International Congress on Science and Engineering
HAMBURG – GERMANY
March 2018

Application of hierarchical clustering algorithms for clustering of macro data

Mohammad reza Assadpour[1], Ali Asghar safaei[2], Mehdi Hossein Zadeh[3]

1-Department of Computer, Qeshm international Branch, Islamic Azad University, qeshm ,Ira Email:M.Assadpour67@yahoo.com

2-Department of Medical Informatics, Tarbiat Modares University, Tehran-Iran Email:aa.safaei@modares.ac.ir

3- Department of Computer, Islamic Azad University Science and Research Branch, Tehran-Iran Emai:mehdi.hoza@gmail.com

Abstract

As you know, medical data is one of the data that should be stored with high speed and accuracy. Since all people have a medical history, the volume of these data is extremely high and management requires the use of appropriate and efficient methods. However, this huge amount of data can really be useful for people and corporations, but also problematic. The problem with this progress is the analysis and analysis of large data. Using data mining techniques, you can extract useful information and hidden relationships between data. The traditional methods of data mining, due to their low speed, cannot directly run on large data, and we must look for a solution that we can analyze with large data. In this paper, the clustering of large medical data has been investigated using a hierarchical clustering algorithm and the results have been compared with some of the methods available in this field. The results show that the proposed method of this paper can cluster with greater accuracy, lower execution time and higher data rates.

Key words: Macro data, clustering, clustering algorithm, large data volume, data mining

1. Introduction

In the current digital age, the pace of change in volume and diversity of data has been significantly different from previous decades. Considering the great progress and development of the Internet and global online technologies, such as powerful and large data servers, we face a huge amount of data and data from different sources and services every day, which did not exist in the past decades [(Dong and Sirostava[1], 2013) and (Tang, Kang[2], 2013)]. However, this huge amount of data can really be useful for people and corporations, but also problematic. The problem with this progress is the analysis and analysis of large data. Using data mining techniques, you can extract useful information and hidden relationships between data. The traditional methods of data mining, due to their low speed, cannot directly run on large data, and we need to look for a solution that we can analyze with them large data (lighter, bigger lady, wow, Harran[3], 2014). In this paper, the clustering of large medical data has been investigated using a hierarchical clustering algorithm and the results have been compared with some of the methods available in this field. The results show that the proposed method of this paper can cluster with greater accuracy, lower execution time and higher data rates.

[1] X. L. Dong and D. Srivastava
[2] H. Tong and U. Kang
[3] 3 A. S. Shirkhorshidi, S. Aghabozorgi ,T. Y. Wah, and T. Herawan