Machine Learning Engineer Nanodegree

Capstone Proposal

Tejash Panchal
June 15th, 2018

Proposal

Domain Background

I have been working in supply chain for past many years as a test engineer for a large networking company in bay area. Various functions in supply chain rely on accurate demand forecast to perform their job accurately. As a test development engineer, I heavily rely on accurate demand forecast to develop test processes and test capacities for each of the products I am responsible for. Setting up test process is an expensive and time-consuming task; thus, if I don't setup test process that has enough capacity, orders will be delayed and we will have angry customers that will be willing to move their business to competitor. If I setup test process with excess capacity, we, as company, would waste tremendous amount of money that will have impact on company's bottom line revenue. At my company, we have tried several methods to receive accurate forecast; however, so far, we haven't been able to find an adequate solution that works for different types of products as well as different volume of product shipment.

The solution I am proposing is an attempt to solve the forecasting problem using some of the latest artificial intelligence concepts and technologies that is available today.

Attempts have been made to solve this problem using various different machine learning methods. During my research, I have found one particular research at jiraset, where the authors had attempted to use neural networks to resolve the problem. This research aligns closely with what I am attempting to do.

Problem Statement

Overestimated forecast in supply chain causes excess inventory that stagnant significant dollar amount; on the other hand, underestimated forecast causes unfulfilled orders, angry customers, as well as lost sales and opportunities. The goal of this project is to develop a solution using the latest concepts of neural network that predicts accurate demand forecast for networking products.

Datasets and Inputs

For this project, I am considering to use dataset "historical product demand", readily available at Kaggle website. From the dataset, I plan to use three fields: Product_Code, Date, and Order_Demand. This dataset has 1048575 rows of data and it has 2160 different products demand forecast. To simplify the process, I intend to use data for only one product (Product_0001) with 597 rows of data to train and test my model.

While working with time series analysis, in many cases, one must account for seasonal spikes or quarter end spikes in demand. There are various ways to adjust that. One of the way that I will be implementing is to use “lag”. Lags are very useful in time series analysis to correlate data from one day, one week, one month, or one quarter to next. Lags allows us to correlate the values within a time series with previous copies of itself. This phenomenon is known as autocorrelation. One benefit to autocorrelation is that we can identify patterns within the time series, which helps in determining seasonality, the tendency for patterns to repeat at periodic frequencies.

Solution Statement

The solution I am proposing attempts to solve the forecasting problem using Long Short-Term Memory (LTSM) concept of recurrent neural network (RNN). Unlike many of the traditional time series models, one of the advantage of using LSTM is that LSTM model is not required to be trained on the data you are intending to predict. You can train the model on series that you have sufficient data for and then use it on a series that you are intending to predict. This is promising approach since in my company, we are always developing new products; thus, we may not have sufficient data to train the model with for new product.

Benchmark Model

There are multiple forecasting techniques being used in different industries to predict the forecast. Some of them are listed as below:

- Naive Approach
- Simple average
- Moving average
- Single Exponential smoothing
- Holt’s linear trend method
- Holt’s Winter seasonal method
- ARIMA

In this project, I will be using ARIMA method as benchmark model. I will be comparing the results that I achieve using LSTM method against my benchmark model to validate my model. I will not only be comparing the predictions but also RMSE for both methods.

Evaluation Metrics

There are various different performance measures can be used to evaluate the time series forecasting model. It can rather get confusing. Time series generally focus on predicting the real values which aligns with regression model; therefore, I will focus on methods for evaluating the real value predictions.

Some of the measures that can be utilized to measure the performance of the model are as below:

- Forecast Error
- Mean Forecast Error
- Mean Absolute Error
- Mean Squared Error
- Root Mean Squared Error

I intend to use Root Mean Squared Error (RMSE) to compare my LTSM model vs. the Benchmark model.

Project Design

For my Forecast Demand Prediction project, below I have listed programing language, libraries needed to develop the project, as well as theoretical workflow of the project that I intend to use during development cycle.

- Programming Language
  - Python 2.7
- Libraries:
  - Scikit-learn
  - Keras
  - TensorFlow
  - Pandas
  - Numpy
  - Math
  - Statsmodels
- Theoretical workflow:
  - Load and prepare the data
  - Normalize the data
  - Split the data into train and test:
    - For train data, use data from date 12/16/2011 to 12/21/2015
    - For test data, use data from date 1/5/2016 to 12/26/2016
  - Create benchmark model
  - Calculate benchmark model Root Mean Squared Error (RMSE)
  - Plot benchmark model predictions
  - Create LSTM network
  - Fit LSTM network
  - Make Predictions
  - Invert Predictions
  - Calculate LSTM model Root Mean Squared Error (RMSE)
  - Plot LSTM baseline and predictions

Reference

- http://www.business-science.io/timeseries-analysis/2017/08/30/tidy-timeseries-analysis-pt-4.html
- https://machinelearningmastery.com/time-series-forecasting-performance-measures-with-python/
- https://www.analyticsvidhya.com/blog/2018/02/time-series-forecasting-methods/
- https://www.kaggle.com/sydjaffy/historical-product-demand
- https://www.ijraset.com/fileserve.php?FID=381
-