Pattern Recognition 46 (2013) 2840–2847

Contents lists available at SciVerse ScienceDirect

Pattern Recognition

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

Multiple graph regularized nonnegative matrix factorization

Jim Jing-Yan Wang [a], Halima Bensmail [b], Xin Gao [a,c,*]

[a] Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia
[b] Qatar Computing Research Institute, Doha 5825, Qatar
[c] Computational Bioscience Research Center, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia

ARTICLE INFO

Article history:
Received 12 June 2012
Received in revised form
11 December 2012
Accepted 5 March 2013
Available online 16 March 2013

Keywords:
Data representation
Nonnegative matrix factorization
Graph Laplacian
Ensemble manifold regularization

ABSTRACT

Non-negative matrix factorization (NMF) has been widely used as a data representation method based on components. To overcome the disadvantage of NMF in failing to consider the manifold structure of a data set, graph regularized NMF (GrNMF) has been proposed by Cai et al. by constructing an affinity graph and searching for a matrix factorization that respects graph structure. Selecting a graph model and its corresponding parameters is critical for this strategy. This process is usually carried out by cross-validation or discrete grid search, which are time consuming and prone to overfitting. In this paper, we propose a GrNMF, called MultiGrNMF, in which the intrinsic manifold is approximated by a linear combination of several graphs with different models and parameters inspired by ensemble manifold regularization. Factorization metrics and linear combination coefficients of graphs are determined simultaneously within a unified object function. They are alternately optimized in an iterative algorithm, thus resulting in a novel data representation algorithm. Extensive experiments on a protein subcellular localization task and an Alzheimer's disease diagnosis task demonstrate the effectiveness of the proposed algorithm.

1. Introduction

Non-negative matrix factorization (NMF) techniques have become popular data representation methods in recent years for a number of problems such as bioinformatics, medical imaging, computer vision, etc. When a non-negative matrix X is given, NMF attempts to find two lower dimensional non-negative matrices H and W, the product of which provides a good approximation of the original product. A standard NMF determines factorization matrices H and W by minimizing the loss function defined by Euclidean distance or the divergence between X and HW. Recently, a great number of studies have been conducted to improve the NMF method. For example, Sandler and Lindenbaum proposed two new NMF algorithms that minimize Earth mover's distance error between data and matrix product [1]. Bonettini applied the cyclic block gradient method on large-scale problems resulting from the NMF approach [2]. Guan et al. proposed a new and efficient NeNMF solution to simultaneously overcome the problems of slow convergence rate, numerical instability, and non-convergence [3]. Cichocki et al. proposed a class of multiplicative algorithms for NMF by formulating a new family of generalized divergences referred to as alpha–beta divergences (AB-divergences) [4]. Guan et al. introduced manifold regularization and margin maximization to NMF, and obtained manifold regularized discriminative NMF (MD-NMF) as a result [5]. Das Gupta and Jing pursued a discriminative decomposition process by coupling NMF objective with a support vector machine (SVM), and proposed an SVM regularizer based on NMF [6]. Hsieh and Dhillon presented a variable selection scheme for NMF that uses the gradient of the objective function to develop a new coordinate descent method [7]. Guan et al. presented a non-negative patch alignment framework to combine popular NMF techniques by proposing a fast gradient descent [8]. Lefevre et al. proposed an unsupervised inference procedure for separating audio sources by automatically grouping the components in an NMF in audio sources via a penalized maximum likelihood approach [9]. Rezaei et al. enhanced NMF performance by using fuzzy c-means clustering as an efficient initialization method for estimating initial NMF factors [10].

By conducting this knowledge in Euclidean space, NMF fails to discover the intrinsic geometrical and discriminating structure of data space [11]. To avoid such limitation, Cai et al. introduced the graph regularized NMF (GrNMF) algorithm by incorporating a geometrically based regularizer [11]. The local geometric structure is modeled by a P-nearest neighbor graph on a scattering of data points. As argued by [11], graph construction is critical for GrNMF. A number of methods on defining the P-nearest neighbor graph

* Corresponding author at: Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia. Tel.: +966 2 808 0323.
E-mail addresses: jimjywang@gmail.com (J.-Y. Wang), hbensmail@qf.org.qa (H. Bensmail), xin.gao@kaust.edu.sa (X. Gao).

0031-3203/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2013.03.007

and its corresponding weight matrix are available. The performance of these methods is known to hinge heavily on the choice of graph. Unfortunately, we do not know which graph is most suitable for a particular task. Moreover, an exhaustive search on a predefined pool of graphs and their parameters will be time consuming. Therefore, efficiently determining an appropriate graph to make the performance of the employed graph-regularized representation method robust, or even better, is crucial.

To address the aforementioned problems, and inspired by ensemble manifold regularization (EMR) [12], we propose a multiple graph regularization framework for GrNMF, which combines automatic intrinsic manifold approximation and NMF. By providing a series of initial guesses of the graph Laplacian, the framework learns how to combine them to approximate the intrinsic manifold. At the same time, factorization matrices for NMF are solved and restricted to run smoothly along the estimated graph.

The rest of the paper is organized as follows. Section 2 introduces our multiple graph regularized NMF (MultiGrNMF) algorithm. Experimental results on two data sets are presented in Section 3. Finally, we give our concluding remarks and suggestions for future work in Section 4.

2. MultiGrNMF

In this section, we present our proposed MultiGrNMF algorithm.

2.1. Objective function

Given N data points in the training set with their non-negative feature vector set $\mathcal{X}=\{x_n\}$, $n=1\ldots,N$, we organize them as a nonnegative data matrix $X=[x_1,\ldots,x_n]\in\mathbb{R}_+^{D\times N}$, where n-th column x_n of X is the feature vector of n-th data point. NMF aims to find two non-negative matrices H and W, the product of which can approximate well the original matrix X as $X\approx HW$, where $H\in\mathbb{R}_+^{D\times R}$ and $W\in\mathbb{R}_+^{R\times N}$. We commonly have $R\ll D$ and $R\ll N$. In reality, each feature vector x_n is approximated by a linear combination of the columns of H, and weighted by the components of W, as

$$x_n\approx\sum_{r=1}^{R}h_r w_{rn} \tag{1}$$

Therefore, H can be regarded as containing a set of basis vectors. Let $w_n=[w_{1n},\ldots,w_{Rn}]^\top$ denote the n-th columns of W. w_n can be regarded as the coding vector or as a new representation of the n-th data point with respect to the basis H. To learn the basis matrix H and the coding matrix W, we propose a novel objective function $O^{MultiGrNMF}$. This objective function considers the local manifold structure of data space for the regularization of NMF and approximates the intrinsic local manifold by combining several initial graphs. The proposed objective function is composed of two terms: (1) the original loss function of NMF and (2) the multiple graph regularization term.

2.1.1. NMF loss term

The most commonly used NMF loss function is based on l_2 norm distance between two matrices [13]:

$$\begin{aligned}O^{NMF}(H,W)&=\|X-HW\|^2\\&=\mathrm{Tr}(X^\top X)-2\,\mathrm{Tr}(X^\top HW)+\mathrm{Tr}(W^\top H^\top HW).\end{aligned} \tag{2}$$

The aforementioned objective function can be minimized by the iterative update algorithm proposed by Lee and Seung [14].

2.1.2. Multiple graph regularization term

In GrNMF, local invariance assumption was imposed to NMF in the following manner. If two feature vectors x_n and x_m are close in the intrinsic geometry of data distribution, then w_n and w_m, the coding vectors of these two feature vectors with respect to the new basis, are also close to each other, and vice versa [11]. The local geometric structure is modeled by a P nearest neighbor graph $\mathcal{G}$ on a scattering of data points [11]. For each feature vector $x_n\in\mathcal{X}$, its P nearest neighbors $\mathcal{N}_n$ in $\mathcal{X}$, is first found. A P nearest neighbor graph is then constructed for $\mathcal{X}$ as $\mathcal{G}=\{\mathcal{V},\mathcal{E},A\}$. The node set $\mathcal{V}$ corresponds to N data points. $\mathcal{E}$ is the edge set, and $(n,m)\in\mathcal{E}$ if $x_m\in\mathcal{N}_n$ or $x_n\in\mathcal{N}_m$. $A\in R^{N\times N}$ is the weight matrix on the graph with A_{nm} equal to the weight of edge (n,m). With weight matrix A, we can use the following **g**raph **r**egularization term to measure the smoothness of low-dimensional coding vector representations in W

$$\begin{aligned}O^{Gr}(W)&=\frac{1}{2}\sum_{n,m=1}^{N}\|w_n-w_m\|^2A_{nm}\\&=\mathrm{Tr}(WUW^\top)-\mathrm{Tr}(WAW^\top)\\&=\mathrm{Tr}(WLW^\top),\end{aligned} \tag{3}$$

where U is a diagonal matrix, the entries of which are column sums of A, $U_{nn}=\sum_{n=1}^{N}A_{nm}$, and $L=U-A$ is the graph Laplacian. By minimizing $O^{Gr}(W;A)$ with regard to W, we can expect that if two feature vectors x_n and x_m are close to one another (i.e., A_{nm} is big), then w_n and w_m will also be close to each other.

A number of techniques for defining weight matrix A are available. Four of the most commonly used methods are as follows:

- 0-1 weighting is used to regularize NMF and sparse coding in [11]. It is defined as

$$A_{nm}=\begin{cases}1, & \text{if } (n,m)\in\mathcal{E},\\ 0, & \text{else.}\end{cases} \tag{4}$$

- *Heat kernel weighting* is also widely used in most manifold learning algorithms [11]. This technique is defined as

$$A_{nm}=\begin{cases}e^{-(\|x_n-x_m\|^2)/\sigma^2}, & \text{if } (n,m)\in\mathcal{E},\\ 0, & \text{else.}\end{cases} \tag{5}$$

- *Histogram intersection kernel weighting* is usually used to construct similarity graphs of scale-invariant feature transform features. It is defined as

$$A_{nm}=\begin{cases}\sum_{d=1}^{D}\min(x_{dn},x_{dm}), & \text{if } (n,m)\in\mathcal{E},\\ 0, & \text{else.}\end{cases} \tag{6}$$

Given such numerous graph weight matrix model definitions with different parameters, we can compute several corresponding graph Laplacians. The number of nearest neighbors should also be noted as a parameter for nearest neighbor graph construction. With different numbers of nearest neighbors, we can produce various graphs. As such, this parameter is not controlled by the user, but is selected automatically by the algorithm introduced as follows, by weighting graphs constructed using different numbers of nearest neighbors. Suppose that we have already computed a set of candidate graph Laplacians $\{L_1,\ldots,L_K\}$. Similar to EMR, we assume that the intrinsic manifold is located in the convex hull of the previously given manifold candidates [12]. This assumption constrains the search space of possible graph Laplacians as linear combinations of M candidate Laplacians,

$$L=\sum_{k=1}^{K}\tau_k L_k,\quad \text{s.t.}\sum_{k=1}^{K}\tau_k=1,\quad \tau_k\geq 0, \tag{7}$$

where τ_k is the combination weight of k-th graph Laplacian. To avoid negative contribution, we further constrain $\sum_{k=1}^{K}\tau_k=1,\ \tau_k\geq 0$. As such, we attempt to determine optimal linear combination weights for a group of pre-computed graph candidates instead of selecting the optimal graph model and estimating its parameters.

By substituting (7) to (3), we obtain the augmented objective function of multiple graph regularization in an enlarged parameter space,

$$O^{MultiGr}(W,\tau)=\mathrm{Tr}\left(W\sum_{k=1}^{K}\tau_k L_k W^{\top}\right)=\sum_{k=1}^{K}\tau_k\mathrm{Tr}(WL_kW^{\top})$$
$$\text{s.t.}\quad \sum_{k=1}^{K}\tau_k=1,\ \tau_k\geq 0, \tag{8}$$

where $\tau=[\tau_1,\ldots,\tau_K]^{\top}$ is the graph weight vector. MultiGr is different from the EMR introduced in [12]. The latter tries to minimize classifier complexity over the composite manifold, whereas the former aims to restrict coding vectors of NMF to enable them to move smoothly along the estimated composite manifold.

2.1.3. Object function of MultiGrNMF

Combining the multiple graph-based regularizer $O^{MultiGr}(W,\tau)$ with the original NMF objective function $O^{NMF}(H,W)$ results in the loss function of MultiGrNMF as,

$$\begin{aligned}O^{MultiGrNMF}(H,W,\tau)&=O^{NMF}(H,W)+\alpha O^{MultiGr}(W)+\beta\|\tau\|^2\\&=\mathrm{Tr}(X^{\top}X)-2\,\mathrm{Tr}(X^{\top}HW)+\mathrm{Tr}(W^{\top}H^{\top}HW)\\&\quad+\alpha\sum_{k=1}^{K}\tau_k\mathrm{Tr}(WL_kW^{\top})+\beta\|\tau\|^2.\end{aligned} \tag{9}$$

To avoid parameter τ overfitting to a single graph, we also introduce a l_2 norm regularization term $\|\tau\|^2$ to τ. In (9), α and β are tradeoff parameters that balance the three terms. Thus, the MultiGrNMF problem turns into a minimization problem as,

$$\min_{H,W,\tau}\ O^{MultiGrNMF}(H,W,\tau)$$
$$\text{s.t.}\quad H\geq 0,\ W\geq 0,\ \sum_{k=1}^{K}\tau_k=1,\ \tau_k\geq 0. \tag{10}$$

The relationship between MultiGrNMF and GrNMF is explained as follows. As we can see from the previously discussed objective function, GrNMF is a special case that occurs when only one graph is present in the graph pool. When only one graph is used, the weight of this graph will be solved as one, and MultiGrNMF will degenerate into GrNMF.

2.2. Optimization

Instead of optimizing (10) directly, we optimize NMF factorization matrices (H, W) and graph weights τ by using an iterative, two-step strategy because direct optimization to (10) is difficult. At each iteration, either (H, W) or τ is optimized first while the other is fixed, and then the roles of (H, W) and τ are reversed. Iterations are repeated until convergence is achieved or a maximum number of iterations are reached.

2.2.1. On optimizing (H, W)

By fixing τ, using the matrix property $\mathrm{Tr}(X)=\mathrm{Tr}(X^{\top})$ and $\mathrm{Tr}(XY)=\mathrm{Tr}(YX)$, and removing irrelevant items, the optimization problem (10) is reduced to

$$\min_{H,W}\ -2\,\mathrm{Tr}(X^{\top}HW)+\mathrm{Tr}(W^{\top}H^{\top}HW)+\alpha\,\mathrm{Tr}(WLW^{\top})$$
$$\text{s.t.}\quad H\geq 0,\ W\geq 0. \tag{11}$$

Let ϕ_{dr} and ψ_{rn} be the Lagrange multipliers for constraints $h_{dr}\geq 0$ and $w_{rn}\geq 0$, respectively, and $\Phi=[\phi_{dr}]$, $\Psi=[\psi_{rn}]$, the Lagrange $\mathcal{L}$ of (11) is

$$\mathcal{L}=-2\,\mathrm{Tr}(X^{\top}HW)+\mathrm{Tr}(W^{\top}H^{\top}HW)+\alpha\,\mathrm{Tr}(WLW^{\top})+\mathrm{Tr}(\Phi H^{\top})+\mathrm{Tr}(\Psi W^{\top}). \tag{12}$$

The partial derivatives of $\mathcal{L}$, with respect to basis matrix H and coding matrix W are

$$\frac{\partial\mathcal{L}}{\partial H}=-2XW^{\top}+2HWW^{\top}+\Phi$$

and

$$\frac{\partial\mathcal{L}}{\partial W}=-2H^{\top}X+2H^{\top}HW+2\alpha WL+\Psi. \tag{13}$$

By using the Karush–Kuhn–Tucker conditions $\phi_{dr}h_{dr}=0$ and $\psi_{rn}w_{rn}=0$, and substituting $L=U-A$ to (13), we obtain the following equations for h_{dr} and w_{rn}:

$$-(XW^{\top})_{dr}h_{dr}+(HWW^{\top})_{dr}h_{dr}=0$$

and

$$-(H^{\top}X+\alpha WA)_{rn}w_{rn}+(H^{\top}HW+\alpha WU)_{rn}w_{rn}=0. \tag{14}$$

These equations lead to the following updating rules:

$$h_{dr}\leftarrow\frac{(XW^{\top})_{dr}}{(HWW^{\top})_{dr}}h_{dr}$$

and

$$w_{rn}\leftarrow\frac{(H^{\top}X+\alpha WA)_{rn}}{(H^{\top}HW+\alpha WU)_{rn}}w_{rn}. \tag{15}$$

2.2.2. On optimizing τ

By fixing (H, W) and removing irrelevant terms, the optimization problem (10) is transformed into

$$\min_{\tau}\ \sum_{k=1}^{K}\tau_k s_k+\gamma\|\tau\|^2$$
$$\text{s.t.}\quad \sum_{k=1}^{K}\tau_k=1,\ \tau_k\geq 0, \tag{16}$$

where $s_k=\mathrm{Tr}(WL_kW^{\top})$ and $\gamma=\beta/\alpha$. Additional constraints $\sum_{k=1}^{K}\tau_k=1,\ \tau_k\geq 0$ cause the optimization presented in (16) to turn into a constrained quadratic programming (QP) problem [15]. As such, the optimization can now be efficiently solved by the algorithm based on coordinate descent [16].

2.3. Algorithm

Listed as Algorithm 1, the procedure for MultiGrNMF requires an initial guess for both τ and (H, W) in the alternating optimization. We have tried the following initialization strategies:

1. τ is initialized by setting all of its elements as $1/K$ to base graphs with equal weights;
2. (H, W) is initialized by performing the original NMF to X.

We average performances obtained by the models with different initializations. In our empirical testing, initialization strategies exhibit stable performances.

Algorithm 1. The training procedure of the proposed MultiGrNMF algorithm.

Input: Training data matrix X;
Input: K graph Laplacians $\{L_1,\ldots,L_K\}$ constructed from X with different graph models and parameters;
Make an initial guess for τ^0 and (H^0,W^0);
for $t=1,\ldots,T$ **do**

Update the basis matrix H^t and coding matrix W^t by (14) while fixing τ^{t-1};
Update graph weights τ^t by (16) while fixing H^t and W^t;
end for
Output: U^T, W^T and τ^T.

3. Experiments

Two experiments have been conducted to demonstrate the effectiveness of the proposed MultiGrNMF algorithm on two challenging tasks: (1) protein subcellular localization and (2) Alzheimer's disease (AD) diagnosis.

3.1. Experiment I: Protein subcellular localization of fluorescence imagery

Predicting protein subcellular locations is crucial in the complete understanding of various protein functions. Currently, fluorescence microscopy is the most suitable method for proteome-wide determination of subcellular location. In this experiment, we evaluate MultiGrNMF as a feature representation method for protein subcellular localization of fluorescence imagery.

3.1.1. Data set and setup

In this empirical evaluation, we use the *2D HeLa image data set* [17]. This data set consists of 862 single-cell images, each measuring 382×382. Each image contains a single cell from one of the 10 major classes of protein localization patterns. Subcellular location patterns in these collections include endoplasmic reticulum (ER), the Golgi complex, lysosomes, mitochondria, nucleoli, actin microfilaments, endosomes, microtubules, and nuclear DNA [17]. Several sample images from this data set is shown in Fig. 1.

In this experiment, we use various texture-based feature extraction strategies, including Haralick textures, local binary patterns, local ternary patterns, etc. The feature vector x of each image is constructed by fusing the aforementioned features as hybrid features. The features are further entered as inputs to MultiGrNMF for feature representation. We used the following graph types to construct multiple graphs for MultiGrNMF: 0-1 weighted graph, heat kernel weighted graph, and histogram intersection kernel weighted graph. By varying neighborhood size parameters for all graph types and the bandwidth parameter σ for heat kernel weighted graph, we have obtained a total of 25 graphs for this experiment. Moreover, a10-fold cross-validation is employed to test the performance of MultiGrNMF.

3.1.2. Results

We first tested the classification performance of MultiGrNMF, GrNMF, and NMF against the number of basis vectors R. Results are shown in Fig. 2(a). Classification accuracies of MultiGrNMF, GrNMF, and NMF have all increased when more basis vectors are used to represent data. This figure shows that the proposed MultiGrNMF coding outperforms the other two algorithms. For this data set, classification performances of the original NMF with l_2 norm metrics are generally inferior to those in low-dimensional subspaces selected by graph-based data representation methods. Based on the results, we arrived at the conclusion that NMF methods based on manifold assumption, such as GrNMF and MultiGrNMF, perform better than the original NMF. The reason is that, after graph regularization, discriminative information is contained by coding vectors embedded in an R-dimensional subspace. Classification result of GrNMF using only one graph is already competitive. We attribute this result to the manifold distribution property of fluorescent images. However, compared with the results of GrNMF shown in Fig. 2(a), one can see that the recognition performance of our proposed MultiGrNMF is generally superior in almost all R-dimension reduced subspaces than that of GrNMF. This finding provides strong evidence that using multiple graphs is more robust than using a single graph for NMF regularization.

Fig. 2(b) shows the performance of MultiGrNMF, GrNMF, and NMF with varying values of parameter α. Selecting the value of α is one of the most difficult aspects of graph regularization [11]. As shown in Fig. 2(b), GrNMF and MultiGrNMF are both relatively

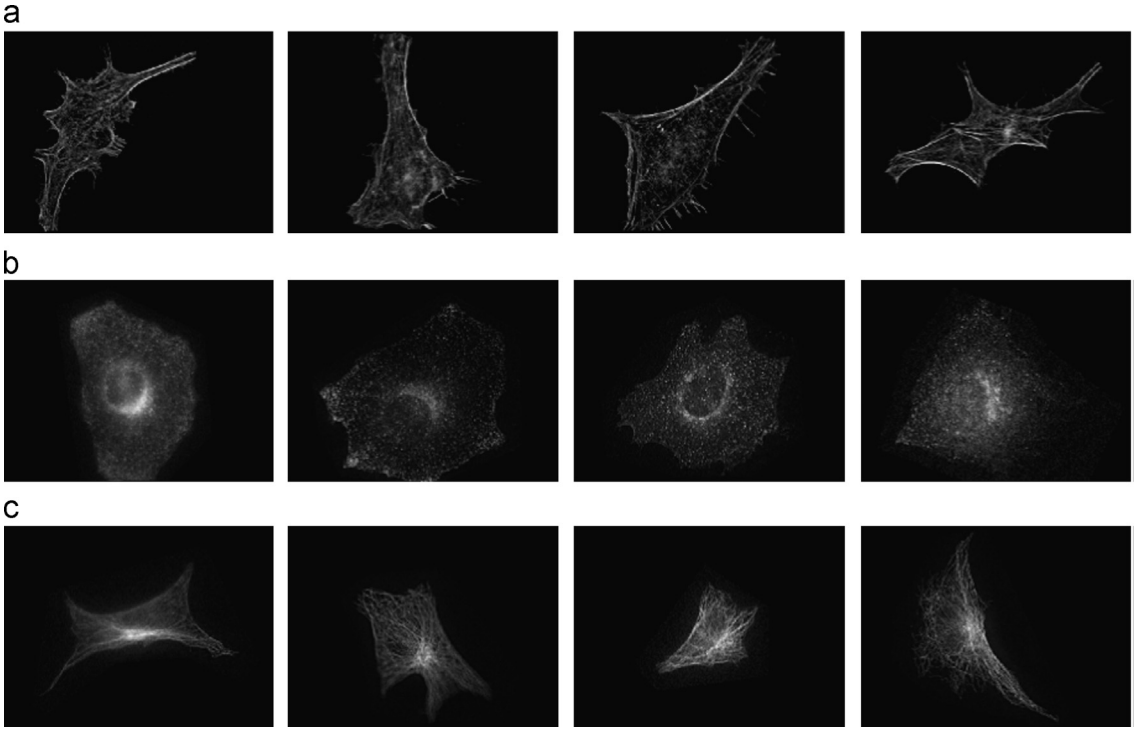

Fig. 1. Sample images from the 2D HeLa image data set. (a) ActinFilaments, (b) endosome and (c) microtubules.

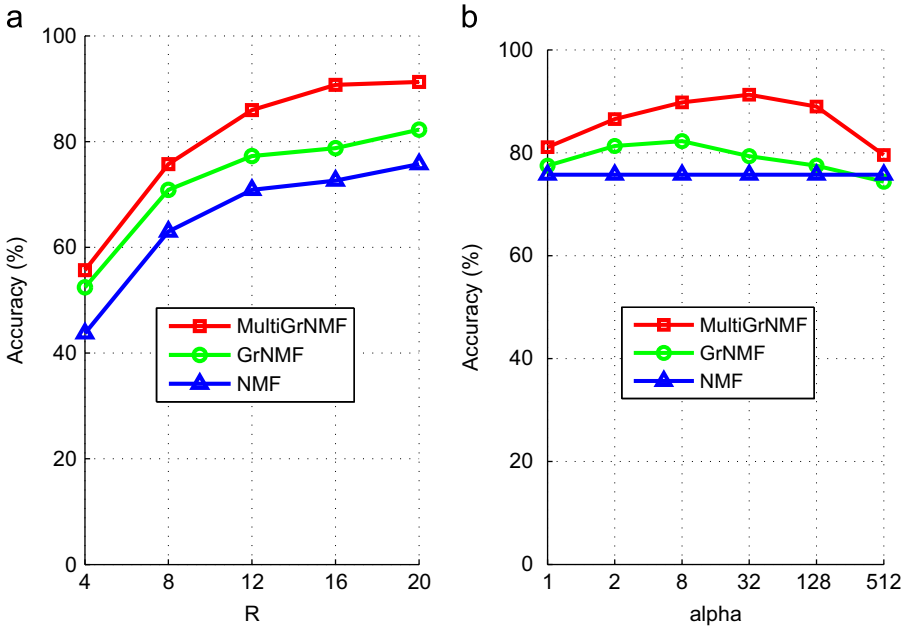

Fig. 2. Classification accuracies of MultiGrNMF, GrNMF, and NMF versus parameters R and α. (a) Accuracies v.s. R and (b) accuracies v.s. α.

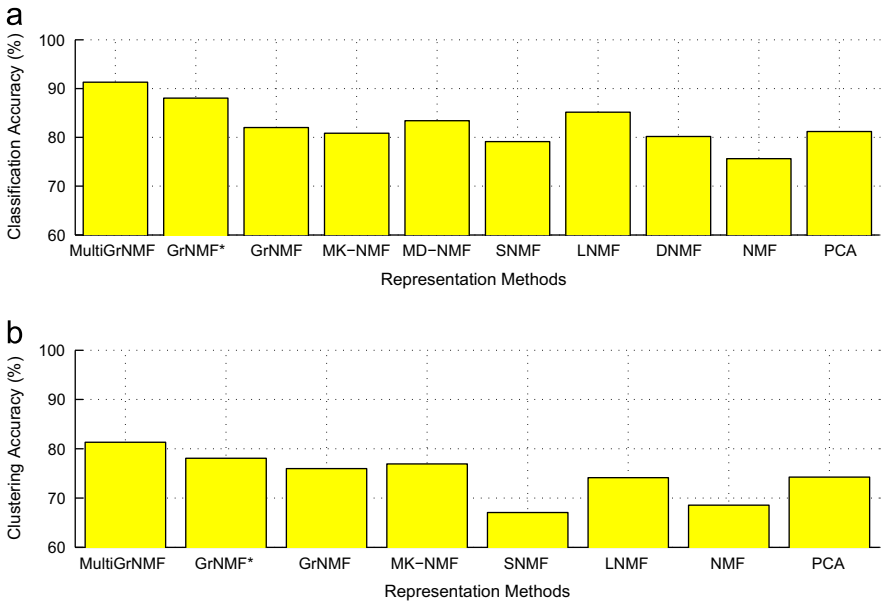

Fig. 3. Comparison of classification and clustering accuracies of different NMF-related data representation methods. (a) Classification results and (b) clustering results.

stable with respect to regularization parameter α. However, based on empirical observations, a choice of α values between 8 and 128 is recommended. Fig. 2(b) shows that, in the case of $\alpha=32$ and 128, MultiGrNMF significantly improves its performance in all classes; however, its performance decreases when α becomes as large as 512. A probable reason for this finding is that when α is too small, the effect of graph regularization is not enough for the matrix factorization procedure, whereas when it is too large, the matrix factorization procedure will be overfitted to the graph. As such, choosing an α value in the middle range, such as 128 for MultiGrNMF, is recommended. However, we can see from this figure that both MultiGrNMF and GrNMF are robust to the α parameter to some extent. Overall, MultiGrNMF performs comparably with GrNMF and significantly better than NMF. GrNMF outperforms NMF significantly on five out of six cases. The poor performance of GrNMF on $\alpha=512$ may be ascribed to the graph used by GrNMF. GrNMF uses a kernel-based graph, which may not be as suitable for protein subcellular-location tasks as a linear kernel. From our experiments, we see that MultiGrNMF is a viable alternative to GrNMF-based data representation approaches.

To compare our method with state-of-the-art NMF-based data representation methods, we match the classification performance of six NMF-based data representation methods with identical initializations:

- MultiGrNMF MultiGrNMF.
- GrNMF [11].
- Multiple kernel NMF (MK-NMF) [18].
- MD-NMF [8].
- Sparse NMF [19].
- Local NMF [20].
- Discriminant NMF (DNMF) [21].

- Original NMF.
- Principal component analysis.

.

For MultiGrNMF, multiple graph regularization is performed to coding vectors in W, whereas for MK-NMF, multiple kernel functions are performed to data vectors X and basis vectors H. Results are given in Fig. 3(a). Moreover, we are also interested in determining whether MultiGrNMF outperforms GrNMG using the best graph selected by MultiGrNMF (denoted as GrNMF* in Fig. 3). An experimental comparison is also conducted and the result of GrNMF* is shown in Fig. 3. Experiments on classification are insufficient in demonstrating the superiority of the novel method. Previous studies have shown that NMF methods demonstrate favorable clustering results in many applications. As such, we also proposed clustering performance of the methods, as presented in Fig. 3(b).

From Fig. 3, we can see the following observations and conclusions:

1. MultiGrNMF outperforms NMF and most of its sparse versions because of the use of multiple graphs, except for MD-NMF. Not surprisingly, MD-NMF exhibits better performance than MultiGrNMF because MD-NMF is a supervised method that uses class label information, whereas MultiGrNMF is an unsupervised method. Interestingly, the performance of DNMF, which is also a supervised algorithm, is per, which means it does not use class information as effectively as MD-NMF.
2. Using the same linear combination of multiple Gaussian kernels as multiple graph and multiple kernel strategies, MultiGrNMF is much better than MK-NMF because of two possible explanations.
   (a) First, formulating the objective function by regularizing coding vectors directly is more effective compared with applying multiple kernels to the original data space and the basis vectors. A possible reason for this result is that when original data vectors and basis vectors are mapped in a non-linear space via multiple kernels, final coding vectors are still determined by matrix factorization procedure, which cannot guarantee that coding vectors lies on a proper manifold. By contrast, when a multiple graph is used to regularize coding vectors directly, the manifold assumption is directly implemented by multiple graph regularization.
   (b) Second, the regularizer proposed in this study can more effectively exploit the intrinsic manifold structure of data space by applying a l_2 norm regularization to linear combination coefficients of initial graphs. The explanation for this finding is that this term can prevent combination coefficients from overfitting in one graph.
3. MultiGrNMF achieves the best result, which further demonstrates that an algorithm based on ensemble manifold regularization outperforms algorithms based on a single manifold, which are widely used in many existing data representation and classification methods.
4. The single-graph method that uses the best graph selected by MultiGrNMF does not outperform MultiGrNMF, but is comparable to MultiGrNMF. The possible reason for this outcome is that MultiGrNMF prevents the algorithm from overfitting to a single graph by introducing the l_2 norm to graph weights.

Lastly, proving the convergence of the proposed algorithm is difficult. As an alternative, we plot the convergence curve in Fig. 4 to show the convergence of the proposed algorithm. As shown in Fig. 4, the objective value appears to be stable after approximately 100 iterations.

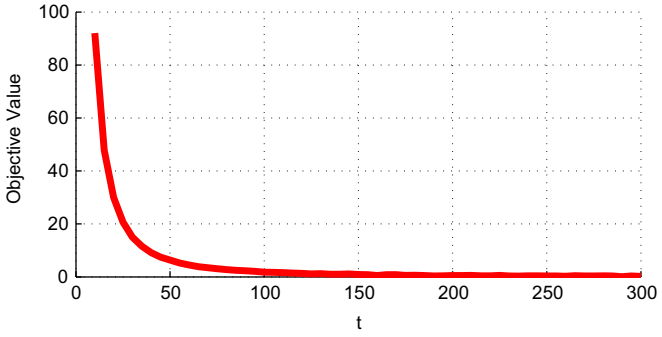

Fig. 4. Convergence curve of MultiGrNMF.

3.2. Experiment II: Diagnosis of AD using positron emission tomography (PET) images

In this section, we will test the proposed MultiGrNMF as a PET image representation method for application in AD diagnosis [22].

3.2.1. Data set and setup

A PET data set is selected from the AD Neuroimaging Initiative (ADNI) database [23] to validating the computer-aided design (CAD) tool using MultiGrNMF as data representation method. The main purpose of ADNI is to measure the progression of AD during its initial stages [23]. A total of 800 participants from the United States and Canada were recruited to develop this database, including approximately 200 normal participants without symptoms, about 400 subjects with cognitive impairment, and 200 patients with early AD symptoms. We have selected PET data from 340 ADNI participants to build our *PET340* data set. The participants in the *PET340* data set are classified into two groups: 168 AD patients and 172 normal control subjects.

The voxels of a brain functional three-dimensional PET image will be organized as data vector x, and further represented into coding vector w by MultiGrNMF. Coding vectors will be used as feature vectors of an SVM classifier to separate AD patients (positive subjects) from normal subjects (negative subjects). These classification results are evaluated using the leave-one-out cross-validation strategy. To evaluate the developed MultiGrNMF-based CAD tool, receiver operating characteristic (ROC) curve and recall–precision curve, are obtained. At the same time, the area under the curve (AUC) of an ROC curve is also computed as single measure of classification performance.

3.2.2. Results

Fig. 5(a) shows the ROC curve of a true positive rate versus a false positive rate by using SVM as the classification algorithm. Each curve in the figure represents a PET image representation algorithm based on NMF. As can be seen from Fig. 5(a), our proposed MultiGrNMF algorithm performs the best, whereas GrNMF outperforms NMF. As decision threshold decreases, MultiGrNMF performs slightly better than GrNMF and NMF. MultiGrNMF and GrNMF perform comparably with each other when decision thresholds are very large or very small. NMF performs the worst for all cases. Fig. 6 shows the AUC for each algorithm. Our MultiGrNMF algorithm yields the highest AUC.

Fig. 5(b) shows recall–precision curves by using MultiGrNMF, GrNMF, and NMF as data representation algorithms. As shown in the figure, our MultiGrNMF algorithm significantly outperforms the other two representation algorithms in all cases. Performance difference gets larger as decision threshold increases. The GrNMF algorithm outperforms NMF in most cases. When a large recall value is obtained, more true positive subjects and more false

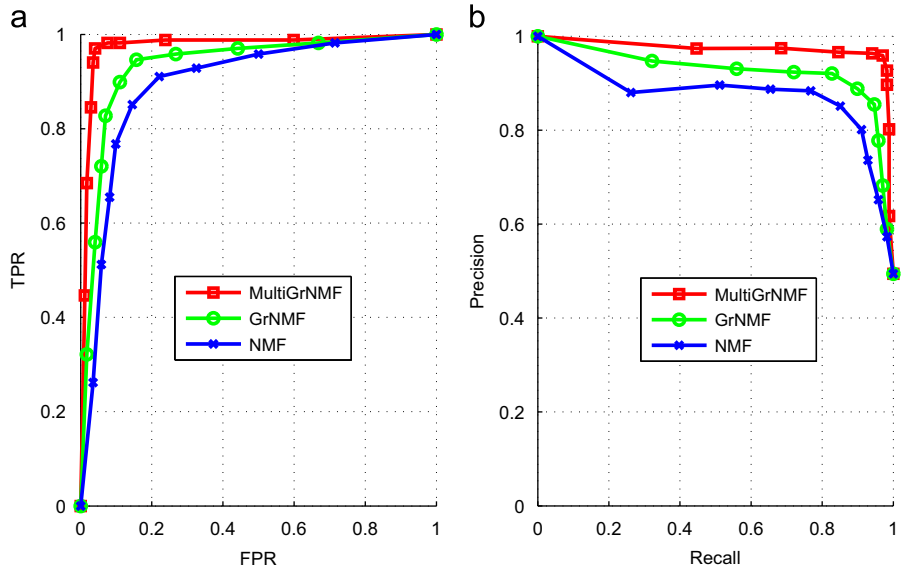

Fig. 5. (a) ROC curve and (b) recall–precision curve of MultiGrNMF, GrNMF, and NMF on ADNI data set.

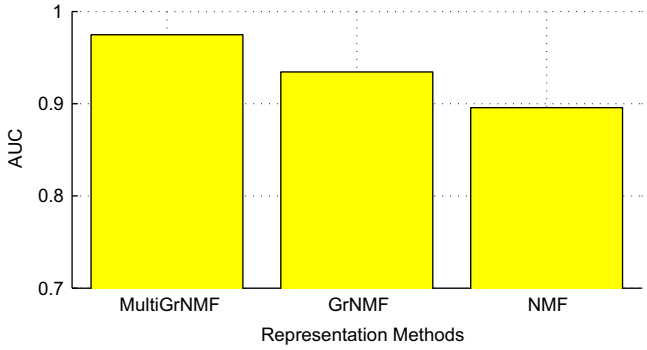

Fig. 6. AUC of MultiGrNMF, GrNMF, and NMF on ADNI data set.

positive subjects exist. Therefore, all the algorithms yield low precision rates in this case. As recall value increases, precision rates of all algorithms decrease. However, the precision rate of MultiGrNMF tends to converge when the recall value is smaller than 0.95, whereas GrNMF and NMF can consistently benefit from a smaller recall value.

4. Conclusion and future work

This paper introduced a new data representation method by improving GrNMF [11]. The method depends on a graph constructed by a linear combination of several initial graphs with different models and parameters. The main idea behind the model is that the intrinsic manifold can be approximated by multiple nearest-neighbor graphs. A unified object function framework is proposed to derive the MultiGrNMF form in terms of NMF loss function and multiple graph regularization. The resulting coding matrix between nodes is a data representation technique regularized by multiple graphs. Graph weights can also be computed efficiently based on the derived coding matrix. Two data classification experiments show that MultiGrNMF performs well compared with other NMF-based data representation methods.

Other optimization methods can also be used with our proposed algorithm instead of multiplicative update rules, such as the optimal gradient method proposed in [3] by Guan et al. However, for a fair comparison, we have used the same multiplicative update rule of GrNMF to demonstrate the advantage of multiple graph regularization. In our future work, we will reconsider optimization algorithms by using more efficient algorithms.

Conflict of interest statement

None declared.

Acknowledgments

The study was supported by grants from 2011 Qatar Annual Research Forum Award (Grant No. ARF2011) and King Abdullah University of Science and Technology (KAUST), Saudi Arabia.

References

[1] R. Sandler, M. Lindenbaum, Nonnegative matrix factorization with Earth mover's distance metric for image analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence 33 (8) (2011) 1590–1602.
[2] S. Bonettini, Inexact block coordinate descent methods with application to the nonnegative matrix factorization, IMA Journal of Numerical Analysis 31 (4) (2011) 1431–1452.
[3] N. Guan, D. Tao, Z. Luo, B. Yuan, NeNMF: an optimal gradient method for non-negative matrix factorization, IEEE Transactions on Signal Processing 60 (6) (2012) 2882–2898.
[4] A. Cichocki, S. Cruces, S. ichi Amari, Generalized alpha–beta divergences and their application to robust nonnegative matrix factorization, Entropy 13 (1) (2011) 134–170.
[5] N. Guan, D. Tao, Z. Luo, B. Yuan, Manifold regularized discriminative non-negative matrix factorization with fast gradient descent, IEEE Transactions on Image Processing 13 (1) (2011) 134–170.
[6] M. Das Gupta, J. Xiao, Non-negative matrix factorization as a feature selection tool for maximum margin classifiers, in: IEEE Conference on Computer Vision and Pattern Recognition 2011, 2011.
[7] C.-J. Hsieh, I.S. Dhillon, Fast coordinate descent methods with variable selection for non-negative matrix factorization, in: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '11, 2011, pp. 1064–1072.
[8] N. Guan, D. Tao, Z. Luo, B. Yuan, Non-negative patch alignment framework, IEEE Transactions on Neural Networks 22 (8) (2011) 1218–1230.
[9] A. Lefevre, C. Bach, F. Fevotte, Itakura-saito nonnegative matrix factorization with group sparsity, in: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2011, pp. 21–24.
[10] M. Rezaei, R. Boostani, M. Rezaei, An efficient initialization method for nonnegative matrix factorization, Journal of Applied Science 11 (2) (2011) 354–359.
[11] D. Cai, X. He, J. Han, T. Huang, Graph regularized nonnegative matrix factorization for data representation, IEEE Transactions on Pattern Analysis and Machine Intelligence 33 (8) (2011) 1548–1560, http://dx.doi.org/10.1109/TPAMI.2010.231.

[12] B. Geng, D. Tao, C. Xu, L. Yang, X.-S. Hua, Ensemble manifold regularization, IEEE Transactions on Pattern Analysis and Machine Intelligence 34 (6) (2012) 1227–1233, http://dx.doi.org/10.1109/TPAMI.2012.57.
[13] A. Srivastava, E. Klassen, S.H. Joshi, I.H. Jermyn, Shape analysis of elastic curves in euclidean spaces, IEEE Transactions on Pattern Analysis and Machine Intelligence 33 (7) (2011) 1415–1428, http://dx.doi.org/10.1109/TPAMI.2010.184.
[14] D.D. Lee, H.S. Seung, Algorithms for non-negative matrix factorization, in: NIPS, 2000, pp. 556–562.
[15] G.R. Mauri, L.A. Nogueira Lorena, Improving a Lagrangian decomposition for the unconstrained binary quadratic programming problem, Computers & Operations Research 39 (7) (2012) 1577–1581, http://dx.doi.org/10.1016/j.cor.2011.09.008.
[16] F. Wei, H. Zhu, Group coordinate descent algorithms for nonconvex penalized regression, Computational Statistics & Data Analysis 56 (2) (2012) 316–326, http://dx.doi.org/10.1016/j.csda.2011.08.007.
[17] K. Huang, R. Murphy, Boosting accuracy of automated classification of fluorescence microscope images for location proteomics, BMC Bioinformatics 5 http://dx.doi.org/10.1186/1471-2105-5-78.
[18] S. An, J.-M. Yun, S. Choi, Multiple kernel nonnegative matrix factorization, in: 2011 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2011, pp. 1976–1979.
[19] N. Gillis, F. Glineur, Using under approximations for sparse nonnegative matrix factorization, Pattern Recognition 43 (4) (2010) 1676–1687, http://dx.doi.org/10.1016/j.patcog.2009.11.013.
[20] S.Z. Li, X. Hou, H. Zhang, Q. Cheng, Learning spatially localized, parts-based representation, in: Proceedings of the IEEE International Conference on Computer Visual and Pattern Recognition 2001, 2001, p. 207–212.
[21] S. Zafeiriou, A. Tefas, I. Buciu, I. Pitas, Exploiting discriminant information in nonnegative matrix factorization with application to frontal face verification, IEEE Transactions on Neural Networks 17 (3) (2006) 683–695.
[22] P. Padilla, M. Lopez, J. Gorriz, J. Ramirez, D. Salas-Gonzalez, I. Alvarez, NMF-SVM based CAD tool applied to functional brain images for the diagnosis of Alzheimer's disease, IEEE Transactions on Medical Imaging 31 (2) (2012) 207–216, http://dx.doi.org/10.1109/TMI.2011.2167628.
[23] S.A. Meda, B. Narayanan, J. Liu, N.I. Perrone-Bizzozero, M.C. Stevens, V. D. Calhoun, D.C. Glahn, L. Shen, S.L. Risacher, A.J. Saykin, G.D. Pearlson, A.D. N. Initia, A large scale multivariate parallel ICA method reveals novel imaging-genetic relationships for Alzheimer's disease in the ADNI cohort, Neuroimage 60 (3) (2012) 1608–1621, http://dx.doi.org/10.1016/j.Neuroimage.2011.12.076.

Jim Jing-Yan Wang received his Ph.D. degree from the Graduate University of Chinese Academy of Sciences, China, 2102. Currently, he is a postdoctoral fellow at the Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology, Saudi Arabia. His research interests are machine learning, data mining, bioinformatics, and biometrics.

Halima Bensmail obtained her Ph.D. degree jointly from the Department of Statistics and Mathematics in Pierre & Marie Curie (Paris IV) University and National institute of Automatics and informatics (INRIA), Paris, France, in 1995. After winning the prestigious French Educational Award, she joined the University of Washington, Seattle, as a visiting scientist. After a short period at the Fred Hutchinson Cancer Research Center, she joined the University of Social and Behavioral Sciences of Leiden as a postdoc. She was appointed the assistant professor position at the University of Tennessee in 2000, was tenured, and promoted to Associate Professor in 2005. She joined the Eastern Virginia Medical School as an Associate Professor of Biostatistics and Bioinformatics in 2006. Currently, she is a senior scientist at the Qatar Computing Research Institute, Qatar where she is leading the Bioinformatics and scientific computing center. She is working broadly on statistical machine learning, applied it to medical areas for research that is referred to as: machine learning in Bioinformatics. Dr. Bensmail has published several articles in peer-reviewed conference proceeding and journals, including JASA, Statistics and Computing Journal, Bioinformatics, Plos One, computational sciences, Biomedicine and Biotechnology.

Xin Gao received the BS degree from Computer Science and Technology Department, Tsinghua University, China, 2004, and the PhD degree from David R. Cheriton School of Computer Science, University of Waterloo, Canada, 2009. He worked as a Lane Fellow in Lane Center for Computational Biology, Carnegie Mellon University, US, from 2009 to 2010. Currently, he is an assistant professor at the Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology, Saudi Arabia. His research interests are bioinformatics and computational biology.