Extracting and Presenting Different Viewpoints from Political News Articles

Mahboobeh Harandi¹, Bei Yu¹, Jon Atle Gulla²
¹Syracuse University
²Norwegian University of Science and Technology

Abstract

This study explores user interest in browsing different viewpoints from politicians and news agencies in news recommendation system. Along with providing personalized news articles to the user, bringing the opinion of politicians and news agencies about political events might be interesting for users. As there is always bias in publishing the news articles, newsreaders try to realize politicians’ opinions manually. We designed a prototype system to extract relevant politicians' viewpoints to a controversial event, and present them along with the news articles in the same user interface. We then observed user behaviour in browsing viewpoints vs. original news articles. According to the users’ clicks pattern analysis and their responses to the questionnaire they are interested in browsing the different viewpoints. This result suggests that news recommendation system could provide alternative recommendation criteria such as different viewpoints when recommending news articles.

Keywords: Opinion mining; political news; news recommender systems; user clicks
doi: 10.9776/16514
Copyright: Copyright is held by the authors.
Contact: mharandi@syr.edu

1 Introduction

Exploring people’s opinion has a significant impact to all the human related sciences such as economics, political studies, management science, psychology and social sciences (Pang & Lee, 2008). In this study, we explore a new feature of news recommendation system by designing a prototype system that presents different politicians' opinion to a controversial event, the nuclear program in Iran. This event involves several countries and parties all around the world, and the news articles included different politicians' perspectives. The research is trying to answer the question whether and how extracting and presenting different viewpoints in news recommender system affects users' news reading behaviour.

2 Methodology

In order to study effect of the new feature in news recommendation system, both systems should have the equal chance to be interacted by online users. Provided information in both sides are the article’s url and title, news agency’s name and the first sentence of the article. But in sentiment part, the user will choose the politician’s name and then see their quotes, their sentiment and mentioned information. In the following sections the methodology is described.

2.1 Link Extraction

In order to work on the chosen topic the csv file of GDELT¹ on the third of April is downloaded. Then among all the links that have nuclear and Iran, thirty of them are chosen randomly and stored in the text file. In order to extract link, title and body text of each article Newspaper² package has been applied. Later the news url, its title, news agency’s name and the first sentence of each article are stored in MySQL database.

2.2 Politicians Name Extraction

In order to provide the sentiment of the politicians’ quotes, the name of politicians should be extracted from the text body of the articles. The package of Name Entity Recognition of Stanford(Sutton, 2012) is applied for this purpose. The package with seven class model has been used in Java in eclipse environment and all the PERSON entities have been extracted.

¹ Global Database of Events, Language and Tone (GDELT) project (team, 2014) is an open database over the news articles in the world based on the date that they are published not the event’s date.
² The Newspaper is written in Python with the license of MIT with the aim of article scraping and curation.
2.3 Quote Extraction

The next step is specifying the desired patterns for the quote extraction. Two explicit patterns and an implicit one are specified as follows:

- The most common verbs such as said, told, warned and says are selected as the verbs set.
- The first pattern is the set of double quotation + politician’s name + verbs set
- The second pattern is the politician’s name + verbs set + colon + set of double quotation
- The third pattern is the politician’s name + verbs set + the remainder of the sentence to the end

All these quote extraction techniques are not covering all the quotes of politicians, since the news articles are not following the specific format of reporting.

2.4 Sentiment of Quotes

Through the Stanford CoreNLP pipeline four different annotators for Stanford Sentiment tools are specified including tokenize, ssplit, parse and sentiment (Christopher D. Manning, 2014). The sentiment prediction is done in a sentence level, which is satisfying the sentiment of the politicians’ quotes (Socher, Pennington, Huang, Ng, & Manning, 2011) (Richard Socher, 2013). Every single quote’s sentiment is presented to the user and they will realize all different sentiments of chosen politician by themselves. The sentiment of each politician over their extracted quotes with the data of the news url, its title, the first sentence of the news and the news agency’s name is stored in the database in MySQL.

2.5 Web Design

In order to present and evaluate the new system, the experiment should be run among different people online. The web application is implemented through the Eclipse Java JEE luna edition. The server is Apache Tomcat version 8.0 that is used locally first and later in the Amazon Web Service. In order to deploy the web application on the Amazon Web Service, the current local database is immigrated to the instance of the Rational Database Service. The user clicks on two different parts of the web page will be stored based on their user account.

On the first page to login, the user should enter their desired, valid email address and preferred password. Then the list of the news articles with mentioned information is retrieved and shown on the left side of the page. On the right side of the page, the list of politician’s names is available through the dropdown list, by choosing the name of the politician, the extracted quotes, its sentiment and other mentioned information will be shown in the table format.

![Web Interface](image.png)

Figure 1. Web Interface

3. Evaluation and Result

The evaluation of the research design is conducted in two ways. The first part of the evaluation is analysing the stored user clicks and the second one is the questionnaire. Twenty people who have attended the experiment are from different countries with different nationalities, educations and interests. Due to convenience of research, they are Facebook friends of the author. The good point of this choice is the availability of attendees for the interview in further analysis.

3.1 Click Rating

Storing the users’ clicks is an implicit feedback which is available as soon as they are making account and navigate between two different parts of the news page. The following chart shows the number of
users with three types of activity\(^3\), including low, medium and high in news reading and sentiment exploring. It shows that numbers of users with low and medium activity are higher in the sentiment exploring part in comparison to the news reading part and numbers of users with high activity are the same in both sides. By accumulating the numbers of users in each part it is observed that the sentiment exploring part has attracted more users; 19 versus 17.

![Figure 2. Users’ activity comparison between news and sentiment of news](image)

<table>
<thead>
<tr>
<th>News Articles</th>
<th>Low activity</th>
<th>Medium activity</th>
<th>High activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentiment Analysis</td>
<td>6</td>
<td>8</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>News Articles</th>
<th>Low activity</th>
<th>Medium activity</th>
<th>High activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentiment Analysis</td>
<td>7</td>
<td>9</td>
<td>3</td>
</tr>
</tbody>
</table>

3.2 Questionnaire Design

The questionnaire is designed to have more exact feedback from users and receive their opinions about the expected requirements. In order to save the time of the users and keep them motivated to respond, seven closed questions are asked with the complete range of the answers; totally disagree, disagree, neutral, agree and totally agree. Besides, three open questions with the aim of more improvement are requested. By reviewing the respond of open questions, it is understood that they would like to have the sentiment over all news events and also the comparison among different politicians’ quotes.

<table>
<thead>
<tr>
<th>Question</th>
<th>Strongly agree</th>
<th>Agree</th>
<th>Neutral</th>
<th>Disagree</th>
<th>Strongly disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>The specified sentiment was the same as I recognized from the quotes.</td>
<td>13%</td>
<td>56%</td>
<td>25%</td>
<td>6%</td>
<td>0%</td>
</tr>
<tr>
<td>I’d like to have the political sentiment analysis in my news recommender application.</td>
<td>31%</td>
<td>50%</td>
<td>6%</td>
<td>13%</td>
<td>0%</td>
</tr>
<tr>
<td>I’d like the way that shows every quote separately and I can decide by myself the overall sentiment of that.</td>
<td>56%</td>
<td>38%</td>
<td>6%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>The sentiment analysis of politicians is completing the news recommender systems.</td>
<td>13%</td>
<td>50%</td>
<td>19%</td>
<td>19%</td>
<td>0%</td>
</tr>
<tr>
<td>I’d like to see the changes of a politician's view of the event in time.</td>
<td>7%</td>
<td>33%</td>
<td>53%</td>
<td>7%</td>
<td>0%</td>
</tr>
<tr>
<td>I’d like to see all the quotes of politicians.</td>
<td>20%</td>
<td>33%</td>
<td>27%</td>
<td>20%</td>
<td>0%</td>
</tr>
<tr>
<td>Would you like to see the prediction of the system about the politicians' sentiment for the specific event in advance?</td>
<td>13%</td>
<td>40%</td>
<td>20%</td>
<td>27%</td>
<td>0%</td>
</tr>
</tbody>
</table>

Table 1. Questionnaire's result

---

\(^3\) Activity is defined according to the clicks pattern of each user in each part; frequency, minimum and maximum numbers of clicks are considered and three intervals are defined as low, medium and high activity.
4 Conclusion and further works

News recommender systems are one of the most popular applications that personalize the interesting articles to users. We designed an online prototype system to present the articles with viewpoints of politicians beside the original recommended articles in the same interface. Then by collecting the data through the implicit and explicit feedback of the users, we recognized they are willing to browse different viewpoints. In future, the comparison among different politicians’ opinions will be available and users will see the changes of their opinions in time. Besides, further quotes’ patterns will be extracted.

5 References

Richard Socher, A. P., Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Ng, Christopher Potts. (2013). recursive deep models for semantic compositionality over a semantic treebank.