Multiview Clustering: A Late Fusion Approach Using Latent Models

Eric Bruno
Viper group, Computer Science dpt.
University of Geneva
Geneva, Switzerland
Eric.Bruno@unige.ch

Stéphane Marchand-Maillet
Viper group, Computer Science dpt.
University of Geneva
Geneva, Switzerland
Stephane.Marchand-Maillet@unige.ch

ABSTRACT

Multi-view clustering is an important problem in information retrieval due to the abundance of data offering many perspectives and generating multi-view representations. We investigate in this short note a late fusion approach for multi-view clustering based on the latent modeling of cluster-cluster relationships. We derive a probabilistic multi-view clustering model outperforming an early-fusion approach based on multi-view feature correlation analysis.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Clustering; I.5.1 [**Models**]: Statistical

General Terms

Algorithms

Keywords

Multi-view clustering, data fusion, latent models

1. INTRODUCTION

Information retrieval has to deal with data presenting many perspectives and usually characterized by various heterogeneous sources of information. Web pages, Web communities, multimedia documents, user profiles are examples of data that can be organized into multi-graphs or decomposed onto multi-modal signals. Statistical analysis of those data collections has to deal efficiently with the various available views to infer internal relationships, structures and categories. Consequently, multi-view clustering approaches have been recently developed to derive convenient mining tools handling these complex representations. Two families of algorithms might be identified: the first aims at doing an early fusion of the multi-view information while the second is based on the late fusion of clusters estimated independently in each view. For early fusion approaches, we can cite [6, 8] who propose to build a convex combination of multi-view similarities, where mixing parameters and clusters are jointly estimated to minimize a given penalty error. Late fusion consists of first taking a decision in each view separately and then fusing all decisions arising from all views in a global model. The global model tends to define a consensus clustering by determining cluster agreements/disagreements [1, 2]. Following this idea, we investigate a latent model framework to explicit cluster-cluster relationships in term of global multi-view clustering model.

SIGIR'09, July 19–23, 2009, Boston, Massachusetts, USA.
ACM 978-1-60558-483-6/09/07.

2. AN INTER-CLUSTERS ANALYSIS

We suppose that M views on N documents are available. Each view $m = 1 \dots M$ gives rise to k^m clusters $\{c_1^m, \dots, c_{k^m}^m\}$ obtained through some clustering methods. The clustering might be represented as a $k^m \times N$ cluster-document matrix C^m, where the binary entry C_{kn}^m indicates whether the document d_n is in the cluster c_k^m. The M independent clusterings C^m are concatenated into a $K \times N$ matrix C, where $K = \sum_m k^m$ is the total number of clusters over all views. The empirical joint cluster-document probability is simply obtained with a proper normalization

$$P(c_k, d_n) = \frac{C_{kn}}{MN}, \ \forall k \in [1, K], \ \forall n \in [1, N]. \tag{1}$$

The inter-cluster analysis we propose relies on the co-occurrence relationships among multiple clusters and views. Exploiting cluster agreement and disagreement, the aim is to derive a multi-view clustering providing a consensus among all information initially available. The cluster agreement is evaluated through the joint cluster-cluster probability derived from (1) by noticing the conditional independence of observing simultaneously c_k and $c_{k'}$ given d_n

$$\begin{aligned} P(c_k, c_{k'}) &= \sum_n P(c_k|d_n) P(c_{k'}|d_n) P(d_n) \\ &= \sum_n \frac{P(c_k, d_n) P(c_{k'}, d_n)}{P(d_n)}. \end{aligned} \tag{2}$$

The joint probability $P(c_k, c_{k'})$ is defined as the fraction of documents falling simultaneously in clusters c_k and $c_{k'}$ (by definition the probability is 0 if the two clusters are from the same view). We will see in section 3 and 4 how a multi-view global clustering might be derived from this empirical joint distribution.

3. THE PROBABILISTIC LATENT CLUSTERING

The Probabilistic Latent Semantic Analysis (PLSA), proposed in [4] and originally designed for term-document matrix analysis, states the existence of binary latent variables $z_i, \ i = 1, \dots L$ in the generative process of $P(c_k, c_{k'})$. It moreover assumes the conditional independence of the co-occurrence of c_k and $c_{k'}$ given z_i

$$P(c_k, c_{k'}) = P(c_{k'}) \sum_{i=1}^{L} P(c_k|z_i) P(z_i|c_{k'}) \quad (3)$$

PLSA establishes a generative relationship between instances of clusters observed in various views and discrete variables z and thus makes explicit the absolute data distribution in a homogeneous latent space. Practically, as the latent model is estimated from the observations, it effectively fuses the sources of information.

The multi-view clustering of the data is given by assigning to every document d_n the variable z_i maximizing the *posterior* probability

$$P(z_i|d_n) = \frac{\sum_k P(z_i|c_k) P(c_k, d_n)}{P(d_n)}. \quad (4)$$

4. LATENT MODEL ESTIMATION

Expectation-Maximization algorithm (EM) is traditionally used to estimate the latent model of equation (3). From initial random probabilities, EM iteratively maximizes the log-likelihood of the model knowing the empirical distribution $P(c_k, c_{k'})$. Please refer to [4] for the complete estimation procedure.

Non-negative Matrix Factorization (NMF) [7] is an alternative solution to get estimation of the PLSA parameters [3]. NMF seeks for the factorization of the empirical distribution matrix $P_{kk'} = P(c_k, c_{k'})$ into two nonnegative matrices W and H such that $P \approx WH$. As detailed in [3], $P(c|z)$ and $P(z)$ can be easily recovered from W and H. We use the NMF implementation minimizing the Frobenius norm between the matrix and its factorization. Again, please refer to [7] for the NMF implementation.

EM and NMF are both considered and compared in our experiments to operate the probabilistic latent clustering.

5. EXPERIMENTS

The experiments have been conducted on three data sets offering a multi-view description:

- UCI Wine data set: 178 wines depicted by 13 chemical numerical attributes and falling into three categories. Each attribute is considered as a view.
- Audio genre clustering [5]: 1858 audio tracks covering 9 music genres (rock, blues, classic,...), 49 low-level audio features grouped into 15 vector spaces are provided. Each vector space is considered as a view.
- Corel image collection: 528 images distributed over 30 categories. Visual content (color, shape and texture histograms) and textual information (tf-idf indexing of tags) are considered to form 4 views on the collection.

Base clusterings $\{c_1^m, \ldots, c_{k^m}^m\}$ are computed for each view using a standard k-means with k corresponding to the right number of categories in the data set (3, 9 and 30 respectively). The number of latent variables in equation (3) is set to the same number of k (3, 9 and 30 respectively).

For comparison purpose, a baseline Mahalanobis k-means is applied over the concatenated features, with again k equals to 3, 9 and 30. Similarly to the probabilistic latent clustering which analyzes cluster co-occurrences, the Mahalanobis clustering estimates feature correlations to infer a suitable fusion model. This approach is representative of a multi-view early fusion clustering.

Table 1: Clustering accuracy (F_1 measure)

	Wine	Audio	Images
NMF	**0.85**	0.19	**0.09**
EM	0.82	**0.20**	**0.09**
Maha. k-means	0.57	0.18	0.05

Clustering accuracy, evaluated with the F_1 measure, is given in table 1. At the light of these results, it appears that the latent clustering outperforms the early fusion approach. The result is particularly clear for the easy Wine data set, but also for the difficult Corel collection. This tends to indicate that base clustering operates a pre-filtering of the data and provides a more appropriate information than the "spherized" multi-view features. As far as the latent model estimation algorithms are concerned, NMF and EM perform comparably over the three data sets.

6. CONCLUSIONS

This communication has presented a latent clustering approach based on the latent modeling of cluster-cluster relationships. A nice feature of this model is the fact that its estimation relies on a very compact coding of the complete database (the cluster-cluster contingency table), and therefore is particularly suited for large scale multi-view clustering. Beside clustering, new multi-view similarity measures based on $P(d_n, d_{n'})$ may be also estimated from the same compact representation.

7. ACKNOWLEDGMENTS

This work is funded by the Swiss National Center of Competence in Research (NCCR) on Interactive Multi-modal Information Management (IM2).

8. REFERENCES

[1] R. Bekkerman, R. El-Yaniv, and A. McCallum. Multi-way distributional clustering via pairwise interactions. In *Proceedings of the 22nd International Conference on Machine Learning (ICML)*, pages 41–48, 2005.

[2] S. Bickel and T. Scheffer. Multi-view clustering. In *ICDM '04: Proceedings of the Fourth IEEE International Conference on Data Mining*, pages 19–26, Washington, DC, USA, 2004. IEEE Computer Society.

[3] E. Gaussier and C. Goutte. Relation between PLSA and NMF and implications. In *SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 601–602, New York, NY, USA, 2005. ACM.

[4] T. Hofmann. Unsupervised learning by Probabilistic Latent Semantic Analysis. *Machine Learning*, 42(1-2):177–196, 2001.

[5] H. Homburg, I. Mierswa, K. Morik, B. Möller, and M. Wurst. A benchmark dataset for audio classification and clustering. In *Proc. ISMIR*, pages 528–531, 2005.

[6] T. Lange and J. Buhmann. Fusion of similarity data in clustering. In Y. Weiss, B. Schölkopf, and J. Platt, editors, *Advances in Neural Information Processing Systems 18*, pages 723–730. MIT Press, Cambridge, MA, 2006.

[7] D. D. Lee and H. S. Seung. Algorithms for Non-negative Matrix Factorization. In *NIPS*, pages 556–562, 2000.

[8] D. Zhou and C. J. C. Burges. Spectral clustering and transductive learning with multiple views. In *ICML '07: Proceedings of the 24th international conference on Machine learning*, pages 1159–1166, New York, NY, USA, 2007. ACM.