Proceedings of the 34th Chinese Control Conference
July 28-30, 2015, Hangzhou, China

Review on spectral methods for clustering

Zhang JingMao[1], Shen YanXia[1]
1. College of Internet of Things engineering, Jiangnan University, Wuxi 214122
E-mail: jmzhang19@sina.cn, shenyx@jiangnan.edu.cn

Abstract: Spectral clustering(SC) is a clustering technology based on graph theory. It becomes one of the most hot topics on clustering because that it can get global optimal solution without any assumptions on data's structure. In this paper, the basic graph theories including some typical graph cut methods for SC are described, then, classic SC algorithms are introduced. Several problems and research topics on SC are also predicted at the end of this paper.

Key Words: Spectral clustering, graph theory, graph cut, similarity matrix, Laplacian matrix

1 Introduction

Clustering is an important unsupervised technology. Clustering algorithm has been applied to many areas, such as data mining, pattern recognition, machine learning, image segmentation and text retrieval. Data is divided into different groups based on some criteria by clustering algorithm, data in the same group is similar to each other and dissimilar otherwise. Clustering methods can be roughly summarized as two categories: hierarchical and partitioning algorithm. Hierarchical algorithms accomplished by creating a tree structure. Partitioning algorithms group data into $k<=n$ parts(n is the number of objects in the dataset). Traditional clustering algorithms, such as k-means, FCM, EM algorithm, perform well when process convex data, but it traps into local optimization when process non-convex data.

Spectral clustering based on graph theory and the SC algorithms treat data as graph's vertex. In clusters of result, there's high relevance within cluster and litter edges between clusters. SC problems actually are graph cut problems, there's many traditional graph cut methods, such as minimum cut, ratio cut, normalized cut, min/max cut etc. In graph cut problems, we can prove that classical information is involved in eigenvectors of similarity matrix of the data[1][2]. SC doesn't make any assumptions on the structure of data. So SC algorithms can find global optimal results when process non-convex data[3].

In the last years, SC algorithms attracted more and more attention, because of its sound theory foundation and good clustering results[4]. In recent years, many SC algorithms have been proposed by researchers. Normalized cut was proposed by Shi and Malik in 2000[5]. This criterion considers both internal and external connections, which result in a more balance clustering result. Ding et al[6] proposed min/max cut. Ng et al[7] proposed the classic NJW algorithm. Now SC algorithms have been applied in many domains, such as computer vision[8][9], integrated circuit design[10], load balancing[11][12], biological information[13][14], text classification[15]. It's very necessary to have a deep research in SC.

This paper describes basic graph theories, typical graph cut methods and introduces classical SC algorithms. In the last, the paper summarizes several problems need to be solved in SC algorithm. The paper is organized as follow: Section 2 introduces the basic graph theory in SC, Section 3 introduces some SC algorithms and Section 4 is the conclusion and summarizes some problems need to be solved.

*This work is supported by the Specialized Research Fund for the Doctoral Program of Higher Education of China (20130093110011), the Integration of Enterprises, Universities and Research of Innovation Foundation of Jiangsu Province (BY2013015-04)

2 Spectral Clustering Theory

Objects of dataset are treated as graph vertexes. Then a weighted matrix is constructed based on the data, $G=(V, E)$. V is a non-empty set involved all vertexes, $V=(v_i , v_j)$. E represents edges between vertexes of V. Let W expresses the weight of edges of E. $W=(w_{ij})_{n*n}$, w_{ij} represents the weight between v_i and v_j. Graph G can be expressed by an adjacent matrix A, $A=[a_{ij}]_{n*n}$ is a binary matrix, $a_{ij}=1$ while there is an edge between v_i and v_j and 0 otherwise. (i, j=1…n).

2.1 Graph Cut Methods

Minimum Cut

Minimum cut was first proposed by Bames[16]. Wu and Leahy[17] applied the criteria in image segmentation. The object function is:

$$cut(A,B) = \sum_{u \in A, v \in B} w(u,v) \qquad (1)$$

This function aims at minimizing the sum of weights between clusters. Graph is divided into subgraph by minimizing this function. This criterion is simple, easy to perform and has a good performence in image segmentation. But it also leads to an uneven clustering result, for example, cluster only has an isolated point. This problem can be solved by ratio cut[18][19][20], average cut[21], normalized cut[5], and min/max cut[6] et al.

Ratio Cut

2-way ratio cut was proposed to solve bipartition problem. It is extended to k-way ratio cut by Chan et al[20]. Subgraph is obtained by minimizing the following object function:

$$Rcut(A,B) = \frac{cut(A,B)}{\min(|A|,|B|)} \qquad (2)$$

$|A|$, $|B|$ represent the number of nodes in subgraph A and B respectively. This criterion can reduce the possibility of over-segmentation. But only the dissimilarity between clusters is considered.

Normalized Cut

Normalized cut was first proposed by Shi and Malik[5], it implemented by minimizing the following equation:

$$Ncut(A,B) = \frac{cut(A,B)}{assoc(A,V)} + \frac{cut(A,B)}{assoc(B,V)} \qquad (3)$$

$$assoc(A,V) = \sum_{u \in A, t \in V} w(u,t) \quad .$$

Minimizing the above equation was known as 2-way normalized cut. Both internal and external connections are considered in this criterion, which result in a more balanced clustering result. But it is hard to deal with big data.

Other Graph Cut Methods

Except for above graph cut methods, there are other criteria to divide graph. Average cut[21] proposed by Sarkar et al can achieve more accurate results, but only focuses on the connections between clusters. Min/max cut[6] can generate more balanced results, but its running time is long. Modularity[22] is a quality validation measure proposed by Newman and Girvan.

2.2 Similarity and Laplacian Matrix

When the structure of data is not a graph, a similarity matrix is needed to be constructed using the raw data. Values in similarity matrix represent the weights between nodes. The matrix can be calculated by the following equation:

$$w_{ij} = e^{\frac{-\|x_i - x_j\|^2}{2\sigma^2}} \qquad (4)$$

While σ is a scale parameter which has a big effect on clustering results. So there are many researches on its optimization now. Degree matrix D is diagonal, the sum of rows of similarity matrix contribute the value of D's diagonal element. Then the Laplacian matrix can be definited by:

$$L=D-W \tag{5}$$

If G is an unweighted graph, W can be replaced by adjacent matrix A.

The eigenvector associated with the second largest eigenvalue named Fiedler eigenvector. Fiedler eigenvector can be used to solve the bipartition problem approximately. The above L is unnormalized Laplacian matrix, the other two normalized Laplacian matrix is:

$$L_N = I - D^{-1/2} W D^{-1/2}$$

$$L_{rw} = I - D^{-1} W$$

3 Spectral Clustering Algorithm[23]

SC algorithm can be grouped as 2-way and k-way algorithm. The following text introduces two classic SC algorithms, corresponding to 2-way and k-way SC respectively.

SM Algorithm

The algorithm proposed by Shi and Malik[5] is a very useful method to image segmentation problem. They proved that when relax the discrete parameter x to continue domain, 2-way *Ncut* problem is relaxed to be a Rayleigh quotient problem. Then,

$$\min Ncut(A,B) = \min \frac{x^T (D-W) X}{x^T Dx}$$

can be translated to

$$\arg \min_{x^T Dl=0} \frac{x^T (D-W) X}{x^T Dx}$$

According to Rayleigh quotient, solving this equation can be seen as solving Fiedler eigenvector problem:

$$Lx = \lambda Dx$$

SM algorithm is composed by these steps:

1) Construct similarity matrix and graph G, let W be its weighted matrix,
2) Compute the Laplacian matrix L,
3) Compute Fiedler eigenvector λ_2 by equation $Lx = \lambda Dx$,
4) Segment G by eigenvector λ_2.

In SM algorithm, both internal and external connections are considered in this criterion, which result in a more balanced clustering result. But it is hard to deal with big data due to its computational complexity.

NJW Algorithm

NJW algorithm is a typical k-way SC algorithm. It is proposed by Ng, Jordan and Weiss[7]. The normalized Laplacian matrix L_N is used by NJW algorithm. NJW algorithm can be described as:

1. Construct similarity matrix and diagonal matrix D,
2. Compute normalized Laplacian matrix L_N,
3. Compute the first k eigenvectors $v_1, v_2, \cdots, v_k$ of L_N,
4. Form matrix V by stacking $v_1, v_2, \cdots, v_k$ in columns,
5. Form matrix U by normalizing the rows of V by $u_{ij} = \dfrac{v_{ij}}{(\sum_k v_{ik}^2)^{1/2}}$,
6. Use other classic clustering algorithm to cluster the points.

NJW algorithm replaces Laplacian with adjacency matrix, which leads to a computational advantage. And as a k-way SC algorithm, it makes full use of eigenvectors.

Other Spectral Clustering Algorithm

Except the above SC algorithms, there are many other SC algorithms. Perona and Freeman[24] propose PF algorithm. Scott et al[25] propose SLH algorithm. Weiss[26] propose a new algorithm by combining SLH algorithm and SM algorithm. Kannan et al[27] propose KVV algorithm. Ding et al[6] propose a new partition criterion called Mcut. Meila et al[28] describe a new SC algorithm in the framework of Markov random walks, which called MS algorithm. Using neighbor relation propagation principle, a new similarity matrix is developed by Li and Guo[29]. Yang et al[30] propose a density sensitive distance based method. A SC algorithm called spectral multi-manifold clustering is proposed by Wang et al[31]. Yang et al[32] propose a SC algorithm named clustering using local discriminant models and global integration. Discriminant cut is presented by Chen and Feng[33]. Linear algebra techniques is used to solve eigenvalue problem by Frederix and Van Barel[34]. An entropy ranking based eigenvector method is proposed by Zhao et al[35].

4 Conclusions and Outlook

Spectral clustering algorithm is an effective clustering algorithm. It never makes any assumption on the structure of data, so the algorithms perform well on non-convex data and get the global optimal result. It can solve clustering problems in polynomial time and has sound theory foundation. Although it has so many advantages, its research is still on the start stage. There are many problems need to be researched[4][36][37].

1) Parameter selection problem. First, when construct similarity matrix, scale parameter σ is selected by manual. Clustering result is effected great by the selection of σ. The selection of parameter σ is an important research direction. Second, how to calculate eigenvector and choose the eigenvector are urgent problems to be solved. Third, the chosen of the number of clusters affect the clustering result directly. Fourth, the selection of Laplacian matrix, three Laplacian matrixes are mentioned above, but the use environment of each matrix is not clearly.
2) Semi-supervise SC algorithm. Limited priori knowledge is easily obtained and it's a kind of supervise information. Now, there are some researches using priori knowledge in SC to improve clustering effect[38][39].
3) Fuzzy SC algorithm. In actual situations, most objects do not have clear attribute. It is necessary to use fuzzy method in SC to solve clustering problems.
4) Applied to big data. SC algorithms also have long run time and large memory space. However, big data has already been a research area in fact. So, SC algorithms applied to big data is a valuable research direction.
5) Kernel spectral clustering algorithm and spectral clustering ensemble. If the data is very complex, SC algorithm does not perform well. This problem can be solved by combining kernel method and SC. Clustering ensemble can deal with scale parameter selection problem and the inherent randomness of SC algorithm.

References

[1] Chung, F. R. (1997). Spectral graph theory (Vol. 92). American Mathematical Soc..

[2] Fiedler, M. (1973). Algebraic connectivity of graphs. Czechoslovak Mathematical Journal, 23(2), 298-305.

[3] Ding SF, Jia HJ, Zhang LW et al (2012) Research of semi-supervised spectral clustering algorithm based on pairwise constraints. Neural Comput Appl. doi:10.1007/s00521-012-1207-8

[4] Nascimento, M. C., & De Carvalho, A. C. (2011). Spectral methods for graph clustering–A survey. European Journal of

Operational Research, 211(2), 221-231.

[5] Shi J, Malik J (2000) Normalized cuts and image segmentation. IEEE Trans Patt Anal Mach Intell 22(8):888–905

[6] Ding, C. H., He, X., Zha, H., Gu, M., & Simon, H. D. (2001). A min-max cut algorithm for graph partitioning and data clustering. In Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on (pp. 107-114). IEEE.

[7] Ng AY, Jordan MI, Weiss Y (2002) On spectral clustering: analysis and an algorithm. Adv Neural Inf Process Syst 14: 849–856

[8] Malik J, Belongie S, Leung T et al (2001) Contour and texture analysis for image segmentation. Int J Comput Vis 43(1):7–27

[9] Zhang XR, Jiao LC, Liu F (2008) Spectral clustering ensemble applied to SAR image segmentation. IEEE Trans Geosci Rem Sens 46(7):2126–2136

[10] Alpert CJ, Kahng AB (1995) Multi-way partitioning via geometric embeddings, orderings and dynamic programming. IEEE Trans Comput-Aaid Des Integr Circuits Syst 14(11):1342–1358

[11] Driessche RV, Roose D (1995) An improved spectral bisection algorithm and its application to dynamic load balancing. Parallel Comput 21(1):29–48

[12] Hendrickson B, Leland R (1995) An improved spectral graph partitioning algorithm for mapping parallel computations. SIAM J Sci Comput 16(2):452–459

[13] Kluger Y, Basri R, Chang JT et al (2003) Spectral biclustering of microarray data: coclustering genes and conditions. Genome Res 13(4):703–716

[14] Paccanaro A, Chennubhotla C, Casbon JA (2006) Spectral clustering of protein sequences. Nucl Acids Res 34(5):1571–1580

[15] Xie YK, Zhou YQ, Huang XJ (2009) A spectral clustering based conference resolution method. J Chin Inf Process 23(3):10–16

[16] Bames ER (1982) An algorithm for partitioning the nodes of a graph. SIAM J Algebraic Discrete Methods 17(5):541–550

[17] Wu Z, Leahy R (1993) An optimal graph theoretic approach to data clustering: theory and its application to image segmentation. IEEE Trans Patt Anal Mach Intell 15(11):1101–1113

[18] Leighton, T., & Rao, S. (1988, October). An approximate max-flow min-cut theorem for uniform multicommodity flow problems with applications to approximation algorithms. In Foundations of Computer Science, 1988., 29th Annual Symposium on (pp. 422-431). IEEE.

[19] Wei, Y. C., & Cheng, C. K. (1989, November). Towards efficient hierarchical designs by ratio cut partitioning. In Computer-Aided Design, 1989. ICCAD-89. Digest of Technical Papers., 1989 IEEE International Conference on (pp. 298-301). IEEE.

[20] Chan, P. K., Schlag, M. D., & Zien, J. Y. (1994). Spectral k-way ratio-cut partitioning and clustering. Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on, 13(9), 1088-1096.

[21] Sarkar, S., & Soundararajan, P. (2000). Supervised learning of large perceptual organization: Graph spectral partitioning and learning automata.Pattern Analysis and Machine Intelligence, IEEE Transactions on, 22(5), 504-525.

[22] Newman, M. E., & Girvan, M. (2004). Finding and evaluating community structure in networks. Physical review E, 69(2), 026113.

[23] Von Luxburg, U. (2007). A tutorial on spectral clustering. Statistics and computing, 17(4), 395-416.

[24] Perona, P., & Freeman, W. (1998). A factorization approach to grouping. InComputer Vision—ECCV'98 (pp. 655-670). Springer Berlin Heidelberg.

[25] Scott G L, Longuet Higgins H C (1990). Feature grouping by relocalization of eigenvectors of the proximity matrix Proc . British Machine Vision Conference. 103-108

[26] Weiss, Y. (1999). Segmentation using eigenvectors: a unifying view. InComputer vision, 1999. The proceedings of the seventh IEEE international conference on (Vol. 2, pp. 975-982). IEEE.

[27] Kannan, R., Vempala, S., & Vetta, A. (2004). On clusterings: Good, bad and spectral. Journal of the ACM (JACM), 51(3), 497-515.

[28] MeilPa, M., & Shi, J. (2001). Learning segmentation by random walks.

[29] Li, X. Y., & Guo, L. J. (2012). Constructing affinity matrix in spectral clustering based on neighbor propagation. Neurocomputing, 97, 125-130.

[30] Yang, P., Zhu, Q., & Huang, B. (2011). Spectral clustering with density sensitive similarity function. Knowledge-Based Systems, 24(5), 621-628.

[31] Wang, Y., Jiang, Y., Wu, Y., & Zhou, Z. H. (2011). Spectral clustering on multiple manifolds. Neural Networks, IEEE Transactions on, 22(7), 1149-1161.

[32] Yang, Y., Xu, D., Nie, F., Yan, S., & Zhuang, Y. (2010). Image clustering using local discriminant models and global integration. Image Processing, IEEE Transactions on, 19(10), 2761-2773.

[33] Chen, W., & Feng, G. (2012). Spectral clustering with discriminant cuts. Knowledge-Based Systems, 28, 27-37.

[34] Frederix, K., & Van Barel, M. (2013). Sparse spectral clustering method based on the incomplete Cholesky decomposition. Journal of Computational and Applied Mathematics, 237(1), 145-161.

[35] Zhao, F., Jiao, L., Liu, H., Gao, X., & Gong, M. (2010). Spectral clustering with eigenvector selection based on entropy ranking. Neurocomputing, 73(10), 1704-1717.

[36] Jia, H., Ding, S., Xu, X., & Nie, R. (2014). The latest research progress on spectral clustering. Neural Computing and Applications, 24(7-8), 1477-1486.

[37] Filippone, M., Camastra, F., Masulli, F., & Rovetta, S. (2008). A survey of kernel and spectral methods for clustering. Pattern recognition, 41(1), 176-190.

[38] Chen WF, Feng GC (2012) Spectral clustering: a semi-supervised approach. Neurocomputing 77(1):229–242.

[39] Jiao LC, Shang FH, Wang F, Liu YY (2012) Fast semi-supervised clustering with enhanced spectral embedding. Patt Recogn 45(12):4358–4369.