Trading on Sentiment:

Using News and Twitter Sentiment to Better Financial Predictions

Siyu Liu, Zinan Chen, Qiuhao Chengyong, Tyler McMurray and Senbo Zhang

Abstract

Greed and **fear** are the two main drivers of the stock market. It turns out that positive and negative emotions in textual data, such as Twitter and News, can be used to predict daily changes or trends in stock prices.

Dataset

Twitter finance tweets of Nasdaq 100 component stocks from March 28, 2016 to June 15, 2016

Stock closing price:
Data pulled from Yahoo Finance

Historical news headlines from Reddit WorldNews Channel. They are ranked by Reddit users' votes, and only the top 25 headlines are considered for a single date

Reference

https://finance.yahoo.com
https://followthehashtag.com/datasets
https://www.kaggle.com/aaron7sun/stocknews

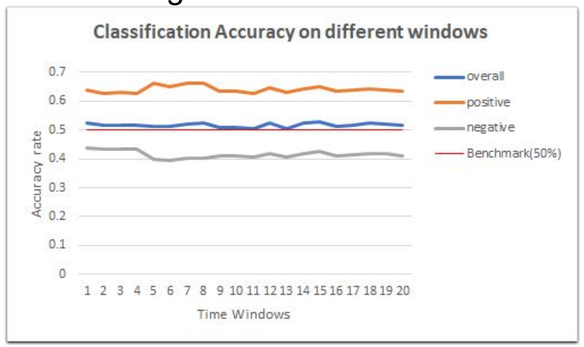
Our Goal and Plan

Analyze and predict stock price or make a Buy/Sell decision by using time series and text sentiment

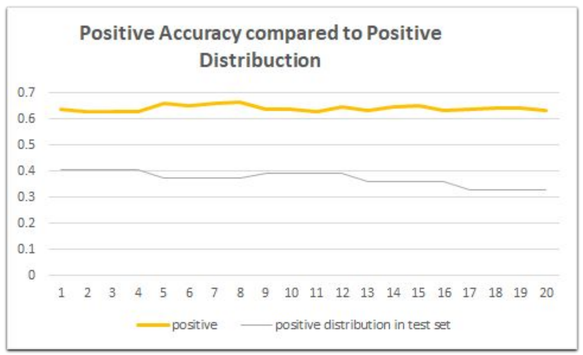
- NLP techniques of various data sources
  - Sentiment polarity score
  - Bert method
- Time series predict the stock price
  - LSTM model

Classification +Sentiment

- Random forest model perform bet with 7-days windows.
- Overall accuracy all higher than 50% but not significant



- Model have higher accuracy when predicting stock increase.



Portofolio Management

- Found simple linear regression with 8 day window had smallest errors for regression
- Combining regression model with classification outputs created best result
- Winning Portfolios beat 10 Random Portfolios and average of the market

Name	Gains_Losses	Average	Beat_Winner	Beat_Aggregate
Winner	0.220	0.044	--	Yes
Basket1	-0.249	-0.050	No	No
Basket2	0.055	0.011	No	Yes
Basket3	-0.153	-0.031	No	No
Basket4	-0.018	-0.004	No	Yes
Basket5	-0.345	-0.069	No	No
Basket6	-0.472	-0.094	No	No
Basket7	-0.400	-0.080	No	No
Basket8	-0.076	-0.015	No	Yes
Basket9	-0.313	-0.063	No	No
Basket10	-0.205	-0.041	No	No
*Numbers were scaled			0%	50%

Time Series: LSTM Model

- Level - Average stock price
- Trend - Increasing or Decreasing
- Seasonality - repeated cycle

LSTM:
Three gate help to store import information and forget unimportant information to make prediction more accurate.

Advanced model: LSTM + Sentiment

- First step: Create basic LSTM model
- Second step: Add extra feature "news sentiment score" to the model
- Result: Prediction fit the actual trend better
- **RMSE** reduce from 120 to **71**

Conclusion

- 60% of random basket portfolios beat winning portfolio predictions using just regression
- When including sentiment classification predictions with regression new winning portfolio beat all random baskets and the average of the market
- Utilizing alternative data sources and types greatly improve financial predictions

Acknowledgements

Professor. Mohammad Soltanieh
Professor. Brock Tibert
Professor. Peter Wysocki