Nuclear Engineering and Technology 53 (2021) 1013–1019

Contents lists available at ScienceDirect

Nuclear Engineering and Technology

journal homepage: www.elsevier.com/locate/net

Original Article

Sentiment analysis of nuclear energy-related articles and their comments on a portal site in Rep. of Korea in 2010–2019

So Yun Jeong [a], Jae Wook Kim [a], Young Seo Kim [a], Han Young Joo [a], Joo Hyun Moon [a, *]

[a] Dankook University, 119 Dandae-ro Dongnam-gu, Cheonan-si, Chungnam, 31116, Republic of Korea

ARTICLE INFO

Article history:
Received 15 June 2020
Accepted 23 July 2020
Available online 2 August 2020

Keywords:
Articles and comments
Nuclear energy
Sentiment analysis
Big data analysis
Public opinion

ABSTRACT

This paper reviewed the temporal changes in the public opinions on nuclear energy in Korea with a big data analysis of nuclear energy-related articles and their comments posted on the portal site NAVER. All articles that included at least one of “nuclear energy,” “nuclear power plant (NPP),” “nuclear power phase-out,” or “anti-nuclear” in their titles or main text were extracted from those posted on NAVER in January 2010–December 2019. First, we performed annual word frequency analysis to identify what words had appeared most frequently in the articles. For that period, the most frequent words were “NPP,” “nuclear energy,” and “energy.” In addition, “safety” has remained in the upper ranks since the Fukushima NPP accident. Then, we performed sentiment analysis of the pre-processed articles. The sentiment analysis showed that positive-tone articles have been reported more frequently than negative-tone over the entire analysis period. Last, we performed sentiment analysis of the comments on the articles to examine the public’s intention regarding nuclear issues. The analysis showed that the number of negative comments to articles each month—irrespective of positive or negative tone—was always larger than that of positive comments over the entire analysis period.

1. Introduction

Nuclear power is double-sided. It generates power economically, emits little carbon dioxide during power generation, and its fuel is easy to store, which is favorable for national energy security. However, it raises fear of severe accidents or radiation exposure among people and there are concerns over the safe management of spent nuclear fuel. As it is clearly double-sided, the public’s perception of nuclear energy may be influenced by whatever information they have gathered or their personal tendency.

As the public acceptance of nuclear energy has a large influence on national policy and the construction of any nuclear facilities, it is necessary to survey the public’s general perception of nuclear energy periodically or their opinions about a particular nuclear issue. Therefore, the survey should be objective and impartial to get their real intention as accurately as possible. Survey methods such as telephone polling and face-to-face interviews have been traditionally used to gather public opinions. However, the surveys’ anonymity means that many of those questioned will likely not give frank answers or make completely different responses that contrast with their real opinion [1].

To supplement their weakness and grasp the overall tendency of public opinion at a single glance, a method of analyzing big data on internet articles or social network services has recently emerged. Natural language processing is important for analyzing the public’s implicit emotions or opinions in big data, and many studies have been performed to improve processing. Natural language processing aims to obtain people’s feelings expressed as positive or negative comments by analyzing a large number of documents [2]. This has been used in various fields such as marketing, travel, and customer satisfaction. In marketing, many companies analyze customer reactions through their posts in social network service and reflect them in the sales strategy [3]. A group of Turkish experts performed sentiment analysis of mentions posted on Twitter with machine learning to investigate personal opinions about their travels [4]. Seo et al. identified meaningful customer responses and requirements after extracted keywords using a Term Frequency - Inverse Document Frequency (TF-IDF) method [5].

In nuclear energy, many studies have been performed that applied big data analysis methods to analyze internet articles or social network services. Park performed sentiment analysis of keywords for nuclear issues from articles posted on social network

* Corresponding author.
E-mail address: jhmoon86@dankook.ac.kr (J.H. Moon).

https://doi.org/10.1016/j.net.2020.07.031
1738-5733/

services such as Facebook, Twitter, Kakaostory, and Naver band [6]. Roh collected data from the Web, Twitter, and NAVER to extract the top-five keywords and identify their perception of nuclear power through the analyzing search words on the Internet from people who live in Seoul and around the Kori nuclear power plant (NPP) [7]. Park et al. analyzed press releases on Korea's energy policy and nuclear energy policy using Latent Semantic Analysis (LSA), which is a natural language processing method [8]. Jang et al. analyzed the public perception of NPP by applying several models based on surveys taken in Korea after the Fukushima accident [9].

If we only performed frequency analysis, we could not be aware of the tendencies of sentiments in articles and the trend of public opinion. Then, sentiment analysis has emerged as a way of studying and construing public opinion [10]. Most previous studies performed frequency analysis and sentimental analysis together. However, few studies have focused on the comments on articles, which is also important since such comments may be regarded as an alternative expression of the public's opinion about a certain issue. Unlike previous studies, we matched the nuclear energy-related articles with the comments written on them so that we could grasp the tendency of articles and comments at a glance. By matching articles with comments, we hope to enhance the reliability of the analysis for the results of public opinion. For this, we applied the big data analysis software R to take natural language processing, frequency analysis, and sentiment analysis.

2. Materials and method

2.1. Analysis objects

First, we collected the URLs of all articles that included at least one of "nuclear energy," "NPP," "nuclear power phase-out," or "anti-nuclear" in their titles or main texts from articles posted on NAVER on a monthly basis in January 2010–December 2019 and extracted the titles and the main texts of those articles whose URLs were collected with help of the web scraping packages "tidyverse" and "rvest" for R language. The articles were of various types: photo news, video clips, press releases, and news.

Fig. 1 shows the whole analysis procedure. We made three types of analysis for the collected articles. First, we identified the frequency with which each word occurred in the articles and sorted the words by their frequency of occurrence in the articles. Then, excluding unnecessary words, we screened the top-25 ranked words in terms of annual occurrence frequency. We used "KoNLP" to analyze morphemes and extract nouns from the words in the articles.

Second, we performed sentiment analysis of all articles to determine if the tone of argument in each article is positive or negative regarding nuclear energy and score the articles according to the degree of their tone of argument. For this, we used the "KNU Korean language sentiment dictionary" that sorts each word into one of two groups of positive and negative sentiment words by analyzing their meaning. Since this includes various words with positive or negative meanings in the form of phrase, sentence, abbreviation, or emoticon, which have also been used in diverse areas such as movies, music, and automobiles, we judged it suitable for sentiment analysis.

Third, we screened out the months whose positive- or negative-tone scores were higher than the average, collected the comments for all the articles issued in those months and tried to identify any relationships between the tendencies of argument tone in the articles and the tendency of argument tone in the comments. We used the same method as those above to extract the URLs of the articles and used the "getAllComment" function in the "N2H3" package to collect comments.

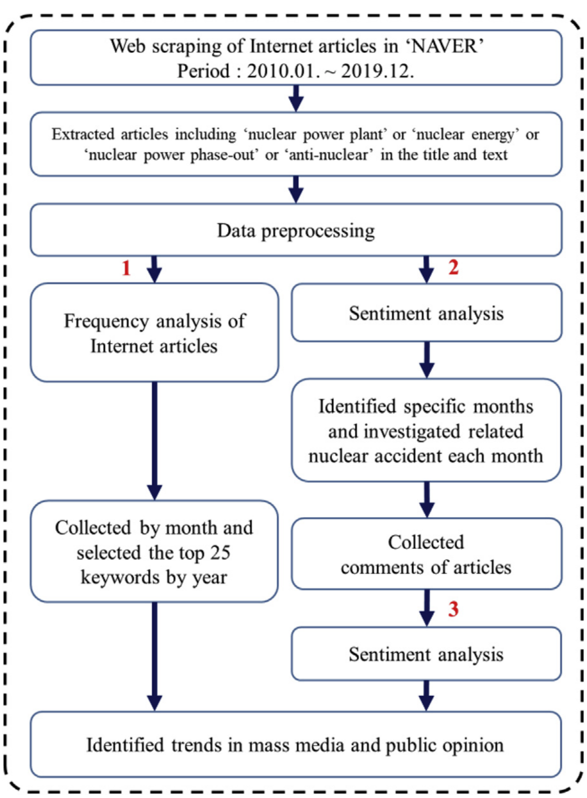

Fig. 1. Overview of the analysis procedure.

2.2. Pre-processing

The collected articles should be pre-processed before the main analysis to obtain the analysis results with reliability. First, we screened out duplicate articles collected through web scraping and removed words, special characters, punctuation, and numbers that were unnecessary for our analysis while extracting nouns from the article sentences. For prompt analysis of annual frequency words, we extracted nouns from the sentences in the articles on a monthly basis for a corresponding year. In doing this, we located and removed the words that had not been removed in the previous step but were meaningless. In addition, we added words such as "nuclear power phase-out" and "anti-nuclear" that were required for the main analysis from the KNU dictionary. In addition, we applied the same procedure to analyze the article comments.

2.3. Frequency analysis

We examined the annual frequencies of words to find the events at issues in the year that were newsworthy. For this, we extracted the 30 most frequently mentioned nouns each month using the "extractNoun" function and "SejongDic" function built into the "KoNLP" package. It was because unnecessary words may still be included. We arranged the extracted nouns by monthly frequency of occurrence in the articles and made a table that included the frequency analysis results. Then, we again removed unnecessary words that had not been removed in previous steps. Lastly, we combined the 12 monthly frequencies of the extracted words into their annual frequencies and selected the final 25 words with the highest annual frequency.

Table 1
Top-25 English words and their corresponding Korean words with highest annual frequency of occurrence in the articles per year.

Rank	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
1	NPP 원전	NPP 원전	NPP 원전	NPP 원전	Nuclear Energy 원자력	NPP 원전	NPP 원전	NPP 원전	NPP 원전	NPP 원전
2	Nuclear Energy 원자력	Nuclear Energy 원자력	Nuclear Energy 원자력	Nuclear Energy 원자력	Safety 안전	Nuclear Energy 원자력	Nuclear Energy 원자력	Nuclear Energy 원자력	Nuclear Energy 원자력	Nuclear Energy 원자력
3	Research 연구	Power generation 발전	Safety 안전	Power generation 발전	Government 정부	Safety 안전	Safety 안전	Energy 에너지	Energy 에너지	Safety 안전
4	Government 정부	Government 정부	Power generation 발전	Safety 안전	NPP 원전	Technology 기술	Construction 건설	Safety 안전	Safety 안전	Energy 에너지
5	Export 수출	Safety 안전	Government 정부	Government 정부	Power generation 발전	Resident 주민	Power generation 발전	Power generation 발전	Power generation 발전	Power generation 발전
6	Overseas contract 수주	Japan 일본	Energy 에너지	Power 전력	Research 연구	Power generation 발전	Government 정부	Construction 건설	Construction 건설	Technology 기술
7	Business 사업	Economy 경제	Japan 일본	KHNP Co. (주)한국수력원자력	Japan 일본	Government 정부	KHNP Co. (주)한국수력원자력	Government 정부	Business 사업	Government 정부
8	Power generation 발전	Reactor 원자로	Technology 기술	Economy 경제	KHNP Co. (주)한국수력원자력	Construction 건설	Technology 기술	Industry 산업	Government 정부	Construction 건설
9	Technology 기술	Construction 건설	Resident 주민	Technology 기술	Construction 건설	Research 연구	Occurrence 발생	Policy 정책	Economy 경제	Economy 경제
10	Plan 계획	Technology 기술	Operation 가동	Energy 에너지	Technology 기술	Energy 에너지	Business 사업	Technology 기술	Technology 기술	Industry 산업
11	Economy 경제	Energy 에너지	Reactor 원자로	Japan 일본	Business 사업	KHNP Co. (주)한국수력원자력	Research 연구	People 국민	Industry 산업	Policy 정책
12	Construction 건설	Occurrence 발생	Power 전력	Construction 건설	Industry 산업	Business 사업	Resident 주민	Business 사업	Research 연구	Business 사업
13	Development 개발	Research 연구	KHNP Co. (주)한국수력원자력	Industry 산업	Samcheok 삼척	Japan 일본	Japan 일본	Declaration 신고	Policy 정책	Research 연구
14	UAE UAE	Development 개발	Economy 경제	Research 연구	Energy 에너지	Industry 산업	Energy 에너지	Research 연구	KHNP Co. (주)한국수력원자력	KHNP Co. (주)한국수력원자력
15	Energy 에너지	Power 전력	Construction 건설	Investigation 수사	Wolseong 월성	Opposition 반대	Earthquake 지진	Economy 경제	Power 전력	Reactor 원자로
16	Safety 안전	Radioactivity 방사능	Occurrence 발생	Test 시험	Resident 주민	Wolseong 월성	Gyeongju 경주	President 대통령	President 대통령	Japan 일본
17	USA 미국	Radioactive 방사성	Research 연구	Operation 가동	Support 지원	Cooperation 협력	Busan 부산	Citizen 시민	Decommissioning 해체	Development 개발
18	Industry 산업	Operation 가동	Kori 고리	Development 개발	Economy 경제	Operation 가동	Development 개발	NSSC 원자력안전위원회	UAE UAE	People 국민
19	Reactor 원자로	Radiation 방사선	Yeonggwang 영광	The Prosecution 검찰	Busan 부산	Development 개발	Facility 시설	Power 전력	People 국민	Contaminated water 오염수
20	Cooperation 협력	Resident 주민	Business 사업	USA 미국	Invitation 유치	USA 미국	Report 신고	Cessation 중단	Promotion 추진	Decommissioning 해체
21	President 대통령	Earthquake 지진	Wolseong 월성	Radioactivity 방사능	Vote of resident 주민투표	Radiation 방사선	Reactor 원자로	Decision 결정	Occurrence 발생	Coal 석탄
22	Turkey 터키	Detection 검출	Reoperation 재가동	Manufacturer 업체	Operation 가동	Reactor 원자로	Operation 가동	Public deliberation 공론화	Visit 방문	NSSC 원자력안전위원회
23	Support 지원	Accident 사고	Discontinue 중단	Contaminated water 오염수	Development 개발	Power 전력	People 국민	Occurrence 발생	Saudi Arabia 사우디	U.S.A 미국
24	Contract 체결	Business 사업	Failure 고장	Counterfeit 위조	President 대통령	Support 지원	Radioactivity 방사능	Renewable energy 신재생에너지	Japan 일본	Fine dust 미세먼지
25	India 인도	Invitation 유치	People 국민	Science 과학	Opposite 반대	Promotion 추진	Ulsan 울산	Earthquake 지진	Transformation 전환	Wolseong 월성

*NPP: Nuclear Power Plant, NSSC: Nuclear Safety and Security Commission.

2.4. Article sentiment analysis

To find the tone of argument in the collected articles, we made a sentiment analysis of the articles' main text. For this, we extracted the positive- and the negative-tone words from the articles using the KNU Korean Sentiment Dictionary. We calculated and scored the difference between the numbers of positive and negative words in each article to check the overall tone of arguments in articles. We set the neutral tone if the score was 0, a positive tone if the score was > 0, and a negative tone if the score was < 0. Then, we

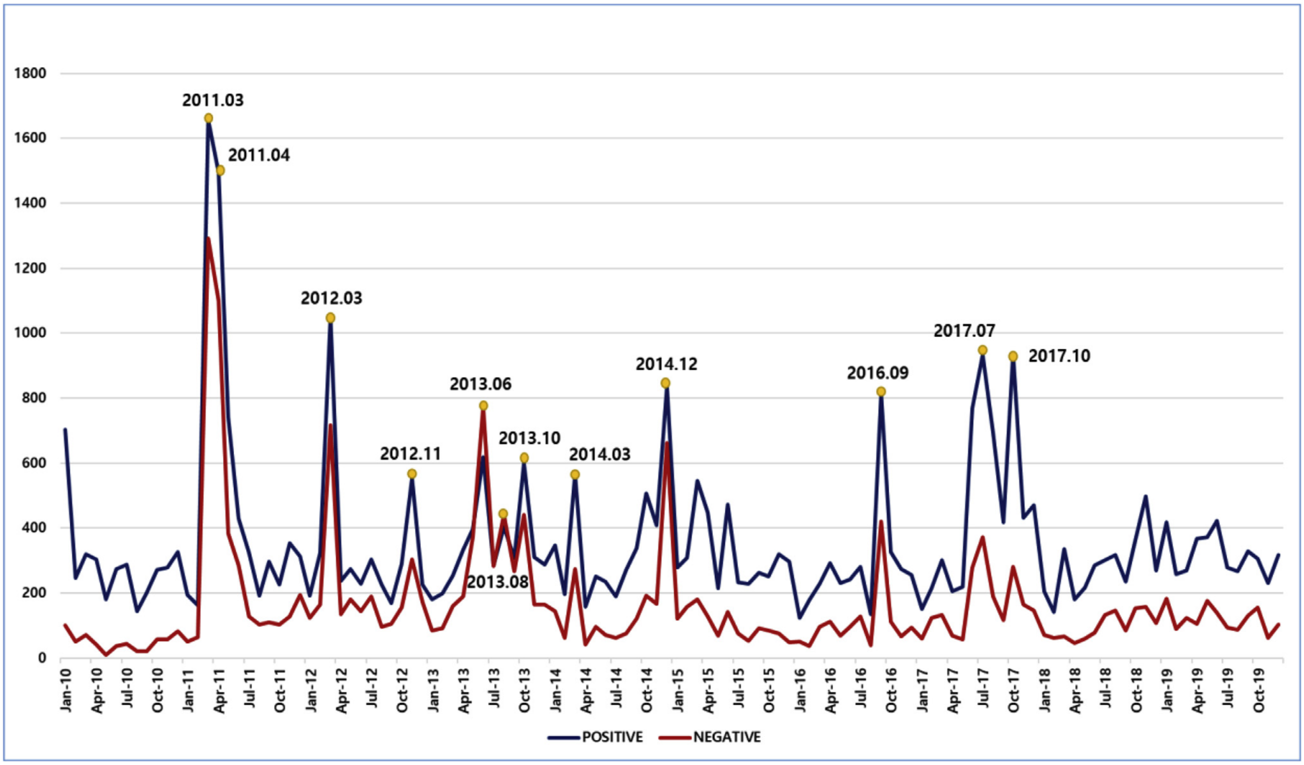

Fig. 2. Temporal changes in the total sentiment scores of articles each month.

Table 2
Major incidents or issues that occurred in the months with high sentiment score.

Time (Month–Year)	Major nuclear-related incidents or issues
March 2011	• Fukushima NPP accident
April 2011	• Amendment of the Atomic Energy Act
March 2012	• The first anniversary of the Fukushima accident
November 2012	• Groundbreaking ceremony of UAE NPPs,
	• The emergence of “nuclear power phase-out” argument
June 2013	• Scandal over fake safety certifications for parts in NPP
August 2013	• Corruption in the Korean nuclear power industry,
	• Release of Fukushima radioactive material into the ocean
October 2013	• Indictment of the people of corruption of in a scandal over fake safety certifications for parts in NPP
March 2014	• Failure to pass the amendment of “the Act on Physical Protection and Radiological Emergency”
December 2014	• Cyberattack on NPP
	• Exposure of sensitive information about an NPP
	• Nitrogen gas leak at Shinkori unit 3
September 2016	• Earthquakes in Gyeongju
July 2017	• Permanent shutdown of Kori Unit 1
	• Public deliberation on the resumption of construction on Shin-kori units 5 and 6
October 2017	• Determination of resumption of construction on Shin-kori units 5 and 6
	• Nuclear power phase-out policy

identified articles with the highest positive or negative scores, respectively, and analyzed the monthly tendency of articles issued for each month.

2.5. Comment sentiment analysis

Finally, we checked the public's perception of nuclear energy through comment sentiment analysis. We found that > 80% of collected articles did not have any comments and only a small portion of articles had a lot of comments. Considering these characteristics, we collected articles with comments and analyzed articles and comments separately to identify the relationship—if any—between the argument tone in the articles and that in the comments. For this, we applied the same procedure and analysis tool in Section 2.4.

First, we screened out articles in the months whose scores were higher than the average and analyzed the article comments. We extracted and collected the URLs of the articles posted on those months and collected the comments to articles by using the “getAllComment” function in the “N2H4” package. We scored the difference between the number of positive words and the number of negative words for each comment similar to the article sentiment analysis. Lastly, we made a scatterplot of the scores for the articles and their comments.

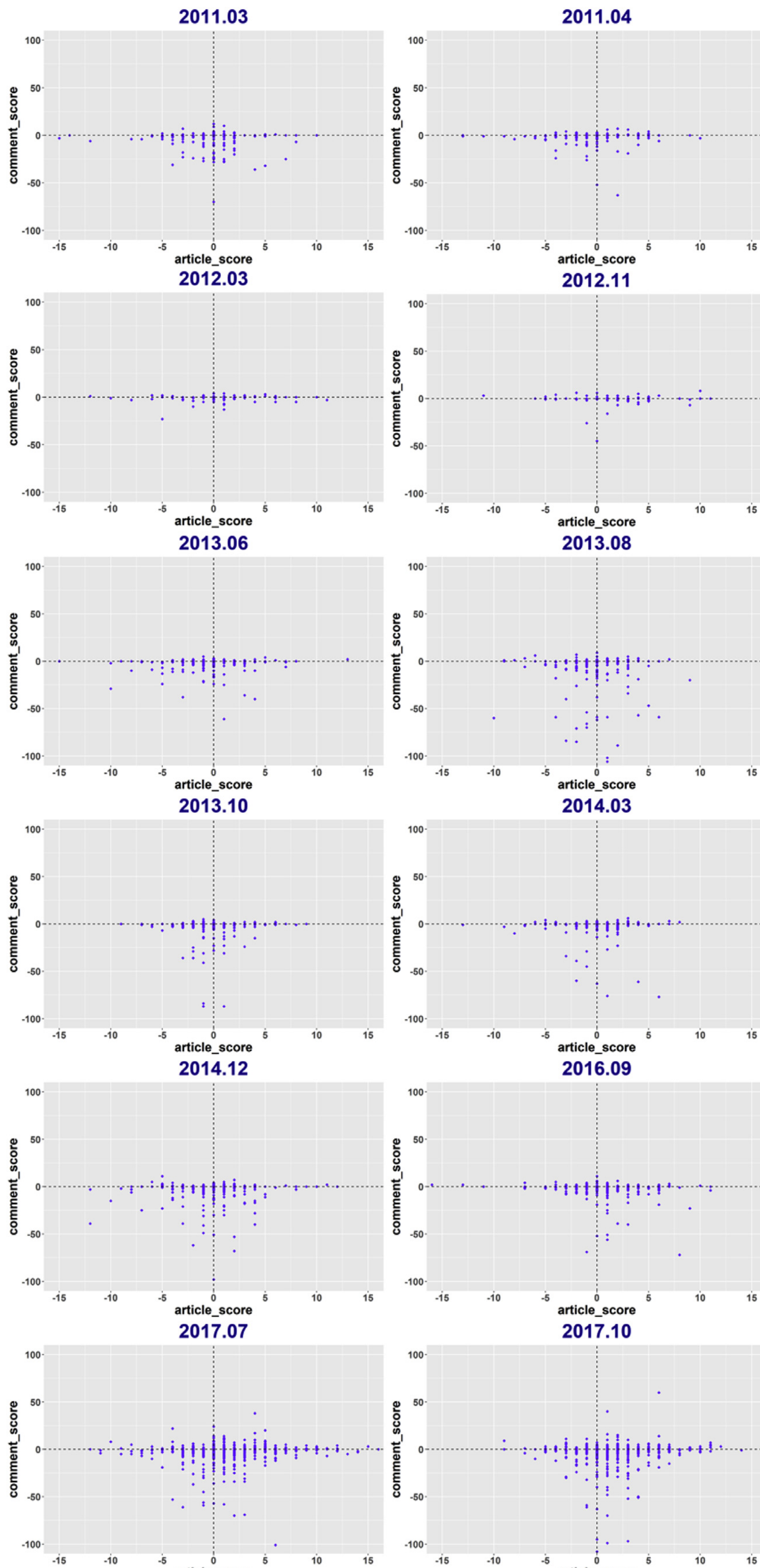

Fig. 3. Sentiment scores of articles and their comments for 12 months in which there was a specific issue.

3. Results

3.1. Frequency analysis

The total number of articles collected for our analysis from those posted on NAVER in January 2010–December 2019 was 215,744, averaging 21,574 articles per year.

Table 1 shows the words with the 25 highest annual occurrence frequency in the articles issued each year. The words that occurred most frequently across the entire analysis period were “NPP,” “nuclear energy,” and “energy.” The word “safety” has remained in the upper ranks since the 2011 Fukushima NPP accident. As nuclear projects are usually government-driven in Korea, words associated with government also appeared frequently.

The frequently mentioned words for each year implied that nuclear issues or incidents associated with those words had happened in that year. For instance, the words most frequently mentioned in 2011 were “radiation,” “radioactivity,” “safety,” and “Japan” due to the Fukushima NPP accident. When the public deliberation on the construction of Shin Kori units 5 and 6 was held in 2017, the most frequently mentioned words were such as “people,” “nuclear safety and security commission (NSSC),” “suspension of construction,” “decision,” and “renewable energy.” Using the frequency analysis results, we can guess what the major nuclear issue was at that time.

3.2. Monthly article sentiment analysis

We made a sentiment analysis of the extracted articles above. We counted the numbers of positive or negative words in each article and assigned +1 to a positive-tone word, −1 to a negative-tone word, and 0 to a neutral-tone word. Then, we converted the difference between the two numbers into a sentiment score for each article. Fig. 2 shows the temporal changes in the total positive-tone and negative-tone sentiment scores of the articles each month. As shown in Fig. 2, more positive-tone articles had been posted than negative-tone articles across the whole analysis period. The monthly negative-tone sentiment scores in June and August 2013 were much higher than those in other months. The blue line in Fig. 2 shows the monthly positive-tone sentiment score and the red line shows the monthly negative-tone sentiment score.

In Fig. 2, the monthly sentiment scores peaked in March and April 2011, March and November 2012, June, August, and October 2013, March and December 2014, September 2016, and July and October 2017. On those peak months, more articles about nuclear energy were posted than in the other months because major incidents or issues about nuclear energy occurred, as summarized in Table 2.

Looking back on Fig. 2, the monthly negative-tone score exceeded the monthly positive-tone score in June 2013, which meant that more negative-tone articles had been issued than positive-tone articles in that month. This was unexpected because the number of positive-tone articles exceeded that of the negative-tone articles even in March 2011 when the Fukushima NPP accident occurred. As shown in Table 2, a scandal over fake safety certifications for NPP parts was in the headlines in June 2013. This meant that the scandal of fake certifications had a more negative influence on the Korean’s perception of nuclear energy. They seemed to consider the fake certification scandal a more serious threat to their safety and property than the NPP accident in Japan. We scrutinized the words in Table 1 and the nuclear-related incidents or issues in Fig. 2 and Table 2 and found that most of the words in Table 2 were related to nuclear-related incidents or issues.

3.3. Comments sentiment analysis

The total number of articles posted in those 12 months listed in Table 2 was 47,020; of them, 4,526 had comments. We filtered out duplicate articles and ruled out comments erased by comment posters themselves or an administrator from further consideration. Like monthly article sentiment analysis, we made a monthly sentiment analysis on the article comments. Instead of a separate analysis of just the comments, we considered the articles along with their comments. Most of the positive-tone comments advocated nuclear power or supported the continuous utilization of nuclear energy to secure national energy security. Most of the negative-tone comments advocated our country increasing its proportion of renewable energy power generation and decreasing its dependence on nuclear energy to reduce fears of radiation from NPP. Fig. 3 contains scatterplots that show the two sentiment scores of the articles and their comments across those 12 months.

As shown in Fig. 3, comment sentiment scores were mostly located in the negative domain for all months, regardless of the article tone. This means that most comments were negative-tone. The highest numbers of comments were posted in July and October 2017, totaling 1,047 and 844, respectively. The nuclear energy-related issues of the time were the permanent shutdown of Kori unit 1 and public deliberation on whether construction of Shinkori units 5 and 6 would resume. Even in these two months, the number of negative-tone comments exceeded those with a positive-tone.

4. Discussion

This paper scrutinized the temporal change in public opinions on nuclear energy by using big data analysis of the tone of argument associated with nuclear-related articles and their comments to them posted on a portal site in January 2010–December 2019. In addition, we identified nuclear incidents or issues addressed in many of the negative-tone articles and comments.

The sentiment analysis showed that, while the total number of positive-tone articles exceeded that of the negative-tone over the entire analysis period, the number of negative-tone articles surged in a particular month when critical incidents such as the scandal over fake safety certifications for NPP parts happened. This meant that Korean society tended to be very sensitive to incidents that people consider more direct threats to their safety, which degraded people’s trust in nuclear power.

We examined the tendency of public opinions on major nuclear incidents or issues by analyzing the comments added to articles posted in the 12 months when the monthly sentiment scores peaked. The analysis results showed that sentiment scores in the comments were mostly in the negative domain for all 12 months, regardless of the article tone. This means that most comments were negative-tone. Looking across all the scatter plots, we can see that many points are concentrated on the borderline between the positive and the negative sentiment score domain. This means that most people are neutral to nuclear energy in ordinary times but would take a position of either positive or negative depending on the issue’s characteristics.

We found that the number of comments per nuclear-related article was much smaller than that of the comments per article in other fields. Hence, we cannot conclude that a small number of comments per article posted on the NAVER site represent the views of the entire populace. To collect more extensive opinions from all walks of life, we believe that it is necessary to expand the scope for analysis to the comments or mentions posted on other social networks.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (Ministry of Science and ICT). (No. 2020M2D2A2062436).

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.net.2020.07.031.

References

[1] K.B. Wright, Researching internet-based populations: advantages and disadvantages of online survey Research, online questionnaire authoring software packages, and web survey services, J. Comput. Mediat. Commun. 10 (2017), https://doi.org/10.1111/j.1083-6101.2005.tb00259.x.
[2] N.M. Shelke, S. Deshpande, V. Thakre, Survey of techniques for opinion mining, Int. J. Comput. Appl. 57 (2012) 30–35.
[3] V.D. Pavaloaia, E.M. Teodor, D. Fotache, M. Danileţ, Opinion mining on social media data: sentiment analysis of user preferences, Sustainability (Switzerland) 11 (2019), https://doi.org/10.3390/su11164459.
[4] A. Karahoca, D. Karahoca, E. Evirgen, Sentiment analysis of Turkish tweets by data mining methods, Int. J. Mech. Eng. Technol. 10 (2019) 915–925.
[5] S. Seo, D. Seo, M. Jang, J. Jeong, P. Kang, Unusual customer response identification and visualization based on text mining and anomaly detection, Expert Syst. Appl. 144 (2020), https://doi.org/10.1016/j.eswa.2019.113111.
[6] E. Park, Positive or negative? Public perceptions of nuclear energy in South Korea: evidence from big data, Nucl. Eng. Technol. 51 (2019) 626–630, https://doi.org/10.1016/j.net.2018.10.025.
[7] S. Roh, Big data analysis of public acceptance of nuclear power in Korea, Nucl. Eng. Technol. 49 (2017) 850–854, https://doi.org/10.1016/j.net.2016.12.015.
[8] C. Park, T. Yong, Prospect of Korean nuclear policy change through text mining, in: International Scientific Conference Environmental and Climate Technologies, Riga, Latvia, May 10–12, 2017.
[9] Y. Jang, E. Park, Social acceptance of nuclear power plants in Korea: the role of public perceptions following the Fukushima accident, Renew. Sustain. Energy Rev. 128 (2020), https://doi.org/10.1016/j.rser.2020.109894.
[10] P.K. Singh, M. Shahid Husain, Methodological study of opinion mining and sentiment analysis techniques, Int. J. Soft Comput. 5 (2014) 11–21, https://doi.org/10.5121/ijsc.2014.5102.