IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)
e-ISSN: 2278-2834,p- ISSN: 2278-8735.Volume 10, Issue 2, Ver. II (Mar - Apr.2015), PP 36-40
www.iosrjournals.org

Curtailment of Clustering Using Soq

K. Dinusha, G. Anjali, Ch. S. Surya Raj, M. Raviteja, CH. Gayatri
(Electronics and Communication Engineering,Lendi Institute of Engineering and Technology,India)
(Electronics and Communication Engineering,Lendi Institute of Engineering and Technology,India)
(Electronics and Communication Engineering,Lendi Institute of Engineering and Technology,India)
(Electronics and Communication Engineering,Lendi Institute of Engineering and Technology,India)
(Asst.professor, Electronics and Communication Engineering,Lendi Institute of Engineering and Technology,India)

Abstract: *This paper is about self organizing queue based clustering and its algorithms. Self organizing clustering is more useful compared to spectral clustering. The draw-backs of graph clustering has been overcome by spectral clustering i.e., similarity matrix of a set of data points or nodes, this problem is commonly referred as k/a graph clustering. Experimental results have been clearly shown their dominance over other existing techniques such as Spectral clustering, Graph clustering and k-means algorithm. The experiment will be done by using matlab software.*

Key Words: *clustering, graph clustering, spectral clustering , k means.*

I. Introduction

This paper is mostly discussing about the problem of graph clustering i.e., given the similarity matrix of a set of points or data .clustering is defined as "It Is the task of grouping a set of objects in such a way that objects in the same group (called a **cluster**) are more similar (in some sense or another) to each other than to those in other groups (clusters)".Clustering is used in wireless sensor networks. Hierarchical clustering in wireless sensor networks is used to increase robustness, system scalability, life time. This is normally used in data aggregation and fusion this will decrease the number of transmitted messages to the base station. WSN is having several challenges in terms of designing and implementation and these are overcome by using clustering techniques. Graph clustering is finding sets of "related" vertices in graphs. Graphs are structures formed by a set of vertices (also called nodes) and a set of edges that are connections between pairs of vertices. The draw-back of spectral clustering overcomes by spectral clustering. Spectral clustering technique outcome is not satisfactory for many applications to improve its outcome we take bio-inspired approach. The main idea is to place all nodes into multiple fictitious queues, each of which corresponds to one cluster; then we enable these fictitious queues with self-organizing capability to group similar nodes into the same cluster; we call the resulting scheme, Self-Organizing-Queue (SOQ) clustering scheme. The Self-Organizing Queue (SOQ) clustering algorithm offers more accurate clustering than spectral clustering or graph clustering algorithms. Based on the concept of self-organization, it has low computational complexity and can be applied to large data sets, making it valuable for data mining applications, social networks such as Facebook and Twitter, machine learning and artificial intelligence. Computer simulation show that SOQ clustering results in a significantly lower error rate than spectral clustering or graph clustering algorithms.

Example: Let us consider an example that how the vehicles do grouping. At the beginning of time slot i.e., t=1, let the vehicles of two diff-erent types are mixed and form a row in parking place. Our main aim is to separate those vehicles into two groups(alike groups).For this procedure we will select one vehicle and we will move that vehicle to any of the side and that vehicle is called as the current vehicle, after that we consider another vehicle and move that vehicle to its corresponding place, and this vehicle is called as next vehicle. The next vehicle and the current vehicle are compared to each other, if the both vehicles are of same type the next vehicle is to move the left side of the current vehicle; otherwise the vehicle is move to the tail of the line. This process will continue in each time slot and only vehicle is allowed to move in that particular time slot. This process is continued until the two different types of vehicles are formed two groups.

There are three algorithms for Self-Organizing Queue based clustering they are:-
1.SOQ 2.CESOQ 3.MSSOQ

Algorithm 1:- This is an important step in SOQ. Current queue is pointing to a queue, current person is pointing to a member in a current queue, where as next queue is pointing to a queue. The current queue, current person ,next person are normally referred as variables. Similarity matrix '**X**' has a dimension of **t x t** and each entry denotes as **Wyz ,** the similarity measure between Node y and Node z . If the input is dissimilar matrix we can convert that dissimilar matrix into similar matrix by adding negatives to the dissimilar matrix.

Steps of dissimilar matrix are:
Input: a set of **t** nodes and similarity matrix **w**
1) Initialization :Divide the set of **t** nodes into k queues.
2)While(Flag)
3) WHO: Choose who in Current Queue as Current Person.
4) HOW: (How to) select a queue as Next Queue for Current Person to join.
5) WHERE: (Where to) place Current Person in Next Queue.
6) Assign Next Queue to Current Queue.
7) WHEN: (When to) let us stop the loop.
8) End while Output: the resulting queues/clusters.
From the algorithm 1, we clearly understand that there are three key features in SOQ those are
1)Self-Organizing i.e., each person or node has the ability to decide where it wants to join.
2) sequential process i.e., person by person at one time only one queue is selected as current-queue.
3) similarity matrix i.e., we already know that it will take even dissimilar matrix and converts them into similarity matrix by adding negative entities to the dissimilar matrix. Note that none of the existing spectral clustering algorithms allows asymmetric similarity matrix and similarity matrix with negative entries .

The steps of 1,3,5,7 can be implemented as in Step1 (Initialization), the default setting is to divide the set of nodes into queues by random selection, and assign the queue with the most members to Current Queue. Another way of obtaining initial queues is to use a spectral clustering technique any spectral clustering technique is applicable here.
In Step 2 (WHO), we can choose who is the current person in a current queue in default it is head of current queue is the current person. In Step 4 (HOW), Current Person can use the following default criterion (called Most Friends) to choose Queue as the Next Queue to join:

$$\widehat{K} = \arg Max \frac{\sum_{j \in C_k}(W_{i,j} + W_{j,i})}{|C_k|}$$

(1) Where is the set of indices of members in Queue. Note that Current Person itself is not counted in its Current_Queue in (1) since Current Person is looking for a queue with most friends excluding itself.
In Step 5 (WHERE) , it is to decide where we have to place current person in the next queue ,in default current person will be at the tail of next queue.
In Step 7 (WHEN),it will decide when we will stop the algorithm ,in default it will stop when none of the queues has membership change Algorithm 1 has a limitation: when the current person joins the next queue the current queue will be empty ,then the number of clusters will be reduced to k-1,instead of target value k. To rectify this , we propose CESOQ (one Cluster being Empty SOQ) as shown in Algorithm 2; specifically, we insert Step 5.1 between Step 5 and Step 6 in Algorithm 1. In Algorithm 2, the Crosstalk Density between Queue ‘s’ and Queue ‘t’ is defined as:

$$\rho(C_s, C_t) = \frac{\sum_{j \in C_s}\sum_{i \in C_t}(W_{i,j} + W_{j,i})}{2 \times |C_s| \times |C_t|}$$

Algorithm 2:(CESOQ)
Input and step 1 to 5 is same as algorithm1.
5.1) If (Current_Queue is empty) Use Algorithm 1 to partition each of the non-empty queues into two clusters (a pair of clusters).
Find the pair of clusters with minimum crosstalk density. One of the two clusters with minimum crosstalk density replaces its original queue, and the other cluster replaces the empty queue.
Endif
Step 6 through 8 and Output are the same as Algorithm 1.
If we are allowing multiple queues to be selected into current queues causes non convergence in all experiments.
Algorithm 2 has a limitation i.e., cluster talk of a cluster must be smaller than the cross talk between clusters.
To rectify this, we are looking towards MSSOQ(Merge Split SOQ) as shown in Algorithm 3.
Specifically , we insert Step 1.1 and 1.2 between Step 1 and Step 2, and append new Steps 9–15 to Step 8 in Algorithm 2.

Algorithm3:(MSSOQ)
Input and Step 1 are the same as Algorithm 2.
1.1) Flag2=1
1.2) While (Flag2)

Step 2 through Step 8 are the same as Algorithm 2.
9) Use (2) to calculate for $\rho(Cs, Ct)$ for s$\neq t$,s $\in \{1, \ldots..k\}$ and t $\in \{1, \ldots \ldots k\}$
10)Compute $\rho between$=max{s,t:s$\neq t\}\rho(Cs, Ct)$ and{s1*,t1*}= arg max{s,t:s$\neq t\}$ $\rho(Cs, Ct)$.
11) Use Algorithm 1 to split each queue into two new queues, denoted by Cs1 and Cs2 .
12) Use (2) to calculate for $\rho(C_s, 1, C_s, 2)$for s $\in \{1, \ldots., k\}$.
13) Compute $\rho_{within} = min_{s\in\{1,\ldots,k\}\rho(Cs,1,Cs,2)}$, and s_2^*arg $min_{s\in\{1,\ldots.k\}\,\rho(Cs,1,Cs,2)}$
and s_2^*=arg $min_{s\,\in\{\,1,\ldots,K\}\rho(}C_{s,1,C_{S,2}})$
14) If ($\rho_{within} < \rho_{between}$)
If($s_1^* == s_2^*$)
If ($\rho(C_s *_2$,1,$C_t *_1) > \rho(\,Cs_2^*$,2, $C_t *_1$)
Merge $C_s *_2$ and $C_t *_1$ into a new Queue t_1^*, and call $C_s *_2$ as a new Queue s_2^* .
Else Merge $C_s *_{2,}$ 2 and $C_t *_1$ into a new Queue , and call as a new Queue s_2^* .
Endif
Else
If ($s_2^* == t_1^*$)
If ($\rho(C_{s_2^*,1}, C_{s_1^*}) > \rho(C_{s_2^*,2}, C_{s_1^*})$)
Merge $C_{s_2^*,1}$ and $C_{s_1^*}$ into a new Queue s_1^* , and call
$C_{s_2^*,2}$ as a new Queue s_2^*. .
Else Merge $C_{s_2^*,2}$ and $C_{s_1^*}$ into a new s_1^* Queue , and call $C_{s_2^*,1}$ as a new Queue s_2^*.
Endif
Else
Merge $C_{s_1^*}$ and $C_{t_1^*}$into a new Queue s_1^* , call $C_{s_2^*,1}$ as a new Queue s_2^*. , and call $C_{s_2^*,2}$as a new Queue t_1^*
Endif
Endif
Else Flag2=0.
Endif
15) Endwhile Output is the same as Algorithm 2.

II. Experimental Results

In this paper, we check the performance of MSSOQ with synthetic data and compared it with the existing main-stream clustering algorithms, i.e., K-means, un-normalized spectral clustering (SC for short), the normalized spectral clustering algorithms.

Experiment with synthetic data set:

1.Description of synthetic data set: Synthetic data is mainly used for the production of data for the required situation that will not be obtained by direct measurement. It can also defined as information can be persistently stored and used by professionals. Synthetic data outcome may not be same as real , original data. Synthetic data can be used in designing of any system because the synthetic data can be used for simulation ,theoretical values, this will produce unexpected results and remedy if the produced results are unsatisfactory. Synthetic data is used to generate an authentic data and also used to protect privacy and confidentiality of authentic data.
The synthetic data consist of 2D Gaussian Data .To simulate that we use four 2D Gaussian distribution

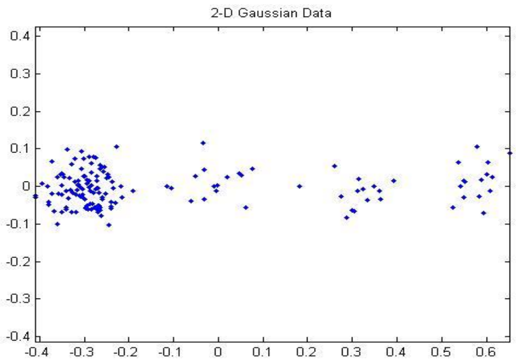

Fig1: 2D Gaussian Data

Let us take queues before randpermed i.e., 1st cluster 15 members,3rd cluster 105 members 2nd and 4th clusters of 15 members and the output has been obtained for the randomly permitted similarity matrix.
Show queues results before randpermed:

The 1st cluster,15 members are:121 122 123 124 125 126 127 128 129 130 131 132 133 134 135
The 2nd cluster,15 members are:136 137 138 139 140 141 142 143 144 145 146 147 148 149 150
The 3rd cluster,105 members are:1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105
The 4th cluster,15 members are:106 107 108 109 110 111 112 113 114 115 116 117 118 119 120

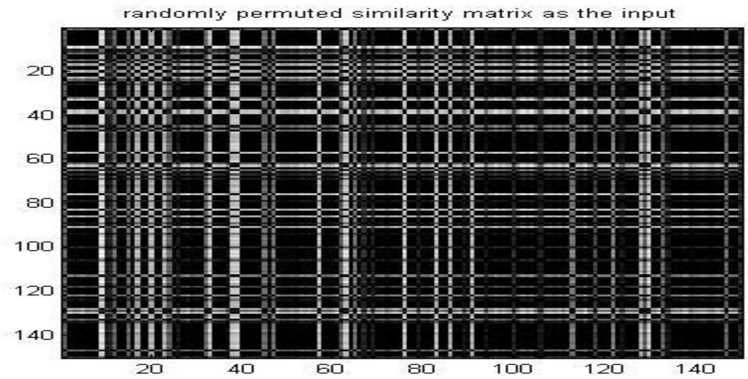

Fig:- Randomly permitted similarity matrix as the input.

Show GroundTruth Results:The ground table rearranged the clusters in an descending order such as 105,15,15,15 and the graph has been shown below
The 1st cluster,105 members are:1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105
The 2nd cluster,15 members are:106 107 108 109 110 111 112 113 114 115 116 117 118 119 120
The 3rd cluster,15 members are:121 122 123 124 125 126 127 128 129 130 131 132 133 134 135
the 4th cluster,15 members are:136 137 138 139 140 141 142 143 144 145 146 147 148 149 150

The confusion matrix is: (Row is the GroundTruth)

0	0	105	0
0	0	0	15
15	0	0	0
0	15	0	0

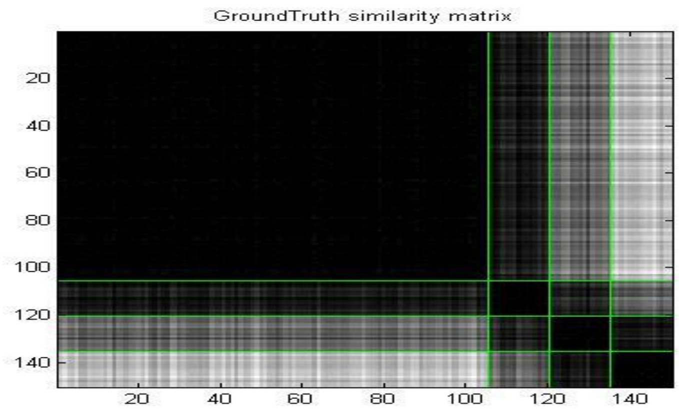

Fig:- Ground Truth similarity matrix

The confusion matrix is: (After Matching.)

105	0	0	0
0	15	0	0
0	0	15	0
0	0	0	15

The final result after matching the matrix is given in an order and the final graph of SOQ is shown below:

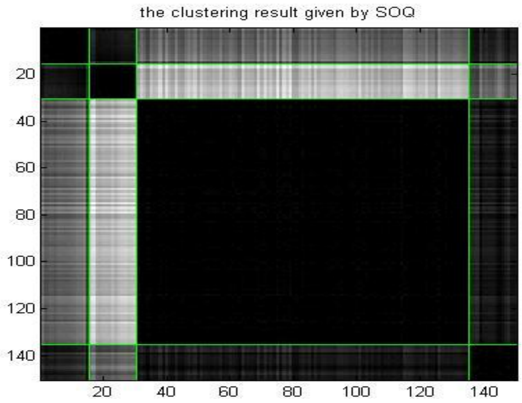

Fig:-The clustering result given by SOQ

III. Conclusion

MSSOQ is a best method compared to all other techniques such as Graph clustering, Spectral clustering, k-means algorithm .It is mostly used to protect privacy and confidentiality of authentic data.

References

[1]. U. Von Luxburg, "A tutorial on spectral clustering," Statist. Comput., vol. 17, no. 4, pp. 395–416, 2007.
[2]. S. Fortunato, "Community detection in graphs," Phys. Rep., vol. 486, no. 3, pp. 75–174, 2010.
[3]. A. Ng, M. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," Adv. Neural Inf. Process. Syst., vol. 2, pp. 849856, 2002.
[4]. J. Shi and J. Malik, "Normalized cuts and image segmentation," IEEE Trans. Patt. Anal. Mach. Intell., vol. 22, no. 8, pp. 888–905, 2000.
[5]. D. Verma and M. Meila, A Comparison of Spectral Clustering Al- gorithms Univ. Washington, Seattle, Tech. Rep. UW-CSE-03-05-01, 2003.
[6]. D.VermaandM.Meila,digit1000.mat2003[Online].Available:http://www.stat.washington.edu/spectral/datasets.html
[7]. Self–Organised queue based clustering by Baohua Sun and Dapeng Wu IEEE SIGNAL PROCESSING LETTERS,VOL.19,NO.12,DECEMBER 2012