Journal of Natural Gas Science and Engineering 21 (2014) 26–39

Contents lists available at ScienceDirect

Journal of Natural Gas Science and Engineering

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

Design of an ensemble neural network to improve the identification performance of a gas sweetening plant using the negative correlation learning and genetic algorithm

Javad Sadeghi Azizkhani [a], Hooshang Jazayeri-Rad [a,*], Nader Nabhani [b]

[a] Department of Automation and Instrumentation, Petroleum University of Technology, Ahwaz, Iran
[b] Department of Mechanical Engineering, Petroleum University of Technology, Ahwaz, Iran

ARTICLE INFO

Article history:
Received 8 May 2014
Received in revised form
10 July 2014
Accepted 12 July 2014
Available online

Keywords:
Ensemble neural network
Negative correlation learning
Genetic Algorithm
Gas sweetening
H_2S content
Alkanoamine

ABSTRACT

This paper presents a combination of negative correlation learning (NCL) and Genetic Algorithm (GA) to create an ensemble neural network (ENN). In this approach the component neural networks (CNNs) of ENN are trained simultaneously. The resulting CNNs negatively correlate together through the penalty terms in their objective functions. The predicted output is obtained by using the weighted averaging of the outputs of CNNs. GA participates in the training of CNNs and assigns proper weights to each trained CNN in the ensemble. The proposed method was tested on a case study in the Gas Treatment Plant (GTP) of the AMMAK project in the Ahwaz onshore field in Iran. The testing results of the model properly follow the experimental data. In addition, the proposed method outperformed the single neural network and some other network ensemble techniques.

1. Introduction

Today system identification is one of the most important fields of research in engineering sciences. The identification is a process of deriving a model of a physical system by using a series of experimental data. One of the most recent and authoritative techniques for identification problems is Artificial Neural Network (ANN). Main reasons of popularity of this approach are simplifying the process modeling and enabling implementation of inclusive tools for control system design. However, most chemical processes are complicated and a Single Neural Network (SNN) cannot identify such a complex system accurately. To overcome this limitation, improve the performance of ANN and reduce the model output variance, Hansen and Salamon (1990) proposed ENN which is a combination of a series of CNNs. The key point in applying these individual networks is their dissimilarity and diversity. In other word, there is no advantage in aggregation of networks which are all identical (Cunningham et al., 2000). There are many ensemble techniques developed by researchers, most of these techniques follow two steps: first generating individual networks, and then combining them (Sharkey, 1996). In such techniques, CNNs are learned independently. Two disadvantages of these approaches are loss of interaction between independent CNNs during training (Liu and Yao, 1999) and possibility of the lack of proper distribution of independent networks throughout the ensemble. Two of the most popular methods of this approach of ENN construction are Bagging (*B*ootstrap *Agg*regating) proposed by Breiman (1996) and Boosting proposed by Freund and Schapire (1996). Another recent pioneer technique in ENN employs the NCL method. NCL is different from previous works which trained the individual networks independently or sequentially (Drucker et al., 1994). Rather than producing unbiased individual networks whose errors are uncorrelated, NCL can create negatively correlated networks to encourage specialization and cooperation among the individual networks (Liu and Yao, 1999). By correlating individual networks through the NCL method, assigning a proper weight to each CNN by GA and combining the CNN outputs through weighted averaging, we create an ENN which accurately characterizes the process. An empirical study was performed in this paper on a regression problem of a

* Corresponding author. Automation and Instrumentation Department, Petroleum University of Technology, 63431 Ahwaz, Iran. Tel.: +98 919 611 3440.
E-mail address: Jazayeri.sadeghi611@gmail.com (H. Jazayeri-Rad).

http://dx.doi.org/10.1016/j.jngse.2014.07.012

chemical process to show the ability of the negative correlation learning and genetic algorithm to identify and model a complicated system. The rest of this paper is organized as follows: Section 2 describes the negative correlation learning and genetic algorithm. As an empirical case study, Section 3 describes an experimental study of the negative correlation learning of a nonlinear chemical process in a refinery in AHWAZ petroleum field. Section 4 specifies the input and output variables of CNNs. Section 5 of the paper presents results of this study. Summary and conclusion are provided in Section 6.

2. Combination of the negative correlation and Genetic Algorithm

To learn an ensemble network, there should be a plan to create dissimilar and diverse networks and another plan to combine the outcomes of these diverse networks in order to strengthen accurate networks and weaken poor ones in the ENN output. Diversity of networks is important because there is no advantage in multiplying similar networks which generalize identically. Different training parameters (Hansen and Salamon, 1990), different training patterns (Bauer and Kohavi, 1999), different feature subsets (Zio et al., 2008) and different learning methods for each network of the ensemble (Xu et al., 1992) are among some techniques of creating diverse networks. In addition, simple averaging and weighted averaging are famous methods of combining CNNs. In this paper, we use the different feature subset technique to generate diversity and the weight averaging method for combining CNNs.

Now suppose that the experimental input matrix, X, and output matrix, $F(X)$, are defined as:

$$X = [X_1, X_2, X_3, X_4, X_5]_{Z\times 5} \tag{1}$$

$$F(X) = [F(X_1, X_2, X_3, X_4, X_5)]_{Z\times 1} \tag{2}$$

where $X_z \in R$ and Z is the size of the data set. This section considers the estimation of $F(x)$ by creating a negative correlation ensemble neural network (NCENN) which increases accuracy and reduces the generalization error. Firstly, we train 'M' individual neural networks and then combine them. These individual networks are dissimilar in their training data sets. This results in different base learners for the individual CNN and increases their diversity. In order to construct these sub-data sets the bootstrap method introduced by Efron (1979) is used. This is a general resampling method for estimating the distributions of data based on independent observations.

The proposed technique in this work simultaneously provides diverse and negative correlations amongst individual component networks. The predicted results are then compared with the results obtained using other methods of designing ENNs. In negative correlation learning, all the individual networks in the ensemble are trained simultaneously through the correlation penalty terms in their error functions (Liu and Yao, 1999). Therefore, by increasing the interactions between the component networks, the method trains each network to obtain the best outcome for the whole of ensemble.

We use the Levenberg–Marquardt Algorithm (LMA) to learn weights of CNNs by assigning negative correlation penalty terms in their objective functions. The LMA is a very precise and popular curve-fitting algorithm used in neural networks for solving generic curve-fitting problems. However, like many fitting algorithms, LMA finds only a local minimum which is not necessarily the global minimum. Our goal is to derive the global minimum of a negative correlation objective function. We employ GA to estimate the primal weights which directs the objective function to its global minimum point and then use LMA to derive accurate weights corresponding to this global minimum.

As shown in Fig. 1, LMA cannot compute weights which globally minimize the objective function, because it cannot distinguish between local and global minima. In such a situation, GA avoids these local minima by estimating the primal weights of each diverse component network. To achieve this goal, each network use 80% of the bootstrap sample sub-data set for training and the remaining 20% for testing. These data are randomly distributed all over the sample sub-data sets. The following equation presents the error function used in this step:

$$E_{iGA}(x) = \frac{1}{Z}\sum \left(f_{iGA}(x) - F(x)\right)^2 \tag{3}$$

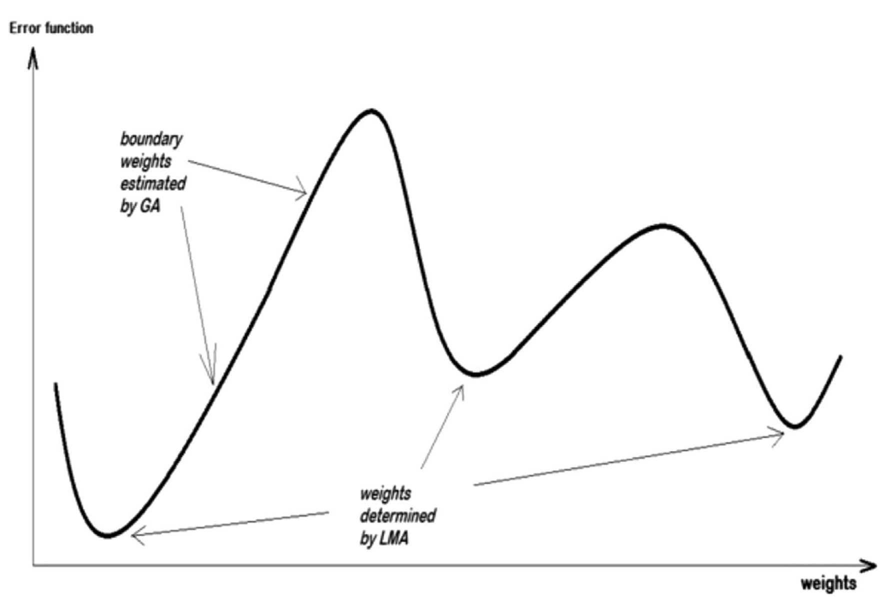

Fig. 1. Boundary of the weights estimated by GA and LMA.