Store Item Demand Forecasting using Deep Learning

Akash Savaliya

ITCS-5156 Spring 2022 - Lee

February 28, 2022.

Based on - Punia, Sushil; Nikolopoulos, Kostas; Prakash Singh, Surya; Madaan, Jitendra K.; Litsiou, Konstantina "Deep learning with long short-term memory networks and random forests for demand forecasting in multi-channel retail". International Journal of Production Research. 2020, 58(16). 4964-4979. https://doi.org/10.1080/00207543.2020.1735666

1. Introduction

This project report is related to the implementation of the work published by Punia Sushil et al. "Deep learning with long short-term memory networks and random forests for demand forecasting in multi-channel retail" [1]. The team formulated a new approach to solve the problem of demand forecasting in the retail industry. The team followed a hybrid modeling technique. The technique involved used the sequential deep learning technique named long-short-term-memory networks (LSTM) and machine learning technique Random forest. The results obtained in this project are similar to the work mentioned in the paper.

1.1. Problem

The uncertainty of the demand for products in the market makes the job of the manufacturing and retail industries very difficult. This uncertainty in the market can be due to customer behavior or marketing efforts. Demand can be forecasted using statistical and mathematical approaches that help a lot to deal with uncertainty. In statistics, any kind of forecasting problem is a time-series problem which is a type of univariate analysis. The literature advocated for a "horses for courses" approach, in which different types of forecasting methods are expected to be better suited to different sorts of data[1].

1.2. Motivation and Challenges

The retail industry is under constant pressure with meeting up the demand. The improper demand estimates lead to “opportunity loss” when the products are not available during high demand and “wastage of resources” when demand is low. An accurate forecast of demand can help these companies to manage their inventory, plan their resources effectively and wisely, and to manage the engaged workforce. These factors are very important for the retail industry to maximize its profits and reduce unwanted expenses.

The time series univariate data be decomposed into the trend, seasonality, and noise, as shown in Figure 1. The traditional methods used for forecasting to work on these time-series data try to find a relationship between the target(here demand) and the decomposition. These methods do help to predict the forecast with effective estimations of trend and seasonality. The noise is hard to estimate for these methods since noise is dependent on external factors and values related to the product like price, quality, and so on.

The proposed method LSTM is used to find the estimation related to trend and seasonality, and the random forest is used to inculcate the noise attribute. The hybrid model LSTM and random forest make the demand forecasting more reliable to overcome the “opportunity loss” and “wastage of resources”.

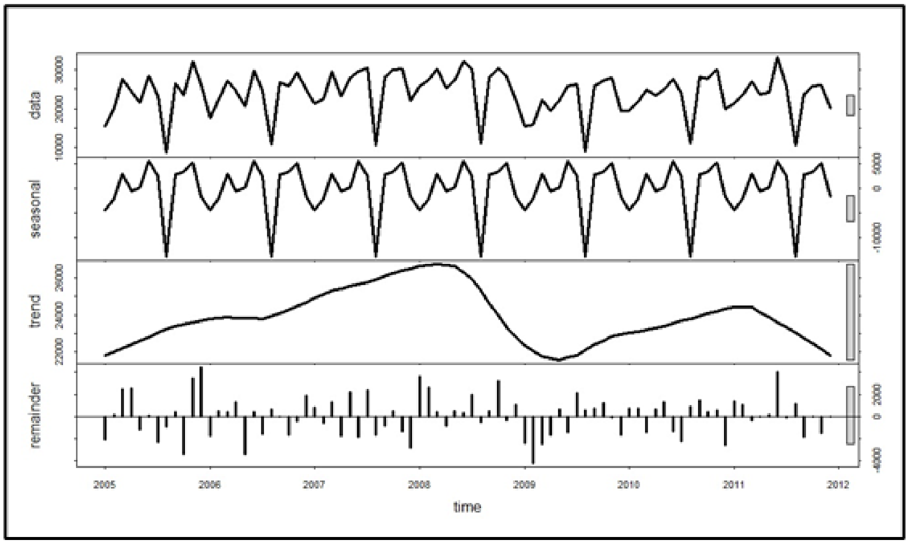

Figure 1 - Time series data decomposition

1.3. Summary of Approach

The overall architecture uses the data from the transactional data which is then converted to a dataset where we have temporal and product explanatory data. The temporal data is actually the demand(quantity) that we are trying to forecast and the product explanatory data is used for the noise in the forecast. At the time of training, the temporal data is fed to the LSTM to build a forecasting model with hyperparameter tuning using a grid search approach. The residuals of this forecast are calculated and it is used as a target variable for random forest regressor. During prediction, the final complete model gives an aggregated value from both the trained models.

2. Related works

2.1 Forecasting aggregate retail sales: a comparison of artificial neural networks and traditional methods

The research held to find out how artificial neural networks(ANN) perform in comparison to the traditional methods like Winters' exponential smoothing, Box-Jenkins ARIMA model, and multivariate regression over aggregated US retail demand data. Also, the paper demonstrates how ANN can be used for forecasting problems.

The methods are compared on the basis of two types of tests, mean absolute percentage error (MAPE), and two different approaches (one step and multiple steps). The Diebold and Mariano (DM) test and Wilcoxon's signed-ranks test (SR) are the tests being used. The DM test is used for the significance of the difference between the squared percentage errors of the two models, whereas the SR test gives an observation with a larger absolute-squared percentage error difference a higher weight than that with a smaller difference[2]. The out-of-sample data is used for the whole comparison test.

The results from these tests concluded that the ANN outperformed the traditional univariate-based methods. These results support the notion that neural networks can be used for demand forecasting for the retail industry.

2.1 An Experimental Review on Deep Learning Architectures for Time Series Forecasting

The paper discusses the efficiency of various deep learning techniques like multilayer perceptron (MLP), Elman recurrent neural network ERNN, long-short term memory (LSTM), gated recurrent unit (GRU), echo state network (ESN), convolutional neural network (CNN) and temporal convolutional network (TCN) for time series forecasting. The deep learning models are evaluated on the basis of WAPE and statistical analysis. WAPE, weighted average percentage error, is a more suitable alternative for intermittent and low-volume data. It rescales the error dividing by the mean to make it comparable across time series of varying scales. [3]

The statistical analysis consists of various tests Friedman test, Holm-Bonferroni’s procedure, and Wilcoxon's signed-ranks test. Friedman test is a non-parametric test that allows detecting global differences and provides a ranking of the different methods.[3] Holm-Bonferroni’s procedure detects significant differences between each pair of models, which allows establishing a statistical ranking[3].

This extensive experimental review over deep learning models concludes that the recurrent neural nets are giving the best performance with less number of training layers whereas convolutional neural nets give better results on the cost of multiple training layers and kernels. Among the recurrent neural nets, the LSTM outperformed GRU. The present approach for demand forecasting is backed by these promising outcomes for LSTM.

3. Method

The transactional data is collected and aggregated in the form of a combined dataset of temporal data and non-temporal data. This complete model training is divided into two stages viz LSTM for univariate analysis and Random Forest for covariate analysis.

3.1 LSTM architecture

LSTM is a type of recurrent neural network which are capable of retaining states of the previous that make them very useful for sequential data like time series data. LSTM is an improved version of the Gated recurrent unit (GRU) based neural network.

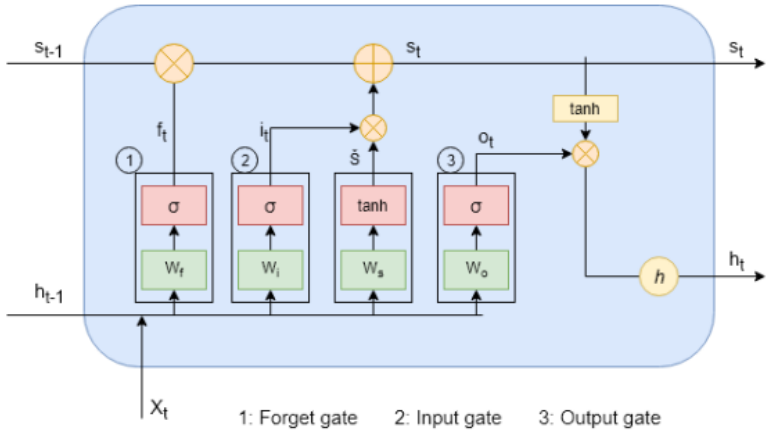

Figure 2 - LSTM architecture [1]

One single LSTM unit (node) consists of three layers: 1) an input layer with a number of neurons equal to the number of input variables, 2) single or multiple hidden layers and 3) an output layer with a number of neurons equal to the number of output variables [1]. The LSTM layer consists of these interconnected units that take in the sequential data to give out a single value which is the predicted value for the next value. After training, residuals are calculated after predicting the forecast for the training data.

3.2 Random Forest

The residuals from the LSTM stage are then used by the random forest regressor model as the target variable. The random forest model uses non-temporal data like price, the number of orders, and so on, to reduce the loss in predicting the residuals. The robust nature of the random forest model helps generalize the noise factor for demand forecasting.

3.3 Final Forecast

In this stage, the trained models are used for the final prediction. At the time of prediction, the forecast from LSTM and the residuals predicted from the random forest are aggregated to forecast with better accuracy.

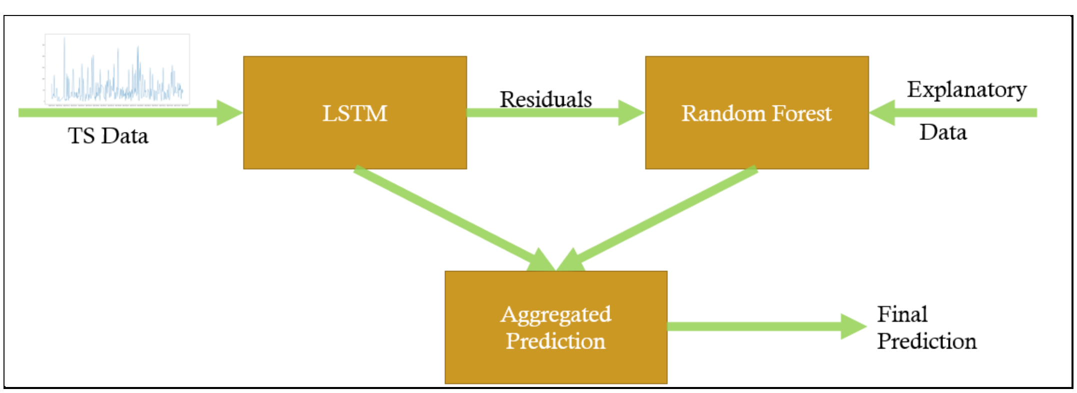

Figure 2 - Complete architecture

4. Experimental setup

4.1 Data

The data is collected from the UCI Machine Learning Repository. The data consists day on day transactions of an online retail store. The features in this data include InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID, and Country. The dates ranged from 12/1/2010 to 12/9/2010.

On counting distinct dates for a particular product sold in particular countries, I found most of the product the products were not sold throughout the period. This experiment was done on only one product “WHITE HANGING HEART T-LIGHT HOLDER”. The values for quantity are aggregated for this product date-wise, where quantity is our demand value for LSTM and other values like relative unit price, the number of customers and the number of invoices are our nontemporal data for the random forest.

4.1 LSTM

The LSTM model is trained with hyperparameter tuning using grid search of parameters:

1. Number of layers - 2 and 3
2. Activation function - 'linear', 'relu', 'tanh'
3. Optimizer algorithms - "rmsprop", "adam", "sgd"

4.1 Random Forest

The random forest model is built using 10-fold cross-validation to avoid overfitting and obtaining a generalized regressor to get better results.

4.1 Test Results

The grid approach built the best model from LSTM. On plotting, actual and predicted values just from the LSTM, the MAE (Mean Absolute Error) was 74.28. After training the random forest model over the residuals from LSTM, the actual and predicted values from the final aggregation of LSTM + RF gave an MAE of 34. These results do not match with the results in the paper. This could be because of noise in the data. Although, the results are pretty satisfying looking at the drastic difference in the results from LSTM and LSTM + RF.

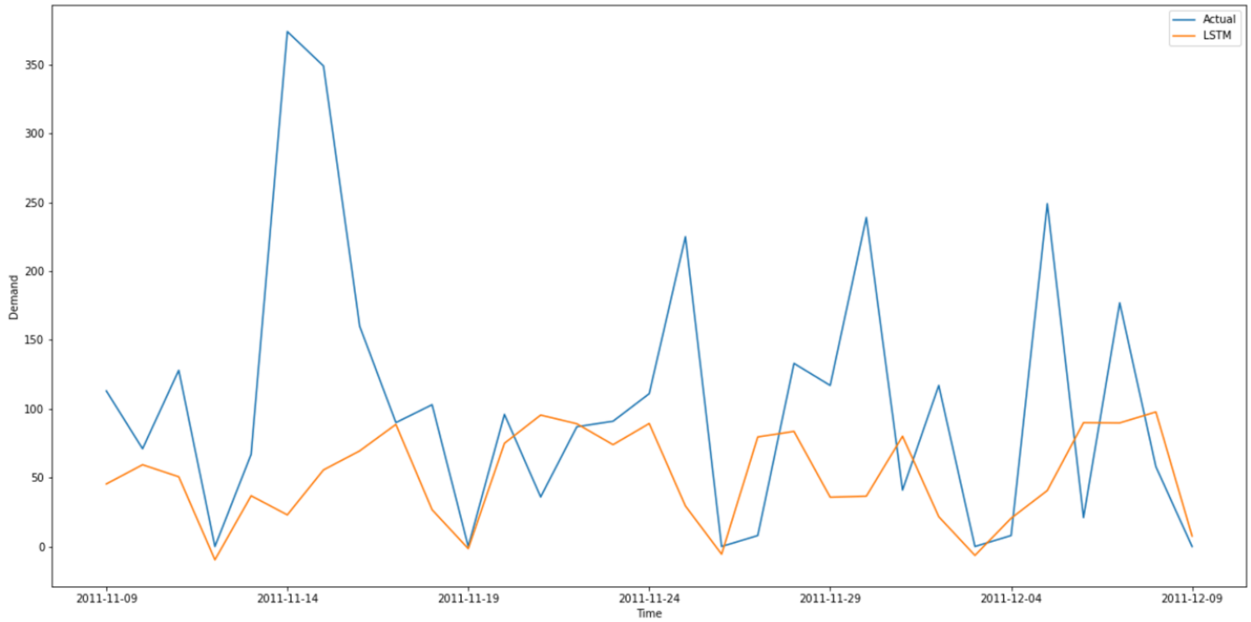

Figure 3 - Actual and Predicted from LSTM

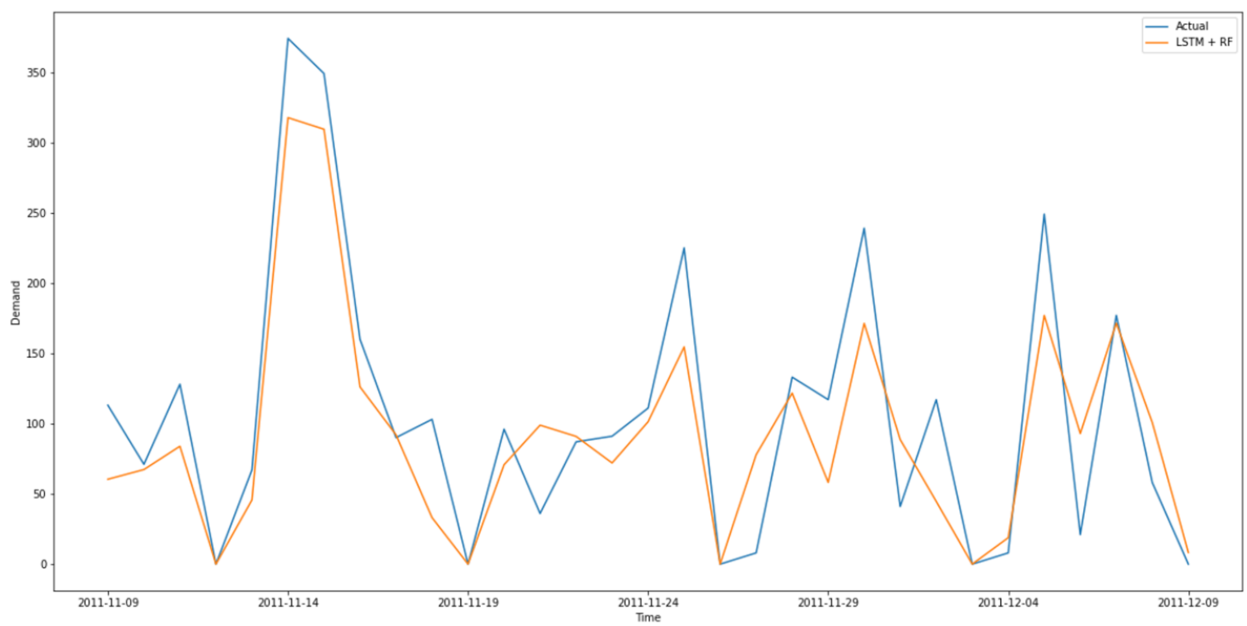

Figure 4 - Actual and Predicted from LSTM + RF

5. Conclusion

As per the results shown above, in conclusion, the hybrid model of LSTM and RF works more efficiently than only LSTM for demand forecasting problems in the retail industry. The challenging part was to decide the size of the sequence but eventually found the weekly drops in the data to zero these days were the Saturdays. The size of the sequence finally was selected as 7 days for training. The learning component from this project is deep learning models can be used for demand forecasting in the retail industry but may differ in results for other industries.

6. My Contribution

The grid search approach and finding the best model for LSTM is completely my contribution. I didn't find any exact similar or existing solution to the current work.

7. References

[1] Punia, Sushil; Nikolopoulos, Kostas; Prakash Singh, Surya; Madaan, Jitendra K.; Litsiou, Konstantina "Deep learning with long short-term memory networks and random forests for demand forecasting in multi-channel retail". International Journal of Production Research. 2020, 58(16). 4964-4979. https://doi.org/10.1080/00207543.2020.1735666

[2] Ilan Alon, Min Qi, Robert J. Sadowski. "Forecasting aggregate retail sales:: a comparison of artificial neural networks and traditional methods" Journal of Retailing and Consumer Services. 2001,8(3). 0969-6989. https://doi.org/10.1016/S0969-6989(00)00011-4

[3] Pedro Lara-Benítez, Manuel Carranza-García, José C. Riquelme." An Experimental Review on Deep Learning Architectures for Time Series Forecasting" International Journal of Neural Systems. 2021,31(3). 2130 - 0011. https://arxiv.org/abs/2103.12057