

Generating Market Comments Referring to External Resources

Tatsuya Aoki^{♣♣} Akira Miyazawa^{◇♡♣} Tatsuya Ishigaki^{♣♣} Keiichi Goshima^{#♣}
Kasumi Aoki[♣] Kobayashi Ichiro[♣] Hiroya Takamura^{♣♣} Yusuke Miyao[♣]

[♣]Tokyo Institute of Technology [◇]The Graduate University for Advanced Studies

[♡]National Institute of Informatics [#]Bank of Japan* [♣]Ochanomizu University

[♣]National Institute of Advanced Industrial Science and Technology [♣]The University of Tokyo

{aoki, ishigaki}@lr.pi.titech.ac.jp miyazawa-a@nii.ac.jp
keiichi.goshima@boj.or.jp {g1120501, koba}@is.ocha.ac.jp
takamura@pi.titech.ac.jp yusuke@is.s.u-tokyo.ac.jp

Abstract

Comments on a stock market often include the reason or cause of changes in stock prices, such as “*Nikkei turns lower as yen’s rise hits exporters.*” Generating such informative sentences requires capturing the relationship between different resources, including a target stock price. In this paper, we propose a model for automatically generating such informative market comments that refer to external resources. We evaluated our model through an automatic metric in terms of BLEU and human evaluation done by an expert in finance. The results show that our model outperforms the existing model both in BLEU scores and human judgment.

1 Introduction

Nikkei Stock Average opens at a high price after Dow Jones Industrial Average closes at a high price.

This is an example of a comment on markets that describes the stock prices shown in Figure 1. The closing price of the Dow Jones Industrial Average at 5 am JST is represented as the right-most point in the figure on the top, while the opening price of the Nikkei Stock Average is represented as the left-most point in the figure at the bottom. While the comment describes the behavior of Nikkei Stock Average (henceforth, *Nikkei 225*), the main indicator of the Japanese stock market, it also refers to an external indicator, the Dow Jones Industrial Average, which represents the US stock market. Such

*Views expressed in this paper are those of the authors and do not necessarily reflect the official views of the Bank of Japan.

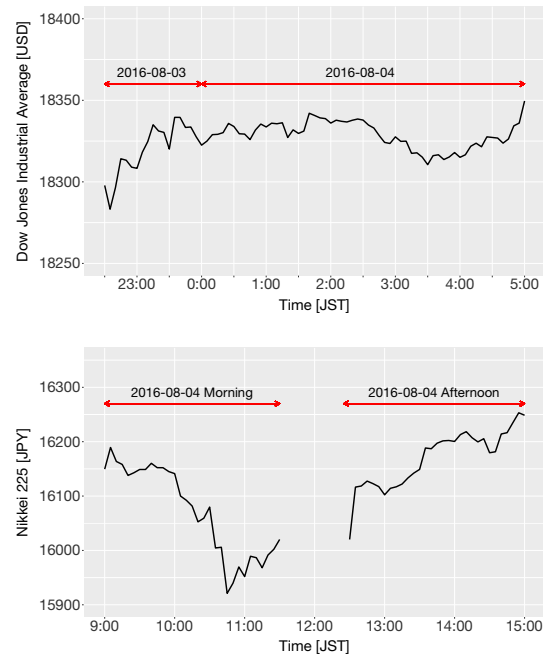


Figure 1: Relationship between Dow Jones Industrial Average and Nikkei 225 (Nikkei Stock Average).

mentions of external resources as a cause of the behavior of a target index are very common in market comments. Comments on the Japanese stock market can also refer to, for example, other stock market indices, foreign exchange rates, and oil prices, and comments that also describe causes will facilitate readers in understanding financial situations.

In this paper, we address the task of generating market comments that refer to external resources. Specifically, we extend the encoder in the encoder-decoder model proposed by Murakami et al. (2017) so that the model can take into account external resources related to the financial do-

main. We encode each of the external resources in addition to the target index, i.e., Nikkei 225, and feed them to the decoder. The experimental results show that our proposed model outperforms the existing single-source model in terms of the fluency and informativeness of human evaluation, in addition to the BLEU score.

2 Related Work

There has been a lot of work on generating text from numerical time series or structured data including weather data (Belz, 2008), healthcare data (Portet et al., 2009), sports data (Liang et al., 2009), and market data (Kukich, 1983). Approaches to such tasks are traditionally dependent on hand-crafted rules (Goldberg et al., 1994; Dale et al., 2003) or are template-based.

Neural encoder-decoders (Sutskever et al., 2014; Bahdanau et al., 2015; Luong et al., 2015) have also been successfully applied to various data-to-text generation tasks. While many generate text from table data, such as reviews from product attributes (Dong et al., 2017) and biographies from the infoboxes of Wikipedia (Lebret et al., 2016), there is an attempt to generate text from numerical data (Murakami et al., 2017), in which market comments are generated from a time-series of stock prices. However, the model of Murakami et al. (2017), which is based on an encoder-decoder, takes only a target time series and ignores the fact that there are many mentions of external resources.

3 Generating Market Comments

We describe our model for generating comments. We extend the encoder part of the model proposed by Murakami et al. (2017), which had a limitation in generating informative market comments due to the lack of a capability to consider multiple data sources as input. We first explain the encoder used in the existing model and then show how we extend it.

3.1 Base Model (BASE)

The existing model by Murakami et al. (2017) takes only a single source of data, a sequence of prices of Nikkei 225, as input. Specifically, the prices are recorded every five minutes in the data. The model first converts the input data into two vectors: a short-term vector $\mathbf{x}_{\text{short}}$ and a long-term vector \mathbf{x}_{long} . The vector $\mathbf{x}_{\text{short}}$ is N -dimensional

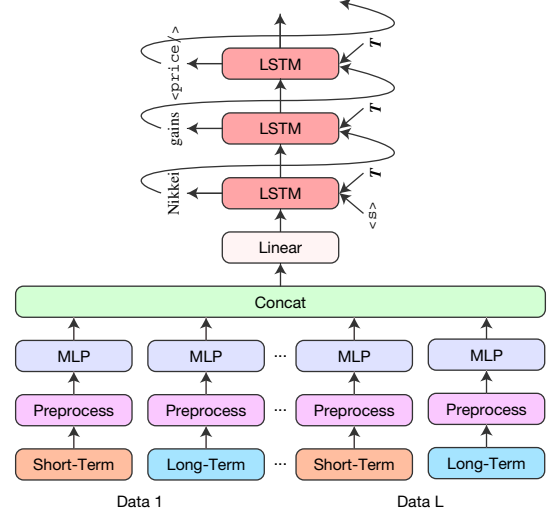


Figure 2: Neural-network architecture of our model. The symbol T denotes embedding of publishing time (see Murakami et al. (2017) for details).

and consists of the N previous stock prices, while \mathbf{x}_{long} is M -dimensional and consists of the closing prices of the M preceding trading days. Thus, $\mathbf{x}_{\text{short}}$ contains short-term changes in the stock price, while \mathbf{x}_{long} contains long-term changes.

In the encoding step, the vectors are passed to multilayer perceptrons (MLPs) with three layers and concatenated as

$$\mathbf{v}_{\text{single}} = [\text{MLP}_{\text{short}}(\mathbf{x}_{\text{short}}); \text{MLP}_{\text{long}}(\mathbf{x}_{\text{long}})], \quad (1)$$

where the semicolon represents the concatenation. The vector $\mathbf{v}_{\text{single}}$ is then transformed to a vector \mathbf{s}_0 by an affine transformation $\mathbf{s}_0 = W_s \mathbf{v}_{\text{single}} + \mathbf{b}_s$, where W_s is a weight matrix and \mathbf{b}_s is a bias term.

In the decoding step, the hidden state of the decoder is initialized by \mathbf{s}_0 , and LSTM cells (Hochreiter and Schmidhuber, 1997) are used following the model by Murakami et al. (2017). Please refer to the original paper for more details on the decoder.

They also replaced numerical values in the training data with placeholders representing arithmetic operations, e.g., rounding down the difference between the latest price and the closing price of the previous day.

3.2 Multiple Source-Aware Model (MULTI)

The architecture of our model is shown in Figure 2. We extend the encoder part of the base model so that the model can take L different sources as input,

including the Dow Jones Industrial Average, and US dollar/Japanese yen exchange rates in addition to Nikkei 225. We convert each input source to a continuous representation \mathbf{v}^i ($1 \leq i \leq L$) as

$$\mathbf{v}^i = \left[\text{MLP}_{\text{short}}^i(\mathbf{x}_{\text{short}}^i); \text{MLP}_{\text{long}}^i(\mathbf{x}_{\text{long}}^i) \right]. \quad (2)$$

Note that the model has $2L$ MLPs; L MLPs are for short-term data and the others are for long-term data. Each $\mathbf{x}_{\text{short}}^i$ is an N -dimensional short-term vector for the i -th data source generated with the same approach as Murakami et al. (2017). Each $\mathbf{x}_{\text{long}}^i$ is an M -dimensional long-term vector for the i -th data source.

The representations $\mathbf{v}^1, \dots, \mathbf{v}^L$ are then concatenated to a representation $\mathbf{v}_{\text{multi}}$ as:

$$\mathbf{v}_{\text{multi}} = [\mathbf{v}^1; \dots; \mathbf{v}^L]. \quad (3)$$

It is then passed to an affine transform function as is done in the base model $\mathbf{s}_0 = W_m \mathbf{v}_{\text{multi}} + \mathbf{b}_m$, and \mathbf{s}_0 is used for the initial state of the decoder.

Our model is clearly a straight extension of the model by Murakami et al. (2017). All the multiple input resources are treated equally with this architecture. The target resource to be described will be determined by the training data. For example, if the comments in the training data describe the behavior of Nikkei 225, the other resources are regarded as causes influencing Nikkei 225.

4 Experiments

4.1 Data

Training the model requires pairs consisting of a time series and a market comment aligned with it. As market comments, we used 20,093 headlines of Nikkei Quick News (NQN) that describe the behavior of Nikkei 225. They are provided by Nikkei, Inc. and written in Japanese. We divided them into three parts on the basis of the period of publication: 16,276 for training (Dec. 2010–Oct. 2015), 1,866 for validation (Oct. 2015–April 2016) and 1,951 for testing (April 2016–Oct. 2016). In addition, we retrieved the five-minute charts of 10 indices from Thomson Reuters DataScope Select¹. They consist of seven stock market indices (Nikkei 225, TOPIX Price Index, S&P 500 Index, FTSE 100 Index, Hang Seng Index, Shanghai SE Composite Index, and Dow Jones Industrial Index),

¹<https://hosted.datascope.reuters.com/DataScope/>

a forward transaction index (Nikkei 225 Future), and two currency exchange rates, USD/JPY and EUR/JPY.

4.2 Preprocessing and Parameters

As a preprocessing procedure, we created short- and long-term sequences of each index from the five-minute charts in the same way as Murakami et al. (2017). The size N of a short-term vector was set to 62, and the size M of a long-term one was set to 7. We used Adam (Kingma and Ba, 2015) for optimization with a learning rate of 0.0001 and a mini-batch size of 100. The dimensions of the three hidden layers in MLPs were all set to 32.

4.3 Evaluation Settings

We compared our model with the model by Murakami et al. (2017). The latter was not provided with external resources as input, but could still refer to them groundlessly simply because mentions of external resources are found in the training data.

We conducted both an automatic evaluation in terms of BLEU scores and a manual evaluation done by a financial expert. The outputs from the proposed model were compared with reference market comments extracted from NQN and comments generated by the base model. In the automatic evaluation by BLEU score, we used the market comments collected from NQN as references. We calculated the BLEU scores for both the base model and our model. In the human evaluation, a human judge (an expert in finance) manually judged the outputs in terms of two criteria: *fluency* and *informativeness*. Specifically, we presented three market comments generated by a human (HUMAN), the base model (BASE), and our model (MULTI). For fluency, the human judge manually selected a label from two labels (*fluent* and *not_fluent*) for each comment. For informativeness, the judge was asked to evaluate whether a comment included a correct mention of an external resource. The human judge was asked to select one out of four labels: *no*, *correct*, *wrong*, and *subtle*. The label *no* means that a comment did not contain a mention of an external resource. The label *correct* means that the comment contained correct mentions of external resources, whereas the label *wrong* means that the comment contained a wrong mention of external resources. The label *subtle* corresponded to the other cases. For example, when a comment contained a mention of an external resource that was

Method	BLEU (%)
BASE	21.88 \pm 0.31
MULTI	23.66 \pm 0.35

Table 1: Result of evaluation in terms of BLEU. Scores were macro-averaged over 5 runs. Each value after \pm is standard deviation.

	fluent	not_fluent
HUMAN	98	2
BASE	95	5
MULTI	96	4

(a) Fluency.

	Mentions of Ext. Resources		
	no	yes	(cr / wr / sb)
HUMAN	54	46	(5 / 0 / 41)
BASE	51	49	(13 / 9 / 27)
MULTI	46	54	(11 / 2 / 41)

(b) Informativeness.

Table 2: Result of human evaluation. In (a), values are number of times that comments were judged fluent or not_fluent. In (b), no indicates number of comments that do not contain any mention of external resources. yes indicates number of comments that contain mention of external resources. yes is divided into correct (cr), wrong (wr), and subtle (sb), which respectively mean numbers of comments with correct, wrong, and subtle mentions.

not any of the L inputs, subtle was assigned. When evaluating the informativeness, the human judge does not simply measure the similarity between the generated comments and the reference comments; he referred to the input data to check the correctness of the generated comments.

5 Results

Table 1 shows the BLEU scores for each model. The scores were calculated by averaging the scores of five trials. By incorporating multiple resources as input, our model outperformed the base model with an improvement of 1.78 points in BLEU. This suggests that integrating multiple resources into the encoder helps to improve the ability to generate comments similar to human generated ones.

Table 2 shows the results of the human evaluation for each model. In terms of fluency, most of the comments generated by all of the methods were judged fluent. BASE and MULTI were slightly worse than HUMAN in fluency because they failed to output the correct placeholders representing arithmetic operations.

In terms of informativeness, our model referred to external resources more often than BASE. Specifically, our model outputs 54 comments with mentions of external resources, while 46 were without the mentions. The method BASE outputs only 49 comments with such a mention. In addition, the proportion of wrong was notably reduced by our model. The results suggest that our proposed model improved the ability to generate more informative sentences including correct mentions of external resources.

We show examples of the generated comments in Table 3. The method BASE erroneously mentioned external information, “US stock rise,” due to the lack of input information. Our method, MULTI, tended to avoid generating clearly erroneous mentions such as “US stock rise.” We also found that HUMAN often referred to important events as in the output example “easing Brexit concerns.” Generating such comments requires yet other external resources such as news streams, which we leave for future work.

6 Conclusion

We proposed an encoder-decoder model for generating market comments that refer to external resources. Our automatic and manual evaluation showed that integrating multiple resources into the encoder improves the ability to include such information in the outputs and to generate more informative comments.

Our code is available at <https://github.com/aistairc/market-reporter>.

Acknowledgements

This paper is based on results obtained from a project commissioned by the New Energy and Industrial Technology Development Organization (NEDO). This work is partially supported by JST PRESTO (Grant Number JPMJPR1655).

Method	Output
HUMAN	Toushou yoritsuki zokushin, agehaba 300 en koeru, ei EU ridatsu kenen-ga koutai TSE opening continual_rise, gain 300 yen jump_over, UK EU leaving concern-NOM retreat “Tokyo stocks open 300 yen higher with a continual rise, due to easing Brexit concerns.”
BASE	Toushou yoritsuki zokushin, agehaba 300 en chou, bei-kabu-daka ya en-yasu-o koukan TSE opening continual_rise, gain 300 yen over US-stock-high and yen-cheap-ACC good_feeling “Tokyo stocks open 300 yen higher with a continual rise, helped by a cheaper yen and US stocks rise.”
MULTI	Toushou yoritsuki zokushin, agehaba 300 en chou, en-yasu-de yushutsu-kabu-ni kai TSE opening continual_rise, gain 300 yen over yen-cheap-INS exporting-stock-DAT purchase “Tokyo stocks open ... a continual rise, thanks to demand for export-related shares boosted by a cheaper yen.”

Table 3: Examples of generated comments. Each example is accompanied by **original Japanese comment** transliterated into English alphabet, its word-for-word translation, and the corresponding English **sentence**. TSE stands for Tokyo Stock Exchange. Abbreviations used in word-for-word translation are as follows. NOM: nominative, ACC: accusative, INS: instrumental, and DAT: dative.

References

- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *Proceedings of International Conference on Learning Representations (ICLR)*.
- Anja Belz. 2008. Automatic generation of weather forecast texts using comprehensive probabilistic generation-space models. *Natural Language Engineering*, 14(4):431–455.
- Robert Dale, Sabine Geldof, and Jean-Philippe Prost. 2003. Coral: Using natural language generation for navigational assistance. In *Proceedings of the 26th Australasian computer science conference-Volume 16*, pages 35–44. Australian Computer Society, Inc.
- Li Dong, Shaohan Huang, Furu Wei, Mirella Lapata, Ming Zhou, and Ke Xu. 2017. [Learning to generate product reviews from attributes](#). In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, volume 1, pages 623–632.
- Eli Goldberg, Norbert Driedger, and Richard I Kit-tredge. 1994. Using natural-language processing to produce weather forecasts. *IEEE Intelligent Systems*, (2):45–53.
- Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation*, 9(8):1735–1780.
- Diederik P. Kingma and Jimmy Ba. 2015. [Adam: A method for stochastic optimization](#). In *Proceedings of International Conference on Learning Representations (ICLR)*.
- Karen Kukich. 1983. Design of a knowledge-based report generator. In *Proceedings of Association for Computational Linguistics (ACL)*, pages 145–150.
- Rémi Lebret, David Grangier, and Michael Auli. 2016. [Neural text generation from structured data with application to the biography domain](#). In *Proceedings of Empirical Methods in Natural Language Processing (EMNLP)*, pages 1203–1213.
- Percy Liang, Michael I. Jordan, and Dan Klein. 2009. [Learning semantic correspondences with less supervision](#). In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL (ACL2009)*, pages 91–99. Association for Computational Linguistics.
- Thang Luong, Hieu Pham, and Christopher D. Manning. 2015. [Effective approaches to attention-based neural machine translation](#). In *Proceedings of Empirical Methods in Natural Language Processing (EMNLP)*, pages 1412–1421.
- Soichiro Murakami, Akihiko Watanabe, Akira Miyazawa, Keiichi Goshima, Toshihiko Yanase, Hiroya Takamura, and Yusuke Miyao. 2017. [Learning to generate market comments from stock prices](#). In *Proceedings of Association for Computational Linguistics (ACL)*, pages 1374–1384.
- François Portet, Ehud Reiter, Albert Gatt, Jim Hunter, Somayajulu Sripada, Yvonne Freer, and Cindy Sykes. 2009. Automatic generation of textual summaries from neonatal intensive care data. *Artificial Intelligence*, 173(7-8):789–816.
- Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems (NIPS)*, pages 3104–3112.