

Event-Driven Stock Prediction using NLP

Anant Jain, Ilya Yudkovich

CS6120 NLP Spring 2019
jain.anan@husky.neu.edu,
yudkovich.i@husky.neu.edu

1 Introduction

It has been stated that the price of a security reflects all of the information available, and that everyone has a certain degree of access to the information: Efficient Market Hypothesis (EMH) [Fama 1965]. Today, a lot of new information related to companies and corporations listed on various stock exchanges like NASDAQ, NYSE and AMEX appears constantly on the web, making an instant impact on the companies' stock value. Surveilling all this information in real time is crucial for big trading firms but is not very much in reach of a lone investor. In this project, we present a news surveillance system targeting financial news to predict its impact on some of the top market cap companies. We aim to tackle the challenge from the position of a single investor without any access to real time trading infrastructure. The input to the system are the financial news articles indicating an event for a particular company and the output is a confident suggestion whether the user should buy or sell his share of stocks for that company.

Examples:

Ticker: AAPL.O

News: "A U.S. federal judge has issued a preliminary ruling that Qualcomm Inc owes Apple Inc nearly \$1 billion in patent royalty rebate payments, though the decision is unlikely to result in Qualcomm writing a check to Apple because of other developments in the dispute."

Prediction: "Buy"

Critique: The above news clipping instantly puts AAPL in good light even though it's unlikely to profit from the dispute. Nevertheless, it won't be a bad decision to Buy.

Ticker: AMZN.O

News: "PARIS Casino's upmarket Monoprix supermarket chain is working to expand its partnership with E-commerce giant Amazon in France, following a successful launch in Paris, Monoprix's Chief Executive said on Thursday."

Prediction: "Strong Buy"

Critique: Definitely a strong buy. This is a no brainer.

Ticker: MSFT.O

News: SAN FRANCISCO Some Microsoft Corp employees on Friday demanded that the company cancel a \$480 million hardware contract to supply the U.S. Army, with 94 workers signing a petition calling on the company to stop developing "any and all weapons technologies."

Prediction: "Sell"

Critique: Above news definitely puts Microsoft in legit pressure. Selling is the right action to take here.

2 Related Work

Despite years of research and studies, there are still debates about what information best predicts the volatility of the stock market. In the field of Artificial Intelligence, there have been previous attempts in designing models based on historical and time series data [Taylor and Xu, 1997; Andersen and Bollerslev, 1997; Taylor, 2007]. However, these methods do not account for one of the key sources of market volatility which can have dramatic effect on a security's share i.e. financial news [Cutler et al., 1998; Tetlock et al., 2008; Luss and d'Aspremont, 2012; Xie et al., 2013; Wang and Hua, 2014].

Recent improvement in computing power and introduction of new NLP techniques enabled to address this issue and make use of financial news. Since then, there have been many attempts to boost the accuracy of market predictions using language features, such as, tree representations of information [Xie et al., 2013], identification of expert investors and risk based on financial reports [Kogan et al. 2009]. The paper by Engelberg (2008) discussed how linguistic information in the text has a greater long-term predictability for prices than using term frequencies. Although very useful, these techniques do not account for structured relations in text which limits their potential.

Structured representations can be found using open information extraction (Open IE) tools and take semantics in account, but that leads to increased sparsity which in turn limits the predictive power.

To this end, this project focuses on improving upon the works of [Luss et al. 2012] and [Ding et al. 2015] and use event embeddings for learning. Event embedding are dense vector matrices which are trained such that similar events would have similar matrices even if they don't comprise of common words.

Along with the traditional deep natural language processing methods like using convolution neural nets and bidirectional gated recurrent units, we also experiment with pre-training of deep bidirectional transformers for language understanding with Google AI's BERT¹.

¹ <https://arxiv.org/abs/1810.04805>

3 Methodology

3.1 Dataset Construction

We created a large-scale financial news dataset by crawling Reuters.com. The dataset comprises of news for the past 1000 days for the top 37% market-cap companies (~2590 tickers) and is stored in a highly compressed pickle of a pandas dataframe. As expected, higher market-cap companies have more news as compared to lower market-cap companies, but since we are not concerned with the companies themselves but the type of events they undergo, we discarded ticker information from the data after aggregating the news by ticker and date (since a company can have multiple news articles in a single day).

To build the labels for each event, we crawled historical stock prices (Open, Close, Adj Close, Date) for each ticker from Yahoo Finance and calculated their relative return w.r.t. S&P for that day. Further depending on the relative return value, we segregated the returns into categories (Strong Sell, Sell, Buy, Strong Buy) using quartile ranges and one-hot encoded them.

3.2 Preprocessing Text

We normalized the text, eliminated stop words and filtered the dataset such that it is only left with words which appear in the vocabulary comprising of the top 20000 most frequently repeated financial words obtained from the Reuters corpus.

Next, in order to process the embeddings on a laptop, we limited the maximum sequence length per event to 512 (something our laptops can handle) and padded the shorter sequences. This step ensured that we have a consistent 2D tensor for each event.

3.3 Feature Engineering

As explained in section 2, we're using event embedding which are dense vector matrices and are trained such that similar events would have similar matrices even if they don't comprise of common words. Event embedding are largely known for capturing semantic information in the text.

For tradition deep nlp models like convolution neural net and bidirectional gated recurrent unit net, we prepared an embedding matrix for words in our financial vocabulary using the 42 billion version of pretrained glove² embeddings each containing 300 dimensions.

For BERT, the embeddings were calculated as the part of the tokenization process using the vocab.txt file which comes as the part of the pretrained model³ and is used to map WordPiece to word id.

² <https://nlp.stanford.edu/projects/glove/>

³ <https://github.com/google-research/bert>

4 Experiments

4.1 Experimental Setting

	Total	Training	Validation
percentage	100	80	20
#events	38,386	30,637	7,749
#words	2,163,758	1726270	437488

Table 1: Statistics of datasets

4.2 Evaluation Metrics

Following Luss et al. [2012] and Ding et al. [2015], the standard measure of accuracy (Acc) and Matthews Correlation Coefficient (MCC) are used to evaluate our experiments.

Since our classes are perfectly balanced (cut by quartiles), accuracy is more than fair metric to be used for evaluation. In addition, MCC measures the quality of the classification irrespective of class balance (just like F1 metric).

4.3 Baselines and Proposed Models

We use two state-of-the-art stock market prediction systems based on financial news from reuters as the baselines: Luss and d'Aspremont et al. [2012] propose using bags-of-words to represent news documents, and constructing the prediction model by using Support Vector Machines (SVMs). Ding et al. [2015] report a system that uses neural tensor networks to learn event embeddings for representing news documents, and build a prediction model based on a deep CNN.

Similar to Ding et al. [2015], we use two kinds of deep neural tensor networks to learn event embeddings for representing news documents:

1D-CNN-GMP: GloVe event embeddings input and 1-dimensional convolutional neural network prediction model with global max-pooling.

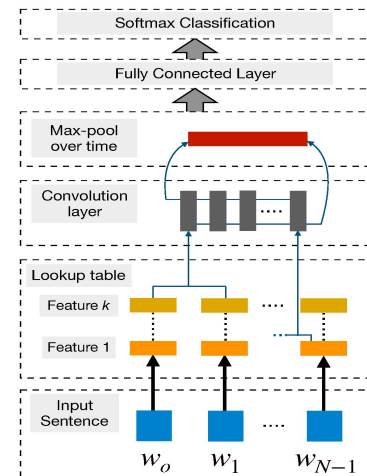


Figure 1: 1D-CNN-GMP Architecture

BI-DIR-GRU: GloVe event embeddings input and bi-directional gated recurrent unit prediction model

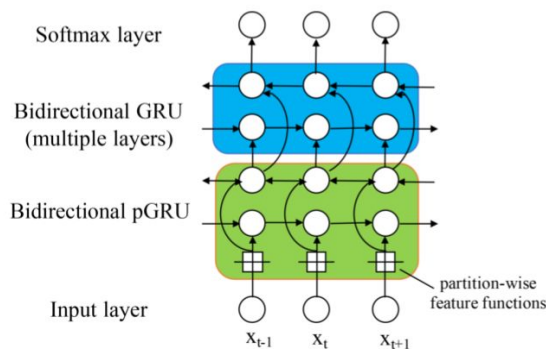


Figure 2: BI-DIR-GRU Architecture

Along with deep neural tensor networks, we also experiment with pre-training of deep bidirectional transformers for language understanding with Google AI’s BERT.

BERT: WordPiece event embeddings input and BERT-Base, Uncased (12-layer, 768-hidden, 12-heads, 110M parameters) fine-tuned prediction model

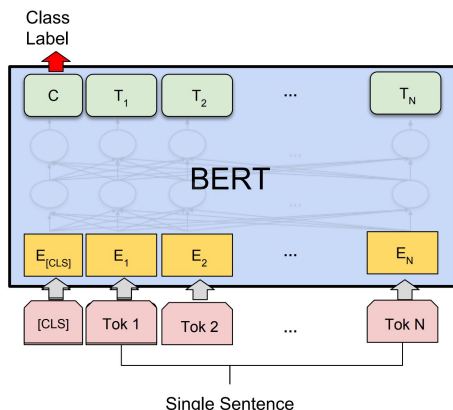


Figure 3: BERT Single Sentence Classification Architecture

4.4 Final Results

To make detailed analysis, we constructed the following five models:

Model	Accuracy	MCC
Luss and d’Aspremont [2012]	56.42%	0.0711
Ding et al. [2015]	65.08%	0.4357
1D-CNN GM (this project)	64.43%	0.2918
BI-DIR-GRU (this project)	63.34%	0.2711
BERT (this project)	66.43%	0.3010

Table 2: Development results

5 Conclusion

We demonstrated that deep natural language processing is useful for event-driven stock price movement prediction by proposing novel neural tensor networks and fine-tuned deep bidirectional transformers for language understanding using event embeddings for short-term news events. Experimental results showed that BERT architecture can give more accurate event-based predictions than deep neural tensor networks. In addition to better performance, BERT also cut the training time it took to train neural tensor networks in half.

6 Future Work

We can construct a better label by calculating relative return by comparing the stock price of the company and the corresponding industry, instead of comparing everything with S&P 500. It is almost like hedging, as long as an investor knows which company does well in some specific industry, he can make a decent prediction.

Also, we plan to experiment with financial news from other platforms like Bloomberg. Different platforms can reflect different opinions and can have an effect on the degree of prediction for better or worse.

Last but not the least, we plan to give more time to BERT for the fine-tuning task and see if the evaluations improve.

7 References

- [Fama, 1965] Eugene F Fama. The behavior of stock-market prices. The journal of Business, 38(1):34–105, 1965.
- [Taylor and Xu, 1997] Stephen J Taylor and Xinzhong Xu. The incremental volatility information in one million foreign exchange quotations. Journal of Empirical Finance, 4(4):317–340, 1997.
- [Taylor, 2007] Stephen J Taylor. Modelling financial time series. 2007.
- [Cutler et al., 1998] David M Cutler, James M Poterba, and Lawrence H Summers. What moves stock prices? Bernstein, Peter L. and Frank L. Fabozzi, pages 56–63, 1998.
- [Tetlock et al., 2008] Paul C Tetlock, Maytal Saar-Tsechansky, and Sofus Macskassy. More than words: Quantifying language to measure firms’ fundamentals. The Journal of Finance, 63(3):1437–1467, 2008.
- [Luss and d’Aspremont, 2012] Ronny Luss and Alexandre d’Aspremont. Predicting abnormal returns from news using text classification. Quantitative Finance, (doi:10.1080/14697688.2012.672762):1–14, 2012.
- [Xie et al., 2013] Boyi Xie, Rebecca J. Passonneau, Leon Wu, and Germán G. Creamer. Semantic frames to predict stock price movement. In Proc. of ACL, pages 873–883, 2013.

[Wang and Hua, 2014] William Yang Wang and Zhenhao Hua. A semiparametric gaussian copula regression model for predicting financial risks from earnings calls. In Proc. of ACL, pages 1155–1165, 2014.

[Kogan et al., 2009] Shimon Kogan, Dmitry Levin, Bryan R. Routledge, Jacob S. Sagi, and Noah A. Smith. Predicting risk from financial reports with regression. In Proc. of NAACL, pages 272–280, 2009.

[Ding et al., 2015] Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. Deep Learning for Event-Driven Stock Prediction: An empirical investigation. Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015).