Available online at www.sciencedirect.com

SCIENCE DIRECT®

Applied Soft Computing 6 (2005) 100–107

Applied Soft Computing

www.elsevier.com/locate/asoc

Stopping criteria for ensemble of evolutionary artificial neural networks

Minh Ha Nguyen *, Hussein A. Abbass, Robert I. McKay

Artificial Life and Adaptive Robotics (A.L.A.R.) Laboratory, School of Information Technology and Electrical Engineering, Australian Defence Force Academy, University of New South Wales, Canberra, ACT 2600, Australia

Received 3 May 2004; received in revised form 8 November 2004; accepted 13 December 2004

Abstract

The formation of ensemble of artificial neural networks has attracted attentions of researchers in the machine learning and statistical inference domains. It has been shown that combining different neural networks could improve the generalization ability of the learning machine. One challenge is when to stop the training or evolution of the neural networks to avoid overfitting. In this paper, we show that different early stopping criteria based on (i) the minimum validation fitness of the ensemble, and (ii) the minimum of the average population validation fitness could generalize better than the survival population in the last generation. The proposition was tested on four different ensemble methods: (i) a simple ensemble method, where each individual of the population (created and maintained by the evolutionary process) is used as a committee member, (ii) ensemble with island model as a diversity promotion mechanism, (iii) a recent successful ensemble method namely ensemble with negative correlation learning and (iv) an ensemble formed by applying multi-objective optimization. The experimental results suggested that using minimum validation fitness of the ensemble as an early stopping criterion is beneficial.
© 2005 Elsevier B.V. All rights reserved.

Keywords: Ensemble; Artificial neural network; Machine learning; Early stopping; Negative correlation learning; Island model; MOP; Evolutionary computation; $(\mu + \lambda)$ evolutionary strategies

1. Introduction

Evolutionary artificial neural networks (EANNs) have been widely studied in the last decade. The main power of neural network lies in its ability to correctly learn the underlined function or distribution from a sample. This ability is called generalization. Mathematically, the generalization ability could be expressed in term of minimizing the recognition error of the neural network. Thus, evolutionary computation (EC), a powerful tool for global optimization, may be employed to solve this optimization problem to find the optimum network. According to the prominent review of Yao [28], evolutionary methods could be applied on different levels of ANN, such as the architecture and connection weights.

* Corresponding author.
E-mail addresses: z2280951@student.adfa.edu.au, minhha_76@yahoo.com (M.H. Nguyen), h.abbass@adfa.edu.au (H.A. Abbass), r.mckay@adfa.edu.au (R.I. McKay).

1568-4946/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.asoc.2004.12.005

Most of the works in the ANN literature concentrate on finding a single optimum solution to solve a problem. However, an optimum network on the training data may not be the best-generalized network. The problem is that the ANN could be either overfitted (memorizing the data rather than learning the correct distribution) or under-trained (have not been trained enough or too simple to generalize well) (see [10] on bias/variance dilemma and generalization problem). Recently it has been found that an ensemble of neural networks could perform better than individual ones [23,25,13,16–18,29–31]. The main argument of ANN ensemble is that different members of the ensemble possess different bias/variance trade-offs, hence, a suitable combination of these biases/variances could result in an improvement in the generalization ability of the whole ensemble [25,31]. Islam et al. [11] constructed an ensemble of ANNs using the cascading method together with partial training and early stopping criteria. The reported results showed an improvement in the testing accuracy over individual networks. In 2003, Islam et al. [12] presented another ensemble constructive method which combined ensemble architecture design with cooperative training for individual neural networks (NNs) in ensembles. Early stopping was used at various stages such as in the training of the individual networks or determining the size of the networks (in terms of the hidden nodes) and the size of the ensemble (in terms of the number of networks in the ensemble). Their method produced NN ensembles with good generalization ability. Cho et al. [8] used speciation together with fitness-sharing-evolution and BP to construct an ensemble that also performs better than the individual ones. Cho and Ann [7] evolve an ensemble of diverse ANNs using fitness sharing. Rosen [23] defined a decorrelation penalty term to be included in the Mean Square Error function of the member of the neural network during training to differentiate the members of the ensemble. Similarly, Yao and Liu [29–31,18] used evolutionary programming and negative correlation to construct the ensemble of neural networks that are negatively correlated to one another. Chandra and Yao [5] attempted to use negative correlation together with multi-objective optimization to construct optimized ensembles with good accuracy and high diversity. McKay and Abbass [19] applied another anti-correlation term and mathematically analysed the effect of these anti-correlation measures on differentiating the networks within the ensemble from their means. Zhou et al. proposed a method called GASEN [32] and later on e-GASEN [27] which used genetic algorithm (GA) to find a set of weights of the ensemble (a weight in this context means how much a member network contributes to the overall ensemble) and selected those networks associated with weights greater than some threshold to include. Opitz and Sharlik [20] used GA to search for an ensemble which generalizes well and disagrees as much as possible.

The aim of generalization could be translated to minimizing the testing errors. However, we have only the training data to train the neural network and the criteria for training is to minimize the training error. As discussed above, a neural network that performs well on the training set can perform badly on the testing data. Thus, it is also desirable to stop the training phase in the right moment before overfitting happens. Nevertheless, most works in the literature simply use the last generation as the solution. The questions are how many generations are considered enough? Does the global optimum found on the training data (the survival individuals in the last generation) is the best-generalized solution? In this paper, we will show that using the population in the last generation as the ensemble is not always the best solution.

2. Methods

2.1. Artificial neural networks

The following notations are used for a single hidden layer ANN:

- I and H are the number of input and hidden units, respectively.
- $\vec{X}^p \in \mathbf{X} = (x_1^p, x_2^p, \dots, x_I^p)$, $p = 1, \dots, P$, is the p th pattern in the input feature space $\mathbf{X}$ of dimension I, and P is the total number of patterns.
- Without any loss of generality, $\mathbf{Y}_o^p \in \mathbf{Y}_o$ is the corresponding scalar of pattern $\vec{X}^p$ in the hypothesis space $\mathbf{Y}_o$.
- w_{ih} and w_{ho}, are the weights connecting input unit i, $i = 1, \dots, I$, to hidden unit h, $h = 1, \dots, H$, and

Download English Version:

https://daneshyari.com/en/article/10349235

Download Persian Version:

https://daneshyari.com/article/10349235

Daneshyari.com