

# Stock Market Prediction with Deep Learning: A Character-based Neural Language Model for Event-based Trading

Leonardo dos Santos Pinheiro  
Macquarie University  
Capital Markets CRC  
lpinheiro@cmcrc.com

Mark Dras  
Macquarie University  
mark.dras@mq.edu.au

## Abstract

In the last few years, machine learning has become a very popular tool for analyzing financial text data, with many promising results in stock price forecasting from financial news, a development with implications for the Efficient Markets Hypothesis (EMH) that underpins much economic theory. In this work, we explore recurrent neural networks with character-level language model pre-training for both intraday and interday stock market forecasting. In terms of predicting directional changes in the Standard & Poor's 500 index, both for individual companies and the overall index, we show that this technique is competitive with other state-of-the-art approaches.

## 1 Introduction

Predicting stock market behavior is an area of strong appeal for both academic researchers and industry practitioners alike, as it is both a challenging task and could lead to increased profits. Predicting stock market behavior from the arrival of new information is an even more interesting area, as economists frequently test it to challenge the Efficient Market Hypothesis (EMH) (Malkiel, 2003): a strict form of the EMH holds that any news is incorporated into prices without delay, while other interpretations hold that incorporation takes place over time.

In practice, the analysis of text data such as news announcements and commentary on events is one major source of market information and is widely used and analyzed by investors (Oberlechner and Hocking, 2004).

Financial news conveys novel information to broad market participants and a fast reaction to the release of new information is an important component of trading strategies (Leinweber and Sisk, 2011).

But despite the great interest, attempts to forecast stock prices from unstructured text data have had limited success and there seems to be much room for improvement. This can be in great part attributed to the difficulty involved in extracting the relevant information from the text. So far most approaches to analyzing financial text data are based on bag-of-words, noun phrase and/or named entity feature extraction combined with manual feature selection, but the capacity of these methods to extract meaningful information from the data is limited as much information about the structure of text is lost in the process.

In recent years, the trend for extracting features from text data has shifted away from manual feature engineering and there has been a resurgence of interest in neural networks due to their power for learning useful representations directly from data (Bengio et al., 2013). Even though deep learning has had great success in learning representations from text data (e.g. Mikolov et al. (2013a), Mikolov et al. (2013b) and Kiros et al. (2015)), successful applications of deep learning in textual analysis of financial news have been few, even though it has been demonstrated that its application to event-driven stock prediction is a promising area of research (Ding et al., 2015).

Finding the most informative representation of the data in a text classification problem is still an open area of research. In the last few years a range of different neural networks architectures have been proposed for text classification, each one with strong results on differ-

ent benchmarks (e.g. [Socher et al. \(2013\)](#), [Kim \(2014\)](#) and [Kumar et al. \(2016\)](#)), and each one proposing different ways to encode the textual information.

One of the most commonly used architectures for modeling text data is the Recurrent Neural Network (RNN). One technique to improve the training of RNNs, proposed by [Dai and Le \(2015\)](#) and widely used, is to pre-train the RNN with a language model. In this work this approach outperformed training the same model from random initialization and achieved state of the art in several benchmarks.

Another strong trend in deep learning for text is the use of a word embedding layer as the main representation of the text. While this approach has notable advantages, word-level language models do not capture sub-word information, may inaccurately estimate embeddings for rare words, and can poorly represent domains with long-tailed frequency distributions. These were motivations for character-level language models, which [Kim et al. \(2016\)](#) and [Radford et al. \(2017\)](#) showed are capable of learning high level representations despite their simplicity. These motivations seem applicable in our domain: character-level representations can for example generalise across numerical data like percentages (e.g. the terms 5% and 9%) and currency (e.g. \$1,29), and can handle the large number of infrequently mentioned named entities. Character-level models are also typically much more compact.

In this work we propose an automated trading system that, given the release of news information about a company, predicts changes in stock prices. The system is trained to predict both changes in the stock price of the company mentioned in the news article and in the corresponding stock exchange index (S&P 500). We also test this system for both intraday changes, considering a window of one hour after the release of the news, and for changes between the closing price of the current trading session and the closing price of the next day session. This comparative analysis allow us to infer whether the incorporation of new information is instantaneous or if it occurs gradually over time. Our model consists of a recurrent neural network pre-trained by a

character level language model.

The remainder of the paper is structured as follows: In Section 2, we describe event-driven trading and review the relevant literature. In Section 3 we describe our model and the experimental setup used in this work. Section 5 presents and discuss the results. Finally, in Section 6 we summarize our work and suggest directions for future research.

## 2 Event-based Trading

In recent years, with the advances in computational power and in the ability of computers to process massive amounts of data, algorithmic trading has emerged as a strong trend in investment management ([Ruta, 2014](#)). This, combined with the advances in the fields of machine learning and natural language processing (NLP), has been pushing the use of unstructured text data as source of information for investment strategies as well ([Fisher et al., 2016](#)).

The area of NLP with the biggest influence in stock market prediction so far has been sentiment analysis, or opinion mining ([Pang et al., 2008](#)). Earlier work by [Tetlock \(2007\)](#) used sentiment analysis to analyze the correlation between sentiment in news articles and market prices, concluding that media pessimism may affect both market prices and trading volume. Similarly, [Bollen et al. \(2011\)](#) used a system to measure collective mood through Twitter feeds and showed it to be highly predictive of the Dow Jones Industrial Average closing values. Following these results, other work has also social media information for stock market forecasting ([Nguyen et al., 2015](#); [Oliveira et al., 2017](#), for example).

With respect to direct stock price forecasting, from news articles, many systems based on feature selection have been proposed in the literature. [Schumaker et al. \(2012\)](#) built a system to evaluate the sentiment in financial news articles using a Support Vector Regression learner with features extracted from noun phrases and scored on a positive/negative subjectivity scale, but the results had limited success. [Yu et al. \(2013\)](#) achieved better accuracy with a selection mechanism based on a contextual entropy model which expanded a set of seed words by discovering similar

emotion words and their corresponding intensities from online stock market news articles. Hagenau et al. (2013) also achieved good results by applying Chi-square and Binormal separation feature selection with n-gram features. As with all sentiment analysis, scope of negation can be an issue: Pröllochs et al. (2016) recently proposed a reinforcement learning method to predict negation scope and showed that it improved the accuracy on a dataset from the financial news domain.

A different approach to incorporate news into stock trading strategies was proposed by Nuij et al. (2014), which used an evolutionary algorithm to combine trading rules using technical indicators and events extracted from news with expert-defined impact scores. While far from using an optimal way to extract information from financial text data their results concluded that the news events were a component of optimal trading strategies.

As elsewhere in NLP, deep learning methods have been used to tackle financial market trading. A key approach that has informed the model and evaluation framework of this paper is that of Ding et al. (2014), which used a two-layer feed forward neural network as well as a linear SVM, treating the question of whether stocks would rise or fall as a classification problem; they found that the deep learning model had a higher accuracy. They also compared bag-of-words as input with structured events extracted from financial news via open information extraction (open IE), with the structured input performing better. They found that prediction accuracy was better for the following day’s price movement than for the following week, which was in turn better than the following year, as expected.

In their subsequent work, Ding et al. (2015) used a neural tensor network to learn embeddings of both words and structured events as inputs to their prediction models. They then applied a multichannel deep convolutional network — the channels corresponding to events at different timescales — to predict changes in the Standard & Poor’s 500 stock (S&P 500) index and in individual stock prices. This work was followed by Vargas et al. (2017) who combined recurrent and convolution layers with pre-trained word vectors to also pre-

dict changes to the S&P 500 index. The architecture here was also multichannel, and incorporated a technical analysis<sup>1</sup> input channel. The results from both pieces of work outperformed the former manual feature engineering approaches.

To the best of our knowledge, character-level sequence modeling has not been applied to stock price forecasting so far; neither has the use of language model pre-training. We note that the event models of Ding et al. (2014) and Ding et al. (2015) make use of generalization and back-off techniques to deal with data sparsity in terms of named entities etc, which as mentioned earlier character-level representations could help address. Also, character-level inputs are potentially complementary to other sorts such as word-level inputs or event representations, in particular with the multichannel architectures used for the work described above: research such as that of Kim (2014) has shown that multiple input representation can be usefully combined, and work using this kind of model such as Ruder et al. (2016) has specifically done this for character-level and word-level or other inputs. In this work, we aim to investigate whether this kind of character-level input may capture useful information for stock price prediction.

### 3 Model design and Training Details

Following Ding et al. (2014), we have a two-part model. The first builds a representation for the input, which for us is the character-level language model. The second is the recurrent neural network used for the prediction, a classifier that takes the input and predicts whether the price will rise or fall in the chosen timeframe. Both models process text as a sequence of UTF-8 encoded bytes.

#### 3.1 Neural language model

Existing pre-trained embeddings typically come from general domains (Google News,

---

<sup>1</sup>Fundamental analysis looks at fundamental properties of companies (e.g. earnings) to predict stock price movements; the previously described work in this section could be seen as carrying out a kind of fundamental analysis based on information from news reports. Technical analysis looks at past price movements as a guide to future ones.

Table 1: Statistics of Dataset

Data	Training	Validation	Test
Time Interval	02/10/2006 - 18/06/2012	19/06/2012 - 21/02/2013	22/02/2013 - 21/11/2013
Documents	157,033	52,344	51,476
Total bytes	736,427,755	232,440,500	245,771,999
News average per day	126	124	124
News average per company	911	878	897

Wikipedia, etc), but these word embeddings often fail to capture rich domain specific vocabularies. We therefore train our own embeddings on financial domain news text consisting of news articles from Reuters and Bloomberg. The data is further described in Section 4.

For the language model we used a character embedding with 256 units followed by a single layer LSTM (Hochreiter and Schmidhuber, 1997) with 1024 units. The characters are first encoded as bytes to simplify the embedding look-up process. The model looks up the corresponding character embedding and then updates its hidden state and predicts a probability distribution over the next possible byte. Individual text paragraphs are prepended with `<s>` to simulate a starting token and appended with `<\s>` to simulate an end token. Figure 1a shows a representation of this network.

The model was trained for 10 epochs on mini-batches of 256 subsequences of length 256. The character embeddings and the LSTM weights are then saved and used to initialize the first two layers of deep neural network for classification. The model is trained with stochastic gradient descent (SGD).

### 3.2 RNN for Stock Prediction

The second neural network has the same two layers as the language model, but with one additional fully connected layer with 512 units using a Leaky Relu activation (Maas et al., 2013). Only the last output of the LSTM layer is used to connect to the fully connected layer, the rationale being that this final state should encode a full representation the text sentence.

After the embedding look-up and hidden state update, the model goes through the fully connected layer and then predicts the probability of a positive direction price change for the stock price. This model is trained with Adam (Kingma and Ba, 2014) for 50 epochs. Figure 1b displays this architecture.

## 4 Experiments

**Data** We evaluated our model on a dataset of financial news collected from Reuters and Bloomberg over the period from October 2006 to November 2013. This dataset was made available by Ding et al. (2014). Stock price data for all S&P 500 companies and for the S&P 500 index were obtained from Thomson Reuters Tick History.<sup>2</sup> Following Radinsky et al. (2012) and Ding et al. (2014) we focus on the news headlines instead of the full content of the news articles for prediction since they found it produced better results. With this data we tested price response to news releases and daily responses, both shortly after (‘intraday’) and at end of day (‘interday’) as described below, and both for the stocks mentioned in the news article and for the index. Summary statistics of the data are shown in Table 1.

For intraday prediction, we filtered the news that contained only the name of one company belonging to the S&P 500 index and conducted our experiments on predicting whether the last price after one hour would be higher than the first price after the news release, using the timestamp for the news release. We also tested the S&P 500 index in the same time window.

In the interday prediction we used a setup similar to Ding et al. (2015) and Vargas et al. (2017), in which we concatenated all news articles from the same company on each day and predicted if the closing price in the day  $t + 1$  would increase when compared with closing price on day  $t$ , and similarly for the S&P 500 index.

**Models** We compare the model described in Section 3 with several baselines. For the other work using the same dataset (Ding et al., 2015; Vargas et al., 2017), we give the results from

<sup>2</sup><https://github.com/philipperemy/financial-news-dataset>

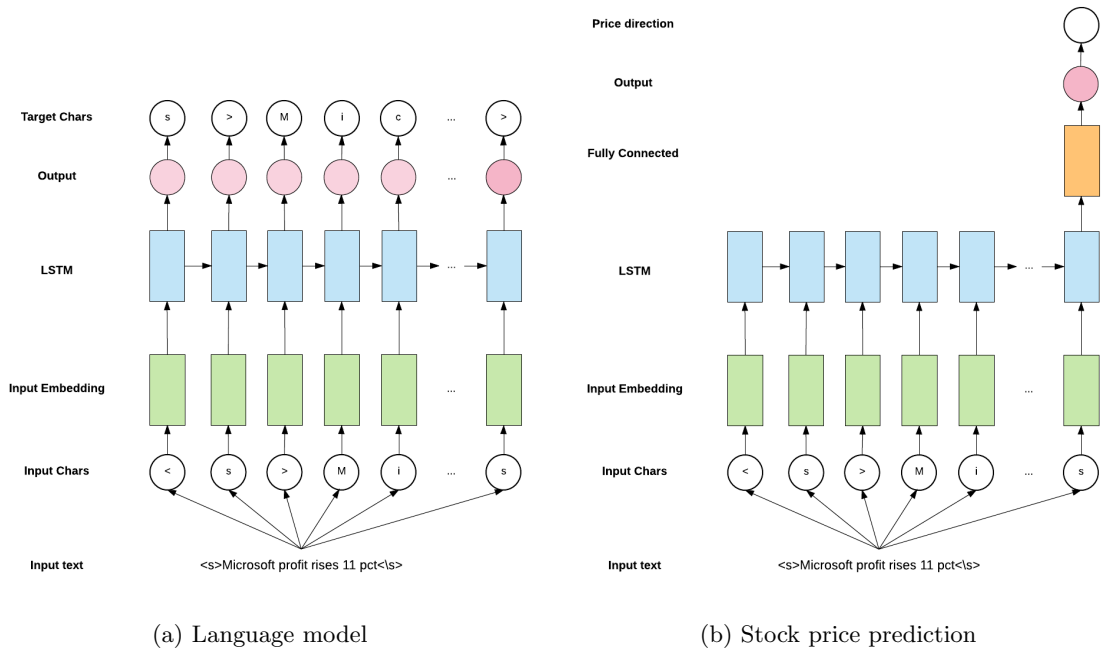


Figure 1: (a) Network architecture for the language model. In each step the output of the LSTM layer predicts the probability distribution of the next character. (b) Networks architecture for the stock prediction network. Only at the final processing of the text the output of the LSTM is used to predict the direction of the stock price.

the respective papers; we use the same experimental setup as they did. These models do not have intraday results, as the authors of those papers did not have stock data at more fine-grained intervals than daily. Only the models presented in Ding et al. (2015) have results for individual stocks, in addition to the S&P 500 Index.

We also reimplemented the model used by Luss and d’Aspremont (2015), which was a competitive baseline for Ding et al. (2015). In this model bags-of-words are used to represent the news documents and Support Vector Machines (SVMs) are used for prediction. We thus have both interday and intraday results for this model.

In the results, the following notation identifies each model:

- CharB-LSTM (ours): character embedding input and LSTM followed by fully connected prediction model.
- WI-RCNN: word embedding and technical indicators input and RCNN prediction model (Vargas et al., 2017).
- SI-RCNN: sentence embedding and technical indicators input and RCNN predic-

tion model (Vargas et al., 2017).

- BW-SVM: bag-of-words and support machines (SVMs) prediction model (Luss and d’Aspremont, 2015).
- E-NN: structured events tuple input and standard neural network prediction model (Ding et al., 2014).
- WB-CNN: sum of each word in a document as input and CNN prediction model (Ding et al., 2015).
- EB-CNN: event embedding prediction model (Ding et al., 2015).

Following Lavrenko et al. (2000) and Ding et al. (2015) we also test the profitability of our proposed model. We follow a slightly different strategy, though. As in Ding et al. (2015) we perform a market simulation considering the behavior of a fictitious trader. This trader will use the predictions of the model to invest \$10,000 worth in a stock if the model indicates the price will rise and will hold the position until the end of the current session, selling at the closing price. The same strategy is used for short-selling if the model indicates that an individual stock price will fall. Differently from

Ding et al. (2015) we do not consider profit taking behavior in our simulation. Rather, we plot the behaviour to visualize what happens over time, instead of presenting a single aggregate number. For this simulation we considered only the predictions on a portfolio consisting of the S&P 500 Index constituent companies. We compare the results of this strategy with the S&P 500 Index performance over the same period.

## 5 Results and Discussion

**Language Model** We first look at the quality of the representations learned by the character language model. Considering it was trained exclusively on a dataset of financial news we wondered how well this model would be able to reproduce the information dependencies present in the data such as the time of the events and currency information. The language model seemed capable of reproducing these dependencies. In Table 2 we show some sample text generated by the language model. While semantically incorrect, the representations learned by the model seem able to reproduce to some extent the grammatical structure of the language, as well as understand the entities present in the training dataset and the structure of numerical data.

**Model Accuracy** Table 3 shows the experimental results of the model S&P 500 Index prediction on the test dataset, in terms of accuracy of predicting stock price movement, while Table 4 shows the test results of individual company predictions. Similarly to Ding et al. (2015), individual stock prediction performs better than index prediction. Overall, our character-level language model pre-training performs at least as well as all of the other models with the exception of EB-CNN, but with the advantage over EB-CNN of being substantially simpler in implementation in terms of not having a module for modelling events. In general, while the other proposed models use more complex architectures and external features such as technical indicators and structured event detection, our approach leverages only on language model pre-training.

The higher performance of the EB-CNN architecture of Ding et al. (2015) is likely to be due to the neural tensor network component

that takes as input word embeddings and returns a representation of events; this provides a boost of 3+% over their comparable CNN model that does not explicitly incorporate events (WB-CNN in our table). This event component is potentially compatible with our approach; they could be combined, for example, by feeding our character-based input to an event component, or as noted earlier via multi-channel inputs, along the lines of Kim (2014).

**Market Simulation** In Figure 2 we report the results of the market simulation. Overall, the model is able to outperform the index consistently, despite having greater variance. While the model do not consider trading frictions such as transaction costs and market impact, we believe these results highlight the viability of the strategy. Exploration of advanced market microstructure implications are beyond the scope of this paper.

**Efficient Markets Hypothesis** One interesting aspect of these results is the superior performance of the daily prediction over intraday prediction.

In terms of what this might suggest for the EMH, Malkiel (2003) notes:

It was generally believed that securities markets were extremely efficient in reflecting information about individual stocks and about the stock market as a whole. The accepted view was that when information arises, the news spreads very quickly and is incorporated into the prices of securities without delay.

In fact, in the original formulation of the EMH, Fama (1970) remarks that “at *any* time prices fully reflect all available information” [*italics added*], implying instantaneous incorporation of information into prices. Other work such as Grossman and Stiglitz (1980) has argued that there are informational inefficiencies in the market that lead to delays in that information being incorporated into prices. Malkiel (2003) reviews some of the reasons for underreacting to new information, which include judgement biases by traders, such as conservatism (Edwards, 1968), “the

Table 2: Random samples from the character language model

---

<s>Copper for the region will increase the economy as rising inflation to recover from the property demand proposals for the region's largest economy and a share price of 10 percent of the nation's bonds to contain the company to spend as much as \$1.3 billion to \$1.2 billion in the same period a year earlier.</s>

<s>(Reuters) - The Bank of America Corp ( NBA.N ) said on Wednesday as proposals are seeking to be completed by the end of the year, the biggest shareholder of the stock of a statement to buy the company's casino and the country's biggest economy. "The U.S. is a way that the credit crisis will be a proposal to get a strong results of the budget deficit in the next month," said Toyota Motor Chief Executive Officer Tom Berry said in a telephone interview.</s>

<s>The U.S. is considering a second straight month in the U.S. and Europe to report the stock of the nation's currency and the previous consecutive month. The company will sell 4.5 billion euros (\$3.6 billion) of bonds in the first quarter of 2012, according to the median estimate of analysts surveyed by Bloomberg.</s>

---

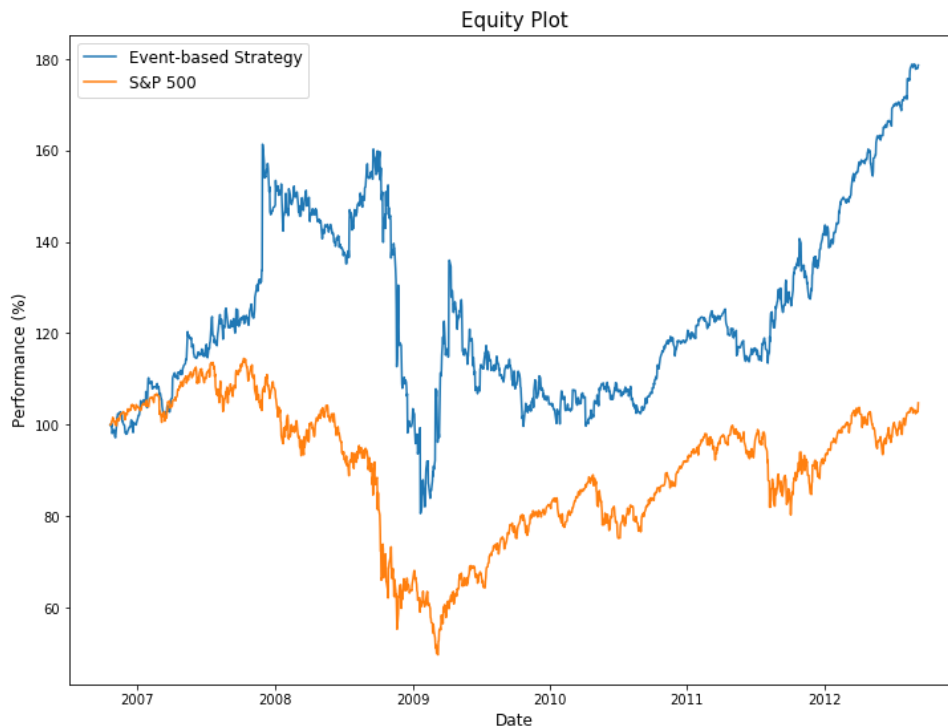


Figure 2: Equity plot of trading using the proposed event-based strategy.

Table 3: Results of S&amp;P 500 Index prediction

Model	Interday	Intraday
BW-SVM	56.42%	53.22%
<b>CharB-LSTM (ours)</b>	63.34%	59.86%
WI-RCNN	61.29%	*
SI-RCNN	63.09%	*
WB-CNN	61.73%	*
E-NN	58.94%	*
EB-CNN	64.21%	*

Table 4: Results of Individual stock prediction

Model	Interday	Intraday
BW-SVM	58.74%	54.22%
<b>CharB-LSTM (ours)</b>	64.74%	61.68%
WB-CNN	61.47%	*
EB-CNN	65.48%	*

slow updating of models in the face of new evidence”.

In terms of looking at the effect of news announcements, there historically haven’t been the tools to analyse vast quantities of text to evaluate effects on stock prices. With deep learning, and the online availability of stock prices at fine-grained intervals, it is now possible to look empirically at how long it takes information to be incorporated by assessing how predictable stock prices are as a function of news announcements. Previous work discussed in Section 2 had observed that it was still possible to predict at levels better than chance for up to a year out, although most strongly as time horizons were shorter. However, our preliminary results from our two models with intraday results show that predictability does not decrease monotonically: information is more typically incorporated later than the first hour.

## 6 Conclusion

This paper presented the use of a simple LSTM neural network with character level embeddings for stock market forecasting using only financial news as predictors. Our results suggest that the use of character level embeddings is promising and competitive with more complex models which use technical indicators and event extraction methods in addition to the news articles.

Character embeddings models are simpler and more memory efficient than word embeddings and are also able to keep sub-word infor-

mation. With character embeddings the risk of seeing unknown tokens in the test set is diminished, since the data sparsity is much lower than with word embeddings.

In the future we consider testing the use of character embeddings with more complex architectures and possibly the addition of other sources of information to create richer feature sets.

In addition, while previous work has found that including the body text of the news performs worse than just the headline, there may be useful information to extract from the body text, perhaps along the lines of Pang and Lee (2004), which improves sentiment analysis results by snipping out irrelevant text using a graph-theoretic minimum cut approach.

Other directions include looking at predicting price movements at a range of time horizons, in order to gauge empirically how quickly information is absorbed in the market, and relate this to the finance literature on the topic.

## Acknowledgements

We thank the Capital Markets CRC for providing financial support for this research.

## References

- Yoshua Bengio, Aaron Courville, and Pascal Vincent. 2013. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35(8):1798–1828.
- Johan Bollen, Huina Mao, and Xiaojun Zeng. 2011. Twitter mood predicts the stock market. *Journal of Computational Science* 2(1):1–8.
- Andrew M Dai and Quoc V Le. 2015. [Semi-supervised sequence learning](http://papers.nips.cc/paper/5949-semi-supervised-sequence-learning.pdf). In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28*, Curran Associates, Inc., pages 3079–3087. <http://papers.nips.cc/paper/5949-semi-supervised-sequence-learning.pdf>.
- Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. 2014. [Using structured events to predict stock price movement: An empirical investigation](https://doi.org/10.3115/v1/D14-1148). In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, pages 1415–1425. <https://doi.org/10.3115/v1/D14-1148>.



- Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. 2015. Deep learning for event-driven stock prediction. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*. pages 2327–2333.
- W. Edwards. 1968. Conservatism in human information processing. In B. Kleinmütz, editor, *Formal Representation of Human Judgement*, Wiley, New York.
- Eugene Fama. 1970. Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance* 25:383–417.
- Ingrid E Fisher, Margaret R Garnsey, and Mark E Hughes. 2016. Natural language processing in accounting, auditing and finance: a synthesis of the literature with a roadmap for future research. *Intelligent Systems in Accounting, Finance and Management* 23(3):157–214.
- Sanford J. Grossman and Joseph E. Stiglitz. 1980. On the Impossibility of Informationally Efficient Markets. *American Economic Review* 70:393–408.
- Michael Hagenau, Michael Liebmann, and Dirk Neumann. 2013. Automated news reading: Stock price prediction based on financial news using context-capturing features. *Decision Support Systems* 55(3):685–697.
- Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation* 9(8):1735–1780.
- Yoon Kim. 2014. [Convolutional neural networks for sentence classification](#) pages 1746–1751. <https://doi.org/10.3115/v1/D14-1181>.
- Yoon Kim, Yacine Jernite, David Sontag, and Alexander M Rush. 2016. Character-aware neural language models. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*. pages 2741–2749.
- Diederik Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*.
- Ryan Kiros, Yukun Zhu, Ruslan R Salakhutdinov, Richard Zemel, Raquel Urtasun, Antonio Torralba, and Sanja Fidler. 2015. [Skip-thought vectors](#). In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28*. Curran Associates, Inc., pages 3294–3302. <http://papers.nips.cc/paper/5950-skip-thought-vectors.pdf>.
- Ankit Kumar, Ozan Irsoy, Peter Ondruska, Mohit Iyyer, James Bradbury, Ishaan Gulrajani, Victor Zhong, Romain Paulus, and Richard Socher. 2016. [Ask me anything: Dynamic memory networks for natural language processing](#). In Maria Florina Balcan and Kilian Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*. PMLR, New York, New York, USA, volume 48 of *Proceedings of Machine Learning Research*, pages 1378–1387. <http://proceedings.mlr.press/v48/kumar16.html>.
- Victor Lavrenko, Matt Schmill, Dawn Lawrie, Paul Ogilvie, David Jensen, and James Allan. 2000. Language models for financial news recommendation. In *Proceedings of the Ninth International Conference on Information and Knowledge Management*. ACM, pages 389–396.
- David Leinweber and Jacob Sisk. 2011. Event-driven trading and the “new news”. *The Journal of Portfolio Management* 38(1):110–124.
- Ronny Luss and Alexandre d’Aspremont. 2015. Predicting abnormal returns from news using text classification. *Quantitative Finance* 15(6):999–1012.
- Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. 2013. Rectifier nonlinearities improve neural network acoustic models. In *Proceedings of the ICML Workshop on Deep Learning for Audio, Speech, and Language Processing*.
- Burton G Malkiel. 2003. The efficient market hypothesis and its critics. *The Journal of Economic Perspectives* 17(1):59–82.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. [Efficient estimation of word representations in vector space](#). *CoRR* abs/1301.3781. <http://arxiv.org/abs/1301.3781>.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013b. [Distributed representations of words and phrases and their compositionality](#). In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*. Curran Associates, Inc., pages 3111–3119. <http://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf>.
- Thien Hai Nguyen, Kiyooki Shirai, and Julien Velcin. 2015. Sentiment analysis on social media for stock movement prediction. *Expert Systems with Applications* 42(24):9603–9611.
- Wijnand Nuij, Viorel Milea, Frederik Hogenboom, Flavius Frasincar, and Uzay Kaymak. 2014. An automated framework for incorporating news into stock trading strategies. *IEEE Transactions on Knowledge and Data Engineering* 26(4):823–835.

- Thomas Oberlechner and Sam Hocking. 2004. Information sources, news, and rumors in financial markets: Insights into the foreign exchange market. *Journal of Economic Psychology* 25(3):407–424.
- Nuno Oliveira, Paulo Cortez, and Nelson Areal. 2017. The impact of microblogging data for stock market prediction: Using twitter to predict returns, volatility, trading volume and survey sentiment indices. *Expert Systems with Applications* 73:125–144.
- Bo Pang and Lillian Lee. 2004. [A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts](#). In *Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL'04), Main Volume*. Barcelona, Spain, pages 271–278. <https://doi.org/10.3115/1218955.1218990>.
- Bo Pang, Lillian Lee, et al. 2008. Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval* 2(1–2):1–135.
- Nicolas Pröllochs, Stefan Feuerriegel, and Dirk Neumann. 2016. Negation scope detection in sentiment analysis: Decision support for news-driven trading. *Decision Support Systems* 88:67–75.
- Alec Radford, Rafal Józefowicz, and Ilya Sutskever. 2017. [Learning to generate reviews and discovering sentiment](#). *CoRR* abs/1704.01444. <http://arxiv.org/abs/1704.01444>.
- Kira Radinsky, Sagie Davidovich, and Shaul Markovitch. 2012. Learning causality for news events prediction. In *Proceedings of the 21st International Conference on World Wide Web*. ACM, pages 909–918.
- Sebastian Ruder, Parsa Ghaffari, and John G. Breslin. 2016. [Character-level and Multi-channel Convolutional Neural Networks for Large-scale Authorship Attribution](#). *CoRR* abs/1609.06686. <http://arxiv.org/abs/1609.06686>.
- Dymitr Ruta. 2014. Automated trading with machine learning on big data. In *IEEE International Congress on Big Data*. IEEE, pages 824–830.
- Robert P Schumaker, Yulei Zhang, Chun-Neng Huang, and Hsinchun Chen. 2012. Evaluating sentiment in financial news articles. *Decision Support Systems* 53(3):458–464.
- Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Ng, and Christopher Potts. 2013. [Recursive deep models for semantic compositionality over a sentiment treebank](#). In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, pages 1631–1642. <https://aclanthology.info/pdf/D/D13/D13-1170.pdf>.
- Paul C Tetlock. 2007. Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance* 62(3):1139–1168.
- Manuel R Vargas, Beatriz SLP de Lima, and Alexandre G Evsukoff. 2017. Deep learning for stock market prediction from financial news articles. In *IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*. IEEE, pages 60–65.
- Liang-Chih Yu, Jheng-Long Wu, Pei-Chann Chang, and Hsuan-Shou Chu. 2013. Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news. *Knowledge-Based Systems* 41:89–97.