Predicting Stock Price Movement Using Social Media Analysis

Dušan Srbulović, Filip Mićić

1. Motivation

Social media platforms provide a wealth of information on real-world patterns and behaviors. This project analzyes aggregated message data from StockTwits, a popular online investor platform, to approach the financial problem of stock price prediction.

2. Research question

We performed analysis on the component stocks of multiple companies. Data was collected for period Jan 2014 to Dec 2016. Two main datasets were used:

- Daily split-adjusted price data, collected via Yahoo finance API

- SotckTwits message data collected and downloaded in raw json fromat, approximately of 100 000 messages

Preprocessing was necessary to generate "bag-of-words" feature vectors, including removing stop-words and company names, removing posts mentioning multiple stocks.

The 2 day future-return, calulated as price difference was used as the prediction target to model the short -term correlation with social media acitvity. A shorter period was not selected to reduce the effects of market noise.

3. Metodology

We use thsi good "off-the-shelf" learning algorithm for regression, which is similar to how SVC finds a decision boundary that maximizes the margin, aims to minimize the e-insensitive loss function created by Vladimir Vapnik.

L2 normalization was used and the SVR e parameter was optimized with 5-fold cross validation.

4. Discussion

Test accuracy for our model is:

SVR - 56% accuracy

For pracical trading purposes, however, the more important metric is profitability. The model prediction for the positive and negative classes were used to generate profit on the test data. If investor trust our model and he invest every time our model precit price growth, he will be in constan profit.

Overall, the regression model proved to be more accurate and more actionable as trading signals compared to other models that we made research on as Pure bag of model and Naive Bayes classification. The test errors rates were below what this project initially aimed to achive, however, the signal's positive performance indicates that the selected features are in fact meaningful, and capture some insight into short-term market movements.

6. Reference

Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques <https://www.sciencedirect.com/science/article/pii/S0957417414004473>

Sentiment analysis on social media for stock movement prediction <https://www.sciencedirect.com/science/article/pii/S0957417415005126>

Short term forecasting with supported vector machines and application to stock price prediction

<https://www.tandfonline.com/doi/abs/10.1080/03081070601068595>

Twitter as a tool for forecasting stock market movements <https://www.sciencedirect.com/science/article/pii/S2405918817300247>