Neurocomputing 197 (2016) 78–85

Contents lists available at ScienceDirect

Neurocomputing

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

Research of multi-sided multi-granular neural network ensemble optimization method

Hui Li [a,b,*], Xuesong Wang [b], Shifei Ding [c,d]

[a] School of Computer Science and Technology, Jiangsu Normal University, Xuzhou, Jiangsu 221116, China
[b] School of Information and Electrical Engineering, China University of Mining and Technology, Xuzhou, Jiangsu 221116, China
[c] School of Computer Science and Technology, China University of Mining and Technology, Xuzhou, Jiangsu 221116, China
[d] Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China

ARTICLE INFO

Article history:
Received 1 July 2015
Received in revised form
27 January 2016
Accepted 12 February 2016
Communicated by: Grana Manuel
Available online 26 February 2016

Keywords:
Neural network ensemble(NNE)
Multi-sided attribute granularity
Feature selection

ABSTRACT

According to the thought “divide and conquer” that human perceives complicated things from multi-side and multi-view and balances the final decision, this paper puts forward the multi-sided multi-granular neural network ensemble optimization method based on feature selection, which divides attribute granularity of dataset from multi-side, and structures multi-granular individual neural networks using different attribute granularity and the corresponding subsets. In this way, we can gain multi-granular individual neural networks with greater diversity, and get better performance of neural network ensemble(NNE). Firstly, use feature selection method to calculate the importance of each attribute, according to the average weight to choose some attributes whose average weight is greater than a certain threshold, to form an attribute granularity and the corresponding sample subset, thus to construct an individual neural network. If samples are not properly identified, this attribute granularity is weak for the generalization ability of the sample. Secondly, again calculate the importance of the attributes of samples not properly identified, choose the attributes that can generalize the corresponding samples better, and add to the last attribute granularity to form a new attribute granularity, and at the same time random choose two-thirds of sample subset to construct an individual neural network. In turn, one can get a series of attribute granularities and the corresponding sample subsets and a series of multi-granular individual neural networks. These attribute granularities and the corresponding sample subsets constructed from multi-side and multi-view with greater diversity can construct multi-granular individual neural networks with greater diversity. This method not only reduces the dimension of the dataset, but also makes the attribute granularity to identify the corresponding sample as large as possible. Finally, by calculating the diversity of each of the two individual neural networks, optimal selects some individual neural networks with greater diversity to ensemble. The simulation experiments show that our proposed method here, multi-side multi-granular neural network ensemble optimization method, can gain better performance.

1. Introduction

With the maturity and popularity of database technology, accumulations of human data are increasing exponentially. Facing the different types of massive high-dimensional data sets, data mining has encountered great difficulties [1]. Massive data contain not only the quantity, but also contains redundancy, uncertainty, even disruptive, so the first thing we have to do is to remove those who have nothing to do with the decision or smaller aspects. This is also a hot research topic in the machine learning, namely, feature selection. Feature selection is attracting more and more extensive attention, and has successfully been applied to the various aspects, such as text classification, image retrieval, biometrics identification, text identification, digital identification and gene analysis. Feature selection method is also a general rule for people to perceive the complicated things, which perceives complicated things from multi-side and multi-view and balancing the final decision, from point to surface, from shallow to deep, divide and rule based on personal experience or knowledge.

Hansen and Salamon proved in 1990 that training multiple neural networks and ensembling the results can significantly improve the generalization ability of the neural network system [2]. Because of excellent performance and accuracy significantly improved in the aspect of the generalization ability, neural network ensemble(NNE), proposed since 1990, has been extensively studied by many scholars [3–11]. Refs. [3,4] pointed out that the

* Corresponding author.

http://dx.doi.org/10.1016/j.neucom.2016.02.013
0925-2312/

effective condition of ensemble learning was that learning error rate of each individual neural network should be less than 0.5, otherwise, the error rate of NNE will increase. Ref. [5] gave the decomposition formula of generalization error of NNE, and showed that as long as the average generalization error of individual neural networks remained the same, increasing diversity could improve the generalization ability of NNE, so, in theory, to structure the neural networks with a bigger diversity is considered as one of the important characteristics of NNE. Generalization error of NNE is equal to the diversity of the average generalization error and the average diversity of individual neural networks [6]. So in order to enhance the generalization ability of NNE, on the one hand, we should improve the network generalization ability of individual neural networks as far as possible; on the other hand, to as far as possible increase the diversity between individual neural networks is one of the important research directions of NNE.

Transforming network training data is one of the main methods of increasing diversity of individual networks. The current commonly used method to generate individual neural network is by perturbation training data to obtain larger diversity of individual networks that mainly include Boosting [12] algorithm and Bagging [13] algorithm. The training set of Boosting is decided by the performance of the network generated before, examples that are error judged by the existing networks will appear in new training set with a larger probability. Boosting has many advantages, it has higher accuracy, it does not need a priori knowledge and only needs to select the appropriate number of iterations, etc., but its speed, to a certain extent, depends on the training dataset and the weak classifier. Insufficient training data or the poor base classifiers will lead to decline of the training precision. The training set of Bagging includes repeated sampling and is randomly selected from the original training set; the size of the training set is usually comparable to the original training set. In this way, some examples of the original training set may appear multiple times in the new training set, and other examples might not appear, and when the data is bigger, the size of the training set is relatively large.

Due to restrictions by attribute selection problem, single classifier is difficult to obtain satisfactory results when dealing with the data with more category and noise.

Based on the above analysis, according to the thought of divide and conquer that human perceives complicated things from multi-side and multi-view and balancing the final decision, this paper puts forward the multi-sided multi-granular neural network ensemble optimization method based on feature selection, which divides attribute granularity of dataset from multi-side, and structures multi-granular individual neural networks using different attribute granularity and the corresponding subsets. In this way, we can gain multi-granular individual neural networks with greater diversity, and get better performance of NNE. Firstly, use feature selection method to calculate the importance of each attribute, according to the average weight to choose some attributes, in which average weight is greater than a certain threshold, to form an attribute granularity and the corresponding sample subset, thus to construct an individual neural network. If samples are not properly identified, this attribute granularity is weak for the generalization ability of the sample. Secondly, again calculate the importance of the attributes of samples not properly identified, choose the attributes that can generalize the corresponding samples better, and add to the last attribute granularity to form a new attribute granularity, and at the same time randomly choose two-thirds of sample subset to construct an individual neural network. In turn, one can get a series of attribute granularities and the corresponding sample subsets and a series of multi-granular individual neural networks. These attribute granularities and the corresponding sample subsets constructed from multi-side and multi-view with greater diversity can construct multi-granular individual neural networks with greater diversity. This method not only reduces the dimension of the dataset, but also makes the attribute granularity to identify the corresponding sample as large as possible. Finally, by calculating the diversity of each two individual neural networks, optimal selects some individual neural networks with greater diversity to ensemble. The simulation experiments show that our proposed method here, multi-side multi-granular neural network ensemble optimization method, can gain better performance.

The following chapters are organized as follows. Chapter 2 introduces the related knowledge, such as NNE and feature selection algorithm et al. Chapter 3 analyzes multi-sided multi-granular neural network ensemble optimization method, which uses the method of divide-and-rule and multi-sided to build diversity characteristic attribute granularity and the corresponding sample subset, on this basis, we give the architecture, algorithm description and analysis of our proposed method here. In chapter 4, through a series of experiments, we verify the performance and effectiveness of multi-sided multi-granular neural network ensemble optimization method our proposed here. Finally, chapter 5 elaborates the overall conclusions, discussion and directions for further research.

2. Related knowledge

2.1. Neural network ensemble(NNE)

NNE uses a finite number of individual neural networks to learn the same question, the output is jointly decided by all the output of individual neural networks under this example [5].

Suppose the input $x \in R^m$ meets distribution $p(x)$, if the output corresponding to x is $d(x)$, the output corresponding to individual neural network $f_i(i=1,2,\ldots,N)$ is $f_i(x)$, the weigh corresponding to $f_i(x)$ is ω_i, then the output $f_{ensemble}(x)$ of NNE corresponding to x can be defined as

$$f_{ensemble}(x) = \sum_{i=1}^{N} \omega_i f_i(x) \tag{1}$$

$$\text{Generalization error of NNE} : E_{ensemble} = \int P(x)(f_{ensemble}(x) - d(x))^2 dx \tag{2}$$

$$\text{Generalization error of } f_i(i=1,2,\ldots,N) : E_i = \int P(x)(f_i(x) - d(x))^2 dx$$

$$\text{Weighted average of } f_i(i=1,2,\ldots,N) : E_{average} = \sum_{i=1}^{N} \omega_i E_i \tag{4}$$

$$\text{Diversity of } f_i(i=1,2,\ldots,N) : A_i = \int P(x)(f_i(x) - f_{ensemble}(x))^2 dx \tag{5}$$

$$\text{Diversity of NNE} : A_{ensemble} = \sum_{i=1}^{N} \omega_i A_i \tag{6}$$

Through the theoretical analysis, Krogh et al. gave the computational formula of NNE:

$$E_{ensemble} = E_{average} - A_{ensemble} \tag{7}$$

Because the diversity of each network is nonnegative, from formula (7), we know that the generalization error $E_{ensemble}$ of NNE is no greater than the weighted average of generalization error $E_{average}$ of individual neural networks. Increasing diversity $A_{ensemble}$ (i. e. increasing the diversity A_i) can effectively reduce the generalization error of NNE.