Abstract

Outlier detection is an extremely important problem with direct application in a wide variety of domains. A key challenge with outlier detection is that it is not a well-formulated problem like clustering. In this paper, discussion on different techniques and then comparison by analyzing their different aspects, essentially, time complexity. Every unique problem formulation entails a different approach, resulting in a huge literature on outlier detection techniques. Several techniques have been proposed to target a particular application domain. The classification of outlier detection techniques based on the applied knowledge discipline provides an idea of the research done by different communities and also highlights the unexplored research avenues for the outlier detection problem. Discussed of the behavior of different techniques will be done, in this paper, with respect to the nature. The feasibility of a technique in a particular problem setting also depends on other constraints. For example, Statistical techniques assume knowledge about the underlying distribution characteristics of the data. Distance based techniques are typically expensive and hence are not applied in scenarios where computational complexity is an important issue.

References
- Markus M. Breunig, Hans-peter Kriege, Raymond T. Ng, Jorg Sander. "LOF: Identifying Density-Based Local Outlier.";
- Hinneburg, C. C. Aggarwal, and D. A. Keim. "What is the Nearest Neighbor in High Dimensional Spaces." In Proc. 26th Int. Conf. on Very Large Databases (VLDB 2000), Cairo, Egypt, 2000.
- C. B. D. Newman and C. Merz. UCI repository of machine learning databases.
- Han and Kamber (2007), Data Mining: Concepts and Techniques Morgan Kaufmann publications
- George Marakas, Data Warehousing, Data Mining and Visualisation, Pearson publications

**Index Terms**

Computer Science  Algorithms

**Keywords**
Outlier  time complexity  statistical techniques  eucledian distance