Compare the effectivenss and efficiency of the algorithms

Short discussion about the reasons why those datasets were choosen.

Description about the algorithms . detailsabout the R package used.it should be very succinct .example: the R package XXX implements the algorithm in paper YYY with the following changesZZZ

Analysis of the result.It should contain effectiveness and efficiency

Format :

Margins at each of the top,bottom,left and right sides is 1.0 inch

Font size is 10pt. font is times new roman

Single line space

For testing accuracy and efficiency of simple k-means and hierarchical clustering algorithm, various datasets with known clustering available at UCI repository of machine learning databases[10].This paper uses Iris[10] and Diabetes[10] datasets and a brief description of datasets used in experiment evaluation Table 1 shows some characteristics of datasets as the test datasets.

TABLE 1. DESCRIPTION OF DATASETS.

Datasets

Number of Attribute

Number of records/Instances IRIS 05 150 DIABETES 09 768 In this paper, we have used weka data mining tool version 3.7 for testing accuracy and running time of simple K- means and Hierarchical clustering algorithm on given datasets. The clustering results for cluster k=3 is shown in Table 2.

TABLE 2.CLUSTERING RESULTS FOR DATASETS.

Datasets

k-means running time(sec)

Hierarchical clustering running time

k-means Accuracy %

Hierarchical clustering Accuracy % IRIS 0.03 0.17 88.667 66 DIABETES 0.06 2.14 51.6927 65.1042

The results for accuracy and running time are shown in Figure 3 and Figure 4.

iris diabetes

0

10

20

30

40

50

60

70

80

90

datasets

accuracy(%)

k-means clustering hierarichal clustering

Figure 3. Accuracy v/s datasets

iris diabetes

0

0.5

1

1.5

2

2.5

datasets

time(Sec)

k-means clustering hierarical clustering

Figure 4. Running time v/s datasets IV.CONCLUSION K-means is a typical clustering algorithm and it is widely used for clustering large sets of data. From experiments, we can conclude that the accuracy of k-means for iris dataset having “real” attributes is much than the hierarchical clustering and for diabetes dataset having “integer, real” attributes accuracy of hierarchal clustering is more than the k-means algorithm (fig 3). Though the time taken to cluster the data sets is less in case of k-means ( fig 4). Hierarchical clustering results are usually presented in dendrogram, the dendrogram for iris dataset is shown in (fig 2).A good clustering method produces high-quality clusters to ensure that the inter-cluster similarity is low and the intra-cluster similarity is high, in other words, members of a cluster are more like each other than they are like members of a different cluster. As we have discussed in this paper k- means algorithm is good for large datasets.