University of Ostrava

Institute for Research and Applications of Fuzzy Modeling

Discovering Linguistic Associations from Numerical Data

Vilém Novák, Irina Perfilieva, Antonín Dvořák, Guoqing Chen, Qiang Wei, Peng Yan

Research report No. 88

Submitted/to appear:

Fuzzy Sets and Systems

Supported by:

Czech-Chinese bilateral cooperation project ME702 and partially also project 1M0572 of the MŠMT ČR

University of Ostrava
Institute for Research and Applications of Fuzzy Modeling
30. dubna 22, 701 03 Ostrava 1, Czech Republic

tel.: +420-597 460 234 fax: +420-597 461 478
e-mail: {Vilem.Novak,Irina.Perfilieva,Antonin.Dvorak}@osu.cz,
{chengq,weiq,yanp}@em.tsinghua.edu.cn

Discovering Linguistic Associations from Numerical Data [1]

Vilém Novák, [a,*] Irina Perfilieva, [a] Antonín Dvořák [a]
Guoqing Chen, [b] Qiang Wei, [b] Peng Yan [b]

[a] *University of Ostrava*
Institute for Research and Applications of Fuzzy Modeling
30. dubna 22, 701 03 Ostrava 1, Czech Republic

[b] *Tsinghua University, Beijing, China*

Abstract

This paper contains a method for direct search of associations from numerical data that are expressed in natural language and so, we call them "linguistic associations". We have applied three theories: the formal theory of evaluating linguistic expressions, the GUHA method and the F-transform. One of essential outcomes of our theory is high understandability of the found associations because when formulated in natural language they are much closer to the way of thinking of experts from various fields. Moreover, associations characterizing real dependencies can be directly taken as fuzzy IF-THEN rules and used as expert knowledge about the problem.

Key words: Evaluating linguistic expressions, GUHA method, F-transform, associations

1 Introduction

This paper turns to a novel field, whose goal is to discover linguistically characterized information from numerical data. To achieve it, we combine three

* *Corresponding author:* Vilém Novák, University of Ostrava, Institute for Research and Applications of Fuzzy Modeling, 30. dubna 22, 701 03 Ostrava 1, Czech Republic

Email addresses: {Vilem.Novak,Irina.Perfilieva,Antonin.Dvorak}@osu.cz, {chengq,weiq,yanp}@em.tsinghua.edu.cn (Peng Yan).

[1] The paper has been supported by the Czech-Chinese bilateral cooperation project ME702 and partially also by the project 1M0572 of the MŠMT ČR.

tools: the theory of evaluating linguistic expressions, GUHA method and fuzzy transform.

The main concept used in this paper is the formal logical theory of *evaluating linguistic expressions* (expressions like "big", "roughly medium", "very small", etc.), that was initiated by V. Novák in [10] and recently elaborated in details in [14] as a part of higher order fuzzy logic. Note that evaluating (linguistic) expressions are values of linguistic variables; the latter concept was introduced by L. A. Zadeh in [20]. It should be stressed that also the well known fuzzy IF-THEN rules can be taken as special *compound evaluating linguistic expressions* because they can be viewed as *linguistically characterized* logical implications. In this paper, we consider more general compound evaluating expressions called *linguistic associations*. These are, in fact, hypotheses about relations between boolean combinations of evaluating linguistic expressions.

Another tool used in this paper is the *GUHA method*. This method has been introduced as the first data-mining method[†)] by P. Hájek in [3] and later extensively developed by several authors. A comprehensive book on the GUHA method is [5]. Surprisingly, this method is now almost unknown. The original presentation of the GUHA method is based on classical logic. A recent paper extending GUHA by fuzzy logic considerations is [4].

The third theory applied for discovering linguistic associations is the *fuzzy transform* (F-transform) developed by I. Perfilieva in [17–19]. This is a new fuzzy approximation technique that can be used for solution of many problems, for example filtering, solution of differential equations, data compressions, etc. In this paper, it is applied to discovering associations consisting of fuzzy numbers (these also belong among evaluating linguistic expressions).

The paper is organized as follows. The next section recalls few basic notions. Section 3 contains parts of the theory of evaluating linguistic expressions essential for this topic of this paper. Section 4 is an overview of parts of the theory of F-transform that are necessary in further explanation. Section 5 is the main contribution of this paper and it presents a method for mining linguistic associations. We also propose several rules using which it is possible to reduce significantly number of the discovered associations. Finally in Section 6 we demonstrate our method on an example.

[†)] The term "data mining" was not known in that time; instead, the authors of GUHA method used the term "exploratory data analysis".

2 Preliminaries

The methods presented in this paper are in large extent based on the results of formal fuzzy logic. Therefore, it is necessary to decide, which algebra of truth values will be used. Recall that this is, in general, a residuated lattice (see [9, 15]). In this paper, we will consider only standard algebra that is based on the interval of reals $[0, 1]$. Moreover, we will suppose it to be the Łukasiewicz MV-algebra

$$\mathcal{L}_{\text{Ł}} = \langle [0, 1], \vee, \wedge, \otimes, \rightarrow, 0, 1 \rangle$$

where $\vee$ is the operation of *maximum*, $\wedge$ that of *minimum*, $a \otimes b = 0 \vee (a+b-1)$ is *Łukasiewcz conjunction* and $a \rightarrow b = 1 \wedge (1-a+b)$ is *Łukasiewicz implication* $(a, b \in [0, 1])$.

The Łukasiewicz implication $\rightarrow$ interprets the implication connective, $\wedge$ interprets ordinary conjunction connective, and the Łukasiewicz conjunction $\otimes$ interprets strong conjunction connective. Since this paper works with semantics of fuzzy logic only, we will skip presentation of syntax. Let us only stress that linguistic considerations in this paper require even fuzzy logic of higher order, that is, fuzzy type theory (see [12]).

Furthermore, we will work also with derived operations, namely *negation* $\neg a = a \rightarrow 0 = 1 - a$ (interpretation of negation) and *Łukasiewicz disjunction* $\oplus = \neg(\neg a \otimes \neg b) = 1 \wedge (a + b)$ (interpretation of strong disjunction connective). Recall that the negation is involutive, i.e. $\neg\neg a = a$ holds for all $a \in [0, 1]$.

A fuzzy set A in the universe V, in symbols $A \underset{\sim}{\subset} V$, is identified with a function $A : V \longrightarrow [0, 1]$ (the function A is also called the *membership function* of the fuzzy set A). By $\mathcal{F}(V)$ we denote the set of all fuzzy sets on V.

3 Theory of Evaluating Linguistic Expressions and Predications

The fundamental concept in many applications of fuzzy logic is that of evaluating linguistic expressions and predications. These are expressions of natural language such as "very large, extremely expensive, roughly one thousand, more or less hot", etc. In this section, we will briefly describe some aspects of their theory and present a mathematical model of their semantics. More about their theory can be found in [11, 13, 15]. A precise logical theory has been developed in [14].

An *evaluating linguistic expression* is either of the following:

(i) *Simple evaluating expression*, which is one of the linguistic expressions:
 (a) ⟨pure evaluating expression⟩ :=
 ⟨linguistic hedge⟩⟨atomic evaluating expression⟩
 (b) ⟨fuzzy number⟩ := ⟨linguistic hedge⟩⟨numeral⟩
 where "numeral" is a name of some number $x_0 \in \mathbb{R}$.

(ii) *Negative evaluating expression*, which is an expression

$$\textsf{not}\ \langle\text{pure evaluating expression}\rangle$$

(iii) *Compound evaluating expression*, which is either of the following:
 (a) ⟨pure evaluating expression⟩ or ⟨pure evaluating expression⟩
 (b) ⟨pure evaluating expression⟩ and/but ⟨negative evaluating expression⟩

Atomic evaluating expressions and *numerals* form the basic component of all kinds of evaluating expressions. They comprise any of the adjectives "small", "medium", or "big". It is important to stress that these words should be taken as canonical and can be replaced by many other kinds of words such as "thin", "thick", "old", "new", etc.

Numerals are basic linguistic expressions characterizing some element x_0 on an ordered scale. Examples are *one million, 1 145*, etc. Note that numbers in common human understanding are almost always understood as imprecise, i.e. for example "one milion" never means the number 1 000 000, but "something close to it". They can be modified using linguistic hedge, for example *roughly 1 000, about 3 millon*, etc.

Atomic evaluating expressions usually form pairs of antonyms such as *thin — thick, old — young, shallow — deep, close — far*, etc. When completed by the middle term, such as *medium, average* [†)], etc. they form the so called *fundamental linguistic trichotomy.*

Linguistic hedges (introduced by L. A. Zadeh, see [20]) are special adverbs which modify the meaning of adjectives before which they stand (cf. also [9]). We distinguish hedges with *narrowing effect* (very, significantly, etc.) and *widening effect* (more or less, roughly, etc.). It is important to realize that missing linguistic hedge is understood as presence of the *empty linguistic hedge* so that all simple evaluating expressions can be treated equally. In the

[†)] Note that natural language is quite rich by the basic pairs of antonyms but the middle term has mostly the form *medium* or *average* completed by the corresponding adjective.

sequel, we will use script letters $\mathcal{A}, \mathcal{B}, \ldots$ to denote evaluating expressions.

We must distinguish between evaluating expressions and predications. While the former characterize linguistically values on an ordered scale in a rather abstract way, the latter do the same in concrete, specific scales.

Evaluating linguistic predications are expressions of the form

$$\langle\text{noun phrase}\rangle \text{ is } \mathcal{A} \tag{1}$$

where $\mathcal{A}$ is an evaluating linguistic expression. In our considerations, we often replace $\langle$noun phrase$\rangle$ by some attribute (variable) X and assume that its interpretation are real numbers. Therefore, we will deal with predications of the form

$$X \text{ is } \mathcal{A}. \tag{2}$$

Expression (1) should be distinguished from "x is $\mathcal{A}$" where x is some concrete value of X (or concrete object denoted by $\langle$noun$\rangle$). For example, if (1) is the expression "house is small" then the concrete object x can be "house of John" so that we obtain the predication "house of John is small". It is important to notice that in the latter case, a concrete context is always considered. Expression (2) is more abstract and it concerns all possible values of the variable X. Because of vagueness of $\mathcal{A}$, the predication "x is $\mathcal{A}$" for concrete x (in the given context) may attain a general truth value.

The predication (2) is usually taken as synonymous with

$$\mathcal{A}\ X. \tag{3}$$

For example, let the noun phrase denoted by attribute X be "pressure in the tank". Then we take the evaluating predication "pressure in the tank is high" as synonymous with "high pressure in the tank" For our purposes, the form (3) is more convenient and so, we will prefer it in the sequel.

Let I, J be two nonempty finite index sets. Then we put

$$\mathcal{C} := \underset{i\in I}{\mathsf{AND}}\,(\mathcal{A}_i\ X_i), \tag{4}$$

$$\mathcal{D} := \underset{i\in I}{\mathsf{OR}}\,(\mathcal{B}_i\ X_i) \tag{5}$$

where $\mathcal{A}_i, \mathcal{B}_i$ are evaluating linguistic expressions. Furthermore, we suppose that each X_i in $\mathcal{C}$ or $\mathcal{D}$, $i \in I$, differs from the other ones. The number of elements of I is called *length* of $\mathcal{C}$ or $\mathcal{D}$.

A special case occurs when the variable X is the same in the disjunction (5). Then the evaluating predication $\mathcal{D}$ reduces to the form

$$\mathcal{D} := (\underset{i\in I}{\mathsf{OR}}\ \mathcal{B}_i\ X). \tag{6}$$

For example, let '*small pressure*' and '*big pressure*' be two linguistic predications (3). Then their disjunction (5) reduced to (6) is '*small or big pressure*' (recall that we take it synonymous with '*pressure is small or big*').

A *compound evaluating (linguistic) predication* is either of

$$\mathcal{E} := \operatorname*{OR}_{j\in J} \mathcal{C}_j, \tag{7}$$

$$\mathcal{F} := \operatorname*{AND}_{j\in J} \mathcal{D}_j \tag{8}$$

where $\mathcal{C}_j$, $\mathcal{D}_j$ are conjunction or disjunction of evaluating predications from (4) and (5), respectively. In other words, a compound evaluating predication is a boolean combination of evaluating linguistic predications. Obviously, the expressions (4) and (5) are special cases of compound evaluating predications (8) and (7), respectively.

A very important special expression that also belongs among compound linguistic predications is a fuzzy IF-THEN rule

$$\mathsf{IF}\ X\ \mathsf{is}\ \mathcal{A}\ \mathsf{THEN}\ Y\ \mathsf{is}\ \mathcal{B}. \tag{9}$$

This is a conditional clause characterizing linguistically some dependence between the variables X and Y. In practice, it is construed either as a logical implication, or as a conjunction.

3.2 Semantics of evaluating expressions

Any model of semantics of natural language expressions must be able to distinguish between intension and extensions in various possible worlds (cf., e.g. [1, 8]). Recall that a *possible world* is in logic understood as a state of our world at a given time moment and place. In linguistics, this is a *particular context* in which a linguistic expression is used. *Intension* of a linguistic expression is an abstract construction which conveys a property denoted by the expression. Linguistic expressions are names of intensions. It is important to notice that intension is invariant towards change of the context (possible world) while *extension* of a linguistic expression is a class of objects determined by its intension in a given context (possible world) and so, it it changes whenever the context (time, place) is changed. For example, the expression "long distance" is the name of an intension being a certain property of length, which in a concrete context may mean 10 m for an ant, 100 km in the Czech Republic, but 1000 km or more in China.

In this section, we outline the main points of the mathematical model of the meaning of evaluating expressions that has been in details described in [14] and which is based on the solution of the sorites paradox †) in fuzzy logic (for the detailed analysis see [6] and also [14]). First we will define meaning of evaluating expressions (not predications!).

The simplest semantic interpretation of the sorites paradox leads to linear functions $LH, MH, RH : [0,1] \longrightarrow [0,1]$ defined by

$$LH(z) = \left(\frac{0.5 - z}{0.5}\right)^*, \qquad RH(z) = \left(\frac{z - 0.5}{0.5}\right)^*,$$
$$MH(z) = \left(\frac{z}{0.5}\right)^* \wedge \left(\frac{1 - z}{0.5}\right)^*$$

where the star '$*$' means cut of all the values to interval $[0,1]$. These functions characterize the idea of running towards horizon: small values start at 0 and run towards 0.5 where they vanish, big values start at 1 and run in an opposite direction also towards 0.5 where they surely vanish and medium values start at 0.5 and vanish on both sides at 0 and 1, respectively.

To define the meaning of fuzzy numbers, we put

$$\delta_z(x) = \begin{cases} 1 & \text{if } x = z, \\ 0 & \text{otherwise} \end{cases}$$

and

$$B_{z_0,h}(z) = \begin{cases} \left(\frac{(z+h)-z_0}{h}\right)^* \wedge \left(\frac{z_0-(z-h)}{h}\right)^*, & \text{if } 0 < h < 1, \\ \delta_{z_0} & \text{if } h = 0, \end{cases} \tag{10}$$

where $z_0 \in [0,1]$.

Linguistic hedges are modeled using continuous functions $\nu_{a,b,c} : [0,1] \longrightarrow [0,1]$ (horizon deformations) where $a < b < c$ are parameters, $\nu_{a,b,c}(y) = 0$ for $y \le a$, $\nu_{a,b,c}(y) = 1$ for $c \le y$ and it is increasing otherwise. We explicitly put †)

$$\nu_{a,b,c}(y) = \begin{cases} 1, & c \le y, \\ 1 - \frac{(c-y)^2}{(c-b)(c-a)}, & b \le y < c, \\ \frac{(y-a)^2}{(b-a)(c-a)}, & a \le y < b, \\ 0, & y < a. \end{cases} \tag{11}$$

†) One grain does not form a heap. Adding one grain to what is not yet a heap does not make a heap. Consequently, there are no heaps.

†) Of course, there are infinitely many possible functions $\nu_{a,b,c}$. Our goal was to consider the simplest non-linear one since according to known psychological investigations, the shapes of membership functions should be nonlinear. However, we cannot reject the function $\nu_{a,b,c}$ to be, possibly even linear.

The following functions model intensions of evaluating expressions:

$$\widetilde{Sm}_\nu(z) = \nu(LH(z)), \tag{12}$$

$$\widetilde{Me}_\nu(z) = \nu(MH(z)), \tag{13}$$

$$\widetilde{Bi}_\nu(z) = \nu(RH(z)), \tag{14}$$

$$\widetilde{\mathrm{Fn}}_{\nu,z_0}(z) = \nu(B_{z_0,h})(z) \tag{15}$$

where $\nu \in \mathbf{Hf}$ interprets $\langle$linguistic hedge$\rangle$ and $z \in [0,1]$. Then we put:

$$\mathrm{Int}(\langle\text{linguistic hedge}\rangle\ \mathsf{small}) = \widetilde{Sm}_\nu, \tag{16}$$

$$\mathrm{Int}(\langle\text{linguistic hedge}\rangle\ \mathsf{medium}) = \widetilde{Me}_\nu, \tag{17}$$

$$\mathrm{Int}(\langle\text{linguistic hedge}\rangle\ \mathsf{big}) = \widetilde{Bi}_\nu. \tag{18}$$

Note that (16)–(18) are, in fact, fuzzy sets in a fixed set $[0,1]$ and so, they are at the same time extensions of the corresponding evaluating expressions. This is not true for evaluating predications — see below.

In the sequel, we will use as a metavariable the symbol $\widetilde{Ev}_\nu$ that will denote either of $\widetilde{Sm}_\nu$, $\widetilde{Me}_\nu$, $\widetilde{Bi}_\nu$, $\widetilde{\mathrm{Fn}}_{\nu,z_0}$ for some $z_0 \in [0,1]$. The latter formula represents abstract concept of a fuzzy number (but not yet an intension of a fuzzy number itself — see below). We will denote by $\widetilde{\mathbf{Ev}}$ the set of intensions $\widetilde{Ev}_\nu$.

3.3 Ordering of evaluating expressions

Important role in manipulation with evaluating expressions is played by natural ordering of them. This is a lexicographic ordering based on two orderings: the first is called the *specificity ordering* and the second is *position ordering*. The former is is determined by the fact that we can distinguish hedges with *narrowing* and *widening* effect where the first make the meaning of the atomic expression before which they stand more precise while the second make it opposite.

In discovering linguistic association rules, we will work with several concrete hedges, namely *"extremely (Ex), significantly (Si), very (Ve), empty hedge, more or less (ML), roughly (Ro), quite roughly (QR), very roughly (VR)"*. Among them, the hedges *Ex, Si*, and *Ve* have narrowing effect and *ML, Ro, QR* and *VR* have widening effect. In [13], we have defined empirical values of the parameters a, b, c of these hedges. These hedges have been chosen because they are very common in ordinary speech. However, our theory is general enough to include many other concrete examples of hedges.

The *specificity ordering* $\preceq$ for the above hedges is defined as follows:

$$\mathrm{Ex} \preceq \mathrm{Si} \preceq \mathrm{Ve} \preceq \langle\text{empty hedge}\rangle \preceq \mathrm{ML} \preceq \mathrm{Ro} \preceq \mathrm{QR} \preceq \mathrm{VR}. \tag{19}$$

This ordering means that all values in some context, that are *extremely* small (or big), are also *significantly* small (or big), etc. However, since there exist also hedges which have neither narrowing, nor widening effect (for example *rather*), the ordering $\preceq$ is in general, only partial. Therefore, we suppose that the set **Hf** is partially ordered by the specificity relation $\preceq$. This ordering induces a partial ordering on evaluating expressions defined by

$$\widetilde{Ev}_{\nu_1} \preceq \widetilde{Ev}_{\nu_2} \quad \text{iff} \quad \nu_1 \leq \nu_2 \tag{20}$$

where $\widetilde{Ev}$ is either of $\widetilde{Sm}, \widetilde{Me}, \widetilde{Bi}$ or $\nu(B_{z_0,h})$ and $\leq$ is pointwise ordering of functions.

The *position ordering* $\lessdot$ corresponds to the position of the evaluating expressions on the scale, i.e.

$$\widetilde{Sm}_{\nu_1} \lessdot \widetilde{Me}_{\nu_2} \lessdot \widetilde{Bi}_{\nu_3}, \tag{21}$$

and

$$\text{if } z_1 \leq z_2 \text{ then } \nu_1(B_{z_1,h_1}) \lessdot \nu_2(B_{z_2,h_2}) \tag{22}$$

where $\nu_1, \nu_2, \nu_3, h_1, h_2$ are arbitrary. In the applications discussed below, we also consider a special expression *zero* (Ze) whose position in (21) is leftmost.

On the basis of (20)–(22), we introduce *natural (partial) ordering*

$$Ev_1 \precapprox Ev_2 \tag{23}$$

of evaluating expressions as lexicographic ordering, where first we order $\widetilde{Ev}_1, \widetilde{Ev}_2$ according to $\lessdot$ from (21) or (22), and then according to $\preceq$ from (20) (provided that $\widetilde{Ev}_1$ and $\widetilde{Ev}_2$ are comparable).

3.4 Semantics of evaluating predications

To formalize semantics of evaluating predications (see Subsection 3.1), we must first explicitly define the concept of context. In general, a context is a $\leq$-homomorphism $w : [0,1] \longrightarrow [0,\infty]$ with three distinct points, namely $w(0) = v_L$, $w(0.5) = v_S$ and $w(1) = v_R$. These points are *left limit* ("most typically small"), *central point* ("most typically medium") and *right limit* ("most typically big"), respectively. Clearly, $\text{rng}(w) = [v_L, v_R]$. The set of all contexts will be denoted by W.

Extensions of the evaluating expressions characterizing small values lay between v_L, v_S and those characterizing big values lay between v_S, v_R (with the direction from v_R to v_S). The expressions characterizing medium values are determined by the point v_S which is the "most typical" medium and their extensions lay around it. The following types of functions will play a role as intensions of evaluating predications:

(i) S-intensions: $Sm_\nu : W \longrightarrow \mathcal{F}(\mathbb{R})$, $\nu \in \mathbf{Hf}$,
where $Sm_\nu(w) = \widetilde{Sm}_\nu(w^{-1}(x))$ for all $w \in W$ and $x \in \mathrm{rng}(w)$.

(ii) M-intensions: $Me_\nu : W \longrightarrow \mathcal{F}(\mathbb{R})$, $\nu \in \mathbf{Hf}$,
where $Me_\nu(w) = \widetilde{Me}_\nu(w^{-1}(x))$ for all $w \in W$ and $x \in \mathrm{rng}(w)$.

(iii) B-intensions: $Bi_\nu : W \longrightarrow \mathcal{F}(\mathbb{R})$, $\nu \in \mathbf{Hf}$,
where $Bi_\nu(w) = \widetilde{Bi}_\nu(w^{-1}(x))$ for all $w \in W$ and $x \in \mathrm{rng}(w)$.

(iv) FN-intensions: $\mathrm{Fn}_{\nu,x_0} : W \longrightarrow \mathcal{F}(\mathbb{R})$, $\nu \in \mathbf{Hf}$
where $\mathrm{Fn}_{\nu,x_0}(w)(x) = \nu(B_{w^{-1}(x_0),h})(w^{-1}(x)))$ for all $x, x_0 \in \mathrm{rng}(w)$ and $h \in [0, 1)$.

It is clear from the previous explanation that S-intensions are intensions of predications '⟨linguistic hedge⟩ small X', M-intensions are intensions of predications '⟨linguistic hedge⟩ medium X' and B-intensions are intensions of predications '⟨linguistic hedge⟩ big X'. FN-intensions are intensions of predications '⟨linguistic hedge⟩ [approximately] x_0 X'.

Similarly as in Subsection 3.2, we introduce a metavariable Ev_ν which denotes intension

$$Ev_\nu : W \longrightarrow \mathcal{F}(\mathbb{R}). \tag{24}$$

of some evaluating predication (3). Its extension in each context $w \in W$ is the fuzzy set

$$Ev_\nu(w) \underset{\sim}{\subset} \mathrm{rng}(w).$$

Clearly,

$$Ev_\nu(w)(x) = \widetilde{Ev}_\nu(w^{-1}(x)) \tag{25}$$

holds for every $x \in \mathrm{rng}(w)$. Hence, by abuse of notation, we will often write the evaluating predication (3) in the form "$\widetilde{Ev}_\nu\ X$" keeping in mind that $\widetilde{Ev}_\nu$ is, in fact, intension of some evaluating expression.

If the hedge ν is unimportant for the given considerations, we may omit it from the symbol. Construction of extensions of evaluating predications is depicted on Figure 1.

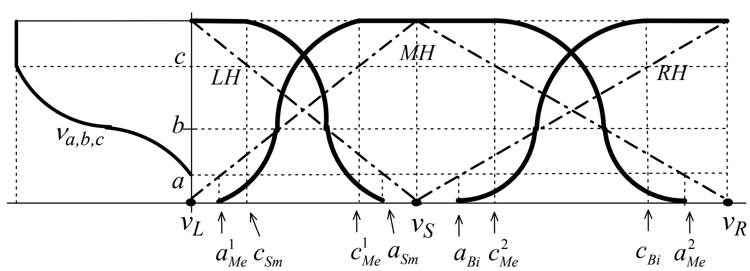

Fig. 1. A scheme of construction of extensions of evaluating expressions where $c_{Sm} = LH^{-1}(c), a_{Sm} = LH^{-1}(a), c^1_{Me} = (\neg LH)^{-1}(c), a^1_{Me} = (\neg LH)^{-1}(a), c^2_{Me} = (\neg RH)^{-1}(c), a^2_{Me} = (\neg RH)^{-}$

We may extend orderings introduced in Subsection 3.3 both to extensions as well intensions of evaluating predications by

$$Ev_1(w) \preceq Ev_2(w) \quad \text{iff} \quad \widetilde{Ev_1} \preceq \widetilde{Ev_2}, \tag{26}$$

and similarly for the position ordering $\lessdot$ and natural ordering $\precapprox$. Note that the context is the same both for Ev_1 as well as for Ev_2. Consequently, $Ev_1 \preceq Ev_2$ iff (26) holds for all $w \in W$. From it follows that we may introduce the specificity ordering of evaluating predications by

$$(\widetilde{Ev_1}\ X) \preceq (\widetilde{Ev_2}\ X) \quad \text{iff} \quad \widetilde{Ev_1} \preceq \widetilde{Ev_2}, \tag{27}$$

and similarly for $\lessdot$ and $\precapprox$. Note that the variable X is the same in both predications. If the variables are different then the predications cannot be compared, even in the case that they contain the same evaluating expression.

Intension of compound evaluating predications is obtained as a combination of intensions Ev. This combination is defined using logical operations on the set of truth values as follows.

Let $\square$ be a binary logical operation on $[0,1]$ and Ev_1 and Ev_2 be intensions of some evaluating predications. Then a compound intension $Ev_1 \square Ev_2$ (intension of a compound predication) is a function assigning to each couple of contexts $w_1, w_2 \in W$ a fuzzy set given by the membership function

$$Ev_1(w_1)(x) \square Ev_2(w_2)(y) = \widetilde{Ev}_1(w_1^{-1}x) \square \widetilde{Ev}_2(w_2^{-1}y), \tag{28}$$

where $x \in \text{rng}(w_1), y \in \text{rng}(w_2)$.

Let the linguistic AND and OR be interpreted by $\wedge$ or $\vee$, respectively (alternatively, AND can be interpreted by $\otimes$). Then intensions of (4) and (5) are defined by

$$\text{Int}(\underset{i \in I}{\mathsf{AND}}\ (\widetilde{Ev_i}\ X_i)) = \bigwedge_{i \in I} Ev_i, \tag{29}$$

$$\text{Int}(\underset{i \in I}{\mathsf{OR}}\ (\widetilde{Ev_i}\ X_i)) = \bigvee_{i \in I} Ev_i, \tag{30}$$

where the conjunction and disjunction on the right hand side of (29) and (30), respectively is defined using (28) when putting $\square = \wedge$ or $\square = \vee$, respectively. Analogously, we may define intensions of (7) and (8). Their extensions can be easily obtained as the corresponding functional values of their intensions for all $w \in W$.

Remark 1
Let us give thought to the question, what kind of conjunction should interpret linguistic "and". The operation $\wedge$ interprets a phrasal conjunction, while $\otimes$ raises when joining formulas in modus ponens and so, it corresponds to a sentential conjunction. Therefore, we should distinguish whether the evaluating

predication is taken as a phrase or a sentence, and when forming compound predication, use the corresponding conjunction . On the other hand, we doubt that such a mechanical solution would really work. There is one, purely technical problem — the $\otimes$ operation is usually quite restricting (in Łukasiewicz algebra it is even nilpotent, so that it quickly sinks to zero) and so, we might get very small truth values when joining many predications. The problem, unfortunately, is far from being solved and so, we will, for the present, prefer phrasal conjunction $\wedge$. Note that the operation $\oplus$ dual to $\otimes$ has no specific role (analogous to the role of $\otimes$ in modus ponens) and so, we do not consider it for disjunction.

3.5 Finding a suitable expression

Let an element $x \in \text{rng}(w)$ in a context w be given. This element becomes an observation which can be in this context evaluated by several evaluating predications. For example, let $x = 9$ in the context $\text{rng}(w) = [0, 10]$. When taking into account that the highest possible value is 10, we may form an evaluating predication "9 is big". Of course, we may form also "9 is very big", "9 is roughly big", etc. Intuitively, we prefer the most precise expression (e.g. "roughly big" is less precise than "very big") but at the same time, the value 9 must be typical for it (of course, still in the fixed context). This gives hint for the following concept.

We introduce a function Suit which assigns to each context $w \in W$ and to each element $x \in \text{rng}(w)$ an evaluating expression with intension $\widetilde{Ev}_\nu$,

$$\text{Suit} : \langle x, w \rangle \mapsto \widetilde{Ev}_\nu, \tag{31}$$

so that $\widetilde{Ev}_\nu$ is the *most specific* (sharpest) in the sense of the natural ordering $\precapprox$ defined in Subsection 3.3, and $x \in \text{rng}(w)$ is *typical* in the extension $Ev_\nu(w)$ of the evaluating predication "$\widetilde{Ev}_\nu\ X$". To be *typical* means that the membership degree $Ev_\nu(w)(x) = \widetilde{Ev}_\nu(w^{-1}(x))$ is greater than some reasonable threshold a^0 (we usually put $a^0 = 0.9$ or even $a^0 = 1$).

Note that (31), in fact, evaluates linguistically the value x in the given context w. When taking into account the above example, $\text{Suit}(9, w)$ is the evaluating expression *very big*.

The definition of Suit can be justified by the empirical finding that given a context, each value of it can be classified by some evaluating predication. Since the expressions can be more, or less specific, the most specific one gives the most precise information. If there is no evaluating predication being most specific and typical then Suit gives nothing. The evaluating predication "$\widetilde{Ev}_\nu\ x$" with the intension given by (31) will be called *perception* of x (in the context

w).

3.6 Linguistic Associations

A *linguistic association* is the expression

$$\mathcal{E} \sim \mathcal{F} \tag{32}$$

where $\mathcal{E}$ and $\mathcal{F}$ are compound evaluating predications of the form (7) or (8) and $\sim$ is a binary quantifier in the sense of GUHA — see [4, 5]. In the GUHA terminology $\mathcal{E}$ is *antecedent* and $\mathcal{F}$ is *succedent*, i.e. a follower of the antecedent (we should not call it consequent because when the association is found in the data, we cannot be sure that it expresses a real dependence between $\mathcal{E}$ and $\mathcal{F}$).

For further explanation, we will write $\mathcal{E}(X_1, \ldots, X_p)$ where $X_1, \ldots, X_p$ are all variables occurring in $\mathcal{E}$. Similarly we write $\mathcal{F}(Y_1, \ldots, Y_q)$. Note that $\mathcal{E}(X)$ is either a simple predication (3) (or (2)), or a compound predication being a disjunction of the form (6).

Let $\mathcal{E}$ and $\mathcal{E}'$ be compound evaluating predications that contain the same variables and have the same boolean structure. We say that $\mathcal{E}$ *is narrower* than $\mathcal{E}'$ ($\mathcal{E}'$ *is wider* than $\mathcal{E}$) if

$$(\widetilde{Ev}\ X) \preceq (\widetilde{Ev}'\ X)$$

holds for all evaluating predications '$\widetilde{Ev}\ X$' from $\mathcal{E}$ and '$\widetilde{Ev}'\ X$' from $\mathcal{E}'$, respectively. Similarly, a linguistic association $\mathcal{C} \sim \mathcal{D}$ *is narrower* (wider) than $\mathcal{C}' \sim \mathcal{D}'$ if $\mathcal{C}$ is narrower (wider) than $\mathcal{C}'$ and $\mathcal{D}$ is narrower than $\mathcal{D}'$.

Linguistic associations are hypotheses about possible validity of fuzzy IF-THEN rules (9). Note that intension of a fuzzy IF-THEN rule is given by (28) where $\Box$ is either $\rightarrow$ or $\wedge$ (possibly also $\otimes$).

4 Fuzzy transform

The fuzzy transform is in details described in [18, 19] where many theorems including full proofs can be found. In this paper, we recall only some of the main points.

Let a continuous function f be defined on an interval of real numbers $w = [v_L, v_R] \subset \mathbb{R}$ and choose some *points* $p_1, \ldots, p_N \in w$ at which the function f is computed. Furthermore, let the interval w be divided into a set of equidistant *nodes* $x_k = v_L + h(k-1)$, $k = 1, \ldots, n$ where $N > n$ and $h = \frac{v_R - v_L}{n-1}$ is the fixed length. Obviously, $x_1 = v_L$ and $x_n = v_R$. The F-transform has two phases.

4.0.0.1 Direct F-transform.

We define n basic functions $A_1, \ldots, A_n$, which cover w and divide it into n vague areas. The basic functions must fulfil the following conditions ($k = 1, \ldots, n$):

1. $A_k : w \longrightarrow [0,1]$, $A_k(x_k) = 1$,
2. $A_k(x) = 0$ if $x \notin (x_{k-1}, x_{k+1})$ where we formally put $x_0 = x_1 = v_L$, $x_{n+1} = x_n = v_R$,
3. $A_k(x)$ is continuous,
4. $A_k(x)$ monotonously increases on $[x_{k-1}, x_k]$ and monotonously decreases on $[x_k, x_{k+1}]$,
5. $\sum_{k=1}^{n} A_k(x) = 1$, for all $x \in w$.

Clearly, there are infinitely many possibilities for choosing the basic functions. In our case, we take as basic functions extensions of fuzzy numbers $\mathrm{Fn}_{\nu,x_0}(w)$ and set explicitly

$$\mathrm{Fn}_{\nu,x_0}(x) = \begin{cases} 1, & x \in [c^L_{x_0}, c^R_{x_0}], \qquad c^L_{x_0} = x_0 - (1-c)h, \\ & \qquad\qquad c^R_{x_0} = x_0 + (1-c)h, \\ 1 - \frac{(c^L_{x_0} - x)^2}{K_1 h^2}, & x \in [b^L_{x_0}, c^L_{x_0}), \qquad b^L_{x_0} = x_0 - (1-b)h, \\ 1 - \frac{(x - c^R_{x_0})^2}{K_1 h^2}, & x \in (c^R_{x_0}, b^R_{x_0}], \qquad b^R_{x_0} = x_0 + (1-b)h, \\ \frac{(x - a^L_{x_0})^2}{K_2 h^2}, & x \in (a^L_{x_0}, b^L_{x_0}), \qquad a^L_{x_0} = x_0 - (1-a)h, \\ \frac{(a^R_{x_0} - x)^2}{K_2 h^2} & x \in (b^R_{x_0}, a^R_{x_0}), \qquad a^R_{x_0} = x_0 + (1-a)h, \\ 0 & x \le a^L_{x_0}, \quad x \ge a^R_{x_0}, \end{cases} \tag{33}$$

where $a, b, c \in [0,1]$ are parameters of the hedge ν. Note that $\mathrm{Fn}_{\nu,x_0}(x_0) = 1$ and $\mathrm{Fn}_{\nu,x_0}(x_0 \pm h) = 0$. Thus, one fuzzy number is spread over three neighboring nodes $x_0 - h, x_0, x_0 + h$. Furthermore,

$$\sum_{k=1}^{n} \mathrm{Fn}_{\nu,x_{0k}}(x) = 1$$

holds for each $x \in w$. Consequently, each $x \in w$ is covered by exactly two neighboring fuzzy numbers $\mathrm{Fn}_{\nu,x_{0k}}, \mathrm{Fn}_{\nu,x_{0k+1}}$. This means that $x_{0k} \le x \le x_{0k+1}$ and $\mathrm{Fn}_{\nu,x_{0k}}(x) + \mathrm{Fn}_{\nu,x_{0k+1}}(x) = 1$.

Using the basic functions, we transform the given function f into n-tuple of components, that are real numbers $[F_1, \dots, F_n]$ defined by

$$F_k = \frac{\sum_{j=1}^{N} f(p_j) A_k(p_j)}{\sum_{j=1}^{N} A_k(p_j)}, \qquad k = 1, \dots, n. \tag{34}$$

4.0.0.2 Inverse F-transform. The result of the direct F-transform is a vector of numbers $[F_1, \dots, F_n]$. This set contains information about the original function f and can be used to obtain a function

$$f_{F,n}(x) = \sum_{k=1}^{n} F_k \cdot A_k(x). \tag{35}$$

The function (35) is called an inverse F-transform. It can be proved that if n increases then $f_{F,n}(p_j)$ converges to $f(p_j)$, $j = 1, \dots, N$. It is clear that the function $f_{F,n}$ is continuous and is a reasonable approximation of the original function f.

The F-transform has (besides others) the following important properties:

(a) It has nice filtering properties.
(b) It is easy to compute.
(c) The F-transform is stable with respect to the choice of the points $p_1, \dots, p_N$, provided that the number of nodes is fixed. This means that when choosing other points p_k (and possibly changing their number N), the resulting function $f_{F,n}$ *does not significantly change*. Note that this is not true for many other classical numerical methods.

5 Discovering linguistic knowledge from numerical data

We suppose to have numerical data in the form

	X_1	$\cdots$	X_i	$\cdots$	X_n
o_1	f_{11}	$\cdots$	f_{1i}	$\cdots$	f_{1n}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
o_j	f_{j1}	$\cdots$	f_{ji}	$\cdots$	f_{jn}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
o_m	f_{m1}	$\cdots$	f_{mi}	$\cdots$	f_{mn}

(36)

where $o_1, \ldots, o_m$ are some objects (processes, transactions, etc.), $X_1, \ldots, X_n$ are variables (attributes). The $f_{ji} \in \mathbb{R}$, $j = 1, \ldots, m$, $i = 1, \ldots, n$ are values of i-th attribute measured on j-th object.

5.1 *Searching pure linguistic associations*

Each attribute X_i can attain values from some range. In other words, for each attribute X_i there exists its context $w_i \in W$. With reference to our theory of evaluating expressions, it is possible to assign to each value f_{ji} the corresponding evaluating expression using the function (31), i.e. we set

$$\widetilde{Ev}_{ji} = \mathrm{Suit}(f_{ji}, w_i). \tag{37}$$

From it follows that (37) provides evaluating linguistic predication of the form

$$\mathcal{A}_{ji}\ X_i \tag{38}$$

where $\mathcal{A}_{ji}$ is some evaluating expression such that $\mathrm{Int}(\mathcal{A}_{ji}) = \widetilde{Ev}_{ji}$. Using (37), we may convert the given numerical data into linguistic ones containing evaluating expressions. In other words, (36) is transformed into

$$\begin{array}{c|ccccc|} & X_1 & \cdots & X_i & \cdots & X_n \\ \hline o_1 & \mathcal{A}_{11} & \cdots & \mathcal{A}_{1i} & \cdots & \mathcal{A}_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ o_j & \mathcal{A}_{j1} & \cdots & \mathcal{A}_{ji} & \cdots & \mathcal{A}_{jn} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ o_m & \mathcal{A}_{m1} & \cdots & \mathcal{A}_{mi} & \cdots & \mathcal{A}_{mn} \end{array} \tag{39}$$

Note that each item in (39) represents predication of the form (38) *in the given context*! This means that for the object o_j, the attribute X_i is $\mathcal{A}_{ji}$ (in the given context).

For example, let X_i be *salary* and object o_j be some person *Hellen*. Let the value f_{ji} from the original table (36) occur in some context w_i so that $\mathrm{Suit}(f_{ji}, w_i) = \widetilde{Ev}_{ji} = \mathrm{Int}(\textit{very small})$. Then the ji-th item in data (39) corresponds to the linguistic predication

"very small salary of Hellen"

(recall that this is synonymous with "the salary of Hellen is very small").

Of course, data (39) is not equivalent with (36) because the latter contains vague linguistic expressions. On the other hand, its size is, in general, smaller than that of (36) because of the fuzziness of the evaluating expressions: it may happen that different values f_{ji} and f_{ki} lead to the same expression. Hence, the number of different rows in (39) can be smaller than m. In the sequel, we will work with the data

$$\begin{array}{c|ccccc|} & X_1 & \cdots & X_i & \cdots & X_n \\ \hline o_1 & \mathcal{A}_{11} & \cdots & \mathcal{A}_{1i} & \cdots & \mathcal{A}_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ o_j & \mathcal{A}_{j1} & \cdots & \mathcal{A}_{ji} & \cdots & \mathcal{A}_{jn} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ o_{m'} & \mathcal{A}_{m'1} & \cdots & \mathcal{A}_{m'i} & \cdots & \mathcal{A}_{m'n} \end{array} \tag{40}$$

where $m' \leq m$. Each expression $\mathcal{A}_{ji}$ in (40) is assigned a number p_{ji} representing the number of objects which have the same linguistic values for *all* attributes $X_1, \dots, X_n$. Clearly,

$$\sum_{j=1}^{m'} p_{ji} = m$$

for arbitrary $i = 1, \dots, n$.

Further step is to discover *linguistically expressed* associations of the form (32). We will deal with certain subsets of the attributes $X_1, \dots, X_n$ from (40). For convenience, we will rename them and write them as sequences of new attributes (variables) $Y_1, \dots, Y_p$ and $Z_1, \dots, Z_q$. We start with simpler associations of the form

$$\mathcal{C}(Y_1, \dots, Y_p) \sim \mathcal{D}(Z_1, \dots, Z_q) \tag{41}$$

where

$$\mathcal{C}(Y_1, \dots, Y_p) = \mathop{\mathsf{AND}}_{i=1}^{p} (\mathcal{A}_i \; Y_i), \tag{42}$$

$$\mathcal{D}(Z_1, \dots, Z_q) = \mathop{\mathsf{AND}}_{j=1}^{q} (\mathcal{B}_j \; Z_j) \tag{43}$$

are conjunctions of evaluating predications (4) and $\{Y_1, \dots, Y_p\} \cap \{Z_1, \dots, Z_q\} = \emptyset$.

The linguistic predications forming the data (40) are vague. However, after assignment to concrete values as their perceptions, they behave as logical data. This means that for each object o_j, *it is true* (or it is not true) that the attribute (variable) X_i has the (vague) property named by the linguistic expression $\mathcal{A}_{ji}$ (i.e. that $\mathcal{A}_{ji}$ is its perception). Consequently, we can count numbers of the corresponding objects and apply standard GUHA quantifiers (see [4, 5]) for discovering associations. This can be done as follows.

Let $\mathcal{C}(Y_1, \ldots, Y_p) \sim \mathcal{D}(Z_1, \ldots, Z_q)$ be a suspected linguistic association (41). Then we can construct a four-fold table

$$\begin{array}{r|c|c|} & \mathcal{D} & \text{not } \mathcal{D} \\ \hline \mathcal{C} & a & b \\ \hline \text{not } \mathcal{C} & c & d \\ \hline \end{array} \tag{44}$$

where a is a number of positive occurrences of $\mathcal{C}$ as well as $\mathcal{D}$, that is, a number of objects, whose attributes $Y_1, \ldots, Y_p$ *are* evaluated by the respective expressions $\mathcal{A}_1, \ldots, \mathcal{A}_p$ from (42) and, at the same time, their attributes $Z_1, \ldots, Z_q$ *are* evaluated by the respective expressions $\mathcal{B}_1, \ldots, \mathcal{B}_q$ from (43).

Further, b is a number of positive occurrences of $\mathcal{C}$ but negative occurrences of $\mathcal{D}$, that is, a number of objects, whose attributes $Y_1, \ldots, Y_p$ *are* evaluated by the respective expressions $\mathcal{A}_1, \ldots, \mathcal{A}_p$ from (42) but their attributes $Z_1, \ldots, Z_q$ *are not* evaluated by the respective expressions $\mathcal{B}_1, \ldots, \mathcal{B}_q$ from (43). Analogous meaning have the numbers c and d.

A fundamental example of the GUHA quantifier is $\sim := \sim_x$ which is a *symmetric associational quantifier* taken as true if $ad > bc$. Another example is the *binary multitudinal quantifier* $\sim := \sqsubset_r^\gamma$ where $\gamma \in [0, 1]$ and $r \in \mathbb{N}$ are parameters, and which relates to implication. We take $\sqsubset_r^\gamma$ as true, if $a > \gamma(a + b)$ and $a > r$. The parameter γ characterizes ratio of objects for which the number of positive occurrences of $\mathcal{C}$ as well as $\mathcal{D}$ is greater than the number of positive occurrences of $\mathcal{C}$ but negative occurrences of $\mathcal{D}$. This definition follows from the truth table of classical implication which gives *false* if the antecedent is *true* but the consequent is *false*, and *true* otherwise. The value of $\sqsubset_r^\gamma$ thus decreases if the number of negative cases of the succedent $\mathcal{D}$ increases.

The second parameter r characterizes relevancy of the portion of the data entering the test. For example, consider the data with $m = 1000$ but $a = 4$ and $b = 1$. Then $\sqsubset_r^\gamma$ becomes true for $\gamma = 0.75$ (quite high number) but in fact, the number of positive cases is negligible with the total number of objects. We will usually specify r as percentage of a w.r.t. m. For the details and properties of the mentioned quantifiers, see [4, 5]. Note that there are many other quantifiers described in [5] and the other related literature.

On the basis of data (40) we can extract a linguistic association

$$\mathcal{C}(Y_1, \ldots, Y_p) \sqsubset_r^\gamma \mathcal{D}(Z_1, \ldots, Z_q) \tag{45}$$

which can be taken as a hypothesis about validity of the fuzzy IF-THEN rule

(9).

5.2 *Reduction of number of the discovered linguistic associations*

5.2.1 *General problem*

One of the encountered practical problems is abundance of the discovered associations. Therefore it is desirable to find methods how to reduce them and, at the same time, not to loose information. The solution is twofold: first, on the basis of general logical properties, we can reduce the number of found associations. Second, on the basis of their semantical meaning we can reduce the number of associations that will be *presented* to the user.

We will consider three sets of associations that can be found in the given data:

(i) The set K of *all* linguistic associations that are true (in the given data).
(ii) The set K^M of all *discovered* linguistic associations.
(iii) The set K^P of all *presented* linguistic association.

Clearly,

$$K^P \subseteq K^M \subseteq K.$$

To generate K effectively, we may proceed as follows. First, we consider two disjoint sets of attributes: the sets $\{Y_1, \ldots, Y_p\}$ of independent and $\{Z_1, \ldots, Z_q\}$ dependent attributes. Then we systematically search the associations (45) starting from the shortest (one element) conjunctions $\mathcal{A}$ and $\mathcal{B}$ having the narrowest $\mathcal{B}$ in the sense of Subsection 3.3. The sets K^M and K^P are obtained using rules of logical entailment and semantic reduction described below.

5.2.2 *Logical entailment*

The rules of logical entailment can be used for reduction of the size of the set K^M. Let $\mathcal{A} \sqsubset_r^\gamma \mathcal{B} \in K$. If this fact necessarily implies that $\mathcal{C} \sqsubset_r^\gamma \mathcal{D} \in K$, i.e. that it must also be true then the former entails the latter and we write

$$(\mathcal{A} \sqsubset_r^\gamma \mathcal{B}) \vdash (\mathcal{C} \sqsubset_r^\gamma \mathcal{D}). \tag{46}$$

The relation $\vdash$ means that if the association on the left-hand side is true in the data then the association on the right-hand side is also true on the basis of general logical properties, and so, we need not test it.

Theorem 1
Let $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}$ be linguistic predications.

(a) If $\mathcal{D} \preceq \mathcal{D}'$ then $(\mathcal{C} \sqsubset_r^\gamma \mathcal{D}) \vdash (\mathcal{C} \sqsubset_r^\gamma \mathcal{D}')$.
(b) $(\mathcal{C} \sqsubset_r^\gamma \mathcal{D}) \vdash (\mathcal{C} \sqsubset_r^\gamma \mathcal{D}$ OR $\mathcal{B})$.
(c) Let $\mathcal{A}$ AND $\mathcal{B} \sqsubset_s^\gamma \mathcal{C}$ be empty if there is no object having $\mathcal{A}$ AND $\mathcal{B}$. Then $(\mathcal{A} \sqsubset_r^\gamma \mathcal{C},\ \mathcal{B} \sqsubset_r^\gamma \mathcal{C},\ \mathcal{A}\ \mathsf{AND}\ \mathcal{B} \sqsubset_s^\gamma \mathcal{C}) \vdash (\mathcal{A}\ \mathsf{OR}\ \mathcal{B} \sqsubset_r^\gamma \mathcal{C})$, where $s \leq r$.

PROOF: (a) Let $M_{\mathcal{C}}$ denote a set of all objects having the property $\mathcal{C}(Y_1, \dots, Y_p)$ and similarly $M_{\mathcal{D}}$, $M_{\mathcal{C}\ \mathsf{AND}\ \mathcal{D}}$, etc., and denote their cardinalities by $|\cdot|$. Then we obtain

$$M_{\mathcal{C}\ \mathsf{AND}\ \mathcal{D}'} = M_{\mathcal{C}\ \mathsf{AND}\ \mathcal{D}} \cup (M_{\mathcal{C}\ \mathsf{AND}\ \mathcal{D}'} - M_{\mathcal{C}\ \mathsf{AND}\ \mathcal{D}}),$$
$$M_{\mathcal{C}\ \mathsf{AND}\ \neg\mathcal{D}} = M_{\mathcal{C}\ \mathsf{AND}\ \neg\mathcal{D}'} \cup (M_{\mathcal{C}\ \mathsf{AND}\ \mathcal{D}'} - M_{\mathcal{C}\ \mathsf{AND}\ \mathcal{D}}).$$

Hence, $a = |M_{\mathcal{C}\ \mathsf{AND}\ \mathcal{D}}| \leq a' = |M_{\mathcal{C}\ \mathsf{AND}\ \mathcal{D}'}|$ and $b' = |M_{\mathcal{C}\ \mathsf{AND}\ \neg\mathcal{D}'}| \leq b = |M_{\mathcal{C}\ \mathsf{AND}\ \neg\mathcal{D}}|$. Hence, $r \leq a'$ and $a'/(a' + b') > \gamma$.

(b) is a consequence of (a) (cf. also [5]).

(c) We have

$$M_{(\mathcal{A}\ \mathsf{OR}\ \mathcal{B})\ \mathsf{AND}\ \mathcal{C}} = M_{\mathcal{A}\ \mathsf{AND}\ \mathcal{C}} \cup M_{\mathcal{B}\ \mathsf{AND}\ \mathcal{C}},$$
$$M_{(\mathcal{A}\ \mathsf{OR}\ \mathcal{B})\ \mathsf{AND}\ \neg\mathcal{C}} = M_{\mathcal{A}\ \mathsf{AND}\ \neg\mathcal{C}} \cup M_{\mathcal{B}\ \mathsf{AND}\ \neg\mathcal{C}}.$$

Denote $a = |M_{\mathcal{A}\ \mathsf{AND}\ \mathcal{C}}|$, $a' = |M_{\mathcal{B}\ \mathsf{AND}\ \mathcal{C}}|$, $a'' = |M_{(\mathcal{A}\ \mathsf{OR}\ \mathcal{B})\ \mathsf{AND}\ \mathcal{C}}|$, similarly for b, b', b'', and $d = |M_{(\mathcal{A}\ \mathsf{AND}\ \mathcal{B})\ \mathsf{AND}\ \mathcal{C}}|$, $e = |M_{(\mathcal{A}\ \mathsf{AND}\ \mathcal{B})\ \mathsf{AND}\ \neg\mathcal{C}}|$. Then $a + a' = a'' + d$ and $b + b' = b'' + e$.

Since, by the assumption, $a > \gamma(a + b)$ and $a' > \gamma(a' + b')$, we get $a'' > \gamma(a'' + b'') + (\gamma(d + e) - d)$ which is fulfilled because $d > \gamma(d + e)$ by the assumption (if $M_{\mathcal{A}\ \mathsf{AND}\ \mathcal{B}} = \emptyset$ then $d = e = 0$). □

The set $K^M \subset K$ is a minimal set of associations from which the associations from $K - K^M$ entail using the relation of logical entailment. We may thus put $K^P = K^M$.

5.2.3 Semantic reduction rules

The set K^P of presented associations may be further reduced using semantic reduction rules. For this purpose, we introduce a relation of semantic entailment $\models$. The basic idea combines discovered associations with their (linguistic) meaning due to the fact that the associations consists of evaluating linguistic expressions. Essential role is played by the theory of their ordering introduced in Subsection 3.3.

Let $H_1, H_2 \subset K$ be two sets of discovered associations. If the meaning of any of the association from H_2 is covered by the meaning of any of the associations from H_1 then we say that associations from H_1 are *more informative* than associations from H_2 (the latter are *less informative* than the former) and write

$$H_1 \models H_2. \tag{47}$$

As a special case, if $H_1 = \{\mathcal{A} \sim \mathcal{B}\}$ and $H_2 = \{\mathcal{C} \sim \mathcal{D}\}$ then we will write

$$(\mathcal{A} \sim \mathcal{B}) \models (\mathcal{C} \sim \mathcal{D}).$$

Unfortunately, the structure of the data (36) may be be such that H_2 in (47) need not be true in it even in case that H_1 is true. Therefore, $\models$ can be introduced only on the set K and so, it is weaker than $\vdash$.

(i) **Rule of strong entailment**

Let $\mathcal{C}, \mathcal{D}, \mathcal{E}, \mathcal{F}$ be arbitrary linguistic predications and suppose that

$$(\mathcal{C} \sim \mathcal{D}) \vdash (\mathcal{E} \sim \mathcal{F}).$$

Then

$$(\mathcal{C} \sim \mathcal{D}) \models (\mathcal{E} \sim \mathcal{F}).$$

(ii) **Rule of specificity**

Let $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}$ be arbitrary linguistic predications such that $\mathcal{C} \preceq \mathcal{A}$ as well as $\mathcal{B} \preceq \mathcal{D}$. Let $\mathcal{A} \sqsubset_r^\gamma \mathcal{B}, \mathcal{C} \sqsubset_r^\gamma \mathcal{D} \in K$. Then

$$(\mathcal{A} \sqsubset_r^\gamma \mathcal{B}) \models (\mathcal{C} \sqsubset_r^\gamma \mathcal{D}).$$

(iii) **Rule of disjunction**

Let $\mathcal{A}_j(X)$, $j \in J$ be a set of simple evaluating predications, $\mathcal{C}$ and $\mathcal{D}$ arbitrary predications and $\mathcal{B} := \ \mathsf{OR}_{j \in J} \mathcal{A}_j$ (note that this is equivalent to (6)). Let $H = \{(\mathcal{A}_j \ \mathsf{AND}\ \mathcal{D}) \sqsubset_r^\gamma \mathcal{C} \mid j \in J\}$ be a set of linguistic associations such that $H \subset K$ as well as $(\mathcal{B}\ \mathsf{AND}\ \mathcal{D}) \sqsubset_r^\gamma \mathcal{C} \in K$. Then

$$(\mathcal{B}\ \mathsf{AND}\ \mathcal{D}) \sqsubset_r^\gamma \mathcal{C} \models H.$$

(iv) **Rule of empty predication**

Let $\mathcal{A}_j$, $j \in J$ be a set of pure evaluating predications of the form

$$\langle\text{linguistic hedge}\rangle\langle\text{atomic evaluating expression}\rangle\ X$$

containing evaluating predications

$$\langle\text{linguistic hedge}\rangle\ \mathsf{small}\ X,$$
$$\langle\text{linguistic hedge}\rangle\ \mathsf{medium}\ X,$$
$$\langle\text{linguistic hedge}\rangle\ \mathsf{big}\ X$$

where ⟨linguistic hedge⟩ is either empty or widening. If

$$H = \{\mathcal{A}_j \sqsubset_r^\gamma \mathcal{C} \mid j \in J\} \subset K$$

then

$$(X \sqsubset_r^\gamma \mathcal{C}) \models H$$

where X represents the pure ⟨noun phrase⟩ without specific predication.

5.2.3.1 Justification. Ad (i): This is a trivial rule stating that if $\mathcal{C} \sim \mathcal{D}$ is a logical consequence of $\mathcal{A} \sim \mathcal{B}$ then it must be also its semantic consequence. The rule can be justified by the fact that logical consequence is always given by general logical properties that are also transferred to semantics of the used evaluating expressions.

Ad (ii): Justification of this rule is based on the fact that narrower antecedent $\mathcal{C}$ is contained (semantically) in the antecedent $\mathcal{A}$, i.e. given a context w, each extension of $\mathcal{C}$ in w is contained in the extension of $\mathcal{A}$. Therefore, whatever we say about the latter we, at the same time, say about the former. On the other hand, narrower succedent is more precise than wider one and so, learning that something implies values that are more specific gives us more information that saying the same about wider succedent.

Ad (iii): It can be demonstrated that the rule of disjunction cannot be introduced on the basis of syntactic entailment. Semantically, however, information contained in the number of associations differing only in the antecedent is the same as that in one rule where its antecedent is union of the antecedents of the rest of the discovered associations.

Ad (iv): The rule of empty predication is justified by the fact that the evaluating trichotomy covers the whole universe. Therefore, if the conditions of this rule are fulfilled then the whole attribute implies $\mathcal{C}$ and so, we need not consider any more specific predication of it.

5.2.3.2 Examples. Let $\mathcal{A}$:= '*small X*', $\mathcal{B}$:= '*big Y*', $\mathcal{C}$:= '*very small X*' and $\mathcal{D}$:= '*roughly big Y*'. Then, by *rule of specificity*,

$$\textit{small } X \sqsubset_r^\gamma \textit{ big } Y \tag{48}$$

is more informative than

$$\textit{very small } X \sqsubset_r^\gamma \textit{ big } Y \tag{49}$$

because everything which is very small is at the same time small and so (49) is already included in (48). On the other hand, by the same rule

$$\textit{small } X \sqsubset_r^\gamma \textit{ big } Y \tag{50}$$

is more informative than

$$\textit{small } X \sqsubset_r^\gamma \textit{ roughly big } Y \tag{51}$$

because (50) tells *more precisely*, what values of Y are related to small values of X than (51).

Let X be "heating" and Y be "temperature of melted metal". If we find associations

$$\begin{aligned} \textsf{weak heating} \sqsubset_r^\gamma \textsf{ medium temperature of melted metal} \\ \textsf{roughly medium heating} \sqsubset_r^\gamma \textsf{ medium temperature of melted metal} \\ \textsf{more or less strong heating} \sqsubset_r^\gamma \textsf{ medium temperature of melted metal} \end{aligned} \tag{52}$$

then, by the rule of *empty predication*,

$$\textsf{heating} \sqsubset_r^\gamma \textsf{ medium temperature of melted metal}$$

is more informative than all the associations (52) together.

The set K^M is obtained from K when omitting all associations that *syntactically entail* from those already discovered. The set K^P of presented associations is obtained from K^M using the relation $\models$. Namely, we may reduce the set K^M by all associations that are less informative. More precisely, if $H_1, H_2 \subseteq K^M$ and $H_1 \models H_2$ then we put into K^P all associations from H_1 and no association from H_2.

5.3 Mining linguistic associations with fuzzy numbers

In this subsection, we will use fuzzy transform described in Section 4 for searching linguistic associations. Let us consider attributes $X_{i_1}, \ldots, X_{i_p}$ and X_k. To simplify the explanation, we will rename them in this section into $Y_1, \ldots, Y_p$ and Z, respectively. The values f_{jk}, $j = 1, \ldots, m$ of the attribute X_k are renamed into $f_{j,Z}$.

The F-transform enables us to seek associations from data (36) in the form

$$(Y_1 \text{ is } \mathcal{A}_{1,l_1}) \ \textsf{AND} \cdots \textsf{AND} \ (Y_p \text{ is } \mathcal{A}_{p,l_p}) \overset{F}{\sim}_\gamma (\mathcal{B} \ \textsf{average} \ Z) \tag{53}$$

where $\mathcal{A}_{i,l_1}, \ldots, \mathcal{A}_{i,l_p}$ are fuzzy numbers with the respective intensions $\mathrm{Int}(\mathcal{A}_{i,l_1}) = \mathrm{Fn}_{\nu,y_{1,l_1}}, \ldots, \mathrm{Int}(\mathcal{A}_{i,l_p}) = \mathrm{Fn}_{\nu,y_{1,l_p}}$ and $\mathcal{B}$ is an evaluating expression (the subscripts $l_1, \ldots, l_p$ will be specified below). The parameter γ is a degree of support defined below.

5.3.0.3 Determining antecedent. The attributes $Y_1, \dots, Y_p$ represent independent variables and Z a dependent one. For each of them, data (36) provides a corresponding context $w_1, \dots, w_p, w_Z$, respectively. Furthermore, for each of these contexts we determine a number $s_1, \dots, s_p, s_Z$ of nodes and find the nodes

$$\begin{aligned} &y_{1,1} \in w_1, \dots, y_{1,s_1} \in w_1, \\ &\dots\dots\dots\dots\dots\dots\dots \\ &y_{p,1} \in w_p, \dots, y_{1,s_p} \in w_p, \\ &y_{Z,1} \in w_Z, \dots, y_{1,s_Z} \in w_Z. \end{aligned}$$

To apply the technique of F-transform, we must construct partition of the respective contexts $w_1, \dots, w_p$ using fuzzy numbers

$$\begin{aligned} &\{\mathrm{Fn}_{\nu,y_{11}}, \dots, \mathrm{Fn}_{\nu,y_{1s_1}}\} \\ &\dots\dots\dots\dots\dots\dots\dots \\ &\{\mathrm{Fn}_{\nu,y_{p1}}, \dots, \mathrm{Fn}_{\nu,y_{ps_p}}\}. \end{aligned} \tag{54}$$

Note that the linguistic hedges ν in (54) are irrelevant because the width of the fuzzy numbers is given by the number of the respective nodes. We may assign ν a standard meaning, such as "approximately".

The result of this step are linguistic expressions

$$(Y_1 \text{ is } \mathcal{A}_{1,l_1}) \ \ \mathsf{AND} \ \cdots \ \mathsf{AND} \ \ (Y_p \text{ is } \mathcal{A}_{p,l_p}) \tag{55}$$

for all $l_1 \in \{1, \dots, s_1\}, \dots, l_p \in \{1, \dots, s_p\}$ where $\mathcal{A}_{1,l_1}, \dots, \mathcal{A}_{p,l_p}$ are linguistically specified fuzzy numbers with the respective intensions taken from (54). Each expression (55) is antecedent of a new discovered linguistic association of the form (53).

5.3.0.4 Determining consequent. The consequent is determined on the basis of components of F-transform obtained in its direct phase:

$$F_{l_1,\dots,l_p} = \frac{\sum_{j=1}^{m} \mathrm{Fn}_{\nu,y_{1,l_1}}(f_{j1}) \cdots \mathrm{Fn}_{\nu,y_{p,l_p}}(f_{jp})\ f_{j,Z}}{\sum_{j=1}^{m} \mathrm{Fn}_{\nu,y_{1,l_1}}(f_{j1}) \cdots \mathrm{Fn}_{\nu,y_{p,l_p}}(f_{jp})} \tag{56}$$

for all $l_1 \in \{1, \dots, s_1\}, \dots, l_p \in \{1, \dots, s_p\}$. Recall that $y_{1,l_1} \in w_1, \dots, y_{1,l_p} \in w_p$ are nodes in the respective contexts $w_1, \dots, w_p$ and $f_{j1}, \dots, f_{jp}, f_{j,Z}$ are values of the respective attributes $Y_1, \dots, Y_p, Z$.

Each $F_{l_1,\dots,l_p}$ is an *average value* of the associated attribute Z, so that we obtain a *new attribute* '$\mathsf{average}$ Z'. Because, $F_{l_1,\dots,l_p} \in w_Z$, it can be evaluated by an evaluating linguistic expression using the function Suit:

$$\widetilde{Ev}_{l_1,\ldots,l_p} = \mathrm{Suit}(F_{l_1,\ldots,l_p}, w_Z). \tag{57}$$

The result of (57) is a consequent (evaluating predication) of the form (38):

$$\mathcal{B}_{l_1,\ldots,l_p} \ \textsf{average} \ Z \tag{58}$$

where $\mathcal{B}_{l_1,\ldots,l_p}$ is a linguistic expression with the intension $\mathrm{Int}(\mathcal{B}_{l_1,\ldots,l_p}) = \widetilde{Ev}_{l_1,\ldots,l_p}$.

From (55) and (58), we obtain for all $l_1 \in \{1,\ldots,s_1\},\ldots,l_p \in \{1,\ldots,s_p\}$ linguistic associations of the form (53).

It remains to specify the parameter γ that characterizes *degree of support* of $l_1 \ldots l_p$-th association. Let $n_{l_1,\ldots,l_p}$ be a number of objects o_j, for which the membership degree

$$\mathrm{Fn}_{\nu,y_{1,l_1}}(f_{j1}) \cdots \mathrm{Fn}_{\nu,y_{p,l_p}}(f_{jp}) > 0.$$

Then we set

$$\gamma = \frac{n_{l_1,\ldots,l_p}}{m}. \tag{59}$$

We will put into the set K of all discovered associations each association (53) for which $\gamma > \alpha$ where α is a specified threshold. To obtain the set $K^P \subset K$, we can use all semantic reduction rules introduced in Subsection 5.2.

6 Some practical experiences and hints for further development

To test the method proposed in the paper, we have developed a program LAMWin that is based on the system LFLC 2000 (University of Ostrava, Czech Republic). For testing, we have chosen the Boston Housing dataset taken from the StatLib library which is maintained at Carnegie Mellon University. The creators are D. Harrison and D. L. Rubinfeld and detailed description of the data including results of their analysis can be found in [7].

The data set concerned housing values in suburbs of Boston. The number of objects is 506 (without missing values). The following 14 attributes are measured on each object:

- X_1 (CRIM) — per capita crime rate by town
- X_2 (ZN) — proportion of residential land zoned for lots over 25,000 sq.ft.
- X_3 (INDUS) — proportion of non-retail business acres per town

X_4 (CHAS) — Charles River dummy variable (= 1 if tract bounds river; 0 otherwise)

X_5 (NOX) — Nitric oxides concentration (parts per 10 million)

X_6 (RM) — average number of rooms in owner units

X_7 (AGE) — Proportion of owner-occupied units built prior to 1940

X_8 (DIS) — Weighted distances to five Boston employment centers in Boston region

X_9 (RAD) — Index of accessibility to radial highways

X_{10} (TAX) — full-value property-tax rate per \$ 10,000

X_{11} (PTRATIO) — Pupil-teacher ratio by town school district

X_{12} (B) — Black proportion of population

X_{13} (LSTAT) — Proportion of population that is lower status = 1/2

Y (MEDV) — Housing value

The attribute Y (MEDV) is dependent, the other 13 attributes are regarded as independent.

During the testing it turned out that the data are not too rich and so, to obtain some results, we had to set the parameters $\gamma = 0.2$ and $r = 0.005$ (the portion of objects having a in the four-fould table (44)). This fact also significantly influences the reduction rate when transforming the data into (40) in comparison with the original data (36). For example, when keeping all 14 attributes then (40) reduced to 500 objects only. When considering 10 attributes, it reduced to 481 objects and with 8 attributes it reduced to 469 objects. There is no general way how to compute the resulting size of (40) since it heavily depends on the data and richness of information they contain.

First, we have tested the method for searching the pure linguistic associations. In 2 runs, we have discovered altogether 233 associations. This number has reduced to 85 using the above semantical reduction rules which demonstrates that they are very effective. Examples of our method are below.

6.0.0.5 Group of associations A. We have tested possible associations of the form

$$(\mathcal{A}\,X_1) \sqsubset^{0.2}_{0.005} (\mathcal{C}\,Y) \tag{60}$$

(per capita crime rate by town "implies" median value of owner-occupied homes in \$ 1000). The context w_1 of the variable X_1 is computed from the smallest and biggest observed values: $v_L = 0.006$, $v_R = 88.976$ and we set $v_S = 35.694$ †) . The context w_Y of Y is $v_L = 5$, $v_S = 23$, $v_R = 50$.

Alltogether 6 linguistic associations have been discovered:

$$A1. \quad (ex\ sm\ X_1) \sqsubset^{0.2}_{0.005} (me\ Y)$$

$$A2. \quad (ex\ sm\ X_1) \sqsubset^{0.2}_{0.005} (ml\ me\ Y)$$

$$A3. \quad (si\ sm\ X_1) \sqsubset^{0.2}_{0.005} (me\ Y)$$

$$A4. \quad (sm\ X_1) \sqsubset^{0.2}_{0.005} (qr\ sm\ Y)$$

$$A5. \quad (sm\ X_1) \sqsubset^{0.2}_{0.005} (ro\ sm\ Y)$$

$$A6. \quad (ve\ sm\ X_1) \sqsubset^{0.2}_{0.005} (vr\ sm\ Y)$$

(for the used shorts — see Subsection 3.3). Using Rule of specificity, we get

$$K^P = \{A3, A5\}. \tag{61}$$

For example,

$$(ex\ sm\ X_1 \sqsubset^{0.2}_{0.005} me\ Y) \models (ex\ sm\ X_1 \sqsubset^{0.2}_{0.005} ml\ me\ Y)$$

because *medium* is more specific than *more or less medium* and so, association A1 is more informative than A2. At the same time,

$$(si\ sm\ X_1 \sqsubset^{0.2}_{0.005} me\ Y) \models (ex\ sm\ X_1 \sqsubset^{0.2}_{0.005} me\ Y)$$

because *significantly small* is less specific than *extremely small* and so, association A3 is more informative than A1. These two remaining associations can be linguistically expressed as follows:

A3: *Significantly small per capita crime rate by town "imply" medium housing value.*

A5: *Small per capita crime rate by town "imply" roughly small housing value.*

It can also be concluded from all 6 associations that the antecedent X_1(CRIM) varies in a way opposite to the succedent Y(MEDV). This corresponds to the observation that CRIM is negatively correlated with MEDV.

---

†) This value is computed as 40% of the length of interval $[0.006, 88.976]$ because people generally distinguish better small values than gib ones. Of course, our theory enables to set v_L arbitrarily.

6.0.0.6 Group of associations B. Other tested associations are

$$(\mathcal{A}\,X_1)\ \mathsf{AND}\ (\mathcal{B}\,X_2) \sqsubset^{0.2}_{0.005} (\mathcal{C}\,Y) \tag{62}$$

(per capita crime rate by town AND proportion of residential land zoned for lots over 25,000 sq.ft. "imply" median value of owner-occupied homes in $ 1000). The context of X_1 and Y is the same as above. The context w_2 of X_2 is $v_L = 0$, $v_S = 40$, $v_R = 100$.

Altogether 13 linguistic associations have been discovered:

B1. $(\mathit{ex\ sm}\ X_1)$ AND $(\mathit{ze}\ X_2) \sqsubset^{0.2}_{0.005} (\mathit{me}\ Y)$

B2. $(\mathit{ex\ sm}\ X_1)$ AND $(\mathit{sm}\ X_2) \sqsubset^{0.2}_{0.005} (\mathit{me}\ Y)$

B3. $(\mathit{ex\ sm}\ X_1)$ AND $(\mathit{ze}\ X_2) \sqsubset^{0.2}_{0.005} (\mathit{ml\ me}\ Y)$

B4. $(\mathit{ex\ sm}\ X_1)$ AND $(\mathit{ro\ sm}\ X_2) \sqsubset^{0.2}_{0.005} (\mathit{vr\ bi}\ Y)$

B5. $(\mathit{ex\ sm}\ X_1)$ AND $(\mathit{qr\ sm}\ X_2) \sqsubset^{0.2}_{0.005} (\mathit{ml\ me}\ Y)$

B6. $(\mathit{ex\ sm}\ X_1)$ AND $(\mathit{qr\ sm}\ X_2) \sqsubset^{0.2}_{0.005} (\mathit{vr\ bi}\ Y)$

B7. $(\mathit{ex\ sm}\ X_1)$ AND $(\mathit{bi}\ X_2) \sqsubset^{0.2}_{0.005} (\mathit{me}\ Y)$

B8. $(\mathit{ex\ sm}\ X_1)$ AND $(\mathit{vr\ sm}\ X_2) \sqsubset^{0.2}_{0.005} (\mathit{ml\ me}\ Y)$

B9. $(\mathit{ex\ sm}\ X_1)$ AND $(\mathit{me}\ X_2) \sqsubset^{0.2}_{0.005} (\mathit{me}\ Y)$

B10. $(\mathit{si\ sm}\ X_1)$ AND $(\mathit{ze}\ X_2) \sqsubset^{0.2}_{0.005} (\mathit{me}\ Y)$

B11. $(\mathit{sm}\ X_1)$ AND $(\mathit{ze}\ X_2) \sqsubset^{0.2}_{0.005} (\mathit{qr\ sm}\ Y)$

B12. $(\mathit{ra\ sm}\ X_1)$ AND $(\mathit{ze}\ X_2) \sqsubset^{0.2}_{0.005} (\mathit{ro\ sm}\ Y)$

B13. $(\mathit{ve\ sm}\ X_1)$ AND $(\mathit{ze}\ X_2) \sqsubset^{0.2}_{0.005} (\mathit{vr\ sm}\ Y)$

(the short "ra" means *rather* †)). Using Rule of specificity, we obtain

$$K^P = \{B2, B6, \ldots, B10, B12\}.$$

Using Rule of empty predication, we furthermore replace associations B2, B7 and B9 by

$$B0 := (\mathit{ex\ sm}\ X_1)\ \mathsf{AND}\ X_2 \sqsubset^{0.2}_{0.005} (\ \mathit{me}\ Y)$$

†) This is a specific hedge between "ve" and "empty" which has been omitted from our discussion above to simplify the explanation.

which can be read as *extremely small per capita crime rate by town AND proportion of residential land zoned for lots over 25,000 sq.ft. "imply" medium housing value*. Therefore, we finally obtain

$$K^P = \{B0, B6, B8, B10, B12\}.$$

We have also tested the method for finding associations with fuzzy numbers using F-transform due to Subsection 5.3. For comparison, we will present associations concerning the same associations as above.

6.0.0.7 Group of associations A'. This group of associations has the structure

$$(X_1 \text{ is } \mathcal{A}_l) \overset{F}{\sim}_{0.1} (\mathcal{B} \textsf{ average } Z).$$

The contexts w_1 and w_Y are the same as in case A. The number of nodes in the context w_1 was 10. The following associations were found:

A'1. $(X_1$ *is about 0.006*$) \overset{F}{\sim}_{0.1}$ *(me average* Y*)*

A'2. $(X_1$ *is about 9.892*$) \overset{F}{\sim}_{0.1}$ *(qr sm average* Y*)*

For comparison, we have changed the number of nodes of w_1 to 8. The results are the following:

A'1. $(X_1$ *is about 0.006*$) \overset{F}{\sim}_{0.1}$ *(me average* Y*)*

A'2. $(X_1$ *is about 12.716*$) \overset{F}{\sim}_{0.1}$ *(qr sm average* Y*)*

One may see that these results are in good accordance with the associations (61). This means that both methods, though of different principles, give analogous results. On the other hand, the method using F-transform provides a more detailed view which can utilized in further precise analysis.

6.0.0.8 Group of associations B'. We have similarly tested associations of the form

$$(X_1 \text{ is } \mathcal{A}_{1l_1}) \textsf{ AND } (X_2 \text{ is } \mathcal{A}_{2l_2}) \overset{F}{\sim}_{0.1} (\mathcal{B} \textsf{ average } Z).$$

The contexts w_1, w_2 and w_Y are the same as in case B. The number of nodes in both contexts was 10. The following associations were found:

B'1. $(X_1$ *is about 0.006)* AND $(X_2$ *is about 0.0)* $\overset{F}{\sim}_{0.1}$ *(me average* $Y)$

B'2. $(X_1$ *is about 9.892)* AND $(X_2$ *is about 0.0)* $\overset{F}{\sim}_{0.1}$ *(qr sm average* $Y)$

B'3. $(X_1$ *is about 0.006)* AND $(X_2$ *is about 22.222)* $\overset{F}{\sim}_{0.1}$ *(vr bi average* $Y)$

B'4. $(X_1$ *is about 9.892)* AND $(X_2$ *is about 22.222)* $\overset{F}{\sim}_{0.1}$ *(ro bi average* $Y)$

Also this test demonstrates good agreement with test B above. Namely, B'1 corresponds to B1, B'2 to B11, B'3 to B4 and B'4 to B6.

Both examples verify that F-transform provides associations that are more detailed because they contain fuzzy numbers that are semantically more precise than rough pure evaluating expressions. On the other hand, the results of both methods are in correspondence. Therefore, when combining them we obtain a very informative complex view on the data.

7 Conclusion

In this paper, we have developed a method for searching associations from numerical data that are expressed in natural language. Such associations are called "linguistic". We have applied three theories, namely the theory of evaluating linguistic expressions, GUHA method and F-transform.

One of essential outcomes of our theory is high understandability of the discovered associations that are formulated in natural language and so, they can serve experts from various fields to discover new relations of dependencies in a way that is much closer to their knowledge and the way of their thinking. Moreover, those discovered associations that characterize real dependencies can be directly taken as fuzzy IF-THEN rules and used as expert knowledge about the problem.

Another outcome can be smaller size of the data, provided that we prefer a global characterization of information contained in the (originally numerical) data (note that this is often the case). The profit, however, depends a lot on richness of the contained information.

The paper is a first attempt to discover linguistic associations directly and it opened various problems that should be studied. One of the first problems is extent of the considered part of natural language and its expansion to other kinds of expressions; possibly, to apply the theory of generalized (linguistic) quantifiers. Further problem is development of other syntactic as well as semantic reduction rules using which we can significantly reduce the amount of information presented to the user without loosing the discovered information.

There are also a lot of results in classical GUHA method that can be considered for the use in our method. Still another problem concerns the question, whether we can generalize GUHA using the formal theory of fuzzy logic (cf. [4]); what would be outcome of such generalization and how can it contribute to our theory of linguistic associations? We can thus conclude that the results are encouraging and they open a vast field for further intensive research.

References

[1] Carnap, R., Meaning and Necessity: A Study in Semantics and Modal Logic, University of Chicago Press, Chicago 1947.

[2] Dvořák, A. and Novák, V., Fuzzy Logic Deduction with Crisp Observations. Soft Computing 8(2004), 256–263.

[3] Hájek, P. The question of a general concept of the GUHA method, Kybernetika 4(1968), 505-515 (in Czech).

[4] Hájek, P. Logics for data mining (GUHA rediviva). Submitted.

[5] Hájek, P., Havránek, T. Mechanizing Hypothesis Formation (Mathematical Foundations for a General Theory). Springer-Verlag, Berlin-Heidelberg-New York 1978.

[6] Hájek, P. and Novák, V., The Sorites paradox and fuzzy logic. Int. J. of General Systems 32(2003), 373-383.

[7] Harrison, D., Rubinfeld, D. L., Hedonic Housing Prices and the Demand for Clean Air. J. of Environmental Economics and Manag. 5(1978), 81–102.

[8] Materna, P., Concepts and Objects. Acta Philosophica Fennica 63, Helsinki 1998.

[9] Novák, V., Fuzzy Sets and Their Applications. Bristol, Adam Hilger 1989.

[10] Novák, V.: A Formal Integrated Theory of Approximate Reasoning. Proc. of V^{th} IFSA World Congress, Seoul, Korea 1993, 399–402.

[11] Novák, V., Antonyms and Linguistic Quantifiers in Fuzzy Logic. Fuzzy Sets and Systems, 124(2001), pp. 335–351.

[12] Novák, V.: On fuzzy type theory. Fuzzy Sets and Systems 149(2005), 235–273.

[13] Novák, V.: Fuzzy logic deduction with words applied to ancient sea level estimation. In: Demicco, R. and Klir, G.J. (Eds), Fuzzy logic in geology. Academic Press, Amsterdam 2003, 301–336.

[14] Novák, V.: A Comprehensive Theory of Evaluating Linguistic Expressions. (submitted)

[15] Novák, V., Perfilieva I. and J. Močkoř, Mathematical Principles of Fuzzy Logic. Kluwer, Boston/Dordrecht 1999.

[16] Novák, V. and I. Perfilieva: Evaluating Linguistic Expressions and Functional Fuzzy Theories in Fuzzy Logic. In: L. A. Zadeh and J. Kacprzyk (eds.): Computing with Words in Information/Intelligent Systems 1.

Springer-Verlag, Heidelberg 1999, 383–406.

[17] Perfilieva I. Fuzzy transform: Application to Reef Growth Problem. In: Demicco R, Klir G. J., editors. Fuzzy logic in geology. Academic Press, Amsterdam 2003, 275–300.

[18] Perfilieva, I. Fuzzy Transforms. In: Dubois, D. et al., editors. Rough and Fuzzy Reasoning: Rough versus Fuzzy and Rough and Fuzzy. Springer-Verlag, Heidelberg 2004.

[19] Perfilieva, I. Fuzzy Transforms: Theory and Applications. Fuzzy Sets and Systems (to appear).

[20] Zadeh LA. The concept of a linguistic variable and its application to approximate reasoning I, II, III. Information Sciences: 8(1975), 199–257, 301–357; 9, 43–80.

[21] Zadeh, L.A., Toward a Logic of Perceptions Based on Fuzzy Logic, in: Novák, V. and I. Perfilieva (eds.), Discovering the World With Fuzzy Logic. Springer-Verlag, Heidelberg 2000 (Studies in Fuzziness and Soft Computing, Vol. 57), 4–28.