The Workshops of the Tenth International AAAI Conference on Web and Social Media
Social Web for Environmental and Ecological Monitoring: Technical Report WS-16-20

Public Discourse on Environmental Pollution and Health in Korea: Tweets Following the Fukushima Nuclear Accident

Seung-Hoi Kim[1] Yu-I Ha[1] Meeyoung Cha[1] Jiyon Lee[2] Byoung-Jik Kim[2] Dong-Myung Lee[2]

1Graduate School of Culture Technology, Korea Advanced Institute of Science and Technology, South Korea

2Korea Institute of Nuclear Safety, South Korea

{skylugia, yuiha, meeyoungcha}@kaist.ac.kr, {jylee, bjkim, k320ldm}@kins.re.kr

Abstract

Public discourse on environmental and health issues has risen on social media. Upon an environmental crisis, various chatters such as breaking news, misinformation, and rumor could aggravate social confusion and proliferate negative public sentiment. In an effort to study public sentiments on environmental issues in South Korea, we analyzed 158,964 tweets generated over a 4-year period following the Fukushima accident in 2011, the largest release of radioactivity to environment in recent history. This event led to a significant increase in public's interest on environmental and nuclear issues in Korea. We employed Bayesian network and recursive partitioning to observe the classification regression tree structure of major topics. Topics on health and environment were interlinked closely and represented both apprehension and concern about health threats and pollution. Our methodology helps analyze large online discourse efficiently and offers insight to crisis response organizations.

Introduction

Environmental issue has never been more important than it is now, and societies around the world are fully aware of its impact and implication. When severe environmental pollution occurs, we see uptakes of online discourse. For instance, oil spills and nuclear leakages that directly influence public health and nature raise the problem to a social dimension.

Soon after such accidents, public opinions are formed through various platforms including social network service (SNS). At the same time, untested stories, misinformation and rumors may spread public anxiety and social confusion. Tremendous scale of social media allows us to observe public perception on certain issues via efficient use of data.

On March 11th 2011, a magnitude 9.0 earthquake hit the Northeastern Honshu Coast in Japan and resulted in meltdown of Fukushima nuclear power plants. This led to a considerable emission of radioactive materials into the environment. This leak is the most serious environmental contamination in recent years and the public awareness toward environmental issues has risen significantly.

South Korea, being Japan's closest neighbor and possessing nuclear power plants, was affected immediately via import of radiation-polluted marine products as well as gradually via the burst of public discussion on nuclear safety that ranged from radio-active contamination to nuclear waste, marine safety, health fears, etc.

We analyzed Twitter data containing the discourse on the Fukushima's radioactive water leak to investigate the public perception towards environmental pollution and health in Korea. We made three key observations:

First, we measured the degree of information propagation by utilizing retweet (RT) function, which is mainly used as an explicit measure of information diffusion. The Fukushima accident generated an average of 100 or more Korean tweets regarding radiation per day. Compared to the number of tweets before the accident, the number after the tweets saw an increase of more than 10 times. In particular, significant issues such as leakage of radioactively contaminated water entailed active information propagation. Therefore, we regarded the Korean public as sensitive to the environmental influence of nuclear radiation.

Second, we identified a set of 43 keywords with 4 topic categories from 500 sample tweets by manually coding them through iterative discussion. The keywords of the health topic included radiation exposure, leukemia, cancer, pregnancy, etc. For the environment topic, keywords such as pollution, fishery product, import, sea and leakage existed. The 19 keywords that belong to the health and environment topics were mainly involved in concerns about the result of radiation exposure and pollution.

Third, we analyzed the health and environment topics with the Bayesian networks and conducted recursive partitioning analysis to investigate online information propagation specifically. Radiation safety and import had a dominant position within each topic, followed by the remaining keywords, such as pollution, fishery product, cancer and groceries strongly influencing each other. Through this

process, we understood local distribution of the keywords and created a specified decision rule.

Our methodology helps to identify main keywords, allows make intuitive and subdivided distribution model about online discourse. Social media facilitate communication and help form social capitals prior to occurrence of such catastrophic situation (Kaigo, 2012). Therefore, our finding can be utilized for providing information strategies to help the public understand about emergency situation and offer insight to crisis response organization.

Methodology

We gathered 158,964 Korean tweets from January 1, 2011, two months prior to the Fukushima accident, until December 14, 2014, by searching for streams containing the following keywords: Radioactive, Radiation, Nuclear Power, as well as their variations. For each tweet, we know the time of posting, username, number of retweets, favorites, and comments.

The data show a critical change in terms of total number of tweets and retweets, where the average number of daily tweets mentioning nuclear topics increased by more than 10 times from 10 tweets per day before the Fukushima accident to over 100 tweets per day after the event. Prior to the crisis, most mentions were not related to any health or environmental threats (e.g., department of radiology).

The frequency of tweets over the 4-year period captures well the major events such as the Tohoku Earthquake, and apology by the Japanese government for the illegal discharge of radioactively contaminated water and entails an active level of information propagation and public attention. This implies that the Korean public became sensitive to potential threat of a nuclear accident and actively propagated information on issues related to nuclear power.

Result

Perception towards the Nuclear Issues

High-quality information and well-organized risk strategies that have been prepared for potentially affected communities increase the level of preparedness, resulting in efficacious disaster management (Svendsen, 2013). Searching for topics of discourse on environmental and nuclear issues should be a sound basis for establishing effective countermeasures upon crisis.

To identify major keywords from data, a random sample of 500 tweets were manually coded, where all of the authors read the online discourse and coded them based on inductive arguments. The obtained keywords were validated through iterative discussion until an agreement was reached among the coders. In order to build a higher-level concept, the authors categorized and organized keywords into hierarchical orders, as described in Table 1.

Table 1 Classification of 43 keywords on nuclear issues

Topic	Keywords
Location (18.6%)	Domestic, Republic of Korea, Our country, Korean Peninsula, Korea, Tokyo, Fukushima, Japan
Health (20.9%)	Thyroid, Health, Leukemia, Maternity, Cancer, Pregnancy, Radiation Exposure, groceries, Food
Environment (23.3%)	Leakage, Sea, Import, Fishery products, Pollution, Environment, Agricultural products, Leakage, Travel, Forest resources
Nuclear Power (37.2%)	Cesium, Iodine, Operation, Alert, Worn out, Security, Black out, Corruption, Accident, Stability, Media control, Risk management, Risk, Concealment, Doubt, Fact

Keywords in the location category were mainly associated with comparison between Japan and Korea. We deduced that Koreans were concerned about the harmful consequences of the nuclear accident from Japan. Health category was divided into two main keywords: food and human body. Keywords such as thyroid, leukemia, and cancer seem to have resulted from an anxiety about exposure to radiation and its effect on health. This tendency also emerged in the environment category as concerns about the contamination of agro-fishery products, trading of contaminated goods, and safety of traveling to Japan.

Hence, which keywords had the highest frequencies and how to relate them to each other are the most pressing tasks for understanding online discourse related to environmental pollution and the proper measures to take for prevention.

Bayesian Network and Recursive Partitioning Analysis on Health and Environment Keyword

Since keywords in health and environment topics directly affect safety of human body and environment, the corresponding topics received much more public attention than the other topics, on location and nuclear power.

We now observe the mentioned rate classified by keywords via Bayesian networks. In this graphical model, nodes represent random variables and arrows represent the probabilistic dependencies among nodes that are identified by recursive partitioning analysis, a statistical method for multivariable analysis on each topic. This model helps us understand a sparse set of keywords of direct dependencies.

Discourse on Health Topics

Figure 1 shows the pie-chart encompassing each keyword in the health topic. Radiation Exposure is the most frequently mentioned keyword, followed by cancer, groceries, health, food, thyroid, and leukemia in order of frequency.

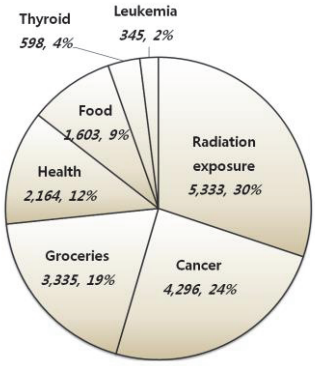

Figure 1 Prominent health topics in Korean tweets

In Figure 2, the boxes at the bottom represent the frequency distribution. The recursive partitioning analysis may proceed until the tree is saturated in the sense that the offspring nodes can no longer be subjected to further division (Zhang and Singer, 2010). The root node, groceries, was split into nodes 2 and 3 with criterion 7.5. In node 2, 34,952 people who mentioned radiation exposure talked about groceries. If the mention of groceries is 7.5 times or more, the node is split into two further parts based on the number of comments about cancer. According to nodes 4 and 5, only 45 people mentioned all of the keywords: radiation exposure, groceries, and cancer.

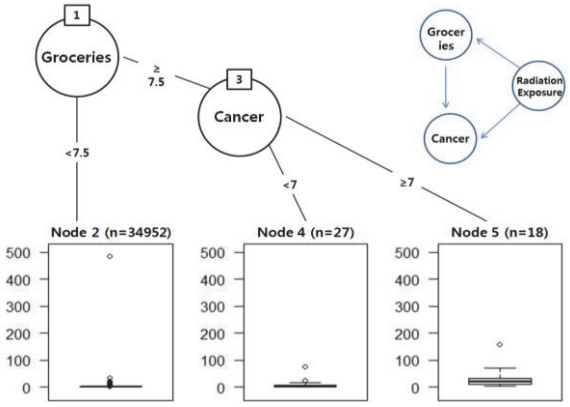

Figure 2 Bayesian network and tree structure of topics regarding radiation exposure

The Bayesian networks drawn at the top right of Figure 2 are representative of the relationships among the three keywords. Radiation exposure is a parent node of both *groceries* and *cancer* nodes, as well as a forebear node of cancer. Also, groceries are a parent node of cancer. Radiation exposure has achieved the most dominant position within the health topic and the remaining keywords have played a subsidiary role. This illustrates the concern that radiation exposure would lead to food contamination, which eventually lead to the topic of cancer. In particular, the public expressed worries about the incidence of diseases such as thyroid and leukemia.

Discourse on Environmental Topics

Figure 3 shows that *pollution* is the most frequently mentioned keyword in the environmental topic. Twitter users exhibited worries about pollution with their opinions (e.g. *"radiation-contaminated scrap metal imported from Japan is okay"*) and unconfirmed information (e.g. *"The Korean Peninsula and coastal waters are heavily contaminated by radiation. Pollution is almost at the same level as that of the central region in Japan."*). Some users expressed emotional reactions and complained about the government.

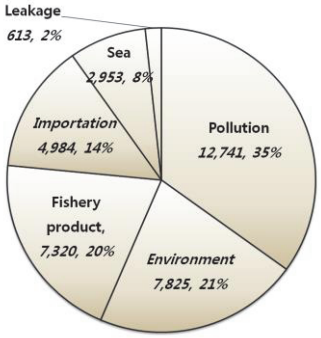

Figure 3 Prominent environmental topics in Korean tweets

The keyword "environment" was mainly used to discuss government policies regarding nuclear plants (*e.g. "The environmental policy of Germany was changed radically after the Fukushima nuclear disaster. All nuclear power plants in Germany will be closed until 2022*."). Some public institutions, such as the Korean Nuclear Environment Corporation, also participated actively on Twitter to distribute news and information about their promotions.

The keyword *fishery product* includes various keywords, such as mackerel, fish and seafood. These keywords were used to show Twitter users' concern over the safety of marine products that are being endangered by marine pollution. Note that the keywords often involve *import*, which shows.

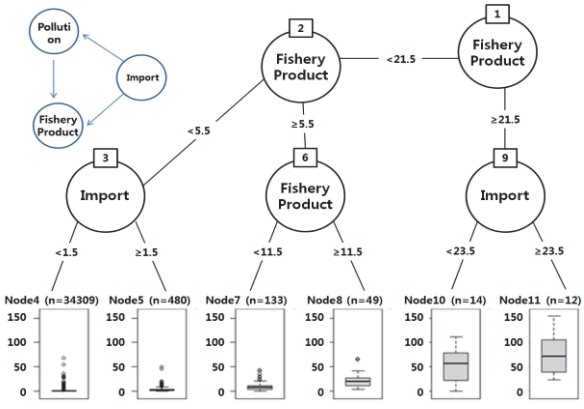

Figure 4 Bayesian network and tree structure of pollution-related topics

Using the discourses that mentioned pollution, we analyzed the Bayesian network and tree-structure of the keywords, *import* and *fishery product* (Figure 4). Node 1 (fishery product) was divided into nodes 2 and 9, with the former subdivided into nodes 3 and 6. Node 3 was the most frequently mentioned node (34,789 tweets), along with pollution, fishery product and import. However, node 9 that mentioned fishery product 21.5 times or more and mentioned importation more than once is only 26 contrastively. Referring to import among pollution discourses was concluded in the case of mentioning fishery product under 5 times or over 21.5 times. According to the Bayesian network, import and pollution influenced mentions of fishery product, while pollution itself influenced mentions of fishery product.

Discussion and Conclusion

The main goal of this research was to measure and analyze the awareness and concern about environmental pollution caused by the Fukushima nuclear accident. We addressed this goal by conducting information propagation analysis, keywords classification, Bayesian networking and recursive partitioning analysis from 158,964 tweets.

We found the following results: First, the temporal change in the total number of tweets and RT counts for each day of the monitored periods are observed. Following the Fukushima accident, an average of a 100 tweets per day have been generated, and in particular, significant issues such as leakage of radioactively contaminated water entailed active information propagation. Despite the increase with time in the general usage of Twitter, the continued information propagation trend about nuclear issue throughout the 4-year study period is noteworthy. Second, iterating topic tagging steps, we identified 43 keywords, which were then categorized into 4 higher-level concepts. The health and environment topics further included 19 keywords that concerned radiation exposure and pollution. The total number of tweets (36,436) on the environment topic was more than twice that of health-related discourse (17,674). In the environment topic, the pollution was used 12,741 times (35%), and environment itself was mentioned 7,825 times (21%). Since these two keywords comprised more than half of the total, we regarded the environment topic after the nuclear disaster as having repercussions on the public's concern. Third, we identified main keywords and observed influencing relationship among the keywords in health and environment topics by employing a network analysis. Topics were interlinked closely and represented both apprehension and concern about health threats and pollution. In the future, we would like to match online discourse with the user profiles to understand the precise relationship between demographics and topics.

Our observations bring insight on the public's discourse on health and environmental issues related to nuclear accidents. Mentions of thyroid cancer and rumors regarding mutations were the most talked-about topics after the nuclear disaster. However, many experts doubt that incidences of cancer are related to the nuclear disaster (Kyodo, 2015), and suggest the possibility of overdiagnosis (Katanoda et al., 2016). It is interesting to confirm that the public continues to propagate the unverified causal relationships among radiation exposure, physical health, and environmental pollution. Furthermore, we found that government organizations did not engage in similar topics as the public. For instance, some government institutions used Twitter to notify the public of their events (e.g., *Can you guess the answer to the quiz in July issue of our magazine?*) rather than addressing the pressing concerns. This discrepancy indicates an opportunity for better communication and a need for a better strategy to handle misinformation. The methodology presented in this case study can be easily applied to characterize online discourse of other health and environmental issues. In the future, we would like to delve deeper into findings by jointly looking at sentiments and user demographics.

Acknowledgements

The first two authors contributed equally to this paper. This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIP) (R0184-15-1037, Development of Data Mining Core Technologies for Real-time Intelligent Information Recommendation in Smart Spaces)

References

Kaigo, M., 2012. *Social media usage during disasters and social capital: Twitter and the Great East Japan earthquake.* Keio Communication Review, 34, pp.19-35.

Breiman, L., Friedman, J., Stone, C.J. and Olshen, R.A., 1984. *Classification and regression trees.* CRC press.

Svendsen, E.R., 2013. *A new perspective on radiation risk communication in Fukushima,* Japan. J. Natl. Inst. Public Health, 62(2), p.196.

Zhang, H. and Singer, B.H., 2010. *A practical guide to tree construction. In Recursive Partitioning and Applications* (pp. 9-22). Springer New York.

Kyodo, 2015, *New report links thyroid cancer rise to Fukushima nuclear crisis.* The Japan Times

Katanoda, K., Kamo, K.I. and Tsugane, S., 2016. *Quantification of the increase in thyroid cancer prevalence in Fukushima after the nuclear disaster in 2011—a potential overdiagnosis?*. Japanese journal of clinical oncology, p.191.