Abstract

Outlier is a data point that deviates too much from the rest of dataset. Most of real-world dataset have outlier. Outlier analysis is one of the techniques in data mining whose task is to discover the data which have an exceptional behavior compare to remaining dataset. Outlier detection plays an important role in data mining field. Outlier Detection is useful in many fields like Medical, Network intrusion detection, Credit card fraud detection, medical, fault diagnosis in machines, etc. In order to deal with outlier, clustering method is used. Outlier detection contains clustering and finding outlier by applying any outlier detection technique. For that K-mean is widely used to cluster the dataset. Different techniques like statistical-based, distance-based, and deviation-based and density based methods are used to detect outlier. The experiment result shows that existing algorithm perform better than proposed cluster-based and distance-based Algorithm.

References
Outlier Detection in Dataset using Hybrid Approach

- Jingke Xi, "Outlier Detection Algorithm in Data Mining", Second International Symposium on Intelligent Information Technology Application, 2008 IEEE, pp. 94-97

Index Terms

Computer Science Databases

Keywords

Data Mining Outlier Clustering Approach k-mean Algorithm Distance Based Approach