

Public Psychological Response to The Government's Medical School Quota Expansion Policy and Doctors' Opposition: A Text Mining and Sentiment Analysis Approach

Jimin Park^{1*}, Sunah Hyun²

¹Seoul International School, Seoul, Republic of Korea, ²Jonathan M Tisch College of Civic Life, Tufts University, Medford, MA, USA

ABSTRACT

In November 2023, the South Korean government announced a policy to increase medical school enrollment by 2000 students each year, ending the freeze that had been in place since. In response, the medical community collectively expressed resistance, resorting to strikes and resignation. This study investigates whether the public shares factual information or primarily expresses personal and psychological reactions to social phenomena. This study employed text mining techniques, including Term Frequency-Inverse Document Frequency (TF-IDF), structured topic modeling, and sentiment word analysis. The analysis focused on 1,10,227 comments from 432 YouTube news articles related to medical school enrollment between November 10, 2023 and June 9, 2024. To mitigate temporal and political biases, the samples were selected by distinguishing between political orientations and time periods. This analysis yielded several key findings. First, the TF-IDF analysis highlighted keywords such as patient, president, strike, and responsibility. Second, structured topic modeling identified three main topics, with topics 1 and 2 focusing on rational discussions concerning government policies, doctors' perspectives, and the medical system. In contrast, topic 3 predominantly centered on the emotional public response to expanding medical school enrollment, capturing both supportive and opposing sentiments. Third, the sentiment word analysis underscored the notable imbalance between negative and positive words. Negative expressions (11,277) such as disgust, infuriation, worry, and threats, outnumbered positive words (6,827) by approximately 1.5 times. The most frequently used positive words were affection and gain. Based on the results, various implications related to the formation of public opinion on the issue, as well as people's linguistic and psychological responses, are discussed.

Keywords: School quota expansion; Text mining; structured topic modeling; Sentiment analysis

INTRODUCTION

In November 2023, the South Korean government announced a plan to increase the number of medical school enrollments. By February 2024, the government proposed a policy to increase the enrollment by 2,000 students annually. South Korea has frozen the number of medical school admissions since 2006, spanning 18 consecutive years. Previous administrations have attempted to increase the number of medical school slots; however, these efforts have been met with strong opposition from vested interest

groups, such as the Korean Medical Association. Consequently, politically sensitive governments have been reluctant to implement these changes. The number of doctors has remained static, with many physicians gravitating toward lucrative urban areas and concentrating on specialties such as dermatology and plastic surgery. This trend has led to a scarcity of essential medical services, such as pediatrics and obstetrics, in smaller cities, rural areas, and mountainous regions. To address this imbalance in the healthcare system, the current government has announced a policy to expand medical school admissions.

Correspondence to: Jimin Park, Seoul International School, Seoul, Republic of Korea, E-mail: jessicapark349@gmail.com

Received: 29-Aug-2024, Manuscript No. JPPT-24-33704; **Editor assigned:** 02-Sep-2024, PreQC No. JPPT-24-33704 (PQ); **Reviewed:** 16-Sep-2024, QC No. JPPT-24-33704; **Revised:** 23-Sep-2024, Manuscript No. JPPT-24-33704 (R); **Published:** 30-Sep-2024, DOI: 10.35841/2161-0487.24.14.491.

Citation: Park J, Hyun S (2024). Public Psychological Response to The Government's Medical School Quota Expansion Policy and Doctors' Opposition: A Text Mining and Sentiment Analysis Approach. J Psychol Psychother. 14:491.

Copyright: © 2024 Park J. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

However, the medical communities, perceiving this as a threat to their interests, have strongly opposed the policy, resulting in a standoff with neither side willing to concede. Medical policies directly affect public health, life, and welfare, leading to varied public opinions on both the government's policies and the medical community's response.

With the widespread use of mobiles and digital devices, news consumption has shifted from print media to digital platforms. In Korea, a significant amount of news is consumed through platforms such as YouTube, portal websites, and mobile apps. Readers often participate in forming public opinion by posting or reading comments on news articles and engaging in two-way communication. Psycholinguistics is a branch of psychology that studies the relationship between language use and psychological traits. Research suggests that people's vocabulary choices and linguistic features in writing or speaking reflect their psychological traits [1]. Therefore, analyzing the comments on the aforementioned platforms provides valuable insights into public psychology and sentiment [2].

Research that uses text mining to analyze online news comments can be categorized into two main areas [3]. The first category focuses on the characteristics of online comments and analyzes them statistically and quantitatively. Studies that quantitatively analyze the comment tree structure of online discussions [4] and compare the characteristics of social and general comments [5], fall into this category. The second category focuses on analyzing comments related to specific topics. These studies aimed to identify the main interests of users in news article comments on specific subjects. Park et al., analyzed comments on articles about the World Cup to identify readers' main interests and notable games or players [6]. This study belongs to the second category as it examines the psycholinguistic characteristics of comments and employs topic modeling techniques to identify comment topics.

Analyzing written text can provide insights into the writer's thoughts, personality, and emotional state. Using certain words and the subject of the writing can gauge the writer's interests, while using words expressing negative or positive emotions, such as "sad" or "happy" reflects their psychological state. People who indirectly express discomfort may be perceived as being less straightforward and considerate of others' feelings. This linguistic analysis approach is not only useful for analyzing the process of public opinion formation and reactions to social issues but also for understanding people's psychology [7].

Therefore, this study investigates the process of public opinion formation and psychological reactions to this social phenomenon by analyzing comments on online news. This study aims to expand the research domain of psycholinguistics through text analysis, which was not feasible before the advent of AI. Accordingly, this study addresses the following research questions by analyzing comments on YouTube:

What words play important roles within sentences based on the relationships of key terms used in the comments?

What themes are primarily addressed in the vast collection of comments?

What are the frequently used positive and negative emotion words?

To answer these questions, this study employed text mining techniques, including keyword analysis, Structured Topic Modeling (STM) Analysis, and sentiment word analysis.

MATERIALS AND METHODS

Data collection

YouTube news comments were chosen as the subject of the analysis to analyze public comments regarding the policy of expanding medical school quotas. From November 2023, when this policy became a prominent social issue, to June 2024, 110,227 comments from 431 YouTube news videos were selected for the analysis. To improve the completeness of the analysis, 3,020 duplicate comments were removed, resulting in a final dataset containing 107,207 comments. To prevent bias due to timing or political inclinations, news articles were selected to ensure a balanced distribution across months, with both pro-government and opposition-leaning news sources represented evenly.

Data cleaning

Data cleaning involves the process of removing unnecessary data and refining the keywords used in the actual analysis. To eliminate words irrelevant to the analysis, particles, conjunctions, and articles were removed first, followed by the exclusion of all parts of speech except nouns, verbs, and adjectives [8].

Analysis tools

Keyword analysis: Keyword analysis is useful for identifying the frequency of each word in large text datasets and for determining the topics to which these words relate [8]. This study extracted key keywords through frequency analysis and calculated Term Frequency-Inverse Document Frequency (TF-IDF) values to assess the importance of the keywords. TF-IDF is a weighted value obtained by multiplying the frequency of a keyword (TF) by the Inverse of the Document Frequency (IDF), which indicates the importance of a keyword within a specific document. The higher the frequency in a particular document and the fewer documents containing the keyword, the higher the TF-IDF value [9].

Structured Topic Modeling (STM): STM is a technique among topic modeling methods that automatically extracts hidden topics within documents. STM is an extension of latent dirichlet allocation, the most widely known topic modeling technique, which considers both document text and metadata (e.g., author, publication year, and document type) [8]. By utilizing metadata, STM can better explain topic distribution, enabling a more detailed analysis of the relationships between topics and documents using information such as author details and document types. In STM topic modeling, determining the appropriate number of topics is important. This study selected the number of topics based on topic interpretability and content

relevance and used the coherence score as an indicator to evaluate the quality of the extracted topics. The coherence score measures how the words in a topic are meaningfully connected, with high scores indicating that the modeling successfully extracted meaningful topics [10].

Sentiment word analysis: Sentiment analysis involves analyzing the emotions embedded in a text to understand the writer's sentiments and how people perceive topics positively or negatively [11]. For sentiment analysis, emotion words or sentiment dictionaries are used to identify emotional expressions and their intensities. This study employed the Kunsan National University (KNU) Korean emotion dictionary developed by the data intelligence lab at KNU, which contains 14,843 emotion words unaffected by specific domains.

RESULTS

Keyword frequency and TF-IDF analysis

The simple frequency analysis and TF-IDF results of the most

frequently mentioned keywords in comments related to medical school quota expansion are presented in Table 1. The most common keywords in the simple frequency analysis were doctor (22,454), government (10,494), increase (7,269), citizen (7,248), patient (4,778), hospital (4,142), policy (2,801), and selfishness (2,314). In contrast, the TF-IDF analysis highlighted patient (6.51), president (6.51), murderer (6.51), strike (6.51), responsibility (6.51), life (6.42), pain (6.42), and doctor (5.82).

STM analysis

To ensure the validity of the results, the number of topics was selected based on the coherence score, which is an objective measurement indicator. The analysis revealed that coherence was highest with three topics, so the number of topics was set to three. When examining the expected probabilities of individual comments appearing in each topic, topic 1 appeared with a probability of 20%, topic 2 with 40%, and topic 3 with 80%. Thus, topic 3 was overwhelmingly higher.

Table 1: Keyword analysis results.

Rank	Frequency analysis		TF-IDF	
	Keyword	Frequency	Keyword	TF-IDF
1	Doctor	22,454	Patient	6.51
2	Government	10,494	President	6.51
3	Increase	7,269	Murderer	6.51
4	Citizen	7,248	Strike	6.51
5	Patient	4,778	Responsibility	6.51
6	Hospital	4,142	Life	6.42
7	Policy	2,801	Pain	6.42
8	Selfishness	2,314	Doctor	5.82
9	Resident	1,887	Hospital	5.13
11	Doctor	1,608	Doctor	4.81
12	Opposition	1,531	License	4.57
13	License	1,450	Citizen	4.12
14	Strike	1,443	Policy	3.87
15	Self	1,372	Expansion	3.34
17	You	1,176	Opposition	3.26
18	Affection	1,148	Government	2.75
19	Solution	1,107	You	2.71
20	Responsibility	1,081	Themselves	2.44

By interpreting the key terms forming each topic, the following meanings were identified: Topic 1 included keywords such as “government”, “people”, “doctors”, and “policy”, focusing on government policies, public reactions, and doctors’ stances. Topic 2 comprised keywords such as “doctor”, “hospital”, “medical”, and “issues”, emphasizing problems within the hospital and healthcare system and the role of doctors. Topic 3 included keywords such as “expansion”, “thoughts”, “people”, and “real”, addressing diverse opinions, thoughts, and public reactions related to the medical school quota expansion.

Sentiment word analysis

The analysis of frequently mentioned sentiment words in the comments is presented in Table 2. Negative words (n=11,277)

outnumbered positive words (n=6,827) by approximately 1.5 times.

DISCUSSION

This study examined public opinion and psychological reactions to the sensitive social issue of expanding medical school quotas in South Korea by analyzing YouTube news comments. The analysis involved text mining techniques, including keyword, topic modeling, and sentiment word analyses.

According to the keyword analysis results, the simple keyword analysis revealed frequent mentions of key public figures and issues related to policies, such as “doctor”, “government”, “increase”, “citizen”, and “patient”. However, in the TF-IDF

Table 2: Sentiment words.

Rank	Negative		Positive	
	Word	Frequency	Word	Frequency
1	Patient	4,778	Affection	1,148
2	Anger	656	Best	593
3	Worry	610	Gain	497
4	Infuriated	563	Ability	375
5	Slave	461	Essential	366
6	Selfishness	443	Popular	352
7	Threat	437	Rational	344
8	Disgusted	355	Improvement	313
9	Insufficient	323	Respect	306
10	Give up	298	Gratitude	288
11	Wrong	273	Active	276
12	Burden	272	Acknowledgment	274
13	Greed	266	Important	258
14	Stubbornness	241	Development	238
15	Pain	235	Benefit	227
16	Nonsense	231	Normal	216
17	Loss	231	Income	206
18	Damage	210	Revenue	196
19	Lie	207	Privilege	182
20	Incompetent	187	Profit	172

analysis, words reflecting public resentment toward doctors, such as “patient”, “president”, “strike” and “life”, appeared prominently, indicating dissatisfaction with doctors’ collective actions, like strikes, that impede seamless patient care. Words criticizing doctors, such as “doctor” and “selfishness”, appeared consistently in both analyses. Overall, people seemed more frustrated by medical strikes and collective actions that hindered patient care rather than the government policy itself.

The topic modeling analysis results indicated that three topics exhibited the highest semantic coherence. Topic 1 addressed issues related to the government, citizens, and doctors, with approximately 20% of the comments expected to cover this subject. Topic 2 was closely associated with issues concerning the healthcare system, including doctors and hospitals, and was expected to be the focus of approximately 40% of the comments. Topic 3 primarily dealt with the diverse opinions and emotional reactions of the public regarding the expansion of medical school admissions, with an estimated 80% of the comments anticipated to relate to this topic. The significantly higher likelihood of topic 3 being discussed compared to other topics suggests that individuals are more inclined to express personal opinions and emotions about the issue rather than exchanging factual information through comments.

An analysis of emotional words revealed that negative words appeared approximately 1.5 times more frequently than positive words. Among the intriguing findings, the positive words “affect” and “gain” were particularly frequent. While these words are classified as positive emotional words, in the context of this issue, they might have been used with negative connotations: “Affection” possibly implying that doctors should return to the hospital out of care for their patients, and “gain” suggesting that doctors should not be preoccupied with their own benefits. Consequently, it appears that people perceive the medical community as a privileged interest group at the top of the social hierarchy. Their actions, such as collective resignations, are seen as leveraging patients’ lives for personal benefit, which is viewed as group selfishness and elicits strong psychological resistance.

Implications

Our findings indicate that people tend to prefer expressing personal preferences about fragmentary facts rather than relying on factual information when forming or articulating their opinions on social issues. Additionally, many individuals perceive medical professionals as a privileged group and tend to indiscriminately label their actions as collective selfishness, while simultaneously exhibiting another form of selfishness by hoping to strip these professionals of their privileges in society.

CONCLUSION

This study explored the formation of public opinion and psychological reactions to the highly debated social issue of expanding medical school quotas in South Korea by analyzing

comments left on YouTube news broadcasts. Keyword analysis, topic modeling analysis, and sentiment word analysis, among various text mining techniques, were used to identify important words, topics, and emotions related to the medical school quota expansion. The analysis, based on approximately 100,000 comments, surpassed the limitations of traditional content analysis, which allowed for a more rigorous examination.

LIMITATIONS

The following limitations must be considered while interpreting the results of this study. First, much of the meaning derived from the keywords revealed through data cleaning and text mining relies on the researcher’s subjective interpretation, necessitating caution in generalizing the findings. Follow-up research involving qualitative methods, such as interviews based on quantitative analysis, could help overcome these limitations. Second, the analysis focused on a single online medium, limiting representation across diverse age groups and demographics owing to each platform’s unique characteristics. Analyzing comments from various media sources, such as portal site news, could offer broader insights for analysis. Future research should aim for a more comprehensive analytical framework and expand the scope of the analysis to further generalize the findings.

REFERENCES

1. Jeong H. How to make use of verbal data in psychological research. *Korean J Psychol Gen.* 2014;33(3): 671-704.
2. Park YS, Kim HS. Analysis of comments emotion in online news using text mining. *Des Conver Study.* 2024;23(2): 41-65.
3. Ahn, H.J. Analysis of reactions to online news comments using word embedding and psycholinguistic variables. *J Korea Inform Technol Soc.* 2020;18(6):107-115.
4. Kim SH, Tak H, Cho HG. User characterization from replying comment structures in online discussion. *J Korea Contents Assoc.* 2018;18(11):135-145.
5. Kim SD, Yang SB. A comparative analysis between general comments and social comments on an online news site. *J Korea Contents Assoc.* 2015;15(4):391-406.
6. Park S, Won G, Lee S. Web news comment-based sentiment analysis of the South Korean national team members in the 2014 Brazil World Cup. *J Korean Soc Sport Manage.* 2015;20(2):13-28.
7. Kim Y, Kim K. Language style of the president of the Republic of Korea: Linguistic and psychological properties in presidents’ address. *Korean J Cogn Sci.* 2019;30(3):105-132.
8. Kim YW. Easy-to-learn R text mining. *Easypublishing.* 2021.
9. Heimerl F, Lohmann S, Lange S, Ertl T. Word cloud explorer: Text analytics based on word clouds. In *2014 47th Hawaii international conference on system sciences 2014*;1833-1842.
10. Syed S, Spruit M. Full-text or abstract? Examining topic coherence scores using latent dirichlet allocation. In *2017 IEEE International conference on data science and advanced analytics.* 2017;165-174.
11. Sohn SJ, Park MS, Park JE, Sohn JH. Korean emotion vocabulary: Extraction and categorization of feeling words. *Sci Emot Sensib.* 2012;15(1):105-120.