Expert Systems with Applications 37 (2010) 4379–4391

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Evolutionarily optimized features in functional link neural network for classification

Satchidananda Dehuri *, Sung-Bae Cho

Soft Computing Laboratory, Department of Computer Science, Yonsei University, 262 Seongsanno, Seodaemun-gu, Seoul 120-749, Republic of Korea

ARTICLE INFO

Keywords:
Classification
Data mining
Functional link neural network
Genetic algorithms
Radial basis function network

ABSTRACT

In this paper, an adequate set of input features is selected for functional expansion genetically for the purpose of solving the problem of classification in data mining using functional link neural network. The proposed method named as HFLNN aims to choose an optimal subset of input features by eliminating features with little or no predictive information and designs a more compact classifier. With an adequate set of basis functions, HFLNN overcomes the non-linearity of problems, which is a common phenomenon in single layer neural networks. The properties like simplicity of the architecture (i.e., no hidden layer) and the low computational complexity of the network (i.e., less number of weights to be learned) encourage us to use it in classification task of data mining. We present a mathematical analysis of the stability and convergence of the proposed method. Further the issue of statistical tests for comparison of algorithms on multiple datasets, which is even more essential in data mining studies, has been all but ignored. In this paper, we recommend a set of simple, yet safe, robust and non-parametric tests for statistical comparisons of the HFLNN with functional link neural network (FLNN) and radial basis function network (RBFN) classifiers over multiple datasets by an extensive set of simulation studies.

1. Introduction

For the past two decades, there have been a lot of studies focused on the classification problem in the field of data mining (Fayyad, Piatetsky-Shapiro, & Smyth, 1996; Kriegel, 2007). The abstract goal of data mining is to discover novel and useful information in databases this is where consensus ends and the means of achieving this goal are as diverse as the communities contributing. Classification is one of the fundamental tasks of data mining. Neural networks (Zhang, 2000, 2007) have emerged as an important tool for classification. The recent vast research activities in neural classification have established that neural networks are a promising alternative to various conventional classification methods. The ANN's are capable of generating complex mapping between the input and the output space and thus these networks can form arbitrarily complex nonlinear decision boundaries. This does not mean that finding such type of networks is easy. On the contrary, problems such as local minima trapping, saturation, weight interference, initial weight dependence, and overfitting make neural network training difficult. Moreover, most neural learning methods, being based on gradient descent, cannot search the non-differentiable landscape of multi-layer architectures. This is the key; since it can be proved that if a network is allowed to adapt its architecture, it can solve any learnable problem in polynomial time.

Myriad combinations of evolutionary algorithms (EAs) and neural networks (NNs) used in classification problem have been proposed (Branke, 1995; Cant-Paz & Kamath, 2005; Yao, 1999). Genetic algorithms have been used to train the networks (Caudell & Dolan, 1989; Fogel, Fogel, & Porto, 1990; Kitano, 1990; Montana & Davis, 1989; Whitley & Hanson, 1989), design their architecture (Miller, Todd, & Hegde, 1989), and select feature subsets (Oh, Lee, & Moon, 2004; Seidlecki & Skalansky, 1989; Yang & Honavar, 1998). However, most of these combinations have been devoted to neural networks with hidden layer but in this paper our primary focus is on combining genetic algorithms with neural networks having no hidden layer. FLNNs (Pao & Philips, 1995; Pao & Takefuji, 1992) a class of higher order neural networks are a network of such type. FLNNs can capture non-linear input–output relationships, provided that they are fed with an adequate set of functional inputs. The primary interest of considering such type of neural networks is to reduce architectural complexity and to achieve faster learning by optimizing less number of weights parameters. Pao (1992) have given a pointer that FLNN may be conveniently used for function approximation and can be extended for pattern classification with faster convergence rate and lesser computational load than a multi-layer perceptron (MLP) structure.

Furthermore, selecting optimal set of features, as the input for functional expansion is another interesting and improvement in this direction. Selecting optimal set of features and fed as the input for the functional expansion has numerous advantages such as: (i)

* Corresponding author.
E-mail addresses: satchi.lapa@gmail.com (S. Dehuri), sbcho@cs.yonsei.ac.kr (S.-B. Cho).

0957-4174/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2009.11.090

reducing or removing the amplification of noisy/irrelevant features, (ii) minimizing the complexity of the architecture, (iii) minimizing the computational load of training the network, etc. Another important point is no matter how intelligent the FLNN is, it will fail to predict the unknown sample if it is applied to low quality data. Hence, to improve the capability of the FLNN for accurate classification and to make more insight in the nature of the problem we rely on the hybridization of genetic algorithms and FLNN. Moreover, genetic selection is taken up due to the intractable nature and plagued by host of local minima in the exponential search space. This method not only has practical time complexity but also achieves good performance.

Over the last years, the machine learning and data mining community has become increasingly alert the need for statistical validation of the results. This can be ascribed to the maturity of the area, increasing the number of real-life applications and the availability of open algorithmic frameworks that make it easy to develop new algorithms or modify the existing methods, and compare them among themselves. In this paper, we used a set of simple, yet safe, robust and non-parametric tests for statistical comparisons of newly proposed method with FLNN and RBFN classifiers over multiple datasets. The reason being chosen these two classifiers is that FLNN is flat net without feature selection and has demonstrated well in the classification task of data mining (Misra & Dehuri, 2007). Light (1992), Powell (1992) has already been proved that RBFN can approximate arbitrarily well any decision boundary if a sufficient number of radial basis function units are given.

The rest of this paper is organized as follows. In Section 2, we discuss the background materials and related works. Section 3 provides the HFLNN for classification with optimal set of features. The robustness of the proposed method is illustrated in Section 4 by a mathematical analysis based on the stability and convergence analysis. In Section 5 we have presented the experimental studies and a parametric and non-parametric statistical comparative performance with other classifiers such as RBFN and FLNN trained by back propagation learning. Section 6 concludes the paper.

2. Background and related work

Adaptation of GAs in NNs has already proven a sound theoretical and empirical results from three aspects such as: GAs for optimizing weights in NNs, feature/instance selection in NNs (the objective is to reduce the size and dimension of the training set), and to design the structure of the network. However, to the best of our knowledge not a single attempt has been made for genetically select the most relevant features for functional expansion which in turns can be fed as the input of the FLNN for classification.

In this section we will discuss the basic background material required for a clear notion of the proposed method and some of the related work.

2.1. Genetic algorithms based feature selection in NNs

GAs (Goldberg, 1989) are stochastic search algorithms characterized by the fact that a number P of potential solutions (called individuals $ch_i \in \aleph$, where $\aleph$ represents the space of all possible individuals) of the optimization problem simultaneously sample the search space. This population $\Omega = \{ch_1, ch_2, \ldots, ch_P\}$ is modified according to the natural evolutionary process: after *initialization*, *selection* $S{:}\Omega \to ch_i$ and recombination $R{:}\Omega \to ch_i$ are executed in a loop until some termination criterion is reached. Each run of the loop is called a generation and $\Omega(t)$ denotes the population at generation t.

The selection operator is intended to improve the average quality of the population by giving fitter individuals a higher probability to be copied into the next generation. Selection thereby focuses on the search of promising regions in the search space. The quality of an individual is measured by a fitness function $f : \Omega \to \Re$. Recombination and mutation change the genetic material in the population in order to obtain new points in the search space.

Besides searching for the weights topology determination, genetic algorithms may be used to select relevant features that are input to the NNs for classification. The training samples may contain features that are irrelevant, noisy or redundant, but it is not known a priori which features are relevant. Avoiding these features is desirable not only because they increase the size of the network and the training time, but also they may reduce the accuracy of the network. Further, among the different categories of feature selection algorithms the GAs is a rather recent development. GA based feature selection is very essential because of the following reasons. Suppose there are m features in the data being mined. Then the total number of candidate feature subsets is 2^m that are the size of search space of the feature selection grows exponentially with the number of features.

The pioneering work by Siedlecki and Skalansky (1988) demonstrated evidence for the superiority of GA compared to representative classical algorithms. Subsequently many literatures were published that have shown advantages of GAs for feature selection in NNs (Brill, Brown, & Martin, 1990; Brotherton & Simpson, 1995; Yang & Honavar, 1998; Ozdemir et al., 2001).

2.2. Functional link neural networks for classification

Classification in the context of data mining is learning a function f that maps (classifies) a data item (i.e., called facts F) into one of several predefined classes. In real-life learning a function means that it has the capability to generate non-linear decision boundaries. It is known that using a number of hyper-planes one can approximate any nonlinear surface. Hence, the problem of classification can be viewed as searching for a number of linear surfaces that can appropriately model the class boundaries while providing minimum number of misclassified data points. Multiple layer neural networks can mostly be used for solving complex classification problems. However, depending on the complexities of the problems, the number of layers and number of neurons in the hidden layer need to be changed. As the number of layers and the number of neurons in the hidden layer increases, training the model becomes further complex. Very often different algorithms fail to train the model for a given problem set.

To overcome the complexities associated with multi-layer neural network, a single layer neural network can be considered as an alternative approach. But the single layer neural network being linear in nature very often fails to solve the complex nonlinear problems. The classification task in data mining is highly nonlinear in nature. Therefore for solving such problems in single layer feed forward artificial neural network is almost an impossible task.

In order to bridge the gap between the linearity in the single layer neural network and the highly complex and computationally intensive multi-layer neural networks, the FLNN architecture with back propagation learning for classifications was proposed (Misra & Dehuri, 2007). In contrast to the linear weighting of the input pattern produced by the linear links of artificial neural network, the functional link acts on an element of a pattern or on the entire pattern itself by generating a set of linearly independent functions, then evaluating these functions with the pattern as the argument. Thus class separability is possible in the enhanced feature space. The FLNN obtains the solution for W iteratively using BP algorithm based on all the training samples.

Learning of a FLNN may be considered as approximating or interpolating a continuous multivariate function $\phi(X)$ by an approximating function $\phi_W(X)$. In FLNN architecture, a set of basis