Neurocomputing 97 (2012) 125–130

Contents lists available at SciVerse ScienceDirect

Neurocomputing

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

Constructing affinity matrix in spectral clustering based on neighbor propagation

Xin-Ye Li *, Li-jie Guo

Department of Electronic and Communication Engineering, North China Electric Power University, Baoding, Hebei 071003,China

ARTICLE INFO

Article history:
Received 22 July 2011
Received in revised form
13 March 2012
Accepted 15 June 2012
Communicated by M. Sato-Ilic
Available online 3 July 2012

Keywords:
Pattern recognition
Spectral clustering
Affinity matrix
Neighbor relation propagation

ABSTRACT

Ng–Jordan–Weiss (NJW) spectral clustering method partitions data using the largest K eigenvectors of the normalized affinity matrix derived from a dataset, but when the dataset is of complex structure, the affinity matrix constructed by traditional Gaussian function could not reflect the real similarity among data points, then the decision of clustering number and selection of K largest eigenvectors are not always effective. Constructing a good affinity matrix is very important to spectral clustering. A new affinity matrix generation method is proposed by using neighbor relation propagation principle and a neighbor relation propagation algorithm is also given. The affinity matrix generated can increase the similarity of point pairs that should be in same cluster and can well detect the structure of data. An improved multi-way spectral clustering algorithm is proposed then. We have performed experiments on dataset of complex structure, adopting Tian Xia and his partner's method for a baseline. The experiment result shows that our affinity matrix well reflects the real similarity among data points and selecting the largest K Eigenvectors gives the correct partition. We have also made comparison with NJW method on some common datasets, the results show that our method is more robust.

1. Introduction

In recent years, spectral clustering [1–7] has attracted more and more interest due to their high performance in data clustering and simplicity in implementation. Spectral clustering methods utilize the eigenvectors of the normalized affinity matrix derived from data to perform data partitioning. Compared with traditional clustering method, spectral clustering method does not need to suppose that data distribution is spheral, so it could recognize the nonspheral distributed clusters. NJW method [3] is one of the most widely used spectral clustering algorithms. For a K clustering problem, this method always partitions data using the largest K eigenvectors of the normalized affinity matrix of a dataset. Although the spectral relaxation solution of normalized cut criteria lies in the subspace spanned by these eigenvectors, it is not guaranteed that the largest K eigenvectors can well detect the structure of the data [1]. To improve spectral clustering algorithm, [8] simplifies the selection of parameters and using neighbor adaptive scale, which simplifies the selection of parameters and makes the improved algorithm insensitive to both density and outliers. [9] proposed a relevance learning method which measures the relevance of an eigenvector according to how well it can separate the dataset into different clusters. [10] proposed a novel eigenvector selection method based on entropy ranking for spectral clustering.

According to [3], in the ideal case in which all points in different clusters are infinitely far apart, the value lying in the ith row and the jth column of affinity matrix is greater than zero if the ith point and the jth point are in same cluster, but zero if the ith point and the jth point are in different cluster. The K largest eigenvalues of the laplacian matrix derived from the affinity matrix are all 1, while the $(K+1)$th largest eigenvalue is far away from 1, then using the K largest eigenvectors to construct matrix Y, we will get K mutually orthogonal points on the surface of the unit K-sphere around which Y's rows will cluster, Y is a new space in NJW method; moreover, these clusters correspond exactly to the true clustering of the original data. In the general case, if the perturbation of affinity matrix is small, the difference between the Kth largest eigenvalue and the $(K+1)$th largest eigenvalue will be the largest, then select the largest K eigenvectors to construct matrix Y, the rows of Y still will form tight clusters around K well-separated points on the surface of the K-sphere according to their "true" cluster.

Typically the distribution of a dataset generated by a real-world system is complex and of an unknown shape, the affinity matrix constructed by Gauss function does not consider the distribution structure of the dataset and could not reflect the real similarity among data points, the eigenvalues of the laplacian

* Corresponding author.
E-mail address: lxyyjh@126.com (X.Y. Li).

0925-2312/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2012.06.023

matrix derived from the affinity matrix do not conform to the above rules again, so it is not guaranteed that the largest K eigenvectors can well detect the structure of the data. In this case, the research of selecting parameters or Eigenvectors would not be effective. Constructing an effective affinity matrix representing the dataset's distribution structure is important. [11] proposed a new definition of affinity graph for spectral clustering from the graph partition perspective, and then it defined the affinity graph respecting two consistencies in a regularization framework of ranking on manifolds. Then it got the affinity matrix defined as $A=(I-\alpha S)^{-1}Y$, where $S=D^{-1/2}WD^{-1/2}$, W is the initial affinity matrix computed by Gaussian function, the degree matrix D is a diagonal matrix whose element D_{ii} is the degree of the point x_i, $\alpha\in(0,1)$, I is an identity matrix. The proposed definition of affinity graph was applicable to both unsupervised and semi-supervised spectral clustering. In unsupervised spectral clustering, $Y=I$, that is, Y is also an identity matrix, then the vector space of A is identical with S, so we think that in unsupervised spectral clustering, affinity matrix A has not been improved radically.

Here, we also propose a construction method of affinity matrix; we use neighbor propagation to get the last affinity matrix that could depict the intrinsic structure of the data.

The rest paper is organized as follows. In Section 2, we review NJW method. In Section 3, we first present some definitions, and then give the construction of affinity matrix based on neighbor propagation, a neighbor propagation algorithm and an improved multi-way spectral clustering algorithm. Experimental results on dataset of complex structure and on several common dataset are given in Section 4, comparing our method with method in Ref. [11] and with the NJW method respectively. Finally, some concluding remarks and issues for future work are given in the Section 5.

2. Ng–Jordan–Weiss (NJW) spectral clustering algorithm

Spectral clustering methods are widely used graph-based approaches for data clustering. Given a dataset $X=\{x_1, x_2,\ldots,x_n\}$ in R^d with K clusters, we can define a $n\times n$ affinity matrix A whose element A_{ij} can be viewed as the weight on the edge connecting the ith and jth data points. The element A_{ij} of the affinity matrix is measured by a typical Gaussian function:

$$A_{ij}=\begin{cases}\exp(-d^2(x_i,x_j)/\sigma^2) & i\neq j\\ 0 & i=j\end{cases} \tag{1}$$

Furthermore, the degree matrix D is a diagonal matrix whose element D_{ii} ($D_{ii}=\sum_{j=1}^{n}A_{ij}$) is the degree of the point x_i.

As a spectral approach to graph partitioning problem, NJW method [3] uses the normalized affinity matrix as the Laplacian matrix and solves the optimization of the normalized cut criterion through considering the eigenvectors associated with the largest eigenvalues. The idea of NJW method is to find a new representation of patterns on the first K eigenvectors of the Laplacian matrix. The details of NJW method are given as follows.

(1) Form the affinity matrix $A\in R^{n\times n}$ defined by Eq. (1).
(2) Compute the degree matrix D and the normalized affinity matrix $L=D^{-1/2}AD^{-1/2}$.
(3) Let $1=\lambda_1\geq\lambda_2\geq\ldots\geq\lambda_K$ be the K largest eigenvalues of L and v^1, v^2,…,v^K be the corresponding eigenvectors. Form the matrix $V=[v^1,v^2,\ldots,v^K]\in R^{n\times K}$ and here v^i is the column vector.
(4) Form the matrix Y, from V by renormalizing each of V's rows to have unit length (i.e.$Y_{ij}=V_{ij}/(\sum_j V_{ij}^2)^{1/2}$).
(5) Treat each row of Y as a point in R^K, and cluster them into K clusters via K-Means algorithm to obtain the final clustering of original dataset.

In step (1), Gaussian function is widely used to construct the affinity matrix for spectral clustering. But Gauss function does not consider the distribution structure of the dataset and could not reflect the real similarity among data points, especially when a dataset generated by a real-world system is complex and of an unknown shape. In this case, it is not guaranteed that the largest K eigenvectors of L in step (3) can well detect the structure of the data.

3. Defining affinity matrix for spectral clustering through neighbor propagation

3.1. Some definitions

Here, we give several definitions.

Given a set of points $S=\{s_1,\ldots,s_n\}$ in R^l, then:

Distance matrix B is a $n\times n$ symmetrical matrix, the element in the ith row and the jth column of B represents the euclidean distance between the ith point s_i and the jth point s_j in S, and is denoted as b_{ij}, as the follow definition:

$$b_{ij}=\sqrt{(s_{i1}-s_{j1})^2+(s_{i2}-s_{j2})^2+\ldots+(s_{il}-s_{jl})^2} \tag{2}$$

Distance threshold ε is defined as

$$\varepsilon=\max_{i=1}^{n}(\min_{j=1}^{n}b_{ij}) \tag{3}$$

Similarity matrix $\boldsymbol{W}$ is a $n\times n$ symmetrical matrix, the element in the ith row and the jth column of W represents the similarity between the ith point s_i and the jth point s_j in S, and is denoted as w_{ij}, as the follow definition:

$$w_{ij}=\exp\left(-\frac{b_{ij}^2}{2\sigma^2}\right) \tag{4}$$

Neighbor relation R: if b_{ij} of distance matrix B is less than distance threshold ε, then we call that points s_i and s_j in S are neighbors, denoted as $(s_i,s_j)\in R$.

Neighbor relation matrix $\boldsymbol{N}$ is a $n\times n$ symmetrical matrix, the value of N's element could only be zero or one. If points s_i and s_j in S are neighbors, then the value of element in the ith row and the jth column of N is one, otherwise, it equals to zero.

Neighbor propagation principle: if $(s_i,s_j)\in R$ and $(s_j,s_k)\in R$, then $(s_i,s_k)\in R$.

3.2. Affinity matrix construction

We construct the affinity matrix according to the following steps:

Step (1). Compute the euclidean distance between each pairs of points in S according to formula (2), get the distance matrix B; and then compute the similarity between each pairs of points in S according to formula (4), get the similarity matrix W.

Step (2). Initiate the neighbor relation matrix N according to distance matrix B and distance threshold ε, that is, if b_{ij} of distance matrix B is less than distance threshold ε, then $n_{ij}=1$ and $n_{ji}=1$.

Step (3). Update neighbor relation matrix N and similarity matrix W according to the neighbor propagation principle:

If $n_{ij}=1$, $n_{jk}=1$ and $n_{ik}=0$, then set $n_{ik}=1$ and $n_{ki}=1$, simultaneously, update w_{ik} and w_{ki} as $\min(w_{ij}, w_{jk})$.

We also propose a neighbor propagation algorithm here, that is:

Input: $n\times n$distance matrix B, $n\times n$similarity matrix W, $n\times n$initial neighbor relation matrix N

Output: affinity matrix A