Neurocomputing xx (xxxx) xxxx–xxxx

Contents lists available at ScienceDirect

Neurocomputing

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

Progressive Operational Perceptrons

Serkan Kiranyaz[a,*], Turker Ince[b], Alexandros Iosifidis[c], Moncef Gabbouj[c]

[a] Department of Electrical Engineering, Qatar University, Qatar
[b] Department of Electrical Engineering, Izmir University of Economics, Izmir, Turkey
[c] Tampere University of Technology, Department of Signal Processing, Tampere, Finland

ARTICLE INFO

Keywords:
Artificial neural networks
Multi-layer perceptrons
Progressive operational perceptrons
Diversity
Scalability

ABSTRACT

There are well-known limitations and drawbacks on the performance and robustness of the feed-forward, fully-connected Artificial Neural Networks (ANNs), or the so-called Multi-Layer Perceptrons (MLPs). In this study we shall address them by Generalized Operational Perceptrons (GOPs) that consist of neurons with distinct (non-)linear operators to achieve a generalized model of the biological neurons and ultimately a superior diversity. We modified the conventional back-propagation (BP) to train GOPs and furthermore, proposed Progressive Operational Perceptrons (POPs) to achieve self-organized and depth-adaptive GOPs according to the learning problem. The most crucial property of the POPs is their ability to simultaneously search for the optimal operator set and train each layer individually. The final POP is, therefore, formed layer by layer and in this paper we shall show that this ability enables POPs with minimal network depth to attack the most challenging learning problems that cannot be learned by conventional ANNs even with a deeper and significantly complex configuration. Experimental results show that POPs can scale up very well with the problem size and can have the potential to achieve a superior generalization performance on real benchmark problems with a significant gain.

1. Introduction

Learning in the broader sense can be in the form of classification, data regression, feature extraction and syntheses, or function approximation. For instance the objective for classification is finding out the right transformation of the input data (raw signal, data or feature vector) of each class to a distinct location in N-dimensional space that is far and well-separated from the others where N is the number of classes. Therefore, the main challenge in learning is to find out the right transformation (linear or nonlinear) or in general, the right set of consecutive transformations so as to accomplish the underlying learning objective. For this purpose most existing classifiers use only one or few (non-)linear operators. The most typical example is Support Vector Machines (SVMs) where one has to make the critical choice of the (non-)linear kernel function that will be used and subsequently define appropriate parameters. Even if one can optimize the performance of the classifier with respect to the kernel function's parameters, choosing an inappropriate kernel function can lead to far inferior performance, when compared to the performance that can be achieved by using the kernel function fitting to the characteristics of the problem at hand. Consider for instance, two sample feature transformations (FS-1 and FS-2) illustrated in Fig. 1 where for illustration purposes features are only shown in 1-D and 2-D, and only two-class problems are considered. In the case of FS-1, the SVM with a polynomial kernel in quadratic form would make the proper transformation into 3-D so that the new (transformed) features are linearly separable. However, for FS-2, a sinusoid with the right frequency, f, should be used instead. Therefore, especially in real and complex problems a high level of operational diversity, which can only enable the right (set of) transformations is of paramount importance.

In biological learning systems, this is addressed in the neurons at the cellular level. As shown in Fig. 2, in the mammalian nervous system, each neuron conducts the electrical signal over three distinct operations: 1) synaptic connections in *Dendrites*: an individual operation over each input signal from the synapse connection of the input neuron's axon terminals, 2) a pooling operation of the *operated* input signals via spatial and temporal signal integrator in the *Soma*, and finally, 3) an activation in the initial section of the *Axon* or the so-called *Axon hillock*: if the pooled potentials exceed a certain limit, it "activates" a series of pulses (called action potentials). As shown in the right side of the figure, each terminal button is connected to other neurons across a small gap called a *synapse*. The physical and neurochemical characteristics of each synapse determine the signal operation which is nonlinear in general [1,2] along with the signal

* Corresponding author.
E-mail addresses: mkiranyaz@qu.edu.qa (S. Kiranyaz), turker.ince@izmirekonomi.edu.tr (T. Ince), alexandros.iosifidis@tut.fi (A. Iosifidis), moncef.gabbouj@tut.fi (M. Gabbouj).

http://dx.doi.org/10.1016/j.neucom.2016.10.044
Received 9 August 2016; Received in revised form 27 October 2016; Accepted 29 October 2016
Available online xxxx
0925-2312/

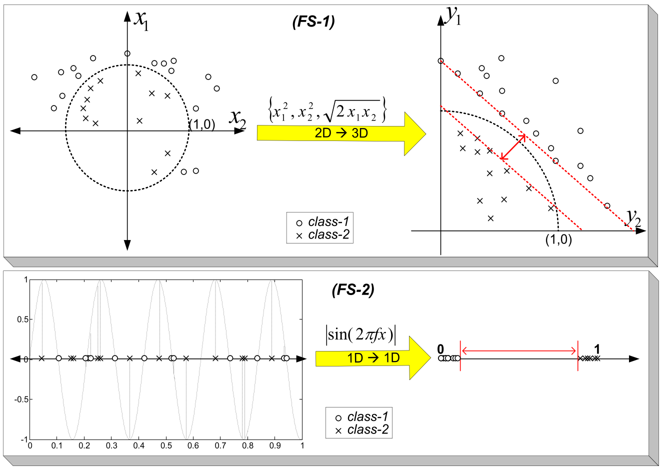

Fig. 1. Two sample feature synthesis performed on 2-D (FS-1) and 1-D (FS-2) feature spaces.

strength and polarity of the new input signal. Information storage or processing is concentrated in the cells' synaptic connections or more precisely through certain operations of these connections together with the connection strengths (i.e., weights) [1]. Such biological neurons or neural systems in general are built from a large diversity of neuron types varying entirely or partially structural, neurochemical and electrophysiological properties [4–9]. For instance in mammalian retina there are roughly 55 different types of neurons to perform the low-level visual sensing [7]. The functions of the 22 of them are already known and a cell defined as a "type" by structural criteria carries out a distinct and individual physiological function (operator). Accordingly in neurological systems, several distinct operations with proper weights (parameters) are created to accomplish such diversity and trained in time to perform or "to learn" many neural functions. Neural networks, both biological and artificial with higher diversity of computational operators have more computational powers [5,10–13] and it is also a fact that adding more neural diversity allows the network size and total connections to be reduced [9].

Conventional ANNs were designed to simulate biological neurons; however, at the best ANN models are based only *loosely* on biology. The most typical ANN neuron model is McCulloch-Pitts [14] which is mainly used in many feed-forward ANNs such as multi-layer perceptrons (MLPs) [33]. As in Eq. (1), in this formal model an artificial neuron performs linear summation scaled with the synaptic weights. So the synaptic connections with distinct neurochemical operations and the integration in the *Soma* are modelled *solely* as a linear transformation (i.e. the linear weighted sum) followed by a possibly nonlinear thresholding function, *f*(.), also called as the activation function.

$$x_k^l = b_k^l + \sum_{i=1}^{N_{l-1}} w_{ik}^{l-1} y_i^{l-1} \text{and } y_k^l = f(x_k^l) \tag{1}$$

It is obvious from Eq. (1) that this model is indeed a limited and crude model of the biological neurons and this is one of the reasons that render ANNs having a high variation on their learning and generalization performances in many problems [15,16]. There have been some attempts in the literature to modify MLPs by changing the neuron model and/or conventional BP algorithm [17–19], however, their performance improvements were not significant in general. Even though the network topology [20–22] or the parameter updates [24] were optimized according to the problem in hand, such approaches still inherit the main drawback of MLPs, i.e., they employ the conventional neuron model described in Eq. (1). This is also true for other ANN topologies such as recurrent neural networks, long short-term memory networks and convolutional neural networks [23]. Another well-known feed-forward and fully-connected ANNs are the Radial Basis Functions (RBFs) [11,12] which employ a set of RBFs each of which is embedded in a hidden neuron. The most typical RBF is Gaussian and, thanks to this nonlinear operator RBF networks promise a faster learning capability than the MLPs. However, they still suffer from the same major problem of incapability to approximate certain functions or discriminate certain patterns [25] unless (sometimes infeasibly) large network configuration is used because they use only one operator, the RBF, regardless of the problem in hand. This is also evident on the recent studies on MLPs [26–28]. Particularly [27] focuses on Deep and Shallow Architecture of Multilayer Neural Networks and their limitations. The two main questions that we want to answer in this paper are: *"Can we learn highly complex transformations by using small, minimal-depth network topologies? "* and *"How can we efficiently form networks that can adapt to both the nature and the complexity of the problem at hand with proper operators? "*

We believe that the answers for these questions are hidden in the nature of the basic processing units (i.e. neurons) used to build our learning model. In order to address the aforementioned drawbacks and especially accomplish a generalized model of biological neurons with superior operational diversity, in this paper we shall first present Generalized Operational Perceptrons (GOPs) that can encapsulate many linear and nonlinear operators. Contrary to MLPs, each neuron in a GOP can perform a distinct operation over its input signals. This mimics a biological neuron cell with a distinct neurochemical characteristics of its synaptic connections each with a certain strength (weight). A neuron (node) has only one operator and hence it is called the *nodal operator* that uses the same function but with a different parameter (weight) for each neuron connection from the previous layer. The outputs of the nodal operators will then be integrated with a *pooling operator*, which contrary to MLPs, can be any proper integrator besides summation. Finally, a similar flexibility is also allowed for the *activation operator* (function). Thus, each GOP neuron can have any operator set (*nodal*, *pool* and *activation*) where each operator is selected among a library of operators to maximize the learning performance.. Finding out the optimal operator set for each neuron is crucial for GOPs. In this study our primary objective is to design a minimal depth GOP with the least number of hidden layers while it can learn a complex problem with the desired accuracy. In order to achieve this we shall then propose the Progressive Operational Perceptrons (POPs). POPs are heterogeneous GOPs that are self-organized and depth-adaptive according to the learning problem. As the name implies they are created progressively, layer by layer, while

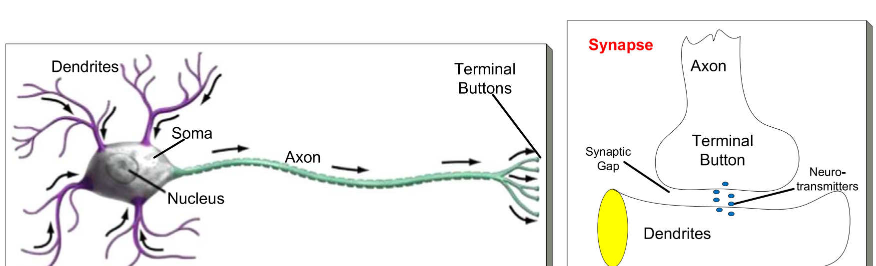

Fig. 2. A biological neuron (left) with the direction of the signal flow and a synapse (right).