Multi-view Clustering using Barycentric Coordinate Representation

Xiaotong Qian, Lili Jin, Guénaël Cabanes, Parisa Rastin, Nistor Grozavu

Multi-view Clustering using Barycentric Coordinate Representation

Xiaotong QIAN [1] Lili JIN [2] Guénaël CABANES [2] Parisa RASTIN [3] Nistor GROZAVU [1]

[1]ETIS, UMR 8051, CY Cergy Paris Université, [2]LIPN, UMR 7030, Université Sorbonne Paris Nord, [3]LORIA, UMR 7503, Université de lorraine

Purpose

- Work on multiple views
- Reduce a significant time loss
- Update weights of different views

Multi-view clustering

Clustering – Data analysis technique that groups similar objects together to uncover patterns and gain insights from complex datasets. e.g. KMeans

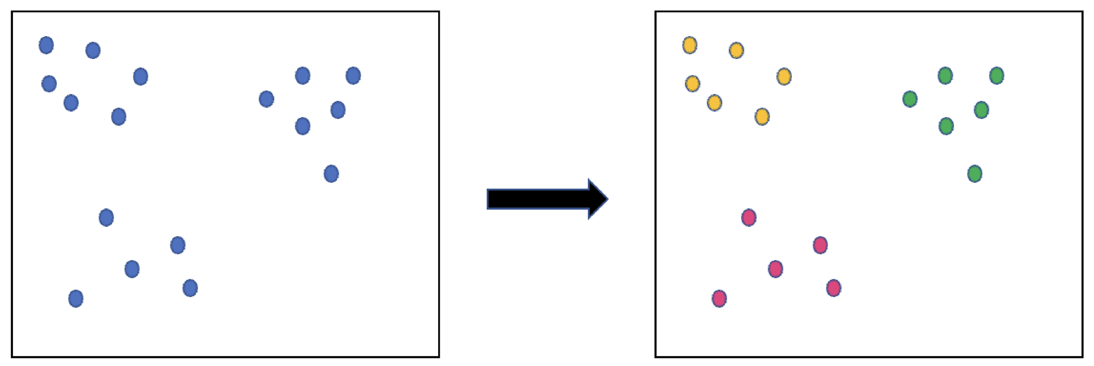


Figure 1. Example of Clustering

Multi-view data/clustering – Looking at an object or dataset from multiple perspectives. By integrating information from different views, multi-view clustering improves the accuracy and robustness of clustering algorithms, enabling a more complete understanding of the data.

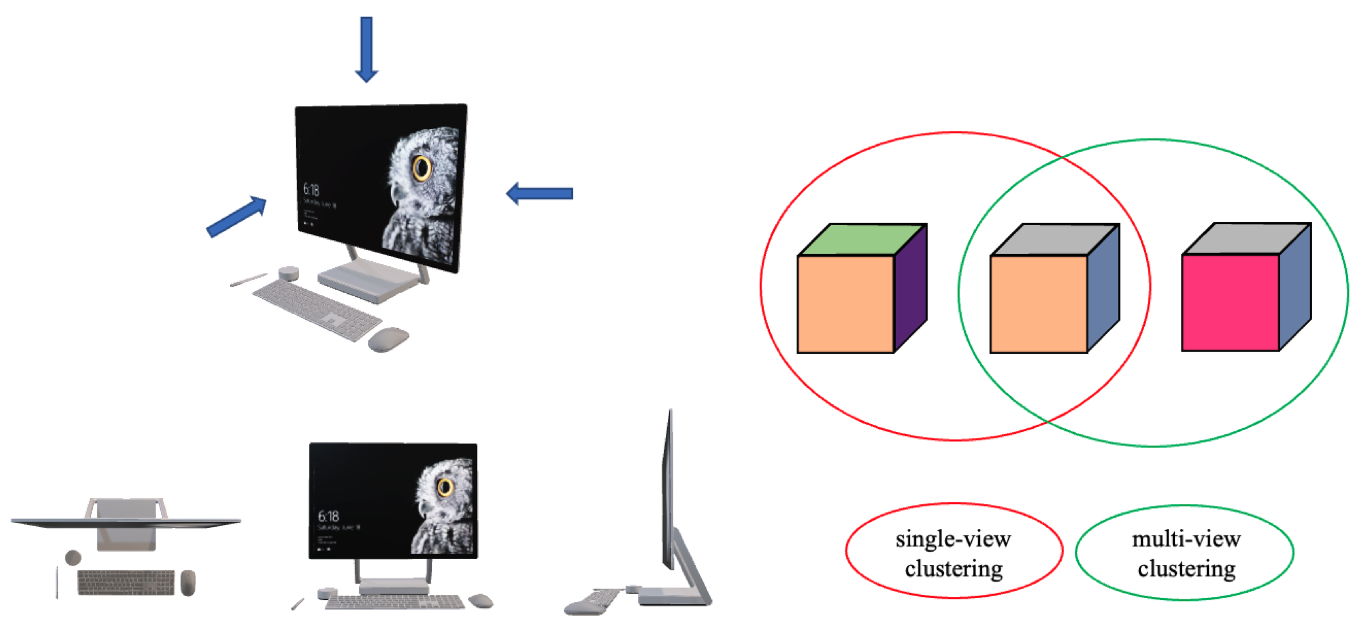


Figure 2. Example of Multi-view data/clustering

RMKMC – Robust Multi-View K-means Clustering [1]

The RMKMC model aims to minimize the weighted distance between each point and its cluster centers across all views, resulting in a global communal cluster indicator matrix G. The weights of the views can be automatically updated using a single parameter to control the weight distribution.

$$\min_{C^{(v)},G,\alpha^{(v)}} \sum_{v=1}^{V} (\alpha^{(v)})^{\gamma} ||X^{(v)} - GC^{(v)}||_{2,1}$$

BC representation

A recent research [2] proposed that a set of fixed support points are chosen among the objects of the dataset and be used as a basis for the definition of a representation space, in order to find barycentric coordinates (BC) of data points in that space, then apply prototype-based clustering on them.

Illustration:

- triangle : $(\lambda_1 + \lambda_2 + \lambda_3) * X = \lambda_1 * A + \lambda_2 * B + \lambda_3 * C$
- p-vertices : $(\lambda_1 +, ..., +\lambda_p) * X = \lambda_1 * S^1 +, ..., +\lambda_p * S^p$

where $(\lambda_1, ..., \lambda_p)$ is called barycentric coordinate representation with respect to the support points $(S^1, ..., S^p)$.

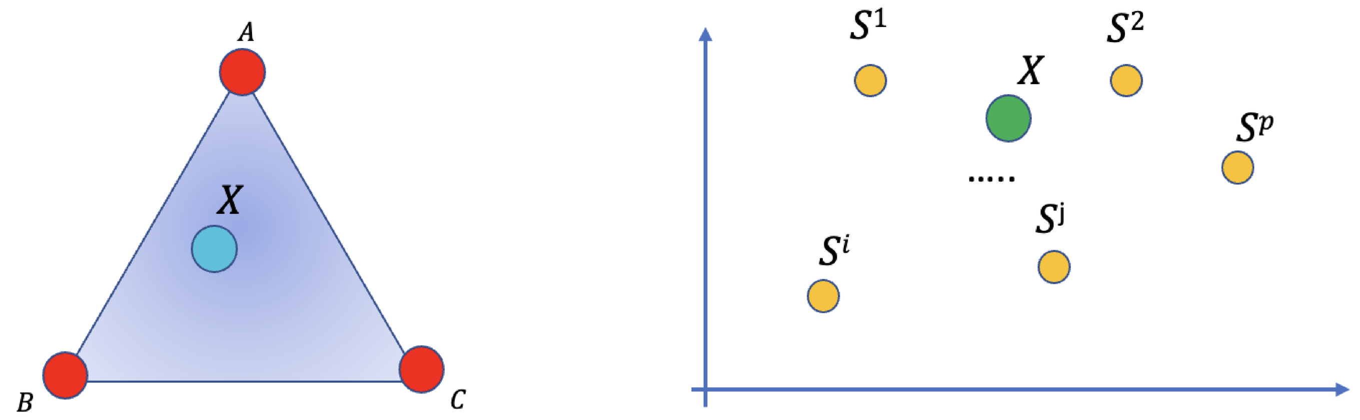


Figure 3. Define barycentric coordinate from triangle to p-vertices

Normalized barycentric coordinates:

$$X = \frac{\lambda_1 * S^1 +, ..., +\lambda_p * S^p}{\lambda_1 +, ..., +\lambda_p} = \sum_{i=1}^{p} \beta_i S^i \text{ with } \sum_{i=1}^{p} \beta_i = 1$$

Finding β:

$$A = \begin{pmatrix} d(s^1,s^1) - d(s^2,s^1) & ... & d(s^1,s^P) - d(s^2,s^P) \\ ... & ... & ... \\ d(s^1,s^1) - d(s^P,s^1) & ... & d(s^1,s^P) - d(s^P,s^P) \\ 1 & ... & 1 \end{pmatrix}, M^n = \begin{pmatrix} d(x^n,s^1) - d(x^n,s^2) \\ ... \\ d(x^n,s^1) - d(x^n,s^P) \\ 1 \end{pmatrix}$$

$$A * \beta^n = M^n => \beta^n = A^{-1} * M^n$$

Distance d^2 between X and a cluster center μ:

$$d^2(x^n, \mu^k) = -\frac{1}{2}(\beta^n - \beta^k)^T * D_s * (\beta^n - \beta^k),$$

where D_s is the dissimilarity matrix between the support points.

References

[1] Xiao Cai, Feiping Nie, and Heng Huang. Multi-view k-means clustering on big data. In *Twenty-Third International Joint conference on artificial intelligence*, 2013.

[2] Parisa Rastin, Guénaël Cabanes, Basarab Matei, Younès Bennani, and Jean-Marc Marty. A new sparse representation learning of complex data: Application to dynamic clustering of web navigation. *Pattern Recognition*, 91:291–307, 2019.

BCmvlearn

Objective function:

$$J = \min_{\mu^{(v)},G,\alpha^{(v)}} \sum_{v=1}^{V} (\alpha^{(v)})^{\gamma} \sum_{n=1}^{N} d^2((x^n)^{(v)}, \mu^{(v)} G_n^T) = \min_{(\beta^{\mu})^{(v)},G,\alpha^{(v)}} -\frac{1}{2} \sum_{v=1}^{V} (\alpha^{(v)})^{\gamma} H^{(v)}$$

where $H^{(v)} = \sum_{n=1}^{N} \Phi_n^T D_s^{(v)} \Phi_n$ with $\Phi_n = (\beta^n)^{(v)} - (\beta^{\mu})^{(v)} G_n^T$

Optimization:

- Updating $(\beta^{\mu})^{(v)}$: $(\beta^{\mu})^{(v)} = \sum_{n=1}^{N} (\beta^n)^{(v)} (\sum_{n=1}^{N} G_n^T G_n)^{-1}$
- Updating G :

$$i = \underset{k \in 1,...,K}{argmin} \left\{ -\frac{1}{2} \sum_{v=1}^{V} (\alpha^{(v)})^{\gamma} \Psi_k^T D_s^{(v)} \Psi_k \right\}$$

where $\Psi_k = (\beta^n)^{(v)} - (\beta^{\mu})_k^{(v)}$, $G_{ni} = 1$ and $G_{nj} = 0$ where $j \in \{1, ..., k\}, i \neq j$

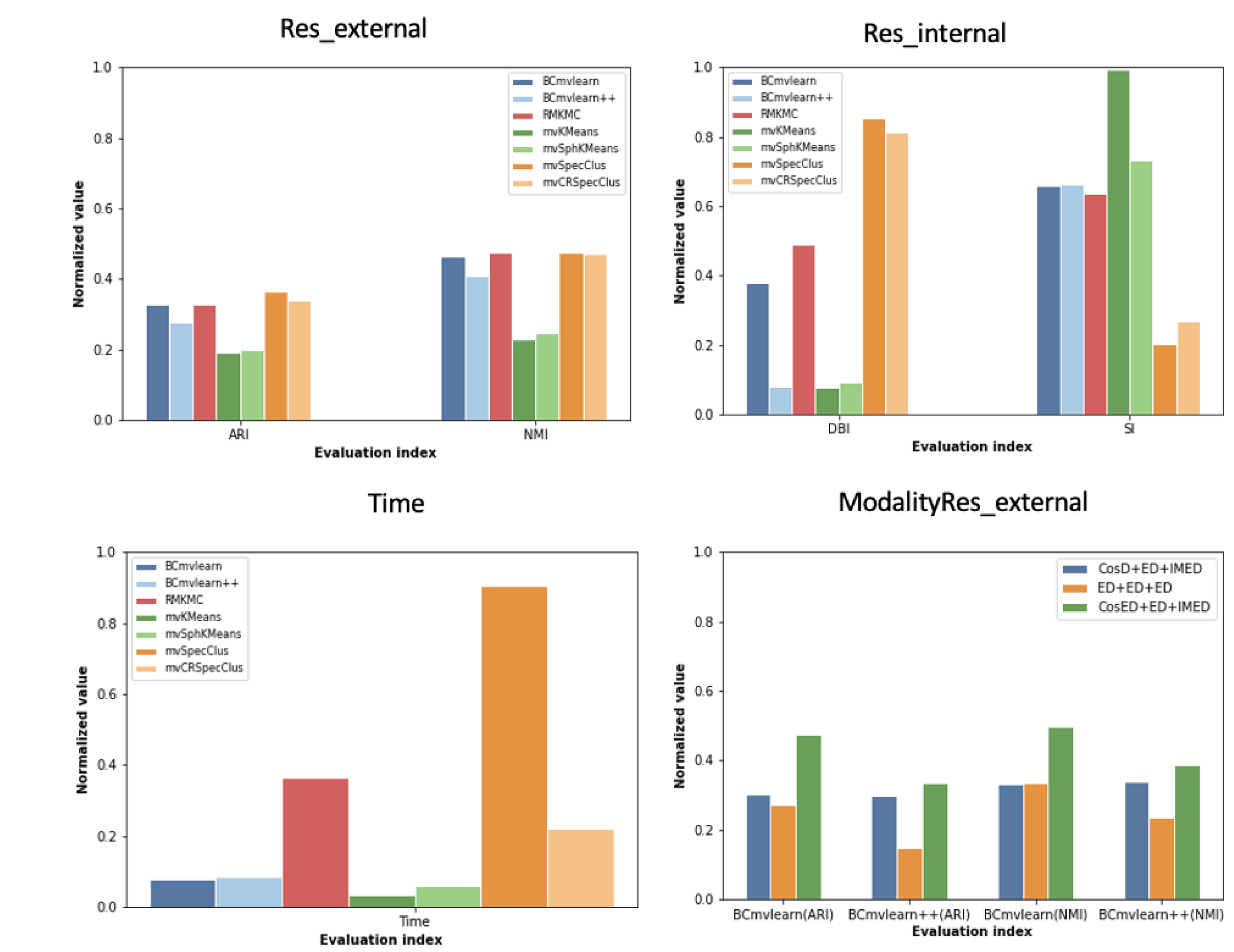
- Updating $\alpha^{(v)}$:

$$\alpha^{(v)} = \frac{(\gamma H^{(v)})^{\frac{1}{1-\gamma}}}{\sum_{v=1}^{V} (\gamma (H^{(v)})^{\frac{1}{1-\gamma}}}$$

Alternative version:

BCmvlearn++ improves initialization by avoiding poor center placement.

Experiment Results



The experimental results on different evaluation metrics show that the proposed approach is competitive with other state-of-the-art multi-view algorithms in terms of quality and time complexity. And also could handle different distance metrics for application to multimodal clustering.

Gold Coast Convention and Exhibition Centre, Queensland, Australia, June 18 - 23, 2023

xiaotong.qian@ensea.fr