Visual Clustering Approaches

Lydia Boudjeloud-Assala, Brieuc Conan-Guez, Alain Gély

Département STatistique et Informatique Décisionnelle (STID), IUT de Metz,

Laboratoire d'Informatique Théorique et Appliquée de Metz, Université de Lorraine,

Contact E-mail : lydia.boudjeloud-assala@univ-lorraine.fr

Abstract

The data stored in the world are rapidly growing. This growth of databases has far outpaced the human ability to interpret this data creating new phenomena of big data. Big data is difficult to work using most existing tools, automatic methods and visualization. Some new methods called Visual data Mining have recently appeared trying to involve more significantly the user in the data mining process and using more intensively the visualization. We think that is important to consider user perception to drawback dimension selection process, select initial seed of clustering algorithm, and select interactively the clusters for example.

Inspired by these ideas, we propose a semi-interactive algorithm (Fig1 [1]) that we have developed, which integrate an automatic algorithm, an interactive evolutionary algorithm and visualization tools. This Approach can be applied on parsimonious clustering where clusters can be overlapping and we can detect interactively the relation between overlapped clusters ([2] Fig 2).

We propose also an interactive clustering approach, where we can identify visually clusters by combining different projections of interactive visualizations. In many situations where visual perception is more effective than classical clustering methods, the proposed approach gives better results ([3] Fig 3). We can apply these techniques combined with a new iterative clustering approach that extract compact clusters one by one, in this aspect the visualization is very important, the user can stop or continue the process according to the obtained information [4].

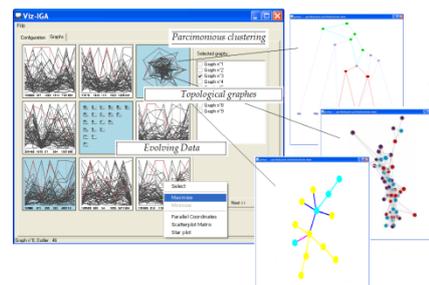These approaches can be applied on different problems on big data: data steam, social networks [5], …
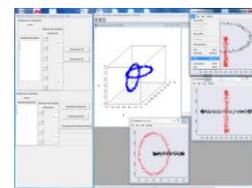
Fig-1

Fig-2

Fig-3

References

1. Lydia Boudjeloud-Assala, "Visual Interactive Evolutionary Algorithm for High Dimensional Outlier Detection and Data Clustering Problems ", International Journal of Bio-Inspired Computation, International Journal of Bio-Inspired Computation - Vol. 4, No.1 pp. 6 – 13, 2012.

2. F. Brucker, A. Gély : Parsimonious cluster systems, Advances in Data Analysis and Classification, Volume 3, Issue 3 (2009), Pages 189-204.

3. L. Boudjeloud-Assala, Kamel Chelghoum, "Approche interactive pour la classification non supervisée", 18èmes Rencontres de la Société Francophone de Classification, pp 67-70, septembre 2011, SFC'11, Orléans.

4. L. Boudjeloud-Assala, A. Blansché: Iterative Evolutionary Subspace Clustering. ICONIP (1) 2012: 424-431

5. Brieuc Conan-Guez, Hoai An Le Thi, Manh Cuong Nguyen and Tao Pham Dinh, Détection de communautés: une approche par programmation DC, Actes des XVIII rencontres de la SFC'11, Orléans, France. Septembre 2011.