Expert Systems With Applications 59 (2016) 60–77

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Period-aware local modelling and data selection for time series prediction

Marcin Bernas*, Bartłomiej Płaczek

Institute of Computer Science, University of Silesia, Będzińska 39, 41-200 Sosnowiec, Poland

ARTICLE INFO

Article history:
Received 25 May 2015
Revised 2 March 2016
Accepted 18 April 2016
Available online 20 April 2016

Keywords:
Local models
Time series prediction
Data reduction
Segmentation
k-nearest neighbours
Soft computing

ABSTRACT

The paper tackles with local models (LM) for periodical time series (TS) prediction. A novel prediction method is introduced, which achieves high prediction accuracy by extracting relevant data from historical TS for LMs training. According to the proposed method, the period of TS is determined by using autocorrelation function and moving average filter. A segment of relevant historical data is determined for each time step of the TS period. The data for LMs training are selected on the basis of the k-nearest neighbours approach with a new hybrid usefulness-related distance. The proposed definition of hybrid distance takes into account usefulness of data for making predictions at a given time step. During the training procedure, only the most informative lags are taken into account. The number of most informative lags is determined in accordance with the Kraskov's mutual information criteria. The proposed approach enables effective applications of various machine learning (ML) techniques for prediction making in expert and intelligent systems. Effectiveness of this approach was experimentally verified for three popular ML methods: neural network, support vector machine, and adaptive neuro-fuzzy inference system. The complexity of LMs was reduced by TS preprocessing and informative lags selection. Experiments on synthetic and real-world datasets, covering various application areas, confirm that the proposed period aware method can give better prediction accuracy than state-of-the-art global models and LMs. Moreover, the data selection reduces the size of training dataset. Hence, the LMs can be trained in a shorter time.

1. Introduction

Time series (TS) prediction is an active research topic, due to its application potential in many areas of science and industry. The TS prediction algorithms play a major role in decision-making processes for various applications, e.g., stock markets, climate changes, industrial management, and transportation. Over the past decade, much effort has been devoted to the fusion and improvement of conventional TS prediction models (Lin, Yeh, & Lee, 2011; Khashei & Bijari, 2011). Machine learning (ML) is an alternative approach to the TS prediction problem (Štěpnička, Cortez, Donate, & Štěpničková, 2013). Recently, special attention was paid to the ML methods that are based on local prediction models. A local model (LM) is built "just in time", i.e., when a prediction is required, by using historical data that are similar to current observations (Kaneko, Matsuzaki, Ito, Oogai, & Uchida, 2010). The LMs take advantage of the divide and conquer principle by splitting the global prediction problem into several sub-problems and adjusting a LM for a specific sub-problem (Martinez-Rego, Fontenla-Romero, & Alonso-Betanzos, 2011 and Wu & Lee, 2015).

These approaches do not take under consideration the periodic character of TS during selection of the historical data (nearest neighbours) that are used for model training. In this paper a novel prediction method is proposed, which detects periodic changes in a TS and utilizes the information about TS periodicity, to extract relevant data for LM training.

The proposed prediction method consists of the following main steps. First, a period of the analysed TS is determined by using autocorrelation function and moving average filter (Box & Jenkins, 2008). Second, a usefulness relation is extracted from the TS. The usefulness relation enables selection of relevant data for training the LM at a given step of the TS period. Third, when a prediction query has to be processed, the similar historical data (k-nearest neighbours) are searched by taking into account a hybrid usefulness-related distance. Forth, the most informative lags for the selected data subseries are extracted in accordance with the Kraskov mutual information criteria (Kraskov & Stogbauer, 2004). Fifth, training of the LM is performed based on the selected data and finally, the prediction is made.

The novelty of the proposed method lies in selection of the training data that are expected to be useful for prediction making

* Corresponding author. Tel.: +48 517 435 509.
E-mail addresses: marcin.bernas@gmail.com (M. Bernas), placzek.bartlomiej@gmail.com (B. Płaczek).

http://dx.doi.org/10.1016/j.eswa.2016.04.024
0957-4174/

at a given step of TS period. There are two main contributions of this work: a ranking-based algorithm for extraction of the usefulness relation from periodical TS, and a definition of the usefulness-related hybrid distance, which enables selection of relevant historical data to train the LMs. Effectiveness of the proposed approach was confirmed in experiments with real-world and synthetic TS. The prediction accuracy was compared with that of the seasonal ARIMA (Box & Jenkins, 2008) as well as the LMs without period-aware training data selection (Wu & Lee, 2015).

The paper is organized as follows. Related works are reviewed in Section 2. Section 3 describes details of the proposed method. An example of training data extraction from periodical TS is presented in Section 4. Section 5 includes presentation and discussion of the experimental results. Finally, conclusions are given and future research directions are outlined in Section 6.

2. Related works and contribution

2.1. Prediction methods

TS prediction has been an active research area over last decade. The variety of applications brings a need for universal prediction tools as well as dedicated models that could be applied for a given problem. The dedicated models are commonly used in such areas as weather forecast (Liang et al., 2012), transportation (Płaczek, 2013) or medicine (Wu, Cheng, Su, & Feng, 2003).

The development of the dedicated models requires detailed knowledge of the predicted processes. However, if detailed knowledge is not available, the development of dedicated model becomes a difficult task. In such case, the prediction model can be constructed based on historical data, by using ML algorithms and statistical data analysis to find a relation that enable prediction of future values. The historical data usually has the form of discrete TS that contains data points observed at constant time intervals.

If the prediction is made one time interval ahead into the future, it is called one-step or single-step forecasting (Gooijer & Hyndman, 2006). This type of prediction is used in real time applications, e.g., stock market exchange (Zatlavi, Kenett, & Ben-Jacob, 2014) or traffic control (Bernaś, Płaczek, Porwik, & Pamuła, 2015). A multi-step prediction refers to estimation of future values for more than one time interval ahead. Such prediction is used in long-term analysis, e.g., to forecast the climate change (Liang et al., 2012).

Over the past decade, major advances have occurred in statistical models and ML methods for TS prediction. In the literature, several linear prediction approaches were proposed. ARIMA model is one of the most popular prediction methods. This method can be used when the considered TS is stationary and no data are missing (Weigend & Gershenfeld, 1993). Several extensions of ARIMA have been proposed that enable applications for different types of TS (Box & Jenkins, 2008). These extensions include the seasonal models (Khashei et al., 2012; Joo & Kim, 2015). The major drawback of those approaches is the pre-assumed linearity of the model and sensitivity to outliers (Khashei & Bijari, 2011).

Other statistical methods, like spectral analysis (Brillinger, 2011), Markov process (Lennox, Dahl, Vannucci, Day, & Tsai, 2010; Zhou, Guo, Gao, Zhao, & Yan, 2014) and Kalman filter (Lin, Chen, & Peng, 2012) are based on the probability theory and require prior knowledge of the underlying process.

The ML methods have been introduced to enable extraction of underlying characteristics for a predicted process without the prior knowledge and human intervention (Štěpnička et al., 2013). The most widely used ML method is based on artificial neural networks (ANNs). ANN became one of the most important nonparametric nonlinear TS prediction models. Main advantage of ANNs is the capability of flexible nonlinear function approximation with a desired accuracy (Cybenko, 1989). As a nonparametric and data-driven model, ANNs do not require additional assumptions before the model generation (Zhang, Patuwo, & Hu, 1998).

Various problems and challenges are associated with ANNs. The selected weights and thresholds have mayor impact on the prediction results. When training the ANN, local optima can be found instead of the global optimum. Wang, Zeng, and Chen (2015) have proposed an adaptive differential evolution algorithm to select appropriate initial connection weights and thresholds for ANN. Kocadağlı and Aşıkgil (2014) have used a Bayesian inference approach to train an ANN. Kourentzes, Barrow, and Crone (2014) have suggested that a hybrid ANNs ensemble may improve robustness and accuracy of prediction at the cost of increased complexity. Nevertheless, ANN is still considered as a 'black-box' and does not provide intuitive description of the prediction process (Lai, Fan, Huang, & Chang, 2009; Du, Leung, & Kwong, 2014).

Other ML techniques that have been successfully applied for TS prediction include the adaptive neural fuzzy inference system (ANFIS) and the support vector machines (SVMs). ANFIS allows a set of IF-THEN rules and membership functions of fuzzy sets to be constructed based on the historical data (Jyh-Shing,1993 and Jang, Sun, & Mizutani, 1997). This inference system integrates the best features of ANNs and fuzzy logic to handle the non-linearity and uncertainty in real-world processes (Piero, 2000 and Lee & Ouyang, 2003). SVMs have found many applications in classification, pattern recognition and regression analysis (Suykens & Vandewalle, 1999). Over the years, multiple variations of this method have been proposed. Partial least squares SVM method combines the partial least squares based feature selection with support vector machine for information fusion (Yang, You, & Ji, 2011). This method was proposed to identify complex nonlinearity and correlations among financial indicators. Ensemble learning proposed by Kang et al. (2010) improves the performance of SVM-based classification and prediction algorithms. Fuzzy sets adaptation (Chaudhuri & Kajal, 2011) is capable of handling uncertainty and imprecision in prediction of corporate data. It is effective in finding a subset of optimal features and parameters. Other examples of SVM applications to financial predictions can be found in (Lin et al., 2011) and (Chen, Chen, & Chang, 2010). Least squares SVM (LS-SVM) uses linear instead of quadratic programming, thus it reduces computational complexity of the original SVM algorithm (Gestel, Baesens, Suykens, & Espinoza, 2003). LS-SVM involves mapping the data to a space of features, in which a function is constructed that can be used for TS prediction (Huang & Shyu, 2010).

The prediction models based on ML can be categorized into two classes: LMs, and global models. A global model is trained only once and then the same model is used for making many predictions (at different time instances). A LM is trained independently for each prediction case (Martinez-Rego et al., 2011). The LMs are usually trained by using a relatively small number of historical data subseries (nearest neighbours) that are similar to an input sequence (query) for which the prediction has to be made. The main issues of local modelling are efficient model building (Kaneko et al., 2010) and selection of lags that provides useful information. In Hastie, Tibshirani, and Friedman (2008) a lag selection method was proposed which uses t-statistics of estimated coefficients. Several distance measures are commonly used: Euclidean distance, hash function transformation (Chang, Lee, Yoon, & Baek, 2012) or fuzzy measures (Smith & Oswald, 2003). Mutual information criteria were used for informative lags selection (Božić, Stojanović, Stajić, & Floranović, 2013 and Wu & Lee, 2015).

2.2. Time series pre-processing

In the related literature, several pre-processing methods have been proposed in order to improve the accuracy of TS prediction.