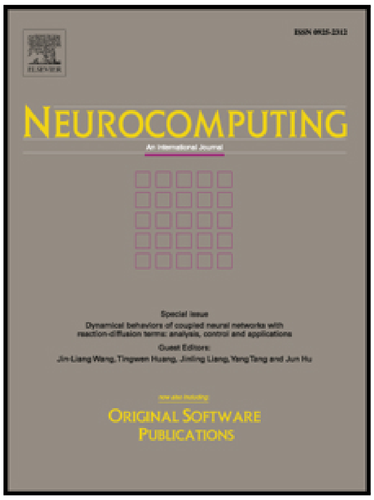

Balance Gate Controlled Deep Neural Network

Anji Liu , Yuanjun Laili

PII: S0925-2312(18)31056-7
DOI: https://doi.org/10.1016/j.neucom.2018.08.075
Reference: NEUCOM 19925

To appear in: *Neurocomputing*

Received date: 21 December 2017
Revised date: 13 April 2018
Accepted date: 10 August 2018

Balance Gate Controlled Deep Neural Network

Anji Liu, Yuanjun Laili*

***Abstract*—By mimicking the information expression ways in human's brain, deep neural network refers to a deeply structured framework with multiple processing layers that attempt to model high-level abstractions of the data. It is becoming a mainstream technology for pattern recognition, data mining and intelligent control at industrial scale. Previous work on constructing very deep networks, such as convolutional neural network and recurrent neural network, makes complex tasks such as image classification feasible. However, they are very limited when dealing with irregular data whose features are unstructured or even unknown. The traditional fully connected neural network is contrarily too shallow to extract high-level features. In this paper, we present a balance gate controlled deep network structure to deepen the fully connected neural network. It uses a new gating strategy to control information flow and increase network stability. Experimental results on both irregular regression and time-series forecasting demonstrate that the proposed network out-performs other ad-hoc models and is easier to train in a deeper form than the fully connected neural network.**

***Index Terms*—Deep Neural Network, Balanced Gate, Light-weight Training, Regression, Time-series Prediction.**

I. Introduction

DEEP neural network (DNN) is a trainable structure with multiple hidden layers of non-linear units between the input and output layers [1]. Unlike traditional shallow networks, it reforms the data features as a hierarchical representation and uses a group of low-dimensional and high-level features to make classification, prediction and recognition, etc. Due to its good performance on image/video/language processing, DNN has become a research hotspot in various domains and led to many breakthroughs. At present, there are mainly two types of successful DNN structures, i.e., convolutional neural network (CNN) [2] and recurrent neural network (RNN) [3].

Inspired by animal visual cortex, CNN uses convolution operation to process the spatial-consistent features among different regions and extract high-level features by widely parameter sharing. The same kernel can be applied to different regions of the data (e.g. different pixel blocks of an image). Convolutional layers are efficient in learning and extracting features from both images and soundwaves with much less training parameters. By applying such hierarchical visual cortex (which is simulated by multiple convolutional layers), CNN is becoming deeper and deeper and it learns low-level features in the first several layers and higher-level features afterward. It can efficiently deal with various signals [4, 5], especially image/video signals [6-11].

Different from CNNs, RNNs are designed to deal with time-series data. Some hidden layers in RNN have additional links to itself. This generates internal states in these layers to form a historical memory, which records dynamic temporal behavior. Salehinejad et al. [12] summarized the recent advances in RNN and reviewed different applications of it. Specifically, RNN is widely used in natural language processing (NLP) [13-15], speech/audio processing [16-18], and image/video processing [19-21].

Looking at the recent advances in deep learning [12, 22], most models are designed for regular data with certain topological or sequential structures. CNN and RNN adopt specific parameter sharing schemes for topological and sequential data respectively. As a result, they achieve state-of-the-art performance in certain tasks. However, when it comes to tasks that are not related to image or time-series data, limited DNN models can be adopted. Yet most of the industrial data is irregular with unexpected noises. Current research is more likely to integrate some de-noising network [23], dimensionality reduction network [24] and other feature extraction network combined with the fully connected neural network (FC net) to form a complex DNN to deal with the irregular data. In addition, regularization and other task-specific approaches such as feature selection and bagging are widely used. On one hand, the pre-processing networks are highly problem-dependent and cannot be extended to different circumstances. On the other hand, some critical information can be filtered out from the FC net due to regularization or feature selection and result in inaccurate learning and analysis. More importantly, the FC net is far from being fully exerted to make an accurate learning without under-fitting or over-fitting.

Although various breakthroughs have been made in terms of activation function [25], regularization rule [26], and initialization scheme [27] for the FC net, it still adopts consecutive fully connected layers as hidden layers. This simple structure greatly restricts its capability and thus makes it hard to learn high-level representations for common irregular data. Many theories and experiments

Anji Liu is with the School of Automation Science and Electrical Engineering, Beihang University, Beijing, 100191, P.R. China. (Email: liuanji@buaa.edu.cn)
Yuanjun Laili is with the School of Automation Science and Electrical Engineering, Beihang University, Beijing, 100191, P.R. China. (*Corresponding author is Yuanjun Laili, Email: llyj0721@gmail.com)