34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Multiview Approach to Spectral Clustering

Samir Kanaan-Izquierdo, Andrey Ziyatdinov, Raimon Massanet and Alexandre Perera

Abstract— **In this paper we propose a generic approach to the multiview clustering problem that can be applied to any number of data views and with different topologies, either continuous, discrete, graphs, or other. The proposed method is an extension of the well-established spectral clustering algorithm to integrate the information from several data views in the partition solution. The algorithm, therefore, resolves a joint cluster structure which could be present in all views, which enables researchers to better resolve data structures in data fusion problems The application of this novel clustering approach covers an extended number of machine learning unsupervised clustering problems including biomedical analysis or machine vision.**

I. INTRODUCTION

A general trend in most research areas is to integrate information from several data sources (referred to as *data views* throughout this paper) in order to attain a richer and deeper insight of the subject under study. Technical advances in instrumentation, information management and processing make this possible. For example, a clinical study on a set of pacients may include sources of information as diverse as biochemical analysis, written tests, image scans or data gathered by mobile devices among many other sources of information.

However many computational tools are constrained to use only a single source of information, which either limits its range of applications or enforces researchers to combine their data using custom functions that are not always possible or desirable. This is certainly the case with clustering algorithms, that are generally designed to operate with a single data view.

Although it is possible to combine several data views into a single space, this may not always serve the desired goal. By simply stacking several variables (i.e. dimensions) of the data it is not generally possible to *obtain a clustering that is compatible with all views individually*. Such operation would yield the most restrictive clustering: two observations would only be assigned to the same cluster if they are close to each other in all data views. On the contrary, an inclusive clustering assigns a pair of observations to the same cluster if they are close to each other in *any data view*. Obtaining an inclusive clustering thus requires to use a custom distance function or an equivalent method to alter the default behaviour of clustering algorithms with stacked data dimensions, which makes them inconvenient for this purpose.

There already exist several techniques that are somehow similar to the concept of *inclusive clustering* explained above. Consensus clustering [1] methods receive a set of clustering assignments, either from different data views, from different clustering algorithms or from different executions of the same clustering algorithm (many clustering algorithms such as k-means have a random step and thus may yield different clustering solutions on the same input data). Consensus clustering is therefore defined as the problem of finding the most compatible clustering assignment with respect to the input clusterings. While some consensus algorithms are restrictive, i.e. they penalize merging two observations that are far apart in some views, others do not.

Clustering of data with multiple views, or simply multiview clustering, is related to consensus clustering but essentially different. Multiview clustering methods simultaneously use several data views to determine the best clustering assignments that are compatible with all or most of the views of the data. According to [2], multiview clustering algorithms attain better results than consensus clustering as the former integrate all the available information. Therefore such algorithms are able to extract finer relationships between the data elements.

In this paper we propose a generic approach to the multiview clustering problem that can be applied to any number of data views, with different topologies (continuous, discrete, graphs, and so on). It is based in well proven techniques that exhibit good performance plus other useful features. This paper is organized as follows. In section II, the methods and techniques on which our method is based are described. Then in section III a set of synthetic examples is proposed and the results of applying our method are presented. Finally section IV contains the conclusions of the work described in this paper.

II. METHODS

Spectral clustering [3] is one of the most powerful clustering algorithms. It is based on the spectral graph theory [4], which studies the properties of graphs with respect to their eigenvalues and eigenvectors. Given a set of points or observations $x_1, x_2, ..., x_N$, a similarity graph is a graph $G = (V, E)$ where each vertex $v_i \in V$ represents a point x_i and the similarity between two points $s_{ij} \in E$ is represented by the edge that links them.

The goal of spectral clustering is to divide the graph into several vertex (i.e. sample) groups so that the samples in each

S. Kanaan-Izquierdo is with the Department of Software, EUETIB, Universitat Politecnica de Catalunya, Barcelona, Spain. A. Ziyatdinov, R. Massanet and A. Perera-Lluna are with the Department of Automatic Control, Universitat Politecnica de Catalunya, Barcelona, Spain, and theBiomedical Research Networking Center in Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN) {samir.kanaan, andrey.ziyatdinov,raimon.massanet,alexandre.perera}@upc.edu

978-1-4577-1787-1/12/$26.00 ©2012 IEEE

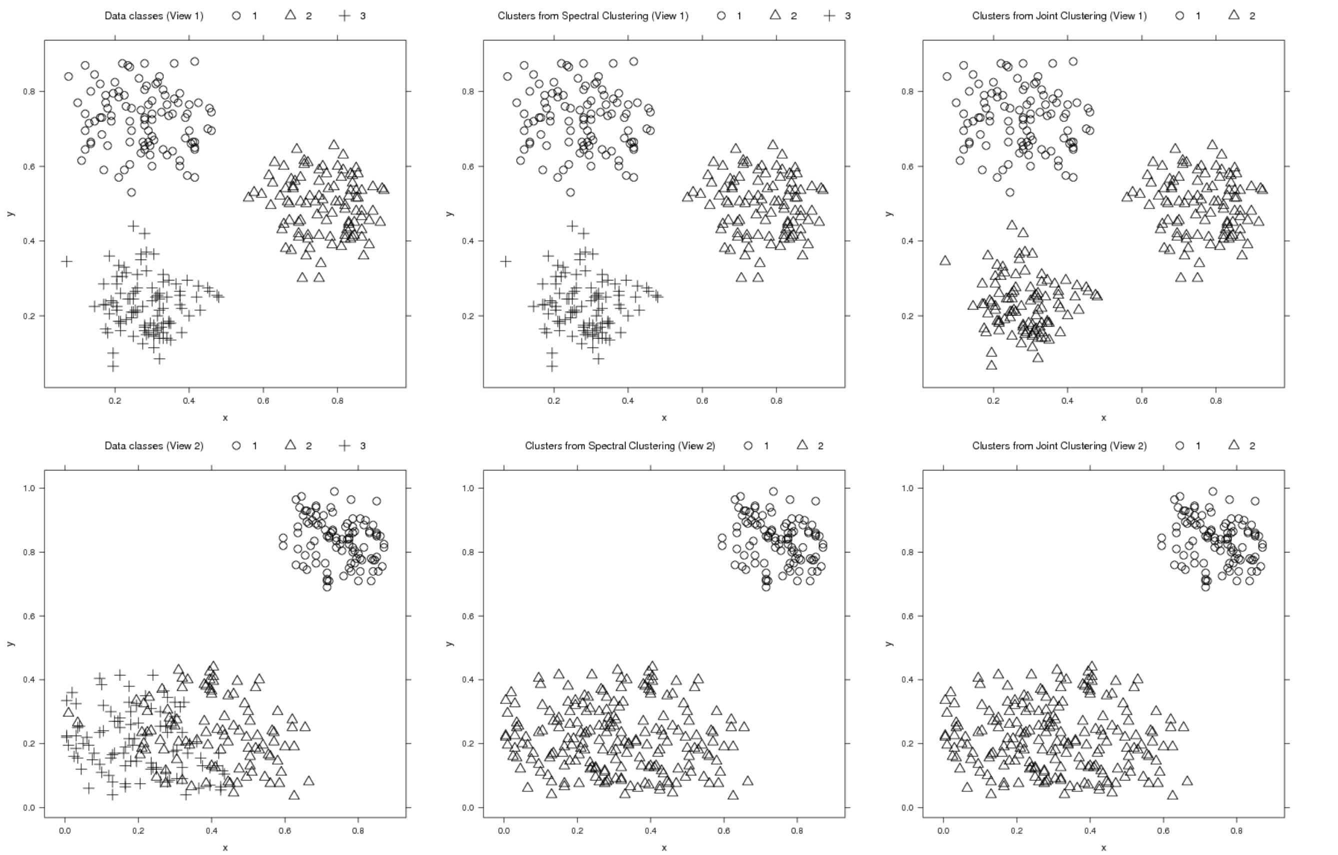

Fig. 1: Data views (first column), clusters computed independently for each view by Spectral Clustering (second column) and clusters computed jointly for all views by Multiview Spectral Clustering (third column). Example 1.

group or cluster have maximum similarity between them and minimum similarity with the samples in other clusters.

Although a comprehensive explanation of spectral clustering is beyond the scope of this paper, an overall description will be given for completeness. The input to the algorithm is a set P of n samples. First, a similarity graph G of P is obtained using a similarity metric (euclidean or Gaussian for example). Then the laplacian matrix L of the graph G is computed. Finally the eigenvalues EV and eigenvectors ET of L are obtained. According to the spectral graph theory, the number of eigenvalues in EV whose value is zero indicates the number of disconnected subgraphs in G, in other words the number of clusters in the dataset. The clustering assignments can be obtained by any ordinary clustering algorithm (k-means for example) using the eigenvector columns whose eigenvalue is zero. The theory includes a relaxation variant to cover graph partitions which do not need to be fully disconnected, which are captured through almost-zero eigenvalues. Spectral clustering outperforms other clustering algorithms when dealing with non-gaussian or, in general, non-convex clusters. While most clustering algorithms find density-based clusters (i.e. points grouped in a similar density area), spectral clustering finds connectivity-based clusters, i.e. groups of points that are connected to each other even if they spread across a wide area that would not be recognizable by density-based methods. This makes it a valuable algorithm for many clustering problems that cannot be solved by other clustering algorithms.

A. *Multiview Spectral Clustering*

This paper proposes an extension of the spectral clustering to a K-views problem through a joint eigendecomposition of K Laplacians $L_k, k = 1, 2, \cdots, K$, where L_k corresponds to the Laplacian constructed in the k^{th} view.

The proposed algorithm is referred to as *Multiview Spectral Clustering*, and takes as an input data points from K data views $x_{1,k}, x_{2,k}, \ldots, x_{i,k}, \ldots, x_{N,k}$, where first index $i = 1, 2, \ldots, N$ denotes the sample number and the second index $k = 1, 2, \ldots, K$ corresponds to one of the data views. The dimensionality and type of data (continuous, binary, graphs, etc) can differ for each data view as long as each view contains the same number of samples. K laplacian matrices L_k are computed independently per data view following one of the established strategies for spectral clustering [3]. For example, it is possible to select a similarity function as the Gaussian similarity function $s(x_{i,k}, x_{j,k}) = exp(|x_{i,k} - x_{j,k}|^2/(2\sigma^2))$ with an adjusted value of σ.

Eigendecomposition of the laplacian matrices from several data views is based on the common principal component analysis proposed by Flury in 1984 [5]. This statistical method, also known as *joint diagonalization*, attempts to diagonalize a set of positive-definite symmetric matrices simultaneously. The hypothesis of common components H_c states that exists an orthogonal matrix V such that the given matrices of K groups have the diagonal form, as formulated in Equation 1.

$$H_c : L'_k = V^T \ L_k \ V, \ k = 1, 2, ..., K \tag{1}$$

where L_k is positive-definite symmetric matrix of group k, and L'_k is its diagonalized form obtained by the linear transformation defined in matrix V. Note that the eigenvectors ET (columns of the matrix V) are common for all the groups, while the eigenvalues EV are group-specific. In practice, the hypothesis H_c is hardly feasible, and the matrices are attempted to be as diagonal as possible. We employ a new algorithm recently published by Trendafilov [6] to find an approximate solution.A the joint diagonalization of the laplacian matrices is therefore computed by performing common principal component analysis from the Equation 1. The resulting eigenvectors ET with near-zero value represent the vectors relevant to the graph partition. The final clustering assignment on the selected eigenvectors is performed by the k-means clustering algorithm. The method of joint diagonalization has already shown a good performance in processing of biomedical signals (ECG, EEG, multi-electrode neural recordings) [7] and gas sensor array data [8].

III. RESULTS

This sections reports some preliminary results of the multiview clustering algorithm applied to three synthetic examples of data points in two dimensions. Note that the strategy of barely stacking variables fails to detect proper clusters meaningful for all data views, and these results are omitted in the section.

The first example is composed of data points from three classes, separable in the first view and mixed in the second view (Figure 1, first column). The standard spectral clustering algorithm yields an independent clustering assignment for each data view, thus suggesting three clusters for the first view and two for the second one (Figure 1, second column). On the contrary, the multiview approach integrates the information from both views, thus assigning all data points to two clusters only (Figure 1, third column), that are compatible with both data views.

The second example deals with non-Gaussian data of four classes (Figure 2). It is intended to demonstrate the ability of the spectral clustering algorithm to discover non-Gaussian clusters, including convex data clusters. The multiview clustering successfully discovers the three clusters that are compatible with both data views (Figure 3).

Finally, an example of application to data with more than two views is presented. While existing multiview clustering algorithms are usually limited to data with two views, like [2], our approach can be applied to any number of data views. In this case a three-view data set is used (Figure 4), where the first view apparently has four clusters, while in the other two views either the upper or the lower clouds of points are merged. The resulting clustering algorithm yields

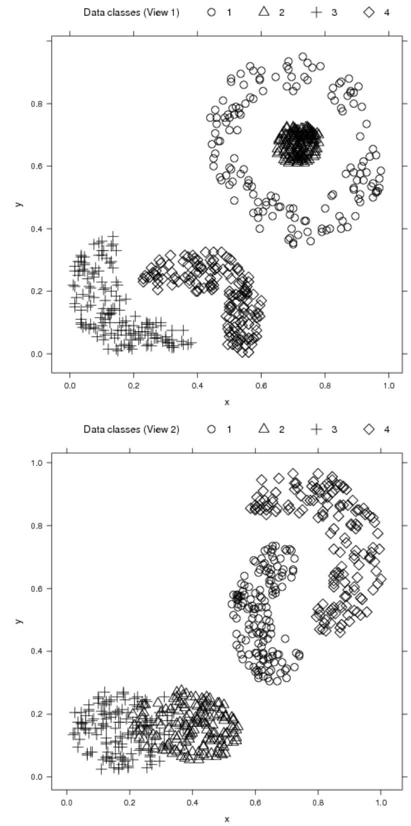

Fig. 2: Data views with four clusters in each. Example 2.

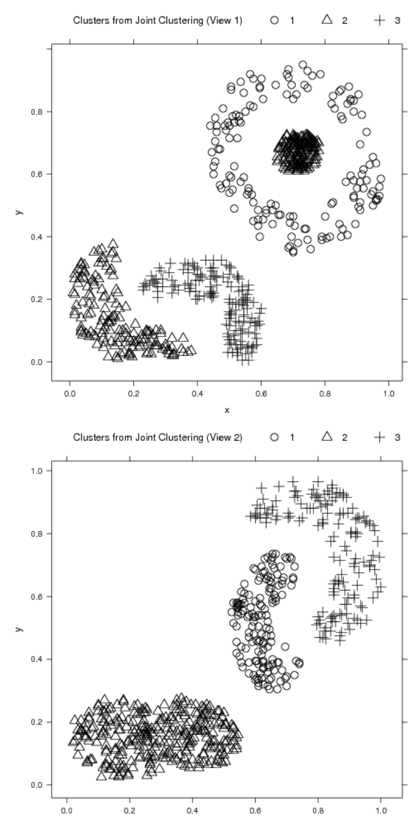

Fig. 3: Three clusters from Multiview Spectral Clustering. Example 2.

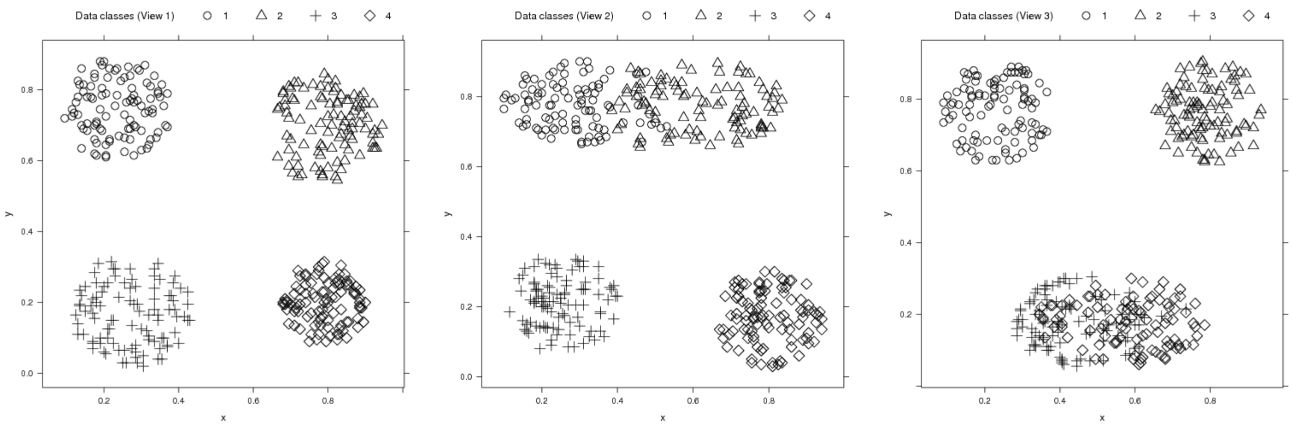

Fig. 4: Three Data views with different groupping of data points per view. Example 3.

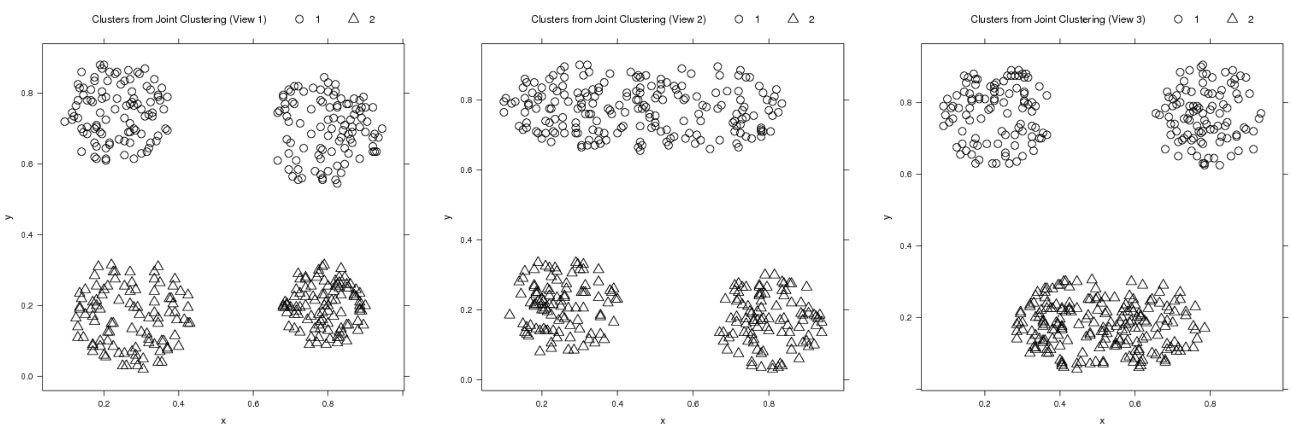

Fig. 5: Two clusters found by Multiview Spectral Clustering, which are common to all Data views. Example 3.

an assignment compatible with all three views, with the upper clouds of points assigned to a single cluster and the lower clouds assigned to another cluster (Figure 5).

IV. CONCLUSIONS

The proposed method can be viewed as a novel generic approach to the multiview clustering problem. As compared to previous proposals, our algorithm is designed to accept any number of data views. We show some results on synthetic examples with two and three views. Indeed, it can be seen as a natural extension of spectral clustering, which is a powerful and proven technique for difficult clustering problems, to the multiple data view scenario, that is increasingly important as more data sources are available to researchers. Another important advantage of this algorithm, inherited from the standard spectral clustering algorithm, is that it gives a suggestion on the number of clusters based on a statistical estimator. Finally, the proposed clustering algorithms can be naturally applied to a variety of problems from different areas, as it does not require to define ad-hoc solutions or to create custom distance functions.

ACKNOWLEDGMENT

This work was partially funded from the European Communitys Seventh Framework Programme under grant agreement no. 216916: Biologically inspired computation for chemical sensing (NEUROCHEM) and the Ramon y Cajal program from the Spanish Ministerio de Educación y Ciencia, TEC2010-20886-C02-02 and TEC2010-20886-C02-01. CIBER-BBN is an initiative of the Spanish ISCIII.

REFERENCES

[1] S. Monti, P. Tamayo, J. Mesirov, and T. Golub, "Consensus Clustering," pp. 1–34, 2003.

[2] V. R. De Sa, "Spectral Clustering with Two Views," in *Proceedings of the Workshop on Learning with Multiple Views, 22 nd ICML*, pp. 20–27, 2005.

[3] U. von Luxburg, "A Tutorial on Spectral Clustering," *Statistics and Computing*, vol. 17, no. 4, pp. 395–416, 2007.

[4] F. R. K. Chung, *Spectral Graph Theory*, vol. 92 of *CBMS Regional Conference Series in Mathematics*. American Mathematical Society, 1997.

[5] B. N. . Flury, "Common Principal Components in K Groups," *Journal of the American Statistical Association*, vol. 79, no. 388, pp. 892– 898, 1984.

[6] N. T. Trendafilov, "Stepwise estimation of common principal components," *Computational Statistics & Data Analysis*, vol. 54, pp. 3446–3457, Dec. 2010.

[7] E. O. Klaus Nordhausen, Jean-Francois Cardoso, Jari Miettinen, Hannu Oja and S. Taskinen, "JADE: JADE and other BSS methods as well as some BSS performance criteria. R package version 1.1-0.," 2012.

[8] A. Ziyatdinov, S. Marco, A. Chaudry, K. Persaud, P. Caminal, and A. Perera, "Drift compensation of gas sensor array data by common principal component analysis," *Sensors and Actuators B: Chemical*, vol. 146, pp. 460–465, Apr. 2010.