Dimensionality Reduction using Relative Attributes

Mohammadreza Babaee[1], Stefanos Tsoukalas[1], Maryam Babaee[2]
Gerhard Rigoll[1], and Mihai Datcu[3]

[1] Institute for Human-Machine Communication, Technische Universität München
{reza.babaee,rigoll}@tum.de, s.tsoukalas@mytum.de
[2] Computer Engineering Dept. University of Isfahan, Iran
babaee@eng.ui.ac.ir
[3] The Remote Sensing Technology Institute (IMF), German Aerospace Center
mihai.datcu@dlr.de

1 Introduction

Visual attributes are high-level semantic description of visual data that are close to the language of human. They have been intensively used in various applications such as image classification [1,2], active learning [3,4], and interactive search [5]. However, the usage of attributes in dimensionality reduction has not been considered yet. In this work, we propose to utilize relative attributes as semantic cues in dimensionality reduction. To this end, we employ Non-negative Matrix Factorization (NMF) [6] constrained by embedded relative attributes to come up with a new algorithm for dimensionality reduction, namely attribute regularized NMF (ANMF).

2 Approach

We assume that $X \in \mathbb{R}^{D\times N}$ denotes N data points (e.g., images) represented by D dimensional low-level feature vectors. The NMF decomposes the non-negative matrix X into two non-negative matrices $U \in \mathbb{R}^{D\times K}$ and $V \in \mathbb{R}^{N\times K}$ such that the multiplication of U and V approximates the original matrix X. Here, U represents the bases and V contains the coefficients, which are considered as new representation of the original data. The NMF objective function is:

$$F = \left\|X - UV^{\mathrm{T}}\right\|_F^2 \qquad \text{s.t. } \begin{array}{l} U = [u_{ik}] \geq 0 \\ V = [v_{jk}] \geq 0. \end{array} \tag{1}$$

Additionally, we assume that M semantic attributes have been predefined for the data and the relative attributes of each image are available. Precisely, the matrix of relative attributes, $Q \in \mathbb{R}^{M\times N}$, has been learned by some ranking function (e,.g, rankSVM). Intuitively, those images which own similar relative attributes have similar semantic contents and therefore belong to the same semantic class. This concept can be formulated as a regularizer to be added to the main NMF objective function. This information can be formulated in a regularization term as

$$R = \alpha \left\| Q - AV^{\mathrm{T}} \right\|^2, \tag{2}$$

where $V = [\boldsymbol{v}_1, ..., \boldsymbol{v}_N]^{\mathrm{T}} \in \mathbb{R}^{N \times K}$ and the matrix $A \in \mathbb{R}^{M \times K}$. The matrix A linearly transforms and scales the vectors in the new representation in order to obtain the best fit for the matrix Q. The matrix A is allowed to take negative values and is computed as part of the NMF minimization. We arrive at the following minimization problem:

$$\min F = \left\| X - UV^{\mathrm{T}} \right\|^2 + \alpha \left\| Q - AV^{\mathrm{T}} \right\|^2 \qquad \begin{aligned} \text{s.t. } U &= [u_{ik}] \geq 0 \\ V &= [v_{jk}] \geq 0. \end{aligned} \tag{3}$$

2.1 Update rules

For the derivation of the update rules we expand the objective to

$$\begin{aligned} O = &\mathrm{Tr}\left(XX^{\mathrm{T}}\right) - 2\mathrm{Tr}\left(XVU^{\mathrm{T}}\right) + \mathrm{Tr}\left(UV^T VU^T\right) \\ &+ \alpha \mathrm{Tr}\left(QQ^{\mathrm{T}}\right) - \alpha 2\mathrm{Tr}\left(QVA^{\mathrm{T}}\right) + \alpha \mathrm{Tr}\left(AV^T VA^T\right) \end{aligned} \tag{4}$$

and introduce Lagrange multipliers $\Phi = [\phi_{ik}]$, $\Psi = [\psi_{jk}]$ for the constraints $[u_{ik}] \geq 0$, $[v_{jk}] \geq 0$ respectively. Adding the Lagrange multipliers and ignoring the constant terms leads to the Lagrangian:

$$\begin{aligned} \mathcal{L} = &- 2\mathrm{Tr}\left(XVU^{\mathrm{T}}\right) + \mathrm{Tr}\left(UV^T VU^T\right) + \mathrm{Tr}\left(\Phi U\right) + \mathrm{Tr}\left(\Psi V\right) \\ &- \alpha 2\mathrm{Tr}\left(QVA^{\mathrm{T}}\right) + \alpha \mathrm{Tr}\left(AV^T VA^T\right). \end{aligned} \tag{5}$$

The partial derivatives of $\mathcal{L}$ with respect to U, V and A are:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial U} &= -2XV + 2UV^{\mathrm{T}}V + \Phi \\ \frac{\partial \mathcal{L}}{\partial V} &= -2X^T U + 2VU^T U - \alpha 2Q^T A + \alpha 2VA^T A + \Psi \\ \frac{\partial \mathcal{L}}{\partial A} &= -2QV + 2AV^{\mathrm{T}}V \end{aligned} \tag{6}$$

For the derivation of the update rules for U and V we apply the KKT-conditions $\phi_{ik} u_{ik} = 0$, $\psi_{jk} v_{jk} = 0$ [7]. For A the update rules can be derived directly by setting its derivative of the Lagrangian to 0. Thus, we arrive at the following equations:

$$u_{ik} \leftarrow u_{ik} \frac{[XV]_{ik}}{[UV^{\mathrm{T}}V]_{ik}} \tag{7}$$

$$v_{jk} \leftarrow v_{jk} \frac{[X^{\mathrm{T}}U + \alpha(VA^{\mathrm{T}}A)^- + \alpha(Q^{\mathrm{T}}A)^+]_{jk}}{[VU^{\mathrm{T}}U + \alpha(VA^{\mathrm{T}}A)^+ + \alpha(Q^{\mathrm{T}}A)^-]_{jk}} \tag{8}$$

$$A \leftarrow QV(V^{\mathrm{T}}V)^{-1} \tag{9}$$

where for a matrix M we define M^+, M^- as $M^+ = (|M| + M)/2$ and $M^- = (|M| - M)/2$. The newly introduced terms depend only on the variables V and A. The proof of convergence for the update rules for V and A can be found here[4].

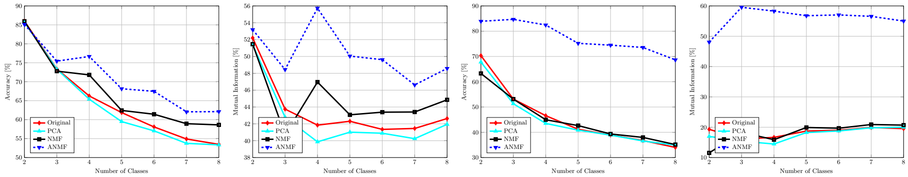
3 Experiments

We perform our experiments by applying the proposed method on two borrowed datasets from [3], namely Outdoor Scene Recognition (OSR) and Public Figure Face Database (PubFig). The OSR dataset contains 2688 images from 8 categories and the PubFig contains 772 images from 8 different individuals. The OSR images are represented by 512-dimensional gist [8] features and PubFig images are represented by a concatenation of gist descriptor and a 45-dimensional Lab color histogram [3]. We also utilize the learned relative attributes for both datasets from [3].

First, we reduce the dimensionality of the data using the proposed method (ANMF) and also PCA, and NMF. Then, we apply K-Means clustering algorithm on the new representation of the data and also on the original data. We perform the experiments with different number of classes, k, extracted from each dataset. In order to obtain representative results, we repeat the experiments 10 times for each k, by selecting a random subset of k classes from the dataset and computing the average results. For the dimensionality reduction techniques (i.e. PCA, NMF, and ANMF), we always set the new dimension equal to the number of classes. We use two metrics to evaluate the performance of the compared algorithms, namely accuracy (AC) and normalized mutual information (nMI) [9]. The K-Means runs 20 times in each experiment and the best result are selected. In ANMF, the regularization parameter is chosen by running cross-validation on each dataset.

The results of our experiments are depicted in Figure 1. The accuracy and normalized mutual information of clustering results for OSR dataset are depicted in Figure 1a and 1b, respectively. Figures 1c and 1d show the accuracy and mutual information of clustering results for PubFig dataset, respectively. As it can be seen, the proposed method that utilizes the relative attribute outperforms largely the other techniques in both datasets. For PubFig dataset we even achieve $75\% - 85\%$ accuracy. Additionally, we reduce the dimensionality of both datasets to 2D for visualization (see Figures 2a and 2b). Here, it is clearly observable that those image which share similar attributes are located closely. For instance, in OSR dataset, the images sharing openness attribute are in left down part of layout. Additionally, the plots of convergence rate of the proposed algorithm applied on both OSR and PubFig datasets are represented in Figures 3a and 3b, respectively. The convergence plots show that the algorithm converges after 20 iterations, which means the running time is quite small. The experimental results confirm, that the proposed method learns the bases with different semantic attributes.

[4] http://www.mmk.ei.tum.de/%7Erez/convergenceproof.pdf

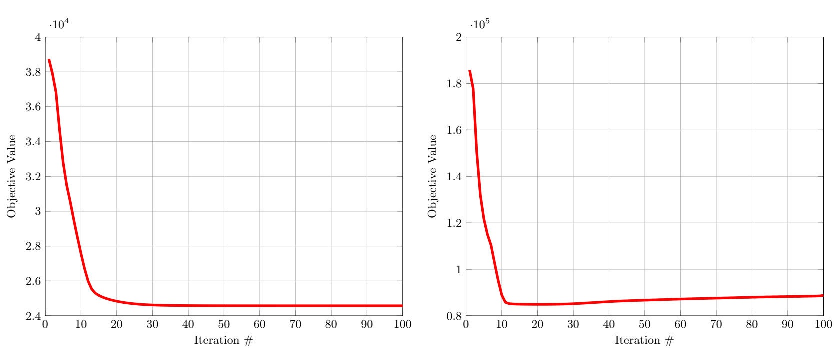


(a) (b) (c) (d)

Fig. 1: Clustering results on new representation computed by PCA, NMF, ANMF and original data evaluated by accuracy (AC) and normalized mutual information (nMI). (a) and (b) show the AC, nMI for the OSR dataset, respectively. (c) and (d) show the AC and nMI for the PubFig dataset, respectively.

(a) (b)

Fig. 2: 2D visualization of the datasets computed by the proposed method (ANMF); (a) OSR dataset; (b) PubFig dataset.



(a) (b)

Fig. 3: Convergence plot of the proposed algorithm applied on (a) OSR dataset and (b) PubFig dataset.

References

1. Kumar, N., Berg, A.C., Belhumeur, P.N., Nayar, S.K.: Attribute and simile classifiers for face verification. In: Computer Vision, 2009 IEEE 12th International Conference on, IEEE (2009) 365–372
2. Farhadi, A., Endres, I., Hoiem, D., Forsyth, D.: Describing objects by their attributes. In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, IEEE (2009) 1778–1785
3. Parikh, D., Grauman, K.: Relative attributes. In: Computer Vision (ICCV), 2011 IEEE International Conference on, IEEE (2011) 503–510
4. Kovashka, A., Vijayanarasimhan, S., Grauman, K.: Actively selecting annotations among objects and attributes. In: Computer Vision (ICCV), 2011 IEEE International Conference on, IEEE (2011) 1403–1410
5. Kovashka, A., Parikh, D., Grauman, K.: Whittlesearch: Image search with relative attribute feedback. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE (2012) 2973–2980
6. Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. Nature **401**(6755) (1999) 788–791
7. Boyd, S., Vandenberghe, L.: Convex optimization. Cambridge university press (2009)
8. Oliva, A., Torralba, A.: Modeling the shape of the scene: A holistic representation of the spatial envelope. International journal of computer vision **42**(3) (2001) 145–175
9. Xu, W., Liu, X., Gong, Y.: Document clustering based on non-negative matrix factorization. In: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval, ACM (2003) 267–273