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
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（イ勢志摩サミット）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

- 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

- 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)

\[
\text{positive\_score} = \frac{\text{positive}}{\text{positive} + |\text{negative}| + \text{neutral}}
\]
Proposed method

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

→ 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.
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:

<table>
<thead>
<tr>
<th>Date</th>
<th>event</th>
<th>Used keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016/05</td>
<td>Ise-Shima Summit</td>
<td>サミット(summit)</td>
</tr>
<tr>
<td>2016/11</td>
<td>presidential election</td>
<td>選挙(election)</td>
</tr>
<tr>
<td>2017/08</td>
<td>geopolitical risk</td>
<td>地政学リスク(geopolitical risk)</td>
</tr>
</tbody>
</table>
### Results

<table>
<thead>
<tr>
<th>Article set conditions</th>
<th>2016/05</th>
<th>2016/11</th>
<th>2017/08</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>percentage</td>
<td>Rise in value</td>
<td>percentage</td>
</tr>
<tr>
<td>F.D</td>
<td>2.419</td>
<td>3.875</td>
<td>5.313</td>
</tr>
<tr>
<td>F.D + senti</td>
<td>2.488</td>
<td>0.069</td>
<td>3.815</td>
</tr>
<tr>
<td>F.D + U.S.</td>
<td>5.484</td>
<td>10.583</td>
<td>10.968</td>
</tr>
<tr>
<td>F.D + U.S. + senti</td>
<td>6.410</td>
<td>0.926</td>
<td>10.823</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Year</th>
<th>F.D</th>
<th>U.S.</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016/05</td>
<td>0.069 →</td>
<td>0.926</td>
<td></td>
</tr>
<tr>
<td>2016/11</td>
<td>-0.061 →</td>
<td>0.239</td>
<td></td>
</tr>
<tr>
<td>2017/08</td>
<td>1.399 →</td>
<td>3.135</td>
<td></td>
</tr>
</tbody>
</table>

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