MULTI-VIEW DATA REPRESENTATION VIA DEEP AUTOENCODER-LIKE NONNEGATIVE MATRIX FACTORIZATION

Haonan Huang[1], *Yihao Luo*[1], *Guoxu Zhou*[1,*], *Qibin Zhao*[1,2]

[1]School of Automation, Guangdong University of Technology, China; [2]RIKEN AIP, Japan

ABSTRACT

Since a large proportion of real-world data is made of different representations or views, learning on data represented with multiple views (e.g., numerous types of features or modalities) has garnered considerable attention recently. Nonnegative matrix factorization (NMF) has been widely adopted for multi-view learning due to its great interpretability. We focus on unsupervised multi-view data representation in this paper and propose a novel framework termed Deep Autoencoder-like NMF (DANMF-MDR), which learns an intact representation by simultaneously exploring multi-view complementary and consistent information. Furthermore, an efficient iterative optimization algorithm is developed to solve the proposed model. Experimental results on three real-world multi-view datasets demonstrate that ours performs better than the SOTA multi-view NMF-based MDR approaches.

Index Terms— nonnegative matrix factorization, multi-view data representation, unsupervised learning, deep matrix factorization.

1. INTRODUCTION

Many data in real life are naturally composed of different types of features or come from multiple sources [1, 2], and these are referred to as multi-view data. Multi-view data representation (MDR) is an emerging topic of interest due to its ability to systematically embed rich information and multi-way interactions in the learning process and has recently become a common challenge in signal processing, pattern recognition, and other fields [3].

Extensive studies have been conducted in order to build successful MDR approaches [4], [5], [6]. Among these approaches, one of the most widely used MDR methods is based on nonnegative matrix factorization (NMF). Liu et al. [7] developed a joint nonnegative matrix factorization multi-view method with consensus term to learn the underlying clustering space embedded in multiple views. Liang et al. [8] designed a novel NMF method with co-orthogonal constraints to capture the multi-view diversity attributes. The general NMF framework is mainly relying on reconstructing the data from the representation of its weight, called the decoder component. To better inherit the representation learning ability of autoencoders, the Non-negative Symmetric Encoder-Decoder (NSED) was proposed with a novel encoder component in [9]. Due to its symmetric and non-negative structure, NSED also implicitly achieves the sparsity and orthogonality of low-dimensional space, thus has a better representation ability than NMF in real-life application.

In this paper, we introduce a new MDR methodology based on NSED. Besides being the first to use NSED in an MDR application, our method differs from most existing NMF-based MDR methods in that we use a deep structure [10] to perform NSED hierarchically. Through the deep NSED structure, we can uncover complicated hierarchical and structural information with implicit lower-level hidden attributes, which cannot be captured using traditional shallow NMF-based methods. More importantly, our method uses a unified framework to perform multi-view representation learning (with the encoder component) and each view information decoding (with the decoder component), allowing us to balance the consistency and complementarity among multiple views and learn a common intact representation. To our knowledge, this is the first work an autoencoder-like nonnegative matrix factorization technique has been applied to MDR with a deep structure.

2. DEEP AUTOENCODER-LIKE NMF

Different from traditional NMF which only considers the loss of the decoder, Sun et al. ([9]) integrates a decoder component and an encoder component into a unified objective function NSED. The NSED can obtain the orthogonal cluster indicator matrix, from which exclusive cluster memberships can be naturally derived as answers for clustering tasks. However, there is only one layer mapping between the original data and the low-dimensional space in NSED, which cannot capture the complex hierarchical patterns of real-world data.

Recently, inspired by the development of deep learning [11], Ye et al. ([12]) extended NSED to a multi-layer version and designed the deep autoencoder-like NMF (DANMF) to automatically uncover the hierarchical mappings of

*Correspondence: gx.zhou@gdut.edu.cn

The work is partially supported by the National Natural Science Foundation of China Grant Numbers 62071132, 62073087 and 61973090.

attributes to facilitate clustering tasks. Mathematically, the data matrix $\mathbf{X} \in \mathbb{R}^{l \times n}$ is factorized into m nonnegative basis matrices $\mathbf{W}$ and a nonnegative representation matrix $\mathbf{H}_m$. Its objective function can be expressed as:

$$\begin{aligned}&\mathcal{O}_{DANMF} = \|\mathbf{X} - \mathbf{W}_1\mathbf{W}_2 \cdots \mathbf{W}_m\mathbf{H}_m\|_F^2 + \\ &\|\mathbf{H}_m - \mathbf{W}_m^T \cdots \mathbf{W}_2^T\mathbf{W}_1^T\mathbf{X}\|_F^2 + \lambda tr(\mathbf{H}_m\mathbf{L}\mathbf{H}_m^T) \\ &\text{s.t. } \mathbf{W}_i, \mathbf{H}_m \geq 0, \forall i = 1, \cdots, m\end{aligned} \tag{1}$$

where $\mathbf{W}_1 \in \mathbb{R}^{l \times p_1}, \mathbf{W}_i \in \mathbb{R}^{p_{i-1} \times p_i}$ denotes the basis matrices and $\mathbf{H} \in \mathbb{R}^{p_m \times n}$ denotes the final layer representation matrix. The first term is the decoder stage which obtains the best basis matrices and low-dimensional representation matrix $\mathbf{H}_m$ to reconstruct the input data $\mathbf{X}$. The second term is the encoder stage which fine-tunes submatrices to transform the original data $\mathbf{X}$ into the distributed representation $\mathbf{H}_m$. The last term is utilized to explore the intrinsic geometric structure of input data. Because the work in [12] is focusing on the problem of community detection, the $\mathbf{L}$ is constructed from adjacency matrix $\mathbf{X}$, i.e., $\mathbf{L} = \mathbf{D} - \mathbf{X}$, and $\mathbf{D}$ is a diagonal matrix with $d_{ii} = \sum_j x_{ij}$. While DANMF has the NSED algorithm capabilities mentioned above, this architecture enables it to better inherit the learning ability of deep autoencoder.

3. DANMF-BASED MULTI-VIEW DATA REPRESENTATION

3.1. Objective Function

For multi-view setting, we design a novel consistent approach to measure the consistency loss between the view-specific matrix $\mathbf{H}^v$ and the consensus matrix $\mathbf{H}^*$. The decoder component successfully extracts information from every single view into a common intact representation, while the encoder component encodes latent representation from common space return to every single view. Thus our approach is adaptable in balancing the consistency and complementarity of different views. Let $\mathbf{X} = \{\mathbf{X}^1, \cdots, \mathbf{X}^V\}$ denote the data of all views, the objective function of DANMF-MDR for multi-view data representation can be formulated as

$$\begin{aligned}\mathcal{O} = \sum_{v=1}^{V}(&\|\mathbf{X}^v - \mathbf{W}_1^v\mathbf{W}_2^v ... \mathbf{W}_m^v\mathbf{H}_m^v\|_F^2 \\ &+ \|\mathbf{H}_m^v - (\mathbf{W}_m^v)^T \cdots (\mathbf{W}_2^v)^T(\mathbf{W}_1^v)^T\mathbf{X}^v\|_F^2 \\ &+ \lambda tr(\mathbf{H}_m^v\mathbf{L}^v(\mathbf{H}_m^v)^T) + (\alpha^v)^\gamma\|\mathbf{H}_m^v - \mathbf{H}^*\|_F^2) \\ \text{s.t. } &\mathbf{W}_i^v, \mathbf{H}_m^v, \mathbf{H}^* \geq 0, \sum_{v=1}^{V}\alpha^v = 1, \forall i = 1, \cdots, m\end{aligned} \tag{2}$$

where α^v is the weight for the v-th view and γ ($\gamma > 1$) is the hyper-parameter to control the weights distribution. The $\mathbf{H}^*$ denotes the consensus matrix, $\mathbf{L}^v$ denotes the graph Laplacian matrix, and the way to define $\mathbf{L}^v$ can be found in [13]. During multi-view learning, we update the weights α^v for different types of features such that the important features acquire big weights. For brevity, we define $\mathbf{Y} = \mathbf{H}_m^v - \mathbf{H}^*$ and use $\mathbf{X}$, $\mathbf{W}_i$, $\mathbf{H}_m$, $\mathbf{L}$, α to denote $\mathbf{X}^v$, $\mathbf{W}_i^v$, $\mathbf{H}_m^v$, $\mathbf{L}^v$, α^v, which means that they are the same concept. The objective function of DANMF-MDR can be further rewritten with the properties of matrix trace, as follows:

$$\begin{aligned}\mathcal{O} = \sum_{v=1}^{V}\{&tr\left(\mathbf{X}^T\mathbf{X} + \mathbf{H}_m^T\mathbf{H}_m - 4\mathbf{X}^T\mathbf{\Psi}_{i-1}\mathbf{W}_i\mathbf{\Phi}_{i+1}\mathbf{H}_m\right. \\ &+ \mathbf{H}_m^T\mathbf{\Phi}_{i+1}^T\mathbf{W}_i^T\mathbf{\Psi}_{i-1}^T\mathbf{\Psi}_{i-1}\mathbf{W}_i^v\mathbf{\Phi}_{i+1}\mathbf{H}_m \\ &\left.+\mathbf{X}^T\mathbf{\Psi}_{i-1}\mathbf{W}_i\mathbf{\Phi}_{i+1}\mathbf{\Phi}_{i+1}^T\mathbf{W}_i^T\mathbf{\Psi}_{i-1}^T\mathbf{X}\right) \\ &+ \lambda tr(\mathbf{H}_m\mathbf{L}\mathbf{H}_m^T) + \alpha^\gamma tr(\mathbf{Y}^T\mathbf{Y})\} \\ &\text{s.t. } \mathbf{W}_1^v, \cdots, \mathbf{W}_m^v \geq 0, \mathbf{H}_m^v \geq 0, \mathbf{H}^* \geq 0, \sum_{v=1}^{V}\alpha^v = 1\end{aligned} \tag{3}$$

where $\mathbf{\Psi}_{i-1}^v = \mathbf{W}_1^v\mathbf{W}_2^v \cdots \mathbf{W}_{i-1}^v$ and $\mathbf{\Phi}_{i+1}^v = \mathbf{W}_{i+1}^v \cdots \mathbf{W}_m^v$. It is worth noting that we denote $\mathbf{\Psi}_0^v = \mathbf{I}$ and $\mathbf{\Phi}_{m+1}^v = \mathbf{I}$. In addition, we need to normalize $\mathbf{H}_m^v$ and change $\mathbf{\Psi}_m^v$ to make the consistency constraint and improve the accuracy of the approximation:

$$\begin{aligned}\mathbf{H}_m^v &\leftarrow \mathbf{H}_m^v\left(\mathbf{Q}^v\right)^{-1} \\ \mathbf{\Psi}_m^v &\leftarrow \mathbf{\Psi}_m^v\mathbf{Q}^v\end{aligned} \tag{4}$$

where $\mathbf{Q}^v = \mathrm{diag}\left(\sum_i \mathbf{H}_{i,1}^v, \sum_i \mathbf{H}_{i,2}^v, \ldots, \sum_i \mathbf{H}_{i,k}^v\right)$.

3.2. Optimization

The update rule for hidden matrix $\mathbf{W}_i^v (m \geq i \geq 0)$: We minimize $\mathcal{O}$ over $\mathbf{W}_i^v$, with $\mathbf{H}_m^v$, $\mathbf{H}^*$ and α^v are fixed. Then, the objective function in Eq. 3 can be reduced to:

$$\begin{aligned}\mathcal{O}\left(\mathbf{W}_i\right) = tr&\left(-4\mathbf{X}^T\mathbf{\Psi}_{i-1}\mathbf{W}_i\mathbf{\Phi}_{i+1}\mathbf{H}_m\right. \\ &+ \mathbf{H}_m^T\mathbf{\Phi}_{i+1}^T\mathbf{W}_i^T\mathbf{\Psi}_{i-1}^T\mathbf{\Psi}_{i-1}\mathbf{W}_i^v\mathbf{\Phi}_{i+1}\mathbf{H}_m \\ &\left.+\mathbf{X}^T\mathbf{\Psi}_{i-1}\mathbf{W}_i\mathbf{\Phi}_{i+1}\mathbf{\Phi}_{i+1}^T\mathbf{W}_i^T\mathbf{\Psi}_{i-1}^T\mathbf{X}\right) \\ &\text{s.t. } \mathbf{W}_i^v \geq 0\end{aligned} \tag{5}$$

Let $\mathbf{\Theta}_i$ be the Langrange multiplier for nonnegative constraint on $\mathbf{W}_i$, resulting in the following Lagrangian function:

$$\begin{aligned}\mathcal{L}\left(\mathbf{W}_i, \mathbf{\Theta}_i\right) = tr&\left(-\mathbf{\Theta}_i\mathbf{W}_i^T - 4\mathbf{X}^T\mathbf{\Psi}_{i-1}\mathbf{W}_i\mathbf{\Phi}_{i+1}\mathbf{H}_m\right. \\ &+ \mathbf{H}_m^T\mathbf{\Phi}_{i+1}^T\mathbf{W}_i^T\mathbf{\Psi}_{i-1}^T\mathbf{\Psi}_{i-1}\mathbf{W}_i\mathbf{\Phi}_{i+1}\mathbf{H}_m \\ &\left.+\mathbf{X}^T\mathbf{\Psi}_{i-1}\mathbf{W}_i\mathbf{\Phi}_{i+1}\mathbf{\Phi}_{i+1}^T\mathbf{W}_i^T\mathbf{\Psi}_{i-1}^T\mathbf{X}\right)\end{aligned} \tag{6}$$

By setting the partial derivative of $\mathcal{L}\left(\mathbf{W}_i, \mathbf{\Theta}_i\right)$ with respect to $\mathbf{W}_i$ to $\mathbf{0}$, we have:

$$\mathbf{\Theta}_i = -4(\mathbf{\Psi}_{i-1})^T\mathbf{X}\mathbf{H}_m^T(\mathbf{\Phi}_{i+1})^T + 2\mathbf{\Pi}_i \tag{7}$$

where

$$
\begin{aligned}
\mathbf{\Pi}_i =& \mathbf{\Psi}_{i-1}^T \mathbf{\Psi}_{i-1} \mathbf{W}_i \mathbf{\Phi}_{i+1} \mathbf{H}_m \mathbf{H}_m^T \mathbf{\Phi}_{i+1}^T \\
&+ (\mathbf{\Psi}_{i-1})^T \mathbf{X}\mathbf{X}^T \mathbf{\Psi}_{i-1} \mathbf{W}_i \mathbf{\Phi}_{i+1} \mathbf{\Phi}_{i+1}^T
\end{aligned}
\tag{8}
$$

From the complementary slackness condition of the Karush-Kuhn-Tucker (KKT) conditions, we obtain:

$$
\mathbf{\Theta}_i \odot \mathbf{W}_i = \left(-4\mathbf{\Psi}_{i-1}^T \mathbf{X}\mathbf{H}_m^T \mathbf{\Phi}_{i+1}^T + 2\mathbf{\Pi}_i\right) \odot \mathbf{W}_i = \mathbf{0} \tag{9}
$$

where $\odot$ denotes the element-wise product. Equation (12) is the fixed point equation that the solution must satisfy at convergence. By solving this equation, we derive the following updating rule for $\mathbf{W}_i$:

$$
\mathbf{W}_i \leftarrow \mathbf{W}_i \odot \frac{2(\mathbf{\Psi}_{i-1})^T \mathbf{X}\mathbf{H}_m^T (\mathbf{\Phi}_{i+1})^T}{\mathbf{\Pi}_i} \tag{10}
$$

The update rule for representation matrix $\mathbf{H}_m^v$: The objective function in Eq. 3 can be reduced to the following Eq. 11 by fixing all the variables except for $\mathbf{H}_m$:

$$
\begin{aligned}
\mathcal{O}(\mathbf{H}_m) = & tr\left(\mathbf{H}_m^T \mathbf{H}_m - 4\mathbf{X}^T \mathbf{\Psi}_m \mathbf{H}_m + \mathbf{H}_m^T \mathbf{\Psi}_m^T \mathbf{\Psi}_m \mathbf{H}_m\right) \\
& + \lambda tr(\mathbf{H}_m \mathbf{L} \mathbf{H}_m^T) + \alpha^{\gamma} tr(\mathbf{Y}^T \mathbf{Y}) \\
& \text{s.t. } \mathbf{H}_m^v \geq 0
\end{aligned}
\tag{11}
$$

We set the graph Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{S}$ and follow the similar derivation process of the *updating rule for* $\mathbf{W}_i$, the updating rule for $\mathbf{H}_m$ is obtained as follows:

$$
\mathbf{H}_m \leftarrow \mathbf{H}_m \odot \frac{2\mathbf{\Psi}_m^T \mathbf{X} + \lambda \mathbf{H}_m \mathbf{S} + \alpha^{\gamma} \mathbf{H}^*}{\mathbf{\Psi}_m^T \mathbf{\Psi}_m \mathbf{H}_m + \lambda \mathbf{H}_m \mathbf{D} + (1+\alpha^{\gamma})\mathbf{H}_m} \tag{12}
$$

The update rule for consensus matrix $\mathbf{H}^*$: Optimizing the consensus matrix $\mathbf{H}^*$ amounts to solving the following function:

$$
\begin{aligned}
\mathcal{O}(\mathbf{H}^*) = & \sum_{v=1}^{V} \alpha^{\gamma} tr\left((\mathbf{H}_m - \mathbf{H}^*)^T (\mathbf{H}_m - \mathbf{H}^*)\right) \\
& \text{s.t. } \mathbf{H}^* \geq 0
\end{aligned}
\tag{13}
$$

We take the derivative of the Eq. 13 with respect to $\mathbf{H}^*$, set it to 0, and obtain an exact solution for $\mathbf{H}^*$:

$$
\mathbf{H}^* = \frac{\sum_v^V \alpha^{\gamma} \mathbf{H}_m}{\sum_v^V \alpha^{\gamma}} \tag{14}
$$

The update rule for weight factors α^v: To optimize the weight factor, we only consider the term that is relevant to α^v, and derive the following solution utilizing a similar proof to [14]:

$$
\alpha^v = \frac{(\mathcal{Y}^v)^{\frac{1}{1-\gamma}}}{\sum_{v=1}^{V} (\mathcal{Y}^v)^{\frac{1}{1-\gamma}}}, \tag{15}
$$

where $\mathcal{Y}^v = \|\mathbf{Y}^v\|_F^2$.

Algorithm 1 DANMF-MDR algorithm

Input: Multi-view Data $\{\mathbf{X}^v\}_{v=1}^V$, the number of layer m, layer sizes $\{p_i\}_{i=1}^m$, hyperparameters λ.
Output: The consensus matrix $\mathbf{H}^*$.

```
Initialization
for v = 1 to V do
    Compute graph Laplacian matrix L^v from X^v;
    for i = 1 to m do
        (W_i^v, H_i^v) ← NMF(H_{i-1}^v, p_i);
    end for
end for
while not converged do
    for v = 1 to V do
        for i = 1 to m do
            Update W_i^v via Eq. 10;
        end for
        Update H_m^v via Eq. 12;
        Normalize W^v and H^v by Eq. 4;
        Update α^v via Eq. 15;
    end for
    Update H^* via Eq. 14;
end while
```

Table 1. Statistics of datasets used in experiments.

Datasets		Sizes	Classes	Views
documents	Washington	230	5	2
images	MSRCV1	210	7	5
	Handwritten	2000	10	2

4. EXPERIMENTS AND ANALYSIS

4.1. Experimental Setup

In this section, we evaluate the effectiveness of our proposed DANMF-MDR on three widely-used multi-view datasets, which are chosen from documents and images, i.e., Washington[1], MSRCV1[15], and Handwritten[2]. The descriptions of these datasets are summarized in Table 1.

Regrading the evaluation metrics, we utilized the three most used evaluation metrics Accuracy (**ACC**), Normalized Mutual Information (**NMI**), and **Purity**. To save space, readers could refer to [16, 17] for details about the metrics. Because each of these evaluation measures reports the clustering attribute in a distinct way, they analyze the clustering performance in a systematic way. And a higher value signifies greater performance for all evaluation metrics.

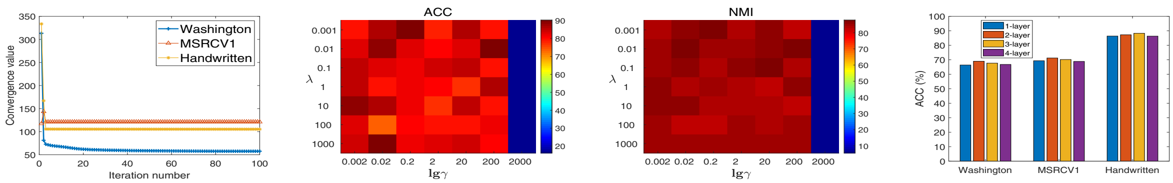
4.2. Performance Evaluation

We consider several state-of-the-art multi-view learning methods as comparing methods, including ANMF [9], DAN-

[1]http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-20/www/data/
[2]https://archive.ics.uci.edu/ml/datasets/Multiple+Features

Table 2. Results on three datasets(mean (%) $\pm$ standard deviation). Higher value indicates better performance and the highest values are in boldface. The last line shows the running time comparison of each iteration of different methods, taking dataset Handwritten as an example.

Datasets		NSED	DANMF	MultiNMF	NMFCC	2CMV	DMVC	PSDMF	Ours
Washington	ACC	59.04±1.49	61.39±2.94	50.17±1.50	62.87±2.01	59.91±2.11	41.65±0.94	67.22±1.33	**67.83±0.43**
	NMI	28.04±5.77	35.50±1.51	21.57±0.75	34.56±3.81	27.71±3.73	4.46±1.27	41.06±1.52	**43.52±0.40**
	Purity	67.48±2.74	69.22±1.46	67.83±1.28	69.39±1.59	65.24±1.65	46.52±0.00	72.52±1.17	**72.87±0.58**
MSRCV1	ACC	53.71±18.81	66.38±0.26	67.98±1.51	62.95±3.15	66.26±5.04	54.38±5.12	67.71±4.40	**71.24±0.93**
	NMI	47.73±16.45	52.07±0.98	58.75±1.77	50.72±3.58	58.39±3.19	49.61±3.12	61.10±2.92	**63.47±1.19**
	Purity	56.67±17.15	68.67±1.40	67.99±1.49	64.76±3.32	68.43±3.35	58.67±3.17	69.33±3.12	**71.52±0.62**
Handwritten	ACC	76.00±0.72	79.80±6.14	71.70±2.77	80.89±7.99	66.02±3.75	73.04±5.78	86.83±7.22	**87.25±5.17**
	NMI	78.61±1.62	81.79±1.64	68.47±2.22	81.16±3.61	64.99±1.01	74.91±1.43	86.98±2.90	**87.95±2.52**
	Purity	80.09±0.81	83.24±4.25	74.33±2.73	82.47±6.59	69.78±2.43	75.06±4.76	**88.68±5.54**	88.10±4.59
	Time	0.0103	0.0189	0.335	0.7705	1.9982	0.0674	0.1061	0.0431



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

Fig. 1. (a). Iterations versus the convergence value of DANMF-MRL on three datasets. (b) and (c). ACC and NMI changes with the alterations of γ and λ on dataset Handwritten, respectively. (d). ACC with different number of layers on three datasets.

MF [12], MultiNMF [7], NMFCC [8], 2CMV [18], D-MVC [19], PSDMF [15]. We directly concatenate different views data as the input of the single-view methods (i.e., ANMF and DANMF). It should be noted that DMVC and PSDMF are two recently presented MDR methods based on deep matrix factorization, but they all ignore the encoder component and nonnegative constraints. Since the clustering performance is affected by initializations, we repeat each method 10 times with random initializations and report the average results. It is noted that, the parameters of all methods, according to the corresponding article, search for the state that provides the best clustering performance. And the clustering results of multi-view representation learning methods are reported in Table 2.

As we can see, our proposed method achieves the better results than the SOTA methods on most of the datasets in terms of ACC, NMI, and Purity, which demonstrate the robustness of ours. For the MSRCV1 dataset, we raise the performance bar by around 3.53% in ACC, 2.37% in NMI, and 2.19% in Purity. Compared with DMVC and PSDMF, our advantages further prove that by integrating encoder components and decoder components, we can better inherit the learning ability of deep autoencoder and obtain a better clustering representation. In addition, we take dataset Handwritten as an example for the comparative experiment of running time. It can be seen that our method is superior to other multi-view methods, and is second only to single-view methods, which demonstrates the effectiveness and efficiency of our method.

4.3. Convergence and Parameters Analysis

From Fig. 1. (a), we can observe that the convergence value decreases steadily on all datasets and achieve fast convergence within about 40 iterations. The effect of parameters γ and λ on Handwritten is shown in Fig. 1. (b) and (c). We can find that DANMF-MDR is robust to λ and the best performance is obtained when $\lg\gamma = 0.2$ and $\lambda = 0.001$. We also investigated the influence of different layers on the ACC as shown in Fig. 1. (d). We observe that the setting of 2-layer performs best on the first two cases and 3-layer achieves best results on the last dataset.

5. CONCLUSION

A novel NMF-based multi-view data representation method, DANMF-MDR, is proposed in this paper. Specifically, DANMF-MDR aims at capturing the multi-view complementarity and consistency through an autoencoder-like NMF framework to obtain an intact representation. An efficient algorithm is developed to solve the optimization problem. Comprehensive experiments on three datasets are constructed to present the representation advantages of the DANMF-MDR against SOTA multi-view learning methods.

6. REFERENCES

[1] Yuning Qiu, Guoxu Zhou, Yanjiao Wang, Yu Zhang, and Shengli Xie, "A generalized graph regularized non-negative tucker decomposition framework for tensor data representation," *IEEE Transactions on Cybernetics*, 2020.

[2] Yuning Qiu, Guoxu Zhou, Zhenhao Huang, Qibin Zhao, and Shengli Xie, "Efficient tensor robust pca under hybrid model of tucker and tensor train," *IEEE Signal Processing Letters*, pp. 1–5, 2022.

[3] Zuyuan Yang, Naiyao Liang, Wei Yan, Zhenni Li, and Shengli Xie, "Uniform distribution non-negative matrix factorization for multiview clustering," *IEEE Transactions on Cybernetics*, vol. 51, no. 6, pp. 3249–3262, 2020.

[4] Changqing Zhang, Huazhu Fu, Si Liu, Guangcan Liu, and Xiaochun Cao, "Low-rank tensor constrained multiview subspace clustering," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1582–1590.

[5] Changqing Zhang, Huazhu Fu, Qinghua Hu, Xiaochun Cao, Yuan Xie, Dacheng Tao, and Dong Xu, "Generalized latent multi-view subspace clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 1, pp. 86–99, 2018.

[6] Ben Yang, Xuetao Zhang, Feiping Nie, Fei Wang, Weizhong Yu, and Rong Wang, "Fast multi-view clustering via nonnegative and orthogonal factorization," *IEEE Transactions on Image Processing*, vol. 30, pp. 2575–2586, 2020.

[7] Jialu Liu, Chi Wang, Jing Gao, and Jiawei Han, "Multi-view clustering via joint nonnegative matrix factorization," in *Proceedings of the SIAM International Conference on Data Mining (SDM)*. SIAM, 2013, pp. 252–260.

[8] Naiyao Liang, Zuyuan Yang, Zhenni Li, Weijun Sun, and Shengli Xie, "Multi-view clustering by non-negative matrix factorization with co-orthogonal constraints," *Knowledge-Based Systems*, vol. 194, pp. 105582, 2020.

[9] Bing-Jie Sun, Huawei Shen, Jinhua Gao, Wentao Ouyang, and Xueqi Cheng, "A non-negative symmetric encoder-decoder approach for community detection," in *Proceedings of the 26th ACM International Conference on Information and Knowledge Management (CIKM)*, 2017, pp. 597–606.

[10] Haonan Huang, Zuyuan Yang, Zhenni Li, and Weijun Sun, "A converged deep graph semi-nmf algorithm for learning data representation," *Circuits, Systems, and Signal Processing*, pp. 1–20, 2021.

[11] Haonan Huang, Zuyuan Yang, Naiyao Liang, and Zhenni Li, "Semi-nmf network for image classification," in *Chinese Control Conference (CCC)*. IEEE, 2019, pp. 8899–8903.

[12] Fanghua Ye, Chuan Chen, and Zibin Zheng, "Deep autoencoder-like nonnegative matrix factorization for community detection," in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management (CIKM)*, 2018, pp. 1393–1402.

[13] Deng Cai, Xiaofei He, Jiawei Han, and Thomas S Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1548–1560, 2010.

[14] Xiao Cai, Feiping Nie, and Heng Huang, "Multi-view k-means clustering on big data," in *Twenty-Third International Joint conference on artificial intelligence (IJCAI)*, 2013.

[15] Haonan Huang, Naiyao Liang, Wei Yan, Zuyuan Yang, Zhenni Lit, and Weijun Sun, "Partially shared semi-supervised deep matrix factorization with multi-view data," in *International Conference on Data Mining Workshops (ICDMW)*. IEEE, 2020, pp. 564–570.

[16] Yuan Xie, Jinyan Liu, Yanyun Qu, Dacheng Tao, Wensheng Zhang, Longquan Dai, and Lizhuang Ma, "Robust kernelized multiview self-representation for subspace clustering," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 2, pp. 868–881, 2020.

[17] Yuyuan Yu, Kan Xie, Jinshi Yu, Qi Jiang, and Shengli Xie, "Fast nonnegative tensor ring decomposition based on the modulus method and low-rank approximation," *Science China Technological Sciences*, vol. 64, no. 9, pp. 1843–1853, 2021.

[18] Khanh Luong and Richi Nayak, "A novel approach to learning consensus and complementary information for multi-view data clustering," in *IEEE International Conference on Data Engineering (ICDE)*. IEEE, 2020, pp. 865–876.

[19] Handong Zhao, Zhengming Ding, and Yun Fu, "Multi-view clustering via deep matrix factorization," in *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2017, vol. 31.