Public Opinion Mining on Social Media

Reema D[1], J Nagesh Babu[2]

Associate Professor, RYMEC[2]

Abstract: Social media, including micro-blogs such as Twitter, have become main channels to communicate and share public social opinion among people. The opinion on nuclear power is an important social issue because nuclear plant construction needs national consensus. This paper proposes an opinion mining approach to monitor public sentiments on nuclear related issues using tweets on Twitter. The proposed process consists on (1) Crawling related tweets (2) Text preparation (3) Sentiment dictionary construction, and (4) Sentimental scoring. Based on experiment using nuclear related tweets in Korean between 2009 and 2013, we verify the usefulness on the proposed approach and confirm the changes on national opinion on nuclear generation depending on critical events such as Fukushima Daiichi nuclear disaster.

Keywords: public opinion mining, sentiment analysis, nuclear power.

INTRODUCTION

The popularity of social media such as Facebook and Twitter turned them into main channels to communicate and share opinions on political, economic, social, and cultural issues. Even though social media contribute to changing consumers to prosumers (producers plus consumers), there are also some drawbacks on public opinion conversation and convergence such as fraudulent and biased messages, witch hunting, and extrusion of personal information. Reflecting the increase of interest on opinion on social media, there has been trials and experiments to monitor and analyze public opinion on specific issues on social media.

Nuclear power is a significant national issue because it is a double-edged sword. Economic efficiency of power generation is the most important benefit of nuclear power. However, the potential risk of radiation leakage is the biggest difficulty of nuclear energy. So, nuclear power plant construction requires national consensus and agreement of residents in construction area.

Traditional survey approach has been used to monitor and investigate public opinion on nuclear power; however, it takes excessive cost and time. Opinion mining approach can provide an alternative way to monitor public opinion on nuclear power. This paper aims to suggest and verify opinion mining approach on nuclear power.

The structure of this topic is as follows: In section 2, the researches of opinion mining approaches to monitor customer opinion and public opinion are described briefly. In section 3, the proposed opinion mining approach will be presented. In section 4, the experimental results using Twitter data between 2009 and 2013 are provided to demonstrate the usefulness of the proposed approach. In section 5, includes conclusion remarks.

RELATED WORKS

Due to the popularity of social media, the power on policy decision making is shifted from mass media and major politicians to citizens. Social media can work as new cheap direct communication channels between governments and citizens. So, there are trials to use social media as new tools to establish public agenda setting and to evaluate the results of policy execution.

For example, USA government established 'Cash for Clunkers Program' in which government supported car renewals as an economic activation policy in 2009. The initial planned execution period of the program was 4 months and the government prepared required budget based on rough demand forecasting. However, due to the unexpected good response, the budget was exhausted within one week and the government needed to prepare additional budget in a rush. However, Google predicted the respond level based on the number of related terms in Google search, and forecasted budget exhaustion time. Base on the results, Google predicted the need for additional budget accurately.

Recently, there have been many researches to monitor public opinion and social trends using social media.

They include:
Election prediction using Twitter data.
Monitoring of customer sentiment on a certain brand.
Movie performance prediction using Twitter.
Disease and disaster tracking using Internet information, and Unemployment benefits prediction using Internet search information.

Especially, public opinion monitoring is useful in sensing public opinion trends and reduction of potential social risks and conflicts. In this study, we mainly focus on public opinion on nuclear power.

PROPOSED APPROACH AND EXPERIMENTAL DESIGN

The proposed procedure for public opinion mining consists of four phases: (1) crawling social media data, (2)

cleansing and pre-processing texts, (3) construction of a sentiment dictionary, and (4) tweets sentiment prediction. The detail of such experimental procedure is shown in Fig. 1.
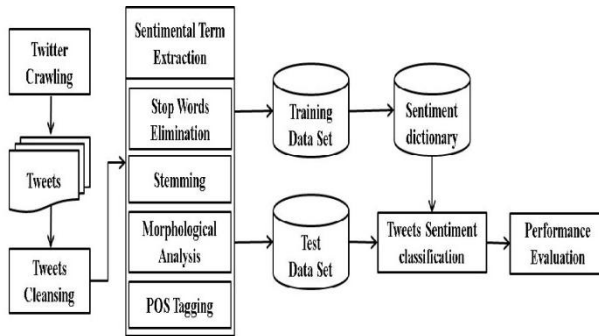

Figure 1. Experimental Procedure

3.1. Crawling and Cleansing on Social Media Data
In the crawling phase, tweets including terms "Nuclear" or "Nuclear power" in Korean are crawled by a crawling tool, LocoySpider[2]. Twitter, with its great number of users having different opinions and real time update, is a useful tool to sense public opinion.

Among the crawled tweet, irrelevant tweets are identified and removed for next analysis. Also tweets on neural news articles and tweet from government institutes are removed because our purpose is to identify social opinions rather than government announcement. Finally, the final data set consists of 237 tweets in 2009, 914 tweet in 2010, 5,705 tweets in 2011, 5,081 tweets in 2012, and 4,655 tweets in 2013.
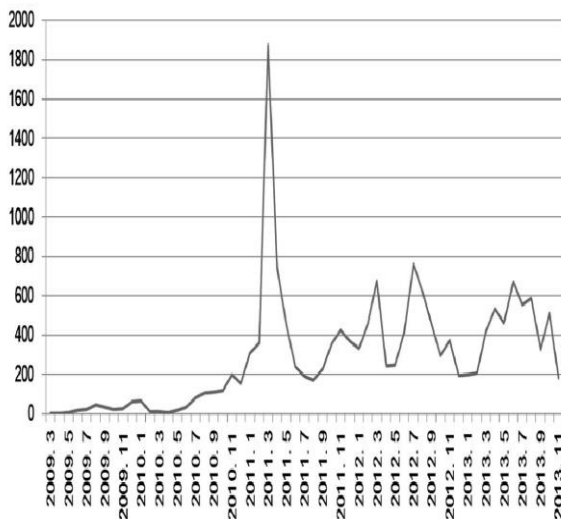

Figure 2. The Number of Tweets

The number of tweets on nuclear power is relatively small in 2009. There were three good news on nuclear power in Korea during 2009 and 2010 (refer Table 2). First, on December 28 2009, there was the first nuclear power plant export to UAE. There were two additional export news in

2010. The most important event during the period was Fukushima nuclear accident in March 2011. The number of tweets in March 2011 made a peak. After the event, there was a good news (Nuclear Security Summit in Korea) in 2012 and several bad news on stops of operations at nuclear power plants during 2012 and 2013. Depending on the events, the monthly tweets made fluctuation.

3.2. Sentimental term Extraction
Due to Korean language characteristics including many variations on roots, importance of distinguish postposition and ending, and compound nouns with space, natural language processing in Korean is more difficult than that in English. Especially, in social media, users use composite terms and abbreviations more frequently, and include many sentences which do not follow grammar rules. So, it requires more effort to extract sentimental terms in social media. In this study, we use text mining package tm[4] in R and KoNLP (Korea Natural Language Processing)[5] which provides functionalities to parse Korean sentences.

Table 2. Important Events on Nuclear Power during 2009 and 2013

Year	date	Event
2009	December 27	First nuclear power plant export to UAE
2010	March 30	First research nuclear reactor export contract to Jordan
	June 15	MOU on nuclear power plant between Turkey and Korea after a summit
2011	March 11	Meeting
	March 26 and 27	Nuclear Security Summit
2012	July	Stop of operation at Youngkoyang nuclear power plant
2013	March, August, and November	Stops of operation at nuclear power Plants

Potential sentimental terms are extracted though stop words elimination, stemming, morphological analysis and POS (Part of Speech) tagging. In stop words elimination, the special characters such as twitter tags (for example, '@' and '#'), punctuation marks, URL (started with 'http://'), and numbers are removed as stop words. Finally, only nouns are extracted potential sentimental terms.

In the previous study, nouns and emoticons are more frequently used to express emotion than adjectives due to the limitation of the number of characters in Twitter. Also, in our case, representative nouns such as 'export', 'safety', 'Mafia', and 'disaster' are used frequently to

express users' emotion. These are the reasons why we consider nouns first as potential sentimental terms.

The followings are several examples of English-translated tweets. First two tweets are examples of positive tweets, and the next two tweets are negative tweets.

(1)   a. "South Korea has a high level of nuclear safety systems" http://xxx/xxxxx
b. "Do you know that Republic of South Korea exports 'nuclear technology' as well as 'nuclear safety'?" http://xxx/xxxx
(2)   a. #xxxx "Government's safety ignorance of nuclear, more Serious than Fukushima.
b. It is fraud of the nuclear mafia!" RT@xxxx

3.3. Construction of a Sentiment Dictionary
Sentimental classification is usually based on the number of positive terms and negatives terms in sentences. The referential positive terms and negative terms are stored in sentiment dictionary. So, the construction of sentimental dictionary is an important part in sentimental analysis. Previous constructed sentiment dictionaries can be used or a new sentiment dictionary can be constructed using gathered sentences. In this study, we use latter to reflect contexts of nuclear power domain.

Table 3. Example of Terms in a Sentiment Dictionary

Polarity	Terms
Positive	Safety, Export, Creative, Clean, Economic, Powerful Nation, Clean Energy, Respect, Profit, Green Growth, Advanced Country, Technology, Excellent, Green, Alternative Energy, Efficiency......
Negative	Disaster, Ignorance, Nuclear Power Mafia, Unrest, Radiation Leak, Fraud, Irresponsibility, Destruction, Tsunami, Opposition, Impeachment, Suspicion, Corruption, Conceit, Opacity, Imprisonment, Dilemma, Catastrophe, Scapegoat......

The tweets from 2009 to 2011 are used to construct a sentimental dictionary. Nouns are extracted after potential sentimental term extraction. Three human evaluators are classified positive, negative, and neutral terms separately. The terms which assigned in positive or negative terms by all of three evaluators are finally assigned as positive or negative terms. The number of positive terms is 1,012 term and that of negative terms is 3,291 (refer Table 3). We can find that the number of negative terms is greater than that of positive terms.

3.4. Tweets Sentiment Classification
The tweets between 2012 and 2013 are used to evaluate the performance of sentimental classification. Sentimental classification is based on sentimental scores of tweets. The

sentimental score of a tweet is calculated based on the number of positive terms and the number of negative terms in the tweet. The range of sentimental score is between -1 and 1. Tweets are classified as positive tweets when sentimental scores are greater than 0, and as negative tweets when those are less than 0. The tweets with 0 sentimental scores which mean that there are no positive terms and negative terms, or the number of positive terms and negative terms are the equal.

$$\text{Sentimental\_score}(t)= \frac{N\big(\text{Positive Terms}(t)\big) - N\big(\text{Negative Terms}(t)\big)}{N\big(\text{Positive Term}(t)\big) + N\big(\text{Negative Terms}(t)\big)}$$

EXPERIMENTAL RESULTS

The performance of proposed approach is tested using tweets between 2012 and 2013. The prediction accuracy is shown in Table 4. In the case of positive tweets, accuracy rates are 51.57% for 2012 tweets and 50.55% for 2013 tweet. Accuracy rates on negative tweets are relatively higher than positive tweets. It is 61.19% for 2012 tweets and 64.08% for 2013 tweets. However the accuracy rates on neutral tweets are relative lower those on positive or negative tweet. It is 38.96% for 2012 neutral tweets and 21.67% for 2013 neutral tweets. The reason is that we focus on positive or negative tweets and construct only positive and negative dictionaries without a neutral dictionary.

Table 4. Sentimental Prediction Accuracy

Sentiment	2012		2013	
	Accuracy Rate	No. of Tweets	Accuracy Rate	No. of Tweets
Positive	51.58%	948	50.55%	991
Negative	61.19%	2067	64.08%	2289
Neutral	38.96%	2066	21.67%	1375

To trace the temporary changes of public opinion on nuclear power, we define Nuclear Opinion Index (NOI) as follows (refer <formula 2>). NOI is determined based on the number of positive tweets, negative tweets, and total tweets of the month. Based on NOIs, we can trace the change of public opinion on nuclear power at a specific period.

$$\text{Nuclear Opinion Index}(m)= \frac{N(\text{Positive tweets }(m)) - N(\text{Negative tweets }(m))}{N(\text{Total tweets }(m))} \times 100 + 100$$

Figure 3 shows the changes of NOIs during 2009 and 2013. Based on positive news on export of nuclear technology, public opinion on nuclear power was positive until 2010. However, it turned to negative from Fukushima Daiichi nuclear disaster and was still located in negative until the end of 2013. The NOI chart visualizes the changes of public opinion on nuclear power.
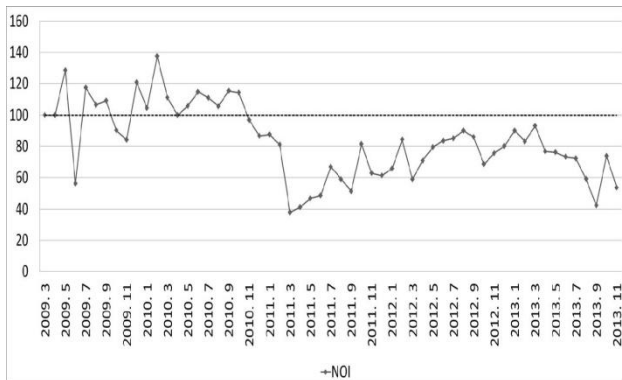
Figure 3. Changes on Monthly Nuclear Opinion Index

CONCLUSIONS

Social media advanced as new communication channels for public opinion because they are open to many citizens, easy to use, and provides online discussion mechanisms. Attempts to use social media to monitor public opinions are in initial stages of studies and practices.

Traditionally, survey-based approach has been used to monitor public opinion on nuclear power, which is expensive and has time delays. Opinion mining approach based on social media contents can provide complementary approach, which is relatively cheap and with no time delays. Even though we focus on public opinion on nuclear power, the proposed approach can be easily applied to other public opinion cases.

In this, we propose public opinion mining approach to monitor nuclear power opinion on Twitter.

The procedure consists of four phases:
(1) Crawling related tweets,
(2) Extracting potential sentimental terms,
(3) Building sentiment dictionary, and
(4) Tweets sentiment scoring and prediction.

The experiments using tweets between 2009 and 2013 showed that the proposed approach provided acceptable performance on sentimental prediction. Also, NOI (Nuclear Opinion Index) is proposed to visualize the sentimental changes on nuclear power opinion.

REFERENCES

[1]   D. S. Kim and J. W. Kim, "Public Opinion Mining on Social Media: A Case Study of Twitter Opinion on Nuclear Power", Proceeding of CES-CUBE 2014, (**2014**); Saipan, USA.
[2]   Akcora, C.G., Bayir, M.A., Demirbas, M., Ferhatosmanoglu, H.: Identifying Breakpoints in Public Opinion. In: 1st Workshop on Social Media Analysis, pp. 62--66. Washington, DC (2010).
[3]   Boutet, H. Kim and E. Yoneki, "What's in Your Tweets? I Know Who You Supported in the UK 2010 General Election", Proceeding of The International AAAI Conference on Weblogs and Social Media, (**2012**).
[4]   Baek, H.M., Ahn, J.H., Oh, S.W.: Impact of Tweets on Movie Sales: Focusing on the Time when Tweets are Written. J. ETRI.
(2014).
[5]   Boutet, A., Kim, H., Yoneki, E.: What's in Your Tweets? I Know Who You Supported in the UK 2010 General Election. In: The International AAAI Conference on Weblogs and Social Media (2012) .
[6]   S. K. Hong, Y. J. Lee and Y. W. Choi, "Fukushima nuclear power plant since the accident, policy direction and future major issue surrounding the nuclear power", Science & Technology Policy, no. 84, (**2011**), pp. 1-32.
[7]   T. Nasukawa and J. Yi, "Sentiment analysis: capturing favorability using natural language processing", Proceeding of the K-CAP-03, 2nd International Conference on Knowledge Capture, (**2003**), pp. 70-77.
[8]   Choi, H., Varian, H.: Predicting the Present with Google Trends. Technical Report, Google (2009).