Research Trend Analysis of Nuclear Technology Using Topic Modeling

Gyu-Ri Bae
Quantum and Nuclear Engineering, Sejong University, 209 Neungdong-ro, Gwangjin-gu, Seoul, Republic of Korea

1. Introduction

Nuclear power plants(NPPs) have been used as one of the primary power sources so far, along with the industrial revolution, because they can produce considerable energy with low mass. Nuclear technology has a significant impact not only on energy production but also on medical, industrial sectors. Since the Fukushima NPP accident, public concerns about nuclear power are growing, so continuous research is encouraged for safe and sustainable use of NPPs.

Recently, data scientists are performing trend analysis using text mining techniques in various fields. Text mining objectively finds topics in many papers and searches for changes in the interest field. Research trends in specific fields or one year of research were analyzed so far, but there was no analysis of the research trends of overall nuclear technologies.

In this paper, we analyzed the research papers of the 'nuclear technology' field to get topics of study and find research trends. We have reviewed whether it is actively studied for the sustainable use of nuclear energy. Section 2 explains the topic modeling technique. Section 3 describes the research methods and results of this study, and Section 4 discusses the meaning and limitations of the study.

2. Method and Result

Text mining processes a lot of information objectively based on natural language processing technology. Topic modeling is a text mining technique that can extract meaningful information from text. The technique is useful for analyzing the subject of a document or predicting trends.

This section describes the topic modeling process: data collection, data preprocessing, vectorization, and Latent Dirichlet Allocation(LDA) model.

2.1 Collect Research Data

Research data is collected from the publisher-independent global citation database 'Web of Science (WOS)' to understand the world's overall research trends. WOS provides the WOS core collection that curated collection by evaluating the quality and impact of the journal. All journals in the WOS core collection are assigned to at least one topic category. We selected papers in the category of 'nuclear technology' as our research data. We chose the abstract and publication year for the analysis data. Since the abstract of the paper must be in the form of text data, we limited the years of published research data to 2001-2019. The total number of data is 160,389.

In the preprocessing step, we removed letters, symbols, and numbers except for English words and combined all upper case letters and lower case letters. We also removed stopwords, which are negligible words in the thesis, such as articles and conjunctions. To get better results, we have also added to stopwords that are frequently used but do not contribute to the topics. Examples of added stopwords are 'analysis', 'research', 'results', 'applied', and 'based'.

2.2 TF-IDF Vectorization

Natural language processing requires converting characters to vector format. Documents are transformed into a Document-Term Matrix(DTM), a matrix representing the frequency of each word appearing in multiple documents. We added TF-IDF weights to indicate the importance of words within a specific document. The scikit-learn library provides the ability to calculate TF-IDFs and generates a vocabulary list. The additional functions exclude words that appear too much or too rarely and provide the n-gram model. The n-gram model uses n linked words as a single word. For example, 'advanced reactor' and 'reactor core' share the word 'reactor' but are commonly used in different topics. These words can be classified effectively as n-grams. In vectorization step, we set n-grams range from 1 to 3 and excluded words that appear to be more than 99.9% or less than 0.5% of the entire documents.

2.3 LDA model

LDA was introduced in 2003 by David Blei, Andrew Ng, and Micheal Jordan [1]. It is an unsupervised machine learning technique for finding topics through words that appear in a set of unstructured documents. Topics found through the LDA are expressed as the probability of occurrence of words. Words with similar meanings are grouped[2]. A series of keywords represent each extracted topic, and we entered the topic names according to the keywords. We have to decide on the number of topics. For detailed topic classification, we set the number of topics to 70, which is bigger than the expected number and used default values on other LDA parameters.

2.4 Results

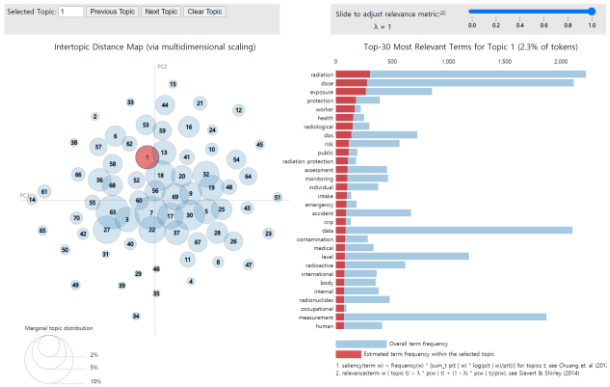

Fig. 1. Result of LDA

The LDA results can be visualized, as shown in Fig.1 using the pyLDAvis tool. The tool also provides the impact of each word on other topics. In the Fig.1 the topic can be called 'radiation protection' because it consists of ‘risk’, ‘health’, 'workers', 'danger', 'public', 'radiation protection', etc.

Table I: Categorized Topic Names

	Topic Name	Main keywords
1	Physics	'capture cross section', 'neutron diffusion', 'transport equation', 'neutron activation analysis'
2	Radiation Application	‘spectrometry’, ‘detector’, ‘target’, ‘scanner’, ‘ct’
3	Radiation Protection'	‘activity concentration’, ‘annual effective’, ‘indoor’, ‘concentration radon’
4	I&C	‘data acquisition’, ‘control’, ‘real time’, ‘signal’
5	Advanced Reactor	‘inertial confinement fusion’, ‘tokamak’, ‘plasma’, ‘high temperature gas’
6	Nuclear Fuel	‘fuel assembly’, ‘spent fuel’, ‘cycle’, ‘repository’
7	Structural & Material	‘strength’, ‘stress’, ‘mechanical’, bismuth’, ‘stainless steel’
8	Policy & Economics	‘energy consumption’, ‘renewable energy’, ‘efficiency’, ‘cost’
9	Chemical	‘LiCl’, ‘cyclic voltammetry’, ‘aqueous solution’
10	Safety	‘probabilistic safety assessment’, ‘Fukushima’, ‘severe accident’
11	Thermal-Hydraulics	‘heat transfer’, ‘fluid mechanics’, ‘critical heat flux’, ‘flow rate’

We grouped similar subjects and gave the same names to them. Table 1 shows the classified topic names and some keywords of each topic that influenced the determination of the topic name.

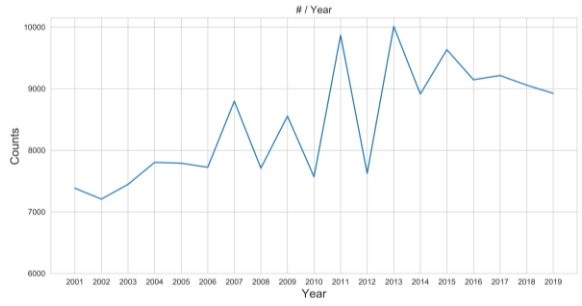

Fig. 2. Number of whole papers by year

Fig. 2 shows the total number of papers per year. The number of papers was the lowest at 6,964 in 2002 and the highest at 9,687 in 2013. The significant increase in the number of papers in 2011 was due to the impact of the Fukushima accident. Since 2013, the number of papers has increased by about 1,000 compare before.

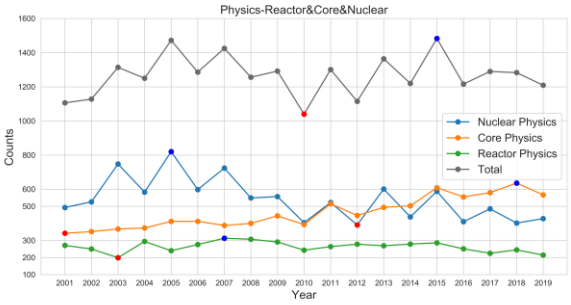

Fig. 3. # of papers in Topic 1 'Physics' by year

Fig. 3 shows the number of papers in Topic 1 'Physics', one of the big topics. These include nuclear reactor physics, nuclear physics, and core physics. Topic 1 contains the study of neutron behavior, atomic reactions, neutron calculations, and kinetic physics. The number of articles on all sub-topics is similar throughout the period.

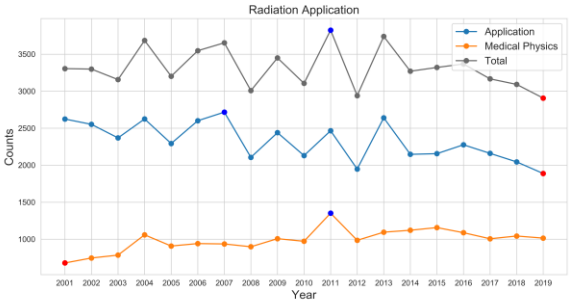

Fig. 4. # of papers in Topic 2 'Radiation Application' by year

The annual trend of Topic 2 'Radiation Application' is shown in Fig. 4. Topic 2 contains keywords for spectrometers, detectors, and other related equipment. Topic 2 covers radiotherapy equipment and medical scanners(CT and MRI). Since radiation technology is used in a variety of fields, Topic 2 covers at least 30% of all periods.

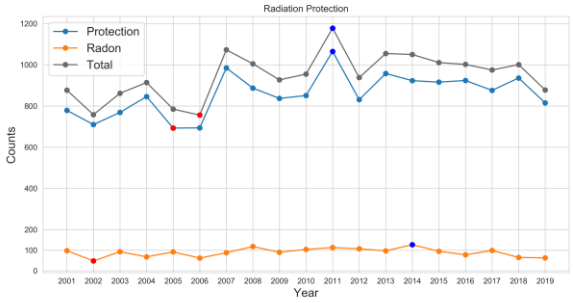

Fig. 5. # of papers in Topic 3 'Radiation Protection' by year

The variation of numbers in Topic 3 'Radiation Protection' is shown in Fig. 5. Topic 3 discusses the absorption dose of workers or the protection of people and the environment from radiation. Research of radon belongs to Topic 3 as well. Radon, a radioactive material that can be exposed in living spaces and a leading cause of lung cancer, so it is also crucial in the study of radiation protection.

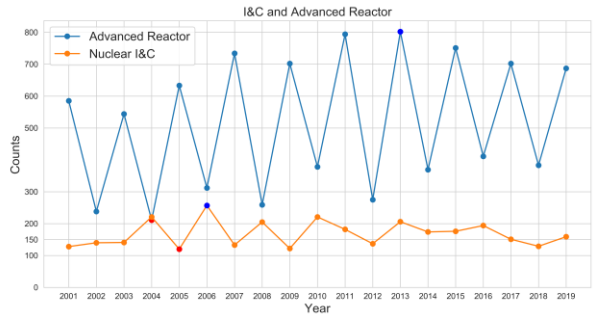

Fig. 6. # of papers in Topic 4 ‘I&C’ & Topic 5 ‘Advanced Reactor’ by year

Fig. 6 shows the change in the number of published papers that belong to Topic 4 ‘I&C’ and 5 ‘Advanced Reactor’.

Topic 4 is instrumentation and control(I&C). The I&C system develops and evaluates a monitoring system that receives the signal from NPPs. The system enables NPPs to operate normally and react effectively to accidents. Topic 4 has a steady trend.

The name of Topic 5 is ‘Advanced Reactor’. It consists of the technology and materials of fusion or gas-cooled reactor. We speculated the reason for the jagged graph shown in Fig. 6 for the impact of the IAEA Convergence Energy Conference(FEC). FEC is the largest fusion energy conference and is held biennially.

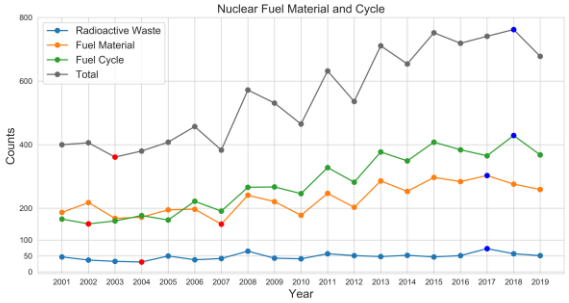

Fig. 7. # of papers in Topic 6 ‘Nuclear Fuel’ by year

The number of papers in Topic 6 ‘Nuclear Fuel’ is shown in Fig. 7. Topic 6 covers specific topics such as fuel cycle, radioactive waste, and fuel materials. Topic 6 seems to have almost doubled compared to the early 2000s. Reducing or disposing of radioactive waste and developing safer NPPs are essential challenges for researchers. Fuel is a related solution to many problems, so we understand the upward trends in topic 6.

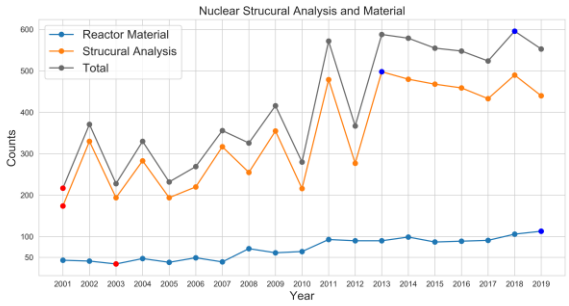

Fig. 8. # of papers in Topic 7 ‘Structural & Material’ by year

The annual trend for Topic 7 ‘Structure and Materials’ is shown in Fig. 8. Topic 7 consists of the words about structure analysis and material. After the 2011 Fukushima accident, the number of papers has increased. If natural disasters or accidents damage the structure of a nuclear power plant, severe accidents and radioactive leakage may occur. The increased research trend indicates efforts to maintain plant structure.

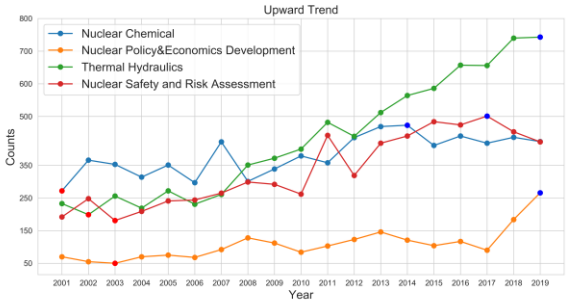

Fig. 9. # of papers in some upward trend topics by year

Fig. 9 shows the trend of Topic 8 ‘Policy & Economics’, Topic 9 ‘Chemical’, Topic 10 ‘Safety’, and

Topic 11 ‘Thermal Hydraulics’. All four of these topics are on the rise.

Topic 8 contains keywords that refer to discussions about the efficiency and economics of various energies. We named Topic 8 ‘Policy & Economy’. Although the percentage of Topic 8 is low, it shows interest in a variety of energy sources, including nuclear power. The trend seems to be to develop cleaner and more economical energy.

Topic 9 of ‘Chemistry’ studies the chemical use of nuclear reaction. Research on methods of recovering reusable elements from spent fuel is included in Topic 9. The research grew with nuclear fuel disposal research.

Topic 10 ‘Safety’ corresponds to safety and risk assessment. As can be expected, Topic 10 rose sharply in 2011 and continues the upward trend.

‘Thermal Hydraulics’ is the name of Topic 11. Nuclear thermal hydraulics is based on thermodynamics and heat transfer. Topic 11 covers frequent issues related to leakage and sealing coolant in NPPs and is essential for both normal and accident conditions. The number of Topic 11 papers increases the most over time.

3. Conclusions

In this paper, we analyzed the research trends of nuclear technology through text-mining techniques. We used LDA topic modeling analysis to extract topics from all nuclear technology papers published from 2001 to 2019. We classified the papers in the ‘nuclear technology’ category into 11 main topics and a few specific topics, then observed the changes in each topic over time.

Most subjects are being studied steadily. The research on the topic 'Nuclear Safety', ‘Thermal Hydraulics’, and ‘Structural & Material’ has especially grown further. This upward trend implies a growing interest in power plant accidents, keeping nuclear power plants safe and developing next-generation reactors.

We limited research data for abstract and WOS core collection. Nevertheless, many papers have been published in numerous journals. Title, the body of text, and keywords are having additional information. If expanding research material or scale, we can get the extensive and immediate attention of the research trend.

The nuclear technology already has safety and efficiency. However, more technical developments are still needed. The results of the topic modeling analysis indicate that researchers have been a steady interest in development.

Acknowledgment

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Science, ICT & Future Planning(NRF-2016R1A5A1013919)

REFERENCES

[1] David M.Blei, Andrew Y.Ng, Michael I.Jordan Latent Dirichlet Allocation, Journal of Machine Learning Research 3, p. 993-1022, 2003.
[2] Benjamin Bengfort, Rebecca Bilbro, Tony Ojeda, Applied Text Analysis with Python, O'Reilly Media, 2018.