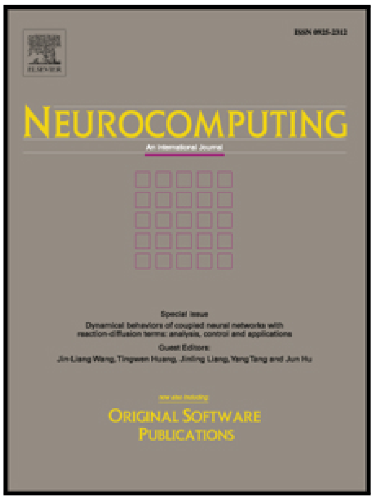
Accepted Manuscript



Deep hybrid neural-kernel networks using random Fourier features

Siamak Mehrkanoon, Johan A.K. Suykens

PII: S0925-2312(18)30210-8
DOI: 10.1016/j.neucom.2017.12.065
Reference: NEUCOM 19354

To appear in: *Neurocomputing*

Received date: 8 July 2017
Revised date: 2 November 2017
Accepted date: 27 December 2017

Please cite this article as: Siamak Mehrkanoon, Johan A.K. Suykens, Deep hybrid neural-kernel networks using random Fourier features, *Neurocomputing* (2018), doi: 10.1016/j.neucom.2017.12.065

Deep hybrid neural-kernel networks using random Fourier features

Siamak Mehrkanoon[1], and Johan A.K. Suykens

KU Leuven, ESAT-STADIUS, Kasteelpark Arenberg 10, B-3001 Leuven (Heverlee), Belgium

Abstract

This paper introduces a novel hybrid deep neural kernel framework. The proposed deep learning model makes a combination of a neural networks based architecture and a kernel based model. In particular, here an explicit feature map, based on random Fourier features, is used to make the transition between the two architectures more straightforward as well as making the model scalable to large datasets by solving the optimization problem in the primal. Furthermore, the introduced framework is considered as the first building block for the development of even deeper models and more advanced architectures. Experimental results show an improvement over shallow models and the standard non-hybrid neural networks architecture on several medium to large scale real-life datasets.

Keywords: Deep learning, neural networks, explicit feature mapping, kernel methods, hybrid models

1. Introduction

Conventional machine learning techniques were limited in processing natural data in their raw forms and a lot of domain experts were required in transforming raw data into meaningful features or representations. Recent years have witnessed considerable interests in models with deep architectures, inspired by the layered architecture of the human visual cortex, due to their successful impact in revolutionizing many application fields ranging from auditory to vision sensory signal processing such as computer vision, speech processing, natural language processing and game playing among others.

Deep Learning is a class of machine learning techniques that belongs to the family of representation learning models [1, 2]. Deep learning models deal with complex tasks by learning from subtasks. In particular, several nonlinear modules are stacked in hierarchical architectures to learn multiple levels of representation (hierarchical features) from the raw input data. Each module transforms the representation at one level into a slightly more abstract representation at a higher level, i.e. the higher-level features are defined in terms of lower-level ones. Deep learning architectures have grown significantly, resulting in different models such as stacked denoising autoencoders [3, 4], Restricted Boltzmann Machines [5, 6, 7], Convolutional Neural Networks [8, 9], Long Short Term Memories [10] among others.

Recent works in machine learning have highlighted the superiority of deep architectures over shallow architectures in terms of accuracy in several application domains [1, 11]. However, training deep neural networks involves costly nonlinear optimization problems and demands huge amount of labeled training data. The generalization performance of deep artificial neural networks largely depends on the parameters of the model of which they can be thousands to learn. Furthermore, finding the right architecture such as the number of layers and hidden units, the type of activation functions among others, as well as the networks associated hyper-parameters become a difficult task with increasing complexity of deep architectures. Most of the developed deep learning models are based on artificial neural networks (ANN) architecture, whereas deep kernel based models have not yet been explored in great detail. On the other hand support vector machines (SVM) and kernel based methods have also made a large impact in a wide range of application domains, with their strong foundations in optimization and learning theory [12, 13, 14] and are able to handle high-dimensional data directly.

Therefore, exploring the existing synergies or hybridization between ANN and Kernel based models can potentially lead to the development of models that have the best of two worlds. One has started already to explore such directions e.g. Kernel Methods for Deep Learning and a family of positive-definite kernel functions that mimic the computation in multilayer neural networks [15], Convolutional kernel networks [16], Deep Gaussian processes [17, 18]. In particular, the authors in [19] introduced a convex deep learning model via normalized kernels. The authors in [20] investigated iterated compositions of Gaussian kernels with an interpretation that resembles a deep neural networks architecture. A kernel based Convolutional Neural Network is introduced in [16] where new representations of the given image are obtained by stacking and composing kernels at different layers. A survey of recent attempts and motivations existing in the community for finding such a synergy between the two frameworks is also discussed in [21].

In this paper, we discuss possible strategies to bridge neural networks and kernel based models. The approach has been originally proposed in our previous work [22] where a two layer

[1]Corresponding author.
E-mail address: {*siamak.mehrkanoon,johan.suykens*}
@esat.kuleuven.be