Multiple Medoids based Multi-view Relational Fuzzy Clustering with Minimax Optimization

Yangtao Wang[1], Lihui Chen[2], Xiaoli Li[1]

[1]Institude for Infocomm Research(I2R), A*STAR, Singapore

[2]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore

{wangyt, xlli}@i2r.a-star.edu.sg, ELHCHEN@ntu.edu.sg

Abstract

Multi-view data becomes prevalent nowadays because more and more data can be collected from various sources. Each data set may be described by different set of features, hence forms a multi-view data set or multi-view data in short. To find the underlying pattern embedded in an unlabelled multi-view data, many multi-view clustering approaches have been proposed. Fuzzy clustering in which a data object can belong to several clusters with different memberships is widely used in many applications. However, in most of the fuzzy clustering approaches, a single center or medoid is considered as the representative of each cluster in the end of clustering process. This may not be sufficient to ensure accurate data analysis. In this paper, a new multi-view fuzzy clustering approach based on multiple medoids and minimax optimization called M4-FC for relational data is proposed. In M4-FC, every object is considered as a medoid candidate with a weight. The higher the weight is, the more likely the object is chosen as the final medoid. In the end of clustering process, there may be more than one medoid in each cluster. Moreover, minimax optimization is applied to find consensus clustering results of different views with its set of features. Extensive experimental studies on several multi-view data sets including real world image and document data sets demonstrate that M4-FC not only outperforms single medoid based multi-view fuzzy clustering approach, but also performs better than existing multi-view relational clustering approaches.

1 Introduction

Multi-view data becomes prevalent nowadays because large amount of data can be easily collected from different sources or represented by different features. Mining valuable information in the prevalent multi-view data is very important for different parties nowadays. Clustering as an excellent tool of data analysis to find the structure of pattern and information underlining the unlabelled data has been widely studied in different areas including artificial intelligence, data mining and machine learning. Based on various theories and applications, different clustering algorithms have been developed to handle different types of data [Jain, 2010; Filippone *et al.*, 2008; Xu and Wunsch, 2005]. As real world data sets may not be well separated, fuzzy clustering algorithms in which each object can be assigned to different clusters with different memberships are widely used in many applications. Two main types of data processed by clustering algorithms are feature data and relational data which are represented as feature vectors and relational (distance) matrix. Relational data is prevalent as it is a more generic data representation in many real world applications. Many clustering approaches have been proposed to handle these two kinds of data. For example, two popular fuzzy clustering approaches Fuzzy c medoids (FCMD)[Krishnapuram *et al.*, 2001] and Fuzzy c means (FCM)[Bezdek, 1981], handle relational data and feature data respectively. Note that FCMD can also be applied on feature data because it can be transferred to relational data easily by calculating the distance of each pair of feature vectors, while feature vector-oriented FCM can only handle feature vector data and can not handle relational data directly. Different multi-view clustering approaches based on FCM and FCMD are also proposed in the literature. However, single center and single medoid is considered for each cluster in FCM and FCMD respectively, which may not be sufficient for various data sets. In this paper, a new multi-view relational fuzzy clustering approach called M4-FC is proposed which is able to achieve the fuzzy partition of objects based on multiple distance matrices. Two mechanisms are introduced into M4-FC to better use the information of different views and improve the clustering performance. First, each object is considered as a medoid candidate by introducing a representative weight to represent the degree of representativeness of each object to its cluster. The higher the weight is, the more probable this object can be considered as a medoid. Second, minimax optimization in which the maximum summation of weighted cost of different views are minimized is applied to find more consensus clustering results of different views.

Several multi-view fuzzy clustering approaches have been proposed in the literature. For example, two FCM based multi-view fuzzy clustering CoFKM [Cleuziou *et al.*, 2009] and WV-Co-FCM [Jiang *et al.*, 2015] are developed. They

use the same two-step strategy. In the first step, the membership of each view is achieved by minimizing the FCM related objective function and the disagreements between different views. Then the extra step is needed to get the final consensus fuzzy membership based on the membership of all views. From an optimization point of view, both CoFKM and WV-Co-FCM formulated the multi-view clustering to a minimization problem. Another optimization method used to formulate multi-view clustering is minimax optimization. Minimax optimization is used in multi-objective optimization to help making decision rules. For multi-view clustering, clustering each view can be considered as one objective and minimax optimization is subsequently used to find consensus clustering results of different views. Recently a multi-view fuzzy clustering based on minimax optimization called MinimaxFCM has been proposed in [Wang and Chen, 2017]. The superior experimental results have shown the effectiveness of minimax optimization on multi-view clustering. Note that CoFKM, WV-Co-FCM and MinimaxFCM can only handle feature data. In order to handle relational data, a multi-view relational fuzzy clustering approach called MVFCMddV based on FCMD is proposed in [de Carvalho *et al.*, 2015]. However, single medoid is identified for each cluster in MVFCMddV which may not be sufficient to capture the pattern structure of various data sets.

In this paper, we propose a new multiple medoids based multi-view fuzzy clustering (M4-FC) with minimax optimization to handle multi-view relational data. Different from MVFCMddV in which a single medoid is identified for each cluster of each view, M4-FC is able to identify multiple medoids to better capture the underlying cluster structure by introducing a representative weight for each object. Inspired by [Wang and Chen, 2017], minimax optimization is adopted in M4-FC to integrate different views to help to generate better consensus clustering results. The experimental results on different real world data sets including image and document data sets show that M4-FC not only outperforms single medoid based multi-view fuzzy clustering approach, but also performs better than other multi-view clustering approaches on all the benchmark data sets.

The rest of the paper is organized as follows: in the next section, some related fuzzy clustering approaches are reviewed. In section 3, the details of the proposed multi-view clustering approach M4-FC are presented. Experiments on five real world multi-view data sets are conducted and the results are analyzed in section 4. Finally, conclusions are drawn in section 5.

2 Related Work

Due to the limited space, we only review some closely related fuzzy clustering based approaches for relational data.

2.1 Fuzzy C Medoids (FCMD)

FCMD [Krishnapuram *et al.*, 2001] is one of the most popular fuzzy clustering approaches for relational data. The following objective function is minimized in FCMD algorithm:

$$J_{FCMD} = \sum_{c=1}^{K} \sum_{i=1}^{N} u_{ci}^m Dis(x_i, \delta_c) \qquad (1)$$

Where K is the cluster number, u_{ci} is the membership degree of object i to cluster c, $m > 1$ is the fuzzification constant called fuzzifier, and δ_c is the medoid (a data object) for cluster c. $Dis(x_i, v_c) = \|x_i - \delta_c\|$ is the distance between object x_i and medoid δ_c. FCMD can generate a fuzzy partition of the data set based on its distance matrix and identify single medoid for each cluster.

2.2 MVFCMddV

MVFCMddV [de Carvalho *et al.*, 2015] is the multi-view extension of the single view FCMD for multi-view relational data clustering. The relevance weights are introduced for each cluster of each view in the objective function as follows:

$$J_{MVFCMddV} = \sum_{c=1}^{K} \sum_{i=1}^{N} u_{ci}^m \sum_{p=1}^{P} w_{cp} Dis^p(x_i, \delta_c^p) \qquad (2)$$

Where u_{ci} is consensus membership of multiple views for object i in fuzzy cluster c, w_{cp} is relevance weight of distance matrix Dis^p of p-th view on cluster c, and $Dis^p(x_i, \delta_c^p)$ is the distance between object x_i and medoid δ_c^p of cluster c under p-th view. By minimizing the weighted summation of the distance as shown in the objective function, MVFCMddV is able to achieve the consensus fuzzy partition and a medoid for each cluster under each view.

As discussed above, FCMD and MVFCMddV identified only one medoid for each cluster. In our method, all objects are considered as medoid candidates by introducing a weight for each object. The higher the weight is, the more probable this object can be considered as a medoid. Moreover, minimax optimization is applied to find consensus clustering results of different views.

3 The Proposed Approach

In this section, we formulate the objective function of the proposed M4-FC by integrating two types of mechanisms: introduction of the representative weight for each object in a cluster and adoption of minimax optimization to find consensus clustering results. In M4-FC, the consensus clustering results are generated based on minimax optimization in which the maximum summation of weighted cost of different views are minimized. Moreover, the weight of each view can be learned automatically in the clustering process.

3.1 Formulation

First the following objective function is formulated for M4-FC J_{M4-FC}:

$$J_{M4-FC} = \min_{U^*, \{V^{(p)}\}_{p=1}^{P}} \max_{\{\alpha^{(p)}\}_{p=1}^{P}} \sum_{p=1}^{P} (\alpha^{(p)})^{\gamma} Q^{(p)} \qquad (3)$$

where

$$Q^{(p)} = \sum_{c=1}^{K} \sum_{i=1}^{N} \sum_{j=1}^{N} u_{ci}^* v_{cj}^{(p)} r_{ij}^{(p)} + \frac{T_u}{2} \sum_{c=1}^{K} \sum_{i=1}^{N} (u_{ci}^*)^2$$
$$+ \frac{T_v}{2} \sum_{c=1}^{K} \sum_{j=1}^{N} (v_{cj}^{(p)})^2 \qquad (4)$$

subject to

$$\sum_{c=1}^{K} u_{ci}^* = 1, \text{ for } i = 1, 2, ..., N \quad (5)$$

$$u_{ci}^* \geq 0, \text{ for } c = 1, 2, ..., K, i = 1, 2, ...N \quad (6)$$

$$\sum_{j=1}^{N} v_{cj}^{(p)} = 1, \text{ for } c = 1, 2, ..., K \quad (7)$$

$$v_{cj}^{(p)} \geq 0, \text{ for } c = 1, 2, ..., K, j = 1, 2, ...N \quad (8)$$

$$\sum_{p=1}^{P} \alpha^{(p)} = 1 \quad (9)$$

$$\alpha^{(p)} \geq 0, \text{ for } p = 1, 2, ..., P \quad (10)$$

In J_{M4-FC}, $r_{ij}^{(p)}$ is the distance between object i and object j in p-th view which is the input. U^* is the $K \times N$ membership matrix whose element in row c and column i is the consensus membership u_{ci}^* for object i of cluster c. $V^{(p)}$ is the $K \times N$ representative weight matrix whose element in row c and column j is the representative weight $v_{cj}^{(p)}$. $v_{cj}^{(p)}$ represents how well an object j can be considered as a medoid for cluster c of p-th view. When weight is higher, this object is more probable to be considered as a medoid. $Q^{(p)}$ is the cost of p-th view and $(\alpha^{(p)})^\gamma$ is the weight of p-th view. The parameter $\gamma \in [0, 1)$ controls the distribution of weights $(\alpha^{(p)})^\gamma$ for different views.

The cost of p-th view $Q^{(p)}$ is composed of three parts. The first part is the distance summation of all clusters weighted by membership and representative weight. The other two parts are applied as regularization terms which are penalties on u_{ci}^* and $v_{cj}^{(p)}$. $T_u > 0$ and $T_v > 0$ control the weight of the two parts in the objective function. As shown in (3), the summation of weighted cost is maximized with respect to the weight $(\alpha^{(p)})^\gamma$ at first. In this step, the view with higher cost will be given higher weight. Then the maximum of the summation of weighted cost is minimized with respect to consensus membership U^* and $V^{(p)}$. In this way, the consensus membership is generated by minimizing the cost under the worst case (maximum summation of weighted cost) to achieve harmonic consensus clustering results.

3.2 Optimization

The clustering goal is to conduct a minimax optimization on the objective function $\sum_{p=1}^{P} (\alpha^{(p)})^\gamma Q^{(p)}$, and subject to the constraints in (5), (6), (7), (8), (9) and (10). It is difficult to solve the variables u_{ci}^*, $v_{cj}^{(p)}$ and $\alpha^{(p)}$ in (3) directly because (3) is nonconvex. Therefore, the alternative optimizaiton (AO) is used to solve the optimization problem by solving one variable with others fixed.

Minimization: Fixing $v_{cj}^{(p)}$, $\alpha^{(p)}$ and Updating u_{ci}^*

The updating rule of u_{ci}^* can be derived based on the Lagrangian Multiplier method and KKT conditions as follows:

$$u_{ci}^* = \begin{cases} 0 & for \quad c \in q^- \\ \dfrac{1}{|q^+|} - \dfrac{1}{T_u \sum_{p=1}^{P} (\alpha^{(p)})^\gamma} (\sum_{p=1}^{P} \sum_{j=1}^{N} (\alpha^{(p)})^\gamma v_{cj}^{(p)} r_{ij}^{(p)} \\ \quad - \dfrac{1}{|q^+|} \sum_{p=1}^{P} \sum_{f \in q^+} \sum_{j=1}^{N} (\alpha^{(p)})^\gamma v_{fj}^{(p)} r_{ij}^{(p)}) \, for \quad c \in q^+ \end{cases} \quad (11)$$

$$\begin{aligned} q^- &= \{c : u_{ci}^* = 0\} \\ q^+ &= \{c : u_{ci}^* > 0\} \end{aligned} \quad (12)$$

Here, q^- is cluster set in which the membership of object i is zero to the clusters. q^+ is cluster set in which the membership of object i is larger than zero. And $|q^+|$, $|q^-|$ are the number of clusters in the set. The sets q^+ and q^- which are updated in each iteration are determined by a method similar to the one used in [Mei and Chen, 2010] which is outlined as follows.

Procedure for determining q^+ and q^-

1 Initialize $q_0^+ = \emptyset, q_0^- = 1, 2, ..., K, s = 0$;

2 $s \leftarrow s + 1, q_s^+ = q_{s-1}^+ + \{m\}, q_s^- = q_{s-1}^- - \{m\}$;

where $m = arg\,min_{c \in q_{s-1}^-} \{\sum_{p=1}^{P} \sum_{j=1}^{N} (\alpha^{(p)})^\gamma v_{cj}^{(p)} r_{ij}^{(p)}\}$;

3 Check whether $u_{fi} > 0$ computed by (11),

where $f = arg\,max_{c \in q_s^+} \{\sum_{p=1}^{P} \sum_{j=1}^{N} (\alpha^{(p)})^\gamma v_{cj}^{(p)} r_{ij}^{(p)}\}$.

If yes, go to step 2, else set $q^+ = q_{s-1}^+, q^- = q_{s-1}^-$ and terminate.

Minimization: Fixing u_{ci}^*, $\alpha^{(p)}$ and Updating $v_{cj}^{(p)}$

The $v_{cj}^{(p)}$ is updated as follows:

$$v_{cj}^{(p)} = \begin{cases} 0 & for \quad j \in m^- \\ \dfrac{1}{|m^+|} - \dfrac{1}{T_v} (\sum_{i=1}^{N} u_{ci}^* r_{ij}^{(p)} - \dfrac{1}{|m^+|} \sum_{f \in m^+} \sum_{i=1}^{N} u_{ci}^* r_{if}^{(p)}) \\ & for \quad j \in m^+ \end{cases} \quad (13)$$

where

$$\begin{aligned} m^- &= \{j : v_{cj}^{(p)} = 0\} \\ m^+ &= \{j : v_{cj}^{(p)} > 0\} \end{aligned} \quad (14)$$

Here, m^+ and m^- are object sets in which the representative weight for cluster c of p-th view is positive and zero respectively. $|m^-|$ and $|m^+|$ represent the size of set m^- and m^+ respectively. Similar to q^+ and q^-, the process for determining $|m^-|$ and $|m^+|$ is presented as follows.

Procedure for determining m^+ and m^-

---

1 Initialize $m_0^+ = \emptyset, m_0^- = 1, 2, ..., N, s = 0$;

2 $s \leftarrow s + 1, m_s^+ = m_{s-1}^+ + \{m\}, m_s^- = m_{s-1}^- - \{m\}$;

  where $m = arg\,min_{j \in m_{s-1}^-}\{\sum_{i=1}^{N} u_{ci}^* r_{ij}^{(p)}\}$;

3 Check whether $v_{cg}^{(p)} > 0$ computed by Eq. 13,

  where $g = arg\,max_{j \in m_s^+}\{\sum_{i=1}^{N} u_{ci}^* r_{ij}^{(p)}\}$.

  If yes, go to step 2, else set $m^+ = m_{s-1}^+, m^- = m_{s-1}^-$
  and terminate.

Maximization: Fixing u_{ci}^*, $v_{cj}^{(p)}$ and Updating $\alpha^{(p)}$

Based on the Lagrangian Multiplier method, the updating rule $\alpha^{(p)}$ is given as follow:

$$\alpha^{(p)} = \frac{(Q^{(p)})^{\frac{1}{1-\gamma}}}{\sum_{j=1}^{P} (Q^{(j)})^{\frac{1}{1-\gamma}}} \quad (15)$$

Here the cost term $Q^{(p)}$ is the weighted distance summation of all pairs of the data objects under p-th view. The larger the value of $Q^{(p)}$ is, the larger cost this view will contribute to the objective function. From (15), we can see that the larger cost of p-th view is, the higher value will be assigned to $\alpha^{(p)}$ which leads to the maximum of the weighted cost. The maximum is then minimized with respect to the memberships u_{ci}^* and $v_{cj}^{(p)}$ in order to suppress the high cost views and achieve harmonic consensus clustering results. The whole algorithm of the proposed approach is listed in Algorithm 1. First, the consensus membership matrix U_0^* and $\alpha_0^{(p)}$ is initialized. $\alpha_0^{(p)}$ is initialized to be $1/P$ to consider each view equally. The initialization method of U_0^* is introduced in Section 4.3. Then in second step, the representative weight matrix $V^{(p)}$ and $\alpha^{(p)}$ for each view are updated using (13) and (15). After updating all the views, the membership matrix U^* is updated based on (11). This step will be continued until the difference between two consecutive membership matrix is less than ϵ.

Algorithm 1:

Input: Distance matrix of P views $R_{n \times n}^1, ..., R_{n \times n}^P$,
    number of clusters K, parameter T_u, T_v, γ
    stopping criterion ϵ

Output: Consensus Fuzzy membership matrix U^*,
    representative weight matrix of each view $V^{(p)}$

Method:

1 Initialize U_0^* and $\alpha_0^{(p)}$, set iteration number $l \leftarrow 0$

2 **Repeat**

    $l \leftarrow l + 1$

    for $p = 1$ to P

      Update $V_l^{(p)}$ using (13) based on U_{l-1}^* and $\alpha_{l-1}^{(p)}$

      Update $\alpha_l^{(p)}$ using (15) based on U_{l-1}^* and $V_l^{(p)}$

    end for

    Update U_l^* using (11) based on $\{V_l^1, ..., V_l^{(P)}\}$
    and $\{\alpha_l^1, ..., \alpha_l^{(P)}\}$

  Until($\| U_l^* - U_{l-1}^* \| < \epsilon$)

The time complexity of M4-FC is $O(P(N^2))$ considering the number views P and the data size N. The small elements in distance matrix can be set to zero or applying sampling approaches to accelerate the computation as reported in [Bezdek *et al.*, 2006].

4 Experiments

In this section, we conduct experimental studies of the proposed approach on five benchmark data sets. Table 1 shows the summarized characteristics of these data sets. Feature Type in Table 1 highlights the number of data objects, and type of the features used in each feature set (or View) of the corresponding data set. The code is implemented in MATLAB and runs on a computer with eight cores and eight gigabytes of memory.

4.1 Data Sets

Multiple features (MF)[1]: This data set is composed of 2000 handwritten digit images (0-9) which is from a collection of Dutch utility maps. It has 10 classes and each class has 200 images. Each image is represented by six different views including Fourier coefficients (FOU), profile correlations (FAC), Karhunen-Love coefficients (KAR), pixel averages (PIX), Zernike moments (ZER), and morphological features (MOR).

Image segmentation (IS) [2]: This data set is composed of 2310 outdoor images which have 7 classes. Each image is represented by 19 features. The features can be considered as two views which are shape view and RGB view. The shape view consists of 9 features which describe the shape information of each image. The RGB view consists of 10 features which describe the RGB values of each image.

Oxford Flowers [3]: This data set consists of 1360 flower images. It has 17 classes and each class has 80 images. Each image is represented by 4 views including Colour, Histogram of Gradients (HOG), SIFT on the foreground region (SIFTreg) and SIFT on the foreground boundary (SIFTbdy).

3-Sources document (3-S) [4]: This data set is composed of 948 news articles containing 416 distinct news stories. They are collected from three online news sources: BBC, Guardian and Reuters. We selected 169 news articles which are reported in all three sources (views). It has 6 topic classes which are entertainment, health, business, politics, sport and technique.

Reuters multilingual: This data set consists of documents written in five different languages (English, French, German, Spanish and Italian) and their corresponding translations [Amini *et al.*, 2009]. It has six classes. The original documents in English are used as one view and their four translations are considered as the other four views. 1500 documents are randomly sampled with each class having 250 documents.

---

[1]https://archive.ics.uci.edu/ml/data sets/Multiple+Features.
[2]https://archive.ics.uci.edu/ml/data sets/Image+Segmentation.
[3]http://www.robots.ox.ac.uk/ vgg/data/flowers/17/index.html.
[4]http://mlg.ucd.ie/data sets/3sources.html.

Table 1: The characteristics of the multi-view data sets

Feature Type	Date set				
	MF	IS	Flowers	3-Sources	Reuters
1	PIX(240)	Shape(9)	Colour(1000)	BBC(3560)	English(21531)
2	FOU(76)	RGB(19)	HOG(1500)	Guardian(3631)	French(24892)
3	FAC(216)	-	SIFTreg(8000)	Reuters(3068)	German(34251)
4	ZER(47)	-	SIFTbdy(3000)	-	Spanish(11547)
5	KAR(64)	-	-	-	Italian(15506)
6	MOR(6)	-	-	-	-
Total data #	2000	2310	1360	169	1500
Classes #	10	7	17	6	6
View #	6	2	4	3	5

4.2 Experimental Settings

In the experiments, we aim to answer the following four questions. First, does the proposed M4-FC perform better than single view Fuzzy c medoids(FCMD)? Second, does M4-FC perform better than FCMD on concatenated features of all the views? Third, does multiple medoids based M4-FC perform better than single medoid based MVFCMddV? Last, how effective is M4-FC compared to other multi-view clustering approaches for relational data? In order to answer the four questions, the following approaches are selected for comparisons.

Single view Fuzzy c medoids (FCMD): Fuzzy c medoids is run on each single view.

Feature Concatenation Fuzzy c medoids (ConFCMD): The features of all the views are concatenated first and then FCMD is run on the resulted features.

MVFCMddV: The single medoid based multi-view fuzzy c medoids is proposed in [de Carvalho et al., 2015].

Multi-view Spectral Clustering (MVSC): One of the state-of-the-art multi-view spectral clustering methods for relational data is proposed in [Wang et al., 2014].

The time complexity of FCMD, MVFCddV and MVSC are $O(N^2)$, $O(K * N^2 * P)$ and $O(N^3 * P)$, respectively. Here N is the number of objects, K is the number of clusters, P is the number of views. For fair comparison, the parameter m in FCMD, ConFCMD and MVFCMddV is searched from the range of [1.1 2] with step 0.1. For MVSC, we follow the authors' experimental setting and parameter selecting methods in their paper. For M4-FC, the parameter γ is searched from [0.1 0.9] with the step 0.1. T_u is searched from [0.001, 0.003, 0.01, 0.03, 0.1, 0.3]. T_v is set by a guideline to make T_u/K and T_v/N close to each other or be in the same order which always produces reasonable results in our experimental study.

4.3 Initialization Method

For FCMD and ConFCMD, the medoids are initialized by using the same method used in [Krishnapuram et al., 2001]. The first medoid is selected as the object which has the minimum distance to all the other objects. We then select the remaining medoids by choosing the objects which maximize their minimal distance to the existing medoids. Selecting meoids by this method helps the algorithm converge to a better position because the meoids are distributed more evenly in the data space than random initialization. The same method is also applied in K-means which is used as the final step of MVSC.

For MVFCMddV and M4-FC, the consensus fuzzy membership U^* is initialized based on this similar method. First, we select medoids for each view by using the same method mentioned above. Then, if object i in p-th view is closest to the selected medoid of cluster m, we set $u_{mi}^p = 1$, and set the membership of this object to other clusters as 0 which is $u_{ci}^p = 0, c = 1, 2...K, c \neq m$. Last, U^* is generated based on the membership of all views U^p by setting $U^* = \sum_{p=1}^{P} U^p/P$.

Note that using this initialization method the initialization of each run is the same, therefore the clustering results are the same for each run.

4.4 Experimental Results

For the experimental results, we report three standard clustering evaluation metrics : Accuracy (ACC) [Cai et al., 2005], Normalized Mutual Information (NMI) [Strehl and Ghosh, 2003] and F-measure [Larsen and Aone, 1999]. For MF and IS data sets, the Euclidean distance is used to calculate the distances. For Flowers data set, the distance matrices are downloaded directly from website. Therefore, there are no results for ConFCMD on Flowers data. For 3-Sources data set, cosine distance is used for calculating the distance. For Reuters data set, the same experimental setting as in [Kumar et al., 2011], Probabilistic Latent Semantic Analysis (PLSA) [Hofmann, 1999] is applied to project the data to a 100-dimensional space and the clustering approaches are conducted on the low dimensional data. Table 2, Table 3 and Table 4 show the clustering ACC, NMI and F-measure respectively.

As we can see from the tables, all multi-view approaches perform better than single view approaches in general. Our proposed M4-FC constantly performs much better than the single view FCMD and FCMD with concatenated features. In addition, multiple medoids based M4-FC constantly outperforms single medoid based MVFCMddV. Moreover, comparing with multi-view spectral clustering, our proposed approach also performs better on all the data sets.

4.5 Parameter Analysis

In our objective function as shown in (3) and (4), we have three parameters which are γ to control the view weight, T_u and T_v to control the weight of the regularization term in the objective function. To show the impact of the parameters on

the performance of M4-FC, we plot the NMI performance curve w.r.t. γ and T_u for each data set in Fig. 1 and Fig. 2,

Table 2: ACC(%) of different methods on the five data sets

Method	Date set				
	MF	IS	Flowers	3-S	Reuters
FCMD(1)	64.90	33.30	22.06	56.21	27.67
FCMD(2)	54.45	54.42	35.37	62.13	41.73
FCMD(3)	55.95	–	24.04	59.76	34.33
FCMD(4)	37.20	–	22.87	–	41.20
FCMD(5)	60.65	–	–	–	19.67
FCMD(6)	49.20	–	–	–	–
ConFCMD	72.40	47.36	–	66.27	45.87
MVFCMddV	78.20	63.77	40.44	66.86	47.00
MVSC	92.00	66.58	54.71	64.50	43.07
M4-FC	**95.45**	**72.77**	**56.91**	**84.62**	**52.53**

Table 3: NMI(%) of different methods on the five data sets

Method	Date set				
	MF	IS	Flowers	3-S	Reuters
FCMD(1)	62.08	19.15	19.35	38.65	11.58
FCMD(2)	54.63	59.37	38.82	47.55	17.65
FCMD(3)	55.30	–	26.19	37.25	12.68
FCMD(4)	48.82	–	27.19	–	14.98
FCMD(5)	58.99	–	–	–	00.49
FCMD(6)	46.37	–	–	–	–
ConFCMD	67.82	45.37	–	45.11	26.41
MVFCMddV	73.77	62.84	42.59	51.09	26.49
MVSC	84.48	60.10	51.09	65.33	28.76
M4-FC	**90.64**	**68.80**	**55.52**	**78.37**	**29.48**

Table 4: F-measure(%) of different methods on the five data sets

Method	Date set				
	MF	IS	Flowers	3-S	Reuters
FCMD(1)	69.18	36.31	25.16	57.25	33.19
FCMD(2)	60.19	59.00	37.59	62.63	43.25
FCMD(3)	61.19	–	26.37	55.60	38.39
FCMD(4)	46.45	–	27.82	–	39.51
FCMD(5)	64.21	–	–	–	23.20
FCMD(6)	52.83	–	–	–	–
ConFCMD	72.68	53.01	–	65.66	48.55
MVFCMddV	80.57	64.59	42.48	70.94	50.59
MVSC	92.00	66.53	56.99	71.06	47.84
M4-FC	**95.44**	**75.54**	**59.39**	**85.66**	**51.01**

respectively. As mentioned above, T_v can be set to make T_u/K and T_u/N close to each other or be in the same order. Here we only show the NMI results, the results of accuracy and F-measure have a similar pattern. Fig. 1 and Fig. 2 are generated as follows. First, for Fig. 1 the value of the T_u and T_v which produces the results in Table 2, Table 3 and Table 4 are fixed. Then, the NMI results are plotted w.r.t the parameter γ with values from [0.1 0.9] with a step of 0.1. Respectively, Fig. 2 shows the NMI results with γ and T_v fixed and T_u taking values from [0.001, 0.003, 0.01, 0.03, 0.1, 0.3].
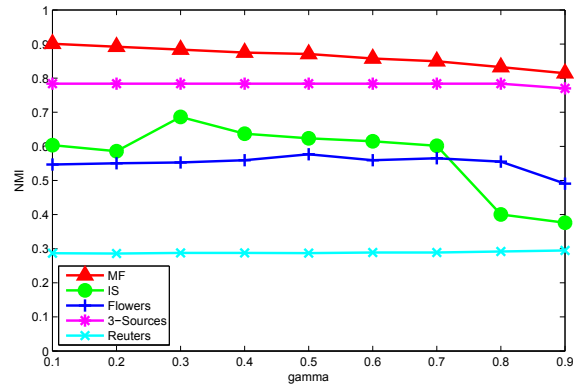

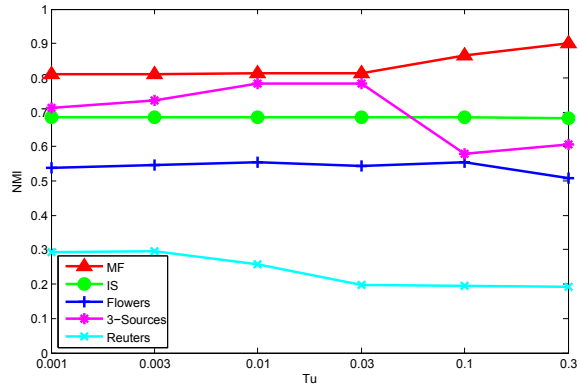Figure 1: Parameter analysis: NMI values with different γ.

Figure 2: Parameter analysis: NMI values with different T_u.

As shown in Fig. 1 and Fig. 2, the NMI results are robust with respect to the parameter γ in the range of [0.1 0.7] and T_u in the range of [0.001 0.03] for most of the data sets. In general, higher γ value near 0.9 may achieve worse performance. Moreover, it shows that choosing γ from [0.1 0.7] may achieve better and robust performance. For different T_u, the performance on IS and Flowers data sets are very stable. Therefore, we recommend to set γ to a value from [0.1 0.7] and T_u from [0.001 0.03] to have robust results in practice.

5 Conclusion

We have proposed M4-FC, a new multiple medoids based multi-view fuzzy clustering with minimax optimization for multi-view data analysis, and applied M4-FC on several real world data sets to demonstrate its effectiveness and potential. In M4-FC, a representative weight for each object is introduced to generate clusters with better quality. In addition, minimax optimization is applied to ensure more harmonic consensus clustering results. Experimental results show that multiple medoids based M4-FC not only outperforms single medoid based multi-view fuzzy clustering, but also achieves better clustering results than the related multi-view clustering approaches on the benchmark data sets used in the study.

References

[Amini *et al.*, 2009] Massih Amini, Nicolas Usunier, and Cyril Goutte. Learning from multiple partially observed views-an application to multilingual text categorization. In *Proceedings of Advances in Neural Information Processing Systems*, pages 28–36, Vancouver, Dec. 2009.

[Bezdek *et al.*, 2006] James C Bezdek, Richard J Hathaway, Jacalyn M Huband, Christopher Leckie, and Ramamohanarao Kotagiri. Approximate clustering in very large relational data. *International Journal of Intelligent Systems*, 21(8):817–841, 2006.

[Bezdek, 1981] James C Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Kluwer Academic Publishers, Norwell, MA, 1981.

[Cai *et al.*, 2005] Deng Cai, Xiaofei He, and Jiawei Han. Document clustering using locality preserving indexing. *IEEE Transactions on Knowledge and Data Engineering*, 17(12):1624–1637, 2005.

[Cleuziou *et al.*, 2009] Guillaume Cleuziou, Matthieu Exbrayat, Lionel Martin, and J Sublemontier. Cofkm: A centralized method for multiple-view clustering. In *Proceedings of the 9th IEEE International Conference on Data Mining*, pages 752–757, Miami, FL, Dec. 2009.

[de Carvalho *et al.*, 2015] Francisco de AT de Carvalho, Filipe M de Melo, and Yves Lechevallier. A multi-view relational fuzzy c-medoid vectors clustering algorithm. *Neurocomputing*, 163:115–123, 2015.

[Filippone *et al.*, 2008] Maurizio Filippone, Francesco Camastra, Francesco Masulli, and Stefano Rovetta. A survey of kernel and spectral methods for clustering. *Pattern Recognition*, 41(1):176–190, 2008.

[Hofmann, 1999] Thomas Hofmann. Probabilistic latent semantic indexing. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 50–57, Berkeley, CA, Aug. 1999.

[Jain, 2010] Anil K Jain. Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*, 31(8):651–666, 2010.

[Jiang *et al.*, 2015] Yizhang Jiang, Fu-Lai Chung, Shitong Wang, Zhaohong Deng, Jun Wang, and Pengjiang Qian. Collaborative fuzzy clustering from multiple weighted views. *IEEE Transactions on Cybernetics*, 45(4):688–701, Apr. 2015.

[Krishnapuram *et al.*, 2001] Raghu Krishnapuram, Anupam Joshi, Olfa Nasraoui, and Liyu Yi. Low-complexity fuzzy relational clustering algorithms for web mining. *IEEE Transactions on Fuzzy Systems*, 9(4):595–607, Aug. 2001.

[Kumar *et al.*, 2011] Abhishek Kumar, Piyush Rai, and Hal Daume. Co-regularized multi-view spectral clustering. In *Proceedings of Advances in Neural Information Processing Systems*, pages 1413–1421, Granada, Spain, Dec. 2011.

[Larsen and Aone, 1999] Bjornar Larsen and Chinatsu Aone. Fast and effective text mining using linear-time document clustering. In *Proceedings of the 5th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 16–22, San Diego, CA, Aug. 1999.

[Mei and Chen, 2010] Jian-Ping Mei and Lihui Chen. Fuzzy clustering with weighted medoids for relational data. *Pattern Recognition*, 43(5):1964–1974, 2010.

[Strehl and Ghosh, 2003] Alexander Strehl and Joydeep Ghosh. Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research*, 3:583–617, 2003.

[Wang and Chen, 2017] Yangtao Wang and Lihui Chen. Multi-view fuzzy clustering with minimax optimization for effective clustering of data from multiple sources. *Expert Systems with Applications*, 72:457–466, 2017.

[Wang *et al.*, 2014] Hongxing Wang, Chaoqun Weng, and Junsong Yuan. Multi-feature spectral clustering with minimax optimization. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 4106–4113, Columbus, OH, Jun. 2014.

[Xu and Wunsch, 2005] Rui Xu and Donald Wunsch. Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, 16(3):645–678, May. 2005.