Investigation into Stock Market Forecasting using Recurrent Neural Networks

Student: Oliver Crow, Supervisor: John Cartlidge, Project Type: Research

University of Bristol, Department of Computer Science

Overview

- Can RNNs (with LSTM) be used to generate accurate predictions of the stock market?
- Are the predictions reliable?
- How is accuracy and reliability of predictions affected when predicting more than 1 day into the future?
- What RNN configurations are optimal?
- How much data is required?
- What type of data best educates predictions (news, fundamentals, technical indicators, etc...)?

Project Outline

- Attempt to replicate findings and results from published articles
- Further improve findings by adding in additional items of data (e.g. news, fundamentals, technical indicators, etc...)
- Test model over configuration space
  - Number of neurons
  - Batch size
  - Epochs
  - Historical window size
  - Prediction days
- Reflect on results
- Investigate potential usefulness of predictions to assess predictive powers of the model

More Data

Giving the model a larger quantity and more diverse range of data can assist in making better predictions.

Unfortuantely, financial data (esp. historical) is very difficult and/or expensive. This is a large limitation in projects such as this one.

The types and quality of the data fed to the network can make or break the results.

Pre-processing is a step which needs to be performed to clean the data, and make it ready for the network (e.g. fill in missing values).

Textual data needs to be pre-processed heavily before being given to the network. This can include performing:

- Sentiment Analysis
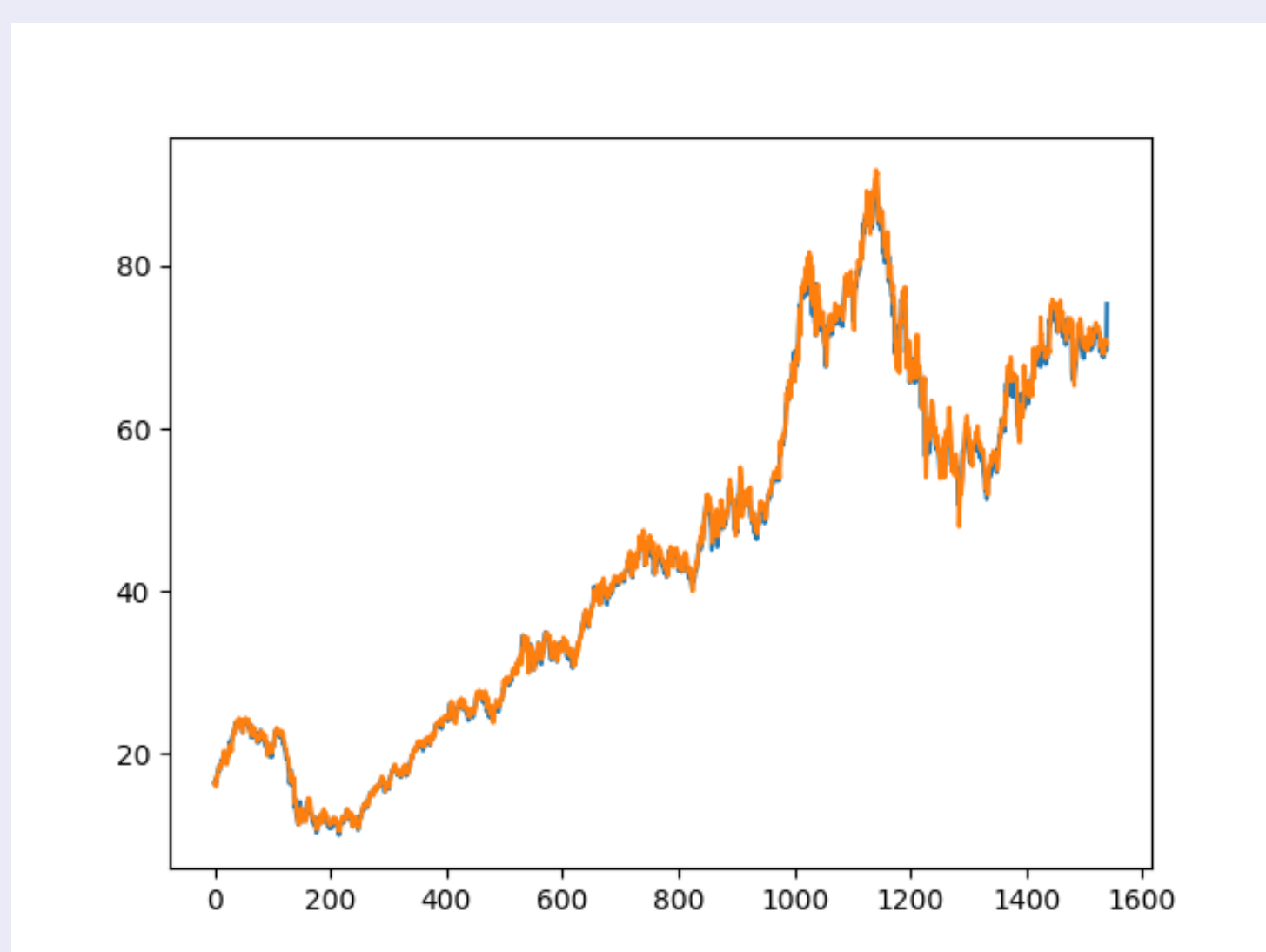- Word Embedding

Preliminary Results

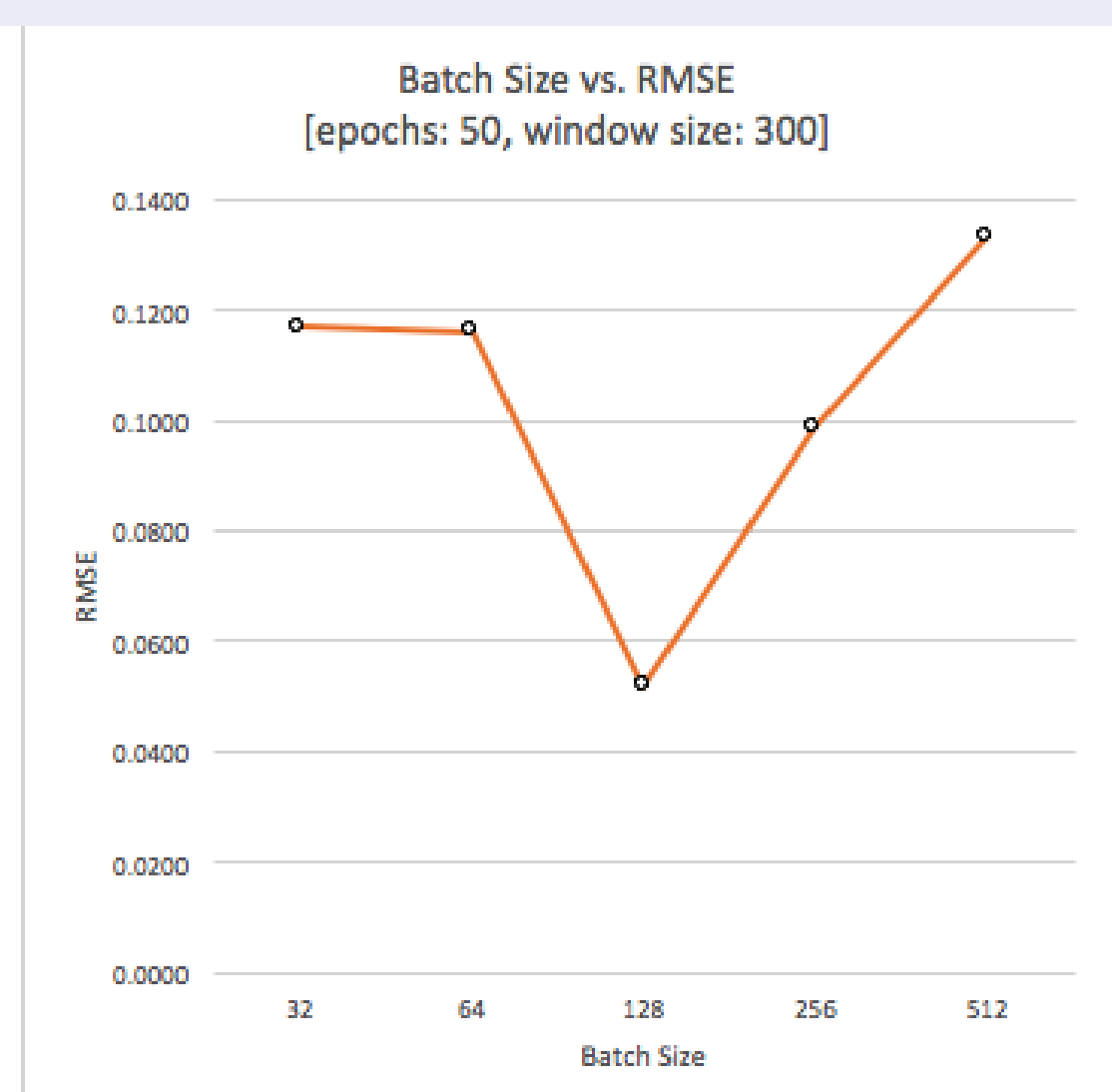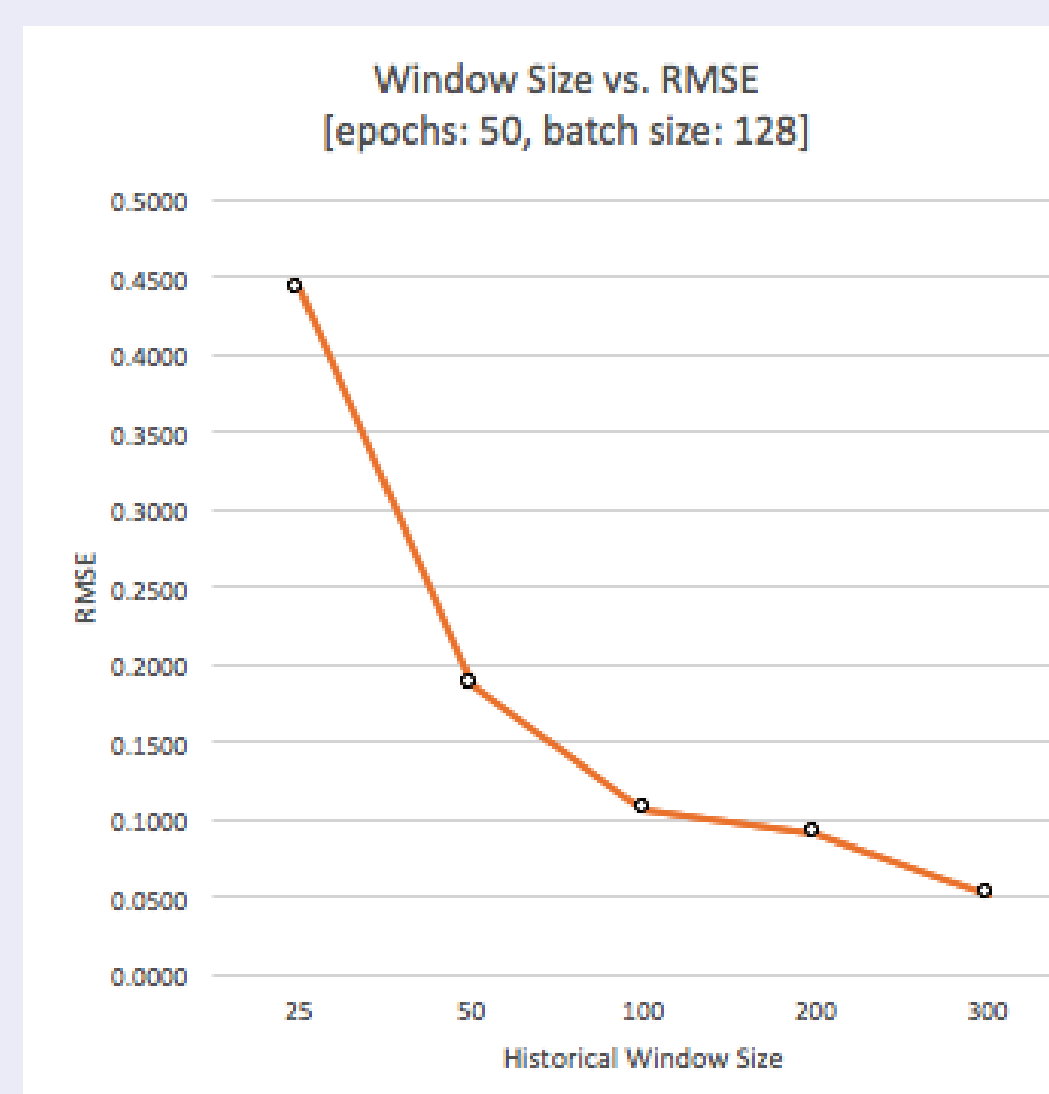Figure: Closing Price Prediction (1-day ahead, 300-day window) [prediction: orange, ground truth: blue]

Figure: Comparison of Configurations