International Conference on Control, Engineering & Information Technology (CEIT'13)
Proceedings Engineering & Technology - Vol.4, pp.176-177, 2013

SB-CLUST: Spectral b-Coloring based Clustering Algorithm[1]

H. Hablal[1], H. Ait Haddadene[2], N. Belacel[3]
[1]University of Bejaia, Technology Department, Technology Faculty, Algeria
e-mail: hablalhouria@yahoo.fr
[2]USTHB-University, Operations Research Department, LaROMaD Laboratory, Algiers, Algeria
e-mail: aithaddadenehacene@yahoo.fr
[3]University of Moncton, Computer Science Department, Moncton, New Brunswick, Canada
e-mail: nbelacel@gmail.com

ABSTRACT

Spectral clustering is a powerful technique in data analysis. It has become one of the most popular modern clustering algorithms and it is widely used in many real world applications. It groups entities by using eigenvalues and eigenvectors of matrices derived from data. It is simple to implement and can be solved efficiently by standard linear algebra software. Another method recently developed for clustering problem is the b-coloring method. In this work we propose a new approach, called SB-CLUST, to improve the performance of spectral clustering, which is based on both graph b-coloring algorithm and the recursive spectral approach. This is the first time a method based on spectral and graph b-coloration algorithm proposed for clustering problems. In this paper we will perform a comparative study between our developed method SB-CLUST and the classical clustering spectral algorithm using UCI data sets. The comparative results show that SB-CLUST algorithm has a great potential to solve a clustering problems.

Keywords—Unsupervised learning, Spectral clustering, b-coloring clustering

I. INTRODUCTION

The clustering approach is the task of grouping a set of entities in such away entities in same cluster are more similar to each other than those belong to other clusters [1]. The spectral clustering methods are based on the calculation of eigenvectors and eigenvalues of the similarity matrix of entities to be clustered. It consists of extracting the eigenvectors associated to the largest eigenvalues (spectrum) of Laplace matrix. These eigenvalues constitute a reducible dimension space where the transformed data are linearly separable. Spectral algorithms differ basically in the number of eigenvectors they use for partitioning and the similarity matrix [3].

In this work we will introduce a new method called sb-clust. This newly proposed method empowers the existing spectral algorithm by the b-coloring algorithm proposed in [4]. The principle of our algorithm is to decompose recursively the set of entities into two subsets "clusters", and then it chose the subset to divide by the spectral clustering algorithm and so on until the number of the subsets will reach the number k of clusters given a priori. At the end of the application of the recursive spectral clustering we will apply the b-coloring algorithm to improve the final clustering results.

II. MATERIALS AND METHODS

In this section we present SB-CLUST method, which is able to solve the clustering problem. The SB-CLUST algorithm is a hybrid method that combines the spectral clustering algorithm and b-coloring technique to improve the clustering results. Before we describe our proposed algorithm, we will provide a brief overview of the both algorithms spectral and b-coloring used by our method.

1. Spectral clustering algorithm:

Spectral clustering is became a popular clustering method [3]. It has been applied to several applications including image segmentation [5], social network analysis [6] and text mining [7]. For more details please refer to von Luxburg [8]. The spectral clustering techniques calculate the eigenvalues of the similarity matrix of entities to perform dimensionality reduction before clustering in fewer dimensions by using the most popular local search heuristic k-means MC-Queen [8]. k-means is an interchange heuristic, where points are reassigned to another cluster than their own, one at time, until a local optimum is reached.

2. b-coloring algorithm:

The b-coloring based clustering method is inspired from the graph theory approach and known as b-coloration algorithm. The algorithm was presented in [4]. It enables to build a fine partition of the data set into clusters. It has several desirable clustering properties: utilization of topological relations between objects, robustness to outliers, all types of data can be accommodated, and identification of each cluster by at least one dominant object.

3. The SB-CLUST: Spectral b-Coloring based Clustering Algorithm:

The different steps of SB-CLUST algorithm are presented in Figure 1. In the first step a recursive spectral algorithm is applied. In each splitting step a validation index was calculated based on the similarity between the entities and the centroids of the clusters. In this work we have choose the Dunn index as presented as [10]. After the recursive spectral algorithm was applied we will use at the end the famous graph theory approach called b-coloring algorithm as developed by Haytham Alghazel [4].

[1] This work was supported by the Algerian PNR (SAD4) of the DGRSDT, Ministry of Higher Education and Scientific Research of Algeria

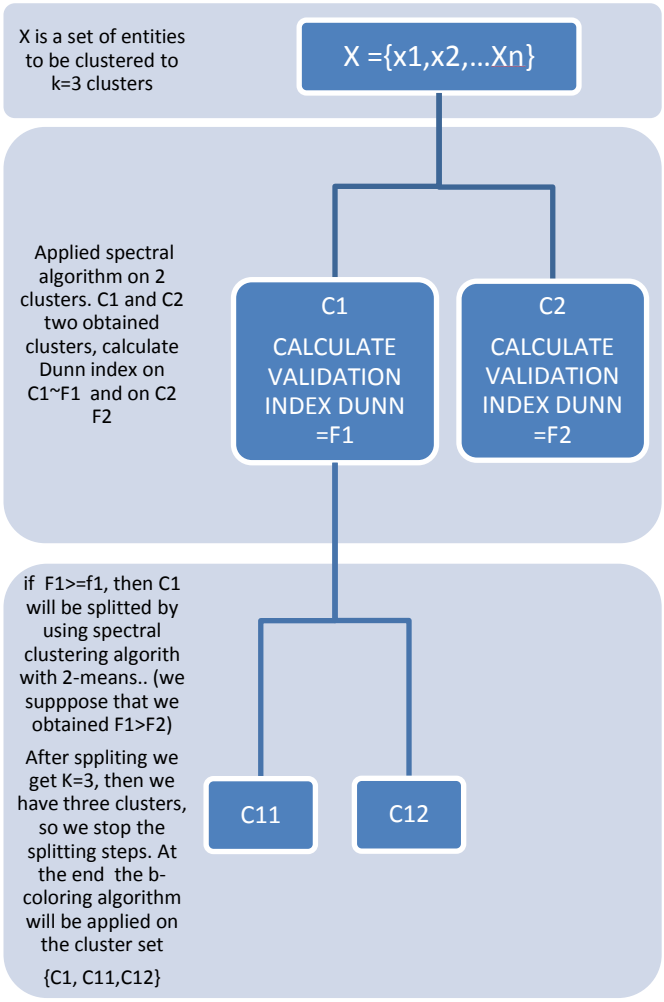

Fig. 1 SB-CLUST ALGORITHM

III. RESULTS

To test the effectiveness of our proposed algorithm, we have compared its performance with the classical spectral clustering algorithm. It is noted that an objective function based on the similarities was used for this comparison.

The preliminary results are encouraged and show that our approach has a great potential for data clustering.

To compare the two methods, we calculate *Generalized Dunn's* index [2]. The highest value of *Generalized Dunn's* indicates the best clustering. The comparison between the classical clustering algorithm and ours shows that our proposed algorithm is more efficient.

Table 1 shows comparative results between the proposed clustering method and spectral algorithm. The three data sets considered in this study are Iris, Breast cancer and Wine obtained from the UCI machine learning repository (http:www.ics.uci.edu/~mlearn/MLrepository.html).

TABLE I. .COMPARAISON RESULTS OF DUNN INDEX

Simple	# clusters	Method	$dunn_G$
1 IRISDATA	3	spectral	3.7674
		sb-clust	3.6588
1 IRISDATA	4	spectral	1.7728
		sb-clust	3.4941
2 BreastCancer	3	spectral	2.4357
		sb-clust	2.4359
2 BreastCancer	4	spectral	1.6262
		sb-clust	2.4273,
2 BreastCancer	5	spectral	1.4416
		sb-clust	2.1128
2 BreastCancer	6	spectral	1.4416
		sb-clust	2.2084
3 WineData	2	spectral	4.5246
		sb-clust	6.3227

IV. DISCUSSION AND CONCLUSIONS

In this work we have proposed a new hybrid clustering approach SB-CLUST that combines spectral algorithm and b-coloring approach for clustering problem. To the best of our knowledge, it is the first time that an hybrid clustering method based on spectral and b-coloring graph are applied as method for solving the clustering problem. The first results shows that our proposed approach has a great potential to solve the clustering problems. In the future work more experiments analysis will be done with a comparison with several other clustering methods. And also we plan to introduce metaheuristics to our local search approach to improve the performances of our proposed algorithm SB-CLUST.

V. REFERENCES

[1] N. Belacel, P. Hansen and N. Mladenovic, Fuzzy J-means: a new heuristic for fuzzy clustering, *Pattern Recognition*, Vol.10, 35, pp. 2193-2200, 2002.

[2] J. C. Bezdek and N. R. Pal, Some new indexes of cluster validity, *IEEE Transactions on Systems, Man and Cybernetics*, 28(3), pp. 301-315, 1998.

[3] A. Ng, M. Jordan, and Y. Weiss, *On spectral clustering: analysis and an algorithm*, Ed. T. Dietterich, S. Becker, and Z. Ghahramani, *Advances in Neural Information Processing Systems* Vol. 14, pp. 849-856, 2002.

[4] H. Elghazel, V. Deslandres, M.S. Hacid, A. Dussauchoy, and H. Kheddouci, A new clustering approach for symbolic data and its validation: Application to the healthcare data, Proceedings of ISMIS2006, Springer Verlag, pp. 473-482, 2006.

[5] J. Shi and J. Malik, Normalized cuts and image segmentation, *IEEE Transactions on Pattern Analysis and Machine Intelligence* vol. 22, pp. 888-905, 2000.

[6] M. Newman, D. Watts, and S. Strogatz, Random graph models of social networks., *Proc.Natl. Acad. Sci.* USA 99 pp.2566-2572, 2000

[7] I. Dhillon, Co-clustering documents and words using bipartite spectral graph partitioning., In Proceedings of the *Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD),*ACM Press, New York269-274, 2001.

[8] U. V. Luxburg, A tutorial on spectral clustering", *Statistics and Computing*, vol. 17, pp. 395 - 416, Dec. 2007

[9] J. B. MacQueen, Some Methods for classification and Analysis of Multivariate Observations, *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press. pp. 281–297, 1967.