Information Fusion 20 (2014) 242–251

Contents lists available at ScienceDirect

Information Fusion

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

Mixture of feature specified experts

Saeed Reza Kheradpisheh [a,b], Fatemeh Sharifizadeh [a], Abbas Nowzari-Dalini [a,*], Mohammad Ganjtabesh [a,b], Reza Ebrahimpour [c]

[a] Department of Computer Science, School of Mathematics, Statistics and Computer Science, University of Tehran, Tehran, Iran
[b] School of Cognitive Sciences (SCS), Institute for Research in Fundamental Sciences (IPM), Tehran, Iran
[c] Brain and Intelligent Systems Research Lab, Department of Electrical and Computer Engineering, Shahid Rajaee Teacher Training University, Tehran, Iran

ARTICLE INFO

Article history:
Received 29 July 2013
Received in revised form 20 February 2014
Accepted 24 February 2014
Available online 17 March 2014

Keywords:
Combining classifiers
Diversity
Mixture of experts
BCI
EEG

ABSTRACT

Mixture of Experts is one of the most popular ensemble methods in pattern recognition systems. Although, diversity between the experts is one of the necessary conditions for the success of combining methods, ensemble systems based on Mixture of Experts suffer from the lack of enough diversity among the experts caused by unfavorable initial parameters. In the conventional Mixture of Experts, each expert receives the whole feature space. To increase diversity among the experts, solve the structural issues of Mixture of Experts such as zero coefficient problem, and improve efficiency in the system, we intend to propose a model, entitled Mixture of Feature Specified Experts, in which each expert gets a different subset of the original feature set. To this end, we first select a set of feature subsets which lead to a set of diverse and efficient classifiers. Then the initial parameters are infused to the system with training classifiers on the selected feature subsets. Finally, we train the expert and the gating networks using the learning rule of classical Mixture of Experts to organize collaboration between the members of system and aiding the gating network to find the best partitioning of the problem space. To evaluate our proposed method, we have used six datasets from the UCI repository. In addition the generalization capability of our proposed method is considered on real-world database of EEG based Brain-Computer Interface. The performance of our method is evaluated with various appraisal criteria and significant improvement in recognition rate of our proposed method is indicated in all practical tests.
© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Pattern recognition is the scientific discipline whose aim is to classify patterns into a set of categories. Classification process is the most significant step of any pattern recognition system. Commonly, the neural networks are the general classifiers for pattern recognition systems. The ensemble classification systems have been widely proposed to achieve high classification performance [1–3]. Ensemble classification is often used to cope with a complex classification problem such as those involving very overlapped classes, finite number of samples, and high-dimensional feature space [4–6].

It is intuitively clear that an ensemble of identical classifiers will be no better than a single classifier thereof. Empirical and theoretical studies have been demonstrated that diversity among the ensemble members is one of the most significant keys to the success of ensemble models [7–10], and by combining the results of diverse and accurate base classifiers the generalization capability is obtained in the system [11,12]. Diversity may be obtained through different approaches in the ensemble systems [13]; these include different learner in ensemble, different representation of input, partitioning the input space, different feature space and adding a penalty to the outputs to encourage diversity [6,14].

In the categorization of ensemble systems, there are generally two types of combining classifiers: *classifier fusion* and *classifier selection* [3]. In the fusion, each classifier is trained on the whole problem space and the final result is determined with considering the results of all classifiers [15–17]. But in the classifier selection, each classifier is trained on a portion of the problem space [18–20] and the final result is determined with aggregating the result(s) of one (or some) classifier(s) [3,21]. There is a combination strategy lying between selection and fusion strategies [22] which the Mixture of Experts (ME) method is its most famous instance [3].

The principle of ME is that the certain experts will be able to specialize to a particular parts of the input space by adopting a gating network who is responsible for learning the appropriate weighted combination of the specialized experts for any given

* Corresponding author. Tel.: +98 2166412478.
E-mail addresses: kheradpisheh@ut.ac.ir (S.R. Kheradpisheh), sharifizade@ut.ac.ir (F. Sharifizadeh), nowzari@ut.ac.ir (A. Nowzari-Dalini), mgtabesh@ut.ac.ir (M. Ganjtabesh), ebrahimpour@ipm.ir (R. Ebrahimpour).

http://dx.doi.org/10.1016/j.inffus.2014.02.006
1566-2535/© 2014 Elsevier B.V. All rights reserved.

input. In this way, the input space is dynamically divided and conquered by the gating network and the experts, respectively [19].

The ME scheme has been widely studied because of its classification performance and transparency. In conventional ME [19], the experts and gating network are linear classifiers. Though, in case of complex problem classifications, in order to enhance the efficiency in this system, the Multilayer Perceptron (MLP) networks are used to form the networks of ME [23,24]. This model is called as Mixture of Multilayer Perceptron Experts (MME).

In [25–27], the idea of ME model was extended with a hierarchical mixture of experts model, where each constituent of ME is substituted with a mixture of experts model. ME has been regarded as a mixture model for estimating conditional probability distributions, and its statistical properties have been investigated in some studies [25,28–32]. These statistical properties have led to the development of various Bayesian training methods [31,32]. In [29], Mossavat et al. used Bayesian learning to train the structure of the hierarchical ME to estimate speech quality. In addition, Expectation Maximization algorithm was employed for adjusting the parameters of ME [25,28,30,33]. In [34,35], the ME model was extended for time series data to handle causal dependencies.

Although the ME is an appropriate ensemble method, it has some deficiencies in partitioning the input space. Its imperfections are affected from the infirmity of the gating network for dividing the input space. It is expected that the gating network stochastically divides the input space according to the spatial similarity of input patterns. But, this may leads to complex and nested partitions, and thus, the experts cannot appropriately model their subproblems [36]. In [11], Hansen reported the zero coefficient problem in competitive learning mechanism of ME. In this problem, some experts are ejected from the competition mechanism of ME and hence, no subspaces dedicated to them. In addition, the gating network may fail to divide the problem space among the experts and hence, all experts learn the overall problem space [37,22]. As a result of mentioned facts, it is observed that a favorite input partitioning method via gating network is required to perform high performance classification. To overcome these problems of ME, several methods are developed to divide the input space of a problem among the experts [36,38–43]. Basically, in these researches a prior knowledge about the input space is required to decompose the input space among the experts [36,38,39], where each expert is localized on a specific area of input space. In [36], the Self Organizing Maps is used as a gating network to divide the input space between the experts according to the underlying probability distribution of data. In other words, they attempt to localized experts on specified subproblems by gathering information about the data distribution. In [38], instead of leaving the gating network to automatically partition the input space, they performed an intelligent contribution to direct each expert towards a specific area of input space using Fuzzy *c*-Means clustering method. In some face processing studies [39,40] the input space is divided among the experts based on the pose of face or facial expression. In [44], a method based on both cooperative coevolution and Mixture of Experts is introduced to decompose problems into different regions and assign the experts to these distinct regions. By considering all mentioned methods, it is observed that the boundaries between the subspaces are distinctive in all of these methods. Thus, the generalization capability of the MME for the complex samples in the overlapped subspaces is omitted and also the interactions between experts in the competitive system of MME are neglected. Regarding to these facts, it is better to follow other methods that try to use more advantages of ME. Some studies attempted to combine the specifications of boosting and ME algorithms [41]. In [42], a new dynamically boosted ME method is proposed that applies a confidence measure as a gating function, which indicates the contribution of each expert to the ensemble output. In [43], a boost wise partitioning procedure was modified using the error and confidence measures as the hardness criterion in the problem space. In [45,46], the comprehensive survey of the ME model and its extensions are provided.

Although diversity in an ensemble is deemed to be a key factor to the performance of ensembles [9,13,47] and many studies on diversity have been conducted, the ME employs simplest method to produce diverse experts using random initial weights. In fact, ME had some defects in creating diverse and efficient experts in its system. In above mentioned methods, it is tried to conquer on shortcomings of ME by exerting different training samples to the experts.

In this paper, we propose a method for inferring most important parameter of MME and making favor diversity between the experts. Unlike the conventional MME which each expert receives the whole feature space, in our proposed method each expert gets a different subset of the original feature set. Our method, which is called Mixture of Feature Specified Experts (MFSE), improves the performance of the MME in three steps. In the first step, optimal feature subsets are selected in a process to create diverse information resources. In the second step, MLP experts are trained on the selected feature subsets and then the initial parameters of MFSE architecture are loaded. Therefore, initial diversity is infused among the experts. Finally, in the third step, the formal algorithm of MME is executed on the loaded architecture of MFSE to organize cooperation between the experts. To evaluate our proposed method, we use six datasets from the UCI repository and a real world dataset of signals from a Brain Computer Interface system. The experimental results using desired datasets support our claim that increasing the diversity among the experts and loading appropriate initial parameters in the system, improve the performance of MFSE with respect to the conventional MME in all tests.

The paper is arranged as follows: in Section 2, a brief description of MME architecture, its learning algorithm, and one of its extended methods are described. In Section 3, our proposed model is discussed in details. Section 4 provides and illustrates the experimental results. Finally, we summarize the paper and draw conclusions in Section 5.

2. Mixture of experts

Mixture of Experts is a widely used paradigm for creating a combination of learners [19]. The ME architecture is composed of a gating network and several experts. In this method, the input signal is included in motivating the integration part of system (i.e. gating network) which unites the outputs of experts [21]. There is a competition between experts to learn the input samples and hence, the input space is decomposed into subspaces and each expert examines different part of the input space. In the ME method, the gating network is responsible for combining the various experts' outputs and thus, intervene the competition among the experts.

Since our suggested method is established upon the MME, this method is briefly described in Section 2.1. Then, we concisely describe one of the most successful techniques that attempt to improve the performance of ME, entitled Intelligent Contribution to Direct Experts [38–40], in Section 2.2.

2.1. Mixture of multilayer perceptron experts

Fig. 1 illustrates the basic structure of MME model. Suppose that N MLP experts, each contains one hidden layer, are used in this model. The ith expert ($i = 1, 2, \ldots, N$) computes the output vector O_i for the input vector x. The gating network divides the problem space by assigning the weight g_i to the output vector of the ith

Download English Version:

https://daneshyari.com/en/article/528138

Download Persian Version:

https://daneshyari.com/article/528138

Daneshyari.com