Fuzzy rule-based ensemble with use of linguistic associations mining for time series prediction

Lenka Štěpničková[1] **Martin Štěpnička**[1] **David Sikora**[2]

[1]CE IT4Innovations - Division of the University of Ostrava - IRAFM, 30. dubna 22, 701 03 Ostrava, CZ
[2]Dept. of Informatics, Faculty of Science, University of Ostrava, 30. dubna 22, 701 03 Ostrava, CZ

Abstract

There are many various methods to forecast time series. However, there is no single forecasting method that generally outperforms any other. Consequently, there always exists a danger of choosing a method that is inappropriate for a given time series. To overcome such a problem, distinct ensemble techniques are being proposed. These techniques combine more individual forecasting methods.

In this contribution, we employ the so called fuzzy rule-based ensemble to determine the weights based on time series features such as trend, seasonality or stationarity. For identification of fuzzy rule base, we use linguistic association mining. An exhaustive experimental justification is provided.

Keywords: Time series, fuzzy rules, ensembles, Fuzzy Rule Based Ensemble, fuzzy GUHA, linguistic associations, perception-based logical deduction

1. Introduction and motivation

1.1. Introduction

Time series prediction is an important tool for support of individual and organizational decision making. It has a wide practical use in economy, industry, demography, and other areas of application. The time series is usually given as a finite sequence $y_1, y_2, \ldots, y_T$ of real numbers and the task is to predict future values $y_{T+1}, y_{T+2}, \ldots, y_{T+h}$ where h denotes so called forecasting horizon. There are many different methods to this task which are nowadays widely used in practise, let us recall e.g. well-known Box-Jenkins methodology [1] which consists of autoregressive and moving average models, decomposition method which is based on a decomposition a given time series into the components, or exponential smoothing methods. Further, a notable number of works aiming at fuzzy approach to time series analysis and prediction has been published. For instance, a study presenting Takagi-Sugeno rules [2] in the view of the Box-Jenkins methodology [3] or the works dealing with the linguistic approach [4, 5] have been published. Analogously, various neuro-fuzzy approaches, which lie on the border between neural networks, Takagi-Sugeno models and evolving fuzzy systems, are very often successfully used

[6, 7]. Unfortunately, there is no single forecasting method that generally outperforms any other. Thus, there is a danger of choosing a method which is inappropriate for a given time series. We note that even searching for methods that outperform any other for narrower specific subsets of time series has not been successful yet, see [8]:

"Although forecasting expertise can be found in the literature, these sources often fail to adequately describe conditions under which a method is expected to be successful".

1.2. Ensembles

In order to eliminate the risk of choosing an inappropriate method, distinct ensemble techniques, ensembles in short, have been designed and successfully applied. The main idea of ensembles consists in an appropriate combination of more forecasting methods. Typically, ensemble techniques are constructed as a linear combination of the individual ones. It can be described as follows. Let us assume that we are given a set of M individual methods and for a given times series $y_1, y_2, \ldots, y_T$ and a given forecasting horizon h, j-th individual method provides us with the following prediction:

$$\hat{y}_{T+1}^{(j)}, \hat{y}_{T+2}^{(j)}, \ldots, \hat{y}_{T+h}^{(j)}, \quad j = 1, \ldots, M.$$

Then the ensemble forecast is given by the following formula:

$$\hat{y}_{T+i} = \frac{1}{\sum_{j=1}^{M} w_j} \cdot \sum_{j=1}^{M} w_j \cdot \hat{y}_{T+i}^{(j)}, \quad i = 1, \ldots, h$$

where $w_j \in \mathbb{R}$ is a weight of the j-th individual method. These weights are usually normalized, that is $\sum_{j=1}^{M} w_j = 1$.

Let us recall that Bates and Granger [9] was one of the first to show significant gains in accuracy through combination. Another early work by Newbold and Granger [10] combined various time series forecasts and compared the combination against the performance of the individual methods. They show that for set of forecasts, a linear combination of these forecasts could be obtained which would also be unbiased and achieve a combined forecast error variance smaller than the individual forecasts. They found that the better combining procedures

did produce an overall forecast superior to individual forecasts on the majority of tested time series.

How to combine methods, i.e., how to determine appropriate weights, is still a relatively open question. For instance, Makridakis et al. [11] show that taking a simple average outperforms taking a weighted average method combination. In other words, the so called "equal-weights combining" [12], that is an arithmetic mean, is a benchmark that is hard to beat and finding appropriate non-equal weights rather leads to a random damage of the main averaging idea that is behind the robustness and accuracy improvements.

1.3. Motivation for the suggested approach

Although the equal-weights performs as accurately as mentioned above, there are works that promisingly show the potential of more sophisticated approaches. We recall Lemke and Gabrys [13] that describes an approach using meta-learning for time series forecasting based on the features of time series such as: measure for the strength of the trend, standard deviation, skewness, etc. Given time series are clustered using the k-means algorithm. Individual methods are ranked according to their performance on each cluster and then three best methods for each cluster are selected. For a given new time series, the closest cluster is determined and the given three best methods are combined.

It should be stressed that this approach performed very well on sufficiently big set of time series. For us, it is one of the main motivations because it demonstrates that there exists a dependence between time series features and success of forecasting method.

The second major motivation stems from the so called *Rule-Based Forecasting* (RBF) developed by Collopy and Armstrong [8, 12]. It is an expert system that uses domain knowledge to combine forecasts from various forecasting methods. Using IF-THEN rules, RBF determines what weights to give to the forecasts.

We follow the main ideas of rule-based forecasting [8] and of using time series features [13] to obtain an interpretable and understandable model.

2. Fuzzy Rule-Based Ensemble

As mentioned above, RBF uses the rules to determine weights [8]. However, only few of these rules are directly used to set up weights. Most of them set up rather a specific model parameters, e.g. the smoothing factors of the Brown's exponential smoothing with trend. Moreover, in antecedents the rules very often use properties that are not crisp but rather vague, e.g. expressions such as: "last observation is unusual; trend has been changing; unstable recent trend" etc., see [12]. For such cases, using crisp rules that are either fired or not and nothing between, seems to be less natural than using fuzzy

rules. Similarly, the use of crisp consequents such as: "add 10% to the weight; subtract 0.4 from beta; add 0.1 to alpha" etc. [12], seems to be less intuitive than using vague expressions that are typical for fuzzy rules.

2.1. General structure of the model

Therefore, our goal was to propose a method that uses fuzzy rules instead of crisp rules in order to capture the omnipresent vagueness in the expressions; to use only quantitative features (no domain knowledge) in the antecedent variables which enable to fully automatize the method; to use only individual forecasting method weights as the consequent variables [14, 15]. The result of such motivated investigation is the Fuzzy Rule-Based Ensemble (FRBE) that is schematically illustrated on Figure 1.
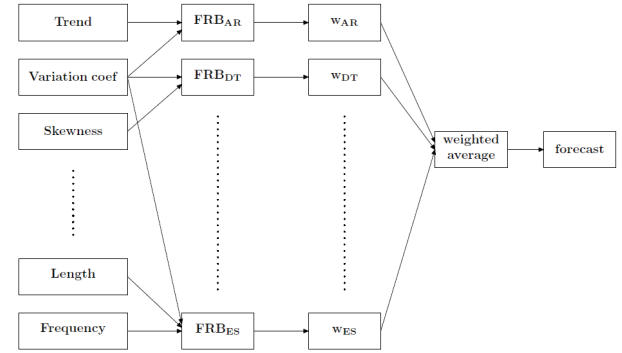

Figure 1: Structure of the FRBE method.

The FRBE method uses a single linguistic description, i.e. fuzzy rule base with evaluative linguistic expressions [16], for each forecasting method. Each of these linguistic descriptions determines the weights of a single individual method based on transparent and interpretable rules, such as:

"**IF** *Strength of Seasonality is Small* **AND** *Coefficient of Variation is Roughly Small* **THEN** *Weight of the j-th method is Big.*"

After an appropriate inference method is applied (see Section 2.2) in order to obtain a fuzzy output, a defuzzification method is employed and thus, a crisp result (weight of a particular method) is determined.

So far, based on experiments and previous publications [13] the following features were considered: *strength of trend, strength of seasonality, length of the time series, skewness, kurtosis, coefficient of variation, stationarity and frequency.* With help of the forward stepwise regression on the training set of the time series, only the statistically significant features were used for each method [14, 15]. Note, that for each method, different features played the significant role.

2.2. Components of the model

In order to estimate (set up) a particular value of the weight of each forecasting method with help of the fuzzy rules, an appropriate fuzzy inference mechanism has to be employed. As mentioned above, the FRBE method employs linguistic descriptions, i.e. fuzzy rule bases with so called evaluative linguistic expressions. These are expressions of natural language that are based on the expressions of the basic trichotomy `Small (Sm)`, `Medium (Me)`, and `Big (Bi)`. The expression of the basic trichotomy may be modified using linguistic hedges either with narrowing or with widening effect, see Table 1.

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.

Such linguistic expressions have their theoretical model of the semantics based on intension, context and extension, which is in detail described in the referred literature [16]. For the purpose of this contribution, it is sufficient to mention that extensions, that model the meaning in a given context $[v_L, v_R]$, are fuzzy sets that are depicted on Figure 2. One may easily see the influence of the modifiers on the shape of the extensions.
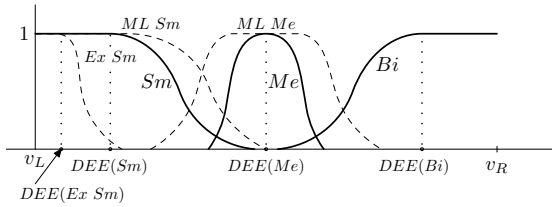

Figure 2: Shapes of extensions (fuzzy sets) of evaluative linguistic expressions.

If a fuzzy rule base is viewed as a linguistic description and thus, uses the above recalled evaluative linguistic expressions with their model of semantics, one can neither model the rules (and consequently the whole description) as a conjunction of implicative rules nor as a disjunction of conjunctions (Mamdani-Assilian model). The used expressions, mainly their full overlapping, require a specific inference method – *Perception-based Logical Deduction* (PbLD) [17]. This method models each fuzzy rule

$$\mathcal{R}_i := \text{IF } X \text{ is } \mathcal{A}_i \text{ THEN } Y \text{ is } \mathcal{B}_i,$$

by a fuzzy relation R_i on $X \times Y$ given as follows:

$$R_i(x, y) = A_i(x) \to_{\text{Ł}} B_i(y)$$

where $\to_{\text{Ł}}$ is the Łukasiewicz implication [18] given by $a \to_{\text{Ł}} b = 1 \wedge (1 - a + b)$. However, unlike in the case of implicative rules, the rules are not aggregated conjunctively. The PbLD uses a specific algorithm (perception) that chooses only some rules to be used in the inference. These are the most specific among the most fired rules. And only the outputs obtained based on these fuzzy rules are aggregated by the intersection at the final stage. For details regarding the algorithm, we refer e.g. to [19, 20].

Finally, the inferred output is defuzzified. This is done by the *Defuzzification of Evaluative Expressions* (DEE) that has been designed specifically for the outputs of the PbLD inference mechanism. 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`, the LOM is applied; if the inferred output is of the type `Medium`, the MOM is applied; and finally, if the inferred output is of the type `Big`, the FOM is applied, see Figure 2. In the case of the FRBE method, the defuzzification DEE is applied after the inference so that, the deduced weights $w_{AR}, w_{DT}, \ldots, w_{ES}$ displayed on Figure 1 are already crisp numbers.

2.3. Fuzzy rule base identification

The last missing point is the identification of the linguistic descriptions. This may be done by distinct approaches. One could expect a deep applicable expert knowledge however, neither our experience nor the experience of others confirms this expectations. Let us once more refer to the observation of Armstrong, Collopy and Adya in [8] already recalled in Section 1.

Because of the missing reliable expert knowledge, we focus on data-driven approaches that may bring us the interpretable knowledge hidden in the data.

However, before we apply any data-mining technique, we have to clarify how we interpret the weights in the data because only the features serving as antecedent variables are measured. Naturally, the individual method weights should be proportionally higher if a given method is supposed to provide lower forecasting error and vice-versa. Thus, it is natural to put

$$w_j = 1 - acc_j$$

where acc_j denotes an appropriate normalized forecasting error of the j-th method, and to estimate directly the weight of the j-th method w_j instead of the value of the forecasting error. Now, any appropriate data-mining technique may be applied in order to determine the dependence between features and the precision (weight) of each method.

3. Fuzzy GUHA – linguistic association mining

In this paper, we employ the so called linguistic associations mining [21] for the fuzzy rule base identification. This approach, firstly introduced as GUHA method [22, 23], 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, $X_1, \ldots, X_m$ 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.

	X_1	$\ldots$	X_m	Z
o_1	a_{11}	$\ldots$	a_{1m}	a_1
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
o_n	a_{n1}	$\ldots$	a_{nm}	a_n

Table 2: Standard GUHA table.

The original GUHA allowed only boolean attributes to be involved, see [24]. Since most of the features of objects are measured on the real interval, standard approach assumed to booleanize the attributes by a partition of the interval into subintervals, see Example 1.

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

$$C(X_1, \ldots, X_p) \simeq D(Z)$$

where C, D are (compound) evaluative predications [16] containing only the connective AND and $X_1, \ldots, X_p$ for $p \leq m$ are all variables occurring in C. The C, D are called the *antecedent* and *consequent*, respectively. Generally, for the GUHA method, the well-known four-fold table is constructed, see Table 3.

	D	not D
C	a	b
not C	c	d

Table 3: Classical GUHA four-fold table.

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, i.e. of 'not D'. Analogous meaning have the numbers c and d. For our purposes, only numbers a and b are important.

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

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

and

$$\frac{a}{m} > r,$$

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.*

	BMI$_{\leq 25}$	BMI$_{> 25}$	Chol$_{> 6.2}$	BP$_{> 130/90}$
o_1	1	0	0	0
o_2	0	1	1	1
o_3	0	1	0	1
o_4	1	0	0	0
o_5	0	1	1	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
o_n	0	0	1	1

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$_{> 6.2}$ denotes Cholesterol higher than 6.2 and BP$_{> 130/90}$ denotes Blood Pressure higher than 130/90. Objects o_i are particular patients.

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

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

In many situations, including our situation, the fuzzy variant of the GUHA method [25, 26] seems to be more appropriate. In the fuzzy variant of the method, the attributes are not boolean but rather vague (such as BMI$_{\text{ExBi}}$, BMI$_{\text{MLBi}}$, Chol$_{\text{VeBi}}$ etc.) and thus, values a_{ij} (or a_i) are elements of the interval $[0, 1]$ that express membership degrees.

The four-fold table analogous to Table 3 is constructed also for the fuzzy variant of the method. The difference is that the numbers a, b, c, d are not summations of 1s and 0s but summations of membership degrees of data into fuzzy sets representing the antecedent C and consequent D or their complements, respectively. Otherwise, the main idea of the method remains the same.

The advantage is that it searches for implicative associations that may be directly interpreted as fuzzy rules.

In our case, for each individual forecasting method, we have transformed the training data set of time series with their normalized features into a table similar to Table 5.

	F_1^{Sm}	$\ldots$	F_3^{Bi}	$W_{\text{ES}}^{\text{Sm}}$	$\ldots$	$W_{\text{ES}}^{\text{Bi}}$
TS$_1$	0.9	$\ldots$	0.7	0	$\ldots$	0.9
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
TS$_{99}$	0.1	$\ldots$	0.2	0.8	$\ldots$	0

Table 5: Transformed training data set for the Exponential Smoothing forecasting method.

Objects $TS_1, \ldots, TS_{99}$ in Table 5 are the time series from the training set, see Table 6; $F_1, \ldots, F_3$ are those three normalized features of given time series that were found to be statistically significant for the performance of the Exponential Smoothing method by the forward stepwise regression, see Section 2.1. Note, that there are significantly more than just 3 columns in this part of Table 5 because each evaluative linguistic expression leads to a single column for a single feature F_i, i.e., for the expression ExSm, there are three columns: F_1^{ExSm}, F_2^{ExSm} and F_3^{ExSm}. Symbol W_{ES} stands for the weight (accuracy) of the Exponential Smoothing method and again, there are as many columns in this part of the Table 5 as there exist so many evaluative linguistic expressions.

The fuzzy GUHA then combinatorically generates hypotheses that are immediately statistically either declined or confirmed as linguistic associations based on the chosen quantifier parameters, see Example 2.

Example 2 *Let F_1 be skewness and F_2 be coefficient of variation. Our fuzzy GUHA approach provided us with the following implicative hypothesis:*

$$C(Skew^{\mathrm{RoMe}}, \mathrm{CV}^{\mathrm{ExSm}}) \ \sqsubset_r^\gamma \ D(W_{\mathrm{DT}}^{\mathrm{MLBi}})$$

where Skew denotes the normalized skewness and CV denotes the coefficient of variation, that was confirmed on the following confidence degree and support degree:

$$\gamma = 0.7, r = 0.04,$$

respectively.

Such a confirmed association may be viewed and thus, directly interpreted, as the following fuzzy rule:

"IF *Skewness is Roughly Medium* AND *Coefficient of Variation is Extremely Small* THEN *Weight of the decomposition method is More or less Big.*"

Note, that the above described application of the fuzzy GUHA generates linguistic description determining the weight of a single method – in our example of the Exponential Smoothing. Thus, the method, including the transformation of training data set into a table similar to Table 5, has to be applied as many times as is the number of methods (and consequently of the linguistic descriptions). In our case, this led to the sevenfold use of the method as we deal with seven individual methods.

4. Implementation

To develop and validate the model we have used 198 time series from the M3 data set repository that contains 3003 time series from the M3-Competition

[27] and that serves as a generally accepted benchmark database provided by the authority of the International Institute of Forecasters. We have chosen the time series from 5 categories (Microeconomy, Macroeconomy, Industry, Finance, Demography) and with different time intervals between successive observations (yearly, quarterly, monthly). This selected data set was divided into 2 sets (training set and testing set) both containing 99 time series, see Table 6 and Table 7.

	Training set		
	Yearly	Quarterly	Monthly
Micro	5	7	7
Macro	7	8	8
Industry	7	7	6
Finance	7	7	5
Demography	7	4	7
Total	33	33	33

Table 6: Training data set.

	Testing set		
	Yearly	Quarterly	Monthly
Micro	5	7	7
Macro	7	8	8
Industry	7	7	6
Finance	7	7	5
Demography	7	4	7
Total	33	33	33

Table 7: Testing data set.

The training set was used for an identification of our model, that is for a fuzzy rule base identification. The testing set was used for a testing whether the determined knowledge encoded in the fuzzy rules works generally also for other time series.

To forecast the future values of the time series, we have chosen the most often used forecasting methods that are at disposal to the widest community of users: seasonal Autoregressive Integrated Moving Average (ARIMA), Decomposition Techniques (DT), Exponential Smoothing (ES), Generalized Autoregressive Conditional Heteroscedasticity models (GARCH), Moving Averages (MA), Random Walk process (RW) and Random Walk process with a drift (RWd). For details about these methods, we only refer to the relevant literature [1, 28, 29].

In order to avoid any bias from a naive implementation of the above listed methods, we adopted implementations of these methods by professional software package such as ForecastPro® for ARIMA, ES and MA; Gretl® for GARCH and RWd and NCSS® for DT. These tools executed fully automatic parameter selection and optimization which made possible to concentrate the investigation purely on the combination technique. Moreover, their arithmetic mean (AM), that represents the equal weights en-

semble, was also determined and used as a valid benchmark.

There are many accuracy measures that are used to analyze the performance of the various forecasting methods. However, very popular measures such as *Mean Absolute Error* or *(Root) Mean Squared Error* are inappropriate for comparison across more time series because they are scale-dependent. Therefore, we use *Symmetric Mean Absolute Percentage Error* (SMAPE) that is scale-independent and thus appropriate to compare methods across different time series [30]. It is defined as follows:

$$\text{SMAPE} = \frac{1}{h} \sum_{t=T+1}^{T+h} \frac{|y_t - \hat{y}_t|}{(|y_t| + |\hat{y}_t|)/2} \times 100\%.$$

For further investigation, some important features need to be extracted from a given time series. We used the following features: *strength of trend, strength of seasonality, length of the time series, skewness, kurtosis, coefficient of variation, stationarity and frequency.* Since most of them are standard and well-known, we do not describe them.

Since the range of some features can significantly vary, it is crucial to normalize the range of features to the interval $[0, 1]$ as well as the SMAPE accuracy measure.

Let us recall, we used the binary multitudinal quantifier $\simeq := \sqsubset_r^\gamma$ to find implicative associations between features of time series occuring in the antecedent and accuracy of individual forecasting method occuring in the consequent. For our purposes, we set up $\gamma = 0.7$ and $r = 0.04$.

5. Results

As mentioned above, the associations generated by GUHA method are implicative. Thus, they may be directly interpreted as fuzzy rules. Due to the big number of such generated rules, a post-processing was applied on these rules. This process consists in search and deletion the rules that are duplicate and redundant [19]. After that, the number of rules was significantly reduced, see Table 8.

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

Table 8: Number of rules generated by fuzzy GUHA method and number of rules after post-processing.

In order to judge its performance, the fuzzy rule-based ensemble was applied on the 99 time series from the testing set. Table 9 and Table 10 show that in the average and the standard deviation of SMAPE forecasting errors over all testing set, the fuzzy rule-based ensemble outperforms all individual forecasting methods. Moreover, the equal-weights, i.e. arithmetic mean (AM), has been outperformed as well.

Methods	Average Error
DT	21.59
GARCH	17.27
RWd	15.95
RW	15.26
MA	15.11
ARIMA	14.44
ES	14.43
AM	14.40
FRBE	**14.14**

Table 9: Average of the SMAPE forecasting errors.

Methods	Error Std.Dev.
DT	24.64
GARCH	21.32
RWd	20.73
RW	19.53
MA	19.36
ARIMA	20.41
ES	18.48
AM	18.51
FRBE	**18.24**

Table 10: Standard deviation of the SMAPE forecasting errors.

Although the improvement does not seem significant, it is evident that the fuzzy rule-based ensemble performs very well even against the equal-weights combining, i.e. a procedure that has performed well in prior studies. Moreover, statistical significance test has been performed. Namely, we have performed the t-test testing the null hypothesis that the mean value of the random variable ($\text{SMAPE}_{\text{AM}} - \text{SMAPE}_{\text{FRBE}}$) equals to zero against the alternative hypothesis that the mean value is positive. The null hypothesis was rejected in the standard significance level $\alpha = 0.05$. Particularly, the obtained p-value was equal to 0.009. Furthermore, we have also applied the Wilcoxon test testing the null hypothesis that median of the above defined random variable was equal to zero, the alternative hypothesis was defined analogously. Again, as in the previous case, we have rejected the null hypothesis on the same significance level with the p-value equal to 0.016 in this case.

Let us stress that the victory has been reached not only in the accuracy but also in the robustness (standard deviation of the SMAPE forecasting errors, see Table 10) which is perhaps even more im-

portant w.r.t. the goals of ensemble methods. This investigation also confirms that there is really a dependence between time series features and success of forecasting method. This fact is good motivation to continue in this topic.

Rule	IF part		THEN part
	Kurtosis	CV	w_{GARCH}
$\mathcal{R}_1$	ML Sm	Ve Sm	Ro Bi
$\mathcal{R}_2$	Ro Me	ML Sm	Ro Bi
$\mathcal{R}_3$	Ro Me	Ve Sm	ML Bi
$\mathcal{R}_4$	Sm	Ro Me	Ro Bi
$\mathcal{R}_5$	Sm	Sm	Ro Bi
$\mathcal{R}_6$	—	Sm	Ro Bi
$\mathcal{R}_7$	—	Ve Sm	ML Bi

Table 11: Fuzzy rules setting up the weight of the GARCH method. CV denotes the coefficient of variation.

In order to emphasize the linguistic nature of the approach, we provide readers with one of the linguistic descriptions generated by the fuzzy GUHA method in Table 11. Because of the small number of generated rules, we choose the linguistic description that set up the weight of the GARCH method. The fuzzy rules symbolically displayed in Table 11, can be easily read as conditional sentences of natural language. For example, let us take the first fuzzy rule $\mathcal{R}_1$:

"**IF** *Kurtosis is More or less Small* **AND** *Coefficient of Variation is Very Small* **THEN** *Weight of the GARCH method is Roughly Big*"

or the rule $\mathcal{R}_7$ can be read as:

"**IF** *Coefficient of Variation is Very Small* **THEN** *Weight of the GARCH method is More or less Big*".

Recall, that we have chosen the weight to be proportional to the expected method accuracy and thus, the weight and the accuracy may be freely replaced. This makes the rule even more interpretable, which underlines the goal of our approach.

6. Conclusions

In this contribution, we have clearly stated the main motivations and ideas for fuzzy rule-based ensemble, i.e. the rule-based forecasting and use of time series features.

The introduced fuzzy rule-based ensemble has been "equipped" with fuzzy rule bases that were generated by the fuzzy GUHA method. The performance of the final model has been verified on the testing set composed of 99 times series from the world-known benchmark data-set M3. The obtained results showed a slight but yet statistically significant improvement in the accuracy compared to the given benchmark ensemble – equal weights. Moreover, the proposed model achieved also lower standard deviation of the accuracy that confirms the improvement in the sense of "robustness".

Let us also recall the linguistic nature of the suggested approach motivated by the aim to obtain an interpretable and understandable model.

We can conclude that we have demonstrated the promising potential of the fuzzy rule-based forecasting that entitles us to continue in the future research. For instance, the other techniques for fuzzy rule base identification, e.g. fuzzy cluster analysis, are supposed to be experimentally evaluated. Furthermore, it is possible to combine the fuzzy rule bases obtained by various techniques in order to improve the results.

References

[1] G. Box and G. Jenkins. *Time Series Analysis: Forecasting and Control.* Holden-Day, San Francisco, 1976.

[2] T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man and Cybernetics*, 15:116–132, 1985.

[3] J. Aznarte, J. Benítez, and J. Castro. Smooth transition autoregressive models and fuzzy rule-based systems: Functional equivalence and consequences. *Fuzzy Sets and Systems*, 158:2734–2745, 2007.

[4] 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.

[5] M. Štěpnička, A. Dvořák, V. Pavliska, and L. Vavříčková. A linguistic approach to time series modeling with the help of the f-transform. *Fuzzy sets and systems*, 180:164–184, 2011.

[6] G. Leng, T. McGinnity, and G. Prasad. An approach for on-line extraction of fuzzy rules using a self-organising fuzzy neural network. *Fuzzy sets and systems*, 150:211–243, 2005.

[7] H. J. Rong, N. Sundararajan, G. B. Huang, and P. Saratchandran. Sequential adaptive fuzzy inference system (safis) for nonlinear system identification and prediction. *Fuzzy Sets and Systems*, 157:1260–1275, 2006.

[8] J. S. Armstrong, M. Adya, and F. Collopy. Rule-based forecasting using judgment in time series extrapolation. In J. S. Armstrong, ed-

itor, *Principles of Forecasting: A handbook for reasearchers and practitioners.* Kluwer Academic Publishers, Boston/Dordrecht/London, 2001.

[9] J. M. Bates and C. W. J. Granger. Combination of forecasts. *Operational Research Quarterly*, 20:451–468, 1969.

[10] P. Newbold and C. W. J. Granger. Experience with forecasting univariate time series and combination of forecasts. *Journal of the Royal Statistical Society Series a-Statistics in Society*, 137:131–165, 1974.

[11] S. Makridakis, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen, and R. Winkler. The accuracy of extrapolation (time-series) methods - results of a forecasting competition. *Journal of Forecasting*, 1:111–153, 1982.

[12] F. Collopy and J. S. Armstrong. Rule-based forecasting: Development and validation of an expert systems approach to combining time series extrapolations. *Management Science*, 38:1394–1414, 1992.

[13] C. Lemke and B. Gabrys. Meta-learning for time series forecasting in the nn gc1 competition. In *Proc. 16th IEEE Int. Conf. on Fuzzy Systems*, pages 2258–2262, Barcelona, 2010.

[14] 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*, volume 190 of *Advances in Intelligent Systems and Computing*, pages 379–387. Springer-Verlag, 2013.

[15] D. Sikora, M. Štěpnička, and L. Vavříčková. On the potential of fuzzy rule-based ensemble forecasting. In *International Joint Conference CISIS'12 - ICEUTE'12 - SOCO'12 SPECIAL SESSIONS*, volume 189 of *Advances in Intelligent Systems and Computing*, pages 487–496. Springer-Verlag, 2013.

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

[17] 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.

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

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

[20] L. Vavříčková, M. Štěpnička, and A. Dvořák. New results on redundancies of fuzzy/linguistic if-then rules. In *Proc. 8th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT-2013)*, page in this issue, Milano, 2013. Atlantic Press.

[21] R. Agrawal and R. Srikant. Fast algorithms for mining association rules. In *Proc. 20th Int. Conf. on Very Large Databases*, pages 487–499, Chile. AAAI Press.

[22] P. Hájek. The question of a general concept of the GUHA method. *Kybernetika*, 4:505–515, 1968.

[23] P. Hájek and T. Havránek. *Mechanizing hypothesis formation: Mathematical foundations for a general theory.* Springer-Verlag, Berlin/Heidelberg/New York, 1978.

[24] P. Hájek, M. Holeňa, and J. Rauch. The GUHA method and its meaning for data mining. *Journal of Computer and Systems Sciences*, 76:34–48, 2010.

[25] J. Kupka and I. Tomanová. Some extensions of mining of linguistic associations. *Neural Network World*, 20:27–44, 2010.

[26] 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*, 48:44–22, 2008.

[27] S. Makridakis and M. Hibon. The m3–competition: results, conclusions and implications. *International Journal of Forecasting*, 16:451–476, 2000.

[28] J. D. Hamilton. *Time Series Analysis.* Princeton University Press, New Jersey, 1994.

[29] S. Makridakis, S. Wheelwright, and R. Hyndman. *Forecasting: methods and applications.* John Wiley & Sons, USA, 2008.

[30] R. Hyndman and A. Koehler. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22:679–688, 2006.