A survey for the essential Data Mining technique: Clustering

Isabel B. Amaro[1], **Mirella M. Moro**[2], **Clodoveu Davis**[3]

[1]Department of Computer Science – Federal University of Minas Gerais (UFMG)
Belo Horizonte, Brazil

{isabel.amaro,mirella,clodoveu}@dcc.ufmg.br

Abstract. *Clustering is a Data Mining technique capable of group data with some non trivial similarity. Because of the advance of technology and data generation, clustering algorithms had to be developed and improved to process and extract useful information from the current large amount of data. Therefore, many areas that use Data Mining are constantly publishing articles with new clustering techniques and algorithms. This survey will study some state-of-the-art clustering algorithms, summarize, organize and present the temporal analysis in order to support the studies of its readers.*

1. Introduction

Clustering is a Data Mining technique capable of group data with some non trivial similarity. Because of the advance of technology and data generation, clustering algorithms had to be developed and improved to process and extract useful information from the current large amount of data. Therefore, many areas that use Data Mining are constantly publishing articles with new clustering techniques and algorithms. This survey will study some state-of-the-art clustering algorithms, summarize, organize and present the temporal analysis in order to support the studies of its readers.

2. Clustering

3. Data Mining

4. Algorithms

4.1. K-Means

4.2. Hierarchical clustering

4.3. DBSCAN

5. Temporal analysis

6. Conclusion