*Research Article*

# Mining Linguistic Associations for Emergent Flood Prediction Adjustment

## Michal Burda, Pavel Rusnok, and Martin Štěpnička

*Institute for Research and Applications of Fuzzy Modeling, National Supercomputing Center IT4Innovations,*
*Division of University of Ostrava, 30. dubna 22, 701 03 Ostrava, Czech Republic*

Correspondence should be addressed to Martin Štěpnička; martin.stepnicka@osu.cz

Floods belong to the most hazardous natural disasters and their disaster management heavily relies on precise forecasts. These forecasts are provided by physical models based on differential equations. However, these models do depend on unreliable inputs such as measurements or parameter estimations which causes undesirable inaccuracies. Thus, an appropriate data-mining analysis of the physical model and its precision based on features that determine distinct situations seems to be helpful in adjusting the physical model. An application of fuzzy GUHA method in flood peak prediction is presented. Measured water flow rate data from a system for flood predictions were used in order to mine fuzzy association rules expressed in natural language. The provided data was firstly extended by a generation of artificial variables (features). The resulting variables were later on translated into fuzzy GUHA tables with help of Evaluative Linguistic Expressions in order to mine associations. The found associations were interpreted as fuzzy IF-THEN rules and used jointly with the Perception-based Logical Deduction inference method to predict expected time shift of flow rate peaks forecasted by the given physical model. Results obtained from this adjusted model were statistically evaluated and the improvement in the forecasting accuracy was confirmed.

## 1. Introduction

Disaster management is generally becoming more and more important task. Among many natural disasters, floods are the one of the most hazardous, and moreover, one of the most frequently occurring in the region of the central Europe. Researchers invest enormous efforts into investigation of distinct flood models that would help to forecast floods and thus provide the disaster management with a reliable decision support that could be used in order to prevent further deceases and material costs.

One of such long-term researches focusing on the disaster management and especially on modeling and forecasting floods gave rise to the creation of the FLOREON, a system for emergent flood prediction [1]. No matter how sophisticated the system is, due to the natural imprecision in data sources (e.g., measuring stations) and due to the natural imprecision in parameter setting (crisp values determined by an expert decision), and having in mind how complicated the whole problem is, it necessarily provides forecasts that are not always precise.

Therefore, it seems to be appropriate to focus on some analysis of the performance of the system that could give at least a vague idea under which conditions the system works, under which conditions it provides us with a certain imprecision, and under which conditions we are able to correct the forecast. Based on the sources of the imprecision, it seems that an appropriate data-mining technique that would involve fuzziness might provide us with promising results and is worth of being attempted. In this investigation, we face the above foreshadowed problem with the help of the fuzzy GUHA method, that is, a specific variant of associations mining technique that allows using the concepts of fuzzy logic in a broad sense [2].

*1.1. Brief Problem Description.* The data being analyzed come from the measures of water flow rate of the Odra River in Ostrava, Czech Republic. Measuring stations provide us with a flow rate $[m^3/s]$ on hourly basis. The goal is to forecast a future flow rate. This is done by the so-called Math-1D model [3] developed for the FLOREON disaster management IT system [1].

The Math-1D model is a differential equation based model of the flow rate. In order to provide us with flow rate forecasts, it uses information about precipitations (past and forecasted), soil type, river bank shape, and other parameters. Although it is a well-established physical model that is empirically examined, it is not sufficiently reliable. The reason does not lie in the model but in the fact that most of the parameters and input data are highly imprecise. For example, the soil type is provided by a hydrologist expert but, due to natural limitations, without a deeper geological analysis and, moreover, the provided soil type is the same for the whole river flow.

Having in mind these limitations, the Math-1D model forecasts depend mainly on the measured past precipitations and flow rates and on the forecasted precipitations. Thus, provided forecast, though often reliable, may be even highly imprecise. The imprecision may be viewed in two perspectives: in the vertical one and the horizontal one. The vertical imprecision actually means either the overestimation or even worse the underestimation of the flow rate in the culminating peak. For our investigation, the second, that is, the horizontal imprecision, is crucial. That is, we focus on the precision in terms of time; that is, we focus on the question whether and under which conditions the model forecasts the peak discharge earlier or later and how big is the time shift of the peak.

The vertical as well as horizontal imprecision may be significant. As one can see from an exemplary forecast in Figure 1, the real culminating peak can appear a few hours sooner than forecasted. Let us note that the Math-1D model does not use the knowledge of the water flow rate in the past and depends mainly on the precipitations. This explains why it may happen that the model does not fit well the past data (from −119th hour to 0th; see Figure 1). On the other hand, precipitations are rather precise compared to the data from the measuring stations that may not be well calibrated or, even worse, the measuring station may be partially damaged or even fully destroyed (occasionally, it happens that even during a massive flood, measuring stations provide a zero flow rate measurements).

Our task is to analyze and forecast the peak shift on the horizontal (time) axis. In other words, the task is to build a model that would (based on the flow rate measurements and the Model-1D performance in the past) provide disaster management with a valuable information about possible horizontal imprecision of the Math-1D model and, moreover, that would additionally provide the disaster management with an estimation about the peak shift. This peak shift estimation could be used in the corrections of the forecasts.

## 2. Theoretical Background

In this Section, we introduce fundamental theoretical background that is used in our investigation. As there is no space to introduce all the theoretical concepts in detail, we will provide readers only with a brief introduction and refer to further valuable sources [4–7].

*2.1. Evaluative Linguistic Expressions.* One of the main constituents of systems of fuzzy/linguistic IF-THEN rules is *evaluative linguistic expressions* [4], in short *evaluative*
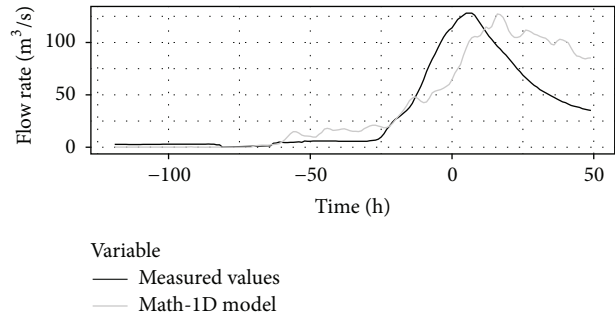


FIGURE 1: Particular example of real measured values (in black) and the Math-1D model simulation (in gray). Flow rate values $[m^3/s]$ on the vertical axis are measured on a particular measuring station placed on the Odra River starting from time −119 up to 0 (horizontal axis, time $[h]$). Starting from the time point "0" to the right, the gray curve denotes the forecast of future flow rate values.

TABLE 1: Linguistic hedges and their abbreviations.

| Narrowing effect | Widening effect |
|---|---|
| Very (Ve) | More or less (ML) |
| Significantly (Si) | Roughly (Ro) |
| Extremely (Ex) | Quite roughly (QR) |
| — | Very roughly (VR) |

*expressions*, for example, *very large*, *more or less hot,* and so forth. They are special expressions of natural language that are used whenever it is important to evaluate a decision situation, to specify the course of development of some process, and in many other situations. Note that their importance and the potential to model their meaning mathematically have been pointed out by Zadeh (e.g., in [8, 9]).

A simple form of evaluative expressions keeps the following structure:

$$\langle \text{ling} \cdot \text{hedge} \rangle \langle \text{atomic evaluative expression} \rangle . \quad (1)$$

*Atomic evaluative expressions* comprise any of the *canonical* adjectives *small*, *medium*, and *big*, abbreviated in the following as Sm, Me, and Bi, respectively.

*Linguistic hedges* are specific adverbs that make the meaning of the atomic expression more or less precise. We may distinguish hedges with *narrowing effect*, for example, *very*, *extremely*, and so forth and with *widening effect*, for example, *roughly*, *more or less* and so forth. In the following text, we, without loss of generality, use the hedges introduced in Table 1 that were successfully used in real applications [10] and that are implemented in the LFLC software package [11]. As a special case, the ⟨linguistic hedge⟩ can be empty. Note that our hedges are of so-called inclusive type [12], which means that extensions of more specific evaluative expressions are included in less specific ones; see Figure 2.

Evaluative expressions of the form (1) will generally be denoted by script letters $\mathscr{A}$, $\mathscr{B}$, and so forth. They are used to evaluate values of some variable $X$. The resulting expressions are called *evaluative linguistic predications* and have the form

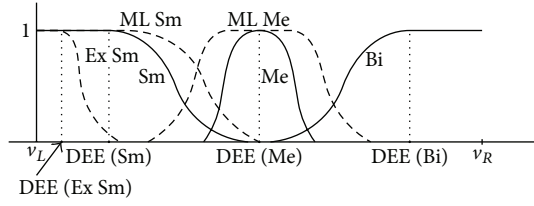$$X \text{ is } \mathscr{A}. \quad (2)$$

FIGURE 2: Shapes of extensions (fuzzy sets) of evaluative linguistic expressions. DEE denotes the defuzzified values obtained using the *Defuzzification of Evaluative Expressions*.

Examples of evaluative predications are "temperature is very high," "price is low," and so forth. The model of the meaning of evaluative expressions and predications makes distinction between *intensions* and *extensions* in various *contexts*. The context characterizes a range of possible values. This range can be characterized by a triple of numbers $\langle v_L, v_M, v_R \rangle$, where $v_L, v_M, v_R \in \mathbb{R}$ and $v_L < v_M < v_R$. These numbers characterize the minimal, middle, and maximal values, respectively, of the evaluated characteristics in the specified context of use. Therefore, we will identify the notion of context with the triple $w = \langle v_L, v_M, v_R \rangle$. By $v \in w$ we mean $v \in [v_L, v_R]$. In the sequel, we will work with a set of contexts $W \subset \{\langle v_L, v_M, v_R \rangle \mid v_L, v_M, v_R \in \mathbb{R}, v_L < v_M < v_R\}$ that are given in advance.

The intension of an evaluative predication "$X$ is $\mathscr{A}$" is a certain formula whose interpretation is a function:

$$\text{Int}(X \text{ is } \mathscr{A}) : W \longrightarrow \mathscr{F}(\mathbb{R}); \qquad (3)$$

that is, it is a function that assigns a fuzzy set to any context from the set $W$.

Given an intension (3) and a context $w \in W$, we can define the extension of "$X$ is $\mathscr{A}$" in the context $w$ as a fuzzy set:

$$\text{Int}(X \text{ is } \mathscr{A})(w) \subseteq [v_L, v_R], \qquad (4)$$

where $\subseteq$ denotes the relation of fuzzy subsethood.

*Convention 1.* For the sake of brevity and simplicity and having in mind that an extension is a fuzzy set on a given context, we will omit the notion of extension from our consideration when appropriate and write only the abbreviated forms:

$$\begin{aligned} A &:= (\text{Int}(X \text{ is } \mathscr{A})(w)), \quad w \in W, \\ A(v_0) &:= (\text{Int}(X \text{ is } \mathscr{A})(w))(v_0), \quad v_0 \in w, \end{aligned} \qquad (5)$$

if there is no danger of any confusion caused by the fact that the left-hand side does not explicitly mention the chosen context $w$ and variable $X$.

### 2.2. Linguistic Descriptions.
Evaluative predications occur in conditional clauses of natural language of the form

$$\mathscr{R} := \text{IF } X \text{ is } \mathscr{A} \text{ THEN } Y \text{ is } \mathscr{B}, \qquad (6)$$

where $\mathscr{A}, \mathscr{B}$ are evaluative expressions. The linguistic predication "$X$ is $\mathscr{A}$" is called the *antecedent* and "$Y$ is $\mathscr{B}$" is called the *consequent* of rule (6). Of course, the antecedent may consist of more evaluative predications, joined by the connective "AND." The clauses (6) will be called fuzzy/linguistic IF-THEN rules in the sequel.

Fuzzy/linguistic IF-THEN rules are gathered in a *linguistic description*, which is a set LD = $\{\mathscr{R}_1, \ldots, \mathscr{R}_m\}$, where

$$\mathscr{R}_1 := \text{IF } X \text{ is } \mathscr{A}_1 \text{ THEN } Y \text{ is } \mathscr{B}_1,$$
$$\vdots \qquad (7)$$
$$\mathscr{R}_m := \text{IF } X \text{ is } \mathscr{A}_m \text{ THEN } Y \text{ is } \mathscr{B}_m.$$

Because each rule in (7) is taken as a specific *conditional sentence of natural language*, a linguistic description can be understood as a *specific kind of a (structured) text*. This text can be viewed as a *model* of specific behavior of the system in concern.

The *intension of a fuzzy/linguistic IF-THEN rule* $\mathscr{R}$ in (6) is a function:

$$\text{Int}(\mathscr{R}) : W \times W \longrightarrow \mathscr{F}(\mathbb{R} \times \mathbb{R}). \qquad (8)$$

This function assigns to each context $w \in W$ and each context $w' \in W$ a *fuzzy relation* in $w \times w'$. The latter is an *extension* of (8).

We also need to consider a linguistic phenomenon of topic-focus articulation (cf. [13]), which in the case of linguistic descriptions requires us to distinguish the following two sets:

$$\begin{aligned} \text{Topic}_{\text{LD}} &= \left\{\text{Int}\left(X \text{ is } \mathscr{A}_j\right) \mid j = 1, \ldots, m\right\}, \\ \text{Focus}_{\text{LD}} &= \left\{\text{Int}\left(Y \text{ is } \mathscr{B}_j\right) \mid j = 1, \ldots, m\right\}. \end{aligned} \qquad (9)$$

The phenomenon of topic-focus articulation plays an important role in the inference method called perception-based logical deduction described below.

*Convention 2.* Besides the above introduced notions of topic and focus, it is sometimes advantageous to introduce the following notation:

$$\text{Topic}_{\text{LD}}^w = \left\{\text{Int}\left(X \text{ is } \mathscr{A}_j\right)(w) \mid j = 1, \ldots, m\right\}, \qquad (10)$$

which will denote the set of extensions of evaluative predications that are contained in $\text{Topic}_{\text{LD}}$ knowing the particular context $w$. This notation will be used later on when defining the function of local perception. In the view of Convention 1 one can also easily introduce the $\text{Topic}_{\text{LD}}^w$ as follows:

$$\text{Topic}_{\text{LD}}^w = \left\{A_j \mid j = 1, \ldots, m\right\}. \qquad (11)$$

### 2.3. Ordering of Linguistic Predications.
To be able to state relationships among evaluative expressions, for example, when one expression "covers" another, we need an ordering relation defined on the set of them. Let us start with the

ordering on the set of linguistic hedges. We may define the ordering $\leq_H$ of examples of hedges as follows:

$$Ex \leq_H Si \leq_H Ve \leq_H \langle empty \rangle \leq_H ML \leq_H Ro \leq_H QR \leq_H VR. \tag{12}$$

We extend the theory of evaluative linguistic expressions by the following *inclusion axiom*. Let $Ker(A)$ denote the kernel of a fuzzy set $A$. For any $w$,

$$Int \left( X \text{ is } \langle hedge \rangle_i \mathscr{A} \right)(w) \subseteq Int \left( X \text{ is } \langle hedge \rangle_j \mathscr{A} \right)(w)$$

$$Ker \left( Int \left( X \text{ is } \langle hedge \rangle_i \mathscr{A} \right)(w) \right)$$

$$\subset Ker \left( Int \left( X \text{ is } \langle hedge \rangle_j \mathscr{A} \right)(w) \right) \tag{13}$$

hold for any atomic expression $\mathscr{A}$ under the assumptions $\langle hedge \rangle_i \leq_H \langle hedge \rangle_j$, $i \neq j$.

Based on $\leq_H$ we may define an ordering $\leq_{LE}$ of evaluative expressions. Let $\mathscr{A}_i$, $\mathscr{A}_j$ be two evaluative expressions such that $\mathscr{A}_i := \langle hedge \rangle_i \mathscr{A}$ and $\mathscr{A}_j := \langle hedge \rangle_j \mathscr{A}$. Then we write

$$\mathscr{A}_i \leq_{LE} \mathscr{A}_j \tag{14}$$

if $\mathscr{A} \in \{Sm, Me, Bi\}$ and $\langle hedge \rangle_i \leq_H \langle hedge \rangle_j$.

In other words, evaluative expressions of the same type are ordered according to their specificity which is given by the hedges appearing in the expressions. If we are given two evaluative predications with an atomic expression of a different type, we cannot order them by $\leq_{LE}$.

Finally, we define the ordering $\leq_{(v_0, w)}$ of evaluative predications wrt. a given observation. Let us be given a context $w \in W$, an observation $v_0 \in w$, and two extensions $A_i$ and $A_j$ from the $Topic^w_{LD}$. We write $A_i \leq_{(v_0, w)} A_j$ either if $A_i(v_0) > A_j(v_0)$ or if $A_i(v_0) = A_j(v_0)$ and $\mathscr{A}_i \leq_{LE} \mathscr{A}_j$.

It should be noted that usually the $Topic_{LD}$ contains intensions of evaluative predications which are compound by a conjunction of more than one evaluative predication. In other words, we usually meet the following situation:

$$\left( X \text{ is } \mathscr{A}_i \right) := \left( X_1 \text{ is } \mathscr{A}_{i_1} \right) AND \cdots AND \left( X_K \text{ is } \mathscr{A}_{i_K} \right),$$

$$\left( X \text{ is } \mathscr{A}_j \right) := \left( X_1 \text{ is } \mathscr{A}_{j_1} \right) AND \cdots AND \left( X_K \text{ is } \mathscr{A}_{j_K} \right). \tag{15}$$

In this case, the ordering $\leq_{LE}$ is preserved with respect to the components:

$$\mathscr{A}_i \leq_{LE} \mathscr{A}_j \text{ if } \mathscr{A}_{i_k} \leq_{LE} \mathscr{A}_{j_k} \quad \forall k = 1, \ldots, K \tag{16}$$

and the extension of the compound linguistic predication is given as follows:

$$A_i(u_1, \ldots, u_K) = \bigwedge_{k=1}^{K} A_{ik}(u_k). \tag{17}$$

Then, the final ordering $\leq_{(v_0, w)}$ is analogous to the one-dimensional one.

### 2.4. Perception-Based Logical Deduction.

*Perception-based Logical Deduction* (abb. PbLD) is a special inference method aimed at the derivation of results based on fuzzy/linguistic IF-THEN rules. A perception is understood as an evaluative expression assigned to the given input value in the given context. The choice of perception depends on the topic of the specified linguistic description. In other words, perception is always chosen among evaluative expressions which occur in antecedents of IF-THEN rules; see [5, 10, 14].

Based on the ordering $\leq_{(v_0, w)}$ of linguistic predications, a special function of *local perception*

$$LPerc^{LD} : w \times W^K \longrightarrow \mathscr{P}\left( Topic_{LD} \right) \tag{18}$$

assigns to each value $v_0 = [v_1, \ldots, v_K] \in w$ for $w = [w_1, \ldots, w_K] \in W^K$ a subset of intensions minimal wrt. the ordering $\leq_{(v_0, w)}$

$$LPerc^{LD}(v_0, w)$$

$$= \left\{ A_i \mid A_i(v_0) > 0, \forall A_j \in Topic^w_{LD} : \left( A_j \leq_{(v_0, w)} A_i \right) \right.$$

$$\left. \Longrightarrow \left( \mathscr{A}_j = \mathscr{A}_i \right) \right\}. \tag{19}$$

Let LD be a linguistic description (7). Let us consider a context $w \in W$ for the variable $X$ and a context $w' \in W$ for $Y$. Let an observation $X = v_0$ in the context $w$ be given, where $v_0 \in w$. Then, the following *rule of perception-based logical deduction* ($r_{PbLD}$) can be introduced:

$$r_{PbLD} : \frac{LPerc^{LD}(v_0, w), LD}{C}, \tag{20}$$

where $C$ is the conclusion which corresponds to the observation in a way described below. Inputs to this inference rule are linguistic description LD and local perception $LPerc^{LD}(v_0, w)$ from (19). This local perception is formed by a set of evaluative expressions from antecedents of IF-THEN rules (i.e., from the topic) of the given linguistic description. Formula (19) chooses these antecedents which best fit the given numerical input $v_0$; in other words, they are the most specific according to the ordering $\leq_{(v_0, w)}$.

Once one or more antecedents $A_{i_\ell} \in Topic^w_{LD}$, $i_\ell = 1, \ldots, L$ are chosen according to (19), we compute for any of them conclusions $C_{i_\ell}$:

$$C_{i_\ell}(v) = A_{i_\ell}(v_0) \longrightarrow B_{i_\ell}(v), \quad v \in w', \tag{21}$$

where $\rightarrow$ is the Łukasiewicz implication [2] given by $a \rightarrow b = 1 \wedge (1 - a + b)$.

Suppose that $LPerc^{LD}(v_0, w)$ is nonempty; that is, $L > 0$. Then the final conclusion $C$ is given as the Gödel intersection of the set of all $L$ conclusions $C_{i_\ell}$ that correspond to $L$ members in $LPerc^{LD}(v_0, w)$; that is,

$$C(v) = \bigwedge_{\ell=1}^{L} C_{i_\ell}(v) = \bigwedge_{\ell=1}^{L} \left( A_{i_\ell}(v_0) \longrightarrow B_{i_\ell}(v) \right). \tag{22}$$

In many application, the inferred output fuzzy set $C$ needs to be defuzzified to a crisp value in $w'$. For this task, a special defuzzification technique called *Defuzzification of Evaluative Expressions* (abb. DEE) has been proposed. In principle, this defuzzification is a combination of *First-Of-Maxima* (FOM), *Mean-Of-Maxima* (MOM), and *Last-Of-Maxima* (LOM) that are applied based on the classification of the output fuzzy sets. Particularly, if the inferred fuzzy set is of the type `Small` (nonincreasing), the LOM is applied; if the inferred output is of the type `Big` (nondecreasing), the FOM is applied; and finally, if the inferred output is of the type `Medium`, the MOM is applied; see Figure 2.

## 3. Fuzzy GUHA: Linguistic Associations Mining

In this paper, we employ the so-called linguistic associations mining [15] for the fuzzy rule base identification. This approach, mostly known as mining association rules [16], was firstly introduced as GUHA method [17, 18]. It finds distinct statistically approved associations between attributes of given objects. Particularly, the GUHA method deals with Table 2 where $o_1, \ldots, o_n$ denote objects, $X_1, \ldots, X_q$ denote independent boolean attributes, $Z$ denotes the dependent (explained) boolean attribute, and finally, symbols $a_{ij}$ (or $a_i$) $\in \{0, 1\}$ denote whether an object $o_i$ carries an attribute $X_j$ (or $Z$) or not.

The original GUHA allowed only boolean attributes to be involved; see [19]. Since most of the features of objects are measured on the real interval, standard approach assumed categorization of quantitative variables and subsequently definition of boolean variables for every category.

The goal of the GUHA method is to search for linguistic associations of the form

$$C\left(X_1, \ldots, X_p\right) \simeq D(Z),  \tag{23}$$

where $C$, $D$ are (compound) evaluative predications [20] containing only the connective AND and $X_1, \ldots, X_p$ for $p \leq q$ are all variables occurring in $C$. The $C$, $D$ are called the *antecedent* and *consequent*, respectively. Generally, for the GUHA method, the well-known fourfold table is constructed; see Table 3.

Symbol $a$, in Table 3, denotes the number of positive occurrences of $C$ as well as $D$; $b$ is the number of positive occurrences of $C$ and of negated $D$, that is, of "not $D$." Analogous meaning has the numbers $c$ and $d$. For our purposes, only numbers $a$ and $b$ are important.

The relationship between the antecedent and the consequent is described by so-called *quantifier* $\simeq$. There are many quantifiers that characterize validity of association (23) in the data [18]. For our task, we use the so-called binary multitudinal quantifier $\simeq := \sqsubset_r^\gamma$. This quantifier is taken as true if

$$\frac{a}{a+b} > \gamma,$$

$$\frac{a}{n} > r,  \tag{24}$$

Table 2: Standard GUHA table.

|  | $X_1$ | $\ldots$ | $X_q$ | $Z$ |
|---|---|---|---|---|
| $o_1$ | $a_{11}$ | $\ldots$ | $a_{1q}$ | $a_1$ |
| $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\vdots$ |
| $o_n$ | $a_{n1}$ | $\ldots$ | $a_{nq}$ | $a_n$ |

Table 3: Classical GUHA fourfold table.

|  | $D$ | Not $D$ |
|---|---|---|
| $C$ | $a$ | $b$ |
| Not $C$ | $c$ | $d$ |

where $\gamma \in [0, 1]$ is a confidence degree and $r \in [0, 1]$ is a support degree.

*Example 1.* For example, let us consider Table 4.

Depending on the chosen confidence and support degree, the GUHA method could generate, for example, the following linguistic association:

$$C\left(\text{BMI}_{>25}, \text{Ch}_{>6.2}\right) \simeq D\left(\text{BP}_{>130/90}\right).  \tag{25}$$

According to [21], there are two approaches in treating quantitative variables in association rules mining. The first one is to categorize the variables using the predefined concept hierarchies (e.g., $\text{BMI}_{\leq 25}$). And the second one is to search for clusters in a variable and discretize it according to the found clusters (distribution of the data). Nevertheless both approaches divide numerical variables into crisp intervals.

In many situations, including our situation, it is better to define fuzzy sets on the numerical variables and use the fuzzy variant of the GUHA method [15, 22]. In this case we have also two possibilities how to treat quantitative variables. Either we will apply fuzzy clustering or we will use some predefined concepts. Because of the well-developed theory of Evaluative Linguistic Expressions (Section 2.1) we chose the latter approach.

In the fuzzy variant of the method, the attributes are not boolean but rather vague. The minimum (resp., maximum) of a particular attribute becomes $v_L$ (resp., $v_R$) and thus we obtain the context $\langle v_L, v_M, v_R \rangle$ for the given attribute ($v_M$ might be median, mean, or other value between $v_L$ and $v_R$). With canonical adjectives Sm, Me, and Bi and seven different linguistic hedges we may define more than 20 fuzzy sets for every quantitative variable (attribute). The values $a_{ij}$ (or $a_i$) are now elements of the interval $[0, 1]$ that express membership degrees.

For example, instead of defining a boolean variable $\text{BMI}_{\leq 25}$ (see Table 4), we take the quantitative variable BMI and generate all the possible evaluative linguistic predications and define fuzzy sets $\text{BMI}_{\text{ExSm}}$, $\text{BMI}_{\text{SiSm}}$, $\text{BMI}_{\text{VeSm}}$, $\text{BMI}_{\text{Sm}}, \ldots, \text{BMI}_{\text{VRSm}}$, so that the first column in Table 4 is replaced with Table 5. In this way we are able to separate a group of malnourished people ($\text{BMI}_{\text{ExSm}}$). Analogically to capture such cases of people, who have almost ideal BMI index, we define $\text{BMI}_{\text{Me}}$, $\text{BMI}_{\text{RoMe}}$. Finally, instead of $\text{BMI}_{>25}$ we define $\text{BMI}_{\text{ExBi}}$, $\text{BMI}_{\text{SiBi}}$, and so forth. Thus, we

Table 4: Example of GUHA table. $BMI_{\leq 25}$ denotes Body-Mass-Index lower or equal to 25, $BMI_{>25}$ denotes the same index above 25, $Chol_{\leq 6.2}$ denotes cholesterol lower or equal to 6.2, $Chol_{>6.2}$ denotes cholesterol higher than 6.2, $BP_{\leq 130/90}$ denotes blood pressure lower or equal to 130/90, and $BP_{>130/90}$ denotes blood pressure higher than 130/90. Objects $o_i$ are particular patients.

| | $BMI_{\leq 25}$ | $BMI_{>25}$ | $Chol_{\leq 6.2}$ | $Chol_{>6.2}$ | $BP_{\leq 130/90}$ | $BP_{>130/90}$ |
|---|---|---|---|---|---|---|
| $o_1$ | 1 | 0 | 1 | 0 | 1 | 0 |
| $o_2$ | 0 | 1 | 0 | 1 | 0 | 1 |
| $o_3$ | 0 | 1 | 1 | 0 | 0 | 1 |
| $o_4$ | 1 | 0 | 1 | 0 | 1 | 0 |
| $o_5$ | 0 | 1 | 0 | 1 | 0 | 1 |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $o_n$ | 0 | 0 | 0 | 1 | 0 | 1 |

Table 5: Example of a part of a fuzzy GUHA table with variable BMI and one canonical adjective Sm. $BMI_{ExSm}$ denotes the fuzzy set "Extremely Small Body-Mass-Index", $BMI_{SiSm}$ denotes the fuzzy set "Significantly Small Body-Mass-Index," and so forth. Objects $o_i$ are particular patients and numbers in the table are their membership degrees in particular fuzzy set.

| | $BMI_{ExSm}$ | $BMI_{SiSm}$ | $BMI_{VeSm}$ | $BMI_{Sm}$ | ... | $BMI_{QRSm}$ | $BMI_{VRSm}$ |
|---|---|---|---|---|---|---|---|
| $o_1$ | 0.5 | 0.6 | 0.7 | 0.8 | ... | 1 | 1 |
| $o_2$ | 0.8 | 0.9 | 1 | 1 | ... | 1 | 1 |
| $o_3$ | 0 | 0 | 0 | 0.1 | ... | 0.3 | 0.4 |
| $o_4$ | 0 | 0 | 0 | 0 | ... | 0 | 0 |
| $o_5$ | 0.6 | 0.9 | 1 | 1 | ... | 1 | 1 |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $o_n$ | 0 | 0.5 | 0.8 | 0.9 | ... | 1 | 1 |

add information that was lost by the transition from the quantitative variable BMI into two boolean variables $BMI_{\leq 25}$ and $BMI_{>25}$. More importantly, we also capture gradual transitions between different groups of people. An object $o_i$ (in our example a patient) might have membership degree to the fuzzy set $BMI_{Sm}$ equal to $0.4 (BMI_{Sm}(o_i) = 0.4)$ and simultaneously $o_i$ belongs to the fuzzy set $BMI_{Me}$ with the degree $0.3 (BMI_{Me}(o_i) = 0.3)$. This way we capture the information about the patient that is on the transition from being underweight to having ideal BMI index. This kind of information cannot be captured by crisp intervals.

In this way we treat every quantitative variable so that the final fuzzy GUHA table will look similarly to Table 6.

The fourfold table analogous to Table 3 is constructed also for the fuzzy variant of the method. The difference is that the numbers $a$, $b$, $c$, and $d$ are not summations of 1 s and 0 s but summations of membership degrees of data into fuzzy sets representing the antecedent $C$ and consequent $D$ or their complements, respectively. Naturally, the fact that antecedent $C$ as well as consequent $D$ holds simultaneously leads to the natural use of a t-norm [23]. In our case, we use the Gödel t-norm that is the minimum operation. For example, if an object $o_i$ belongs to a given antecedent in a degree 0.7 and to a given consequent in a degree 0.6, the value that enters the summation equals to $\min\{0.7, 0.6\} = 0.6$. Summation of such values over all the objects equals the value $a$ in Table 3; the other values from the table are determined analogously. The rest of the ideas of the method remain the same.

By using fuzzy sets, we generally get more precise results, and, more importantly, we avoid undesirable threshold effects

[24]. The further advantage is that the method searches for implicative associations that may be directly interpreted as fuzzy rules for the PbLD inference system.

*Example 2.* A confirmed association as

$$C\left(BMI_{ExBi}, Chol_{VeBi}\right) \sqsubset_r^\gamma D\left(BP_{MLBi}\right) \qquad (26)$$

may be directly interpreted as the following fuzzy rule:

"*IF Body-Mass-Index is Extremely Big AND Cholesterol level is Very Big THEN Blood Pressure is More or Less Big.*"

This approach has been found very efficient and reasonable, for example, for the identification of the so-called Fuzzy Rule Base Ensemble [25] which is a special ensemble technique for time series forecasting [26] that uses fuzzy rules to determine weights of individual forecasting methods. Naturally, the overlapping of extensions of linguistic expressions causes a massive generation of redundant associations. However, there exist efficient methods that detect and remove these redundancies automatically; see [6, 7].

In Section 4.3, we apply this method to artificial variables computed from the measures of water flow in order to obtain interesting descriptions of water flow rate peak time shift.

## 4. Data Analysis

*4.1. Data Description.* As mentioned in the introduction, we are provided only with the data from the measuring stations and from the Math-1D model implemented in the FLOREON system. Unlike the Math-1D model, we are neither provided

TABLE 6: Example of fuzzy GUHA table (compared with Table 4).

| | $BMI_{ExSm}$ | ... | $Chol_{ExSm}$ | ... | $Chol_{ExBi}$ | $BP_{ExSm}$ | ... | $BP_{ExBi}$ |
|---|---|---|---|---|---|---|---|---|
| $o_1$ | 0.5 | ... | 0.7 | ... | 0 | 1 | ... | 0 |
| $o_2$ | 0.8 | ... | 1 | ... | 0 | 0.4 | ... | 0 |
| $o_3$ | 0 | ... | 0 | ... | 0.1 | 0 | ... | 0.4 |
| $o_4$ | 0 | ... | 0 | ... | 0.4 | 0 | ... | 0.3 |
| $o_5$ | 0.6 | ... | 1 | ... | 0 | 1 | ... | 0 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| $o_n$ | 0 | ... | 0.8 | ... | 0 | 0.5 | ... | 0 |

with the measured precipitations nor with their forecasts nor with other physical attributes or their estimations. The reason is that this is the domain for the physical model Math-1D and our task is not to build another competitive physical model but to concentrate on the analysis of the existing one. However, in order to deal with the (fuzzy) GUHA method, we need to generate several features (artificial variables) and investigate the question, which of those variables have some influence on the performance of the model.

For the purpose of this investigation, we were provided with the data set collected from different *events* (floods) on the measuring station *Svinov* placed on the Odra River (Svinov is a part of the Ostrava city through which the Odra River flows. Naturally, the measuring station carries the same name). The whole data set is divided into 57 *simulations*. Each simulation captures a state of the system (provided real values and model values) at some time point $t$ that is for each simulation denoted by zero ($t = 0$). Each simulation can be further divided into past and future data measured or simulated on the hourly basis.

So, we can introduce the following two sets:

$$\text{Past} = \{t \mid -119, -118, \ldots, 0\},$$
$$\text{Future} = \{t \mid 1, 2, \ldots, 48\}, \tag{27}$$

and the two time dependent variables, namely, the real water flow rate at time $t$ and the originally modelled flow rate at time $t$, denoted by $r_t$ and $m_t$, respectively. Thus, we can also introduce the following sets:

$$r_{\text{Past}} = \{r_t \mid t \in \text{Past}\}, \qquad r_{\text{Future}} = \{r_t \mid t \in \text{Future}\}, \tag{28}$$

and analogously

$$m_{\text{Past}} = \{m_t \mid t \in \text{Past}\}, \qquad m_{\text{Future}} = \{m_t \mid t \in \text{Future}\}. \tag{29}$$

Indeed, the values $r_{\text{Future}}$ had been unknown at the time point $t = 0$ and they were added to the data later on only for the comparison and efficiency evaluation purposes. The values $m_{\text{Future}}$ are forecasts made by the original Math-1D model that were at disposal at the time point $t = 0$.

The aim is to analyze associations between input variables that were at disposal at the time $t = 0$ ($r_{\text{Past}}$, $m_{\text{Past}}$, and $m_{\text{Future}}$) and the dependent variable which was (for this stage of investigation) chosen to be the *peak-time* $r_{\text{Peak}}$, that is, the time of maximum water flow rate:

$$r_{\text{Peak}} = \arg \max_{t \in \text{Future}} r_t. \tag{30}$$

For the sake of result quality evaluation, the data was split into a training set and a testing set in the ratio of 2 : 1, that is, 38 simulations for the training and 19 simulations for the testing.

*4.2. Features Generation and Reduction.* For each simulation $s$, a set of features was extracted by applying several statistical characteristics on different vectors of data that were derived from $r_{\text{Past}}$ and $m_{\text{Past}}$. Namely, the following statistics were utilized: mean $\bar{u}$, standard deviation $\sigma_u$, median $\tilde{u}$, minimum $\min(u)$, maximum $\max(u)$, range $(\max(u) - \min(u))$, interquartile range $(q_{0.75}(u) - q_{0.25}(u))$, difference of the last value and the mean $(u_0 - \bar{u})$, coefficient of variation $CV_u$, difference of the mean and the median $(\bar{u} - \tilde{u})$, absolute difference of the mean and the median $|\bar{u} - \tilde{u}|$, skewness $\text{Skew}(u)$ and its absolute value $|\text{Skew}(u)|$, kurtosis $\text{Kurt}(u)$ and its absolute value $|\text{Kurt}(u)|$, slope $\beta_u$ computed from linear regression of $u_t = \iota_u + \beta_u t + \varepsilon_u$ (where $\iota_u$ is the intercept and $\varepsilon_u$ is the residual error), and trend strength $\text{Trend}(u)$ computed as a $P$ value of the hypothesis $\beta_u = 0$.

All the statistics listed above were computed for each of the following data $u \in \{r_{\text{Past}}, m_{\text{Past}}\}$. Additionally, the same statistics were determined for the following further newly created data vectors:

$$\log(u) = \{\log |u_t| \mid t \in \{-119, -118, \ldots, 0\}\},$$
$$\text{diff} = \{u_{t+1} - u_t \mid t \in \{-119, -118, \ldots, -1\}\},$$
$$\log(\text{diff}) = \{\log |u_{t+1} - u_t| \mid t \in \{-119, -118, \ldots, -1\}\}, \tag{31}$$

where again $u \in \{r_{\text{Past}}, m_{\text{Past}}\}$.

Analogously, the same statistics have been utilized also for $u = m_{\text{Future}}$ with the only difference stemming from the different time values; that is, they were applied to

$$\log(m) = \{\log |m_t| \mid t \in \{1, 2, \ldots, 48\}\},$$
$$\text{diff}_m = \{m_{t+1} - m_t \mid t \in \{2, 3, \ldots, 48\}\}, \tag{32}$$
$$\log(\text{diff}_m) = \{\log |m_{t+1} - m_t| \mid t \in \{2, 3, \ldots, 48\}\}.$$

Finally, the time point of the forecasted peak,

$$m_{\text{Peak}} = \arg \max_{t \in \text{Future}} m_t, \tag{33}$$

was also added as an additional feature. It means that a total amount of 205 new features were generated.

TABLE 7: Example of a part of the fuzzy GUHA table for peak shift of PS of the peak forecasted by the Math-1D model. Objects $s_i$ are particular simulations.

| | $\sigma_{\log(r_{Past})\,ExSm}$ | ... | $m_{Peak\,ExBi}$ | $PS_{ExSm}$ | ... | $PS_{ExBi}$ |
|---|---|---|---|---|---|---|
| $s_1$ | 0.97 | ... | 0.62 | 0.45 | ... | 0 |
| $s_2$ | 0 | ... | 0.2 | 0 | ... | 0.58 |
| $s_3$ | 0.75 | ... | 0.97 | 0.66 | ... | 0 |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | |
| $s_{38}$ | 0.66 | ... | 0.74 | 0.69 | ... | 0 |

TABLE 8: Examples of fuzzy rules found by the fuzzy GUHA method.

| Rule | IF part | | | THEN part |
|---|---|---|---|---|
| | $\sigma_{\log(r_{Past})}$ | $CV_{\log(r_{Past})}$ | $m_{Peak}$ | PS |
| $\mathcal{R}_1$ | Sm | VR Sm | Ve Bi | QR Sm |
| $\mathcal{R}_2$ | VR Sm | Ve Sm | Ve Bi | Ro Sm |
| $\mathcal{R}_3$ | Ve Sm | — | Ex Bi | ML Sm |
| ... | ... | ... | ... | ... |
| $\mathcal{R}_{69}$ | VR Me | QR Sm | VR Bi | Ro Me |

From the pool of features, a regression method [27] was utilized to select those, which had the highest significance for a regression model. Particularly, the dependent variable

$$PS = \left( r_{Peak} - m_{Peak} \right), \qquad (34)$$

denoting the peak shift, was modelled with the linear regression of all the generated features. After that, statistical significance of all the regression coefficients was tested and only features with $P$-value below 0.05 were selected.

In this way, we ended the feature selection with the following three features: $\sigma_{\log(r_{Past})}$: standard deviation of $\log(r_{Past})$; $CV_{\log(r_{Past})}$: coefficient of variation of $\log(r_{Past})$; and finally, $m_{Peak}$: time point of the forecasted peak given by (33).

*4.3. Fuzzy GUHA Application.* All computed features, which were found statistically significant, as described in the previous subsection, are viewed as quantitative variables. In order to use them in mining linguistic associations, we had to convert them into fuzzy attributes. More specifically, we generated all the possible linguistic expressions (see Section 2) and determined appropriate contexts for each of the variables, and finally, for each simulation, we determined the degrees of membership of the given simulation into the extensions of the linguistic expressions for each variable. Such process turned the three antecedent variables into 63 fuzzy attributes—each related certain evaluative linguistic expression (21 expressions for each variable; see Section 2.1).

The above introduced variable $PS = (r_{Peak} - m_{Peak})$ is the dependent variable whose dependence on the generated attributes appearing in antecedents is being "explained" with the help of the fuzzy GUHA method and the generated linguistic associations; see Section 3.

Part of the resulting fuzzy GUHA table that contained 84 columns, 63 for antecedent attributes, and 21 for consequent attributes is shown in Table 7.

Upon the choice of the multitudinal implicational quantifier and the degree of confidence $\gamma = 0.7$ and the degree of support $r = 0.1$, the fuzzy GUHA generated considerably many linguistic associations. After the application of the redundancy detection and removal algorithm [7], we have obtained 69 fuzzy rules Table 8 that have a twofold importance:

(i) they describe the situations, under which the disaster management may expect some time shift of the water flow rate peak, which is essential for precise warning and evacuation of people or other preparations works that may save material costs of the approaching disaster;

(ii) connected to the PbLD inference mechanism, they may be directly used to forecast the time shift of the peak originally forecasted by the Math-1D model and, thus, to directly correct and precisiate the forecast by the physical model.

## 5. Prediction, Results, and Evaluation

*5.1. Results and Evaluation.* The prediction model was evaluated on a testing dataset, that is, on data previously hidden during the whole data-mining procedure. The testing dataset consists of 19 simulations, each simulation containing hourly flow rates for five days in the past and two days of predictions for the future.

On the testing simulations, the prediction accuracy of the time of culminating-peak was compared between the original Math-1D model and the Math-1D model newly adjusted with GUHA association rules.

For each testing simulation $s$ and model—either original (by Math-1D) or adjusted (with the help of fuzzy rules)—a prediction error $e(s)$ was evaluated as follows:

$$e(s) = m_{Peak}(s) - r_{Peak}(s), \qquad (35)$$

TABLE 9: A comparison of peak forecast errors $e_{m,s}$ among the original and the adjusted model. The table presents basic statistics computed from $e_{m,s}$ evaluated on testing data.

| Model | Min | 1st quart. | Mean | Stdev | Median | 3rd quart. | Max |
|---|---|---|---|---|---|---|---|
| Original | −0.33 | 0.31 | 0.603 | 0.52 | 0.46 | 0.88 | 1.92 |
| Adjusted | −1.33 | −0.69 | −0.205 | 0.65 | −0.12 | 0.07 | 1.02 |

TABLE 10: Statistical tests of hypotheses of $\bar{e}_m = 0$: Both Wilcoxon signed rank sum test and one sample $t$-test indicate the error $e_{\text{original}}$ to be very likely nonzero. For adjusted model, the variability of $e_{\text{adjusted}}$ can be explained by randomness.

| Model | Wilcoxon test | | One sample $t$-test | | |
|---|---|---|---|---|---|
| | $V$ | $P$ value | $t$ | df | $P$ value |
| Original | 166 | 0.0004871 | 5.042 | 18 | 0.00008486 |
| Adjusted | 61 | 0.1776 | −1.373 | 18 | 0.1866 |

where $m_{\text{Peak}}(s)$ is the peak time forecasted for simulation $s$ by a given model and $r_{\text{Peak}}(s)$ is the time of real occurrence of the peak in simulation $s$; see also formulas (30)–(33). Summary of the comparison can be found in Table 9.

Briefly, it can be stated that the original model expects the flood peaks approximately a half an hour later than in reality, on the testing dataset. After adjustments made by our GUHA model, the estimates become more accurate. More precisely, the original (Math-1D) model error is on average 0.603 days (with standard deviation 0.521). The error of the adjusted model is −0.205 days (with standard deviation 0.65).

A bias towards positive values of the original model was also justified by the one sample Wilcoxon rank sum test [28]: a null hypothesis of zero shift was rejected with $P$ value = 0.000487, on the original model. On the other hand, the same hypothesis cannot be rejected for the adjusted model (with $P$ value = 0.1776). Similar results were also obtained with the one sample $t$-test (see Table 10).

## 6. Conclusion

In this paper, we attempted to deal with an adjustment of a physical model of water flow rate during floods with the help of linguistic associations mining. As any physical model based on differential equations (the Math-1D model, in our case) is highly dependent on many unreliable parameters, it seems reasonable to perform some real data analysis that would inform us, when and under which conditions the model is (in terms of the culminating water flow rate peak) time lagged or vice-versa too much ahead.

We approached the task with the help of the fuzzy GUHA method that automatically generates linguistic associations. The provided data was firstly extended by a creation of artificial variables describing various features of the data. The resulting variables were later on translated into fuzzy GUHA table using the so-called Evaluative Linguistic Expressions. This table was used to mine the associations that may be directly interpreted as fuzzy IF-THEN rules. Such interpretation is beneficial not only because of its interpretability but it can also be used jointly with the Perception-based Logical Deduction inference method in order to predict expected time shift of the flood peaks originally forecasted by the physical model. Results obtained from this adjusted

model were statistically evaluated in order to confirm the improvement in the forecasting accuracy.

Let us note that the data-mining analysis as well as experimental evaluation was performed only on a single measuring station Svinov placed on Odra River. Indeed, as the physical model depends on many imprecise and estimated parameters that may differ over the river flow, each station would require its own analysis. However, as the number of stations in the whole region is rather low (9 stations placed on four main rivers), such approach is obviously feasible. Thus the promising results give chance for further and deeper analysis that could enhance the disaster management by more accurate physical models with forecasts adjusted by fuzzy IF-THEN rules. On the other hand, there is a serious complication in the lack of the past data that could be analyzed. The high number of previous floods is unfortunately not accompanied by a sufficiently high number of precise data. As we have mentioned, there was, for example, a problem of measured zero water flow rates even during massive floods due to uncalibrated measuring stations or due to other unspecified reasons. This lack of reliable data may significantly complicate the situation.

As the first step for future research, we plan to extend our investigation by using measured past precipitations and possibly also the forecasted future precipitations that are already at disposal to the Math-1D model but that were not at disposal to our data analysis presented in this paper.

## Acknowledgment

## References

[1] J. Martinovič, S. Štolfa, J. Kožusznik, J. Unucka, and I. Vondrák, "FLOREON—the system for an emergent flood prediction," in *ECECFUBUTEC- EUROMEDIA*, Porto, Portugal, 2008.

[2] V. Novák, I. Perfilieva, and J. Močkoř, *Mathematical Principles of Fuzzy Logic*, Kluwer Academic Publishers, Boston, Mass, USA, 1999.

[3] P. Kubíček and T. Kozubek, "Mathematical-analytical solutions of the flood wave and its use in practice," (in czech). VŠB-TU Ostrava, 150.

[4] V. Novák, "A comprehensive theory of trichotomous evaluative linguistic expressions," *Fuzzy Sets and Systems*, vol. 159, no. 22, pp. 2939–2969, 2008.

[5] V. Novák and A. Dvořák, "Formalization of commonsense reasoning in fuzzy logic in broader sense," *Applied and Computational Mathematics*, vol. 10, no. 1, pp. 106–121, 2011.

[6] A. Dvořák, M. Štěpnička, and L. Vavříčková, "Redundancies in systems of fuzzy/linguistic if-then rules, pages," in *Proceedings of the 7th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT '11)*, Advances in Intelligent Systems Research, pp. 1022–1029, Atlantic Press, Paris, France, 2011.

[7] L. Štěpničková, M. Štěpnička, and A. Dvořák, "New results on redundancies of fuzzy/linguistic if-then rules," in *Proceedings of the 8th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT '13)*, pp. 400–407, Atlantic Press, Milano, Italy.

[8] L. A. Zadeh, "Precisiated natural language (PNL)," *AI Magazine*, vol. 25, no. 3, pp. 74–91, 2004.

[9] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning—I," *Information Sciences*, vol. 8, no. 3, pp. 199–249, 1975.

[10] M. Štěpnička, A. Dvořák, V. Pavliska, and L. Vavříčková, "A linguistic approach to time series modeling with the help of F-transform," *Fuzzy Sets and Systems*, vol. 180, no. 1, pp. 164–184, 2011.

[11] A. Dvořák, H. Habiballa, V. Novák, and V. Pavliska, "The concept of LFLC 2000—its specificity, realization and power of applications," *Computers in Industry*, vol. 51, no. 3, pp. 269–280, 2003.

[12] M. De Cock and E. E. Kerre, "Fuzzy modifiers based on fuzzy relations," *Information Sciences*, vol. 160, no. 1–4, pp. 173–199, 2004.

[13] E. Hajičová, B. Partee, and P. Sgall, *Topic-Focus Articulation, Tripartite Structures, and Semantic Content*, Kluwer, Dordrecht, The Netherlands, 1998.

[14] V. Novák, "Perception-based logical deduction," in *Computational Intelligence, Theory and Applications, Advances in Soft Computing*, B. Reusch, Ed., pp. 237–250, Springer, Berlin, Germany, 2005.

[15] J. Kupka and I. Tomanová, "Some extensions of mining of linguistic associations," *Neural Network World*, vol. 20, no. 1, pp. 27–44, 2010.

[16] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in *Proceedings of the 20th International Conference on Very Large Databases*, pp. 487–499, AAAI Press, Santiago de Chile, Chile, 1994.

[17] P. Hájek, "The question of a general concept of the GUHA method," *Kybernetika*, vol. 4, pp. 505–515, 1968.

[18] P. Hájek and T. Havránek, *Mechanizing Hypothesis Formation: Mathematical Foundations for a General Theory*, Springer, Berlin, Germany, 1978.

[19] P. Hájek, M. Holeňa, and J. Rauch, "The GUHA method and its meaning for data mining," *Journal of Computer and System Sciences*, vol. 76, no. 1, pp. 34–48, 2010.

[20] V. Novák, "A comprehensive theory of trichotomous evaluative linguistic expressions," *Fuzzy Sets and Systems*, vol. 159, no. 22, pp. 2939–2969, 2008.

[21] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, Morgan Kaufmann, 2006.

[22] V. Novák, I. Perfilieva, A. Dvořák, G. Chen, Q. Wei, and P. Yan, "Mining pure linguistic associations from numerical data," *International Journal of Approximate Reasoning*, vol. 48, no. 1, pp. 4–22, 2008.

[23] E. P. Klement, R. Mesiar, and E. Pap, *Triangular Norms*, vol. 8 of *Trends in Logic*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2000.

[24] T. Sudkamp, "Examples, counterexamples, and measuring fuzzy associations," *Fuzzy Sets and Systems*, vol. 149, no. 1, pp. 57–71, 2005.

[25] L. Štěpničková, M. Štěpnička, and D. Sikora, "Fuzzy rule-based ensemble with use linguistic associations mining for time series prediction," in *Proceedings of the 8th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT '13)*, pp. 408–415, Atlantic Press, Milano, Italy, 2013.

[26] D. Sikora, M. Štěpnička, and L. Vavříčková, "Fuzzy rule-based ensemble forecasting: Introductory study," in *Synergies of Soft Computing and Statistics for Intelligent Data Analysis*, vol. 190 of *Advances in Intelligent Systems and Computing*, pp. 379–387, Springer, 2013.

[27] J. M. Chambers, A. Freeny, and R. M. Heiberger, "Statistical models in S," in *Linear Models*, Wadsworth & Brooks/Cole, 1992.

[28] M. Hollander and D. A. Wolfe, *Nonparametric Statistical Methods*, A Wiley Publication in Applied Statistics, Wiley, New York, NY, USA, 1973.

Advances in
**Multimedia**

The Scientific
**World Journal**

International Journal of
**Distributed
Sensor Networks**

Journal of
Industrial Engineering

**Applied
Computational
Intelligence and Soft
Computing**

Advances in
**Fuzzy
Systems**

**Modelling &
Simulation
in Engineering**

Journal of
**Computer Networks
and Communications**

Advances in
**Artificial
Intelligence**

Submit your manuscripts at
http://www.hindawi.com

Advances in
**Computer Engineering**

International Journal of
**Computer Games
Technology**

International Journal of
Biomedical Imaging

Advances in
**Artificial
Neural Systems**

Advances in
**Software Engineering**

Journal of
**Robotics**

Advances in
**Human-Computer
Interaction**

**Computational
Intelligence and
Neuroscience**

International Journal of
**Reconfigurable
Computing**

Journal of
**Electrical and Computer
Engineering**