Abstract

Clustering high dimensional data has been a big issue for clustering algorithms because of the intrinsic sparsity of the data points. Several recent research results signifies that in case of high dimensional data, even the notion of proximity or clustering possibly will not be significant. K-Means is one of the basic clustering algorithm which is commonly used in several applications, but it is not possible to discover subspace clusters. The subspaces are explicit to the clusters themselves. In this paper, an algorithm called Modified Projected K-Means Clustering Algorithm with Effective Distance Measure is designed to generalize K-Means algorithm with the objective of managing the high dimensional data. The experimental results confirm that the proposed algorithm is an efficient algorithm with better clustering accuracy and very less execution time than the Standard K-Means and General K-Means algorithms.

References

A Modified Projected K-Means Clustering Algorithm with Effective Distance Measure

- Charu C. Aggarwal, J. Han, J. Wang, and S. Philip Yu, "A Framework for Projected Clustering of High Dimensional Data Streams", Proc. Very Large Data Base (VLDB '04), Pp. 852-863, 2004.
- Ada Wai-chee Fu and E. Ng Ka Ka, "Efficient algorithm for projected clustering", Proceedings. 18th International Conference on Data Engineering, DOI: 10.1109/ICDE.2002.994727, 2002.

Index Terms

Computer Science
Information Sciences
Keywords
Data Mining  Projected Clustering  K-means  High Dimensional Data  General K-means  Efficient Projected Clustering (epc)