Co-Regularized PLSA for Multi-View Clustering

Yu Jiang, Jing Liu, Zechao Li, Peng Li, Hanqing Lu

National Laboratory of Pattern Recognition, Institute of Automation,
Chinese Academy of Sciences, Beijing 100190, China
{yjiang, jliu, pli, luhq}@nlpr.ia.ac.cn, zechao.li@gmail.com

Abstract. Multi-view data is common in a wide variety of application domains. Properly exploiting the relations among different views is helpful to alleviate the difficulty of a learning problem of interest. To this end, we propose an extended Probabilistic Latent Semantic Analysis (PLSA) model for multi-view clustering, named Co-regularized PLSA (CoPLSA). CoPLSA integrates individual PLSAs in different views by pairwise co-regularization. The central idea behind the co-regularization is that the sample similarities in the topic space from one view should agree with those from another view. An EM-based scheme is employed for parameter estimation, and a local optimal solution is obtained through an iterative process. Extensive experiments are conducted on three real-world datasets and the compared results demonstrate the superiority of our approach.

1 Introduction

Most learning problems in real world applications often involve rich data comprising multiple information modalities. These different modalities, which reveal the fundamental attributes and properties of the studied objects from different perspectives, are generally considered as different views of the objects. A classic example is a document or a news report in multiple languages. Image is another example, which can be described with many types of visual features, such as color, texture and shape. However, the information from individual view is usually a kind of unilateral or partial reflection of data properties. We believe that the proper exploration of the complementary information (or relations) among different views is vertical to boost the performance of learning problems.

As one of the fundamental learning problems, clustering can also benefit from the rich information across multiple views. The goal of clustering is to determine the intrinsic grouping in a set of unlabeled data. Traditional clustering methods usually take multiple views as a flat set of variables and ignore the complementary information among them. Different from traditional clustering methods, the so-called multi-view clustering [1], [2], [3], [4], [5], [6], [7], which exploits relations among multiple views in order to improve the clustering performance, is gradually on the rise in the past decade. Three families of algorithms might be identified: the first aims at doing an early fusion of the multi-view information; the second is based on the late fusion of clusters estimated independently in each view; and the third attempts to build a clustering model with integrating the multi-view information into the model. To some extent, the third one is more reasonable and possibly effective to achieve a better clustering performance from a global view. It is also our focus in this paper.

Probabilistic Latent Semantic Analysis (PLSA) [8], [9] is a kind of generative model, in which latent topics are imported to model the document-word co-occurrence. When the latent topics correspond to the document clusters, the occurrence of a document given a topic can be used to solve the document clustering problem [10], [6]. Although PLSA achieves great success as a topic model, it cannot deal with multi-view data well. How to effectively explore multi-view data in PLSA-based framework motivates this work.

In this paper, we propose Co-regularized Probabilistic Latent Semantic Analysis (CoPLSA), a novel topic model for multi-view clustering. The proposed CoPLSA attempts to perform PLSAs in different views collaboratively through a constraint of pairwise co-regularization. The underlying assumption behind our algorithm is that the pairwise similarities of samples on probability distribution of topics are consistent across the different views. That is to say, if two samples are close (or far away) to each other in the topic space of one view, they should be also close (or far away) in the topic space of another view. An EM-based alternative optimization algorithm is adopted to solve the joint optimization problem. Experimental results on three real-world datasets verify the effectiveness of our method.

The remainder of this paper is organized as follows. In Section 2, we overview some related work on multi-view clustering. In Section 3, we give a brief review of PLSA. Section 4 elaborates the formulation and the model estimation for CoPLSA. The experimental evaluations and discussions are presented in Section 5. Section 6 concludes this paper with future research directions.

2 Related Work

A number of multi-view clustering algorithms have been proposed in the past decade. According to the different level where the algorithms exploit the multi-view information, these proposed algorithms can be categorized as feature-level, decision-level, and model-level multi-view clustering algorithms.

Feature-level algorithms try to extract a set of fused features from multiple views and then perform an off-the-shelf clustering algorithm such as K-means with this feature set. Conventional feature fusion methods simply concatenate or integrate several kinds of features together. Due to ignoring the complementary information from multiple views, performance adopting such feature fusion may be not better (or even worse) than that using single-view features. In order to exploit the relationship of multiple views, Chaudhuri et al. [2] considered clustering on lower dimensional subspace of the multi-view data, projected via Canonical Correlation Analysis (CCA). Feature-level algorithms just do an early fusion of the multi-view information, while decision-level algorithms do a late fusion. Decision-level algorithms first perform clustering with each view data respectively, and then combine the individual results by a certain strategy to produce a final partition. Greene and Cunningham [4] developed a new unsupervised algorithm based on matrix factorization to group related clusters produced on individual views. Long et al. [5] introduced a mapping function to make the different patterns from different spaces comparable and hence an optimal pattern can be learned. Kim et al. [6] presented a PLSA-based multi-view clustering approach, which first groups documents with voting patterns assigned by view-specific PLSA, and then assigns unclustered documents to the groups using a constrained PLSA.

Different from feature-level and decision-level algorithms, model-level algorithms integrate the multi-view information into the process of model estimation. In such a way, the complementary information may be exploited better. Bickel and Scheffer [1] proposed a general multi-view EM algorithm based on the co-EM framework. However, this method cannot guarantee to converge. Sa [3] approached the problem of two-view clustering based on "minimizing-disagreement" idea and performed spectral clustering on a bipartite graph. Kumar et al. [7] proposed a multi-view clustering approach in the framework of spectral clustering. The approach uses the philosophy of co-regularization to make the clusterings in different views agree with each other. It's worth to mention that our proposed method CoPLSA is also a model-level multi-view clustering algorithm. CoPLSA combines individual PLSAs and models documents collaboratively in different views by a constraint of pairwise co-regularization which contains the structure information from multiple views.

3 A Brief Review of PLSA

Probabilistic Latent Semantic Analysis (PLSA) models each word in a document as a sample from a mixture model, where the mixture components are multinomial random variables that can be viewed as representations of "topics". Thus each word is generated from a single topic, and different words in a document may be generated from different topics. Each document is represented as a list of mixing proportions for these mixture components and thereby reduced to a probability distribution on a fixed set of topics.

As a generative model for document-word co-occurrences, PLSA is defined by the following scheme:

1) select a document d_i with probability $p(d_i)$,
2) pick a latent topic z_k with probability $p(z_k|d_i)$,
3) generate a word w_j with probability $p(w_j|z_k)$.

As a result one obtains an observation pair (d_i, w_j), while the latent topic variable z_k is discarded. Translating the data generation process into a joint probability model results in the following expressions:

$$p(d_i, w_j) = p(d_i)p(w_j|d_i), \tag{1}$$

$$p(w_j|d_i) = \sum_{k=1}^{K} p(w_j|z_k)p(z_k|d_i). \tag{2}$$

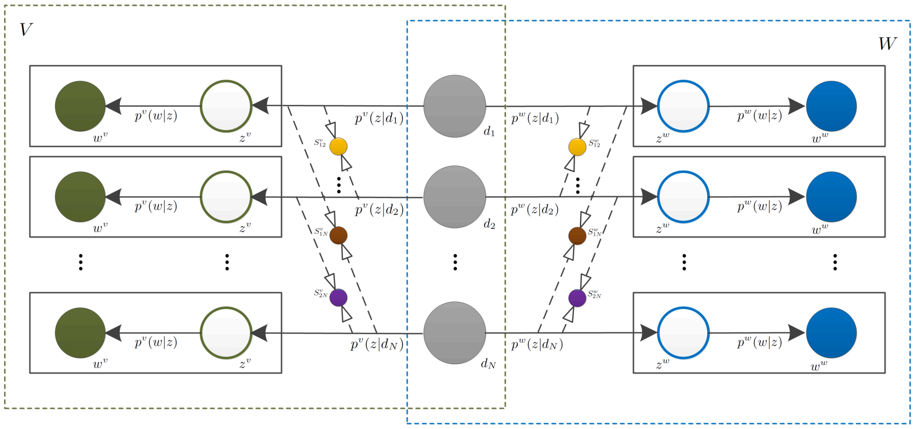

Fig. 1. The structure of CoPLSA. The small circles with same colors means CoPLSA encourages the pairwise similarities based on topic distribution as consistent as possible across different views.

Following the likelihood principle, one determines $p(d_i)$, $p(z_k|d_i)$, and $p(w_j|z_k)$ by maximizing the following log likelihood function:

$$\begin{aligned}\mathcal{L}(\Psi) &= \sum_{i=1}^{N}\sum_{j=1}^{M} c(d_i, w_j)\log p(d_i, w_j) \\ &\propto \sum_{i=1}^{N}\sum_{j=1}^{M} c(d_i, w_j)\log \sum_{k=1}^{K} p(w_j|z_k)p(z_k|d_i)\end{aligned} \tag{3}$$

where $\Psi = \{p(w_j|z_k), p(z_k|d_i)\}$ is the set of model parameters, and $c(d_i, w_j)$ denotes the term frequency, i.e., the number of times w_j occurred in d_i.

The above optimization problem can be solved by standard Expectation Maximization (EM) algorithm [11].

4 Co-Regularized PLSA

PLSA is a typical generative model for document-word co-occurring analysis, which is originally designed for the case of single-view data. To collaboratively leverage information contained in multiple views, we propose our method Co-Regularized PLSA (CoPLSA). We first start with two-view data for simplicity. Then we further extend the two-view case to a more general case where data has more than two views.

4.1 Two-View CoPLSA

Figure 1 shows the structure of CoPLSA in the two-view case. There are N samples (also called documents). Suppose we have two document-word co-occurrence tables, i.e., $C^v_{N\times M^v}$ and $C^w_{N\times M^w}$, which are representations of documents on view V and view W individually. CoPLSA models documents in different views collaboratively with the constraint that pairwise similarities of documents across different view topic spaces are encouraged as consistent as possible. Formally, CoPLSA is modeled as maximizing the following objective function:

$$\mathcal{O}(\Psi^v, \Psi^w) = \tau^v \mathcal{L}(\Psi^v) + \tau^w \mathcal{L}(\Psi^w) - \lambda \mathcal{R} \tag{4}$$

where $\mathcal{L}(\Psi^v)$ and $\mathcal{L}(\Psi^w)$ are the log likelihood functions of PLSA on view V and view W, respectively. The parameter $\tau^v(\tau^w)$ is the weight of view $V(W)$, which satisfy the constraint

$$\tau^v + \tau^w = 1 \tag{5}$$

The parameter λ trades off two PLSA log likelihood objectives and the pairwise co-regularization $\mathcal{R}$ which bridges these two individual views together. The form of $\mathcal{R}$ is defined as

$$\mathcal{R} = ||S^v - S^w||_F^2 \tag{6}$$

where $S^v(S^w)$ is a $N \times N$ pairwise similarity matrix in the topic space of view V (W). The element S^v_{ij} (similar to S^w_{ij}) in S^v which measures the similarity between the i-th document and the j-th document in the topic space of view V could be defined by many forms. Here we adopt Euclidean distance to measure the distance between two documents in the topic space and choose Gaussian kernel as our similarity measure defined in Eq. 7.

$$S^v_{ij} = exp(-\frac{\sum_k (p^v(z_k|d_i) - p^v(z_k|d_j))^2}{\sigma}) \tag{7}$$

The joint optimization problem given by Eq. 4 can be iteratively solved with the following two maximization problems:

1) Problem P_1: fix $\Psi^w = \hat{\Psi}^w$, and solve the reduced problem $\mathcal{O}(\Psi^v, \hat{\Psi}^w)$;

2) Problem P_2: fix $\Psi^v = \hat{\Psi}^v$, and solve the reduced problem $\mathcal{O}(\hat{\Psi}^v, \Psi^w)$.

We summarize the parameter estimation algorithm in Alg. 1. Since P_1 and P_2 are completely symmetrical, we give a detailed analysis of P_1 for concise. For a given $\Psi^w = \hat{\Psi}^w$, we get the following reduced optimization problem in view V:

$$\mathcal{O}(\Psi^v, \hat{\Psi}^w) = \tau^v \mathcal{L}(\Psi^v) - \lambda \mathcal{R} \tag{8}$$

The optimization problem $\mathcal{O}(\Psi^v, \hat{\Psi}^w)$ adopts the same generative scheme as that of PLSA. The standard procedure for maximum likelihood estimation in latent variable models such as PLSA is the Expectation Maximization (EM) algorithm. EM alternates two steps:

1) an expectation (E) step, where posterior probabilities are computed for the latent variables, based on the current estimates of the parameters;

2) a maximization (M) step, where parameters are updated based on maximizing the so-called expectation of complete-data log likelihood which depends on the posterior probabilities computed in the E-step.

Algorithm 1 Algorithm of CoPLSA

Input:

Two view feature matrices $C^v_{N\times M^v}$ and $C^w_{N\times M^w}$,

Termination condition value ε

Output:

$\Psi^v_n = \{p^v_n(z_k|d_i), p^v_n(w_j|z_k)\}$,

$\Psi^w_n = \{p^w_n(z_k|d_i), p^w_n(w_j|z_k)\}$

```
Initialize Ψ_0^v, Ψ_0^w;
n ← 1;
while true do
    P_1: Keep Ψ_{n-1}^w fixed, update Ψ_n^v as Alg. 2
    P_2: Keep Ψ_n^v fixed, update Ψ_n^w similar as Alg. 2
    if O(Ψ_n^v, Ψ_n^w) − O(Ψ_{n-1}^v, Ψ_{n-1}^w) ≤ ε then
        break;
    else
        n ← n + 1;
    end if
end while
```

E-step:

The optimization problem $\mathcal{O}(\Psi^v, \hat{\Psi}^w)$ has exactly the same E-step as that of PLSA. The posterior probabilities for the latent variables $p^v_n(z_k|d_i, w_j)$ are computed by simply applying Bayes' formula:

$$p^v_n(z_k|d_i, w_j) = \frac{p^v_{n-1}(w_j|z_k)p^v_{n-1}(z_k|d_i)}{\sum_k p^v_{n-1}(w_j|z_k)p^v_{n-1}(z_k|d_i)} \tag{9}$$

M-step:

Using a similar derivation to PLSA, we have the following relevant part of the expectation of complete-data log likelihood function for $\mathcal{O}(\Psi^v, \hat{\Psi}^w)$, where for convenience of discussion, we also add the Lagrange multipliers corresponding to the constraints on parameters:

$$\begin{aligned}\mathcal{Q}(\Psi^v, \hat{\Psi}^w) &= \tau^v \mathcal{F}(\Psi^v) - \lambda\mathcal{R} \\ &\quad + \sum_i \alpha_i(1 - \sum_k p^v(z_k|d_i)) + \sum_k \beta_k(1 - \sum_c p^v(w_c|z_k))\end{aligned} \tag{10}$$

$$\mathcal{F}(\Psi^v) = \sum_{i=1}^{N}\sum_{j=1}^{M^v} c^v(d_i, w_j) \sum_{k=1}^{K}(p^v(z_k|d_i, w_j)\log(p^v(w_j|z_k)p^v(z_k|d_i))) \tag{11}$$

where $\mathcal{F}(\Psi^v)$ is the expectation of complete-data log likelihood function of PLSA in view V. $\sum_i \alpha_i(1 - \sum_k p^v(z_k|d_i))$ and $\sum_k \beta_k(1 - \sum_c p^v(w_c|z_k))$ are the Lagrange multipliers corresponding to the constraints that $\sum_k p^v(z_k|d_i) = 1$ and $\sum_c p^v(w_c|z_k) = 1$. Through derivation, we obtain the solutions to re-estimate the parameter $\Psi^v_n = \{p^v_n(w|z), p^v_n(z|d)\}$ as follows:

$$p^v_n(w_j|z_k) \propto \sum_i c^v(d_i, w_j)p^v_n(z_k|d_i, w_j) \tag{12}$$

$$p_n^v(z_k|d_i) \propto \tau^v \sum_j c^v(d_i, w_j) p_n^v(z_k|d_i, w_j) - \lambda \frac{\partial R}{\partial p_{n-1}^v(z_k|d_i)} p_{n-1}^v(z_k|d_i) \quad (13)$$

The estimation of $p_n^v(w_j|z_k)$ does not rely on the pairwise co-regularization, thus the solution is the same as PLSA. For $p_n^v(z_k|d_i)$, there is an extra part $-\lambda \frac{\partial R}{\partial p_{n-1}^v(z_k|d_i)} p_{n-1}^v(z_k|d_i)$ compared to PLSA. This extra part propagates complementary information between view V and view W, where $\frac{\partial R}{\partial p^v(z_k|d_i)}$ is the derivative of pairwise co-regularization R w.r.t $p^v(z_k|d_i)$ and defined in Eq. 14.

$$\begin{aligned} \frac{\partial R}{\partial p^v(z_k|d_i)} &= 2\sum_j \frac{\partial (S_{ij}^v - S_{ij}^w)^2}{\partial p^v(z_k|d_i)} \\ &= -\frac{8}{\sigma} \sum_j S_{ij}^v (S_{ij}^v - S_{ij}^w)(p^v(z_k|d_i) - p^v(z_k|d_j)) \end{aligned} \quad (14)$$

The updating rules of Ψ_n^v are detailed in Alg. 2.

Algorithm 2 Update rule of Ψ_n^v

Input:

Feature matrix of V $C_{N \times M^v}^v$,

The result of last iteration Ψ_{n-1}^v, Ψ_{n-1}^w

Output:

$\Psi_n^v = \{p_n^v(z_k|d_i), p_n^v(w_j|z_k)\}$

1: **E step:**
2: Computer $p_n^v(z_k|d_i, w_j)$ as in Eq. 9
3: **M step:**
4: Computer $p_n^v(w_j|z_k)$ as in Eq. 12;
5: Computer $p_n^v(z_k|d_i)$ as in Eq. 13;
6: Make normalization to $p_n^v(w_j|z_k)$ and $p_n^v(z_k|d_i)$.

4.2 Extension to Multiple Views

CoPLSA proposed in the Section 4.1 can be extended to more than two views, which is a more general situation in practice. This can be done by employing pair-wise co-regularization on each two-view pair. For M views, we have

$$\mathcal{O}(\Psi^1, ..., \Psi^M) = \sum_{v=1}^{M} \tau^v L^v - \lambda \sum_{1 \le v \ne w \le M} R^{vw} \quad (15)$$

We use a common λ for all pair-wise co-regularizations for simplicity of exposition, however different λ's can be used for different pairs of views. Similar to the two-view case, we can optimize it by alternative maximization. With all but one Ψ^v fixed, we have the following optimization problem:

$$\mathcal{O}(\hat{\Psi}^1, ...\Psi^v, ..., \hat{\Psi}^M) = \tau^v \mathcal{L}(\Psi^v) - \lambda \sum_{1 \le v \ne w \le M} R^{vw} \quad (16)$$

The re-estimation equations of $p_n^v(z_k|d_i, w_j)$, $p_n^v(w_j|z_k)$ are quite the same as Eq. 9 and Eq. 12. As for $p_n^v(z_k|d_i)$, we have the updating solution as:

$$p_n^v(z_k|d_i) \propto \tau^v \sum_j c^v(d_i, w_j) p_n^v(z_k|d_i, w_j) - \lambda \sum_{1 \le v \ne w \le M} \frac{\partial R^{vw}}{\partial p_{n-1}^v(z_k|d_i)} p_{n-1}^v(z_k|d_i) \tag{17}$$

This optimization process is cycled over all views.

The computational complexity is very important for the algorithm efficiency. For CoPLSA, the complexity of each iteration is $O(KNM(\bar{W} + N))$, where K is the number of topics, N is the number of samples, M is the number of views and $\bar{W}$ is the average size of different views' dictionaries.

5 Experiment

5.1 Datasets

To investigate the performance of CoPLSA for clustering, experiments are conducted on three real-world datasets. We give a brief description of each dataset as follows:

Reuters Multilingual dataset [12]: This collection contains documents originally written in five different languages (English, French, German, Spanish and Italian), and their translations, over a common set of 6 categories. We random sample 10% documents from 18,758 documents originally in English. We use their original representation as the first view, their Spanish translation as the second view and their French translation as the third view. The vocabulary size of English is 21,531, while that of Spanish and French are 11,547 and 24,893 respectively.

NUS-Wide Object dataset [13]: As a subset of NUS-Wide data, it consists of 31 object categories (e.g., computer, toy and bear) and 30,000 images in total. We eliminate images which belong to more than one category. Then we randomly sample 20% images from the remaining 23,953 images for model learning. 216-dimensional RGB histogram and 500-dimensional SIFT bag-of-words are extracted as the first and second view. 256-dimensional LBP feature are used as the third view.

Corel5K dataset [14]: This set contains 5,000 images with 50 groups, such as train, mountain and tiger. Each group is composed of 100 images. We extract the same features as NUS-Wide Object dataset.

5.2 Evaluation Criterion

Two metrics, the accuracy (ACC) and the normalized mutual information (NMI) are used to measure the clustering performance [15] [16].

Accuracy (ACC): Given a document d_i, let l_i and r_i be the cluster label and the label provided by the document corpus, respectively. The ACC is defined as follows:

$$ACC = \frac{\sum_{i=1}^{n} \delta(r_i, map(l_i))}{n} \tag{18}$$

where n denotes the total number of documents in the test, $\delta(x, y)$ is the delta function that equals one if $x = y$ and equals zero otherwise, and $map(l_i)$ is the mapping function that maps each cluster label l_i to the equivalent label from the document corpus. The best mapping can be found by using the Kuhn-Munkres algorithm [17].

Normalized Mutual Information (NMI): Let C denote the set of clusters obtained from the ground truth and $\hat{C}$ obtained from our algorithm. Their mutual information metric $MI(C, \hat{C})$ is defined as follows:

$$MI(C, \hat{C}) = \sum_{c_i \in C, \hat{c}_j \in \hat{C}} p(c_i, \hat{c}_j) \log_2 \frac{p(c_i, \hat{c}_j)}{p(c_i)p(\hat{c}_j)} \tag{19}$$

where $p(c_i)$ and $p(\hat{c}_j)$ are the probabilities that a document arbitrarily selected from the corpus belongs to the clusters c_i and $\hat{c}_j$, respectively, and $p(c_i, \hat{c}_j)$ is the joint probability that the arbitrarily selected document belongs to the clusters c_i as well as $\hat{c}_j$ at the same time. In our experiments, we use the normalized mutual information NMI as follows:

$$NMI(C, \hat{C}) = \frac{MI(C, \hat{C})}{\max(H(C), H(\hat{C}))} \tag{20}$$

where $H(C)$ and $H(\hat{C})$ are the entropies of C and $\hat{C}$, respectively. It is easy to check that $NMI(C, \hat{C})$ ranges from 0 to 1. $NMI = 1$ if the two sets of clusters are identical, and $NMI = 0$ if the two sets are independent.

5.3 Compared Scheme

To validate the performance of the proposed CoPLSA, we compare it with a number of baselines, which are listed as follows:

K-means with the best one view feature(K-means + BOVF): running K-means with the view which contains most information.

K-means with concatenation of two view features(K-means + CTVF): running K-means with concatenation of the first two view features.

PLSA with the best one view feature(PLSA + BOVF): running PLSA with the view which contains most information.

PLSA with concatenation of two view features(PLSA + CTVF): running PLSA with concatenation of the first two view features.

CCA-based Multi-View Clustering [1] (CCAMC) [2]: applying CCA to fuse features from the first two views and then running K-means.

Vote-based PLSA [2] (VPLSA) [6]: a PLSA-based multi-view clustering method which uses clustering results obtained on each view as a voting pattern.

Pairwise Co-regularized Spectral Clustering (PCoSC) [7]: an improved multi-view spectral approach with pairwise co-regularization.

[1] Since the feature dimension of the Reuters Multilingual dataset is too high for CCA, we first reduce it to 100-dimension by PLSA before performing CCAMC.

[2] Due to working in two view case, in the stage of single-view clustering, only when two vote patterns are agreed, the document is assigned.

Table 1. Clustering performance

	Reuters Multilingual		NUS-Wide Object		Corel5K	
	ACC(%)	NMI(%)	ACC(%)	NMI(%)	ACC(%)	NMI(%)
K-means+BOVF	33.6 ± 4.0	7.2 ± 5.6	12.5 ± 0.5	10.0 ± 0.3	18.4 ± 0.5	28.2 ± 0.4
K-means+CTVF	29.9 ± 2.2	4.0 ± 3.7	12.6 ± 0.6	9.8 ± 0.2	21.1 ± 0.5	32.4 ± 0.3
PLSA+BOVF	45.4 ± 3.5	27.1 ± 2.6	12.2 ± 0.4	10.5 ± 0.2	21.3 ± 0.2	31.1 ± 0.8
PLSA+CTVF	42.3 ± 5.0	24.3 ± 2.9	10.9 ± 0.2	10.1 ± 0.3	21.9 ± 0.3	33.2 ± 0.7
CCAMC	32.4 ± 3.0	10.9 ± 2.7	13.1 ± 0.4	8.8 ± 0.3	18.7 ± 0.8	24.8 ± 0.6
VPLSA	46.8 ± 5.0	29.8 ± 2.6	12.4 ± 0.5	11.2 ± 0.3	22.8 ± 0.1	$\mathbf{34.7} \pm 0.9$
PCoSC	42.8 ± 1.5	32.0 ± 0.9	10.4 ± 0.1	9.9 ± 0.1	18.0 ± 0.5	34.2 ± 0.4
CoPLSA(2)	49.4 ± 3.6	31.8 ± 2.7	13.8 ± 0.6	$\mathbf{11.6} \pm 0.4$	$\mathbf{23.5} \pm 0.7$	33.6 ± 0.6
CoPLSA(3)	$\mathbf{50.2} \pm 4.3$	$\mathbf{33.8} \pm 3.6$	$\mathbf{14.5} \pm 0.8$	10.7 ± 0.2	22.9 ± 0.3	32.3 ± 0.7

5.4 Experimental Analysis

In this paper, we focus on clustering problem. The true underlying topics extracted by CoPLSA can be regarded as clusters. The estimated conditional probability distribution function $p(z_k|d_i)$ can be used to infer the cluster label of each document. In order to speed up, in implementation of CoPLSA, we randomly select 5% sample pairs to construct the similarity matrix $S^v(S^w)$. In other words, only 5% pairwise similarities of samples are kept consistent across different views in parameter estimation. In addition, all the view weight τ's in our experiment are set to 0.5 in two-view case, and 0.33 in three-view case for simplicity.

For the multi-view clustering algorithms (i.e., CCAMC, PCoSC and CoPLSA), we can get one cluster result for each individual view. And we report the best view [3] results in Table 1. For the Reuters Multilingual dataset, the best view is the representation of Spanish, while for the NUS-Wide Object and Corel5K datasets, the best views are both the representations of 500-dimensional SIFT bag-of-words. Moreover, the final performance scores for all the methods are obtained by averaging the scores over 20 tests with random initializations.

From the compared results in Table 1, we can obtain some observations as follows. For the Reuters Multilingual dataset, CoPLSA with two views achieves significant improvement compared with all the baselines according to ACC, and the performance of CoPLSA with three views is further promoted in terms of ACC and NMI. Besides, rather than improving the performance, the simple feature concatenation results in even worse scores than that of single view. For the NUS-Wide Object dataset, the proposed CoPLSA with two views outperforms all the baselines in terms of ACC and NMI. The second best performance on ACC is attained by CCAMC. Furthermore, addition of the third view also leads to further improvement than the two view case according to ACC. For the Corel5K dataset, CoPLSA with two views performs better than others on the basis of ACC. Moreover, we note that the addition of the third view does not

[3] In our experiments, we decide which view is the best view according to the single view performance of K-means. In practice, it may be judged by some prior information.

always improve the performance. This may be attributed to the reason that the added view is not informative or contains too much noise. To sum up, all of these experiments demonstrate the effectiveness of our approach.

The regularization parameter λ balances the weight of PLSA log likelihood objectives and the pairwise co-regularization. We also conduct an experiment by changing the values of parameter λ to validate its impact on the clustering performance. From Fig. 2, it can be observed that our CoPLSA achieves consistent good performance with λ varying in a certain range. Best performance can be achieved when λ are around 0.07, 0.03, and 0.005 for these three datasets, respectively.

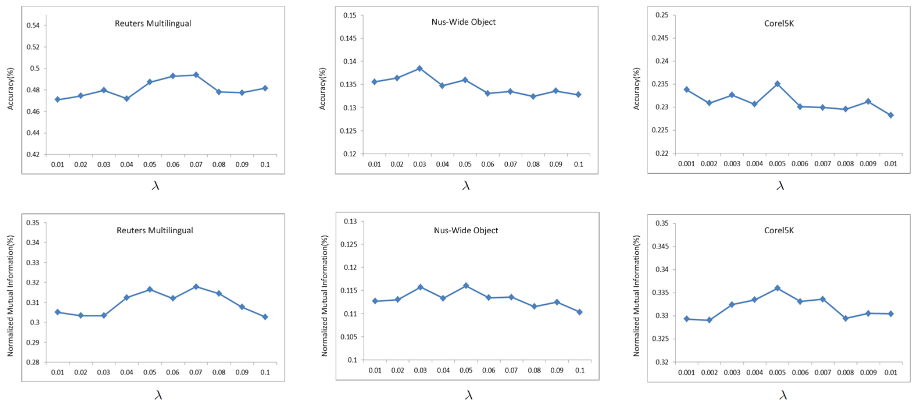

Fig. 2. The performance of CoPLSA with varying λ

6 Conclusion

In this paper, we propose Co-Regularized PLSA, a novel clustering algorithm based on PLSA for multi-view data. Our approach adopts a constraint of pairwise co-regularization to connect different PLSAs which work on different views. Given multi-view data, CoPLSA can model documents in different views simultaneously with the constraint of encouraging the pairwise similarities of samples consistent across different views. We first work with two-view case and then extend the algorithm to more than two views cases. An alternative maximization framework is adopted to solve the joint optimization problem. Experimental results on three real-world datasets demonstrate the promising performance of our method.

In future, we will extend the proposed framework to the case where some of views have missing data. Moreover, CoPLSA can also be applied to other unsupervised problems such as dimensionality reduction, topic extraction, etc.

Acknowledgement. This work was supported by 973 Program (Project No. 2012CB316304) and National Natural Science Foundation of China (Grant No. 60835002, 61070104, 60903146).

References

1. Bickel, S., Scheffer, T.: Multi-View Clustering. In: IEEE International Conference on Data Mining. (2004) 19–26
2. Chaudhuri, K., Kakade, S.M., Livescu, K., Sridharan, K.: Multi-view clustering via canonical correlation analysis. In: International Conference on Machine Learning. (2009) 17–136
3. de Sa, V.R.: Spectral clustering with two views. In: International Conference on Machine Learning. (2005) 20–26
4. Greene, D., Cunningham, P.: A Matrix Factorization Approach for Integrating Multiple Data Views. In: Principles of Data Mining and Knowledge Discovery. (2009) 423–438
5. Long, B., Yu, P.S.: A General Model for Multiple View Unsupervised Learning. In: SIAM International Conference on Data Mining. (2008) 822–833
6. Kim, Y.M., Amini, M.R., Goutte, C., Gallinari, P.: Multi-view clustering of multilingual documents. In: Research and Development in Information Retrieval. (2010) 821–822
7. Kumar, A., Rai, P., III, H.D.: Co-regularized multi-view spectral clustering. In: Advances in Neural Information Processing Systems. (2011) 1413–1421
8. Hofmann, T.: Probabilistic Latent Semantic Analysis. In: Uncertainty in Artificial Intelligence. (1999) 289–296
9. Hofmann, T.: Probabilistic latent semantic indexing. In: Research and Development in Information Retrieval. (1999) 50–57
10. Cai, D., Mei, Q., Han, J., Zhai, C.: Modeling hidden topics on document manifold. In: International Conference on Information and Knowledge Management. (2008) 911–920
11. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the em algorithm. (1977)
12. Amini, M.R., Usunier, N., Goutte, C.: Learning from multiple partially observed views - an application to multilingual text categorization. In: Advances in Neural Information Processing Systems. (2009) 28–36
13. Chua, T.S., Tang, J., Hong, R., Li, H., Luo, Z., Zheng, Y.: NUS-WIDE: a real-world web image database from National University of Singapore. In: Conference on Image and Video Retrieval. (2009) 1–9
14. Duygulu, P., Barnard, K., Freitas, J.F.G.D., Forsyth, D.A.: Object Recognition as Machine Translation: Learning a Lexicon for a Fixed Image Vocabulary. (2002)
15. Xu, W., Liu, X., Gong, Y.: Document clustering based on non-negative matrix factorization. In: Research and Development in Information Retrieval. (2003) 267–273
16. Zechao Li, Yi Yang, J.L.X.Z.H.L.: Unsupervised Feature Selection Using Nonnegative Spectral Analysis. In: National Conference on Artificial Intelligence. (2012) 1026–1032
17. Lovsz, L., Plummer, M.D.: Matching Theory. (1986)