Pattern Recognition 52 (2016) 113–134

Contents lists available at ScienceDirect

Pattern Recognition

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

Distance metric learning for soft subspace clustering in composite kernel space

Jun Wang [a,*], Zhaohong Deng [a], Kup-Sze Choi [b], Yizhang Jiang [a], Xiaoqing Luo [d], Fu-Lai Chung [c], Shitong Wang [a]

[a] School of Digital Media, Jiangnan University, China
[b] Centre for Smart Health, School of Nursing, Hong Kong Polytechnic University, Hong Kong, China
[c] Department of Computing, Hong Kong Polytechnic University, Hong Kong, China
[d] School of Internet of Things, Jiangnan University, China

ARTICLE INFO

Article history:
Received 11 November 2014
Received in revised form
30 September 2015
Accepted 26 October 2015
Available online 9 November 2015

Keywords:
Fuzzy clustering
Soft subspace clustering
Composite kernel space
Distance metric learning

ABSTRACT

Soft subspace clustering algorithms have been successfully used for high dimensional data in recent years. However, the existing algorithms often utilize only one distance function to evaluate the distance between data items on each feature, which cannot deal with datasets with complex inner structures. In this paper, a composite kernel space (CKS) is constructed based on a set of basis kernels and a novel framework of soft subspace clustering is proposed by integrating distance metric learning in the CKS. Two soft subspace clustering algorithms, i.e., entropy weighting fuzzy clustering in CKS for kernel space (CKS-EWFC-K) and feature space (CKS-EWFC-F) are thus developed. In both algorithms, the prototype in the feature space is mapped into the CKS by multiple simultaneous mappings, one mapping for each cluster, which is distinct from existing kernel-based clustering algorithms. By evaluating the distance on each feature in the CKS, both CKS-EWFC-K and CKS-EWFC-F learn the distance function adaptively during the clustering process. Experimental results have demonstrated that the proposed algorithms in general outperform classical clustering algorithms and are immune to ineffective kernels and irrelevant features in soft subspace.

1. Introduction

Clustering has a wide range of applications, including statistics, data mining, and database. It has been extensively studied and many algorithms have been developed [1–7]. Among the studies, soft subspace clustering has emerged as a hot research topic in the fields of data mining in recent years [8–17,39]. Under the classical framework of *k*-means or fuzzy c-means clustering algorithms, data objects in the entire data space are grouped but assigned with different weights for different dimensions of the clusters. The assignment is based on the importance of the features in identifying the corresponding clusters. For datasets with different clusters correlating to different subsets of features, soft subspace clustering is a more suitable approach since different vectors of feature weights are assigned to each cluster.

According to the ways of dataset partitioning, soft subspace clustering algorithms [8–20] can be divided into two categories, namely, soft subspace hard clustering and soft subspace fuzzy clustering. For the former, each data object belongs to only one cluster [8,11–13], while for the latter, each data object belongs to every cluster to a certain degree [10,17]. Besides, soft subspace fuzzy clustering can deal with overlapping cluster boundaries. On the other hand, according to the way of soft subspace weighting, soft subspace clustering can also be classified into fuzzy weighting subspace clustering and entropy weighting subspace clustering [10]. Typical fuzzy weighting subspace clustering algorithms include attributes-weighting algorithm (AWA) [8], fuzzy weighting *k*-means (FWKM) [12], fuzzy subspace clustering (FSC) [11] and partition-indexed soft subspace clustering (PI-SSC) [17]. The algorithms assign a fuzzy weight w_{jh}^{α} to the *h*th feature of the *j*th cluster and adjust the feature weights for each cluster automatically during the clustering process. Entropy weighting subspace clustering algorithms include entropy weighting *k*-means (EWKM) [13], clustering objects on subsets of attributes (COSA) [20] and enhanced soft subspace clustering (ESSC) [10]. The algorithms utilize entropy to control the feature weights in each cluster.

Although many soft subspace clustering algorithms have been developed for different application areas, there are still rooms to further improve the performance. A major weakness of soft subspace clustering is the lack of algorithms that are universal for

* Corresponding author. Tel.: +86 13771123570.
E-mail address: wangjun_sytu@hotmail.com (J. Wang).

http://dx.doi.org/10.1016/j.patcog.2015.10.018
0031-3203/

various real world applications. In other words, given a particular soft subspace clustering algorithm, the clustering results can be satisfactory for some datasets while inferior for others. This is because existing soft subspace clustering algorithms utilize only one fixed distance function to evaluate the relationships between data items in two patterns during the clustering process. However, data items in two patterns of different datasets could exhibit different and complex relationships which cannot be described simply by a distance function. Moreover, as the clustering process proceeds, the relationships between data items may change from time to time while the existing soft subspace clustering techniques cannot adapt to the change by updating the distance computation, thereby leading to performance degradation.

To improve the performance of soft subspace clustering, it is necessary to evaluate the relationship between data items adaptively and a distance metric learning strategy is thus in demand. Recent studies have shown that learning the distance function from the data can improve the performance effectively. Depending on the availability of the training data, algorithms for distance metric learning can be divided into supervised and unsupervised approaches. In supervised distance metric learning algorithms, labeled data or side information are utilized to learn the distance function such that data points from the same class are put closely together whereas those from different classes to moved far apart. Representative approaches include convex optimization approach [21], information-theoretic approach [22], smooth optimization approach [23] and alternating optimization approach [40]. On the other hand, unsupervised distance metric learning is a more challenging approach due to the lack of any prior knowledge. In the absence of constraint or class label information, most unsupervised distance metric learning algorithms are in general developed to exploit the underlying manifold structure of the data. Typical unsupervised approaches include adaptive metric learning algorithm (AML) [29], nonlinear adaptive distance metric learning algorithm (NAML) [25], adaptive metric learning for self-organizing incremental neural network (SOINN-AML) [27], locally linear metric adaptation (LLMA) [24]. However, all these clustering algorithms are developed based on distance computation in full space, which is different from the situation in soft subspace clustering algorithms where distance computation is performed based on data items along with each feature. Thus, it is necessary to develop distance metric learning approach so that the most suitable relationship between data items along with each feature can be learned in an unsupervised way.

In this paper, a distance metric learning mechanism for soft subspace clustering is investigated. First, a composite kernel space (CKS) is constructed by linear combination of a set of basis kernel mappings. With the mechanism of distance metric learning, the distance between data items on each feature can be learned adaptively in this CKS. Accordingly, a novel framework of soft subspace clustering is proposed by integrating distance metric learning in the CKS. Especially, two novel soft subspace clustering algorithms, i.e., entropy weighting fuzzy clustering in CKS for kernel space (CKS-EWFC-K) and feature space (CKS-EWFC-F) are proposed, with suffixes K and F in the abbreviations standing for the kernel space and feature space respectively. In both algorithms, the prototype in the feature space is mapped into the CKS by a class of mappings simultaneously, one mapping for each cluster. The mechanism is different from existing kernel-based clustering algorithms. Based on fuzzy partition of the datasets, the proposed algorithms simultaneously locate clusters in CKS and identify the optimal kernel weights for a combination of kernel sets. The incorporation of soft subspace and the automatic adjustment of kernel weights in CKS enable adaptive computation of the distance between data items. Hence, the clustering quality of CKS-EWFC-K and CKS-EWFC-F can be improved for various applications. For easy reference and to enhance the readability of the paper, the major notations used in this paper are summarized in Table 1.

The rest of the paper is organized as follows. In Section 2, related work on soft subspace clustering is reviewed. In Section 3, the composite kernel space is presented, followed by the discussion of the CKS-EWFC-K and CKS-EWFC-F algorithms and their properties. The experiment results are reported and analyzed in Section 4. Conclusions are given in Section 5.

2. Related work

Soft subspace clustering has been a hot research topic in recent years [8–20]. Many algorithms have been developed and the ultimate goal, generally speaking, is to find the local minimum of the objective function J below

$$J(\mathbf{U},\mathbf{W},\mathbf{Z}) = \sum_{j=1}^{c}\sum_{i=1}^{n} u_{ji}^{m} \sum_{h=1}^{s} w_{jh}^{\alpha} d^{2}(x_{ih}, z_{jh}) + H(\mathbf{U},\mathbf{W}), \tag{1}$$

under the constraints $\sum_{j=1}^{c} u_{ji} = 1$ and $\sum_{h=1}^{s} w_{jh} = 1$. In the equation, the first term $\sum_{j=1}^{c}\sum_{i=1}^{n} u_{ji}^{m} \sum_{h=1}^{s} w_{jh}^{\alpha} d^{2}(x_{ih}, z_{jh})$ is interpreted as the total weighted distance between each data object $\mathbf{x}_i$, $i=1, 2, \ldots, n$, and the cluster centers $\mathbf{z}_j$, $j=1, 2, \ldots, c$; and the second term $\mathbf{H}(\mathbf{U,W})$ is a penalty term which is often used to optimize the performance of the algorithm. The term $d(x_{ih}, z_{jh})$ in Eq. (1) is a dissimilarity measure between x_{ih} and z_{jh}, which is often taken as the Euclidean distance, i.e. $d(x_{ih},z_{jh})=\|x_{ih}-z_{jh}\|$, in the original feature space. Other distance functions have also been used in some recent studies, e.g. Minkowski distance function [30], alternative distance function [15], ε-insensitive distance [10] and the Euclidean distance function in kernel space [16]. In this paper, we present a new taxonomy for soft subspace clustering based on the distance function adopted.

2.1. Euclidean distance

The attribute weighting algorithm proposed by Chan et al. is one of the earliest soft subspace clustering algorithms. It adopts the Euclidean distance function [8] and the fuzzy weighting strategy is incorporated into the learning criterion. The objective function of AWA J_{AWA} is formulated as follows:

$$J_{AWA}(\mathbf{U},\mathbf{W},\mathbf{Z}) = \sum_{j=1}^{c}\sum_{i=1}^{n} u_{ji} \sum_{h=1}^{s} w_{jh}^{\alpha} \left(x_{ih} - z_{jh}\right)^{2} \tag{2a}$$

Table 1
Notations used in this paper.

Notations	Descriptions
c	Cluster number
m	Fuzziness of membership
n	Size of dataset
s	Number of features
p	Number of mappings or kernels
u_{ik}	Fuzzy memberships
w_{ih}	Feature weight
$\mathbf{Z}$	Cluster center matrix
$\mathbf{W}$	Fuzzy weighting matrix
$\mathbf{U}$	Fuzzy partition matrix in fuzzy clustering algorithms, or hard partition matrix in hard clustering algorithms
$\mathbf{V}$	Kernel weights matrix
α	Fuzziness of $\mathbf{W}$
η, γ, ε, ε_u, ε_w	Coefficients for penalty terms
x_{ih}	The hth feature of data point $\mathbf{x}_i$
z_{jh}	The hth feature of cluster center $\mathbf{v}_j$