Human Input about Linguistic Summaries in Time Series Forecasting

Katarzyna Kaczmarek
Systems Research Institute
Polish Academy of Sciences
Newelska 6
01-447 Warsaw, Poland
K.Kaczmarek@ibspan.waw.pl

Olgierd Hryniewicz
Systems Research Institute
Polish Academy of Sciences
Newelska 6
01-447 Warsaw, Poland
Olgierd.Hryniewicz@ibspan.waw.pl

Rudolf Kruse
Faculty of Computer Science
Otto-von-Guericke-University Magdeburg
Universitaetsplatz 2
D-39106 Magdeburg, Germany
Rudolf.Kruse@ovgu.de

***Abstract*—Finding an appropriate predictive model for time series and formulating its assumptions may become very challenging task. We propose to represent time series in a human-consistent way using linguistic summaries. Such summaries describe general trends in time series and are easily interpretable for decision makers. The aim of this contribution is to show that the linguistic summaries may be successfully applied to support the analysis and forecasting of time series. Information about trends is first retrieved from experts, and then, processed with soft computing tools. The performance of the approach is verified on the real-world datasets from the M3-Competition. Users are asked to evaluate linguistic summaries that are intuitive and easy for interpretation. This paper shows that human-consistent summaries deliver new knowledge for forecasting.**

Keywords–information retrieval; human-computer interaction; time series and sequence models; Bayesian methods; supervised learning.

I. INTRODUCTION

Practitioners are very often posed to the dilemma of choice between the wealth of mathematic models for forecasting for the imprecise real-world data. For a recent review of competitive forecasting models and methods, see e.g., Gooijer and Hyndman [1]. Within this research, the Box-Jenkins models are adapted. They are simple, and though, have been proven successful in various practical applications [2]–[4].

An important task of the Box-Jenkins time series analysis is the estimation of the unknown variables. One of the potential approach to this estimation is the Bayesian inference, that enables the inclusion of subjective prior information. Following Geweke [5], definitions for the prior probability distributions are usually assumed basing on expert's experience and intuitions, and normal or uniform distributions are often appropriate. However, experts may fail to adequately establish the prior distributions for the unknown variables and models, and then the problem arises. To conclude, the ability to describe the data imprecision in terms of prior probability distributions is one of the main advantages of the Bayesian approach, and at the same time, the main challenge for practitioners, because models may be difficult to understand for non-mathematician experts. Also, the proper selection of prior probability distributions is essential for the satisfactory forecasting performance.

Therefore, we propose to retrieve from experts the information about the expected trends in time series, and then, formulate the prior probability distributions automatically basing on this natural language information. The proposed approach realizes innovative incorporation of linguistic summaries into time series analysis.

The objective of this paper is to present this approach that consists of the human-computer interaction for the information retrieval, and then, its incorporation into the time series analysis and forecasting process. It assumes employing techniques from the following research fields: time series analysis and forecasting, the fuzzy set theory, the time series summarization and pattern mining, classification methods and the Bayesian analysis.

The comparative analysis of the forecasting accuracy is performed on time series from the M3-Competition by Makridakis and Hibon [6]. Experiments prove that the human-consistent summaries deliver new knowledge for forecasting.

The structure of this paper is as follows. Next section introduces basic definitions of the time series analysis. In Section III, the description of the human-computer interaction related to the linguistic summaries is explained. Section IV presents the proposed approach to incorporate the retrieved human-consistent knowledge into the Bayesian forecasting. Numerical results are gathered in Section V. In Section VI, conclusions are presented.

II. PRELIMINARIES: TIME SERIES ANALYSIS

Discrete time series $y = \{y_t\}_{t=1}^{n} \in Y$ is a sequence of observations measured at successive $t \in T = \{1, ..., n\}$ moments and at uniform time intervals, e.g., a sequence of monthly sales for a specific product builds up a time series.

A. Box-Jenkins Methodology

Due to the Box-Jenkins methodology [7], the time series analysis starts with the identification of the probabilistic model that generated the observed time series.

One of the most popular stationary models for forecasting, and though, very successful in applications are autoregressive and moving average (ARMA) models, defined as follows:
ARMA(p,q) [7]

$$\tilde{y_t} = \sum_{i=1}^{p} \phi_i \tilde{y_{t-i}} + \sum_{i=1}^{q} \alpha_i a_{t-i} + a_t; a_t \sim N(0, \sigma^2); \tilde{y_t} = y_t - \mu \quad (1)$$

where $\theta = \{\phi_1, ..., \phi_p, \alpha_1, ..., \alpha_q, \mu, \sigma^2\}$ *is the vector of unknown variables.*

ARMA models are based on the concept of a linear filter assuming that the observations are generated by the sequence $\{a_t\}_{t=1}^n$ of values taken by the independent and identically distributed random variable (white noise).

The class of autoregressive and moving average processes is rich, and in many contexts, it is usually possible to find a process or a combination of processes which provide an adequate description to the considered real-life time series data.

The Box-Jenkins approach to the time series analysis is an iterative process. After the identification of the probabilistic model, the following steps are performed: estimation of its parameters, verification techniques and finally, prediction. For further details, we refer to [2].

B. Bayesian Model Averaging

To diminish the risk of selecting one non-adequate model, multiple models may be combined through the Bayesian inference. Clemen and Winkler [8] show that combining various methods for forecasting on average leads to better results than applying individual ones.

The Bayesian model averaging enables to include multiple models and the posterior density $p(y_{n+h}|y, M)$ is a weighted average of the posterior densities of models $\{M_1, M_2, ..., M_J\}$:

$$p(y_{n+h}|y, M) = \sum_{j=1}^{J} p(y_{n+h}|y, M_j)p(M_j|M) \qquad (2)$$

The Bayesian averaging requires defining the prior model probability distributions $p(M_j|M)$. In [9], Ley and Steel show by theoretical and empirical evidence the critical importance of prior assumptions for the Bayesian model averaging. Within the proposed approach, these prior model probability distributions are automatically generated basing on the human-computer interaction.

III. HUMAN INPUT ABOUT LINGUISTIC SUMMARIES

In many domains, it is important to deliver results that are simple and easy to interpret by user. One may provide various forms of human-consistent descriptions of large datasets with the use of data mining and knowledge discovery techniques, and the literature on discovery of different information granules about time series data is extensive [4], [10]–[12]. Linguistic summaries are an example of information granules, and mining for linguistic summaries has also gained a lot of attention in the literature [13]–[15].

Within the proposed approach, we adapt the linguistic summaries in the sense of Yager [16] developed by Kacprzyk et al. [13], and we use the fuzzy set theory as introduced by Zadeh [17] to model the data imprecision.

A. Linguistic Summary Definition

Linguistic summaries describe general trends about the evolution of time series with quasi natural language, e.g., *Most increasing trends are short*.

Linguistic summary [16]
Let $A = \{a_1, a_2..., a_u\}$ denote a finite set of attributes (e.g., dynamics of change, duration). $S = \{l_1, l_2..., l_s\}$ is a finite set of imprecise labels for attributes (e.g., *quickly increasing trends, short trend*). The protoform-based linguistic summary

$$LS : Q\ R\ trends\ are\ P \qquad (3)$$

consists of quantifier Q (e.g., *most, among all*), summarizer P (attribute together with an imprecise label), qualifier R.

The imprecise labels refer to linguistic values of either qualitative or quantitative measurements for attributes (e.g., *low, increasing, short*). The interpretation for imprecise labels is modeled as fuzzy trapezoidal numbers. For further definitions, refer to e.g., [18], [19].

The quality of a linguistic summary is evaluated with **degree of truth (validity)** T due to [20], defined as follows:

$$T(LS) = \mu_Q\left(\frac{\sum_{i=1}^{n} (\mu_R(y_n) \wedge \mu_P(y_n))}{\sum_{i=1}^{n} \mu_R(y_n)}\right) \qquad (4)$$

where $\mu_R(y_n), \mu_P(y_n)$ are the membership functions $\mu_R, \mu_P : \Re \to [0, 1]$ determining the degree to which R, P, respectively, are satisfied for the time series y at the given moment n.

B. Linguistic Summary Retrieval

The following attributes and labels defining trends are considered to build up the linguistic summaries: duration (*short, medium, long*), dynamics (*increasing, constant, decreasing*) and variability (*low, moderate, high*). The resulting set of linguistic summaries may be exemplified by:
Most decreasing trends are medium;
Most trends are constant;
Most trends are decreasing.

If a time series is long, then the linguistic summaries are generated and evaluated automatically, e.g., with the Trend Analysis System [21]. However, at the beginning of the data collection process, if the available time series is very short, then the automatic results may be unreliable. Therefore, experts could be employed to validate the quality of linguistic summaries.

Let $T_E : LS \to [0, 1]$ denote subjectively defined quality evaluation function that maps linguistic summaries to the interval [0,1].

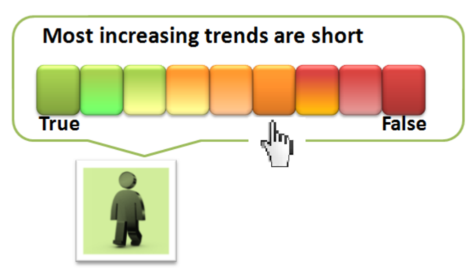

Figure 1. Evaluating quality of linguistic summaries.

As presented in Figure 1, the simple natural language expressions, e.g., *Most increasing trends are short* are presented to the decision maker who points his confidence that this summary is true about the considered time series. The values of T_E are interpreted as the expert's degree of confidence that the linguistic summary is true.

IV. FORECASTING WITH LINGUISTIC SUMMARIES

Algorithm 1 described below presents a high-level description of the proposed *Forecasting with Linguistic Summaries* (F-LS) approach.

Algorithm 1 Forecasting with Linguistic Summaries (F-LS) provides prediction y_{n+1}

Input:
y: $y = \{y_t\}_{t=1}^{n}$, $n \in \{n_{min}, ..., n_{max}\} \subseteq N$, $y \in Y$, where Y is a space of discrete time series
S: set of imprecise labels
$M = \{M_1, M_2, ..., M_J\} \subseteq M$: template probabilistic models where M is a set of stationary autoregressive processes
Output: y_{n+1}
Algorithm:

1: *Defining of imprecise concepts:*
2: build_fuzzy_numbers (S)
3: *Data preprocessing:*
4: **repeat** difference(y) **until** y *is* validated
5: min-max normalization(y)
6: *Supervised learning for the training database:*
7: **while** $i \in J$ **do**
8: $\quad T_m^s, C^s$ = generate k sample time series (M_i, k, m)
9: $\quad LI^s$ = discover_linguistic_summaries (T_m^s)
10: $\quad V^s$ = calculate_degree_of_truth (LI^s)
11: CL = supervised_learning_withSVM (C^s,V^s)
12: *Imprecise knowledge retrieval from humans:*
13: LI^E = create_provisional_linguistic_summaries (y)
14: v^E = calculate_degree_of_truth (LI^E)
15: T^E = expert_evaluation (LI^E, v^E) *EXPERT INPUT*
16: **while** $i \in J$ **do**
17: $\quad Sc^{M_i}$ = estimate_classification_scores (T^E, CL)
18: *Posterior simulation and forecasting:*
19: P=construct_prior_prob_distr (M, Sc^M)
20: y_{n+1}=MCMC_posterior_simulation (P, y)

The input for the algorithm is the discrete time series for prediction y and the set of template probabilistic models M, that need to be defined a priori. Within this research, we focus on supporting forecasting of short time series assuming that $n_{min} = 10, n_{max} = 20$.

The algorithm starts from the definition of imprecise concepts that describe the trends and linguistic summaries (Line #1). Secondly, the preprocessing of the time series data (Line #3) is performed to ensure that they are normalized and without missing values. Next, the supervised learning of the probabilistic models (Line #6) is executed. Its goal is to build the training database and to discover rules enabling the classification of the probabilistic models based on the sets of linguistic summaries describing the evolution of time series.

Then, the mining for the human-consistent prior information (Line #12) is performed. Its goal is to discover and validate with experts the linguistic summaries about the expected evolution of the predicted time series. Next, the prior probability distributions are calculated (Line #19). Finally, Markov Chain Monte Carlo Posterior Simulation is run (Line #20) to simulate the posterior probability distributions for the vector of interest and calculate the forecast y_{n+1}.

V. NUMERICAL RESULTS

The experimental study aims at showing the forecasting accuracy of the proposed *Forecasting with Linguistic Summaries* (F-LS) approach among other forecasting methods. The results are presented for the real-life benchmark time series data.

We use the subset of the 10 first yearly time series (N1-N10) that have length 20 from the M3-Competition Datasets Repository [6]. The performance of the proposed F-LS approach is compared to best 13 benchmark methods studied in [6]. These methods are briefly presented in Table I. Methods marked with * are commercially available in forecasting packages. The forecast accuracy is measured by Symmetric Mean Absolute Percentage Error (sMAPE).

Table II shows the medal classification based on sMAPE. Is is observed that the proposed F-LS forecast is number one for two series and has never performed worst. Only the Robust-Trend forecast has also been number one for two series, and number two for another two time series. Nonetheless, it has also been the worst for one series.

TABLE II. MEDAL CLASSIFICATION. TOP-3 AND THE WORST FORECASTING METHOD FOR N1-N10 TIME SERIES FROM M3-C DATASET.

	TOP-3				WORST
Method	**I**	**II**	**III**	**...**	
ForecastX	1	0	1		0
F-LS	**2**	**0**	**1**		**0**
Comb S-H-D	0	0	0		0
Robust-Trend	2	2	0		1
Theta	0	1	3		0
RBF	0	1	1		0
Auto-ANN	2	0	0		1
ForecastPro	0	0	1		1
B-J Auto	0	0	0		1
Naive2	1	1	1		0
Single	0	1	1		0
SmartFcs	1	1	0		2
ARARMA	0	1	1		1
Flores /Pearce2	0	1	0		2

Details about sMAPE for the 1 step horizon are gathered in Table III.

As demonstrated by the results in Table III, none of the benchmark methods outperforms or dominates the proposed F-LS method for all 10 time series. The best average sMAPE result of 6.1 is achieved by ForecastX method. At the same time, it is observed that for 4 (N1, N6, N7 and N9) out of all 10 time series the proposed F-LS approach delivers more accurate forecast than the ForecastX.

F-LS provides forecasts which are similarly accurate to the ones provided by Comb S-H-D, Robust-Trend, Theta and RBF methods. The average sMAPE amounts to 6.7 for F-LS, and 6.7, 6.8, 6.8, 6.9 for the other methods, respectively.

We conclude that the proposed approach delivers very competitive results in terms of the forecasting accuracy.

VI. CONCLUSION AND FUTURE WORK

In this paper, we have discussed the human-consistent input to support the analysis and forecasting of time series. We have proposed a new approach consisting of the human-computer interaction for the retrieval of the natural language summaries and their application for the Bayesian forecasting.

One of the main advantages of the proposed solution is its interpretability, which is of special importance for experts

TABLE I. SELECTED BENCHMARK METHODS FROM THE M3-COMPETITION.

Method	Author	Description
Naive2	M. Hibon	Deseasonalized Naive (Random Walk)
Robust-Trend	N. Meade	Trend model - Non-parametric version of Holt's linear model with median based estimate of trend
Flores /Pearce2	B.Flores, S. Pearce	Expert system that chooses among four methods based on the characteristics of the data
SmartFcs*	C. Smart	Expert System - conducts a forecasting tournament among four exponential smoothing and two moving average methods
Theta	V. Assimakopoulos	Decomposition technique - projection and combination of the individual components
Comb S-H-D	M. Hibon	Trend model - combining three methods: Single / Holt/ Dampen
ARARMA	N. Meade	ARIMA models - Automated Parzen's methodology with Auto regressive filter
Single	M. Hibon	Single Exponential Smoothing
ForecastX*	J. Galt	Expert System - selects from among several methods
RBF	M. Adya, S. Armstrong, F. Collopy, M. Kennedy	Rule-based forecasting: using random walk, linear regression and Holt's to estimate level and trend, involving corrections, simplification, automatic feature identification and re-calibration
ForecastPro*	R. Goodrich, E. Stellwagen	Expert System - Expert System - selects from among several methods
Auto-ANN	K. Ord, S. Balkin	Automated Artificial Neural Networks
B-J Auto	M. Hibon	ARIMA models - Box-Jenkins methodology of 'Business Forecast System'

TABLE III. SMAPE FORECASTING ACCURACY FOR N1-N10 TIME SERIES FROM THE M3-COMPETITION. F-LS IS THE PROPOSED METHOD, OTHER ARE BENCHMARK.

Method	TS-N 1	TS-N 2	TS-N 3	TS-N 4	TS-N 5	TS-N 6	TS-N 7	TS-N 8	TS-N 9	TS-N 10	Avg sMAPE
ForecastX	1.8	10.6	15.7	4.1	4.0	1.7	5.0	0.7	16.9	0.7	6.1
F-LS	**0.2**	**10.9**	**18.6**	**7.5**	**6.3**	**0.7**	**1.6**	**10.4**	**9.2**	**1.0**	**6.7**
Comb S-H-D	2.0	12.8	14.8	3.0	3.2	3.5	1.6	7.4	17.3	1.5	6.7
Robust-Trend	3.3	6.1	19.9	8.7	8.4	0.2	5.3	4.2	11.5	0.1	6.8
Theta	0.6	7.1	21.4	4.5	2.5	4.6	0.7	13.3	12.8	0.6	6.8
RBF	3.1	12.0	17.2	8.5	2.3	0.8	3.0	8.5	12.2	1.3	6.9
Auto-ANN	1.4	8.7	5.1	11.9	5.3	9.7	0.3	6.1	20.3	3.3	7.2
ForecastPro	2.0	12.6	13.9	0.5	4.0	1.7	4.5	14.3	20.1	0.8	7.4
B-J Auto	2.0	12.6	18.2	0.5	5.0	2.1	0.7	6.1	22.3	5.1	7.4
Naive2	8.6	12.5	13.8	0.5	4.4	5.8	0.7	6.1	20.1	5.0	7.7
Single	8.6	12.5	13.8	0.5	4.4	5.8	0.7	6.1	20.1	5.0	7.7
SmartFcs	2.3	1.3	24.4	9.3	5.0	1.4	5.0	1.1	30.8	4.0	8.5
ARARMA	3.2	11.1	14.7	4.7	4.1	0.4	24.3	3.5	17.9	3.1	8.7
Flores /Pearce2	10.3	11.1	13.8	20.5	3.1	5.8	1.2	9.7	16.0	1.3	9.3

involved in the forecasting process. Instead of providing definitions of prior probability distributions, users are asked to evaluate linguistic summaries that are intuitive and easy for interpretation.

The performance of the proposed approach is illustrated with the experimental study for benchmark datasets. The numerical results of the forecast accuracy show that the proposed approach of combining human input about linguistic summaries and Box-Jenkins models through the Bayesian averaging may lead to the increase of the accuracy compared to the competitive methods. Although the human input is highly subjective, it helps to eliminate the need to express the assumptions as prior probability distributions, which may be difficult to understand for non-mathematician decision makers.

Further experiments on other benchmark datasets are planned to analyze all advantages and disadvantages of the proposed approach. Future research also assumes the analysis of other forms of imprecise information like fuzzy classification rules and frequent temporal patterns, and the modeling of multiple imprecise labels interpretations.

ACKNOWLEDGEMENTS

The authors thank Janusz Kacprzyk and Anna Wilbik for their support and advice concerning the Trend Analysis. Katarzyna Kaczmarek is supported by the Foundation for Polish Science under International PhD Projects in Intelligent Computing financed from the European Union within the Innovative Economy Operational Programme 2007-2013 and European Regional Development Fund.

REFERENCES

[1] J. D. Gooijer and R. J. Hyndman, "25 years of time series forecasting," International Journal of Forecasting, vol. 22, 2006, pp. 443–473.

[2] G. Box, G. Jenkins, and G. Reinsel, Time Series Analysis: Forecasting and Control, 4th Edition. Wiley, 2008.

[3] P. D'Urso, D. Lallo, and E. Maharaj, "Autoregressive model-based fuzzy clustering and its application for detecting information redundancy in air pollution monitoring networks," Soft Computing, 2013, pp. 83–131.

[4] O. Hryniewicz and K. Kaczmarek, "Bayesian analysis of time series using granular computing approach," Applied Soft Computing, 2014.

[5] J. Geweke, "Contemporary bayesian econometrics and statistics," Wiley series in probability and statistics, 2005.

[6] S. Makridakis and M. Hibon, "The m3-competition: results, conclusions and implications," International Journal of Forecasting, 2000, pp. 451–476.

[7] G. Box and G. Jenkins, Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco, 1970.

[8] R. Clemen and R. Winkler, "Combining probability distributions from experts in risk analysis," Risk Analysis, vol. 19(2), 1999, pp. 187–203.

[9] E. Ley and M. Steel, "On the effect of prior assumptions in bayesian model averaging with applications to growth regression." Journal of Applied Econometrics, vol. 24, 2009, pp. 651–674.

[10] D. Nauck and R. Kruse, "Obtaining interpretable fuzzy classification rules from medical data," Artificial Intelligence in Medicine, vol. 16(2), 2014, pp. 149–169.

[11] S. Kempe, J. Hipp, C. Lanquillon, and R. Kruse, "Mining frequent temporal patterns in interval sequences," Fuzziness and Knowledge-Based Systems in International Journal of Uncertainty, vol. 16 (5), 2008, pp. 645–661.

[12] J. Yao, A. Vasilakos, and W. Pedrycz, "Granular computing: Perspectives and challenges," IEEE Transactions on Cybernetics, vol. 43(6), 2013, pp. 1977–1989.

[13] J. Kacprzyk, "Linguistic summarization of time series using a fuzzy

quantifier driven aggregation," Fuzzy Sets Syst, vol. 159 (12), 2008, pp. 1485–1499.

[14] C. Moewes and R. Kruse, "Zuordnen von linguistischen ausdrücken zu motiven in zeitreihen (matching of labeled terms to time series motifs)," Automatisierungstechnik, 2009, pp. 146–154.

[15] K. Kaczmarek and O. Hryniewicz, "Linguistic knowledge about temporal data in bayesian linear regression model to support forecasting of time series," in Proc. of Federated Conference on Computer Science and Information Systems, 2013, pp. 655 – 658.

[16] R. Yager, "A new approach to the summarization of data," Information Science, vol. 28 (1), 1982, pp. 69–86.

[17] L. Zadeh, "Fuzzy sets," Information and Control, 1965, pp. 338–353.

[18] M. Gil and O. Hryniewicz, "Statistics with imprecise data," Encyclopedia of Complexity and Systems Science, 2009, pp. 8679–8690.

[19] R. Kruse, C. Borgelt, F. Klawonn, C. Moewes, M. Steinbrecher, and P. Held, Computational Intelligence. Texts in Computer Science. Springer London, 2013, ch. Fuzzy Sets and Fuzzy Logic.

[20] L. A. Zadeh, "A computational approach to fuzzy quantifiers in natural languages," Computers and Maths with Applications, 1983, pp. 149–184.

[21] J. Kacprzyk, A. Wilbik, A. Partyka, and A. Ziółkowski, Trend Analysis System. Systems Research Institute, Polish Academy of Sciences, Warsaw, 2011.