Specialized software for fuzzy natural logic and fuzzy transform applications, in: 2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), IEEE, 2014, pp. 2337–2344.

[47] V. Novák, I. Perfilieva, On the semantics of perception-based fuzzy logic deduction, International Journal of Intelligent Systems 19 (11) (2004) 1007–1031.

[48] V. Novák, I. Perfilieva, A. Dvořák, Insight Into Fuzzy Modeling, John Wiley & Sons, 2016.

[49] V. Novák, I. Perfilieva, A. Dvořák, G. Chen, Q. Wei, P. Yan, Mining pure linguistic associations from numerical data, Int. J. Approx. Reasoning 48 (1) (2008) 4–22.

[50] V. Novák, I. Perfilieva, J. Močkoř, Mathematical principles of fuzzy logic, Vol. 517, Springer Science & Business Media, 2012.

[51] D. P Pancho, J. M. Alonso, J. Alcalá-Fdez, L. Magdalena, Interpretability analysis of fuzzy association rules supported by fingrams, in: 8th conference of the European Society for Fuzzy Logic and Technology (EUSFLAT-13), Atlantis Press, 2013, pp. 469–474.

[52] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria (2015).
URL `http://www.R-project.org/`

[53] M. Ralbovský, Fuzzy guha, Ph.D. thesis, University of Economics, Prague (2009).

[54] J. Rauch, Classes of association rules: an overview, in: Data Mining: Foundations and Practice, Springer, 2008, pp. 315–337.

[55] J. Rauch, Observational calculi and association rules, Vol. 469, Springer, 2012.

[56] P. Rusnok, Linguistic descriptions as a modeling tool for multivariate time series, in: Proceedings of the 2015 Conference of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology, 2015, pp. 981–986.

[57] P. Rusnok, Probabilistic coverage of linguistic if-then rules, in: IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), IEEE, 2016, pp. 798–803.

[58] M. Štěpnička, A. Dvořák, On perception-based logical deduction and its variants, in: 2015 Conference of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology (IFSA-EUSFLAT-15), 2015, pp. 341–350.

[59] M. Štěpnička, A. Dvořák, V. Pavliska, L. Vavříčková, A linguistic approach to time series modeling with the help of f-transform, Fuzzy Sets and Systems 180 (1) (2011) 164–184.

[60] M. Štěpnička, M. Burda, L. Štěpničková, Fuzzy rule base ensemble generated from data by linguistic associations mining, Fuzzy Sets and Systems 285 (2015) 141–161.

[61] G. I. Webb, Opus: An efficient admissible algorithm for unordered search, Journal of Artificial intelligence Research 3 (1995) 431–465.

[62] A. Wilbik, J. M. Keller, A distance metric for a space of linguistic summaries, Fuzzy Sets and Systems 208 (2012) 79–94.

[63] R. R. Yager, A new approach to the summarization of data, Information Sciences 28 (1) (1982) 69–86.

[64] L. A. Zadeh, Fuzzy sets, Information and control 8 (3) (1965) 338–353.

[65] L. A. Zadeh, A fuzzy-set-theoretic interpretation of linguistic hedges, Journal of Cybernetics (1972) 4–34.

[66] L. A. Zadeh, Outline of a new approach to the analysis of complex systems and decision processes, Systems, Man and Cybernetics, IEEE Transactions on 3 (1) (1973) 28–44.

[67] L. A. Zadeh, Precisiated natural language (pnl), AI magazine 25 (3) (2004) 74.