

Implementing Agglomerative hierarchical clustering using multiple attribute

¹R.B Ramadevi (PG student) , ²H. Prabha (Assistant professor)
^{1,2}Department of computer science and engineering

ABSTRACT: Agglomerative hierarchical clustering algorithm used with top down approach. It implement with multiple attributes. In multiple attributes frequency calculation is allocated. Memory requirements are less in this process. Hierarchical clustering produce accurate result than any other algorithm. This is very less time consuming process.

Keywords: Hierarchical clustering, multiple attributes

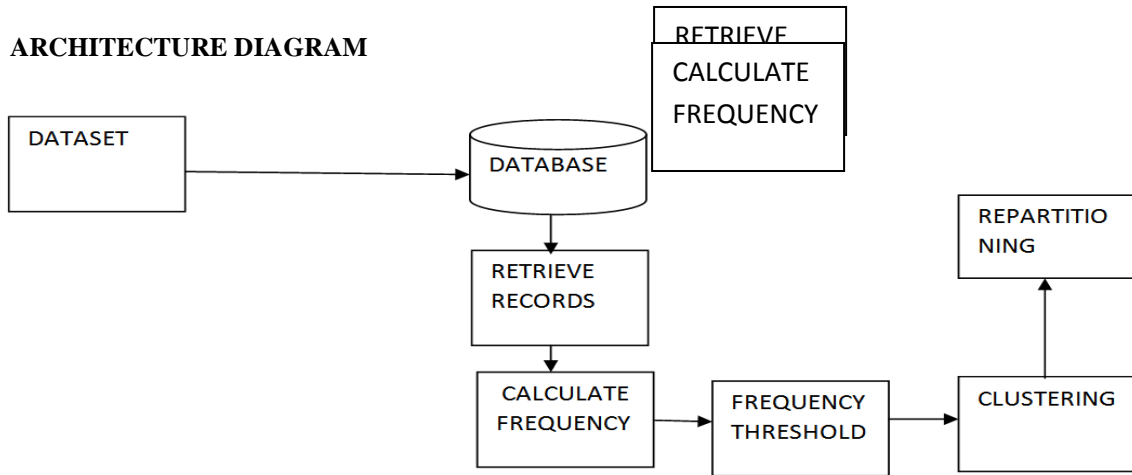
I. INTRODUCTION

Data mining is the process of mining of data. Data mining used over in several fields such as medical field, engineering field, in space research it is implemented. Data mining developed in tremendous manner. In data mining Field such privacy is provided. It denote a vector space model. Vector space model represent vector numbers .Using vector numbers spatial representation are provided. In this vector numbers cosine similarity distance are calculated. The main function of data mining is information retrieval, relevancy rankings. It mainly enhances its function in genetic engineering. In genetic engineering calculation of genes are processed, Laterally all discovery of similarity facts are existed. Using this similarity facts several data mining analyzing algorithms are used. Using this new technique several important high dimensional problems can be solved. Easy retrieval of data's are included. In this retrieval of data's fast retrieval techniques are used. Several calculations also processed over in this algorithm. In data mining several important identifications also provided. some of them are classifications and predictions. In classifications stepwise classifications are provided. In this approach several important ranking and analysis are used. In this several other mining techniques are also provided. Term of occurrence is also calculated. On this basis frequency subdivisions are provided. Mainly algorithm used over in this part Is agglomerative hierarchical clustering algorithm. In this mainly arrangement of lowest to highest rankings are provided. In this rankings special applications are provided. In this application several important activities are provided. High dimensional complexity is reduced in this part. Computational parts are extended over in this phase. Using this computerized can demolish their activities in efficient manner. Various other important specifications are provided. Certain enormous representations are provided over in this process. Using this term of occurrence, we can predict the accurate result. In the proposed term this algorithm is used for certain calculation. In this certain outlier problems are also removed. Frequency threshold calculations are also included. Threshold mainly includes the allocation of objects. Several important facts and rules are established in this algorithm to produce the computational result efficiently.

II. RELATED WORK AND REVIEWS

In this related work certain advanced technology is found out for clearance of certain facts. Data's are selected for finalize of the reviews. These data's contain certain field properties. These attributes are selected for the mining of data's. Terms and frequencies are characterized on this way. Several Terms and conditions are provided in this process. In this terms and conditions certain rules are applied. Frequency selection is the main application in the multiple attribute usage. In data, we have to find the frequency of specific attribute. The frequency is based on occurrence of each entry in attributes. After calculating the frequency compare the similarity between frequencies of entity. Next further process the initial clustering process based on frequency. We have to perform the initial clustering based frequency similarity between entities. First of all we have to find the frequency threshold value for all entities. We have to perform the initial partition based on the clustering. We are using the hierarchical clustering for removing the repartitioning the outliers in the initial partitioning. We have to perform of initial clustering and produce the subsets based on hierarchical clustering. In this module we are getting the best fit cluster. We perform the repartition in top down manner. In this clustering, we removing the outliers in final clustering result.

ARCHITECTURE DIAGRAM



NAME	SURNAME	COMPANY NAME	GENDER	AGE
PRANEETHA	NONE	INFOSYS	FEMALE	24
RATHIKA	DEVI	HCL	FEMALE	25
SARANYA	NONE	TCS	FEMALE	25

TABLE 1.1

III. RESULTS AND CONCLUSION

Thus the all attributes are formalized according with their frequencies these frequencies are arranged in the linear manner. This algorithm provide more accuracy than K-mean algorithm. Time complexity is reduced over in this process. In this level 1 , level 2 clustering are provided for clustering process. Clustering accuracy is improved in efficient way. In future this method tempt to be a important method for further reference.

IV. ACKNOWLEDGEMENT

This work is submitted under the guidance of my guide, and with the help of all my colleagues. My parents help me a lot to do this process. Thanking you. This project is submitted for the further enhancement of the knowledge of the readers.

REFERENCES

- [1]. Iterative visualization and hierarchical neural projections for data mining, IEEE transactions ,vol 2 may 2000.
- [2]. Hierarchical clustering for large scale short conversation based on domain ontology 2008 international symposium on computer science and computational technology
- [3]. A hierarchical approach to represent relational data applied to clustering tasks, proceedings of international joint conference on neural networks, sanjose California.
- [4]. Analysis of document using sample agglomerative technique, omar h karam, Ahmed m.hamad