

Adaptive Spectral Co-clustering for Multiview Data

Jeong-Woo Son, Junekey Jeon, Sang-Yun Lee, Sun-Joong Kim

Intelligent Convergence Technology Research Department

Broadcasting & Telecommunications Media Research Laboratory

Electronics and Telecommunications Research Institute

[jwson, jkjeon, syllee, kimsj]@etri.re.kr

Abstract— Spectral clustering is a typical unsupervised machine learning technique and it has widely adopted in various fields. This paper proposes a novel spectral clustering technique to handle the characteristics of multiview data. In the proposed method, co-training approach is adopted in the spectral clustering. When an instance has more than three views, it is difficult to handle different dependencies among views in ordinary co-training. To overcome this, the proposed method reflects these different dependencies among views when the information is propagated in the training phrase. In the experiment, the proposed method is evaluated with the synthetic data whose instances are represented with three views. The proposed method achieves up to 8.25% better ARI (Adjusted Rand Index) than those of five algorithms.

Keywords— Spectrum clustering, co-training, unsupervised learning, weight estimation, co-efficiency

I. INTRODUCTION

Clustering is a typical unsupervised machine learning techniques and it has been widely applied in various tasks like document clustering and image processing [1]. Especially for multimedia data, various tasks like image segmentation, scene boundary detection and so on basically utilize diverse clustering algorithms. Among various clustering techniques, the spectral clustering [2] has shown better performances in diverse tasks comparing with other clustering methods, since the spectral clustering can handle efficiently non-convex characteristics of clusters observing in real-world tasks [3]. In the spectral clustering, the similarity graph is constructed with similarities between all pairs of data. The spectral embedded data is obtained by determining Top-k eigen vectors from the graph laplacian of the similarity graph. Then, a simple clustering like k-NN is performed with the spectral embedded data. Since the spectral embedding is a kind of non-linear vector transformation, the spectral clustering can handle non-convex clusters.

When an instance has Multiview features as usually observed in real-world problems like video processing (vision, audio, and sometimes texts) and text clustering (multilingual or structured text), a clustering should reflect information both in a single view and among views. Co-training is a well-known method to handle this case, multiview data [4, 5]. In co-training, models are trained separately with each view. The trained models share the information obtained from their own

view. Then, adjusted a model to reflect the information acquired from other models (views). In many cases, co-training shows its superiority comparing with a single view based method or simple view concatenation method. Kumar and Daume III have proposed the spectral co-clustering converging with a co-training for multiview clustering problems [6].

The spectral co-clustering showed better performances to deal with multiview data. However, the spectral co-clustering assumes that independent among all the views. Thus, when an instance has more than three views, the spectral co-clustering cannot reflect the different dependencies among views. Dependencies among views are crucial points to be concerned in co-training. A co-training improves the performance of ordinary methods when views are independent and sufficient. Thus, ignorance of dependencies in the spectral co-clustering can affect the clustering performance with more than three views.

This paper proposes a novel spectral clustering algorithm, Adaptive Spectral Co-clustering (ASC) for multiview data, especially data with more than three views. In the proposed method, different dependencies among views are determined and reflected through its spectral embedding step. When a target view is selected, a new data is constructed with other views to minimize correlation coefficient by using a simple linear model. As a result, the new data has the information in other views and at the same time, it is independent with the target view. A new data is adopted in the spectral embedding of the target view. Thus, the information in other views is efficiently propagated to the target view. After the spectral embedding, a simple clustering algorithm is applied as ordinary spectral clustering.

Experiments are performed with a synthetic data sampled from the mixture of three Gaussian distributions. In experiments, ASC shows 98.4% of ARI (Adjusted Rand Index) and it is the best performance among five experimental methods. The results imply the efficiency of ASC in multiview data clustering.

II. MULTIVIEW SPECTRAL CO-CLUSTERING

The proposed method, ASC is an extended version of the multiview spectral co-clustering proposed by Kumar and Daume III [6]. Thus, the multiview spectral co-clustering is

first introduced in this section. Multiview spectral co-clustering is a spectral clustering with co-training. The similarity graphs are constructed with each view. Then, graph laplacians are determined and eigen vectors are constructed separately. Constructed eigen vectors are applied to the other view. By applying the eigen vectors of a single view to the similarity graph of the other view, information is propagated between views. Table 1 shows the algorithm of the multiview spectral co-clustering. In this algorithm, $sym()$ is the function that makes the input matrix to be symmetric. While, $concat()$ is a column-wise concatenation of input matrixes.

Table 1. The algorithm of multiview spectral co-clustering

Input: similarity matrixes K_1, K_2 Max iteration $iter$ Output: k clusters Initialization: Graph laplacians L_1, L_2 Eigen vectors U_1^0, U_2^0
for $i = 1$ to $iter$ do 1: $S_1 = sym(U_2^{i-1} U_2^{i-1T} K_1)$ 2: $S_2 = sym(U_1^{i-1} U_1^{i-1T} K_2)$ 3: compute U_1^i, U_2^i with S_1, S_2 End for 4: normalize U_1^i, U_2^i 5: $V = concat(U_1^i, U_2^i)$ 6: perform k -means clustering with V 7: return the clustering result

When more than three views are given, the Multiview spectral co-clustering considers them equally. By assuming independences among views, the 1 and 2 steps in the algorithm in Table 1 is modified as

$$S_v = sym((\sum_{v' \in -V} U_{v'}^{i-1} U_{v'}^{i-1T}) K_v), \quad (1)$$

where v is the target view, while $-V$ is the set of other views.

III. ADAPTIVE SPECTRAL CO-CLUSTERING

Equation (1) shows how the multiview spectral clustering handles information in data with more than three views. As shown in Equation (1), views are regarded equally as information sources to be shared across other views. However, in real-world tasks, even though domain experts want to design every feature to be independent, it is almost impossible to design every feature in different views have the independent characteristics. Thus, due to Equation (1), ordinary multiview spectral clustering can be affected from the dependencies among views. To resolve this, this paper proposes a new spectral co-clustering method, Adaptive Spectral Co-clustering (ASC). For the simplicity, in this paper, we address ASC with three views but the extension of ASC

for data with arbitrary numbers of views can be easily achieved with the same manner explained in this section.

Table 2. The algorithm of Adaptive Spectral Co-clustering

Input: similarity matrixes K_1, K_2, K_3 Max iteration $iter$ Output: k clusters
1: estimate w_1, w_2, w_3 , 2: construct K'_1, K'_2, K'_3 3: determine U_1^0, U_2^0, U_3^0 with K'_1, K'_2, K'_3 for $i = 1$ to $iter$ do for j in $ V $ do 4: $S_j = sym(U_j^{i-1} U_j^{i-1T} K_j)$ 5: estimate w_1, w_2, w_3 , 6: construct S'_1, S'_2, S'_3 7: compute U_1^i, U_2^i, U_3^i with S'_1, S'_2, S'_3 End for 8: normalize U_1^i, U_2^i, U_3^i 9: $V = concat(U_1^i, U_2^i, U_3^i)$ 10: perform k -means clustering with V 11: return the clustering result

Given an instance $x \in X$ which is composed of three similarity vectors based on views $\{x_1, x_2, x_3\}$, ASC generate a new similarity vector x'_1, x'_2, x'_3 for each view. In the new data generation step, for the target view v , x'_v is constructed with the weighted summation of others in $-V = \{v' | v \text{ and } v \in V, v \neq v'\}$. The weights are estimated to the new instance should be maximally independent with v . That is,

$$x'_v = \sum_{v' \in -V} w_{v'} x_{v'},$$

where $w_{v'}$ is the weight for v' . The weights are estimated based on correlation coefficient between v and v' . The correlation coefficient can be represented how much two vectors are independent. The value of the correlation coefficient is from -1 to 1. Since -1 is also equivalent meaning with 1 in our task, the weights are estimated with the squared correlation coefficient. When the target view x_v and the set of other views X_{-v} are given, the objective function for the weight estimation can be written as,

$$w_{v'}^* = \arg \min_{w \in \mathcal{W}} (bA \cdot w^T)^2 + \frac{C}{2} \|w\|^2,$$

where $\mathcal{W} \subset R^{[|V|-1, n]}$ is the set of weights and $|V|$ is the number of views, C is the user parameter for the L_2 regularizer in the right side of the equation. In this equation, b

and A is are newly introduced to measure correlation co-efficiency and they are defined as

$$b = x_v - E[x_v],$$

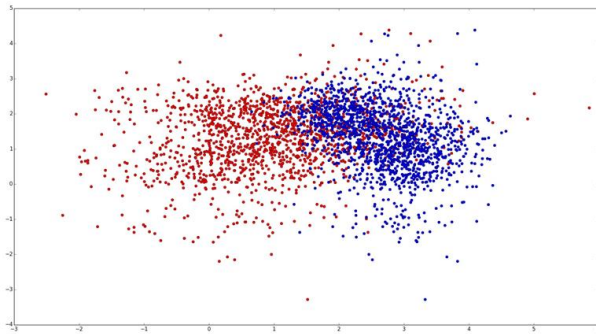
$$A = X_{-v} - E[X_{-v}].$$

Note that from the definition of X that x_v should be a similarity vector. Thus, to keep this, two constraints are defined as

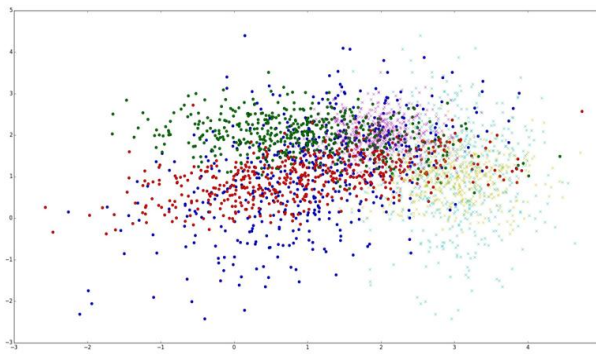
$$\forall i, 0 \leq w_{v,i} \leq 1.0,$$

$$\sum_i^{V-1} w_{v,i} = 1.0.$$

These constrains keep the value in x_v should be a values from 0.0 to 1.0. Table 2 shows the algorithm of ASC. As shown in this table, ASC is a modified version of the multiview spectral co-clustering basically. To generate new data for each view, the algorithm is expended with the weight estimation and data generation from the algorithm in Table 1.



(a) Distributions of two clusters



(b) Distribution of six Gaussian components

Figure 1. The distribution of experimental data

IV. EXPERIMENTS

A. Dataset

ASC is evaluated with the synthetic data sampled from two Gaussian distributions. Thus, the number of clusters is two and each cluster is represented with three Gaussian

components. Each Gaussian component is corresponding to a single view for an instance. Gaussian components are defined as

$$\mu_1^1 = (1, 1), \mu_1^2 = (1, 2), \mu_1^3 = (1, 1),$$

$$\Sigma_1^1 = \begin{pmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{pmatrix}, \Sigma_1^2 = \begin{pmatrix} 1.0 & -0.2 \\ -0.2 & 1.0 \end{pmatrix}, \Sigma_1^3 = \begin{pmatrix} 1.2 & 0.2 \\ 0.2 & 1.0 \end{pmatrix},$$

$$\mu_2^1 = (3, 4), \mu_2^2 = (2, 2), \mu_2^3 = (3, 3),$$

$$\Sigma_2^1 = \begin{pmatrix} 0.3 & 0.2 \\ 0.2 & 0.6 \end{pmatrix}, \Sigma_2^2 = \begin{pmatrix} 0.6 & 0.1 \\ 0.1 & 0.5 \end{pmatrix}, \Sigma_2^3 = \begin{pmatrix} 1.0 & 0.4 \\ 0.4 & 0.7 \end{pmatrix}.$$

Figure 1 shows the distributions of sampled instances for two clusters and that of three views.

Five hundred instances are sampled for a cluster and these experimental settings are exact same with Kumar and Daume III [6]. In the experiment, five methods are compared based on ARI (Adjusted Rand Index). Experimental methods are a single view k-means clustering (Singleview), kernel summation based multiview k-means (KS), kernel product based multiview k-means (KP), the multiview spectral co-clustering (MSC), the proposed method ASC.

B. Experimental results

Table 3 shows the experimental result. As shown in this table, ASC achieved 98.4% of ARI and it is the best performance among five methods. In this experiment, all the multiview based methods show better performances than that of Singleview. As mentioned in previous studies, when an instance is composed of multiple views, Multiview techniques can be improved their performances in most cases. With more than three views, ASC can handle the dependencies among views and the dependency is one of the most crucial factors in co-training. This characteristic of ASC gives rise to achieve the best performance.

Table 3. ARIs of five experimental methods

	ARI
Singleview	90.9% (0.302)
KS	97.7% (0.035)
KP	97.6% (0.035)
MSC	97.2% (0.056)
ASC	98.4% (0.069)

V. CONCLUSION

This paper proposes a novel spectral clustering to handle multiview data. Especially, when data with more than three views are given, the proposed method generates data corresponding to each view by estimating weights for views. Then, information of other views is efficiently propagated to the target view. Through experiments, the proposed method shows consistently better performances than four ordinary clustering algorithms.

ACKNOWLEDGMENT

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No. B0126-15-1002, Development of smart broadcast service platform based on the watching situation using semantic cluster to construct an open-media ecosystem).

REFERENCES

- [1] C. Bishop, "Pattern Recognition and Machine Learning," Springer, 2007.
- [2] Y. Ng, M. Jordan, and Y. Weiss. "On spectral clustering: Analysis and an algorithm." *Advances in Neural Information Processing Systems*, vol. 2 pp. 849-856, 2002.
- [3] U. Luxburg, "A tutorial on spectral clustering." *Statistics and computing*, vol. 17, no. 4, pp. 395-416, 2007.
- [4] A. Blum and T. Mitchell. "Combining labeled and unlabeled data with co-training." in *Proceedings of the eleventh annual conference on Computational learning theory*, pp. 92-100, 1998.
- [5] A. Levin, P. Viola, and Y. Freund, "Unsupervised improvement of visual detectors using cotraining," In *Proceedings of the Ninth IEEE International Conference on Computer Vision*, pp. 626-633, 2003.
- [6] A. Kumar and H. Daumé III, "A Co-training Approach for Multi-view Spectral Clustering," In *Proceedings of the 28th International Conference on Machine Learning*, pp. 393-400, 2011.



Jeong-Woo Son received his MS and Ph.D. degrees in computer engineering from Kyungpook National University, Daegu, Rep. of Korea in 2007 and 2012 respectively. Since 2013, he has been with Broadcasting & Telecommunications Media Research Lab., Electronics and Telecommunications Research Institute (ETRI), Daejeon, Rep. of Korea. He focuses on machine learning, NLP, and information retrieval.



Junekey Jeon received his BS degrees in the field of electrical engineering from KAIST, Rep. of Korea, in 2015. Since 2015, he has been a researcher with the Electronics and Telecommunications Research Institute, Daejeon, Rep. of Korea. His research interests include statistics, estimation theory, computer vision, machine learning, and information theory.



and Web technology.

Sang-Yun Lee received his MS and Ph.D. degrees in electronics and telecommunications engineering from Hanyang University, Seoul, Rep. of Korea in 1996 and 2007 respectively. Since 2011, he has been with Broadcasting & Telecommunications Media Research Lab., Electronics and Telecommunications Research Institute (ETRI), Daejeon, Rep. of Korea. He focuses on embedded software, smart TV software platform,



Sun-Joong Kim received her BS degree in computational statistics and her MS degree in computer science from Chungnam National University, Daejeon, Rep. of Korea, in 1989 and 2000 respectively. In February 1989, she joined ETRI, Daejeong, Rep. of Korea, where she is currently principal researcher and director. Her research interests includes convergence service control, smart TV, content knowledge mining.