2009 Ninth IEEE International Conference on Data Mining

CoFKM: a centralized method for multiple-view clustering

Guillaume Cleuziou, Mathieu Exbrayat, Lionel Martin, Jacques-Henri Sublemontier
LIFO
University of Orleans, France
firstname.lastname@univ-orleans.fr

***Abstract*—This paper deals with clustering for multi-view data, *i.e.* objects described by several sets of variables or proximity matrices. Many important domains or applications such as Information Retrieval, biology, chemistry and marketing are concerned by this problematic. The aim of this data mining research field is to search for clustering patterns that perform a consensus between the patterns from different views. This requires to merge information from each view by performing a fusion process that identifies the agreement between the views and solves the conflicts. Various fusion strategies can be applied, occurring either before, after or during the clustering process. We draw our inspiration from the existing algorithms based on a centralized strategy. We propose a fuzzy clustering approach that generalizes the three fusion strategies and outperforms the main existing multi-view clustering algorithm both on synthetic and real datasets.**

***Keywords*-multi-view data; fuzzy clustering; collaborative clustering; centralized clustering;**

I. Introduction and related works

The ever growing complexity of data constitutes a central challenge for the data mining community. This complexity may concern various aspects, such as the size of the dataset, the complexity of the features, the temporality or scalability and more generally the multiplicity of data. Amongst the recent advances in data mining, the one we will focus on in this paper, known as Multiple View Learning, aims at dealing with objects described by multiple views, each one providing its own light on the data to be mined.

A. Multiple views in real world data

Data with multiple representations are quite common in scientific, economical and social domains such as biology, chemistry, medicine, marketing and social networks. For instance, biologists may consider simultaneously gene activity (expression), phylogenetic and location data in order to detect gene interaction or regulation [1]. Concerning chemistry, molecules can be described using both fingerprints and spatial structure [2]. In the context of marketing, information on a same set of customers are currently available from different databases (bank, store, administration, *etc.*). Building and mining a social network suppose to assemble various data (*e.g.* emails, collaborations, organization memberships, *etc.*). From a transverse point of view, multiple views are also frequently provided for textual data analysis in Computational Linguistic and Information Retrieval. Information Retrieval applications currently deal with web pages observed according to various angles (text, links, structure, *etc.*) and several description levels of texts must similarly be considered for Computational Linguistic analysis (lexicon, morphology, syntax, *etc.*).

B. Context and motivations

The present study is focused on the specific task of data clustering. Recent work has highlighted the fact that multiple-view clustering could be of interest for various clustering issues [3].

Knowledge Reuse consists in considering several different clusterings performed previously on a single dataset (considered as the knowledge on this dataset), in order to process this dataset with a new description. Each clustering is then considered as one view in the multi-view clustering process.

Distributed Computing occurs when data (objects and/or features) is scattered among several (geographical) sites, with the impossibility to collect them into a single site, due to technical or confidential reasons [4], [3].

Quality and Robustness improvement can be the objective sought with a multi-view clustering. The diversity of the descriptions provided by the different views could help in learning a better clustering solution that simultaneously satisfies each representation space (view). Some experimental results presented in the last part of the present study tend to reinforce this hypothesis.

C. Strategies for multiple-view clustering

Several supervised or semi-supervised classification approaches have been designed through the past decade which learn models in presence of several views [5], [6], [7]. These approaches sought to take advantage of multiplicity, by using the agreement of the different views as a pseudo supervisor that complements the original supervisor (labels).

The process can also be transposed, in fully unsupervised situations, by using the agreement as a pseudo supervisor which complements usual unsupervised quality measures (let us remind that such quality measures guide traditional single-view clustering processes so that similar objects tend to appear in the same cluster while dissimilar ones do not [8]).

In the particular context of multi-view clustering, we distinguish three combination strategies depending on whether

1550-4786/09 $26.00 © 2009 IEEE
DOI 10.1109/ICDM.2009.138

the view combination is performed before, during or after the clustering process.

Concatenation strategy. This first strategy (also denoted as "*a priori* fusion" thereafter), consists in the concatenation of the views into a single one, either directly by juxtaposing the sets of features or undirectly by combining the proximity matrices derived from each view. For instance in [9], the authors combine the lexical and contextual description of web pages into a single similarity matrix that results from linear combination of similarities between term vectors and hyperlinks contents respectively. In [1], proximity between genes is stored into a single kernel obtained by the sum of multiple kernels, each one corresponding solely on expressions, locations, phylogenetic profiles or chemical compatibilities. The concatenation strategy is a natural solution that leads to good results in practice. Unfortunately, as we mentioned previously, such a concatenation is impossible in some contexts: when data comes from different sources with confidentiality restrictions, or several geographical sites with bandwidth and storage limitations; also when the concatenation would result in feature vectors with a dimensionality level that makes any clustering process untractable.

Distributed strategy. Also known as *clustering ensembles* problematic or *a posteriori* fusion, this strategy proceeds by first clustering the objects from each view independently and then searching a solution that represents a consensus among the set of clusterings. For instance, Strehl&Ghosh [3] propose several search heuristics in order to provide the best consensus according to a given objective function (average normalized mutual information). Long et al. [10] define a quality measure based on the generalized I-divergence associated with an optimization algorithm. A more complete panorama of clustering ensembles techniques is available in [11].

The main drawback of these methods lies in the fact that they do not reconsider the clusterings previously performed. It is thus possible to imagine a third (centralized) strategy where clusterings are not given as input, but are rather learned in a collaborative process.

Centralized strategy. This last strategy makes use of the multiple views simultaneously in order to mine hidden patterns from the data. It represents an important challenge because it requires to modify deeply the clustering process. As far as we know, two main centralized methods have been proposed in the literature, first by Pedrycz [4] and then by Bickel&Sheffer [12], that show improvements with respect to both single distributed or concatenation strategies. In this paper we present (Section 2) some limitations of the existing centralized algorithms, concerning the applicability of these models as well as their theoretical definition. We propose a new multi-view fuzzy clustering solution (Section 3) such that each of the three combination strategies can be instantiated depending on a single parameter to adjust. In addition, we propose the first experimental comparison among the existing centralized methods, based on both synthetic and real datasets (Section 4).

II. RELATED WORKS ON CENTRALIZED MULTI-VIEW CLUSTERING

In this section, we introduce all necessary notations and briefly present the two centralized multi-view clustering "Collaborative Fuzzy Clustering" and "Co-EM" proposed by Pedrycz and Bickel&Sheffer respectively.

A. Notations

In the following we consider a set of data $X = \{x_1, \ldots, x_n\}$ to cluster. The data are described in $|R|$ different views such that in view r, x_i is defined by the vector denoted as $x_{i,r}$. In the fuzzy clustering framework, we use $u_{i,k,r} \in [0,1]$ to denote the membership degree of $x_{i,r}$ to cluster k, $c_{k,r}$ to denote the center of cluster k and $d_{i,k,r}$ to denote the euclidean distance between x_i and c_k in view r. Similarly for the generative framework we denote as $P(k|x_{i,r}, \Theta_r)$ the posterior probability of x_i to belong to cluster k in view r and Θ_r the parameters of the distribution in view r. We use $\overline{r}$ to represent each r' with $r' \neq r$.

B. Collaborative Fuzzy Clustering

In [4], Pedrycz uses the fuzzy k-means model [13] and derives the collaborative variant CoFC for the multiview context. The collaboration between the views only concerns the membership degrees $u_{i,k,r}$; in this way data confidentiality is satisfied and bandwidth or storage costs are strongly reduced. A local inertia term (to minimize) is defined (1) by the inertia of the fuzzy clusters in view r (first term), "penalized" by a disagreement with the other views r' (second term). The disagreement between views r and r' is weighted *via* $\alpha_{r,r'}$ (fixed) that denotes *a priori* information about the desirable collaboration between the two views.

$$Q_{CoFC}(r) = \sum_{k=1}^{K} \sum_{x_i \in X} u_{i,k,r}^2 d_{i,k,r}^2 \qquad (1)$$
$$+ \sum_{r'=1}^{|R|} \alpha_{r,r'} \sum_{k=1}^{K} \sum_{x_i \in X} (u_{i,k,r} - u_{i,k,r'})^2 d_{i,k,r}^2$$

Let us notice that the original fuzzy-k-means contains a fuzzyfier parameter $\beta > 1.0$ that is fixed at 2.0 in the CoFC model to make the optimization process possible. By the way, CoFC is less general than fuzzy-k-means; we show in the experiments the influence of the fuzzifier β and observe significant improvements with our new model allowing other values for β. In addition the minimization of $Q_{CoFC}(r)$ results in non-intuitive updates, in particular the computation of a new membership degree $u_{i,k,r}$ depending of the $u_{j,k,r}$ $(j \neq i)$ from the other views only.

C. Co-EM

Bickel&Sheffer propose in [12] a process very similar to CoFC but using the framework of mixture models. They propose a collaborative variant (Co-EM) of the EM algorithm commonly used for clustering [14]. The objective function (to maximize) combines the local log-likelihoods ($Q_{EM}(r)$) from any views and a disagreement term Δ:

$$Q_{Co-EM} = \sum_{r=1}^{|R|} Q_{EM}(r) - \eta\Delta \quad (2)$$

In (2), η parametrizes the contribution of the disagreement Δ that can be interpreted almost as a Kullback-Leibler divergence between the posterior distributions on any pair of views:

$$\begin{aligned}\Delta &= \frac{1}{|R|-1} \times \quad (3)\\ &\sum_{r \neq r'} \sum_{x_i \in \mathcal{X}} \sum_{k=1}^{K} P(k|x_{i,r}, \Theta_r^t) \log \frac{P(k|x_{i,r}, \Theta_r)}{P(k|x_{i,r'}, \Theta_{r'})}\end{aligned}$$

The authors give an other formulation of (2)[1] with a sum of log-likelihoods on each views where posteriors are obtained by weighted averages of the posteriors on the different views. Unfortunately, by using the new formalization of the posteriors, the objective criterion cannot be maximized on the whole in anything but to annihilate the disagreement with $\eta \to 0$ through the iterations. The model CoFKM we propose in this study enriches of the previous approaches and passes through the observed practical and theoretical limitations.

III. The CoFKM model

The approach we propose here is an extension of the fuzzy k-means method FKM [13]. We try to obtain in each view a specific organization, but as done in CoEM, we introduce a penalty term which aims at reducing the disagreement between organizations on the different views.

Let us recall that the goal of FKM [13] is to optimize a weighted inertia criterion:

$$Q_{FKM} = \sum_{k=1}^{K} \sum_{x_i \in X} u_{i,k}^{\beta} d_{i,k}^2 \quad (4)$$

with $\forall x_i \in X$, $\Sigma_{k=1}^{K} u_{i,k} = 1$, where variables of the problem are centers of clusters (c_k) and membership degrees ($u_{i,k}$).

A. Objective criterion

Given a set R of views, we note $Q_{FKM}(r)$ the above criterion associated to the view $r \in R$. In any view, objects are described by a vector in $\mathbb{R}^{N_r}$, where N_r is the dimensionality of the view r.

[1] by introducing the disagreement under the sum on each view.

We propose a collaborative approach, based on FKM, which aims at minimizing $Q_{FKM}(r)$ in each view, and penalizing the disagreement between any pairs of views. Thus, the criterion to be minimized can be written:

$$\begin{aligned} Q_{CoFKM} &= \left(\sum_{r \in R} Q_{FKM}(r) \right) + \eta\Delta \quad (5)\\ &= \sum_{r \in R} \sum_{x_i \in X} \sum_{k=1}^{K} u_{i,k,r}^{\beta} d_{i,k,r}^2 + \eta\Delta \end{aligned}$$

where views are normalized in order to get comparable inertia for all views: each variable is normalized to unit variance, and is applied a weight equal to $\frac{1}{\sqrt{N_r}}$ (if the variable belongs to view r).

In the previous definition of Q_{CoFKM} (5), Δ is a disagreement term: when the same organization is obtained in all the views, this term must be equal to zero. We propose:

$$\Delta = \frac{1}{|R|-1} \sum_{r \neq r'} \sum_{x_i \in X} \sum_{k=1}^{K} (u_{i,k,r'}^{\beta} - u_{i,k,r}^{\beta}) d_{i,k,r}^2$$

In this expression, we sum the differences between organizations obtained from views r and r', for any couple of views (r, r'). The previous expression can be written as a sum on pairs (r, r') such that $r > r'$:

$$\Delta = \frac{1}{|R|-1} \sum_{r > r'} \sum_{x_i \in X} \sum_{k=1}^{K} (u_{i,k,r'}^{\beta} - u_{i,k,r}^{\beta})(d_{i,k,r}^2 - d_{i,k,r'}^2)$$

This disagreement term is made to penalize our criterion. It can be considered as a divergence between organizations since, the lower is $(u_{i,k,r}^{\beta} - u_{i,k,r'}^{\beta})$, the lower the disagreement. The normalization we apply implies that $d_{i,k,r'}$ and $d_{i,k,r}$ are comparable. $d_{i,k,r}$ being inversely proportional to $u_{i,k,r}$, one can consider the term $(d_{i,k,r'} - d_{i,k,r})$ being comparable to $(u_{i,k,r} - u_{i,k,r'})$, so the disagreement can be seen as a distance between the local organizations $(u_{i,k,r})$ and $(u_{i,k,r'})$. The advantage is that our disagreement term has the same order of magnitude than local inertia, then the sum of these expressions can be considered as a coherent global criterion Q_{CoFKM}:

$$Q_{CoFKM} = \sum_{r \in R} \sum_{x_i \in X} \sum_{k=1}^{K} u_{i,k,r,\eta} d_{i,k,r}^2 \quad (6)$$

where $u_{i,k,r,\eta} = (1-\eta) u_{i,k,r}^{\beta} + \frac{\eta}{|R|-1} (\sum_{\overline{r}} u_{i,k,\overline{r}}^{\beta})$ and η is a parameter which allows to control the penalty associated to the disagreement. This criterion is then a weighted inertia where, in each view, $u_{i,k,r,\eta}$ is a weighted mean of usual membership degrees $(u_{i,k,r'}^{\beta})$ obtained from each view.

As done in the case of FKM, our goal is to find a solution minimizing the global criterion (6) and satisfying $\Sigma_k u_{i,k,r} = 1$ for any view r and any object x_i. We can solve this constrained optimization problem by considering the Lagrangian. As done in the case of FKM, we propose an algorithm (1) alternating the computation of the optimal centers $c_{k,r}$ with fixed degrees $u_{i,k,r}$, and the computation of the optimal $u_{i,k,r}$ with fixed centers $c_{k,r}$.

These updates are given by :

$$c_{k,r} = \frac{\sum_{x_i \in X} u_{i,k,r,\eta} x_{i,r}}{\sum_{x_i \in X} u_{i,k,r,\eta}} \tag{7}$$

$$u_{i,k,r} = \frac{\left((1-\eta) d^2_{i,k,r} + \frac{\eta}{|R|-1} \sum_{\overline{r}} d^2_{i,k,\overline{r}}\right)^{1/(1-\beta)}}{\sum_{k'=1}^{K} \left((1-\eta) d^2_{i,k',r} + \frac{\eta}{|R|-1} \sum_{\overline{r}} d^2_{i,k',\overline{r}}\right)^{1/(1-\beta)}} \tag{8}$$

Thus, the criterion (6) decreases at any step of the process, which ensure the convergence (to a local optimum).

B. Assignment rule

The method proposed above ensures that we get an optimum of the global criterion, however, it produces for each view a fuzzy partition. In order to get a global clustering, we have to merge these results. We propose to build a global partition of objects using an assignment rule. This rule consists in computing, for each object and for each cluster, a global membership degree, as the geometric mean of membership degrees (local to views):

$$\hat{u}_{i,k} = \sqrt[|R|]{\prod_{r \in R} u_{i,k,r}} \tag{9}$$

and assign the object x_i to the cluster k maximizing $\hat{u}_{i,k}$.

This rule requires that we can associate a cluster in a view to a cluster in another view. We consider here that a cluster is identified by the same index $k \in \{1..K\}$ in all the views. The consistency of this identification of clusters is not addressed neither in CoFC nor in CoEM. We propose here, a particular initialization which consists in randomly choosing an object as center for all clusters, for all index k, centers $c_{k,r}$ correspond to the views of the same objects in each view.

C. Comparisons

We show here that our model CoFKM is a generalization of both the model FKM applied to the concatenation of the views (*a priori* fusion), and a simple *a posteriori* model where FKM is applied independently on each view.

Let us consider expressions of $c_{k,r}$ and $u_{i,k,r}$ proposed in equations (7) and (8), where terms corresponding to different views are associated to the same weight, i.e. $(1-\eta) = \frac{\eta}{|R|-1}$ or $\eta = \frac{|R|-1}{|R|}$.

Algorithm 1 CoFKM

Input: Dataset X, Number of clusters K, Number of views $|R|$
Initialize the same K cluster centers for all views r.
repeat
  for all $r \in R$, k **do**
    update $c_{k,r}$ using (7)
  end for
  for all $x_i \in X$, $r \in R$, k **do**
    update $u_{i,k,r}$ using (8)
  end for
until convergence
assign x_i to the cluster C_k using (9)

In this case, $u_{i,k,r,\eta} = \frac{1}{|R|} u^{\beta}_{i,k,r} + \frac{1}{|R|} \sum_{\overline{r}} u^{\beta}_{i,k,\overline{r}} = \frac{1}{|R|} \sum_{r'} u^{\beta}_{i,k,r'}$ and then $u_{i,k,r,\eta}$ does not depend on r, and the expression of $c_{k,r}$ corresponds exactly to the one obtained for FKM applied to the concatenation of the views.

In the same way, $d^2_{i,k,r} + \sum_{\overline{r}} d^2_{i,k,\overline{r}}$ is the distance between x_i and c_k associated to the concatenation of the views, the expression of $u_{i,k,r}$ corresponds exactly to the one obtained by FKM from the concatenation of the views. Then, CoFKM is a generalization of FKM applied on the concatenation of views, where we can force to find a solution corresponding to a consensus by choosing a η value different than $\frac{(|R|-1)}{|R|}$.

Experiments show that the sign of the disagreement term depends on the η value. When $\eta > \frac{(|R|-1)}{|R|}$, the disagreement term is negative, which is not expected. For this reason, we propose to choose $0 \leq \eta \leq \frac{(|R|-1)}{|R|}$.

If we consider now the CoFKM model with $\eta = 0$, the criterion Q_{CoFKM} can be rewritten as the sum of FKM criterion for all the views: $Q_{CoFKM_{\eta=0}} = \sum_{r \in R} Q_{FKM(r)}$ This criterion is the sum of local inertia, which are optimized independently with FKM method. The *a posteriori* fusion is realized by the assignment rule. Our collaborative model CoFKM is then a generalization of an *a posteriori* fusion, by choosing $\eta = 0$.

We compare now our model to the CoEM approach. The main theoretical drawback of CoEM is the non-convergence of the algorithm: to ensure convergence, [12] proposes to decrease the parameter η down to 0, which corresponds to the optimization of local criterion, independently in all the views. CoEM can then be viewed as a 2-steps method: during the first phase ($\eta > 0$) parameters are estimated in order to increase the consensus but with no convergence guaranteed; during the second phase ($\eta = 0$) the global criterion converges by local convergence in all views, but the penalty term is not considered. Our model is defined such that, whatever the η value, the convergence is ensured since the global criterion decreases at each step of the algorithm.

IV. Experimental results

We conduct experiments and valid our approach on two datasets. The first one is *multiple features* [2] available in the UCI Machine Learning repository. The second one is an artificial dataset used in [3] [3].

A. Datasets

The *multiple features* dataset consists in a set of $2,000$ handwritten numbers (digitalized pictures) described by 6 different views (Fourier coefficients, profile correlations, Karhunen-Loève coefficients, pixel averages, Zernike moments, morphological features). 10 homogeneous classes (200 objects per class) have to be recovered.

The *2D2K* dataset contains $1,000$ objects generated by a mixture model of two 2-dimensional gaussians with diagonal variance matrices. Three views are constructed artificially, as proposed in [3] (the first dimension gives the first and third views, the second dimension give the second view).

In order to better distribute the information contribution among the views, we did some classical statistical pre-processing, as explained in Section III.*A*.

B. Evaluation

The evaluation of a clustering result is still an open problem, since one does not always know the true label of objects. Indeed, if all classes labels are available, we can use an external evaluation criterion to measure how well the organization produced by the clustering process corresponds to these labels. We chose here to measure the result's quality of the compared approaches by two well-known different evaluations: average entropy, and normalized mutual information.

The average entropy (AvgEnt) uses class labels to compute the mean of all clusters' impurities:

$$\text{AvgEnt} = \sum_{k=1}^{K} \frac{n_k}{n} \left(- \sum_{c=1}^{C} p_{ck} \times log(p_{ck}) \right)$$

where K refers to the number of clusters, C refers to the number of classes, n_k (resp. n) is the number of objects in the cluster k (resp. the total number of objects). p_{ck} corresponds to the proportion of objects of class c in the cluster k.

The normalized mutual information (NMI) quantifies the statistical information shared between two distributions, here, the distribution of cluster labels and the distribution of class labels:

$$\text{NMI} = \frac{2}{n} \sum_{k=1}^{K} \sum_{c=1}^{C} n_{kc} \times log_{k.c} \left(\frac{n_{kc} \times n}{n_k \times n_c} \right)$$

where n_{kc} refers to the number of objects simultaneously in the cluster k and in the class c.

[2] *http://archive.ics.uci.edu/ml/*

[3] the dataset is called *2D2K*, available for download at *http://strehl.com/*

Note that a high (resp. low) value of NMI (resp. AvgEnt) is obtained for a good clustering scheme as regards to the expected classes.

C. Experiments

The results we obtained correspond to a mean of 20 runs for *multiple features* and 100 runs for *2D2K*. The methods were compared with the same initialization, the parameters were set to $\beta = 1.25$ and $\eta = \frac{|R|-1}{2\times|R|}$ which is an heuristic between the *a priori* and the *a posteriori* versions of CoFKM. Results are summarized in Tab.I.

We empirically observed that the general Gaussian mixture model estimation with CoEM was inefficient so we instantiated this model by some classical parsimonious models ($\sigma.I$, $\sigma_k.I$ and diagonal variance matrices).

In Tab.I FKM and EMgmm refer to the best results obtained running these single-view approaches over all the views and, in the case of EM, over all the three models. CoFC refers to the result on the best view it is applied. CoEMgmm refers to the best result obtained from the three models.

Table I
Experiments on the *2D2K* and the *multiple features* datasets.

	2D2K		*multiple features*	
Method	AvgEnt	NMI	AvgEnt	NMI
CoFKM	0.18 ± 0.00	0.82 ± 0.00	**0.29 ± 0.00**	**0.91 ± 0.00**
CoEMgmm	0.15 ± 0.00	0.85 ± 0.00	0.50 ± 0.09	0.85 ± 0.03
CoFC	0.19 ± 0.00	0.81 ± 0.00	2.45 ± 0.00	0.26 ± 0.00
CoFKM-*a posteriori*	0.34 ± 0.27	0.66 ± 0.27	1.21 ± 0.12	0.64 ± 0.04
CoEMgmm-*a posteriori*	0.32 ± 0.29	0.68 ± 0.29	1.18 ± 0.14	0.65 ± 0.04
FKMconcat	**0.13 ± 0.00**	**0.87 ± 0.00**	0.33 ± 0.07	0.90 ± 0.02
EMgmmconcat	**0.13 ± 0.00**	**0.87 ± 0.00**	0.56 ± 0.12	0.83 ± 0.04
FKM	0.40 ± 0.19	0.60 ± 0.19	0.88 ± 0.06	0.74 ± 0.02
EMgmm	0.40 ± 0.19	0.60 ± 0.19	0.99 ± 0.13	0.70 ± 0.04

Firstly, we can observe for *multiple features* that CoFKM outperforms state of art approaches CoEM and CoFC. We investigate by further experiments (not shown here) why CoFC runs poorly in this case. Our conclusion is that the fixed choice of the fuzzifier $\beta = 2$ in CoFC criterion seems to be the reason leading to so much degenerated result. The results obtained for *2D2K* are strongly differents, this time CoFKM performs a little bit lower than CoEM for a spherical ($\sigma.I$ matrix) gaussian mixture model, which is very close to FKM objective. We can also refer to [3] for the results of cluster ensembles approaches (NMI of 0.69) for the *2D2K* dataset and notice that in this dataset CoFKM outperforms cluster ensembles heuristics too.

Secondly, we see that CoFKM performs better than *a posteriori* solutions of CoFKM and CoEM.[4] Each view discovers its own organisation independently, then the final assignment rule merges all the local results into a consensual one. We can also note that CoFKM succeed in improving *a priori* solutions (concatenation) of FKM and EM in the real dataset, which is a very promising result, indeed this is not the object of this field study.

[4] imposing the null value to η.

Finally, for the two datasets we observe a very strong improvement over the single view approaches, which is currently a known result, and the main motivation of multi-view approaches.

We also test our model to see how the parameters η and β affect the quality of the result. The curve (Fig.1) presents the behavior of CoFKM for the real dataset. We can note that an appropriate value for β should be close to 1.2 (1.25 in our experiments), the value $\beta = 2$ (as fixed in CoFC) gave very bad results. We can also note that our heuristic choice of $\eta = \frac{|R|-1}{2\times|R|}$ (≈ 0.42 for *multiple features*) generally leads to good results.

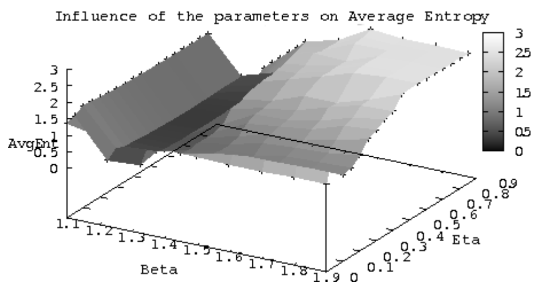

Figure 1. Influence of the parameters η and β in CoFKM for multiple features corresponding to a mean of 5 runs.

V. Conclusion and Future Works

In this work, we studied the clustering problem on complex multi-view data. We presented some state of the art approaches and proposed a new centralized model to extend and improve the centralized approaches clustering initiated by Pedrycz [4] and Bickel&Sheffer [12]. We provided a clustering criterion that extends the standard Fuzzy minimum-square error and that allows collaboration between the views following the disagreement-based strategy. We showed that this model generalizes different fusion solutions and we also derived an algorithm which locally optimizes our formulated criterion. This paper also presents the first experimental comparison between centralized multi-view methods on real and synthetic datasets. We observed empirically the benefits of the CoFKM model proposed and these results motivate our approach by showing significant performances. Many perspectives and generalizations will be proposed in future works. For example the possibility to deal with different number of clusters in each data set, the formalization of other disagreement terms, the integration of a semi-supervised treatment of data through a kernelized variant of our model by using the approach of [15].

References

[1] Y. Yamanishi, J. P. Vert, and M. Kanehisa, "Protein network inference from multiple genomic data: a supervised approach," *Bioinformatics*, vol. 20, no. 1, p. i363i370, 2004.

[2] P. Labute, "Quasar-cluster: A different view of molecular clustering," *Chemical Computing Group, Inc.*, 1998.

[3] A. Strehl and J. Ghosh, "Cluster ensembles — a knowledge reuse framework for combining multiple partitions," *J. Mach. Learn. Res.*, vol. 3, pp. 583–617, 2003.

[4] W. Pedrycz, "Collaborative fuzzy clustering," *Pattern Recogn. Lett.*, vol. 23, no. 14, pp. 1675–1686, 2002.

[5] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *COLT: Proceedings of the Workshop on Computational Learning Theory, Morgan Kaufmann Publishers*, 1998. [Online]. Available: http://citeseer.ist.psu.edu/blum98combining.html

[6] R. Ghani, "Combining labeled and unlabeled data for multi-class text categorization," in *Proceedings of ICML*, 2002, pp. 187–194.

[7] K. Ganchev, J. Graça, J. Blitzer, and B. Taskar, "Multi-view learning over structured and non-identical outputs," in *UAI*, D. A. McAllester and P. Myllymäki, Eds. AUAI Press, 2008, pp. 88–96.

[8] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," *ACM Computing Surveys*, vol. 31, no. 3, pp. 264–323, 1999. [Online]. Available: citeseer.nj.nec.com/jain99data.html

[9] J. Heer and E. H. Chi, "Mining the Structure of User Activity using Cluster Stability," in *proceedings of the Web Analytics Workshop, SIAM Conference on Data Mining*, 2002.

[10] B. Long, P. S. Yu, and Z. M. Zhang, "A general model for multiple view unsupervised learning," in *SDM*. SIAM, 2008, pp. 822–833.

[11] G. Reza, S. Md. Nasir, I. Hamidah, and M. Norwati, "A survey: Clustering ensembles techniques," *Proceedings of World Academy of Science, Engineering and Technology*, vol. 38, pp. 644–653, 2009.

[12] S. Bickel and T. Scheffer, "Estimation of mixture models using co-em," in *ECML*, 2005, pp. 35–46.

[13] J. C. Bezdek, "Pattern Recognition with Fuzzy Objective Function Algorithms," *Plenum Press, New York*, 1981.

[14] A. Dempster, N. Laird, and D. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of Royal Statistical Society B*, vol. 39, pp. 1–38, 1977.

[15] B. Kulis, S. Basu, I. Dhillon, and R. Mooney, "Semi-supervised graph clustering: a kernel approach," in *ICML '05: Proceedings of the 22nd international conference on Machine learning*. New York, NY, USA: ACM, 2005, pp. 457–464.