Extraction of News Articles Related to Stock Price Fluctuation Using Sentiment Expression

Kazuto Tanaka, Minoru Sasaki

Kazuto Tanaka

Ibaraki University

20nm715g@vc.ibaraki.ac.jp

茨城大学
Ibaraki University

IARIA

Short resume

Kazuto Tanaka

He received his B.A. degree from Ibaraki University, Japan in 2020.

He is currently a master's-degree student at the Graduate School of Science and Engineering, Ibaraki University.
His research interests include financial text processing.

Research Background

Reports refer to a variety of sources, including news articles.
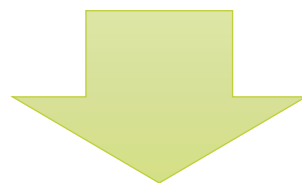→The amount of data is too large to refer to everything.
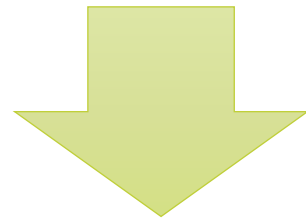
Extract important articles from a large amount of article data.

Related Works

In a previous study, Sakai et al proposed to use Japanese financial news to extract information on the causes of fluctuations in corporate performance.

→But They're not paying attention to the Nikkei Stock Average.

So we conducted an experiment by focusing on the Nikkei Stock Average.

Purpose of the research

- Purpose

  To extract articles related to changes in the Nikkei Stock Average every month.

- Subtarget

  Extract articles related to one sentence in the report.

2016年5月の日経平均株価

　5月の日経平均株価は月末終値が１万7234円98銭と前月末に比べ568円93銭（3.4％）上昇した。日銀が４月28日の金融政策決定会合で追加緩和を見送った流れを受けて月初は弱含み、５月２日の取引時間中に１万6000円を割り込んだ。中旬に主要企業の決算発表が一巡し、主要国首脳会議（伊勢志摩サミット）が開かれた26、27日に向け、政府の財政出動への期待が高まった。月末には外国為替相場が円安・ドル高に動き、輸出関連株に見直し買いが入って約１ヶ月ぶりに１万7000円の大台を回復した。
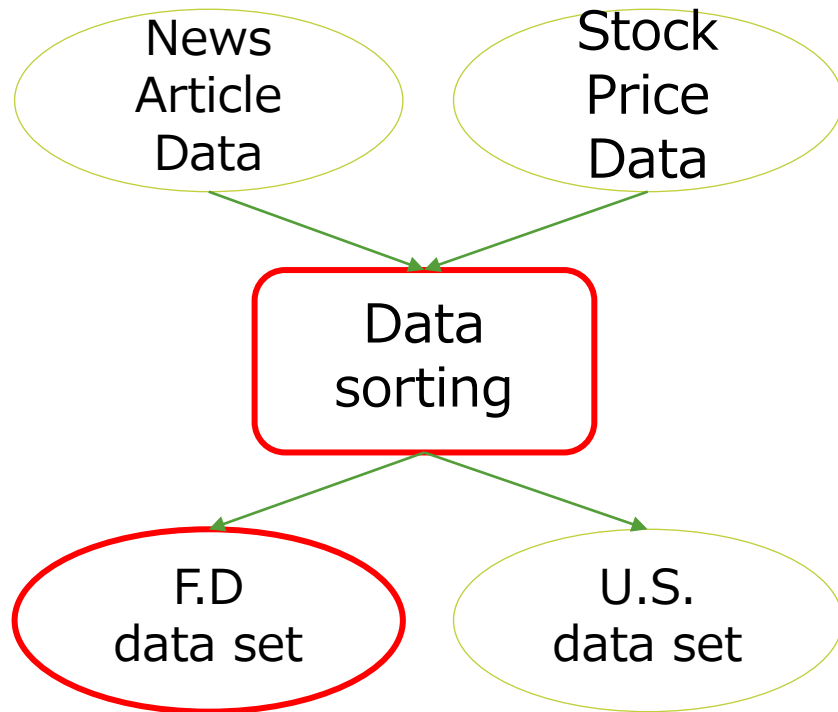
Methods and data used

- Methods used

  Sentiment Expression: Converting the impression of a word into a numerical value

  Ex）幸福(happy) → １，減少(decrease) → -0.5

- Data used

  Nikkei Stock Average Information, Nikkei QUICK News, sentiment dictionary

Proposed method

News Article Data

Stock Price Data
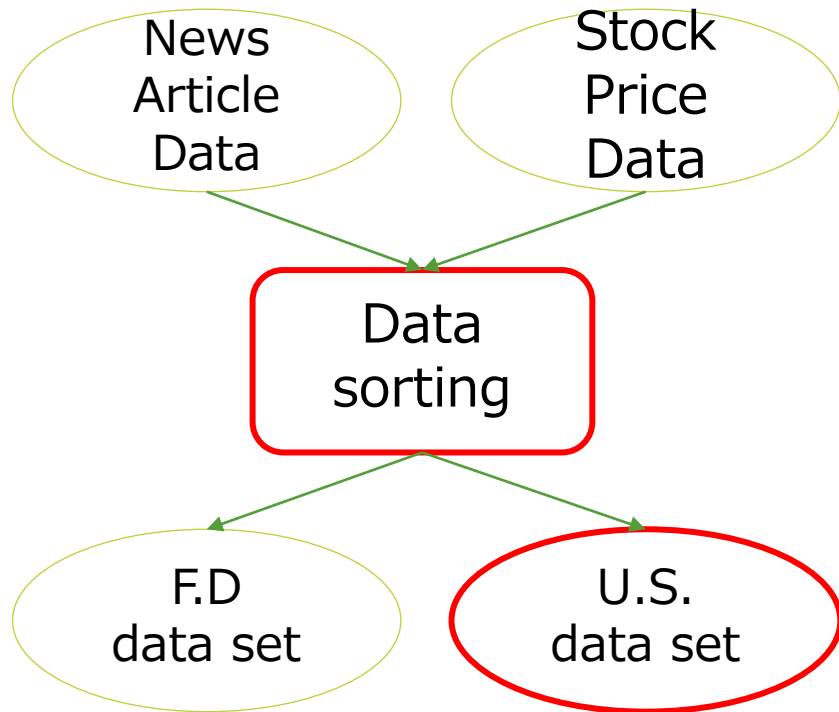
Data sorting

F.D data set

U.S. data set

- Extracting articles related to stocks

- Extracting articles of Fluctuation Day

  Fluctuation Day ‥‥
  days when the fluctuation between the closing price of the previous day and the current day is ±1% or more.

This set is called F.D data set

Proposed method

```
  ┌─────────┐        ┌─────────┐
  │  News   │        │  Stock  │
  │ Article │        │  Price  │
  │  Data   │        │  Data   │
  └────┬────┘        └────┬────┘
       └────────┐  ┌──────┘
            ┌───▼──▼───┐
            │   Data   │
            │ sorting  │
            └───┬──┬───┘
         ┌──────┘  └──────┐
     ┌───▼───┐        ┌───▼───┐
     │  F.D  │        │ U.S.  │
     │data set│       │data set│
     └───────┘        └───────┘
```

- The United States is often involved in the fluctuations of the Nikkei Stock Average

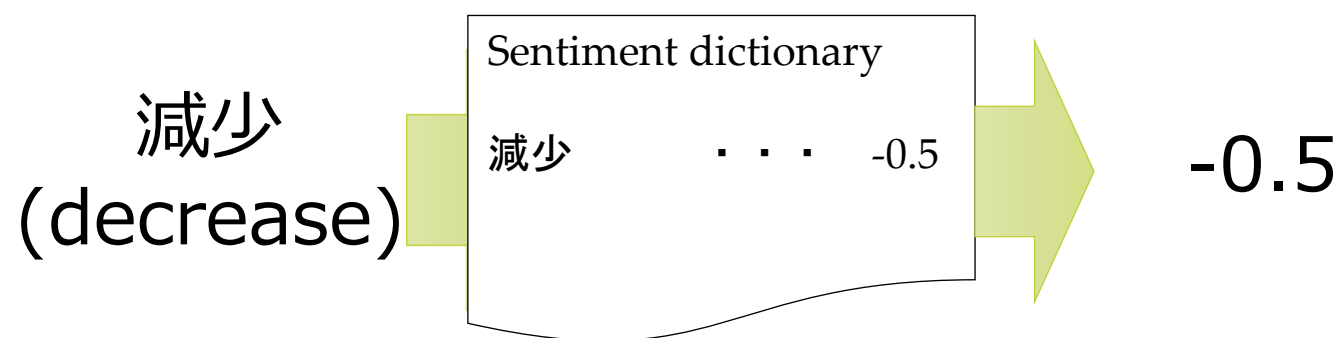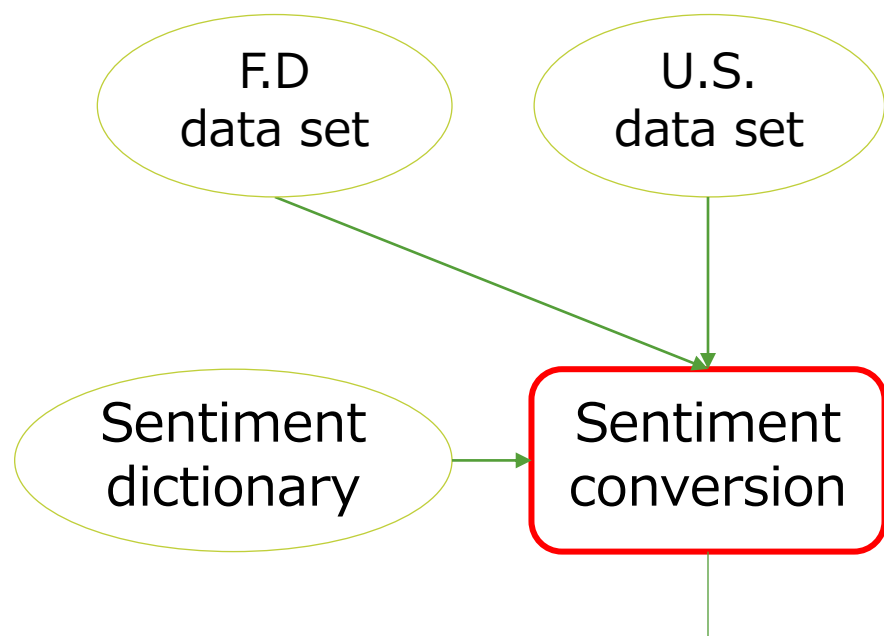→Create a separate set of articles that contain the word "US" in the article
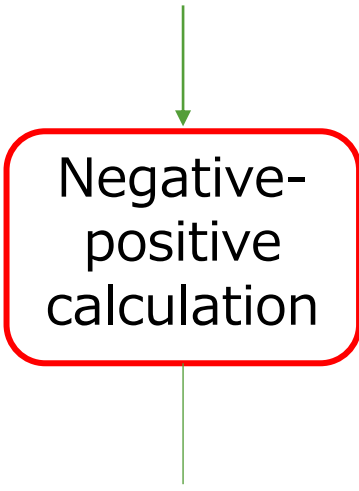
This set is called U.S. data set

Proposed method

- Sentiment Conversion

  Convert keywords in articles to numerical values from 1 to -1

F.D data set

U.S. data set

Sentiment dictionary → Sentiment conversion

減少
(decrease)

Sentiment dictionary

減少　　　・・・　　-0.5

-0.5

Proposed method

- Negative-positive calculation

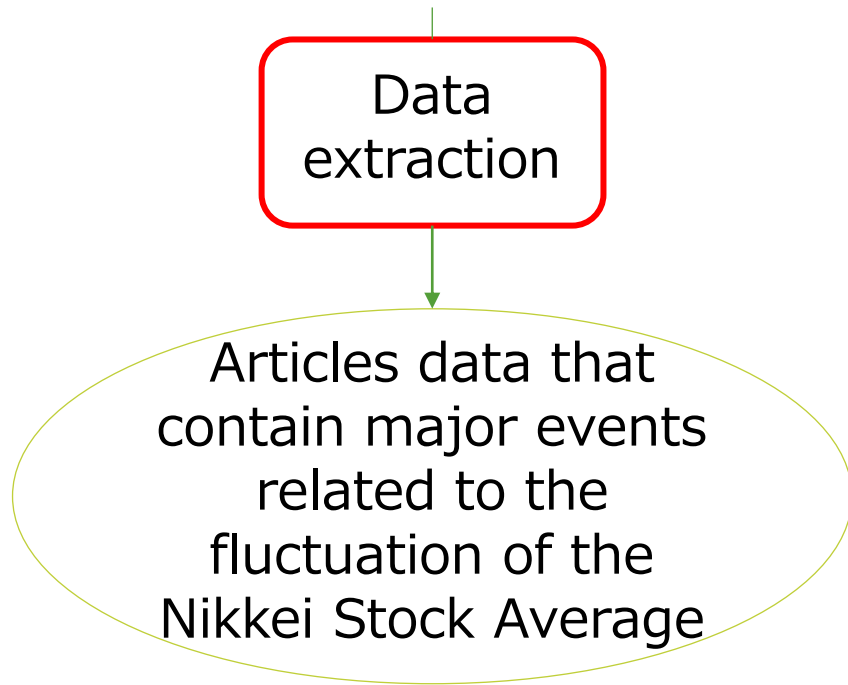 Use the average of each positive and negative value to calculate the percentage.

  Ex)

$$positive\_score = \frac{\overline{positive}}{\overline{positive} + |\overline{negative}| + \overline{neutral}}$$

Negative-positive calculation

Proposed method

Data extraction

Articles data that contain major events related to the fluctuation of the Nikkei Stock Average

- Extract the top 75% with the highest absolute value of positive and the top 75% with the highest absolute value of negative

$\rightarrow$Reduce the article data by about half

Experimental Methods

- This time we'll cover three months, 2016/05, 2016/11, and 2017/08.

- We'll put these together as a data set for each month, and conduct experiments on each of these three patterns.

News Article Data = 2016/05 News Article   2016/11 News Article   2017/08 News Article
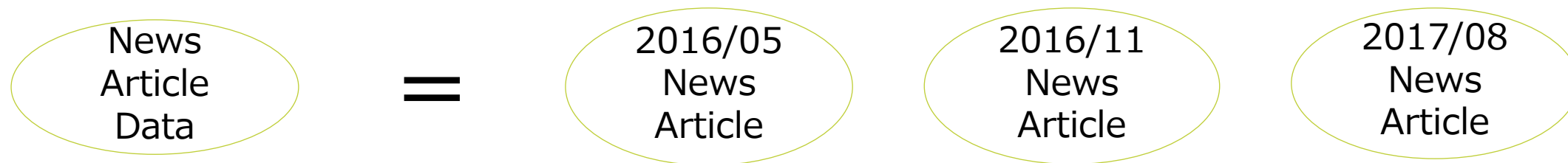
Verification method

- Compare the percentage of correct articles in the extracted article set with the original article set

- Articles containing the following keywords are considered as correct articles

Date	event	Used keywords
2016/05	Ise-Shima Summit	サミット(summit)
2016/11	presidential election	選挙(election)
2017/08	geopolitical risk	地政学リスク(geopolitical risk)

Results

Article set conditions	2016/05		2016/11		2017/08	
	percentage	Rise in value	percentage	Rise in value	percentage	Rise in value
F.D	2.419		3.875		5.313	
F.D + senti	2.488	0.069	3.815	-0.061	6.713	1.399
F.D + U.S.	5.484		10.583		10.968	
F.D + U.S. + senti	6.410	0.926	10.823	0.239	14.103	3.135

Discussions

- Reasons for the result of 2016/11

  Existence of important articles that do not include the election
  Ex) email issues, election status report

- Increase in the percentage of important articles included

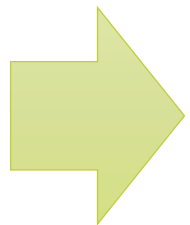- Reduce the number of articles by about half
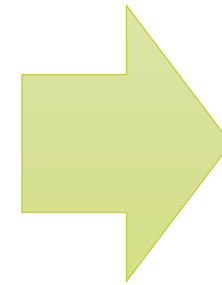
Sentiment expressions are effective for article extraction

Discussions

- Compare the rise in F.D data set with U.S. data set

	F.D		U.S.
	F.D		U.S.
2016/05 :	0.069	→	0.926
2016/11 :	-0.061	→	0.239
2017/08 :	1.399	→	3.135

U.S. is more effective

➡ Selection method based on F.D only is insufficient

Conclusion

- Purpose

  Extraction of important articles using polarity expressions

- Result

  Reduced the number of articles by almost half and improved the percentage of important articles included

  →Discovered the effectiveness of polar expressions.

- Problem

  Insufficient data selection method