Hybrid Deep Sequential Modeling for Social Text-Driven Stock Prediction

Huizhe Wu
East China Normal University
51164500253@stu.ecnu.edu.cn

Wei Zhang*
East China Normal University
zhangwei.thu2011@gmail.com

Weiwei Shen
GE Global Research Center
realsww@gmail.com

Jun Wang
East China Normal University
jwang@sei.ecnu.edu.cn

ABSTRACT

In addition to only considering stocks' price series, utilizing short and instant texts from social medias like Twitter has potential to yield better stock market prediction. While some previous approaches have explored this direction, their results are still far from satisfactory due to their reliance on performance of sentiment analysis and limited capabilities of learning direct relations between target stock trends and their daily social texts. To bridge this gap, we propose a novel Cross-modal attention based Hybrid Recurrent Neural Network (CH-RNN), which is inspired by the recent proposed DA-RNN model. Specifically, CH-RNN consists of two essential modules. One adopts DA-RNN to gain stock trend representations for different stocks. The other utilizes recurrent neural network to model daily aggregated social texts. These two modules interact seamlessly by the following two manners: 1) daily representations of target stock trends from the first module are leveraged to select trend-related social texts through a cross-modal attention mechanism, and 2) representations of text sequences and trend series are further integrated. The comprehensive experiments on the real dataset we build demonstrate the effectiveness of CH-RNN and benefit of considering social texts.

KEYWORDS

deep sequential modeling, stock prediction, social text

ACM Reference Format:
Huizhe Wu, Wei Zhang, Weiwei Shen, and Jun Wang. 2018. Hybrid Deep Sequential Modeling for Social, Text-Driven Stock Prediction. In *The 27th ACM International Conference on Information and Knowledge Management (CIKM '18), October 22–26, 2018, Torino, Italy*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3269206.3269290

1 INTRODUCTION

Stock trend prediction has already been researched for decades [1], due to its great value in seeking to maximize stock investment profit. Early approaches are mainly based on historical stock price time series and use time series analysis methods such as autoregressive model [3]. However, due to the excess volatility of stock prices [11], it is hard to solely utilize them for prediction. To alleviate this issue, financial news from traditional news media [4, 5] and more recently, social messages from Twitter [6, 9], have been largely explored to verify their predictive power. Since Twitter spreads information faster and meanwhile ensures high coverage of information contained in traditional medias [14], we focus on utilizing social texts from Twitter to help stock trend prediction in this paper.

As tweets largely involve users' opinions, stock-dependent tweet analysis is promising to benefit stock trend prediction. On the whole, humans may express their attitudes towards specific stocks by directly mentioning the corresponding stock codes followed by cash-tags ("$"). While a few previous studies have investigated using knowledge from social media for trend prediction [6, 8, 9, 13], most of their methods heavily depend on the good results of sentiment analysis and are not trained in an end-to-end fashion. Thus a natural research question arises: can we learn the relation between the target stock trend and the representations of corresponding social short texts in an end-to-end fashion?

To this end, we propose a novel deep learning model called Cross-modal attention based Hybrid Recurrent Neural Network (CH-RNN), partially inspired by the recent proposed DA-RNN model [7] which is originally developed to utilize attention mechanism [12] to fuse multiple time series to predict a central target time series (e.g, using major corporations' price series to predict the index values of NASDAQ 100). Our model involves two main modules (see Figure 1) for modeling stock price trends and social short texts simultaneously. The first one utilizes DA-RNN to learn stock trend representations. The other one utilizes recurrent neural network to model social texts, where a simple text modeling method is used to gain daily aggregated social text representation. These two modules interact seamlessly by the following two manners to form a unified framework: 1) the daily representations of stock trends from the first module are leveraged to attend daily representations of social texts through a cross-modal attention mechanism, ensuring to give more weights to the trend-related daily social texts; and 2) CH-RNN further combines the representations of stock trend and social text to enable joint learning for target stock trend prediction.

As no publicly available datasets exist for this problem, we build a real dataset[1] and the empirical experiments on it demonstrate

*Wei Zhang is the corresponding author.

[1] https://github.com/wuhuizhe/CHRNN

CIKM '18, October 22–26, 2018, Torino, Italy

ACM ISBN 978-1-4503-6014-2/18/10…$15.00
https://doi.org/10.1145/3269206.3269290

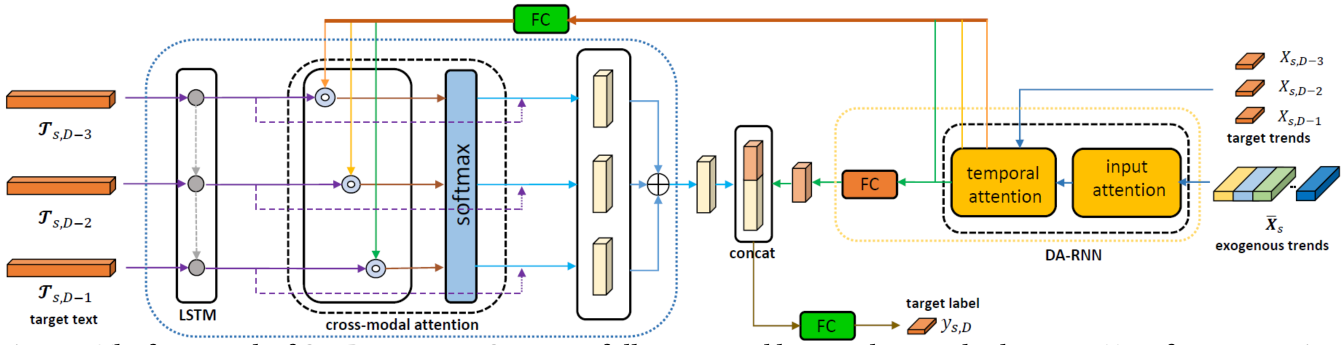

Figure 1: The framework of CH-RNN. Note: FC means a fully-connected layer and we set the day count $T = 3$ for presentation.

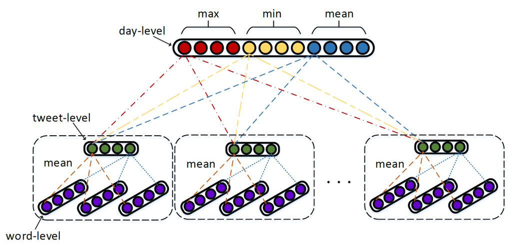

Figure 2: The architecture of social text representation.

CH-RNN achieves the best performance among several baselines and further verify the rationality of the model design.

2 PRELIMINARIES

2.1 Notations and Problem Statement

Following the settings of [4, 6], we suppose a stock market trend takes a binary value. When the closing price of stock s for the next day $(d + 1)$ is greater than the closing price for today, we let market trend $\mathbf{X}_{s,d+1} = +1$. Otherwise, we let $\mathbf{X}_{s,d+1} = -1$.

Suppose $\mathcal{S}$ to be the stock set where each stock $s \in \mathcal{S}$. We utilize $\mathbf{X}$ to denote a stock trend matrix where $\mathbf{X}_{s,d}$ is the market trend of stock s for the day d. For convenience of later problem statement, we use $\mathbf{X}_{.,d}$ to represent the market trends of all stocks for the day d. Similarly, We define $\mathcal{T}$ to be the social text set. $\mathcal{T}_{s,d}$ denotes the text set of stock s in day d. Then the problem is stated as follows:

Problem 1 (**Social Text-Driven Stock Prediction**). *For a target day D, we are given the stock market trends $\mathbf{X}_{.,D-T:D-1}$ and the corresponding social texts $\mathcal{T}_{.,D-T:D-1}$. The aim is to learn a function $f(\mathbf{X}_{.,D-T:D-1}, \mathcal{T}_{.,D-T:D-1}) \longrightarrow \mathbf{X}_{.,D}$ to predict the trends.*

2.2 Dual-stage Attention-based RNN

2.2.1 Input attention in the encoder. With regard to the target stock s and M $(|\mathcal{S}|-1)$ exogenous trend series, the encoder is to determine the attention weights of the different exogenous stocks for different days. Particularly, the attention weight α_d^m of stock m for the day d is calculated as follows:

$$g_d^m = \mathbf{v}_{en}^\top \tanh(\mathbf{W}_{en}\mathbf{h}_{d-1}^{en} + \mathbf{U}_{en}\mathbf{X}_{m,.}), \quad \alpha_d^m = \frac{\exp\left(g_d^m\right)}{\sum_{s=1}^{M}\exp\left(g_d^s\right)}. \quad (1)$$

Then we can get the updated trend representation for each day, e.g., $\mathbf{X}_{.,d}^* = (\alpha_d^1\mathbf{X}_{1,d}^*, \ldots, \alpha_d^M\mathbf{X}_{M,d}^*)$ and the hidden state $\mathbf{h}_d^{en}$ for the day d is obtained by $\mathbf{h}_d^{en} = \mathrm{LSTM}(\mathbf{X}_{.,d}^*, \mathbf{h}_{d-1}^{en})$.

2.2.2 Temporal attention in the decoder. The decoder is to determine the importance weights of different hidden states from encoders for each day. The importance weight β_d^t of hidden state $\mathbf{h}_t^{en}$ for the day d is given by:

$$l_d^t = \mathbf{v}_{de}^\top \tanh(\mathbf{W}_{de}\mathbf{h}_{d-1}^{de} + \mathbf{U}_{de}\mathbf{h}_t^{en}), \quad \beta_d^t = \frac{\exp\left(l_d^t\right)}{\sum_{t'=1}^{T}\exp\left(l_d^{t'}\right)}. \quad (2)$$

Thus a context vector $\mathbf{c}_d = \sum_{t'=1}^{T}\beta_d^{t'}\mathbf{h}_{t'}^{en}$ is gained and further used to compute the decoder hidden state $\mathbf{h}_d^{de} = \mathrm{LSTM}(\mathbf{h}_{d-1}^{de}, \mathbf{X}_{s,d-1}, \mathbf{c}_d)$. Finally, the decoder state $\mathbf{h}_D^{de}$ is utilized to predict the stock trend.

3 COMPUTATIONAL MODEL

The overall framework of CH-RNN is shown in Figure 1. The left part learns social text representation while the right utilizes DA-RNN to model stock trend series. Then we introduce the input representation and the cross-modal attention, follower by the description of how stock trend is predicted by integrating representations of social text and stock trend. To make the model clarification clearer, we take the stock s as an example.

3.1 Input Representation

Stock trend representation. The exogenous stock trends and the target stock trends are defined as $\bar{\mathbf{X}}_s = [\mathbf{X}_{m,D-T:D-1}]_{m=1,m\neq s}^{m=|\mathcal{S}|}$ and $\mathbf{X}_{s,D-T:D-1}$, respectively.

Daily aggregated social text representation. We aggregate all tweets belonging to one day to construct a daily text representation for avoiding the noise of a single tweet [2]. As Figure 2 shows, the adopted method first takes a mean pooling operation over all word embeddings to obtain tweet-level embeddings. Then the max, mean, and min poolings are all applied to daily tweet level embeddings. After concatenating all the pooled embeddings, we get daily aggregated social text representation, denote as $\mathbf{E}_{s,d}$ for $\mathcal{T}_{s,d}, \forall d \in \{d\}_{d=D-T}^{d=D-1}$.

3.2 Cross-modal Attention

Since there exists a temporal sequential relation between daily representations, we utilize LSTM to associate them together through sequential modeling, which is given by: $\mathbf{h}_{s,d}^{st} = \mathrm{LSTM}(\mathbf{E}_{s,d}, \mathbf{h}_{s,d-1}^{st})$, where $\mathbf{h}_{s,d}^{st}$ is the hidden state of daily aggregated social text corresponding to the stock s and day d.

To determine the different importance weights for each daily social text presentation, we propose a cross-modal attention computational method to leverage the representations from stock trend series to attend social texts. Specifically, we employ the generated hidden state $\mathbf{h}_d^{de}$ by DA-RNN, which could be regarded as a high-level state of stock trend for the day d. We first change $\mathbf{h}_d^{de}$ to $\mathbf{h}_{s,d}^{de}$ for satisfying our setting. Afterwards, we define the score γ_d^s to measure the degree of relevance between $\mathbf{h}_{s,d}^{de}$ and $\mathbf{h}_{s,d}^{st}$ and further apply the softmax function to the gotten score:

$$\gamma_d^s = \mathbf{h}_{s,d}^{st\top}(\mathbf{W}_1^{FC}\mathbf{h}_{s,d}^{de}), \quad \eta_d^s = \frac{\exp\left(\gamma_d^s\right)}{\sum_{d'=D-T}^{D-1}\exp\left(\gamma_{d'}^s\right)}, \qquad (3)$$

where $\mathbf{W}_1^{FC}$ is the parameter matrix of a fully connected (FC) layer, η_d^s is the attention weight of the social text for the day d.

The intuition for explaining why price representations can assist to select trend-relevant daily social texts is that even social texts are daily aggregated, they might not be equally informative and have different degrees of relevance to stock market trend. This phenomenon is caused by the fact that social texts are not informal and involve users' own statements and other non-instant events.

Consequently, daily aggregated text representations should have different importance weights due to the different degree of information instantaneity and relevance. Based on the cross-modal attention weight, with attaching higher weights to more relevant text traits, CH-RNN computes the social text embedding $\bar{\mathbf{h}}_{s,D}$ as a weighted sum of all the recent T daily representations: $\bar{\mathbf{h}}_{s,D} = \sum_{d=D-T}^{D-1} \eta_d^s \cdot \mathbf{h}_{s,d}^{st}$.

3.3 Trend Prediction

To generate good target trend prediction results, we combine the knowledge mined from already existed social text and stock trend. More specifically, $\bar{\mathbf{h}}_{s,D}$ and $\mathbf{h}_{s,D}^{de}$ are concatenated to form an integrated representation $\mathbf{h}_{s,D}^{de}$ for generating the stock trend prediction:

$$\mathbf{h}_{s,D} = [\bar{\mathbf{h}}_{s,D}; \mathbf{W}_2^{FC}\mathbf{h}_{s,D}^{de}], \;\; y_{s,D} = \sigma(\mathbf{W}_3^{FC}\mathbf{h}_{s,D}), \qquad (4)$$

where $\sigma(\cdot)$ is the sigmoid function, $\mathbf{W}_2^{FC}$ gives a linear transformation to make a flexible fusion.

4 EXPERIMENTS

4.1 Dataset and Implementation

We build a real-world dataset by crawling stock prices from Yahoo Finance and social texts from Twitter using Tweepy. It ranges from January 2017 to November 2017 and choose 47 stocks which have sufficient tweets from the Standard & Poor's 500 list. The basic statistics of the dataset are shown in Table 1. To evaluate our model and other baselines, we split the dataset with the ratio of approximately 5: 1: 1 in chronological order.

Table 1: Basic statistics of the dataset.

Data	#Stocks	#Days	#Tweets	#Words
Twitter	47	231	746,287	137,052

The dimensions of hidden states are set to 64 and 16 for price and social text modules. Word embeddings, with 50 dimensions, are initialized with the pre-trained ones [10]. The accuracy metric [4] is adopted for our evaluation. Adam is used for optimization.

4.2 Model Comparison

To validate the effective of CHRNN, we have chose three categories of algorithms for evaluation: (i) traditional machine learning model. We select autoregressive model (AuReg [3]) and feature based classification method (FeaCla) which constructs features from stock price and multiple related tweets; (ii) topic modeling based model. The semantic stock network (SSN [9]) is chosen, using a labeled Latent Dirichlet Allocation (LDA) to model texts of stocks (labels) for acquiring sentiment scores to replace price sequence features; (iii) deep learning network model. We choose dual-stage attention RNN (DA-RNN [7]) and the coupled LSTM (CLSTM) where one LSTM is used to model stock trend series and the other one is leveraged to model daily aggregated social texts.

To ensure robust comparison, we consider different lengths of sequences (days), i.e., from 3 to 8, to test the sequential models. Table 2 shows the results of CH-RNN and the other adopted baselines with several different sequence lengths.

Table 2: Accuracies of different models on Twitter. Note: (3-8) corresponds the average results of length from 3 to 8.

Model	AuReg	FeaCla	SSN	CLSTM	DA-RNN	**CH-RNN**
(3-8)	53.02	53.32	55.42	56.00	56.20	**59.15**

We first compare AutoReg and FeaCla. Although FeaCla takes the features of social text into consideration, it does not improve AutoReg significantly, showing that it might be hard to mine effective knowledge from social text based only on simple features. We find SSN outperforms the first two models obviously, demonstrating the benefit of the customized model for social-text driven stock prediction. As a simple alternative deep learning based approach, CLSTM presents good performance, showing the good capability of deep sequential modeling. By comparing DA-RNN with SSN and CLSTM, we see that DA-RNN performs better than the other two models, even it does not consider social text information. This motivates us to use DA-RNN to model stock trend series. Compared with all the other models, no matter what the length of the modeled sequence is, CH-RNN improves their results by a large margin, revealing the advantages of CH-RNN and benefits of considering social text for trend prediction.

4.3 Ablation Study

We further conduct ablation study of CH-RNN. We adopt "CH-RNN (w/o cm_att)" to denote the variant of CH-RNN which does not adopt the cross-model attention. Instead, the average of the hidden states is used. Besides, we propose an alternative attention method, i.e., defining a global parameter vector and use it to replace $\mathbf{W}_1^{FC}\mathbf{h}_{s,d}^{de}$ in Equation 3. We name it as "CH-RNN (go_att)".

Table 3 shows the performance of different variants of CH-RNN. We can see that CH-RNN (w/o cm_att) gets the worst results, which shows that utilizing attention computation to get a weighted sum of daily aggregated social text representations is necessary. Moreover, CH-RNN outperforms CH-RNN (go_att), indicating the cross-modal attention computation can benefit our model.

Table 3: Ablation study of CH-RNN.

Models	len=4	len=6	len=8	Ave (3-8)
CH-RNN (w/o cm_att)	58.7	56.7	57.7	57.75
CH-RNN (go_att)	58.8	58.6	57.6	58.40
CH-RNN	**59.0**	**59.5**	**58.1**	**59.15**

4.4 More Experimental Analysis

4.4.1 Market simulation. Following [4], we specify a market simulation strategy to evaluate the stock prediction performance of CH-RNN through a standard way of making profits. Compared with the opening price, we set threshold of price fluctuation to 2% as a signal to finish stock trading before the end of the day.

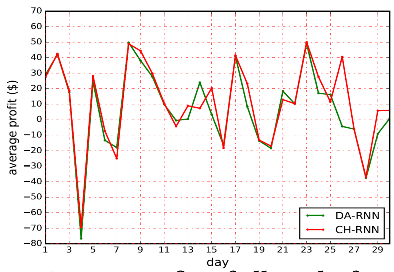

Figure 3: Average profits of all stocks for each day.

Table 4: Profit comparison between CH-RNN and DA-RNN.

Stock	DA-RNN	CH-RNN	Stock	DA-RNN	CH-RNN
ABBV	$1327	$1396	CVX	$673	$940
BMY	$912	$1022	F	$820	$1128
CELG	$602	$802	WMT	$710	$717

Figure 3 shows the average daily profit for all stocks. Obviously, CH-RNN gains more profits than DA-RNN on many days of the month, although its profit curve fluctuates because of the irregularities in stock prices. The one-month cumulative profits of CH-RNN are 37% more than DA-RNN's. Specifically, Table 4 presents the profits of the six companies, i.e., AbbVie, Bristol-Myers Squibb, Celgene, Chevron, Ford Motor, and Wal-Mart Stores, in the month.

4.4.2 Case study. In order to give qualitative analysis to the cross-modal attention, we select some stock text presentations for showing the ability of attention weights. Figure 4 shows several tweet examples of two days with attention weights corresponding to 0.6 and 0.09. The tweets on the day with higher attention weight contain more indicative words such as "Increase", "nice" and "Grows", which are not only with instantaneity but also have impacts on stock price series. The tweets on another day consist of neutral words with less indicative opinions or with neural descriptions themselves, thus the cross-modal attention attaches lower weight to them.

4.4.3 Error analysis. We compare the predictions of DA-RNN and CH-RNN, analyze the cases where trends are wrongly predicted by CH-RNN but well predicted by DA-RNN, and summarize two common situations. First, a tweet may talk events regarding one company but mention more than one stocks which might have competitive relationship. For example, as the first tweet in the figure 5 shows, it talks the event of Samsung ($SSNLF). However, as we collect tweets for each stock based on the existence of stock name, this tweet is also used for Apple ($AAPL), which harms the

Figure 4: Daily examples with different attention weights.

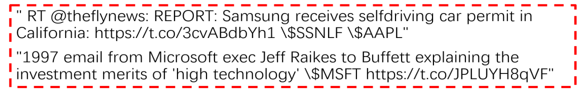

Figure 5: Error analysis.

prediction for Apple. Second, a tweet may talk things happened past. As the second tweet in Figure 5 shows, the event happened in 1997 and is apparently not relevant to current stock trend. But it is hard for our models to capture this knowledge.

5 CONCLUSION

In this paper, we propose a novel deep learning model CH-RNN which can leverage stock price trend representations to attend daily social text representations through a cross-modal attention interaction. We build a real-world Twitter dataset and the extensive experiments show that CH-RNN is effective for jointly modeling stock trend and social text for the studied problem.

ACKNOWLEDGMENTS

This work was supported in part by Shanghai Chenguang Program (16CG24), Shanghai Sailing Program (17YF1404500), and NSFC (61702190, 61672231, 61672236).

REFERENCES

[1] Werner FM Bondt and Richard Thaler. 1985. Does the stock market overreact? *The Journal of finance* 40, 3 (1985), 793–805.
[2] Zhiyuan Cheng, James Caverlee, and Kyumin Lee. 2010. You are where you tweet: a content-based approach to geo-locating twitter users. In *CIKM*. 759–768.
[3] Michael P Clements and David F Hendry. 1969. Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics* 21, 1 (1969), 243–247.
[4] Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. 2015. Deep learning for event-driven stock prediction.. In *IJCAI*. 2327–2333.
[5] Shimon Kogan, Dimitry Levin, Bryan R Routledge, Jacob S Sagi, and Noah A Smith. 2009. Predicting risk from financial reports with regression. In *NAACL*. 272–280.
[6] Thien Hai Nguyen and Kiyoaki Shirai. 2015. Topic modeling based sentiment analysis on social media for stock market prediction. In *ACL*, Vol. 1. 1354–1364.
[7] Yao Qin, Dongjin Song, Haifeng Chen, Wei Cheng, Guofei Jiang, and Garrison W. Cottrell. 2017. A Dual-Stage Attention-Based Recurrent Neural Network for Time Series Prediction. In *IJCAI*. 2627–2633.
[8] Jianfeng Si, Arjun Mukherjee, Bing Liu, Qing Li, Huayi Li, and Xiaotie Deng. 2013. Exploiting Topic based Twitter Sentiment for Stock Prediction. In *ACL*. 1779–1787.
[9] Jianfeng Si, Arjun Mukherjee, Bing Liu, Sinno Jialin Pan, Qing Li, and Huayi Li. 2014. Exploiting social relations and sentiment for stock prediction. In *EMNLP*. 1139–1145.
[10] Duyu Tang, Furu Wei, Nan Yang, Ming Zhou, Ting Liu, and Bing Qin. 2014. Learning sentiment-specific word embedding for twitter sentiment classification. In *ACL*. 1555–1565.
[11] Allan Timmermann. 1996. Excess volatility and predictability of stock prices in autoregressive dividend models with learning. *The Review of Economic Studies* 63, 4 (1996), 523–557.
[12] Wei Zhang, Wen Wang, Jun Wang, and Hongyuan Zha. 2018. User-guided Hierarchical Attention Network for Multi-modal Social Image Popularity Prediction. In *WWW*. 1277–1286.
[13] Bo Zhao, Yongji He, Chunfeng Yuan, and Yihua Huang. 2016. Stock market prediction exploiting microblog sentiment analysis. In *IJCNN*. 4482–4488.
[14] Wayne Xin Zhao, Jing Jiang, Jianshu Weng, Jing He, Ee-Peng Lim, Hongfei Yan, and Xiaoming Li. 2011. Comparing twitter and traditional media using topic models. In *ECIR*. 338–349.