Improving Sentiment Analysis for Stock Trend Prediction

Scientific Advisors - Alexey Stankus, Evgeny N. Pavlovskiy

Novosibirsk State University

May 18, 2021

Introduction

- Stock market prediction is the act of trying to determine the future value of a company stock. Apparently, the successful prediction of a stock's future price can yield significant profit, making the prediction problem an area of strong appeal for both academic researchers and industry practitioners.
- Stock market prediction is usually considered as one of the most challenging issues among time series predictions due to its noise and volatile features.
- During the past decades, machine learning models, such as Support Vector Regression (SVR) and Support Vector Machines(SVM), have been widely used to predict financial time series and gain high predictive accuracy

Introduction

- The well-known efficient-market hypothesis (EMH) suggests that stock prices reflect all currently available information and any price changes based on the newly revealed relevant information.
- However, due to the implicit correlations between daily events and their effect on the stock prices, finding relevant information that contribute to the change of price for each individual stock are difficult.
- In this work, our goal is to leverage public released financial news and train a model named CNN-LSTM Hybrid to make prediction on directional changes for Standard Poor's 500 index(SP500)

Description of Data

- SP500 Historical Data of Duration from 20 March 2018 to 14 July 2020.
- Financial News from Reuters and Guardian of Duration 20 March 2018 to 14 July 2020.
- Stock Trend is calculated as 0 for decrease in Closing Price and 1 as increase in Closing Price.
- Technical Indicators such as Moving Average of Close Price for 3 days and Exponential Moving Average of Close Price is calculated.
- News for last 3 Days, Close Price for last 3 days along with Moving Average and Exponential Moving Average for last 3 days are used to predict stock trend rise of failure in Day 4.

Model Architecture

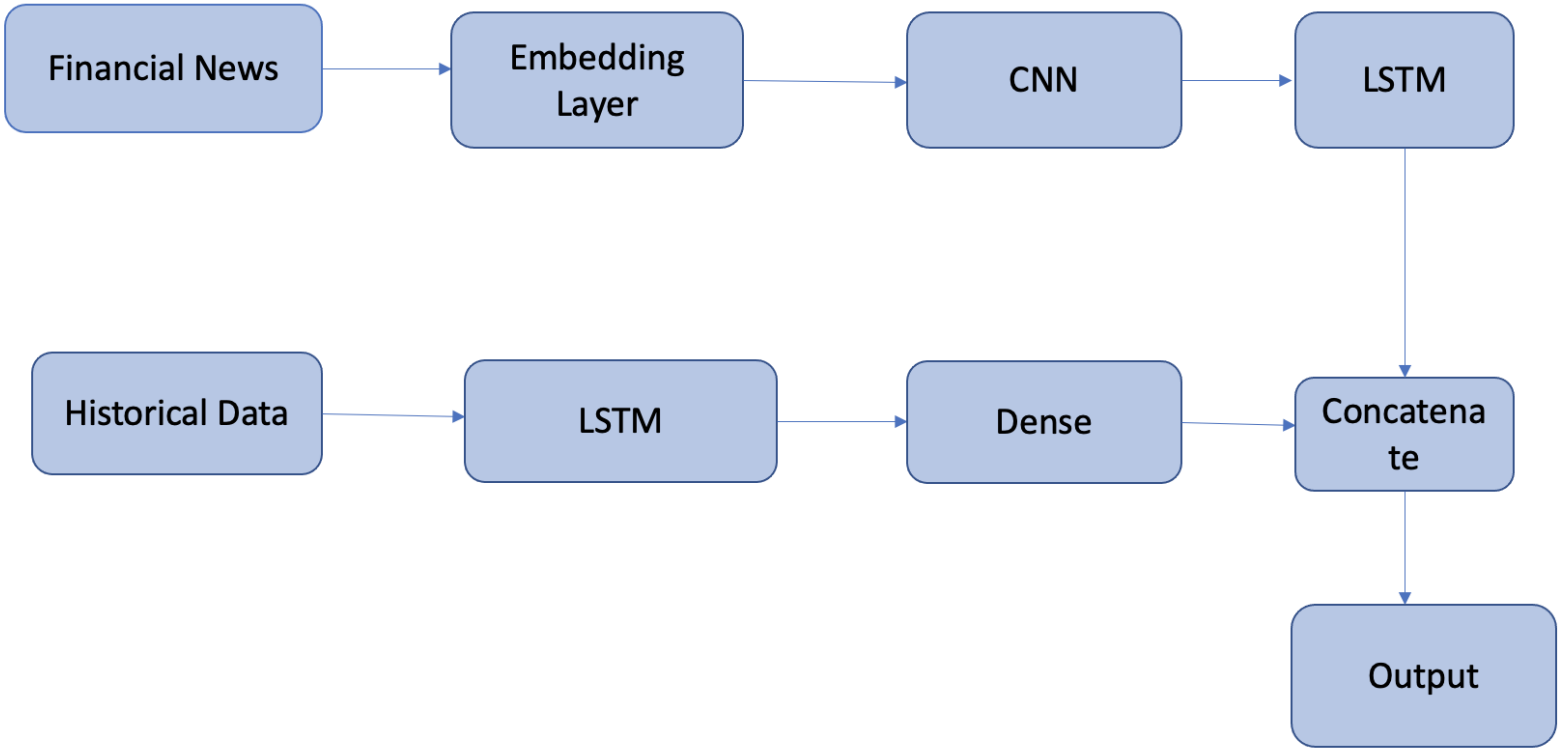

Figure 1: CNN-LSTM Hybrid Model

f

Trading Goal as Return Maximization

- Activation Function for LSTM Layers - Sigmoid
- Optimiser - Adam with learning rate 0.01
- Loss Function - Binary Crossentropy Loss
- Epochs - 100
- Training Data - 436 Days
- Testing Data - 109 Days

Baseline Results

MODEL	ACCURACY	F1 SCORE
RAND	0.509 ± 8e−4	0.502 ± 8e−4
ARIMA (Brown, 2004)	0.514 ± 1e−3	0.513 ± 1e−3
SVM	56.38%	NA
Bag-At-LSTM	61.93%	NA
Ding et al. (2014)	58.83%	NA
WB-CNN	60.57%	NA
Ding et al. (2015)	64.21%	NA
KGEB-CNN	**66.93%**	NA
CNN-LSTM (Our Results)	66.05	0.7752

Figure 2: Results Comparison

THANK YOU