Modeling Approaches for Time Series Forecasting and Anomaly Detection

Shuyang Du[1], Madhulima Pandey[2], and Cuiqun Xing[3]

Background

With the rapid rise of real time data sources, prediction of future trends and the detection of anomalies is becoming increasingly important. Accurate time series forecasting is critical for business operations for optimal resource allocation, budget planning, anomaly detection and tasks such as predicting customer growth, or understanding stock market trends.

Problem Statement

This project focuses on prediction of time series data for Wikipedia page accesses for a period of over eighteen months. We use a number of different machine learning methods for the web traffic prediction. Unexpected deviations in web-traffic or anomalies can be an indication of an underlying issue. Detecting such deviation is non-trivial because streams are variable and noisy with outlier spikes. This problem requires methods that work on temporal sequences rather than single points. We explore three different approaches including K-Nearest Neighbors (KNN), Recurrent Neural Networks (LSTM), and Sequence to Sequence model with Causal Convolutional Neural Networks. These approaches will be evaluated based on Symmetric Mean Absolute Percent Error (SMAPE) and time series prediction versus actual value plot.

Datasets

- The dataset we used is provided by Kaggle and Google from their Web Traffic Time Series Forecasting Challenge. The dataset consists of approximately 145k time series. Each of these time series represent a number of daily views of a different Wikipedia article, starting from July, 1st, 2015 up until September, 1st , 2017. We split our dataset into **train** and **test** by July, 9th, 2017.

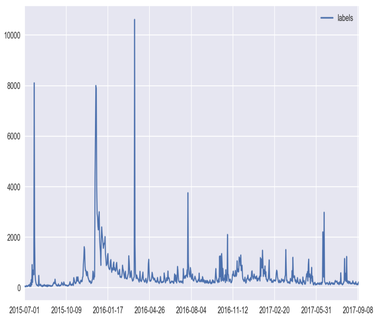

Fig. 1. Typical Periodicity

- There are significant spikes in the data, where values have a broad range from one to hundreds for several web pages. We normalize this data by adding 1 to all entries and taking log of the values and setting the mean to 0. We have the results of fourier analysis for exploring periodicity on a weekly/monthly basis. Fig 1 shows the typical periodicity.

Models-KNN

KNN is the simplest of the nearest neighbor algorithms and has been combined with some additional features including weekly seasonality, country, and traffic type. Among the different distance metrics, the Canberra followed by the Minkowski were the most effective. The Canberra distance aligns with our data which has high variances and positive values.

Models-LSTM

We use RNN with LSTM as it is well suited for learning from long variable sequence time series data because of better gradient flow topology. Following is the many to one LSTM topology we have used. We extract several additional features from the input such as days of the week, country code, auto-correlation of series quarterly and yearly. We tuned the model with different hyper-parameters such as learning rate and learning decay rate. The best results came from back-propagation length between 30-60 and LSTM state size between 30-45.

Fig. 2. LSTM

Models-Seq2seq CNN

We combine the sequence to sequence model with causal convolutions. Following is the graph for a typical model. Based on the input time series, the model will learn a set of parameters for encoding convolutional layers. Then for the prediction period, the model will learn another decoding convolutional layers based on inputs, previous encoding hidden units, previous decoding hidden units and predictions.

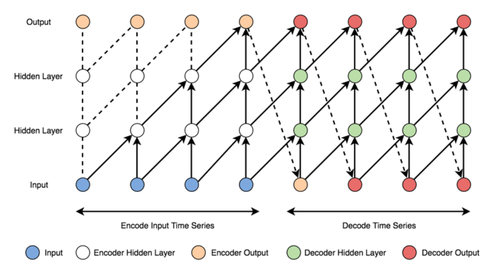

Fig. 3. Seq2seq Causal CNN Structure

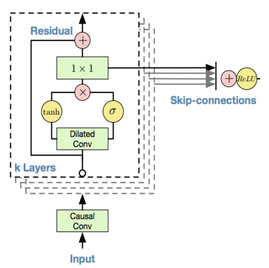

Fig. 4. Gated Cell and Skip Connection

Results and Discussion

- Time series plot for prediction versus actual values

Fig. 5. Time series plot for prediction versus actual values

- Test Results

	KNN	LSTM	Seq2seq CNN
SMAPE	165.51	143.12	42.37

Conclusions and Future Work

- Conclusions
  - Seq2seq CNN approach has the best out sample performance
  - Models can capture the stable trend but have difficulties in anomaly cases
  - Regular (non-deep) machine learning approaches like KNN can achieve comparable results
- Future work
  - LSTM enhancements to include multi-layer and sequence to sequence topologies. Tune performance with large GPUs
  - Based on the prediction create anomaly measures
  - Investigate the scenarios where some models outperform others
  - Ensemble different approaches together

[1] shuyangd@stanford.edu [2] mpandey8@stanford.edu [3] carriex@stanford.edu