Formatting instructions for NIPS 2017

Forest Kobayashi
Department of Mathematics
Harvey Mudd College
Claremont, CA 91711
fkobayash@hmc.edu

Jacky Lee
Department of Mathematics
Department of Computer Science
Harvey Mudd College
Claremont, CA 91711
jaclee@hmc.edu

Abstract

Stock data is very difficult to analyze using classical methods, in large part due to heavy involvement of humans in stock pricing. In this paper, we examine a new approach to analyzing time-series stock data, particularly in the context of grouping correlated stocks.

1 Introduction

Here is a rough table of contents for our paper, so that the reader has a rough idea of where we'll be going:

1. Motivation (challenges to modelling stock data)
2. Long-term goal for the model / ideal model architecture
3. Classification
4. Findings
5. Conclusion

2 Motivation

Stock prices can be influenced by many factors such as other stock prices, anticipation of capital gains or losses, global politics, earnings reports, and etc. These factors may each have a different weighting on the stock depending on time and which stock is being analyzed in question. For example, earnings reports might have significantly more impact on stock prices immediately after release as opposed to two months after the report comes out. Another example would be that global politics might influence the stock price of **LMT** (Lockheed Martin), an aerospace and defense company, more than the stock price of **BKC** (Burger King), a fast food restaurant chain.

Ultimately, humans are the ones who choose to buy, sell, or short these stocks based on some analysis, usually not completely comprehensive, of the factors they observe. Since humans are often unaware or unable to analyze all these competing factors at once, their decisions are often unpredictable or at least very difficult to predict. As such, predicting the price of stocks by modeling these interactions between stock price factors and humans is difficult and unwise. Therefore, we have decided to perform stock analysis by only analyzing past stock trends. This approach, although loses out on a lot of information, is computationally less demanding and easier to find data for performing analysis.

2nd Course in Mathematics of Big Data (MATH189 2018), Claremont, CA, USA.

3 Goals

Our ultimate goal is to be able to use past stock data to train our model. Then given some stock trend data for some stock, use our model to predict the upcoming trends and return a vector of decisions. An example of the output is shown below.

$$\mathbf{y} = \left\langle P_{buy}(\mathbf{x}), P_{sell}(\mathbf{x}), P_{short}(\mathbf{x}), P_{nothing}(\mathbf{x}) \right\rangle$$

Here, $P_{buy}(\mathbf{x})$ represents how confident the model is in recommending the user to buy the stock. Similarly, $P_{sell}(\mathbf{x})$ represents selling the stock, $P_{short}(\mathbf{x})$ represents shorting the stock, and $P_{nothing}(\mathbf{x})$ represents doing nothing. All these probabilities should sum to 1.

4 Model Overview

Our large-scale goal for the model is as follows:

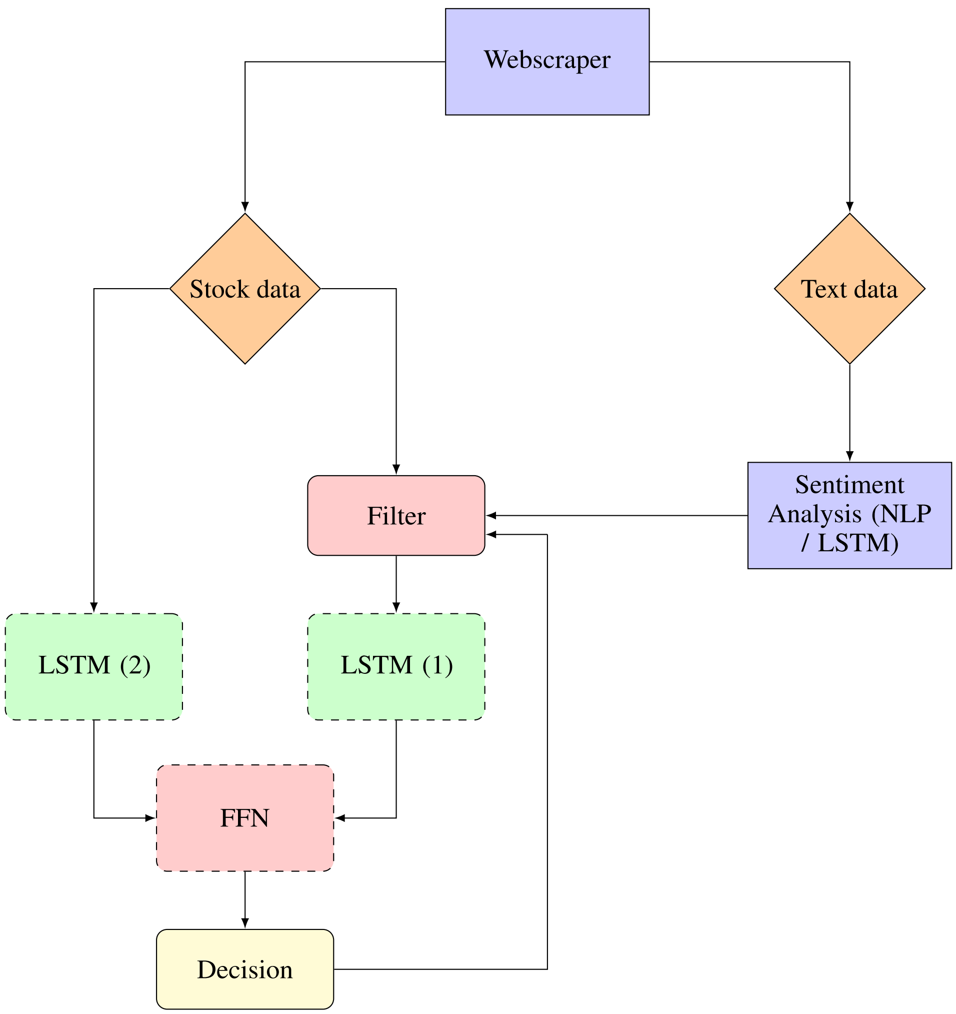

Figure 1: Model overview

5 Classification

Using LSTMs to predict stock pricing is, for the most part, a solved problem. Implementations and hyperparameters may differ, of course, but underneath, the structure of the model is largely the

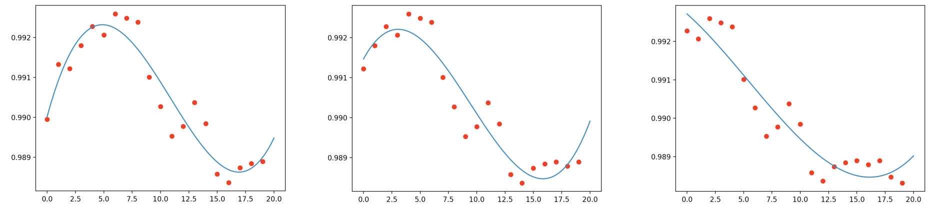

Figure 2: Sliding windows of data fitted with polynomials

same. Hence, we decided to focus our work on augmenting the filtering stage shown above. We were particularly interested in finding new representations of market trends that we could feed into the network.

5.1 Polynomial Fitting on Windowed Data

5.2 Frequency Analysis