Neurocomputing 81 (2012) 86–96

Contents lists available at SciVerse ScienceDirect

Neurocomputing

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

Meta-cognitive Neural Network for classification problems in a sequential learning framework

G. Sateesh Babu, S. Suresh *

School of Computer Engineering, Nanyang Technological University, Singapore

A R T I C L E I N F O

Article history:
Received 29 June 2011
Received in revised form
7 October 2011
Accepted 13 December 2011
Communicated by G.-B. Huang
Available online 24 December 2011

Keywords:
Meta-cognitive learning
Self-regulatory thresholds
Radial basis function network
Multi-category classification problem
Sequential learning

A B S T R A C T

In this paper, we propose a sequential learning algorithm for a neural network classifier based on human meta-cognitive learning principles. The network, referred to as Meta-cognitive Neural Network (McNN). McNN has two components, namely the cognitive component and the meta-cognitive component. A radial basis function network is the fundamental building block of the cognitive component. The meta-cognitive component controls the learning process in the cognitive component by deciding *what-to-learn*, *when-to-learn* and *how-to-learn*. When a sample is presented at the cognitive component of McNN, the meta-cognitive component chooses the best learning strategy for the sample using estimated class label, maximum hinge error, confidence of classifier and class-wise significance. Also sample overlapping conditions are considered in growth strategy for proper initialization of new hidden neurons. The performance of McNN classifier is evaluated using a set of benchmark classification problems from the UCI machine learning repository and two practical problems, viz., the acoustic emission for signal classification and a mammogram data set for cancer classification. The statistical comparison clearly indicates the superior performance of McNN over reported results in the literature.
© 2011 Elsevier B.V. All rights reserved.

1. Introduction

In supervised learning, artificial neural networks became powerful tools to model complex input–output relationships by learning. Hence they are being increasingly employed to solve classification tasks to learn the decision surface that maps set of input features to class labels [1,2]. In most of the practical applications especially in medical diagnosis, the complete training data describing the input–output relationship is not available a priori. For these problems, classical batch-learning algorithms are rather infeasible and instead sequential learning is employed [3].

In a sequential learning framework, the training samples arrive one-by-one and the samples are discarded after the learning process. Hence, it requires less memory and computational time during the learning process. In addition, sequential learning algorithms automatically determine the minimal architecture that can accurately approximate the true decision function described by a stream of training samples. Radial basis function networks have been extensively used in a sequential learning framework due to its universal approximation ability and simplicity of architecture. Many sequential learning algorithms in radial basis function framework are available in the literature to solve function approximation or classification problems [4–10].

Resource Allocation Network (RAN) [4] was the one of the first sequential learning algorithm introduced in the literature. RAN evolves the network architecture required to approximate the true function using novelty based neuron growth criterion. The Minimal Resource Allocation Network (MRAN) [5] and the Extended Minimal Resource Allocation Network (EMRAN) [6] use a similar approach, but these algorithms incorporate error based neuron growing/pruning criterion. Hence, they determine compact network architecture than RAN algorithm. In Growing and Pruning Radial Basis Function Network (GAP-RBFN) [7], growing/pruning criteria of the network is selected based on the significance of a neuron. Aforementioned algorithms were developed for the function approximation problems, and may not work well for classification problems [9]. On the other hand, in Sequential Multi-Category Radial Basis Function Network (SMC-RBFN) [9], the similarity measures within class, misclassification rate and prediction error are used in neuron growing and parameter update criterion. It has been shown in [9] that updating the nearest neuron parameters in the same class as that of current sample helps in improving the performance than updating a nearest neuron in any class.

Another well known paradigm in fast learning neural network is extreme learning machine (ELM) [11]. It is a batch learning algorithm for a single-hidden layer feed forward neural network.

* Corresponding author. Tel.: +65 6790 6185.
E-mail address: ssundaram@ntu.edu.sg (S. Suresh).

0925-2312/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2011.12.001

ELM chooses input weights randomly and analytically determines the output weights using minimum norm least-squares. It has been extended to sequential framework using recursive least squares in [8]. In case of sparse and imbalance data sets, the random selection of input weights in the ELM and its variants affects the performance significantly [12]. A complete survey of research works in ELM framework are presented in [13].

Another widely used algorithm to solve classification problems is support vector machines (SVM). A sequential learning version in support vector machine framework is called an incremental and decremental support vector machine learning was presented in [14]. It uses an on-line recursive algorithm for training support vector machines, and it handles one sample at a time by retaining Karush–Kuhn–Tucker conditions on all previously seen training data. Recently in [15], a multiple incremental decremental learning of support vector machines was proposed. Here, multiple samples are added or removed simultaneously and is faster than the conventional incremental decremental support vector machines presented in [14].

Aforementioned sequential learning algorithms in radial basis function network, ELM and SVM frameworks address *how-to-learn* the decision function from stream of training samples. In Self-adaptive Resource Allocation Network (SRAN) [10] significant samples are selected using misclassification error and hinge loss function. It has been shown in [10] that the selection of appropriate samples by removing repetitive samples helps in achieving better generalization performance. Therefore, an efficient classifier must also be capable of judging what samples to learn and when to learn, during the training process. Hence, there is a need to develop a learning algorithm which automatically selects appropriate samples for learning and adopt best learning strategy to learn them accurately. Similar observations in complex domain are also reported in the literature [16,17].

Recent studies in human learning suggested that the learning process is effective when the learners adopt self-regulation in learning process using meta-cognition [18]. The term *meta-cognition* is defined in [19] as 'one's knowledge concerning one's own cognitive processes or anything related to them'. Precisely the learner should control the learning process, by planning and selecting learning strategies and monitor their progress by analyzing the effectiveness of the proposed learning strategies. When necessary, these strategies should be adapted appropriately. Meta-cognition present in human-being provides a means to address *what-to-learn*, *when-to-learn* and *how-to-learn*, i.e., the ability to identify the specific piece of required knowledge, judge when to start and stop learning by emphasizing best learning strategy. Hence, there is a need to develop a Meta-cognitive Neural Network classifier that is capable of deciding *what-to-learn*, *when-to-learn* and *how-to-learn* the decision function from the training data.

In this paper, we introduce a Meta-cognitive Neural Network (McNN) classifier which employs human-like meta-cognition to regulate the sequential learning process. McNN has two components namely the cognitive component and the meta-cognitive component. The cognitive component of McNN is a single hidden layer radial basis function network. The cognitive component adds neurons and updates the parameters of the network so as to approximate the decision surface described by the stream of training data. The meta-cognitive component of McNN measures the knowledge contained in the current training sample with respect to the cognitive component using estimated class label, maximum hinge error, confidence of classifier and class-wise significance. Class-wise significance is obtained from spherical potential, which is used widely in kernel methods to determine whether all the data points are enclosed tightly by the Gaussian kernels [20]. Here, the squared distance between the current sample and the hyper-dimensional projection helps in measuring the novelty in the data. Since, McNN address classification problems, we redefine the spherical potential in class-wise and is used in devising the learning strategy. In addition, the meta-cognitive component identifies the overlapping/non-overlapping criterion by measuring the distance from nearest neuron in the inter/intra-class. Using the above-mentioned measures the meta-cognitive component constructs two sample based learning strategies and two neuron based learning strategies. One of these strategies is selected for the current training sample such that the cognitive component learn them accurately and achieves better generalization performance.

Performance of the proposed McNN classifier is evaluated using set of benchmark binary/multi-category classification problems from the University of California, Irvine (UCI) machine learning repository [21]. For this purpose, we consider five multi-category classification problems and five binary classification problems with varying values of imbalance factor. The training sample imbalance between different classes are measured using imbalance factor [9]. The performance of McNN classifier on these benchmark data sets are compared with the existing sequential learning algorithms in the literature using class-wise performance measures like overall/average efficiency and a non-parametric statistical significance test [22]. The non-parametric Friedman test based on the mean ranking of each algorithm over multiple data sets [22] indicate the statistical significance of the proposed McNN classifier. Finally, the performance of McNN classifier has also been evaluated using two practical classification problems viz., the acoustic emission signal classification problem [23] and the mammogram classification problem for breast cancer detection [24]. The results clearly highlight that McNN classifier provides a better generalization performance than the results reported in the literature.

This paper is organized as follows: Section 2 describes the proposed McNN classifier. Section 3 presents performance evaluation of the proposed classifier for benchmark multi-category problems, binary classification problems and practical applications. Section 4 summarizes the conclusions from this study.

2. Meta-cognitive Neural Network (McNN) classifier

In this section, we describe the Meta-cognitive Neural Network classifier for solving multi-category classification problems. First, we define the classification problem. Next, we present the Meta-cognitive Neural Network architecture. Finally, we present the learning algorithm which addresses the three fundamental principles of meta-cognition in the learning process.

2.1. Problem definition

Let us consider the sequence of training samples $\{(\mathbf{x}^1,c^1),\ldots,(\mathbf{x}^i,c^i),\ldots,(\mathbf{x}^N,c^N)\}$, where $\mathbf{x}^i=[x_1^i,\ldots,x_m^i]^T\in\Re^m$ be the m-dimensional feature vector of ith training sample and c^i be its class label. Note, N is the total training samples available in the data stream. In a standard classification problem, the class label (c^i) of the sample i is converted into a coded class label $(\mathbf{y}^i=[y_1^i,\ldots,y_j^i,\ldots,y_n^i]^T)$ as

$$y_j^i=\begin{cases}1 & \text{if } j=c^i,\\ -1 & \text{otherwise,}\end{cases}\quad j=1,2,\ldots,n \tag{1}$$

where n is the number of distinct class labels present in the training samples.

The goal is to determine a suitable discriminant function $f(\mathbf{x}):\mathbf{x}\in\Re^m\to\mathbf{y}\in\Re^n$, from a class of functions $\mathbb{F}$, such that $f(\mathbf{x})$ accurately predicts the class labels with a certain degree of confidence. In this paper, we employ a meta-cognitive neural network as a classifier. The learning algorithm finds an

Download English Version:

https://daneshyari.com/en/article/410078

Download Persian Version:

https://daneshyari.com/article/410078

Daneshyari.com