GENERATING MARKETS COMMENTS ON STOCK PRICE FLUCTUATIONS USING NEURAL NETWORKS

IBARAKI UNIVERSITY

IBUKI SEKINO , 21NM734A@VC.IBARAKI.AC.JP

MINORU SASAKI , MINORU.SASAKI.01@VC.IBRAKI.AC.JP

RESUME

- 2021 March Graduated from Ibaraki University, Faculty of Engineering, Department of Information Engineering.
- 2021 April Entered Ibaraki University, Graduate School of Science and Engineering, Department of Information Technology

PROJECTS

- I am currently attempting to use Japanese stock price data and articles to generate market commentary.
- In this study, I have succeeded in generating some of the market commentary using the Nikkei Stock Average and an article called NQN.
- Next, I would like to proceed with the research with a view to factor analysis.
- From the next page will be the research I did as an undergraduate student, "Generating Expressions on Stock Price Fluctuations Using the Nikkei Stock Average and NQN.

TABLE OF CONTENTS

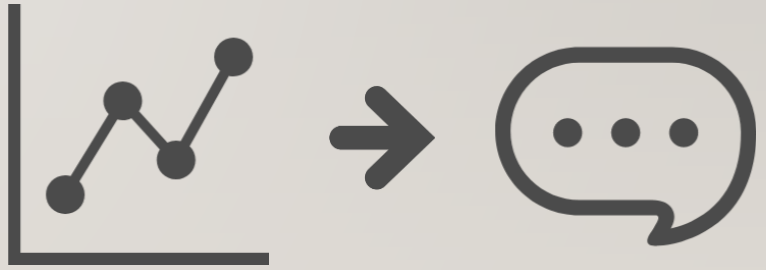
1. BACKGROUND AND PURPOSE OF THE STUDY

- In recent years, there have been more opportunities to handle data in various fields.
- If the data is large and complex, it takes time to analyze.
- One of the effective ways to use data is the "data-to-text" technology.

1. BACKGROUND AND PURPOSE OF THE STUDY

- " Data-to-text" is a method of expressing a summary of data in text.

- The task of generating market price comments from stock prices, which is dealt with in this study, is also a type of Data-to-text.



1. BACKGROUND AND PURPOSE OF THE STUDY

- Currently, analysts write full market comments.
- →But it takes too much effort.

- Therefore, we propose a method for generating a part of the market comment in order to reduce the effort required for analysts to generate market comment.

2. PROPOSED METHOD

- **Outline**
- 1. Extract expressions related to stock price fluctuation and their fluctuation ranges from articles.
- 2.Learn stock price fluctuation and expressions of stock prices.
- 3. Generate a part of the market comment.
- 4.Combine a part of the market comments into completed market comments. (But, in this study, I evaluated up to step 3.)

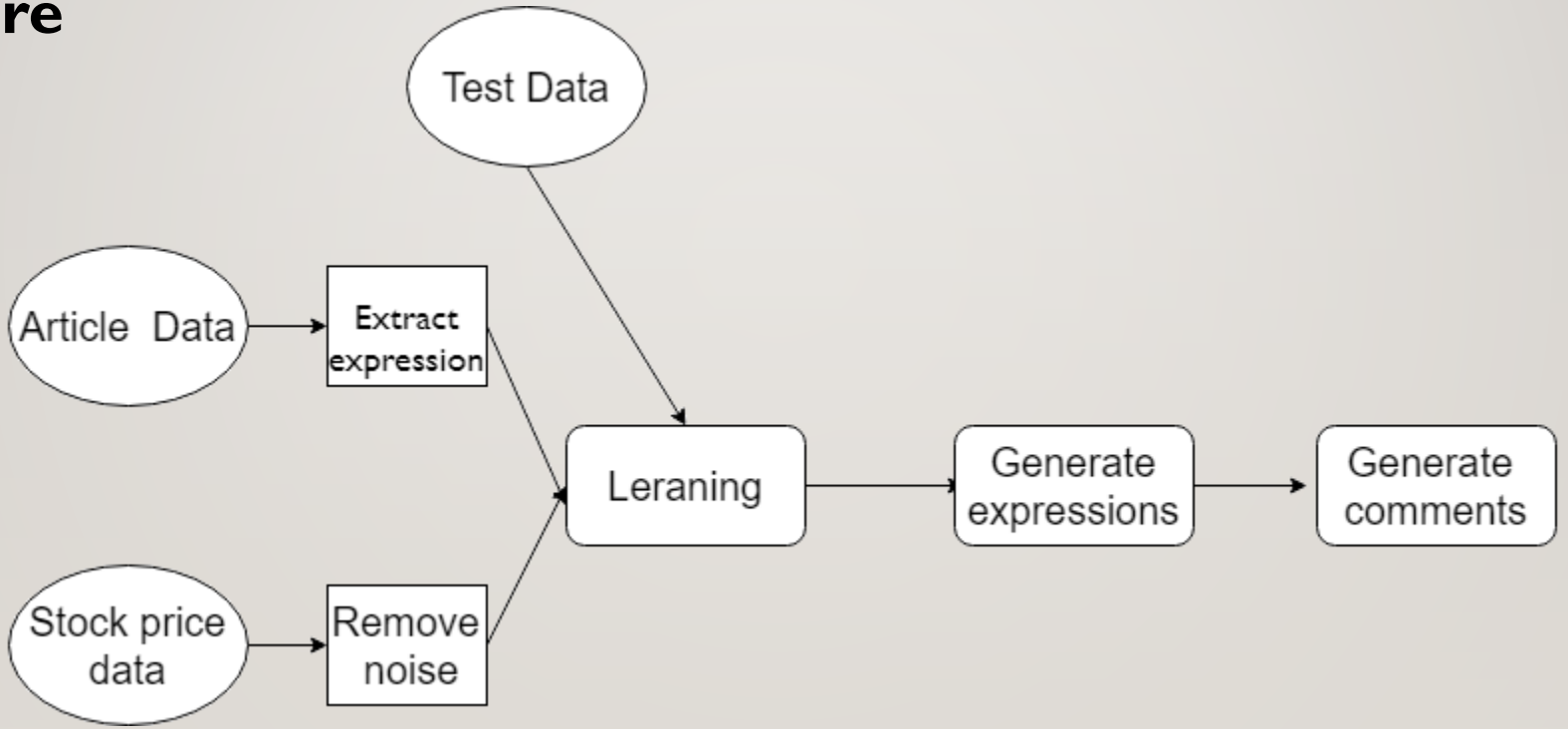
2. PROPOSED METHOD

- **Extract expressions**

Expressions	Explanation
Continue to rise	The stock prices continued to rise.
Continue to decline	The stock prices continued to decline.
Reactionary fall	A sudden drop in a stock price that has been rising.
Rebound	A sudden rise in a stock price that has been declining.
Large or Small, continue to rise	The stock price continued to rise significantly or slightly.
Large or Small, continue to decline	The stock price continued to decline significantly or slightly.
Large or Small, reactionary fall	The stock price reactionary fall significantly or slightly.
Large or Small, rebound	The stock price rebound significantly or slightly.

2. PROPOSED METHOD

- **Figure**



2. PROPOSED METHOD

- **Used Data**/Article data(Nikkei Quick News form 2014-2017.)
- Red text was extracted.

Time_Stamp_ Original(JST)	Headline
2017-01-04 15:00:00	<NQN>◆日経平均、４日ぶり大幅反発　大引けは479円高の１万9594円　昨年来高値を更新 (Nikkei 225 rebounds larger for the first time in four days, closing up 479 yen to 19,594 yen, a new record high for last year.)
2017-01-04 15:03:00	<NQN>◇日経平均大引け、４年ぶり大発会高値　昨年来高値を更新、景況感改善が支え (Nikkei 225 closes at 4-year high, supported by improved business confidence)

2. PROPOSED METHOD

- **Used Data**/Stock price data(Nikkei 225 5-minute chart from 2014-2017.)

Day	Time	Open price	High price	Low price	Close price
2017/1/4	9:00:00	19298.68	19351.47	19277.93	19351.47.1
2017/1/4	9:05:00	19354.42	19362.40	19335.90	19352.44
2017/1/4	9:10:00	19358.04	19390.47	19358.04	19387.76

2. PROPOSED METHOD

- **How to remove noise on stock price (pre-processing).**
- The following equations were used to remove noise from the stock price.

$$x_{std} = \frac{x_i - \mu}{\theta}$$

$$x_{move} = x_i - r_i$$

- To capture long-term and short-term fluctuations in stock prices, we prepared data Xlong, which uses the closing price (15:00) as input, and Xshort, which consists of 62 time steps (for one day).

2. PROPOSED METHOD

- **Learning**
- I used MLP.
- Max_iter (the maximum number of learning iterations) is set to 10000, and the other settings are default.
- Examples of generated expressions….続伸(continue to rise)…

*The generated expressions are in Japanese.

2. PROPOSED METHOD

- **Evaluation**
- Compare the generated expressions with the expressions extracted from the article.
- As a comparison, the four types of data used during training were used.

Data name	Explanation
Xlong_move	Xlong with pre-processing“move”.
Xlong_std	Xlong with pre-processing“std”.
Xlong_move_std	Xlong with pre-processing“move” and “std”.
Xshort_move	Xshort with pre-processing“move”.

2. PROPOSED METHOD

- **About short**
- Since A was unable to extract expressions due to the form of the article, the results of comment generation using the rule base are shown.
- **Previous Study**
- As a comparison, we describe some of the results of a previous study, Murakami et al. This result was the highest value in the previous study.
- *Neither Short nor Previous study generated any expression for the stock price fluctuation range, so I wrote "-"in the results.

3. EXPERIMENTAL RESULTS

- L…Large, S…Small
- Red…Highest one except short.
- Blue…Highest one, but short.

expression	Xlong_move	Xlong_std	Xlong_move_std	Xshort_move	Previous Study
Rebound	0.90	0.85	0.91	0.98	0.803
Reactionary fall	0.94	0.90	0.90	0.98	0.748
Rebound L.	0.62	0.38	0.60	-	-
Reactionary fall L.	0.55	0.60	0.44	-	-
Continue to rise L.	0.00	0.77	0.00	-	-
Continue to decline L.	0.60	0.69	0.63	-	-
Rebound S.	0.00	0.00	0.00	-	-
Reactionary fall S.	0.00	0.00	0.00	-	-
Continue to rise S.	0.00	0.00	0.46	-	-
Continue to decline S.	0.00	0.00	0.50	-	-
Continue to rise.	0.90	0.89	0.88	1.00	0.814
Continue to decline.	0.89	0.87	0.90	1.00	0.753

4.DISUCSS

- **Why are there few expressions in SHORT?**
- Short was created to view stock price fluctuations for a day, but NQN did not generate market comments containing the expression multiple times a day, so it was not possible to extract the expression.
- So I tried to generate comments using a rule base, but the article did not specify what criteria to use to generate Lage or Small.
- The high results are due to the small number of expressions and the method of setting the threshold needs to be considered if it is to be generated on a rule-based.

DISCUSS

- **Reasons for the improved results compared to previous studies.**
- Results compared with previous studies.(This study refers to Xlong_move_std)

Expressions	This study	Previous dtudy
Continue to decline	0.91	0.803
Continue to rise	0.90	0.748
Rebound	0.88	0.814
Reactionary fall	0.90	0.753

- This is because this study treats the similar expressions in the previous studies as identical.
- Ex)" start to move up","Reactionary fall"→ "Reactionary fall"

DISCUSS

- **Why there were zeroes in the results.**
- Looking at the results, there are a few that have a generation rate of 0.
- The rate of stock price fluctuation generation. Number of expression in the article..

expression	Xlong_move_std
Rebound L.	0.60
Reactionary fall L.	0.44
Continue to rise L.	0.00
Continue to decline L.	0.63
Rebound S.	0.00
Reactionary fall S.	0.00
Continue to rise L.	0.46
Continue to decline S.	0.5

Expressions	number
Continue to rise	184
Rebound	150
Reactionary fall	147
Continue to decline	111
Rebound L.	24
Continue to decline L.	24
Continue to decline S.	20
Reactionary fall L.	17
Continue to rise S.	17
Reactionary fall S.	14
Continue to rise L.	13
Rebound S.	8

DISCUSS

- Comparing the tables, we can see that the comment generation rate is lower when there are fewer expressions appearing in the article.
- This is natural in machine learning, and if we want to remove 0,we need better training data or we need to review the expressions to be extracted.

CONCLUSION

- I used Nikkei QUICK News and Nikkei Stock Average to generate market comment.
- High values were obtained in the results of data with two types of preprocessing implemented.
- In order to produce "expressions representing the fluctuation range of stock prices" with short-term data, it is necessary to study the method of determining the threshold.
- The reason why the generation rate of the expression was higher than that of the previous study is thought to be because similar expressions were grouped together.
- Better training data is needed to produce "expressions representing the fluctuation range of stock prices. Or, the expression needs to be revised.