

ATTRIBUTE CONSTRAINED SUBSPACE LEARNING

Mohammadreza Babaei¹, Maryam Babaei², Daniel Merget¹, Philipp Tiefenbacher¹, Gerhard Rigoll¹

¹Institute for Human-Machine Communication, Technische Universität München, Germany

²Dept. of Artificial Intelligence, Faculty of Computer Engineering, University of Isfahan, Iran

ABSTRACT

Visual attributes are high-level semantic descriptions of visual data that are close to the human language. They have been used intensively in various applications such as image classification, active learning, and interactive search. However, the usage of attributes in subspace learning (or dimensionality reduction) has not been considered yet. In this work, we propose to utilize relative attributes as semantic cues in subspace learning. To this end, we employ Non-negative Matrix Factorization (NMF) constrained by embedded relative attributes to learn a subspace representation of image content. Experiments conducted on two datasets show the efficiency of attributes in discriminative subspace learning.

Index Terms— Subspace learning, non-negative matrix factorization, relative attributes

1. INTRODUCTION

Image content representation plays a key role in computer vision and pattern recognition. Modern techniques such as Bag-of-Words models of local features (e.g., SIFT [1], Weber [2], Gabor [3]) represent an image by a very high-dimensional feature vector. Although this representation leads to relatively high accuracy in visual recognition and search, it increases the computation time and, consequently, is improper for real-time applications. Therefore, developing new algorithms that generate a compact and informative representation of image content is highly needed. Perhaps, the most common way to tackle this problem is learning a subspace of the original feature space and using this representation for the recognition tasks. For instance, Principal Component Analysis (PCA) [4] is the most commonly used algorithm to reduce the dimensionality of feature space. However, the obtained results are not necessarily discriminative.

Non-negative Matrix Factorization (NMF) [5, 6] is a parts-based representation technique that has been widely used in different data mining applications such as blind source separation and document clustering [7, 8, 9]. In many cases, NMF is employed to generate a new representation of the image content from high-dimensional features. The beauty of this technique is the ability to constrain the NMF to generate a new representation with desirable properties. For instance, in

Graph regularized NMF (GNMF) [10], a regularizer based on the Laplacian of the original neighborhood graph is coupled with the main objective function of NMF, in order to preserve locality within the subspace representation. The label information of a fraction of data points is used in Constrained NMF (CNMF) [11] to generate a discriminative subspace.

For several years, the representation of images with visual attributes has been studied intensively by researchers in the fields of clustering, classification, object recognition, and face verification [12, 13, 14, 15]. Generally, these attributes are observable semantic cues, which can be learned from low-level features. For example, “smiling” and “dry” can be considered as attributes of a face or a scene, respectively. Recently, it has been proposed that *relative attributes* provide a richer source of semantic information in images [15] than binary attributes. They depict the strength of attributes in an image and can be predicted by pre-learned rank functions. For each attribute, a single rank function, which is a rank-SVM, is learned from a set of training data [15]. In this work, we use predicted relative attributes, as discriminant constraints to guide a NMF to generate a new subspace of images. More precisely, the relative attributes are embedded in a regularizer coupled with the NMF objective function. We call our proposed method Attribute constrained NMF (ANMF).

The rest of the paper is organized as follows: In Section 2, we briefly recapitulate NMF. In Section 3, we describe our proposed method. We describe and discuss our conducted experiments in Section 4. Finally, we draw our conclusion in Section 5.

2. A REVIEW OF NMF

We consider a data matrix $X = [\mathbf{x}_1, \dots, \mathbf{x}_N] \in \mathbb{R}^{M \times N}$, where $\mathbf{x}_i$ is a feature vector, N is the number of samples and M is the dimension of the feature vectors. Given a new reduced dimension K , the NMF algorithm approximates the matrix X by a product of two non-negative matrices $U = [u_{ik}] \in \mathbb{R}^{M \times K}$ and $V = [v_{jk}] \in \mathbb{R}^{N \times K}$:

$$X \approx UV^T \quad (1)$$

In the new representation, U can be considered as a set of basis vectors and $V = [\mathbf{v}_1, \dots, \mathbf{v}_N]^T$ as the coordinates of

each sample with respect to these basis vectors. One popular cost function that quantifies the quality of the approximation is the Frobenius norm of the matrix differences [5]. For the Frobenius norm, the problem is stated as:

$$\begin{aligned} \min O_F &= \|X - UV^T\|^2 \\ \text{s.t. } U &= [u_{ik}] \geq 0 \\ V &= [v_{jk}] \geq 0 \end{aligned} \quad (2)$$

While the minimization problem is convex with respect to U and with respect to V , it is not convex in both variables together. Therefore, there typically exist many local minima. In [5], Lee and Seung present update rules for the two minimization problems and prove their convergence. These update rules are

$$u_{ik} \leftarrow u_{ik} \frac{(XV)_{ik}}{(UV^TV)_{ik}}, \quad v_{jk} \leftarrow v_{jk} \frac{(X^TU)_{jk}}{(VU^TU)_{jk}} \quad (3)$$

3. ATTRIBUTE CONSTRAINED NMF

Relative attributes are a rich source of semantic information. They indicate the strength of an attribute in an image with respect to the others. Let $Q \in \mathbb{R}^{S \times N}$ hold the relative attributes learned from the images in a separate step, where S is the number of attributes and N is the number of images. Based on the introduced matrix, we define a regularizer to be added to the main objective function of NMF:

$$O_L = \alpha \|Q - AV^T\|^2, \quad (4)$$

where $V \in \mathbb{R}^{N \times K}$ is a K -dimensional subspace representation of image content to be learned and the matrix $A \in \mathbb{R}^{S \times K}$ linearly transforms and scales the vectors V in order to obtain the best match 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:

$$\begin{aligned} \min O &= \|X - UV^T\|^2 + \alpha \|Q - AV^T\|^2 \\ \text{s.t. } U &= [u_{ik}] \geq 0 \\ V &= [v_{jk}] \geq 0 \end{aligned} \quad (5)$$

3.1. Update rules

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

$$\begin{aligned} O &= \text{Tr}(XX^T) - 2\text{Tr}(XVU^T) + \text{Tr}(UV^TVU^T) \\ &\quad + \alpha \text{Tr}(QQ^T) - \alpha 2\text{Tr}(QVA^T) + \alpha \text{Tr}(AV^TVA^T) \end{aligned} \quad (6)$$

We 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:

$$\begin{aligned} \mathcal{L} &= -2\text{Tr}(XVU^T) + \text{Tr}(UV^TVU^T) + \text{Tr}(\Phi U) + \text{Tr}(\Psi V) \\ &\quad - \alpha 2\text{Tr}(QVA^T) + \alpha \text{Tr}(AV^TVA^T). \end{aligned} \quad (7)$$

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

$$\frac{\partial \mathcal{L}}{\partial U} = -2XV + 2UV^TV + \Phi \quad (8)$$

$$\frac{\partial \mathcal{L}}{\partial V} = -2X^TU + 2VU^TU - \alpha 2Q^TA + \alpha 2VA^TA + \Psi \quad (9)$$

$$\frac{\partial \mathcal{L}}{\partial A} = -2QV + 2AV^TV \quad (10)$$

For the derivation of the update rules for U and V we solve the equations $\frac{\partial \mathcal{L}}{\partial U} = 0$ and $\frac{\partial \mathcal{L}}{\partial V} = 0$ in terms of Φ and Ψ , respectively, and apply the Karush-Kuhn-Tucker (KKT) conditions $\phi_{ik}u_{ik} = 0$, $\psi_{jk}v_{jk} = 0$ [16]. For A , since we have no Lagrange multiplier, the update rules can be derived directly by setting $\frac{\partial \mathcal{L}}{\partial A} = 0$ and solving for A . We arrive at the following equations:

$$u_{ik} \leftarrow u_{ik} \frac{[XV]_{ik}}{[UV^TV]_{ik}} \quad (11)$$

$$v_{jk} \leftarrow v_{jk} \frac{[X^TU + \alpha(VA^TA)^- + \alpha(Q^TA)^+]_{jk}}{[VU^TU + \alpha(VA^TA)^+ + \alpha(Q^TA)^-]_{jk}} \quad (12)$$

$$A \leftarrow QV(V^TV)^{-1} \quad (13)$$

where for a matrix M we define $M_{ij}^+ = (|M_{ij}| + M_{ij})/2$ and $M_{ij}^- = (|M_{ij}| - M_{ij})/2$. As expected, the update rule for U remains the same as in the original algorithm [5], since the newly introduced terms depend only on the variables V and A .

4. EXPERIMENTS

We perform our experiments by applying the proposed method on two image datasets, namely Outdoor Scene Recognition (OSR) and Public Figure Face Database (PubFig) [15]. The OSR dataset contains 2688 images from 8 categories and PubFig contains 772 images from 8 different individuals. The OSR images are represented by 512-dimensional GIST [17] features, while PubFig images are represented by a concatenation of GIST descriptors and a 45-dimensional Lab color histogram [15]. We also utilize the learned relative attributes for both datasets from [15].

4.1. Evaluation metrics

We use two metrics to evaluate the performance of k-means clustering on the learned subspaces, each computed by a different method. These metrics are accuracy (AC) and normalized Mutual Information (nMI) [18]. The accuracy indicates

the percentage of correctly clustered data points, whereas the normalized mutual information depicts how well the obtained clusters match the ground truth. The details of these evaluation metrics can be found in [10, 11].

4.2. Setting

We apply the proposed method (ANMF), PCA, and NMF on the original representations of both datasets to generate (learn) different subspaces of the data. Then we apply k-means clustering on the new subspaces and also on the original data with k equal to the dimension of subspaces. We perform the experiments with k different classes, sampled from each dataset. In order to obtain representative results, we repeat the experiments 10 times for each k . The k-means runs 20 times per experiment and the best result is selected. For the subspace learning techniques (i.e., PCA, NMF, ANMF), we always set the new dimension equal to the number of classes. In ANMF, the regularization parameter is chosen by running a cross-validation on each dataset. For the OSR dataset and the PubFig dataset, this parameter was 10 and 100, respectively.

4.3. Results

To visualize the datasets in 2D, we set the subspace dimension to two. The results are depicted in Figs. 1(a) and 1(b) for OSR and PubFig datasets, respectively. Here, it is clearly observable that those images which share similar attributes are located close to each other. For instance, the OSR images sharing the openness attribute reside in the bottom left part of the layout.

The clustering results on the learned subspaces are depicted in Fig. 2. Figs. 2(a) and 2(b) show the accuracy and normalized Mutual Information (nMI) of the clustering results for the PubFig dataset. The OSR results are depicted in Figs. 2(d) and 2(e). It can be seen that the proposed method outperforms the other techniques significantly on both datasets. For the PubFig dataset we even achieve 75% – 85% accuracy. The convergence rate of the proposed algorithm applied on each dataset is presented in Figs. 2(c) and 2(f). The algorithm converges quickly after 20 iterations and therefore can be considered computationally efficient.

The experimental results confirm that the proposed method learns the bases with different semantic attributes successfully.

5. CONCLUSIONS

We proposed to use relative attributes in the context of NMF to learn different discriminative subspaces of image content. In contrast to other works, where label information is used to increase the discriminative property of the subspace representation, the proposed algorithm uses predicted relative attributes. We embedded these relative attributes in a regular-



Fig. 1. 2D visualization of the datasets computed by the proposed method (ANMF); (a) OSR dataset; (b) PubFig dataset. Images sharing the same attribute are located close to each other.

izer coupled with the main objective function of NMF. We solved the optimization problem by deriving an updating rule for the subspace representation. Our experiments on two image datasets confirm that the learned subspaces are very discriminative and produce good clustering results in comparison to other state-of-the-art methods. It will be interesting to examine whether the usage of relative attributes is also beneficial in other domains such as dictionary learning, kernel learning, automatic search and recognition.

ACKNOWLEDGMENTS

The authors would like to thank Munich Aerospace for the financial support.

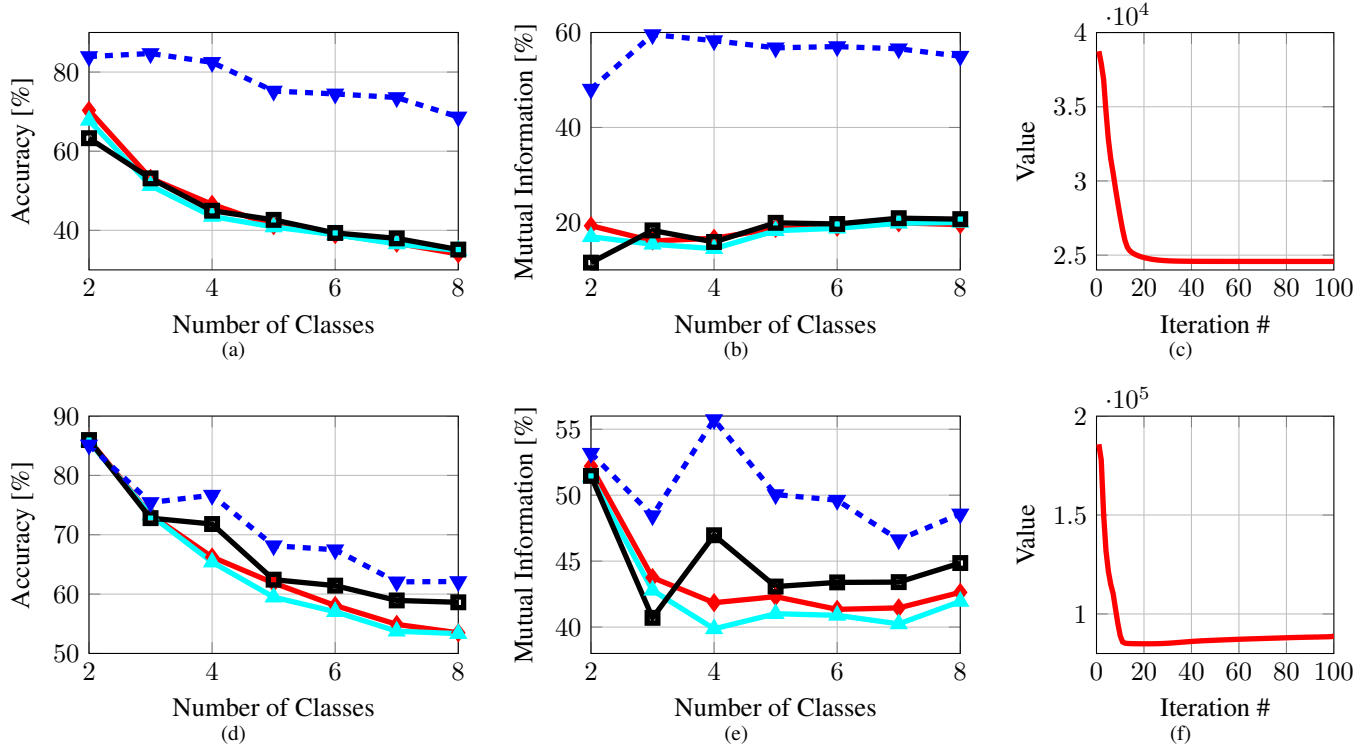


Fig. 2. Clustering results computed by PCA (cyan), NMF (black), ANMF (blue) and original data (red) evaluated by accuracy (AC) and normalized mutual information (nMI). (a) and (b) show the AC and nMI for the OSR dataset, respectively. (c) and (d) show the AC and nMI for the PubFig dataset, respectively.

Table 1. Clustering results of different subspace learning methods on the PubFig dataset

k	Accuracy (AC)				normalized Mutual Information (nMI)			
	KMeans	PCA	NMF	ANMF	KMeans	PCA	NMF	ANMF
2	86.0 ± 14.5	85.7 ± 14.6	86.0 ± 14.2	85.1 ± 16.1	52.2 ± 27.8	51.4 ± 28.1	51.5 ± 27.7	53.2 ± 34.0
3	73.4 ± 12.6	73.3 ± 12.4	72.8 ± 13.2	75.4 ± 11.7	43.8 ± 16.1	42.8 ± 16.0	40.7 ± 18.6	48.4 ± 15.3
4	66.2 ± 8.6	65.4 ± 7.1	71.8 ± 9.0	76.7 ± 9.4	41.8 ± 11.3	39.9 ± 9.4	47.0 ± 11.1	55.7 ± 10.5
5	61.9 ± 8.6	59.5 ± 10.2	62.4 ± 8.9	68.2 ± 8.8	42.3 ± 7.7	41.0 ± 7.2	43.1 ± 8.1	50.0 ± 9.3
6	58.1 ± 9.6	57.0 ± 9.4	61.4 ± 8.2	67.5 ± 7.4	41.3 ± 9.5	40.9 ± 9.4	43.4 ± 7.5	49.6 ± 6.6
7	54.9 ± 5.0	53.7 ± 4.5	58.9 ± 3.5	62.1 ± 5.1	41.5 ± 4.3	40.2 ± 4.1	43.4 ± 2.8	46.6 ± 4.2
8	53.5 ± 2.2	53.3 ± 2.4	58.6 ± 2.9	62.1 ± 3.1	42.6 ± 1.6	41.9 ± 1.1	44.9 ± 1.6	48.6 ± 3.3

Table 2. Clustering results of different subspace learning methods on the OSR dataset

k	Accuracy (AC)				normalized Mutual Information (nMI)			
	KMeans	PCA	NMF	ANMF	KMeans	PCA	NMF	ANMF
2	70.3 ± 13.6	67.8 ± 14.2	63.3 ± 13.7	83.9 ± 14.8	19.3 ± 21.7	17.0 ± 21.7	11.5 ± 21.1	48.1 ± 26.9
3	53.1 ± 11.1	51.3 ± 9.1	53.1 ± 6.7	84.7 ± 6.9	16.1 ± 14.3	15.4 ± 12.3	18.3 ± 8.7	59.6 ± 9.7
4	46.6 ± 7.5	43.5 ± 6.3	45.0 ± 7.2	82.4 ± 4.1	16.7 ± 8.7	14.5 ± 6.6	15.9 ± 8.9	58.3 ± 6.8
5	41.2 ± 5.0	40.8 ± 4.8	42.6 ± 6.4	75.2 ± 9.2	18.7 ± 6.2	18.2 ± 5.8	19.9 ± 6.4	56.8 ± 7.9
6	38.9 ± 4.6	38.9 ± 4.2	39.3 ± 4.3	74.5 ± 5.1	19.0 ± 4.0	18.8 ± 3.7	19.7 ± 3.5	57.0 ± 4.8
7	36.7 ± 3.9	36.6 ± 3.6	38.0 ± 3.7	73.6 ± 4.3	19.9 ± 4.4	19.8 ± 4.0	20.9 ± 3.3	56.6 ± 3.1
8	34.0 ± 1.6	34.9 ± 1.4	35.1 ± 1.6	68.7 ± 4.2	19.5 ± 1.1	20.0 ± 1.7	20.7 ± 1.2	55.0 ± 3.2

6. REFERENCES

- [1] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [2] J. Chen, S. Shan, G. Zhao, X. Chen, W. Gao, and M. Pietikainen, "A robust descriptor based on weber's law," in *Computer Vision and Pattern Recognition, CVPR 2008. IEEE Conference on*, pp. 1–7.
- [3] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *Image processing, IEEE Transactions on*, vol. 11, no. 4, pp. 467–476, 2002.
- [4] I. Jolliffe, *Principal component analysis*, Wiley Online Library, 2005.
- [5] D. Lee and H.S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in neural information processing systems*, 2001, pp. 556–562.
- [6] M. Babae, S. Tsoukalas, G. Rigoll, and M. Datcu, "Discriminative feature learning from sar images," in *International Conference on Big Data from Space*, 2014.
- [7] W. Xu, X. Liu, and Y. Gong, "Document clustering based on non-negative matrix factorization," in *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2003, pp. 267–273.
- [8] A. Cichocki, R. Zdunek, A. H. Phan, and S.-i. Amari, *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*, John Wiley & Sons, 2009.
- [9] C. Vollmer, S. Hellbach, J. Eggert, and H. Gross, "Sparse coding of human motion trajectories with non-negative matrix factorization," *Neurocomputing*, vol. 124, pp. 22–32, 2014.
- [10] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 8, pp. 1548–1560, 2011.
- [11] H. Liu, Z. Wu, X. Li, D. Cai, and T. S. Huang, "Constrained nonnegative matrix factorization for image representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 7, pp. 1299–1311, 2012.
- [12] A. Farhadi, I. Endres, D. Hoiem, and D. Forsyth, "Describing objects by their attributes," in *Computer Vision and Pattern Recognition, CVPR 2009. IEEE Conference on*, pp. 1778–1785.
- [13] N. Kumar, A. C Berg, P.N. Belhumeur, and S. Nayar, "Attribute and simile classifiers for face verification," in *Computer Vision, 12th IEEE International Conference on*, 2009, pp. 365–372.
- [14] A. Kovashka and K. Grauman, "Attribute pivots for guiding relevance feedback in image search," in *Computer Vision (ICCV), 2013 IEEE International Conference on*. IEEE, 2013, pp. 297–304.
- [15] D. Parikh and K. Grauman, "Relative attributes," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 503–510.
- [16] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge university press, 2009.
- [17] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *International journal of computer vision*, vol. 42, no. 3, pp. 145–175, 2001.
- [18] M. Babae, R. Bahmanyar, G. Rigoll, and M. Datcu, "Farness preserving non-negative matrix factorization," in *IEEE International Conference on Image Processing (ICIP)*, 2014, pp. 3023–3027.