Spectral clustering

Listenzcc

August 8, 2020

Abstract

In a graph, known as $G(V, E)$, an important problem is how to clustering the vertex *automatically*. That means that the clustering method should be *unsupervised*. The spectral clustering method is a powerful solution.

1 Laplacian matrix and graph

The weight matrix is defined as

$$W = [w_{ij}]$$

in which, $w_{ij} = w_{ji}$ is the measurement weight of the edge, refers the distance between V_i and V_j.

The degree matrix is diagonal matrix

$$d_{ii} = \sum_{j=1}^{N} w_{ij}$$

where N is the number of vertex.

The subtraction is *Laplacian* matrix L

$$L = D - W \tag{1}$$

For any N dimensional vector $f \in R^N$

$$f^T L f = \frac{1}{2} \sum_{i,j=1}^{N} w_{ij} (f_i - f_j)^2 \tag{2}$$

2 Ratio Cut

For graph $G(V, E)$, separate vertex into k sets $A_1, A_2, \cdots, A_k$, the union is V. For sets of A and B, the weight between them is

$$W(A, B) = \sum_{i \in A, j \in B} w_{ij}$$

For a cut protocol *cut*, the weight is defined as

$$Cut(A_i, A_j, \cdots, A_k) = \frac{1}{2} \sum_{i=1}^{k} W(A_i, \bar{A}_i) \tag{3}$$

where $\bar{A}_i$ refers the supplementary set.

Minimization of Eq.3 refers the *cut* that use *Smallest* cut. However, to get the *Best* cut, it has to minimize the *RatioCut*

$$Ratio(A_i, A_j, \cdots, A_k) = \frac{1}{2}\sum_{i=1}^{k}\frac{W(A_i, \bar{A}_i)}{|A_i|} \tag{4}$$

where the denominator is the sum of weights in A_i.

2.1 Computation

Direct computing *RatioCut* is NP-hard, use clustering vectors $h_1, h_2, \cdots, h_k$ as alternative method

$$h_{ij} = \begin{cases} 0 & , v_i \notin A_j \\ \frac{1}{\sqrt{|A_j|}} & , v_i \in A_j \end{cases}$$