

Graph Regularized Semi-supervised Concept Factorization

Yuqing Shi^{1,a}, Shiqiang Du^{2,b} and Weilan Wang^{2,c}

¹ School of Electrical Engineering, Northwest University for Nationalities, Lanzhou, China

² School of Mathematics and Computer Science, Northwest University for Nationalities, Lanzhou, China

^aadusq@163.com, ^bshqiangdu@gmail.com, ^cweilanchina@gmail.com

Corresponding author: E-mail address: weilanchina@gmail.com (W. Wang)

Keywords: Image Clustering; Semi-supervised Learning; Graph Regularized; Concept Factorization (CF)

Abstract. Concept Factorization (CF) is a new matrix decomposition technique for data representation. A modified CF algorithm called Graph Regularized Semi-supervised Concept Factorization (GRSCF) is proposed for addressing the limitations of CF and Local Consistent Concept Factorization (LCCF), which did not consider the geometric structure or the label information of the data. GRSCF preserves the intrinsic geometry of data as regularized term and use the label information as semi-supervised learning, it makes nearby samples with the same class-label are more compact, and nearby classes are separated. Compared with Non-Negative Matrix Factorization (NMF), CNMF, CF and LCCF, experiment results on ORL face database and Coil20 image database have shown that the proposed method achieves better clustering results.

Introduction

Machine learning, pattern recognition and data mining constitute areas of great development during the past decade. One of the primary goals of many data mining and machine learning systems is dimensionality reduction. The goal of dimensionality reduction is to reduce the number of features of data in order to perform tasks like clustering and/or training a classifier. Matrix factorization techniques such as eigendecomposition, Singular Value Decomposition (SVD) have been widely applied in dimensionality reduction algorithms, including Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) and graph embedding.

Recently, Nonnegative Matrix Factorization (NMF, [1]) has been proposed to incorporate the non-negativity constraints and obtain a parts-based representation. It has been widely applied to image processing and pattern recognition applications, because the learned bases can be interpreted as a natural parts-based representation of data. In particular, it 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, one drawback of NMF is that it can only be performed in the original feature space of the data points, and thus cannot make use of the power of kernelization. To overcome the limitations of NMF while inheriting all its strengths, Xu and Gong developed Concept Factorization (CF) for image clustering [2]. In CF, each basis vector (named concept) is modeled as a linear combination of the data points, and each data point as a linear combination of the basis vectors. With this model, CF is accomplished by computing the two sets of linear coefficients, which is carried out by finding the non-negative solution that minimizes the reconstruction error of the data points. After some mathematical transformations, it can be shown that the computation of CF only involves the kernel matrix of the data. The capability of using any kernel function for defining the kernel matrix, dramatically increases the power of CF.

Motivated by recent progress in semi-supervised and manifold learning [3,4,5,6,7,8], in this paper we proposed a novel algorithm, called Graphed Regularized Semi-supervised Concept Factorization (GRSCF), which explicitly considers the local geometrical information of the data and takes the label information of two images from each category as additional hard constraints. The central idea of our approach is that the data points from the same class should be merged together in the new

representation space, and we want to find a parts-based representation space in which two data points are sufficiently close to each other, if they are connected in the nearest neighbor graph. In this way, GRSCF learns a concept factorization of the data matrix which has more discriminating power. The objective function of GRSCF is not convex in its variables jointly, but convex in them individually. This fact suggests an optimization scheme that minimizes the objective function alternatively, each time optimizing one variable while keeping the others fixed. Thus, we can find a local optimal solution for GRSCF through alternative convex programming. We present some initial results of the proposed method using ORL and Coil20 images databases.

Non-negative Matrix Factorization (NMF)

Suppose that the data matrix $X = [x_1, x_2, \dots, x_n] \in R_+^{m \times n}$ consists of m -dimensional nonnegative data vectors, each of which is a sample vector. NMF aims to find two nonnegative matrices $W \in R_+^{m \times r}$ and $H \in R_+^{r \times n}$ which minimize the following objective function:

$$O = \|X - WH\|^2 \quad s.t. \, W \geq 0, H \geq 0, \quad (1)$$

where $\| \cdot \|$ denotes the matrix Frobenius norm.

The multiplicative updating rules minimizing the above objective function as follows:

$$w_{ik} \leftarrow w_{ik} \frac{(XH^T)_{ik}}{(WHH^T)_{ik}}, \quad h_{kj} \leftarrow h_{kj} \frac{(W^T X)_{kj}}{(W^T WH)_{kj}}, \quad (2)$$

where $W = [w_{ik}]$, $H = [h_{kj}]$.

It has been proved that the above update rules will find a local minimum of the objective function.

Graph Regularized Semi-supervised Concept Factorization (GRSCF)

Concept Factorization (CF). Concept Factorization (CF) is developed on the basis of NMF, it models basis vector matrix U of NMF as a linear combinations of the data points, that is $U = XW$, then the data matrix is factorized as $X \approx XWH$, the optimization problem of CF is given by

$$O = \|X - XWH\|^2 \quad s.t. \, W \geq 0, H \geq 0. \quad (3)$$

The iterative update algorithm as follows:

$$w_{ik} \leftarrow w_{ik} \frac{(KH^T)_{ik}}{(KWHH^T)_{ik}}, \quad h_{kj} \leftarrow h_{kj} \frac{(W^T K)_{kj}}{(W^T KWH)_{kj}}, \quad (4)$$

where $K = X^T X$, which computes the inner product in the original data space.

Locally Consistent Concept Factorization (LCCF) . CF fails to discover the local geometrical of the data space, which is essential to the clustering problem. Recent studies on manifold learning theory have demonstrated that the local geometric structure can be efficiently modeled through a nearest neighbor graph on a scatter of data points. Let $G = \{X, S\}$ be an undirected weighted graph and X denotes the vector set, $S \in R^{n \times n}$ is the similarity matrix, which is symmetric matrix and its element measures the similarity between a pair of vertices. Define the edge weight matrix S as follows:

$$S_{ij} = \begin{cases} \frac{x_i^T x_j}{\|x_i\| \|x_j\|}, & \text{if } x_i \in N_p(x_j) \text{ or } x_j \in N_p(x_i), \\ 0, & \text{otherwise,} \end{cases} \quad (5)$$

where $N_p(x_i)$ denotes the set of nearest neighbors of x_i . Then, the following term can be used to measure the smoothness of the low-dimensional representations.

$$R = \frac{1}{2} \sum_{i,j=1}^n \|h_i - h_j\|^2 S_{ij} = \sum_{i=1}^n h_i^T h_i D_{ii} - \sum_{i,j=1}^n h_i^T h_j S_{ij} = \text{Tr}(HDH^T) - \text{Tr}(HSH^T) = \text{Tr}(HLH^T) \quad (6)$$

where $\text{Tr}(\cdot)$ denotes the trace of a matrix, $h_i = [h_{i1}, \dots, h_{ki}]^T$ is low-dimensional representation of x_i , D is a diagonal matrix, and $D_{jj} = \sum_i S_{ij}$, $L = D - S$, which is called graph Laplacian.

The objective function of LCCF as follows:

$$O = \|X - XWH\|^2 + \lambda \text{Tr}(HLH^T), \quad \text{s.t. } W \geq 0, H \geq 0 \quad (7)$$

where the $\lambda \geq 0$ is the regularization parameter.

We can get updating rules as follows:

$$w_{ik} \leftarrow w_{ik} \frac{(KH^T)_{ik}}{(KWHH^T)_{ik}}, \quad h_{kj} \leftarrow h_{kj} \frac{(W^T K + \lambda HS^T)_{kj}}{(W^T KWH + \lambda HD^T)_{kj}}. \quad (8)$$

Graph Regularization Semi-supervised Concept Factorization (GRSCF). As NMF, CF and LCCF are unsupervised learning algorithm. There always have a small amount of labeled data in real-world, however, many researchers have found it can produce considerable improvement in learning accuracy [8].

Consider the data sets consisting of n data points $x_i, i=1,2,\dots,n$, which belong to c classes, x_1, x_2, \dots, x_l are the first l data points, and the rest x_{l+1}, \dots, x_n are unlabeled.

Define the indicator matrix A , where $a_{ij}=1$ if x_j is labeled with the i -th class; $a_{ij}=0$, otherwise. A label information matrix C as follows:

$$C = \begin{pmatrix} A_{c \times l} & 0 \\ 0 & I_{n-l} \end{pmatrix} \quad (9)$$

We hope the data points from the same class can be merged together in the new representation space, thus the obtained low-dimensional representation has the same label with the original data, and therefore can have more discriminating power. Similar to CNMF, we consider label information in LCCF, and get the following objective function of GRSCF:

$$O = \|X - XWHC\|^2 + \lambda \text{Tr}(HCLC^T H^T), \quad \text{s.t. } W \geq 0, H \geq 0. \quad (10)$$

We rewrite the objective function O :

$$\begin{aligned} O &= \text{Tr}((X - XWHC)^T (X - XWHC)) + \lambda \text{Tr}(HCLC^T H^T) \\ &= \text{Tr}(K) - 2\text{Tr}(C^T H^T W^T K) + \text{Tr}(C^T H^T W^T KWHC) + \lambda \text{Tr}(HCLC^T H^T) \end{aligned} \quad (11)$$

Let ϕ_{ik} and φ_{kj} be the Lagrange multiplier for $w_{ik} \geq 0$ and $h_{kj} \geq 0$, respectively. We define matrix $\Phi = [\phi_{ik}]$ and $\Psi = [\varphi_{kj}]$, then the Lagrangian function L

$$L = \|X - XWHC\|^2 + \lambda \text{Tr}(HCLC^T H^T) + \text{Tr}(\Phi W^T) + \text{Tr}(\Psi HC) \quad (12)$$

The partial derivatives of L with respect to W and H are

$$\begin{aligned}\frac{\partial L}{\partial W} &= -2KC^T H^T + 2KWHCC^T H^T + \Phi, \\ \frac{\partial L}{\partial H} &= -2W^T KC^T + 2W^T KWHCC^T + 2\lambda HCLC^T + \Psi^T C^T.\end{aligned}\quad (13)$$

Using the KKT conditions $\phi_{ik} w_{ik} = 0$ and $\varphi_{kj} h_{kj} = 0$, we get

$$\begin{aligned}-(KC^T H^T)_{ik} w_{ik} + (KWHCC^T H^T)_{ik} w_{ik} &= 0 \\ -(W^T KC^T)_{kj} h_{kj} + (W^T KWHCC^T)_{kj} h_{kj} + \lambda (HCLC^T)_{kj} h_{kj} &= 0\end{aligned}$$

The above equations lead to the following updating formulas

$$w_{ik} \leftarrow w_{ik} \frac{(KC^T H^T)_{ik}}{(KWHCC^T H^T)_{ik}}, \quad h_{kj} \leftarrow h_{kj} \frac{(W^T KC^T + \lambda HCSC^T)_{kj}}{(W^T KWHCC^T + \lambda HCDC^T)_{kj}}. \quad (14)$$

When $\lambda = 0$ or $C = I$, it is easy to get CF and LCCF are special case of GRSCF.

Numerical Experiments

Cluster is an important application of NMF and its variants. To evaluate the clustering performance of the proposed algorithm, we compare GRSCF with NMF [1], CF [2], LCCF [6], CNMF [7] and k-means. Two metrics, purity [9] and normalized mutual information (NMI) [2] are adopted to measure the clustering performance. All these algorithms were implemented in Matlab 7.10 and all experiments were run on Inter core2 Quad 2.2 GHz processor with 2 GB memory under Windows XP. We set the maximum iterations number of these algorithms as 1000 and keep it constant in all the following experiments.

Our empirical studies on clustering were accomplished on two images database: ORL and Coil20. Preprocessing to locate the faces was applied. Original images were normalized, aligned and cropped image is 32×32 pixels. Fig.1 shows some sample images.

For each data set, the evaluations are conducted with different numbers of clusters varying from two to ten. For the semi-supervised algorithm of GRSCF and CNMF, we randomly picked up two images from each category and use their label information as the semi-supervised information. We set the dimensionality of the new space to be the same as the number of clustering. The input matrix X for all methods has the same arrangement: each column denotes an instance. The experiments are performed 10 times and the average results are reported. The number of nearest neighbors p is set 9, and the regularized parameter λ is searched from the grid: $\{10^{-5}, \dots, 10^0, \dots, 10^5\}$, and $\lambda = 10$ is used for LCCF, CNMF and GRSCF, Fig.2 and Fig.3 shows the average clustering performance of these two databases.

It can be seen from Fig.2 and Fig.3, our proposed GRSCF algorithms consistently outperform all the other algorithms.



(a) Examples of images from the ORL database



(b) Examples of images from the Coil20 database

Fig. 1 Examples of experiment images

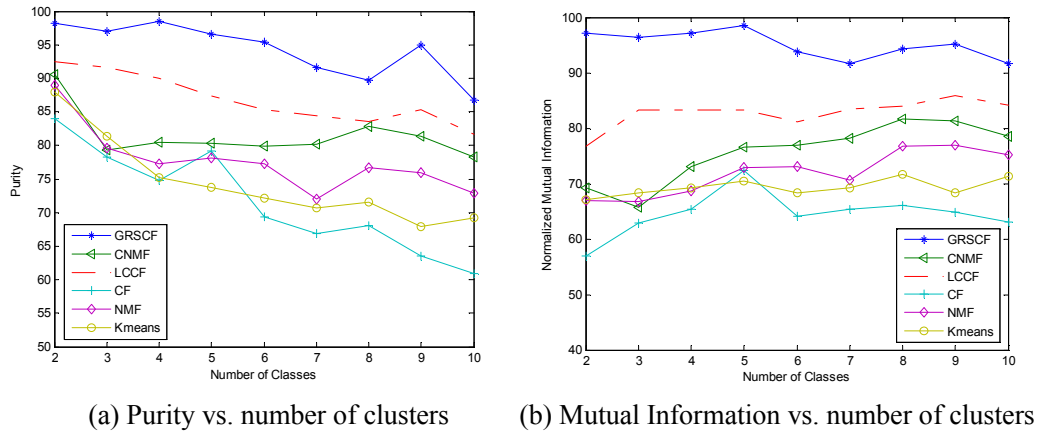


Fig. 2 Clustering Performance on ORL Database

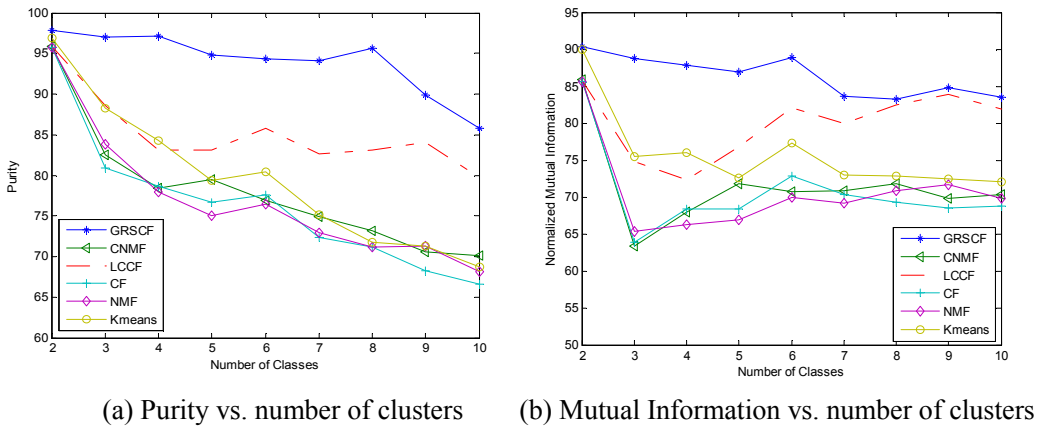


Fig. 3 Clustering Performance on Coil20 Database

Conclusions

A novel method Graph Regularization Semi-supervised Concept Factorization (GRSCF) to incorporate local geometrical structure of the data and the label information into concept factorization is presented. Specifically, GRSCF models the data space as a submanifold embedded in the ambient space and projects the two samples with same label of each class to one point in this space as semi-supervised learning. As a result, it makes nearby samples with the same class-label are more compact, and nearby classes are separated, the new representations of the data points can have more discriminating power. Compared with NMF, CNMF, CF, LCCF and K-means, the experimental results on two standard image databases have demonstrated the effectiveness of our approach.

Acknowledgments

This work was supported in part by Youth Foundation of Northwest University for Nationalities under Grant No.12xb30, Science and Research Program of The Nature Science Foundation of China under Grant No.61162021, Science Foundation of Ministry of Culture (2011) 820, and Innovative Team Subsidize of Northwest University for Nationalities.

References

- [1] D. D. Lee and H.S. Seung.: Algorithms for Nonnegative Matrix Factorization. *Adv. Neural Inf. Process. Syst.*, vol. 2(2000) 556-562.
- [2] W. Xu and Y. Gong.: Document Clustering by Concept Factorization. in *Proc. 2004 Annual ACM SIGIR Conference* (2004)
- [3] D. Cai, X. He, X. Wu, and J. Han.: Graph Regularized Non-negative Matrix Factorization for Data Representation. *IEEE Trans. Patt. Anal. Mach. Intel.*, vol.33, no.8 (2011), 1548-1560.
- [4] X. Liu, S. Yan, and H. Jin.: Projective Nonnegative Graph Embedding. *IEEE Trans. Image Process.*, vol.19, no.5 (2010) 1126-1137.
- [5] Z. Yuan and E. Oja. : Projective Nonnegative Matrix Factorization for Image Compression and Feature Extraction. in *Proc. Scand. Conf. Image Analysis*(2005) 333-342.
- [6] D. Cai, X. He and J. Han.: Locally Consistent Concept Factorization for Document Clustering, *IEEE Trans. Know. Data Eng.*, vol. 23, no.6, June (2011) 902-913.
- [7] H. Liu, Z. Wu, D. Cai and T. S. Huang.: Constrained Non-negative Matrix Factorization for Image Representation. *IEEE Trans. Pattern Anal. Mach.*, vol.34, no.7 (2012) 1299-1311.
- [8] X. Zhu, Z. Ghahramani, and J. Lafferty.: Semi-supervised Learning Using Gaussian Fields and Harmonic Functions. in *Proc. of the twentieth International Conference on Machine Learning*, (2003)
- [9] C. Ding, T. Li, W. Peng and H. Park.: Orthogonal Nonnegative Matrix Tri-factorizations for Clustering. in *Proc. 12th ACM SIGKDD int. Conf. Know. Disc. Data Mining* (2006) 126-135.