

Abstract

Outlier is data which have different characteristic when compared with the large amount of data. Outliers often contain unexpected knowledge. Because of that in many Knowledge Discovery, finding outliers is more interesting than finding inlier in dataset. Outlier Detection is one of data mining's functionalities that aims to find outlier in dataset. There many methods to detect outlier, but most of them faced the problems of handling the scalability of dataset. Scalability problem had caused the using of distances of points inappropriate to discover outliers in high dimensional. RSetOF (Rough Set Outlier Factor) is a method o detecting outlier in high dimensional dataset based on Non-Reduct from Rough Set approach. A RsetOF value calculated for each data based on rules from Non-Reduct, whether outlier data or not. RsetOF can detect outliers with relatively good accuracy in some test scenarios based on measurement parameters RsetOF value, top n outlier and parameter evaluation of detection rate and false positive rate.

Keywords: *outlier, RSetOF, data mining, scalability*