Multiview Learning via Non-negative Matrix Factorization for Clustering Applications

Jiajia Chen[1(✉)], Ao Li[1], Jie Li[2], and Yangwei Wang[2]

[1] School of Computer Science and Technology, Harbin University of Science and Technology, Harbin, China
544953065@qq.com

[2] Shandong Provincial Innovation and Practice Base for Postdoctors, Weihaizhenyu Intelligence Technology Co., Ltd., Weihai, China

Abstract. Multiview clustering is to more fully use the information between views to guide the division of data points, and multiview data is often accompanied by high-dimensionality. Since non-negative matrix factorization can effectively extract features while reducing dimensionality, this paper proposed a multi-view learning method based on non-negative matrix factorization. Compared with other NMF-based multiview learning methods, the proposed method has the following advantages: 1) graph regularization is added to traditional NMF to explore potential popular structures, so that the learned similarity graph contains more potential information. 2) A common graph learning strategy is designed to integrate hidden information from different views. 3) Put the NMF-based similarity graph learning and common graph learning strategies into a unified framework, and optimize the similarity graph and common graph at the same time, so that the two promote each other. Experiments on three public datasets show that the proposed method is more robust than the existing methods.

Keywords: Non-negative matrix factorization · Multiview clustering · Similarity learning · Spectral clustering

1 Introduction

Due to the rapid development of multimedia technology, in practical applications, more and more data show high-dimensional and unlabeled. Therefore, how to deal with this kind of data effectively has become the current research hotspot, and dimensionality reduction and clustering also highlight the application value they contain. The main purpose of dimension reduction is to explore a low dimensional structure hidden in high dimensional space, which will contain useful information in high dimensional space as much as possible [1]. The purpose of clustering is to divide samples into different clusters according to the similarity of data [2–6]. Therefore, how to combine the two sufficiently to obtain a clustering model with high robustness is a challenge.

S. Shi et al. (Eds.): 6GN 2021, LNICST 439, pp. 354–361, 2022.
https://doi.org/10.1007/978-3-031-04245-4_31

In recent years, non negative matrix factorization (NMF) has become an effective dimension reduction method for multi view clustering. For example, Zong [7] et al. proposed a NMF based clustering framework to solve the problem of clustering unmapped data in multiple views. In order to understand the influence of the orthogonality of the vectors in the division matrix and the representation matrix, a new NMF model with co-orthogonality constraint was proposed by Liang et al. [8]. Similarly, Liang et al. [9] proposed a semi supervised multi view clustering method to solve the impact of dimensionality reduction on label data categories. In addition, Zhou et al. [10] proposed a grid sparsity based multi popular regularized multi NMF method for multi view clustering to capture the shared cluster structure among different views.

Although the above methods have achieved good performance, they adopt a two-step strategy for the information within and between views, and do not consider the possibility of mutual guidance between the two. The influence of noise on model learning is not considered. Our proposed a multiview learning strategy based on non-negative matrix factorization. This method can not only use the representation ability of non-negative matrix to retain the effective information between views, but also use a graph regularization constraint to learn the similarity relationship within views. To sum up, this paper has the following advantages:

(1) A non-negative matrix graph learning mechanism is constructed to explore the hidden lines in the view, which makes the common graph more robust.
(2) A joint optimization framework is designed, in which the similarity graph learning and common graph learning based on non-negative matrix are put into a unified framework for joint optimization. In the iterative process, hidden information can be transmitted among variables to increase the accuracy of clustering model.
(3) A numerical method is developed to obtain the optimal solution for each variable.

2 Related Work

2.1 Nonnegative Matrix Factorization

Given an original data $X \in R^{n\times d}$, where m is the sample dimension and n is the number of samples. The purpose of nonnegative matrix factorization is to find two nonnegative matrices $U \in R^{m\times r}$ and $P \in R^{n\times r}$, which can be as close as possible to the original data X, where $r < d$. The objective function of NMF can be expressed as follows:

$$\begin{aligned} &\min \left\| X - UP^T \right\|^2 \\ &s.t. U \geq 0, P \geq 0 \end{aligned} \tag{1}$$

where U is the base matrix and P is the coefficient matrix. In order to optimize formula (1), an iterative multiplication updating method is proposed as follows:

$$U_{ij} \leftarrow \mathbb{U}\frac{(XP)_{ij}}{\left(UP^TP\right)_{ij}} \tag{2}$$

$$P_{ij} \leftarrow \mathbb{P}\frac{\left(X^TU\right)_{ij}}{\left(PU^TU\right)_{ij}} \tag{3}$$

2.2 Similarity Graph Learning

Inspired by locality preserving projection method, Nie et al. [11] proposed an adaptive nearest neighbor graph learning method. Suppose that s_{ij} represents the probability similarity of two samples x_i and x_j, then the probability of similarity between two samples can be defined by the following formula:

$$\min_{s_i} \sum_{j=1}^{n} \left(\frac{1}{2} \left\| x_i - x_j \right\|_2^2 s_{ij} + \gamma s_{ij}^2 \right) \tag{4}$$

where $\|\cdot\|_2^2$ is the l_2-norm and γ is the trade-off parameter. By defining the graph Laplacian matrix $L = W - \left(S + S^T/2\right)$, formula (4) can be rewritten as follows:

$$\begin{aligned} &\min_{s} \ \mathrm{Tr}\left(XLX^T\right) + \gamma \|s\|_F^2 \\ &s.t. \ \ \mathbf{S1} = 1, 0 \leq s \leq 1. \end{aligned} \tag{5}$$

where W is a diagonal matrix satisfying $W_{ii} = 0.5\left(\sum S_{i*} + \sum S_{*i}\right)$. By optimizing formula (5), a similarity graph S can be adaptively learned from the data to represent the similarity relationship between samples.

3 Our Proposed

3.1 Similarity Graph Learning Based on Non-negative Matrix

The traditional graph construction process is independent of NMF. In the process of matrix decomposition, graph S is fixed. In order to overcome this problem, this paper integrates the non-negative matrix factorization process into the graph construction, so that the graph construction is based on the matrix factorization, rather than two independent processes. Given a set of multi view data $X = \left\{X^1 X^2, \ldots, X^v, v = 1, \ldots, m\right\}$ and v denote the number of views, the multi view graph learning objective function based on non-negative matrix is as follows:

$$\begin{aligned} &\min_{U,P,S} \sum_{v=1}^{m} \left\| X^v - U^v (P^v)^T \right\|_F^2 + \alpha Tr\left(P^v L^v (P^v)^T\right) + \beta \|s^v\|_F^2 \\ &s.t. \ \ \mathrm{U} \geq 0, \mathrm{P} \geq 0, 0 \leq \mathrm{S} \leq 1, \ \mathbf{S1} = 1 \end{aligned} \tag{6}$$

where α and β are trade-off parameters, $\|\cdot\|_F$ is Frobinus norm and $Tr(\cdot)$ is trace. By optimizing the above formula, the coefficient matrix P and the similarity graph S will transfer hidden information to each other in the iterative process, so that the coefficient matrix and the similarity graph of the next view contain more hidden cues.

3.2 Objective Function

This paper designs a common graph learning strategy, so that the learned common graph can fuse the effective information from different views, and use the common graph in the subsequent clustering process to increase the accuracy of the model. In order to get

better consistency information between views, the above models are put into a unified framework to increase the robustness of the model. The specific formula is as follows:

$$\begin{aligned} &\min_{U,P,S,G} \sum_{v=1}^{m} \left\| X^v - U^v (P^v)^T \right\|_F^2 + \alpha Tr\left(P^v L^v (P^v)^T \right) + \beta \|s^v\|_F^2 + \|C - S^v\|_F^2 \\ &s.t.\ \ \mathrm{U} \geq 0, \mathrm{P} \geq 0, 0 \leq \mathrm{S} \leq 1,\ \ \mathbf{S1} = 1 \end{aligned} \tag{7}$$

From formula (7), we can see that common graph C contains consistency clues and latent clues from various views. In the joint framework, the representation matrix P and the similarity graph S can be alternately optimized to obtain high quality common graph.

3.3 Optimization Strategy of Objective Function

It can be seen from Eq. (7) that all variables are coupled in the objective function. Therefore, in order to achieve the optimal value of each variable, this paper uses an alternative optimization strategy to solve the objective function.

Through some algebraic formulas, formula (7) is rewritten as:

$$\begin{aligned} &\min_{U,P,S,G} \sum_{v=1}^{m} \left\| X^v - U^v (P^v)^T \right\|_F^2 + \alpha Tr\left(P^v L^v (P^v)^T \right) + \beta \|S^v\|_F^2 + \|C - S^v\|_F^2 \\ &= tr\left(XX^T \right) - 2tr(XPU^T) + tr\left(UP^T PU^T \right) + \alpha tr\left(P^v L^v (P^v)^T \right) + \beta \|S^v\|_F^2 + \|C - S^v\|_F^2 \end{aligned} \tag{8}$$

Using Eq. (8), the objective function is divided into the following sub optimization problems.

Update U^v: Fix other variables and update U^v with formula (9):

$$\min_{U^v} tr\left(UP^T PU^T \right) - 2tr(XPU^T) \tag{9}$$

By derivation of the above formula, we can get the following formula:

$$\left[2UP^T P - 2XP \right]_{ij} U_{ij} = 0 \tag{10}$$

According to formula (10), the update rule of U^v is as follows:

$$U_{ij} = U_{ij} \frac{[XP]_{ij}}{\left[UP^T P \right]_{ij}} \tag{11}$$

Update P^v: Fix other variables and update P^v with formula (12):

$$\min_{P^v} tr\left(UP^T PU^T \right) - 2tr(XPU^T) + \alpha tr\left(P^v L^v (P^v)^T \right) \tag{12}$$

By deriving the above formula, we can get the following formula:

$$\left[-X^T U + PU^T U + \alpha LP \right]_{ij} P_{ij} = 0 \tag{13}$$

Since $L = D - S$, according to formula (12), the update rule of P^v is as follows:

$$P_{ij} = P_{ij} \frac{\left[X^T U + \alpha SP\right]_{ij}}{\left[PU^T U + \alpha DP\right]_{ij}} \tag{14}$$

Update S^v: To fix other variables, S^v can be updated with the following formula:

$$\begin{aligned} &\min_{s_i} \sum_{j=1}^{n} \left(\frac{\beta}{2} \left\| p_i - p_j \right\|^2 s_{ij} + \gamma s_{ij}^2 + \left(c_{ij} - s_{ij} \right) \right) \\ &s.t.\ \ s_i^T \mathbf{1} = 1,\ 0 \le s_{ij} \le 1 \end{aligned} \tag{15}$$

Define $f_{ij} = \left\| p_i - p_j \right\|^2 - \frac{4}{\beta} c_{ij}$ and $f_i \in R^{n \times 1}$, then formula (15) can be rewritten as:

$$\min_{s_i^T \mathbf{1} = 1, 0 \le s_{ij} \le 1} \left\| s_i + \frac{\beta}{4\gamma} f_i \right\|_2^2 \tag{16}$$

By removing the constraints of the above equation, the Lagrangian form of Eq. (16) is given as follows:

$$\begin{aligned} \mathcal{L}(s_i, \eta, \xi) = &\left\| s_i + \frac{\beta}{4\gamma} f_i \right\|_2^2 \\ &- \eta \left(s_i^T \mathbf{1} - 1 \right) - \xi_i^T s_i \end{aligned} \tag{17}$$

where η and $\xi \in R^{n \times 1}$ are Lagrange multipliers. Under the KTT condition, the solution is obtained by the following formula:

$$\begin{cases} s_{ik} = \dfrac{\eta}{2} - \dfrac{\beta f_{ik}}{4\gamma_i} > 0 \\ s_{i,k+1} = \dfrac{\eta}{2} - \dfrac{\beta f_{i,k+1}}{4\gamma_i} \le 0 \\ s_i^T \mathbf{1} = \sum_{j=1}^{k} \left(\dfrac{\eta}{2} - \dfrac{\beta f_{ij}}{4\gamma_i} \right) = 1 \end{cases} \Rightarrow \begin{cases} s_{ij} = \dfrac{f_{i,k+1} - f_{ij}}{k f_{i,k+1} - \sum_{r=1}^{k} f_{ir}}, j \le k \\ \gamma_i = \dfrac{\beta}{4} \left(k f_{i,k+1} - \sum_{j=1}^{k} f_{ij} \right) \\ \eta = \dfrac{2}{k} + \dfrac{\beta}{2k\gamma_i} \sum_{j=1}^{k} f_{ij} \end{cases} \tag{18}$$

Update C: To fix other variables, C can be updated with the following formula:

$$\min_{C} \sum_{v=1}^{m} \left\| C - S^v \right\|_F^2 \tag{19}$$

The optimal solution can be obtained by solving each row of matrix C:

$$\min_{c_i^T \mathbf{1} = 1, c_{ij} > 0} \sum_{v=1}^{m} \left\| c_i - s_i^v \right\|_F^2 \tag{20}$$

Formula (19) can be solved by an iterative algorithm, which was proposed in reference [12].

4 Experiments

4.1 Datasets

COIL20 dataset: COIL20 is a collection of grayscale images, including 20 objects shot from different angles. The images of the objects are taken every 5 degrees. There are 72 images of each object, 1440 images in total. According to the shooting angle, this article divides it into four different viewing angle data: V1[0°–85°], V2[90°–175°], V3[180°–265°], V4[270°–360°].

UCI dataset: This data set contains 10 digital handwritten images from 0, 1, 2,…, 9. Each number has 200 samples, and the entire data set has a total of 2000 samples. This paper selects 500 samples from them, and extracts three different features from each sample as three different perspectives. The first perspective extracts 216-dimensional profile-related features, the second perspective extracts 76-dimensional Fourier coefficients, and the third perspective extracts a 6-dimensional morphological feature.

YALE dataset: Yale collected facial images from 15 different people. Each of these people has 11 photos with different light, pose and expression, for a total of 165 images. This paper will extract three features of gray intensity (Gray), local binary pattern (LBP) and Gabor for each sample in the data set as three different perspective data.

4.2 Experimental Results and Analysis

In the experiment, in order to fully prove the effectiveness of the proposed method, this paper will compare with four advanced methods on the above three datasets. The four methods are: the classical k-nearest neighbor graph construction method via Gaussian distance kernel (GCG), rbust graph learning from noisy data (RGC) in reference [13], parameter-free auto-weighted multiple graph learning(AMGL) [14], and Multi-Graph Fusion for multi-view Spectral Clustering(GFSC) in reference [15].

In the experiment, the missing value rate is used to generate noise data. The rate of missing values changed from 0.1 to 0.5, and the interval step was 0.1. It can be seen from Table 1, 2 and 3 that with the increasing noise intensity, the ACC index of this method is basically higher than other comparison methods. Only in UCI dataset, when the noise intensity is 0.4 and 0.5, our method is lower than GFSC method, but it can be seen that our method is still above the average baseline, which proves the robustness of our proposed model.

Table 1. Accurate results of clustering on UCI dataset.

Missing rate	GFSC	AMGL	RGC	GCG	Ours
0.1	60.16	72.92	41.16	37.65	65.13
0.2	57.26	70.38	35.6	42.31	63.35
0.3	59.60	57.59	36.72	32.45	60.23
0.4	59.30	55.84	30.27	28.64	55.98
0.5	54.52	48.74	20.15	25.56	32.91

Table 2. Accurate results of clustering on YALE dataset.

Missing rate	GFSC	AMGL	RGC	GCG	Ours
0.1	48.78	56.84	51.91	54.79	65.26
0.2	41.12	55.73	58.74	56.26	63.98
0.3	30.20	59.75	45.06	52.16	60.05
0.4	24.65	58.62	43.20	52.91	59.19
0.5	19.38	55.04	24.86	46.42	58.94

Table 3. Accurate results of clustering on COIL20 dataset.

Missing rate	GFSC	AMGL	RGC	GCG	Ours
0.1	65.47	57.25	77.58	66.84	81.12
0.2	60.58	60.51	79.13	65.04	80.24
0.3	55.42	70.52	74.35	66.46	77.56
0.4	50.19	70.28	71.65	63.14	73.48
0.5	42.01	63.14	23.31	54.82	67.26

5 Conclusions

Aiming at the negative influence of noise on model learning, our proposed a multiview learning strategy via non-negative matrix factorization. This method combines non negative matrix factorization with graph learning model, so that they can transfer hidden information to each other and better explore the potential structure of data. In addition, a common graph learning strategy is proposed to fuse the effective information from different views, and a joint strategy is proposed to make the variables promote each other in the iterative process and increase the robustness of the model. Experimental results on three open datasets show that the proposed method is superior to the existing excellent multi view learning methods.

Acknowledgment. This work was supported in part by the National Natural Science Foundation of China under Grant 62071157, Natural Science Foundation of Heilongjiang Province under Grant YQ2019F011 and Postdoctoral Foundation of Heilongjiang Province under Grant LBH-Q19112.

References

1. Yi, Y., Wang, J., Zhou, W., et al.: Non-Negative Matrix Factorization with Locality Constrained Adaptive Graph. IEEE Trans. Circuits Syst. Video Technol. 1 (2019)
2. Wang, Q., Dou, Y., Liu, X., Lv, Q., Li, S.: Multi-view clustering with extreme learning machine. Neurocomputing **214**, 483–494 (2016)

3. Zhang, C., Hu, Q., Fu, H., Zhu, P., Cao, X.: Latent multi-view subspace clustering. In: Computer Vision and Pattern Recognition, pp. 4333–4341 (2017)
4. Li, B., et al.: Multi-view multi-instance learning based on joint sparse representation and multi-view dictionary learning. IEEE Trans. Pattern Anal. Mach. Intell. **39**, 2554–2560 (2017)
5. Jing, X., Wu, F., Dong, X., Shan, S., Chen, S.: Semi-supervised multi-view correlation feature learning with application to webpage classification. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, pp. 1374–1381 (2017)
6. Wu, J., Lin, Z., Zha, H.: Essential tensor learning for multi-view spectral clustering. IEEE Trans. Image Process. **28**, 5910–5922 (2019)
7. Zong, L., Zhang, X., Liu, X.: Multi-view clustering on unmapped data via constrained non-negative matrix factorization. Neural Netw. **108**, 155–171 (2018)
8. Liang, N., Yang, Z., Li, Z., et al.: Multi-view clustering by non-negative matrix factorization with co-orthogonal constraints. Knowl.-Based Syst. **194**, 105582 (2020)
9. Liang, N., Yang, Z., Li, Z., et al.: Semi-supervised multi-view clustering with graph-regularized partially shared non-negative matrix factorization. Knowl.-Based Syst. **190**, 105185 (2020)
10. Zhou, L., Du, G., Lü, K., et al.: A network-based sparse and multi-manifold regularized multiple non-negative matrix factorization for multi-view clustering. Expert Syst. Appl. **174**, 114783 (2021)
11. Nie, F., Wang, X., Huang, H.: Clustering and projected clustering with adaptive neighbors. In: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 977–986 (2014)
12. Duchi, J.C., Shalevshwartz, S., Singer, Y., Chandra, T.D.: Efficient projections onto the l1-ball for learning in high dimensions. In: International Conference on Machine Learning, pp. 272–279 (2008)
13. Kang, Z., Pan, H., Hoi, S.C.H., Xu, Z.: Robust graph learning from noisy data. IEEE Trans. Cybern. 1–11 (2019)
14. Nie, F., Li, J., Li, X.: Parameter-free auto-weighted multiple graph learning: a framework for multiview clustering and semi-supervised classification. In: International Joint Conference on Artificial Intelligence, pp. 1881–1887 (2016)
15. Kang, Z., et al.: Multi-graph fusion for multi-view spectral clustering. Knowl. Based Syst. **189**, 102–105 (2020)