Time Series Anomaly Detection

AI & CYBERSECURITY (DSCI 6015)

Final Project Report – Mahidar Reddy Uppalapati

Submitted to

Dr. Vahid Behzadan

ASSISTANT PROFESSOR

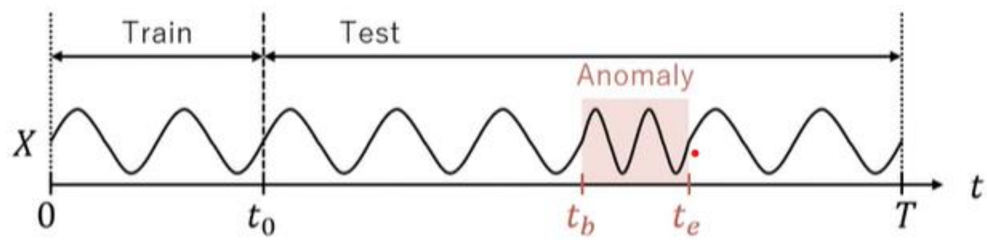
UNIVERSITY OF NEW HAVEN

Fall 2021

Introduction

Given Dataset

- Time Series $X= [x_t]$
- Split location between train and test



Evaluation

- Ground Truth Location of anomaly in test dataset i.e., $t_0 <= t_b < t_e < T$
- Assuming every timeseries has exactly one anomaly.
- Identifying the anomaly location

Goal

- Producing a single model to find anomalies for entire dataset.

Overview

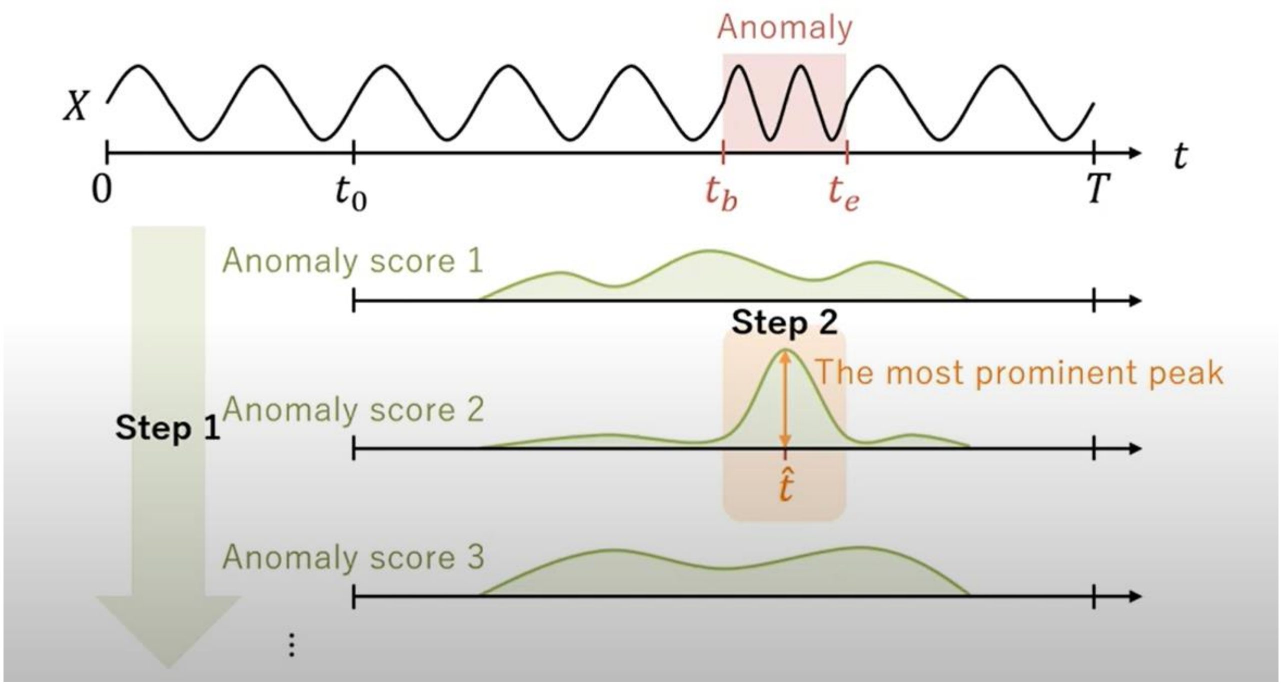
The algorithm we defined is 2-step time series anomaly detection.

Step 1.

Computing multiple non-negative anomaly scores.

Step 2.

Selecting Anomaly score with the most prominent peak.

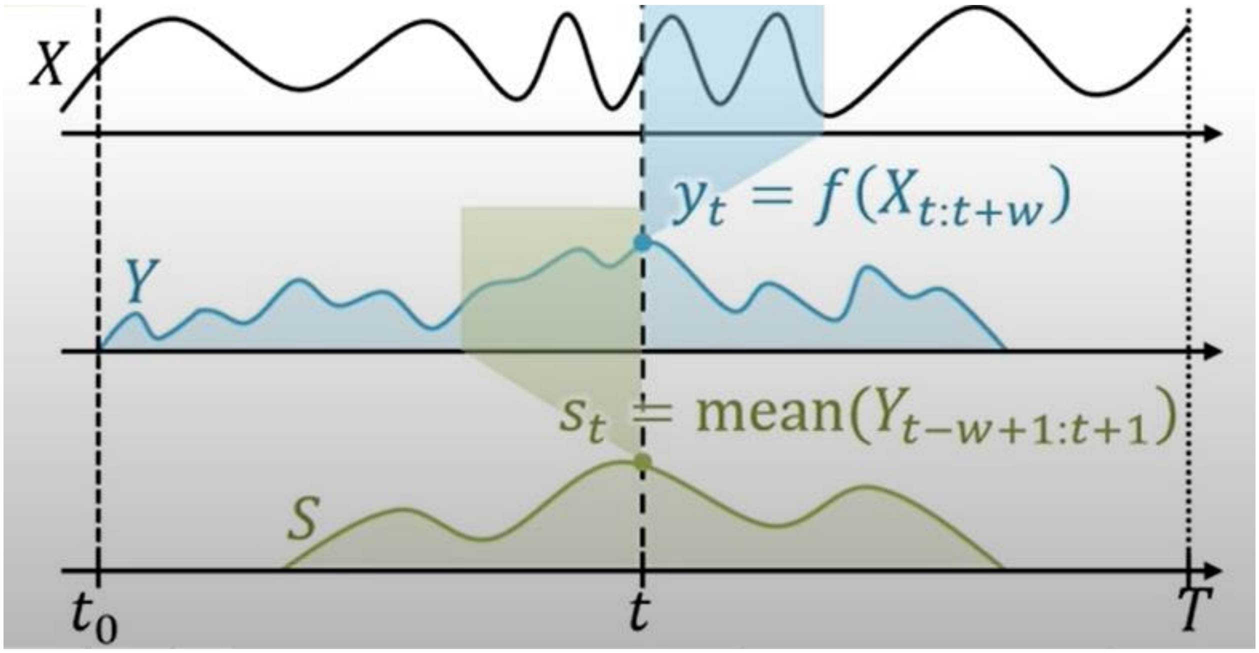


Step 1.

- Scoring along Sliding Window

$$Y_t = f(X_{t:t+w}) >= 0 \quad (t_0 <= t < T-w) \text{ , where } w = \text{Window}$$

- To deal with various lengths of time series and different anomalies we are opting for different window lengths and to detect anomalies we are finding 11 anomaly functions scores.
- Taking Moving Average along the time series to reduce the noise of scores to improve stability.
- Averaging the anomaly score within the window.



Step 2.

- Computing the prominence for each anomaly score which is "Ratio of 1st Peak to the 2nd Peak in the test data
- Selecting the anomaly score that maximizes the prominence.
- After selecting we deduce the 1st Peak location as the Anomaly Location.

Anomaly Score Functions Used:

- Orig_p2p: Peak to peak value of the original time series.
- Diff_p2p: Peak to peak value of the first order difference of time series.
- Acc_p2p: Peak to peak value of the second order difference of time series.
- Orig_p2p_inv: inverse of Orig_p2p.
- Diff_small: Percent5age of time when the absolute value of difference is small.
- Acc_std: standard deviation of acc
- Acc_std_inv: Inverse of Acc_std.
- Orig_mp_novelty: Matrix Profile (AB-join) of original Time series.

- Orig_mp_outlier: Matrix Profile (Self join) of original Time series.
- Orig_np_novelty: Normalized Matrix Profile (AB-join) of original Time series.
- Orig_np_outlier: Normalized Matrix Profile (Self join) of original Time series.

Conclusion:

So, Using the above model I build a Time series anomaly detector and deduced where the anomalies present as I don't have the labels for the given dataset I wasn't able to determine its accuracy.

Time series anomaly Detection is very important because this is used in various fields such as medicine, Incremental mode for streaming data analysis, semantic segmentation and snippet discovery.