Abstract

Outlier is defined as an observation that deviates extensively from other observations. The identification of outliers can lead to the discovery of useful and meaningful knowledge. Outlier detection has been widely studied in the past decades. Most refined methods in data mining address this issue to some extent, but not fully, and can be improved by addressing the problem more directly. The detection of outliers can lead to the invention of unpredicted facts in areas such as credit card fraud detection, calling card fraud detection, discovering criminal behaviors, discovering network intrusions, etc. This paper mainly discusses and compares approach of different outlier detection from data mining perspective, which can be grouped into distance-based approach and density-based approach.

References


Index Terms

Computer Science Information Sciences

Keywords

Outlier detection, Distance-based approach, Density-based approach