Neurocomputing 275 (2018) 1333–1349

Contents lists available at ScienceDirect

Neurocomputing

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

Recurrent neural system with minimum complexity: A deep learning perspective

Xiaochuan Sun [a,b,*], Tao Li [b], Yingqi Li [a], Qun Li [b], Yue Huang [b], Jiayu Liu [a]

[a] School of Information Engineering, North China University of Science and Technology, Tangshan 063009, PR China
[b] Jiangsu BDSIP Key Lab, School of Computer science and Technology, Nanjing University of Posts and Telecommunications, Nanjing 210023, PR China

ARTICLE INFO

Article history:
Received 7 April 2017
Revised 4 July 2017
Accepted 17 September 2017
Available online 28 September 2017

Communicated by Deng Cheng

Keywords:
Echo state network
Deep belief network
Time series prediction
Memory capacity

ABSTRACT

This paper proposes a novel echo state network (ESN) architecture in a deep learning framework for time series prediction. The architecture is a uniform and consistent system with functional parts of the pre-training input network that effectively captures information with different degrees of abstraction in the observed data, and the minimum complexity ESN that possesses the powerful nonlinear approximation capability and highly efficient training. To our best knowledge, this is the first systematic model attempting to introduce the deep learning methodology to the ESN modeling, which provides a more robust alternative to the conventional shallow ESNs. Extensive experiments on various widely used benchmarks of different origins and features show that our model achieves a great enhancement in the prediction accuracy and short-term memory capacity, without significant tradeoff in the model's computational efficiency.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Recurrent neural network (RNN) offers an outstanding nonlinear approach for modeling dynamic systems, which is characterized by the recurrent connections between neurons. The special architecture is capable of directly processing temporal dependencies related to systems. Generally, RNNs can approximate arbitrary nonlinear dynamical system with arbitrary precision [1]. However, the early RNN architectures suffer from their limited memory capacity, due to the vanishing or exploding gradient problem [2], especially when the information involved in past inputs need to be recovered over a long time interval. In other words, the conventional RNNs poorly handle the long-term temporal dependencies. In contrast, the Long Short Term Memory (LSTM) architecture is capable of dealing with this problem by designing special memory cells that is actually a gated access mechanism to the neurons states [3]. Unfortunately, as a typical gradient-based network, LSTM is still unable to get rid of the drawbacks [2,4], such as slow convergence, excessive calculation and suboptimal solutions. These issues are mainly attributed to the unfolding and backpropagation through time procedure.

An extremely efficient network structure of RNN, called echo state network (ESN), was designed independently by Jaeger [5,6] for solving all the aforementioned issues. ESN is viewed as a powerful tool to model temporal correlations between the input and output sequences, whose kernel part is a single reservoir consisting of a great many neurons that are randomly interconnected and/or self-connected. The reservoir itself remains unchanged, once it is selected. During the ESN training, only the output weights need be computed through offline linear regression or online methods, such as the recursive least square [5–7], which considerably reduces computational complexity. Consequently, the ESN paradigm completely escapes these shortcomings of gradient-descent RNNs (e.g., LSTM) listed above. Until now, ESN has been successfully applied in various research areas, e.g., noise modeling [6], pattern recognition [8], robot control [9], reinforcement learning [10] and time series prediction [11–14].

As a result of these merits, ESN has captured widespread attentions of the computational intelligence community, and many ESN extensions have been explored. The existing ESN implementations mostly concentrate on the design of the network topologies, the selection of the neuron types, and the proposal of training algorithm. For example, Rodan and Tino [11] proposed a minimum complexity network structure for ESN, and through an exhaustive experimental and theoretical analysis, demonstrated that a reservoir could be simplified as much as possible, but not compromising the model's performance. Moreover, the most simplified ESN possessed the memory capacity being arbitrarily close to the

* Corresponding author at: School of Information Engineering, North China University of Science and Technology, Tangshan 063009, PR China.
E-mail address: sunxiaochuan@njupt.edu.cn (X. Sun).

https://doi.org/10.1016/j.neucom.2017.09.075
0925-2312/© 2017 Elsevier B.V. All rights reserved.

proved optimal value. Qiao et al. [12] constructed a growing ESN with a multiple subreservoirs in an incremental way, leading to superior prediction performance and learning speed, and gave the proof of the convergence. Holzmann and Hauser [13] introduced general infinite impulse response filter neurons instead of original sigmoid ones as well as a delay&sum readout for ESNs, which significantly outperformed the standard ESN and other state-of-the-art models for nonlinear dynamical system modeling. Li et al. [14] proposed a robust ESN in a Bayesian framework that replaced the original linear regression with the Bayesian regression, so that the resulted model was capable of dealing with outliers in the training data set. Although these studies have given the remarkable advantages in the modeling performance, the extended ESN models are just shallow neural networks, characterized by a single nonlinear transformation of the input data into a feature space, followed by a linear mapping. A number of recent theoretical results have demonstrated that the shallow structures are insufficient at representing some functions [15–17], since they do not fully consider the diversity of the space distributions of the features in observed data, which greatly affects nonlinear approximation capability. Conversely, deep networks typically exhibit more powerful representation capacity than shallow networks with the same number of parameters for certain types of problems [18,19]. Hence, we would like to refine ESNs for superior nonlinear approximation capability from a deep learning perspective.

Generally, a deep architecture is built-in layers, each of which consists of feature detector units responsible for feature extraction [20,21]. Lower layers extract simple features and inject into higher layers, which successively perceive more abstract features. Deep belief network (DBN) [22–24] is one of the most important multiple-layer deep network architectures as well as a powerful probabilistic generative model. It can be trained to extract a deep hierarchical representation of input data by maximizing the likelihood of training data. Compared with the traditional shallow models, such as support vector machine, DBN can express highly variant functions, discover the potential laws existing in multiple features, and have a better generalization capacity, since *"functions that can be compactly represented by a depth k architecture might require an exponential number of computational elements to be represented by a depth $k-1$ architecture"* [15]. Especially, the DBN model, proposed by Hinton et al. [24], is a most promising alternative for deep networks. Structurally, it is viewed as stacked restricted Boltzmann machines (RBMs), each of which contains a visible layer representing observed data and a hidden layer learning to represent features that capture higher-order correlations in the data [25]. This promising architecture has been successfully applied to various domains, such as natural language understanding [18], hand-written character recognition [26], acoustic modeling [27], action recognition [28,29], super-resolution image processing [30,31], information retrieval [32]. Inspired by this, we investigate the possibility of introducing the DBN methodology to the ESN paradigm so that the resulting model is able to obtain more powerful nonlinear approximation capability.

In this paper, we propose a recurrent neural system with minimum complexity from a deep learning perspective. A unified and consistent architecture is built by integrating DBN and minimum complexity ESN, where a pre-training input network (PIN) is fully connected to the considered ESN structure. The special component is able to learn the multi-level features related to the modeled sequential data, providing a powerful support for the subsequent nonlinear approximation. Similar to DBN, PIN is also composed of stacked RBMs, whereas its learning in a greedy layer-wise mode based on contrastive divergence theory (CD) is completely irrelevant to the following ESN training. Extensive experiments demonstrate the superiority of the proposed model. In summary, the contributions of this paper include three aspects.

(1) To the best of our knowledge, this is the first attempt to introduce the deep learning methodology to the ESN modeling. In theory, the powerful capability of PIN on feature extraction can make the minimum complexity ESN a promising method for time series prediction.
(2) The short-term memory (STM) of our model is redefined as the ability of recovering the visible inputs from the whole network output, and the relationship between the STM capacity and the prediction performance is further given.
(3) The efficacy of the proposed model is evaluated considering a number of well-known benchmark datasets for prediction. Compared to the state-of-the-art models, our model obtains better prediction accuracy, while the computational burden is not significantly increased.

The remaining of the paper is organized as follows. Section 2 provides some basic background on RBM, and DBN. Section 3 elaborates the architecture and learning algorithm of the proposed model. Experiments and evaluation results on the benchmark datasets are given in Section 4. We empirically analyze the STM capacity of our deep ESN model in Section 5. We discuss our DSCR model in Section 6. Finally, this paper is concluded in Section 7.

2. Background

In this section, we give a brief introduction to the concepts related to DBNs, but particularly focus on the theoretical background of RBMs, which are the foundation of DBN understanding.

2.1. Restricted Boltzmann machines

A restricted Boltzmann machine [15,25,32] is a two-layer, undirected, bipartite graphical model composed by two parts, i.e., visible layer and hidden layer, which is trained in an unsupervised mode. Similar to the classical Boltzmann machine, the visible layer and hidden layer are fully connected via symmetric undirected weights, but there is no any intra-layer connection within both visible and hidden layer. The architecture of a typical RBM model is shown in Fig. 1, where v denotes the visible layer, h denotes the hidden layer, and w_{ij} denotes the connection weight between the visible unit i and hidden unit j.

The surprising advantage of RBM is embodied in the idea of reconstruction oriented learning. Just the information in hidden units, learnt as features, can be used to reconstruct the input. Once the original input is recovered perfectly during reconstruction, it implies that the hidden units reserve input information as much as possible, and the updated weights and biases are capable of effectively measuring the input data.

2.2. Deep belief network

In general, as a deep feedforward network, DBN is built by a set of stacked RBMs, and trained by a layer-by-layer learning algorithm in an unsupervised greedy fashion. Especially, the features obtained by a RBM are viewed as the input data for a next RBM. Thereby, RBMs in a DBN are trained one by one, proceeding from the lower-level RBM and progressively shifting up in the hierarchy. In this way, DBN can increasingly capture deep features of input data. A typical DBN architecture is shown in Fig. 2, where it contains four layers: one visible input layer, three RBM hidden layers and one output layer, and the visible input layer is the input layer of the first RBM.

Generally speaking, the DBN learning consists of pre-training and fine-tuning. During the pre-training, input data is loaded to the visible input layer, and then the first RBM maps it to own

متن کامل مقاله ← دریافت فوری

ISIArticles
مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات