Adobe Anomaly Detection - RNN and FGSS

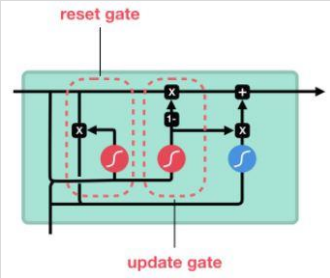
Broderick Prows, Wilson Redd, Alan Hamson, Ethan Pedersen - Chris Challis

Introduction

Adobe has servers all over the world that collect data at given time intervals. Having anomaly detection software in place to monitor this data is a valuable tool, as it can help determine and even predict when problems occur or when there is suspicious behavior. Our mentor at Adobe met Dr. Edward McFowland and found his methods of anomaly detection useful. While most anomaly detection focuses on identifying times or observations that are out of place, we were asked to implement the Fast Generalized Subset Scan (2013) to detect anomalous times and features.

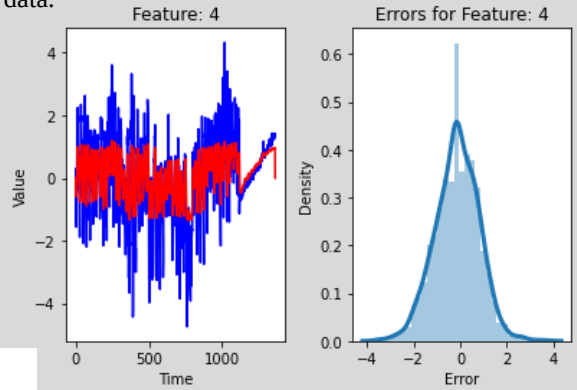
Model

The effectiveness of the Fast Generalized Subset Scan (FGSS) is dependent on a good model for the distribution of the data. In the original paper, researchers used a Bayesian Network to model the categorical data they were dealing with. We decided to use Recurrent Neural Networks (GRU) to model the distribution of each feature. We did some dimension reduction to limit the number of RNNs to train.



Likelihood

After training each RNN, we used them to obtain the likelihood for each value in the training and test data. We assumed that the errors from the training data were normally distributed, which was confirmed by looking at the density plots of the errors. We used these normal distributions to obtain p-values for how likely each observed error was in the test data. This yields a matrix with an entry (p-value) for each observation in the test data.



This is the output from the RNN, which is impressive given the noisiness of the time series.

This is the distribution of the errors from a feature in the training set. We use this normal distribution to obtain p-value for the test errors.

Subset Scanning

With a matrix of p-values that is the same shape as the test data, we are ready to begin subset scanning. This method was proposed by Dr. McFowland. Start by defining alpha max, and the set of all p-values in the matrix that are less than alpha max. Then for each alpha:

1. Convert p-values to 1 if less than alpha, else 0.
2. Sum across the rows of the matrix and sort in descending order. Score each subset starting with the first row, then the first and second row, etc until you find the maximum score using Kullback-Liebler divergence.
3. Using that subset of rows from the p-value matrix with the maximum score, repeat step 2 column-wise.
4. Repeat steps 2 and 3 until convergence.

This will leave use with the subset of rows and columns for the most anomalous data.

Results

To test this method, we added some noise to some of the test data points and tried to identify them using the subset scanning.

Simulation	# Generated Anomalies	% Found
1	50	100%
2	50	98%
3	100	100%
4	100	100%
5	200	99%