Abstract

Clustering is one of the major techniques in data mining. Clustering data streams have drawn attentions in the past few years because of their ever-growing presence. Data streams add more challenges to clustering such as limited time, limited memory and one pass clustering. Further, discovering clusters with arbitrary shapes is important in data stream applications.

Now a few clustering techniques for data streams exist in multidimensional spaces and the technique of "clustering projected or subspace" is used. Therefore, the task of projected clustering (or subspace clustering) has to be defined.

PreDeConStream is a density-based data stream clustering algorithm for clustering high-dimensional data streams. In this paper, PPreDeConStream is present as a parallel version of PreDeConStream algorithm in the shared memory model. The theoretical and experimental results show that PPreDeConStream offers nearly linear speedup while keeps other advantages of PreDeConStream.
References


Index Terms

Computer Science

Algorithms
Keywords

Clustering data stream algorithms, parallel algorithms, microcluster, density-based clustering, shared memory model.