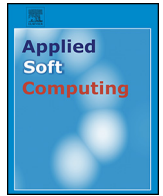




Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc



Bayesian analysis of time series using granular computing approach

Olgierd Hryniewicz, Katarzyna Kaczmarek*

Systems Research Institute, Polish Academy of Sciences, Newelska 6, 01-447 Warsaw, Poland

ARTICLE INFO

Article history:

Received 8 January 2014
Received in revised form
19 November 2014
Accepted 22 November 2014
Available online xxx

Keywords:

Time series forecasting
Granular computing
Soft computing
Data mining
Bayesian methods
Linguistic summaries

ABSTRACT

The soft computing methods, especially data mining, usually enable to describe large datasets in a human-consistent way with the use of some generic and conceptually meaningful information entities like information granules. However, such information granules may be applied not only for the descriptive purposes, but also for prediction. We review the main developments and challenges of the application of the soft computing methods in the time series analysis and forecasting, and we provide a conceptual framework for the Bayesian time series forecasting using the granular computing approach. Within the proposed approach, the information granules are successfully incorporated into the Bayesian posterior simulation process. The approach is evaluated with a set of experiments on the artificial and benchmark real-life time series datasets.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

The Bayesian methods for the time series analysis are proven successful in many practical applications, e.g., [1–3]. Nonetheless, the Bayesian time series analysis and probability theory are not meant to process directly the information described in natural language, and according to Zadeh [4], people granulate information and operate mainly on words and propositions easy to express in natural language. Hopefully, the soft computing methods, especially data mining, enable to retrieve and process the conceptually meaningful human-consistent information from large datasets. Such information entities have given rise to the general framework of Granular Computing [5,6].

Although, time series data mining with the use of the soft computing tools has gained a lot of attention in the literature, the interdisciplinary combination of soft computing and Bayesian approach does not seem to be extensively investigated. As presented in Table 1, according to the keyword and abstract search, there are over 2.7k articles on ‘Bayesian’ AND ‘time series’, and over 2.3k results about ‘granular computing’. However, only 1 result combining these keywords. Furthermore, the Scopus search engine returns only 6 positions for the keywords ‘soft computing’ AND ‘time series’ AND ‘Bayesian’.

The objective of this paper is to review the related work on the application of the soft computing methods for the time series analysis, and to propose the conceptual *Bayesian Granular Computing (B-GC) framework for Time Series Forecasting*. The goal is to incorporate the human-consistent soft computing methods, especially data mining and classification techniques, in the construction of the prior model probability distributions.

The proposed approach assumes employing techniques from the following research fields: the fuzzy sets theory, the knowledge discovery from sequential data, especially the time series abstraction and mining for linguistic summaries, Support Vector Machines for the classification problem and the Markov Chain Monte Carlo methods for the posterior simulation and the Bayesian inference. The approach is in line with the Generalized Theory of Uncertainty [7].

We provide an application example as a demonstration of concept and the numerical results of forecasting real-life data. It is observed that the granular computing approach may successfully support the probabilistic time series analysis and forecasting. In the opinion of the authors of this paper, the intelligent combination of the granular computing methods and the Bayesian modeling is a promising direction for the future research on the time series analysis and forecasting.

The structure of this paper is as follows. Next section summarizes state-of-the-art of the soft computing techniques supporting the time series analysis and forecasting. Section 3 recalls basic definitions related to the time series analysis and granular computing. Section 4 presents the proposed *Bayesian Granular Computing*

* Corresponding author. Tel.: +48 603692082.
E-mail address: K.Kaczmarek@ibspan.waw.pl (K. Kaczmarek).

Table 1
Results from document search according to keywords, title and abstract in the Scopus database of research (<http://www.scopus.com>).

Search field by article title, keywords, abstract	Results
'granular computing' AND 'Bayesian' AND 'time series'	1
'soft computing' AND 'Bayesian' AND 'time series'	6
'granular computing' AND 'time series'	40
'soft computing' AND 'time series'	239
'Bayesian' AND 'time series'	2704
'granular computing'	2303
'soft computing'	9633
'time series'	127,933

(B-GC) method for Time Series Forecasting. The description of the experiments is gathered in Section 5. This paper concludes with discussion in Section 6.

2. State-of-the-art

Within this review, the focus lays on the approaches processing information that can be easily interpreted in natural language. First, we review the main developments and objectives of the time series data mining. Secondly, methods for the human-consistent time series forecasting are reviewed. Finally, we shortly discuss the Bayesian approach to the time series analysis.

2.1. Time series data mining

The Data Mining and Knowledge Discovery adapt the soft computing techniques to provide the human-consistent description of large datasets, that can be easily understood by users. Following Zadeh [8], the Soft Computing is based on the Computational Intelligence, and together with the Hard Computing (that is based on the Artificial Intelligence) form the Machine Intelligence. The field of Computational Intelligence was formally initiated in 1994 during the IEEE World Congress on Computational Intelligence in Orlando. Computational Intelligence (CI) is defined as *a methodology involving computing that exhibits an ability to learn and/or deal with new situations such that the system is perceived to possess one or more attributes of reason, such as generalization, discovery, association, and abstraction* cf. [9].

Data mining usually leads to some generic and conceptually meaningful information entities like **information granules**. The theory of information granulation [5] is inspired by the ways in which people granulate information and reason about it. The general framework of Granular Computing [6,10–12] studies the information granulation and the information granules. It endows the information processing with a facet of human-centricity.

As defined by Zadeh [5], the information granule is a clump of points (objects) grouped together by indistinguishability, similarity, proximity or functionality. The information granules are formally described with the use of fuzzy sets, intervals, rough sets, shadowed sets, probabilistic sets, etc. Pedrycz and contributors [13,14] study the optimal granular representation of time series. In [15], the authors review the granular computing developments in general. In [16], the fast interval predictors for large-scale, nonlinear time series with noisy data using fuzzy granular support vector machines are presented.

The knowledge discovery process from sequential data may be divided into segmentation, clustering, classification of identified meaningful intervals or patterns, detecting anomalies, frequent patterns and discovery of association rules. For the review of the time series data mining methods, see e.g., Fu [17]. The comparative analysis of the time series representations and the similarity measures is provided in [18]. Batyrshin et al. [19] introduce an interesting perception-based approach to time series data mining.

Linguistic summaries in the sense of Yager [20] are an example of information granules. Linguistic summaries describe general facts about evolution of time series with (quasi) natural language e.g., *Among all increasing segments, majority are long*, and are intuitive and easily interpretable for people. Kacprzyk et al. [21–23] employ the classic calculus of linguistically quantified propositions for the development of the linguistic summaries.

Another example of the information granules are **linguistic descriptions** based on the Computational Theory of Perceptions [4] and the Systemic Functional Linguistics by Halliday. The linguistic descriptions [24–28] are characterized by more complex semantic and lexico-grammar structures than linguistic summaries, e.g., *Before the knee lesion, the gait quality is high because the gait symmetry is medium and the gait homogeneity is high* cf. [26]. In [29], the authors further develop the concept of the Granular Linguistic Model of a Phenomenon and provide basic architecture of a computational system to generate the linguistic descriptions.

Apart from the linguistic summaries and descriptions, another example of imprecise information entities mined from the time series datasets are **temporal patterns**. Höppner et al. [30] propose the approach for the identification of frequent patterns based on the Allen's temporal logic [31]. In [32], the authors focus on enhancing patterns with a context information and operating on block constraints instead of Allen's relations e.g., *If A happens before B and in the meantime we do not observe C, then we have a failure of class X*. Schockaert et al. [34] fuzzyfy Allen's temporal interval relations and propose a complete framework to represent, calculate and reason about the temporal relationships for fuzzy intervals.

Temporal patterns are used to build the **association rules**, that may be exemplified by *If exchange rate for currency X decreases slowly, the number of international travels will increase rapidly*. In [33], Agrawal et al. propose an interesting algorithm for discovery of rules in sets of items. Methods for the frequent patterns and association rules recognition have been successfully applied to describe huge datasets in different contexts, e.g., [35–39].

The primer goal of the data mining in most of the cited papers is the interpretation of huge datasets by generation of meaningful information granules. The retrieval of such information helps to increase the understanding of the dataset and may reduce the storage space. In the opinion of the authors of this paper, the generic methodology on how to benefit in statistical forecasting from the results of the time series data mining shall be further investigated in the literature.

2.2. Time series forecasting

Traditional autoregressive and moving average (ARMA) processes [40] are the most popular simple probabilistic models for forecasting, and though, very successful in applications. In the recent application [41], D'urso et al. apply the autoregressive estimates of models for the time series classification and clustering. One of the attractive features of the Box–Jenkins approach to forecasting is that the class of its processes is rich, and it is usually possible to find a process or a combination of processes, which provide an adequate description to the dataset. Nonetheless, the identification of the mathematical model and the specification of its required parameters may be time consuming and difficult to interpret for practitioners (experts of the field) involved in the forecasting process.

To alleviate this problem and increase the overall understanding of the probabilistic time series analysis, there have been proposed some **hybrid systems** supporting the model selection like e.g., [42,43]. In [44], the autoregressive and moving average models are combined with the fuzzy rule based approach using Mamdani inference mixed with some techniques of counting in fuzzy sets. Chen et al. [45] propose an approach for the automatic generation

of the fuzzy rule base. In [46], the imprecise linguistic information is incorporated into the forecasting process by means of the linear regression. In [47], the authors propose a hybridization of intelligent techniques such as ANNs, fuzzy systems and evolutionary algorithms and the use of fuzzy rules to identify the ARMA model for prediction.

Alternatively, to reflect the imprecision about the dataset and the model itself, there have been studied fuzzy forecasting models. Tanaka [48] suggests the fuzzy regression. Tseng et al. [49,50] propose to use the fuzzy triangular numbers instead of crisp parameters in the ARIMA models. Helin et al. [51] suggest employing the fuzzy conditional distributions in the GARCH model. Song and Chissom introduce fuzzy time series [52], that are based on the concepts of the linguistic variables and fuzzy relations. The approach assumes the split of the universum into intervals, defining the linguistic variables and relations, fuzzification of the observations, forecasting and then, defuzzification of the outputs. Further modifications to the Song and Chissom's approach are proposed in e.g., [53–56]. Chen et al. [57] introduce the fuzzified variation and the fuzzy logical relationship groups. The fuzzy time series are linked to the linguistic concepts, and therefore, are usually easy for interpretation. However, fuzzy time series scheme of representation of the fuzzy logical relationships may not be relevant for the complex forms of the information granules like the linguistic summaries or linguistic descriptions.

Hopefully, the intelligent combination of the soft computing methods with the statistical time series analysis is gaining attention in the literature. Other examples of the applications of the soft computing methods for the time series forecasting include e.g., the Fuzzy Transform by Perfilieva [58,59], the neural networks approaches [60–62], fuzzy cognitive maps [63], rough sets [64], genetic algorithms and support vector regression [65,66] or fuzzy clustering and fuzzy rule interpolation techniques [67].

Within this paper, we introduce an innovative approach to identify the a priori probability distributions for the Bayesian forecasting with the use of the information granules mined from the time series datasets.

2.3. Bayesian approach to time series analysis

The practitioners are very often posed to the dilemma of choice between the wealth of mathematic models for forecasting for the imprecise real-world data. Clemen [68] review the aggregation methods used for the predictive purposes concluding that the Bayesian approach in the risk analysis is a very powerful tool. One of the main advantages of the Bayesian approach is the possibility to describe the data imprecision, also about the model selection, in terms of the a priori probability distributions. However, the proper selection of a priori distributions is essential for the overall performance.

Geweke [1] describes the Bayesian inference for the time series and provides a detailed review of the Bayesian forecasting elements with numerous references. Prior distributions are usually assumed subjectively. Doan, Litterman and Sims [69] introduce a method to calculate the priors based on the simple statistics. The Bayesian approach for inconsistent information is discussed also by Beer and contributors in [70,71].

3. Basic definitions

In this section we recall basic definitions related to both, the Bayesian time series analysis and the granular computing.

3.1. Bayesian time series analysis

The purpose of the time series analysis is to describe the model (stochastic process) that produced the time series.

Discrete time series [40]

Let $O = \{o_1, o_2, \dots, o_q\}$ denote a finite set of objects in a considered domain. The properties of objects are measured by observables $P = \{p_1, p_2, \dots, p_r\}$. Discrete time series $\mathbf{y} = \{y_t\}_{t=1}^n \in \Psi_n$ is a sequence of observations of given object's property (o, p) such that $o \in O$ and $p \in P$ measured at successive $t \in T = \{1, \dots, n\}$ moments and at uniform time intervals. For each $t \in T$ the observation y_t is a realization of the random variable Y_t defined on the probability space (Ω, A, P) . A sequence of Y_t formulates a stochastic process.

The main principles for the Bayesian analysis are to express all assumptions using probability statements, and then, to design the distributions for the future events conditional on the observed values and a certain loss function. The reflect the uncertainty about identification of the model (process) that generated the time series, we average multiple models.

Bayesian averaging [1]

The posterior density of the vector of interest ω conditional on multiple competitive probabilistic models for forecasting $M = \{M_1, M_2, \dots, M_J\}$ is defined as follows

$$p(\omega|y, M) = \sum_{j=1}^J p(M_j|y, M)p(\omega|y, M_j) \quad (1)$$

This posterior density $p(\omega|y, M)$ is a weighted average of the posterior densities of models $\{M_1, M_2, \dots, M_J\}$ which are defined as follows:

$$p(M_j|y, M) = \frac{p(M_j)p(y|M_j)}{p(y|M)} = \frac{p(M_j)p(y|M_j)}{\sum_{j=1}^J p(M_j)p(y|M_j)} \quad (2)$$

We propose the approach to construct the prior model distributions $p(M_i)$, $i \in \{1, \dots, J\}$ with the support of the intelligent techniques.

3.2. Granular computing

We now recall definitions related to the selected granular computing aspects.

Information granule [5]

Information granule is a clump of points (objects) grouped together by indistinguishability, similarity, proximity or functionality.

A simple example of an information granule is a labeled interval. Labeled intervals are used to build e.g., linguistic summaries.

Linguistic summary [20]

Let $A = \{a_1, a_2, \dots, a_u\}$ denote a finite set of attributes (eg. dynamics, duration, variability), $S = \{s_1, s_2, \dots, s_l\}$ denotes a finite set of imprecise labels for attributes (eg. low, increasing, short). The protoform-based linguistic summary

$$Q \ R \ y\text{'s are } P \quad (3)$$

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

For further details see also [21,22].

Another constructs created from the labeled intervals are temporal patterns. The relations between labeled intervals in temporal patterns may be expressed according to Allen's interval temporal logic [31]. For any 2 intervals, at least one of the available 13 relationships, e.g., before, meets, during, overlaps is defined. The set of all possible relationships is denoted as I .

Temporal pattern [30]

A pair $P := (s, R)$ such that $s : \{1, \dots, n\} \in S$ and $R \in I^{n \times n}$ is called a temporal pattern if there is a labeled interval sequence $(A_1, s_1), (A_2, s_2), \dots, (A_n, s_n)$ such that for every $i \in \{1, \dots, n\} s(i) = s_i$ and $R[i, j] = ir(A_i, A_j)$

Frequent temporal patterns are building elements for the association rules.

Association rule [33]

An association rule is an implication of the form $P \Rightarrow C$ such that P, C are sets of items representing premise and conclusion, respectively.

There have been proposed in the literature different measures to evaluate the quality of the information granules depending on their type and the considered evaluation purposes. Table 2 presents exemplary definitions of quality measures.

4. Bayesian approach meets granular computing

The information granules like linguistic summaries, frequent patterns or association rules are usually in line with the visual pattern recognition and perception capabilities of humans, and therefore, seem appealing to support not only the interpretation of time series, but also the time series forecasting. From the range of the statistical methods, we adapt the Bayesian approach, which has been proven successful in many real-life applications. Multiple models are included in the forecasting process through the Bayesian averaging and the focus of the approach is to establish the prior probabilities for the competitive models.

The input data for the proposed Bayesian Granular Computing (B-GC) method for forecasting are multivariate time series $Y_n^k = \{\{y_t^1\}_{t=1}^n, \dots, \{y_t^k\}_{t=1}^n\}$ being a k -vector of discrete time series. We assume that the preprocessing, that is the normalization of the time series and handling of the missing data, is completed beforehand.

The output data are both, forecasts $Y_n^k = \{\{y_t^1\}_{t=n+1}^{n+h}, \dots, \{y_t^k\}_{t=n+1}^{n+h}\}$ and the descriptions of the imprecise information used in the forecast generation procedure.

The overview of the proposed Bayesian Granular Computing (B-GC) method for forecasting is presented in Fig. 1. The procedure starts with the definition of the linguistic concepts, then mining for the information granules, identification of the probabilistic model, and finally, the Bayesian posterior simulation with the prediction generation procedure.

We show that the imprecise information processed with the selected soft computing methods is supportive for the construction of the a priori model probability distributions required for the Bayesian forecasting.

4.1. Defining of concepts

We start from defining the linguistic variables related to the imprecise information. Then, we define the selected template probabilistic predictive models used to build the training database.

Elements of the algorithm for the definition phase are as follows:

- 1 Define linguistic variables for imprecise labels
- 2 Define probabilistic models and build training dataset

4.1.1. Define linguistic variables for imprecise labels

The first step of the definition phase requires providing the specification of trend attributes A and the fuzzy sets for the interpretation of imprecise labels S . Membership functions for all the linguistic variables are established. We assume that the interpretation of linguistic variable for a given property $p \in P$ of an object $o \in O$ in a given domain is known and does not change in time.

4.1.2. Define probabilistic models and build training dataset

The second step requires defining the template probabilistic predictive models $M = \{M_1, M_2, \dots, M_J\}$. Then, the training time series dataset is built based on models from M . The definitions of the a priori distributions for all parameters related to the models are saved as $M^\theta = \{M_1^\theta, M_2^\theta, \dots, M_J^\theta\}$. $T_m^s = \{\{y_t^1\}_{t=1}^m, \dots, \{y_t^s\}_{t=1}^m\}$ is a resulting s -vector of the sample time series generated from the template models M , and $C_T^J = \{c_1, \dots, c_s\}$ is the respective s -vector denoting the template models that produced the time series from T_m^s .

4.2. Time series data mining

The data mining phase aims at processing of the crisp time series datasets into meaningful information objects. Elements of the algorithm are as follows:

- 1 Mine for the information granules
- 2 Discover linguistic information
- 3 Calculate the validity matrix

4.2.1. Mine for the linguistic information granules

Data mining is performed successively on the training dataset T_m^s and then, for the time series to be predicted Y_n^k to find meaningful information granules. As a result, e.g., labeled intervals are mined from the time series datasets.

4.2.2. Discover linguistic information

Secondly, human-consistent and more complex information granules are constructed like linguistic summaries, descriptions, frequent patterns or association rules. The selection of the considered types of information granules depends on the application context. As a result, the $k \times l$ matrix LI_Y of the information entities discovered for Y_n^k time series is created and the $s \times l$ matrix LI_T for T_m^s , respectively. The linguistic information may also be formulated with the support of experts of the field thanks to the human-computer interaction.

4.2.3. Calculate validity matrix

Finally, the set of the considered quality measures is defined $QM = \{q_1, \dots, q_w\}$ and for each instance of the information granule in LI_Y and LI_T , their quality is measured. The applied quality measures shall be chosen adequately depending on the considered types of linguistic information. The results form the validity matrices V_Y and V_T of size $k \times z$ and $s \times z$, respectively, where $1 \leq z \leq \text{card}(QM)$.

4.3. Probabilistic model identification with soft computing

The prior model distributions $p(M_i), i \in \{1, \dots, J\}$ are established with the use of the classification methods. The classification problem is defined and solved for the vectors of the information granules.

The algorithm for the probabilistic model identification phase consists of the following elements:

- 1 Classify the probabilistic models
- 2 Identify prior distribution of the unobservables

4.3.1. Classify the probabilistic models

The rows of the validity matrices V_T and V_Y are feature vectors in the classification tasks. The classifier is trained on V_T , and then applied for V_Y to identify the probabilistic models.

First, we assign categories (classes) related to T_m^k that correspond to the template probabilistic models M indicated by C_T^J . Support Vector Machines (SVM) [72] are adapted to analyze the training

Table 2
Exemplary quality measures for the entities of information granules.

Linguistic information	Quality measure
Linguistic summary	<p>- the degree of truth [20] $T = \mu_Q \left(\frac{\sum_{i=1}^n (\mu_R(y_n) \wedge \mu_P(y_n))}{\sum_{i=1}^n \mu_R(y_n)} \right)$ (4) where $\mu_R(y_n), \mu_P(y_n)$ are the membership functions $\mu_R, \mu_P : \mathfrak{X} \rightarrow [0, 1]$ determining the degree to which R, P, respectively, are satisfied for the time series y at the given moment n.</p> <p>- the degree of imprecision (fuzziness) [21] $T_2 = 1 - \sqrt[m]{\prod_{j=1, \dots, m} \frac{card\{x \in X_j : \mu_{s_j}(x) > 0\}}{card X_j}}$ (5) where summarizer P is given as a family of fuzzy sets $P = \{s_1, s_2, \dots, s_m\}$ and $card$ denotes the cardinality of the corresponding (non-fuzzy) set.</p>
Temporal pattern	<p>- the support of temporal pattern [30] $supp$ - the total time in which the instances of P can be observed in the sliding window.</p> <p>- the frequency of occurrences of temporal pattern $freq$</p>
Association rule	<p>- confidence $conf = \frac{supp_P}{supp_C}$ (6) where $supp$ represents the support of the temporal pattern.</p> <p>- j-measure compares a priori distribution of X with a posteriori distribution of X given $Y=y$, and is defined by the following sum: $\sum_{z \in \{x, \bar{x}\}} Pr(X = z Y = y) \log_2 \frac{Pr(X = z Y = y)}{Pr(X = z)}$ (7) where 'IF $Y=y$ THEN $X=x$' is the rule, X, Y denote random variables, X takes values $\{x, \bar{x}\}$ representing two events: the rule is true or false.</p>

data (V_T, C_T^j) and learn the classifier. Cross validation is performed to estimate the classification accuracy.

4.3.2. Identify prior distribution of the unobservables

Having trained the SVM classifier, we apply it to classify the sets of the information granules describing the time series Y_n^k . The rows of the validity matrix V_Y are the feature vectors. The classification assumes that a score is assigned to each of J possible categories (models).

$$Score(V_Y(i), j) = B_j V_Y(i) \tag{8}$$

where $V_Y(i)$ is the feature vector being i -th row of V_Y , B_j is the vector of weights corresponding to category j , and $Score(V_Y(i), j)$ is the score associated with assigning instance i to the category (model) j . The resulting scores for all time series to be predicted form the k -vector $C_Y^j = \{Score(V_Y(1))^j, \dots, Score(V_Y(k))^j\}$. The prior model distributions $p(M_j)$ are estimated with the scores from C_Y^j . Finally, the a priori distributions for the parameters $p(\theta_{M_j} | M_j)$ for $M_j \in M$ are adapted from M^θ .

4.4. Bayesian prediction

The last phase of the proposed approach assumes prediction basing on the identified prior distributions for the probabilistic models with the Bayesian inference. The elements of the algorithm for the prediction phase are as follows:

- 1 Simulate posterior distributions
- 2 Generate forecast and describe its components

4.4.1. Simulate posterior distribution of the unobservables and the vector of interest

In line with [1], the posterior density of the vector of interest

$$p(\omega | y^o, M) = \sum_{j=1}^J p(M_j | y, M) p(\omega | y^o, M_j) \tag{9}$$

$$p(\omega | y^o, M_j) = \int_{\Theta_{M_j}} p(\omega | y^o, \theta_{M_j}, M_j) p(\theta_{M_j} | y^o, M_j) dv(\theta_{M_j}) \tag{10}$$

and the posterior density of the unobservables θ_{M_j}

$$p(\theta_{M_j} | y^o, M_j) = \frac{p(\theta_{M_j} | M_j) p(y^o | \theta_{M_j}, M_j)}{p(y^o | M_j)} \tag{11}$$

in which $\theta_{M_j} \in \Theta_{M_j}$ is a $k_{M_j} \times 1$ vector of unobservables. We adapt the Markov Chain Monte Carlo (MCMC) posterior simulator that yields pseudo-random sequence of the vector of interest to estimate the posterior moments.

4.4.2. Generate forecast and describe its components

The estimates from the formula (9) are used to generate point forecasts and their errors according to the model definitions M . Once the prediction is calculated, additional information that supported the forecasting process is provided. This information consists of the objects with the information granules and the quality measures, and therefore, shall be easily understood and helpful for the experts of the field responsible of decision making based on the forecasting results.

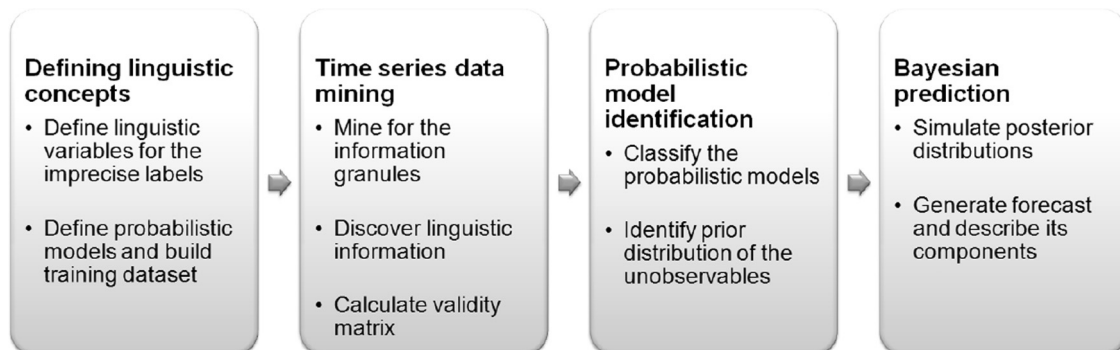


Fig. 1. Overview of the proposed Bayesian Granular Computing (B-GC) method for forecasting.

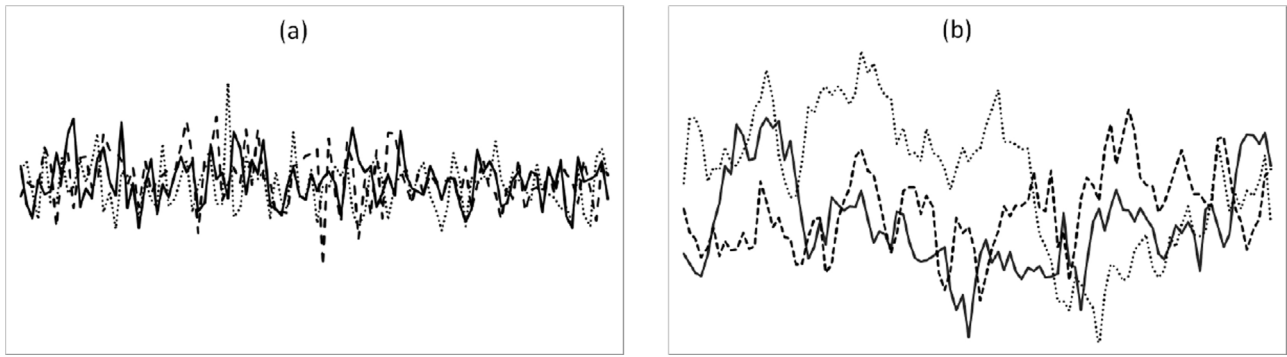


Fig. 2. Sample time series generated from the autoregressive process (a) M_1 (b) M_{10} .

5. Experimental results

The performance of the proposed *Bayesian Granular Computing (B-GC) method for Time Series Forecasting* is illustrated with the experiments on the artificial and real-life time series datasets. The following forecast accuracy measures are adapted to evaluate the performance:

- Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{1}{h} \sum_{i=1}^h \left| \frac{100e_{t+i}}{X_{t+i}} \right| \quad (12)$$

- Symmetric Mean Absolute Percentage Error (sMAPE)

$$sMAPE = \frac{1}{h} \sum_{i=1}^h \left| \frac{200|e_{t+i}|}{(X_{t+i} + \hat{F}_{t+i})} \right| \quad (13)$$

The goal of the experiments is to provide an illustrative demonstration of the proposed approach and to evaluate its forecasting accuracy.

The program has been implemented in Python with the support of NumPy, SciPy and the PyMc extension modules. Linguistic summaries are generated with the Trend Analysis System [22].

5.1. Artificial datasets

An illustrative demonstration of the proposed approach is provided for the autoregressive datasets. This simulation study enables to investigate in detail the theoretical aspects, like importance of the prior knowledge, of the proposed approach.

Time series datasets are created from the autoregressive processes $M = \{M_1, M_2, \dots, M_{10}\}$ according to the following formula:

$\hat{y}_t = \sum_{i=1}^p \phi_i \hat{y}_{t-i} + a_t$, $a_t \sim N(0, \sigma^2)$, $\hat{y}_t = y_t - \mu$, $\phi_1 \in [0, 1]$, $\sigma^2 = 0.1$, and the autoregressive coefficient $\phi_1 \subseteq [0.0, 0.1]$ for M_1 , $\phi_1 \subseteq [0.1, 0.2]$ for M_2, \dots , $\phi_1 \subseteq [0.9, 1]$ for M_{10} . As a result, 80 time series are created with 100 observation in each as training dataset

T_{80}^{100} , and another 20 time series are generated randomly as dataset for prediction (testing dataset) Y_{90}^{20} .

Fig. 2 shows the exemplary time series generated from the 2 competitive probabilistic models from the testing dataset Y_{90}^{20} . As observed by visual inspection, the characteristics of trends presented in (a) and (b) are different.

The set of attributes A consists of *length, dynamics of change, variability*, and the imprecise labels for the attributes S are as follows: *short, medium, long, increasing, constant, decreasing, low, moderate, high*. The information granules considered in this experiment are the extended linguistic summaries, that may be exemplified by *Most increasing trends are short*. We apply the quantifier $Q = \{most\}$. The resulting sets of linguistic summaries Ll_k and Ll_s consist of linguistic summaries for each time series. The degree of true $QM = \{T\}$ is applied to evaluate the quality of the considered information granules and to calculate the validity matrices V_k, V_s .

The crisp time series datasets are segmented, then summarization is performed. We group the sets of the information granules by the template models that produced the time series. It is observed that some types of linguistic summaries may be good discriminant for the model classification analysis. Table 3 presents average degree of truth T within these groups. For example, the degree of truth for the summary *Among all trends, most are short* amounts to 1 for time series generated from M_1 and to 0.49 for time series generated from M_{10} . At the same time, there exist linguistic summaries, e.g., *Among all long trends, most are constant* that do not bring value to the classification in this context because the degree of truth is 1 for any model.

The next step of the proposed approach is the classification of the time series by probabilistic models based on the sets of the linguistic granules. For the clarity reasons, we group the template probabilistic models into 3 classes as follows: *AR with weak positive autocorrelation* $C_1 = \{M_1, M_2, M_3\}$, *AR with medium positive autocorrelation* $C_2 = \{M_4, M_5, M_6, M_7\}$ and *AR with strong positive autocorrelation* $C_3 = \{M_8, M_9, M_{10}\}$, and these classes are processed in the remaining of this experiment.

The classification scores based on the linguistic summaries for time series Y_{90}^{20} are presented in Table 4. For example, for TS1, which

Table 3
Avg degree of truth T for sample linguistic summaries in groups of time series generated from 10 models: $M_1, M_2, \dots, M_9, M_{10}$ with autoregressive coefficient $\phi_1 = 0.0, \phi_1 = 0.1, \dots, \phi_1 = 0.9$ respectively.

Linguistic summary description	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}
Among all trends, most are short	1.00	1.00	0.95	0.97	0.83	0.71	0.62	0.69	0.57	0.49
Among all trends, most are low	0.47	0.61	0.45	0.43	0.26	0.22	0.21	0.18	0.21	0.14
Among all trends, most are moderate	0.47	0.35	0.53	0.51	0.74	0.81	0.77	0.77	0.88	0.86
Among all trends, most are medium	0.51	0.51	0.64	0.65	0.76	0.88	0.90	0.90	0.95	0.98
Among all decreasing trends, most are moderate	0.34	0.05	0.25	0.39	0.37	0.44	0.45	0.64	0.71	0.71
Among all decreasing trends, most are medium	0.17	0.14	0.24	0.28	0.21	0.46	0.41	0.54	0.58	0.69
Among all long trends, most are constant	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 4

The SVM classification scores based on the sets of linguistic summaries for the time series Y_{90}^{20} (dataset for prediction) in the 3-class problem.

TS	1	2	3	4	5	6	7	8	9	10
True class	C_1	C_1	C_1	C_1	C_1	C_1	C_2	C_2	C_2	C_2
Score(C_1)	0.68	0.64	0.55	0.88	0.79	0.27	0.43	0.34	0.66	0.25
Score(C_2)	0.20	0.24	0.35	0.10	0.14	0.54	0.44	0.41	0.27	0.54
Score(C_3)	0.12	0.12	0.10	0.02	0.06	0.19	0.12	0.25	0.08	0.21
TS	11	12	13	14	15	16	17	18	19	20
True class	C_2	C_2	C_2	C_2	C_3	C_3	C_3	C_3	C_3	C_3
Score(C_1)	0.02	0.30	0.24	0.02	0.03	0.09	0.01	0.01	0.02	0.05
Score(C_2)	0.39	0.53	0.28	0.55	0.50	0.48	0.41	0.43	0.23	0.54
Score(C_3)	0.59	0.17	0.48	0.42	0.47	0.43	0.58	0.56	0.75	0.41

Table 5

Forecast accuracy by Mean Absolute Percentage Error (MAPE) for the set of the testing time series Y_{90}^{20} .

Method	MAPE
Bayesian Granular Computing (B-GC) method for forecasting	0.0528
Bayesian averaging with uniform distribution for the competitive models	0.0563
Naive method(the last observed is the first predicted)	0.3661

was generated from C_1 , the classification scores assigned to models C_1 , C_2 and C_3 are 0.68, 0.20 and 0.12, respectively. It is observed that the classifier yields very good results for the identification of C_1 , in 5/6 cases the highest score indicates the true model. The overall SVM classification accuracy (and its error) measured by 5-fold-cross validation for the training dataset Y_{100}^{80} amounts to 0.55 (0.09).

The final step of the approach is the posterior simulation and the verification of forecasting accuracy. The summary of the forecasting accuracy is demonstrated in Table 5. As demonstrated in Table 5, the forecasting error for the Bayesian Granular Computing (B-GC) method is 7 times smaller than for the naive method. More importantly, for the considered artificial dataset, the approach also outperforms the traditional Bayesian averaging with uniform prior distributions. The mean absolute percentage error for the proposed approach and the approach with uniform distributions are 0.0528 and 0.0563, respectively.

5.2. Real-life datasets

The comparative forecasting analysis is performed for the real-life time series from the M3-Competition Datasets Repository by Makridakis and Hibon [73]. We use the subset of the 5 first yearly time series that have length 47. The performance of the proposed approach is compared to the selected benchmark methods that have been claimed as one of the best according to the Makridakis' study. These methods are briefly presented in Table 6.

The results presenting the forecasting accuracy of the proposed method and the competitive ones are gathered in Table 7. As demonstrated in Table 7, none of the benchmark methods outperforms the proposed B-GC method for all time series. On average, the proposed method scored a very good sMAPE result of 3.81, which is the best among the competitive methods. At the same time, it is observed that for some time series there exist methods that deliver more accurate forecast, so there is still potential for improvement. For example, when considering the time series N156, the sMAPE amounts to 2.33 and 1.9 for the proposed B-GC and the Theta method, respectively.

We conclude that the proposed approach delivers very competitive results in terms of the forecasting accuracy. However, it should be noted, that the obtained results are limited to the autoregressive models, and therefore are of preliminary character. In the future,

Table 6

Methods from the M3-Competition that are selected for the comparative analysis of the forecasting accuracy within this research due to their very good performance in the M3-Competition. Methods marked with * are commercially available in forecasting packages. Source: Table 2, The 24 methods included in the M3-Competition classified into six categories in [73]

Method	Author	Description
Theta	V. Assimakopoulos	Decomposition model - projection and combination of the individual components
Robust-Trend	N. Meade	Trend model - Non-parametric version of Holt's linear model with median based estimate of trend
SmartFcst*	C. Smart	Expert System - conducts a forecasting tournament among four exponential smoothing and two moving average methods
ForecastPro*	R. Goodrich and E. Stellwagen	Expert System - selects from among several methods

Table 7

sMAPE forecasting accuracy for 5 benchmark real-life time series by: the proposed Bayesian Granular Computing (B-GC) method and 4 competitive - some of the best methods from the M3 Competition

	B-GC	Theta	RobustTrend	SmartFcst	ForecastPRO
N 156	2.33	1.90	6.80	6.39	17.58
N 157	2.60	2.85	1.11	14.92	17.97
N 158	3.41	7.70	8.12	2.20	3.04
N 159	5.12	31.12	31.27	31.50	19.58
N 160	5.60	2.66	6.37	9.97	10.08
Avg	3.81	9.25	10.73	13.00	13.65

they should be confirmed also on an extensive set of benchmark datasets to yield a complete empirical validation of the general approach.

6. Conclusion

In this paper we have reviewed the main challenges and the recent developments of incorporating the soft computing techniques into the statistical time series analysis and forecasting. We have introduced the *Bayesian Granular Computing (B-GC) method for Time Series Forecasting*, that aims at the identification of the a priori model distributions with the use of the information granules mined from the time series datasets with soft computing techniques.

One of the main advantages of the proposed solution is its human-centricity, being a result of including the imprecise information granules. This is especially important in practice for experts involved in the forecasting process for large datasets. The data mining and classification methods employed in the work provide useful information granules. It is shown that the considered linguistic summaries constitute intuitive and very informative knowledge,

and may successfully support the Bayesian time series forecasting process.

The performance of the approach is illustrated with the artificial and real-life datasets. The simulations confirm that the construction of the prior model probabilities with support of the selected granular computing techniques may lead to the increase of forecast accuracy compared to the uniform Bayesian averaging and the competitive benchmark methods. The numerical results prove that the proposed approach delivers very competitive results, not only in terms of its human-centricity, but also in terms of the forecasting accuracy.

Future research assumes the incorporation of other types of information granules and probabilistic models into the forecasting framework. We conclude that the intelligent combination of the granular computing achievements and the Bayesian modeling is a promising direction for the future research on the time series analysis and forecasting.

Acknowledgements

The authors thank Janusz Kacprzyk and Anna Wilbik for their advice and support 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. Geweke, Contemporary Bayesian econometrics and statistics, in: Wiley Series in Probability and Statistics, John Wiley, 2005, <http://dx.doi.org/10.1002/0471744735>.
- [2] G. Koop, Bayesian Econometrics, Wiley, 2003.
- [3] V. Petridis, A. Kehagias, L. Petrou, A. Bakirtzis, S. Kiartzis, H. Panagiotou, N. Maslaris, A Bayesian multiple models combination method for time series prediction, *J. Intell. Robot. Syst.* 31 (1–3) (2001) 69–89, doi: 10.1.1.104.9879.
- [4] L. Zadeh, From computing with numbers to computing with words – from manipulation of measurements to manipulation of perceptions Lecture Notes in Computer Science. Intelligence Systems and Soft Computing, 1804, Springer, 2000, pp. 3–40, http://dx.doi.org/10.1007/978-3-7908-1792-8_5.
- [5] L. Zadeh, Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, *Fuzzy Sets and Systems.* 90 (1997) 111–127, doi: 10.1016/S0165-0114(97)00077-8.
- [6] W. Pedrycz, Granular Computing: An Emerging Paradigm Studies in Fuzziness and Soft Computing, Physica-Verlag, Heidelberg, 2001, pp. 71.
- [7] L. Zadeh, Generalized theory of uncertainty (GTU) – principal concepts and ideas, *Comput. Stat. Data Anal.* 51 (2006) 15–46, <http://dx.doi.org/10.1016/j.csda.2006.04.029>.
- [8] A. Lotfi, Zadeh, Communications of the ACM fuzzy logic, *Neural Netw. Soft Comput.* 77–84 (1994) 37, <http://dx.doi.org/10.1145/175247.175255>.
- [9] R. Eberhart, P. Simpson, R. Dobbins, Computational Intelligence, PC Tools Academic Press, Boston, 1996.
- [10] A. Bargiela, W. Pedrycz, Human-Centric Information Processing Through Granular Modelling, Studies in Computational Intelligence, Springer, Berlin/Heidelberg, 2009, pp. 182, <http://dx.doi.org/10.1007/978-3-540-92916-1>.
- [11] W. Pedrycz, A. Gacek, Temporal granulation and its application to signal analysis, *Inform. Sci.* 143 (1–4) (2002) 47–71, [http://dx.doi.org/10.1016/S0020-0255\(02\)00179-2](http://dx.doi.org/10.1016/S0020-0255(02)00179-2).
- [12] Yiyu Yao, Granular computing for data mining, in: B.V. Dasarathy (Ed.), Proceedings of the SPIE Conference on Data Mining, Intrusion Detection, Information Assurance, and Data Networks Security, 2006.
- [13] R. Al-Hmouz, W. Pedrycz, A. Balamash, A. Morfeq, Granular representation schemes of time series: a study in an optimal allocation of information granularity, *Proc. IEEE Symp. Found. Comput. Intell.* 4 (2013) 4–51, <http://dx.doi.org/10.1109/FOCI.2013.6602454>.
- [14] F. Yu, W. Pedrycz, The design of fuzzy information granules: tradeoffs between specificity and experimental evidence, *Appl. Soft Comput. J.* 9 (1) (2009) 264–273, <http://dx.doi.org/10.1016/j.asoc.2007.10.026>.
- [15] J. Yao, A. Vasilakos, W. Pedrycz, Granular computing: perspectives and challenges, *IEEE Trans. Cybern.* 43 (6) (2013) 1977–1989, <http://dx.doi.org/10.1109/TSMCC.2012.2236648>.
- [16] J. Ruan, X. Wang, Y. Shi, Developing fast predictors for large-scale time series using fuzzy granular support vector machines, *Appl. Soft Comput. J.* 13 (9) (2013) 3981–4000, <http://dx.doi.org/10.1016/j.asoc.2012.09.005>.
- [17] T. Fu, A review on time series data mining, *Eng. Appl. Artif. Intell.* 24 (2011) 164–181, <http://dx.doi.org/10.1016/j.engappai.2010.09.007>.
- [18] X. Wang, A. Mueen, H. Ding, G. Trajcevski, P. Scheuermann, E. Keogh, Experimental comparison of representation methods and distance measures for time series data, *Data Min. Knowl. Discov.* 26 (2013) 275–309, <http://dx.doi.org/10.1007/s10618-012-0250-5>.
- [19] I. Batyrshin, L. Sheremetov, Perception-based approach to time series data mining, *Appl. Soft Comput.* 8 (2008) 1211–1221, <http://dx.doi.org/10.1016/j.asoc.2007.02.020>.
- [20] R. Yager, A new approach to the summarization of data, *Inform. Sci.* 28 (1) (1982) 69–86.
- [21] J. Kacprzyk, S. Zadrozny, Linguistic database summaries and their protoforms: towards natural language based knowledge discovery tools, *Inform. Sci.* 173 (2005) 281–304, <http://dx.doi.org/10.1016/j.ins.2005.03.002>.
- [22] J. Kacprzyk, A. Wilbik, S. Zadrozny, Linguistic summarization of time series using a fuzzy quantifier driven aggregation, *Fuzzy Sets Syst.* 159 (12) (2008) 1485–1499, <http://dx.doi.org/10.1016/j.fss.2008.01.025>.
- [23] J. Kacprzyk, S. Zadrozny, Comprehensiveness of linguistic data summaries: a crucial role of protoforms, *Stud. Comput. Intell.* 445 (2013) 207–221, http://dx.doi.org/10.1007/978-3-642-32378-2_14.
- [24] L. González-Villanueva, A. Alvarez-Alvarez, L. Ascari, G. Trivino, A tool for linguistic assessment of rehabilitation exercises, *Appl. Soft Comput. J.* 14 (Part A) (2014) 120–131, <http://dx.doi.org/10.1016/j.asoc.2013.07.010>.
- [25] D. Sanchez-Valdes, A. Alvarez-Alvarez, G. Trivino, Linguistic description about circular structures of the Mars' surface, *Appl. Soft Comput. J.* 13 (12) (2013) 4738–4749, <http://dx.doi.org/10.1016/j.asoc.2013.08.003>.
- [26] A. Alvarez-Alvarez, G. Trivino, Linguistic description of the human gait quality, *Eng. Appl. Artif. Intell.* 26 (1) (2013) 13–23, <http://dx.doi.org/10.1016/j.engappai.2012.01.022>.
- [27] L. Eciolaza, M. Pereira-Fariña, G. Trivino, Automatic linguistic reporting in driving simulation environments, *Appl. Soft Comput. J.* 13 (9) (2013) 3956–3967, <http://dx.doi.org/10.1016/j.asoc.2012.09.007>.
- [28] M. Alvarez, P. Felix, P. Cari nena, Discovering metric temporal constraint networks, *Artif. Intell. Med.* 58 (3) (2013) 139–154, <http://dx.doi.org/10.1016/j.artmed.2013.03.006>.
- [29] G. Trivino, M. Sugeno, Towards linguistic descriptions of phenomena, *Int. J. Approx. Reason.* 54 (1) (2013) 22–34, <http://dx.doi.org/10.1016/j.ijar.2012.07.004>.
- [30] F. Höppner, F. Klawonn, Finding informative rules in interval sequences, *Intell. Data Anal.* 6 (3) (2002) 237–256, doi: 10.1.1.29.3138.
- [31] J. Allen, Maintaining knowledge about temporal intervals, *Commun. ACM* 26 (11) (1983) 832–843, <http://dx.doi.org/10.1145/182.358434>.
- [32] F. Höppner, S. Peter, M. Berthold, Enriching multivariate temporal patterns with context information to support classification Computational Intelligence in Intelligent Data Analysis, Studies in Computational Intelligence, 445, 2013, pp. 195–206, http://dx.doi.org/10.1007/978-3-642-32378-2_13.
- [33] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, Fast discovery of association rules, in: U.M. Fayyad, et al. (Eds.), Advances in Knowledge Discovery and Data Mining, Menlo Park Calif.: AAAI Press, 1996, pp. 307–328.
- [34] S. Schockaert, M. De Cock, Temporal reasoning about fuzzy intervals, *Artif. Intell.* 172 (2008) 1158–1193, <http://dx.doi.org/10.1016/j.artint.2008.01.001>.
- [35] S. Peter, F. Höppner, Finding temporal patterns using constraints on (partial) absence, presence and duration, in: R. Setchi, I. Jordanov, R. Howlett, L. Jain (Eds.), in: KES (1). Lecture Notes in Computer Science, 6276, Springer, 2010, pp. 442–451, http://dx.doi.org/10.1007/978-3-642-15387-7_48.
- [36] F. Klawonn, R. Kruse, Derivation of fuzzy classification rules from multidimensional data Advances in Intelligent Data Analysis. Windsor, Ontario, 9, 1995, pp. 0–94, doi: 10.1.1.31.2479.
- [37] F. Moerchen, A. Ullsch, Efficient mining of understandable patterns from multivariate interval time series, *Data Min. Knowl. Discov.* 15 (2) (2007) 181–215, doi: 10.1.1.95.3573.
- [38] P. Dash, H. Behera, I. Lee, Time sequence data mining using time-frequency analysis and soft computing techniques, *Appl. Soft Comput. J.* 8 (1) (2008) 202–215, <http://dx.doi.org/10.1016/j.asoc.2007.01.001>.
- [39] C. Chen, T. Hong, V. Tseng, Fuzzy data mining for time-series data, *Appl. Soft Comput. J.* 12 (1) (2012) 536–542, <http://dx.doi.org/10.1016/j.asoc.2011.08.006>.
- [40] G. Box, G. Jenkins, Time Series Analysis: Forecasting and Control, Holden-Day, San Francisco, 1970.
- [41] P. D'Urso, D. Lallo, E. Maharaj, Autoregressive model-based fuzzy clustering and its application for detecting information redundancy in air pollution monitoring networks, *Soft Comput.* 17 (2013) 83–131, <http://dx.doi.org/10.1007/s00500-012-0905-6>.
- [42] C. Toro, S. Gómez Meire, J. Gálvez, F. Fdez-Riverola, A hybrid artificial intelligence model for river flow forecasting, *Appl. Soft Comput. J.* 13 (8) (2013) 3449–3458, <http://dx.doi.org/10.1016/j.asoc.2013.04.014>.
- [43] M. Khashei, M. Bijari, A novel hybridization of artificial neural networks and ARIMA models for time series forecasting, *Appl. Soft Comput. J.* 11 (2) (2011) 2664–2675, <http://dx.doi.org/10.1016/j.asoc.2010.10.015>.
- [44] D. Petrovic, Y. Xie, K. Burnham, Fuzzy decision support system for demand forecasting with a learning mechanism, *Fuzzy Sets Syst.* 157 (12) (2006) 1713–1725, <http://dx.doi.org/10.1016/j.fss.2006.03.011>.
- [45] D. Chen, J. Wang, F. Zou, H. Zhang, W. Hou, Linguistic fuzzy model identification based on PSO with different length of particles, *Appl. Soft Comput. J.* 12 (11) (2012) 3390–3400, <http://dx.doi.org/10.1016/j.asoc.2012.07.014>.
- [46] K. Kaczmarek, 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. 651–654.
- [47] O. Valenzuela, I. Rojas, F. Rojas, H. Pomares, L. Herrera, A. Guillen, L. Marquez, M. Pasadas, Hybridization of intelligent techniques and ARIMA models for time series prediction, *Fuzzy Sets Syst.* 159 (2008) 821–845, <http://dx.doi.org/10.1016/j.fss.2007.11.003>.
- [48] H. Tanaka, S. Uejima, K. Asai, Linear regression analysis with fuzzy model, *IEEE Trans. Syst. Man Cybern.* 12 (1982) 903–907, <http://dx.doi.org/10.1109/TSMC.1982.4308925>.
- [49] F. Tseng, G. Tzeng, H. Yu, B. Yuan, Fuzzy ARIMA model for forecasting the foreign exchange market, *Fuzzy Sets Syst.* 118 (2001) 9–19, [http://dx.doi.org/10.1016/S0165-0114\(98\)00286-3](http://dx.doi.org/10.1016/S0165-0114(98)00286-3).
- [50] F. Tseng, G. Tzeng, A fuzzy seasonal ARIMA model for forecasting, *Fuzzy Sets Syst.* 126 (2002) 367–376, [http://dx.doi.org/10.1016/S0165-0114\(01\)00047-1](http://dx.doi.org/10.1016/S0165-0114(01)00047-1).
- [51] T. Helin, H. Koivisto, The GARCH-Fuzzy Density method for density forecasting, *Appl. Soft Comput.* J. 11 (6) (2011) 4212–4225, <http://dx.doi.org/10.1016/j.asoc.2011.03.015>.
- [52] Q. Song, B. Chissom, Fuzzy time series and its models, *Fuzzy Sets Syst.* 54 (3) (1993) 269–277, [http://dx.doi.org/10.1016/0165-0114\(93\)90372-0](http://dx.doi.org/10.1016/0165-0114(93)90372-0).
- [53] K. Huarng, Effective lengths of intervals to improve forecasting in fuzzy time series, *Fuzzy Sets Syst.* 123 (2001) 387–394, [http://dx.doi.org/10.1016/S0165-0114\(00\)00057-9](http://dx.doi.org/10.1016/S0165-0114(00)00057-9).
- [54] H. Liu, An integrated fuzzy time series forecasting system, *Expert Syst. Appl.* 36 (2009) 10045–10053, <http://dx.doi.org/10.1016/j.eswa.2009.01.024>.
- [55] S. Li, S. Kuo, Y. Cheng, C. Chen, A vector forecasting model for fuzzy time series, *Appl. Soft Comput.* J. 11 (3) (2011) 3125–3134, <http://dx.doi.org/10.1016/j.asoc.2010.12.015>.
- [56] M. Chen, B. Chen, Online fuzzy time series analysis based on entropy discretization and a Fast Fourier Transform, *Appl. Soft Comput.* J. 14 (PART B) (2014) 156–166, <http://dx.doi.org/10.1016/j.asoc.2013.07.024>.
- [57] S. Chen, C. Chen, TAIEX forecasting based on fuzzy time series and fuzzy variation groups, *IEEE Trans. Fuzzy Syst.* 19 (1) (2011) 1–12, [10.1109/TFUZZ.2010.2073712](http://dx.doi.org/10.1109/TFUZZ.2010.2073712).
- [58] I. Perfilieva, Fuzzy transforms: theory and applications, *Fuzzy Sets Syst.* 157 (2006) 993–1023, <http://dx.doi.org/10.1016/j.fss.2005.11.012>.
- [59] E. Bautu, A. Barbulescu, Forecasting meteorological time series using soft computing methods: an empirical study, *Appl. Math. Inform. Sci.* 7 (4) (2013) 1297–1306, <http://dx.doi.org/10.12785/amis/070404>.
- [60] A. Jain, A. Kumar, Hybrid neural network models for hydrologic time series forecasting, *Appl. Soft Comput.* J. 7 (2) (2007) 585–592, <http://dx.doi.org/10.1016/j.asoc.2006.03.002>.
- [61] A. Tewari, M. Macdonald, Knowledge-based parameter identification of TSK fuzzy models, *Appl. Soft Comput.* J. 10 (2) (2010) 481–489, <http://dx.doi.org/10.1016/j.asoc.2009.08.034>.
- [62] W. Yap, V. Karri, Comparative analysis of artificial neural networks and dynamic models as virtual sensors, *Appl. Soft Comput.* J. 13 (1) (2013) 181–188, <http://dx.doi.org/10.1016/j.asoc.2012.08.022>.
- [63] W. Froelich, E. Papageorgiou, M. Samarinas, K. Skriapas, Application of evolutionary fuzzy cognitive maps to the long-term prediction of prostate cancer, *Appl. Soft Comput.* J. 12 (12) (2012) 3810–3817, <http://dx.doi.org/10.1016/j.asoc.2012.02.005>.
- [64] J. Yao, J. Herbert, Financial time-series analysis with rough sets, *Appl. Soft Comput.* J. 9 (3) (2009) 1000–1007, <http://dx.doi.org/10.1016/j.asoc.2009.01.003>.
- [65] C. Huang, A hybrid stock selection model using genetic algorithms and support vector regression, *Appl. Soft Comput.* J. 12 (2) (2012) 807–818, <http://dx.doi.org/10.1016/j.asoc.2011.10.009>.
- [66] A. Kazem, E. Sharifi, F. Hussain, M. Saberi, O. Hussain, Support vector regression with chaos-based firefly algorithm for stock market price forecasting, *Appl. Soft Comput.* 13 (2013) 947–958, <http://dx.doi.org/10.1016/j.asoc.2012.09.024>.
- [67] S. Chen, Y. Chang, Multi-variable fuzzy forecasting based on fuzzy clustering and fuzzy rule interpolation techniques, *Inform. Sci.* 180 (24) (2010) 4772–4783, <http://dx.doi.org/10.1016/j.ins.2010.08.026>.
- [68] R. Clemen, R. Winkler, Combining probability distributions from experts in risk analysis, *Risk Anal.* 19 (2) (1999) 187–203, <http://dx.doi.org/10.1111/j.1539-6924.1999.tb00399.x>.
- [69] T. Doan, R. Litterman, C. Sims, Forecasting and conditional projection using realistic prior distributions, *Econ. Rev.* 3 (1984) 1–100.
- [70] M. Beer, S. Ferson, V. Kreinovich, Imprecise probabilities in engineering analyses, *Mech. Syst. Signal Process.* 37 (1–2) (2013) 4–29, <http://dx.doi.org/10.1016/j.jymssp.2013.01.024>.
- [71] M. Stein, M. Beer, V. Kreinovich, Bayesian approach for inconsistent information, *Inform. Sci.* 245 (2013) 96–111, <http://dx.doi.org/10.1016/j.ins.2013.02.024>.
- [72] V. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998.
- [73] S. Makridakis, M. Hibon, The M-3 competition: results, conclusions, and implications, *Int. J. Forecast.* 16 (2000) 451–476, [http://dx.doi.org/10.1016/S0169-2070\(00\)00057-1](http://dx.doi.org/10.1016/S0169-2070(00)00057-1).