Neurocomputing 155 (2015) 43–52

Contents lists available at ScienceDirect

Neurocomputing

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

Learning latent features by nonnegative matrix factorization combining similarity judgments

Jiang-She Zhang [a,*], Chang-Peng Wang [a], Yu-Qian Yang [b]

[a] School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China
[b] School of Mathematics and Statistics, Xidian University, Xi'an, Shaanxi 710075, China

ARTICLE INFO

Article history:
Received 25 April 2014
Received in revised form
17 October 2014
Accepted 15 December 2014
Communicated by S. Choi
Available online 5 January 2015

Keywords:
Nonnegative matrix factorization
Additive clustering
Feature extraction
Face recognition
Document clustering

ABSTRACT

Nonnegative matrix factorization (NMF) is a popular method for learning low-rank approximation of nonnegative matrix. However, aiming at seeking the low-rank approximation from the viewpoint of data reconstruction, NMF may produce unfavorable performances in classification and clustering tasks. In this paper, we develop a novel modification of NMF (called NMFCSJ) by incorporating the similarity judgments of data points into NMF, and then performs a collective factorization on the data matrix and a weighted similarity matrix with a closely related factor matrix. With the superiority of additive clustering, the proposed method NMFCSJ exploits the latent features hidden in the original data. Experiments show that NMFCSJ improves the classification performance on two face databases and achieves better clustering accuracy for semi-supervised or unsupervised document clustering on 9 documents datasets from CLUTO toolkit.

1. Introduction

During the past several years, nonnegative matrix factorization (NMF) [1] has been shown to be useful for many applications in pattern recognition [2,3], text mining [4–7], and DNA gene expressions [8,9] and so on. NMF attempts to decompose a matrix into a product of two nonnegative matrices. Due to the nonnegative constraints imposed on both factors, NMF only allows additive, not subtractive linear combinations of vectors to reconstruct the original vectors. Consequently, NMF can be interpreted as a parts-based representation of the data.

When the whole data or parts are labeled, it is desirable to incorporate the information on class labels into NMF to improve the classification performance. This leads to the supervised or semi-supervised modifications of NMF. One kind of supervised NMF [10,11] incorporate the Fisher linear discriminant constraints into NMF to extract more discriminant features than the original NMF.

Supervised NMF were shown to perform better than the standard NMF in classification tasks. However, in many learning tasks, there is a large supply of unlabeled data but insufficient labeled data since it can be expensive to generate. Semi-supervised learning, which has received more and more attention in recent years, is preferred. Semi-supervised NMF (SSNMF) [12] is one of these methods. SSNMF is a modification of NMF which jointly incorporates the data matrix and the (partial) class label matrix into NMF. SSNMF performs a collective factorization of both data matrix and class label matrix with a common factor matrix. By this way, matrix factorization is guided by the prior information contained in labeled data so that the classification performance of NMF is improved. Another semi-supervised nonnegative matrix factorization (SS-NMF) proposed by Chen et al. [13] exploited the data similarity matrix to infer the clusters.

Similarity matrix is a nonnegative matrix that measures the relationship among a collection of objects. The analysis of the subjective similarity judgments is one of the central problems in cognitive science to determine the mental representations that underlie human inferences. Since subjective similarity cannot be derived from a straightforward analysis of objective stimulus characteristics [14], it is believed that the similarity judgments rely close to the underlying representations on latent features [15].

Additive clustering [16,17] provides a method of assigning a set of latent features to a collection of objects, based on the observable similarities. The observed similarity between two different objects relies on the fact that whether they share the same feature and whether the common feature is salient. The similarity is represented as a weighted linear combination of features. This is a form of nonnegative factorization. Additive clustering aims to uncover a feature matrix and saliency vector that provides a good approximation to the empirical similarities. However, additive clustering is a discrete combinatorial optimization which is unable to accommodate continuously varying properties. Additive fuzzy clustering proposed by Sato et al. [18,19] is a natural extension of the additive clustering model which allows a fuzzy degree of belongingness of

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

http://dx.doi.org/10.1016/j.neucom.2014.12.050
0925-2312/

feature to a cluster. This model has an explicitly probabilistic interpretation [20].

In this paper, we propose a new procedure to learn latent features by nonnegative matrix factorization combining similarity judgments (NMFCSJ). Our new method aims at improving the classification performance by exploiting the available information in similarity judgments. NMFCSJ transfers the columns of encoding vectors into probabilistic vectors, feeds them into the additive fuzzy model and performs a collective factorization of both data matrix and a weighted similarity judgments matrix to learn the latent features. Taking advantage of the weighted matrix, NMFCSJ can be treated as a semi-supervised or supervised variants of NMF which is relied on the fact that whether the similarity matrix is incomplete (i.e., some of pairwise similarities are missing or unobserved.) or not. Furthermore, we incorporate the classic clustering algorithms into NMFCSJ to construct an ensemble clustering solution. In this case, NMFCSJ is an unsupervised document clustering algorithm. Experiments on two face databases (the PIE and Yale database) and document datasets from CLUTO toolkit have shown that NMFCSJ could improve the classification performance and clustering performance by incorporating similarity judgments of data points.

The paper is organized as follows: In Section 2, the related methods including the classical NMF problem, additive clustering model and additive fuzzy clustering model are described briefly. Section 3 presents the proposed NMFCSJ and its learning procedure. Section 4 presents some factorization examples on face databases and document clustering examples on document datasets to demonstrate the advantage of our method. Finally, Section 5 contains some concluding remarks.

The notation adopted in this paper obeys the following rule to facilitate presentation: for any matrix A, A_i means the ith row of A, its corresponding lowercase version a_i means the ith column vector of A, and A_{ij} denotes the elements of A at the ith row and jth column.

2. Related works

In this section, we first briefly review the classical nonnegative matrix factorization [1,21], and then additive clustering model [16,17] and additive fuzzy clustering model [18–20].

2.1. Nonnegative matrix factorization (NMF)

Assume that n nonnegative input p-dimensional vectors $v_i (i = 1, 2, \ldots, n)$, stored in a matrix $V = [v_1, v_2, \ldots, v_n]$. The NMF method is to find a nonnegative $p \times q$ matrix W and another nonnegative $q \times n$ matrix H such that the product of W and H is approximately equal to V, denoted by

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

where W is the basis matrix and H is the encoding matrix.

To obtain such an approximation for V, one needs a cost function to measure the quality of reconstruction error. Based on the Gaussian likelihood, the conventional approach to find W and H is by minimizing the difference between V and WH as follows:

$$\min_{W,H} \quad D(V, WH) = \frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{n}\left(V_{ij} - (WH)_{ij}\right)^2$$
$$\text{s.t.} \quad W_{ia} \geq 0, \quad H_{bj} \geq 0, \quad \forall i, a, b, j \tag{2}$$

The multiplicative update rules proposed by Lee and Seung for solving (2) are given by

$$H_{ab} \leftarrow H_{ab}\frac{(W^T V)_{ab}}{(W^T WH)_{ab}}, W_{ca} \leftarrow W_{ca}\frac{(VH^T)_{ca}}{(WHH^T)_{ca}} \tag{3}$$

Lee and Seung [21] have shown that the multiplicative update rules are a good compromise between speed and ease of implementation for solving optimization problem (2) and the objective function is nonincreasing under the update rules.

2.2. Semi-supervised nonnegative matrix factorization (SSNMF)

NMF is an unsupervised learning algorithm for feature extraction. When the data are labeled, supervised NMF incorporates the information on class labels into NMF to improve the classification and clustering performance of NMF. Different from supervised NMF, SSNMF [12] is a semi-supervised modification of NMF which exploits the information on class labels contained in only a small number of labeled examples.

Supposed that each column of V belongs to one of r classes and the associated labels are encoded in the label matrix $L = [l_1, l_2, \ldots, l_n] \in R^{r \times n}$, where each l_i is a binary vector such that only jth entry is 1 and remaining elements are zero if v_i belongs to class j. SSNMF consider a joint factorization of V and L, sharing a common factor matrix H

$$D = \sum_{i,j}\tilde{Y}_{ij}\left(V_{ij} - (WH)_{ij}\right)^2 + \lambda\sum_{i,j}\widehat{Y}_{ij}\left(L_{ij} - (ZH)_{ij}\right)^2 \tag{4}$$

where λ is a tradeoff parameter determining the importance of the supervised term, $Z \in R^{r \times q}$ is the basis matrix for L, and $\tilde{Y} \in R^{p \times n}$ and $\widehat{Y} \in R^{r \times n}$ are weight matrices used to handle missing data and labels respectively. $\tilde{Y}$ and $\widehat{Y}$ are defined as follows:

$$\tilde{Y}_{ij} = \begin{cases} 1 & \text{if } V_{ij} \text{ is observed;} \\ 0 & \text{if } V_{ij} \text{ is unobserved.} \end{cases} \quad \widehat{Y}_{:j} = \begin{cases} \mathbf{1}_r & \text{if the label of } v_j \text{ is known;} \\ 0 & \text{otherwise.} \end{cases}$$

where $\mathbf{1}_r = [1, \ldots, 1]^T \in R^r$. The detailed multiplicative update rules for SSNMF and the other definition for $\widehat{Y}$ can be found in [12].

2.3. Additive clustering model and additive fuzzy clustering

In cognitive science, there is an important problem that how the subjective similarity judgments of different objects produce. Since subjective similarity cannot be derived from a straightforward analysis of objective stimulus characteristic [14], it is believed that the similarity judgments rely on the latent features which are used to represent the stimulus [15]. Additive clustering [16,17] is a method of assigning a set of latent features to a collection of objects, based on the observable similarities between them.

Given the similarity matrix $S \in R^{n \times n}$, where S_{ij} measures the observed similarity between object i and j $(i, j = 1, 2, \ldots, n)$, and supposed that there are q unobservable latent features among the objects, additive clustering model assigns the features to objects based on S. In the additive clustering model, S has an approximate decomposition form as

$$S \approx FUF^T \tag{5}$$

where $F_{ik} = 1$ if the ith object possesses the kth feature and $F_{ik} = 0$ if it does not, and U is a diagonal matrix whose diagonal element U_{kk} denotes the nonnegative saliency weight applied to the feature. The additive clustering model a special case of nonnegative matrix factorization. It aims at learning a feature assigned matrix and saliency vector that provide a good approximation to the empirical similarity. Additive clustering is one of the overlapping clustering methods. However, since the entries of F are only 0 or 1, additive clustering model disregards the influences among distinctive features and is unable to accommodate continuously vary;ing properties.

To overcome the limitations of additive clustering model, Sato et al. [18,19] introduced a new simple additive fuzzy clustering model. Additive fuzzy clustering models the similarity S_{ij} as

$$S_{ij} \approx \alpha\sum_{k=1}^{q} G_{ik}G_{jk} \tag{6}$$