Supplemental Materials for Incorporating Fine-grained Events in Stock Movement Prediction

Deli Chen[1]*, Yanyan Zou[2], Keiko Harimoto[3], Ruihan Bao[3], Xuancheng Ren[1], Xu Sun[1]

[1]MOE Key Lab of Computational Linguistics, School of EECS, Peking University
[2]StatNLP Research Group, Singapore University of Technology and Design
[3]Mizuho Securities Co., Ltd.
{chendeli,renxc,xusun}@pku.edu.cn, yanyan_zou@mymail.sutd.edu.sg,
{keiko.harimoto,ruihan.bao}@mizuho-sc.com

A TOPIX Finance Event Dictionary

Table 7 shows the TOPIX Finance Event Dictionary we used in this paper for generating fine-grained events automatically. Table 1 displays all the 32 event types in 8 categories as well as their trigger words and event roles.

B Error Analysis

We observe that there are significant gaps in the model performance on different times' news. The Figure 1 shows the percentage of the three different times' news in dataset. The trade time news (44%) means news happens in trade time (9:00 AM - 15:00 PM in trade day); the out of trade time news (24%) represents the news that takes place in trade day, but not in trade time; the out of trade day news (32%) means the news happens in weekend or holiday. The experiment result shows that SSPM performs best on trade time news (Acc:68.0%, Mcc:0.361) followed by out of trade time news (Acc:65.8%, Mcc:0.312). SSPM performs worst on out of trade day news (Acc:64.6%, Mcc: 0.293). We summarize this result as the closer the news to trade time, the better our method works. It can be explained that the trade time news have the most direct influence on the stock trading because traders can make immediate reactions to the released news. As for the out of trade time news and the out of trade day news, stock traders can not make immediate reactions and their attitudes may be influenced by other factors, such as the trend of overseas exchanges and the movement of the commodities. The out of trade day news is further away from trade time compared to out of trade time news, so there are more uncertain factors which result in the decrease of model performance.

*This work is done when Deli Chen is a intern at Mizuho Securities.

C Stock Movement Label Set

For those samples in trade time, we compare the stock close price with the price of the minute news happens; for those samples out of trade time, we compare the next trade day's open price with the last trade day's close price. We only use the micro news (specific stock related news) and ignore the macro news, so we use the stock related TOPIX sector index to correct the stock change rate:

$$R_f = R_{cmp} - R_{sec}$$

R_f means the final change rate; R_{cmp} means the raw company stock change rate and R_{sec} means the stock related sector index change rate. For example, for Toyota's stock movement, we use the TPTRAN Index[1] which Toyota belongs to for correction. Then the sample is given an up/down label depends on the R_f is positive/negative.

All the sector indexs used in TOPIX are listed here: TPTRAN Index, TPNBNK Index, TPELMH Index, TPCOMM Index, TPWSAL Index, TPLAND Index, TPSERV Index, TPNCHM Index, TPRETL Index, TPPROD Index, TPINSU Index, TPMACH Index, TPFOOD Index, TPPHRM Index, TPREAL Index, TPRUBB Index, TPPREC Index, TPFINC Index, TPOIL Index, TPCONT Index, TPIRON Index, TPSECR Index, TPELEC Index, TPAIR Index, TPMINN Index, TPTEXT Index, TPNMET Index, TPGLAS Index, TPPAPR Index, TPMETL Index, TPMART Index, TPWARE Index, TPFISH Index.

D Experiment Details

The word vocabulary is extracted from the training set and we keep the top 50000 most frequent words. We use the pretrained ELMo[2] embedding with dimension of 512 and Glove pretrained

[1]TOPIX Transport Index, which indicate the overall trend of the transport market.
[2]https://github.com/allenai/allennlp

Category	Event	Trigger Word(s)	Event Role
Affairs	Announcement	announcement,announce	Who Content Time
	Legal Issues	court,litigation,lawsuit	Who Target Reason Location Requirement
	Personnel Affairs	appint,name...as	Who Title Person
	Recall	recall	Firm Amount Target Time Location
Business	Buy	buy	Firm Target Price Time
	Cooperation	alliance,partnership	Firm With-Firm Filed
	Deals	order,deal,trade	Firm Provider Product Value Number
	Demand/Supply	demand,supply	Who Type Content Time
	Investment	take...charge,invest, investment	Firm Target Money
	Sales	sale	Firm Result-number Result-rate Comparason Direction Location Time
Corporate Action	Bond Issurance	share issuance	Firm Time Type How
	Dividend	div,dividend	Firm Type Time
	Fundings	fund,funding	Who Action Target Time Location
	IPO	ipo	Firm Market IBD Value Time
	Joint venture	jv,joint venture	Firm Target-Firm Target-Rate Time
	M&A	acquisition,merge,acquire	Firm Time Method
	Share Buyback	buy...back,share repurchase	Who Share-number Money
	Share Issurance	share issuance	Who Money Purpose
	Stock Split	stock split	Firm Before-Price After-Price Change-Rate Time
	Projects&Productions	product,production,project	Firm Projects/Products Time
Eearnings	Eearnings Profit	profit,group result,earnings parent result,financial result,	Firm Type Value Change-Rate Time
	Eearnings Adjustment	earnings adjust,profit adjust	Firm Before-Value After-Value Reason
	Eearnings Forecast	group forecast,parent forecast	Firm Value Change-Rate Reason Time
ESG	Energy	solar power,plant	Who Type Movement Location
	Social	social,socirty	Who Action Purpose Time
	Government	gove,government	Country Action Target Time
Market	Market Movement	shares down ,shares up, trade hult	Firm Direction Movement
	Share Holders Acion	holder,investor	Firm Action Reason Time
Other	Comodities	oil ,coal,gassteel ,fuel crude, copper	Comodity Markey Before-Price After-Price Change-Direction Change-Rate Time
	Oversea	u.s.,us,american,china,uk,eu, europe,aisa,asian,indonesia, india,australia	Region Action Time
Ratings	Broker Ratings	target price,rating raise...price swhitch...to,scah...to	Reviewer Target-Price Directtion FirmNew Old-Price Change-Rate
	Credit Ratings	moody	Reviewer New-Rating Before-Rating Target Direction

Table 1: TOPIX Finance Event Dictionary

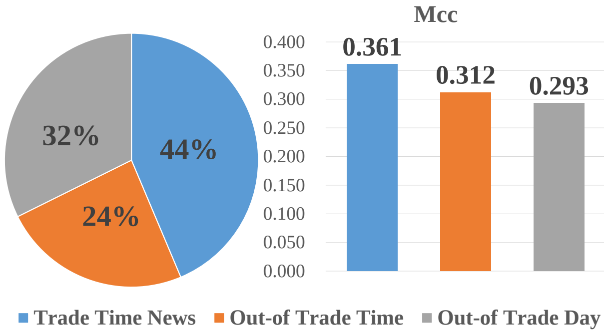

Figure 1: The percentage (all dataset) and performance (SSPM on test set) of different times' news.

embedding[3] with dimension of 300. We fine-tune the embedding weight in the training process. All the LSTMs (Graves et al., 2013) use 3 bi-directional layers and the hidden size is 256. We use the dropout regularization with the dropout probability of 0.2 to reduce overfitting. We use the Adam (Kingma and Ba, 2014) optimizer with the initial learning rate of 0.001 and the learning decay rate of 0.0005. The batch size is 128. The training epoch is 60 and we use early stop. We conduct all the experiments on 4 Nvidia Titan P100 GPUs.

We regard news before 9:10 AM as out of trade news to avoid that there is no input for trade data Besides, we regrade news after 14:50 PM as out of trade news to avoid that the input trade data is too close to the compared price. As for the minute that no trade happens, we pad it with last trade minute's value.

We do not study the multi-news for days on end, which are studied in (Xu and Cohen, 2018; Hu et al., 2018) because we find the situation that news about same stock happens in several continuous days is very sparse in real data.

We make daily prediction in this work because we find stock movement is most sensitive to daily stock trade data. Besides, daily prediction is most valuable for stock investment because clients usually use daily stock data to evaluate the performance of stocks.

We observe that one piece of news may be related to more than one stocks, so we match this kind of news with each related stock and get multi-samples. For the circumstance that there are more than one news occuring in one trade day about the same stock, which occupies a very small proportion (1.4%) in all the data samples, we connect all the news together as input.

[3] https://nlp.stanford.edu/projects/glove/

References

Alex Graves, Abdel-rahman Mohamed, and Geoffrey E. Hinton. 2013. Speech recognition with deep recurrent neural networks. In *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2013, Vancouver, BC, Canada, May 26-31, 2013*, pages 6645–6649.

Ziniu Hu, Weiqing Liu, Jiang Bian, Xuanzhe Liu, and Tie-Yan Liu. 2018. Listening to chaotic whispers: A deep learning framework for news-oriented stock trend prediction. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, WSDM 2018, Marina Del Rey, CA, USA, February 5-9, 2018*, pages 261–269.

Diederik P. Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980.

Yumo Xu and Shay B. Cohen. 2018. Stock movement prediction from tweets and historical prices. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers*, pages 1970–1979.