What is the relationship between news and stock price trends?

TIANXIN ZHENG
MACSS, University of Chicago
06/02/2019

Background

1. News and Financial Markets

- Market is not predictable:

• Efficient Market Hypothesis (Fama, 1964)

• Random walk theory (Malkiel, 1973)

- Empirical studies conducted on the relationship between news and financial markets

• Engelberg et al. (2011) - analyzed the causal impact of media in Financial Markets by analyzing "behaviors of investors with access to different media coverage" (Engelberg et al., 2011, pp 4) and proved that media coverage could strongly predict trading locally.

• Birz et al. (2011) - chose newspaper stories as their measure of news to investigate the impact of macroeconomics news on stock returns. Specifically, they found that news about GDP and unemployment could influence stock returns.

2. Natural Language based Financial Forecasting

- More new researches focusing on the analysis of the news content on stock volatilities and returns.

• Uhl (2014) - investigated into the relationship between Thomson Reuters Datastream and stock returns and proved there is correlation between them.

• Yoshihara et al. (2016) - used bag-of-words approach to represent news text and then feed it into a recurrent neural network and RBM model to investigate into the temporal properties of news events for stock market prediction.

• Ding et al. (2015) - utilized an event-embedding approach to extract events from news text data and then used a deep convolutional neural network to build a model that could both predict short-term and long-term influence of events on stock market trends.

• Schumaker et al. (2009) - built a text analysis system using bag of words, Noun Phrases and Named Entities as different word representations to conduct news analysis and predict for the discrete stock price.

Specific Aims

My paper will focus on the comparison of different word vectorization techniques and machine learning methods and investigate into the prediction power of different approaches.

Data Overview

- Data Source

• News text data from Reddit WorldNews Channel by Web Scraping

• Dow Jones Industrial Average (DJIA) data from Yahoo finance

- Data Preprocessing

Aggregate the top 25 news headlines into a single text string

↓

Regular Expression to remove some garbled characters

↓

Label the DJIA data for binary classification

↓

Split the dataset into training set and test set

Methods

• Converting Text Data into Vectors

- Bag-of-words model
- N-grams model
- TF-IDF model

• Machine Learning Methods

- Logistic regression
- Support Vector Machine
- Random Forest
- Naïve Bayes

Comparison of different word vectorization methods and machine learning methods

	Bag-of-words	Bigrams	Trigrams	TF-IDF
Logistic Regression	0.8360	0.8412	0.8624	0.8174
SVM Linear	0.8254	0.8598	0.8598	0.8307
Random Forest	0.8439	0.8571	0.8519	0.8307
Naïve Bayes	0.8228	0.8254	0.8254	0.8228

Table 1 Prediction accuracy with different word vectorization methods and machine learning methods

Conclusion

News with positive sentiments helps increase the stock price and news with negative sentiments helps decrease the stock price.

- Word vectorization methods:

• Bigrams and trigrams model work better on news text data than the simple bag-of-words model as they take contexts into consideration.

• When comparing between bigrams and trigrams, neither one outperforms the other in all situations.

• TF-IDF doesn't work as well as the other methods.

- Machine learning methods,

• Random forest has the best accuracy with bag-of-words model while support vector machine outperforms other methods with bigrams and trigrams.

Future Improvements

- More data sources

- Deep learning methods

- Including the effect of news into time series model

Acknowledgements

I sincerely thank Dr. Richard Evans and all my fellow students in the MACS 30250 class for your helpful comments and suggestions.

Contact Info

Email:

txzheng@uchicago.edu

Github:

https://github.com/TianxinZheng

References

Fama, E., The Behavior of Stock Market Prices, in Graduate School of Business. 1964, University of Chicago.
Malkiel, B.G., A Random Walk Down Wall Street. 1973, New York: W.W. Norton & Company Ltd.
Engelberg, J. E. and Parsons, C. A. (2011), The Causal Impact of Media in Financial Markets. The Journal of Finance, 66: 67-97.
Gene Birz, John R. Lott (2011), The effect of macroeconomic news on stock returns: New evidence from newspaper coverage, Journal of Banking & Finance, Vol 35, Issue 11, page 2791-2800.
Uhl M (2014) Reuters sentiment and stock returns. J Behav Finance 15(4):287–298.
Yoshihara A, Seki K, Uehara K (2016) Leveraging temporal properties of news events for stock market prediction. Artif Intell Res 5(1):103– 110.
Ding X, Zhang Y, Liu T, Duan J (2015) Deep learning for event-driven stock prediction. In: International joint conference on artificial intelligence.
Robert P. Schumaker and Hsinchun Chen (2009). Textual analysis of stock market prediction using breaking financial news: The AZFin text system. ACM Trans. Inf. Syst. 27, 2, Article 12 (March 2009), 19 pages.