Abstract

Anomaly is a data point that does not follow to the normal points describing the data. Anomaly detection has vital applications in data cleaning and moreover in the mining of anomalous points for fraud recognition, intrusion detection, stock market analysis, customized marketing, system sensors, and email spam detection. Discovering anomalous points among the data points is the fundamental thought to find out an anomaly. Anomaly detection signals out the objects mostly deviating from a particular data set. The majorities of the anomaly detection techniques are normally implemented in batch mode, and so cannot be essentially reached out to large-scale problems without giving up the reckoning and memory requirements. To address this problem, existing online anomaly detection technique with oversampling and Principal Component Analysis (PCA), aim at identifying presence of anomalies from a big amount of data via an online updating technique. The oversampling strategy is used to duplicate the target instance and balance the data set and the PCA is used as a renowned unsupervised dimension reduction method, which determines the principal directions of the data distributions. Inspite of the duplication, the computation and memory requirement is reduced using an online updation technique. Thus present framework is ideal for online applications which have reckoning or memory restrictions.
References

- V. Chandola, A. Banerjee, and V. Kumar, 2009 Anomaly Detection: A Survey.
- F. Angiulli, S. Basta, and C. Pizzuti, 2006 Distance-Based Detection and Prediction of Outliers.
- Edwin M. Knox and Raymond T. Ng, 1998 Algorithms for Mining Distance-Based Outliers in Large Datasets.
- Jiawei Han, Micheline Kamber, Data Mining: Concepts and Techniques.

Index Terms

Computer Science

Information Sciences

Keywords

Anomaly detection online updation technique oversampling principal component analysis unsupervised dimension reduction.