

# Transfer Nonnegative Matrix Factorization for Image Representation

Tianchun Wang<sup>1</sup> (✉), TengQi Ye<sup>2</sup>, and Cathal Gurrin<sup>2</sup>

<sup>1</sup> School of Software, TNList, Tsinghua University, Beijing, China  
wtc13@mails.tsinghua.edu.cn

<sup>2</sup> Insight Centre for Data Analytics, Dublin City University, Dublin, Ireland  
yetengqi@gmail.com, cgurrin@computing.dcu.ie

**Abstract.** Nonnegative Matrix Factorization (NMF) has received considerable attention due to its psychological and physiological interpretation of naturally occurring data whose representation may be parts-based in the human brain. However, when labeled and unlabeled images are sampled from different distributions, they may be quantized into different basis vector space and represented in different coding vector space, which may lead to low representation fidelity. In this paper, we investigate how to extend NMF to cross-domain scenario. We accomplish this goal through TNMF - a novel semi-supervised transfer learning approach. Specifically, we aim to minimize the distribution divergence between labeled and unlabeled images, and incorporate this criterion into the objective function of NMF to construct new robust representations. Experiments show that TNMF outperforms state-of-the-art methods on real datasets.

**Keywords:** Nonnegative matrix factorization · Transfer learning · Image representation

## 1 Introduction

The development of online images and videos has created a compelling demand for advanced technologies for organizing and analyzing the multimedia content. As a powerful technique for finding succinct representations of stimuli and capturing high-level semantics of visual data, *Nonnegative Matrix Factorization* (NMF) can represent images using the combination of nonnegative low-dimensional “basis” vectors. This makes NMF an ideal dimensionality reduction algorithm for image processing [1], face recognition [2,3], and image clustering [4], where it is natural to consider the object as a combination of the parts.

One major problem of NMF is how to improve the quality of the part-based representation while maximally preserving the signal fidelity. To achieve this goal, many works have been proposed to leverage more knowledge. Kuang et al. [4] provide a symmetric NMF method for clustering based on graph. Ding et al. [15] proposed a semi-nonnegative NMF while relaxing the constraint on

the basis vectors. Ding et al. [16] also show that when the Frobenius norm is used as divergence, NMF is equivalent to a relaxed form of K-means clustering, which justifies the use of NMF for data clustering. However, when unlabeled and labeled data are sampled from different distributions, they may be quantized into different coding vectors corresponding to different low-dimensional feature representations. In this case, the basis vectors learned from the feature space of labeled images cannot effectively indicate the features of unlabeled images with high fidelity, and also the unlabeled images may reside far away from the labeled images under the new representation. This distribution difference will greatly challenge the robustness of existing NMF algorithms for cross-distribution image classification problems [23].

Transfer learning [5] has recently become a hot research topic that the labeled training data and unlabeled test data are sampled from different probability distributions. This is a very common scenario in web applications, since training and test data are usually collected in different time periods, or under different conditions. In this case, standard classifiers trained on the labeled data may fail to make reasonable predictions on the unlabeled data. To boost the generalization performance of supervised classifiers across different distributions, Pan et al. [6, 7] proposed to extract a low-dimensional feature representation through which the probability distributions of labeled and unlabeled data are drawn close and achieves much better classification performance by explicitly reducing distribution divergence.

Inspired by recent efforts in both matrix factorization and transfer learning, this paper presents a novel Transfer Nonnegative Matrix Factorization (TNMF) algorithm to construct robust representations for classifying cross-domain images accurately. We aim to minimize the distribution divergence between labeled and unlabeled images with a nonparametric distance measure. Specifically, we incorporate this criterion into the objective function of TNMF to make the new representations of the labeled and unlabeled images close to each other. In this way, the induced representations are made robust for cross-distribution image classification problems. Moreover, to enrich the new representations with the intrinsic nonlinear structure of data space, we also incorporate the Hessian regularization term of coefficients [9, 11, 12] in our objective function. Different from Laplacian term [8, 14], Hessian has a richer null space and drives the solution varying smoothly along the manifold. Experimental results verify the effectiveness of the TNMF approach.

## 2 Related Work

In this section, we review related efforts including NMF and transfer learning.

Recently, NMF has been a hot research topic of computer vision. Ding et al. [15] proposed a semi-nonnegative algorithm where only one matrix factor is restricted to contain nonnegative entries, while it relaxes the constraint on the basis vectors. Ding et al. [16] also show that when the Frobenius norm is used as a divergence, NMF is equivalent to a relaxed form of K-means clustering, which

justifies the use of NMF for data clustering. Heiler et al. [17] derive optimization schemes for NMF based on sequential quadratic and second order cone programming. Our work aims to discover a shared basic matrix which can encode both labeled and unlabeled data sampled from different probability distributions. To improve the quality of low-rank representation, researchers have modified the nonnegative constraint [18], adding graph regularization [8], etc. Our approach aims to construct robust low-dimensional representations for image classification problems across domains, which is different from the previous works.

Aiming at transfer knowledge between the labeled and unlabeled data sampled from different distributions, transfer learning [5] has received extensive research focus. To achieve this goal, Pan et al. [7] proposed a Transfer Component Analysis (TCA) method to minimize the reconstruction error of the input data by reducing the discrepancy between the labeled and unlabeled data. Quanz et al. [19] have explored knowledge transfer of sparse feature, which is a more restricted procedure and prone to overfitting. In addition, Wang et al. [20] extends NMF to cross-domain scenario. They assume that a small set of the target domain data are labeled and an SVM classifier is learned across domains. Different from them, our approach is based on the setting that there is no labeled data in the target domain, which is more restricted. Moreover, our work additionally incorporates the Hessian term of coefficients [9] in the objective function, which can discover more discriminating representations for classification tasks.

### 3 Preliminaries

In this section, we briefly introduce a basic knowledge of NMF and Hessian Regularization.

#### 3.1 Nonnegative Matrix Factorization

Non-negative Matrix Factorization (NMF) [2] is a matrix factorization algorithm that focuses on the analysis of data matrices whose elements are nonnegative. Given data matrix  $X = [x_1, \dots, x_N] \in R^{M \times N}$ , each column of  $X$  is a sample vector. The goal of NMF is to find a basic matrix  $U = [u_{ik}] \in R^{M \times K}$  and a coding matrix  $V = [v_{jk}] \in R^{N \times K}$  and the production of them can well approximate the original matrix  $X$ :

$$X \approx UV^T.$$

The commonly used cost function is the square of the Euclidean distance between two matrices give by:

$$\min_{U, V} \| X - UV^T \|^2 = \sum_{i,j} \left( x_{ij} - \sum_{k=1}^K u_{ik}v_{ik} \right)^2. \quad (1)$$

The objective function in Eq. (1) is convex in either U or V. Therefore, it can be solved by alternately optimizing one variable while fixing the other one.

### 3.2 Hessian Regularization

To make the basis vectors respect the intrinsic geometric structure underlying the input data, Cai et al. [8] proposed a Laplacian regularized NMF (we call it LapNMF) method which is based on the assumption that if two data points  $x_i$  and  $x_j$  are close in the intrinsic geometry of data distribution, then their coding vectors  $v_i$  and  $v_j$  are also close.

**Table 1.** Notations and descriptions used in this paper

Notation	Description
X	Input data matrix
U	Basic matrix
V	Coding matrix
M	MMD matrix
H	Hessian regularization matrix
$D_l, D_u$	Labeled/unlabeled data
$n_l, n_u$	Labeled/unlabeled example index
$m$	Dimensionality of shared feature space
$\lambda$	Trade-off parameter of Hessian regularization
$\mu$	Trade-off parameter of MMD regularization

However, Laplacian regularization is short of extrapolating power. The null space of the graph Laplacian is a constant function along the compact support of the marginal distribution and thus the solution of the Laplacian regularization is biased toward a constant function. In contrast to Laplacian, Hessian can properly exploit the intrinsic local geometry of the data manifold and has a richer nullspace to make the learned function vary linearly along the underlying manifold [9, 10]. The discretization of the Hessian regularization for encoding the local geometry of unlabeled samples is achieved as follows [9, 11, 12]:

- (1) Finding the  $k$ -nearest neighbours  $N_p$  of the  $j$ -th unlabeled sample  $x_j$  and centralizing the neighbourhood by taking  $x_j$  off from the  $k$ -nearest neighbours. This centralization makes  $x_j$  to be the origin of the tangent space  $T_{x_j}(M)$ .
- (2) Estimating the orthonormal coordinate system of the tangent space  $T_{x_j}(M)$  by the eigenspace  $U$  of the neighborhood  $N_p$  of  $p$  associated with the largest  $d$  eigenvalues. This can be implemented by conducting singular value decomposition on  $X^j = [x_i - x_j]_{i=1}^k$ , where in  $x_i$  is the  $i$ -th sample in the  $k$ -nearest neighbours  $N_p$ .
- (3) Using Gram-Schmidt orthonormalization, we take the  $(d + 1)$ -dimensional nullspace off from the matrix  $H^j = [1, u_1, \dots, u_1 u_1, \dots, u_d u_d]$  and get  $\hat{H}^j$ . Therefore, the Frobenius norm of  $\hat{H}^j$  is given by  $(\hat{H}^j)^T \hat{H}^j$ .

- (4) Accumulating  $(\hat{H}^j)^T \hat{H}^j$  over all images and then resulting the Hessian regularization matrix  $H$ .

Inspired by the recent progress of Hessian regularization method [11–13], in this paper, we integrate the minimizing criterion  $tr(VHV^T)$  for preserving the geometric structure into Eq. (1) and get the Hessian Regularized NMF (HeNMF) given by:

$$\min_{U,V} \|X - UV^T\|^2 + \lambda tr(VHV^T) \quad (2)$$

where  $\lambda$  is the regularization parameter to trade off the weight between NMF and geometric preservation.

## 4 Transfer Nonnegative Matrix Factorization

In this section, we will present the Transfer Nonnegative Matrix Factorization (TNMF) algorithm for image representation, which extends HeNMF by taking into account the minimization of distribution divergence between labeled and unlabeled data.

### 4.1 Problem Definition

Given labeled data  $\mathcal{D}_l = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_{n_l}, y_{n_l})\}$  with  $n_l$  examples, unlabeled data  $\mathcal{D}_u = \{\mathbf{x}_{n_l+1}, \dots, \mathbf{x}_{n_l+n_u}\}$  with  $n_u$  examples, denote  $X = [\mathbf{x}_1, \dots, \mathbf{x}_n] \in \mathbb{R}^{m \times n}$ ,  $n = n_l + n_u$  as the input data matrix. Assume that the labeled and unlabeled data are sampled from different probability distributions in an  $m$ -dimensional feature space. Therefore, we can define the problem our problem as follows:

*Problem 1 (Transfer Nonnegative Matrix Factorization):* Given labeled data  $\mathcal{D}_l$  and unlabeled data  $\mathcal{D}_u$  under different distributions, we aim to learn a coding matrix  $V$ , and basic matrix  $U$ , to construct a robust representation for original images sampled from  $\mathcal{D}_l$  and  $\mathcal{D}_u$ .

Notations and descriptions used frequently in this paper are summarized in Table 1.

### 4.2 Proposed Approach

In order to make NMF capture the representation across different distributions, we expect that the basis vectors can represent the shared knowledge underlying both labeled and unlabeled domains. However, the difference of extracted coding vector spaces between labeled and unlabeled data will still be significantly large. Thus a major computational problem is to reduce the distribution difference by explicitly minimizing some predefined distance measures. To achieve this goal, a common strategy is to make the probability distributions of labeled and unlabeled data close to each other in the low-rank representation. Therefore, by representing all data points  $X$  with the learned coding matrix  $V$ , the

probability distributions of the coding vectors for the labeled and unlabeled data should be close enough. In this paper, we adopt the joint distribution adaptation *Maximum Mean Discrepancy* [14, 23] as the nonparametric distance measure to compare different distributions, which minimize both the distance of marginal and conditional distribution divergence between labeled and unlabeled data in the coding vector spaces:

$$\left\| \frac{1}{n_l^{(c)}} \sum_{x_i \in \mathcal{D}_l^{(c)}} v_i - \frac{1}{n_u^{(c)}} \sum_{x_j \in \mathcal{D}_u^{(c)}} v_j \right\|^2 = \sum_{i,j=1}^n v_i^T v_j M_{ij}^{(c)} = \text{tr} \left( V M^{(c)} V^T \right)$$

where  $M^{(c)}$ ,  $c \in \{1, 2, \dots, C\}$  are MMD matrices given by:

$$M_{ij}^{(c)} = \begin{cases} \frac{1}{n_l^{(c)} n_l^{(c)}}, & x_i, x_j \in D_l^{(c)} \\ \frac{1}{n_u^{(c)} n_u^{(c)}}, & x_i, x_j \in D_u^{(c)} \\ -\frac{1}{n_l^{(c)} n_u^{(c)}}, & \begin{cases} x_i \in D_l^{(c)}, x_j \in D_u^{(c)} \\ x_j \in D_l^{(c)}, x_i \in D_u^{(c)} \end{cases} \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

Therefore, the joint distribution adaptation MMD matrix can be computed as  $M = \sum_{c=0}^C M^{(c)}$ , where  $n_l^{(0)} = n_l$ ,  $n_u^{(0)} = n_u$ ,  $D_u^{(0)} = D_u$ ,  $D_l^{(0)} = D_l$ . Integrating the MMD regularization into Eq. (2), we obtain the objective function of TNMF:

$$\mathcal{O} = \min_{U, V} \|X - UV^T\|^2 + \text{tr} (V(\lambda H + \mu M)V^T) \quad (4)$$

where  $\mu$  is the trade-off adaptation regularization parameter weighting between HeNMF and distribution matching. Setting  $\mu$  to 0, the TNMF degenerates to HeNMF. Hence, the geometric preservation and distribution matching is unified into the NMF objective.

Moreover, the adaptation regularization in Eq. (4) is significant to make TNMF robust across different probability distributions. According to [25], MMD will asymptotically approach zero if and only if the two distributions are the same. By minimizing this adaptation regularization, TNMF can match the distributions between labeled and unlabeled data based on the coding vector space.

### 4.3 Optimization

We use gradient descent as optimization algorithm for minimizing the objective function in Eq. (4). For this problem, gradient descent leads to the following update rules:

$$u_{ik} \leftarrow u_{ik} + \eta_{ik} \frac{\partial \mathcal{O}_1}{\partial u_{ik}} \quad (5)$$

$$v_{jk} \leftarrow v_{jk} + \delta_{jk} \frac{\partial \mathcal{O}_1}{\partial v_{jk}} \quad (6)$$

where  $\eta_{ik}$  and  $\delta_{jk}$  referred as step size parameters. As long as  $\eta_{ik}$  and  $\delta_{jk}$  are sufficiently small, the objective would reduce gradually to the desired minimum. We can use some tricks to set size parameters automatically. Following [8], we let  $\eta_{ik} = -u_{ik}/2 (UV^TV)_{ik}$ , hence, we can write Eq. (5) as

$$\begin{aligned}
 & u_{ik} + \eta_{ik} \frac{\partial \mathcal{O}_1}{\partial u_{ik}} \\
 = & u_{ik} - \frac{u_{ik}}{2 (UV^TV)_{ik}} \frac{\partial \mathcal{O}_1}{\partial u_{ik}} \\
 = & u_{ik} - \frac{u_{ik}}{2 (UV^TV)_{ik}} (-2 (XV)_{ik} + 2 (UV^TV)_{ik}) \\
 = & u_{ik} \frac{(XV)_{ik}}{(UV^TV)_{ik}}
 \end{aligned} \tag{7}$$

Similarly, the update formulation of Eq. (6) can also be get easily. It is clear that the updating rules in this problem are special cases of gradient descent with an automatic step parameter selection. The multiplicative updating rules guarantees that  $U$  and  $V$  are non-negative and Eqs. (5) and (6) would ultimately converge to a local minimum.

## 5 Experiments

In this section, we perform image classification experiments on benchmark datasets widely used for visual domain adaptation to evaluate TNMF.

### 5.1 Dataset Description

**USPS**<sup>1</sup> dataset consists 9,298 images of size  $16 \times 16$ . **MNIST**<sup>2</sup> dataset has 70,000 examples of size  $28 \times 28$ . From Fig. 1, we see that USPS and MNIST follow very different distributions. They share 10 semantic classes, each corresponding to one digit. Following [14], we randomly sample 1,800 images in USPS to form the training data and 2,000 images in MNIST as test data to construct a dataset *USPS vs MNIST*. We rescale all images to size  $16 \times 16$  and represent each image by a 256-dimensional feature vector.

**MSRC**<sup>3</sup> dataset contains of 4,323 images and 18 classes, provided by Microsoft Research Cambridge. **VOC2007**<sup>4</sup> dataset contains 5,011 images annotated with 20 concepts, collecting from digital photos on Flickr<sup>5</sup>. The MSRC and VOC2007 datasets used in our experiments share the following 6 semantic classes: aeroplane, bicycle, bird, car, cow, sheep. Following [21], all 1,269

<sup>1</sup> <http://www-i6.informatik.rwth-aachen.de/~keyzers/usps.html>.

<sup>2</sup> <http://yann.lecun.com/exdb/mnist>.

<sup>3</sup> <http://research.microsoft.com/en-us/projects/objectclassrecognition>.

<sup>4</sup> <http://pascallin.ecs.soton.ac.uk/challenges/VOC/voc2007>.

<sup>5</sup> <https://www.flickr.com>.



**Fig. 1.** First line shows examples of USPS, MNIST, MSRC, VOC2007 Second line shows examples of Caltech256, Amazon, Dslr and Webcam

images in MSRC and all 1,530 images in VOC2007 are selected to construct the dataset *MSRC vs VOC*. We uniformly rescale all images to be 256 pixels in size and extract 128-dimensional dense SIFT features with VLFeat [22]. A 240-dimensional codebook is created, where K-means clustering is used to obtain the codewords.

**Caltech256**<sup>6</sup> is a standard database for object recognition consists of 30,607 images and 256 categories. **Office** [24] dataset consists of three domains: Amazon (images from online merchants), Webcam (low-resolution images by a web camera), and DSLR (high-resolution images by a digital SLR camera). It has 4,652 images with 31 classes. In these experiments, we adopt the 800-dimensional SURF features released by Gong et al. [24]. By randomly selecting two different domains as source domain and target domain, we construct 12 cross-domain datasets, *e.g.*, *Caltech vs Amazon*, *Caltech vs Webcam*, *...*, *Dslr vs Webcam*.

## 5.2 Performance on Cross-Domain Datasets

We compare our TNMF approach with three NMF related competing rivals and basic NN classifier for image recognition problems:

- (1) 1-Nearest Neighbor Classifier (NN)
- (2) Nonnegative Matrix Factorization (NMF) + NN
- (3) Laplacian regularized NMF (LapNMF) [8] + NN
- (4) Hessian regularized NMF (HeNMF) + NN

In our experiments, we choose NN as the base classifier since it does not require tuning parameters for cross-validation. NN classifier is trained on the labeled source data for classifying the unlabeled target data. Since labeled and unlabeled data are sampled from different distributions, it is impossible to tune

<sup>6</sup> <https://www.eecs.berkeley.edu/~jhoffman/domainadapt>.



**Table 2.** Accuracy (%) on cross-domain datasets

	NN	NMF	LapNMF	HeNMF	TNMF
USPS vs MNIST	36.50	29.45	31.01	37.45	<b>39.20</b>
MSRC vs VOC	25.95	26.86	26.67	27.06	<b>28.82</b>
Caltech vs Amazon	21.82	29.85	29.96	30.27	<b>31.00</b>
Caltech vs Webcam	13.90	16.27	17.63	18.98	<b>22.03</b>
Caltech vs Dslr	8.92	19.75	20.38	21.66	<b>23.57</b>
Amazon vs Caltech	24.31	26.98	27.43	28.41	<b>29.03</b>
Amazon vs Webcam	<b>24.07</b>	21.69	17.29	21.02	21.69
Amazon vs Dslr	18.47	19.11	21.66	23.57	<b>28.03</b>
Webcam vs Caltech	<b>24.31</b>	20.48	20.93	22.80	22.44
Webcam vs Amazon	27.35	27.66	27.56	28.39	<b>29.12</b>
Webcam vs Dslr	56.05	54.14	52.87	52.87	<b>57.32</b>
Dslr vs Caltech	22.26	25.73	25.20	25.91	<b>26.36</b>
Dslr vs Amazon	21.71	24.74	24.63	25.89	<b>26.30</b>
Dslr vs Webcam	44.07	51.19	54.24	49.15	<b>55.59</b>

the model parameters using cross-validation. Therefore, we evaluate all methods by grid search for the optimal parameter settings which gives the highest classification accuracy on all datasets, and report the best results of each method. We set the trade-off parameter of Hessian regularization  $\lambda$  by searching  $\lambda \in \{10^{-2}, 10^{-1}, 1, 10, 10^2\}$  and adaptation regularization parameter  $\mu$  by searching  $\mu \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3, 10^4\}$ . Our proposed TNMF is run 5 times independently and we report their average. The classification accuracy on test data is used as the evaluation metric, which reads

$$Accuracy = \frac{|x : x \in \mathcal{D}_t \wedge f(x) = y(x)|}{|x : x \in \mathcal{D}_t|} \quad (8)$$

where  $\mathcal{D}_t$  is the set of test data,  $y(x)$  is the ground truth of  $x$ ,  $f(x)$  is the label predicted by the classification algorithm.

The classification performance of TNMF as well as baselines on all datasets are listed in Table 2. When comparing with the baseline methods, we set the Hessian regularization trade-off parameter  $\lambda$  as 0.1, and the MMD regularization trade-off parameter  $\mu$  as 1. The number of iterations is set to 300. From the results, we can find that on all datasets TNMF gaining an improvement of 6.55 % compared to HeNMF and significantly improves 19.36 % than NN, which illustrates the significant role of adaptation regularization. Moreover, the classification accuracy of HeNMF improves 4.02 % than LapNMF which reflects that Hessian can exploit more intrinsic geometry than Laplacian.

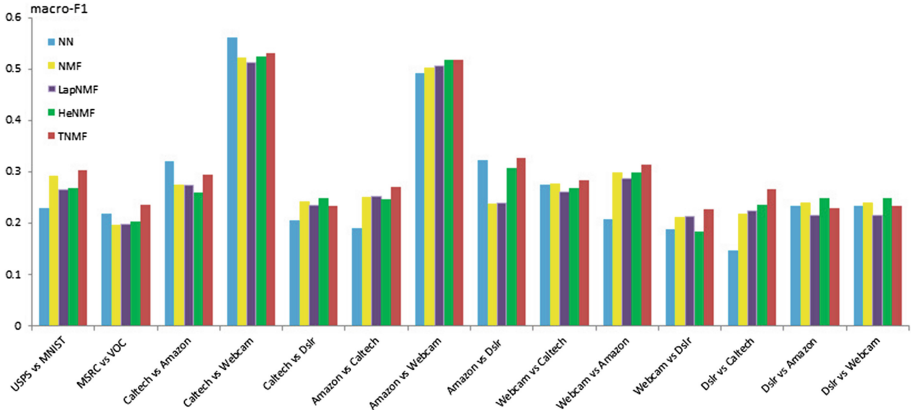


Fig. 2. Macro-F1 measure on cross-domain datasets

Moreover, we choose F1 measure to evaluate the performance of our proposed method as well as baselines. Due to the fact that the datasets have more than 2 classes, macro-averaged evaluation is considered, which is given as follows:

$$Macro - F1 = \frac{2 * Macro - Precision * Macro - Recall}{Macro - Precision + Macro - Recall} \quad (9)$$

where Macro-Precision and Macro-Recall are obtained respectively as:

$$Macro - Recall = \frac{\sum_{i=1}^c R_i}{c} \quad (10)$$

and

$$Macro - Precision = \frac{\sum_{i=1}^c P_i}{c} \quad (11)$$

where  $c$  is the number of classes,  $R_i$  and  $P_i$  are recall and precision for the  $i$ -th class. The results of Macro-F1 score of each methods on all the cross-domain datasets are shown in Fig. 2.

From Fig. 2, we can observe that our method perform better than all the baseline methods as a whole, while a little weak on several datasets. On average, the Macro-F1 measure of TNMF is 1.49% higher than HeNMF and 3.17% than the basic NN classifier.

## 6 Conclusion

In this paper, we extend NMF to cross-domain scenario and propose a novel Transfer Nonnegative Matrix Factorization (TNMF) approach for low-dimensional image representation. An important advantage of TNMF is that it can explore as many necessary learning objectives as possible in cross-domain image representation task, yet still remain simple to implement practically.

Experimental results on several benchmark datasets show that TNMF outperforms state-of-the-art NMF methods.

**Acknowledgements.** The research was supported by the Irish Research Council (IRCSET) under Grant Number GOIPG/2013/330.

## References

1. Luong, T.X., kyeong Kim, B., Lee, S.-Y.: Color image processing based on nonnegative matrix factorization with convolutional neural network. In: 2014 International Joint Conference on Neural Networks (IJCNN), pp. 2130–2135 (2014)
2. Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. *Nature* **401**, 788 (1999)
3. S. Z., Li, X., Hou, Zhang, H., Cheng, Q.: Learning spatially localized, parts-based representation. *CVPR. IEEE Comput. Soc.* **1**, 207–212 (2001)
4. Kuang, D., Park, H., Ding, C.H.Q.: Symmetric nonnegative matrix factorization for graph clustering. *SDM* **12**, 106–117 (2012)
5. Pan, S.J., Yang, Q.: A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **22**(10), 1345–1359 (2010)
6. Fox, D., Gomes, C.P. (eds.): In: Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence, AAAI 2008. AAAI Press, Chicago, 13–17 July 2008
7. Pan, S.J., Tsang, I.W., Kwok, J.T., Yang, Q.: Domain adaptation via transfer component analysis. *IEEE Trans. Neural Netw.* **22**(2), 199–210 (2011)
8. Cai, D., He, X., Han, J., Huang, T.S.: Graph regularized non-negative matrix factorization for data representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(8), 1548–1560 (2011)
9. David, L.D., Carrie, G.: Hessian eigenmaps: new locally linear embedding techniques for high-dimensional data (2003)
10. Bengio, Y., Schuurmans, D., Lafferty, J.D., Williams, C.K.I., Culotta, A. (eds.): Advances in Neural Information Processing Systems 22. In: 23rd Annual Conference on Neural Information Processing Systems 2009. Proceedings of a meeting held 7–10 Dec 2009. Vancouver, British. Curran Associates Inc (2009)
11. Liu, W., Tao, D.: Multiview hessian regularization for image annotation. *IEEE Trans. Image Process.* **22**(7), 2676–2687 (2013)
12. Tao, D., Jin, L., Liu, W., Li, X.: Hessian regularized support vector machines for mobile image annotation on the cloud. *IEEE Trans. Multimedia* **15**(4), 833–844 (2013)
13. Liu, W., Liu, H., Tao, D., Wang, Y., Lu, K.: Multiview Hessian regularized logistic regression for action recognition. *Signal Process.* **110**, 101–107 (2015)
14. Long, M., Wang, J., Ding, G., Pan, S.J., Yu, P.S.: Adaptation regularization: a general framework for transfer learning. *IEEE Trans. Knowl. Data Eng.* **26**(5), 1076–1089 (2014)
15. Ding, C.H.Q., Li, T., Jordan, M.I.: Convex and semi-nonnegative matrix factorizations. *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(1), 45–55 (2010)
16. Ding, C., Li, T., Peng, W.: On the equivalence between non-negative matrix factorization and probabilistic latent semantic indexing. *Comput. Stat. Data Anal.* **52**(8), 3913–3927 (2008)

17. Heiler, M., Schnörr, C.: Learning sparse representations by non-negative matrix factorization and sequential cone programming. *J. Mach. Learn. Res.* **7**, 1385–1407 (2006)
18. Liu, H., Wu, Z., Cai, D., Huang, T.S.: Constrained non-negative matrix factorization for image representation. *IEEE Trans. Pattern Anal. Mach. Intell.* (2012). To appear
19. Quanz, B., Huan, J., Mishra, M.: Knowledge transfer with low-quality data: a feature extraction issue. *IEEE Trans. Knowl. Data Eng.* **24**(10), 1789–1802 (2012)
20. Wang, J.J., Sun, Y., Bensmail, H.: Domain transfer nonnegative matrix factorization. In: 2014 International Joint Conference on Neural Networks, IJCNN, pp. 3605–3612. Beijing, 6–11 July 2014
21. Wang, C., Blei, D., Li, F.-F.: Simultaneous image classification and annotation (2009)
22. Vedaldi, A., Fulkerson, B.: VLFeat: an open and portable library of computer vision algorithms. In: Proceedings of the International Conference on Multimedia, MM 2010, pp. 1469–1472. ACM, New York (2010)
23. Long, M., Wang, J., Ding, G., Sun, J., Yu, P.S.: Transfer feature learning with joint distribution adaptation. In: 2013 IEEE International Conference on Computer Vision (ICCV), pp. 2200–2207. IEEE (2013)
24. Gong, B., Shi, Y., Sha, F., Grauman, K.: Geodesic flow kernel for unsupervised domain adaptation. In: CVPR, pp. 2066–2073. IEEE (2012)
25. Gretton, A., Borgwardt, K.M., Rasch, M.J., Schölkopf, B., Smola, A.: A kernel two-sample test. *J. Mach. Learn. Res.* **13**, 723–773 (2012)