Multi-view clustering via structured low-rank representation

Dong Wang, Qiyue Yin, Ran He, Liang Wang, Tieniu Tan
Center for Research on Intelligent Perception and Computing (CRIPAC)
Institute of Automation, Chinese Academy of Sciences
{dwang, qyyin, rhe, wangliang, tnt} @nlpr.ia.ac.cn

ABSTRACT

In this paper, we present a novel solution to multi-view clustering through a structured low-rank representation. When assuming similar samples can be linearly reconstructed by each other, the resulting representational matrix reflects the cluster structure and should ideally be block diagonal. We first impose low-rank constraint on the representational matrix to encourage better grouping effect. Then representational matrices under different views are allowed to communicate with each other and share their mutual cluster structure information. We develop an effective algorithm inspired by iterative re-weighted least squares for solving our formulation. During the optimization process, the intermediate representational matrix from one view serves as a cluster structure constraint for that from another view. Such mutual structural constraint fine-tunes the cluster structures from both views and makes them more and more agreeable. Extensive empirical study manifests the superiority and efficacy of the proposed method.

Categories and Subject Descriptors

H.3 [**Information search and retrieval**]: Clustering

General Terms

Algorithms, Design, Experimentation

Keywords

Multi-view clustering, multi-modal learning, structure regularizer

1. INTRODUCTION

Multi-view clustering concerns the problem of partitioning data points into a series of subsets in an unsupervised way given their feature representations under different views. Here in the context of this paper, a view simply refers to one feature modality of the data rather than the physical view angle such as front-view versus side-view face images [23]. In many applications, the data points being processed are collected from multiple sources and thus have different view-specific attributes. For instance, in image classification an image can be either represented using the traditional handcrafted feature like SIFT or automatically learned feature obtained from deep learning techniques [7]. Even though the information from any view is somehow sufficient for the clustering task, taking advantage of the complementary information across views is beneficial and can better facilitate the clustering process in most cases.

In the literature, a spectrum of methods are proposed to seek better clustering results by capturing the view complementarity. Among them, one line of research is to directly unify the multi-view information in the clustering process. For instance, in [15] a co-training flavored spectral clustering algorithm was proposed to encourage the clustering agreement between views. Another one is [16] which attempted to regularize on the eigenvectors of view-specific graph laplacians and achieve consistent clusters across views. Another line of research is to first learn a latent representation for multi-view data and then perform clustering on such representation to learn the partition. A notable one is [19] which employed matrix factorization to discover a common latent structure shared by all views and give rise to compatible clustering results. Besides, CCA based multi-view clustering methods also fall into this category [9] [5]. Yet another line of research is by fusing the clustering results obtained from individual views toward a consensus [8] [13].

Different from the aforementioned methods, we tackle the multi-view clustering problem from the perspective of structured low-rank representation. Similar to [24] [21], for each view we also assume similar samples can be used to linearly reconstruct each other. The resulting representational matrix reflects the cluster structure which should be ideally block diagonal. As in [18] [17], low-rankness is a nice property favored by many subspace clustering algorithms due to its better grouping effect. So the key idea of our method is on one hand to impose a low-rank constraint on such representational matrix to let similar samples stay together. On the other hand, since the ideal clustering result should be unanimous irrespective of views, the representational matrix derived from each view is further asked to conform with one another as much as possible through a mutual structure constraint.

Concretely, when alternatively optimizing the objective with respect to one of the representational matrices once at a time while keeping the others fixed, our method actually solves a low-rank linear regression problem. And it happens to enforce that the to-be-determined representational matrix for another view should refer to that fixed intermediate grouping structure. Over time the complementary information among different views is communicated and shared. The grouping structure for one view derived from the previous step helps rectify and fine-tune the to-be-decided grouping structure for a different view in the current step.

In summary, the contributions of this paper are highlighted as

CIKM'15 October 19 - 23, 2015, Melbourne, VIC, Australia

ACM 978-1-4503-3794-6/15/10 ...$15.00.
DOI: http://dx.doi.org/10.1145/2806416.2806629.

follows:

• We propose a novel multi-view clustering method based on structured low-rank representation. Such a joint regularization framework explicitly minimizes the grouping differences across views and gives rise to better clustering performance.

• We develop an effective algorithm inspired by iterative reweighted least squares for solving our formulation. And extensive experimental results on benchmark datasets validate the usefulness of our method.

2. PROPOSED MODEL AND SOLUTION

In our model, under each view similar data points are used to linearly reconstruct each other. Given the data matrix $X_v \in \mathbb{R}^{d_v \times n}$ and reconstruction coefficient matrix $Z_v \in \mathbb{R}^{n \times n}$ where v denotes one of the two views A and B, d_v is the feature dimension of view v and n is the number of data points, our proposed method is formulated as follows.

$$\min_{Z_A, Z_B} \sum_{v \in \{A,B\}} (\frac{1}{2}||X_v - X_v Z_v||_F^2 + \lambda||Z_v||_*) + \frac{\beta}{2}||Z_A - Z_B||_F^2 \tag{1}$$

As in [24] [21], (1) is called a structured low-rank representation since when fixing Z_B, the cluster structure from view A is contrained to agree with a latent cluster structure from view B or vice versa. The first term in this objective is the linear reconstruction term and in particular we seek a lowest rank representation by imposing a trace norm regularizer on Z_v. By virtue of the grouping effect of such a low-rank constraint, the underlying intrinsic cluster structure can be unveiled.

Apart from pursuing better clustering structure for any single view, our model also favors a consistent cluster membership across views. In multi-view clustering, it is usually anticipated that a data point should be assigned to the same cluster irrespective of views. To this end, the last term in our objective is designed to minimize the difference of cluster structures from different views. When treating the candidate samples used to reconstruct the target sample as a dictionary, Z_A and Z_B are the view-specific representational responses on the dictionary, which indicates the sample affinity and cluster structure.

λ and β are the hyper-parameters that control the tradeoff between corresponding terms.

Note that our model can be easily generalized to more than three views by summing the reconstruction and cluster difference terms over all views. Given the space limit and for simplicity, in this section we base our introduction on a two-view scenario.

When it comes to solving the proposed objective, it is not easy to optimize (1) directly given the existence of a trace norm regularizer. So we reformulate our objective by following a well established variational formulation for trace norm [6] [14], in which the statement below holds true for the representational matrix Z_A (Similar for Z_B)

$$\begin{aligned} ||Z_A||_* &= \frac{1}{2} \inf_{S_A \succcurlyeq 0} \; tr(Z_A^T S_A^{-1} Z_A) + tr(S_A) \\ &= \frac{1}{2} \inf_{S_A \succcurlyeq 0} \sum_{i=1}^{n} Z_{Ai}^T S_A^{-1} Z_{Ai} + tr(S_A) \end{aligned} \tag{2}$$

where the infimum is obtained for $S_A = (Z_A Z_A^T)^{1/2}$. Z_{Ai} is the i-th column of matrix Z_A. Here S_A can be seen as an intermediate variable during the optimization procedure.

In the outer loop of our algorithm, we alternatively solve for one of the representational matrix Z_A or Z_B while keeping the other one fixed. In light of the results from (2), when we optimize the objective with respect to Z_A in a column-wise fashion, (1) can be simplified into the following:

$$\begin{aligned} \min_{Z_{Ai}} \inf_{S_A \succcurlyeq 0} \sum_{i=1}^{n} &||X_{Ai} - X_A Z_{Ai}||_2^2 + \lambda Z_{Ai}^T S_A^{-1} Z_{Ai} \\ &+ \beta(Z_{Ai}^T Z_{Ai} - 2 Z_{Ai}^T Z_{Bi}) + \lambda tr(S_A) \end{aligned} \tag{3}$$

The above objective is jointly convex in (Z_{Ai}, S_A). And we solve it in the inner loop of our algorithm. In order to optimize this objective function by alternating the minimization over (Z_{Ai}, S_A), we need to add a term $\lambda \mu_A tr(S_A^{-1})$ which ensures S_A is invertible and thus the infimum can be attained [6] [14]. Here μ_A is a small scaler. And S_A is then given by:

$$S_A = (Z_A Z_A^T + \mu_A I)^{1/2} \tag{4}$$

When S_A is given, (3) becomes an iterative re-weighted least squares problem whose solution is the following:

$$Z_{Ai} = (X_A^T X_A + \beta I + \lambda S_A^{-1})^{-1} (X_A^T X_{Ai} + \beta Z_{Bi}) \tag{5}$$

Similarly, when we solve for Z_B while fixing Z_A, based on the same optimization strategy S_B is given by

$$S_B = (Z_B Z_B^T + \mu_B I)^{1/2} \tag{6}$$

and then we obtain the solution for each column of the representational matrix Z_B as follows:

$$Z_{Bi} = (X_B^T X_B + \beta I + \lambda S_B^{-1})^{-1} (X_B^T X_{Bi} + \beta Z_{Ai}) \tag{7}$$

When optimizing our method following the procedure as shown in Algorithm 1, empirically it can quickly converge after five to ten iterations. And one of the advantages for solving the objective in a column-wise fashion is that we can select a few nearest neighbors to approximately reconstruct a target sample. This makes sense because Z should ideally be block diagonal which implies that candidate samples less similar to the target play insignificant roles in the reconstruction. In such case we update Eqn.(4)-(7) only using smaller data matrices or representational matrices whose entries are extracted from the nearest neighbor positions in the original large matrices. This strategy alleviates the burden of high computational cost due to the large number of samples in the databases. The most time-consuming part is computing (4) and (6) which involves Singular Value Decomposition (SVD) [12]. If we use k nearest neighbors ($k \ll n$) for the linear reconstruction, we only need to decomposite multiple much smaller $k \times k$ matrices rather than the original large $n \times n$ matrix, which reduces the complexity from $O(n^3)$ to $O(nk^2)$. Another advantage is that it is convenient to develop a paralleled solution which may further speed up the algorithm. Once Z_A and Z_B are obtained, we average them by letting $Z = (|Z_A| + |Z_B|)/2$. Then a spectral clustering algorithm like [20] is applied on Z to achieve the final clustering results.

3. EXPERIMENTAL RESULTS

In this section, we test our method on widely used benchmark databases and compare with a series of baselines in order to validate the usefulness of the proposed model.

3.1 Databases

UCI Handwritten Digit dataset [1] consists of features of handwritten digits (0–9). The dataset is represented in terms of six features and contains 2000 samples with 200 in each category. Similar

Algorithm 1 Multi-view clustering via structured low-rank representation (MVCSL)

Input:
Data matrices X_A and X_B, parameters λ and β. Initial guesses Z_A, Z_B, number of clusters c, number of nearest neighbors k for reconstruction and $\mu_A = \mu_B = 10^{-5}$

while not converged **do**
  // Solve Z_A with Z_B fixed
  for $i = 1 : n$ **do**
    Update S_A using Equation (4);
    Update Z_{Ai} using Equation (5) ;
  end for
  // Solve Z_B with Z_A fixed
  for $i = 1 : n$ **do**
    Update S_B using Equation (6);
    Update Z_{Bi} using Equation (7);
  end for
end while

Output: Z_A, Z_B and final clustering results

to [16], we select the 76 Fourier coefficients of the character shapes and the 216 profile correlations as two views of the original dataset.

Movies617 dataset [3] consists of 617 movies with 17 labels extracted from IMDb. The two views are the 1878 keywords and the 1398 actors with a keyword used for at least 2 movies and an actor appeared in at least 3 movies.

Animal dataset [2] consists of 30475 images of 50 animals with six pre-extracted features for each image. Three kinds of features, namely PyramidHOG (PHOG), colorSIFT and SURF, are chosen as three views. We select the first ten categories with each including randomly chosen 50 samples as a subset for evaluation.

Pascal VOC 2007 dataset [4] consists of 20 categories with a total of 9,963 images. We use the Color feature and Bow feature as two-view visual representation. Furthermore, those images with multiple categories are removed, thus leaving 5,649 images for evaluation.

NUS WIDE dataset [11] consists of 269,648 images of 81 categories collected from Flickr. In our experiments, We select 500 images from each of the five classes with the most number of images for evaluation. Six types of low level features are given and we use color correlogram and wavelet texture as two-view representations for multi-view clustering.

3.2 Experimental settings

We extensively compare our method with many representative baselines including 1) *S_Spectral*: Use spectral clustering in [20] to cluster each view's data and select the best clustering result. 2) *S_LowRank*: Use only single-view low-rank representation to construct the affinity matrix and then apply spectral clustering in [20] to cluster the dataset. We also report the best clustering results. 3) *Combined*: Concatenate features from two views and apply low-rank representation without the mutual structural constraint on the combined feature to perform clustering. 4) *PairwiseSC, CentroidSC*: [16] Two objectives for co-regularizing the eigenvectors of all views' Laplacian matrices. 5) *Co_Training*: [15] Alternately modify one view's graph structure using the other view's information. 6) *Multi_NMF*: [19] A multi-view non-negative matrix factorization method to group the multi-view data. Note that this method is not applicable on NUS dataset since it requires all non-negative input features. 7) *Multi_SS*: [22] A structure sparsity based multi-view clustering and feature learning framework. The parameters in these methods are carefully selected in order to achieve their best results.

Whenever K-means is involved, it is run 20 times with random initialization. To speedup the optimization process, during the linear reconstruction we select 100 nearest neighbors of a sample point for its reconstruction. To measure the clustering results, we use accuracy (ACC) and normalized mutual information (NMI). Readers can refer to [10] for more details about such measures. Both mean and standard deviation are reported.

3.3 Experimental results and analysis

It can be seen from Table 1 and 2 that our proposed method (MVCSL) consistently outperforms other baselines using both measures. First of all, comparing with single-view methods like either $S\_Spectral$ or $S\_LowRank$, our method always has an upper hand, which evidences the necessity of utilizing the complementary information among different views and exploring intrinsic group structure. Second, a naive concatenation of features from multiple views as the baseline $Combined$ does is somehow ineffective. However our method explicitly asks the view-specific cluster structure, which is manifested in the representational matrices arising from the data reconstruction, to agree with each other as much as possible. Therefore the additional complementary information across views is shared and thus more accurate clustering results can be obtained. Besides, our method also beats other baselines by a considerable margin. The baseline $Multi\_SS$ puts the data from all views together and explores its global structure. It enforces the sparsity between views while somehow neglecting the intrinsic structure for any individual view. But this is where our low-rank constraint imposed on each view stands out. Our method takes into account both the intra-view partition and inter-view association, which proves that such structured low-rank framework is quite helpful in the multi-view clustering problem.

As mentioned previously, our method can be extended to scenarios involving three or more views, The superior results of our method on the three-view Animal dataset proves full well that the proposed method is also workable beyond two views.

When selecting the parameters λ and β, we empirically grid-search in the interval [0.001, 10]. And their influences on the clustering performance are shown in Figure 1 and 2. By pairing proper λ and β, it is not difficult to get satisfactory results. Given the space limit, only results on the Movies617 dataset are reported and similar trends can be observed on the other datasets as well.

4. CONCLUSION

We have proposed a novel multi-view clustering method through a structured low-rank representation. On one hand, with the help of better grouping effect of a low-rank regularizer, similar data points are assigned together with higher accuracy. On the other hand, a mutual structure constraint is imposed to achieve consistent cluster memberships across views. The view-specific representational matrices resulting from the data reconstruction alternatively serve as the structural reference for one another. The experimental results demonstrate the effectiveness of our proposed method.

5. ACKNOWLEDGEMENTS

This work is jointly supported by National Basic Research Program of China (No.2012CB316300) and National Science Foundation of China (No.61175003, No.61135002 and No.61403390).

6. REFERENCES

[1] http://archive.ics.uci.edu/ml/datasets/Multiple+Features.
[2] http://attributes.kyb.tuebingen.mpg.de/.

ACC(%)	Digits	Movies617	VOC	Animal	NUS
S_Spectral	66.37(4.44)	25.70(1.13)	15.64(0.43)	27.21(1.50)	33.85(0.18)
S_LowRank	66.53(5.31)	30.02(1.05)	17.09(0.41)	31.71(1.88)	34.74(0.02)
Combined	71.83(6.19)	32.93(1.53)	13.98(0.37)	32.51(1.00)	30.28(0.14)
PairwiseSC	80.82(6.30)	27.89(1.64)	11.93(0.14)	31.65(1.59)	33.07(0.14)
CentroidSC	82.77(7.14)	28.57(1.17)	13.98(0.39)	31.06(2.02)	34.80(0.99)
Co_Training	80.22(6.84)	30.74(1.28)	14.84(0.33)	30.35(1.48)	35.15(1.03)
Multi_NMF	69.24(6.28)	26.99(1.19)	12.57(0.23)	30.56(1.02)	N/A
Multi_SS	72.45(4.10)	29.60(1.10)	12.47(0.26)	32.11(1.86)	32.39(0.05)
MVCSL	**85.10(4.89)**	**33.51(1.13)**	**17.16(0.68)**	**32.61(1.34)**	**36.33(0.04)**

Table 1: Clustering results in terms of accuracy on five benchmark databases.

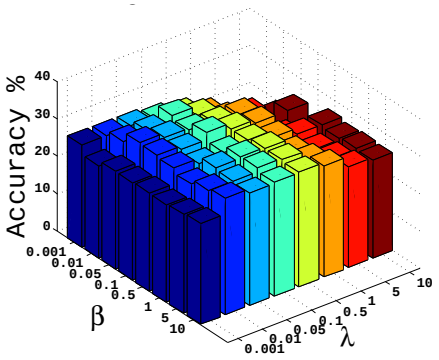

Figure 1: Accuracy vs. parameters λ and β on the Movies617 database

NMI(%)	Digits	Movies617	VOC	Animal	NUS
S_Spectral	62.30(1.85)	25.47(0.85)	9.34(0.19)	15.70(0.65)	6.64(0.15)
S_LowRank	69.79(1.81)	30.15(0.80)	7.05(0.18)	18.42(1.10)	7.68(0.01)
Combined	70.92(2.03)	32.11(0.89)	9.33(0.30)	20.09(0.58)	5.14(0.27)
PairwiseSC	75.84(2.37)	28.04(0.73)	6.07(0.12)	19.90(1.51)	6.87(0.01)
CentroidSC	76.76(2.58)	28.02(0.72)	9.50(0.28)	18.50(1.48)	7.84(0.26)
Co_Training	75.90(2.27)	30.74(1.28)	9.88(0.21)	18.98(0.73)	8.51(0.16)
Multi_NMF	65.05(2.30)	27.45(0.55)	6.40(0.19)	18.77(0.71)	N/A
Multi_SS	74.55(2.49)	30.09(1.32)	6.76(0.11)	21.25(1.76)	5.63(0.02)
MVCSL	**79.95(1.58)**	**34.79(0.92)**	**10.31(0.21)**	**22.11(0.83)**	**8.60(0.01)**

Table 2: Clustering results in terms of NMI on five benchmark databases.

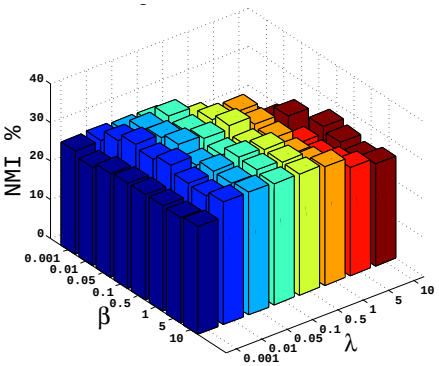

Figure 2: NMI vs. parameters λ and β on the Movies617 database

[3] http://membres-lig.imag.fr/grimal/data.html.

[4] http://pascallin.ecs.soton.ac.uk/challenges/VOC/voc2007/.

[5] U. Ahsan and I. Essa. Clustering social event images using kernel canonical correlation analysis. In *IEEE Conference on CVPR Workshops*, pages 814–819, 2014.

[6] A. Argyriou, T. Evgeniou, and M. Pontil. Multi-task feature learning. In *NIPS*, 2007.

[7] Y. Bengio, A. C. Courville, and P. Vincent. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35:1798–1828, 2013.

[8] E. Bruno and S. Marchand-Maillet. Multiview clustering: A late fusion approach using latent models. *In ACM Conference on Research and Development in Information Retrieval*, pages 736–737, 2009.

[9] K. Chaudhuri, S. M. Kakade, K. Livescu, and K. Sridharan. Multi-view clustering via canonical correlation analysis. *In International Conference on Machine Learning*, pages 129–136, 2009.

[10] W.-Y. Chen, Y. Song, H. Bai, C.-J. Lin, and E. Y. Chang. Parallel spectral clustering in distributed systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(3):568–586, 2011.

[11] T.-S. Chua, J. Tang, R. Hong, H. Li, Z. Luo, and Y.-T. Zheng. Nus-wide: A real-world web image database from national university of singapore. In *ACM Conference on Image and Video Retrieval*, 2009.

[12] E. Grave, G. Obozinski, and F. Bach. Trace lasso: a trace norm regularization for correlated designs. In *NIPS*, pages 2187–2195, 2011.

[13] D. Greene and P. Cunningham. A matrix factorization approach for integrating multiple data views. *In European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 423–438, 2009.

[14] Y. Guo. Convex subspace representation learning from multi-view data. In *AAAI*, 2013.

[15] A. Kumar and H. D. Iii. A co-training approach for multi-view spectral clustering. *In International Conference on Machine Learning*, pages 393–400, 2011.

[16] A. Kumar, P. Rai, and H. D. Iii. Co-regularized multiview spectral clustering. *In Neural Information Processing Systems*, pages 1413–1421, 2011.

[17] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma. Robust recovery of subspace structures by low-rank representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):171–184, 2013.

[18] G. Liu, Z. Lin, and Y. Yu. Robust subspace segmentation by low-rank representation. *In International Conference on Machine Learning*, pages 663–670, 2010.

[19] J. Liu, C. Wang, J. Gao, and J. Han. Multi-view clustering via joint nonnegative matrix factorization. *In SIAM International Conference Data Mining*, pages 252–260, 2013.

[20] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

[21] W. Tang, R. Liu, Z. Su, and J. Zhang. Structure-constrained low-rank representation. *IEEE Transactions on Neural Networks and Learning Systems*, 2014.

[22] H. Wang, F. Nie, and H. Huang. Multi-view clustering and feature learning via structured sparsity. *In International Conference on Machine Learning*, pages 352–360, 2013.

[23] C. Xu, D. Tao, and C. Xu. A survey on multi-view learning. *arXiv preprint arXiv:1304.5634*, 2013.

[24] Y. Zhang, Z. Jiang, and L. S. Davis. Learning structured low-rank representations for image classification. *In Computer Vision and Pattern Recognition*, pages 676–683, 2013.