

Machine Learning Analysis of Company Earnings Releases

Team

Thomas Ulrich (tmulrich@stanford.edu)
Chaz Pratt (cp3@stanford.edu)
Philipp Thun-Hohenstein (pthun@stanford.edu)

Predicting

The stock market is incredibly complex system offering great financial returns to those who can predict its behavior.

Therefore, as investors, we try to acquire all available data about companies of interest and use this information to make decisions about whether to buy or sell stock. But companies usually prefer to communicate with the public only in carefully managed press releases. Therefore, we wondered if the published transcripts of CEOs' unscheduled quarterly conference calls with investors could be useful for predicting future stock prices. Traditionally, such information has been processed by human analysts; but could machine learning automate the process and help us find overlooked stocks with high potential?

Data

We downloaded a large number of earnings call transcripts from the Factiva database. Stock prices were obtained using the Yahoo! Finance API. We ran the analysis based on the difference in stock prices over several different time periods after the release of the earnings transcripts, including 3 months, 1 month, and 2 weeks.

Features

For features, we used the frequencies of words in the earnings transcripts.

Models

Following the Naïve Bayes algorithm, we assume the word frequencies are independent given y . That is,

$$\begin{aligned} p(x_1, \dots, x_{50000}|y) &= p(x_1|y)p(x_2|y, x_1)p(x_3|y, x_1, x_2) \cdots p(x_{50000}|y, x_1, \dots, x_{49999}) \\ &= p(x_1|y)p(x_2|y)p(x_3|y) \cdots p(x_{50000}|y) \\ &= \prod_{i=1}^n p(x_i|y) \end{aligned}$$

Here y is 1 if the stock price rose and 0 if it fell. We then use Bayes' rule to predict the direction of other stock prices.

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)}$$

Results

We consistently obtained about 70-75% accuracy in predicting the trend direction of stock prices in our test sets. But the scale of the profits we were able to generate in our tests was surprising. Our model performed best if we made all trading decisions within a relatively short period after the release of the earnings call transcripts. Interestingly, a Bernoulli Naïve Bayes model consistently performed a bit better than our multinomial model. That's probably because of the relatively short pieces of text we considered.

Table of results

Time Period	APY (multinom.)	APY (Bernoulli)
2 weeks	31.0%	33.5%
1 month	20.7%	21.3%
3 months	8.6%	16.3%

Discussion

Obviously, the historical performance of our proposed investment strategy is outstanding, easily beating the rest of the stock market. It appears that the stock market reacts to earnings calls much less efficiently than we had expected. However, we suspect that if this algorithm works, other financial firms have already noticed it and the results simply aren't showing up due to our use of older data. Therefore, it may no longer be possible to make significant money by using this as an investment strategy.

Future directions

Our original plan was to use a more complicated natural language processing scheme for this project, but we moved away from this due to technical difficulties and the outstanding performance of the simple Naïve Bayes strategy. However, applying sentiment analysis techniques to the earnings call transcripts would still be interesting. Our approach could also be extended by using the actual differences in stock prices rather than our binary direction indicator.