2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)
April 4-7, 2018, Washington, D.C., USA

MULTIPLE INCOMPLETE VIEWS CLUSTERING VIA NON-NEGATIVE MATRIX FACTORIZATION WITH ITS APPLICATION IN ALZHEIMER'S DISEASE ANALYSIS

Kai Liu[1], *Hua Wang*[1†], *Shannon Risacher*[2], *Andrew Saykin*[2], *Li Shen*[2,3], *for the ADNI*[‡]

[1]Department of Computer Science, Colorado School of Mines, Golden, CO 80401
[2]Radiology and Imaging Sciences, BioHealth Informatics, Indiana University, IN 46202
[3]Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, PA 19104

ABSTRACT

Traditional neuroimaging analysis, such as clustering the data collected for the Alzheimer's disease (AD), usually relies on the data from one single imaging modality. However, recent technology and equipment advancements provide with us opportunities to better analyze diseases, where we could collect and employ the data from different image and genetic modalities that may potentially enhance the predictive performance. To perform better clustering in AD analysis, in this paper we conduct a new study to make use of the data from different modalities/views. To achieve this goal, we propose a simple yet efficient method based on Non-negative Matrix Factorization (NMF) which can not only achieve better prediction performance but also deal with some data missing in some views. Experimental results on the ADNI dataset demonstrate the effectiveness of our proposed method.

Index Terms— Multi-View Clustering, Non-negative Matrix Factorization, Incomplete Views, Alzheimer's Disease (AD)

1. INTRODUCTION

Alzheimer's disease (AD) is a neurodegenerative disorder characterized by progressive impairment of memory and other cognitive functions. Thus, in AD studies there can be three diagnostic groups based on its progress: AD, mild cognitive impairment (MCI), and health control (HC). As a result, early detection and diagnosis of AD have been routinely modeled as a supervised classification problem. Supervised classification models require data labels, *i.e.*, the diagnostic group information of the participants. Statistically speaking, the more labeled data one can have, the better a supervised learning model can be trained. However, because data labeling is costly, developing unsupervised learning models that do not rely on data labels has attracted more and more attentions [1].

On the other hand, with recent developments on technologies, advances in acquiring genome-wide array data and multimodal brain imaging data have made it possible and practical to study the influence of genetic variation on brain structures and functions. Research in this promising field, known as *imaging genetics*, plays an important role in simulating biology mechanism of the brain to better understand complex neurobiological systems, from genetic determinants to the interplay of brain structure, function, behavior and cognition. Study and analysis of such multimodal data may possibly deepen our mechanistic understanding of diseases, facilitate early diagnosis, thus improving the treatment of brain disorders. For example, regional imaging biomarkers obtained from magnetic resonance imaging (MRI), such as the voxel-based morphometry (VBM) [2] features, can characterize the structural variations and those obtained from fluorodeoxyglucose positron emission tomography (FDG-PET) can characterize molecular variations on brains. In addition, genome-wide association studies (GWAS) have been increasingly performed to correlate high-throughput single nucleotide polymorphism (SNP) data to large-scale image data. Recent studies [3, 4] examined these associations at the whole genome and entire brain level, which have shown that the variations on genotypic biomarkers as SNPs are strongly associated with the variations on phenotypic measures. Therefore it is beneficial to take advantage of the multimodal phenotypic and genotypic biomarkers when we study the AD patterns.

To address the above two challenges, in this paper we propose a simple, yet computationally efficient, unsupervised clustering method to detect AD patterns, *i.e.*, the diagnostic group memberships of the participants. Our new method is built under the framework of nonnegative matrix factorization (NMF) [5]. Due to its connection to K-means clustering [6], NMF has been broadly used to solve a variety of clustering problems [5, 7, 8]. In the proposed new method, when

†Correspondence to Hua Wang (huawangcs@gmail.com). This research was partially supported by NSF-IIS 1423591 and NSF-IIS 1652943; NIH R01 EB022574, R01 LM011360, R01 AG19771, U19 AG024904, and P30 AG10133.

‡Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how to apply/ADNI Acknowledgement List.pdf.

978-1-5386-3636-7/18/$31.00 ©2018 IEEE

performing clustering, we learn the centroids, *i.e.*, the representative data points, for each individual data modality separately, and fix the cluster memberships of the data points across different modalities by learning a unified clustering indication collectively. As a result, the representations of a same data point in multiple modalities compensate each other and improved clustering results have been achieved. We performed extensive experiments on the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort and the promising results validated our proposed new method.

2. CLUSTERING AND NMF

Our goal is to cluster the multi-view data into three clusters: AD, MCI and HC. K-means is one of the most popular methods for clustering, and it has been proved that under some circumstance, NMF is equivalent to K-means [6]. Due to the mathematical elegance, NMF has been broadly used for clustering. The error function of traditional NMF [5] is:

$$\min\ J = \|X - FG\|_F^2, \quad s.t. \quad F, G \geq 0, \tag{1}$$

where $X \in \Re^{N \times d}$ is the input data and $\|X\|_F^2 = \sum X_{ij}^2$ is a Frobenius form of a matrix. In this paper and our study, it could be any arbitrary view of data, such as SNP data, VBM data, FDG-PET data, *etc*., with N instances and d features. The solution to J in Eq. 1 is given by the updating F and G as following [5]:

$$F_{ik} \longleftarrow F_{ik}\frac{(XG^T)_{ik}}{(FGG^T)_{ik}}, \quad G_{kj} \longleftarrow G_{kj}\frac{(F^TX)_{kj}}{(F^TFG)_{kj}}. \tag{2}$$

The resulted F gives the centroids of the three clusters including AD, MCI and HC, and the resulted G indicates the clustering membership of input data.

3. OUR NEW METHOD

In this section, we will first formalize the problem. Then the background knowledge on weighted nonnegative matrix factorization will be introduced to motivate our new method.

3.1. Instance Weight Matrix

We first summarize the notations used in this paper in Table 1. Suppose we are given a dataset of N instances with n_v views $\left\{X^1, X^2, ..., X^{n_v}\right\}$ where $X^i \in \Re^{N \times d_i}$ denotes the data from i-th view. We define an indicator matrix $M \in \Re^{n_v \times N}$, where $M_{i,j} = 1$ if j-th instance appears in i-th view, otherwise $M_{i,j} = 0$. It can be verified that when all elements in M is 1, the clustering problem becomes a traditional complete multi-view clustering problem. Traditional multi-view clustering [9] defines the objective function to sum up the loss over different views, inspired by giving different weights to different views, we introduce a diagonal weight matrix

Table 1. Notation Summary

Notation	Description
N	total number of data instances
n_v	total number of views
$X^{(i)}$	data matrix from i-th view
d_i	feature dimension in the i-th view
M	indicator matrix
$W^{(i)}$	weight matrix for the i-th view
$U^{(i)}$	clustering matrix for the i-th view
$V^{(i)}$	feature matrix from i-th view
U^*	consensus clustering matrix
α_i	trade-off parameter for the i-th view

$W^{(i)} \in \Re^{N \times N}$ for each view i by

$$W_{j,j}^{(i)} = \begin{cases} 1 & \text{if } M_{ij} = 1, \\ \frac{\sum_{j=1}^{N} M_{i,j}}{N} & \text{otherwise} \end{cases} \tag{3}$$

$W^{(i)}$ gives lower weights to the missing instances than the presented instances in the i-th view. For different views with different incomplete rates, the weights for missing instances are also different.

3.2. Objective Function

Given the representations of the input data in n_v different modalities $\left\{X^{(1)}, X^{(2)}, ..., X^{(n_v)}\right\}$, for each modality of $X^{(i)}$ we can factorize it as $X^{(i)} \approx U^{(i)} * V^{(i)}$. Thus, a simple objective function to combine multiple incomplete views can be formulated as following:

$$\min\ J = \sum_{i=1}^{n_v} \|W^{(i)}\left(X^{(i)} - U^{(i)}V^{(i)}\right)\|_F^2, \tag{4}$$

The objective function is to minimize the sum of the loss over different views, where $X^{(i)} \in \Re^{N \times d_i}$ is the input data, $U^{(i)} \in \Re^{N \times K}$ contains the centroids of the $K = 3$ clusters, one for each column, $V^{(i)} \in \Re^{K \times d_i}$ indicates the clustering membership of the input data for a specific modality/view. It is worthy to notice that when $n_v = 1$, the multimodal data clustering problem is reduced to traditional single view clustering in Eq. (1); when $W^{(i)}$ is an identity matrix, the objective function solves the complete multi-view clustering.

Because the objective in Eq. (4) can be decoupled into n_v optimization subproblems, the clustering results from multiple views may not be consistent. Thus, we further develop the objective in Eq. (4) to enforce the constraint that the clustering indications in different views should be close to a consensus clustering result, which leads to the following objective:

$$\begin{aligned} \min\ J = \sum_{i=1}^{n_v} \Big(& \|W^{(i)}(X^{(i)} - U^{(i)}V^{(i)})\|_F^2 \\ & + \alpha_i \|W^{(i)}(U^{(i)} - U^*)\|_F^2 \Big). \end{aligned} \tag{5}$$

3.3. Algorithm

To solve the above optimization problem, we derive an iterative update algorithm as summarized in Algorithm 1, whose correctness and convergence can be rigorously proved. Due to space limit, we provide the detailed analysis on the algorithm and its convergence proof in the extended journal version of this paper.

Algorithm 1: Incomplete Multi-view Algorithm

Data: Multi-view data: $\left\{X^{(1)}, X^{(2)}, ..., X^{(n_v)}\right\}$, Indicator Matrix M, $\{\alpha_1, \alpha_2, ..., \alpha_{n_v}\}$,Number of Clusters K.

Result: Clustering matrices: $\left\{U^{(1)}, U^{(2)}, ..., U^{(n_v)}\right\}$, Feature matrices: $\left\{V^{(1)}, V^{(2)}, ..., V^{(n_v)}\right\}$ and consensus clustering matrix U^*.

1. Fill the missing instances in each incomplete view with average feature values.
2. For each view of X, do normalization by feature.
3. Initialize $U^{(i)}$ and $V^{(i)}$ as in [6].
4. For each view i, $\tilde{W}^{(i)} = {W^{(i)}}^T W^{(i)}$.

repeat

5. Fixing $U^{(i)}$ and $V^{(i)}$, update U^* as $\left(\sum_{i=1}^{n_v} \alpha_i \tilde{W}^{(i)}\right)^{-1} \left(\sum_{i=1}^{n_v} \alpha_i \tilde{W}^{(i)} U^{(i)}\right)$.
6. For each view i, update $U^{(i)}$ as $U^{(i)}_{j,k} \leftarrow U^{(i)}_{j,k} \frac{(\tilde{W}^{(i)} X^{(i)} {V^{(i)}}^T + \alpha_i \tilde{W}^{(i)} U^*)_{j,k}}{(\tilde{W}^{(i)} U^{(i)} V^{(i)} {V^{(i)}}^T + \alpha_i \tilde{W}^{(i)} U^{(i)})_{j,k}}$ and $V^{(i)}_{j,k} \leftarrow V^{(i)}_{j,k} \frac{({U^{(i)}}^T \tilde{W}^{(i)} X^{(i)})_{j,k}}{({U^{(i)}}^T \tilde{W}^{(i)} U^{(i)} V^{(i)})_{j,k}}$.

until *Converges*

7. Get the Clustering result from U^*.

4. EXPERIMENTS

In this section, we perform the experiments on the ADNI cohort to demonstrate the effectiveness of our new method.

4.1. AD Dataset

Data used in the preparation of this paper were obtained from the ADNI database[1]. Our goal to use the ADNI data set is to test whether serial MRI, FDG-PET, FreeSurfer and SNP markers can be combined to predict the progress stages of AD. Following the prior imaging genetics study [3], 733 non-Hispanic Caucasian participants were included in this study. In this study, we have include 204 HC, 354 MCI and 175 AD participants. Every participant is described by the above four modalities of imaging and genetic data, some of which have incomplete descriptions. The features in every modality of the involved subject samples are summarized in Table 2.

[1] http://adni.loni.usc.edu/

Table 2. Multimodal feature in multiview learning

View ID (feature set ID)	Modality	No. of features
VBM	MRI	86
FreeSurfer	MRI	56
FDG-PET	FDG-PET	26
SNPs	Genetics	1224

Table 3. Clustering performance in AD Dataset

Method	Accuracy(%)	NMI(%)
ConvexSub	39.2 ± 0.1	33.2 ± 0.3
ConcatNMF	34.1 ± 0.2	27.0 ± 0.1
PVC	42.3 ± 0.1	30.6 ± 0.3
MultiNMF	45.2 ± 0.1	38.7 ± 0.1
Ours	$\mathbf{48.9} \pm 0.2$	$\mathbf{39.5} \pm 0.2$

4.2. Baseline Algorithms

To demonstrate how the clustering performance can be improved by the proposed approach, we compared it against the following algorithms:

ConvexSub: A subspace-based multi-view clustering method is proposed by [10]. In our comparison experiments, we set β to be 1 for all different views.

Feature Concatenation (ConcatNMF): A simple and direct way is to concatenate all the features from different views, and run classical NMF on the concatenated representation.

PVC: [11] proposes the partial multi-view clustering (PVC) method, which deals with incomplete views. The idea in PVC is the instances correspond to the same example in different views should be close to each other. We set the parameter Λ to be 0.01 in the experiments.

MultiNMF: by introducing a consensus clustering matrix and optimizing each matrices, MultiNMF [12] could not only give the clustering for each view but also give the optimized consensus clustering matrix.

The clustering accuracy (AC) and the normalized mutual information (NMI) are used to measure the clustering performance for each method and comparison [13].

4.3. Result

Table 3 shows the clustering performance (including accuracy and NMI) of different algorithms on the AD datasets. In order to randomize the experiments, 20 test runs with different random initializations were conducted and the average performance are reported. Obviously, we new method outperforms its competing counterparts with a clear margin.

We further study the results of different rates of data incompleteness. For this stage, first we only selects the complete data with all four views, and we have 345 subject samples (83 HC, 174 MCI and 88 AD). Then we artificially re-

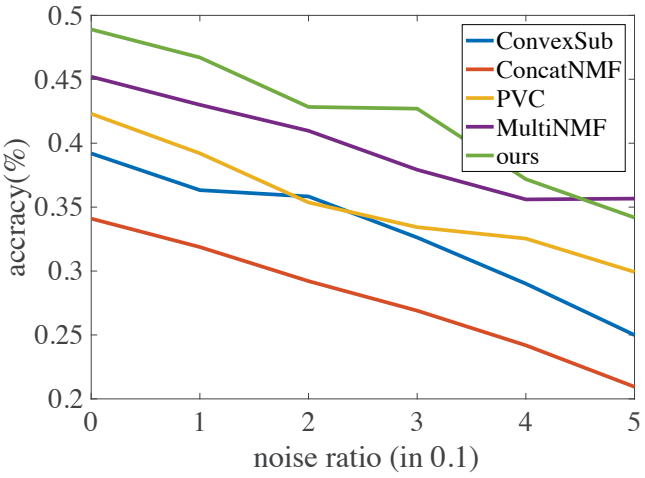

Fig. 1. Clustering Accuracies with different incompleteness rates varies from 0.1 to 0.5, it is obvious that our proposed method has an advantage over other methods

move some views from some data with various rates (from 0 to 0.5 with 0.1 as interval).

5. CONCLUSION

In this paper, we introduced an efficient algorithm for incomplete multi-view clustering based on nonnegative matrix factorization. In order to efficiently learn the underlying clustering structure embedded in multiple views, we require coefficient matrices learned from factorizations of different views to be close by introducing a consensus clustering matrix. Moreover, we propose weight matrices to solve incomplete data problem and give the effective algorithm. We also show that our proposed method converges fast. Experiments on real world AD datasets demonstrate that the proposed method could achieve good clustering accuracy with various incompleteness rates.

6. REFERENCES

[1] Dragan Gamberger, Nada Lavrač, Shantanu Srivatsa, Rudolph E Tanzi, and P Murali Doraiswamy, "Identification of clusters of rapid and slow decliners among subjects at risk for alzheimers disease," *Scientific Reports*, vol. 7, 2017.

[2] Cynthia M Stonnington, Carlton Chu, Stefan Klöppel, Clifford R Jack, John Ashburner, Richard SJ Frackowiak, Alzheimer Disease Neuroimaging Initiative, et al., "Predicting clinical scores from magnetic resonance scans in alzheimer's disease," *Neuroimage*, vol. 51, no. 4, pp. 1405–1413, 2010.

[3] Li Shen, Sungeun Kim, Shannon L Risacher, Kwangsik Nho, Shanker Swaminathan, John D West, Tatiana Foroud, Nathan Pankratz, Jason H Moore, Chantel D Sloan, et al., "Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in mci and ad: A study of the adni cohort," *Neuroimage*, vol. 53, no. 3, pp. 1051–1063, 2010.

[4] Hua Wang, Feiping Nie, and Heng Huang, "Multi-view clustering and feature learning via structured sparsity," in *International Conference on Machine Learning*, 2013, pp. 352–360.

[5] Daniel D Lee and H Sebastian Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788, 1999.

[6] Chris Ding, Xiaofeng He, and Horst D Simon, "On the equivalence of nonnegative matrix factorization and spectral clustering," in *Proceedings of the 2005 SIAM International Conference on Data Mining*. SIAM, 2005, pp. 606–610.

[7] Hua Wang, Heng Huang, Feiping Nie, and Chris Ding, "Cross-language web page classification via dual knowledge transfer using nonnegative matrix tri-factorization," in *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*. ACM, 2011, pp. 933–942.

[8] Hua Wang, Feiping Nie, Heng Huang, and Fillia Makedon, "Fast nonnegative matrix tri-factorization for large-scale data co-clustering," in *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, 2011, vol. 22, p. 1553.

[9] Eric Bruno and Stephane Marchand-Maillet, "Multi-view clustering: A late fusion approach using latent models," in *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2009, pp. 736–737.

[10] Yuhong Guo, "Convex subspace representation learning from multi-view data.," in *AAAI*, 2013, vol. 1, p. 2.

[11] S-YLYJ Zhi and Hua Zhou, "Partial multi-view clustering," in *AAAI Conference on artificial intelligence*, 2014.

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

[13] Wei Xu, Xin Liu, and Yihong Gong, "Document clustering based on non-negative matrix factorization," in *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*. ACM, 2003, pp. 267–273.