Applied Soft Computing 62 (2018) 524–535

Contents lists available at ScienceDirect

Applied Soft Computing

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

A novel echo state network for multivariate and nonlinear time series prediction

Lihua Shen [a], Jihong Chen [a,*], Zhigang Zeng [b], Jianzhong Yang [a], Jian Jin [a]

[a] School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, Hubei, China
[b] School of Automation, Huazhong University of Science and Technology, Wuhan 430074, Hubei, China

ARTICLE INFO

Article history:
Received 5 December 2016
Received in revised form 22 October 2017
Accepted 23 October 2017
Available online 16 November 2017

Keywords:
Robust echo state network
Multivariate time series
Variational inference
Prediction
Adaptive

ABSTRACT

A robust and adaptive multivariate nonlinear time series prediction model is proposed based on echo state network and variational inference and we call it robust variational echo state network (RVESN). RVESN uses a heavy-tailed and more robust Gaussian mixture distribution as the likelihood function of the model output. The variational inference procedure which has an advantage over the Laplace approximation is utilized to handle the marginal likelihood function of the model output which is analytically intractable for the mixture distribution. RVESN not only has strong capability of nonlinear approximation as echo state network and avoids the cross-validation process of estimating model parameters but also is more robust to outliers compared to the traditional Bayesian learning with a single Gaussian distribution as the likelihood function of the model output. And furthermore, the Gaussian mixture distribution can describe the underlying dynamic characteristics of the multivariate time series more comprehensively and practically than a single Laplace or Gaussian one by an adaptive parameter. The experimental results of artificial and real-world multivariate nonlinear time series with and without outliers demonstrate that RVESN has better prediction performance.

1. Introduction

In most practical situations, the time series derived from complex systems generally comprise multiple variables, which contain more useful information than in univariate time series [1–3] and nonlinear signal processing is generally needed in real world [4–6], so multivariate and nonlinear time series prediction has attracted substantial attention for dynamic system modeling in recent years [7–9]. For their strong capability of nonlinear approximation, neural network and support vector machine methods such as multilayer perceptrons (MLPs) [10], radial basis function (RBF) neural network [11,12], self-organization map (SOM) [13], back propagation [14], fuzzy interval neural networks [15], co-evolutionary neural networks [16], support vector machine [5,17,18], fuzzy support vector machine [19], support vector echo state machine (SVESM) [20,21], extreme learning machine (ELM) [22] and its variants including kernel-based ELM [9] and support vector extreme learning machine (SVELM) [23], echo state network (ESN) [24], have been increasingly applied for multivariate and nonlinear time series prediction and have good performance [25,26]. ESN is a kind of recurrent neural network with a large number of randomly generated neurons (called the reservoir) [24]. It avoids the disadvantages of the traditional gradient-based neural network models, has faster convergence speed and can achieve the global optimization.

ESN has a large and sparsely connected reservoir, which can map the input variables to high-dimensional space. ESN learning firstly initializes the input weights and the internal weights of the reservoir randomly, and then only adjust the readout weights during training the neural network [24]. Training the ESN is essentially a linear task so as to be able to obtain the global optimal solution, have faster convergence speed, and overcome the shortcomings of traditional neural network. Pseudoinverse is an algebraic method for training ESN. However, the pseudoinverse method is liable to produce ill-posed problems. To avoid the ill-posed problems, regularization methods, such as truncated singular value decomposition [27] and Tikhonov-type regularization [28] for ESN training, have been developed. Ridge regression ESN (RRESN) is the traditional Tikhonov regularization that solves the ill-posed problems by introducing the regularization term into the objective function

* Corresponding author.
E-mail addresses: D201477195@hust.edu.cn (L. Shen), 13308656728@189.cn (J. Chen), zgzeng@hust.edu.cn (Z. Zeng), yangjz@mail.hust.edu.cn (J. Yang), jinjian1964@hust.edu.cn (J. Jin).

https://doi.org/10.1016/j.asoc.2017.10.038
1568-4946/

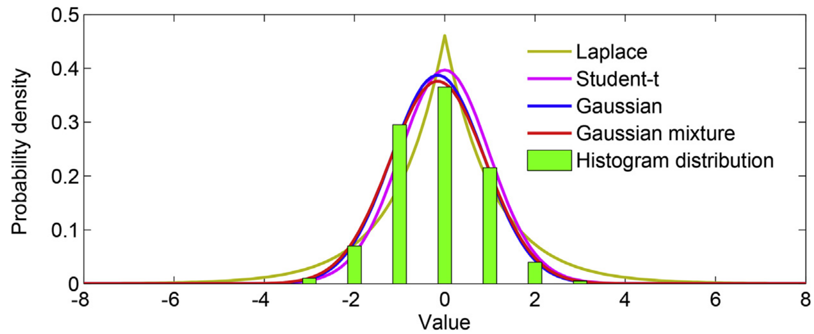

Fig. 1. Probability density curve of different distribution without outliers.

to penalize larger readout weights. The parameter C of the regularization term is needed to be determined to obtain a model with good performance. The commonly used method of determining C is cross-validation, but this method is time-consuming. And furthermore, RRESN is sensitive to the outliers.

ESN based on Bayesian regression (BESN) [29] is on the assumption that the likelihood function of the model output obeys a Gaussian distribution, while this assumption also leads to a model of lacking robustness. If the training dataset is contaminated by outliers, the prediction accuracy of the model based on Bayesian regression can be significantly compromised.

Because the dataset acquisition circumstances are complex and the dataset is affected by a variety of noise, the training dataset is always contaminated by outliers in practice. Building a robust echo state network model is consequently much needed. The robust echo state network (RESN) [30] replaces the commonly used Gaussian distribution with a Laplace one as the likelihood function of the model output. And the likelihood function is finally approximated as a zero-mean Gaussian one by utilizing bound optimization algorithm. However, when the training datasets are not contaminated by the outliers, the likelihood function of the RESN is still set as a Laplace distribution and then approximated as a Gaussian one, which actually lead to information loss.

Take univariate distribution as an example. The probability density curve of different distribution with and without outliers are shown in Figs. 1 and 2. Histogram distribution of 200 integer data points drawn from a Gaussian distribution, together with the maximum likelihood fit obtained from a Laplace, Student-t, Gaussian and Gaussian mixture distribution are shown in Fig. 1. The same dataset but with 20 outliers are shown in Fig. 2. The sensitivity of Gaussian distribution to outliers and the information loss of Laplace distribution can be seen from Figs. 1 and 2.

In order to achieve a robust model which can be adaptive to the datasets with or without outliers and describes the dynamic information more comprehensively, a more practical heavy-tailed distribution is needed to be as the likelihood function of model output. Student-t distribution is a typical heavy-tailed distribution [31,32]. However, using Student-t distribution as the output likelihood function makes the model learning computationally complex. And as observed from Figs. 1 and 2, Laplace distribution leads to loss of information. Consequently as an approximate Student-t distribution, Gaussian mixture distribution is robust to outliers and has the advantages of convenient calculation and retaining more information. So the proposed model utilizes Gaussian mixture distribution as the likelihood function of model output. And it can capture the underlying dynamic characteristics of the multivariate time series and reduce the loss of information by an adaptive parameter with the training datasets contaminated or not. The advantages of Gaussian mixture distribution can be seen from Figs. 1 and 2.

For the marginal likelihood function of the model output is analytically intractable for the replaced distribution, the variational Bayesian inference procedure [33,34] is utilized for the ESN training. The variational approach has an advantage over the Laplace approximation, in particular, the use of variational parameters gives the variational approach greater flexibility. And the greater flexibility of the variational approximation translates into improved accuracy [35]. Thus a robust variational echo state network (RVESN) prediction model is achieved. The proposed RVESN not only has the capability of nonlinear approximation as ESN but also is more robust to outliers and it can represent dynamic characteristics of the multivariate time series more comprehensively. The simulation results also demonstrate that RVESN has better prediction performance.

The rest of this paper is structured as follows. Section 2 gives a brief review of the preliminary works. Section 3 presents the RVESN model. Section 4 gives the simulation results of Lorenz chaotic multivariate dynamic system, Rossler time series, the annual runoff and sunspot bivariate dynamic system and the monthly average temperature and rainfall series in Dalian. In Section 5, the conclusions are given.

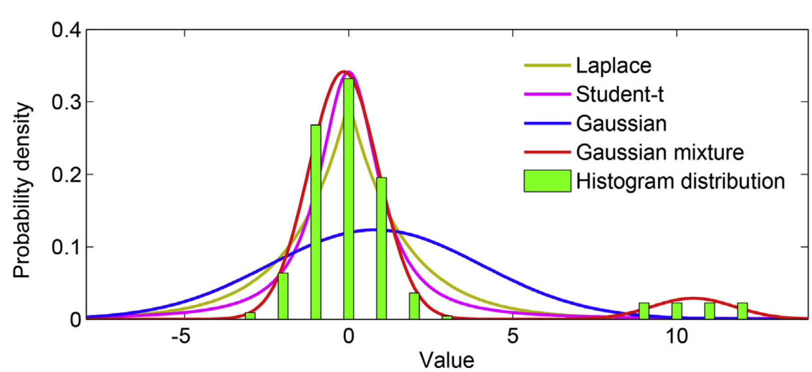

Fig. 2. Probability density curve of different distribution with outliers.