

A Survey on Outlier Detection in Data Streams[1]

Aditi Singla (2014CS50277)
Ankush Phulia (2014CS50279)
Vaibhav Bhagee (2014CS50297)

1 Introduction

In data mining, the task of Outlier Detection is to identify those data points in a dataset which differ significantly from the rest of the points in the dataset and occur in rarity. There are various examples of scenarios where detecting outlier data samples is of use for eg. fraud detection in the financial sector, separation of “noisy” samples from the normal ones to refine the dataset for further processing, detection of “anomalous” event samples in system logs, to prevent large scale system failures etc.

Traditional algorithms have looked at outlier detection in scenarios where the dataset is “static”, i.e the data does not change over the period of time and the algorithm has the access to the entire dataset, in memory. However, as the size of the datasets grows large, it might not be always easy to get the entire dataset in memory, for processing. Moreover, there are many scenarios like prevention of system failure, where the data samples like logs, are generated temporally, in a continuous fashion. In such cases, the outlier detection algorithm can never have access to the complete dataset and the analysis for outliers needs to be performed over the “seen” data.

As a part of this survey, we present and discuss various algorithms and techniques, which have been proposed in context of detecting outliers in the setting of streaming data. In particular, we discuss the challenges which are posed when detecting outliers in data streams what approaches are followed to overcome these.

2 Motivation

Placeholder for Motivation

3 Problem Formulation

Placeholder for problem formulation

References

- [1] MANZOOR, E., LAMBA, H., AND AKOGLU, L. xstream: Outlier detection in feature-evolving data streams. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (New York, NY, USA, 2018), KDD '18, ACM, pp. 1963–1972.