Abstract

Clustering is an unsupervised learning technique which aims at grouping a set of objects into clusters so that objects in the same clusters should be similar as possible, whereas objects in one cluster should be as dissimilar as possible from objects in other clusters. Cluster analysis aims to group a collection of patterns into clusters based on similarity. A typical clustering technique uses a similarity function for comparing various data items. This paper covers the survey of various clustering techniques, the current similarity measures based on distance based clustering, explains the limitations associated with the existing clustering techniques and propose that the combination of the advantages of the existing systems can help overcome the limitations of the existing systems.

References


3. Haixun Wang, Wei Wang, Jiong Yang, Philip S. Yu, Clustering by Pattern Similarity in Large Data Sets, Proceeding SIGMOD '02 Proceedings of the 2002 ACM SIGMOD international conference on Management of data, Pages 394-405, ACM.


17. Xianchao Zhang, Xiaotong Zhang, Han Liu, Multi-Task Multi-View Clustering for Non-Negative Data, Proceedings of the Twenty-Fourth International Joint Conference on Articial
Intelligence, IJCAI 2015.


27. Alina Ene, Sungjin Im, Benjamin Moseley, Fast clustering using MapReduce, Proceeding KDD 2011 Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, Pages 681-689, ACM.


30. Hwanjo Yu, Duane Searsmith, Xiaolei Li, Jiawei Han, Scalable Construction of Topic Directory with Nonparametric Closed Termset Mining, Data Mining, 2004. ICDM '04. Fourth IEEE International Conference, Pages 563-566.


Clustering Techniques and the Similarity Measures used in Clustering: A Survey

pp. 69-84.

34. Fast Distance Metric Based Data Mining Techniques Using P-trees: k-Nearest-Neighbor classification and k-Clustering: A Thesis Submitted to the Graduate Faculty Of the North Dakota State University.


Index Terms

Computer Science Networks

Keywords

pattern based similarity, negative data clustering, similarity measures.