CHAPTER 5

CONCLUSION AND FUTURE ENHANCEMENT

In this research a dynamic density-based clustering algorithm was successfully implemented and evaluated. The normal Dynamic DBSCAN clustering algorithm was able to insert a point dynamically and re-estimate the cluster IDs. The Fast Dynamic DBSCAN algorithm was able to insert a set of data objects in bulk and re-estimate the cluster IDs of all the existing and new points. This algorithm is capable of creating, modifying and inserting clusters over time and can do it in a very fast manner. The performance of the algorithm tested with Wine Dataset, TIC2000 dataset, Zoo Dataset, Wisconsin cancer dataset as well as synthetic data set. The Dunn Index, Generalized Dunn Index (GDI), Davies-Bouldin Index (DBI) as the cluster validation metrics and the CPU time as a metric to measure the clustering speed of the algorithm. As compared with the results in the previous section, the proposed Fast Dynamic DBSCAN algorithm performed significantly well in terms of clustering accuracy as well as speed.

While comparing the performance of the normal Dynamic DBSCAN algorithm with Fast Dynamic DBSCAN algorithm with synthetic data, a peculiar relationship of the algorithm is observed with the input order of the data. If the data is given in random order, then the performance of the normal dynamic DBSCAN algorithm gets reduced. But amazingly the performance of the fast dynamic DBSCAN algorithm was more stable and in fact, gets very much improving in terms of time if the data is given in random order. Further the performance of the previously proposed normal Dynamic DBSCAN algorithm was very much affected by the size of the input dataset. But
proposed fast dynamic DBSCAN was not much affected with the size of the dataset.

This work addresses the problem of clustering incremental data set in which only a bulk data is added over time. The future work may address all the other possibilities of dynamic operations like bulk deletions and bulk modifications of data points and remodel the algorithm to cluster the data during this dynamically changing dataset. Future works may explore these possibilities and address hybrid and more efficient dynamic clustering algorithm.