Engineering Applications of Artificial Intelligence 26 (2013) 2315–2326

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

DE+RBFNs based classification: A special attention to removal of inconsistency and irrelevant features

Ch. Sanjeev Kumar Dash [a], Aditya Prakash Dash [a], Satchidananda Dehuri [b,*], Sung-Bae Cho [c], Gi-Nam Wang [d]

[a] Silicon Institute of Technology, Silicon Hills, Patia, Bhubaneswar 751024, Odisha, India
[b] Department of Systems Engineering, Ajou University, San 5, Woncheon-dong, Yeongtong-gu, Suwon 443-749, South Korea
[c] Soft Computing Laboratory, Department of Computer Science, Yonsei University, 134 Shinchon-dong, Sudaemoon-gu, Seoul 120-749, South Korea
[d] Department of Industrial Engineering, Ajou University, San 5, Woncheon-dong, Yeongtong-gu, Suwon 443-749, South Korea

ARTICLE INFO

Article history:
Received 2 January 2013
Received in revised form
7 May 2013
Accepted 23 August 2013
Available online 21 September 2013

Keywords:
Classification
Radial basis function neural networks
Differential evolution
Feature selection
Dataset consistency

ABSTRACT

A novel approach for the classification of both balanced and imbalanced dataset is developed in this paper by integrating the best attributes of radial basis function networks and differential evolution. In addition, a special attention is given to handle the problem of inconsistency and removal of irrelevant features. Removing data inconsistency and inputting optimal and relevant set of features to a radial basis function network may greatly enhance the network efficiency (in terms of accuracy), at the same time compact its size. We use Bayesian statistics for making the dataset consistent, information gain theory (a kind of filter approach) for reducing the features, and differential evolution for tuning center, spread and bias of radial basis function networks. The proposed approach is validated with a few benchmarked highly skewed and balanced dataset retrieved from University of California, Irvine (UCI) repository. Our experimental result demonstrates promising classification accuracy, when data inconsistency and feature selection are considered to design this classifier.

1. Introduction

Classification is one of the fundamental tasks in data mining (Dehuri and Ghosh, 2013) and pattern recognition (Nanda and Panda, 2013). Over the years many models have been proposed. (Huang and Wang, 2006; Chatterjee and Bhattacherjee, 2011) However, it is a consensus that the accuracy of the discovered model (i.e., neural networks (NNs) (Haykin, 1994; Yaghini et al., 2013), rules (Das et al., 2011), and decision tree (Carvalho and Freitas, 2004)) strongly depends on the quality of the data being mined. Hence inconsistency removal and feature selection brings lots of attention of many researchers (Battiti, 1994; Yan et al., 2008; Keynia, 2012; Ebrahimzadeh and Ghazalian, 2011; Liu et al., 2010). If the inconsistent data are simply deleted or classified as a new category then inevitably some useful information will be lost. The method used in this paper for making the dataset consistent is based on the Bayesian statistical method (Wu, 2007). Here the inconsistent data are classified as the most probable one and the redundant data records are deleted as well. So the loss of information due to simple deletion or random classification of inconsistent data is reduced and the size of the dataset is also reduced.

Feature selection is the process of selecting a subset of available features to use in empirical modeling. Like feature selection, instance selection (Liu and Motoda, 2002) is to choose a subset sample to achieve the original purpose of a classification task, as if the whole dataset is used. Many variants of evolutionary and non-evolutionary based approaches are discussed in Derrac et al. (2010). The ideal outcome of instance selection is a model independent, minimum sample of data that can accomplish tasks with little or no performance deterioration. Unlike feature selection and instance selection, feature extraction at feature level fusion recently attracts data mining/machine learning researchers (Liu, et al., 2011) to give special focus while designing a classifier. However, in this work, we restrict ourselves with feature selection and inconsistency removal only.

Feature selection can be broadly classified into two categories: (i) filter approach (it depends on generic statistical measurement), and (ii) wrapper approach (based on the accuracy of a specific classifier) (Aruna et al., 2012). In this work, the feature selection is performed based on information gain theory (entropy) measure with a goal to select a subset of features that preserve the relevant information found in the entire set of features as much as possible. After the selection of the relevant set of features the fine tuned

* Corresponding author. Tel.: +81 10 5590 1980.
E-mail addresses: sanjeev_dash@yahoo.com (Ch.S.K. Dash), adityaprakashdash@gmail.com (A.P. Dash), satchi@ajou.ac.kr (S. Dehuri), sbcho@yonsei.ac.kr (S.-B. Cho), gnwang@ajou.ac.kr (G.-N. Wang).

0952-1976/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.engappai.2013.08.006

radial basis function network is modeled using differential evolution for the classification of both balanced and unbalanced datasets. In imbalance classification problems the number of instances of each class that occur can be very different (Perez-Godoy et al., 2010).

Over the decade radial basis function (RBF) networks have attracted a lot of interest in various domains of interest (Haykin, 1994; Novakovic, 2011; Naveen et al., 2010; Liu et al., 2005). The reason is that they form a unifying link between function approximation, regularization, noisy interpolation, classification, and density estimation. Moreover, training RBF networks is usually faster than training multi-layer perceptron networks. RBF network training usually proceeds in two steps: first, the basis function parameters (corresponding to hidden units) are determined by clustering. Second, the final-layer weights are determined by a least square method which reduces to solve a simple linear system. Thus, the first stage is an unsupervised method which is relatively fast, and the second stage requires the solution of a linear problem, which is also fast.

The other advantages of RBF neural networks, compared to multi-layer perceptron networks, is the possibility of choosing suitable parameters for the units of hidden layer without having to perform a nonlinear optimization of the network parameters. However, the problem of selecting the appropriate number of basis functions remains a critical issue for RBF networks. The number of basis functions controls the complexity, and hence the generalization ability of RBF networks. An RBF network with too few basis functions gives poor predictions on new data, i.e., poor generalization, since the model has limited flexibility. On the other hand, an RBF network with too many basis functions also yields poor generalization since it is too flexible and fits the noise in the training data. A small number of basis functions yield a high bias, low variance estimator, whereas a large number of basis functions yield a low bias but high variance estimator. The best generalization performance is obtained via a compromise between the conflicting requirements of reducing bias while simultaneously reducing variance. This trade-off highlights the importance of optimizing the complexity of the model in order to achieve the best generalization. However, choosing an optimal number of kernels is beyond the focus of this paper.

In the training procedure of RBFNs revealing center of gravity and width is of particular importance for the improvement of the performance of the networks. There are many approaches along the line with their own merits and demerits (Stron and Price, 1995, 1997; Price et al., 2005). This paper discusses the use of differential evolution to reveal hidden centers and spreads. The motivation using differential evolution (DE) over other EAs (Michalewicz, 1996) such as GAs (Goldberg, 1989) is that in DE string encoding is typically represented as real valued vectors, and the perturbation of solution vectors is based on the scaled difference of two randomly selected individuals of the current population. Unlike GA, the resulting step size and orientation during the perturbation process automatically adopt to the fitness function landscape. The justification behind combining the idea of feature selection, data inconsistency removal with classification is to reduce the space, time, and thereby enhancing accuracy.

This paper is set out as follows. Section 2 gives an overview of the RBF network, feature selection, feature consistency, and differential evolution. In Section 3, the proposed method is discussed. Experimental setup, results, and analysis are presented in Section 4. Section 5 concludes the paper with a future line of research.

2. Background

The background of this research work is presented in this section. In Section 2.1, RBF network classifier is discussed. Feature selection and its importance are the focus of Section 2.2. Feature inconsistency and the method used in this paper is the topic of discussion in Section 2.3. Differential evolution of a new meta-heuristic computing paradigm is discussed in Section 2.4.

2.1. RBF network classifier

In the RBF network a supervised learning algorithm is adopted to build novel and potentially useful models (Naveen et al., 2010). RBF network is a popular artificial neural network architecture that has found wide applications in diverse fields of engineering. It is used in pattern recognition, function approximation, and time series prediction. Further, it is the most widely used alternative neural network model with certain advantages over multi-layered feed forward neural network (MFN) for the task of pattern classification.

RBF network is a three layer feed forward network where each hidden unit implements a radial activation function and each output unit implements a weighted sum of hidden units' output. This network is a special class of neural network in which the activation of a hidden neuron is determined by the distance between the input vector and a prototype vector. Prototype vectors refer to centers of clusters formed by the patterns or vectors in the input space. The centers are determined during RBF training. Fig. 1 shows a typical RBF network.

In the input layer '*n*' number of input neurons exist that connect the network to the environment. The second layer consists of a set of kernel units that carry out a nonlinear transformation from the input space to the hidden space. Some of the commonly used kernel functions are defined in Table 1. Usually, a nonlinear transformation is made based on Gaussian kernel as described in the following equation:

$$\varphi_i(x) = \exp\left(-\frac{||x-\mu_i||^2}{2\sigma_i^2}\right), \tag{1}$$

where ||..|| represents Euclidean norm, μ_i ,σ_i and ϕ_i are center, spread and the output of *i*th hidden unit, respectively. The interconnection between the hidden and the output layers forms weighted

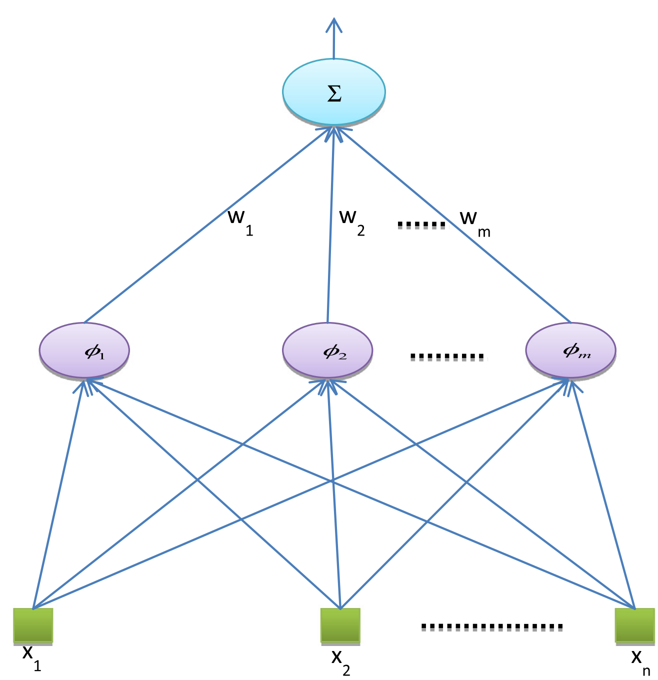

Fig. 1. Pictorial representation of a radial basis function network.