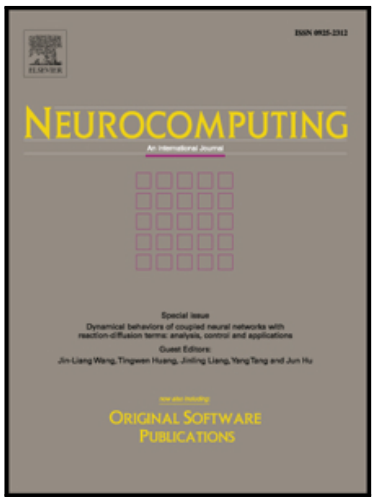

Online Sequential ELM Algorithm with Forgetting Factor for Real Applications

Haigang Zhang, Sen Zhang, Yixin Yin

PII: S0925-2312(17)30220-5
DOI: 10.1016/j.neucom.2016.09.121
Reference: NEUCOM 18017

To appear in: *Neurocomputing*

Received date: 3 May 2016
Revised date: 27 August 2016
Accepted date: 22 September 2016

Please cite this article as: Haigang Zhang, Sen Zhang, Yixin Yin, Online Sequential ELM Algorithm with Forgetting Factor for Real Applications, *Neurocomputing* (2017), doi: 10.1016/j.neucom.2016.09.121

Online Sequential ELM Algorithm with Forgetting Factor for Real Applications

Haigang Zhang, Sen Zhang, and Yixin Yin

Sequential learning algorithms are a good choice for learning data one-by-one or chunk-by-chunk. Liang et. al. has proposed OS-ELM algorithm based on the ordinary ELM algorithm, which produces better generalization performance than other famous sequential learning algorithms. One of the deficiencies of OS-ELM is that all the observations are weighted equally regardless of the acquisition time. However, the training data often have timeliness in many real industrial applications. In this paper, we propose a modified online sequential learning algorithm with the forgetting factor (named WOS-ELM algorithm) that weights the new observations more. Then a convergence analysis is presented to make sure the estimation of output weights tend to converge at the exponential speed with the arriving of new observations. For the determination of the value of forgetting factor, it would change with the forecast error automically and get rid of excessive human interference. We employ several applications in the simulation part including time-series predication, time-variant system identification and the weather forecast problem. The simulation results show that WOS-ELM is more accurate and robust than other sequential learning algorithms.

Index Terms—**Extreme learning machine, online learning, forgetting factor, sequential learning.**

I. Introduction

Extreme learning machine (ELM) proposed by Huang in 2006 is a fast machine learning algorithm based on the generalized single-hidden layer feedforward networks(SLFNs)[1]. The key advantages of ELM compared with other famous neural network algorithms are that the learning parameters in the neural model are generated randomly without human tuning or iterative method[2,3]. The output weights are determined by the method of least square (LS). Nowadays, it has been widely used in many real applications including both regression and classification problems[4-7].

In many real applications, data are obtained one by one or chunk by chunk. Online sequential machine learning is a model of induction that learns one instance or some instances at a time[8,9]. Liang et. al. has proposed a fast and accurate online sequential learning algorithm (OS-ELM) for SLFNs based on ELM network with additive or radial basis function(RBF) hidden nodes[10]. In OS-ELM, the newly generated observations can be trained one-by-one or chunk-by-chunk with fixed or varying data size, while the output weights will be updated analytically simultaneously. Then many modified OS-ELM algorithms have been proposed, such as EOS-ELM[11], OS-ELMK[12], OL-ELM-TV[13] et. al. However, the above listed online sequential learning methods do not take timeliness aspect of the problem into consideration. Timeliness problem extensively exists in our daily life, such as weather forecast and stock forecast[14,15]. With the time passing by, the distribution of data changes and shows much non-stationary phenomenon. In such cases, the old data should contribute lesser and lesser so that the model represents the most recent behavior[16]. Broadly speaking, in the training of ELM model, we should allocate high weights for new data and low weights for old ones.

The corresponding author is Sen Zhang, email: zhangsen@ustb.edu.cn; Haigang Zhang, Sen Zhang and Yixin Yin are with the School of Automation and Electrical Engineering, University of Science and Technology Beijing, No. 30, Xueyuan road, Haidian District, Beijing, China. This work has been supported by the National Natural Science Foundation of China (NSFC grant No. 61333002,61673056,61673055 and 61671054).

There are many ELM-related online learning algorithms subject to nonstationary applications. FOS-ELM aims to learn the sequential data with timeliness, where a removable sliding-window is employed to limit the active area during the process of data acquisition[17]. Zhou employed the same forgetting mechanism in the regularized and kernelized ELM algorithms[18]. In addition, Wang proposed OS-ELMK algorithm, and combined it with a sliding window for nonstationary time series prediction[19]. With the arriving of new observations, the sliding window would move forward in order to forget the 'old' samples. Another strategy to deal with nonstationary data is based on the introduction of forgetting factor. From the extreme point of view, the method of sliding window can be seen as a special case of the method of forgetting factor. Matias introduced the forgetting factor into OS-ELM algorithm[20]. However, the authors did not present the method to choose an approximate value for the forgetting factor. Lim presented a relatively complex mechanism to determine the value of forgetting factor based on the gradient descent method[21]. The additionally required computational complexity would increase at remarkable speed with the increase of the number of hidden nodes, which is time-consuming and can not meet the needs of online implementation.

In this paper, we propose a novel modified online sequential ELM algorithm named WOS-ELM. WOS-ELM algorithm introduces the forgetting factor in the performance indicator. For the online sequential learning, old data are gradually being forgotten, while new coming data gets more emphasis. Then we present a convergence theorem to ensure the estimation of output weights converges to the true value with the arriving of new observations at the exponential speed. In addition, the forgetting factor can be set to be variable according to the output prediction error automatically. Thus the model can ensure the output error fluctuates around the set point. This automatic updating strategy for the forgetting factor is easy and time-saving to implement, which would not affect the advantage of rapid training speed of ELM algorithm. In addition, we present a mechanism to deal with the contaminated industrial data inspired by the introduction of forgetting factor. More