Contents lists available at ScienceDirect

Applied Soft Computing

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

Extended semi-supervised fuzzy learning method for nonlinear outliers via pattern discovery

Xiaoning Song[a,b,c,*], Zi Liu[d], Xibei Yang[c,d], Jingyu Yang[d], Yunsong Qi[c]

[a] School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China
[b] Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford GU2 7XH, UK
[c] School of Computer Science and Engineering, Jiangsu University of Science and Technology, Zhenjiang 212003, China
[d] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

ARTICLE INFO

Article history:
Received 24 July 2013
Received in revised form 2 December 2014
Accepted 24 December 2014
Available online 19 January 2015

Keywords:
Fuzzy discriminant analysis
Parameter estimation
Outlier instances
Semi-supervised clustering
Image recognition

ABSTRACT

This article presents an extended Parameterized Fuzzy Semi-supervised learning (PFSL) method, in which the key innovation is the capability of separating a sample set into two independent subsets: outlier sample subset and regular sample subset. In our proposed PFSL, we first develop an improved parameterized Fuzzy Linear Discriminant Analysis (F-LDA) algorithm to classify regular samples, in which the distribution information of each sample in terms of fuzzy membership degree is incorporated with the redefined within-class and between-class scatter matrices. To achieve good parameter estimation for this improved F-LDA, we advocate the use of Hopfield Neural Networks (HNN) due to its efficiency. Second, a new semi-supervised Fuzzy C-Means (S-FCM) algorithm is designed using pre-computed cluster number and cluster centers in the supervised pattern discovery stage. It is applied to classify the remaining outlier samples and generate the final classification result. Third, since Kernel Fisher Discriminant (KFD) is an efficient way to extract nonlinear discriminant features, a kernel version of the proposed PFSL (K-PFSL) is discussed. Extensive experiments on the ORL, NUST603, FERET and Yale face datasets show the effectiveness and the superiority of the proposed algorithm.

1. Introduction

Semi-supervised learning (SSL) aims at using the combination of labeled and unlabeled samples to construct a classifier. As labeled data is insufficient in many scenarios and unlabeled samples are abundant, SSL is very helpful for better exploiting the information of all these samples. From the previous studies [1,2], we conclude that the information obtained from the distribution of unlabeled data improves the potential of learning from the labeled samples of classifiers. Despite the great success of SSL methods in many applications, some issues are not properly solved. (1) The selections of the number of clusters and the initial center of each cluster are still non-trivial tasks. Many unsupervised clustering algorithms also require high computational load because they involve an extensive search process such as forward selection or backward elimination [3,4]. Therefore, these issues mentioned above are major limitations of the unsupervised learning algorithms. (2) Though SSL methods are specifically trained to identify a set of patterns, they often fail in the cases where some samples called outliers with uncommon behavior are difficult to be accurately classified. The detection of outlier samples is an unavoidable task because traditional SSL methods are dependent on the availability of the training set consisting of regular samples, which might misclassify a normal observation that falls outside the trained boundary. Therefore, the major drawback of traditional SSL methods is that the training set must represent all possible classes [5].

The importance of outlier samples analysis in pattern recognition is that the useful anomalous feature information hidden in large data sets could be discovered by evaluating the correlation between each pattern to identify unusual samples. A good image feature representation method could significantly reduce the computational complexity of a model. The methods that utilize the different results of output variables to identify the best subset of given features in a dataset can be divided into supervised and unsupervised methods [6–10]. Despite extensive studies have used supervised or unsupervised models to exploit effective feature representations, few attempts have been made to identify important outlier instances by means of the SSL methods [11–13].

* Corresponding author at: School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China. Tel.: +86 13905280582.
E-mail addresses: xnsong@hotmail.com, xnsong@aliyun.com (X. Song).

http://dx.doi.org/10.1016/j.asoc.2014.12.034
1568-4946/© 2015 Elsevier B.V. All rights reserved.

In order to classify high-dimensional images, it is necessary to extract features that enable different feature regions to be well separated. In fact, image recognition can be either supervised or unsupervised, depending on whether the prior knowledge or the label of a training sample is available or not. In many learning problems, the difficulty most often encountered is prior knowledge analysis about pattern variations [14,15]. A supervised learning method identifies and separates regions that match feature properties previously learned from training samples. On the contrary, unsupervised feature segmentation has to distinguish the feature classes as well as to separate them into different regions. Although some attempts have been made to incorporate a supervised classifier into an unsupervised strategy [16,17], those previous studies only classify a small number of features that correspond to patterns with low-confidence obtained using an unsupervised algorithm. In fact, if supervised classifier fails in the remaining features, which is mainly due to its unsupervised attributes.

In this study, we propose an extended parameterized fuzzy semi-supervised learning method that separates data into two independent regions, including outlier instances or regular samples, contained in a feature space. More specifically, firstly, an improved supervised F-LDA algorithm for regular samples is proposed. It achieves the distribution information of each sample that is represented with fuzzy membership degree, and then the membership grade is incorporated into the redefinition of scatter matrices. As a result, the initial fuzzy classification of whole regular feature space is obtained. Moreover, the need for such a novel F-LDA model construction is reduced to parameter estimation when the structure of learning model is given beforehand, that is, the parameter estimation method must recursively process the measured data as they become available. Secondly, a new semi-supervised fuzzy C-means (S-FCM) algorithm is presented on the basis of precise number of clusters and initial pattern centers that obtained in the pattern discovery stage, which are then applied to perform the outlier instances classification and to yield the final classification result. Thirdly, since the Kernel Fisher Discriminant (KFD) algorithm is an effective way to extract nonlinear discriminative information of the feature space by using the kernel trick [18–21], a kernel version of our method is presented subsequently, which has the potential to outperform the traditional learning algorithms, especially in the case of nonlinear small sample size. We compared the proposed method PFSL and the kernel version of PFSL (K-PFSL) with various face recognition methods including Fisherface [22], direct LDA (D-LDA) [23], complete PCA plus LDA (C-LDA) [24], random discriminant analysis (R-DA) [25], fuzzy linear discriminant analysis (F-LDA) [26], fuzzy local discriminant embedding (F-LDE) [27], fuzzy maximum margin criterion (F-MMC) [28], reformative fuzzy linear discriminant analysis (RF-LDA) [29], bias corrected FCM (BC-FCM) [30], K-means-based support vector machine (K-Means + SVM) [31], Mean shift-based support vector machine (Mean Shift + SVM) [32], Gaussian mixtures-based support vector machine (Gaussian Mixtures + SVM) [33] and Density-based spatial clustering of applications with noise based K-nearest neighbor classification (DBSCAN + KNN) [34], on the ORL [35], NUST603 [36], FERET [37] and Yale [38] face image databases.

The motivations of this article are summarized as follows: (1) The proposed method validates the effectiveness of a supervised pattern discovery model associated with a semi-supervised manner. A semi-supervised fuzzy clustering algorithm is developed on the basis of prior information that obtained in supervised learning stage, which is then applied to perform outliers classification. (2) The objective of the proposed supervised learning step is to determine a set of suitable patterns, but this step could not accurately distinguish the samples from abnormal training regions such as outlier zones. That is the reason why we propose a new semi-supervised method of clustering to address this issue. (3) We can find that the proposed method is stable for the cases with only a small number of training samples, especially for the non-linearly separable problems, which again validates the advantage of the proposed algorithm in alleviating the nonlinear small sample size problem.

This article is organized in the following manner. Section 2 presents a fuzzy supervised learning method with parameter estimation. Extended semi-supervised fuzzy clustering for outlier samples is proposed in Section 3, meanwhile, a kernel version of our method is presented in this section. Section 4 reports comprehensive classification results on several commonly used face databases, including NUST603, ORL, Yale and FERET. Finally, concluding comments are included in Section 5.

2. Fuzzy supervised learning method with parameter estimation

In our previous work [29], we extended F-LDA [26] by including complete fuzziness in the calculations of the between-class and within-class scatter matrices. By this means, a relaxed normalized condition is presented to achieve the distribution information of each sample in terms of fuzzy membership degree.

Nevertheless, how can we dynamically assign a particular value of the offset of membership grade? There is not explained in our previous work [29]. Moreover, the need for such a novel F-LDA model construction is reduced to parameter estimation when the structure is given beforehand. The parameter estimates should then be based on observations up to the current time, and therefore the parameter estimation methods must recursively process the measured data as they become available. Traditional methods include recursive least-squares and the Kalman filter for parameter estimation [39]. Newer alternative methods include HNN for parameter estimation [40–42]. Therefore, in this Section, the HNN has been further considered in the context of the problem of the proposed fuzzy LDA parameter estimation.

The objective of the proposed initial step is to find out a set of suitable patterns, but this step does not accurately define all sample regions. This process is carried out by using all of training samples and separates training set that matches different sample properties (including outlier instances and regular ones). Henceforth, a pattern discovery stage that relies on an improved supervised fuzzy LDA approach is proposed for determining the original patterns of training samples, which is based on the outcome of identification of outlier instances and their counterparts.

2.1. Traditional LDA and F-LDA

The conventional LDA methods [43–47] aim at maximizing the ratio of between-class scatter matrix to within-class scatter matrix. Given a set of n_i samples belonging to class c_i, we can define the mean of each class as:

$$m_i = \frac{1}{n_i}\sum_{x_k \in c_i} x_k \tag{1}$$

where $i = 1, 2, \ldots, C$, C is the number of classes. The within-class scatter matrix is then defined as:

$$S_w = \frac{1}{N}\sum_{i=1}^{C}\sum_{x_k \in c_i}(x_k - m_i)(x_k - m_i)^T \tag{2}$$

where N is the total number of image samples $N = \sum_{i=1}^{C} n_i$. The between-class scatter matrix is defined as:

$$S_b = \frac{1}{C}\sum_{i=1}^{C}(m_i - \bar{m})(m_i - \bar{m})^T \tag{3}$$