Neurocomputing 190 (2016) 197–208

Contents lists available at ScienceDirect

Neurocomputing

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

Class-driven concept factorization for image representation

Huirong Li [a,b], Jiangshe Zhang [a,*], Junmin Liu [a]

[a] School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China
[b] School of Mathematics and Computer Application, Shangluo University, Shangluo 726000, China

ARTICLE INFO

Article history:
Received 17 August 2015
Received in revised form
4 January 2016
Accepted 14 January 2016
Communicated by Jun Yu
Available online 2 February 2016

Keywords:
Concept factorization
Semi-supervised learning
Label information
Class-driven constraint

ABSTRACT

Recently, concept factorization (CF), which is a variant of nonnegative matrix factorization, has attracted great attentions in image representation. In CF, each concept is modeled as a nonnegative linear combination of the data points, and each data point as a linear combination of the concepts. CF has impressive performances in data representation. However, it is an unsupervised learning method without considering the label information of the data points. In this paper, we propose a novel semi-supervised CF method, called class-driven concept factorization (CDCF), which associates the class labels of data points with their representations by introducing a class-driven constraint. This constraint forces the representations of data points to be more similar within the same class while different between classes. Thus, the discriminative abilities of the representations are enhanced in the image representation. Experimental results on several databases have shown the effectiveness of our proposed method in terms of clustering accuracy and mutual information.

1. Introduction

In the past years, image representation has become a fundamental problem in various research fields such as machine learning, computer vision and pattern recognition [1–4]. To do image representation, we have to deal with high dimensional data. In general, the essential structures of images are laid in a relatively low dimensional space. However, traditional methods cannot infer the low dimensional structures well from their high dimensional data space. Many methods have been proposed to find the better representations of images. Among them, matrix factorization, which aims to find two or more low rank matrices whose product provides a good approximation to the original data, is one of the most popular techniques.

Nonnegative matrix factorization (NMF) [1,2] can obtain interpretable representations of the data points with the nonnegative constraints. These nonnegative constraints lead NMF to a part-based representation, because they allow only additive, not subtractive, combinations of the original data points. The advantages of the part-based representation are benefit in various real world applications such as document clustering [5–7], face recognition [8,9], image representation [10], and DNA gene expression analysis [11]. In particular, NMF represents data as a linear combination of a set of basis vectors, in which both the combination coefficients and the basis vectors are nonnegative. However, in many practical applications, NMF fails to be performed on negative data owing to the nonnegative constraints. In addition, NMF can only be used in original feature space of the data, and thus cannot make use of the power of kernelization [12].

To overcome the drawbacks of NMF, Xu and Gong proposed the concept factorization (CF) algorithm [13]. With this model, the data clustering task is accomplished by computing the two sets of linear coefficients, and the computation of linear coefficients is carried out by minimizing the reconstruction error of the data points. Compared with NMF, CF can not only be kernelized, but also be performed [14]. Recently, Cai et al. [15] developed a locally consistent concept factorization (LCCF) algorithm which encoded the geometrical information of data space by constructing a nearest neighbor graph to model the local manifold structure. The main drawback of LCCF is that it only focuses on the local structure in the data, which may often lead to over-fitting. Liu et al. [16] proposed a novel semi-supervised matrix decomposition method call constrained concept factorization (CCF) that extracts the image concepts consistent with the known label information. The CCF model can guarantee that the data points sharing the same label are mapped into the same concept in the low dimensional space. However, since CCF maps the images with the same label onto the same concept, it is infeasible when there is only one labeled data point to rely on. Therefore, Xiao et al. [17] developed a semi-supervised class-driven NMF method to associate a class label with each basis vector by introducing an inhomogeneous

* Corresponding author.
E-mail address: jszhang@mail.xjtu.edu.cn (J. Zhang).

http://dx.doi.org/10.1016/j.neucom.2016.01.017
0925-2312/

representation constraint. Additionally, some further studies of the matrix factorization have also been developed [14,18–21] based on different constraints during the last few years.

However, CF is an unsupervised learning algorithm. That is, CF cannot differentiate the labeled data from the unlabeled data, and thus cannot take advantage of the label information when such information is available. Therefore, how to use the label information of the data in hand to improve the performance of the algorithms has become a hot topic in machine learning. Many machine learning researchers have pointed out that when a small amount of labeled data is used in conjunction with unlabeled data, it can produce encouraging improvement in learning performance [10,12,18,19,22,23]. In the many practical learning problems, such as image processing [24,25], text classification [26] and bioinformatics [27], we usually need to handle situations when a small size of labeled with a large amount of unlabeled data is available. The unlabeled data are usually much easier to obtain but quite expensive to identify their labels. So it is of great interest both in theory and in practice if we can extend CF to a semi-supervised concept factorization version.

In this paper, we propose a novel semi-supervised concept factorization method by employing the class-driven constraint, which is referred to class-driven concept factorization (CDCF). Class-driven constraints have been used in semi-supervised learning [17], however, to our knowledge, they have not been incorporated into the CF framework. This constraint forces the representations of data points to be more similar within the same class while different between classes. Thus, data points belonging to the same class will have similar representations, and the obtained new representations can have more discriminative power. To achieve this, we carefully design a new concept factorization objective function incorporating the class-driven constraints into it. We also develop an optimization scheme for the objective function to derive the iterative updating rules of two matrices **W** and **V**, the convergence proof of our algorithm is provided. Experiments show that our proposed CDCF method achieves better recognition accuracy and normalized mutual information than some recent variants of NMF. It is worth noting that the proposed CDCF model is closely related to the transductive learning (TL) concept, which was proposed by Vapnik [28] in 1998. Given a testing data set with a small size of labeled data and a large amount of unlabeled data, different from semi-supervised learning (SSL) which usually attempts to obtain a decision rule that can be used for classification of unlabeled data that are different from the testing data, TL aims at identifying the labels of the unlabeled testing data [29,30]. Our proposed CDCF model first tries to learn the discriminative efficient representations of the unlabeled testing data with the help of labeled testing data, and then exploits the labels of the unlabeled data belonging to the testing data set based on their representations. In this sense, the proposed CDCF model can be grouped into TL.

The rest of this paper is organized as follows. We provide a brief review of NMF and CF in Section 2. The proposed CDCF model as well as the optimization method are described in Section 3. Experimental results are reported in Section 4 with considerable analysis. Finally, we draw a conclusion in Section 5 and the proof of convergence of the algorithm is provided in Appendix.

2. Related work

2.1. Nonnegative matrix factorization (NMF)

Given a nonnegative matrix $\mathbf{X} = [x_1, x_2, \ldots, x_n] \in \mathbf{R}^{m \times n}$, where each column vector $x_i \in \mathbf{R}^m$ represents a data point. NMF aims to find two nonnegative matrix factors $\mathbf{U} \in \mathbf{R}^{m \times k}$ and $\mathbf{V} \in \mathbf{R}^{n \times k}$ to approximate the original matrix $\mathbf{X}$ as follows:

$$\mathbf{X} \approx \mathbf{U}\mathbf{V}^T$$

In [31], the square of the Euclidean distance and generalized Kullback–Leibler divergence are used to measure the quality of the approximation. For brevity, we just give a discussion of the former one. In this case, the optimization problem becomes:

$$\min_{U,V} O = \|\mathbf{X} - \mathbf{U}\mathbf{V}^T\|_F^2 \quad \text{s.t.} \quad \mathbf{U} \ge 0, \ \mathbf{V} \ge 0 \tag{1}$$

where $\|\cdot\|_F$ is the matrix Frobenius norm denoting the square root of the squared sum of all the entries in the matrix. It is easy to check that the objective function of NMF is not convex in both variables **U** and **V** together. Therefore, it is hard to find the global solution of the objective function of NMF. The most popular optimization method of NMF is the multiplicative updating algorithm proposed by Lee and Seung [1]. The objective function of NMF in Eq. (1) can be solved by the following updating rules:

$$u_{ij} \leftarrow u_{ij} \frac{(\mathbf{X}\mathbf{V})_{ij}}{(\mathbf{U}\mathbf{V}^T\mathbf{V})_{ij}}$$

$$v_{ij} \leftarrow v_{ij} \frac{(\mathbf{X}^T\mathbf{U})_{ij}}{(\mathbf{V}\mathbf{U}^T\mathbf{U})_{ij}}$$

In NMF, each column vector of **U** can be regarded as a basis and each data point x_i is approximated by a linear combination of these k bases, weighted by the components of **V**. Therefore, NMF maps the data matrix **X** to **V** from m-dimensional space to k-dimensional space ($k \ll m$).

2.2. Concept factorization (CF)

NMF can be used in the original feature space of the data. However, it cannot make use of the power of kernelization when the data are highly non-linear distributed. To overcome this drawback of NMF, Xu and Gong proposed concept factorization (CF) which is an extension of NMF [13]. In CF, each basis vector u_j can be represented as a linear combination of the data samples x_i:

$$u_j = \sum_i w_{ij} x_i$$

where $w_{ij} \ge 0$. Let $\mathbf{W} = [w_{ij}] \in R^{n \times k}$, CF is to factorize the data matrix **X** as:

$$\mathbf{X} \approx \mathbf{X}\mathbf{W}\mathbf{V}^T$$

The quality of the approximation can be quantified by using a cost function with Euclidean distance metric. Then it turns to minimize the following objective function:

$$\min_{W,V} O = \|\mathbf{X} - \mathbf{X}\mathbf{W}\mathbf{V}^\mathbf{T}\|_F^2 \quad \text{s.t.} \quad \mathbf{W} \ge 0, \ \mathbf{V} \ge 0 \tag{2}$$

According to [13], the updating rules of CF in Eq. (2) can be introduced as follows:

$$w_{ij} \leftarrow w_{ij} \frac{(\mathbf{K}\mathbf{V})_{ij}}{(\mathbf{K}\mathbf{W}\mathbf{V}^\mathbf{T}\mathbf{V})_{ij}}$$

$$v_{ij} \leftarrow v_{ij} \frac{(\mathbf{K}\mathbf{W})_{ij}}{(\mathbf{V}\mathbf{W}^\mathbf{T}\mathbf{K}\mathbf{W})_{ij}}$$

where $\mathbf{K} = \mathbf{X}^\mathbf{T}\mathbf{X}$. We can construct a kernel matrix by using a kernel function. Thus, CF can easily be kernelized to enhance the performance in some case.