Information Processing and Management 52 (2016) 387–398

Contents lists available at ScienceDirect

Information Processing and Management

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

Bi-level weighted multi-view clustering via hybrid particle swarm optimization

Bo Jiang[a,*], Feiyue Qiu[a], Liping Wang[b], Zhenjun Zhang[c,*]

[a] College of Education Science and Technology, Zhejiang University of Technology, Hangzhou, 310023, China
[b] College of Business and Administration, Zhejiang University of Technology, Hangzhou, 310023, China
[c] College of Electrical and Information Engineering, Hunan University, Changsha, Hunan, 410082, China

ARTICLE INFO

Article history:
Received 8 July 2014
Revised 22 July 2015
Accepted 24 November 2015
Available online 23 December 2015

Keywords:
Multi-view clustering
Feature weighting
k-means
Particle swarm optimization

ABSTRACT

Many problems in data mining involve datasets with multiple views where the feature space consists of multiple feature groups. Previous studies employed view weighting method to find a shared cluster structure underneath different views. However, most of these studies applied gradient optimization method to optimize the cluster centroids and feature weights iteratively and made the final partition local optimal. In this work, we proposed a novel bi-level weighted multi-view clustering method with emphasizing fuzzy weighting on both view and feature. Furthermore, an efficient global search strategy that combines particle swarm optimization and gradient optimization was proposed to solve the induced non-convex loss function. In the experimental analysis, the performance of the proposed method was compared with five state-of-the-art weighted clustering algorithms on three real-world high-dimensional multi-view datasets.

1. Introduction

In real-world applications, data are often collected from multiple sources or represented by multiple feature spaces. For example, web pages usually consist of both page-text and the hyper-links pointing to them, as well as images are often represented by multiple color and texture descriptors (Blum & Mitchell, 1998; Kriegel et al., 2011). Multiple views often provide compatible and complementary information for pattern discovery, so it is natural to integrate them together to obtain better performance rather than relying on a single view (Liu et al., 2013). On the other hand, multiple views are derived from integration of multiple types of measurements from different perspectives, different views could have very different statistical properties and produce different partitions, so tackling disagreement among views is crucial for multi-view clustering (Christoudias et al., 2008).

Surprisingly, some previous work on multi-view clustering relies on a core assumption that all views are *compatible* to each other, i.e. all views are considered as equally important and different views agree on a consensus partition (Bickel & Scheffer, 2004; Chaudhuri et al., 2009; Kumar & Daumé III, 2011; Long et al., 2008; Zhao et al., 2014). For example, the co-training based methods are based on the assumption that the true underlying clustering would assign corresponding points in each view to the same cluster (Kumar & Daumé III, 2011; Zhao et al., 2014). Observing this assumption may lead to performance degradation when noisy or irrelevant views exist, many recent work applied a simple yet efficient method *view weighting*, i.e. each view is assigned a positive weight to express its importance (Cai et al., 2013; Chen et al., 2013; 2012; Liu et al., 2013; Tzortzis & Likas, 2012; 2010; Yin et al., 2015).

* Corresponding author. Tel.: +86 15869025003, +86 18684822786.
E-mail addresses: bjiang@zjut.edu.cn (B. Jiang), qfy@zjut.edu.cn (F. Qiu), wlp@zjut.edu.cn (L. Wang), zhenjun@hnu.edu.cn (Z. Zhang).

http://dx.doi.org/10.1016/j.ipm.2015.11.003
0306-4573/

Although various existing methods indeed improve the multi-view clustering performance, they often only capture view-wise relationship and ignore the feature-wise importance. For example, Tzortzis and Likas (2012) applied a kernel matrix to each view and then assigned a weight for each kernel; Liu et al. (2013) assigned a weight for each view in the non-negative matrix factorization(NMF) based multi-view clustering; Cai et al. (2013) improved the performance of k-means on multi-view big data also by view weighting. Fortunately, some most recent work discriminated view and weight simultaneously. For example, each view and feature were endowed a weight to represent their respective importance and a negative entropy constraint was imposed on each of them (Chen et al., 2013). The similar idea was implemented using joint structured sparsity-inducing norms, which is often used in multi-task learning (Wang et al., 2013). However, the negative entropy constraints and regulation terms of sparsity-inducing norms make the objective function non-convex and non-smooth, so it is difficult to solve them in general.

Inspired by recent advance in multi-view clustering and heuristic optimization, this work proposes a novel weighting method with emphasizing fuzzy weighting on both view and feature. In addition, an efficient hybrid optimization strategy based on particle swarm optimization (PSO) and gradient-descent method is introduced to solve this problem. PSO is a classical global search algorithm that emulates social interaction and individual cognition of bird flocks foraging (Kennedy & Eberhart, 1995). Compared with complex genetic algorithm, PSO has less algorithm parameters yet provides comparable optimization performance. In our previous work, a cooperative PSO is designed for high-dimensional data clustering (Jiang & Wang, 2013). In this paper, this method is extended to the scenario of multi-view clustering. The main contributions of this work are summarized as follows.

- In problem modelling, a fuzzy weighting schema is proposed to assign weights for both views and features simultaneously. Then, these weights are coupled within the Euclidean distance function to compute the intra-cluster similarity.
- In problem solving, an iterative hybrid algorithm based cooperative particle swarm optimization and gradient descent optimization is developed to find the best cluster centroids and weight vectors.
- A parameter selection method inspired by Clustering Using REpresentative (CURE) (Guha et al., 2001) is proposed and evaluated in this work.
- In experimental evaluation, we verify the effectiveness of our method with comparison experiments against five state-of-the-art weighting-based clustering algorithm on three real world datasets.

The rest of this paper is organized as follows. The related work is given in Section 2. Then, the proposed clustering objective function and the hybrid optimization algorithm are given in Sections 3 and 4, respectively. The parametric study and performance evaluation are provided in Sections 5 and 6, respectively. Finally, Section 7 concludes this paper.

2. Related work

Existing work in multi-view clustering can be broadly classified into three categories. Algorithms in the first category utilize all views simultaneously to discover hidden patterns (Abhishek Kumar, 2011; Bickel & Scheffer, 2004; Kumar & Daumé III, 2011; Liu et al., 2013; de Sa, 2005). In contrast, approaches from the second category first cluster each view independently and then combine the individual clustering results to produce a final partition (Bruno & Marchand-Maillet, 2009; Greene & Cunningham, 2009; Long et al., 2008). The methods in the last category first find a lower space through space transformation and then apply clustering algorithm to learn a partition (Blaschko & Lampert, 2008; Chaudhuri et al., 2009; Cui et al., 2007; Qi & Davidson, 2009).

Centralized approaches. Bickel and Scheffer (2004) proposed a two-view EM algorithm that based on co-EM algorithm. Their method first computes expectations for hidden variables in one view and then use these value in the M-step for the other view. Unfortunately, this algorithm is not guaranteed to converge. Kumar and Daumé III (2011) proposed a multi-view spectral clustering algorithm that extend idea of co-regulation to multi-view environment to make the clusterings in different views agree with each other. Abhishek Kumar (2011) also adopted a co-training framework such that the similarity matrix in one view is affected by the similarity estimated based on the eigenvectors of Laplacian matrix in the other view. Their method is based on the assumption that the true underlying clustering would assign corresponding points in each view to the same cluster.

Distributed approaches. Long et al. (2008) proposed a distributed multi-view clustering approach that first learns hidden patterns individually from each representation of multi-view data and then finds a optima hidden pattern from those multiple patterns. A main challenge of distributed approach is how to make the patterns in different views comparable. They introduce the concept of mapping function to seek an optimal pattern which is close to all patterns as much as possible under a certain distance measure. Greene and Cunningham (2009) applied NMF to lately integrate the original membership matrix that produced on each view separately. The factorization process is regarded as an optimization problem that minimizes the error between the original membership matrix and the reconstructed one. Bruno and Marchand-Maillet (2009) introduced a probabilistic latent semantic analysis(PLSA) clustering approach that based on the latent modelling of inter-cluster relationships. Both EM and NMF were applied to estimate the PLSA parameters and obtained comparable clustering quality.

Space transformation approaches. Blaschko and Lampert (2008) proposed a correlational spectral clustering algorithm, in which kernel canonical correlation analysis (CCA) is used to find a projection of the multiple views. Chaudhuri et al. (2009) also applied CCA to project multi-view data into a lower-dimensional subspace. Moreover, the empirical results in both of them indicated that CCA-based algorithms consistently provided better performance than PCA-based methods on several quite different multi-view datasets. Qi and Davidson (2009) developed a data transformation approach that converts the original data using a distance metric represented by a transformation matrix, which is obtained by minimizing the Kullback–Leibler divergence between the probability density functions of the original data and the transformed data.