Stock Market Trend Prediction Using Recurrent Convolutional Neural Networks

Bo Xu, Dongyu Zhang, Shaowu Zhang, Hengchao Li, and Hongfei Lin[✉]

School of Computer Science and Technology,
Dalian University of Technology, Dalian, China
hflin@dlut.edu.cn

Abstract. Short-term prediction of stock market trend has potential application for personal investment without high-frequency-trading infrastructure. Existing studies on stock market trend prediction have introduced machine learning methods with handcrafted features. However, manual labor spent on hand-crafting features is expensive. To reduce manual labor, we propose a novel recurrent convolutional neural network for predicting stock market trend. Our network can automatically capture useful information from news on stock market without any handcrafted feature. In our network, we first introduce an entity embedding layer to automatically learn entity embedding using financial news. We then use a convolutional layer to extract key information affecting stock market trend, and use a long short-term memory neural network to learn context-dependent relations in financial news for stock market trend prediction. Experimental results show that our model can achieve significant improvement in terms of both overall prediction and individual stock predictions, compared with the state-of-the-art baseline methods.

Keywords: Stock market prediction · Embedding layer
Convolutional neural network · Long short-term memory

1 Introduction

Financial information on the internet has increased explosively with the rapid development of the internet. Daily financial news, as an important resource of financial information, contains a large amount of valuable information, such as the changes in senior management of listed companies and the releases of new products. The information is highly useful for investors to make crucial decisions on their personal investment. The key issue on generating a high return on the stock market lies in how well we are able to successfully predict the future movement of financial asset prices. Therefore, it is necessary to fully exploit the financial information from news for stock market trend predictions.

Existing studies have addressed stock market trend prediction using various machine learning methods, such as Support Vector Machines (SVMs) [1–4], Least Squares Support Vector Machines (LS-SVMs) [5–7] and Artificial Neural Networks (ANNs) [8–10]. Most of these studies have focused on extracting effective features for

training a good prediction model [11–14]. Feature selection and feature engineering have been fully investigated for the improvement of performance. For example, Kogan et al. [11] addressed the volatility of stock returns using regression models based on financial reports to reduce the financial risk. Schumaker et al. [12] used SVM with several different textual representations, bag of words, noun phrases, and named entities for financial news article analysis. However, handcrafted features cost much manual labor and partly limit the scalability of the learned models. How to automatically generate effective features for stock market predictions remains a great challenge.

Recently, deep learning models have exhibited powerful capability in automatically generating effective features, and been successfully applied in different natural language processing tasks. Existing studies on stock trend prediction have also focused on automatically generating effective features based on deep neural network models. For example, Hsieh et al. [13] integrated bee colony algorithm into wavelet transforms and recurrent neural networks for stock prices forecasting. Ding et al. [14] proposed convolutional neural network to model the influences of events on stock price movements. However, how to accurately model the relationship between financial news and stock market movement poses a new challenge for stock market trend prediction.

In this paper, we attempt to introduce deep neural network based models for stock market trend prediction. Deep neural network models, such as convolutional neural network and long short-term memory network, have been widely used in natural language processing tasks [15–19]. We address two key issues for applying deep neural network based models in this task. One is how to generate effective entity embedding based on the contents of financial news, and the other is how to incorporate the complex mutual relationship between financial news and stock market movement into the prediction model.

In order to construct useful word embedding for financial news contents, we introduce an entity embedding method [20] to represent financial news. This method actually introduce an entity embedding layer between one-hot input layer and neural network model for automatically learning entity embedding for news contents. To predict stock market trend of the listed companies, we first extract key influential information from daily financial news. We then propose a convolutional recurrent neural network to represent the news for extracting key information. The proposed network can capture the context-dependence relations in financial news, and use their internal memory to process arbitrary sequences of inputs for better prediction. Experimental results show that the proposed model outperforms other baseline models and effectively predicts the stock market trends.

The main contribution of this paper is as follows: (1) We introduce an entity embedding layer to automatically learn distributed representation of financial news contents without any handcrafted feature. (2) We propose a recurrent convolutional neural network to extract the key information from financial news and model context-dependent relation for predicting stock market movements. (3) We conduct extensive experiments to evaluate the proposed model. Experimental results show that our model achieves significant improvement in terms of the prediction accuracy.

The rest of the paper is organized as follows: Sect. 2 introduces the overall framework and details of our model; Sect. 3 provides our experimental results and comparative analysis of the experimental results; Sect. 4 concludes the paper and introduces our future work.

2   Methodology

In this section, we introduce details about the proposed model. We first illustrate the overall framework of our model for stock market trend prediction shown in Fig. 1. The whole framework includes four modules: the financial data acquisition module, the data preprocessing module, the data labeling module and the model training module.
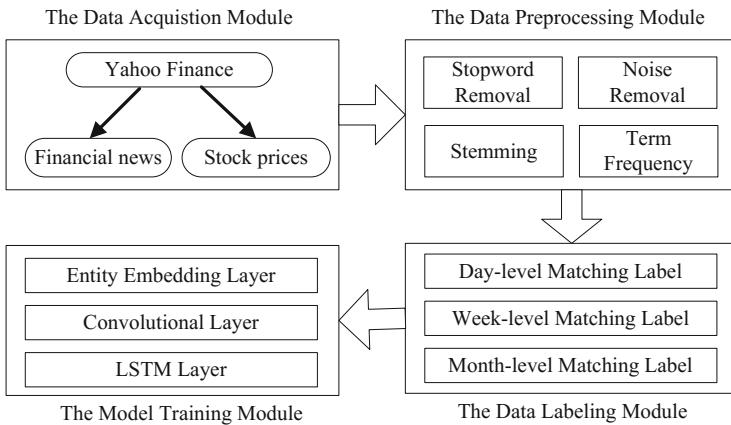
Fig. 1. The overall framework of our model

The financial data acquisition module crawls financial data from Yahoo Finance[1]. We acquire two types of data, financial news and stock prices, for model training. The financial news are used as the information source of model inputs, and the stock prices are used as the source of the ground truth labels for model targets.

The data preprocessing module transforms the webpages into texts by removing useless data, such as images and links. This module also preprocesses the stock prices data by removing stopwords, stemming the contents, and counting the term frequency in news for subsequent processing.

The data labeling module then matches the financial news with stock prices based on their timestamps, which is used to generate ground truth labels for model training at different levels, including the day-level, week-level and month-level matching labels.

The model training module is the core of our predictive model, the recurrent convolutional neural network model (RCNN). This module includes three layers, the embedding layer, the convolutional layer and the long short-term memory (LSTM)

---

[1] https://finance.yahoo.com/.

layer. We illustrate these layers in Fig. 2. The embedding layer learns entity embedding based on the financial news contents, the convolutional layer extracts the key local information of news, and the LSTM layer captures the relationship of dependency context for final prediction of stock market movements by a dense neural network (NN) layer. We introduce the details about each layer in the following subsections.
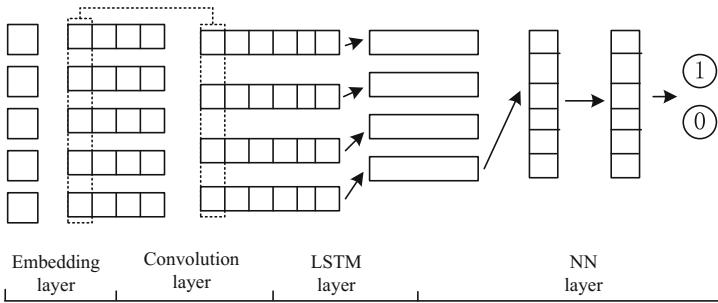

Embedding layer   Convolution layer   LSTM layer   NN layer

Fig. 2. Recurrent convolutional neural network

2.1 The Embedding Layer

For the embedding layer, we first count the term frequency of the crawled financial news to build a financial entity dictionary with high frequency entity terms. We then align the input sentences with diverse lengths using the financial dictionary as the inputs of the embedding layer. We adopt a state-of-the-art embedding method [20] to map the words to matrix. The used embedding method can represent key financial entities into vectors in Euclidean spaces, and map similar values close to each other in the embedding space to reveal the intrinsic properties of the categorical variables. The method can effectively represent financial entities into a vector space as the inputs of the following convolutional layer. Specifically, we first map each state of a discrete variable based on term frequency to a vector for learning vector representations of entities as follows.

$$e_i : x_i \rightarrow \mathbf{x}_i \tag{1}$$

The mapping is equivalent to build an extra layer on top of the input one-hot representations. We encode the inputs as follows.

$$u_i : x_i \rightarrow \delta_{x_i \alpha} \tag{2}$$

where $\delta_{x_i \alpha}$ is Kronecker delta and the range of α is the same as x_i. If m_i is the number of possible values of x_i, then $\delta_{x_i \alpha}$ becomes a vector of length m_i, where the element is non-zero when $\alpha = x_i$. Given the input x_i, the output of this fully connected layer is defined as follows.

$$\mathbf{x}_i \equiv \sum_{\alpha} w_{\alpha\beta} \delta_{x_i \alpha} = w_{x_i \beta} \tag{3}$$

where β is the index of the embedding layer and $w_{\alpha\beta}$ is the weight between the one-hot encoding layer and the embedding layer. It can be seen that the mapped vector representations of the entities are actually the weights of embedding layer. The weights are the parameters in the neural network, and can be learned during model training.

We provide a toy example of our data in Fig. 3 for better understanding the embedding layer. The example is taken from Apple news in April 26, 2016. The raw input of the example is "apple becomes the dow's worst performer". We preprocess the sentences by removing stopwords and stemming, and then we obtain the sentence "appl becom dow worst perform". Based on pre-built financial entity dictionary, we map the input sentence to the matrix using the entity embedding method, which will be taken as the inputs for the convolutional layer.
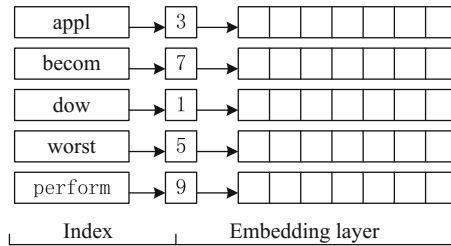

Fig. 3. A toy example for using the embedding layer to automatically learn distributed representation of financial entities.

2.2   The Convolutional Layer

Convolutional neural networks (CNN) are inspired by biological processes and are designed to use minimal amounts of preprocessing for encoding abundant semantic information in different natural language processing tasks. Convolutional neural networks, as variations of multilayer perceptrons, include three characteristics, local connectivity, parameter sharing and pooling. These characteristics make CNN an effective network model in extracting key information in texts

In our model, we use CNN as the convolutional layer, which treats the outputted matrix of the embedding layer as inputs for modeling the key information of financial news. The number of columns of the matrix is the dimensionality of the entity embedding, which is taken as the number of the feature maps of the convolutional layer. The number of rows of the matrix is taken as the number of convolutional kernels. We perform convolutional operation on the columns of the input matrix using max pooling to extract the critical information affecting stock movements. The outputs of the convolutional layer are then regarded as the inputs of the following LSTM layer.

We illustrate our convolutional layer in Fig. 4. The convolutional layer uses convolutional operation with max pooling to extract semantic and context information from financial news, and embeds the information into low dimensional representations for tracking the stock market movement.
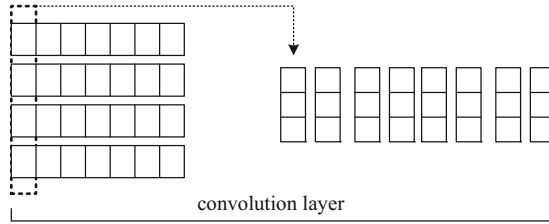
Fig. 4. Using convolution layer to extract the key information from financial news

2.3   The LSTM Layer

Recurrent neural network model (RNN) is widely used in NLP tasks, which equivalents to the multilayer feedforward neural network. Long short-term memory network (LSTM), as a variation of RNN, avoids the gradient vanish issue in RNN and uses historical information through the input, forget and output gate.

We adopt LSTM as a layer in our model. Our model takes the outputted matrix of the convolutional layer as the inputs of the LSTM layer for capturing the relationship of dependency contexts for final prediction of stock market movements. The rows of the matrix are taken as the hidden units of the LSTM layer, and the last hidden unit of the LSTM layer is then regarded as the inputs of the LSTM layer. LSTM has been proved to be effective in capturing temporal sequential information in other natural language processing tasks. Since financial data comprises abundant temporal information, we use LSTM to capture latent information in financial news, particularly to model the relationship between the stock market movement and the news. We illustrate the LSTM layer used in our model in Fig. 5.
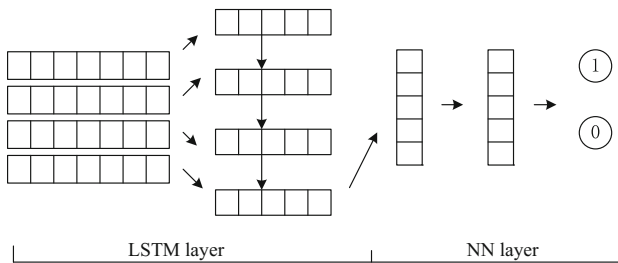
Fig. 5. Using LSTM to extract the context-dependent relation from financial news

Finally, we use a dense neural network layer to classify the financial news for predicting stock movements. We then evaluate our model using extensive experiments.

3   Experiments

3.1   Experimental Setup

In this section, we introduce the experimental setup and report the experimental results for evaluating the proposed model. We fetch the financial news using a web crawler from Yahoo Finance focusing on the shares of listed companies. The date range of the fetched news is from October, 2014 to May, 2016. The obtained data involves 447 listed companies, such as Apple Inc. (AAPL), Google Inc. (GOOG) and Microsoft Inc. (MSFT). We provide the statistics of our data in Table 1.

Table 1.   Statistics of the used data

Statistics	Quantity
The number of listed companies	447
Date range of financial news	2014.10–2016.05
The number of the news	322,694

We crawl historical stock prices from Yahoo Finance website, and use the prices to generate ground truth labels of the financial news. Specifically, if the stock price moves up in the next day, we label the current day's financial news as 1, indicating it is useful. Otherwise, if stock price moves down in the next day, we labeled the current day's news as 0, indicating it is useless for stock market trend prediction.

In addition, we use the headlines of financial news as the training data in our experiments following the work by Ding et al. [14] and Tetlock et al. [21], which showed that news titles are more useful than news contents for the prediction of stock market trend. In order to detect diminishing effects of reported events on stock market volatility, we label news at day level, week level and month level, respectively. Our preliminary experimental results show that week-level and month-level labels are of little use for stock trend prediction. Therefore, we adopt the day-level labels in the following experiments, which is the same setting as other existing studies on stock movement prediction [22–24].

In our experiments, we compare our model with two state-of-the-art baseline models. One is to represent financial news using bag of words features and SVM classifier proposed by Luss et al. [23], denoted as SVM. The other adopted neural tensor network to learn distributed representations of financial events, and used convolution neural network model for predicting the stock market [14], denoted as E-CNN. We evaluate the performance of prediction in terms of two standard evaluation metrics, the accuracy (Acc) and the Matthews correlation coefficient (MCC). We conduct 5-fold cross validations to evaluate the results, and report the average results for fair comparison.

3.2   Experimental Results and Analysis

Hyper-parameter Selection

Compared to the baseline models, we introduce the embedding layer to automatically learn the entity embedding of financial news, and then used the convolutional layer and the LSTM layer to extract critical information and the context-dependent relation for final prediction. There are six hyper-parameters used in our model, including the length of the inputs, the dimensionality of the embedding layer, the length of each CNN kernel, the number of CNN kernels, the dimensionality of the LSTM layer and the number of iterations. We switch these parameters for the baseline models and our model on the development set for selecting the optimal parameters. We report the selected optimal parameter in Table 2.

Table 2. The hyper-parameters of our models

Hyper-parameters	E-CNN	EB-CNN	E-CNN-LSTM	EB-CNN-LSTM
Length of the inputs	30	30	30	30
Dim. of embedding layer	50	128	50	128
Length of CNN kernel	3	3	3	3
Number of CNN kernels	250	250	64	64
Dim. of LSTM layer	–	–	70	70
Number of iterations	30	30	30	30

Experimental Results

We report our experimental results in this section. In the experiments, we train four different neural network models to demonstrate the effectiveness of the proposed embedding layer and the LSTM layer. We introduce these models as follows and report the experimental results in Table 3.

Table 3. The results of experiments

Experiments	Accuracy	MCC
SVM [23]	58.42%	0.1425
E-CNN [14]	63.44%	0.4198
EB-CNN	64.56%	0.4265
E-CNN-LSTM	65.19%	0.4356
EB-CNN-LSTM	**66.31%**	**0.4512**

- **E-CNN:** The model proposed by Ding et al. [14], which is one of the state-of-the-art models for predicting stock market trend. The model includes an event embedding layer and one CNN layer.
- **EB-CNN:** We use the model to examine the effectiveness of the proposed embedding layer for financial entity representation. The model includes the proposed embedding layer and the CNN layer.

- **E-CNN-LSTM:** We use this model to examine the effectiveness of the LSTM layer. The model includes the event embedding, the CNN layer and the LSTM layer.
- **EB-CNN-LSTM:** This model is the proposed model, including the entity embedding layer, the CNN layer and the LSTM layer.

From the table, we observe that the EB-CNN model achieves better prediction performance than the E-CNN model, which indicates the effectiveness of entity embedding used in our model. One possible explanation for this finding is that the entity embedding layer better encodes semantic information of financial entities for word and entity representations for financial news, while the event embedding layer used in E-CNN is designed to solve the sparsity of data and used to extract the key elements of events for the representation. Therefore, we obtain better performance using the CB-CNN model.

Furthermore, we observe that the E-CNN-LSTM model outperforms the E-CNN model, which indicates the effectiveness of the LSTM layer used in our model. We believe that this is because the LSTM layer contributes to extracting the context-dependent relationship between the financial news and the stock market trends. The proposed model finally achieves the best performance among all the baseline models, which demonstrates that our model is effective in capturing the stock market movement and predicting the stock market trends. We illustrate the experimental results with the change of the number of iterations in Fig. 6. The figure clearly shows that our model outperforms other baseline models with the number of iterations changing from 0 to 30.
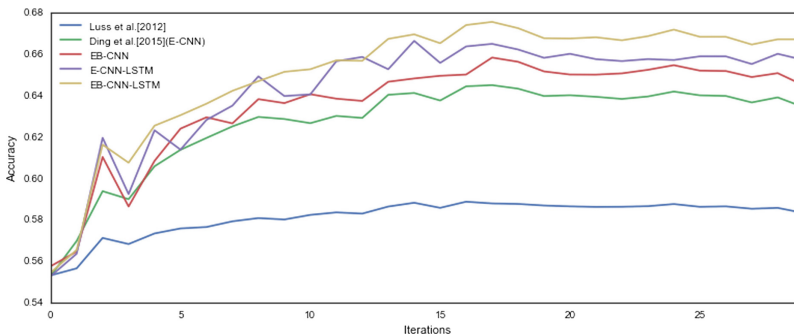

Fig. 6. Comparison of different models

Comparisons on Individual Stock Predictions

To further evaluate our model, we compare our approach with the baseline models in terms of individual stock predictions. We select nine companies as the individuals from our dataset. These companies cover high-ranking companies (GOOG, MSFT, AAPL), middle-ranking companies (AMAT, STZ, INTU), and low-ranking companies (HST, ALLE, JBHT). The ranking of companies are based on the S&P 500 from the Fortune Magazine[2]. We report the accuracy of individual stocks in Fig. 7.

---

[2] http://fortune.com/.
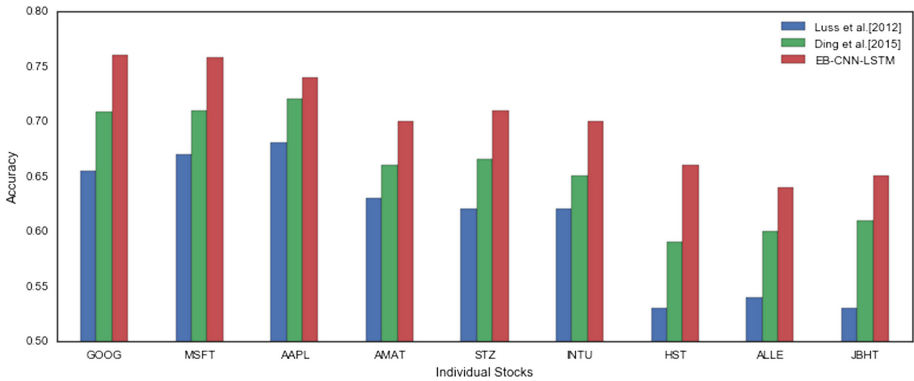
Fig. 7. Comparisons on individual stock prediction. Companies are named by ticker symbols.

From the figure, we can observe that our model achieves robust results in terms of the selected individual stocks. In addition, our model achieves relatively higher improvements on those lower fortune ranking companies, for which fewer pieces of news are available. For the baseline methods, the prediction results of low-ranking companies dramatically decrease. In contrast, our model achieves more stable performance. This is because our model uses the entity embedding layer to learn powerful distributed representations based on the news from these low-ranking companies. Hence, our model yields relatively high accuracy on prediction even without large amounts of daily news.

Diminishing Effects of the News
In order to detect diminishing effects of the news on stock market volatility, we label news in the next one day, next two day, and next three day, respectively. We train our model and the baseline models based on the different levels of labels, and report the experimental results in Fig. 8.
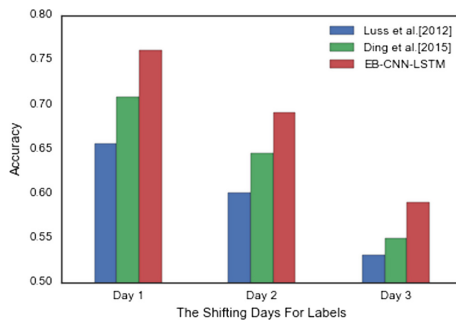
Fig. 8. Development results of different labels for the models

From the figure, we observe that our model achieves the best performance at different levels of labels compared to the baseline models. This finding exhibits the robustness and stability of our model. Besides, we observe that the effects of news on stock market prediction weakened over time, which indicates that daily prediction on stock market trend is necessary. We also use the news at a level of more than 3 days. The experimental results show that the influence of financial news is almost disappeared and useless for the prediction.

4   Conclusion and Future Work

In this paper, we propose a novel recurrent convolutional neural network model to predict the stock market trends based on financial news. In our model, we introduce an entity embedding layer to automatically learn distributed representation of financial entities without any handcrafted feature. We propose a recurrent convolutional neural network to extract the key information from financial news and model context-dependent relation for predicting stock market movements. The proposed network includes a convolutional layer and a long short-term memory layer for capturing abundant semantic information from the financial news. We conduct extensive experiments to evaluate the proposed model. Experimental results show that our model achieves significant improvement in terms of the prediction accuracy. In our future work, we will explore more effective model for predicting the stock market trend in consideration of temporal characteristics of news. We will also attempt to integrate external financial knowledge to optimize our model and improve the performance of stock trend prediction.

References

1. Huang, W., Nakamori, Y., Wang, S.Y.: Forecasting stock market movement direction with support vector machine. Comput. Oper. Res. **32**(10), 2513–2522 (2005)
2. Lee, M.C.: Using support vector machine with a hybrid feature selection method to the stock trend prediction. Expert Syst. Appl. **36**(8), 10896–10904 (2009)
3. Ni, L.P., Ni, Z.W., Gao, Y.Z.: Stock trend prediction based on fractal feature selection and support vector machine. Expert Syst. Appl. **38**(5), 5569–5576 (2011)
4. Yu, L., Wang, S., Lai, K.K.: Mining stock market tendency using GA-based support vector machines. In: Deng, X., Ye, Y. (eds.) WINE 2005. LNCS, vol. 3828, pp. 336–345. Springer, Heidelberg (2005). https://doi.org/10.1007/11600930_33
5. Chai, J., Du, J., Lai, K.K., et al.: A hybrid least square support vector machine model with parameters optimization for stock forecasting. Math. Probl. Eng. **2015**, 1–7 (2015)

6. Marković, I., Stojanović, M., Božić, M., Stanković, J.: Stock market trend prediction based on the LS-SVM model update algorithm. In: Bogdanova, A.M., Gjorgjevikj, D. (eds.) ICT Innovations 2014. AISC, vol. 311, pp. 105–114. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-09879-1_11

7. Yu, L., Chen, H., Wang, S., et al.: Evolving least squares support vector machines for stock market trend mining. IEEE Trans. Evol. Comput. **13**(1), 87–102 (2009)

8. Crone, S.F., Kourentzes, N.: Feature selection for time series prediction – a combined filter and wrapper approach for neural networks. Neurocomputing **73**(10), 1923–1936 (2010)

9. Dai, W., Wu, J.Y., Lu, C.J.: Combining Nonlinear Independent Component Analysis and Neural Network for the Prediction of Asian Stock Market Indexes. Pergamon Press Inc., Tarrytown (2012)

10. Kara, Y., Acar Boyacioglu, M., Baykan, Ö.K.: Predicting direction of stock price index movement using artificial neural networks and support vector machines. Expert Syst. Appl. **38**(5), 5311–5319 (2011)

11. Kogan, S., Levin, D., Routledge, B.R., et al.: Predicting risk from financial reports with regression. In: North American Chapter of the Association for Computational Linguistics, pp. 272–280 (2009)

12. Schumaker, R.P., Chen, H.: Textual analysis of stock market prediction using financial news articles. In: Americas Conference on Information Systems (2006)

13. Hsieh, T.J., Hsiao, H.F., Yeh, W.C.: Forecasting stock markets using wavelet transforms and recurrent neural networks: an integrated system based on artificial bee colony algorithm. Appl. Soft Comput. **11**(2), 2510–2525 (2011)

14. Ding, X., Zhang, Y., Liu, T., et al.: Deep learning for event-driven stock prediction. In: Ijcai, pp. 2327–2333 2015

15. dos Santos, C.N., Gatti, M.: Deep convolutional neural networks for sentiment analysis of short texts. In: COLING, pp. 69–78 (2014)

16. Xie, B., Passonneau, R.J., Wu, L., Creamer, G.G.: Semantic frames to predict stock price movement. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, pp. 873–883 (2013)

17. Martin, L., Lars, K., Amy, L.: A review of unsupervised feature learning and deep learning for time-series modeling. Pattern Recognit. Lett. **42**, 11–24 (2014)

18. Ding, X., Zhang, Y., Liu, T., Duan, J.: Using structured events to predict stock price movement: an empirical investigation. In: EMNLP, pp. 1415–1425 (2014)

19. Schmidhuber, J., Hochreiter, S.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)

20. Guo, C., Berkhahn, F.: Entity embeddings of categorical variables. arXiv preprint arXiv:1604.06737 (2016)

21. Tetlock, P.C., Saar Tsechansky, M., Macskassy, S.: More than words: quantifying language to measure firms' fundamentals. J. Finance **63**(3), 1437–1467 (2008)

22. Radinsky, K., Davidovich, S., Markovitch, S.: Learning causality for news events prediction. In: Proceedings of the 21st International Conference on World Wide Web, pp. 909–918. ACM (2012)

23. Luss, R., d'Aspremont, A.: Predicting abnormal returns from news using text classification. Quant. Finance **15**(6), 999–1012 (2015)

24. Bordes, A., Weston, J., Collobert, R., Bengio, Y.: Learning structured embeddings of knowledge bases. In: Conference on Artificial Intelligence (No. EPFL-CONF-192344) (2011)