A survey of stochastic anomaly detection approaches in time-series data.

Sanket S Mehrotra

Colorado State University

Author Note

This work was done to supplement Hajar Houmayouni’s PhD Dissertation.

Abstract

This document is a survey of widely used methods of time-series analysis. I have used some reference texts and tutorials regarding outlier detection in time series and read several papers about recent advances in the field to try to replicate these approaches on a common base dataset. The performance, advantages, and limitations of the approaches are documented and used to compare against the performance of a Bi-directional LSTM Autoencoder in anomaly detection on time-series data. We explore benchmarks to compare performance such as F1 score, MSE, MAE and ROC.

Keywords: Time-series, Outlier detection, stochastic methods, machine learning, clustering, anomaly detection, anomalous sequence detection.

Table of Contents

[Abstract 2](#_Toc50646868)

[A Survey of stochastic anomaly detection approaches in time-series data. 4](#_Toc50646869)

[Our Datasets 4](#_Toc50646870)

[Yahoo S5 DataSet 4](#_Toc50646871)

[Features. 5](#_Toc50646872)

[NASA Statlog Shuttle Dataset – UCI 5](#_Toc50646873)

[[Heading 3]. 6](#_Toc50646874)

[References 8](#_Toc50646875)

[Footnotes 9](#_Toc50646876)

[Tables 10](#_Toc50646877)

[Figures title: 11](#_Toc50646878)

A survey of stochastic anomaly detection approaches in time-series data.

Time series data is data which includes measures of a value of a variable(or variables) with respect to time. It is a combination of a modelling task and a classification task. We explore stochastic methods to model the data and apply simple criterion to act as classifiers(manual thresholds, k-standard deviations, etc).

## Our Datasets

We have used two popular datasets to apply these methods on, namely the Yahoo S5 Dataset, and the NASA Statlog Shuttle dataset. These are explored below:

## Yahoo S5 DataSet

Released Date: 2015

Designed for: Benchmarking Anomaly Detection

Size: 16 MB

5 folders of Time Series, both real and synthetic.

Folder 1: The real TS – Metrics of Yahoo services (Server Traffic)

Folders 2-4: Synthetic TS with varying anomaly types including outliers and change-points.

Folders 1-2 are Univariate containing only the ‘value’ feature.

Folders 3-4 are Multivariate, containing features such as ‘noise’,’trend’ and three different measures of seasonality: ‘seasonality1’, ‘seasonality2’, ‘seasonality3’

This dataset consists of 367 time series in total. Out of these, 67 are real and the others are synthetically generated to simulate various possible time-series.

### Features.

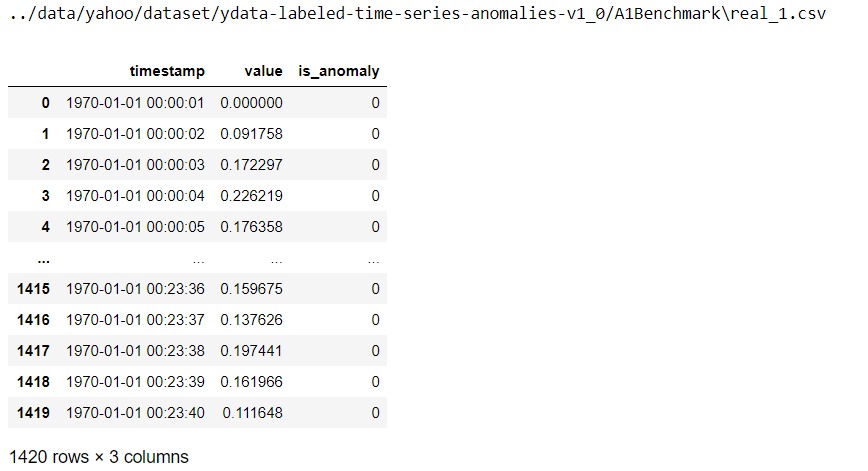
****

Figure 1: A sample of one of the real univariate time-series

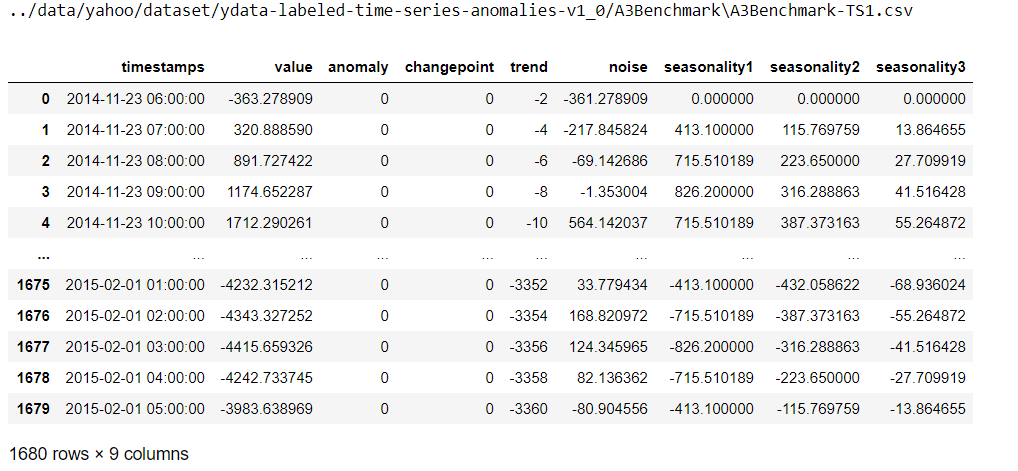
****

Figure 2: A sample of one of the synthetically generated time-series

## NASA Statlog Shuttle Dataset – UCI1

Released Date: 2013(to confirm)

Designed for : Supervised Anomaly detection.

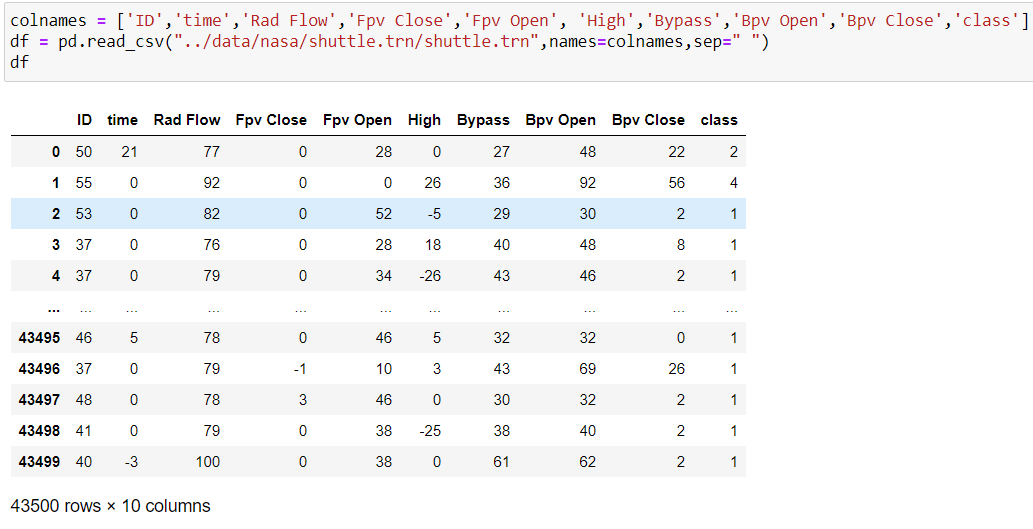
Size: 1.77 MB

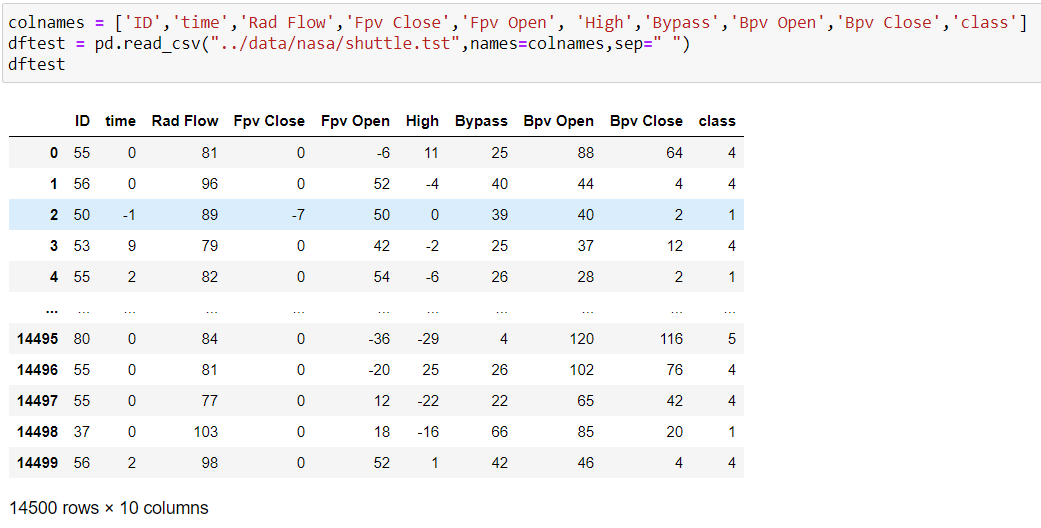
The shuttle dataset describes radiator positions in a NASA space shuttle with 9 attributes and was designed for supervised anomaly detection.

Besides the normal “radiator flow” class, about 20% of the original data describe abnormal situations.

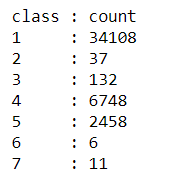
Training and test set are combined in a single big dataset, which has as a result 46,464 instances with 1.89% anomalies. (cite from Goldstein, Markus, and Seiichi Uchida. "A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data." PloS one 11.4 (2016): e0152173.)

### Features.

****

****

Here the class column is used as the outlier/not-outlier label, it’s range of values is as follows:



#### [Heading 4].

[When using headings, don’t skip levels. If you need a heading 3, 4, or 5 with no text following it before the next heading, just add a period at the end of the heading and then start a new paragraph for the subheading and its text.] (Last Name, Year)

##### [Heading 5].

[Like all sections of your paper, references start on their own page. The references page that follows is created using the Citations & Bibliography feature, available on the References tab. This feature includes a style option that formats your references for APA 6th Edition. You can also use this feature to add in-text citations that are linked to your source, such as those shown at the end of this paragraph and the preceding paragraph. To customize a citation, right-click it and then click Edit Citation.] (Last Name, Year)

References

Last Name, F. M. (Year). Article Title. *Journal Title*, Pages From - To.

Last Name, F. M. (Year). *Book Title.* City Name: Publisher Name.

Footnotes

1[Add footnotes, if any, on their own page following references. For APA formatting requirements, it’s easy to just type your own footnote references and notes. To format a footnote reference, select the number and then, on the Home tab, in the Styles gallery, click Footnote Reference. The body of a footnote, such as this example, uses the Normal text style. (Note: If you delete this sample footnote, don’t forget to delete its in-text reference as well. That’s at the end of the sample Heading 2 paragraph on the first page of body content in this template.)]

Tables

Table 1

[Table Title]

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Column Head | Column Head | Column Head | Column Head | Column Head |
| Row Head | 123 | 123 | 123 | 123 |
| Row Head | 456 | 456 | 456 | 456 |
| Row Head | 789 | 789 | 789 | 789 |
| Row Head | 123 | 123 | 123 | 123 |
| Row Head | 456 | 456 | 456 | 456 |
| Row Head | 789 | 789 | 789 | 789 |

Note: [Place all tables for your paper in a tables section, following references (and, if applicable, footnotes). Start a new page for each table, include a table number and table title for each, as shown on this page. All explanatory text appears in a table note that follows the table, such as this one. Use the Table/Figure style, available on the Home tab, in the Styles gallery, to get the spacing between table and note. Tables in APA format can use single or 1.5 line spacing. Include a heading for every row and column, even if the content seems obvious. A default table style has been setup for this template that fits APA guidelines. To insert a table, on the Insert tab, click Table.]

Figures title:

Figure 1. [Include all figures in their own section, following references (and footnotes and tables, if applicable). Include a numbered caption for each figure. Use the Table/Figure style for easy spacing between figure and caption.]

For more information about all elements of APA formatting, please consult the APA Style Manual, 6th Edition.