Sentiment Analysis of News Reports in the Period of New Coronavirus

Guangjun Li¹,²,*, and Lei Pan²

¹College of Sports Engineering & Information Technology, Wuhan Sports University, Wuhan, Hubei 430079, China
²School of Information Technology, Deakin University, Geelong, VIC 3220, Australia

Abstract. The spread of new coronavirus has a significant impact on the world. To protect people from viruses, a large number of personnel and substances in research, production, medical treatment, and other aspects are used to deal with the virus. News reporting is an important way for people to obtain information about the coronavirus. Objective and immediate news reports can provide people with accurate information and correct response methods, thus playing their due role in defeating the coronavirus. In order to compare the emotional characteristics of news reports, the methods of natural language processing are used to conduct sentiment analysis on news reports, including the subjectivity and the polarity of news reports. We select the news websites of BBC and CNN, that have an important influence in the world, for comparative analysis. The results show that BBC news reports are more positive than CNN, and in terms of subjective, BBC News is more factual than CNN.

1 Introduction

The global spread of COVID-19 has caused various effects on the world's economy, health, education, sports, transportation, and political landscape. News from various countries in the world has reported on coronaviruses and related events. At the same time, different news reports on the same incident have also appeared different. Viewing news reports correctly and rationally can enable us to have more reasonable countermeasures when dealing with events during the epidemic. To evaluate the differences of news reports from different sources, we adopt a method based on natural language processing to analyze and compare the news reports.

2 Related works

Sentiment analysis is an important part of natural language research. It can help us discover many important issues. For example, in the car forum, in the comment area of a particular model, through the sentiment analysis of various posts, we can know the overall image of the model in the eyes of users, as well as the ratio of positive and negative reviews, so that we can facilitate our targeted improvement. Sentiment analysis is also used in literary works to evaluate various character roles and to analyze the characteristics of characters.
In the field of sentiment analysis, there are three main methods [1], namely rule-based, traditional machine learning-based, and deep learning-based approaches. Yadollahi et al. [3] introduce some useful resources needed for rule-based sentiment analysis, such as dictionaries and data sets. Sailunaz et al. [2, 6] apply traditional machine learning to sentiment analysis of Twitter text. Stojanovski al. [1] propose a deep learning network for sentiment analysis. There are also some research and application of sentiment analysis, such as Sánchez et al. [4] introduce an open-source service framework for sentiment analysis, Huddar et al. [5] propose a multi-level feature optimization model for sentiment analysis, Chen et al. [7] propose a personalized recommendation model based on sentiment analysis, Machová et al. [8] study sentiment analysis in conversation content.

Due to the widespread news media, it is easy to cause huge repercussions in society, and some scholars have begun to study sentiment analysis on news reports. For example, some hoax news will have a huge negative effect on society. In order to identify this kind of news, Heru et al. [9] propose a method for hoax classification and sentiment analysis on Indonesian news. They have achieved a high accuracy of identification for hoax news. Seifollahi et al. [10] conduct sentiment analysis on news headlines and apply them to market predictions. In this paper, we apply sentiment analysis to the comparison of news reports of two important international news media.

![System architecture of sentiment analysis to news reports.](image)

**Figure 1.** System architecture of sentiment analysis to news reports.

### 3 Methodology

The system architecture of sentiment analysis in this paper is shown in Figure 1. Our process includes text acquisition of news reports, data cleaning, corpus creation, and sentiment analysis.
3.1 Data collection

In order to obtain representative analysis results, we select two news websites that have an important influence in the world, BBC and CNN. BBC is the largest news media in the UK and one of the largest news media in the world. CNN is an international news media with world-class influence. CNN news is broadcast or rebroadcast in more than 210 countries and regions around the world. During the data collection, web crawlers are used to periodically obtain news report data from the homepages of these two news websites.

3.2 Data cleaning

Before processing text data, we first clean these data. Data cleaning includes deleting punctuation marks, deleting numeric values, deleting meaningless characters (’n), deleting stop words (’a’, ’the’, etc.), marking text, converting all characters to lower case, etc.

3.3 Create a corpus

After the data cleaning is completed, the news text is saved as a text corpus, and the text is converted into a document-term matrix for subsequent statistical use. The document-term matrix is shown in Figure 2.

3.4 Word statistics and sentiment analysis

After creating corpus, word statistics and sentiment analysis are performed. We use TextBlob in sentiment analysis. TextBlob is an open-source library for text processing written in Python. It can be used to perform many tasks of natural language processing, such as part-of-speech tagging, noun component extraction, sentiment analysis, text translation, and so on.

We use a rule-based sentiment analysis model. In the model, there is a sentiment lexicon first. The lexicon is marked by vocabulary experts for each word in the corpus according to polarity and subjectivity, where polarity represents the positive or negative attribute of a word, and its value is between -1 and +1, -1 means very negative, and +1 means very positive. The subjective value is between 0 and 1, where 0 means objective and 1 means

---

**Figure 2.** Document-term matrix.
subjective. In the text analysis, we average the polarity and subjectivity of all words in the news text, and we can get the sentiment analysis results of the news reports.

4 Experimental results

Table 1. The average number of words per news.

<table>
<thead>
<tr>
<th></th>
<th>13-Apr</th>
<th>14-Apr</th>
<th>15-Apr</th>
<th>16-Apr</th>
<th>17-Apr</th>
<th>18-Apr</th>
<th>19-Apr</th>
<th>20-Apr</th>
<th>21-Apr</th>
<th>22-Apr</th>
</tr>
</thead>
<tbody>
<tr>
<td>BBC</td>
<td>293</td>
<td>382</td>
<td>323</td>
<td>367</td>
<td>504</td>
<td>334</td>
<td>474</td>
<td>348</td>
<td>311</td>
<td>276</td>
</tr>
<tr>
<td>CNN</td>
<td>9301</td>
<td>4661</td>
<td>9324</td>
<td>9554</td>
<td>5849</td>
<td>5369</td>
<td>5876</td>
<td>9726</td>
<td>5829</td>
<td>4963</td>
</tr>
</tbody>
</table>

Figure 3. Sentiment analysis results of news reports of BBC and CNN one day, 12 news reports each of BBC and CNN.

In the paper, we select the news reports on the homepages of BBC and CNN. The time is from 13 April 2020 to 22 April 2020. A crawler is used to grab 12 news reports from the homepage of each news website every day, for a total of 240 news reports. The average number of words in each news of BBC and CNN is shown in Table 1.

It can be seen from Table 1 that the number of words in each news of CNN is much higher than that of BBC news.

Figure 3 shows the sentiment analysis scatter plot of BBC and CNN news reports. If the abscissa value of 0.06 is used as the dividing line, and the polarity is less than 0.06, it is negative news, and if it is greater than 0.06, it is positive news. It can be seen that the number of CNN negative news is significantly more than that of BBC news.

In order to better distinguish the polarity bias of BBC and CNN news, we take the average polarity value of 240 news reports 0.04349 as the cut-off value. The news with a polarity lower than this value is negative news, the news with a higher value is positive news. Counting the number of negative news each day, the results are shown in Figure 4.
It can be seen from Figure 4 that CNN has significantly more negative news than BBC, almost every day.

Another indicator of sentiment analysis is the subjectivity of the text. In the paper, we take the average subjectivity value of 240 news reports 0.4155 as the cut-off value. If the subjectivity is lower than this value, it is a factual news report, and if the subjectivity is higher than this value, it is an opinion news report. Count the number of fact news reports each day, and the results are shown in Figure 5.

As can be seen from Figure 5, the factual news in the BBC is significantly higher than CNN, and CNN news is more biased towards opinion news, that is, news has more subjective opinions.

5 Conclusion

We face a large number of news reports every day; however, news reports are not always factual and positive. Our sentiment analysis of 240 news reports from BBC and CNN shows that although they are also world-class news sites, due to various reasons, they show
different degrees of subjective in news reports. Overall, the BBC news is more positive than CNN. In terms of the subjective news, BBC News is more factual than CNN.

References