Knowledge-Based Systems 130 (2017) 17–29

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Deep belief echo-state network and its application to time series prediction

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

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

ARTICLE INFO

Article history:
Received 19 November 2016
Revised 19 May 2017
Accepted 23 May 2017
Available online 25 May 2017

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

ABSTRACT

Deep belief network (DBN) has attracted many attentions in time series prediction. However, the DBN-based methods fail to provide favorable prediction results due to the congenital defects of the back-propagation method, such as slow convergence and local optimum. To address the problems, we propose a deep belief echo-state network (DBEN) for time series prediction. In the new architecture, DBN is employed for feature learning in an unsupervised fashion, which can effectively extract hierarchical data features. An innovative regression layer, embedding an echo-state learning mechanism instead of the traditional back-propagation method, is built on top of DBN for supervised prediction. To our best knowledge, this is the first paper that applies the echo state network methodology to deep learning. The resulted model, combining the merits of DBN and ESN, provides a more robust alternative to conventional deep neural networks for the superior prediction capacity. Extensive experimental results show that our DBEN can achieve a significant enhancement in the prediction performance, learning speed, and short-term memory capacity.

1. Introduction

Currently, cognitive neuroscience discoveries [1–5] have provided a profound understanding of the principles on how to govern information representation in the mammal brain, developing new ideas of designing a system to effectively represent information. One of the crucial discoveries is that the mammal brain is organized in a deep architecture, in which the input percept can be represented by means of multiple levels of abstraction, each level corresponding to a different area of the cortex. Based on the learnt abstract features, human is capable of performing a wide variety of real-world tasks such as object detection, image recognition, prediction, and visualization. Hence, the machine learning community expresses the tremendous interest in this hierarchical architecture of the mammal brain, and attempts to mimic it in order to enhance the performance of a learning algorithm.

Currently, deep learning [6,7] has become one of the most sophisticated machine learning techniques to learn perfect information representation that provides similar functionality to that of the mammal brain. It exerts a tremendous fascination on researchers for establishing hierarchical representations from data. A general deep architecture is built in layers, each of which consists of feature detector units. Lower layers extract simple features and inject into higher layers, which can successively perceive more abstract features. In particular, deep belief network (DBN) [8], proposed by Hinton et al., is a powerful hierarchical generative model for feature extraction. Compared with the training methods of traditional deep models, such as multilayer perceptron, DBN can effectively obviate over-fitting to the training set by means of a distinctive unsupervised pre-training. Compared with the traditional shallow models, such as support vector machine (SVM), 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"* [7]. Furthermore, Deng et al., defined a hybrid deep neural network (DNN) that was built by adding a discriminative model component over DBN, namely DBN-DNN [9]. In a typical DBN-DNN, DBN serves as feature extraction, while the back-propagation (BP) algorithm in discriminative model component is used to fine-tune the whole network. This hybrid architecture has been successfully applied to many real-world classification problems from various domains, such as hand-written character recognition [10], acoustic modeling [11], health state diagnosis [12], hyperspectral data classifica-

* Corresponding author at: School of Computer Science, 11200 SW 8th Street, ECS 354, Miami, FL 33199-0001, United States.
E-mail addresses: sunxiaochuan@njupt.edu.cn (X. Sun), taoli@cs.fiu.edu (T. Li).

http://dx.doi.org/10.1016/j.knosys.2017.05.022
0950-7051/

tion [13], fingerprint liveness detection [14], natural language understanding [15], information retrieval [16].

Recently, the hybrid DBN-DNN also shows great potential for time series prediction. For example, Kuremoto et al. [17] applied the typical DBN-DNN to time series prediction. Huang et al. [18] proposed a deep architecture consisting of a traditional DBN and a multitask regression layer for traffic flow prediction. In order to accelerate the learning, Shen et al. [19] introduced a conjugate gradient method to DBN-DNN for exchange rate prediction. Although the proposals have exhibited favorable prediction performance for various considered tasks, the used BP algorithm in the logistic regression (LR) layer (i.e., discriminative model component), that performs the global weight fine-tuning, limits their further development in time series prediction. On the one hand, massive iterative computation results in a slow convergence rate (i.e., learning speed). On the other hand, the BP algorithm based on gradient descent can be easily trapped in to a local optimum [20,21], leading to an unsatisfactory prediction accuracy. Thus further studies are required to explore more effective regression method instead of the much-maligned BP algorithm for superior prediction performance.

Echo state network (ESN), proposed by Jeager et al. [22,23], is a typical paradigm of recurrent neural networks (RNNs). It is viewed as a powerful tool to model temporal correlations between the input and output sequences. The learning can be realized through offline linear regression or online methods [22–25], such as the recursive least square, providing optimal weights for the given ESN. ESNs offer some attractive advantages over classic RNNs using the BP algorithm, such as the faster training speed and stronger nonlinear approximation capacity [25–29]. Inspired by it, we investigate the possibility of combining the training method of ESN and feature extraction of DBN so that the resulting model is able to achieve superior prediction performance.

In this paper, we propose a novel deep prediction framework, termed the deep belief echo-state network (DBEN). In the new structure, a DBN is connected to an ESN-based regression layer. For the learning of DBEN, the contrastive divergence method is used to train multiple RBMs in DBN, followed by the local weight adjustment in the LR layer via a supervised echo-state mechanism. The contribution of this paper could be summarized in three aspects.

(1) To the best of our knowledge, this is the first work with implementation details of combining the deep learning and ESN methodology. The merits of both DBN, in terms of the high-efficient feature extraction from data, and ESN, in terms of the exceptional performance in modeling dynamical data, make DBEN a promising method for time series prediction.
(2) The efficacy of the proposed approach is evaluated considering a number of widely used time series benchmarks, and compared with the baseline models, such as the DBN-based method, classical ESN and SVM, to demonstrate its superior performance.
(3) We define the short-term memory (STM) of DBEN as the capacity that it can reconstruct the inputs of visible layer from the outputs, and further show the relationship between the STM and prediction performance.

The remaining of the paper is organized as follows. Section 2 provides some basic background on RBM, DBN-DNN and ESN. Section 3 elaborates the architecture and learning algorithm of DBEN. Experiments and evaluation results on the benchmark datasets are given in Section 4 with respect to prediction accuracy and STM. We discuss the proposed DBEN in Section 5. Finally, this paper is concluded in Section 6.

2. Theoretical background

In this section, we briefly summarize the theoretical background of the considered models, namely RBM, hybrid DBN-DNN and ESN, which are the basis of the following DBEN understanding.

2.1. RBM

RBM [7,16] is an energy-based stochastic neural network composed by two parts, i.e., visible layer and hidden layer, in which the training process is done in an unsupervised mode. Similar to the classical Boltzmann machine, the visible layer and hidden layer are fully connected via symmetric undirected weights, except that there exist no intra-layer connections within either the visible or hidden layer. The architecture of a typical RBM model is depicted 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. DBN-DNN

The hybrid DBN-DNN [9] consists of both generative and discriminative model components corresponding to a traditional DBN and a LR layer, respectively, as shown in Fig. 2. Here, DBN is a probabilistic generative model with a stack of restricted Boltzmann machines (RBMs) [7,8,13–15], which can effectively model the structure in the input data (feature representation) [30]. In order to solve classification and regression problems, a LR layer is further added above DBN for supervised learning. Accordingly, the training of DBN-DNN includes two phrases: pre-training and fine-tuning.

In the first phrase, DBN is pre-trained in a greedy layer-wise unsupervised fashion, starting from the lowest RBM and using its outputs (after the current RBM training is completed) as the inputs for training the following RBM. Once the training of all RBMs is finished, the outputs of the final one is the learnt features in the pre-training procedure. It is widely agreed that the fast and effective pre-training is extremely beneficial to deep learning [9,30,31].

Then, in the second phrase, the BP algorithm is employed to fine-tune the whole pre-trained network to integrate the layers of neural networks by utilizing the learnt features. It can find a minimum in a peripheral region of parameters initialized by DBN [13]. Finally, the trained DBN-DNN can be used for classification and regression.

2.3. Echo state network

Echo state networks (ESN) establish an efficient and powerful approach to recurrent neural network (RNN) training. Unlike the traditional RNNs, such as Elman networks, that are organized in layers and contain feedback connections. The core part of ESN is a single reservoir consisting of a mass of neurons that are randomly interconnected and/or self-connected. The reservoir itself remains unchanged, once it is selected. The efficient learning can be achieved by determining the weights of the connections between the reservoir and the output layer. ESNs overcome the slow convergence combined with high computational requirements and suboptimal estimates of the model parameters, shown by RNN