Detection of anomalies in large collections of time series

Rob J Hyndman
Monash University
Victoria, Australia
Rob.Hyndman@monash.edu

Earo Wang
Monash University
Victoria, Australia
yiru.wang@monash.edu

Nikolay Laptev
Yahoo Research
California, USA
nlaptev@yahoo-inc.com

ABSTRACT

It is becoming increasingly common for organizations to collect very large amounts of data over time, and to need to detect unusual or anomalous time series.

For example, Yahoo has banks of mail servers that are monitored over time. Many measurements on server performance are collected every hour for each of thousands of servers. We wish to identify servers that are behaving unusually.

We compute a vector of features on each time series, measuring characteristics of the series. For example, the features may include lag correlation, strength of seasonality, spectral entropy, etc. Then we use a robust principal component decomposition on the features, and use various bivariate outlier detection methods applied to the first two principal components. This enables the most unusual series, based on their feature vectors, to be identified. The bivariate outlier detection methods used are based on highest density regions and α-hulls. Data from Yahoo mail servers will be used to illustrate the proposed methods.

Keywords

Multivariate Anomaly Detection