AN ANALYSIS OF TERM LIFE CYCLE ON SOCIAL MEDIA: A CASE STUDY OF RENEWABLE ENERGY ON TWITTER

[1]DONG SUNG KIM, [2]JONG WOO KIM

School of Business, Hanyang University, Korea
E-mail: {paulus82, kjw}@hanyang.ac.kr

Abstract- It is possible to see the changes of major social interests by analyzing various subjects and issues discussed in social media. Furthermore, these changes can be used in predicting future trend. This paper aims to investigate changes on the interests on renewable energy on social media. As using Relative Term Occurrence and Term Life Cycle concepts, we tries to identify changes on the interests of renewable energy including solar, wind, water, and biomass. To this end, tweets related to the renewable energy written from 2013 to 2014 in Twitter, one of the representative social media channels, were collected. As a result of the analysis, keyword occurrence frequency of solar heat is on a gradually decreasing trend while water power and biomass maintains a steady frequency of keyword occurrence. Above this, occurrence frequency of wind tends to sharply increase after certain period.

Index Terms- Social Media, Trend Detection, Term Life Cycle, Renewable Energy

I. INTRODUCTION

With the development of social media such as Twitter or Facebook, online users are actively participating in sharing information and proposing various opinions with each other. Opinions written by users in social media includes variety of subject from personal interests to socially discussed issues[1].

Terms mentioned in online posts and their frequencies are diversely shown according to the social climate or public interests. If a particular social issue receives a great attention among public, it is identified that keywords related to this issue are frequently appeared online. Also, these keywords may be continuously mentioned depending on the importance of the issue or persistency of interests, or neologism related to this keyword can be coined. On the other hand, due to the change in the interests of users or the occurrence of new social issues, there are cases that the frequency of previously created keyword is gradually decreased or the keyword is not mentioned any more. Like this, 'keyword' used in online communication media has its life cycle, grown and dissipated according to the social issue or user interest. This life cycle plays an important role in representing the level of interests while identifying the change of interests among participants. That is, trend analysis of online keywords can help understand important social issues and predict the future changes in public opinions[2]. Therefore, in this research, tweets related to the renewable energy are collected to extract major keywords. These keywords are then analyzed to see the trend of public interests.
The renewable energy has recently received high attention due to the climate change and natural resource depletion, with various technologies and policies related to the solar, wind, water and bio fuels. Globally, advanced nations are gradually increasing the portion of renewable energy and related technology development[3]. By checking and analyzing online public opinions about renewable energy, policy makers can reflect them into new energy policies in national level.
The formation of this study is as follows. In section 2, previous studies related to the renewable energy, and issue analysis in social media is reviewed. In section 3, we propose our approach and experimental design. In the study, main keywords are collected from renewable energy related tweets mentioned in Twitter to extract Relative Term Occurrence (RTO) per keyword, and to identify the method of analyzing life cycle. In section 4, analysis result is described. Finally, in section 5, conclusion and further study ideas are proposed.

II. RELATED WORKS

A. Social Media and Trend Detection

As interest toward the analysis and utilization of high volume data, many researches are conducted to explore main issues and keywords discussed in social media. Twitter, one of the representative social media, provides a platform in which various user opinions can be rapidly created and disseminated. Using the tweet data analysis, it is possible to extract trend of public interest [4,5]. Many studies has been made in a variety of fields such as public survey, disease prediction and stock market analysis[6].
The representative trend exploration researches using Twitter includes frequency variation of particular keywords, relationship analysis of keywords appeared together with the particular keyword, and study using Latent Dirichlet Allocation (LDA) which is one of the famous topic models[7,8]. In addition, for the purpose of marketing or public opinion research, web services

are also developed to use extracted keywords in Twitter [5].

Many researches related to the online issue exploration mostly include studies of finding emerging keywords and their utilization plan. It is important in understanding current issues and social opinion, but it is difficult to see the change in trend by time series analysis of keywords. Accordingly, this research aims to analyze Term Life Cycle of main keywords in Twitter by utilizing measuring method of Relative Term Occurrence. Using the method, it is possible to understand and predict the changes in trend of key interests.

B. Renewable energy

Renewable energy is one of the promising filed in low-carbon green growth with active global investment and development of related technologies. International Energy Agency defines renewable energy as 'an energy endlessly supplied from nature', and electric and heat energy from solar, wind, water and bio fuels is classified into the renewable energy[9]. Major factors increasing interest in renewable energy is energy scarcity due to the existing fossil fuel and greenhouse gas reduction for climate change. Therefore, as renewable energy being a core element of national economic growth to substitute existing fossil fuel, various efforts of technology investment and legal system improvement are being made to change new paradigm of energy.

III. PROPOSED APPROACH AND EXPERIMENTAL DESIGN

The research process of this study is as follows (refer <Fig. 1>). (1) We collected tweets related to renewable energy from 2013 to 2014. (2) After sentence parsing of tweets and term extraction, (3) Relative Term Occurrence is calculated for extracted keywords. (4) Finally, term life cycle is analyzed for primary keywords by checking Relative Term Occurrence previously calculated.
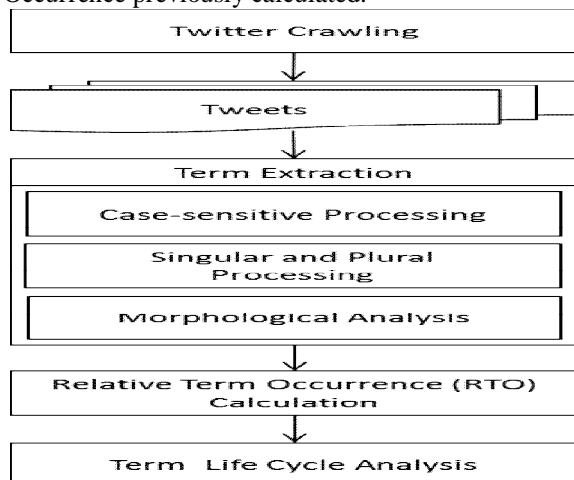

Fig. 1. Experimental Procedure

C. Term Extraction

In this research, crawling of tweets including terms related to the renewable energy is carried out by Java programming. Then primary keywords are extracted through preprocessing and syntax analysis of collected tweets. The analysis tool used for eliminating stopwords and extracting term from tweets is done with 'tm'[*] package in R.

D. Relative Term Occurrence (RTO)

Term life cycle analysis in this research utilizes measurement method of Relative Term Occurrence (RTO), and the value is calculated as follows(refer <formula 1>).

$$RTO_i(t) = \frac{TF_i(t)}{TF(t)} \times \frac{TF_i(t)}{\sum_{t=all} TF_i(t)} \tag{1}$$

$RTO_i(t)$: relative occurrence frequency rate of the term t in period i

$TF_i(t)$: frequency of the term t occurring in period i

$TF(t)$: frequency of the term t occurring in the entire period

$\sum_{t=all} TF_i(t)$: total frequency of all term t in period i

$$CRTO_i(t) = \frac{RTO_i(t)}{RTO_{i-1}(t)} \tag{2}$$

$CRTO_i(t)$: rate of change of RTO

This research imparts a meaning to such terms by classifying the life cycle of the terms into growth, stagnation and deterioration based on Relative Term Occurrence. Formula (2) is calculated based on the value derived from formula (1). If the Relative Term Occurrence at i divided by the Relative Term Occurrence at i-1 is larger than 1, it means that the term is actively used after creation (growth). It the Relative Term Occurrence at i-1 divided by the Relative Term Occurrence at i is smaller than 1, it means that the term is deteriorated. On the other hand, if the Relative Term Occurrence at i is same as the Relative Term Occurrence at i-1, making divided value 1, it means that the term is in the condition of maintenance.

IV. EXPERIMENTAL RESULTS

Accordingly, in this chapter, the Relative Term Occurrence is calculated for keywords related to renewable energy from 2013 to 2014, and trend of terms in relevant period is analyzed. The Relative Term Occurrence is calculated in units of 6-months period and trend analysis is performed by connecting the change ratio of term(refer <formula 2>).

[*] Ingo Feinerer (2014). tm: Text Mining Package. R Package Version 0.6. http://CRAN.R-project.org/package=tm

We have identified the keywords like "Solar", "Wind", "Water", "Biomass" that are related to renewable energy. Total frequency per 6 months of each term is as follows(refer <Fig. 2>). The frequency sum of each term is the absolute difference. We converted this frequency into the relative value by using formula (1), and the formula (2) is calculated by using the value of formula (1).
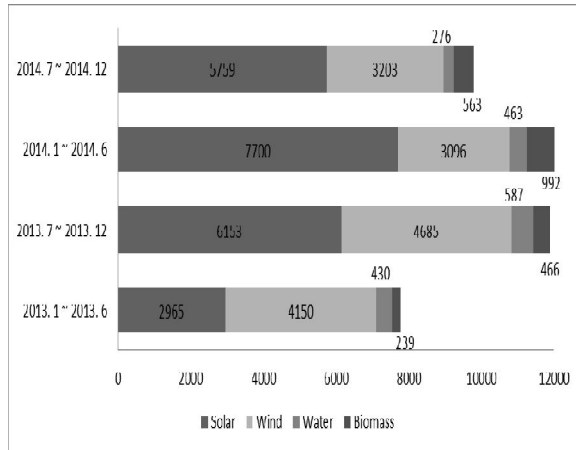

Fig. 2. Term Frequency within the Period

The relative occurrence frequency of "biomass" is shown that it sharply increase until mid 2014. However, the term seems a tendency to decline rapidly at late 2014(refer <Fig. 3>).
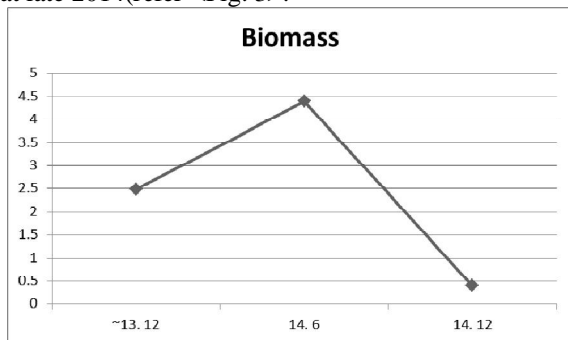

Fig. 3. 'Biomass' – Trend of Relative Term Occurrence

"Water" seems continuous decline. Because the CRTO has more than 1 value until the late 2013, "Water" is verified as growing trend. But it shows tendency to decline since then(refer <Fig. 4>).
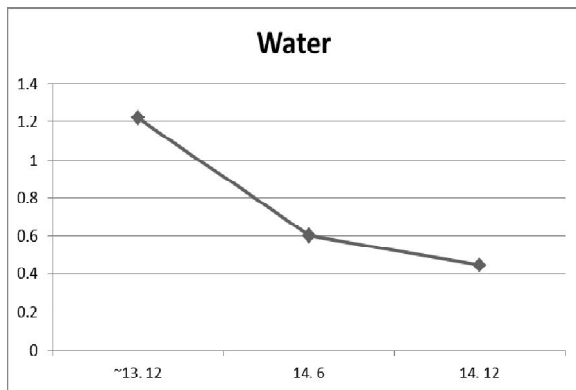

Fig. 3. 'Water' – Trend of Relative Term Occurrence

A Result of examining the Relative Term Occurrence for the keywords, "Solar" is reduced until mid 2014 and went under slow deterioration after mid 2014. Because the CRTO value of "Solar" is falling down at late 2014, it is seen as trend of deterioration since the time (refer <Fig. 5>).
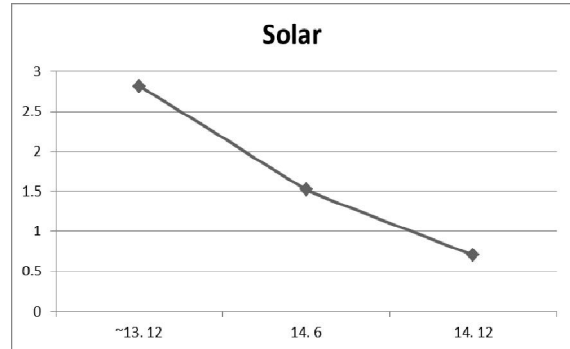

Fig. 5. 'Solar' – Trend of Relative Term Occurrence

"Wind" is marginally decreased until mid 2014 but sharply increase after mid 2014. Because the "Wind" frequency is increased in the period in which other terms rapidly decrease, Relative Term Occurrence showed a significantly high value. In other words, "wind" is identified as a growing trend since mid 2014(refer <Fig. 6>).
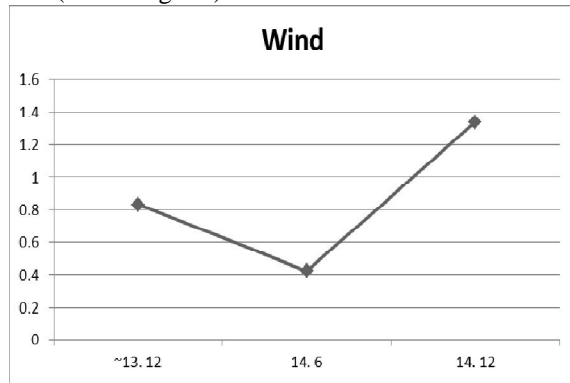

Fig. 4. 'Wind' – Trend of Relative Term Occurrence

CONCLUSIONS AND FUTURE WORK

This research identifies changes of primary keywords related to the renewable energy in Twitter through term life cycle analysis using Relative Term Occurrence. Unlike general way of exploring rapidly emerging terms in social media at issue detection, this research analyzes trend of terms by subdividing the life cycle of terms into growth, deterioration, and maintenance. The limitation of this research is a difficulty to clearly understand the trend of terms due to the 2 years of short collection period for tweet data to be analyzed. Therefore, it is recommended to precede the further study of trend analysis, including creation, growth and maintenance of terms, through additional collection and analysis of long-term data. In further research, it is expected to group keywords with similar life cycle pattern to create predictive model for change of user interest.

REFERENCES

[1] B. J., Jansen, M. Zhang, K. Sobel, and A. Chowdury, "Twitter power: Tweets as electronic word of mouth," Journal of the American society for information science and technology, vol. 60, no. 11, pp. 2169-2188, 2009.

[2] M. Cataldi, L. Di Caro, and C. Schifanella, "Emerging topic detection on twitter based on temporal and social terms evaluation," In Proceedings of the 10th International Workshop on Multimedia Data Mining, p. 4-13, July. 2010.

[3] I. Dincer, "Renewable energy and sustainable development: a crucial review," Renewable and Sustainable Energy Reviews, vol. 4, no. 2, pp. 157-175, 2000.

[4] J. Weng, and B. S. Lee, "Event Detection in Twitter," 5th International AAAI Conference on Weblogs and Social Media, pp. 401-408, May 2011.

[5] M. Mathioudakis, and N. Koudas, "Twittermonitor: trend detection over the twitter stream," Proceedings of the 2010 ACM SIGMOD International Conference on Management of data, pp.1155-1157, June 2010.

[6] D. S. Kim, and J. W. Kim, "Public Opinion Sensing and Trend Analysis on Social Media: A Study on Nuclear Power on Twitter," International Journal of Multimedia and Ubiquitous Engineering, vol. 9, no. 11, pp. 373-384, 2014.

[7] W. X. Zhao, J. Jiang, J. Weng, J. He, E. P. Lim, H. Yan, and X. Li, "Comparing twitter and traditional media using topic models," Advances in Information Retrieval, pp. 338-349, 2011.

[8] Z. Xu, R. Lu, L. Xiang, and Q. Yang, "Discovering user interest on twitter with a modified author-topic model," Web Intelligence and Intelligent Agent Technology (WI-IAT), 2011 IEEE/WIC/ACM International Conference on. IEEE, pp. 422-429, August 2011.

[9] P. Sadorsky, "Renewable energy consumption, CO_2 emissions and oil prices in the G7 countries," Energy Economics, vol. 31, no. 3, pp. 456-462, 2009.

★★★