Forecasting Product Demand from Historical Data Using Machine Learning Methods

Devon K. Reed[1 3] Ayodeji Odetola[2 3] Mahfuza Farooque[1] Suman Saha[1]

https://www.overleaf.com/project/5f15cef7d798b300017ce965 [1]Department of Computer Science [2]Department of Data Science [3]Department of Mathematics

Introduction

- In our global climate food waste has become an increasing issue. According to the Food and Agriculture Organization of the United Nations, global food waste is estimated at 1.3 billion tons of food.
- One of the most popular topics in the state of statistics and computer science is forecasting time-series models. We believe that an accurate time-series model could be used to accurately predict food sales and reduce waste.
- A time-series is any set of data points that are indexed in time order. For example, Temperature, Birth Rates, and Gross Domestic Product can be expressed as a time series.

Method

- Statistical Baseline
  - Moving Average (MA)
    We use the Moving average as our baseline for future models. The moving average (MA) uses a sliding window to take the average over a set number of time periods. For our case, it will be taking the average of our unit sale data points, $u_0, u_1, ..., u_{1678}$. If M is the size of our window then the moving average over our first time interval will be, $\bar{a}_{MA} = \frac{1}{M}\sum_{i=0}^{1677} u_{M-i}$
    We then calculate each succeeding MA using the following formula $\bar{a}_{MA} = \bar{a}_{MA_{prev}} + \frac{1}{1678}(u_M - u_{M-1678})$.
- Feed Forward Neural Network (FFNN)
  - We used a shallow feed forward neural network as our baseline ML 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.
- Recurrent Neural Network (RNN)
  - The RNN is a generalization of the FFNN with internal memory. The architecture for our RNN Model involves 1 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.
- Mean Absolute Error (MAE)
  - An important part of building these models is having a metric to test them by. We decided that the Mean Absolute Error (MAE) would be the best option for this project. MAE is defined as $MAE = \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y}_j|$

Data

We implemented the Corporación Favorita Grocery Sales data set from Kaggle to test our prediction models. 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.

Results

For our moving average statistical baseline, we observed the predictions with both 9 and 21 day windows (MA_9 and MA_{21}) respectively to see how the model would perform.

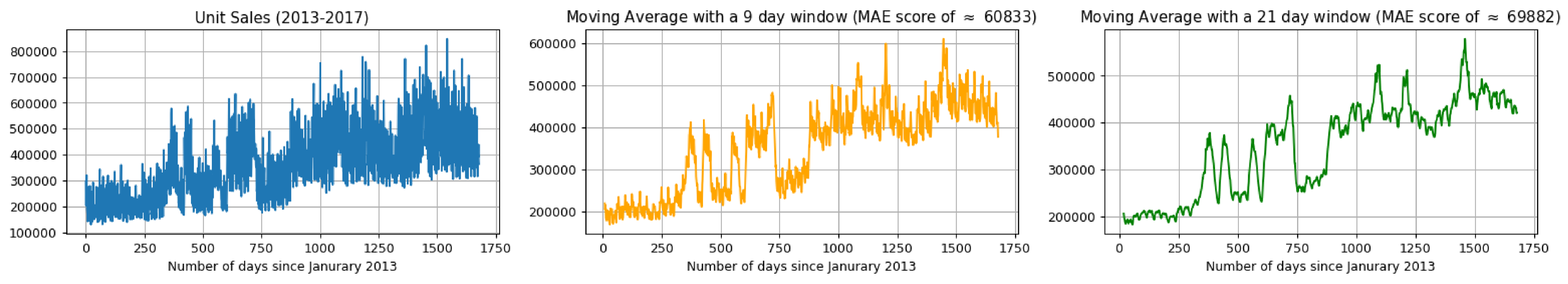

Figure 1: Moving Average Baselines

The mean absolute error for the 9 and 21 day moving averages were $MA_9 = 60834$ and $MA_{21} = 69883$ measured in unit sales. Next, we implemented a shallow Neural Network model as a naive machine learning baseline. Below is the loss iteration for the FFNN.

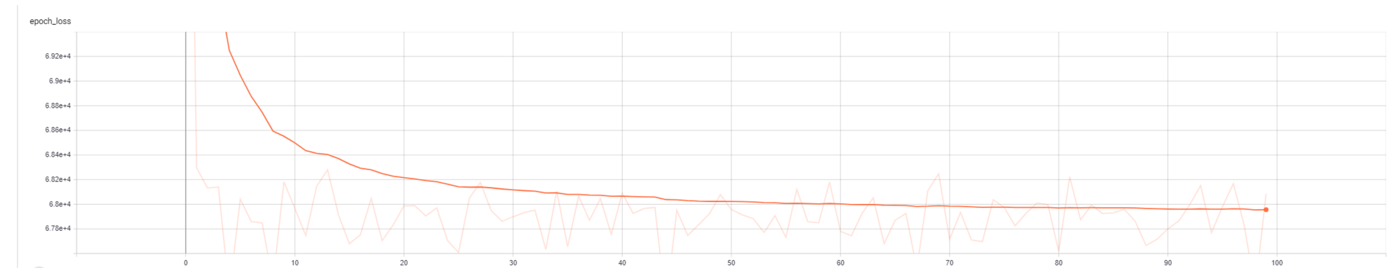

Figure 2: FNNN Loss iteration over epochs

The train MAE for the FFNN is 67366 while the test MAE is 94221. The final and most promising model we tested was a RNN with a LSTM cell. This outperformed both the MA and the NN by a significant margin. The train MAE was 43184 and the test MAE was 51614. Below we compare the LSTM's prediction of the train and test (orange and red respectively) data set and the actual unit sales amount.

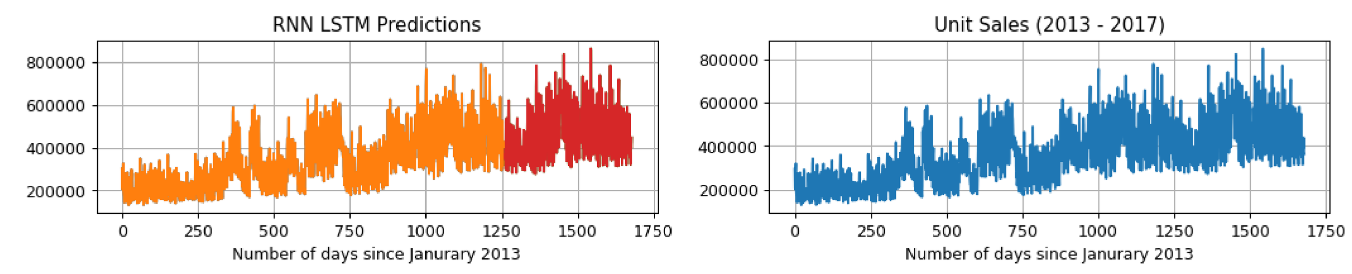

Figure 3: LSTM Prediction (Left) vs Actual Unit Sales (Right)

Conclusions

- Naive FNNN dramatically under performed when compared to the statistical baseline.
- LSTM showed the most improvement with an MAE improvement of 9,220 unit sales.
- Models that incorporate a memory element perform better on time-series data.

Future Work

- Represent more data in a multivariate time-series.
- Explore the application of a graph wavelet neural network on this prediction problem.

References

1 Puchalsky, W et al. (2018). I J of Prod Econ 203: 0925–5273.
2 Wang, H et al. (2018). I Sciences 438: 0020–0255.
3 Yi-Chung, H et al. (2017). PLoS One 12: 19326203.
4 Tsoumakas, G. (2019). Artif Intel Rev 52: 1573-7462
5 Huber, J et al. (2017). Exp Sys with App 76: 0957–4174.

Acknowledgements

We extend our gratitude to Erin Hostetler, Cindy Reed, and all those involved in allowing the MCREU program to take place this summer despite the unforeseen challenges that had to be overcome to make this experience possible.