LABIN: Balanced Min Cut For Large Scale Data

Group 3 :
CS22B2012
CS22B2011
CS22B2010
CS22B2009

December 8, 2023

Introduction

The topic of our presentation is Large Scale Balanced Minimum Cut. Many spectral clustering algorithms proposed in the past decades are not scalable for large-scale data due to their high computational complexities.

- Spectral Clustering methods usually involve eigen-decomposition of the similarity matrix and then the usage of k-means or spectral rotation to obtain final clustering results from the eigenvectors.
- Spectral Clustering generally has a high computational complexity of $O(n^3)$ for this reason.

Three main methods were proposed to solve the scalability problem:

- Nyström method.
- K-means based Approximate Spectral Clustering (KASP).
- Landmarks-based spectral clustering.

Spectral Clustering

Given an affinity matrix $A \in \mathbb{R}^{n \times n}$ constructed from a data-set $X = \{x_1, \ldots, x_n\}$. Let $\psi^{n \times c}$ be the cluster indicator matrix. The objective function of the classical Rcut is written as:

$$\min_{Y^T Y = I} \mathrm{Tr}(Y^T L_A Y) \tag{1}$$

The Objective function of Ncut is as follows:

$$\min_{Y^T D_A Y = I} \mathrm{Tr}(Y^T L_A Y) \tag{2}$$

where L_A is the Laplacian matrix and D_A is the corresponding degree matrix, where the i^{th} diagonal entry of the matrix D_A is the degree of the i^{th} node in the graph. $L_A = D_A - A$ where A is the adjacency matrix with weights. Problems (1) and (2) can be solved by performing eigen-decomposition on L_A first and then by clustering eigenvectors using k-means or spectral rotation.

Constrained Laplacian Rank

A new spectral clustering method was proposed, which learns a probability matrix $S \in \mathbb{R}^{n \times n}$ where s_{ij} is the connected probability of x_i and x_j from the following function.

$$\min_{S} \sum_{i,j=1}^{n} \quad \text{s.t. } S1 = 1 \text{ and } s_{ij} \in (0,1), \text{rank}(L_s) = n - c \qquad (3)$$

where a rank constraint $\text{rank}(L_s) = n - c$.

Spectral Rotation

Spectral rotation is the solution to the k-way NCut problem.

$$\max_{Y \in \psi^{n \times c}, Z = Y(Y^T D_A Y)^{-1/2}} \mathrm{Tr}(Z^T A Z) \tag{4}$$

where Z is the scaled partition matrix. It is difficult to solve this problem. A common method used for the relaxation of matrix Z which means to replace the constraint that Z is a partition matrix with a weaker constraint that Z is a matrix with non-negative entries and unit row sums. This relaxation allows for easier optimization of the objective function.

$$\max_{Z^T D_A Z = I_c, Z = Y(Y^T D_A Y)^{-1/2}} \mathrm{Tr}(Z^T A Z) \tag{5}$$

Spectral Rotation

The optimal solution of Z is $Z^*R : R^TR = I_c$ where $Z^* \in \mathbb{R}^{n \times c}$ consists of c column vectors of the eigenvectors of $D_A^{-1}A$ which correspond to the c greatest eigenvalues.
To obtain the discrete clustering assignment matrix Y, they first compute an approximate solution and then Y can be learned by solving the following problem

$$\min_{Y \in \psi^{n \times c}, R \in \mathbb{R}^{n \times c}, YI_{c \times 1}, R^TR = I_c} \mathrm{Tr}\,||Y - Y^*R||_F^2 \tag{6}$$

However, since Y^* is an approximate solution, the final discrete solution may deviate from the result by optimizing the original objective function.

Self Balanced Min Cut : The Predecessor

This algorithm can simultaneously minimize graph cut while allowing a balanced partition across the clusters which form as a result of the spectral clustering process.

Balance here means that, the two subsets which will be formed as result of the cut partition, should be having approximately the same size.

In this method, they have proposed the following objective function:

$$\min_{Y \in \psi^{n \times c}, s} ||A - sYY^T||_F^2 \tag{7}$$

where s>0 is a balance parameter. An iterative method is used to solve the problem.. To solve the problem, the initial problem has to be re-written as the following problem which is much easier to solve.

$$\min_{Y \in \psi^{n \times c}, s>0} 2s\,\mathrm{Tr}(Y^T AY) - s^2 ||Y||_b \tag{8}$$

Self Balanced Min Cut

$||Y||_b$ is the balance regularization used to obtain balanced partition which is defined as:

$$||Y||_b = \sum_{j=1}^{c}(\sum_{i=1}^{n} y_i j)^2 = \mathrm{Tr}(Y 1_n 1_n^T Y^T) \tag{9}$$

$$\max_{Y \in \psi^{n \times c}, s} , s(\mathrm{Tr}(Y^T A Y) - \frac{s}{2}||Y||_b) \tag{10}$$

It is pointed that solving the above-mentioned problem will automatically adjust the balance parameter s according to the learned cluster indicator matrix Y. On the other hand, maximizing s will maximize $\mathrm{Tr}(Y^T A Y)$, which is the objective function of min-cut.

Anchor Based Strategy

- The proposed solution to handle the scalability problem of spectral clustering, an anchor based strategy was proposed, which is called landmarks based method.
- Given a data-set $X \in R^{d \times n}$ in which d represents the number of features, and n represents the number of objects. The anchor based strategy first seeks m anchors to represent all the n objects (m $<<$ n), and then constructs a sample-anchor similarity matrix by calculating the distance between anchors and original objects. Anchor generation can be done by random selection or by k-means generation. K-means generation is preferred because, cluster centers have more validity than randomly selected data.
- The resultant m anchors represented by $W \in R^{d \times m}$. To compute the sample-anchor similarity matrix $B \in R^{d \times m}$ such that B1 = 1, with B we can obtain an affinity matrix A as

$$A = B\Delta^{-1}B^T \tag{11}$$

Anchor Based Strategy

Δ here represents a diagonal matrix, $\Delta \in R^{d \times m}$, whose j^{th} diagonal entry is defined as $\Delta_{jj} = \sum_{i=1}^{n} b_{ij}$. A can be re-written as

$$A = PP^{-1} \tag{12}$$

where $P \in R^{n \times m} = B\Delta^{-1/2}$. A local anchor embedding method was proposed to learn each b_i for $x_i \in X$ by solving the following problem:

$$\min_{b_i^T 1=1, b_i >=0} \frac{1}{2} ||x_i - U_k(x_i) b_i||_2^2 \tag{13}$$

where the k column vectors of $U_k(x_i) \in R^{d \times k}$ are the k nearest anchors of x_i. With the learned bipartite graph B we acquire a semi-supervised learning algorithm.

LABIN : Balanced Min Cut for Large Scale Data

The aim is to cluster the data matrix X into c clusters and let $Y \in \psi^{n \times c}$ be the cluster indicator matrix. we first construct an affinity matrix A. Substituting the affinity matrix gives:

$$\max_{Y \in \Psi^{n \times c}, s>0} 2s \, \mathrm{Tr}(Y^T P P^T Y) - s^2 ||Y||_b \tag{14}$$

where s>0 is the balance parameter. $P \in R^{n \times m} = B\Delta^{-1/2}$ and $\Delta \in R^{m \times m}$ is a diagonal matrix in which the j^{th} diagonal entry is defined as $\Delta_{jj} = \sum_i^n b_{ij}$ Θ
This new optimization problem can be solved by using an alternative optimization approach. The approach involves solving for Y by fixing s.

LABIN : Balanced Min Cut for Large Scale Data

If s is fixed, the problem can be re-written as:

$$\max_{Y \in \psi^{n \times c}} \mathrm{Tr}(Y^T \Theta Y) \tag{15}$$

where $\Theta = PP^T - s/2 \times 1_n 1_n^T$.
This above problem can be solved using the Augmented Lagrangian method. But ALM of the SBMC performs badly on large scale data since it is difficult to adjust the balance and regularization parameters.
In this algorithm we use improved spectral rotation to solve this problem.
Specifically we involve the usage of relaxation method to relax the matrix Y from the discrete values to continuous values.

$$\max_{Y^T Y = I_c} \mathrm{Tr}(Y^T \Theta Y). \tag{16}$$

Singular Value Decomposition : A little detour

The optimal solution of Y is $Y = Y * R : R^T R = I_c$ where Y* is the c column vectors of the eigen-vectors of Θ which correspond to the c biggest eigen-values.
It is a matrix factorization method that breaks down a matrix A into three constituent matrices, U, Σ and V^T such that:

$$A = U\Sigma V^T \tag{17}$$

U and V are orthogonal matrices and Σ is the diagonal matrix containing the singular values of A.

- The columns of U are the eigen-vectors of AA^T.
- The columns of V are the eigen-vectors of $A^T A$.
- The diagonal elements of Σ are the square roots of the eigen values of AA^T or $A^T A$.

LABIN : Balanced Min Cut for Large Scale Data

To find the solution $Y*$ and consequently Y for the matrix Θ, the algorithm has a time complexity of $O(n^3)$, where $\Theta \in R^{n \times n}$ and $Y* \in R^{n \times c}$.

This algorithm propose an efficient method to solve the eigen-vectors of Θ which correspond to the c greatest eigen-values. We first re-write Θ in matrix form as follows:

$$\Theta = Q \begin{bmatrix} I_m & 0_m \\ 0_m^T & -\frac{s}{2} \end{bmatrix} Q^T \tag{18}$$

where $Q \in r^{n \times (m+1)}$ is defined as $Q = [P1_n]$. Since $n >> m$ the matrix Q is no more a square matrix so, we need to go for Singular Value Decomposition(SVD) on Q such that :

$$Q = U_Q \Sigma_Q V_Q^T \tag{19}$$

where $U_Q \in R^{n \times (m+1)}$ and $V_Q \in r^{(m+1) \times (m=1)}$ are two unitary matrices and $\Sigma_Q \in R^{(m+1) \times (m+1)}$.

LABIN : Balanced Min Cut for Large Scale Data

The matrix Θ is further re-written as :

$$\Theta = U_Q \Phi U_Q^T \tag{20}$$

where $\Phi \in r^{(m+1)\times(m+1)}$ is defined as:

$$\Phi = \Sigma_Q V_Q^T \begin{bmatrix} I_m & 0_m \\ 0_m^T & -\frac{s}{2} \end{bmatrix} V_Q \Sigma_Q \tag{21}$$

Now eigen-decomposition is performed on matrix Φ.

$$\Theta = U_Q V_\Phi \Sigma_\Phi V_\Phi^{-1} U_Q^T \tag{22}$$

$$\Theta U_Q V_\Phi = U_Q V_\Phi \Sigma_\Phi \tag{23}$$

It is known that $U_Q V_\Phi$ consists of $(m+1)$ eigen-vectors of Θ and the diagonal entries of Σ_Φ are the eigen-values. Therefore to find the optimal solution to the cluster indicator matrix, are the c column vectors of $U_Q V_\Phi$ which correspond to the c biggest eigen-values in Σ_Φ.

LABIN: Part A Summary

- To summarize the approach through fixing the balance parameter s, and then solving for the cluster indicator matrix which is done using an alternative optimized approach unlike the classic Augmented Lagrangian and Anchor Based Strategies.
- In this alternative method we only perform the reduced *SVD* of matrix Q and eigen-decomposition of matrix Φ.
- Since $\Phi \in R^{(m+1)\times(m+1)}$ is matrix of order $(m+1) \times (m+1)$, the cost for eigen decomposition is reduced to $O(n^3)$.
- It can be verified that the singular values in matrix Σ_Q are square roots of the eigen-values in matrix Σ_Φ. Then U_Q can be obtained in $O(nm^2)$ time. Which is a significant reduction from the $O(n^3)$ of the conventional methods, keeping in mind that $n >> m$. The current time complexity can be called pseudo-linear with respect to the number of data-points in input matrix X.

LABIN : Part B

Solving s with Y fixed. Unlike the previous method where we fix the balance parameter s, this approach modifies the optimization algorithm to:

$$min_{s>0}(s - \frac{\mathrm{Tr}(Y^T AY)}{||Y||_b})^2 \tag{24}$$

$$\tag{25}$$

The optimal solution of s becomes:

$$s = \frac{\mathrm{Tr}(Y^T AY)}{||Y||_b} \tag{26}$$

Algorithm Summary

To summarize our LABIN algorithm:

- Given a data-set, a sample anchor similarity matrix is constructed first and then the balance parameter and cluster indicator matrix are iteratively updated till convergence conditions are met.
- This algorithm takes $O(r_1(nm^2 + r_2mc^2))$ time to iteratively solve for s and Y.
- r_1 is the number of iterations to update s and r_2 is the average number of iterations to update Y.
- Note that Y converges very fast due to a limited solution space so r_2 is usually very small.
- Therefore LABIN has a time complexity of $O(nm^2)$, $(m >> c)$. Which is a significant reduction compared to the conventional methods of spectral clustering.

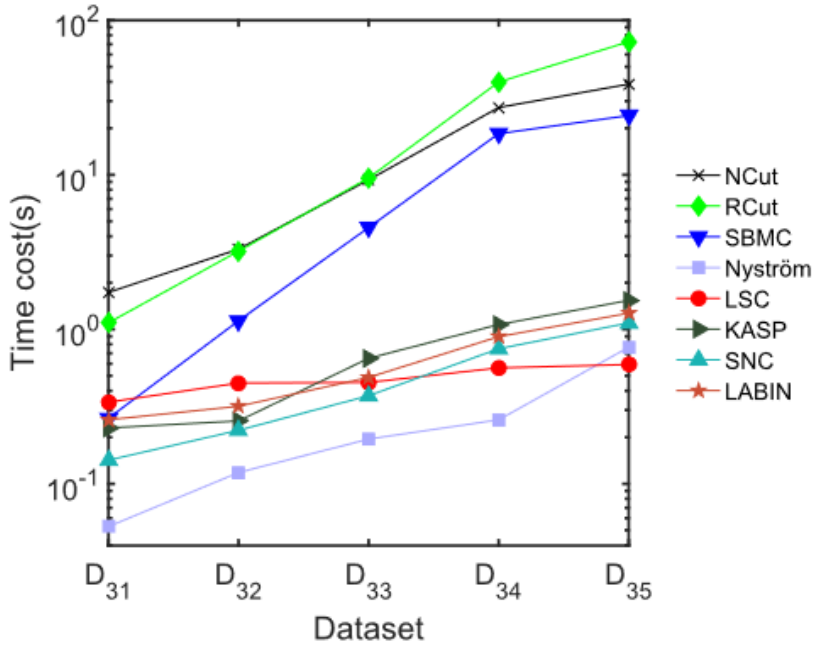
Results and Analysis

- A total of 11 data-sets were taken to test the validity of the new algorithm, the testing was carried out for these data-sets by LABIN and 4 approximation based spectral clustering methods: KASP, Nyström, Landmarks based SC, Scalable Normalized Cut, along with NCut with k-means and RCut with k-means to check if LABIN can produce comparable results with non-approximate methods.

Results show that LABIN:

- Outperformed all four approximation based spectral clustering methods.
- Achieved 18% improvement compared to Scalable Normalized Cut in terms of accuracy.
- Showed a 18% improvement in terms of assigning the data-points to their true clusters.
- Outperformed SBMC, LABIN's parent algorithm on their performance on large-scale datasets.

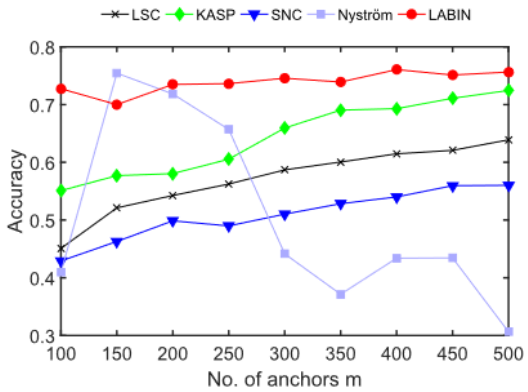
Results and Analysis



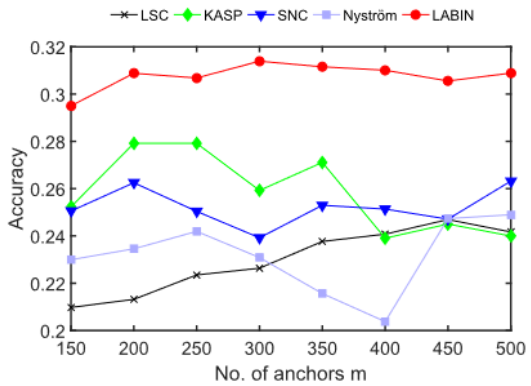
The above graph shows the time-costs of the eight clustering methods on 5 large data-sets where number of anchors was selected to be 300 and number of nearest neighbours was selected to be 10.

Results and Analysis

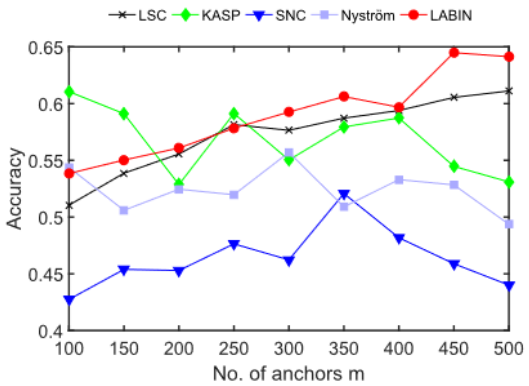
Let us observe how the clustering results vary with change in parameters. The first picture shows accuracy v/s number of anchors, the second shows accuracy v/s no. of nearest neighbours, LABIN v/s the approximation based methods.



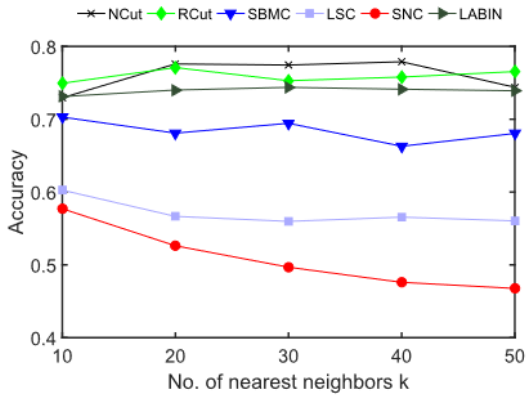
(a)

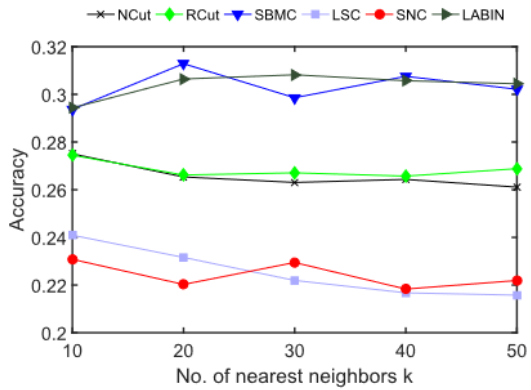


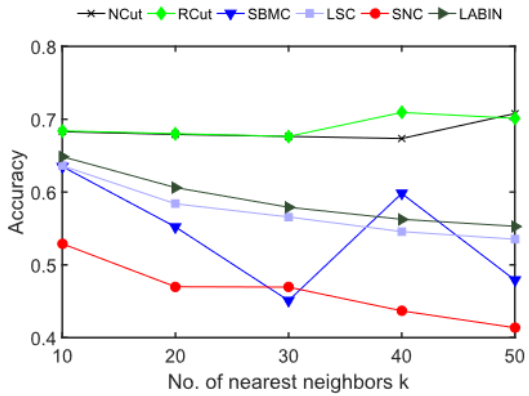
(b)



(c)

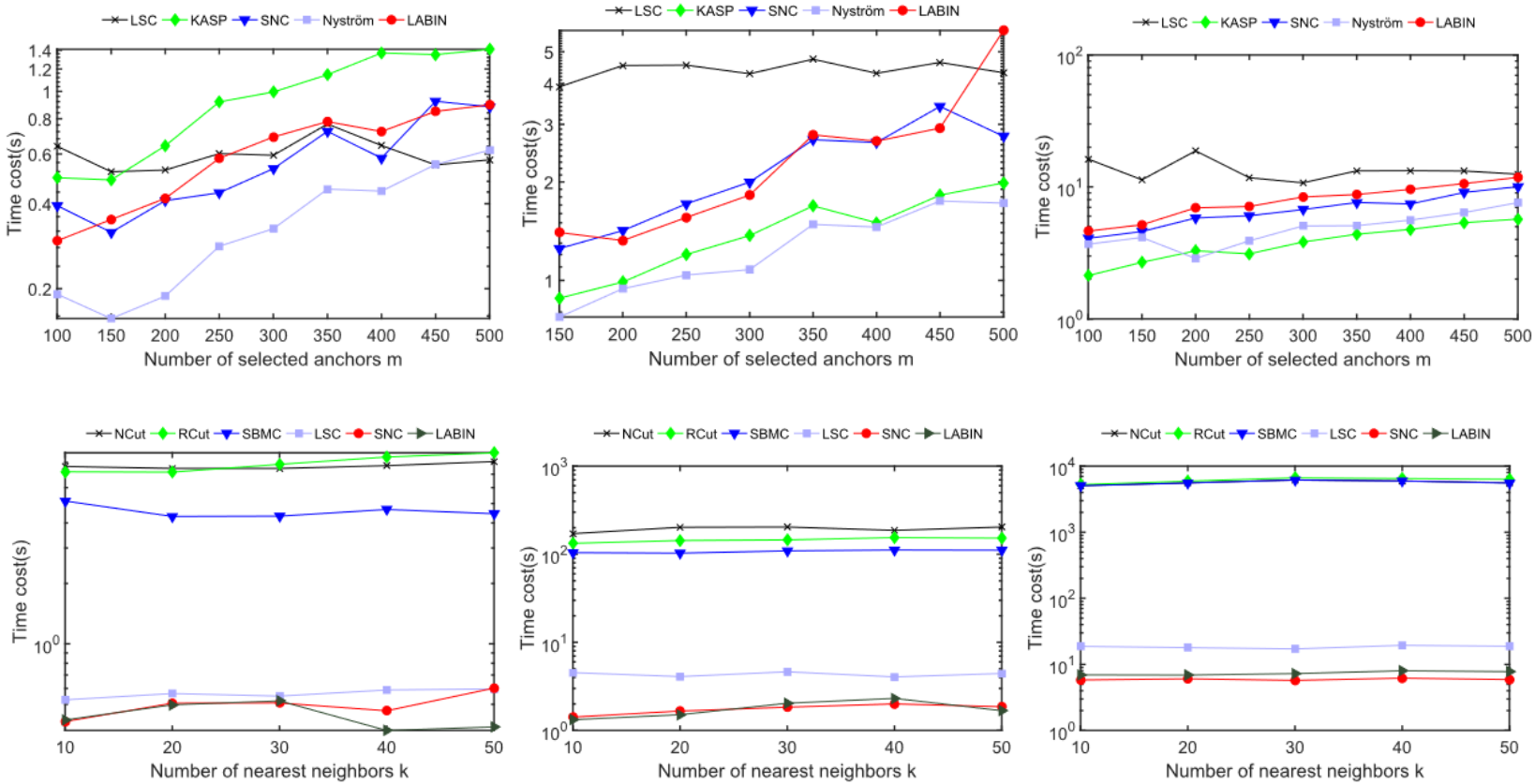






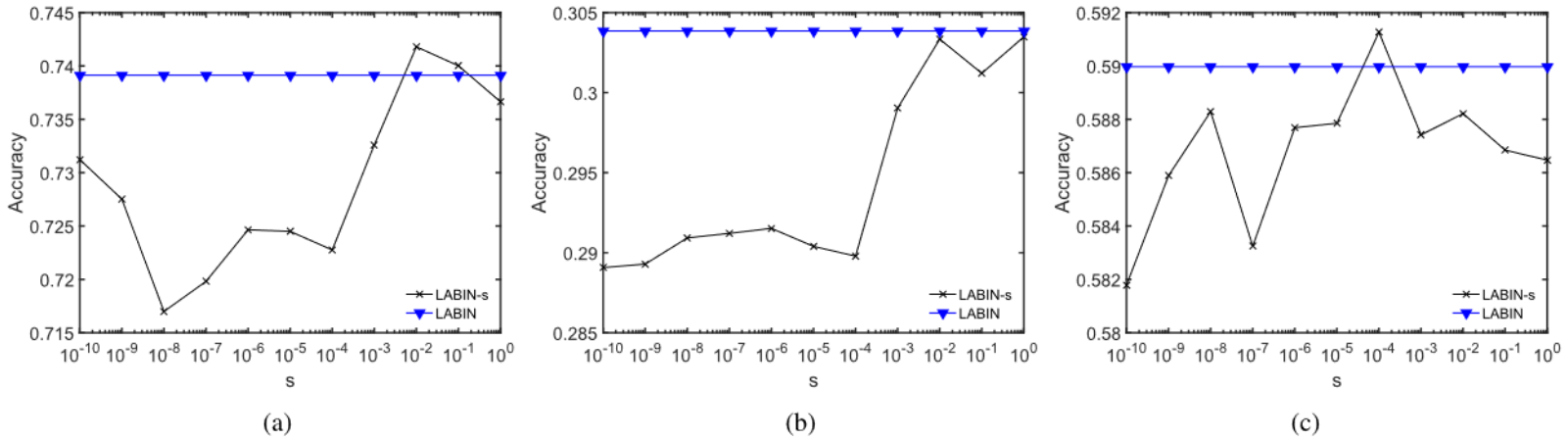
Results and Analysis

Variation of clustering results with change in parameters. The first picture shows time cost v/s number of anchors, the second shows time-cost v/s no. of nearest neighbours, LABIN v/s the non-approximation based methods.



Results and Analysis

With respect to the balance parameter, it has been found through these experiments the fixing the cluster indicator matrix and then solving for s, is very time consuming.
However the clustering result obtained is closest to the optimal solution, when we fix s and Y can be obtained in at-least 1 iteration.



Accuracy v/s variation in balance parameter graphs

Conclusion

- In the research paper, a new spectral clustering algorithm has been proposed for large-scale data, which involved the construction of a sample-anchor similarity matrix, instead of the full similarity matrix, in which improved Spectral Rotation methods are used to optimize the new model.
- This method is efficiently able to perform eigen-decomposition with a pseudo-linear time and space computational complexities.
- Compared with the original SBMC, this method has a significant reduction in space-complexity due to ISR(Improved spectral Rotation).
- Extensive experiments on large-scale data show the effectiveness and efficiency of LABIN compared to the conventional methods.

Thank You