Fast Iterative Normalized Cut

Zhicong Xiao, Xiaojun Chen, Zitong Zhang and Feiping Nie

September 23, 2019

Fast Iterative Normalized Cut

Zhicong Xiao[1], Xiaojun Chen[1*], Zitong Zhang[1], Feiping Nie[2]
[1]College of Computer Science and Software, Shenzhen University, Shenzhen 518060, P.R. China
[2]School of Computer Science and OPTIMAL, Northwestern Polytechnical University, Xi′an 710072, P. R. China
xjchen@szu.edu.cn, xiaozhicong2018@email.szu.edu.cn, 407933368@qq.com, feipingnie@gmail.com

Abstract

Normalized cut is a popular spectral clustering method and has been widely used in many applications. In this paper, we propose a novel Fast Iterative Normalized Cut (FINC) algorithm to solve the classic normalized cut problem in a fast way. In the new method, we rewrite the classical normalized cut problem as a new problem and propose an iterative method with proved convergency to effectively solve the new model without eigendecomposition. Theoretical analysis reveals that solving the new method is equivalent to solving the classic normalized cut. Extensive experimental results show the superior performance of the new method.

Introduction

Spectral clustering is a hot topic and many spectral clustering algorithms have been proposed during the past decades. Given a dataset, spectral clustering usually constructs a weighted undirected graph from the pair-wise similarity matrix known as the affinity matrix. The commonly-used spectral clustering is normalized cut, that is formalized as a min-cut problem to partition the vertices in a graph into several disjoint sets such that the total weight of the set of cut edges is minimized (Ng et al. 2002). Since it is difficult to directly solve the discrete cluster indicator matrix, the normalized cut is usually solved in a two-stage process: 1) relax the discrete cluster indicator matrix into continuous one and solve the relaxed problem with eigendecomposition, and 2) obtain the final discrete cluster indicator matrix with k-means or spectral rotation. However, there is no guarantee on the convergence since the two stages aim to solve different objective functions. Recently, Chen et al. proposed an iterative method, named as Direct Normalized Cut, to directly solve the k-way normalized cut model without relaxation (Chen et al. 2018). However, their method is slow since it employs an inner iterative method to solve the cluster indicator matrix object by object, i.e., assign the cluster membership for one object by fixing the cluster memberships of all other objects.

In this paper, we propose a novel Fast Iterative Normalized Cut (FINC) algorithm to solve the classic normalized cut problem. In the new method, we rewrite the classical normalized cut problem as a new problem and propose an iterative method with proved convergency to effectively solve the new model without eigendecomposition. Theoretical analysis reveals that solving the new method is equivalent to solving the classic normalized cut. Moreover, the new method is able to simultaneously obtain the cluster memberships of all objects so we can use parallel technique to accelerate it. Experimental results on 5 real-life datasets show the superior performance of the new method.

Proposed Method

Given the undirected weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ in which the vertices V represent n samples $\mathbf{X} = \{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ and the edges E is associated with the affinity matrix $\mathbf{A}$. Suppose the vertices $\mathcal{V}$ in $\mathcal{G}$ is partitioned into c components and let $\mathbf{Y} \in \Psi^{n\times c}$ be the cluster indicator matrix, in which $y_{il} = 1$ indicates that $\mathbf{x}_i$ is assigned to the l-th cluster. In this paper, we propose to solve a new problem as follow

$$\max_{\mathbf{Y}\in\Psi^{n\times c},\ \mathbf{s}\in\mathbf{R}^{c\times 1}} \sum_{l=1}^{c} 2s_l\sqrt{\mathbf{y}_l^T\mathbf{A}\mathbf{y}_l} - s_l^2\mathbf{y}_l^T\mathbf{D}_A\mathbf{y}_l \tag{1}$$

where $\mathbf{s} \in \mathbf{R}^{c\times 1}$ contains c balance parameters in order to balance the volume of these clusters. Problem (1) can be solved with an alternative optimization approach as follows.

Update $\mathbf{Y}$ with $\mathbf{s}$ Fixed When $\mathbf{s}$ is fixed, it is difficult to directly solve problem (1) so we rewrite it as a new problem

$$\max_{\mathbf{Y}\in\Psi^{n\times c}} \sum_{l=1}^{c} 2\alpha_l s_l\mathbf{y}_l^T\mathbf{A}\mathbf{y}_l - s_l^2\mathbf{y}_l^T\mathbf{D}_A\mathbf{y}_l \tag{2}$$

where

$$\alpha_l = \frac{1}{\sqrt{\mathbf{y}_l^T\mathbf{L}_A\mathbf{y}_l}} \tag{3}$$

and propose an iterative method to solve $\mathbf{Y}$ in problem (2) with with fixed $\mathbf{s}$. In each iteration, α_l is updated according to Eq. (3) after $\mathbf{Y}$ is updated. Suppose the optimal solution of $\mathbf{Y}$ in the r-th iteration is $\mathbf{Y}^r$, $\mathbf{Y}^{r+1}$ is solved from the following problem

$$\max_{\mathbf{Y}\in\Psi^{n\times c}} \sum_{l=1}^{c} \mathbf{y}_l^T(2\alpha_l s_l\mathbf{A} - s_l^2\mathbf{D}_A)\mathbf{y}_l^r - \frac{\eta}{2}\left\|\mathbf{Y}-\mathbf{Y}^r\right\|_F^2 \tag{4}$$

where $\eta > 0$ is a constant used to make $\mathbf{Y}$ different from $\mathbf{Y}^r$ in order to jump out some stationary point and can be

[*]Xiaojun Chen is the corresponding author.

updated according to

$$\eta = \max_{l}(s_l^2 + 2\alpha_l s_l)d_{ii} \tag{5}$$

It can be verified that problem (4) has the following optimal solution

$$y_{ij} =< j = \arg\max_{j' \in [1,c]} (\mathbf{M}^{(j')})^i \mathbf{y}_{j'}^r > \tag{6}$$

where $< . >$ is 1 if the argument is true or 0 otherwise and $\mathbf{M}^{(l)}$ is defined as

$$\mathbf{M}^{(l)} = 2\alpha_l s_l \mathbf{A} - s_l^2 \mathbf{D}_A + \eta \mathbf{I} \tag{7}$$

Note that all rows of $\mathbf{Y}$ can be simultaneously solved, we can use the parallel technique to accelerate this procedure.

Update s with Y Fixed When $\mathbf{Y}$ is fixed, it can be verified that problem (1) is independent between different s_l, so we can solve the following problem individually for each s_l. It can be verified the following optimal solution of s_l

$$s_l = \frac{\sqrt{\mathbf{y}_l^T \mathbf{A} \mathbf{y}_l}}{\mathbf{y}_l^T \mathbf{D}_A \mathbf{y}_l} \tag{8}$$

Optimization Algorithm If we construct a k-nn affinity matrix $\mathbf{A}$, the new algorithm needs $O(r_1(nkc + r_2 nkc))$ time to iteratively solve $\mathbf{s}$ and $\mathbf{Y}$, where r_1 is the number of iterations to update $\mathbf{s}$ and r_2 is the average number of iterations to update $\mathbf{Y}$. Here, the discrete solution $\mathbf{Y}$ converges very fast due to its limited solution space so r_2 is usually very small. Therefore, isc has a time complexity of $O(nkc)$. The convergency of the above algorithm is ensured (see supplemental file for proof).

Connection to the classic normalized cut

Substituting s_l in Eq. (8) into problem (1) gives

$$\max_{\mathbf{Y} \in \Psi^{n \times c}} \sum_{l=1}^{c} \frac{\mathbf{y}_l^T \mathbf{A} \mathbf{y}_l}{\mathbf{y}_l^T \mathbf{D}_A \mathbf{y}_l} \Longleftrightarrow \min_{\mathbf{Y} \in \Psi^{n \times c}} \sum_{l=1}^{c} \frac{\mathbf{y}_l^T \mathbf{L}_A \mathbf{y}_l}{\mathbf{y}_l^T \mathbf{D}_A \mathbf{y}_l} \tag{9}$$

which is exactly the normalized cut problem. Therefore, solving problem (1) is equivalent to solving the classic normalized cut problem.

Experiments on Real-World Datasets

Benchmark datasets

Five real-world benchmark datasets were used in these experiments, i.e., the **Corel**, **ORL**, **segment**, **USPS20** and **uspst** datasets. In this experiment, we compared FINC with 4 optimization methods for solving the normailized cut problem, including Normalized Cut (NCut) (Ng et al. 2002), Multiclass Spectral Clustering (MSC) (Yu and Shi 2003), improved spectral clustering (ISC) (Chen et al. 2017) and DNC (Chen et al. 2018). To perform fair comparisons on all five datasets, we first constructed a sparse 10 nearest neighbors affinity matrix for each of the first 8 datasets to run all comparison methods. We ran each of these methods on each dataset 100 times and selected the best clustering result according to their objective functions. Finally, we used the clustering results

Table 1: Average accuracies by 5 spectral clustering methods on 5 datasets. The best result on each dataset is highlighted in bold.

Method	Metric	corel	ORL	segment	USPSdata	uspst
NCut	ACC	0.177	0.593	0.565	0.728	0.656
	NMI	0.280	0.779	0.544	0.756	0.742
	RI	0.962	**0.975**	0.848	0.938	0.924
MSC	ACC	0.178	0.535	0.509	0.675	0.643
	NMI	0.272	0.728	0.511	0.753	0.734
	RI	0.956	0.967	0.812	0.927	0.924
ISC	ACC	0.113	0.340	0.450	0.466	0.484
	NMI	0.173	0.560	0.423	0.392	0.418
	RI	0.958	0.959	0.773	0.854	0.863
DNC	ACC	0.167	0.553	0.571	0.560	0.630
	NMI	0.262	0.748	**0.569**	0.628	0.621
	RI	0.955	0.970	0.829	0.905	0.907
FINC	ACC	**0.188**	**0.625**	**0.600**	**0.764**	**0.700**
	NMI	**0.289**	**0.784**	0.565	**0.793**	**0.762**
	RI	**0.963**	0.973	**0.868**	**0.945**	**0.933**

in terms of accuracy (**ACC**), normalized mutual information (**NMI**) and rand index (**RI**) to evaluate the clustering results. A careful examination of the results in Table 1 shows that the new method outperforms almost all other methods. Although FINC solves the same normalized cut problem which is also used by NCut, MSC, ISC and DNC, it outperfotmances them on almost all results, indicating the superior performance of the new optimization method.

Acknowledgments

This research was supported by NSFC under Grant no. 61773268, 61502177, U1636202 and Natural Science Foundation of SZU (grant no. 000346) and the Shenzhen Research Foundation for Basic Research, China (Nos. JCYJ20180305124149387).

References

[Chen et al. 2017] Chen, X.; Nie, F.; Huang, J. Z.; and Yang, M. 2017. Scalable normalized cut with improved spectral rotation. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, 1518–1524.

[Chen et al. 2018] Chen, X.; Hong, W.; Nie, F.; He, D.; Yang, M.; and Huang, J. Z. 2018. Spectral Clustering of Large-scale Data by Directly Solving Normalized Cut. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1206–1215.

[Ng et al. 2002] Ng, A. Y.; Jordan, M. I.; Weiss, Y.; et al. 2002. On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems* 2:849–856.

[Yu and Shi 2003] Yu, S. X., and Shi, J. 2003. Multiclass spectral clustering. In *Proceedings of IEEE International Conference on Computer Vision*, 313–319 vol.1.