Pattern Recognition 53 (2016) 25–35

Contents lists available at ScienceDirect

Pattern Recognition

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

Weighted Multi-view Clustering with Feature Selection

Yu-Meng Xu [a], Chang-Dong Wang [b,*], Jian-Huang Lai [a]

[a] School of Information Science and Technology, Sun Yat-sen University, Guangzhou, PR China
[b] School of Mobile Information Engineering, Sun Yat-sen University, Zhuhai, PR China

ARTICLE INFO

Article history:
Received 12 November 2014
Received in revised form
24 September 2015
Accepted 16 December 2015
Available online 23 December 2015

Keywords:
Data clustering
Multi-view
Feature selection
Weighting

ABSTRACT

In recent years, combining multiple sources or views of datasets for data clustering has been a popular practice for improving clustering accuracy. As different views are different representations of the same set of instances, we can simultaneously use information from multiple views to improve the clustering results generated by the limited information from a single view. Previous studies mainly focus on the relationships between distinct data views, which would get some improvement over the single-view clustering. However, in the case of high-dimensional data, where each view of data is of high dimensionality, feature selection is also a necessity for further improving the clustering results. To overcome this problem, this paper proposes a novel algorithm termed Weighted Multi-view Clustering with Feature Selection (WMCFS) that can simultaneously perform multi-view data clustering and feature selection. Two weighting schemes are designed that respectively weight the views of data points and feature representations in each view, such that the best view and the most representative feature space in each view can be selected for clustering. Experimental results conducted on real-world datasets have validated the effectiveness of the proposed method.

1. Introduction

Clustering is one of the most important methods to explore the underlying (cluster) structure of data [1]. The basic idea is to partition a set of data objects according to some criterion such that similar objects can be grouped into the same cluster, and dissimilar objects are separated into different clusters. To achieve this goal, we usually conduct clustering by maximizing the intra-cluster similarity and the inter-cluster dissimilarity. After several decades' development, a number of clustering algorithms have been developed [1], such as *k*-means clustering [2], spectral clustering [3], kernel-based clustering [4], graph-based clustering [5] and hierarchical clustering [6].

With the development of hardware technology, a huge amount of multi-view data with various representations have been generated in real-world applications [7–14]. For example, in web clustering, different types of data, such as images, videos, hyperlinks and texts, can be taken into consideration as they are different views of web pages (as shown in Fig. 1). In multi-view data, different views are different representations of the same set of instances. It is a significant research challenge to combine together multiple views or sources of the same set of instances to get a better clustering performance. The existing clustering algorithms designed for single-source data cannot be applied directly to the data consisting of multiple views or in various representations as they often vary greatly from traditional single-source data. Data in different views or sources are always not comparable to each other due to their dimensions and semantic representations are always different.

In addition, some views of data may be of high dimensionality which leads to high computational complexity and possibly low clustering accuracy. For example, when it comes to biomedicine, we can get different types of information for a patient, including magnetic resonance images, cerebrospinal fluid test data, blood test data, protein expression data, and genetic data, each of which is taken as a distinct view of patient data. However, some view of data may be of high dimensionality which would lead to a large amount of calculation. For some specific views, only a portion of features are needed for improving the clustering results. In other words, feature selection is a way which can both simplify the calculation and help to get an accurate data model in data clustering [15,13,16].

In order to solve this problem, we propose a novel algorithm, termed Weighted Multi-view Clustering with Feature Selection (WMCFS), which can simultaneously perform multi-view data clustering and feature selection. A global objective function is proposed, which takes into consideration both of the multi-view learning and feature selection in the process of data clustering. In the global objective function, two weighting schemes are designed

* Corresponding author. Tel.: +86 20 84110175.
E-mail addresses: yumengxu@hotmail.com (Y.-M. Xu), changdongwang@hotmail.com (C.-D. Wang), stsljh@mail.sysu.edu.cn (J.-H. Lai).

http://dx.doi.org/10.1016/j.patcog.2015.12.007
0031-3203/

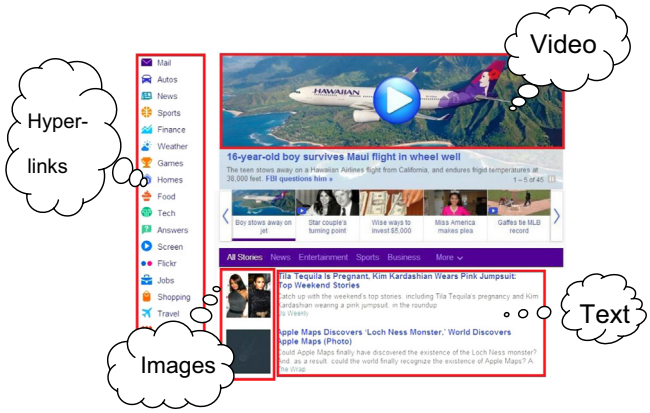

Fig. 1. Multi-view data of web page.

that respectively weight the views of data points and feature representations in each view, such that the best view and the most representative feature space in each view can be selected for clustering. To solve the objective function, we design an EM (Expectation Maximization)-like iteration, which can converge to the acceptable clustering results. Experimental results conducted on real-world datasets have validated the effectiveness of the proposed method.

The rest of the paper is organized as follows. Section 2 briefly overviews the previous work on multi-view data clustering. The proposed WMCFS algorithm and its foundations are described in detail in Section 3. To demonstrate the performance of our algorithms, we have conducted extensive experiments, the experimental results of which are reported in Section 4. The conclusion is drawn in Section 5.

2. Related work

For clustering multi-view or multi-source datasets, some algorithms have been proposed recently which take different factors into consideration, e.g. the differences and relationships between data from various views. Most of the earlier methods extend the traditional single-source clustering algorithms to the multi-view situation by simply minimizing the disagreement between different views, i.e., by minimizing the difference of the clustering results generated from different views. Two early works [17,18] developed two-view algorithms by combining EM, *k*-means and spectral clustering algorithms simultaneously. In [19,20], Kumar et al. used the spectral embedding from one view to conduct clustering of the other views which enforces the clustering results in different views to agree with each other. Wang et al. designed a multi-view spectral clustering, which relies on Pareto optimization to find the best common cuts across all views [21]. However, the above methods only focus on the relationships between various views and ignore the characteristics of distinct views in data. Tzortzis and Likas [22] proposed an multi-view kernel *k*-means (MVKKM) algorithm which assigns a weight for each view according to the view's contribution to the clustering result and then combines the kernels derived from the weighted views together. However, it is based on the inner product kernels for all views, and has no explicit mechanism for feature selection.

To address the above issues, there are some other efforts that investigate feature selection in multi-view data clustering. A framework was proposed in [14], which constructs models respectively for the multi-source learning and feature selection. However, this work is designed for supervised learning and cannot deal with the unsupervised situation. In particular, a model is first trained based on the supervision information, during which the relatively more important features for each cluster can be selected. In this way, feature selection can be accomplished under the criterion to enforce the correct class labels and the important features discovered by this process will be assigned with high weights. However, when it comes to the unsupervised situation, where the labeled samples are not available, this method is no longer applicable, since the importance of features cannot be evaluated due to the lack of the ground-truth labeling. Similarly, Zhao et al. [23] proposed an algorithm combining LDA with co-training, i.e., exploiting labels learned in one view to learn discriminative features in another view. In [24], Wang et al. developed an algorithm to do feature learning for multi-view clustering. However, this method cannot deal with the noisy data in each view. When some of the views are noisy, the result might become unsatisfactory. Chen et al. [25] proposed an automated two-level variable weighting clustering algorithm for multi-view data termed TW-*k*-Means, which can simultaneously compute weights for views and individual variables. However, the same weighting scheme is used for both view weighting and feature selection, which is not able to explore more possibilities. Cai et al. [26] also focused on multi-view clustering based on *k*-means which would be applicable for multi-view data but did not really do feature selection so that their clustering model will degenerate in the case of high dimensionality.

In this paper, inspired by the multi-view kernel *k*-means algorithm proposed by Tzortzis and Likas [22], we design an algorithm termed Weighted Multi-view Clustering with Feature Selection (WMCFS), that can simultaneously perform multi-view data clustering and feature selection. Instead of integrating a feature selection mechanism into multi-view kernel *k*-means, we use a simple yet effective formula based on the original *k*-means algorithm. This is because multi-view kernel *k*-means relies on a kernel mapping in which the kernel selection itself is a challenging issue in the unsupervised learning case.

3. Weighted Multi-view Clustering with Feature Selection

To make this paper clear, Table 1 summarizes the symbols used in this paper.

3.1. Problem formulation

Consider a dataset consisting of N instances represented by V views. Let $\mathcal{X}=\{\mathbf{x}_1^1,\mathbf{x}_2^1,\ldots,\mathbf{x}_N^V\}$ denote the dataset, where $\mathbf{x}_i^v$ is the i-th instance from the v-th view. In this way, the multi-view data

Table 1
Symbols used in this paper.

Symbol	Meaning
$\mathcal{X}$	The whole dataset
X^v	The v-th view dataset
$\mathbf{x}_i^v$	The i-th instance of the v-th view dataset and $\mathbf{x}_i^v \in \mathbb{R}^{d^v}$
$\mathbf{m}_k^v$	The cluster center of the k-th cluster in the v-th view
N	Number of instances in each view
M	Number of clusters
l^v	Number of features in the v-th view
ε_H	Objective function which denotes the sum of intra-class distances
ω_v	Weight for the v-th view
τ_l^v	Weight for the l-th feature of the v-th view
δ_{ik}	Indicator variable showing whether the i-th instance belongs to the k-th cluster
p	Exponential parameter controlling the sparsity of view weight vector
β	Parameter controlling the sparsity of feature weight vectors