

Anomalous score and gamma-anomalous detection algorithm

Sooppawat Thipyarat and Krung Sinapiromsaran

Department of Mathematics and Computer Science, Faculty of Science, Chulalongkorn University,
Tel: 092-2578077, e-mail: T.sooppawat@hotmail.com

Abstract: Outlier analysis is one of the popular data mining topics in Computer Science which involves the automatic extraction of hidden knowledge within data. Various algorithms dealing with this problem can be categorized into two types which are scoring and labeling. This project attacks both types by developing two algorithms for each type. The first algorithm assigns a score to a point using the anomalous score or the largest radius to a nearest point. The second algorithm labels an anomalous point using the ratio of the mean with the parameter γ called the gamma-anomalous algorithm. The γ times its anomalous score that is bigger than the grand mean will be identified as an outlier. In this project, we define the anomalous score of all points in a finite dimensional real-value dataset and prove some properties about the anomalous score. It also includes the implementation of the gamma-anomalous algorithm on the cloud to detect outliers within a dataset.

Finally, we evaluate our algorithm on 3000 synthetic datasets and determine the appropriate value of the γ parameter for a typical dataset.