Pattern Recognition Letters 000 (2017) 1–10

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Data-driven approaches for meteorological time series prediction: A comparative study of the state-of-the-art computational intelligence techniques

Monidipa Das*, Soumya K. Ghosh

Department of Computer Science and Engineering, Indian Institute of Technology, Kharagpur 721302, India

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Computational intelligence
Data-driven modeling
Bayesian network
Time series prediction
Meteorology

ABSTRACT

With the proliferation of sensor generated weather data, the data-driven modeling for prediction of meteorological time series has gained increasing research interest in current years. The recent advancement in machine learning and artificial intelligence paradigm has made such data analysis process more effective, flexible and sound. This paper attempts to provide a comparative study of the state-of-the art computational intelligence (CI) techniques, which have been successfully applied for meteorological time series prediction purpose. The study has been carried out considering eleven distinct variants of CI techniques, especially based on artificial neural network (ANN), fuzzy logic, Bayesian network (BN) and other probabilistic models. Further, one more hybrid CI technique (SpaFBN), derived from the existing approaches, has been proposed in the present work. All these CI techniques have been empirically studied with respect to a multivariate meteorological time series prediction problem, in comparison with three benchmark statistical approaches. Overall, the experimental results demonstrate the superiority of the BN-based models in meteorological prediction. The presently proposed *spatial fuzzy Bayesian network* (SpaFBN) is also found to be an effective tool, especially for predicting humidity and precipitation rate time series. Moreover, the proposed SpaFBN is a generic CI technique which can be applied for predicting spatial time series from the domains other than meteorology as well.

1. Introduction

Meteorological time series prediction is important not only for our day-to-day planning, but also for long term decision making, which can have grave influence on the economic development of any country. Traditionally, the meteorological predictions are made mainly based on physics driven approaches as followed by various global circulation models or numerical weather prediction (NWP) models [2]. However, the recent data explosion has led to the emergence of a new paradigm, termed as *data driven modeling* or DDM [29], which aims at extensively analyzing historical data for generating insights, and utilizing those in further studies. The mathematical equations underlying such approaches are not derived from physical processes. Rather, these approaches are mainly based on various *computational intelligence* (CI) techniques, like artificial neural network (ANN), Bayesian/Belief network (BN), fuzzy logic (FL), genetic algorithms (GAs) and so on.

In this work, an attempt has been made to provide a comprehensive study of the various CI techniques that have been applied for *meteorological time series prediction*. The focus is kept on prediction of spatial time series data obtained from the spatially distributed sensors, and on the CI techniques which have been either recently proposed or most widely used in the recent past.

A number of works reporting the application of CI techniques in time series prediction can be found in literature. Sapankevych and Sankar [28] have presented an exhaustive survey on time series prediction using support vector machines (SVMs). SVM based prediction of air quality, rainfall, environment pollution etc. has been studied here along with other applications. A survey of the wind speed and wind power prediction has been provided by Lei et al. [16]. In their work, Thissen et al. [31] have found that the SVMs perform better than recurrent neural network (RNN) and statistical autoregressive model, in predicting nonlinear chaotic time series. The existing prediction approaches on SVM, ANN, fuzzy logic or combined neuro-fuzzy techniques have been reported here. Prediction of daily precipitation time series, considering variants of ANN models, has been discussed in the work by Partal et al. [24]. Similar study on rainfall prediction can be found in the work by Wu et al. [33]. A comparative study of traditional statistical autoregres-

* Corresponding author.
E-mail addresses: monidipadas@hotmail.com, monidipa42.das@rediffmail.com (M. Das).

http://dx.doi.org/10.1016/j.patrec.2017.08.009
0167-8655/

sive models and autoregressive NN model, with respect to univariate prediction of rainfall time series, has been reported in the work of Chattopadhyay and Chattopadhyay [7]. The study shows superiority of NN model over the traditional statistical approaches.

However, most of the above-referred works either focus on a single computational intelligence family or consider a single meteorological/environmental parameter. Moreover, none of these works has studied on the Bayesian network-based approaches, which also belong to the computational intelligence family and have proved to show encouraging performance in environmental modeling [1].

In contrast, the present work provides a comparative study considering ANN, SVM, hybrid fuzzy-based model, BN and other probabilistic models together. The novelty lies here in accounting for the Bayesian network and its recently proposed variants. Moreover, one of the major contributions in this work is to propose SpaFBN as new CI technique, extending *spatial Bayesian network with incorporated fuzziness*. Further, in the present work, the comparative study has been made with respect to prediction of *three* primary meteorological variables, namely *temperature, relative humidity* and *precipitation rate*, from two separate climate regions.

In the present context of meteorological time series prediction, the overall prediction problem and the associated challenges have been discussed in the subsequent part of this section. Incidentally, the proposed SpaFBN and the other CI techniques discussed in this paper are applicable for predicting not only the meteorological time series but also the spatial time series from various other disciplines.

1.1. Problem statement and challenges

The problem of meteorological time series prediction, with respect to which the comparative empirical study has been made, can be formally stated as follows:

- Given, the historical daily time series data set over n meteorological parameters in $Z = \{z_1, z_2, \cdots, z_n\}$, corresponding to a set of L locations: $Loc = \{l_1, l_2, \cdots, l_L\}$ for previous t years: $\{y_1, y_2, \cdots, y_t\}$. Also given, the spatial attributes $SA = \{sa_1^{l_i}, sa_2^{l_i}, \cdots, sa_p^{l_i}\}$ for each location $l_i \in Loc$. The problem is to determine the daily time series of the variables in Z for any location $x \in Loc$ for future q years $\{y_{(t+1)}, y_{(t+2)}, \ldots, y_{(t+q)}\}$, when the spatial attributes of x is observed as $\{sa_1^x, sa_2^x, \cdots, sa_p^x\}$. Here, q is a positive integer, i.e. $q \in \{1, 2, 3, \cdots\}$.

The key challenges in such meteorological prediction mainly arise due to the spatio-temporal nature of the data. Unlike the classical data, the spatio-temporal data are highly autocorrelated, that means, the data from nearby locations are more likely to have similar values than those from locations that are far apart. Besides, such data are not independent; rather these are dependent on various co-located variables. Therefore, the conventional statistical methods, which assume that the data are independent and identically distributed, are not very suitable for analyzing such kind of data. Moreover, the meteorological data are non-linear, inherently chaotic, and full of uncertainty.

Research efforts have been made to extend existing traditional statistical and artificial intelligence techniques to cope up with these special properties of meteorological time series data. The present paper aims at summarizing the well-used and recently proposed variants of computational intelligence (CI) techniques, along with their pros and cons in respect of *meteorological time series prediction*. Comparative study has also been carried out empirically, with consideration to three meteorological variables.

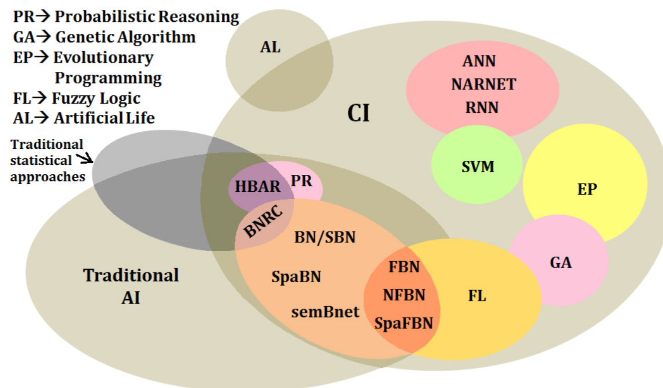

Fig. 1. Relationship between the various CI techniques.

1.2. Contributions

The key contributions of the present study are as follows:

- providing a compact discussion on the various CI techniques (including ANN, SVM, fuzzy logic, BN etc.) used for meteorological time series prediction;
- taking into account the recently proposed variants of Bayesian networks (BNs) which has not been covered by the earlier studies;
- proposing a *new extension of spatial Bayesian network*, namely SpaFBN, that can aid in meteorological prediction with reduced parameter uncertainty;
- performing a comparative empirical study of all the discussed CI techniques, in the light of predicting *temperature*, *humidity* and *precipitation* time series for *two* separate climate regions in India.

1.3. Organization of the paper

The rest of the paper is organized as follows. A comprehensive overview of all the considered CI techniques has been provided in Section 2. The theoretical foundation of the proposed hybrid CI technique (SpaFBN) has been presented in Section 3. The comparative study of all the considered and proposed CI techniques has been extensively discussed in Section 4, with respect to a multivariate time series prediction problem. Finally, the concluding remarks have been made in Section 5.

2. Overview of the state-of-the-art computational intelligence (CI) techniques

Computational intelligence (CI) is a set of nature-inspired computational methods to address the real world problems. It is based on the hypothesis that ‘reasoning is computation’. According to Konar [15], the CI family consists of *granular computing* (fuzzy sets, probabilistic reasoning etc.), *neural computing* (e.g. artificial neural network or ANN), *evolutionary computing* (genetic algorithm, genetic programming etc.) and their *interaction* with artificial life, chaos theory and others.

The recent advancement of CI has greatly influenced the data-driven modeling, since CI techniques are capable of modeling the complex relationships among the parameters without knowing actual natural processes. A brief overview of all the CI techniques, which have been dealt with in the present study, is presented subsequently. The relationships among these CI techniques are depicted in Fig. 1.

Please cite this article as: M. Das, S.K. Ghosh, Data-driven approaches for meteorological time series prediction: A comparative study of the state-of-the-art computational intelligence techniques, Pattern Recognition Letters (2017), http://dx.doi.org/10.1016/j.patrec.2017.08.009