

# Adaptive Multi-View Semi-Supervised Nonnegative Matrix Factorization

Jing Wang<sup>1</sup>, Xiao Wang<sup>2</sup>, Feng Tian<sup>1\*</sup>, Chang Hong Liu<sup>3</sup>,  
Hongchuan Yu<sup>4</sup>, and Yanbei Liu<sup>5</sup>

<sup>1</sup> Faculty of Science and Technology, Bournemouth University, UK

<sup>2</sup> Department of Computer Science and Technology, Tsinghua University, China

<sup>3</sup> Department of Psychology, Bournemouth University, UK

<sup>4</sup> National Centre for Computer Animation, Bournemouth University, UK

<sup>5</sup> School of Electronic Information Engineering, Tianjin University, China

jwang@bournemouth.ac.uk, wangxiao\_cv@tju.edu.cn,

{ftian, liuc, hyu}@bournemouth.ac.uk, liuyanbei@tju.edu.cn

**Abstract.** Multi-view clustering, which explores complementary information between multiple distinct feature sets, has received considerable attention. For accurate clustering, all data with the same label should be clustered together regardless of their multiple views. However, this is not guaranteed in existing approaches. To address this issue, we propose Adaptive Multi-View Semi-Supervised Nonnegative Matrix Factorization (AMVNMF), which uses label information as hard constraints to ensure data with same label are clustered together, so that the discriminating power of new representations are enhanced. Besides, AMVNMF provides a viable solution to learn the weight of each view adaptively with only a single parameter. Using  $L_{2,1}$ -norm, AMVNMF is also robust to noises and outliers. We further develop an efficient iterative algorithm for solving the optimization problem. Experiments carried out on five well-known datasets have demonstrated the effectiveness of AMVNMF in comparison to other existing state-of-the-art approaches in terms of accuracy and normalized mutual information.

**Keywords:** Nonnegative Matrix Factorization, Multi-view learning, Semi-supervised learning

## 1 Introduction

Real data are often comprised of multiple views (features) [3]. For example, color and texture information can be utilized as different views of images and videos; a document may be translated into multiple languages, and a web page may be represented by multiple contents and hyperlinks. In these examples, each view describes a specific perspective of the data. Therefore, it becomes natural to integrate multiple views and discover the underlying data structures.

---

\* Corresponding author.

Multi-view clustering has attracted increasing interests and been explored in several studies. For example, Kumar et al. [7] proposed two objectives to regularize the Laplacian embeddings between different views to be similar and spectral analysis is employed for parameter learning. Tzortzis et al. [14] learned a unified kernel through a weighted combination of kernels of all the views. MultiNMF [12] was proposed to obtain a common consensus matrix, which was designed to reflect the latent clustering structure shared by different views. Subsequently, Wang et al. [15] proposed a regression-like objective, which conducts multi-view clustering and feature selection at the same time. Zhang et al [16] also developed MMNMF which attempted to preserve intrinsic geometrical structure of data across multiple views. Cao et al. [5] utilized a diversity constraints on subspaces to enhance the complementarity among multiple views.

Various existing methods indeed improve the clustering performance for multi-view data, nevertheless, some challenges remain. Firstly, in reality, supervised information, e.g., the labels of data or the pairwise information (must-link and cannot-link constraints) between data, are often available. They have been integrated into the single-view learning and demonstrated the effectiveness. While for multi-view learning, we notice that the supervised information usually has consistency across multiple views. If we can guarantee data with same label but come from various views are still grouped into the same cluster, this will improve the clustering accuracy [1], [17]. Therefore, how to utilize this discriminative information for guiding the multi-view learning is of great value. Secondly, when we cluster data across multiple views, each view may have different contributions. We can consider each view has the same weight in a straightforward way, but this oversimplified assumption may be not always satisfied in the real-world application. Taking the face clustering as an example, a frontal or a three-quarter view is a better representation for faces than a profile view [2], [9]. So, the weight of each view should be determined automatically rather than being treated equally. Finally, outliers or noisy data are ubiquitous, and thus, a robust multi-view learning approach is required for practical applications.

To address these challenges altogether, we propose a new multi-view clustering approach based on non-negative matrix factorization (NMF) [8], called Adaptive Multi-View Semi-Supervised Nonnegative Matrix Factorization (AMVNMF). The overall advantages of this approach are as follows:

1. By taking the label information as hard constraints, AMVNMF guarantees that data sharing the same label will have the same new representation and be mapped into the same class in the low-dimensional space regardless whether they come from the same view.
2. To our best knowledge, this is the first attempt to introduce a single parameter to control the distribution of weighting factors for NMF-based multi-view clustering. Consequently, the weight factor of each view can be assigned automatically depending on the dissimilarity between each new representation matrix and the consensus matrix.
3. Using the structured sparsity-inducing,  $L_{2,1}$ -norm, AMVNMF is robust against noises and hence can achieve more stable clustering results.

## 2 A Brief Review of NMF and CNMF

Given  $N$  data  $\mathbf{X} = [x_1, x_2, \dots, x_N] \in \mathbb{R}^{P \times N}$ , each data  $x_i$  is represented by  $P$ -dimensional feature vector. NMF [8] aims to find two nonnegative matrix factors  $\mathbf{W} \in \mathbb{R}^{P \times K}$  and  $\mathbf{H} \in \mathbb{R}^{N \times K}$  where the product of the two factors can well approximate the original matrix, represented as  $\mathbf{X} \approx \mathbf{W}\mathbf{H}^T$ . In particular,  $\mathbf{H}$  can be considered as the new representations of data in terms of the basis  $\mathbf{W}$ . NMF measures the dissimilarity between  $\mathbf{X}$  and  $\mathbf{W}\mathbf{H}^T$  by  $F$ -norm, which is defined as

$$\|\mathbf{X} - \mathbf{W}\mathbf{H}^T\|_F^2, \quad (1)$$

To extend the traditional unsupervised NMF to a semi-supervised learning approach, CNMF [10] builds a label constraint matrix which incorporates the label information as hard constraints so that the data sharing the same label have the same new representation. In particular, assuming the first  $l$  data points are labeled with  $c$  classes, then an indicator matrix  $\mathbf{C}$  can be constructed, where  $c_{i,j} = 1$  if  $v_i$  is labeled with  $j$ th class; or  $c_{i,j} = 0$  otherwise. Then, the label constraint matrix  $\mathbf{A}$  can be defined as follows,

$$\mathbf{A} = \begin{pmatrix} \mathbf{C}_{l \times c} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{N-l} \end{pmatrix}, \quad (2)$$

where  $\mathbf{I}_{N-l}$  is a  $(N-l) \times (N-l)$  identity matrix. Recall that NMF maps each data point  $x_i$  to its new representation  $h^i$  from  $P$ -dimensional space to  $K$ -dimensional space, where  $h^i$  represents the  $i$ th row of  $\mathbf{H}$ . To incorporate label information, we introduce an auxiliary matrix  $\mathbf{Z}$  with  $\mathbf{H} = \mathbf{A}\mathbf{Z}$ . As we can see from  $\mathbf{A}$ , if  $x_i$  and  $x_j$  have the same label, then the  $i$ th row and  $j$ th row of  $\mathbf{A}$  must be the same, and so  $h^i = h^j$ , which guarantees that data sharing the same label have the same new representation. Therefore, the objective function can be written as follows,

$$\min_{\mathbf{W} \geq 0, \mathbf{Z} \geq 0} \|\mathbf{X} - \mathbf{W}\mathbf{Z}^T \mathbf{A}^T\|_F^2. \quad (3)$$

## 3 Adaptive Multi-View Semi-Supervised Nonnegative Matrix Factorization (AMVNMF)

### 3.1 AMVNMF model

Let  $\mathbf{X}^{(v)} \in \mathbb{R}^{P_v \times N}$  denote the features in  $v$ th view,  $\mathbf{W}^{(v)} \in \mathbb{R}^{P_v \times K}$  and  $\mathbf{Z}^{(v)} \in \mathbb{R}^{K \times N}$  be the basis and auxiliary matrix in  $v$ th view, respectively. Since the matrix  $\mathbf{A}$  above is constructed only based on the label information and consistent for different features, which means different features share the same constraint matrix  $\mathbf{A}$ . Thus, given  $M$  types of heterogeneous features,  $v = 1, 2, \dots, M$ , we naturally integrate all these view together and propose the objective function as follows,

$$\min_{\mathbf{W} \geq 0, \mathbf{Z} \geq 0} \sum_{v=1}^M \|\mathbf{X}^{(v)} - \mathbf{W}^{(v)}(\mathbf{Z}^{(v)})^T \mathbf{A}^T\|_F^2. \quad (4)$$

Assuming that, new representation matrices of  $M$  views are regularized towards a common consensus matrix  $\mathbf{H}^*$ , we aim to obtain  $\mathbf{H}^*$ , which uncovers the common latent structure shared by multiple views. With the constraint matrix  $\mathbf{A}$  and a consensus auxiliary matrix  $\mathbf{Z}^*$ , we have  $\mathbf{H}^* = \mathbf{A}\mathbf{Z}^*$ . Since  $\mathbf{A}$  is known, we turn the problem of finding  $\mathbf{H}^*$  into the problem of finding  $\mathbf{Z}^*$ . The objective function can be rewritten as follows,

$$\min_{\mathbf{W}^{(v)}, \mathbf{Z}^{(v)}, \mathbf{Z}^* \geq 0} \sum_{v=1}^M \|\mathbf{X}^{(v)} - \mathbf{W}^{(v)}(\mathbf{Z}^{(v)})^T \mathbf{A}^T\|_F^2 + \sum_{v=1}^M \lambda_v \|\mathbf{Z}^{(v)} - \mathbf{Z}^*\|_F^2, \quad (5)$$

where  $\lambda_v$  is the weight factor for  $v$ th view.

Note that different views may not be comparable at the same scale, hence, without loss of generality, we assume  $\|\mathbf{X}^{(v)}\|_1 = 1$ . Also, in order to make different  $\mathbf{Z}^{(v)}$  comparable and meaningful, we need to constrain  $\|\mathbf{W}\|_1 = 1$ . To do so, we introduce

$$\mathbf{Q}^{(v)} = \text{Diag}\left(\sum_{i=1}^M \mathbf{W}_{i,1}^{(v)}, \sum_{i=2}^M \mathbf{W}_{i,2}^{(v)}, \dots, \sum_{i=1}^M \mathbf{W}_{i,K}^{(v)}\right) \quad (6)$$

to normalize  $\mathbf{W}$  by using  $\mathbf{W} = \mathbf{W}\mathbf{Q}^{-1}$ . In this way, we can approximately constrain  $\|(\mathbf{Z}^{(v)})^T \mathbf{A}^T\|_1 = 1$  so that  $\mathbf{Z}^{(v)}$  is within the same range [12]. Due to  $\mathbf{W}^{(v)}\mathbf{Z}^{(v)T} \mathbf{A}^T = \mathbf{W}^{(v)}\mathbf{Q}^{-1}(\mathbf{Z}^{(v)}\mathbf{Q})^T \mathbf{A}^T$ , (5) could then be written as

$$\min_{\mathbf{W}^{(v)}, \mathbf{Z}^{(v)}, \mathbf{Z}^* \geq 0} \sum_{v=1}^M \|\mathbf{X}^{(v)} - \mathbf{W}^{(v)}(\mathbf{Z}^{(v)})^T \mathbf{A}^T\|_F^2 + \sum_{v=1}^M \lambda_v \|\mathbf{Z}^{(v)}\mathbf{Q}^{(v)} - \mathbf{Z}^*\|_F^2. \quad (7)$$

Normally, for all  $M$  views, each parameter  $\lambda_v$  need to be specified manually which reflects each view's importance. Apparently, it is hard to decide which view contributes more or less with no prior knowledge. To address this limitation, we use a single parameter  $\gamma$  to control the distribution of weight factors  $\alpha^{(v)}$  in all  $M$  views, such that the important views are assigned bigger weights. Then we have

$$\begin{aligned} J = \min_{\mathbf{W}^{(v)}, \mathbf{Z}^{(v)}, \mathbf{Z}^*, \alpha^{(v)} \geq 0} & \sum_{v=1}^M \|\mathbf{X}^{(v)} - \mathbf{W}^{(v)}(\mathbf{Z}^{(v)})^T \mathbf{A}^T\|_F^2 \\ & + \sum_{v=1}^M (\alpha^{(v)})^\gamma \|\mathbf{Z}^{(v)}\mathbf{Q}^{(v)} - \mathbf{Z}^*\|_F^2 \\ & \text{s.t. } \sum_{v=1}^M \alpha^{(v)} = 1. \end{aligned} \quad (8)$$

Note that the first term in (8) is the least square loss function, which is very sensitive to outliers in real word data, as the error for each data is squared and can easily dominate the objective function.. Instead, we propose a more robust

formulation as the following:

$$\begin{aligned}
J = \min_{\mathbf{W}^{(v)}, \mathbf{Z}^{(v)}, \mathbf{Z}^*, \alpha^{(v)} \geq 0} & \sum_{v=1}^M \|\mathbf{X}^{(v)} - \mathbf{W}^{(v)}(\mathbf{Z}^{(v)})^T \mathbf{A}^T\|_{2,1} \\
& + \sum_{v=1}^M (\alpha^{(v)})^\gamma \|\mathbf{Z}^{(v)} \mathbf{Q}^{(v)} - \mathbf{Z}^*\|_F^2 \\
& s.t. \sum_{v=1}^M \alpha^{(v)} = 1.
\end{aligned} \tag{9}$$

In this objective function, the error for each data is not squared squared and thus the impact of large errors is weakened greatly. Correspondingly, the effects of data outliers are reduced and the robustness of NMF is improved.

### 3.2 Algorithm of AMVNMF model

To solve the optimization problem (9), we propose an iterative update procedure. When  $\mathbf{Z}^*$  is fixed, for each given  $v$ , the computation of  $\mathbf{W}^{(v)}$  or  $\mathbf{Z}^{(v)}$  does not depend on  $\mathbf{W}^{(v')}$  or  $\mathbf{Z}^{(v')}$ , where  $v \neq v'$ . Therefore, we use  $\mathbf{X}$ ,  $\mathbf{W}$ ,  $\mathbf{Z}$ , and  $\mathbf{Q}$  to represent  $\mathbf{X}^{(v)}$ ,  $\mathbf{W}^{(v)}$ ,  $\mathbf{Z}^{(v)}$  and  $\mathbf{Q}^{(v)}$  for brevity. The objective function can be simplified as

$$J = \min_{\mathbf{W}, \mathbf{Z}, \mathbf{Z}^*, \alpha^{(v)} \geq 0} \|\mathbf{X} - \mathbf{WZ}^T \mathbf{A}^T\|_{2,1} + (\alpha^{(v)})^\gamma \|\mathbf{ZQ} - \mathbf{Z}^*\|_F^2. \tag{10}$$

The following multiplicative updating rules for  $\mathbf{W}$ ,  $\mathbf{Z}$  and  $\mathbf{D}$  are applied to update their values sequentially and iteratively.

(1) Fixing  $\mathbf{Z}^*$ ,  $\mathbf{Z}$ ,  $\mathbf{D}$  and  $\alpha^{(v)}$ , update  $\mathbf{W}$

Let  $\Phi_{i,k}$  be the Lagrange multiplier matrix for the constraint  $\mathbf{W}_{i,k} \geq 0$ , and  $\Phi = [\Phi_{i,k}]$ . The Lagrange function is  $L_1 = J + Tr(\Phi \mathbf{W})$ , we only care the terms that are relevant to  $\mathbf{W}^{(v)}$ .

$$\begin{aligned}
L_1 = & Tr(-2\mathbf{XD}\mathbf{A}\mathbf{Z}\mathbf{W}^T + \mathbf{W}\mathbf{Z}^T \mathbf{A}^T \mathbf{D}\mathbf{A}\mathbf{Z}\mathbf{W}^T) \\
& + (\alpha^{(v)})^\gamma Tr(\mathbf{Z}\mathbf{Q}\mathbf{Q}^T \mathbf{Z}^T - 2\mathbf{Z}\mathbf{Q}(\mathbf{Z}^*)^T) + Tr(\Phi \mathbf{W}).
\end{aligned} \tag{11}$$

Taking the derivatives of  $L_1$  with respect to  $\mathbf{W}$  and using the Karush-Kuhn-Tucker condition  $\Phi_{i,k} \mathbf{W}_{i,k} = 0$ , we get the update rule as follows,

$$\mathbf{W}_{i,k} = \mathbf{W}_{i,k} \frac{(\mathbf{XD}\mathbf{A}\mathbf{Z})_{i,k} + (\alpha^{(v)})^\gamma \sum_{j=1}^{N-l+c} \mathbf{Z}_{j,k} \mathbf{Z}_{j,k}^*}{(\mathbf{W}\mathbf{Z}^T \mathbf{A}^T \mathbf{D}\mathbf{A}\mathbf{Z})_{i,k} + (\alpha^{(v)})^\gamma \sum_{f=1}^{d_v} \mathbf{W}_{f,k} \sum_{j=1}^{N-l+c} \mathbf{Z}_{j,k}^2}. \tag{12}$$

(2) Fixing  $\mathbf{Z}^*$ ,  $\mathbf{W}$ ,  $\mathbf{D}$  and  $\alpha^{(v)}$ , update  $\mathbf{Z}$

For each view, we first normalize the column vectors of  $\mathbf{W}$  using  $\mathbf{Q}$  as in (6), then

$$\mathbf{W} \leftarrow \mathbf{W}\mathbf{Q}^{-1}, \mathbf{Z} \leftarrow \mathbf{Z}\mathbf{Q}. \tag{13}$$

Thus, the object function equals

$$\min_{\mathbf{W}, \mathbf{Z}, \mathbf{Z}^*, \alpha^{(v)} \geq 0} \|\mathbf{X} - \mathbf{WZ}^T \mathbf{A}^T\|_{2,1} + (\alpha^{(v)})^\gamma \|\mathbf{Z} - \mathbf{Z}^*\|_F^2. \quad (14)$$

Let  $\Psi$  be the Lagrange multiplier matrix for the constraint  $\mathbf{Z} \geq 0$ , and  $\Psi = [\Psi_{j,k}]$ . Similarly,

$$\begin{aligned} L_2 = & Tr(-2\mathbf{XDAZ}\mathbf{W}^T + 2\mathbf{WZ}^T \mathbf{A}^T \mathbf{DAZ}\mathbf{W}^T) \\ & + (\alpha^{(v)})^\gamma Tr(2\mathbf{ZZ}^T - 2\mathbf{Z}(\mathbf{Z}^*)^T) + Tr(\Psi\mathbf{Z}). \end{aligned} \quad (15)$$

Taking derivative of  $L_2$  with respect to  $\mathbf{Z}$  and using the Kuhn-Tucker condition  $\Psi_{j,k} \mathbf{Z}_{j,k} = 0$ , we have

$$\mathbf{Z}_{j,k} = \mathbf{Z}_{j,k} \cdot \frac{(\mathbf{A}^T \mathbf{D} \mathbf{X}^T \mathbf{W})_{j,k} + (\alpha^{(v)})^\gamma \mathbf{Z}_{j,k}^*}{(\mathbf{A}^T \mathbf{D} \mathbf{A} \mathbf{Z} \mathbf{W}^T \mathbf{W})_{j,k} + (\alpha^{(v)})^\gamma \mathbf{Z}_{j,k}}. \quad (16)$$

(3) Fixing  $\mathbf{Z}^*$ ,  $\mathbf{W}$ ,  $\mathbf{Z}$  and  $\alpha^{(v)}$ , update  $\mathbf{D}$

$\mathbf{D} \in \mathbb{R}^{N \times N}$  is the diagonal matrix with the diagonal elements given by

$$D_{ii} = \frac{1}{\|\mathbf{X}_i - \mathbf{W}(\mathbf{Z}^T \mathbf{A}^T)_i\|}. \quad (17)$$

(4) Fixing  $\mathbf{W}$ ,  $\mathbf{Z}$ ,  $\mathbf{D}$  and  $\alpha^{(v)}$ , update  $\mathbf{Z}^*$

Taking the derivative of the objective function  $J$  in (8), we get

$$\mathbf{Z}^* = \frac{\sum_{v=1}^M (\alpha^{(v)})^\gamma \mathbf{Z}^{(v)} \mathbf{Q}^{(v)}}{\sum_{v=1}^M (\alpha^{(v)})^\gamma}. \quad (18)$$

(5) Fixing  $\mathbf{Z}^*$ ,  $\mathbf{W}$ ,  $\mathbf{Z}$  and  $\mathbf{D}$  update  $\alpha^{(v)}$

We only consider the term that relevant to  $\alpha$ , thus, it is equal to minimize

$$J = \sum_{v=1}^M (\alpha^{(v)})^\gamma \|\mathbf{Z}^{(v)} \mathbf{Q}^{(v)} - \mathbf{Z}^*\|_F^2. \quad (19)$$

By setting  $\mathbf{G}^{(v)} = \|\mathbf{Z}^{(v)} \mathbf{Q}^{(v)} - \mathbf{Z}^*\|_F^2$ , and due to  $\sum_{v=1}^M \alpha^{(v)} = 1$ , Then, the Lagrange function of (19) is

$$J = \sum_{v=1}^M (\alpha^{(v)})^\gamma \mathbf{G}^{(v)} - \lambda \sum_{v=1}^M (\alpha^{(v)} - 1). \quad (20)$$

The solution can be obtained as

$$\alpha^{(v)} = \frac{(\gamma \mathbf{G}^{(v)})^{\frac{1}{1-\gamma}}}{\sum_{v=1}^M (\gamma \mathbf{G}^{(v)})^{\frac{1}{1-\gamma}}}. \quad (21)$$

## 4 Experiments

We use five benchmark datasets, SensIT<sup>1</sup>, ORL[6], Reuters<sup>2</sup>, CiteSeer<sup>3</sup> and Cora<sup>4</sup> to assess the performance of AMVNMF. SensIT uses two sensors to classify three types of vehicle. We randomly sample 100 data for each class, and then conduct experiments on 2 views and 3 classes. ORL contains 10 different images of each of 40 people. The images are represented by two views, raw pixel values and GIST [13]. Reuters contains 1200 documents over the 6 labels. Each sample is translated into five languages. We experiment on the English, French and German views same as [12]. Citeseer and Cora are composed of publications. These publications are linked via citations. Both of them take contents and citations as two views.

We compare our AMVNMF with several representative multi-view clustering methods and their modifications.

1. Best Single View (BSV): Using the most informative view which achieves the best performance with our AMVNMF.
2. ConCNMF: The method firstly concatenates the features of all views and applies CNMF [10] to extract the low dimensional subspace representation.
3. MultiNMF: The NMF-based multi-view clustering method proposed in [12].
4. RMKMC : The multi-view  $k$ -means proposed in [4].
5. CoRegSPC : The co-regularized pairwise multi-view spectral clustering method proposed in [7].

Since clustering performances depend on the initializations, we repeat each method 10 times with random initializations and report the average performance. The performances are measured with two widely used evaluation metrics, accu-

Table 1: Clustering results on five real-world datasets (%)

Metrics	Datasets	BSV	ConCNMF	MultiNMF	RMKMC	CoRegSPC	AMVNMF
AC	SensIT	69.66	52.30	55.04	60.07	61.67	<b>71.33</b>
	ORL	74.3	49.59	54.6	45.5	78.20	<b>80.5</b>
	Reuters	57.50	49.59	51.87	39.80	54.40	<b>59.88</b>
	Citeseer	50.08	40.70	34.36	43.21	47.42	<b>53.14</b>
	Cora	33.42	32.42	44.83	43.90	37.20	<b>48.71</b>
NMI	SensIT	30.14	15.67	19.87	14.84	17.75	<b>31.73</b>
	ORL	89.29	51.32	75.23	65.34	90.84	<b>91.73</b>
	Reuters	41.95	30.37	36.14	21.82	36.57	<b>42.75</b>
	Citeseer	21.38	13.34	20.97	20.61	21.10	<b>26.13</b>
	Cora	26.73	9.87	27.95	21.27	15.44	<b>34.59</b>

racy (AC) and normalized mutual information (NMI) [10]. For all the metrics, the higher value indicates better clustering quality. To compare the performance

<sup>1</sup> <https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/multiclass.html>

<sup>2</sup> <http://multilingreuters.iit.nrc.ca>

<sup>3</sup> <http://linqs.umi.acs.umd.edu/projects//projects/lbc/index.html>

<sup>4</sup> <http://linqs.umi.acs.umd.edu/projects//projects/lbc/index.html>

of semi-supervised approaches, same as [11], 30% of labeled data are randomly picked up as priors.

Table 1 summarizes the clustering performances of different algorithms on the five datasets. It is clear to see that AMVNMF outperforms the second best algorithm in all cases. Furthermore, BSV always gets the second best performance. It outperforms other multi-view methods greatly, i.e., 7.99%/10.27% on SensIT and 3.10%/5.38% on Reuters in terms of AC and NMI, respectively. This is mainly due to AMVNMF guarantees that all the data sharing the same labels are grouped together, regardless they are come from the same or different views. Therefore, both AMVNMF and BSV ( running AMVNMF with single view) produce superior results.

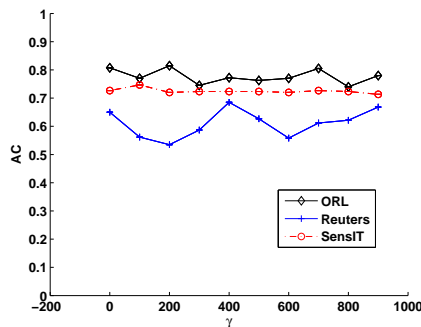


Fig. 1: Performance of AMVNMF w.r.t. parameter  $\gamma$ .

We show the parameter tuning and algorithm convergence on SensIT, ORL and Reuters as examples in Figure 1 and Figure 2, respectively. The parameter  $\gamma$  controls the distribution of weight factors  $\alpha^{(v)}$  for different views. More precisely, when  $\gamma \rightarrow \infty$ , the weight for all views is equal. When  $\gamma \rightarrow 1$ , the weight factor of 1 is assigned to the most important view whose  $\mathbf{G}^{(v)}$  value is the smallest and 0 is assigned to the weights of the other views. Hence, this strategy allows well adjusting the ratio of each view and saves the cost of tuning multiple parameters. As shown in Figure 1, AMVNMF performs stably with varying  $\gamma$  (from 2 to 902). Please note that even the worst results of AMVNMF are always better than other approaches in most cases. The corresponding convergence curves together with performances are shown in Figure 2. The blue solid line shows the value of the objective function and the red dashed line indicates the accuracy. It can be seen that the value of the objective function decreases steadily with more iterations and converges after around 20 times.

Since AMVNMF is a semi-supervised method, we also randomly pick up 10% and 20% labeled data to further demonstrate the benefits of priors. Notice that ORL has only 10 images for each category, thus 10% gives one image only. However, one label is meaningless for AMVNMF since this algorithm maps the



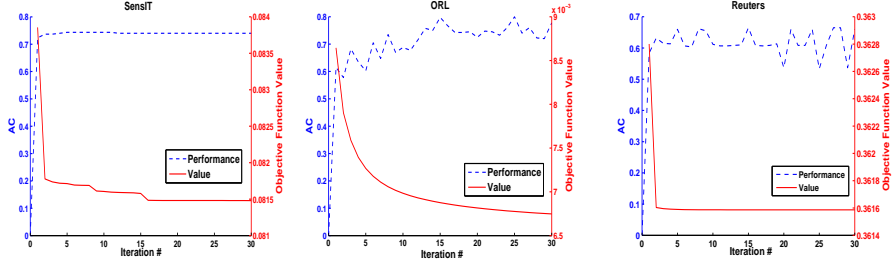


Fig. 2: Convergence and corresponding performance curve.

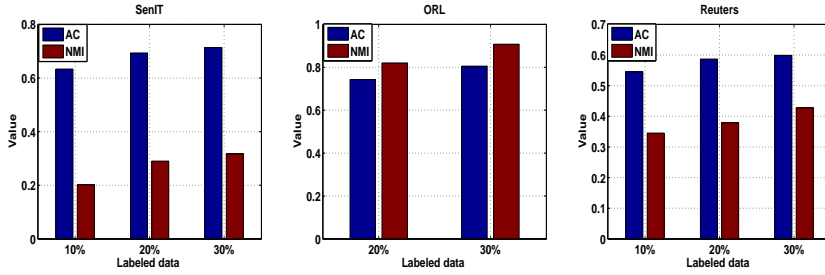


Fig. 3: Performance of AMVNMF w.r.t. labeled data.

images with the same label onto the same coordinate in the new representation space. Thus, we omit the result with 10% labeled data. From Figure 3, it can be seen that both AC and NMI are improved with more labeled data. Also, it is worth pointing out that even with only 10% labeled data, AMVNMF performs better than other approaches when 30% labeled data are applied. For example, for the SensIT dataset, AMVNMF achieves 62% AC and 20% NMI with 10% labeled data, which is better than the best performance of other approaches, i.e., 61.67% AC and 19.87% NMI (as shown in Table 1).

## 5 Conclusion

A novel NMF-based multi-view method, AMVNMF, is proposed in this paper. It efficiently learn the underlying clustering structure embedded in multiple views, by regularizing the new representation matrices learnt from different views towards a common consensus. The advantages of AMVNMF are shown in three aspects. First, it guarantees that labeled data come with multiple views can be clustered into the same low-dimension space. Second, it learns each view's corresponding weight adaptively with a single parameter  $\gamma$ . Third, it handles the noises more effectively. For future work, a sparse regulation may be introduced into AMVNMF to obtain more accurate new representation matrix, with which the clustering performance is expected to be further improved.

## References

1. Belkin, M., Niyogi, P., Sindhwani, V.: Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *The Journal of Machine Learning Research* 7, 2399–2434 (2006)
2. Blanz, V., Tarr, M.J., Bülthoff, H.H., Vetter, T.: What object attributes determine canonical views? *Perception-London* 28(5), 575–600 (1999)
3. Blum, A., Mitchell, T.: Combining labeled and unlabeled data with co-training. In: *Proceedings of the eleventh annual conference on Computational learning theory*. pp. 92–100. ACM (1998)
4. Cai, X., Nie, F., Huang, H.: Multi-view k-means clustering on big data. In: *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*. pp. 2598–2604. AAAI Press (2013)
5. Cao, X., Zhang, C., Fu, H., Liu, S., Zhang, H.: Diversity-induced multi-view subspace clustering. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 586–594 (2015)
6. Hidru, D., Goldenberg, A.: Equinmf: Graph regularized multiview nonnegative matrix factorization. arXiv preprint arXiv:1409.4018 (2014)
7. Kumar, A., Rai, P., Daume, H.: Co-regularized multi-view spectral clustering. In: *Advances in Neural Information Processing Systems*. pp. 1413–1421 (2011)
8. Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. *Nature* 401(6755), 788–791 (1999)
9. Liu, C.H., Chaudhuri, A.: Reassessing the 3/4 view effect in face recognition. *Cognition* 83(1), 31–48 (2002)
10. Liu, H., Wu, Z., Li, X., Cai, D., Huang, T.S.: Constrained nonnegative matrix factorization for image representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 34(7), 1299–1311 (2012)
11. Liu, H., Yang, G., Wu, Z., Cai, D.: Constrained concept factorization for image representation. *Cybernetics, IEEE Transactions on* 44(7), 1214–1224 (2014)
12. Liu, J., Wang, C., Gao, J., Han, J.: Multi-view clustering via joint nonnegative matrix factorization. In: *Proc. of SDM*. vol. 13, pp. 252–260. SIAM (2013)
13. Oliva, A., Torralba, A.: Modeling the shape of the scene: A holistic representation of the spatial envelope. *International journal of computer vision* 42(3), 145–175 (2001)
14. Tzortzis, G., Likas, A.: Kernel-based weighted multi-view clustering. In: *Data Mining (ICDM), 2012 IEEE 12th International Conference on*. pp. 675–684. IEEE (2012)
15. Wang, H., Nie, F., Huang, H.: Multi-view clustering and feature learning via structured sparsity. In: *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*. pp. 352–360 (2013)
16. Zhang, X., Zhao, L., Zong, L., Liu, X., Yu, H.: Multi-view clustering via multi-manifold regularized nonnegative matrix factorization. In: *Data Mining (ICDM), 2014 IEEE International Conference on*. pp. 1103–1108. IEEE (2014)
17. Zhu, X., Ghahramani, Z., Lafferty, J., et al.: Semi-supervised learning using gaussian fields and harmonic functions. In: *ICML*. vol. 3, pp. 912–919 (2003)