Stock Market Trends Prediction after Earning Release

Chen Qian, Wenjie Zheng, Ran An (cqian23 / zhwj814 / anran)@stanford.edu

Introduction

Public companies release quarterly earnings report to present financial status to investors. Short term stock prices can be significantly affected by company’s quarterly performance. Our project aims to explore the impact of financial report factors, earning per share (EPS), consensus forecast and the earning report related articles on short term stock market trends. We narrow down our targeted company to Silicon based public technology companies.

Data Selection

- **Financial factors**
  - Total asset
  - Total equity
  - Total revenue
  - Net income

- **News and analytic articles**
  We would like to analyze if these articles from mainstream media could influence stock market and investor’s trading strategies.

- **EPS vs Forecast**
  How much the actual EPS bits consensus expectation is also a key factor to be considered for investment strategy.

- **Training labels**
  Label 1: Next day opening price vs release date close price
  Label 2: Next day closing price vs release date close price

Data Preprocessing

- **Relativisation**
  Calculate the growth rate of each financial factor from last quarter, allows us to combine companies data together.

- **Sentiment Analysis**
  Interact with Stanford NLP toolkit to obtain sentiment analysis matrix from news and articles.

- **Normalization**
  Normalized input features simplifies model design.

Features

- **Relativized financial data**
  - EPS surprise
  - Total asset
  - Total equity
  - Total revenue
  - Net income
  - Total asset: 15.30%
  - Total equity: 5.52%
  - Total revenue: 7.82%
  - Net income: -3.45%

- **Normalized sentiment analysis features**
  - Very negative: 2.564%
  - Negative: 8.974%
  - Neutral: 62.820%
  - Positive: 20.512%
  - Very positive: 5.128%

Data Visualization

- **Financial features from earning report**
- **Sentiment features from articles**

Model Selection & Evaluation

<table>
<thead>
<tr>
<th>Model</th>
<th>Equation</th>
<th>Parameters</th>
<th>Smoothing Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Locally Weighted LR</td>
<td>$\beta = (X'WX)^{-1}X'WY$</td>
<td>Bandwidth $T$</td>
<td>$D = ae^{-\gamma T^2}$</td>
</tr>
</tbody>
</table>

Results

- **All error analysis are validated using K-Fold cross validation** ($K=5$)

Discussion

Result: Stock market is known as a chaotic system and our model analysis depicts that even building our model with empirical key features could still result in low accuracy. However, by limiting our scope to earning release day, we are able to build the prediction model of around 70% accuracy.

Models and features: SVM seems to give us a better prediction than NB and other generalized LR algorithms. Further, it is observed that data visualization from different companies shows that data set from a certain company is much more distinguishable than mixing data together. The commercial correlations between companies are far less than previous expectation and the features we collected are far from enough to predict the stock trend.

Limitation: Due to the limited number of company choices, we have small data size (~300 samples) which could lead to high bias and over-fitting. Stock price is not only affected by certain financial features, consensus news, but also company direction and future business guidance, which are difficult to be digitized.

Future

Feature selections: Write script to auto collect/process data to increase the learning set size and use feature ranking to analysis the covariance between features and labels and pick the most top ranked one as the model inputs.

NLP analysis: The sentiment result from Stanford NLP toolkit is too general to apply for financial news analysis and we would spend more time to develop a financial specific sentiment analysis algorithm.

Model Improvement: We would spend more time on SVM algorithm and explore the most appropriate kernel to separate features in high dimension.