# Learning to Generate Market Comments from Stock Prices

National Institute of Advanced Industrial Science and Technology (AIST), Japan

Soichiro Murakami, Akihiko Watanabe, Akira Miyazawa, Keiichi Goshima, Toshihiko Yanase, Hiroya Takamura, Yusuke Miyao

Introduction

- A novel encoder-decoder model for automatically generating market comments from stock prices.
- Characteristic problems of comment generation for time-series of stock prices.
    - a. Content varies depending on short- or long-term changes of the time-series data.
    - b. Time-dependent expressions are used.
    - c. Numerical values obtained through arithmetic operations are often described.
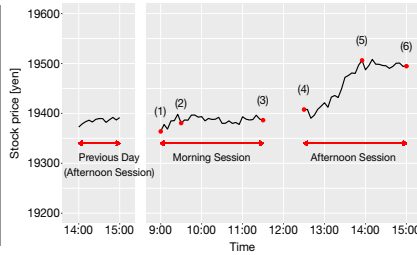
Time	Comment	Type of problem
(1) 09:00	Nikkei opens with a continual fall.	a, b
(2) 09:29	Nikkei turns to rise.	a
(3) 11:30	Nikkei continues to fall. The closing price of the morning session decreases by 5 yen to 19,386 yen.	a, b, c
(4) 12:30	Nikkei rises at the beginning of the afternoon session.	b
(5) 13:54	Nikkei gains more than 100 yen.	c
(6) 15:00	Nikkei rebounds and closes up 102 yen to 19,494 yen.	a, b, c

Generating Market Comments

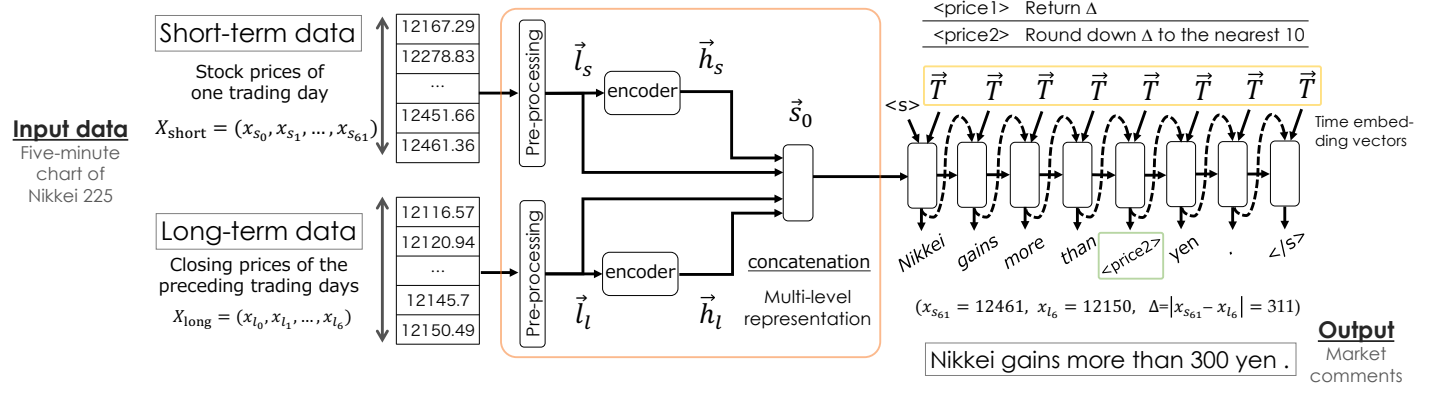### a. Encoding Numerical Time-Series Data
- Short- and long-term data to capture various types of change in different time-scales.
- Two preprocessing methods: *standardization* and *moving reference*.

b. Incorporating Time Embedding
- *Time embedding* vectors on the basis of the time when the comment is delivered.
- A time embedding vector $\vec{T}$ is added to each hidden state in decoding.

c. Estimation of Arithmetic Operations
- Each generalization tag represents a type of arithmetic operation.
- The model performs the operations on the designated values in accordance with the tag.

Input data — Five-minute chart of Nikkei 225

Short-term data — Stock prices of one trading day
$X_{\text{short}} = (x_{s_0}, x_{s_1}, \ldots, x_{s_{61}})$

12167.29
12278.83
...
12451.66
12461.36

Long-term data — Closing prices of the preceding trading days
$X_{\text{long}} = (x_{l_0}, x_{l_1}, \ldots, x_{l_6})$

12116.57
12120.94
...
12145.7
12150.49

Pre-processing → $\vec{l}_s$ → encoder → $\vec{h}_s$

Pre-processing → $\vec{l}_l$ → encoder → $\vec{h}_l$

$\vec{s}_0$ — concatenation — Multi-level representation

Example of the generalization tags

<price1>	Return Δ
<price2>	Round down Δ to the nearest 10

$\vec{T}$ $\vec{T}$ $\vec{T}$ $\vec{T}$ $\vec{T}$ $\vec{T}$ $\vec{T}$ $\vec{T}$ — Time embedding vectors

<s> Nikkei gains more than <price2> yen . </s>

$(x_{s_{61}} = 12461,\ x_{l_6} = 12150,\ \Delta = |x_{s_{61}} - x_{l_6}| = 311)$

Output — Market comments

Nikkei gains more than 300 yen .

Experiments

- Dataset
    - Five-minute chart of Nikkei 225 from March 2013 to October 2016.
    - 7,351 descriptions as market comments.
        - 5,880 for training, 730 for validation, and 741 for testing.
        - 100 descriptions for a human evaluation.
- Evaluation metrics
    - BLEU
    - F-measures for time-dependent expressions
    - Human evaluation in terms of informativeness or fluency
- Model
    - Three types of models:
        - Baseline
        - Full models (e.g., *mlp-enc*)
        - Ablated models (e.g., *-short*)

|	Encoder	x_{short}	x_{long}	Standardization	Moving reference	Multi-level representation	Operation	Time-embedding
baseline	MLP	✓	—	✓	✓	✓	—	—
mlp-enc	MLP	✓	✓	✓	✓	✓	✓	✓
cnn-enc	CNN	✓	✓	✓	✓	✓	✓	✓
rnn-enc	RNN	✓	✓	✓	✓	✓	✓	✓
-short	MLP	—	✓	✓	✓	✓	✓	✓
-long	MLP	✓	—	✓	✓	✓	✓	✓
-std	MLP	✓	✓	—	✓	✓	✓	✓
-move	MLP	✓	✓	✓	—	✓	✓	✓
-multi	MLP	✓	✓	✓	✓	—	✓	✓
-num	MLP	✓	✓	✓	✓	✓	—	✓
-time	MLP	✓	✓	✓	✓	✓	✓	—

Experimental Results
- BLEU
    - The model using both MLP as encoders and all the techniques we developed, *mlp-enc*, outperformed *baseline* and the other models.

Model	baseline	mlp-enc	cnn-enc	rnn-enc	-short	-long
BLEU	0.243	**0.464**	0.449	0.454	0.380	0.433

| Model	-std	-move	-multi	-num	-time
BLEU	0.455	0.393	0.435	0.318	0.395

Experimental Results
- F-measures
    - *mlp-enc* output phrases mentioning changes more appropriately.
    - The model achieved the higher values than *baseline*.

- Human evaluation
    - *mlp-enc* achieved a quality comparable even to *Human*.
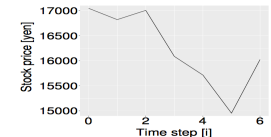   - *mlp-enc* significantly outperformed *baseline* in terms of informativeness.
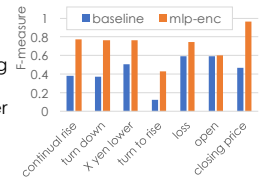
Model	Info.	Fluency	External
Human	95	95	25
mlp-enc	85	93	1
baseline	28	100	6
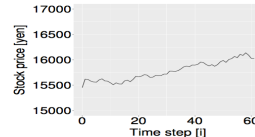
Model	Fluency	Informativeness	Output
baseline	1	0	*Nikkei rebounds. The closing price of the morning session is <unk> yen, which is 81 yen higher.*
mlp-enc, Human	1	1	*Nikkei significantly rebounds. The closing price is 16,022 yen, which is 1,069 yen higher.*

Conclusion

- A novel encoder-decoder model to automatically generate market comments from numerical time-series data of stock prices.
- We developed approaches for generating comments that have these characteristics and showed the effectiveness of the proposed model.