Comparing Social Media and News Articles on Climate Change: Different Viewpoints Revealed

Kang Nyeon Lee[1], Haein Lee[2], Jang Hyun Kim[3], Youngsang Kim[4] and Seon Hong Lee[2*]
[1] Department of Interaction Science, Sungkyunkwan University, Seoul, 03063, Korea
[e-mail: knlee999@naver.com]
[2] Department of Applied Artificial Intelligence/Department of Human-Artificial Intelligence Interaction, Sungkyunkwan University, Seoul, 03063, Korea
[e-mail: lhi00034, dltjsghd12@g.skku.edu]
[3] Department of Interaction Science/Department of Human-Artificial Intelligence Interaction, Sungkyunkwan University, Seoul, 03063, Korea
[e-mail: alohakim@skku.edu]
[4] Department of Human-Artificial Intelligence Interaction, Sungkyunkwan University, Seoul, 03063, Korea
[e-mail: mynameis03@skku.edu]
[*]Corresponding author: Seon Hong Lee

Received June 16, 2023; revised September 15, 2023; accepted October 25, 2023; published November 30, 2023

Abstract

Climate change is a constant threat to human life, and it is important to understand the public perception of this issue. Previous studies examining climate change have been based on limited survey data. In this study, the authors used big data such as news articles and social media data, within which the authors selected specific keywords related to climate change. Using these natural language data, topic modeling was performed for discourse analysis regarding climate change based on various topics. In addition, before applying topic modeling, sentiment analysis was adjusted to discover the differences between discourses on climate change. Through this approach, discourses of positive and negative tendencies were classified. As a result, it was possible to identify the tendency of each document by extracting key words for the classified discourse. This study aims to prove that topic modeling is a useful methodology for exploring discourse on platforms with big data. Moreover, the reliability of the study was increased by performing topic modeling in consideration of objective indicators (i.e., coherence score, perplexity). Theoretically, based on the social amplification of risk framework (SARF), this study demonstrates that the diffusion of the agenda of climate change in public news media leads to personal anxiety and fear on social media.

Keywords: climate change, big data, natural language processing, sentiment analysis, topic modeling

http://doi.org/10.3837/tiis.2023.11.004 **ISSN : 1976-7277**

1. Introduction

Climate change is one of the most noteworthy global issues with serious consequences. The sixth report of the Intergovernmental Panel on Climate Change (IPCC), released in August 2021, stated that human activity can change the climate at an unprecedented rate [1]. According to the report, the global average temperature will likely rise by 1.5 ℃ by 2040, 10 years earlier than the projection of research conducted three years ago. Considering this challenge, researchers have warned of an increase in catastrophic events, such as heatwaves, droughts, floods, and wildfires. To address these problems, government agencies worldwide have proposed policies to implement action in public policy on climate change depending on public opinion [2]. However, previous approaches to measuring public opinion on climate change have been based on limited data such as surveys or have mostly used data from a single type of platform. Brulle et al. [3] used survey data of public perception of the climate change between 2002 and 2010, performing a time series analysis of factors including extreme weather events, public attention to scientific information, and media coverage. Mildenberger et al. [4] conducted data analysis of surveys from 2011 to 2015 to investigate the regional distribution and public perception of climate change. These surveys were conducted to examine the diversity of opinions on Canadian climate and energy at the regional level. Chen et al. [5] used Twitter data and Deep Neural Network (DNN) to predict public opinion of climate change. However, these studies are insufficient to represent public opinion because of the limited number of survey participants and type of data. Therefore, big data from news and social media are used to obtain public opinion [6,7].

In this study, news data from LexisNexis and text data from Reddit, a globally popular social media platform, were considered as an indicator of public awareness of climate change. Subsequently, text mining, a methodology of natural language processing, was used as a big data analysis technology to extract information and derive essential meaning from unstructured data [8,9]. Among the studies in machine learning and statistics, topic modeling automatically patterning words with a polynomial distribution has been used as a significant methodology [10]. Assuming that each document is a function of a latent variable, topic modeling can automatically find topics and focus on grouping text documents. Unstructured data, such as news, posts, emails, and reviews, can be used for topic modeling [11].

Therefore, this study explores the perspective of text mining related to climate change through topic modeling. Specifically, the authors analyze the data using topic modeling and perform discourse analysis on climate change based on various topics in the text. Before applying topic modeling, sentiment analysis was performed to discover differences in discourse according to emotions on climate change. The contribution of this study is the approach to global awareness of climate change in time series and differences according to the source of discourse.

2. Related Works

2.1 Climate Change Research Using Data Analysis

With the increase in climate crisis that threaten the human life, various studies related to climate change have been conducted. Hickman et al. [12] surveyed 10,000 children and adolescents on climate change and investigated their thoughts and feelings about the government’s response to climate change. The participants in this survey had negative thoughts about the government's response, which caused personal suffering and anxiety. Chen et al. [13] used Twitter data to build a model that can distinguish perceptions in text data regarding user attitudes to climate change. In this study, 2,000 datasets of climate change were labeled as denier and non-denier. Loureiro et al. [14] assessed public sentiment toward climate policies in the UK and Spain over six months in 2019 using natural language processing (NLP) tools and Twitter data. Through these data, it was possible to analyze the sentiments and preferences of the British and Spanish public regarding climate change policies. Stoddart et al. [15] investigated the discourse on climate change using Canadian news reports as data. This analysis found that media reports of the climate change focus on policymaking, management, and mitigation of issues. Treen et al. [16] collected posts containing keywords related to climate change from a subreddit on Reddit and used these data to implement polarity analysis for each community. This study confirmed that the form of discussion can differ depending on the platform structure. Furthermore, Fownes et al. [17] explained the advantages and limitations of discussing climate change on Twitter. Specifically, this discussion concerned the possibility that a wide range of discourse activities related to climate change could be transformed offline. Thus, discourse research on climate change has been conducted through surveys, news, and social networking platforms, and analyzed results of these approaches have proven useful in discourse analysis.

2.2 Topic Modeling and Sentiment Analysis

Topic modeling is an unsupervised learning and statistical modeling methodology for discovering the topic of a document set. This methodology is used to find important words in textual data [18]. In addition, sentiment analysis in discourse research is a model in the field of computational linguistics used to predict the emotion of the text [19]. Melton et al. [20] investigated the public perception of the COVID-19 vaccine using text data from Reddit. In this study, sentiment analysis was performed using the TextBlob library to simultaneously consider the emotions of the general public. Subsequently, using classified text data by TextBlob, insights into public opinion were presented through latent dirichlet allocation (LDA) topic modeling. Mishra et al. [21] used topic modeling to analyze user situations, sentiments, and opinions as expressed on social media services about the tourism industry affected by the COVID-19 pandemic. In this process, Twitter data were collected, and the Valence Aware Dictionary and sEntiment Reasoner (VADER) tool was used to calculate sentiment scores; the groups classified by VADER were clustered using LDA modeling. As a result, the clusters of industries affected by COVID-19 (e.g., hotels, medical, tourism) were important for analyzing information flow. Cirqueira et al. [22] noted that universities worldwide use the Facebook platform for public relations, implementing a combination of sentiment analysis and topic modeling to find a perception of the university by students and faculty using the platform. Consequently, this method was proven to be valuable for people management on social media. Furthermore, with the increasing number of airport service users, Kiliç et al. [23] investigated customer perceptions of various airport services. By applying topic modeling to passenger comments, topics for ten services were identified, and sentiments for each topic were extracted. These results suggested guidelines for improving existing services or introducing new service processes for managing airport services. In addition, topic modeling can facilitate analyzing important keywords for various topics, such as tourism and economy domains [24,25]. Based on these previous studies, topic modeling has demonstrated the ability to derive the potential meaning of the text used in various domains.

Table 1. Summary of research in climate change, topic modeling and sentiment analysis

Authors	Research topic	Method	Data source
Hickman et al. (2021) [12]	Climate anxiety	Descriptive statistics	Kantar
Chen et al. (2019) [13]	Climate change	Deep neural network	Twitter
Loureiro and Alló (2020) [14]	Sentiments related to climate change	Natural language processing tools	Twitter
Stoddart et al. (2016) [15]	Climate change	Textual analysis	Canadian national newspapers
Treen et al. (2022) [16]	Studies of climate discourse	Topic modeling	Reddit
Melton et al. (2021) [20]	COVID-19 pandemic	Sentiment analysis and topic modeling	Reddit
Mishra et al. (2021) [21]	Tourism during COVID-19 pandemic	Deep learning-based sentiment analysis and topic modeling	Twitter
Cirqueira et al. (2017) [22]	Relationship management in universities	Sentiment analysis and topic modeling	Facebook
Kiliç and Çadirci (2022) [23]	Airport service experience	Sentiment analysis and topic modeling	Skytrax
Kim et al. (2023) [24]	British Museum	Topic modeling	LexisNexis
Lee et al. (2023) [25]	ESG discourse analysis	Topic modeling	LexisNexis

3. Method

3.1 Data Collection

The authors collected 198,717 and 61,370 data on climate change from news data (i.e., LexisNexis) and social media data (i.e., Reddit), respectively, for the period January 1, 2018 to August 31, 2022. As climate change has a wide range of meanings, four queries ("climate change", "climate crisis", "heatwave", and "wildfire") including all these meanings were selected as keywords in Lexis [26–28]. Similarly, the subreddits r/climate, r/climatechange, and r/wildfire were used to gather review data.

3.2 Data Preprocessing

To achieve an improved representation of the data, the authors applied a lemmatization to news and social media data. Unnecessary words in each data point were removed with a consideration of stopwords in spaCy, an open source Natural Language Processing library. Subsequently, the authors divided the positive and negative tones of groups using Linguistic Inquiry and Word Count (LIWC)-2022, which is a dictionary-based software designed for natural language processing tasks [29]. The "tone_pos" and "tone_neg" features were selected from the results of LIWC-2022 for news and social media data to classify positive and negative groups. These features belong to the Affect category, which reflects the sentiment of text data as a result of LIWC-2022 [29]. Text data having a higher tone_pos score than tone_neg score was classified as part of the positive group, while text data fulfilling the opposite condition was classified as part of the negative group. In addition, in order to know the importance of positive and negative words in each data, the top 50 words can be extracted through TF-IDF and identified through word cloud (**Fig. 1** and **Fig. 2**).

Fig. 1. Word cloud of positive and negative data from LexisNexis

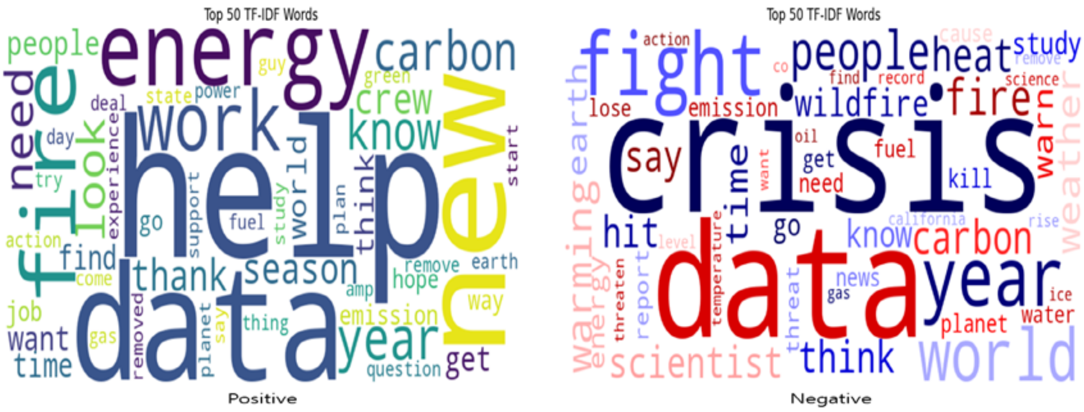

Fig. 2. Word cloud of positive and negative data from Reddit

3.3 Dirichlet Multinomial Regression (DMR)

With positive and negative group data, the authors used Dirichlet Multinomial Regression (DMR), which is an extension of existing LDA. LDA produces the distribution for each topic of the document using a word matrix and conditional probability. Because DMR can utilize specific features of documents with prior distribution, it is more comprehensively used than LDA [30]. Therefore, in this study, "years" was set as a variable to analyze the topic using a time series.

3.4 Coherence and Perplexity

Topic coherence is a measure of the number of semantically similar words within a topic [31]. Therefore, to associate similar words with each topic cluster, a higher coherence score is preferable. The authors have used a tomotopy library to calculate the coherence class. Although there are various methods to measure coherence, the authors have selected C_{uci} based on pointwise mutual information (PMI). C_{uci} computes the PMI of all word pairs in a sliding window and provides N top words [32]. The related equation is as follows:

$$C_{uci} = \frac{2}{N \cdot (N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} PMI(w_i, w_j) \tag{1}$$

$$PMI(w_i, w_j) = log \frac{P(w_i, w_j) + \epsilon}{P(w_i) \cdot P(w_j)} \tag{2}$$

The probability is calculated based on the number of word co-occurrences. ϵ is a smoothing count value to prevent the numerator from becoming zero.

Another important indicator, perplexity (PPL) is a value used in information theory to evaluate the predictive outcome of a natural language statistical model within a dataset. A lower perplexity value indicates a better probabilistic model [33]. The related equation is as follows:

$$Perplexity(D_{test}) = exp\left\{-\frac{\sum_{d=1}^{M} log p(w_d)}{\sum_{d=1}^{M} N_d}\right\} \tag{3}$$

While M is the number of test datasets (i.e., D_{test}), N_d is the total number of words. $p(w_d)$ is the probability that a word is allocated to a specific topic. To calculate the perplexity, an inherited member from the tomotopy library, "perplexity," was utilized. Consequently, in the range where a high coherence score and a low PPL score commonly overlap, the number of topics was determined.

4. Results

To analyze the difference between positive and negative opinions on climate change, the authors used LIWC-2022 to group news and social media data. As a result, four groups were created based on the LIWC-2022 features, and the values of coherence and perplexity for each group were calculated. Subsequently, the optimal number of topics could be determined using these values (**Figs. 3–6**). The values of coherence, perplexity, and the number of topics are shown in **Table 2**. The number of words per topic was set to 25 and words included in the list of stopwords were removed.

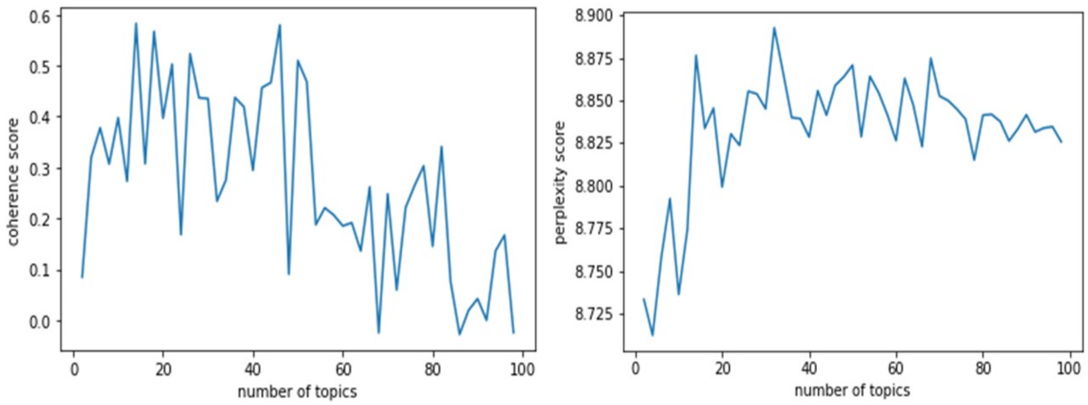

Fig. 3. Coherence and perplexity score on positive data from LexisNexis

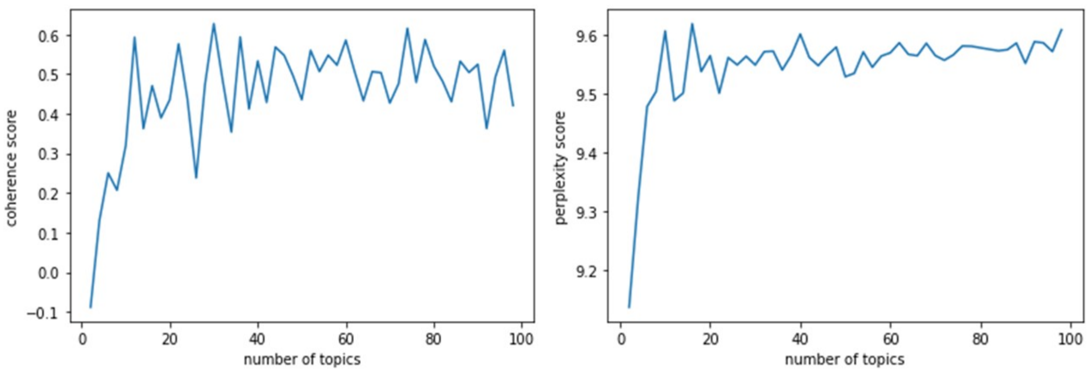

Fig. 4. Coherence and perplexity score on negative data from LexisNexis

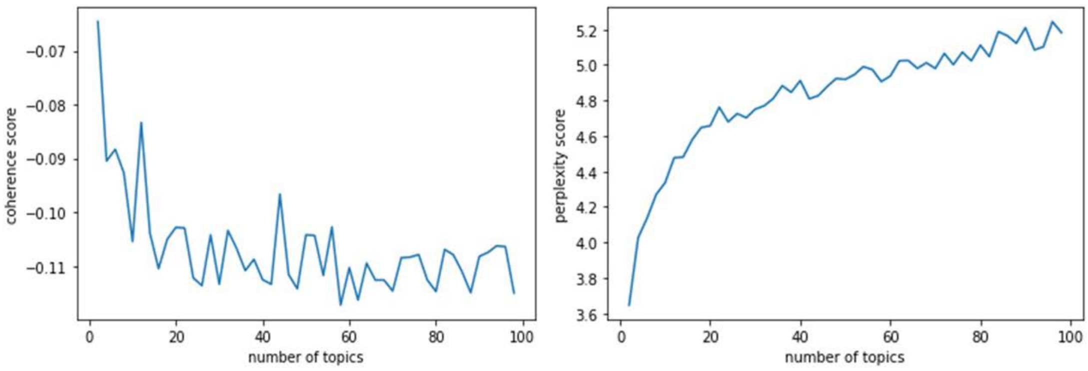

Fig. 5. Coherence and perplexity score on positive data from Reddit

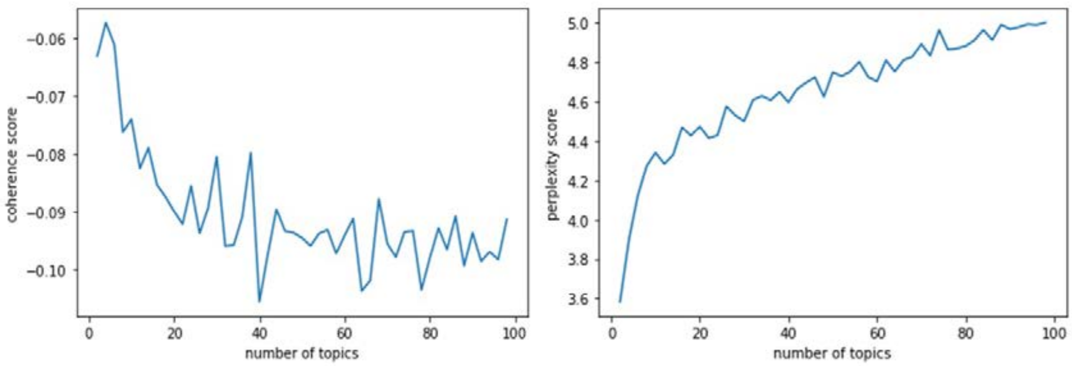

Fig. 6. Coherence and perplexity score on negative data from Reddit

Table 2. Optimal value of coherence score and perplexity

Group	Coherence Score	Perplexity	Topic Number
LexisNexis (Negative)	-0.10387420094779624	4.48108954	8
LexisNexis (Positive)	0.5830874479380691	8.87642067	8
Reddit (Negative)	-0.07397898455409299	4.34063737	6
Reddit (Positive)	-0.10544398879857367	4.26976188	6

4.1 DMR Result from LexisNexis

4.1.1 Positive Group

Among the topics in the positive group of LexisNexis, Topics 2 and 3 accounted for a higher proportion of topic importance than other topics from 2018 to 2022. Topic 2 focuses on changes in educational methods in response to climate change. Specifically, Eames [34] indicated that, although New Zealand, an island nation, knew its ecological vulnerability, there was a lack of education about climate change in its education system. Everth et al. [35] found that the New Zealand government recognized the seriousness of climate change by declaring an emergency regarding climate change and passing carbon neutrality legislation. In addition, they argued that it is important to develop the educational capabilities of school leaders and teachers to teach children about the risks of climate change [35]. Furthermore, Stevenson et al. [36] suggested that educators and students can take a critical and creative approach to develop a curriculum for climate change. In Australia, a survey on climate change found that only one in three students understood the meaning of the greenhouse effect and climate change. Consequently, many studies continue to point out the problem of existing education system.

Topic 3 is a summary of climate change reflected in various arts. The recent rapid increase in the consumption of online content on climate change indicates that videos with such element as music can encourage public participation in climate change awareness [37]. Banchero et al. [38] confirmed that movie directors play a representative role in delivering messages related to climate change to viewers. These results suggest that various media can affect public awareness of climate change.

Topic 6 contains topics related to trees and forests, occupying a rate of approximately 10% every year. Thuiller et al. [39] proposed that climate change affects tree vitality. These results confirmed that evolutionary history has been weakened and phylogeny has led to the homogenization of species. Additionally, models for predicting biodiversity and species threats according to climate change have been presented [40]. Consequently, they suggested that diversity of biological evolution changed in climate change. Therefore, from forests to trees and biodiversity, climate change can affect everyday life.

Topic 7 is summarized as Energy and Technology. Energy generation worldwide is largely dependent on fossil fuels, which have a significant impact on carbon emissions [41]. Therefore, renewable energy is essential for environmental sustainability and thereby addressing global climate change [42]. Notably, attention to energy system innovation has increased in recent years through the reduction of energy technology costs. In addition, some companies that have tried to mitigate the effects of climate change have reduced energy consumption and carbon emissions by adopting changes in supply methodologies, such as the sharing of energy performance contracts [43]. The distribution of the topics is shown in **Fig. 7** and **Table 3**.

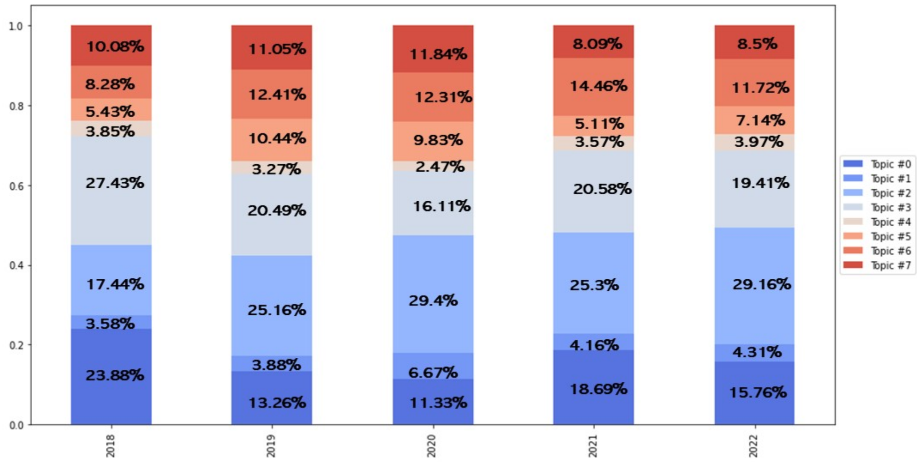

Fig. 7. DMR result on positive data from LexisNexis

Table 3. Topics related to positive perceptions of climate change from LexisNexis

Number	Topic	Words
0	LifeStyle	wine, food, heat, buy, meat, fashion, summer, dress, christmas, wear, eat, bottle, skin, restaurant, vegan, sleep, grape, drink, fan, waste
1	United Kingdom and Brexit Change	service, brexit, johnson, uk, labour, ireland, scotland, eu, community, say, boris, prince, northern, corbyn, london, deal, queen, royal, harry, tory
2	Education of Climate Crisis	people, change, zealand, student, school, work, new, australia, say, think, world, country, china, need, university, science, health, government, ardern, thing
3	Climate and Art	book, film, music, love, festival, art, write, story, life, award, woman, play, song, feel, star, child, york, artist, read, know
4	Affection of Sports by Climate Change	time, block, bst, publish, update, timeupdate, england, game, australia, temperature, player, team, ball, weather, win, sports, play, test, day, case
5	Leaders of Each Country	party, mr, labor, election, biden, trump, vote, candidate, government, campaign, voter, say, morrison, seat, president, ms, minister, leader, greens, democrats
6	Tree and Forest	tree, city, water, plant, fire, garden, new, say, park, island, travel, land, sea, forest, council, specie, river, area, farmer, road
7	Energy and Technology	energy, company, emission, business, carbon, singapore, gas, cost, investment, market, cent, fund, technology, bank, coal, tax, oil, economy, price, risk

4.1.2 Negative Group

In the negative group of LexisNexis, the ratio of almost all topics was uniform. Among these, topic 6 accounted for a large proportion. This topic was summarized with content related to the UK and Australian governments. Australia has experienced extreme climate change, with effects such as wildfires and floods, due to global warming. To prepare for such disasters, studies have been conducted on governance that allows for dynamic interaction with the government [44]. Additionally, Crowley [45] reviewed Australia's climate change policy from 2015 to 2020 and noted that the conservative government of Prime Minister Morrison placed political constraints on efforts to lower the usage of fossil fuels. Frost et al. [46] confirmed that, in the UK, climate change was not considered in legislation related to biodiversity, and Romsdahl et al. [47] noted that the UK government did not prioritize critical issues such as climate change. To address climate change challenges, local governance practitioners must reshape their strategies. Similarly, challenges have been constantly raised in countries that have not established proper policies on climate change in the past few years. One of the topics that did not appear in the positive group was related to forest fires. Topic 4 describes the increasing number of wildfires, and Topic 7 is related to biodiversity. Australia experienced unprecedented wildfires that destroyed large forests between 2019 and 2020. In addition, drylands increased due to climate change have a significant impact on wildfires [48,49].

Topic 5 is summarized as war and policy. Zhang et al. [50] suggested that global war, population, and price cycles are correlated with climate change. Moreover, the ongoing Russia-Ukraine War has influenced economic and geographic factors [51]. Specifically, the war has directly threatened sustainable development, climate change, and conservation. Therefore, as the climate crisis has grown, concerns about the war were considered. Cartwright [52] discussed how former President Trump's environmental restrictions had been altered since Joe Biden was elected. Regulations related to climate change differ according to policy and government.

Topic 0 accounted for a relatively high proportion in 2020. This increase coincided with the outbreak of the pandemic caused by COVID-19. Climate change has reduced biodiversity, and global warming has changed the habitat of species. These alterations have shortened the evolutionary gap between animals and humans, resulting in global epidemics [53]. Based on the similarities between the global climate emergency and Covid-19, Manzanedo et al. [54] presented insights into managing global climate problems. Consequently, climate change can be a serious issue that spans national and social scope. From LexisNexis data, the authors discovered that the negative group mainly focused on the problems caused by climate change (such as, pandemic, wildfires, war) and the issue that has continued from the past (i.e., global warming, policy). The distribution of the topics is shown in **Fig. 8** and **Table 4**.

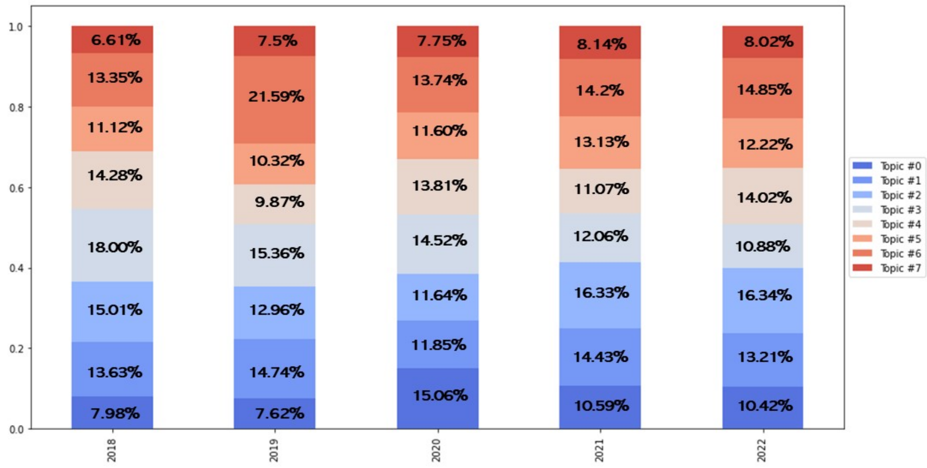

Fig. 8. DMR result on negative data from LexisNexis

Table 4. Topics related to negative perceptions of climate change from LexisNexis

Number	Topic	Words
0	Covid-19 and Climate Change	health, school, student, coronavirus, child, covid, pandemic, people, disease, care, vaccine, university, education, virus, space, worker, word, work, send, vaccination
1	Global Warming	food, change, warming, emission, pollution, carbon, waste, world, country, plastic, meat, agriculture, greenhouse, earth, environmental, global, research, farmer, environment, report
2	Industry	energy, gas, oil, company, power, coal, carbon, price, bank, industry, electric, fuel, investment, emission, vehicle, fund, business, utility, plant, market
3	Art	film, book, police, protest, story, woman, fashion, man, news, london, prince, life, people, art, music, think, feel, write, family
4	Increasing of Wildfires	fire, california, weather, heat, city, temperature, wildfire, disaster, flood, say, usa, home, storm, record, area, burn, county, rain, state
5	War and Policy	trump, mr, biden, china, president, candidate, trade, democrats, party, state, donald, states, russia, administration, united, joe, security, ukraine, war
6	United Kingdom and Australia	australia, government, minister, party, labor, morrison, johnson, policy, election, prime, mr, brexit, labour, uk, say, queensland, elections, leader, coalition, change
7	Wildlife and Conservation	water, canada, specie, tree, wildlife, trudeau, forest, conservation, land, sea, bird, reef, river, fish, animal, ocean, marine, fishing, alberta, ontario

4.2 DMR Result from Reddit

4.2.1 Positive Group

The topics in Reddit's positive group include "wildland Firefighter job", "Careers in wildland Fire," "election and green energy", and "global warming". Topic 0 showed a percentage increase between 2020 and 2022. The distribution is shown in **Fig. 9** and **Table 5**. Owing to climate change, firefighters' activities continue to increase [55]. In the case of Topic 2, words such as "biden," "energy," "vote," "green," and "oil" are included. With Biden's executive order related to climate change in 2021, topic 2 has a higher proportion than in other years [56]. In topic 4, the topic was named "Sea level rise" based on words such as "temperature," "ocean," "sea," "level," "rise," "warming," and "increase." The damage caused by sea-level rise has continued to increase. Dasgupta [57] proved that sea-level rise due to climate change will have an adverse effect by encroaching on the territories of developing countries. In addition, for scientific development related to sea-level rise, the concept and meaning are presented by declaring a unified term instead of ambiguous words [58].

Topic 5, carbon footprint, is an indicator that quantitatively indicates the impact of climate change on the production process. Various companies use the carbon footprint to induce consumers to purchase products and services with low environmental impacts [59,60].

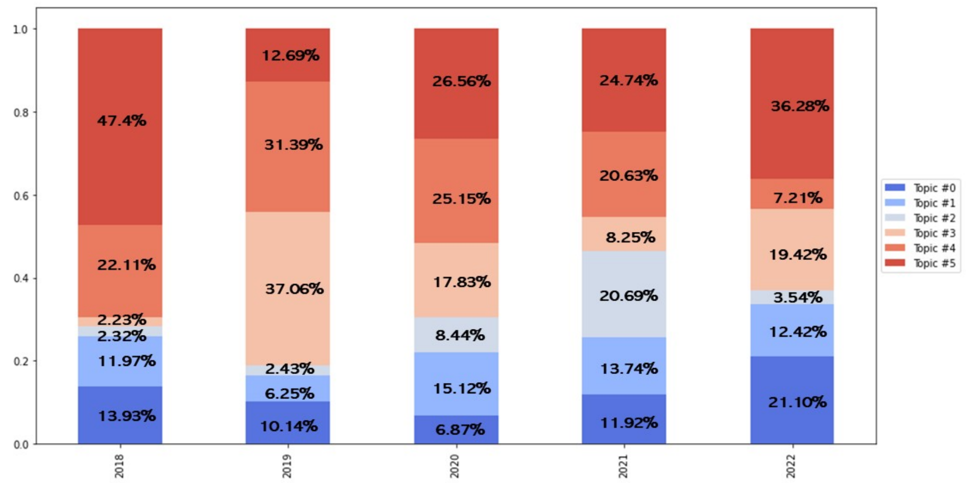

Fig. 9. DMR result on positive data from Reddit

Table 5. Topics related to positive perceptions of climate change from Reddit

Number	Topic	Words
0	Wildland Firefighter job	boot, season, work, fire, crew, get, job, pay, look, thank, know, want, question, time, guy, week, year, go, firefighter, offer, help, engine, wildland, experience
1	Careers in Wildland Fire	crew, fire, season, job, work, position, experience, hire, get, apply, gs, offer, year, thank, look, engine, know, wildland, want, forest, start, advice, resume, time
2	Election and green energy	new, energy, bill, data, change, deal, biden, support, company, green, oil, plan, action, democrats, gas, fund, emission, house, fuel, goal, clean, carbon, vote, panel

3	Interest in climate change	change, help, people, look, know, think, video, book, thank, survey, find, wildfire, post, want, science, feel, thing, fire, research, year, weather, idea, project, read
4	Sea level rise	change, new, temperature, co, earth, energy, water, ice, carbon, ocean, plant, data, rise, year, sea, level, emission, warming, power, world, gas, help, green, study, increase
5	Carbon footprint	carbon, change, energy, amp, help, tree, emission, data, fuel, co, need, world, gas, people, reduce, oil, action, plant, company, planet, way, use, footprint, power

4.2.2 Negative Group

The topics of Reddit's negative group are listed in **Table 6**. Based on the words of topic 0, the authors named the topic “California Wildfire.” These issues have been noted as important in previous studies. In California, wildfires destroy forests every year, causing casualties [61,62]. The structure of the meteorological system associated with wildfires in western America in July 2019 has been studied [61]. These researchers analyzed the data based on the link between atmospheric weather conditions and climate change. In addition, according to the results of these studies, climate change has caused more severe droughts and increased the frequency of wildfires [62].

Topic 1 was named “Attention on Carbon emission using social media.” Treen et al. [16] analyzed Reddit data that discuss climate change. As a result, social media discourse has proven to play an important role in online climate discussions. Uldam et al. [63] explored the potential of YouTube as a communication space for discussion and confirmed that video media is a forum for public debate. In particular, social media has been a place for several groups to discuss climate change qualitatively and raise awareness.

Topic 2, “UN climate change,” can be an important trend in the period 2021–2022. UN negotiations related to climate change are essential for implementing fundamental changes in global economics [64]. This topic has accounted for the largest proportion in 2022 (**Fig. 10**). Topic 5 was summarized as “Concern for climate change.” The number of people who feel anxious about the threat of climate change continues to increase over time [65].

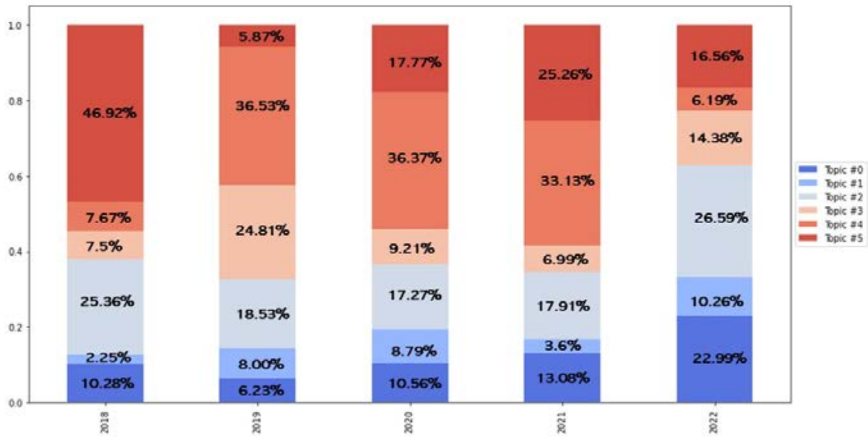

Fig. 10. DMR result on negative data from Reddit

Table 6. Topics related to negative perceptions of climate change from Reddit

Number	Topic	Words
0	California Wildfire	fire, wildfire, season, forest, california, change, crew, fight, work, get, data, job, know, firefighter, crisis, year, anxiety, state, go, people, start, service, burn, day, gt
1	Attention on Carbon emission using social media	amp, youtube, watch, carbon, gt, gas, change, type, co, emission, video, company, national, pay, energy, medium, oil, reddit, gov, earth, be, source, car, greenhouse
2	UN climate change	energy, change, coal, power, fight, fuel, crisis, data, plant, world, warn, report, carbon, fossil, threaten, policy, plan, un, news, water, oil, the, emission, heat
3	Global warming	ice, temperature, year, co, ocean, heat, change, weather, sea, water, earth, level, warming, increase, record, arctic, rise, melt, cause, loss, study, atmosphere, event, hit, planet
4	Academic interest	change, people, crisis, amazon, scientist, data, science, want, co, know, work, find, get, think, year, argument, carbon, go, level, question, fire, need, come, thing, warming
5	Concern for climate change	world, think, feel, data, planet, go, say, know, child, fight, want, life, threat, need, action, change, people, crisis, year, earth, stop, paper, be, emission, future

5. Conclusion

The climate change crisis has increased worldwide as a challenge that requires the efforts of countries and industries [65]. Considering this issue, many researchers have analyzed discourse to investigate changes in perceptions related to climate change. As various technologies that process big data in discourse have been used, it is important to analyze them and find new meanings. Among these technologies, text mining, which analyzes unstructured data, is focused on an important method for discovering latent meanings in discourse. Similarly, sentiment analysis and unsupervised learning research based on news and social media data have been recognized as significant methods for investigating public perception.

In this study, to analyze the public perception of the climate change issue, the authors collected 198,717 news data from LexisNexis' major media worldwide between 2018 and 2022, using keywords such as “climate change,” “climate crisis,” “heatwave,” and “wildfire.” In addition, the authors collected data from the subreddits r/climate, r/climatechange, and r/wildfire from the social media website Reddit, where people can discuss specific topics. Subsequently, a dictionary-based sentiment analysis, LIWC-2022, was performed to classify positive and negative data. To explore public perception over time, the authors utilized the DMR with metadata of “year.” The number of appropriate topics was determined based on coherence and perplexity. The results of the positive group on news data focused on systems that change perceptions of climate change. Countries sensitive to climate change, such as Australia and New Zealand, continue to have an interest in the education curriculum on climate change. In addition, the elements of climate change reflected in the various arts are constantly receiving attention. In the case of the negative group of news data, all topics showed an almost uniform distribution over the years. Among the topics, problems such as increased wildfires due to climate change, which were not in the positive group, emerged. Moreover, global war and climate change policies, according to the leaders of each country, were also considered as major implications. Owing to the nature of news media, information was expressed from a public point of view. Therefore, climate change can be considered a policy agenda.

Second, the results of the social media data also showed that the seriousness of climate change is continuously recognized. However, social media tends to relate personal life to climate change. In the positive group, topics such as occupation (e.g., firefighters) emerged. Through the topic of carbon footprint, the authors focus on the fact that various companies encourage consumers to purchase environmentally friendly services. In addition, there are topics of concern about the continued damage caused by sea level rise. The negative group clearly showed that the discussion of United Nations negotiations on climate change is important. This viewpoint can be interpreted as a lot of public attention being paid to discussions about the climate crisis by the leaders of each country. In addition, anxiety regarding the threat of climate change has increased over time.

The practical implications of this study are as follows: First, coherence and perplexity are considered when determining the number of topics. Previous studies utilized either value or determined the number of topics by multi-step verification [67–69]. To reduce the bias in these methodologies, the authors selected the optimal values for the two variables. As a result of applying the topic number obtained by considering these variables, there were few duplicated words between the topics. Additionally, news and social media data were used to examine public opinion. The collected data showed different characteristics depending on the platform. Therefore, it would be useful to investigate perceptions using multiple platforms.

The theoretical implications of this study are as follows: Public perception of climate change can be interpreted by the Social Amplification of Risk Framework (SARF), in which the media interacts with social, psychological, and institutional processes in ways that can amplify or undermine the public response [70]. In other words, the public receiving information with a danger signal is affected by psychological, social, and cultural dimensions. Based on these influences, an individual's perception of risk is expanded or reduced [71]. This means that the social risks of climate change increased in the news media and were amplified as a public agenda. This can describe the personal anxiety and fear of social media as secondary effects.

The limitations of this study are as follows. First, it solely focused on major news media worldwide as designated by LexisNexis. Future studies should consider the media of each country to understand the specific topic of climate change in each country. Second, only English data were used. Therefore, a study that considers data in various languages is proposed.

Acknowledgements

This study was supported by a National Research Foundation of Korea (NRF) (http://nrf.re.kr/eng/index) grant funded by the Korean government (RS-2023-00208278).
We appreciate Editage (www.editage.co.kr) for their English editing service.

References

[1] S. L. Harper, A. Cunsolo and S. Clayton, "Including mental health as part of climate change impacts and adaptation assessment: A critical advance in IPCC AR6," *PLOS Climate*, vol. 1, no. 5, pp. e0000033, 2022. Article (CrossRef Link)

[2] R. L. Shwom, A. M. McCright, S. R. Brechin, R. E. Dunlap, S. T. Marquart-Pyatt et al., "Public opinion on climate change," *Climate Change and Society: Sociological Perspectives*, pp. 269–298, 2015.

[3] R. J. Brulle, J. Carmichael and J. C. Jenkins, "Shifting public opinion on climate change: an empirical assessment of factors influencing concern over climate change in the US, 2002–2010," *Climatic Change*, vol. 114, no. 2, pp. 169-188, 2012. Article (CrossRef Link)

[4] M. Mildenberger, P. Howe, E. Lachapelle, L. Stokes, J. Marlon et al., "The distribution of climate change public opinion in Canada," *PloS one*, vol. 11, no. 8, pp. e0159774, 2016. Article (CrossRef Link)
[5] X. Chen, L. Zou and B. Zhao, "Detecting climate change deniers on twitter using a deep neural network," in *Proc. of 2019 International Conference on Machine Learning and Computing (ICMLC)*, Zhuhai, China, pp. 204-210, 2019. Article (CrossRef Link)
[6] L. M. Smuts, "Social Networking Sites as a New Public Sphere: Facebook and its Potential to Facilitate Public Opinion as the Function of Public Discourse–A Case Study of the 2008 Obama Campaign," Ph.D. dissertation, University of Stellenbosch, Republic of South Africa, 2010.
[7] M. A. Beam, M. J. Hutchens and J. D. Hmielowski, "Clicking vs. sharing: The relationship between online news behaviors and political knowledge," *Computers in Human Behavior*, vol. 59, pp. 215-220, 2016. Article (CrossRef Link)
[8] M. Rajman and R. Besançon, "Text mining: natural language techniques and text mining applications," *Data mining and reverse engineering*, pp. 50-64, 1998. Article (CrossRef Link)
[9] L. Hickman, S. Thapa, L. Tay, M. Cao and P. Srinivasan, "Text preprocessing for text mining in organizational research: Review and recommendations," *Organizational Research Methods*, vol. 25, no. 1, pp. 114-146, 2022. Article (CrossRef Link)
[10] R. Alghamdi and K. Alfalqi, "A survey of topic modeling in text mining," *Int. J. Adv. Comput. Sci. Appl.(IJACSA)*, vol. 6, no. 1, 2015.
[11] D. Buenano-Fernandez, M. González, D. Gil and S. Luján-Mora, "Text mining of open-ended questions in self-assessment of university teachers: An LDA topic modeling approach," *IEEE Access*, vol. 8, pp. 35318-35330, 2020. Article (CrossRef Link)
[12] C. Hickman, E. Marks, P. Pihkala, S. Clayton, R. E. Lewandowski et al., "Climate anxiety in children and young people and their beliefs about government responses to climate change: a global survey," *The Lancet Planetary Health*, vol. 5, no. 12, pp. e863-e873, 2021. Article (CrossRef Link)
[13] X. Chen, L. Zou and B. Zhao, "Detecting climate change deniers on twitter using a deep neural network," in *Proc. of the 2019 11th International Conference on Machine Learning and Computing*, Zhuhai, China, pp. 204-210, 2019. Article (CrossRef Link)
[14] M. L. Loureiro and M. Alló, "Sensing climate change and energy issues: Sentiment and emotion analysis with social media in the UK and Spain," *Energy Policy*, vol. 143, pp. 111490, 2020. Article (CrossRef Link)
[15] M. C. Stoddart, R. Haluza-DeLay and D. B. Tindall, "Canadian news media coverage of climate change: historical trajectories, dominant frames, and international comparisons," *Society & Natural Resources*, vol. 29, no. 2, pp. 218-232, 2016. Article (CrossRef Link)
[16] K. Treen, H. Williams, S. O'Neill and T. G. Coan, "Discussion of Climate Change on Reddit: Polarized Discourse or Deliberative Debate?," *Environmental Communication*, vol. 16, no. 5, pp. 680-698, 2022. Article (CrossRef Link)
[17] J. R. Fownes, C. Yu and D. B. Margolin, "Twitter and climate change," *Sociology Compass*, vol. 12, no. 6, pp. e12587, 2018. Article (CrossRef Link)
[18] I. Vayansky and S. A. Kumar, "A review of topic modeling methods," *Information Systems*, vol. 94, pp. 101582, 2020. Article (CrossRef Link)
[19] Y. Mejova, "Sentiment analysis: An overview," University of Iowa, Computer Science Department, 2009.
[20] C. A. Melton, O. A. Olusanya, N. Ammar and A. Shaban-Nejad, "Public sentiment analysis and topic modeling regarding COVID-19 vaccines on the Reddit social media platform: A call to action for strengthening vaccine confidence," *Journal of Infection and Public Health*, vol. 14, no. 10, pp. 1505-1512, 2021. Article (CrossRef Link)
[21] R. K. Mishra, S. Urolagin, J. A. Jothi, A. S. Neogi and N. Nawaz, "Deep learning-based sentiment analysis and topic modeling on tourism during Covid-19 pandemic," *Frontiers in Computer Science*, vol. 3, pp. 775368, 2021. Article (CrossRef Link)

[22] D. Cirqueira, M. Pinheiro, T. Braga, A. Jacob Jr, O. Reinhold et al., "Improving relationship management in universities with sentiment analysis and topic modeling of social media channels: learnings from ufpa," in *Proc. of the International Conference on Web Intelligence*, Leipzig, Germany, pp. 998-1005, 2017. Article (CrossRef Link)
[23] S. Kiliç and T. O. Çadirci, "An evaluation of airport service experience: An identification of service improvement opportunities based on topic modeling and sentiment analysis," *Research in Transportation Business & Management*, vol. 43, pp. 100744, 2022. Article (CrossRef Link)
[24] M. Kim, H. Lee, S. H. Lee and J. H. Kim, "What rather than how: A DMR topic modeling analysis of news coverage on the British Museum," in *Proc. of 2023 IEEE International Conference on Big Data and Smart Computing (BigComp)*, JeJu, Korea, pp. 167-173, 2023. Article (CrossRef Link)
[25] H. Lee, S. H. Lee, K. R. Lee and J. H. Kim, "ESG Discourse Analysis Through BERTopic: Comparing News Articles and Academic Papers," *Computers, Materials & Continua*, vol. 75, no. 3, pp. 6023-6037, 2023. Article (CrossRef Link)
[26] R. Pierrehumbert, "There is no Plan B for dealing with the climate crisis," *Bulletin of the Atomic Scientists*, vol. 75, no. 5, pp. 215-221, 2019. Article (CrossRef Link)
[27] I. C. Cornes and B. Cook, "Localising climate change: heatwave responses in urban households," *Disaster Prevention and Management: An International Journal*, vol. 27, no, 2, pp. 159-174, 2018. Article (CrossRef Link)
[28] A. L. Westerling and B. P. Bryant, "Climate change and wildfire in California," *Climatic Change*, vol. 87, no. 1, pp. 231-249, 2008. Article (CrossRef Link)
[29] R. L. Boyd, A. Ashokkumar, S. Seraj and J. W. Pennebaker, *The development and psychometric properties of LIWC-22*, Austin, TX: University of Texas at Austin, 2022.
[30] D. Mimno and A. McCallum, "Topic models conditioned on arbitrary features with dirichlet-multinomial regression," *arXiv preprint arXiv:1206.3278*, 2012. Article (CrossRef Link)
[31] J. Chang, S. Gerrish, C. Wang, J. Boyd-Graber and D. Blei, "Reading tea leaves: How humans interpret topic models," *Advances in Neural Information Processing Systems*, vol. 22, 2009.
[32] M. Röder, A. Both and A. Hinneburg, "Exploring the space of topic coherence measures," in *Proc. of the Eighth ACM International Conference on Web Search and Data Mining*, Shanghai, China, pp. 399-408, 2015. Article (CrossRef Link)
[33] D. M. Blei, A. Y. Ng and M. I. Jordan, "Latent dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, no. Jan, pp. 993-1022, 2003.
[34] C. Eames, "Climate change education in New Zealand," *Curriculum Perspectives*, vol. 37, no. 1, pp. 99-102, 2017. Article (CrossRef Link)
[35] T. Everth, R. Bright, C. Morey, T. dePetris, S. Gaze et al., "Building capacity for climate-change education in Aotearoa New Zealand schools," *Journal Issue*, vol. 2, pp. 40-45, 2021.
[36] R. B. Stevenson, J. Nicholls and H. Whitehouse, "What is climate change education?," *Curriculum Perspectives*, vol. 37, no. 1, pp. 67-71, 2017. Article (CrossRef Link)
[37] J. Ettinger, P. Walton, J. Painter and T. DiBlasi, "Climate of hope or doom and gloom? Testing the climate change hope vs. fear communications debate through online videos," *Climatic Change*, vol. 164, no. 1, pp. 1-19, 2021. Article (CrossRef Link)
[38] P. Banchero, T. A. Rector and J. VanBallenberghe, "Best practices in climate change communication as applied to an informal education documentary about Alaska," *Journal of Geoscience Education*, vol. 69, no. 2, pp. 138-149, 2021. Article (CrossRef Link)
[39] W. Thuiller, S. Lavergne, C. Roquet, I. Boulangeat, B. Lafourcade et al., "Consequences of climate change on the tree of life in Europe," *Nature*, vol. 470, no. 7335, pp. 531-534, 2011. Article (CrossRef Link)
[40] J. A. Aguirre-Liguori, S. Ramírez-Barahona and B. S. Gaut, "The evolutionary genomics of species' responses to climate change," *Nature Ecology & Evolution*, vol. 5, no. 10, pp. 1350-1360, 2021. Article (CrossRef Link)
[41] M. S. Chowdhury, K. S. Rahman, V. Selvanathan, N. Nuthammachot, M. Suklueng et al., "Current trends and prospects of tidal energy technology," *Environment, Development and Sustainability*, vol. 23, no. 6, pp. 8179-8194, 2021. Article (CrossRef Link)

[42] Z. Ahmed, M. Cary, M. Shahbaz and X. V. Vo, "Asymmetric nexus between economic policy uncertainty, renewable energy technology budgets, and environmental sustainability: evidence from the United States," *Journal of Cleaner Production*, vol. 313, pp. 127723, 2021. Article (CrossRef Link)
[43] S. Xu, L. Fang and K. Govindan, "Energy performance contracting in a supply chain with financially asymmetric manufacturers under carbon tax regulation for climate change mitigation," *Omega*, vol. 106, pp. 102535, 2022. Article (CrossRef Link)
[44] L. Head, M. Adams, H. V. McGregor and S. Toole, "Climate change and Australia," *Wiley Interdisciplinary Reviews: Climate Change*, vol. 5, no. 2, pp. 175-197, 2014. Article (CrossRef Link)
[45] K. Crowley, "Fighting the future: The politics of climate policy failure in Australia (2015–2020)," *Wiley Interdisciplinary Reviews: Climate Change*, vol. 12, no. 5, pp. e725, 2021. Article (CrossRef Link)
[46] M. Frost, G. Bayliss-Brown, P. Buckley, M. Cox, S. R. Dye et al., "A review of climate change and the implementation of marine biodiversity legislation in the United Kingdom," *Aquatic Conservation: Marine and Freshwater Ecosystems*, vol. 26, no. 3, pp. 576-595, 2016. Article (CrossRef Link)
[47] R. J. Romsdahl, A. Kirilenko, R. S. Wood and A. Hultquist, "Assessing national discourse and local governance framing of climate change for adaptation in the United Kingdom," *Environmental Communication*, vol. 11, no. 4, pp. 515-536, 2017. Article (CrossRef Link)
[48] T. M. Ellis, D. M. Bowman, P. Jain, M. D. Flannigan and G. J. Williamson, "Global increase in wildfire risk due to climate-driven declines in fuel moisture," *Global Change Biology*, vol. 28, no. 4, pp. 1544-1559, 2022. Article (CrossRef Link)
[49] R. J. Keenan, C. J. Weston and L. Volkova, "Potential for forest thinning to reduce risk and increase resilience to wildfire in Australian temperate Eucalyptus forests," *Current Opinion in Environmental Science & Health*, vol. 23, pp. 100280, 2021. Article (CrossRef Link)
[50] D. D. Zhang, P. Brecke, H. F. Lee, Y. Q. He and J. Zhang, "Global climate change, war, and population decline in recent human history," *Proceedings of the National Academy of Sciences*, vol. 104, no. 49, pp. 19214-19219, 2007. Article (CrossRef Link)
[51] D. Rawtani, G. Gupta, N. Khatri, P. K. Rao and C. M. Hussain, "Environmental damages due to war in Ukraine: A perspective," *Science of The Total Environment*, vol. 850, pp. 157932, 2022. Article (CrossRef Link)
[52] E. D. Cartwright, "Climate and clean energy to take center stage," *Climate and Energy*, vol. 37, no. 6, pp. 13-14, 2021. Article (CrossRef Link)
[53] S. M. Zang, I. Benjenk, S. Breakey, E. Pusey-Reid and P. K. Nicholas, "The intersection of climate change with the era of COVID-19," *Public Health Nursing*, vol. 38, no. 2, pp. 321-335, 2021. Article (CrossRef Link)
[54] R. D. Manzanedo and P. Manning, "COVID-19: Lessons for the climate change emergency," *Science of the Total Environment*, vol. 742, pp. 140563, 2020. Article (CrossRef Link)
[55] R. Agoston, "The effects of global climate change on fire service Human resource view," *Procedia Engineering*, vol. 211, pp. 1-7, 2018. Article (CrossRef Link)
[56] D. South, S. Vangala and K. Hung, "The Biden Administration's Approach to Addressing Climate Change," *Climate and Energy*, vol. 37, no. 9, pp. 8-18, 2021. Article (CrossRef Link)
[57] S. Dasgupta, B. Laplante, C. Meisner, D. Wheeler and J. Yan, "The impact of sea level rise on developing countries: a comparative analysis," *Climatic Change*, vol. 93, no. 3, pp. 379-388, 2009. Article (CrossRef Link)
[58] J. M. Gregory, S. M. Griffies, C. W. Hughes, J. A. Lowe, J. A. Church et al., "Concepts and terminology for sea level: Mean, variability and change, both local and global," *Surveys in Geophysics*, vol. 40, no. 6, pp. 1251-1289, 2019. Article (CrossRef Link)
[59] K. H. Lee, "Integrating carbon footprint into supply chain management: the case of Hyundai Motor Company (HMC) in the automobile industry," *Journal of Cleaner Production*, vol. 19, no. 11, pp. 1216-1223, 2011. Article (CrossRef Link)

[60] Z. Qi, C. Gao, H. Na and Z. Ye, "Using forest area for carbon footprint analysis of typical steel enterprises in China," *Resources, Conservation and Recycling*, vol. 132, pp. 352-360, 2018. Article (CrossRef Link)

[61] S. J. Prichard, P. F. Hessburg, R. K. Hagmann, N. A. Povak, S. Z. Dobrowski et al., "Adapting western North American forests to climate change and wildfires: 10 common questions," *Ecological Applications*, vol. 31, no. 8, pp. e02433, 2021. Article (CrossRef Link)

[62] A. Terrier, M. P. Girardin, C. Périé, P. Legendre and Y. Bergeron, "Potential changes in forest composition could reduce impacts of climate change on boreal wildfires," *Ecological Applications*, vol. 23, no. 1, pp. 21-35, 2013. Article (CrossRef Link)

[63] J. Uldam and T. Askanius, "Online civic cultures? Debating climate change activism on YouTube," *International Journal of Communication*, vol.7, pp. 20, 2013.

[64] R. S. Dimitrov, "Inside UN climate change negotiations: The Copenhagen conference," *Review of Policy Research*, vol. 27, no. 6, pp. 795-821, 2010. Article (CrossRef Link)

[65] P. Bergquist and C. Warshaw, "Does global warming increase public concern about climate change?," *The Journal of Politics*, vol. 81, no. 2, pp. 686-691, 2019.

[66] S. B. Pralle, "Agenda-setting and climate change," *Environmental Politics*, vol. 18, no. 5, pp. 781-799, 2009. Article (CrossRef Link)

[67] B. Dahal, S. A. Kumar and Z. Li, "Topic modeling and sentiment analysis of global climate change tweets," *Social Network Analysis and Mining*, vol. 9, no. 1, pp. 1-20, 2019. Article (CrossRef Link)

[68] D. J. Lemay, C. Baek and T. Doleck, "Comparison of learning analytics and educational data mining: A topic modeling approach," *Computers and Education: Artificial Intelligence*, vol. 2, pp. 100016, 2021. Article (CrossRef Link)

[69] T. Ylä-Anttila, V. Eranti and A. Kukkonen, "Topic modeling for frame analysis: A study of media debates on climate change in India and USA," *Global Media and Communication*, vol. 18, no. 1, pp. 91-112, 2022. Article (CrossRef Link)

[70] R. E. Kasperson, O. Renn, P. Slovic, H. S. Brown, J. Emel et al., "The social amplification of risk: A conceptual framework," *Risk Analysis*, vol. 8, no. 2, pp. 177-187, 1988. Article (CrossRef Link)

[71] J. X. Kasperson, R. E. Kasperson, N. Pidgeon and P. Slovic, "The social amplification of risk: Assessing 15 years of research and theory," *Social Contours of Risk*, pp. 217-245, 2012.

Kang Nyeon Lee is PhD student at the Department of Interaction Science at SungKyunKwan University (S. Korea). His research focuses on Social Media and Public's perceptions related issues and policies by employing Semantic Network Analysis and Natural Language Processing.

Haein Lee is PhD student at Department of Applied Artificial Intelligence/Department of Human-Artificial Intelligence Interaction, Sungkyunkwan University. Her research focuses on Natural Language Processing, Deep Learning, and Machine Learning.

Jang Hyun Kim is a professor at Department of Human-Artificial Intelligence Interaction/Department of Interaction Science/Department of Applied Artificial Intelligence, Sungkyunkwan University. His research focuses on social/semantic data analysis, social media, and future media. He has authored over 50 papers in major journals such as Information Processing & Management, Telematics & Informatics, Cities, Government Information Quarterly, Technological Forecasting and Social Change, and Journal of Computer-Mediated Communication.

Youngsang Kim is a Postdoctoral Fellow at the Department of Human-Artificial Intelligence Interaction, Sungkyunkwan University, specializing in corpus linguistics, pragmatics, and human-computer interaction. Holding a Ph.D. degree from Yonsei Institute of Language and Information Studies, Yonsei University, Kim's research focuses on the role of language and communication patterns in human-computer interaction, with a specific emphasis on face threat in Korean communication and polite communication. He investigates how language usage and pragmatics impact the patterns and dynamics of effective and culturally appropriate human-computer interaction, particularly in the Korean context. Kim's expertise in corpus linguistics and pragmatics allows him to delve into the intricacies of language usage, politeness strategies, and the management of face threat in digital communication. His work, published in esteemed journals, sheds light on the complexities of language-mediated human-computer interaction and contributes to the development of culturally sensitive interactive systems.

Seon Hong Lee is PhD student at Department of Applied Artificial Intelligence/Department of Human-Artificial Intelligence Interaction, Sungkyunkwan University. His research focuses on Natural Language Processing, Deep Learning, and Machine Learning. Seon Hong Lee is the corresponding author and can be contacted via dltjsghd12@g.skku.edu.