ARTICLE IN PRESS
JID: ESWA [m5G;November 1, 2016;11:42]
Expert Systems With Applications 000 (2016) 1–10
Contents lists available at ScienceDirect
Expert Systems With Applications
journal homepage: www.elsevier.com/locate/eswa

Multi-view fuzzy clustering with minimax optimization for effective clustering of data from multiple sources

Yangtao Wang[a,*], Lihui Chen[b]

[a] Connexis(South Tower), Data Analytics Department, Institute for Infocomm Research, A*STAR, 1 Fusionopolis Way, Singapore 138632, Republic of Singapore
[b] INFINITUS Centre for Infocomm Technology, School of Electrical and Electronic Engineering, Nanyang Technological University, 50 Nanyang Ave, Singapore 639798, Republic of Singapore

ARTICLE INFO

Article history:
Received 11 July 2016
Revised 14 September 2016
Accepted 4 October 2016
Available online xxx

Keywords:
Multi-view clustering
Fuzzy clustering
Soft clustering
Minimax optimization

ABSTRACT

Multi-view data clustering refers to categorizing a data set by making good use of related information from multiple representations of the data. It becomes important nowadays because more and more data can be collected in a variety of ways, in different settings and from different sources, so each data set can be represented by different sets of features to form different views of it. Many approaches have been proposed to improve clustering performance by exploring and integrating heterogeneous information underlying different views. In this paper, we propose a new multi-view fuzzy clustering approach called MinimaxFCM by using minimax optimization based on well-known Fuzzy c means. In MinimaxFCM the consensus clustering results are generated based on minimax optimization in which the maximum disagreements of different weighted views are minimized. Moreover, the weight of each view can be learned automatically in the clustering process. In addition, there is only one parameter to be set besides the fuzzifier. The detailed problem formulation, updating rules derivation, and the in-depth analysis of the proposed MinimaxFCM are provided here. Experimental studies on nine multi-view data sets including real world image and document data sets have been conducted. We observed that MinimaxFCM outperforms related multi-view clustering approaches in terms of clustering accuracy, demonstrating the great potential of MinimaxFCM for multi-view data analysis.

1. Introduction

Multi-view data becomes common nowadays because data can be collected from different sources or represented by different features. For example, the same news can be reported in different articles from different news sources, one document can be translated into different kinds of languages and one image can be represented with different kinds of features. Learning and analysing multi-view data has become a hot research topic in recent years and attracted many researchers in, to name a few, the areas of data mining, machine learning, information retrieval and cybersecurity. Many multi-view learning approaches based on different strategies including co-training (Blum & Mitchell, 1998), multiple kernel learning (Lanckriet, Cristianini, Bartlett, Ghaoui, & Jordan, 2004) and subspace learning (Jia, Salzmann, & Darrell, 2010) have been proposed in the literature (Xu, Tao, & Xu, 2013). In Xu, Tao, and Xu (2015), a multi-view learning approach based on subspace learning was proposed to discover a latent intact representation of the data. In Wang, Arora, Livescu, and Bilmes (2015), deep neural networks were used to learn representations (features) for multi-view data. Multi-view learning approaches can be divided into supervised learning and unsupervised learning approaches. In this paper, we focus on one of the unsupervised learning techniques which is clustering for multi-view data analysis. As a promising data analysis tool, clustering is able to find the pattern structure and information underlying the unlabelled data. Clustering algorithms based on different theories have been proposed in various applications in the literature (Filippone, Camastra, Masulli, & Rovetta, 2008; Jain, 2010; Xu & Wunsch, 2005). Multi-view clustering approaches are able to mine valuable information underlying different views of data and integrate them to improve clustering performance which have wide applications. For example, in news articles categorization, each article may be written in different languages or collected from different news sources. In an e-learning education system, students' behaviour and performance in study may be analysed based on some features collected from various sources. Students may be clustered into different groups based on several sets of features, for example, how they approach the exercises, and how they

* Corresponding author.
E-mail addresses: wangyt@i2r.a-star.edu.sg (Y. Wang), ELHCHEN@ntu.edu.sg (L. Chen).

http://dx.doi.org/10.1016/j.eswa.2016.10.006
0957-4174/© 2016 Elsevier Ltd. All rights reserved.

interact with the tutorial videos, those form two different sets of features.

Many multi-view clustering approaches have been proposed in the literature. For clustering multi-view data, roughly three strategies are applied among the existing approaches. The first strategy is to integrate multi-view data into a single objective function which is optimized directly during the clustering process. The consensus clustering result is generated directly without one more step to combine the clustering result of each view. For example, in Kumar, Rai, and Daume (2011), two co-regularized multi-view spectral clustering algorithms were proposed. The pairwise disagreement term and centroid based disagreement term for different views are added into the objective function of spectral clustering. The clustering results which are consistent across the views are achieved after the optimization process. In Tzortzis and Likas (2012), a kernel-based weighted multi-view clustering approach was presented. In particular, each view is expressed by a kernel matrix. The weight of each view and consensus clustering result are learned by minimizing the disagreements of different views. In Cai, Nie, and Huang (2013), a multi-view clustering approach based on K-means was proposed. The consensus cluster indication is integrated in the objective function directly. The second strategy includes two steps as follows. First, a unified representation (view) is generated based on multiple views. Then the existing clustering algorithm such as K-means (MacQueen, 1967) or spectral clustering (Ng, Jordan, & Weiss, 2002) is applied to achieve the final clustering result. For example, Huang, Chuang, and Chen (2012) propose an affinity aggregation spectral clustering in which an aggregated affinity matrix is found first by seeking the optimal combination of different affinity matrices. Then spectral clustering is applied on the new affinity matrix to get the final clustering result. In Guo (2013), a common subspace representation of the data shared across multiple views is first learned. Then K-means is applied on the learned subspace representation matrix to generate the clustering result. In the third strategy, each view of the data is processed independently and an additional step is needed to generate the consensus clustering result based on the result of each view. For example, in Bruno and Marchand-Maillet (2009) and Greene and Cunningham (2009), the consensus clustering result was achieved by integrating the previously generated clusters of individual views based on the latent modelling of cluster-cluster relationships and matrix factorization respectively.

The above multi-view clustering approaches are all based on hard clustering in which each object can only belong to one cluster. Since the real world data sets may not be well separated, different approaches have been proposed based on soft or fuzzy clustering algorithms (Anderson, Zare, & Price, 2013; Aparajeeta, Nanda, & Das, 2016; Kannan, Devi, Ramathilagam, Hong, & Ravikumar, 2015) in which each object can belong to all the clusters with various degrees of memberships. The memberships used in soft clustering help to describe the data better and have many potential applications in the real world. For example, soft clustering approaches can better capture the topics of each document which belongs to several topics with different degrees. Liu, Wang, Gao, and Han (2013) propose a joint Nonnegative Matrix Factorization (NMF) (Lee & Seung, 1999) approach for multi-view clustering in which a disagreement term is introduced in the objective function. Besides NMF based multi-view clustering approaches, several multi-view fuzzy clustering algorithms based on the well known Fuzzy c means (FCM) algorithm (Bezdek, 1981) have been developed. For example, in Cleuziou, Exbrayat, Martin, and Sublemontier (2009), CoFKM is proposed to handle multi-view data by minimizing the objective function of FCM of each view and penalizing the disagreement between any pairs of views. In Jiang et al. (2015), a multi-view fuzzy clustering with weighted views called WV-Co-FCM was proposed. In WV-Co-FCM, the clustering process is based on optimizing the objective function which highlights the fuzzy partition and the weight of each view is achieved by introducing the entropy regularization term.

Both hard and soft approaches discussed above all formulate the multi-view clustering to an optimization problem in which the disagreement of the views is minimized. In Wang, Weng, and Yuan (2014), a minimax optimization based multi-view spectral clustering approach was proposed to handle multi-view relational data. However, as pointed out in Cai et al. (2013), the spectral clustering based multi-view clustering approaches have two drawbacks. One is that the clustering performance is sensitive to the choice of the kernel to build the graph. The other is that they are not suitable for large scale data clustering because of the high time computational cost on kernel construction as well as Eigen decomposition. Fuzzy c means (FCM) is widely applied in many applications because of its effectiveness and low time complexity. To combine the advantages of minimax optimization and FCM, in this paper we propose MinimaxFCM for multi-view data clustering. In MinimaxFCM, the goal is to achieve the consensus clustering result of multi-view data by minimizing the maximum disagreement of the weighted views. Except for the fuzzifier which is one parameter in all FCM based approaches, there is only one extra parameter in MinimaxFCM to control the distribution of each view. Moreover, the time complexity of MinimaxFCM is similar to FCM. The experiments with MinimaxFCM on nine real world data sets including image and document data sets show that MinimaxFCM achieves better clustering performance than the related clustering approaches.

The rest of the paper is organized as follows: in the next section, the highlights of the related multi-view clustering approaches reported in the literature are given. In Section 3, the details of the proposed multi-view fuzzy clustering approach MinimaxFCM are described. Experiments on the real world data sets are conducted and the results are analysed in Section 4. Finally, conclusions are drawn in Section 5.

2. Related work

In this section, five related multi-view clustering approaches including two hard clustering approaches and three soft clustering approaches are reviewed. Two hard clustering approaches are a K-means based multi-view clustering and the minimax optimization based multi-view spectral clustering. Three soft approaches include one Nonnegative Matrix Factorization based approach and two fuzzy clustering based approaches are reviewed.

2.1. Notations

Throughout this paper, the following notations are used unless otherwise stated: we denote the data set which has N objects and K classes as $X = \{x_1, \ldots x_N\}$. The data set is represented by P different views such that the ith object in pth view is denoted as x_i^p. We use u_{ci}^p to denote the fuzzy membership which represents the degree of object i belongs to cluster c in pth view and u_{ci}^* to denote the consensus membership of object i to cluster c shared across different views. The centroid of cluster c of the pth view is denoted as $v_c^{(p)}$. $d_{ic}^p = \|x_i^{(p)} - v_c^{(p)}\|$ is used to denote the distance between centroid $v_c^{(p)}$ and object i in pth view and m is used to denote the fuzzifier.

2.2. RMKMC

RMKMC (Cai et al., 2013) is a multi-view clustering approach based on K-means. The first strategy as discussed in Section 1 is used by RMKMC in which a single objective function is formulated and the consensus clustering result is generated directly after the