Abstract

The task of outlier detection is to find the small groups of data objects that are exceptional to the inherent behavior of the rest of the data. Detection of such outliers is fundamental to a variety of database and analytic tasks such as fraud detection and customer migration. There are several approaches [10] of outlier detection employed in many study areas amongst which distance based and density based outlier detection techniques have gathered most attention of researchers. In information theory, entropy is a core concept that measures uncertainty about a stochastic event, and it means that entropy describes the distribution of an event. Because of its ability to describe the distribution of data, entropy has been applied in clustering applications in data mining. In this project work, we have developed a robust supervised outlier detection algorithm using hybrid approach (RODHA) which incorporates both the concept of distance and density along with entropy measure while determining an outlier. We have provided an empirical study of different existing outlier detection algorithms and established the effectiveness of the proposed RODHA in comparison to other outlier detection algorithms.