Comprehensive Study and Analysis of Partitional Data Clustering Techniques

Aparna K., Department of Master of Computer Applications, B. M. S. Institute of Technology, Bangalore, India

Mydhili K. Nair, Department of Information Science and Engineering, M. S. Ramaiah Institute of Technology, Bangalore, India

ABSTRACT

Data clustering has found significant applications in various domains like bioinformatics, medical data, imaging, marketing study and crime analysis. There are several types of data clustering such as partitional, hierarchical, spectral, density-based, mixture-modeling to name a few. Among these, partitional clustering is well suited for most of the applications due to the less computational requirement. An analysis of various literatures available on partitional clustering will not only provide good knowledge, but will also lead to find the recent problems in partitional clustering domain. Accordingly, it is planned to do a comprehensive study with the literature of partitional data clustering techniques. In this paper, thirty three research articles have been taken for survey from the standard publishers from 2005 to 2013 under two different aspects namely the technical aspect and the application aspect. The technical aspect is further classified based on partitional clustering, constraint-based partitional clustering and evolutionary programming-based clustering techniques. Furthermore, an analysis is carried out, to find out the importance of the different approaches that can be adopted, so that any new development in partitional data clustering can be made easier to be carried out by researchers.

Keywords: Bisecting K-Means, Constraint-Based Partitional Clustering, Evolutionary Programming-Based Clustering, Fuzzy Clustering Algorithm, Partitional Clustering

1. INTRODUCTION

Clustering algorithms are based upon the index of similarity or dissimilarity between pairs of data points (Isa, Salamah, & Ngah, 2009). Traditionally, clustering is viewed as an unsupervised learning method which groups the data objects based on the information presented in the dataset without any external label information (Zhu, Wang, & Li, 2010). Clustering methods are classified into Partitioning approaches, Hierarchical methods, Density-based algorithms, Probabilistic techniques, Graph theoretic, Grid based algorithms, Model-based approaches, Genetic algorithms, Fuzzy methods, Rough set methods, etc.

DOI: 10.4018/ijban.2015010102
Many partitional clustering algorithms that have been introduced in recent years are based on evolutionary algorithms, such as Genetic Algorithms (GA), which are stochastic search heuristics (randomized search methods) inspired by Darwinian evolution and genetics. Partitional clustering algorithms determine a grouping solution by maximizing other similarities among objects within the same groups while minimizing the dissimilarities between the different groups. The K-Means algorithm is one of the most popular partitional clustering algorithms. The advantages of this class of algorithms are that they are relatively very easy to implement and efficient, due to its linear time complexity. But they often terminate at a local optimum which is usually seen as one of the drawbacks. The global optimum may be found by using techniques such as deterministic annealing and Genetic Algorithms. Another drawback is that it cannot deal well with non-spherical shaped clusters (Krink & Paterlini, 2006). Based on the strong positive empirical results that have been reported, the opinion of the research community is that constraints help to improve clustering performance with respect to accuracy, as measured on the set of extrinsic labels used to generate the constraints (Davidson, Wagstaff, & Basu, 2006).

2. LITERATURE SURVEY OF DIFFERENT PARTITIONAL DATA CLUSTERING ALGORITHMS

The survey of partitional data clustering algorithms is conducted with respect to two different aspects – the technical aspect and the application aspect as shown in Figure 1 in the Appendix. The survey proceeds further as follows:

2.1. Technical Aspect

- K-Means Algorithm
- Other Partitional Clustering algorithms
- Constraint based Clustering
- General Clustering Algorithm through Evolutionary Programming

2.2. Application Aspect

- Applications based on Constraint based Clustering
- Application oriented Clustering algorithm through Evolutionary Programming
Related Content

Advance Information Sharing in Supply Chains
Qiannong Gu, Xiuli He and Satyajit Saravane (2014). Encyclopedia of Business Analytics and Optimization (pp. 46-56).
www.igi-global.com/chapter/advance-information-sharing-in-supply-chains/107213?camid=4v1a

What the Future Holds for Data Mining
www.igi-global.com/chapter/future-holds-data-mining/7510?camid=4v1a

Measuring Effectiveness: A DEA Approach Under Predetermined Targets
www.igi-global.com/article/measuring-effectiveness/107067?camid=4v1a
Agile Development in Data Warehousing
www.igi-global.com/article/agile-development-data-warehousing/55589?camid=4v1a