Pattern Recognition 60 (2016) 531–542

Contents lists available at ScienceDirect

Pattern Recognition

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

Fuzzy based affinity learning for spectral clustering

Qilin Li [a,*], Yan Ren [b], Ling Li [a], Wanquan Liu [a]

[a] Department of Computing, Curtin University, Perth, Australia
[b] College of Automation, Shenyang Aerospace University, Shenyang, PR China

ARTICLE INFO

Article history:
Received 1 January 2016
Received in revised form
13 June 2016
Accepted 15 June 2016
Available online 18 June 2016

Keywords:
Similarity measure
Affinity learning
Spectral clustering
AFS theory

ABSTRACT

Spectral clustering makes use of spectral-graph structure of an affinity matrix to partition data into disjoint meaningful groups. It requires robust and appropriate affinity graphs as input in order to form clusters with desired structures. Constructing such affinity graphs is a nontrivial task due to the ambiguity and uncertainty inherent in the raw data. Most existing spectral clustering methods typically adopt Gaussian kernel as the similarity measure, and employ all available features to construct affinity matrices with the Euclidean distance, which is often not an accurate representation of the underlying data structures, especially when the number of features is large. In this paper, we propose a novel unsupervised approach, named Axiomatic Fuzzy Set-based Spectral Clustering (AFSSC), to generate more robust affinity graphs via identifying and exploiting discriminative features for improving spectral clustering. Specifically, our model is capable of capturing and combining subtle similarity information distributed over discriminative feature subspaces to more accurately reveal the latent data distribution and thereby lead to improved data clustering. We demonstrate the efficacy of the proposed approach on different kinds of data. The results have shown the superiority of the proposed approach compared to other state-of-the-art methods.

1. Introduction

Unsupervised data analysis using clustering algorithms provides a useful tool to explore data structures. Clustering methods [1,2] have been studied in many contexts and disciplines such as data mining, document retrieval, image segmentation and pattern classification. The aim of clustering is to group pattern on the basis of similarity (or dissimilarity) criteria where groups (or clusters) are sets of similar patterns. Traditional clustering approaches, such as the k-means and Gaussian mixture models, which are based on estimating explicit models of the data, provide high quality results when the data is distributed according to the assumed models. However, when data appears in more complex or unknown manners, these methods tend to fail. An alternative clustering approach that was shown to handle such structured data is spectral clustering. It does not require estimating an explicit model of data distribution, rather a spectral analysis of the pairwise similarities needs to be conducted.

Spectral clustering normally contains two steps: constructing an affinity graph based on appropriate metric and establishing an appropriate way to "cut" the graph. Plenty of approaches exist to address the graph cut problem, such as minimal cut [3], ratio cut [4] and normalized cut [5], etc. For constructing affinity graph, there are mainly three popular approaches: (1) *The ε-neighborhood graph*: This kind of graph is constructed by connecting all points whose pairwise distances are smaller than a pre-set constant ε. (2) *The k-nearest neighbor graph*: Here the goal is to connect vertex v_i and v_j if v_j is among the k-nearest neighbors of v_i. (3) *The fully connected graph*: Here all vertices are connected and the edges are weighted by the positive similarities between each pair of vertices. According to Luxburg in [6], all three types of affinity graphs mentioned above are regularly used in spectral clustering, and there is no theoretical analysis on how the choice of the affinity graph would influence the performance of spectral clustering.

The crucial problem of constructing *the fully connected graph* is to define the pairwise similarity. The notion of data similarity is often intimately tied to a specific metric function, typically the ℓ_2-norm (e.g. the Euclidean metric) measured with respect to the whole feature space. However, defining the pairwise similarity for effective spectral clustering is fundamentally a challenging problem [7] given complex data that are often of high dimension and heterogeneous, when no prior knowledge or supervision is available. Trusting all available features blindly for measuring pairwise similarities and constructing data graphs is susceptible to unreliable or noisy features [8], particularly so for real-world visual data, e.g. images and videos, where signals can be intrinsically inaccurate and unstable owing to uncontrollable sources of variations and changes in illumination, context, occlusion and background clutters etc. [9].

* Corresponding author.
E-mail address: kylinlovesummer@gmail.com (Q. Li).

http://dx.doi.org/10.1016/j.patcog.2016.06.011
0031-3203/

Moreover, confining the notion of similarity to the ℓ_2-norm metric implicitly imposes unrealistic assumption on complex data structures that do not necessarily possess the Euclidean behavior [8].

In this paper, our aim is to deduce robust pairwise similarity so as to construct more meaningful affinity graph, yielding performance improvement of spectral clustering. To achieve this goal, we first formulate a unified and generalized data distance inference framework based on AFS fuzzy theory [10] with two innovations: (1) Instead of using the complete feature space as a whole, the proposed model is designed to avoid indistinctive features using fuzzy membership function, yielding similarity graphs that can better express the underlying semantic structure in data; this will significantly reduce the number of features used in the clustering process. (2) The Euclidean assumption for data similarity inference is relaxed using fuzzy logic operations defined in AFS. The data distance is then put into the Gaussian kernel to enforce locality. It is worth mentioning that the distinctive features used to represent samples may be different from one another, e.g., every sample could have its own feature subspace. Accordingly the distance measured is dependent on the pairwise feature subspace. A similar idea was presented in [11], which states that different similarities can be induced from a given sample pair if distinct propositions are taken or different questions are asked about data commonalities. In our proposed model, the assumption is that there is no optimal feature subspace which works well for all samples. Each sample pair has its own best feature subspace in terms of distance measure. In terms of AFS clustering, we propose a new method to solve the similarity matrix instead of using the Transitive closure, which needs additional evaluation criteria to obtain clustering result. Extensive experiments have demonstrated that the proposed method is superior compared to both the original spectral clustering and the AFS clustering when the number of features is large.

The rest of this paper is organized as follows. Section 2 presented some previous work on spectral clustering. The main ideas of AFS theory are described in Section 3. In Section 4 we propose a novel approach for generating robust affinity graphs. Experimental results on UCI datasets, USPS handwritten digits and face images are presented in Section 5 and we conclude our work in Section 6.

2. Related work

Large amount of work has been conducted on spectral clustering [5,12–16]. Generally, existing approaches for improving spectral clustering performance can be classified into two paradigms: (1) How to improve data grouping while the affinity graph is fixed [5,12,15]. For example, Xiang and Gong [15] proposed to identify informative and relevant eigenvectors of a data affinity matrix. (2) How to construct appropriate affinity graphs so as to improve the clustering results with standard spectral clustering algorithms [13,17–21]. For example, Chang and Yeung [20] proposed to use path-based similarity to construct robust affinity graph. In this paper, we concentrate on the second paradigm.

Many approaches have been proposed for improving the robustness of affinity graphs adapting to the local data structures [5,17,22]. Particular attention has been focused on learning an adaptive scaling factor σ for the Gaussian kernel $\exp\left(-\frac{dist^2(x_i,x_j)}{\sigma^2}\right)$, when computing the similarity between samples x_i and x_j. For example, Zelnik-Manor and Perona [13] proposed a local scale similarity measure by adjusting the scaling factor as follows:

$$A_{ij} = \exp\left(-\frac{dist^2(x_i, x_j)}{\sigma_i\sigma_j}\right) \tag{1}$$

where σ_i is the distance between point x_i and its k-th nearest neighbor. Yang [23] proposed a similar local scaling factor, which is the mean distance of the k nearest neighbors rather than just considering the k-th neighbor. These methods, however, are still susceptible to the presence of noisy and irrelevant features [8].

To deal with this issue, Pavan and Pelillo [18] proposed a graph-theoretic algorithm for forming tight neighborhoods by selecting the maximal cliques (or maximizing average pairwise affinity), with the hope of constructing graphs with fewer false affinity edges between samples. A kNN based graph generation method was proposed in [19] where the consensus information from multiple kNN is used for discarding noisy edges and identifying strong local neighborhoods. More recently, a random forest based approach was proposed in [8]. This method exploits similarity information from tree hierarchy, leading to a non-linear way of affinity construction. Meanwhile with the random forest framework, the model is capable of removing noisy features. The same idea is proposed in different ways in our approach. Instead of blindly trusting all available variables, our proposed graph inference method exploits discriminative features for measuring more appropriate data pairwise similarities. The affinity matrix created is thus more robust against noisy real-world data.

AFS theory based clustering has been attempted in [24–26]. Instead of using the popular Euclidean metric, AFS clustering approaches capture the underlying data structure through fuzzy membership function, and the distances between samples are represented by membership degree. Furthermore, by extracting the description of samples, those methods are able to establish discriminative feature subspaces for distance measure, which provides a way to deal with commonly existed noise in real-world data. However, in the original AFS clustering, the similarity matrix $S = (s_{ij})_{N\times N}$ does not necessarily satisfy the fuzzy transitive condition $s_{ij} \geq \vee_k (s_{ik} \wedge s_{jk})$, where $\vee$ and $\wedge$ stand for *max* and *min*, respectively. Usually an object is considered similar to another if and only if the degree of similarity between them is greater than or equal to a predefined threshold α. Therefore, the transitive condition states that, for any three objects i, j and k, if object i is similar to object k ($s_{ik} \geq \alpha$) and object k is similar to object j ($s_{kj} \geq \alpha$), object i is similar to object j ($s_{ij} \geq \alpha$) as well. Since the transitive condition is indispensable for clustering, the matrix can always be transformed into its Transitive Closure (denoted by $TC(S) = (t_{ij})_{N\times N}$). $TC(S)$ is defined as a minimal symmetric and transitive matrix. Usually, $TC(S)$ is obtained by searching for an integer k such that $(S^k)^2 = S^k$. With a given α, objects can now be partitioned into different clusters. The problem here is, each threshold α leads to a particular clustering result therefore an evaluation criteria is necessary to obtain a crisp result. It is nontrivial to build such criteria especially in fuzzy clustering. Furthermore, the similarity matrix may not be reflexive (e.g. $s_{ii} = 1$ does not always hold), which means some samples cannot be clustered with certain α (when $s_{ii} < \alpha$). Therefore, a re-clustering process is needed for the original AFS clustering [26] (e.g., to pick up samples which are not clustered in the previous clustering process). The above processes are nontrivial and time-consuming.

3. AFS theory

The proposed affinity matrix construction approach is built based on the AFS theory. AFS theory was originally proposed in [10] and then extensively developed in [27,24,28], etc. AFS fuzzy sets determined by membership functions and their logic operations are algorithmically determined according to distributions of the original data and the semantics of the fuzzy sets. The AFS framework enables the membership functions and fuzzy logic operations to be created based on information within a database,