Information Processing and Management 52 (2016) 618–627

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/ipm

Multi-view clustering via spectral partitioning and local refinement

Nacim Fateh Chikhi*

Department of Computer Science, Faculty of Sciences, University of Blida 1, BP 270 Route de Soumaa, 09000 Blida, Algeria

ARTICLE INFO

Article history:
Received 30 January 2015
Revised 24 November 2015
Accepted 1 December 2015
Available online 15 January 2016

Keywords:
Multi-view clustering
Spectral clustering
Local refinement
Normalized cuts

ABSTRACT

Cluster analysis using multiple representations of data is known as multi-view clustering and has attracted much attention in recent years. The major drawback of existing multi-view algorithms is that their clustering performance depends heavily on hyperparameters which are difficult to set.

In this paper, we propose the Multi-View Normalized Cuts (MVNC) approach, a two-step algorithm for multi-view clustering. In the first step, an initial partitioning is performed using a spectral technique. In the second step, a local search procedure is used to refine the initial clustering.

MVNC has been evaluated and compared to state-of-the-art multi-view clustering approaches using three real-world datasets. Experimental results have shown that MVNC significantly outperforms existing algorithms in terms of clustering quality and computational efficiency. In addition to its superior performance, MVNC is parameter-free which makes it easy to use.

1. Introduction

In many real-world applications, datasets are characterized by multiple sets of features. Web pages and scientific papers are typical examples of such datasets where documents can be represented using not only their textual content but also other modalities such link information. Cluster analysis using multiple representations (or views) of data is known as multi-view clustering and has attracted much attention in recent years (see, e.g., Cai, Nie, & Huang, 2013; Chaudhuri, Kakade, Livescu, & Sridharan, 2009; Greene & Cunningham, 2009; Liu, Wang, Gao, & Han, 2013; Zhao, Evans, & Dugelay, 2014; Zhuang, Karypis, Ning, He, & Shi, 2012). Multi-view clustering seeks to take advantage of the complementarity of views to achieve better clustering performance than when relying on a single view. Bickel and Scheffer (2004), for example, show that exploiting both the textual content of web pages and the anchor text of inbound links improves clustering quality over the use of a single modality; Chikhi, Rothenburger, and Aussenac-Gilles (2008) show that combining text and citation information improves document clustering.

To cluster multi-view data using single-view clustering techniques (such as K-means), one has first to combine the available sets of features in an ad-hoc way to form a single view. This can be achieved either by concatenating the sets of features into a single set, or by building a similarity matrix from each view and then computing the overall affinity matrix by averaging the different similarity matrices. In practice, though, these simple combination techniques have been shown to give poor results in comparison to more elaborate techniques such as the convex K-means algorithm described in

* Tel.: +21 3557172109.
E-mail address: nacim.chikhi@univ-blida.dz

http://dx.doi.org/10.1016/j.ipm.2015.12.007
0306-4573/

(Modha & Spangler, 2003). Convex K-means is a generalized version of the classical K-means algorithm which combines views, in a convex fashion, during the assignment step. In the same vein, Zhou and Burges (2007) proposed a multi-view extension of the spectral clustering algorithm of Meila and Shi (2000), where views are combined using a mixture of Markov chains. In (Kumar, Rai, & Daume, 2011), a co-regularized approach to multi-view clustering is presented. Co-regularization consists in introducing constraints in the clustering process to ensure that the clusterings on different views agree with each other. In (Liu et al., 2013), the authors proposed an adaptation of the non-negative matrix factorization technique to work with multiple sets of features. Their algorithm uses a joint factorization process to find a consensus clustering across the views. More recently, Xia, Pan, Du, and Yin (2014) proposed a multi-view spectral algorithm based on Markov chains and noise handling. The basic idea of their algorithm is to combine the transition probability matrices constructed from each view into a shared transition probability matrix via low-rank and sparse decomposition.

The major drawback of existing multi-view clustering algorithms is that they have hyperparameters which are difficult to set and which affect significantly the clustering performance. For instance, the convex K-means algorithm (Modha & Spangler, 2003) and the mixture model of Zhou and Burges (2007) use a weighting parameter to balance the importance of each view. The approach of (Kumar et al., 2011) has a co-regularization parameter which trades-off a spectral (dis)agreement term and a spectral clustering objective during the optimization process. There is also a regularization parameter in the multi-view non-negative matrix factorization (Liu et al., 2013) and the robust multi-views spectral clustering (Xia et al., 2014) algorithms.

In this paper, we propose the Multi-View Normalized Cuts (MVNC) approach, a parameter free multi-view spectral clustering algorithm. MVNC is described in Section 2. Section 3 describes the experimental environment, while Section 4 reports and discusses the experimental results. Section 5 concludes the paper and gives an outlook to future work.

2. Multi-View Normalized Cuts (MVNC)

In this section, we present MVNC, a new multi-view clustering algorithm which works in two phases. In the first phase, an initial partitioning is performed using a spectral technique. In the second phase, a local search procedure is used to refine the initial clustering.

2.1. Spectral clustering

Given a set of N data points $X = \{x_1, x_2, \ldots, x_N\}$, the single-view normalized cut algorithm proposed by Ng, Jordan, and Weiss (2001) partitions X into K clusters by solving the following minimization problem:

$$\min_{\mathbf{U}\in\mathbb{R}^{N\times K}} tr(\mathbf{U}^\top \mathbf{L}\mathbf{U}), \quad s.t. \quad \mathbf{U}^\top\mathbf{U} = I \tag{1}$$

where tr is the matrix trace and $\mathbf{L}$ is the normalized Laplacian. Cluster memberships are then obtained by clustering the rows of matrix $\mathbf{U}$ using the K-means algorithm.

When the dataset X is represented using V different sets of features (i.e. views), the co-regularized multi-view spectral clustering algorithm of Kumar et al. (2011) divides X into K clusters by solving the following joint optimization problem:

$$\min_{\mathbf{U}^{(1)},\ldots,\mathbf{U}^{(V)}\in\mathbb{R}^{N\times K}} \sum_{v=1}^{V} tr(\mathbf{U}^{(v)\top}\mathbf{L}^{(v)}\mathbf{U}^{(v)}) + \lambda \sum_{\substack{1\le i,j\le V\\ i\neq j}} D(\mathbf{U}^{(i)}, \mathbf{U}^{(j)}) \quad s.t. \quad \mathbf{U}^{(v)\top}\mathbf{U}^{(v)} = I, \forall\, 1 \le v \le V \tag{2}$$

where $\mathbf{L}^{(v)}$ is the normalized Laplacian constructed from view v, $D(\mathbf{U}^{(i)}, \mathbf{U}^{(j)})$ is a measure of disagreement between the clusterings of views i and j, and λ is a hyperparameter to be set by the user.

If we constrain the clusterings of all views to be identical, i.e. $\mathbf{U}^{(1)} = \mathbf{U}^{(2)} = \ldots = \mathbf{U}^{(V)}$, then Eq. (2) reduces to the following minimization problem:

$$\min_{\mathbf{U}\in\mathbb{R}^{N\times K}} \sum_{v=1}^{V} tr(\mathbf{U}^\top\mathbf{L}^{(v)}\mathbf{U}), \quad s.t. \quad \mathbf{U}^\top\mathbf{U} = I \tag{3}$$

or, equivalently, to

$$\min_{\mathbf{U}\in\mathbb{R}^{N\times K}} tr\left(\mathbf{U}^\top\left(\sum_{v=1}^{V}\mathbf{L}^{(v)}\right)\mathbf{U}\right), \quad s.t. \quad \mathbf{U}^\top\mathbf{U} = I \tag{4}$$

The motivation behind the imposed constraint on $\mathbf{U}^{(v)}$, $1 \le v \le V$ is twofold. First, it allows us to get rid of the co-regularization parameter λ, since the disagreement term in Eq. (2) vanishes. Second, the optimization problem is simplified, as it involves a single matrix, in contrast to the original co-regularization framework which involves V matrices.

Eq. (4) is similar to the single-view spectral clustering problem of Eq. (1), where the Laplacian is formed by the sum of the normalized Laplacians constructed from each view. This suggests that the algorithm of (Ng et al., 2001) can be easily extended to multi-view data. Based on this idea, we propose a new multi-view spectral clustering algorithm. The proposed algorithm, summarized in Algorithm 1, is used in the first phase of MVNC.