Engineering Applications of Artificial Intelligence 32 (2014) 203–212

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Neural network ensembles based on copula methods and Distributed Multiobjective Central Force Optimization algorithm

Meng Chao [a], Sun Zhi Xin [a,*], Liu San Min [b]

[a] Key Laboratory of Broadband Wireless Communication and Sensor Network Technology, Nanjing University of Posts and Telecommunications, Ministry of Education, Nanjing 210003, China
[b] College of Computer and Information, Anhui Polytechnic University, Wuhu 241000, China

ARTICLE INFO

Article history:
Received 4 September 2013
Received in revised form
17 January 2014
Accepted 13 February 2014
Available online 16 April 2014

Keywords:
Copula function
Ensemble neural network
Central force optimization

ABSTRACT

Copula is a function with multivariate distribution. It has uniformly distributed marginals. Central Force Optimization is a new algorithm which is based on kinematics. It has been illustrated that this algorithm is better than other heuristic methods, when these techniques are applied to the classification problems. This paper proposes a technique of neural network ensembles which use the distributed Central Force algorithm to optimize each individual component network, simultaneously. The distributed Central Force algorithm incorporates an additional regularization term and utilizes the multiobjective architectures to design component networks. Furthermore it proposes that a new method of combining the component networks is to use Copula function theory as an effective design tool which generates the combining weights. The experimental results show that the copula-based ensemble network achieves better performance than other ensemble methods and that Distributed Multiobjective Central Force Optimization is capable of achieving better solutions in the light of converging speed and local minima. In the experimental discussion, the paper gives several reasons why the proposed method outperforms others.

1. Introduction

Ensemble Neural Network (ENN) is a learning architecture in which many individual networks are jointly used to solve the same problems. In an ENN, every component network is a feed-forward network trained with the same learning algorithms. The ENN (Granitto et al., 2005) demonstrates that the learning ability of an NN system can be improved through ensembling many of NNs. Since this method behaves very well, it has been applied to research areas, such as pattern recognition (Green et al., 2009), medical diagnosis (Zhou and Zhi-Hua, 2003; Minku and Ludermir, 2008).

In general, an ENN is created through two steps, constructing component networks and combining these networks. But most algorithms train the individual network sequentially or independently, so that the advantages of cooperation and interaction among the networks are not utilized. However, García-Pedrajas (2005) the author shows that the cooperation with component networks can obtain better ensembles.

Copula theory concerns dependence among random variables (Nelsen, 2006; Mayor, 2007). Satish and Iyengar (2011) and Sundaresan et al. (2011) use a copula-based solution for binary classification problems. In Gill (2012) and Stephen (2011), because of the complex form between wind speed and active power, the application of copulas is proposed. Application of copulas for the other research area has been shown in Hu et al. (2012) and De Baets et al. (2007), respectively.

Central Force Optimization (CFO) (Formato, 2007) is a novel deterministic search heuristic algorithm based on the gravitational kinematics. Compared with the stochastic optimization algorithms such as Ant Colony Optimization (Li-Ning et al., 2011) and Particle Swarm Optimization (Yu et al., 2011), CFO is a deterministic algorithm because the universal law of gravitation is deterministic. Deterministic algorithms have some significant advantages. For instance, every algorithm run beginning with the same parameters obtains the same results, that may be especially useful in using as a feedback during an run.

CFO is in its infancy and originates from Formato's work (Formato, 2007), which showed that this algorithm models "probes" that move through the search space, which is similar to masses moving under gravity. There are some applications of the CFO in some specific fields of research. Green et al. (2012) apply the algorithm to a basic neural networks. In Mahmoud (2011), an

* Corresponding author.
E-mail address: sunzx@njupt.edu.cn (S.Z. Xin).

http://dx.doi.org/10.1016/j.engappai.2014.02.009
0952-1976/

efficient global hybrid optimization method is proposed combining central force optimization as a global optimizer and the Nelder–Mead algorithm as a local optimizer. Haghighi and Ramos (2012) apply the Central Force Optimization algorithm to Inverse Transient Analysis. Qubati and Dib (2010) evaluates CFO's performance and applies the improved CFO algorithm to the optimal design of two different wideband microstrip patch antennas.

Recently, heuristic algorithms are considered as a good tools for optimization problems. The heuristic algorithms do not demand that the constraints and the objective functions have to be continuous and differentiable. A CFO is such a very young algorithm which is applied to optimization problems.

However, it is observed that the training of CFO corresponds to considering the entire ENN as a single estimator and the entire training errors without regularization.

In that case, CFO only reduces the entire mean squared error (MSE) of the ENN, and it does not regularize the complexity of the ENN, so that it overfits the noises in training sets. Additionally, the combining weights and the correlation of members are also considered for achieving better ensembles. Copula function is the tool which determines the best combining weights. In a ENN, the combining weights of the individual neural networks can be regarded as random variables. The input training sets for the member networks of the ENN are similar or identical, so that the combining weights of the outputs data are not independent. Thus it needs that some mechanisms and theories describe the correlation among the combining weights. Copula function is a kind of effective tool.

In this paper, the proposed solution uses the copula theory to model the joint distribution of random variables, combine the component networks and determine the combining weights. In order to solve the problem of overfitting, the paper also proposes a Distributed Multiobjective Central Force Optimization algorithm that incorporates an regularization term to train the member networks (Chen and Yao, 2010). Because CFO is a new algorithm, there is no research on the proof of convergence which is the basis of wide application. This paper, firstly, uses the gravitational kinematics theory to analyze CFO convergence. Then, it extends basic CFO to the distributed multiobjective architecture.

The basic CFO explores a decision space by “flying” a group of “probes” whose trajectories are governed by Newton's laws. However, DMCFO is a multi-group variate of CFO. It is inspired by the evaluated particle swarm algorithm (Vlachogiannis and Lee, 2005). In DMCFO, the concept of multi-group is analogous to multi-swarm in VEPSO. This algorithm uses “probes” of multi-group to search the optimum of the objective function.

In DMCFO, every group is evaluated by using one of the objective functions of the considering problem, and the information of its own objective function that it holds is transmitted to the other groups through the interchange of the better experience. The DMCFO can be parallelized by distributing the number of groups in the same number of PCs. This method can accelerate the convergence time (Vlachogiannis and Lee, 2005).

The key contributions of this paper are

- The employ copula theory to generate the combining weights.
- Apply the gravitational kinematics theory to the proof of the CFO algorithm convergence.
- Implement the CFO algorithm using distributed multiobjective architecture.
- Use the DMCFO to train the component network.

For the problem of combining component networks and creating component networks in an ENN, the rest of the paper is organized as follows. In Section 2, the copula theory and the copula-based ensemble method are proposed to combine component networks. The DMCFO algorithm is introduced to train the component networks in Section 3. Experimental results and discussions are showed in Section 4. Finally, Section 5 concludes this paper.

2. Copula-based ensemble network

This section mainly describes how to use copula function to combine the member networks. Currently, there are two main methods of combining individual networks: one is averaging weights and the other is voting. Simple averaging weights have a big drawback that this approach is not able to reflect the correlation of individual networks. Because the training data sets for the individual networks are similar or identical, the correlation among the individual networks must be considered. The combining weights of the output data sets can reflect the correlation. In this paper, an approach based on copula function is proposed to design a combining weight of every member in an ENN.

2.1. Copula function

Copulas functions can couple multivariate joint distributions to their component marginal distribution functions. The Sklar theorem (Mayor, 2007) is central to copulas.

Definition [Multivariate Gaussian copula]. $\mathbf{R}$ be a symmetric, positive definite matrix with $\mathrm{diag}(\mathbf{R})=(1,1,.....1)^T$ and $\Phi_{\mathbf{R}}$ the standardized multivariate normal distribution with correlation matrix $\mathbf{R}$. The Gaussian copula is defined as follows:

$$C(u_1,\ldots,u_n)=\Phi_{\mathbf{R}}^n(\Phi^{-1}(u_1),\ldots,\Phi^{-1}(u_n)) \tag{1}$$

where $\Phi_{\mathbf{R}}^n$ denotes the joint distribution function of the n-variate standard normal distribution function with linear correlation matrix $\mathbf{R}$, whose elements are used to describe the dependence between couples of variables and Φ^{-1}, as usual, is the inverse of the standard univariate normal distribution function Φ.

$\rho_{X_iY_j}(-1\le\rho_{X_iY_j}\le 1\ \ ,i,j=1,\ldots,n)$ is the element of matrix $\mathbf{R}$. $\rho_{X_iY_j}$ is the correlation coefficient. If $\rho_{X_iY_j}=0$, the variables are independent. If $\rho_{X_iY_j}=1$, the variables are the positive linear correlation. If $\rho_{X_iY_j}=-1$, the variables are the negative linear correlation (Ahcène et al., 2011).

2.2. Combining the component networks with the copula

As for combining component networks, majority voting and plurality voting are the most prevailing methods (Sylvester and Nitesh, 2006; Gheyas and Smith, 2011) for data classification. There are many other methods (Islam et al., 2008; Chan and Kasabov, 2005; Zheng, 2009). These works do not consider correlation among the individual networks. The combining weights of the output data sets can reflect the correlation.

In a ENN, the combining weights of the individual neural networks can be regarded as random variables. This paper uses Gaussian copulas to describe the correlation among random variables. The description of the optimizing model can be formulated as follows:

$$\left.\begin{array}{ll}\text{Problem} & H(w_1,w_2,\ldots,w_n)=C(w_1,w_2,\ldots,w_n)\\ \text{subject to} & \sum_{i=1}^{n} w_i=1, 0<w_i\le 1\end{array}\right\} \tag{2}$$

where $w_i\ \ i=1,2,...,n$ is the combining weight of the component networks. The H is the joint distribution function of the weight vector. C is Gaussian copula which is defined in Minku and Ludermir (2008) and García-Pedrajas (2005).