Clustering: The essential technique for Data Mining

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 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. Text Mining, which is a subarea of Data Mining, for example, may help to synthesize and associate terms of knowledge of a biomedical data base, or extract useful business information from social media data. This article will study some state-of-the-art algorithms used for clustering text data, summarize and present the temporal analysis in order to support the studies of its readers.*

1. Introduction

2. Clustering

3. Data Mining

4. Algorithms

4.1. K-Means

4.2. Hierarchical clustering

4.3. DBSCAN

5. Temporal analysis

6. Conclusion

References

Boulic, R. and Renault, O. (1991). 3d hierarchies for animation. In Magnenat-Thalmann, N. and Thalmann, D., editors, *New Trends in Animation and Visualization*. John Wiley & Sons ltd.

Knuth, D. E. (1984). *The TEX Book*. Addison-Wesley, 15th edition.

Smith, A. and Jones, B. (1999). On the complexity of computing. In Smith-Jones, A. B., editor, *Advances in Computer Science*, pages 555–566. Publishing Press.