Forecasting Product Demand from Historical Data Using Machine Learning Methods.

Devon K. Reed[1 3 5], **Ayodeji Odetola**[2 3 5], **Mahfuza Farooque**[1 4], **and Suman Saha**[1 5]

[1]Department of Computer Science
[2]Department of Data Science
[3]Department of Mathematics
[4]Penn State University Park, State college, PA
[5]Penn State Altoona, Altoona, PA

ABSTRACT

Recently, time series forecasting has become a subject of interest for many data scientists and researchers alike. Both statistical and machine learning methods have been applied to time series data in order to predict future phenomena like the weather and stock prices. In our global climate, one avenue of time series forecasting that continues to become more imperative is the prediction of the demand for perishable goods. Each year, vast quantities of food are wasted globally with the number continuing to grow. To abate this crisis, an accurate time series model could be used to predict the number of resources needed at the store level. However, current forecasting models are plagued with inaccuracies and are only utilized as decision-making tools for upper-level management who ultimately determine the final decision. We propose an exploratory approach by evaluating different models on their ability to predict market demand. As a baseline, a statistical model is included to measure if the machine learning models show superior predictive capability. We observed that the machine learning models performed slightly better than the statistical baseline. This shows that machine learning has the capability of better predicting market demand than statistical models.

Introduction

In our current global climate, food waste and shortage has become a problem of increasing importance. Many perishable items on grocery store shelves expire and must be thrown away never to be used.[1,2] On the other hand, there are food shortages world wide. Yet, according to the Food and Agriculture Organization of the United Nations, global food waste is estimated at 1.3 billion tons of food. Therefore, it is clear a more optimized system for allocating our resources is vital for the future of agricultural sustainability.

Recently, time-series forecasting has become a topic of interest in many fields of study. time-series forecasting has important applications in prediction models that consume temporal data.[1] For example, weather data can be intuitively represented as a time-series where temperature is reported along with a date. Using this weather time series, future data points can be extrapolated using various models and techniques; thus, giving a weather forecast of the future. Currently, the most cutting edge models use machine learning to make meaningful inferences from the data.[1,2] We believe such an approach could be used to accurately predict future market demand allowing countries and business to more efficiently allocate food resources reducing both waste and shortages.

However, current models are often plagued with inaccuracies and are only practically used as a decision making aid for upper-level management who ultimately choose how to allocate resources.[2] Thus, our work focuses on exploring different machine learning models and specifically test which ones perform the best and the worst on a time-series prediction task.

Methods

In order to gain an idea for how our prediction algorithm would perform we decided to first look at a baseline line procedure. The algorithm we went with was the Moving Average (MA) algorithm which is a generic forecasting method for time series.The (MA) uses a sliding window to take the average over a set number of time periods.[2] For our case, it will be taking the average of our unit sale data points, $u_0, u_1, ..., u_n$ where n is the total number of data points that we have. If M is the size of our window then the moving average over our first time interval will be, $\bar{a}_{MA} = \frac{1}{M}\sum_{i=0}^{n-1} u_{M-i}$, for our purposes $n = 1678$. The data set we used was the Corporación Favorita Grocery Sales Forecasting on kaggle, since the large stores of Corporación Favorita are located in the Pichincha State, in an effort to save some computer memory we decided to only view the unit sales of stores in that region. Next we decided to implement a Feed-Forward Neural Network (NN) as our baseline machine learning model. The architecture we used involved 2 hidden layers of size 8 and 16 respectively. We used the relu activation function defined as

$relu(x) = max(0,x)$ with the output vector of the lth layer being $X^l = relu(W^l X^{l-1} + b^l)$ with W^l and b^l being the weights and bias vectors in the lth layer.Finally the last model we implement is the Recurrent Neural Network (RNN) with a Long Short Term Memory cell (LSTM). The architecture for our RNN Model involves1 Long Short Term Memory (LSTM) cell, 2 dropout layers and several Dense layers. The current "state" (denoted by out_t) of the LSTM is given by, $out_t = relu(W_{out}[h_{t-1}, x_t] + b_{out})$ where h_{t-1} is the previous LSTM block. Below in Figure 2 we look at the visualization of the RNN. In order to evaluate our

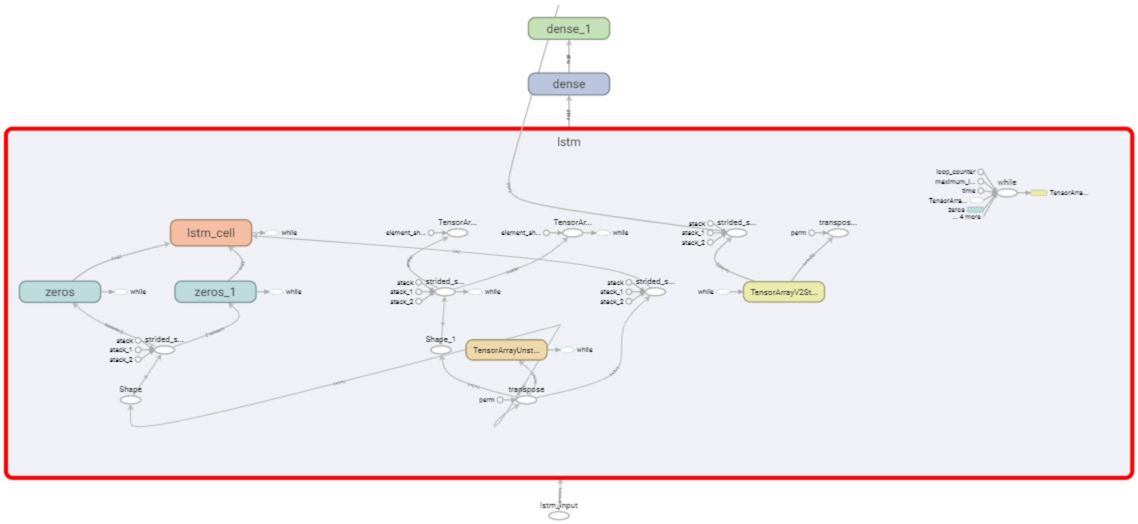

Figure 1. Recurrent Neural Network with LSTM Cell

model we decided to go with the Mean Absolute Error (MSE) metric defined as $MAE = \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y}_j|$. This metric is fairly easy to implement and reports on the raw difference of our prediction and the target value making it the optimal choice for this project.

Results

In order to get a more holistic evaluation of our MA baseline model, we decided to test the model with two windows of size 9 and 21. We denote the MAE score of the MA with the 9 and 21 day window as MA_9 and MA_{21} respectively and observe that $MA_9 = 60834$ and $MA_{21} = 69883$ measured in unit sales. Next, for our Neural Network we split our data set into a train and

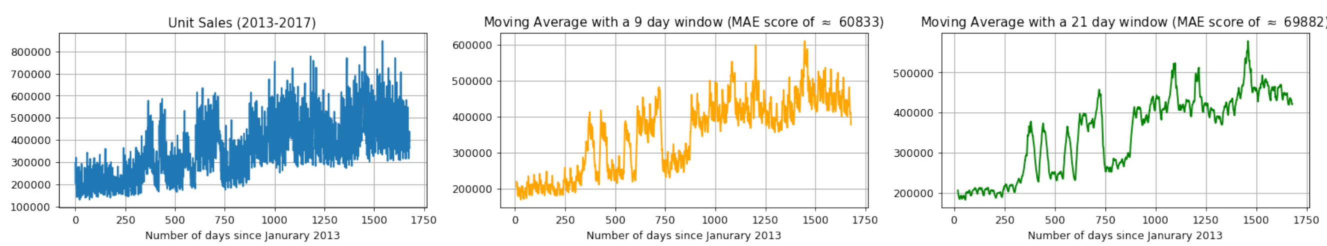

Figure 2. Moving Average with 9 and 21 day window

test subset. The iteration of the loss per epoch of the NN left us with an exponential decay trend leaving the MAE for the train data set to be 67366 and the MAE for the test data set to be 94221. Lastly for the RNN with an $LSTM$ cell, the data shows us that it overcomes the statistical baseline and massively outperforms the Neural Network. The MAE for the RNN was 43184 on the train data set and 51614 on the test data set.

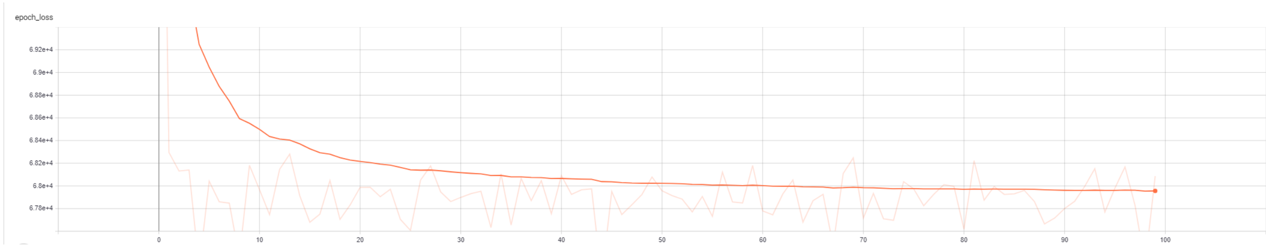

Figure 3. Loss Iteration over epochs

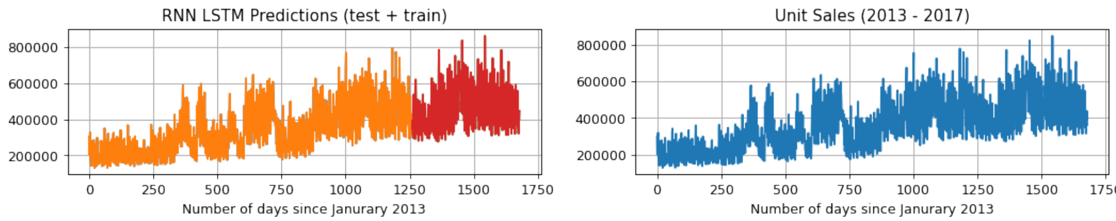

Figure 4. RNN Predictions

Conclusion

Summarizing our results, there are a few key findings that we will highlight. First, we observed that the FFNN dramatically under performed compared to the the statistical baseline. We believe this is the case for several reasons. For one, we implemented an univariate time-series model. This means that a lot of the extra information a neural network can digest was cut out of our data set due to time constrictions. Additionally, it was also not feasible for us to take time to fine tune the models which could have greatly improved its performance.

Secondly, the LSTM model demonstrated the most potential from the statistical baseline with an MAE improvement of 9,220 unit sales. As the most advanced model we tested, we expected the LSTM to be more suitable for the task since LSTM's are inherently spatially oriented, meaning that the order in which the data is in matters which makes sense for a prediction problem. Furthermore, LSTM's incorporate a memory element which allows the model to take into account previous iterations of the model. It is for these reasons we believe that models that incorporate a form of memory element will perform well on temporal data.

Future Work

To continue our work, we plan to explore several more models including a graph wavelet neural network. With our work on only a few models it is already clear that machine learning shows increased potential for this application. Additionally, we would like to rework our data set to implement more data in the form of a multivariate time-series which more complex models can take advantage of for state of the art performance.

References

1. Huber, J., Gossmann, A. & Stuckenschmidt, H. Cluster-based hierarchical demand forecasting for perishable goods. *Expert. Syst. with Appl.* **76**, 140 – 151, DOI: https://doi.org/10.1016/j.eswa.2017.01.022 (2017).

2. Tsoumakas, G. A survey of machine learning techniques for food sales prediction. *Artif. Intell. Rev.* **52**, 441 – 447, DOI: https://doi.org/10.1007/s10462-018-9637-z (2019).