

Adarsh Goyal, Shreyas Krishnamurthy, Shubhda Kulkarni, Ruthwik Kumar, Madhurima Vartak, Matthew A. Lanham
Purdue University, Department of Management, 403 W. State Street, West Lafayette, IN 47907

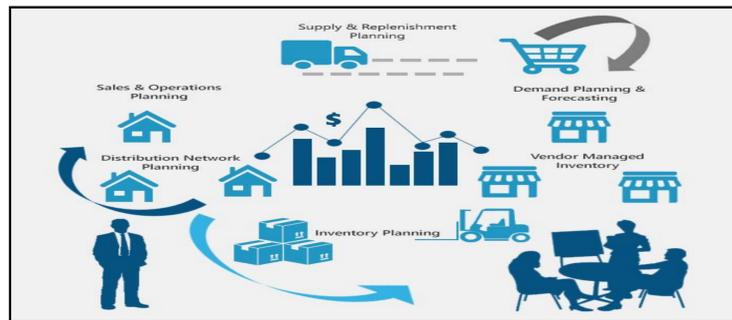
goyal45@purdue.edu; krishn96@purdue.edu; kulkar38@purdue.edu; kumar306@purdue.edu; vartak@purdue.edu; lanhamm@purdue.edu

Abstract

This study focuses on predicting demand based on data collected which spans across many periods. To help our client build a solution to forecast demand effectively, we developed a model using Long Short Term Memory (LSTM) Networks, a type of Recurrent Neural Network, to estimate demand based on historical patterns. While there may be many available models for dealing with a time series problem, the LSTM model is relatively new and highly sophisticated to its counterparts. By comparing this study which works excellently for sequential learning, to the other existing models and techniques, we are now closer to solving at least one of many complications apparent across industries. The study becomes all the more important for supply chain professionals, especially those in purchasing, as they can now rely on a highly accurate model instead of basing their forecasts purely on intuition and recent customer behavior.

Introduction

No matter how good the predictive model is, one is never going to achieve 100 percent accuracy. However, the cost savings that can be achieved by continuously predicting/ forecasting demand better is what separates an average company from the market leader. No matter how strong the company's supplier/ distributor network is, not being able to predict stock accurately, can be very costly. Whether losing customers by not being able to meet their demand due to understocking, or incurring ridiculous costs by overstocking, and thereby blocking working capital in the process, the importance of forecasting demand cannot be underemphasized. The biggest companies of the world like Walmart, Amazon, and Apple are all investing heavily in analytics and especially supply chain analytics to get their demands and sales predictions correct.

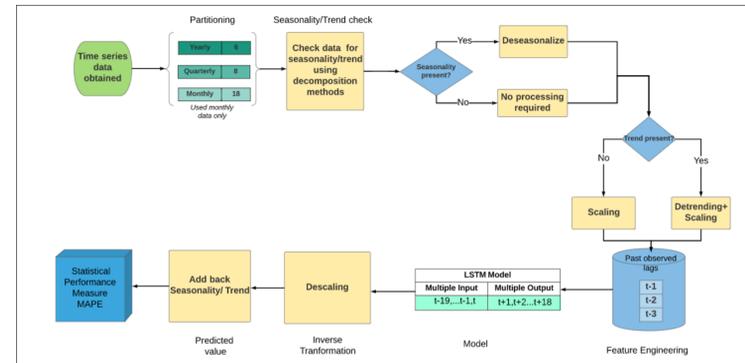


The objective of this study was to help our client provide better solutions to their customers when it came to demand forecasting. Our client had previously run other models such as 'Feedforward Neural Nets' and the 'Theta Exponential smoothing' to work on the dataset and their results are used as a benchmark to check the performance of our LSTM model.

Methodology

Data: The data used in the study was provided by the client and comprises of just one feature which is the value (demand quantity). This was a time series forecasting problem involving predicting demand for the next few periods based on the available data for earlier periods. However, there are various batch ID's (product categories) spanning over three different frequencies or time series (monthly, quarterly and yearly).

- **Partitioning data:** Our data consists of yearly, quarterly and monthly frequencies. Since quarterly and yearly data have less number of data points, LSTM trains ineffectively on this set and poses the issue of overfitting. Hence, we forecast on monthly time series only, which is 18 time steps.
- **Check seasonality and Trend (Deseasonalize /Detrend if required):** LSTM or any neural network struggles when working with non-stationary data. We use STL decomposition to separate seasonal, trend and residual components and LSTM model is then applied on the residual part to learn long term dependencies.
- **Scaling:** As the data values may vary across a wide scale, we perform min-max normalization to ensure they lie within a fixed range (0 to 1) for better forecast.



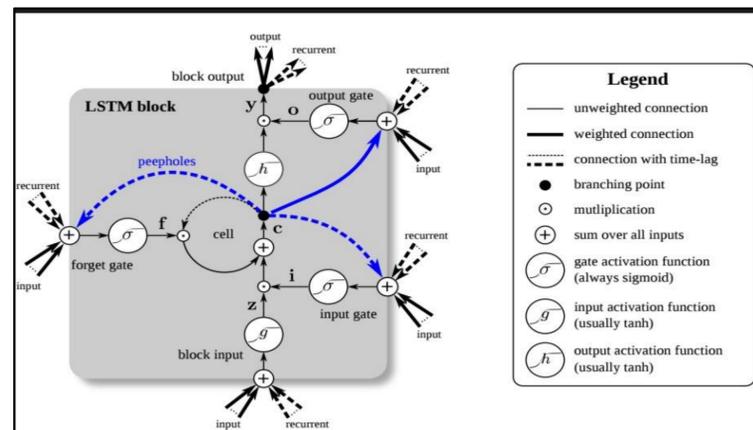
- **Feature Engineering:** Since we do not have parameters for business unit of the data, we only use past 20 observed lags as features to our LSTM model for forecasting.
- **LSTM Model:** Long short-term memory network is a type of recurrent neural network, specifically designed to learn long term dependencies, overcoming the problems of vanishing and exploding gradient. The current model works on the Many-In-Many-Out mechanism, that is it predicts multiple forecast outputs using multiple inputs (lag variables).
- **Descaling:** The output of the LSTM network is inverse transformed to obtain the original range of values.
- **Adding back the seasonality and trend:** We add back the seasonal and trend components to the forecast output from the model.
- **Statistical performance measures:** The performance of the LSTM model is judged over MAPE (Mean Absolute Percentage Error) across all the monthly time series.

Model Formulation

LSTM networks are a type of recurrent neural networks (RNNs), i.e., neural networks where connections between units form a directed cycle. This allows them to retain memory i.e. exhibit temporary dynamic behavior. LSTM networks are capable of learning long-term dependencies and can overcome the previously inherent problems of RNNs, i.e., vanishing and exploding gradients.

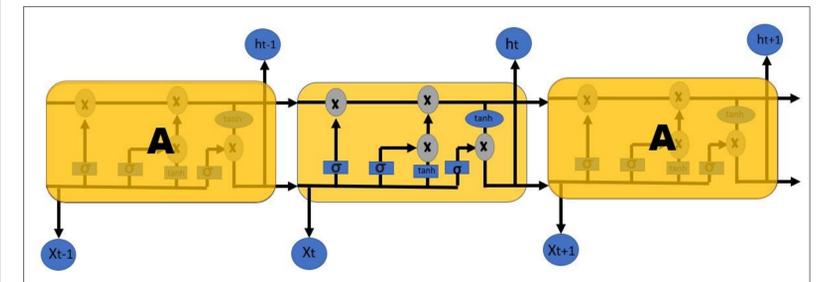
LSTM networks, like dense layers, have an input layer, one or more hidden layers, and an output layer. The main characteristic of the model is contained in the hidden layer(s) which consists of memory cells. Three gates in each memory cell maintain a cell state s_t : a forget gate (f_t), an input gate (i_t), and an output gate (o_t).

The structure of the memory cell is illustrated in the figure below.



- Forget gate: Defines which information is removed from the cell state.
- Input gate: Specifies which information is added to the cell state.
- Output gate: Specifies which information from the cell state is used as output.

At every timestep t , each gate is presented with the current input and the output h_{t-1} of the memory cells at the previous timestep $t - 1$. Each x_t gate has a bias vector associated with it which adds to its calculated value after every input. The working of a LSTM layer can be summarized by the following figure:

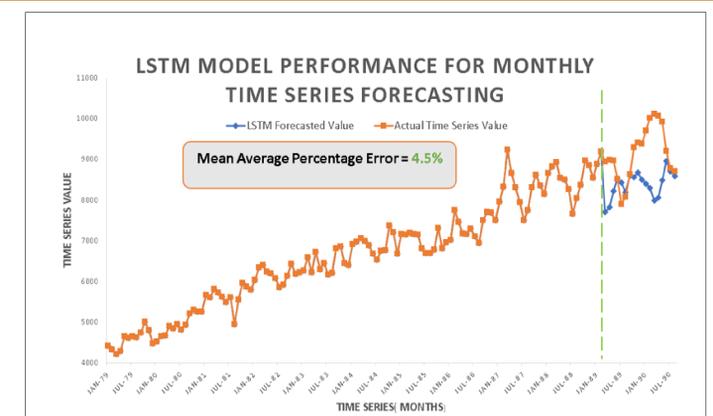


Training a neural network requires multiple iterations called epochs. The weights and bias vectors keep adjusting so that the loss of the specified function is minimized across the training data set.

In our case, we make use of Adam optimizer (commonly used), as optimizer via **keras** for training of the LSTM network. The specified topology of our trained LSTM network is hence as follows:

The LSTM Input layer has one feature and 20 timesteps corresponding past observed time lags. The hidden layer has 15 neurons and the output layer (dense layer) has 18 neurons corresponding to our forecast for next 18 time steps.

Results



The last 18 data points have been forecasted using LSTM for 1400 monthly univariate time series.

We use **MAPE (Mean Average Percentage Error)** to test the performance of the forecasted model with respect to the actual data points. On running the LSTM model for 1400 series, we observe most MAPE values between 4% to 35% with the average being around 24%.

Conclusion

Inventory Management is the core of operational performance in most industries and thus an efficient solution that can help organizations predict future demands is imperative. It was observed that LSTM Neural Network is one such model that performs better (lower MAPE) than other baseline models and is much simpler to implement as it requires minimal feature engineering. However, with more data points in each time series the LSTM will train better and the accuracy of forecasts will be enhanced.

Acknowledgement

We are thankful to our mentor Professor Matthew Lanham and our industry partner for providing us with this opportunity. The Purdue BIAC partially funded this work.