Outlier Detection in Non-Stationary Data Streams

Priyanga Dilini Talagala
Department of Econometrics and Business Statistics, Monash University, Australia
and
Rob J Hyndman
Department of Econometrics and Business Statistics, Monash University, Australia
and
Kate Smith-Miles
School of Mathematics and Statistics, University of Melbourne, Australia

October 25, 2018

Abstract

This work develops a framework for detecting outlying series within a large collection of time series in the context of non-stationary streaming data. We define an outlier as an observation that is very unlikely given the recent distribution of a given system. In this work we make two fundamental contributions. First, we propose a framework that provides early detection of anomalous behaviour within a large collection of streaming time series data using extreme value theory. Second, we propose a novel approach for early detection of non-stationarity (also called "concept drift" in the machine learning literature.) The proposed algorithm uses time series features as inputs, and a density-based comparison to detect any significant change in the distribution of the features. Using various synthetic and real world datasets, we demonstrate the wide applicability and usefulness of our proposed framework. This framework is implemented in the open source R package *oddstream*. We show that the proposed algorithm can work well in the presence of noisy non-stationarity data within multiple classes of time series.

Keywords: Anomaly Detection; Extreme value theory; Time series features; Kernel-based density estimation; Nonstationary temporal data.