ACADEMIC SERMINAR PAPER PRESENTATION

TSVAKI JETINA JULIET

Time-series Extreme Event Forecasting with Neural Networks at Uber

20 DECEMBER 2019

INTRODUCTION

1. Extreme event prediction has become a popular topic for estimating peak electricity demand, traffic jam severity and surge pricing for ride sharing and other applications

2. Motivated by the recent resurgence of Long Short Term Memory networks a proposal on end to-end recurrent neural network architecture that outperforms the current state of the art event forecasting methods on Uber data and generalizes well to a public M3 dataset used for time-series forecasting competitions.

3. There are a number of state of art models used before such as models (e.g., Holt-Winters) and machine learning methods (e.g., random forest). Such a system, however, is hard to tune, scale and add exogenous variables

MOTIVATION

The goal of the work was to develop an end-to-end forecast model for multi-step time series forecasting that can handle multivariate inputs (e.g. multiple input time series).

- The intent of the model was to forecast driver demand at Uber for ride sharing, specifically to forecast demand on challenging days such as holidays where the uncertainty for classical models was high.

- Generally, this type of demand forecasting for holidays belongs to an area of study called extreme event prediction.

DEFINITION OF TERMS

Extreme events - are characterized by the very largest (or smallest) values in a time series, namely those values that are larger (or smaller) than a given threshold.

State of the art - the most recent stage in the development of a product, incorporating the newest ideas and features.

Sparse data - any data which as very large zero value and very little no zero value

End-to-end - describes a process that takes a system or service from beginning to end and delivers a complete functional solution, usually without needing to obtain anything from a third party

PROBLEM STATEMENT

Accurate demand time-series forecasting during high variance segments (e.g., holidays, sporting events), is critical for anomaly detection, optimal resource allocation, budget planning and other related tasks.

This problem is challenging because extreme event prediction depends on numerous external factors that can include weather, city population growth or marketing changes (e.g., driver incentives) (Horne Manzenreiter, 2004).

PROPOSED SOLUTION

A new LSTM-based architecture and train a single model using heterogeneous time-series.

Recently, time-series modeling based on Long Short Term Memory (LSTM) technique gained popularity due to its end-to-end modeling, ease of incorporating exogenous variables and automatic feature extraction abilities

By providing a large amount of data across numerous dimensions it was shown that an LSTM approach can model complex nonlinear feature interactions which is critical to model complex extreme events.

* IT IS A MODIFICATION TO THE USUAL LSTM

DATA

Anonymized access to the rider and driver data from hundreds of cities.

Arise due to the data sparsity found in new cities and for special events. To circumvent the lack of data we use additional features including weather information (e.g., precipitation, wind speed, temperature) and city level information (e.g., current trips, current users, local holidays)

The model was fit in a propitiatory Uber dataset comprised of five years of anonymized ride sharing data across top cities in the US.

A five year daily history of completed trips across top US cities in terms of population was used to provide forecasts across all major US holidays

DATA VISUALISATION

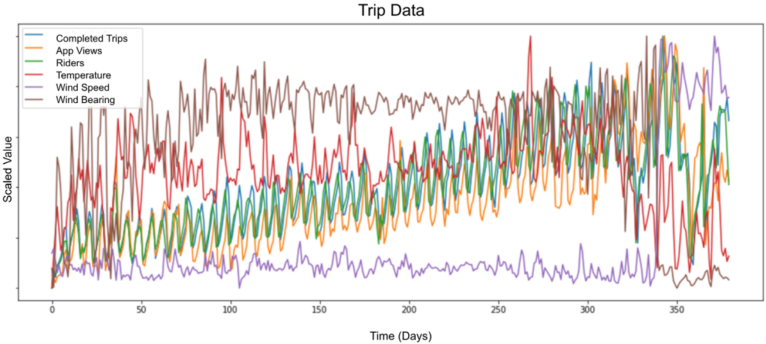


Figura: Time series data

DATA PROCESSING

Training dataset - a sliding window X (input) and Y (output)
X, Y are comprised of (batch, time, features).

Neural networks are sensitive to unscaled data (Hochreiter Schmidhuber, 1997), therefore we normalize every minibatch
De-trending the data, as opposed to de-seasoning, produces better results.

DATA SPLITING

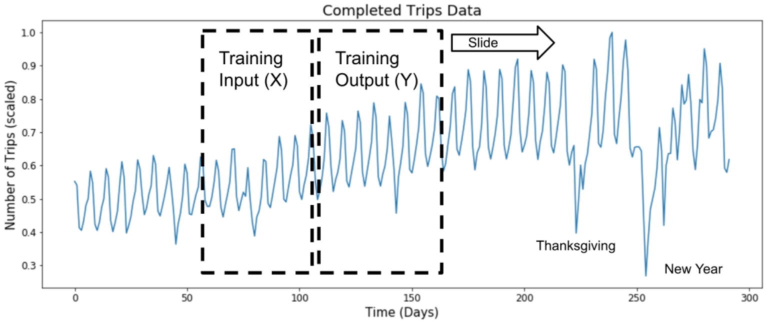


Figura: Time series data

MODEL

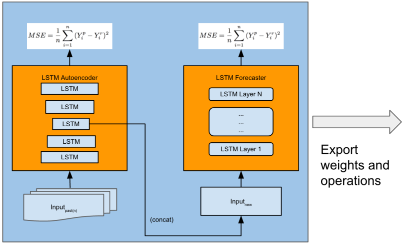
Two existing approaches were described:

Classical Forecasting Methods: Where a model was developed per time series, perhaps fit as needed.

Two-Step Approach: Where classical models were used in conjunction with machine learning models.

The difficulty of these existing models motivated the desire for a single end-to-end model

PROPOSED MODEL




```
func (tt timeTravel) nnForecast(x []float64) ([]float64, error) {
  layerWeights := tt.layerWeights
  layerBias := tt.layerBias
  var h1 []float64
  if layerWeights == nil || layerBias == nil {
    return h1, errNNForecast
  }
  prev := x

  for i := range layerWeights {
    h1, _ = stats.ElemSum(stats.DotProduct(prev,
                      layerWeights[i]), layerBias[i])
    prev = h1
  }
  return h1, nil
}
```

Figura: Generalized LSTM Model

PROPOSED MODEL

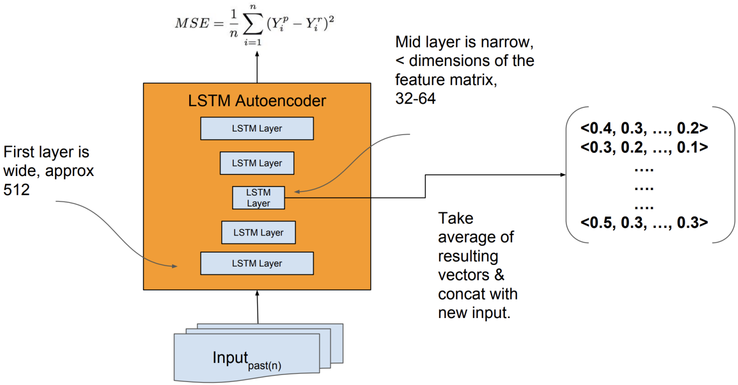


Figura: Details of the LSTM Autoencoder

RESULTS

The new generalized LSTM forecast model was found to outperform the existing model used at Uber, which may be impressive if we assume that the existing model was well tuned.

The results presented show a 2-18 percent forecast accuracy improvement compared to the current proprietary method comprising a univariate timeseries and machine learned model.

RESULTS

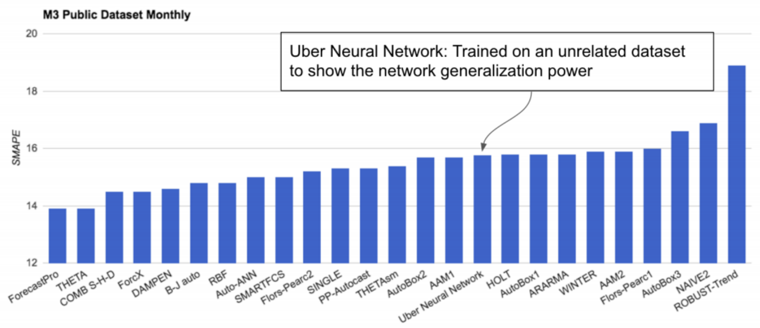


Figura: Comparisons of the Proposed Vs Existing

CONCLUSION

From our experience there are three criteria for picking a neural network model for time-series:

(a) number of timeseries

(b) length of time-series and

(c) correlation among the time-series.

If (a), (b) and (c) are high then the neural network might be the right choice, otherwise classical timeseries approach may work best.

CONCLUSION

Time-Series Type	RNN Performance	Classical Model Performance
Short Time-Series	Not enough data to train.	Symbolic Regression, HMMs perform well.
Long Time-Series	Able to optimize.	Classical Model Performance is Equivalent to RNN.
Multivariate Short Time-Series	Not enough data. While RNNs able to represent any function, need a lot of data.	Multi-varaite regression, Symbolic regression, Hierarchical forecasting perform well.
Multivariate Long Time-Series	RNN is able to model nonlinear relationships among features. Computationally efficient. Automatic feature selection.	Computation efficiency may not be optimal. Feature selection challenging.

Figura: Proposed Model selection Criterion

ABOUT M COMPETITIONS

The Makridakis Competitions (also known as the M-Competitions) are a series of open competitions organized by teams led by forecasting researcher Spyros Makridakis and intended to evaluate and compare the accuracy of different forecasting methods.

Russian Federation Chirikhin, K. Ryabko, B. University Student Novosibirsk State University Combination (S) participated and produced better results than benchmark

The biggest surprise was a "hybrid" approach utilizing both Statistical and ML features produced the most accurate forecasts as well as the most precise PIs and was submitted by Slawek Smyl, Data Scientist at Uber Technologies

The M5 is the latest of the M Competitions and will start on February 1, 2020, and end in June 30 same year using real-life data from Walmart and will be ran on Kaggle's Platform offering substantial prizes totalling 100,000 US dollars to the winners