Use of LSTM for determining the anomolus nature of the time series

Priyavrata Tiwari
AE22B007

June 24, 2024

Contents

0.1 Overview of the Work

I have tried using LSTM models to train on randomly generated time series, and have calcualted the anomaly score described in paper.

0.2 Procedure of Dataset Generation

0.2.1 Training data

The training data consists of 50 time series datasets each having 3600 data points (1 hour of data at 1 second interval).
The elements of the datasets are randomly chosen from a Gaussian Distribution with mean 0 and standard deviation 1.

0.2.2 Training of LSTM model

- The model has been trained with a sequence length of 5 used to predict the value of the series at 6th data point. With a window stride length of 1

- Only the first two data sets were used for training, as beyond that not much change in the loss was observed.

0.2.3 Generation of testing data

Three sets of time series were made in order to test the model.

- F: The set has 50 time series generated using same principle as the training dataset.

- G: The set is generated on same principles as F with random anomalous spikes in each of the time series

- H: The dataset has been generated following the same principle as X, with $\sin(t)$ added to each $h_i(t)$

0.3 Anomalous score

The parameter of anomalous score is defined as

$$A \;=\; 2C(\theta^*) - 2P_\tau(\sigma) - C_2(y, \theta)$$

where
$C(\theta^*)$ = Loss at the current time series
$P_\tau(\sigma) = \frac{1}{\pi Q} \sum_{m-1}^{Q} arctan(2\pi\tau_m \, |\sigma_m|)$
$C_2(y, \theta)$ = Previous Loss

0.4 Observations

Average anomolous score different testing datasets are:
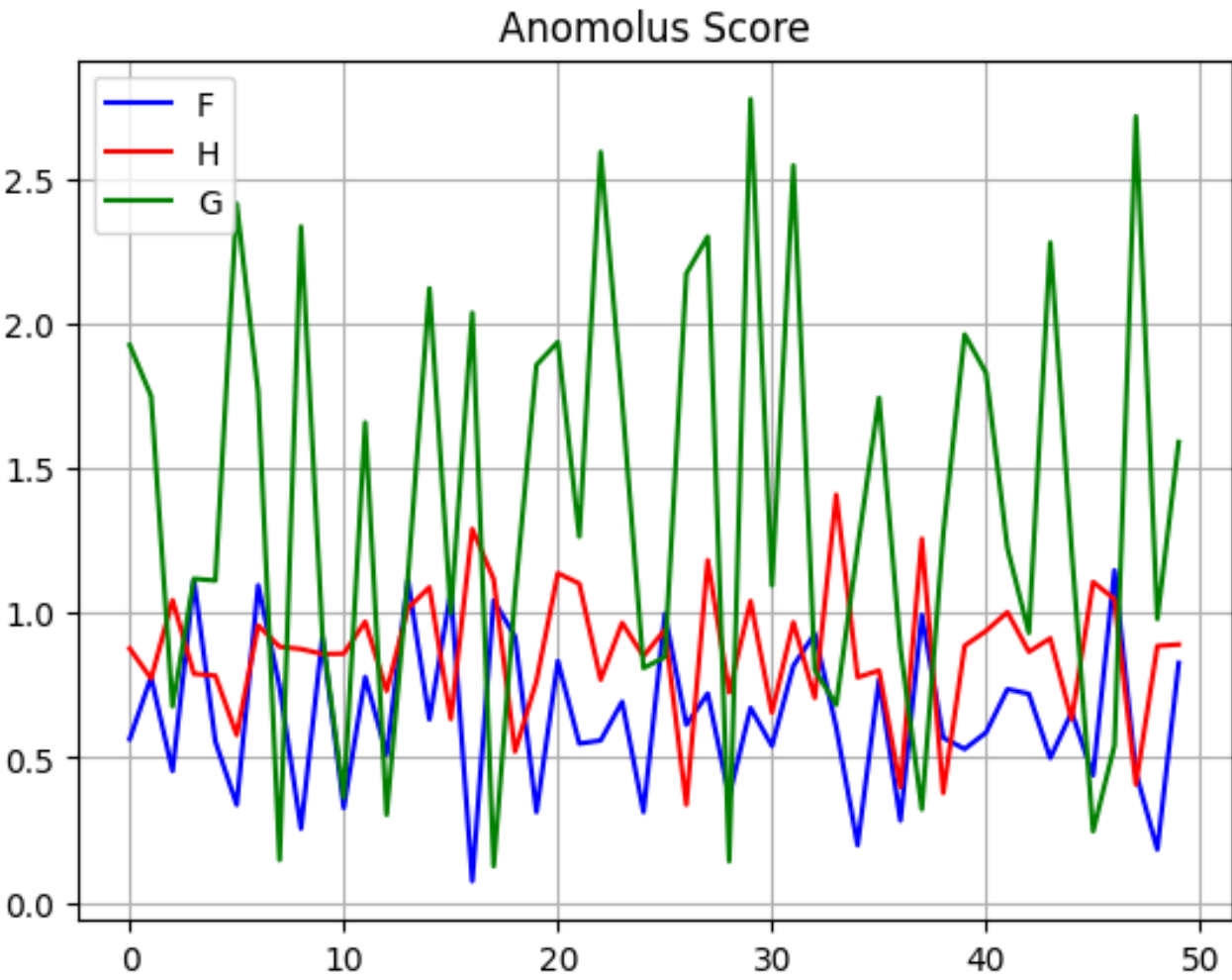F = 0.64
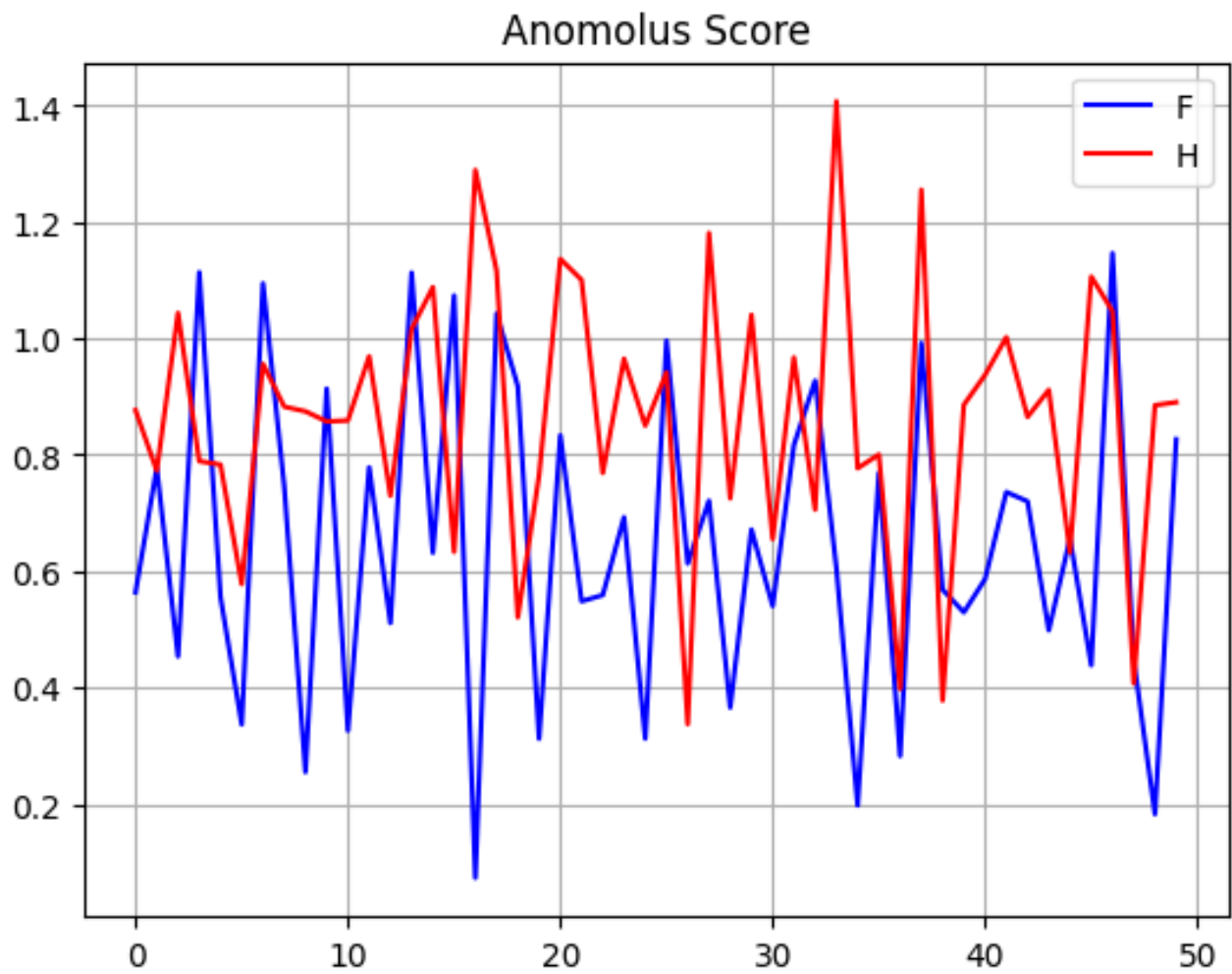G = 1.36
H = 0.83

Figure 1: Anomolus scores

Figure 2: Anomolus scores(F,H)

0.5 Results

1. Average anomaly scores gives us a sense of anomalous nature of the time series.

2. Over the dataset in the given manner anomolous score tends to show random nature, but overall character of the plots depends on the anomality present in the time series.

Github Link for the jupyter notebook