Supplementary Material for Adaptive Graph Completion Based Incomplete Multi-view Clustering

Jie Wen, Ke Yan, Zheng Zhang, Yong Xu*, *Senior Member*, *IEEE*, Junqian Wang, Lunke Fei, Bob Zhang

I. Time complexity analysis

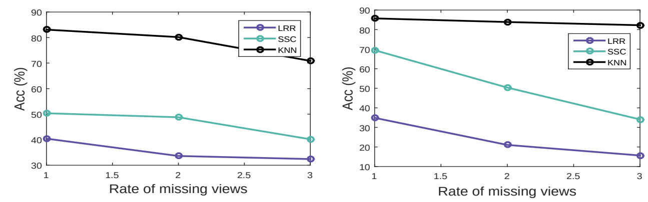
The computational complexities and running times of different incomplete multi-view clustering (IMC) methods are listed in Table I. All methods are implemented on the computer with Win10 system, Inter i7-7700 CPU and 32GB RAM. From this table, we can observe that although the proposed method has a higher computational complexity than the other methods, its practical running times are comparative to those of MIC and OMVC. In addition, OPIMC is the most efficient IMC method, but its performance is far below our method. For instance, on the Animal and 3Sources datasets, the clustering accuracies obtained by our method are about 7%-11% and 17%-25% higher than those of OPIMC, respectively. Besides this, on the 3Sources dataset which has high feature dimensions, the proposed method is more efficient than DAIMC.

TABLE I
RUNNING TIME (SECONDS) OF DIFFERENT METHODS ON THE ABOVE FIVE DATASETS WITH A MISSING-VIEW RATE OR PAIRED-VIEW RATE OF 30%. 'T' DENOTES THE ITERATIONS OF INNER LOOP. m_{max} DENOTES THE LARGEST DIMENSIONALITY OF VIEWS. 't' DENOTES THE ITERATIONS OF k-MEANS CLUSTERING. 'd' IS THE FEATURE DIMENSION OF THE COMMON REPRESENTATION OR CLUSTER NUMBER. n_v IS THE NUMBER OF AVAILABLE INSTANCES IN THE vTH VIEW.

	Running time (seconds)					
Methods	BBCSport	3Sources	Handwritten	Caltech20	Animal	Computational complexity
BSV	0.949	2.574	1.155	73.232	7.425×10^3	$O\left(t\sum_{k=1}^{l} m_k^2 n\right)$
Concat	1.092	2.994	1.355	46.680	2.809×10^3	$O\left(tn\left(\sum_{k=1}^{l} m_k\right)^2\right)$
GPMVC	5.269	6.696	107.871	275.442	234.019	$O\left(\tau\sum_{k=1}^{l} \max\left(m_k, n_k\right) n_k d + tnd^2\right)$
MIC	5.018	9.560	812.353	2.605×10^3	9.079×10^3	$O\left(\tau T\sum_{k=1}^{l} m_k dn + tnd^2\right)$
DAIMC	224.416	807.385	34.309	241.437	5.043×10^3	$O\left(\tau\left(Tndm_{\max} + lm_{\max}^3\right) + tnd^2\right)$
OMVC	5.577	23.097	730.635	1.059×10^3	9.381×10^4	$O\left(\tau T\sum_{k=1}^{v} m_k cn + tnd^2\right)$
OPIMC	0.036	0.062	0.113	0.488	6.255	$O\left(\tau lcnm_{\max}\right)$
Ours	5.521	15.034	1.056×10^3	1.755×10^3	7.998×10^4	$O\left(\tau dn^2 + tnd^2\right)$

II. Analysis of the pre-constructed graphs

In this section, we mainly consider three kinds of popular graphs, *i.e.*, binary KNN-graph constructed via K-nearest-neighbor (KNN) approach, LRR-graph constructed via the low-rank representation (LRR) [1], and SSC-graph adopted in the sparse subspace clustering (SSC) method [2]. We conduct experiments to analyze which graph is more suitable to our IMC model. Fig.1 shows the experimental results of the proposed method with the above three graphs on the BBCSport and Handwritten datasets. It is obvious that using KNN-graph can obtain a better clustering performance than LRR-graph and SSC-graph, which demonstrates that KNN-graph is more suitable to our method for IMC tasks.



(a) BBCSport (b) Handwritten

Fig. 1. Clustering Acc (%) of the proposed method with different pre-constructed graphs on the BBCSport and Handwritten datasets.

III. Proof of proposition 1

To prove proposition 1, we first introduce the following two interesting theorems, *i.e.*, **Theorem 1** and **Theorem 2**.

Theorem 1: For the vth view, suppose $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of Laplacian matrix $L_{S^{(v)}}$, where the eigenvalues are arranged as $0 \le \lambda_1 \le \lambda_2 \le \ldots \le \lambda_n$. For any q ($q \le n$) orthonormal vectors h_i ($i = 1, \ldots, q$), the inequality $\sum_{i=1}^{q} h_i^T L_{S^{(v)}} h_i \ge \sum_{i=1}^{q} \lambda_i$ holds [3, 4].

Proof: Let $\psi_1, \psi_2, \ldots, \psi_n$ be the orthonormal eigenvectors of $L_{S^{(v)}}$ and satisfy $L_{S^{(v)}}\psi_i = \lambda_i \psi_i$. For each j, we have:

$$
\begin{aligned}
& h_j^T L_{S^{(v)}} h_j \\
= & h_j^T \left(\sum_{i=1}^{n} \lambda_i \psi_i \psi_i^T\right) h_j \\
= & \sum_{i=1}^{n} \lambda_i \left(h_j^T \psi_i\right)^2 \\
= & \sum_{i=1}^{n} \left(\lambda_i - \lambda_q\right)\left(h_j^T \psi_i\right)^2 + \lambda_q \sum_{i=1}^{n} \left(h_j^T \psi_i\right)^2 \\
= & \sum_{i=1}^{q} \left(\lambda_i - \lambda_q\right)\left(h_j^T \psi_i\right)^2 + \sum_{i=q+1}^{n} \left(\lambda_i - \lambda_q\right)\left(h_j^T \psi_i\right)^2 \\
& + \lambda_q \sum_{i=1}^{n} \left(h_j^T \psi_i\right)^2
\end{aligned}
\tag{1}
$$

For orthonormal vectors ψ_i and h_j, it is obvious that $\sum_{i=1}^{q}\left(h_j^T \psi_i\right)^2 \le \sum_{i=1}^{n}\left(h_j^T \psi_i\right)^2 = 1$ holds. Therefore, (1) can be further transformed as:

$$
\begin{aligned}
& h_j^T L_{S^{(v)}} h_j \\
= & \sum_{i=1}^{q} \left(\lambda_i - \lambda_q\right) \left(h_j^T \psi_i\right)^2 + \sum_{i=q+1}^{n} \left(\lambda_i - \lambda_q\right) \left(h_j^T \psi_i\right)^2 + \lambda_q \\
\geq & \sum_{i=1}^{q} \left(\lambda_i - \lambda_q\right) \left(h_j^T \psi_i\right)^2 + \lambda_q
\end{aligned}
\tag{2}
$$

According to (2), we have

$$
\begin{aligned}
& \sum_{i=1}^{q} \lambda_i - \sum_{j=1}^{q} h_j^T L_{S^{(v)}} h_j \\
\leq & \sum_{i=1}^{q} \left(\lambda_i - \lambda_q\right) \left(1 - \sum_{j=1}^{q} \left(h_j^T \psi_i\right)^2\right) \leq 0
\end{aligned}
\tag{3}
$$

From (3), we obtain $\sum\limits_{i=1}^{q} \lambda_i \leq \sum\limits_{j=1}^{q} h_j^T L_{S^{(v)}} h_j$. Thus **Theorem 1** is proved.

As demonstrated in many works, the following **Theorem 2** exists in the graph theory [5-7].

Theorem 2: For a non-negative graph $S^{(v)} \in R^{n \times n}$, the number of connected components is equal to the multiplicity of the eigenvalue zero of its Laplacian matrix $L_{S^{(v)}}$, where the Laplacian matrix $L_{S^{(v)}} = D_{S^{(v)}} - \left(S^{(v)T} + S^{(v)}\right)\Big/2$ and the ith diagonal element of $D_{S^{(v)}}$ is $\left(D_{S^{(v)}}\right)_{i,i} = \sum\limits_{j=1}^{n} \left(S_{i,j}^{(v)} + S_{j,i}^{(v)}\right)\Big/2$.

Please refer to [7] for the detailed proof process for **Theorem 2**.

Theorem 1 indicates that by minimizing the objective sub-term $Tr\left(U^T L_{S^{(v)}} U\right)$ until $\sum\limits_{i=1}^{c} \lambda_i = 0$, we can obtain a graph $S^{(v)}$ whose Laplacian matrix satisfies $rank\left(L_{S^{(v)}}\right) = n - c$, where $rank\left(L_{S^{(v)}}\right)$ denotes the rank of matrix $L_{S^{(v)}}$. Since $rank\left(L_{S^{(v)}}\right) = n - c$ means that Laplacian matrix $L_{S^{(v)}}$ has c zero eigenvalues, combing **Theorem 2**, we can conclude that minimizing the objective sub-term $Tr\left(U^T L_{S^{(v)}} U\right)$ of (#9)[1] until $\sum\limits_{i=1}^{c} \lambda_i = 0$ enables graph $S^{(v)}$ to have exactly c connected components.

Overall, we have completed the proof of proposition 1.

References

[1] 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*, vol. 35, no. 1, pp. 171–184, 2012.

[2] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2765–2781, 2013.

[3] K. Zhan, F. Nie, J. Wang, and Y. Yang, "Multiview consensus graph clustering," *IEEE Transactions on Image Processing*, vol. 28, no. 3, pp. 1261–1270, 2018.

[4] K. Fan, "On a theorem of weyl concerning eigenvalues of linear transformations i," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 35, no. 11, pp. 652–655, 1949.

[5] F. Nie, X. Wang, M. I. Jordan, and H. Huang, "The constrained laplacian rank algorithm for graph-based clustering," in *AAAI Conference on Artificial Intelligence*, 2016, pp. 1969–1976.

[6] B. Mohar, Y. Alavi, G. Chartrand, and O. Oellermann, "The laplacian spectrum of graphs," *Graph Theory, Combinatorics, and Applications*, vol. 2, no. 12, pp. 871–898, 1991.

[7] D. Cvetkovic, S. Simic, and P. Rowlinson, *An introduction to the theory of graph spectra*. Cambridge University Press, 2009.

[1] The numbers of equations with the mark '#' refer to the equations in the main manuscript.