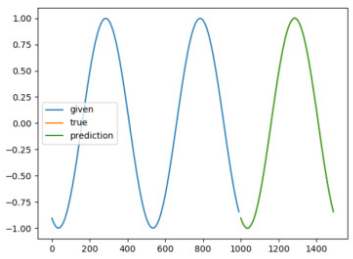
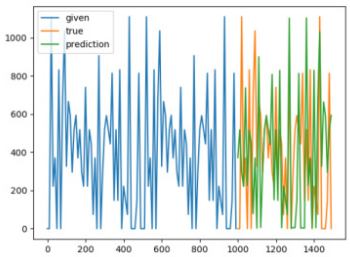
Network Traffic Prediction with Neural Networks





Project Manager
Patrick Jahnke

Principal Investigator
Prof. Patrick Eugster

Project Term
2019 - 2020

Clusters
Lichtenberg Cluster Darmstadt

Software
TensorFlow

Institute
Department of Computer Science

University
Technische Universität Darmstadt

Introduction

The bursty nature of network traffic is one of the main reasons for congestion in data centers, since traffic loads are not know in advance. Congestion reduces the network throughput and can even lead to packet loss. Therefore, it is one of the main constraints to the performance capabilities of applications and services running in the network. A traffic engineering system that could predict the load of individual network flows could reroute traffic to prevent congestion and thereby raise the performance of network applications and services. While several approaches have shown it is possible to predict network traffic in a coarse-grained manner by aggregating flows on a temporal and spatial scale, which greatly reduces the bursty characteristics of the traffic flows, a fine-grained prediction of individual flows is widely considered impossible. In the context of this work such a fine grained approach to network traffic using Neural Networks (NNs) was evaluated. We investigated if the trajectory of individual unseen flows can be predicted based on observations of the previous network traffic. For the prediction Long Short-Term Memory Neural Networks (LSTMs) were used and various possibilities to transform the input were be explored. For Deep Learning with Tensorflow and large data sets many GPU cores are needed for matrix operations, therefore we needed a High Performance Computer.

Methods

Besides normalization there are two important data preprocessing steps used in the prediction experiments. One the is clustering of similar training data to reduce the complexity of the training and the other is the transformation of the input time series to the frequency domain via the Short-Time Fourier Transformation since we observed that data which seams arbitrary in the time domain shows distinct patterns in the frequency domain. We used LSTMs to learn features from traffic time series and predict the time series of unseen traffic flows by providing the Network with a part of time series as input and the network should predict the following part of the time series.

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN
TECHNISCHE UNIVERSITÄT DARMSTADT
UNIKASSEL VERSITÄT
Philipps Universität Marburg

Results

The LSTMs were not able to learn parameters to predict the time series of unseen traffic flows for complex time series. If the train and test data were not extremly similar the LSTMs always fell back to a mean prediction or somthing similar.

Discussion

The result that the LSTMs were not at all able to grasp a concept for the traffic flows came as a surprise to use and we verified that the Networks are actually able to predict time series at all and we did not have a major fault in our implementation. The Networks were able to predict simple periodic curves like a sin wave with easy and even complex non-periodic time series could be fit some what properly if there was not not much variance in training data itself and between training and test data. To verify this we took a part of a real traffic flow time series, duplicated it and added noise. However, once we went back to our original data set of real traffic flow time series the LSTMs were not able to predict something other then a mean or something close to the mean, even if we clustered similar flows together to produce similar data sub-sets for training and testing. In the futur we are trying to eveluate why the LSTMs are not possible to predict the traffic flows more precisely. That is why we request a another project period.

Figures

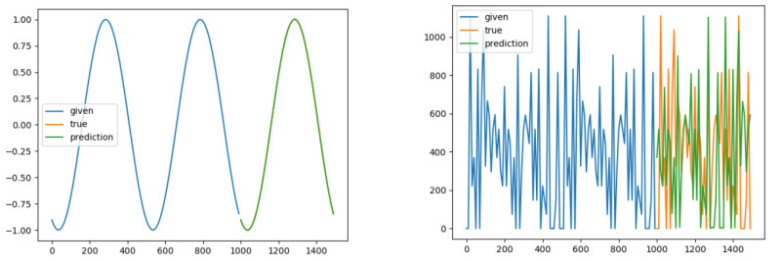


Figure 1: (left) LSTM trained on a sin wave; (right) LSTM trained on a multiplied real flow snippet with some added random noise.

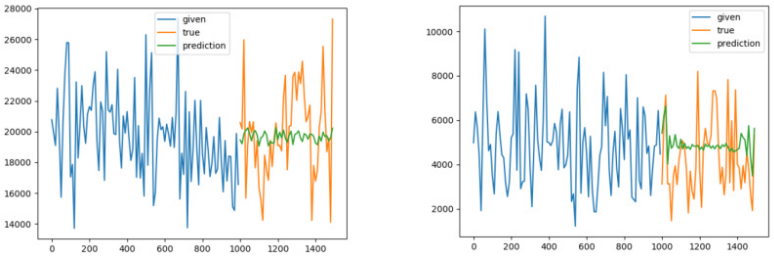


Figure 2: (left) LSTM trained on real flow time series. No FFT was used in the preprocessing. Training and test data comes from the same cluster of similar flows; (right) LSTM trained on real flow time series. FFT was used in the preprocessing. Training and test data comes from the same cluster of similar flows.

Last Update: 2022-04-22 12:17