Information Sciences 444 (2018) 1–19

Contents lists available at ScienceDirect

Information Sciences

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

Flexible semi-supervised embedding based on adaptive loss regression: Application to image categorization

Y. El Traboulsi[a], F. Dornaika[a,b,*]

[a] Department of Computer Science and Artificial Intelligence, University of the Basque Country UPV/EHU, San Sebastian, SPAIN
[b] IKERBASQUE, Basque Foundation for Science, Bilbao, SPAIN

ARTICLE INFO

Article history:
Received 24 July 2017
Revised 11 February 2018
Accepted 18 February 2018
Available online 21 February 2018

Keywords:
Graph-based semi-supervised embedding
Flexible embedding
Feature extraction
Out-of-sample extension
Adaptive loss function
Regression
Image categorization

ABSTRACT

This paper introduces two graph-based semi-supervised embedding methods for generic classification and recognition tasks. These proposed methods combine the merits of Flexible Manifold Embedding, non-linear graph based embedding, and adaptive loss function. The adaptive loss function seems to be a good choice for reaching a flexible and adaptive regressor in the sense that the effect of outliers is reduced and the regression function is more elastic and more robust. Furthermore, unlike label propagation approaches, the proposed methods provide a data embedding into a space whose dimension is not limited to the number of classes. The adaptive loss function combines the merits of the $\ell_{1,2}$ and ℓ_2 matrix norms. It can handle the Laplacian distribution of outliers and the Gaussian distribution of samples with small loss. We provide extensive experiments on eight benchmark datasets in order to study the performance of the proposed methods. These experiments show that the proposed methods can be more discriminative than other state-of-the-art methods.

1. Introduction

Dimension reduction methods have greatly attracted the attention of researchers [5,11,34] in recent times due to their ability to accelerate many real applications. By adopting these paradigms, pattern recognition methods become practical even if they are applied on very high-dimensional data [18,45].

Principal Component Analysis (PCA) is the most popular dimensionality reduction method. PCA is an unsupervised method that aims to represent high dimensional samples in a lower dimensional space through a linear transform. Local Discriminant Analysis (LDA) [13] is a supervised method that projects samples from their original space to another low dimensional space, and brings, at the same time, samples from the same class as close as possible to each other.

In many real-world problems, collecting a large number of labeled samples is practically impossible. The reason is twofold. First, these labeled samples can be very few. Second, for some applications (biometrics, biology) it is very expensive to obtain labels in a manual way. On the other hand, unlabeled samples are easily accessible [14,16,40]. Semi-supervised methods benefit from labeled samples and unlabeled samples [50]. The labeled samples explicitly contribute to the discrimination ability of the final model, and the unlabeled ones maintain the geometrical structure of data. In recent years, graph-based semi-supervised learning methods have become the subject of many research works because of their success

* Corresponding author at: University of the Basque Country UPV/EHU, Manuel Lardizabal 1, 20018 San Sebastian, SPAIN.
E-mail address: fadi.dornaika@ehu.es (F. Dornaika).

https://doi.org/10.1016/j.ins.2018.02.044

in many domains [2,10,32,43,46,48,49]. In this context, the supervised Local Discriminant Embedding (LDE) [5] and LDA methods are extended to their semi-supervised versions. Namely, Semi-supervised Discriminant Embedding (SDE) [17] and Semi-supervised Discriminant Analysis (SDA) [3], by adding a regularizer whose role is to maintain the data smoothness.

The above mentioned property imposes that close data should remain close after embedding (projection). Graph-based semi-supervised embedding techniques receive a lot of attention due to the recent progresses made in the field of graph construction [27,39,44,47].

The linear techniques such as PCA, LDA, LDE, SDA and SDE have been increasingly important in pattern classification since they provide a relatively simple projection of data onto a lower-dimensional subspace, leading to computationally efficient classification schemes. However, these methods can have some limitations in the sense that they are strictly linear methods. To overcome this shortcoming, researchers proposed many non-linear methods (e.g., [19,30,31]). Locally Linear Embedding (LLE) [33] and Laplacian Eigenmap (LE) [1] are recently proposed as non-linear methods. These methods can beat other methods by the fact of non-linearity, but they suffer, on the other hand, from the out-of-sample problem. In other words, these methods cannot explicitly provide the projection or the label of an unseen sample. In [25], the authors proposed the Locality Preserving Projections (LPP) method which is a linear approximation to the non-linear Laplacian Eigenmaps. These methods attempt to preserve data smoothness among nearby points. We can thus understand the requirement of a graph-theoretic learning framework for these methods. Indeed, graph methods are based on an adjacency weight matrix. In general, the construction of this matrix depends on the computed distances between samples (this is very often given by the Euclidean distance). Qiao et al. proposed the Sparsity Preserving Projections (SPP) algorithm [29]. They propose to construct an adjacency weight matrix based on a modified sparse representation framework. They devise a linear embedding that preserves the sparsity relations among all samples. In [9], we proposed the Semi-supervised Flexible Feature Extraction (SFFE) framework. This framework predicts the embeddings of unlabeled data in a non-linear manner. It estimates, at the same time, a linear regression function that can easily transform test (unseen) samples to their new subspace, thus estimating their labels.

In a semi-supervised context, the way the regression function is computed can be crucial in order to get a good model that does not suffer from the out-of-sample problem [12]. This involves the regression model (linear or non-linear) as well as the loss function used for estimating the model. While these concepts are clear for a cascaded estimation, they become more challenging for a simultaneous estimation of the non-linear embedding and the regression model. In general, the regression model is estimated by minimizing a loss function between the regressed data and the embeddings (usually assumed to be non-linear projections consistent with the semi-supervised criterion over the training data). In most cases, this minimization is made following a rigid norm, namely ℓ_1 or ℓ_2 norm. In this paper, inspired by our previous work on simultaneous semi-supervised embedding and flexible regression, we propose an Adaptive Semi-supervised Flexible Feature Extraction (ASFFE). There are two main differences with our previous approach. Firstly, our approach incorporates non-linear sparsity preserving property instead of the classic locality preserving criterion. Secondly, in our proposed Adaptive Semi-supervised Flexible Feature Extraction (ASFFE), the loss function is an adaptive norm that is in between ℓ_1 norm and ℓ_2 norms. This norm takes into account the outlier samples (distributed according to a Laplacian distribution) as well as normal samples (distributed according to a Gaussian distribution). We propose two versions of ASFFE: ASFFE-LE and ASFFE-SPP. The first one is based on Laplacian smoothness and the second one is based on non-linear sparsity preserving. We will show that adaptive method is more robust than the traditional approach and the state-of-the-art methods.

The contributions of the proposed methods can be summarized as follows:

- In addition to predicting embeddings of unlabeled samples in a non-linear way, our proposed methods make it possible to predict embeddings for all unseen samples in an easy and adaptive linear way.
- The proposed methods predict embeddings of samples and not only their labels. Thus, we have flexibility in the choice of the classifier that can be deployed after the embedding is estimated.
- In our proposed methods, the loss function of the regression error (i.e., residual between non-linear data projection and their regressed version) is set to an adaptive loss function which combines the merits of ℓ_1 and ℓ_2 norms. The mainstream for regression loss function is to use the ℓ_2 norm which cannot behave well in the presence of outlier labels or large discrepancies associated with the regressed data. The exploited loss function makes the final embedding flexible and adaptive (not very sensitive to large residuals).
- As already mentioned, ASFFE-SPP is based on sparse representation. This sparsity preserving property provides an additional discriminative power to the proposed framework since the sparse representation is somehow discriminative by its nature.
- The proposed methods inherit the properties of SFFE, since SFFE can be considered as a special case of ASFFE.

For the sake of clarity, Table 1 summarizes the main acronyms used in this paper.

The remainder of this paper is organized as follows. In Section 2, we provide a brief review of the main semi-supervised learning methods. Our proposed methods are introduced in Sections 3 and 4. In Section 5, we present experimental results made on eight benchmark datasets. We finally present our conclusion in Section 6. In this paper, capital bold letters denote matrices and small bold letters denote vectors.