Expert Systems With Applications 86 (2017) 307–320

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Multi-view collaborative locally adaptive clustering with Minkowski metric

Guang-Yu Zhang[a], Chang-Dong Wang[a,*], Dong Huang[b], Wei-Shi Zheng[a]

[a] School of Data and Computer Science, Sun Yat-sen University, No. 132, East Waihuan Road, Guangzhou Higher Education Mega Center, Guangzhou, Guangdong, China

[b] College of Mathematics and Informatics, South China Agricultural University, No. 483, Wushan Road, Tianhe District, Guangzhou, China

ARTICLE INFO

Article history:
Received 25 December 2016
Revised 28 April 2017
Accepted 29 May 2017
Available online 2 June 2017

Keywords:
Clustering
Multi-view clustering
Subspace clustering
Locally adaptive clustering
Collaborative strategy
Minkowski metric

ABSTRACT

Recently, many heterogeneous but related views of data have been generated in a number of applications. Different views may represent distinct aspects of the same data, which often have the same or consensus cluster structure. Discovering cluster structure in multi-view data has become a hot research topic and significant progress has been made in multi-view clustering. However, it remains a challenging issue to exploit the diversity within each view and investigate the relationship across multiple views simultaneously. To address the above issues, in this paper, we extend locally adaptive clustering into a multi-view framework with Minkowski metric and propose a novel approach termed multi-view collaborative locally adaptive clustering with Minkowski metric (MV-CoMLAC). Different from the existing multi-view subspace clustering methods, the proposed approach is capable of simultaneously taking into account the subspace diversity within each view as well as the knowledge across different views. A collaborative strategy is designed to exploit the complementary information from different low-dimensional subspaces. Furthermore, Minkowski metric is utilized to take into account the influence of the L-p distance ($p \geq 0$), making our method adaptive to different application tasks. Extensive experiments have been conducted on several multi-view datasets, which demonstrate the superiority of our approach over the existing multi-view clustering methods.

1. Introduction

With the rapid development of information technology, an increasing amount of data can be collected from multiple sources (or views). For instance, data collected from Twitter consist of multiple views, such as users, tweets, users' friend lists and communities; the news can be reported by a combination of texts, images and videos; Wikipedia articles can be written in multiple languages, etc. Conventional clustering algorithms (Domeniconi et al., 2007; Huang, Lai, & Wang, 2016a; 2016b; Wang, Lai, & Huang, 2011; Wang, Lai, Suen, & Zhu, 2013; Weiss, 2004) are generally designed for single-view data and often neglect the fact that data may be collected from different sources which contain complementary information. To address this issue, multi-view clustering technique has become a popular research topic in recent years due to its ability to exploit information from multiple views to improve the clustering results. Many multi-view clustering methods have been developed from different perspectives, such as the *centralized* strategy (Bickel & Scheffer, 2004; Blaschko & Lampert, 2008; Chaudhuri, Kakade, Livescu, & Sridharan, 2009; Zhou & Burges, 2007), the *distributed* approach (Greene & Cunningham, 2009; Long, Philip, & Zhang, 2008) and the *collaborative* methods (Cleuziou, Exbrayat, Martin, & Sublemontier, 2009; Ghassany, Grozavu, & Bennani, 2013; Jiang et al., 2015). Despite the significant success, it remains a challenging problem how to simultaneously exploit the diversity within each view and investigate the relationship across multiple views to achieve better clustering results for multi-view data.

Aiming to tackle the above issue, in this paper, we propose a novel multi-view subspace clustering approach termed multi-view collaborative locally adaptive clustering with Minkowski metric (MV-CoMLAC). Our approach is able to take into account both the subspace diversity within each view as well as the knowledge across different views. The local subspace constructed in each view is robust to the noisy features and missing values in the real-world multi-view data. Additionally, Minkowski metric is extended to the multi-view clustering framework so as to make our method adap-

* Corresponding author.
E-mail addresses: guangyuzhg@foxmail.com (G.-Y. Zhang), changdongwang@hotmail.com (C.-D. Wang), huangdonghere@gmail.com (D. Huang), wszheng@ieee.org (W.-S. Zheng).

http://dx.doi.org/10.1016/j.eswa.2017.05.076
0957-4174/

tive to different application tasks. Compared to the existing multi-view clustering approaches, our approach mainly has the following four advantages:

- We utilize a collaborative strategy to exploit the complementary information from different low-dimensional subspaces to improve the accuracy of the clustering result.
- Our approach conducts subspace clustering in a locally adaptive manner, which can enhance its robustness to outliers (or noise) in datasets.
- To reflect the different importance of multiple views, we assign weights to different views based on the entropy theory to exploit the multi-view information in a more effective way.
- We introduce the new Minkowski parameter p which adapts to any Minkowski-p metric in different measuring space.

The rest of this paper will be organized as follows. In Section 2, we will review the existing multi-view clustering work. Section 3 will briefly make an overview of locally adaptive clustering. The proposed method will be described in detail in Section 4. Experimental results will be reported and analyzed in Section 5. We conclude this paper in Section 6.

2. Related work

Clustering multi-view data has become a popular research topic in the field of machine learning and pattern recognition in recent years (Feng, Cai, Liu, & Liu, 2016; Wang, Lai, & Philip, 2016a; Xu, Wang, & Lai, 2016). A variety of new multi-view clustering approaches have been developed which can be mainly divided into the following three main categories: *centralized, distributed* and *collaborative*. In what follows, we will briefly review the related works about the aforementioned three multi-view clustering categories.

Early work in multi-view clustering usually concentrates on the *centralized* strategy (Bickel & Scheffer, 2004; Blaschko & Lampert, 2008; Chaudhuri, Kakade, Livescu, & Sridharan, 2009; Zhou & Burges, 2007), which extends the single-view clustering approach to the multi-view framework by taking the mutual link information among the various views into consideration. Most of the existing multi-view clustering approaches follow this strategy due to its robust performance. For example, Tzortzis and Likas (2010) extended the convex mixture models (CMMs) to the multi-view framework by assigning each view with a weight so as to reflect their importance. This approach assumes that the data of each view can be represented by the mixture distribution model, which has been proven to be effective on single-view datasets. Furthermore, Tzortzis and Likas (2012) adopted the same weighting scheme that extends the kernel k-means and spectral clustering from single-view to the multi-view framework. The above two *centralized* multi-view clustering approaches consider the mutual links among different views and obtain the robust clustering results on both artificial and real-world datasets. However, the main drawback of this strategy is that they neglect the diversity of different features within each view and the independence of different views.

Another strategy in multi-view clustering is the *distributed* approach (Greene & Cunningham, 2009; Long, Philip, & Zhang, 2008). This strategy first uses an appropriate single-view clustering approach to partition the data in each view and then finds a consensus rule to combine the clusterings from multiple views to generate a better clustering result. Tang, Lu, and Dhillon (2009) proposed a graph-based framework for multi-view clustering, called Linked Matrix Factorization (LMF). This method first uses matrix factorization to approximate each graph (view), and then extracts the common factor from multiple graphs (views) to generate the consensus clustering. Similarly, Wang, Dou, Liu, Lv, and Li (2016b) proposed a two-step framework for multi-view clustering with extreme learning machine (ELM) and implemented three ELM-based multi-view clustering approaches. At the first stage of these approaches, the original data in each view are projected into a higher feature space by using extreme learning machine. After that an appropriate multi-view clustering is performed on the mapping space to obtain the final clustering result. However, the *distributed* multi-view clustering methods generally lack the ability to consider the relation among different views which may be beneficial to the final clustering results.

Collaborative strategy was first investigated for single-view clustering under the framework of fuzzy k-means clustering (FKM) (Mitra, Banka, & Pedrycz, 2006; Pedrycz, 2002). Cleuziou, Exbrayat, Martin, and Sublemontier (2009) extended the conventional (single-view) collaborative clustering into the multi-view clustering framework. Ghassany, Grozavu, and Bennani (2013) exploited the probabilistic theory and proposed a new collaborative multi-view clustering approach. But these collaborative multi-view clustering approaches fail to take into consideration the hidden information in different subspaces.

Very recently, some other approaches have been developed. Kumar, Rai, and Daume (2011) proposed a multi-view spectral clustering approach based on the idea of co-regularization which assumes that different views have the same cluster assignments. This method has demonstrated competitive clustering performance which however ignores the cluster structure differences among the views and the importance differences among the features. Jiang, Liu, Li, and Lu (2012) proposed a multi-view document clustering approach based on the PLSA model that utilizes a regularizer to combine individual PLSA models in different views. However, the limitation of this approach is that it is only applicable to two-view documents. Cai, Nie, and Huang (2013) proposed a multi-view k-means clustering algorithm called robust multi-view k-means clustering (RMKMC), which uses the cluster indicator to reach the consensus result across multiple views. However, it lacks the ability to exploit the diversity within each view. Chen, Xu, Huang, and Ye (2013) introduced a multi-view clustering algorithm that automatically computes the weights for views and features, called TW-k-means, which, however, is incapable of utilizing the knowledge across different views so that in some cases TW-k-means may degenerate into conventional k-means. Zhao, Evans, and Dugelay (2014) proposed a co-training based multi-view subspace clustering algorithm that projects the features into a subspace using LDA and PCA iteratively in each view and then performs k-means to obtain the final clustering result. However it is only able to deal with the multi-view datasets consisting of no more than two views.

3. Overview of locally adaptive clustering

Our approach extends the locally adaptive clustering (LAC) algorithm into a multi-view framework with Minkowski metric. In this section, we briefly introduce the LAC algorithm (Domeniconi et al., 2007; Weiss, 2004).

Suppose we are given a dataset $\mathcal{X} = \{x_1, \ldots, x_N\}$ in a q-dimensional space and the goal is to cluster the dataset into K clusters $\mathcal{S} = \{S_1, \ldots, S_K\}$ where S_k denotes the set of the data points belonging to the kth cluster. The minimization objective function of LAC is as follows (Domeniconi et al., 2007):

$$J_{LAC} = \sum_{k=1}^{K}\sum_{j=1}^{q}\frac{w_{kj}}{|S_k|}\sum_{x_i\in S_k}(x_{ij}-\theta_{kj})^2 + \alpha\sum_{k=1}^{K}\sum_{j=1}^{q} w_{kj}\log w_{kj} \tag{1}$$

$$\text{s.t.} \quad \sum_{j=1}^{q} w_{kj} = 1, \forall k = 1, \ldots, K,$$

where $\Theta = \{\theta_1, \ldots, \theta_K\}$ denotes the set of cluster centers, $\mathcal{W} = \{w_1, \ldots, w_K\}$ with $w_k \in \mathbb{R}^q$ storing the weights of the kth cluster,