Neurocomputing 128 (2014) 22–30

Contents lists available at ScienceDirect

Neurocomputing

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

Dissimilarity based ensemble of extreme learning machine for gene expression data classification ☆

Hui-juan Lu [a,b,*], Chun-lin An [a], En-hui Zheng [c], Yi Lu [d]

[a] College of Information Engineering, China Jiliang University, Hangzhou 310018, China
[b] School of Information and Electric Engineering, China University of Mining and Technology, Xuzhou 221008, China
[c] College of Mechanical and Electric Engineering, China Jiliang University, Hangzhou 310018, China
[d] Department of Computer Science, Prairie View A&M University, Prairie View 77446, USA

A R T I C L E I N F O

Article history:
Received 18 September 2012
Received in revised form
4 February 2013
Accepted 11 February 2013
Available online 8 November 2013

Keywords:
Extreme learning machine
Dissimilarity ensemble
Double-fault measure
Majority voting
Gene expression data

A B S T R A C T

Extreme learning machine (ELM) has salient features such as fast learning speed and excellent generalization performance. However, a single extreme learning machine is unstable in data classification. To overcome this drawback, more and more researchers consider using ensemble of ELMs. This paper proposes a method integrating voting-based extreme learning machines (V-ELMs) with dissimilarity (D-ELM). First, based on different dissimilarity measures, we remove a number of ELMs from the ensemble pool. Then, the remaining ELMs are grouped as an ensemble classifier by majority voting. Finally we use disagreement measure and double-fault measure to validate the D-ELM. The theoretical analysis and experimental results on gene expression data demonstrate that (1) the D-ELM can achieve better classification accuracy with less number of ELMs; (2) the double-fault measure based D-ELM (DF-D-ELM) performs better than disagreement measure based D-ELM (D-D-ELM).

1. Introduction

Human genome project (HGP) was officially launched in 1990. In the short span of 20 years, gene technology obtained rapid development. Golub et al. [1] were the first to use gene chips to study the human acute leukemia, and found two subtypes of acute lymphoblastic leukemia: T2Cell ALL and B2Cell ALL. The classification methods that were used on gene expression data early include the support vector machine (SVM) [2], artificial neural networks (ANNs) [3], and probabilistic neural network (PNN) [4]. Jin et al. [5] used the partial least squares method to establish a classification model. Zhang et al. [6] applied non-negative matrix factorization (NMF) for the gene expression data classification. Yang et al. [7] used a binary decision tree to classify gene expression data of tumor.

The extreme learning machine (ELM) [8] was proposed as an efficient learning algorithm for single-hidden layer feedforward neural networks (SLFNs). It increases learning speed by means of randomly generating weights and biases for hidden nodes rather than iteratively adjusting network parameters which is commonly adopted by gradient based methods.

However, the stability of single ELM can be improved. To achieve better generalization performance, Lan et al. [9] proposed an ensemble of online sequential extreme learning machine (EOS-ELM) which is more stable and accurate than the original OS-ELM.

Motivated by the ensemble idea, in 2009 van Heeswijk et al. [10] proposed an adaptive ensemble model of ELM which is adaptive and has low computational cost. In 2010, Tian and Meng proposed a bagging ensemble scheme to combine ELMs [11], and another ELM ensemble method based on the modified AdaBoost.RT algorithm [12]. In the same year, an ensemble based ELM (EN-ELM) algorithm was proposed by Liu and Wang [13] which uses the cross-validation scheme to create an ensemble of ELM classifiers for decision making. Wang and Li [14] proposed a dynamic Adaboost ensemble ELM which has been successfully applied to problem of function approximation and classification application. Zhai et al. [15] proposed a dynamic ensemble of sample entropy based extreme learning machines, which can alleviate some extent of instability and over-fitting problem, and increase the prediction accuracy. In 2011, Heeswijk et al. [16] proposed a method which is based on GPU-accelerated and parallelized ELM ensemble, and is used in large-scale regression. In 2012, Wang and Alhamdoosh [17] proposed an algorithm which employs the model diversity as a fitness function to direct the selection of base learners, and produces an optimal solution with ensemble size control. It improved the generalization

☆This work was supported by the National Natural Science Foundation of China (Nos. 61272315, 60842009, and 60905034), Zhejiang Provincial Natural Science Foundation (Nos. Y1110342, Y1080950) and the Pao Yu-Kong and Pao Zhao-Long Scholarship for Chinese Students Studying Abroad.

* Corresponding author at: College of Information Engineering, China Jiliang University, Hangzhou 310018, China. Tel.: +86 57186914580;
fax: +86 57186914573.
E-mail addresses: hjlu@cjlu.edu.cn, huijuanlu29@gmail.com (H.-j. Lu).

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.02.052

power. Cao et al. [18] proposed an improved learning algorithm for classification that is referred to as the voting based extreme learning machine (V-ELM) which is adopted widely.

The ensemble classifiers have already been used in gene expression data classification. Chen et al. [19] used an artificial neural network ensemble to classify the gene expression data. Liao and Li [20] established a gene expression data classifier by using an ensemble of SVM. However, the classification accuracy has not reached the level expected.

In this paper, we propose a data classification method based on the V-ELM with dissimilarity (D-ELM). First, we calculate the dissimilarity of extreme learning machines with a dissimilarity measure, and remove ELMs based on the dissimilarity. Then, we group the remaining ELMs through majority voting. We use two dissimilarity measures to validate the D-ELM.

This paper is organized as follows. In Section 2, we briefly review the basic concept of V-ELM. In Section 3, we introduce the dissimilarity ensemble method and two dissimilarity measures. In Section 4, we first integrate the dissimilarity ensemble method with V-ELM to reduce the redundancy of classifiers, and present the D-ELM. Then we introduce how to use two dissimilarity measures with ELM. Finally, we prove the effectiveness of the D-ELM from mathematics point of view. Simulation results and comparisons are provided in Section 5. Discussions on the performance of D-ELM are given in Section 6. The experiments demonstrate better classification accuracy of D-ELM with less number of ELMs. Between the two D-ELMs with different dissimilarity measures, the double-fault measure based D-ELM (DF-D-ELM) performs better than disagreement measure based D-ELM (D-D-ELM).

2. Dissimilarity measure

The dissimilarity is a very important research topic in an ensemble system. To make ensemble practically significant, the ELMs that we use must have difference from each other. Otherwise the results of ELMs will be the same, and the ensemble will be meaningless. It simply just increases the time complexity and space complexity without any benefit. We want to design an ELM with large dissimilarity and strong generalization performance, because this is the key to successfully build an ensemble system of multiple ELMs. However, what standard should we take to measure the dissimilarity among the ELMs? What is the effect of dissimilarity? These are unresolved issues [21].

In real life, we cannot design a perfect ELM that meets all requirements. We just expect to get a better result from a certain aspect. For a specific sample, if the decisions from a certain number of ELMs are wrong, but the decisions from majority ELMs are correct, we will very likely obtain a correct decision by a dissimilarity ensemble of ELMs. However, if the ELMs we used have no dissimilarity, and decisions obtained from the ELMs are wrong, we will definitely have a wrong result [22].

An effective dissimilarity measure can play an important role in dissimilarity ensemble systems. So far, a lot of methods has been proposed to estimate the dissimilarity among the classifiers qualitatively or quantitatively. Most of the measures are based on the output of classifiers. In the next section, we will introduce two dissimilarity measures. Let us use the following notation for discussion in the next section: the number of training samples is M; the number of classifiers is N; the output of the i-th classifier for the k-th sample is f_{ik}, where $i \in [1, \ldots, N]$, $k \in [1, \ldots, M]$.

2.1. Disagreement measure

Disagreement measure is based on the outputs of classifiers [23]. This measure is proportional to the difference of classifiers' outputs. We set the dissimilarity of two classifiers as $D_{i,j}$, where $j \in [1, \ldots, N]$. This measure can be calculated as follows:

$$D_{i,j} = \sum_{k=1}^{M} d(f_{ik}, f_{jk}) \tag{1}$$

where $d(f_{ik}, f_{jk})$ is the dissimilarity of the two classifiers on the k-th sample. If the outputs of the classifiers are the same, then $d(f_{ik}, f_{jk}) = 0$, else $d(f_{ik}, f_{jk}) = 1$. $D_{i,j}$ is proportional to the dissimilarity between f_i and f_j.

2.2. Double-fault measure

The double-fault measure is also based on the outputs of classifiers. In this method, it is more important to find that both outputs are wrong than both outputs are correct or one of the outputs is wrong. The first case means that we will have wrong result when both outputs are wrong [24]. This measure can be calculated as follows:

$$DF_{i,j} = \sum_{k=1}^{M} df(f_{ik}, f_{jk}) \tag{2}$$

where $df(f_{ik}, f_{jk}) = 0$ when both outputs are wrong at the same time, else $df(f_{ik}, f_{jk}) = 1$. $DF_{i,j}$ is proportional to the dissimilarity between f_i and f_j.

Both measures are based on the dissimilarity of the classifiers' outputs. In order to measure the dissimilarity of an ensemble system, we must sum all outputs of the classifiers.

Besides the above-mentioned measures, researchers also proposed some other measures based on overall classifier, e.g. entropy based measure [25], Kohavi–Wolpert variable [26] and difficult measure [27], etc. These measures, which are based on the global system, can be used to calculate the dissimilarity of an ensemble system and estimate the dissimilarity from different perspectives.

3. Voting based extreme learning machine

In this section, we first review the basic concept of the V-ELM algorithm in Section 3.1. Then, we analyze an issue that exists in V-ELM in Section 3.2.

3.1. Review of V-ELM

V-ELM was proposed by Cao et al. [18] in 2011. For a training set $M' = \{(x_k, t_k) | x_k \in R^d, t_k \in R^m\}_{k=1}^{M'}$, we use N basic ELMs with the same number, L of hidden nodes and same activation function $G(a, b, x)$. We randomly assign learning parameters (a_l^i, b_l^i) to each ELM, where $l = 1, 2, \ldots, L$, $i = 1, 2, \ldots, N$. Then we calculate the hidden layer output matrix H^i and the output weight $\beta^i : \beta^i = (H^i)^+ T$, where T is the target output matrix.

Using each trained basic ELM with learning parameters $(a_k^i, b_k^i, \beta_k^i)$, we can predict the label of the testing sample x^{test} and record it as label c, where $c \in [1, 2, \ldots, C]$. By accumulating the vote of $x^{test} S_{N,x^{test}}(c)$, we have the final class label of the testing sample x^{test} as

$$c^{test} = \arg \max_{c \in [1, \ldots, C]} \{S_{N,x^{test}}(c)\}.$$

3.2. An issue with V-ELM

V-ELM already has complete theoretical and practical foundation, and experiments have proven the validity in classification applications. However, we observe that some of the ELMs are similar to each other, which indicates the ensemble system including some redundant ELMs. So if we can remove the