Advanced Science and Technology Letters
Vol.78 (MulGrab 2014), pp.90-94
http://dx.doi.org/10.14257/astl.2014.78.18

Social Big Data Analysis on Perception Level of Electromagnetic Field

Jwageun Kim[1], Jonghwa Na[2],

[1]Department of Business Data Convergence, Chungbuk National University, Gaesin-dong, Cheongju, Chungbuk, Korea
Jwageun Kim, fedomoon@chungbuk.ac.kr
[2]Department Information and Statistics/BusinessData Convergence, Chungbuk National University, Gaesin-dong, Cheongju, Chungbuk, Korea
Jonghwa Na, cherin@chungbuk.ac.kr

Abstract. The issue of opposition to transmission tower construction is continuously affecting the economy and society as a whole. In order to verify the effect of policies to settle conflicts implemented by the government and create measures to establish new policies to settle conflicts, a survey on the level of perception about the electromagnetic field is necessary. Existing investigations on the level of perception about the electromagnetic field were conducted in the form of surveys centered on direct interested parties and therefore there was a limitation in the scope of surveys. In order to complement the limitation on cost and the number of population, this study conducted analysis of electromagnetic field-related issues, analysis of social networks, and positive and negative analysis based on social big data.

Keywords: social big data, electromagnetic field, issue analysis, social network analysis, positive and negative analysis

1 Introduction

Campaigns against transmission tower construction have been brought into relief as issues of social conflict to be resolved on a government level since they were reported in the press. Therefore, for effective policy establishment, investigations on the level of perception about electromagnetic waves from transmission towers, in other words, the electromagnetic field, should be performed together. Existing investigations in the form of surveys on interested parties have limitation in their investigation scope not sufficiently considering the population. In order to complement such problem, indirect investigation on ordinary people's level of perception about electromagnetic waves needs to be performed. Accordingly, this study collected social information (news, tweets, and comments related to the electromagnetic field), and based on the information conducted analysis of the level of perception and issues about the electromagnetic field, analysis of social networks, and positive and negative analysis.

ISSN: 2287-1233 ASTL

2 Relevant Works

Isaac (2011) presented a keyword analysis method for analyzing social networks [1] and Jo In-dong et al. (2012) came up with relations among keywords in the form of a network as a method to overcome the limitation of keyword searching [2]. Park Ji-hye (2013) studied network influence analysis and marketing utilization through social network analysis in a social network environment [3], and Yeon Jong-heum et al. (2011) examined sensitivity analysis processing modeling with data on reviews about goods [4].

3 Analysis Procedure and Data Collection

3.1 Analysis Process

Ten keywords related to the electromagnetic field were excavated based on home pages of institutions associated with the electromagnetic field and articles and surveys on the electromagnetic field and based on the excavated keywords, data was collected using News RSS Crawler and Twitter Application Programming Interface (API). The collected data was pre-processed and loaded on the database, and issue, social network, and negative and positive analyses of the data loaded on the database were carried out using R, a statistical analysis and visualization tool.

3.2 Data Collection Method

News: A JAVA open source RSS Crawler was modified and made into a collection program and then collection was conducted.

Twitter: Twitter API provided by Twitter was used. In order to easily use Twitter API, twitter4j, a Java API, was used.

Comments related to the electromagnetic field in ordinary home pages: Crawler 4j, a JAVA open source Crawler, should be used but because it was difficult to find comments from people related to the electromagnetic field, passive collection of comments in home pages related to the electromagnetic field or large community sites which people frequented was largely made.

4 Analysis

For the experiment, data was collected from April 2014 to August 2014 (five months) and the number of data by each type was 2372 in news, 542 in tweets, and 2570

comments related to the electromagnetic field in ordinary home pages; A total of 5,483 data were collected.

4.1 Issue Analysis and Social Network Analysis

For issue analysis, nouns were extracted using KoNLP, R's Korean text mining package. Table 1 displays the nouns which appeared most in news, tweets, and electromagnetic field-related comments. The nouns which appeared most are issues in which ordinary people had most interest. In addition, social network analysis using an apriori algorithm of association analysis was performed to identify the most central words and related words and visualize them. Fig 1 visualizes the result of social network analysis in tweets. The most central worlds were transmission towers, Miryang transmission tower, and family, and terms related to transmission towers were sit-in, demolition, opposition, residents, and mud huts and they were those with negative meanings about transmission towers. In addition, terms related to Miryang transmission tower and family included Saenuri Party, missing persons, and apology; There were many tweets about the ruling party Saenuri Party relating the Sewol ferry disaster with Miryang transmission tower.

Table 1. The result of issue analysis by social information

	News	Tweets	Comments
1	transmission tower	transmission tower	transmission tower
2	habitant	demolition	residents
3	police	Miryang transmission tower	transmission electric wire
4	opposition	sit-in	construction
5	sit-in	today	village
6	construction	family	electricity
7	demolition	missing person	we
8	village	pretend	nuclear Power
9	deputy	opposition	person
10	construction	residents	electromagnetic

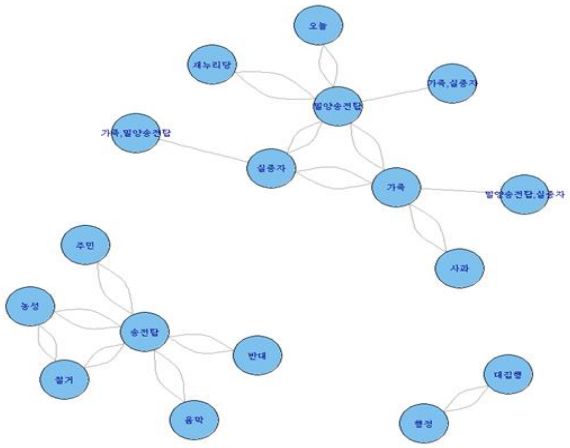

Fig. 1. The visualization of social network analysis of tweets**.**

4.2 Positive and Negative Analysis

Positive and negative analysis was carried out based on a sensitivity classification dictionary. News was excluded from analysis because it mostly delivered facts without sensitivity and therefore positive and negative analysis of it was not able to be made. Sentences in tweets were classified into positive, negative, and objective facts. Among them, objective facts denoted facts where value judgment was not involved at all and they were excluded from analysis. Among 294 facts excluding 248 objective facts from the total 542 facts, the number of negative facts was 255 (87%) and the number of positive facts was 39 (13%); Most facts showed negative attitudes toward keywords related to the electromagnetic field. This was slightly lower than 189 people who responded "the electromagnetic field was harmful" in a survey by the Policy Research Headquarters in 2013 on the electromagnetic fields' harmfulness to the human body with 200 adults across the nation as respondents.

5 Conclusion

This paper selected keywords related to the electromagnetic field in order to investigate people's level of perception about the electromagnetic field. In addition, unstructured data such as news, tweets, and comments related to the keywords were collected and processed, and issue analysis, social network analysis, and positive and negative analysis of them were performed. The most frequent issues related to the electromagnetic field were transmission towers and residents and according to the social network analysis result, many terms related to transmission towers were those

with negative meanings about the towers such as sit-in, opposition, and resistance. In the positive and negative analysis, the rate of negative thoughts (87%) was higher than that of positive thoughts (13%), although the rate of the former was slightly lower than the result of existing surveys.

In case the amount of collected data related to the keywords is increased and a morpheme analyzer for Korean language with better performance than the KoNLP package morpheme analyzer used in this paper is applied, analysis performance of future research is expected to be heightened.

Acknowledgment. This research was supported by the MSIP(The Ministry of Science,ICT and Future Planning), Korea, under the "SW master's course of a hiring contract" support program (NIPA-2014-HB301-14-1011) supervised by the NIPA(National IT Industry Promotion Agency)

References

1. Scott Isaacs, "The Analytic Difference - Better Performance and Deeper Insight", SAS forum KOREA, 2011.
2. Indong Cho, Namgyu Kim, Keeyoung Kwahk, “Recommending Core Research Keywords Using Social Network and Data Mining Analysis”, Entrue Hournal of Information Technology, Vol 11, No 1, 2012. 04.
3. JiHye Park, A Study of Network Influence Analysis by Using Social Network Analysis Methods and Marketing application in Social Network Environment, Graduate School of Sookmyung Women's University, 2013.
4. Jongheum Yeon, Dongjoo Lee, Junho Shim, Sang-goo Lee, "Product Review Data and Sentiment Analytical Processing Modeling", Korea Electronic Commerce Journal: Vol. 16, No. 4, pp1-13, 2011.
5. To cite package ‘KoNLP’ in publications use: Heewon Jeon (2012). KoNLP: Korean NLP Package. R package version 0.76.5. http://CRAN.R-project.org/packge=KoNLP.

