
Predicting Stocks Trends Based on News

Harsh Dedhiya, Raghav Malhotra, Phumin Walaipatchara
Department of Computer Science
Boston University
hdedhiya@bu.edu, raghav20@bu.edu, phuminw@bu.edu

Abstract

To add abstract bibibikb

1 Background

Stocks market has been known to be volatile and sensitive to factors, including news and statistics. Speculation on stocks movement requires complicated techniques and models, still the result is not satisfactory.

1.1 Indicators

Nothing here for now

2 Dataset

Some text for second section

3 Naïve Bayes

Describe the result from Naïve Bayes approach

4 Logistic Regression

Describe the result from logistic regression

5 Sentiment Analysis

6 Recurrent Neural Network

In order to capture the objective of accurate prediction, Recurrent Neural Network (RNN) is introduced because its ability to exhibit internal state (memory). Specifically, Long short-term memory (LSTM), a special kind of RNN, is used for implementation as LSTM can deal with vanishing gradient problems and is capable of learning long-term dependencies.

Before begin training the model, the dataset, 100 samples in this case, must be preprocessed introduced boolean vector through *TfidfVectorizer* from *sklearn.feature_extraction.text*. The *max_features*, which is also the size of the boolean vector is specified to be 10. Each sample (now a boolean vector) is paired up with the stock (AMZN in this case) movement, 1 for going up and 0 otherwise. In one epoch, we use *KFold* from *sklearn.model_selection* to split the dataset

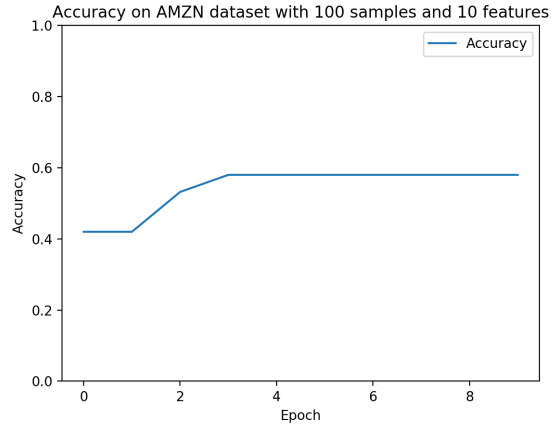


Figure 1: Accuracy of the model

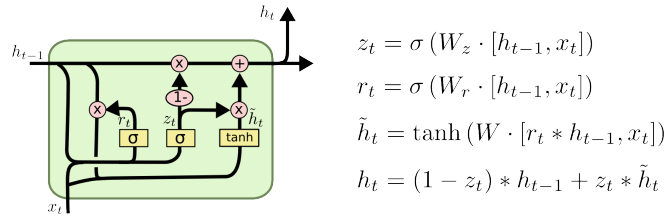


Figure 2: LSTM structure

into 2 groups, one for training and one for testing.

Regarding network structure, the LSTM network has 10 input nodes, which corresponds to the size of the boolean vector. It has 4 hidden nodes between the input layer and the LSTM and use sigmoid as an activation function. After LSTM module, the output layer consisting of 2 nodes uses softmax as an activation function in order to represent output as a probability of classification.

References

if needed