

# New results on redundancies of fuzzy/linguistic IF-THEN rules

Lenka Štěpničková<sup>1</sup> Martin Štěpnička<sup>1</sup> Antonín Dvořák<sup>1</sup>

<sup>1</sup>CE IT4Innovations - Division of the University of Ostrava - IRAFM, 30. dubna 22, 701 03 Ostrava, CZ

## Abstract

In this paper we present new results on detection and removal of redundancies of IF-THEN rules in so-called linguistic descriptions (systems of such rules). We introduce an algorithm for removal of redundancies and describe a practical application.

**Keywords:** Redundant rule, IF-THEN rule, fuzzy inference

## 1. Introduction and motivation

In this contribution, we outline our approach to detection of so-called redundancies in systems of fuzzy/linguistic IF-THEN rules (we call such systems *linguistic descriptions*). Preliminary results can be found in [1].

Notions of redundancy and inconsistency of fuzzy IF-THEN rules were studied by several authors [2, 3, 4, 5]. But, for our so-called linguistic approach originated by V. Novák [6] these questions were not adequately investigated. We were able to successfully apply our linguistic approach in analysis and forecasting of time series [7, 8], decision making [9] or data mining [10]. We found out that this approach has distinct advantages in good interpretability, great robustness, foundations in a strong formal logical system etc. However, if we learn our rules from real-world data, as is the case of time series analysis or data mining, then interpretability can be seriously harmed by the presence of redundant rules. Users of these rules can be then buried under a high number of seemingly very similar rules and consequently would not be able to comprehend and modify them. Therefore, theoretically well-founded and efficient algorithm for detection of redundant rules is necessary.<sup>1</sup>

The inference mechanism tailored for our fuzzy/linguistic IF-THEN rules is called *perception-based logical deduction* (PbLD). Quite informally, it, for a given input  $u_0$ , picks and fires that rule from the linguistic description, whose antecedent (IF part) has a maximal membership degree at  $u_0$ .<sup>2</sup>

<sup>1</sup>Inconsistencies in systems of IF-THEN rules are also unpleasant for the interpretability, but they can harm the performance of an inference mechanism, too. Their detection and elimination will be a topic of our further research.

<sup>2</sup>More precisely, the antecedent is a linguistic expression, e.g. *small*. This expression is, after some formal steps, interpreted by some fuzzy set, and the membership degree in this fuzzy set is referred to here.

If there are several such rules, then it picks that whose antecedent is most specific (for example, *very small* is more specific than *small*, etc.).

A fuzzy IF-THEN rule  $\mathcal{R}_1$  is usually thought to be redundant with respect to rule  $\mathcal{R}_2$  if their consequents are identical and their antecedents are different but not contradictory. For example, let  $\mathcal{R}_1$  and  $\mathcal{R}_2$  be

$\mathcal{R}_1 := \text{IF } X \text{ is } \textit{very small} \text{ THEN } Y \text{ is } \textit{big},$

$\mathcal{R}_2 := \text{IF } X \text{ is } \textit{roughly small} \text{ THEN } Y \text{ is } \textit{big}.$

If the interpretation of linguistic expressions *roughly small* and *very small* is such that every element which is very small is also roughly small in at least the same degree, then rule  $\mathcal{R}_1$  is redundant — nothing changes if it is left out. Initially it seemed to us that elimination of redundancies would consist simply in detection of pairs of rules such as  $\mathcal{R}_1$  and  $\mathcal{R}_2$  and deletion of  $\mathcal{R}_1$ . However, it turned out that it is not that simple. We accepted a natural definition that, informally speaking, a rule is redundant if results of inference mechanism are exactly the same if we leave this rule out. Then, the presence of other rules (with different consequents) can cause that a rule such as  $\mathcal{R}_1$  is all of a sudden not redundant at all. Imagine that third rule is present, namely

$\mathcal{R}_3 := \text{IF } X \text{ is } \textit{small} \text{ THEN } Y \text{ is } \textit{small}$

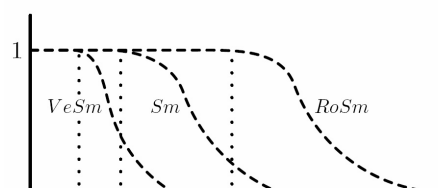


Figure 1: Graphical presentation of extensions (fuzzy sets) that interpret linguistic expressions *very small*, *small* and *roughly small*.

If the interpretation of *small* is such that it lies “in-between” *roughly small* and *very small* (see Figure 1), then we cannot leave out rule  $\mathcal{R}_1$ , because for values of  $X$  for which it is fired (result of inference should be “big”) suddenly, after  $\mathcal{R}_1$  is left out,  $\mathcal{R}_3$  would be fired (and result of inference would be “small”). We call  $\mathcal{R}_1$  to be suspicious to be redundant with respect to  $\mathcal{R}_2$  and search mathematical

results stating conditions under which it is indeed redundant and under which it is not.

Let us remark that our results apply also to systems of IF-THEN rules and inference mechanisms which use compatible design choices (overlapping interpretations of linguistic expressions like *very small* and *small*, inference mechanisms firing rules based on the best fit of inputs and highest specificity, etc.) Particular shapes of fuzzy sets interpreting evaluative expressions and details of inference mechanisms are not crucial from this point of view.

## 2. Theoretical background

Because of space limitations, we introduce basic notions here only. For details and discussions, see [11, 12, 1].

### 2.1. Evaluative linguistic expressions

One of main constituents of systems of fuzzy/linguistic IF-THEN rules are *evaluative linguistic expressions* [11], in short *evaluative expressions*, e.g. *very large*, *more or less hot*, etc. 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 L. A. Zadeh (e.g., in [13, 14]).

A simple form of evaluative expressions keeps the following structure:

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

*Atomic evaluative expressions* comprise any of the *canonical* adjectives *small*, *medium*, *big*, abbreviated in the following as Sm, Me, 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*, e.g. *very*, *extremely*, etc. and with *widening effect*, e.g. *roughly*, *more or less*, etc. In the following text, we, without loss of generality, use the hedges introduced in Table 1 that were successfully used in real applications [8] and that are implemented in the LFLC software package [15]. As a special case, the  $\langle \text{linguistic hedge} \rangle$  can be empty. Note that our hedges are of so-called inclusive type [16], which means that extensions of more specific evaluative expressions are included in less specific ones, see Figure 1.

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

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

Narrowing effect	Widening effect
very (Ve)	more or less (ML)
significantly (Si)	roughly (Ro)
extremely (Ex)	quite roughly (QR)
–	very roughly (VR)

Table 1: Linguistic hedges and their abbreviations.

Examples of evaluative predications are “temperature is very high”, “price is low”, etc.

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  $\langle v_L, v_M, v_R \rangle$ . By  $u \in w$  we mean  $u \in [v_L, v_R]$ . 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  $\mathcal{A}$ ” is a certain formula whose interpretation is a function

$$\text{Int}(X \text{ is } \mathcal{A}) : W \longrightarrow \mathcal{F}(\mathbb{R}), \quad (2)$$

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

Given an intension (2) and a context  $w \in W$ , we can define the *extension* of “ $X$  is  $\mathcal{A}$ ” in the context  $w$  as a fuzzy set

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

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

We extend the theory of evaluative linguistic expressions by the following *partition axiom*: There does not exist any context  $w \in W$  in which there would exist some  $u_0 \in w$  such that

$$(\text{Int}(X \text{ is } \mathcal{A})(w))(u_0) = (\text{Int}(X \text{ is } \mathcal{B})(w))(u_0) = 1 \quad (3)$$

for  $\mathcal{A}, \mathcal{B}$  with different atomic evaluative expressions. Indeed, no element  $u_0$  in any world is naturally assumed to belong in the degree one to a fuzzy set of small objects as well as of medium or big objects - no matter the influence of linguistic hedges.

### 2.2. Fuzzy IF-THEN rules, linguistic description

Evaluative predications occur in conditional clauses of natural language of the form

$$\mathcal{R} := \text{IF } X \text{ is } \mathcal{A} \text{ THEN } Y \text{ is } \mathcal{B} \quad (4)$$

where  $\mathcal{A}, \mathcal{B}$  are evaluative expressions. The linguistic predication “ $X$  is  $\mathcal{A}$ ” is called the *antecedent* and

“Y is  $\mathcal{B}$ ” is called the *consequent* of the rule (4). Of course, the antecedent may consist of more evaluative predications, joined by the connective “AND”. The clauses (4) 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  $LD = \{\mathcal{R}_1, \dots, \mathcal{R}_m\}$  where

$$\begin{aligned} \mathcal{R}_1 &:= \text{IF } X \text{ is } \mathcal{A}_1 \text{ THEN } Y \text{ is } \mathcal{B}_1, \\ &\dots\dots\dots (5) \\ \mathcal{R}_m &:= \text{IF } X \text{ is } \mathcal{A}_m \text{ THEN } Y \text{ is } \mathcal{B}_m. \end{aligned}$$

Because each rule in (5) 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  $\mathcal{R}$  in (4) is a function

$$\text{Int}(\mathcal{R}) : W \times W \longrightarrow \mathcal{F}(\mathbb{R} \times \mathbb{R}). \quad (6)$$

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 (6).

We also need to consider a linguistic phenomenon of topic-focus articulation (cf. [17]), which in the case of linguistic descriptions requires us to distinguish the following two sets:

$$\begin{aligned} \text{Topic}_{LD} &= \{\text{Int}(X \text{ is } \mathcal{A}_j) \mid j = 1, \dots, m\}, \\ \text{Focus}_{LD} &= \{\text{Int}(Y \text{ is } \mathcal{B}_j) \mid j = 1, \dots, m\}. \end{aligned}$$

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

### 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 mentioned in Section 2.1 as follows:

$$\begin{aligned} \text{Ex} \leq_H \text{Si} \leq_H \text{Ve} \leq_H \langle \text{empty} \rangle \leq_H \\ \leq_H \text{ML} \leq_H \text{Ro} \leq_H \text{QR} \leq_H \text{VR}. \end{aligned}$$

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

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

and

$$\begin{aligned} \text{Ker}(\text{Int}(X \text{ is } \langle \text{hedge} \rangle_i \mathcal{A})(w)) \\ \subset \text{Ker}(\text{Int}(X \text{ is } \langle \text{hedge} \rangle_j \mathcal{A})(w)) \end{aligned}$$

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

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

$$\mathcal{A}_i \leq_{LE} \mathcal{A}_j$$

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

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 of evaluative predications wrt. a given observation. Let us be given a context  $w \in W$ , an observation  $u_0 \in w$  and two evaluative predications  $(X \text{ is } \mathcal{A}_i)$  and  $(X \text{ is } \mathcal{A}_j)$  from the  $\text{Topic}_{LD}$ . We write  $(X \text{ is } \mathcal{A}_i) \leq_{(u_0, w)} (X \text{ is } \mathcal{A}_j)$  either if  $\text{Int}(X \text{ is } \mathcal{A}_i(w))(u_0) > \text{Int}(X \text{ is } \mathcal{A}_j(w))(u_0)$  or if  $\text{Int}(X \text{ is } \mathcal{A}_i(w))(u_0) = \text{Int}(X \text{ is } \mathcal{A}_j(w))(u_0)$  and  $\mathcal{A}_i \leq_{LE} \mathcal{A}_j$ .

It should be noted that usually the  $\text{Topic}_{LD}$  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

$$\begin{aligned} (X \text{ is } \mathcal{A}_i) &:= (X_1 \text{ is } \mathcal{A}_{i_1}) \text{ AND } \dots \text{ AND } (X_K \text{ is } \mathcal{A}_{i_K}), \\ (X \text{ is } \mathcal{A}_j) &:= (X_1 \text{ is } \mathcal{A}_{j_1}) \text{ AND } \dots \text{ AND } (X_K \text{ is } \mathcal{A}_{j_K}). \end{aligned}$$

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

$$\mathcal{A}_i \leq_{LE} \mathcal{A}_j \quad \text{if} \quad \mathcal{A}_{i_k} \leq_{LE} \mathcal{A}_{j_k} \quad \text{for all } k = 1, \dots, K$$

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

$$\begin{aligned} &(\text{Int}(X \text{ is } \mathcal{A}_i)(w_1, \dots, w_K))(u_1, \dots, u_K) \\ &= \bigwedge_{k=1}^K (\text{Int}(X_k \text{ is } \mathcal{A}_{i_k})(w_k))(u_k). \end{aligned}$$

Then, the final ordering  $\leq_{(u_0, w)}$  is analogous to the one-dimensional one.

On Figure 2, we provide readers with a visualization of two fuzzy rules with two input variables. Note that the rectangles denote areas where the antecedent of the given rule is minimal wrt.  $\leq_{LE}$ . Each rectangle is also denoted by a respective consequent  $B_i$ . Thus, for the sake of brevity, we will use only the rectangles to display the areas covered by antecedents jointly with the labels denoting the respective consequents, as on Figure 3.

### 2.4. Perception-based logical deduction

This is a special inference method aimed at the derivation of results based on fuzzy/linguistic IF-THEN rules. A perception is understood as an

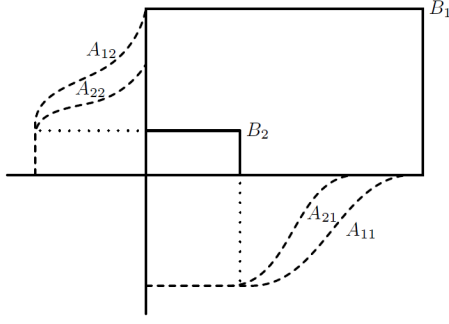


Figure 2: Two fuzzy rules visualization: rectangles denote areas covered by antecedents of given rules.

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 [6, 8, 12].

Based on the ordering  $\leq_{(u_0, w)}$  of linguistic predications we define a special function of *local perception*

$$LPerc^{LD} : w \times W^K \longrightarrow \mathcal{P}(\text{Topic}_{LD})$$

assigning to each value  $u_0 = [u_1, \dots, u_K] \in w$  for  $w = [w_1, \dots, w_K] \in W^K$  a subset of intensions minimal wrt. the ordering  $\leq_{(u_0, w)}$

$$LPerc^{LD}(u_0, w) = \{ \text{Int}(X \text{ is } \mathcal{A}_i) \mid \text{Int}(X \text{ is } \mathcal{A}_i)(w)(u_0) > 0 \ \& \ \forall \text{Int}(X \text{ is } \mathcal{A}_j) \in \text{Topic}_{LD} : ((X \text{ is } \mathcal{A}_j) \leq_{(u_0, w)} (X \text{ is } \mathcal{A}_i)) \Rightarrow ((X \text{ is } \mathcal{A}_j) = (X \text{ is } \mathcal{A}_i)) \} \quad (7)$$

Let  $LD$  be a linguistic description (5). Let us consider a context  $w \in W$  for the variable  $X$  and a context  $w' \in W$  for  $Y$ . Let an observation  $X = u_0$  in the context  $w$  be given, where  $u_0 \in w$ . Then, the following *rule of perception-based logical deduction* ( $r_{PbLD}$ ) can be introduced:

$$r_{PbLD} : \frac{LPerc^{LD}(u_0, w), LD}{C} \quad (8)$$

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  $LPerc^{LD}(u_0, w)$  from (7). 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 (7) chooses these antecedents which best fit the given numerical input  $u_0$ , in other words, they are most specific according to the ordering  $\leq_{(u_0, w)}$ .

Once one or more antecedents  $\text{Int}(X \text{ is } \mathcal{A}_{i_\ell}) \in \text{Topic}_{LD}$ ,  $i_\ell = 1, \dots, L$  are chosen according to (7), we compute for any of them conclusions  $C_{i_\ell}$ :

$$C_{i_\ell}(v) = (\text{Int}(X \text{ is } \mathcal{A}_{i_\ell})(w))(u_0) \rightarrow (\text{Int}(Y \text{ is } \mathcal{B}_{i_\ell})(w'))(v).$$

Suppose that  $LPerc^{LD}(u_0, w)$  is non-empty, i.e.,  $L > 0$ . Then the final conclusion  $C$  is given as a set of all  $L$  conclusions  $C_{i_\ell}$  that correspond to  $L$  members in  $LPerc^{LD}(u_0, w)$ , i.e.,

$$C = \{C_{i_\ell} \mid \ell = 1, \dots, L\}.$$

Usually,  $L = 1$ , i.e., there is one element in  $LPerc^{LD}(u_0, w)$ . In this case, the only element in  $C$  will be denoted by the same letter. If  $\text{Int}(X \text{ is } \mathcal{A}_i)(w)(u_0) = 0$  for all  $\text{Int}(X \text{ is } \mathcal{A}_i) \in \text{Topic}_{LD}$ , then, according to (7),  $C$  is the empty set.

**Remark 1** Let us note that usually the final inference output is aggregated using the intersection of all elements in  $C$ . Thus, it is easy to see that whenever an  $LD$  contains two rules:

$$\begin{aligned} \mathcal{R}_i &:= \text{IF } X \text{ is } \mathcal{A}_i \text{ THEN } Y \text{ is } \mathcal{B}_i, \\ \mathcal{R}_j &:= \text{IF } X \text{ is } \mathcal{A}_j \text{ THEN } Y \text{ is } \mathcal{B}_j, \end{aligned}$$

such that  $\mathcal{A}_i = \mathcal{A}_j$  and  $\mathcal{B}_i \leq_{LE} \mathcal{B}_j$ , the rule  $\mathcal{R}_j$  is trivially redundant.

This fact may be used in the preprocessing of linguistic descriptions in order to efficiently decrease the number of investigated rules. For the formal investigation of redundancy it suffices to deal with the non-aggregated  $C$ .

### 3. Redundancy

#### 3.1. Basic concepts

Let us fix the notation for the rest of the paper. Let us consider a linguistic description  $LD$  and let us be given an observation  $u_0$  in a given context  $w \in W$ . Let  $C$  be the conclusion derived from  $u_0$  based on  $LD$  using the rule of perception based logical deduction given by (8). Then this fact will be denoted by

$$r_{PbLD} \left( LPerc^{LD}(u_0, w) \right) : C. \quad (9)$$

Note that  $C$  is a set of fuzzy sets, in general. By writing, e.g.,  $C = D$  we are expressing the fact that sets  $C$  and  $D$  are equal, i.e., they have precisely the same elements (fuzzy sets).

**Definition 1** Let  $LD = \{\mathcal{R}_1, \dots, \mathcal{R}_m\}$  be a linguistic description (5). Rule  $\mathcal{R}_i$  is redundant in  $LD$  if  $D_1 = D_2$  for each value  $u_0 \in w$ ,  $w \in W$ , where

$$\begin{aligned} r_{PbLD} \left( LPerc^{LD}(u_0, w) \right) &: D_1, \\ r_{PbLD} \left( LPerc^{LD'}(u_0, w) \right) &: D_2 \end{aligned}$$

and  $LD' = LD \setminus \{\mathcal{R}_i\}$ .

As we have mentioned, redundancy is observed as an existence of fuzzy rules with distinct overlapping antecedents and identical consequents. But as

we will show below, sometimes such an intuitively redundant fuzzy rule does not have to be always redundant with respect to a formal definition of the redundancy. Therefore, such a rule will be called *suspicious of redundancy* and a further analysis of its potential redundancy turns out to be necessary.

**Definition 2** Let  $LD$  be a linguistic description (5), let  $\{\mathcal{R}_i, \mathcal{R}_j\} \subseteq LD$ . Rule  $\mathcal{R}_i$  is suspicious of redundancy with respect to  $\mathcal{R}_j$  (denoted by  $\mathcal{R}_i \hookrightarrow \mathcal{R}_j$ ) if  $C_1 = C_2$  for each value  $u_0 \in w$ ,  $w \in W$ , where

$$r_{PbLD} \left( LPerc^{\{\mathcal{R}_i, \mathcal{R}_j\}}(u_0, w) \right) : C_1$$

and

$$r_{PbLD} : \left( LPerc^{\{\mathcal{R}_j\}}(u_0, w) \right) : C_2.$$

**Theorem 1** Let  $LD$  be a linguistic description (5), let  $\{\mathcal{R}_i, \mathcal{R}_j\} \subseteq LD$ . Rule  $\mathcal{R}_i$  is suspicious of redundancy with respect to  $\mathcal{R}_j$  if and only if  $\mathcal{A}_i \leq_{LE} \mathcal{A}_j$  and  $B_i = B_j$ .

Theorem 1 claims that a fuzzy rule with an antecedent overlapped by an antecedent of another rule with the identical consequent is suspicious of the redundancy w.r.t. that rule. Furthermore, there are no other fuzzy rules that could be suspicious of redundancy w.r.t. another fuzzy rule besides those that meet the above mentioned situation. Thus, Theorem 1 specifies fuzzy rules that makes sense to investigate.

### 3.2. Detection of suspicious rules and their cancellation

Due to the involvement of other rules the suspicious rules do not have to be necessarily redundant which may be demonstrated easily. Let us consider a linguistic description  $LD$  with  $\{\mathcal{R}_i, \mathcal{R}_j, \mathcal{R}_k\} \subseteq LD$  where  $\mathcal{R}_i \hookrightarrow \mathcal{R}_j$ , antecedents are ordered as follows  $\mathcal{A}_i \leq_{LE} \mathcal{A}_k \leq_{LE} \mathcal{A}_j$  and where the consequent  $B_k$  is different from the consequents  $B_i = B_j$ . Then fuzzy rule  $\mathcal{R}_k$  “cancels” the redundancy of which  $\mathcal{R}_i$  was suspicious, see Figure 3 for a visualization.

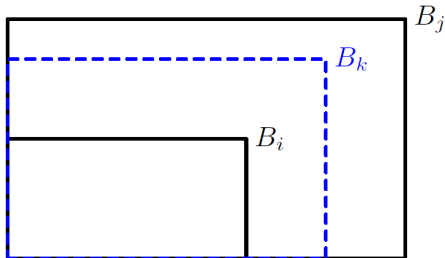


Figure 3: Fuzzy rule  $\mathcal{R}_k$  “cancels” the potential redundancy of fuzzy rule  $\mathcal{R}_i$  w.r.t.  $\mathcal{R}_j$ . Rectangles denoting  $\mathcal{R}_i, \mathcal{R}_j$  are black and solid to symbolize that  $B_i = B_j$ . Area where  $\mathcal{R}_k$  fires is displayed by blue dashed line in order to symbolize that  $B_k \neq B_i(B_j)$ .

Naturally, one could release hypotheses stating the situation when fuzzy rule  $\mathcal{R}_i$  that is suspicious of redundancy wrt.  $\mathcal{R}_j$  is not redundant. Two such hypotheses were formulated in [1]. However, the authors also showed that such hypotheses are not generally valid because there might be some other rules that “cancel the cancellation”. For a visualization of such cancellation of a cancellation we refer to Figure 4 and Figure 5.

Nevertheless, the non-valid hypotheses may be rewritten into two valid theorems if we consider linguistic description that have only three rules [1].

**Theorem 2** Let  $LD = \{\mathcal{R}_i, \mathcal{R}_j, \mathcal{R}_k\}$  and let  $\mathcal{R}_i$  be suspicious of the redundancy with respect to  $\mathcal{R}_j$ . If

- (1)  $B_k \neq B_i$ ,
- (2)  $\mathcal{A}_k \leq_{LE} \mathcal{A}_j$ ,

and either

- (3a)  $\mathcal{A}_i \leq_{LE} \mathcal{A}_k$ ,

or

- (3b)  $\mathcal{A}_i \parallel_{LE} \mathcal{A}_k$ , ( $\parallel_{LE}$  stands for incomparability)

then  $\mathcal{R}_i$  is NOT redundant in  $LD$ .

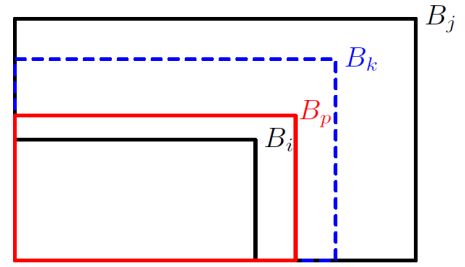


Figure 4: Fuzzy rule  $\mathcal{R}_p$  with  $B_p = B_i(B_j)$  cancels the cancellation by  $\mathcal{R}_k$ .

**Theorem 3** Let  $LD = \{\mathcal{R}_i, \mathcal{R}_j, \mathcal{R}_k\}$  and let  $\mathcal{R}_i$  be suspicious of the redundancy with respect to  $\mathcal{R}_j$ . If

- (4)  $B_k \neq B_i$ ,
- (5)  $\mathcal{A}_k \parallel_{LE} \mathcal{A}_j$ , but  $\mathcal{A}_k, \mathcal{A}_j$  have the same atomic expression,
- (6)  $\mathcal{A}_i \leq_{LE} \mathcal{A}_k$ ,

then  $\mathcal{R}_i$  is NOT redundant in  $LD$ .

Theorems 2 and 3 were formulated for a linguistic description that consist of only three rules, which makes their importance from a practical point of view rather low. Nevertheless, their existence is justified by the following theorem that stems from them. This theorem already provides us with a general result for an arbitrary number of fuzzy IF-THEN rules.

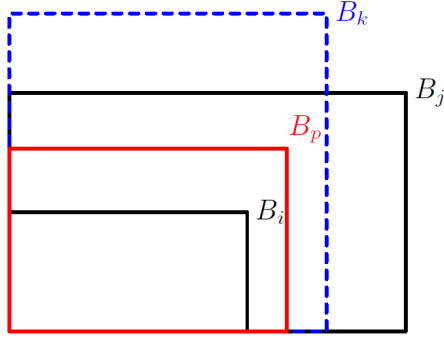


Figure 5: Fuzzy rule  $\mathcal{R}_p$  with  $\mathcal{B}_P = \mathcal{B}_i(\mathcal{B}_j)$  cancels the cancellation by  $\mathcal{R}_k$ .

**Theorem 4** Let  $LD = \{\mathcal{R}_1, \dots, \mathcal{R}_m\}$  be a linguistic description (5) and let  $\mathcal{R}_i \hookrightarrow \mathcal{R}_j$ . If there exists no rule  $\mathcal{R}_k \in LD$  such that either (1)-(3a), (1)-(3b) or (4)-(6) holds, then  $\mathcal{R}_i$  is redundant in  $LD$ .

The main contribution of Theorem 4 is that it provides a full classification of such rules that may cancel the suspicion of redundancy and that no other fuzzy rules may be responsible for this. Hence, we may introduce the notion of *cancelling rule*.

**Definition 3** Let  $LD = \{\mathcal{R}_1, \dots, \mathcal{R}_m\}$  be a linguistic description (5) and let  $\mathcal{R}_i \hookrightarrow \mathcal{R}_j$ . If either (1)-(3a), (1)-(3b) or (4)-(6) holds for  $\mathcal{R}_k \in LD$  then the rule  $\mathcal{R}_k$  is called *cancelling*.

Theorem 4 then actually states that if some suspicion  $\mathcal{R}_i \hookrightarrow \mathcal{R}_j$  exists and no cancelling rule exists in  $LD$ , then  $\mathcal{R}_i$  is redundant.

#### 4. Complex answer

Section 3 provided a full classification of cancelling rules. Hence, we know in which situation a given suspicion may be cancelled. However, we also know that even such cancellation may be also eliminated (by another rule) and the suspicious rule may be really redundant even if there exists a cancelling rule.

In this section we attempt to obtain a more complex answer on a given question whether a rule is redundant in a given linguistic description or not.

**Theorem 5** Let  $LD$  be a linguistic description, let  $\{\mathcal{R}_i, \mathcal{R}_j, \mathcal{R}_k\} \subseteq LD$  and let  $\mathcal{R}_i \hookrightarrow \mathcal{R}_j$ . Furthermore, let (1) – (3a) or (4) – (6) hold for  $\mathcal{R}_k$  and no further cancelling rule related to  $\mathcal{R}_i \hookrightarrow \mathcal{R}_j$  exists in  $LD$ . Then it holds that if  $\mathcal{R}_i$  is redundant in  $LD$  then there exists a rule  $\mathcal{R}_p \in LD$ ,  $\mathcal{R}_p \neq \mathcal{R}_k$  such that

- a)  $\mathcal{R}_i \hookrightarrow \mathcal{R}_p$ ,
- b)  $\mathcal{A}_p \leq_{LE} \mathcal{A}_k$ .

Theorem 5 claims that if we have a cancelling rule fulfilling (1) – (3a) or (4) – (6), we do not have to investigate this pair  $\mathcal{R}_i \hookrightarrow \mathcal{R}_j$  anymore because the influence of the cancellation rule may be *eliminated* only by another rule  $\mathcal{R}_p$  with respect to which  $\mathcal{R}_i$  is suspicious of being redundant and moreover, the cancelling rule does not have the cancellation property with respect to this “*eliminating*” rule  $\mathcal{R}_p$ . It means that in order to detect the redundancy of  $\mathcal{R}_i$  it is sufficient to investigate this new suspicion  $\mathcal{R}_i \hookrightarrow \mathcal{R}_p$ .

Theorem 5 is crucial, but we should investigate also the situation when a cancelling rule fulfils properties (1) – (3b), which is the most complicated case. The reason is that the elimination is not always done by a rule to which the investigated rule  $\mathcal{R}_i$  would be also suspicious. However, a satisfactory answer is obtained even for this case.

**Theorem 6** Let  $LD$  be a linguistic description, let  $\{\mathcal{R}_i, \mathcal{R}_j, \mathcal{R}_k\} \subseteq LD$  and let  $\mathcal{R}_i \hookrightarrow \mathcal{R}_j$ . Furthermore, let (1) – (3b) holds for  $\mathcal{R}_k$  and no further cancelling rule related to  $\mathcal{R}_i \hookrightarrow \mathcal{R}_j$  exists in  $LD$ . Then it holds that  $\mathcal{R}_i$  is redundant in  $LD$  if and only if there exists a rule  $\mathcal{R}_p \in LD$ ,  $\mathcal{R}_p \neq \mathcal{R}_k$  such that

- (a)  $\mathcal{A}_k \not\leq_{LE} \mathcal{A}_p$ ,
- (b)  $\text{Ker}(\text{Int}(X \text{ is } \mathcal{A}_i)(w)) \cap \text{Ker}(\text{Int}(X \text{ is } \mathcal{A}_k)(w)) \subseteq \text{Ker}(\text{Int}(X \text{ is } \mathcal{A}_p)(w))$ , for any  $w \in W$ ,
- (c)  $\mathcal{B}_p = \mathcal{B}_i$  or  $\mathcal{A}_p \leq_{LE} \mathcal{A}_i$ .

Based on all the results introduced in Sections 3 and 4, we may design the following algorithm that searches for redundant rules in a given linguistic description and removes them.

**Algorithm:** input  $LD$ ,

- 1) Preprocessing (using Remark 1).
- 2) Search for all pairs  $\mathcal{R}_i \hookrightarrow \mathcal{R}_j$  in  $LD$ , denote them as Investigated Pairs  $IP$ .
- 3a) For a pair  $\mathcal{R}_i \hookrightarrow \mathcal{R}_j \in IP$  search for an  $\mathcal{R}_k \in LD$  (using Theorem 4). If there is no such  $\mathcal{R}_k$ , delete  $\mathcal{R}_i$  from  $LD$  and delete all pairs containing  $\mathcal{R}_i$  from  $IP$ .
- 3b) If there is such an  $\mathcal{R}_k \in LD$  for which either (1) – (3a) or (4) – (6) holds, the pair  $\mathcal{R}_i \hookrightarrow \mathcal{R}_j$  is deleted from  $IP$ .
- 4) Repeat step 3) for all the pairs from  $IP$ .
- 5) For a pair  $\mathcal{R}_i \hookrightarrow \mathcal{R}_j \in IP$  search for an  $\mathcal{R}_k \in LD$  for which (1) – (3b) holds. If there is no other cancelling rule related to this pair, search for an  $\mathcal{R}_p \in LD$  fulfilling (a) – (c) from Theorem 6. If there is such an  $\mathcal{R}_p$ , then delete  $\mathcal{R}_i$  from  $LD$  and delete all pairs containing  $\mathcal{R}_i$  from  $IP$ . Otherwise delete only the pair  $\mathcal{R}_i \hookrightarrow \mathcal{R}_j$  from  $IP$ .
- 6) Repeat step 5) for all the pairs from  $IP$ .



## 5. Application

In [18], the authors studied an ensemble approach to time series forecasting which should avoid the danger of choosing an inappropriate forecasting method for a given time series by a combination of several methods. The combination is defined as a weighted mean of forecasts by individual methods and the goal of [18] was to determine appropriate weights of individual methods using fuzzy rules. Particularly, there were 7 individual methods combined in the overall forecast. The weight of each method for each time series prediction was determined with help of a linguistic description using quantitative features<sup>3</sup> of the given time series as antecedent variables. Thus, 7 linguistic descriptions, each determining a weight of an individual method, had to be determined.

A detailed analysis of the dependence of the precision of each method on the chosen features plays an essential role. In [19], the authors have shown that this exploration may be done with help of linguistic associations mining, particularly, with a fuzzy variant of the GUHA method originally proposed by P. Hájek [20]. However, the fuzzy GUHA method [10, 21] necessarily produces lots of approved yet redundant implicative associations that may be viewed as fuzzy IF-THEN rules. Obviously, an efficient method that significantly decreases the number of fuzzy rules in the generated linguistic descriptions but without any influence on their behavior, is highly desirable.

The seven chosen individual methods were the following ones: Seasonal ARIMA, Exponential Smoothing (abb. ES), Decomposition Technique (DT), Random Walk (RW), Random Walk with a drift (RWd), GARCH and Moving Averages (MA), see Table 2. Expected precision of each of the method is dependent on various features and thus, the different dimensionality led to different numbers of generated rules.

Methods	No. of generated rules	Reduced no.
ARIMA	7240	141
DT	9	3
ES	686	31
GARCH	17	7
MA	324	25
RW	234	23
RWd	152	20

Table 2: Number of rules generated by GUHA and number of rules after post-processing.

As we may see from Table 2, the theoretical research that led to a design of the algorithm introduced in Section 4 significantly reduced the number of fuzzy rules in these linguistic descriptions.

Additionally, we provide readers with one of the linguistic descriptions before and after the redun-

<sup>3</sup>E.g., seasonality, frequency, kurtosis, skewness or coefficient of variation (CV).

Rule	IF part		THEN part
	Kurtosis	CV	$w_{\text{GARCH}}$
$\mathcal{R}_1$	Me	Sm	Ro Bi
$\mathcal{R}_2$	ML Me	Sm	Ro Bi
$\mathcal{R}_3$	ML Me	Ve Sm	Ro Bi
$\mathcal{R}_4$	ML Sm	Ve Sm	Ro Bi
$\mathcal{R}_5$	Ro Me	Ex Sm	Ro Bi
$\mathcal{R}_6$	Ro Me	Ex Sm	ML Bi
$\mathcal{R}_7$	Ro Me	ML Sm	Ro Bi
$\mathcal{R}_8$	Ro Me	Sm	Ro Bi
$\mathcal{R}_9$	Ro Me	Ve Sm	Ro Bi
$\mathcal{R}_{10}$	Ro Me	Ve Sm	ML Bi
$\mathcal{R}_{11}$	Sm	Ro Me	Ro Bi
$\mathcal{R}_{12}$	Sm	Sm	Ro Bi
$\mathcal{R}_{13}$	—	Ex Sm	Ro Bi
$\mathcal{R}_{14}$	—	Ex Sm	ML Bi
$\mathcal{R}_{15}$	—	Sm	Ro Bi
$\mathcal{R}_{16}$	—	Ve Sm	Ro Bi
$\mathcal{R}_{17}$	—	Ve Sm	ML Bi

Table 3: Fuzzy rules setting up the weight of the GARCH method. Red color denotes redundant rules, blue color denotes rules remaining in the description.

dancy analysis (Table 3). Because of the space requirements, we choose the linguistic description that sets-up the weight of the GARCH method and thus, implicitly, determines a class of time series (with particular features) for which this method usually works. One may easily see, that e.g.  $\mathcal{R}_1 \leftrightarrow \mathcal{R}_7$ , but, there is a cancelling rule  $\mathcal{R}_6$  fulfilling (1) – (3b) and no eliminating rule. However,  $\mathcal{R}_1$  is redundant anyhow because  $\mathcal{R}_1 \leftrightarrow \mathcal{R}_8$  also holds and there is no cancelling rule related to this suspicion. Later on, also  $\mathcal{R}_8$  is deleted as redundant because of the suspicion  $\mathcal{R}_8 \leftrightarrow \mathcal{R}_7$  and no cancelling rule.

## 6. Conclusions and future work

As we have shown, intuitively redundant rules are not always redundant and thus, a deeper and formally correct approach had to be introduced.

Our approach is based on detecting the rules that are suspicious of redundancy and their further investigation. Full classification of the rules that are suspicious of redundancy has been provided. We also obtained a full classification of rules that may cancel the suspicion of redundancy, so called “cancelling rules”. Finally, we have presented theoretical results that allowed us to construct an algorithm that detects and deletes redundant rules. It works in such a way that the behavior of the linguistic description is preserved but, as our application show, the number of rules is reduced significantly. This formal understanding of redundancy, which stresses the fact that original and new linguistic descriptions are equivalent from the point of view of their behavior, is significantly different in comparison with

other approaches aiming mainly at a simplification of linguistic descriptions [3, 4] that use various techniques, e.g. rules merging. Of course, their use may be also beneficial. However, there is no guarantee that the output of simplified linguistic descriptions is equivalent with the output of the original one.

It should be recalled, that especially if fuzzy IF-THEN rules are generated automatically from data, then redundant rules can occur quite often. Higher dimensionality may even strengthen this unwanted effect. The detection and removal of such rules can be really useful from the point of view of performance and interpretability. We have presented one of such real-life examples where the introduced redundancy analysis made possible to apply so far totally unapplicable yet theoretically approved fuzzy rules generated by the linguistic associations mining procedure.

## Acknowledgements

This work was supported by the European Regional Development Fund in the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070). Furthermore, we gratefully acknowledge partial support of project KONTAKT II - LH12229 of MŠMT ČR and of SGS06/PřF/2013 of the University of Ostrava.

## References

- [1] A. Dvořák, M. Štěpnička, and L. Vavříčková. Redundancies in systems of fuzzy/linguistic IF-THEN rules. In *Proc. of EUSFLAT-LFA 2011*, pages 1022–1029, Aix-les-Bains, July 2011.
- [2] R. Babuška and M. Setnes. Data-driven construction of transparent fuzzy models. In H.B. Verbruggen, H.-J. Zimmermann, and R. Babuška, editors, *Fuzzy Algorithms for Control*, pages 83–106. Kluwer, Boston, 1999.
- [3] M. Setnes. *Fuzzy rule base simplification using similarity measures (MSc Thesis)*. Delft University of Technology, Delft, Netherlands, 1995.
- [4] M. Setnes, V. Lacroze, and A. Titli. Complexity reduction methods for fuzzy systems. In H.B. Verbruggen, H.-J. Zimmermann, and R. Babuška, editors, *Fuzzy Algorithms for Control*, pages 185–218. Kluwer, Boston, 1999.
- [5] S. Galichet and L. Foulloy. Size reduction in fuzzy rulebases. In *Proc. IEEE International Conference On Systems, Man and Cybernetics*, pages 2107–2112, San Diego, 1998.
- [6] V. Novák. Perception-based logical deduction. In B. Reusch, editor, *Computational Intelligence, Theory and Applications*, Advances in Soft Computing, pages 237–250, Berlin, 2005. Springer.
- [7] V. Novák, M. Štěpnička, A. Dvořák, I. Perfilieva, V. Pavliska, and L. Vavříčková. Analysis of seasonal time series using fuzzy approach. *International Journal of General Systems*, 39:305–328, 2010.
- [8] 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*, 180:164–184, 2011.
- [9] V. Novák, I. Perfilieva, and N. G. Yarushkina. A general methodology for managerial decision making using intelligent techniques. In E. Rakus-Andersson, editor, *Recent Advances in Decision Making*. Springer, Berlin Heidelberg, 2009.
- [10] V. Novák, I. Perfilieva, A. Dvořák, Q. Chen, Q. Wei, and P. Yan. Mining pure linguistic associations from numerical data. *International Journal of Approximate Reasoning*, 48:4–22, 2008.
- [11] V. Novák. A comprehensive theory of trichotomous evaluative linguistic expressions. *Fuzzy Sets and Systems*, 159(22):2939–2969, 2008.
- [12] V. Novák and A. Dvořák. Formalization of commonsense reasoning in fuzzy logic in broader sense. *Applied and Computational Mathematics*, 10:106–121, 2011.
- [13] L. A. Zadeh. Precisiated natural language. *AI Magazine*, 25:74–91, 2004.
- [14] L. A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning I–III. *Inform. Sci.* 8 (1975) 199–250, 8 (1975) 301–357, 9 (1975) 43–80, 1975.
- [15] A. Dvořák, H. Habiballa, V. Novák, and V. Pavliska. The software package LFLC 2000 - its specificity, recent and perspective applications. *Computers in Industry*, 51:269–280, 2003.
- [16] M. De Cock and E.E. Kerre. Fuzzy modifiers based on fuzzy relations. *Information Sciences*, 160:173–199, 2004.
- [17] E. Hajičová, B. Partee, and P. Sgall. *Topic-focus Articulation, Tripartite Structures, and Semantic Content*. Kluwer, Dordrecht, 1998.
- [18] 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 (Advances in Intelligent Systems and Computing 190)*, pages 379–387. Springer-Verlag, 2013.
- [19] L. Štěpničková, M. Štěpnička, and D. Sikora. Fuzzy rule-based ensemble with use of linguistic associations mining for time series prediction. In *Proc. of EUSFLAT 2013*, in this volume, Milano, 2013.
- [20] P. Hájek. The question of a general concept of the GUHA method. *Kybernetika*, 4:505–515, 1968.
- [21] J. Kupka and I. Tomanová. Some extensions of mining of linguistic associations. *Neural Network World*, 20:27–44, 2010.