Neurocomputing 149 (2015) 65–70

Contents lists available at ScienceDirect

Neurocomputing

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

Ensemble of extreme learning machine for remote sensing image classification

Min Han *, Ben Liu

Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116023, China

ARTICLE INFO

Article history:
Received 27 June 2013
Received in revised form
5 August 2013
Accepted 8 September 2013
Available online 1 August 2014

Keywords:
Remote sensing classification
Nonnegative matrix factorization (NMF)
Extreme learning machine (ELM)
Ensemble learning
Feature extraction

ABSTRACT

There are only a few of labeled training samples in the remote sensing image classification. Therefore, it is a highly challenging problem that finds a good classification method which could achieve high accuracy to deal with these data. In this paper, we propose a new remote sensing image classification method based on extreme learning machine (ELM) ensemble. In order to promote the diversity within the ensemble, we adopt feature segmentation and then feature extraction with nonnegative matrix factorization (NMF) to the original data firstly. Then ELM is chosen as base classifier to improve the classification efficiency. The experimental results show that the proposed algorithm not only has high classification accuracy, but also handles the adverse impact of a few of labeled training samples in the classification of remote sensing image well both on the remote sensing image and UCI data.

1. Introduction

Remote sensing (RS) image classification is an approach to distinguish class attributes and distribution of ground objects based on the feature of material electromagnetic radiation information in the remote sensing images [1]. It's a hot topic in the field of remote sensing. Though the remote sensing image has large number of data samples, the data types are complex and few data samples are available for training. So it is difficult to be classified well only by a single classifier [2–5]. Some researchers have proposed ensemble algorithms to solve the problem. Chi [6] proposed an ensemble learning algorithm which combined generative and discriminative models for remote sensing image classification. But it worked only for hyperspectral remote sensing image which has low generalization. Pan [7] integrated the rough set with the genetic algorithm in order to reduce the number of input features to a single classifier and to avoid bias caused by feature selection.

There are two main factors affecting the performance of ensemble learning algorithms: the accuracy of base classifiers and the diversity among the base classifiers [8–12]. Therefore, how to increase the diversity within the ensemble and keep a high accuracy in the base classifiers is an urgent problem. To promote the diversity, García-Pedrajas [13] gave weights to every base classifier for each training phase. However, the algorithm was sensitive to mis-indexing data which could cause over fitting. Rodriguez [14] proposed rotation forest (ROF) of which the diversity was improved through applying feature extraction to subsets of features for each base classifier by PCA. Decision trees method were chosen as base classifiers in ROF. But they were not suitable for remote sensing image classification. Zhou [15] put forward UDEED which used unlabeled data to enhance ensemble diversity. The basic idea of UDEED was to maximize the proper on the labeled data and the diversity of the unlabeled data. However, it can only be used for binary classification problems.

In this study, we present a remote sensing image classification method based on NMF [16,17] and ELM [18–20] ensemble (NMF–ELM). Diversity is improved by feature segmentation and then feature extraction with NMF for each base classifier. Due to the fast capability and good generalization performance of ELM, we choose it as base classifier to promote the efficiency of remote sensing image classification. Simulation results substantiate the proposed method on both remote sensing image data sets and UCI data sets.

The rest of this paper is organized as follows. Section 2 presents the basic idea of proposed algorithm NMF–ELM in detail. Section 3 gives two simulation results. The conclusions are given in Section 4.

2. ELM based ensemble algorithm for RS classification

NMF–ELM is an algorithm running in parallel. It means that every base classifier in NMF–ELM can be trained simultaneously. In NMF–ELM, to create the training data for a base classifier, the feature set is randomly split into some subsets and then NMF is applied to each subset. Because the fast capability and good generalization, ELM is chosen as base classifier for NMF–ELM. The proposed algorithm will be described in detail in next three subsections.

* Corresponding author.
E-mail address: minhan@dlut.edu.cn (M. Han).

http://dx.doi.org/10.1016/j.neucom.2013.09.070
0925-2312/

2.1. Nonnegative matrix factorization

The remote sensing image data has a characteristic of nonnegative. When we process these data by linear representation method, decomposition results are required to be nonnegative. In this case, if we use the traditional factor analysis method (such as PCA) to process these data, it may lose the physical meaning, because its results may contain negative numbers. But the use of nonnegative matrix factorization can avoid this problem effectively.

NMF [16] is a matrix factorization method which gives a nonnegative constraint to each element in the treated matrix. Let F be a $M \times N$ matrix where each element is nonnegative. Then to decompose F into W and H:

$$F \approx WH \tag{1}$$

Donate W as basic matrix in the form of a $M \times T$ matrix. Donate H as coefficient matrix in the form of a $T \times N$ matrix. When T is smaller than M, we can choose the coefficient matrix to replace the original data matrix in order to achieve dimensionality reduction. At the same time, because of the nonnegative constraint of each element in the decomposition process, they exists additive joint only. After decomposition, the matrix W and H can maintain the feature information of the original matrix well.

For more detail, in order to find the approximate decomposition of $F \approx WH$, we should define a cost function to measure the similarity between F and WH. Usually the Squared Euclidean distance is used to measure it:

$$E(W,H) = \|F - WH\|^2 = \sum_{i,j} (f_{ij} - (WH)_{ij})^2 \tag{2}$$

where f_{ij} is the value of F in ith row and jth column, $1 \le i \le M$, $1 \le j \le N$.Then, NMF aims to minimize the cost function $E(W,H)$:

$$\min\ E(W,H),\quad W \ge 0,\quad H \ge 0 \tag{3}$$

In order to implement Eq. (3), we adopt Eq. (4) and Eq. (5) to solve W and H iteratively. According to this rule, we could ensure that W and H converge to a local optimal solution.

$$\overline{H}_{ij} = H_{ij} \frac{(W^T F)_{ij}}{(W^T W H)_{ij}} \tag{4}$$

$$\overline{W}_{ij} = W_{ij} \frac{(F H^T)_{ij}}{(W H H^T)_{ij}} \tag{5}$$

In this paper, NMF is used to promote diversity for each base classifier in NMF–ELM. In order to keep all data information, we don't reduce the dimension of the original data. In other words, the dimension of F and H are equal. This will not only increase the diversity, but also retain the original data as much as possible.

2.2. Extreme learning machine

There are many neural networks could be improved by ensemble methods, such as radial basis function neural network [21] and standard neural network model [22]. But ELM has a unique advantage. ELM is a feedforward neural network with a simple three-layer structure: input layer, hidden layer and output layer. Let n be the input layer node number, let r be the hidden layer node number, and let c be the output layer node number. For N different samples (x_i, l_i), $1 \le i \le N$, where $x_i = [x_{i1}, x_{i2}, ..., x_{in}]^T \in R^n$, $l_i = [l_{i1}, l_{i2}, ..., l_{ic}]^T \in R^c$ the mathematical expression of ELM is shown in formula (6):

$$o_k = \sum_{i=1}^{r} \beta_i g(w_i \cdot x_k + b_i), k = 1, 2, ..., N \tag{6}$$

where $o_k = [o_{k1}, o_{k2}, ..., o_{kc}]^T$ is the network output value, $w_i = [w_{i1}, w_{i2}, ..., w_{in}]$ is the weight vector connecting the ith hidden node and the input nodes, $\beta_i = [\beta_{i1}, \beta_{i2}, ..., \beta_{ic}]^T$ is the weight vector connecting the ith hidden node and the output nodes, $g(x)$ is activation function, generally set as Sigmoid function, and b_i is the threshold of the ith hidden node.

At the beginning of training, w_i and b_i are generated randomly and kept unchanged. And β is the only parameter to be trained. The mathematical expression is shown in formula (7) and (8):

$$H = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & \dots & g(w_r \cdot x_1 + b_r) \\ \vdots & \cdots & \vdots \\ g(w_1 \cdot x_N + b_1) & \dots & g(w_r \cdot x_N + b_r) \end{bmatrix}_{N \times r} \tag{7}$$

$$\beta = H^{\dagger} L \tag{8}$$

where H is called the hidden layer output matrix of the neural network, $\beta = [\beta_1, \beta_2, ..., \beta_r]^T$ is the output weight vector, $L = [l_1, l_2, ..., l_N]^T$ is the desired output vector. When β is solved, ELM network training process is completed.

From the above we can see ELM just need to set the number of hidden nodes but don't change the input weights and the threshold of hidden layer. And only one optimal solution is generated. So it has good generalization performance and fast learning speed.

2.3. NMF–ELM algorithm

NMF–ELM is an ensemble algorithm based on NMF and ELM. The main idea of NMF–ELM is to achieve good classification result through promoting diversity. The structure of NMF–ELM is show in Fig. 1.And the structure of ELM base classifier is shown in Fig. 2. Specifically, the diversity is improved by feature segmentation and then feature extraction with NMF.

Let $p = [p_1, p_2, ..., p_n]^T$ be a sample point described by n features. Let P be the sample set containing the training objects in the form of a $n \times N$ matrix. Let $Y = [y_1, y_2, ..., y_N]$ be a vector with class labels for the data, where y_i takes a value from the set of class labels $\{l_1, l_1, ..., l_c\}$, and c is the number of labels. Denote by $B_1, B_2, ..., B_q$ the classifier in the ensemble and by F, the feature set, where q is the number of classifiers. NMF–ELM is described as below:

(1) Split the feature set F randomly into K disjointed subsets. And each feature subset contains m features. Let m be an adjustable variable. If m is divided exactly by F, then$K = F/m$. If F divided by m gives out a quotient of U with a remainder of V, then $K = U + 1$ and U feature subsets contains m features and the last one feature subset contains V features.

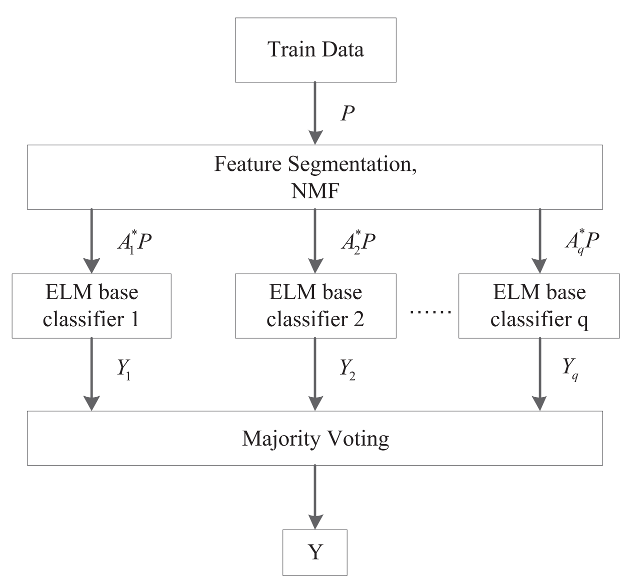

Fig. 1. The structure of NMF–ELM.