Privacy Preserving Database and Classification of Multidatabase Mining

1Keerthana G., 2Sandhya P. and 3Varun R.

1PG student, Computer science department, Aksheyaa College of Engineering, Kancheepuram, India
2Assistant Professor, Electronics and Communication Engineering, Aksheyaa College of Engineering, Kancheepuram, India.
3Assistant Professor, Electronics and Communication Engineering, Aksheyaa College of Engineering, Kancheepuram, India.

ABSTRACT

Big Data is nothing but collection of large data set value. Mainly used in biomedical and biological science. Database mining helps to protect the data from unauthorized parties. K-MEANS algorithm reconstructed with not only clustering the data but also fixing the size of clusters. This in turn help in time consumption, avoids deadlocks, ignores network traffics, can store and retrieve huge data.

INTRODUCTION

1] works only in an offline fashion and incapable of handling the bigdata scenario in real time due to less storage, impossible multitasking, huge data not possible and more expensive. proposed system build a bigdata analytic framework for fast response and real time decision making.

1. Research Background:

With the help of K-means algorithm user clustered the data, but so far size of the cluster haven’t fixed. Fixing the cluster size results in less processing time is the major advantage and it is user friendly. This is not limited to limited user finding but also finding it for huge data collection and can also be implemented in all kind of environments such as commercial, social networks, etc., fig.1 shows the general architecture for fixing the cluster size.

Fig. 1: Architecture for fixing the cluster size.

2. K-Means Algorithm:

Once the data is collected which is supposed to be shared with few more system hence distributed processing is carried out for storing and retrieving the information without use of decentralized database environment is the unique factor. Once it is done deadlock factor has to be analysed using complex relationship. This builds relationship between multiuser and multiple server throughout the network.

Based on the requirement, the data can be separated with various factors based on the environment it is used. Factors with similarity comes under the homogeneity and different comes under heterogeneity and thus results in the huge data collection and calculates the size of the cluster and big data memory was evolved.

Results And Analysis:

Using Hadoop, cluster size has been fixed and huge data are stored.
Fig. 2: modules involved in algorithm.

**Conclusion:**
K-MEAN algorithm helped to store huge data and fixed the cluster size thus helps in time consumption, user friendly, error free process, limited bandwidth and easier way to extract the information in online is the special advantage.

**REFERENCES**


