Knowledge-Based Systems xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

A novel transfer learning framework for time series forecasting

Rui Ye, Qun Dai*

College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

ARTICLE INFO

Keywords:
Time series prediction
Transfer learning
Extreme learning machine (ELM)
Online learning
Ensemble learning

ABSTRACT

Recently, many excellent algorithms for time series prediction issues have been proposed, most of which are developed based on the assumption that sufficient training data and testing data under the same distribution are available. However, in reality, time-series data usually exhibit some kind of time-varying characteristic, which may lead to a wide variability between old data and new data. Hence, how to transfer knowledge over a long time span, when addressing time series prediction issues, poses serious challenges. To solve this problem, in this paper, a hybrid algorithm based on transfer learning, Online Sequential Extreme Learning Machine with Kernels (OS-ELMK), and ensemble learning, abbreviated as TrEnOS-ELMK, is proposed, along with its precise mathematic derivation. It aims to make the most of, rather than discard, the adequate long-ago data, and constructs an algorithm framework for transfer learning in time series forecasting, which is groundbreaking. Inspired by the preferable performance of models ensemble, ensemble learning scheme is also incorporated into our proposed algorithm, where the weights of the constituent models are adaptively updated according to their performances on fresh samples. Compared to many existing time series prediction methods, the newly proposed algorithm takes long-ago data into consideration and can effectively leverage the latent knowledge implied in these data for current prediction. In addition, TrEnOS-ELMK naturally inherits merits of both OS-ELMK and ensemble learning due to its incorporation of the two techniques. Experimental results on three synthetic and six real-world datasets demonstrate the effectiveness of the proposed algorithm.

1. Introduction

As an important direction of dynamic data analysis and processing, time series prediction has stirred up broad attention in many research fields. For the pressing demand of forecasting the future data trend based on historical information, time series prediction comes into wide use in a variety of practical situations, involving engineering, econometrics and natural sciences, etc. [1,2]. Over the past decades, much excellent effort has been contributed to the development of time series prediction, such as traditional linear models [3], e.g., exponential smoothing, autoregressive integrated moving average, and nonlinear models [4–7], e.g., neural networks, support vector machines with nonlinear kernel, extreme learning machines and other machine learning methods.

Due to the attractive properties and effective performances, Extreme Learning Machines (ELMs) have been widely applied to many fields [8]. ELMs are originally proposed for single hidden layer feedforward networks (SLFNs) [8]. Unlike conventional neural networks, the weights connecting input layer and hidden layer and the bias values in SLFNs can be randomly generated in ELMs, where they need not to be adjusted during the training process. Besides, according to the theory of Moore-Penrose generalized matrix, ELMs can effectively obtain the output weights and the output matrix of hidden layer, which vests the algorithm with extremely fast learning capability and good generalization performance. However, since conventional ELMs essentially rely on the linear combination of a fixed number of nonlinear expansions of input vector, presupposing the number of hidden nodes before learning is of great importance [8,9]. Admittedly, how to appropriately determine the hidden-node number poses serious challenge for ELMs [10,11].

To make up for the deficiency of classical ELMs, an extended method, namely kernel extreme learning machine (ELMK), was proposed recently by Huang et al. [12,13]. It is derived from the modeling and solution process of support vector machines (SVMs) and other kernel methods. Instead of using random feature mappings, ELMK incorporates kernel functions with ELMs and forms new kernel mappings, which can conquer the difficulty of determining appropriate number of hidden nodes [14,15]. ELMK eliminates the defect of classical ELMs and has been verified to have similar or even better prediction capacity.

Nevertheless, since ELMK is essentially based on batch learning, it adopts the offline learning strategy, that means, ELMK takes all the available data into account at once. With the sequential arrival of new samples, it is difficult for ELMK to real-time update the output weights,

* Corresponding author.
E-mail address: daiqun@nuaa.edu.cn (Q. Dai).

https://doi.org/10.1016/j.knosys.2018.05.021
Received 15 December 2017; Received in revised form 14 May 2018; Accepted 16 May 2018

which inevitably imposes limitations on addressing real-time situations. In allusion to this problem, some extensions have been proposed to strengthen the robustness of ELM and ELMK for dealing with online problems, such as online sequential ELM (OS-ELM) [16–18], online sequential fuzzy ELM (OS-fuzzy-ELM) [19] and the recently proposed algorithm called online sequential extreme learning machine with kernels (OS-ELMK) [20,21], where samples arrive in a sequential form, one by one or chunk by chunk.

Empirical simulations demonstrate that the aforementioned algorithms usually have considerable performances in the applications of time series prediction issues. However, most algorithms designed for time series prediction are based on the assumption that training samples and testing samples are drawn from the same distribution, and abundant data for training are available. Nevertheless, in many practical situations, time series data are most likely to vary over time, resulting in a wide variability between old data and new data. In addition, in some real-time cases, size of sequentially arrived samples may be relatively small at a time, which may lead to insufficient data for training in the early period. Hence, when tackling the issue of scant fresh training samples, as is often the case in time series prediction, how to transfer knowledge of previous data, whose observation time is far apart from new samples, to the predictive process is of great significance. However, up till now, little research with respect to this aspect has been conducted. In our paper, we lay much emphasis on how to deal with the aforementioned problem.

Since the conundrum we now face is how to make the most of old data and maximally leverage their knowledge for implementing the current new prediction task, this problem, to some extent, is similar to that of transfer learning. Different from the traditional machine learning paradigm, transfer learning breaks the constraint that training data and test data should be under the same distribution [22–24]. It refers to the problem of reusing and transferring knowledge of one domain to other different but related domains. Generally speaking, the primary objective of transfer learning is to retain experience learnt from past data and apply it to current problems. It is desirable to learn an efficient model with only a handful of fresh samples and relatively abundant old data for time series prediction issues. Hence, based on the appealing properties of transfer learning, we incorporate it with online sequential extreme learning machines with kernels (OS-ELMK) for time series prediction.

Inspired by the fact that performance of an ensemble composing of multiple models is generally superior to that of individual ones, the ensemble learning idea is incorporated in our implementation [25,26]. Consequently, a hybrid algorithm incorporating transfer learning, OS-ELMK and ensemble learning, abbreviated as TrEnOS-ELMK, is proposed ultimately in this paper. Compared to conventional time series prediction methods, TrEnOS-ELMK takes previous data, which possess a long time span with current data, into consideration, and first forays into constructing a novel framework for time series prediction based on transfer learning, where knowledge learnt from old data can be effectively leveraged for the present predictive task. While few related investigations involving this aspect have been carried out till now. Detailed descriptions of the proposed algorithm are specified in the sequel.

Fig. 1 presents the construction of the proposed method. In the first step, as shown in Fig. 1, abundant past data having a long time span with current new samples are extracted. Due to the property of time-varying, a wide variability usually lays athwart old samples and new samples. With the application of transfer learning, knowledge learnt from old data is retained and tactfully leveraged for a future predictive process. In particular, kernel matrix and the corresponding weight matrix of ELMK are firstly generated over the old data.

Next, by virtue of the past experience and a handful of fresh samples, new kernels and weight matrix can be obtained for the following prediction. In the second step, strategy imitating OS-ELM is implemented over the sequentially arriving time series data. Specifically, kernels and weight matrix are online updated with the arrival of new samples.

Motivated by the preferable performance of ensemble learning, several models are generated and updated to construct an ensemble. Besides, to maintain a satisfactory performance, weights assigned to different models are adaptively updated according to their corresponding performances on the available new data, where models with higher efficacy are usually endowed with heavier weights. With the arrival of fresh samples, if the ensemble performance falls short of standards, the poorest model will be replaced with a newly trained one.

The main contributions of this work can be summarized as follows:

Firstly, most methods designed for time series prediction are based on the assumption that sufficient training data and testing data drawn from the same distribution are available. Nevertheless, in some practical situations, this assumption may fail to be met because of the time-varying property of time series data. However, up to now, few researches regarding this aspect have been put forward. To alleviate this problem, a new hybrid algorithm incorporating transfer learning with OS-ELMK, abbreviated as TrOS-ELMK, is proposed with a mathematical derivation. It tactfully employs transfer learning for time series prediction and first attempts to make the most of long-ago data rather than throwing away it.

Due to the incorporation of kernel trick, OS-ELMK does not need to determine the number of hidden nodes, which, to some extent, enhances its robustness and prediction ability. Admittedly, our newly proposed algorithm inherits the advantages of OS-ELMK and irons out its flaw of failing to leveraging knowledge of previous data for new predictive tasks. Compared to existing methods, the newly proposed algorithm can make the most of past experience, and can get a preferable performance, even though scarce fresh samples are available.

Secondly, inspired by the considerable performance of ensemble learning, a collection of models endowed with weights is generated in our algorithm. Instead of equally treating all the components in the ensemble, weights allocated for different models are updated with the arrival of new data, and superior models are remained while inferior ones are eliminated. Thus, the ensemble alters with new time series rather than remaining unchanged, aiming to highlight the effects of well-performed models. More specifically, to ensure the ensemble's performance, weights assigned to different models are adaptively adjusted according to their prediction capacities over new coming data. Moreover, models with poor performance would be replaced if the ensemble's prediction ability was substandard. Therefore, the final ensemble can maintain the preferable performance adaptively, owning good characters of better variety and higher feasibility.

The rest of this paper is organized as follows. In Section 2, we briefly review some previous work regarding time series prediction and transfer learning. In Section 3, exhaustive analysis of TrOS-ELMK is deduced. Then detailed elucidations of our newly proposed TrEnOS-ELMK are presented in Section 4. Experiments over synthetic and real-world datasets are carried out in Section 5, where corresponding experimental results are reported and discussions are provided. Finally, a summary of our work is displayed in Section 6.

2. Related work

2.1. Existing methods for time series prediction

Researches about time series prediction turn out to have a long history. Remarkable results and assorted breakthroughs have been made during past decades. Put crudely, the existing time series prediction methods can be roughly divided into the following two categories:

Linear predicting methods: With these methods, time series prediction models are constructed on basis of linear functional forms. For example, the earliest proposed autoregressive (AR) models aimed to