

# Stock Prediction using LSTM on RNN

Gonsoo Moon  
Post Master Student of Computer Science  
University of Texas at Dallas,  
[gonsoomoon@gmail.com](mailto:gonsoomoon@gmail.com)

## 1. Introduction

The LSTM (Long short-term Memory) on RNN (Recurrent Neural Network) can be applied to a series of sequence data. A stock price is not exceptional, which is a time series data. I used one level of LSTM and fully connected layer using five input features: open, high, low, volume and close daily price of a sequence length of 3000 that is interpreted as a number of days that is difference from end date to end date. The toolkit Tensorflow 1.0.0 and python are used on the Mac pro.

## 2. Data Preprocessing

The input data is downloaded from Yahoo stock web site. Specifically, Amazon is from May 15, 1997 to Aug 2, 2017, Google is from Aug 19, 2004 to Aug 2, 2017 and Samsung Electronics is from Jan 4, 2000 to Aug 2, 2017. The close price is also used as target that is a label. It has five input features by day so that its shape is [n, 5]. Specifically, for Google data, data shape is [3262, 5]. If I use sequence length of 3000, its batch size is 262. The first batch is 1 to 3000 and second batch is 2 to 3001 and final batch is 262 to 3261. The 3262 row as the last was used to be target of a prediction of the final 262nd batch. Before making the splits of batches, I used the <sup>1</sup>min-max normalization. After the prediction, de-normalization function was used to revert to original stock price.

## 3. LSTM RNN Training

For training and validation, I used a ratio of 8 to 2, train and test batches. The architecture of one LSTM and one fully connected layer was used. The loss function is sum of the squares of (Target - Prediction). A 0.01 of learning rate, 10 of the hidden size, number of neurons, 3,000 of sequence length and a range of epoch from 1,200 to 2,000 were used.

## 4. Experiment

The <Figure1> shows RMSE (Root Mean Square Error) of Google with the hyperparameters. At the 2,000 of epoch, as

minimum, RMSE was measured as 12.48. The <Figure2> describes average absolute difference by companies. For Google, \$ 9.24 is absolute average difference compared with the real target price. The difference is about 1 percentage of the last stock price, \$ 930.39.

<Figure 1>

Sequence Length	Epoch	RMSE
3000	2000	12.48
3000	1500	13.34
3000	1200	14.321

<Figure2>

Company	Average Absolute Difference	Last Stock Price (Aug 2, 2017)
Amazon	8.64 Dollar	995.89 Dollar
Google	9.24 Dollar	930.39 Dollar
Samsung	33,646 Won	2,450,000 Won

The <Figure3> shows a target, prediction and difference on two axes for Amazon, Google and Samsung Electronics.

## 5. Analysis

The predictions have a pattern to follow the previous targets so that sudden sharp changes were not mostly predicted. However, steady targets were mostly predicted. The future work might be better to solve the problem, adding other features like sentimental analysis from news and blogs.

## 6. References

[1] Sung Kim, DeepLearningZeroToAll (2016), <https://github.com/hunkim/DeepLearningZeroToAll>, lab-12-5-rnn\_stock\_prediction.py

---

<sup>1</sup> Min-max normalization,  
[http://sebastianraschka.com/Articles/2014\\_about\\_feature\\_scaling.html](http://sebastianraschka.com/Articles/2014_about_feature_scaling.html)

< Figure 3>

