준정부호 스펙트럼의 군집화

김재환[O] 최승진
포항공과대학교 컴퓨터학과
{dingkim[O], seungjin}@postech.ac.kr

Semidefinite Spectral Clustering

Jaehwan Kim[O] Seungjin Choi
Department of Computer Science, POSTECH, Korea

Abstract

Graph partitioning provides an important tool for data clustering, but is an NP-hard combinatorial optimization problem. *Spectral clustering* where the clustering is performed by the eigen-decomposition of an affinity matrix [1,2]. This is a popular way of solving the graph partitioning problem. On the other hand, semidefinite relaxation, is an alternative way of relaxing combinatorial optimization, leading to a convex optimization[4]. In this paper we present a semidefinite programming (SDP) approach to graph equi-partitioning for clustering and then we use eigen-decomposition to obtain an optimal partition set. Therefore, the method is referred to as *semidefinite spectral clustering* (SSC). Numerical experiments with several artificial and real data sets, demonstrate the useful behavior of our SSC, compared to existing spectral clustering methods.

1. Introduction

Partitioning a graph with its edges assigned by pairwise similarities, serves as an important tool for data clustering which is a common goal of considerable research in machine learning, data mining, and pattern recognition communities. In general, graph partitioning is an NP-hard combinatorial problem. Spectral relaxation is a popular way of solving the graph partitioning problem, which leads to spectral clustering methods. Successful applications of spectral clustering methods are found in image segmentation, perceptual grouping, biological sequence clustering, document clustering and so on [1,2,3]. As above, spectral methods have been used in various graph partitioning ways. Despite its popularity, the method does not directly relate to an optimization criterion, are somewhat relied on heuristics, so that it is difficult to obtain an ideal solution. Hence, we know the facts that, if we formulate a non-convex combinatorial problem as a convex optimization problem, then we can obtain a globally optimal solution, and semidefinite program is one of the convex optimization problems. When the generalized inequalities are defined over the cone of positive semidefinite(psd), the associated cone program is called a SDP [4].

In this paper, we present a semidefinite relaxation for the multi-way graph equi-partitioning problems, i.e., the sizes of the subgraphs are all equal. We use the geometrical structure of the feasible set, in order to hold a sufficient condition for Strong duality. This tightens the resulting bounds very efficiently. We apply our method to the graph equi-partitioning problems and show its usefulness and robustness in the task of clustering. We show that our approach outperforms the spectral relaxation based methods in various clustering problems.

2. Graph Partitioning Problem

Consider a fully connected and undirected graph $G(V,E)$ where V and E denote a set of vertices and a set of edges, respectively, with pairwise similarity values being assigned as edge weights.

2.1. Classical Graph Bipartitioning

Two-way graph partitioning involves taking the set V apart into two coherent groups, S_i and $\overline{S_i}$, satisfying $V = S_i \cup \overline{S_i}$, ($|V| = n$), and $S_i \cap \overline{S_i} = \emptyset$, by simply cutting edges connecting the two parts. The degree of dissimilarity between S_i and $\overline{S_i}$ can be computed total weight of the edges that have been removed. The cut criterion is given by

$$Cut(S_i, \overline{S_i}) = \frac{1}{4}\tau^T L \tau, \qquad (1)$$

where denote $x^{(j)} = (x_1^{(j)},...,x_n^{(j)})^T \in \Re^{n\times 1}$ as indicator vector for the set S_j, j is class label about each subgraph, i.e.,

$$x_i^{(j)} = \begin{cases} +1, \text{if } i \in S_j \\ 0, \ \ \text{if } i \in \bar{S}_j \end{cases}, \text{ and } x^{(i)T} x^{(j)} = 0. \qquad (2)$$

moreover, $\tau = (x^{(i)} - x^{(j)}) \in \{-1,+1\}^n$ is indicator vector. $W \in \Re^{n\times n}$ is called affinity matrix for the graph G and $D \in \Re^{n\times n}$ is a diagonal matrix with its diagonal elements, $D_i = \sum_{j\in V} W_{ij}$ where D_i is the degree of node i. Matrix $L = D - W$ is the graph Laplacian which is a psd matrix. The cut criterion (1) can be defined as *generalized eigenproblem*. That is, a minimized cut criterion value is the smallest eigenvalue zero. But we can notice some relations between the optimum value and graph' s structure. If graph contains only one subgraph, or two disjointed subgraphs are disconnected, then the smallest eigenvalue of L is zero or the second smallest eigenvalue of that is also zero, repectively. However, above mentioned cases violate our assumptions. Therefore, the eigenvalue of interest to us is the second smallest eigenvalue of L. Objective function of (1) can be defined again as the following form

$$\min \tau^T L\tau, \ e_n{}^T \tau = 0, \ \tau \in \{-1,+1\}^n, \qquad (3)$$

where $e_n = (1,...,1)^T \in \Re^n$. This problem is NP-hard and non-convex problem.

2.2. Multi-way Graph Equi-Partitioning

We present multi-way graph equi-partitioning criterion as a extension of two-way cut with a constraint, which the sizes of the subgraphs are all equal. For multi-way graph partitioning, we involve taking the set V apart into k disjoint groups, $S_1, S_2, ..., S_k, (\forall i, |S_i| = m)$, satisfying $|V| = n$, that is $n = m \times k$. So multi-way cut criterion is to minimize the total weight of the edges connecting nodes in distinct groups of the graph. We just call this criterion *MECut* and that is given by

$$MECut(S_1,...,S_k) = \frac{1}{2}\sum_{i=1}^{k} x^{(i)T}(D-W)x^{(i)}. \qquad (4)$$

Finally, the problem (4) results in a summation of the k smallest eigenvalues of $(D-W)$ and their corresponding eigenvectors are $x^{(i)}$ for $i \in \{1,...,k\}$. Let $X = (x^{(1)},...,x^{(k)}) \in \Re^{n\times k}$ be an indicator matrix. We can derive (4) again as the following problem

$$\min tr(X^T LX), \quad X \text{ is a partition matrix.} \qquad (5)$$

This problem is NP-hard as well as non-convex problem such as (3). Spectral relaxation has been widely used to solve this difficult problems. Next section presents a SDP method for multi-way graph equi-partitioning.

3. SDP Approach

We introduce the optimization approach to solve the problem which is shown in the previous section. *Semidefinite program* is one of the convex optimization problems, which has generalized inequalities defined over the cone of psd. Its *standard form semidefinite program* has the following form

$$\begin{aligned} &\min tr(CX) \\ &s.t. \ tr(A_i X) = b_i, i = 1,...,p, \ X \succeq 0, \end{aligned} \qquad (6)$$

where $C, A_i \in S^n$, the set of symmectric matrices. The most feature between semidefinite programs and linear programs may be consider the constraint of linear matrix inequality instead of the scalar inequality.

3.1. SDP For Multi-way Equi-Partitioning

In order to derive a semidefinite relaxation, we reformulate (5) as a quadratic constrained quadratic programming

$$\begin{aligned} &\min tr(X^T LX) \\ &s.t. \ X : (x_i^{(j)})^2 = x_i^{(j)}, \forall i, j \\ &\quad X : \|Xe_k - e_n\|^2 + \|X^T e_n - me_k\|^2 = 0 \end{aligned} \qquad (7)$$

We first obtain Lagrangian from (7) through considering the lagrangian multipliers $\omega \in \Re^{nk}$ and $\upsilon_0 \le 0$ .The Lagrangian

$$\Lambda(X, \hat{\omega}, \upsilon_0) = (1, x^T)\{L_e + Arrow(\hat{\omega}) + \upsilon_0 C\}\begin{pmatrix}1\\x\end{pmatrix} - \omega_0,$$

$$L_e = \begin{pmatrix} 0 \,|\, 0.......0 \\ \hline 0 \,|\, I_k \otimes L \end{pmatrix}, Arrow(\hat{\omega}) = \begin{pmatrix} \omega_0 & | -0.5\omega^T \\ \hline -0.5\omega & | \, Diag(\omega) \end{pmatrix},$$

$$C = \begin{pmatrix} n & | -e_k^T \otimes e_n^T \\ \hline -e_k \otimes e_n & | \, (e_k e_k^T) \otimes I_n \end{pmatrix} + \begin{pmatrix} m^2 e_k^T e_k & | -me_k^T \otimes e_n^T \\ \hline -me_k \otimes e_n & | \, I_k \otimes (e_n e_n^T) \end{pmatrix} \qquad (12)$$

we define $\hat{\omega} = (\omega_0, \omega)^T \in \Re^{nk+1}$, and $Diag(\omega)$ is the diagonal matrix formed from the vector ω. The symbol, $\otimes$, denotes the Kronecker product. Now we obtain dual and primal problem from (12). We attain dual function which yields lower bound on the optimal value of (5)

$$\begin{aligned} &\pi_d \equiv \max_{\omega_0} - \omega_0 \\ &\quad s.t. \ L_e + Arrow(\hat{\omega}) + \upsilon_0 C = Z \in S_+^{nk+1}. \end{aligned} \qquad (13)$$

Above dual problem (13) is convex optimization problem. From *self-dual* and *minimax inequality* properties, we obtain the primal

$$\begin{aligned} &\pi_r \equiv \min_Y tr(L_e Y) \\ &\quad s.t. \ diag(Y) - (0, Y_{0,1:nk})^T = (1,0,...,0)^T \in \Re^{nk+1} \\ &\quad tr(CY) = 0, Y \succeq 0. \end{aligned} \qquad (14)$$

But, note that C is positive definite, therefore, in order to satisfy $tr(CY)=0$, Y must be a singular, which means that the feasible set of the semidefinite relaxed primal problem (14) is not strictly feasible. That is, due to the Y' s singularity, Slater' s constraint qualification for the primal (14) fails. A sufficient condition for Strong duality to hold is the existence of strictly feasible points for both the primal and the dual problem. In order to overcome, we expolit the geometrical structure of the feasible set, F, of the semidefinite relaxation (14). We need only consider the intersection of faces of F which contain all of these extreme points. The following theorem characterizes the minimal face by finding a point, be called *barycenter point*, in its relative interior.

Theorem 1.(Barycenter point)

$$\hat{Y}=\frac{(m!)^k}{n!}\sum_{partitionsX}\begin{pmatrix}1 \mid x^T \\ x \mid xx^T\end{pmatrix}\in\Re^{(nk+1)\times(nk+1)}. \qquad (15)$$

$rank(\hat{Y})=(k-1)(n-1)+1$, its null space $N(\hat{Y})=k+n-1$. Through (15) and property of which the points in minimal faces' s relative interior can be expressed as $\hat{V}S_+^k\hat{V}^T$, if the matrix $\hat{V}$ and its range $R(\hat{V})=R(\hat{Y})$, we obtain the following semidefinite relaxation in the end

$$f\pi_r \equiv \min_R tr(\hat{V}^T L_e \hat{V}R)$$

$$s.t.\ diag(\hat{V}R\hat{V}^T)=(1,(1/k)e_{nk}{}^T)^T \qquad (16)$$

$$R\in S_+^{(n-1)(k-1)+1}$$

Both (16) and (16)' s dual problems hold Slater' s condition, therefore their optimal values are equal, i.e., guarantee the *strong duality*.

3.2. Finding an Optimal Partition Matrix & Clustering

We use interior-point method to obtain an optimal feasible set Y^*. The Y^*' s repeated diagonal blocks are psd. By Mercer' s theorem, we can think of that matrix as completely determining the kernel. Finally, we can attain an optimal partition matrix $X^*\in\Re^{n\times k}$ defined over feature space using eigen-decomposition of kernel matrix.

We treat each row X^* as a vertice defined in the k dimensional space, cluster them into k cluster, i.e., by projecting onto k dimensional space defined by the rows of X^*. Finally, we assign the original vertice v_i to cluster j if and only if row i of X^* was assigned to cluster j.

4. Experimental Results

We carry out an experiment with some artificial data and real clustering problems. The results are compared with the previous spectral methods, especially in a method by proposed Ng. Jordan,and Weiss[2]. We just call it *NJWCut*.

4.1. Artificial and Real Data Clustering Experiments

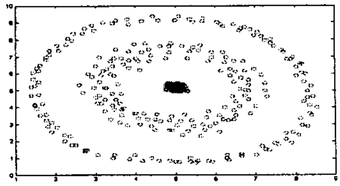

Fig.1. Graph consisting of three subgraphs.

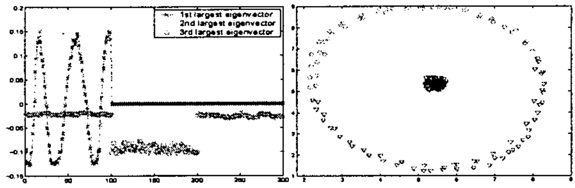

Fig. 2.(L) k Three principle vectors (R)Result via *NJWCut*.

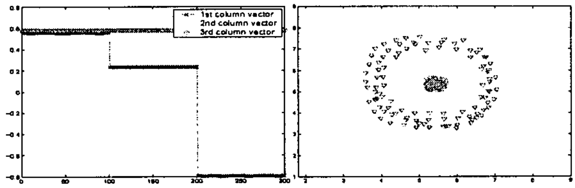

Fig. 3.(L) k Three principle vectors (R)Result via our method

We apply our method to real clustering problem, document clustering problem. Here we do experiment on 20 newsgroup articles(see [3]). Note that some of newsgroups are quitely related, e.g., NG18 and NG19.

Dataset	Balanced case		Unbalanced case	
	Our method	*NJWCut*	Our method	*NJWCut*
NG18	96.040(±0.518)	60.200(±9.539)	96.952(±0.437)	72.000(±4.849)
NG19	95.280(±0.856)	97.067(±0.611)	75.641(±0.235)	95.230(±0.218)
Total	95.660(±0.666)	78.633(±4.652)	86.297(±0.295)	83.615(±2.533)

Table.1. Accuracy(%) of 2groups, std in parenthesis.

5. Conclusion

In this paper, we present a SDP for the graph equi-partitioning problem. We can obtain very tight lower bound where is satisfied with the existence of strictly feasible points for both the primal and dual. The results significantly improve in many problems.

6. Reference

[1]J. Shi, J. Malik, " Normalized cuts and image segmentation," IEEE Trans. PAMI 22(8), 888-905.

[2]A. Ng, M. Jordan, and Y. Weiss " On spectral clustering: Analysis and an algorithm," in Advanced NIPS, vol. 14.

[3] M. Gu, H. Zha, C. Ding, X. He, and H. Simon.," Spectral relaxation models and structure analysis for K-way graph clustering and bi-clustering" , Tech. TR CSE-01-007, Penn State Univ.

[4] Boyd, S,. Vandenberghe, L., " Convex optimization," Cambridge Univ. Press.