Nonparametric Orthogonal NMF and Its Application in Cancer Clustering

Andri Mirzal

Faculty of Computing, N28-439-03
Universiti Teknologi Malaysia
81310 UTM Johor Bahru, Malaysia
andrimirzal@utm.my

Abstract. Orthogonal nonnegative matrix factorization (NMF) is an NMF objective function that enforces orthogonality constraint on its factor. There are two challenges in optimizing this objective function: the first is how to design an algorithm that has convergence guarantee, and the second is how to automatically choose the regularization parameter. In our previous work, we have been able to develop a convergent algorithm for this objective function. However, the second challenge remains unsolved. In this paper, we provide an attempt to answer the second challenge. The proposed method is based on the L-curve approach and has a simple form which is preferable since it introduces only a small additional computational cost. This method transforms the algorithm into nonparametric, and is also extendable to other NMF objective functions as long as the functions are differentiable with respect to the corresponding regularization parameters. Numerical results are then provided to evaluate the feasibility of the method in choosing the appropriate regularization parameter values by utilizing it in cancer clustering tasks.

Keywords: cancer clustering, gene expression, nonnegative matrix factorization, nonparametric learning, orthogonality constraint

1 Introduction

The nonnegative matrix factorization (NMF) is a recent development in matrix decomposition and factor analysis. The NMF was first introduced by Paatero & Anttila [1, 2] and made popular by Lee & Seung [3, 4] in which the latter authors proposed a simple NMF algorithm and showed its uses in image and document analysis. The NMF has been successfully applied in many application domains including document clustering [5, 6], spectral analysis [7, 8], image processing [9–13], blind source separation [14–18], and cancer clustering and classification [19–30].

Orthogonal NMF was introduced by Ding et al. [31] to improve clustering capability of the NMF. There are two challenges in optimizing this objective function that have not been addressed by the authors: (1) how to design a convergent algorithm, and (2) how to develop a method to automatically choose

T. Herawan et al. (eds.), *Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013)*, Lecture Notes in Electrical Engineering 285, DOI: 10.1007/978-981-4585-18-7_21,
© Springer Science+Business Media Singapore 2014

the regularization parameter. In our previous work [32], we have been able to answer the first challenge. In this paper, we provide an attempt to answer the second challenge. The proposed method is based on the L-curve approach and has a simple form which is preferable since it introduces only a small additional computational cost. Numerical results are then provided to evaluate the feasibility of the method in choosing the appropriate regularization parameter values by utilizing it in cancer clustering tasks.

2 Orthogonal NMF

Since there are two factors produced by the NMF, i.e., the basis matrix $\mathbf{B}$ and the coefficient matrix $\mathbf{C}$, orthogonality constraints can be imposed on rows or columns of $\mathbf{B}$ and/or $\mathbf{C}$. Here we only discuss orthogonality constraint on rows of $\mathbf{C}$, similar results can be obtained for other cases by following the same procedure. The objective function of orthogonal NMF thus can defined with the following:

$$\begin{aligned} &\min_{\mathbf{B},\mathbf{C}} J(\mathbf{B},\mathbf{C}) = \frac{1}{2}\|\mathbf{A}-\mathbf{B}\mathbf{C}\|_F^2 + \frac{\alpha}{2}\|\mathbf{C}\mathbf{C}^T - \mathbf{I}\|_F^2 \\ &\text{s.t. } \mathbf{B} \geq \mathbf{0}, \mathbf{C} \geq \mathbf{0}, \end{aligned} \tag{1}$$

where the first component of the right hand side part denotes the approximation error, the second component denotes the orthogonality constraint, and α denotes the regularization parameter.

An algorithm based on the multiplicative update rules (MUR) [3, 4] for minimizing eq. 1 can be derived by utilizing the Karush-Kuhn-Tucker (KKT) optimality conditions. The KKT function of the objective can be defined with:

$$L(\mathbf{B},\mathbf{C}) = J(\mathbf{B},\mathbf{C}) - \text{tr}\left(\mathbf{\Gamma_B}\mathbf{B}^T\right) - \text{tr}\left(\mathbf{\Gamma_C}\mathbf{C}\right),$$

where $\mathbf{\Gamma_B} \in \mathbb{R}_+^{M\times R}$ and $\mathbf{\Gamma_C} \in \mathbb{R}_+^{N\times R}$ denote the KKT multipliers, and tr $(\mathbf{X})$ denotes trace of $\mathbf{X}$. Partial derivatives of L with respect to $\mathbf{B}$ and $\mathbf{C}$ are:

$$\begin{aligned} \nabla_{\mathbf{B}} L(\mathbf{B}) &= \nabla_{\mathbf{B}} J(\mathbf{B}) - \mathbf{\Gamma_B}, \text{ and} \\ \nabla_{\mathbf{C}} L(\mathbf{C}) &= \nabla_{\mathbf{C}} J(\mathbf{C}) - \mathbf{\Gamma_C}^T, \end{aligned}$$

with

$$\begin{aligned} \nabla_{\mathbf{B}} J(\mathbf{B}) &= \mathbf{B}\mathbf{C}\mathbf{C}^T - \mathbf{A}\mathbf{C}^T, \text{ and} \\ \nabla_{\mathbf{C}} J(\mathbf{C}) &= \mathbf{B}^T\mathbf{B}\mathbf{C} - \mathbf{B}^T\mathbf{A} + \alpha\mathbf{C}\mathbf{C}^T\mathbf{C} - \alpha\mathbf{C}. \end{aligned}$$

Then, the KKT optimality conditions can be written with the following:

$$\begin{aligned} \mathbf{B}^* &\geq \mathbf{0}, & \mathbf{C}^* &\geq \mathbf{0}, \\ \nabla_{\mathbf{B}} J(\mathbf{B}^*) = \mathbf{\Gamma_B} &\geq \mathbf{0}, & \nabla_{\mathbf{C}} J(\mathbf{C}^*) = \mathbf{\Gamma_C}^T &\geq \mathbf{0}, \\ \nabla_{\mathbf{B}} J(\mathbf{B}^*) \odot \mathbf{B}^* &= \mathbf{0}, & \nabla_{\mathbf{C}} J(\mathbf{C}^*) \odot \mathbf{C}^* &= \mathbf{0}, \end{aligned} \tag{2}$$