Abstract

The Web has enormous, various and knowledgeable data for data mining research. Clustering web usage data is useful to discover interesting patterns pertaining to user traversals, behaviour and their usage characteristics. Moreover, users accesses web pages in an order in which they are interested and hence incorporating sequence nature of their usage is crucial for clustering web transactions. In this paper we present OPTICS (“Ordering Points To Identify the Clustering Structure”) algorithm to find density based clusters on a web usage data on MSNBC.COM website which is a free news data website with so different categories of news. The clusters are generated by OPTICS algorithm. The average of inter cluster and intra cluster are Calculated. the results are compared with different similarity measures like Euclidean, Jaccard, projected Euclidean, cosine and fuzzy similarity. Finally showed behavior of clusters that made by OPTICS algorithm on a sequential data in a web usage domain. we
performed a variety of experiments in the context of density based clustering, quantify our results by the way of explanations and list conclusions.

Reference

- “Deepak P, Shourya Roy” IBM India Research Lab, “OPTICS on Text Data: Experiments and Test Results”.
- “Dimitris K. Tassoulis, Gordon Ross, and Niall M. Adams “Department of Mathematics Imperial College London, “Visualising the Cluster Structure of Data Streams”.
- “Markus M. Breunig, Hans-Peter Kriegel, Jörg Sander”, “Fast Hierarchical Clustering Based on Compressed Data and OPTICS” Proc. 4th European Conf. on Principles and Practice of Knowledge Discovery in Databases (PKDD 2000), Lyon, France.
- Martin Ester, Hans-Peter Kriegel, Jorg Sander, Xiaowei Xu (1996). “A density-based algorithm for discovering clusters in large spatial databases with noise”. In Evangelos Simoudis, Jiawei Han, Usama M. Fayyad. Proc. 2 International Conference on Knowledge Discovery and Data Mining (KDD-96). pp.226-231.

Index Terms

Computer Science Knowledge Management

Key words

Clustering algorithm OPTICS Ordering Points To Identify the
Clustering Structure
Sequence mining

Average Inter cluster
Intra cluster