New ℓ_1-Norm Relaxations and Optimizations for Graph Clustering

Feiping Nie[1], **Hua Wang**[2], **Cheng Deng**[3], **Xinbo Gao**[3], **Xuelong Li**[4], **Heng Huang**[1*]
[1]Department of Computer Science and Engineering, University of Texas at Arlington, USA
[2]Department of Electrical Engineering and Computer Science, Colorado School of Mines, USA
[3]School of Electronic Engineering, Xidian University, Xi'an, China
[4]Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, China
feipingnie@gmail.com, huawang@mines.edu, {chdeng,xbgao}@mail.xidian.edu.cn, xuelong_li@opt.ac.cn, heng@uta.edu

Abstract

In recent data mining research, the graph clustering methods, such as normalized cut and ratio cut, have been well studied and applied to solve many unsupervised learning applications. The original graph clustering methods are NP-hard problems. Traditional approaches used spectral relaxation to solve the graph clustering problems. The main disadvantage of these approaches is that the obtained spectral solutions could severely deviate from the true solution. To solve this problem, in this paper, we propose a new relaxation mechanism for graph clustering methods. Instead of minimizing the squared distances of clustering results, we use the ℓ_1-norm distance. More important, considering the normalized consistency, we also use the ℓ_1-norm for the normalized terms in the new graph clustering relaxations. Due to the sparse result from the ℓ_1-norm minimization, the solutions of our new relaxed graph clustering methods get discrete values with many zeros, which are close to the ideal solutions. Our new objectives are difficult to be optimized, because the minimization problem involves the ratio of nonsmooth terms. The existing sparse learning optimization algorithms cannot be applied to solve this problem. In this paper, we propose a new optimization algorithm to solve this difficult non-smooth ratio minimization problem. The extensive experiments have been performed on three two-way clustering and eight multi-way clustering benchmark data sets. All empirical results show that our new relaxation methods consistently enhance the normalized cut and ratio cut clustering results.

Introduction

Clustering is an important task in computer vision and machine learning research with many applications, such as image segmentation (Shi and Malik 2000), image categorization (Grauman and Darrell 2006), scene analysis (Koppal and Narasimhan 2006), motion modeling (P.Ochs and T.Brox 2012), and medical image analysis (Brun, Park, and Shenton 2004). In the past decades, many clustering algorithms have been proposed. Among these approaches, the use of manifold information in graph clustering has shown the state-of-the-art clustering performance. The graph based clustering methods model the data as a weighted undirected graph based on the pair-wise similarities. Clustering is then accomplished by finding the best cuts of the graph that optimize the predefined cost functions. Two types of graph clustering methods, normalized cut (Shi and Malik 2000) and ratio cut (Cheng and Wei 1991; Hagen and Kahng 1992), are popularly used to solve the clustering problems due to their good clustering performance.

Solving the graph clustering problem is a difficult task (NP-hard problems). The main difficulty of the graph clustering problem comes from the constrains on the solution. It is hard to solve the graph clustering problems exactly. However, the approximation solutions are possible with spectral relaxations. The optimization usually leads to the computation of the top eigenvectors of certain graph affinity matrices, and the clustering result can be derived from the obtained eigen-space. However, the traditional spectral relaxations lead the non-optimal clustering results. The spectral solutions don't directly provide the clustering results and the thresholding post-processing has to be applied, such that the results often severely deviate from the true solution. More recently, tight relaxations of balanced graph clustering methods were proposed (Bühler and Hein 2009; Luo et al. 2010; Hein and Setzer 2011), and gradient based method was used to solve the problem, which is time consuming and slow to converge in practice.

In order to solve the above challenging issues, in this paper, we revisit the normalized cut and ratio cut methods, and propose new relaxations for these methods to achieve the discrete and sparse clustering results which are close to the ideal solutions. Instead of minimizing the projected squared clustering indictors distance, we minimize the ℓ_1 distance. Meanwhile, our new relaxations also use the ℓ_1-norm for the normalization terms. Due to the ℓ_1-norm minimization, most elements of each clustering indictor are enforced to be zero and hence the clustering results are close the ideal solutions.

However, our new relations introduce a difficult optimization problem which optimizes the ratio of two nonsmooth terms. The standard optimization methods for sparse learning, such as Proximal Gradient, Iterative Shrinkage-

*To whom all correspondence should be addressed. This work was partially supported by the following grants: NSF-IIS 1117965, NSF-IIS 1302675, NSF-IIS 1344152, NSF-DBI 1356628, NSF-IIS 1423591, NIH R01 AG049371.

Thresholding, Gradient Projection, Homotopy, and Augmented Lagrange Multiplier methods, cannot be utilized to solve such an ℓ_1-norm ratio minimization problem. We propose a new optimization algorithm to solve this difficult problem with theoretically proved convergence, and our algorithm usually converges within 10 iterations. The extensive clustering experiments are performed on three two-way clustering data sets and eight multi-way clustering data sets to evaluate our new relaxed normalized cut and ratio cut methods. All empirical results demonstrate our new relaxations consistently achieve better clustering results than the traditional relaxations.

Graph Clustering Revisit

Given a graph $G = (V, E)$ and the associated weight matrix W, we partition it into two disjoint sets A and B, $A \cup B = V$, $A \cap B = \emptyset$, Two types of graph clustering methods, normalized cut (Shi and Malik 2000) and ratio cut (Cheng and Wei 1991; Hagen and Kahng 1992), are usually applied to measure the quality of the partition. The main task is to minimize the defined graph cut to obtain a satisfied partition.

Normalized Cut and Relaxation

The normalized cut (Shi and Malik 2000) is defined as

$$Ncut = \frac{cut(A, B)}{assoc(A, V)} + \frac{cut(A, B)}{assoc(B, V)}, \quad (1)$$

where $cut(A, B) = \sum_{i\in A,j\in B} W_{ij}$ and $assoc(A, V) = \sum_{i\in A,j\in V} W_{ij}$

Denote a vector $y \in \Re^{n\times 1}$ as follows

$$y = [\underbrace{1, ..., 1}_{n_1}, \underbrace{r, ...r}_{n_2}]^T. \quad (2)$$

Denote $d_1 = \sum_{i\in A} D_{ii}$, $d_2 = \sum_{i\in B} D_{ii}$, (Shi and Malik 2000) proved that when $r = -\frac{d_1}{d_2}$, the normalized cut defined in Eq. (1) can be written as

$$Ncut = \frac{\frac{1}{2}\sum\limits_{i,j} W_{ij}(y_i - y_j)^2}{\sum\limits_i D_{ii}y_i^2} = \frac{y^T Ly}{y^T Dy}, \quad (3)$$

where $L = D-W$ is the Laplacian matrix, D is the diagonal matrix with the i-th diagonal element as $D_{ii} = \sum_j W_{ij}$. Previous paper (Shi and Malik 2000) provided proof, but here we provide a much more concise proof as follows.

Let $c = \sum_{i\in A,j\in B} W_{ij}$, then we have

$$\frac{\frac{1}{2}\sum\limits_{i,j} W_{ij}(y_i - y_j)^2}{\sum\limits_i D_{ii}y_i^2} = \frac{(1 - r)^2 c}{d_1 + r^2 d_2}. \quad (4)$$

On the other hand, according to Eq. (1), we have

$$Ncut = \frac{c}{d_1} + \frac{c}{d_2}. \quad (5)$$

Combining the above equations, we have:

$$\frac{(1 - r)^2 c}{d_1 + r^2 d_2} = \frac{c}{d_1} + \frac{c}{d_2} \Leftrightarrow \frac{1 - 2r + r^2}{d_1 + r^2 d_2} = \frac{d_1 + d_2}{d_1 d_2}$$
$$\Leftrightarrow (d_1 + rd_2)^2 = 0 \Leftrightarrow r = -\frac{d_1}{d_2},$$

which completes the proof. In order to minimize the normalized cut to obtain a satisfied partition, we need to solve the following problem:

$$\min_{y=[1,...,1,-\frac{d_1}{d_2},...,-\frac{d_1}{d_2}]^T} \frac{\frac{1}{2}\sum\limits_{i,j} W_{ij}(y_i - y_j)^2}{\sum\limits_i D_{ii}y_i^2} \quad (6)$$

Due to the constraint on y, the problem is NP-hard. In order to solve this problem, usually we need to relax the constraint. The constraint in Eq. (6) indicates that $1^T Dy = 0$, thus the problem can be relaxed by using the constraint $1^T Dy = 0$ to replace the constraint in Eq. (6). The relaxed problem is as follows:

$$\min_{1^T Dy=0} \frac{\frac{1}{2}\sum\limits_{i,j} W_{ij}(y_i - y_j)^2}{\sum\limits_i D_{ii}y_i^2} \quad (7)$$

The optimal solution to the relaxed problem is the eigenvector of $D^{-1}L$ corresponding to the second smallest eigenvalue. However, this relaxation makes the solution y deviate from the constraint in Eq. (6) so much. The eigenvector of $D^{-1}L$ usually take on continuous values while the real solution of y should only take on two discrete values. As suggested in (Shi and Malik 2000), One can take 0 or the median value as the splitting point or one can search for the splitting point such that the resulting partition has the best normalized cut value.

Ratio Cut and Relaxation

The ratio cut (Cheng and Wei 1991; Hagen and Kahng 1992) is defined as

$$Rcut = \frac{cut(A, B)}{|A|} + \frac{cut(A, B)}{|B|}, \quad (8)$$

where $|A|$ denotes the number of points in A. Similarly, it can be easily proved that when $r = -\frac{n_1}{n_2}$ in Eq. (2), the ratio cut defined in Eq. (8) can be written as

$$Rcut = \frac{\frac{1}{2}\sum\limits_{i,j} W_{ij}(y_i - y_j)^2}{\sum\limits_i y_i^2} = \frac{y^T Ly}{y^T y}. \quad (9)$$

In order to minimize the normalized cut to obtain a satisfied partition, we solve the following problem

$$\min_{y=[1,...,1,-\frac{n_1}{n_2},...,-\frac{n_1}{n_2}]^T} \frac{\frac{1}{2}\sum\limits_{i,j} W_{ij}(y_i - y_j)^2}{\sum\limits_i y_i^2} \quad (10)$$

Due to the constraint on y, it was also proved that this problem is NP-hard. The constraint in Eq. (10) indicates that $1^T y = 0$, thus the problem can be relaxed by using the constraint $1^T y = 0$ to replace the constraint in Eq. (10). The relaxed problem is as follows:

$$\min_{1^T y=0} \frac{\frac{1}{2}\sum\limits_{i,j} W_{ij}(y_i - y_j)^2}{\sum\limits_i y_i^2} \quad (11)$$

The optimal solution to the relaxed problem is the eigenvector of L corresponding to the second smallest eigenvalue. The relaxation also makes the solution y deviate from the constraint in Eq. (10), and the final partition can be obtained by the same strategies as in the case of normalized cut.

New Graph Clustering Relaxations and Optimization Algorithms

As discussed in the above section, the traditional graph clustering relaxations make the solution y deviate from the ideal solution. In this section, we will propose the new relaxations for normalized cut and ratio cut, to which the solutions are discrete and close to the ideal ones. We will also provide new optimization algorithms to solve the proposed problems.

New Relaxation of Normalized Cut

First, we have the following theorem for normalized cut:

Theorem 1 *Denote* $y = [1, ..., 1, -\frac{d_1}{d_2}, ..., -\frac{d_1}{d_2}]^T$, *then* $\frac{\frac{1}{2}\sum_{i,j} W_{ij} |y_i - y_j|}{\sum_i |D_{ii} y_i|} = \frac{1}{2} Ncut$

Proof: As before, denote $c = \sum_{i \in A, j \in B} W_{ij}$, then we have

$$\frac{\frac{1}{2}\sum\limits_{i,j} W_{ij} |y_i - y_j|}{\sum\limits_i |D_{ii} y_i|} = \frac{(1 + \frac{d_1}{d_2})c}{2d_1} = \frac{(d_1 + d_2)c}{2d_1 d_2}$$
$$= \frac{1}{2}(\frac{c}{d_1} + \frac{c}{d_2}) = \frac{1}{2} Ncut,$$

which completes the proof. □

Based on Theorem 1, the problem (6) is equivalent to the following problem with the same constraint but different objective function:

$$\min_{y=[1,...,1,-\frac{d_1}{d_2},...,-\frac{d_1}{d_2}]^T} \frac{\frac{1}{2}\sum\limits_{i,j} W_{ij} |y_i - y_j|}{\sum\limits_i |D_{ii} y_i|} \tag{12}$$

Accordingly, we can relax the problem as the following one:

$$\min_{1^T D y = 0} \frac{\frac{1}{2}\sum\limits_{i,j} W_{ij} |y_i - y_j|}{\sum\limits_i |D_{ii} y_i|} \tag{13}$$

Note that problem (13) minimizes a ℓ_1-norm, which usually results in sparse solution (Nie et al. 2011b). That is to say, $|y_i - y_j| = 0$ for many (i, j)-pairs, which indicates the solution y will take on discrete values. Therefore, the solution to the relaxed problem (13) is close to the ideal solution.

New Relaxation of Ratio Cut

Similarly, we have the following theorem for ratio cut:

Theorem 2 *Denote* $y = [1, ..., 1, -\frac{n_1}{n_2}, ..., -\frac{n_1}{n_2}]^T$, *then* $\frac{\frac{1}{2}\sum_{i,j} W_{ij} |y_i - y_j|}{\sum_i |y_i|} = \frac{1}{2} Rcut$

Proof: As the above proof, denote $c = \sum_{i \in A, j \in B} W_{ij}$, then we have:

$$\frac{\frac{1}{2}\sum\limits_{i,j} W_{ij} |y_i - y_j|}{\sum\limits_i |y_i|} = \frac{(1 + \frac{n_1}{n_2})c}{n_1 + \frac{n_1}{n_2} n_2} = \frac{(n_1 + n_2)c}{2n_1 n_2}$$
$$= \frac{1}{2}(\frac{c}{n_1} + \frac{c}{n_2}) = \frac{1}{2} Rcut,$$

which completes the proof. □

Based on Theorem 2, the problem (10) is equivalent to the following problem with the same constraint but different objective function:

$$\min_{y=[1,...,1,-\frac{n_1}{n_2},...,-\frac{n_1}{n_2}]^T} \frac{\frac{1}{2}\sum\limits_{i,j} W_{ij} |y_i - y_j|}{\sum\limits_i |y_i|} \tag{14}$$

Accordingly, we can relax the problem as the following one:

$$\min_{1^T y = 0} \frac{\frac{1}{2}\sum\limits_{i,j} W_{ij} |y_i - y_j|}{\sum\limits_i |y_i|} \tag{15}$$

Similarly, the relaxed problem (15) will result in sparse solution, *i.e.*, $|y_i - y_j| = 0$ for many (i, j)-pairs. Therefore, the solution to the relaxed problem (15) is a good approximation to the ideal solution.

Relation to Cheeger cut

In spectral graph theory (Chung 1997), the Cheeger cut is defined as

$$Ccut = \frac{cut(A, B)}{\min\{|A|, |B|\}} \tag{16}$$

As pointed by (Chung 1997; Hein and Buhler 2010), the optimal Cheeger cut is the same as the value obtained by optimal thresholding the optimal solution to the following problem:

$$\min_{y \neq 0, median(y) = 0} \frac{\frac{1}{2}\sum\limits_{i,j} W_{ij} |y_i - y_j|}{\sum\limits_i |y_i|}. \tag{17}$$

Comparing Eq. (17) and Eq. (14), it is interesting to see that the optimal Cheeger cut and the optimal ratio cut can be obtained with the same objective function but under different constraints. Note that the feasible solution y to problem (17) can be continuous values according to the constraint in Eq. (17), thus one can reasonably conjecture that the value obtained by optimal thresholding of the optimal solution to problem (15) is close to the optimal ratio cut in Eq. (14).

Algorithms to Solve New Relaxation Problems

Our new relaxed graph clustering methods introduce a difficult optimization problem, *i.e.* minimize the ratio of non-smooth terms. The standard optimization methods for sparse learning, such as Proximal Gradient, Iterative Shrinkage-Thresholding, Gradient Projection, Homotopy, and Augmented Lagrange Multiplier methods, cannot be utilized to solve such ℓ_1-norm ratio minimization problem. In this section, we will propose a new optimization algorithm to solve this challenging optimization problem. We first introduce the solution to a general problem, and then provide the solutions to problems in Eqs. (13) and (15), respectively.

A General Framework Before solving the new relaxations of graph clustering methods, we solve the following general problem first:

$$\min_{x \in C} \frac{\sum_i |f_i(x)|}{\sum_i |g_i(x)|} . \tag{18}$$

Motivated by (Nie et al. 2009; 2010; 2011a; Nie, Yuan, and Huang 2014), we give an algorithm to solve this problem, which is very easy to implement. The detailed algorithm is described in Algorithm 1. In the following, we will prove that the algorithm will monotonically decrease the objective value of problem (18) until converges.

Algorithm 1 Algorithm to solve the general problem (18).

Initialize $x \in C$
while not converge **do**
1. Calculate the objective value $\lambda = \frac{\sum_i |f_i(x)|}{\sum_i |g_i(x)|}$. For each i, calculate $s_i = \frac{1}{2|f_i(x)|}$ and $b_i = sign(g_i(x))$
2. Update x by $\arg\min_{x \in C} \sum_i s_i f_i^2(x) - \lambda \sum_i b_i g_i(x)$

end while

Theorem 3 *The procedure of Algorithm 1 will monotonically decrease the objective value of problem (18) until converges.*

Proof: Denote the updated x by $\tilde{x}$. According to step 2,

$$\sum_i s_i f_i^2(\tilde{x}) - \lambda \sum_i b_i g_i(\tilde{x}) \le \sum_i s_i f_i^2(x) - \lambda \sum_i b_i g_i(x)$$

Notice the definitions of s_i and b_i in step 1, we have

$$\begin{aligned} &\sum_i \frac{f_i^2(\tilde{x})}{2|f_i(x)|} - \lambda \sum_i sign(g_i(x)) g_i(\tilde{x}) \\ \le &\sum_i \frac{|f_i(x)|}{2} - \lambda \sum_i |g_i(x)| \end{aligned}$$

It can be checked that the following two inequalities hold:

$$\sum_i \left(|f_i(\tilde{x})| - \frac{f_i^2(\tilde{x})}{2|f_i(x)|} \right) \le \sum_i \frac{|f_i(x)|}{2} \tag{19}$$

$$\sum_i \left(sign(g_i(x)) g_i(\tilde{x}) - |g_i(\tilde{x})| \right) \le 0 \tag{20}$$

Adding the above three inequalities in Eqs. (19-20), we have $\sum_i |f_i(\tilde{x})| - \lambda \sum_i |g_i(\tilde{x})| \le 0$, which indicates

$$\frac{\sum_i |f_i(\tilde{x})|}{\sum_i |g_i(\tilde{x})|} \le \lambda = \frac{\sum_i |f_i(x)|}{\sum_i |g_i(x)|} \tag{21}$$

Therefore, the algorithm will monotonically decrease the objective value until converges. □

Algorithm 2 Algorithm to solve the problem (13).

Initialize y such that $1^T D y = 0$
while not converge **do**
1. Calculate $\lambda = \frac{\frac{1}{2}\sum_{i,j} W_{ij}|y_i - y_j|}{\sum_i |D_{ii} y_i|}$; the matrix S, where the (i,j)-th element is $S_{ij} = \frac{1}{2|y_i - y_j|}$; and the vector b, where the i-th element is $b_i = sign(D_{ii} y_i)$
2. Update y by $y = \arg\min_{1^T D y = 0} y^T \hat{L} y - \lambda b^T y$, where $\hat{L} = \hat{D} - \hat{W}$, $\hat{W} = W \circ S$ and $\hat{D}$ is a diagonal matrix with the i-th element as $\hat{D}_{ii} = \sum_j \hat{W}_{ij}$

end while

Solutions to Problem (13) and Problem (15) We can use the algorithm framework in Algorithm 1 to solve the proposed problem (13) and (15). The detailed algorithm to solve the problem (13) is described in Algorithm 2. The algorithm to solve the problem (15) is similar, we omit the detailed algorithm here during to space limitations.

In Step 2 of the Algorithm 2, we need to solve the problem $\min_{1^T D y = 0} y^T \hat{L} y - \lambda b^T y$. Solving this problem seems time consuming because of the constraint in the problem. Fortunately, the problem is equivalent to the following problem $\min_y y^T \hat{L} y - \lambda b^T y + \eta y^T D 1 1^T D y$ with a large enough η. This problem has a closed form solution $y = \lambda(\hat{L} + \eta D 1 1^T D)^{-1} b$ and can be efficiently solved by using Woodbury matrix identity and solving a very sparse system of linear equations.

Extension to Multi-Way Partitioning

The Algorithm 2 partitions the graph into two parts, we can recursively run the algorithms to obtain the desired number of partitions. Specifically, when the graph is divided into k parts, the $k+1$ part can be obtained by running the algorithms on the k parts individually, and select the one that the defined cut is minimal when this part is divided into 2 parts.

Another method to perform the multi-way partitioning is as follows. After we obtain k vectors by the algorithms, the $k+1$ vector y is obtained by running the algorithms with an additional constraint that the vector y is orthogonal to the pervious k vectors. Recursively run the algorithms, we can obtain the desired number of vectors, and then run K-means clustering on the vectors to obtain the final partitioning of the graph as in (Nie et al. 2011b).

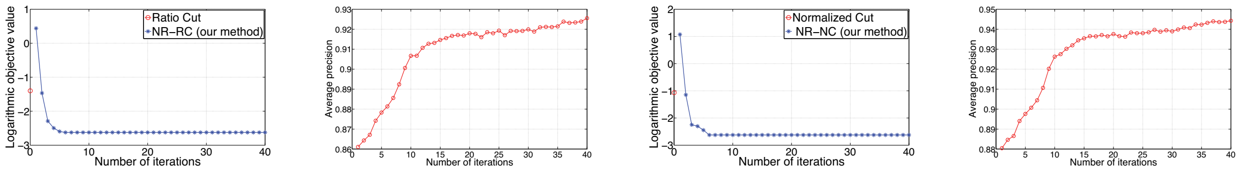
Experimental Results

In this section, we experimentally evaluate the two proposed graph clustering methods in both two-way and multi-way clustering tasks. We abbreviate the proposed new relaxation of the normalized cut as NR-NC, and abbreviate the proposed new relaxation of the ratio cut as NR-RC.

To evaluate the clustering results, we adopt the two widely used standard metrics: clustering accuracy and normalized mutual information (NMI) (Cai et al. 2008).

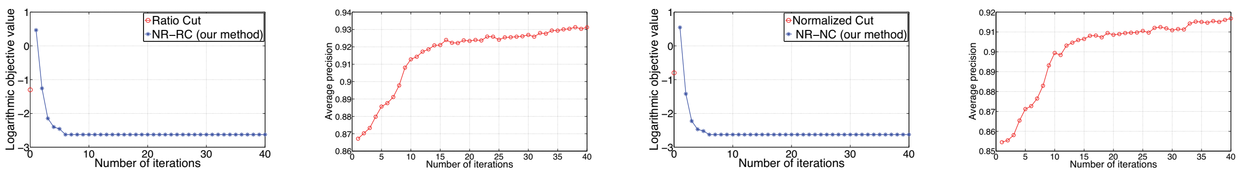
Table 1: Performance and objective value comparison of the proposed methods against their traditional counterparts.

Data	Ratio Cut		NR-RC		Normalized Cut		NR-NC	
	Acc	NMI	Acc	NMI	Acc	NMI	Acc	NMI
Hepatitis	0.894	0.773	0.924	0.797	0.919	0.804	0.944	0.813
ionosphere	0.903	0.812	0.931	0.844	0.897	0.801	0.915	0.824
breast cancer	0.851	0.678	0.913	0.766	0.872	0.703	0.938	0.812



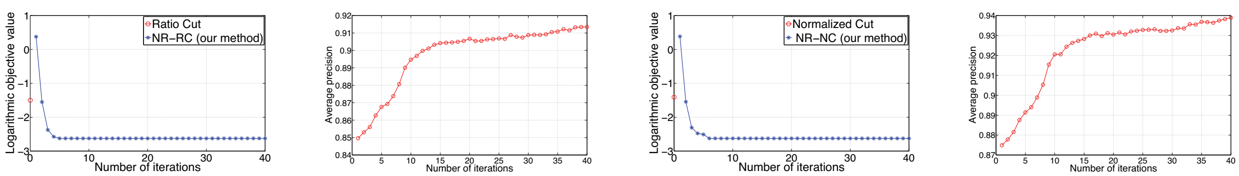
(a) Objective value *vs*. iteration.
(b) Clustering accuracy of NR-RC *vs*. iteration.
(c) Objective value *vs*. iteration.
(d) Clustering accuracy NR-NC *vs*. iteration.

Figure 1: Convergence analysis of 2-way clustering on hepatitis data set.



(a) Objective value *vs*. iteration.
(b) Clustering accuracy of NR-RC *vs*. iteration.
(c) Objective value *vs*. iteration.
(d) Clustering accuracy NR-NC *vs*. iteration.

Figure 2: Convergence analysis of 2-way clustering on ionosphere data set.



(a) Objective value *vs*. iteration.
(b) Clustering accuracy of NR-RC *vs*. iteration.
(c) Objective value *vs*. iteration.
(d) Clustering accuracy NR-NC *vs*. iteration.

Figure 3: Convergence analysis of 2-way clustering on breast cancer data set.

Two-Way Clustering Using NR-RC and NR-NC Methods

We first evaluate the two proposed methods in two-way clustering, and compare them against their respective traditional counterparts. Three benchmark data sets from UCI machine learning repository[1] are used in our experiments, including **hepatitis** database with 155 instances and 20 attributes, **ionosphere** database with 351 instances and 34 attributes, **breast cancer** database with 286 instances and 9 attributes. All these three data sets have only 2 classes, therefore we can perform two-way clustering on them. We construct nearest-neighbor graph for each data set following (Gu and Zhou 2009).

The clustering results by the compared results are shown in Table 1, from which we can see that the proposed new relaxation graph clustering methods consistently outperforms their traditional counterparts, sometimes very significantly. These results clearly demonstrate the advantage of the proposed methods in terms of clustering performance.

Because our methods employ iterative algorithms, we investigate the convergence properties of our algorithms with some details. Given the output vertex ranking from each iteration of the algorithms, we compute the objective value by Eq. (8) for the NR-RC method and by Eq. (1) for the NR-NC method, which are plotted in Figure 1(a) and Figure 1(c) for hepatitis data, Figure 2(a) and Figure 2(c) for ionosphere data, Figure 3(a) and Figure 3(c) for breast cancer data, respectively. The clustering accuracy with respect each iteration of the two proposed methods are also plotted in Figure 1(b) and Figure 1(d) for hepatitis data, Figure 2(b) and Figure 2(d) for ionosphere data, Figure 3(b) and Figure 3(d)

[1]http://archive.ics.uci.edu/ml/

Table 2: Clustering accuracy (%) comparison of multi-way clustering on the eight data sets.

DATA SET	KM	PCA+KM	LDA-KM	RC	NR-RC	NC	NR-NC
DERMATOL	75.96	75.96	71.58	81.52	82.67	81.9	**83.52**
ECOLI	62.91	63.99	62.80	44.74	**64.35**	46.81	63.22
COIL20	64.10	67.92	62.01	78.74	**79.94**	77.51	78.42
BINALPHA	42.95	46.30	46.58	46.41	**48.12**	45.15	47.21
UMIST	45.39	45.39	48.52	60.95	62.31	61.59	**64.20**
AR	27.98	29.17	24.17	37.93	**38.91**	37.01	38.88
YALEB	11.06	12.26	12.43	39.21	45.12	42.22	**46.07**
PIE	18.86	18.56	22.20	42.04	46.74	44.51	**47.18**

Table 3: NMI (%) comparison of multi-way clustering on the eight data sets.

DATA SET	KM	PCA+KM	LDA-KM	RC	NR-RC	NC	NR-NC
DERMATOL	86.18	86.18	85.51	83.21	86.38	84.51	**87.19**
ECOLI	49.27	55.53	52.50	36.01	51.20	40.21	**57.50**
COIL20	77.46	77.14	74.85	87.23	**87.88**	86.14	87.67
BINALPHA	58.52	59.74	59.51	57.21	59.64	58.41	**60.30**
UMIST	66.08	66.08	65.03	76.45	**77.19**	71.52	75.41
AR	61.50	63.18	58.11	71.52	**72.63**	70.63	72.17
YALEB	16.09	17.22	18.74	57.36	**59.15**	53.42	57.93
PIE	38.78	39.02	39.55	56.14	59.21	57.15	**60.32**

for breast cancer data, respectively. From these figures we can see that our algorithms converge very fast with typically no more than 20 iterations, which concretely confirm their computational efficiency.

Moreover, as shown in Figure 3(a) and Figure 3(c), in contrast to the objective values of the traditional graph clustering methods, the objective values at convergence of our new relaxed graph clustering methods are much smaller, which provide another evidence to support the correctness of both our objectives and algorithms.

Multi-Way Clustering Using NR-RC and NR-NC Methods

Now we evaluate the proposed methods in multi-way clustering. In our experiments, we implement our methods using the second strategy introduced in Section . Eight benchmark data sets are used in the experiments, including two UCI data sets, dermatology and ecoli, one object data set, COIL-20 (Nene, Nayar, and Murase 1996), one digit and character data sets, Binalpha, and four face data sets, Umist (Graham and Allinson 1998), AR (Martinez and Benavente 1998), YaleB (Georghiades, Belhumeur, and Kriegman 2001), and PIE (Sim and Baker 2003).

Beside comparing our methods to their traditional counterparts, we also compare to K-means (denoted by Km), PCA+K-means (denoted by PCA+Km), LDA-Km (Ding and Li 2007) methods. Again, we construct nearest-neighbor graph for each data set and set the neighborhood size for graph construction as 10 (Gu and Zhou 2009). The dimension of PCA+K-means is searched from five candidates ranging from 10 to the dimension of data.

The results of all clustering algorithms depend on the initialization. To reduce statistical variety, we independently repeat all clustering algorithms for 50 times with random initializations, and then we report the results corresponding to the best objective values.

The clustering performance measured by clustering accuracy and NMI are reported in Table 2 and Table 3, from which we can see that the proposed methods still perform the best among all compared methods. In addition, our methods are always better their respective traditional counterparts. These advantages validate the effectiveness of the proposed methods and justify our motivations.

Conclusions

In this paper, we proposed new relaxations for normalized cut and ratio cut methods. The ℓ_1-norm distances are utilized in the relaxed graph clustering formulations. Such ℓ_1-norm based relaxations can naturally get the discrete and sparse clustering solutions (with many zeros) which are close to the optimal ones. Moreover, we proposed a new optimization algorithm to address the minimization problem of a ratio of non-smooth terms which cannot be solved by other standard sparse learning optimization algorithms. The validations were performed on both two-way and multi-way clustering problems. On all eleven benchmark data sets, our new relaxed normalized cut and ratio cut methods consistently outperform the traditional ones.

References

Brun, A.; Park, H.-J.; and Shenton, M. E. 2004. Clustering fiber traces using normalized cuts. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)* 368–375.

Bühler, T., and Hein, M. 2009. Spectral clustering based on the graph p-laplacian. In *Proceedings of the 26th Annual International Conference on Machine Learning*, 81–88.

Cai, D.; He, X.; Wu, X.; and Han, J. 2008. Non-negative matrix factorization on manifold. In *ICDM*.

Cheng, C.-K., and Wei, Y.-C. A. 1991. An improved two-way partitioning algorithm with stable performance. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 10(12):1502–1511.

Chung, F. R. K. 1997. *Spectral Graph Theory*. CBMS Regional Conference Series in Mathematics, No. 92, American Mathematical Society.

Ding, C. H. Q., and Li, T. 2007. Adaptive dimension reduction using discriminant analysis and -means clustering. In *ICML*, 521–528.

Georghiades, A.; Belhumeur, P.; and Kriegman, D. 2001. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on PAMI* 23(6):643–660.

Graham, D. B., and Allinson, N. M. 1998. Characterizing virtual eigensignatures for general purpose face recognition. in face recognition: From theory to applications. *NATO ASI Series F, Computer and Systems Sciences* 163:446–456.

Grauman, K., and Darrell, T. 2006. Unsupervised learning of categories from sets of partially matching image features. In *CVPR (1)*, 19–25.

Gu, Q., and Zhou, J. 2009. Co-clustering on manifolds. In *SIGKDD*.

Hagen, L. W., and Kahng, A. B. 1992. New spectral methods for ratio cut partitioning and clustering. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 11(9):1074–1085.

Hein, M., and Buhler, T. 2010. An inverse power method for nonlinear eigenproblems with applications in 1-spectral clustering and sparse PCA. In *NIPS*.

Hein, M., and Setzer, S. 2011. Beyond spectral clustering - tight relaxations of balanced graph cuts. In *NIPS*, 2366–2374.

Koppal, S., and Narasimhan, S. 2006. Clustering Appearance for Scene Analysis. *CVPR*.

Luo, D.; Huang, H.; Ding, C. H. Q.; and Nie, F. 2010. On the eigenvectors of p-laplacian. *Machine Learning* 81(1):37–51.

Martinez, A. M., and Benavente, R. 1998. The ar face database. In *CVC Technical Report*.

Nene, S. A.; Nayar, S. K.; and Murase, H. 1996. *Columbia object image library (COIL-20), Technical Report CUCS-005-96*.

Nie, F.; Xiang, S.; Jia, Y.; and Zhang, C. 2009. Semi-supervised orthogonal discriminant analysis via label propagation. *Pattern Recognition* 42(11):2615–2627.

Nie, F.; Huang, H.; Cai, X.; and Ding, C. 2010. Efficient and robust feature selection via joint $\ell_{2,1}$-norms minimization. In *NIPS*.

Nie, F.; Huang, H.; Ding, C.; Luo, D.; and Wang, H. 2011a. Robust principal component analysis with non-greedy l1-norm maximization. In *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, volume 22, 1433.

Nie, F.; Wang, H.; Huang, H.; and Ding, C. 2011b. Unsupervised and semi-supervised learning via l1-norm graph. In *IEEE International Conference on Computer Vision (ICCV)*, 2268–2273.

Nie, F.; Yuan, J.; and Huang, H. 2014. Optimal mean robust principal component analysis. In *Proceedings of the 31st International Conference on Machine Learning (ICML)*, 1062–1070.

P.Ochs, and T.Brox. 2012. Higher order motion models and spectral clustering. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*.

Shi, J., and Malik, J. 2000. Normalized cuts and image segmentation. *IEEE Transactions on PAMI* 22(8):888–905.

Sim, T., and Baker, S. 2003. The cmu pose, illumination, and expression database. *IEEE Transactions on PAMI* 25(12):1615–1617.