Abstract

The boosted demand for immense information, the enhanced data acquisition and so do the size and number of dimensions of data is a big challenge for the data mining algorithms. Clustering exercise to collect the data with same characteristics together, for better performance of knowledge based systems. High dimensional and large size data results in declined performance of existing clustering algorithms. PROCLUS is an efficient high dimensional clustering algorithm; consist of significant issues like inconsistency in results and expert supervised subspaces. MPROCLUS: a modified PROCLUS algorithm is proposed, aimed at improving the running time and consistency as well as the unsupervised selection of the parameter like, average number of dimensions. The promising and consistent results of MPROCLUS has open the sky wide open for further research for usage of MPROCLUS in stream Data Mining.

References
Towards Unsupervised and Consistent High Dimensional Data Clustering

- Aggarwal, C. C., Joel, L. W., Philip, S. Yu, Cecilia, P., and Jong, S. P., "Fast algorithms for projected clustering." A CM SIGMOD international conference on Management of data (May 1999), 28(2), 61-72
- Hans-Peter, K., Peer K., and Arthur, Z., "Clustering high dimensional data: A survey on subspace clustering, pattern based clustering, and correlation clustering," ACM Transactions on Knowledge Discovery from Data (April 2009), 3(1)
- Hall, M., A., and Holmes, G. "Benchmarking attribute selection techniques for discrete class data mining," IEEE Transactions on Knowledge and Data Engineering (Nov 2003), 15(6), 1437-1447
- Kevin, Y. and David W., "Harp: A practical projected clustering algorithm." IEEE Transactions on Knowledge and Data Engineering (Nov 2004), 16(11), 1387-1397

Index Terms

Computer Science

Information Sciences

Keywords

High dimensional clustering  Unsupervised and consistent clustering  PROCLUS