Research Trend Analysis of Nuclear Technology Using Topic Modeling

Gyu-Ri Bae

Quantum and Nuclear Engineering, Sejong University

1. Introduction

- 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 overall nuclear technologies' research trends.

- Nuclear technology has a significant impact on energy production and the medical, industrial sectors. Since the Fukushima NPP accident, public concerns about nuclear power are growing, so continuous research is encouraged for the safe and sustainable use of NPPs.

- In this paper, we analyzed the 'nuclear technology' field's research papers to get topics of study and find research trends. We have reviewed whether it is actively studied for the sustainable use of nuclear energy.

2. Method
2.1 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. 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 paper's abstract 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.

- Text preprocessing steps need for transferring text to computer-readable data format for further processing.

  1. Removed letters, symbols, and numbers except for English words and combined all upper case letters and lower case letters.

  2. Remove stopwords, which are negligible words in the thesis, such as articles and conjunctions.

  3. Added to stopwords that are frequently used but do not contribute to the topics.

    Example of added stopwords are

    a) Words for explaining topic : 'result', 'paper', 'study', 'article', 'system', 'method', 'applied', 'shown', 'based',

    b) Words representing unit : 'bq', 'msv', 'kev', 'kg', 'mm'

- TF-IDF Vectorization

  (scikit-learn library)

Fig. 1. Encoding documents as vectors [1]

Fig. 2. TF-IDF encoding [1]
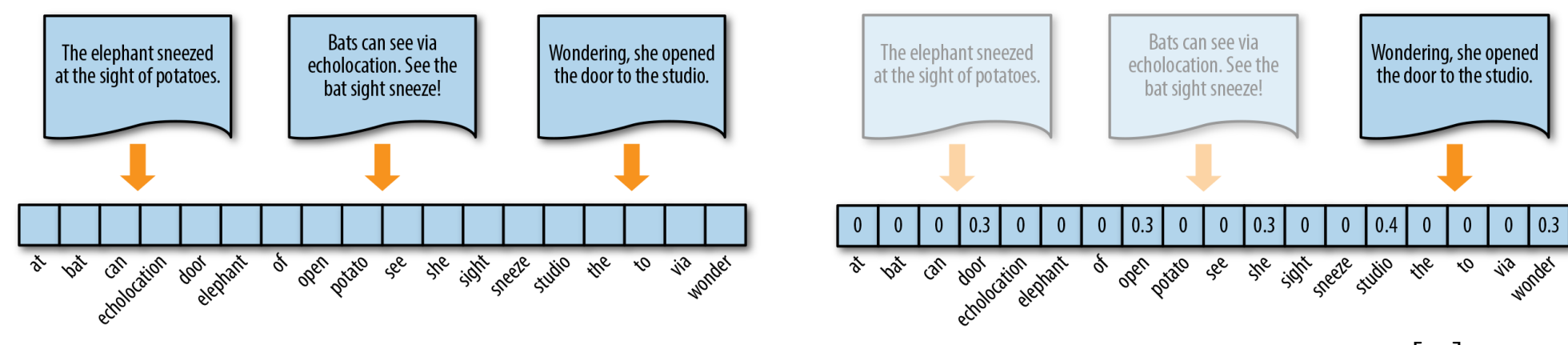
  Natural language processing requires converting characters to vector format. Documents are transformed into a Document-Term Matrix(DTM), a matrix representing each word's frequency appearing in multiple documents. We added TF-IDF weights to indicate the importance of words within a specific document.

  Excluded words that appear to be more than 99.9% or less than 0.5% of the entire documents and set n-grams range from 1 to 3. (n-gram model uses n linked words as a single word.)

2.2 LDA topic modeling

- 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.

- LDA was introduced in 2003 by David Blei, Andrew Ng, and Micheal Jordan [2]. 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[1]. 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. Fig.1 shows LDA results using 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.

Fig. 3. Result of LDA
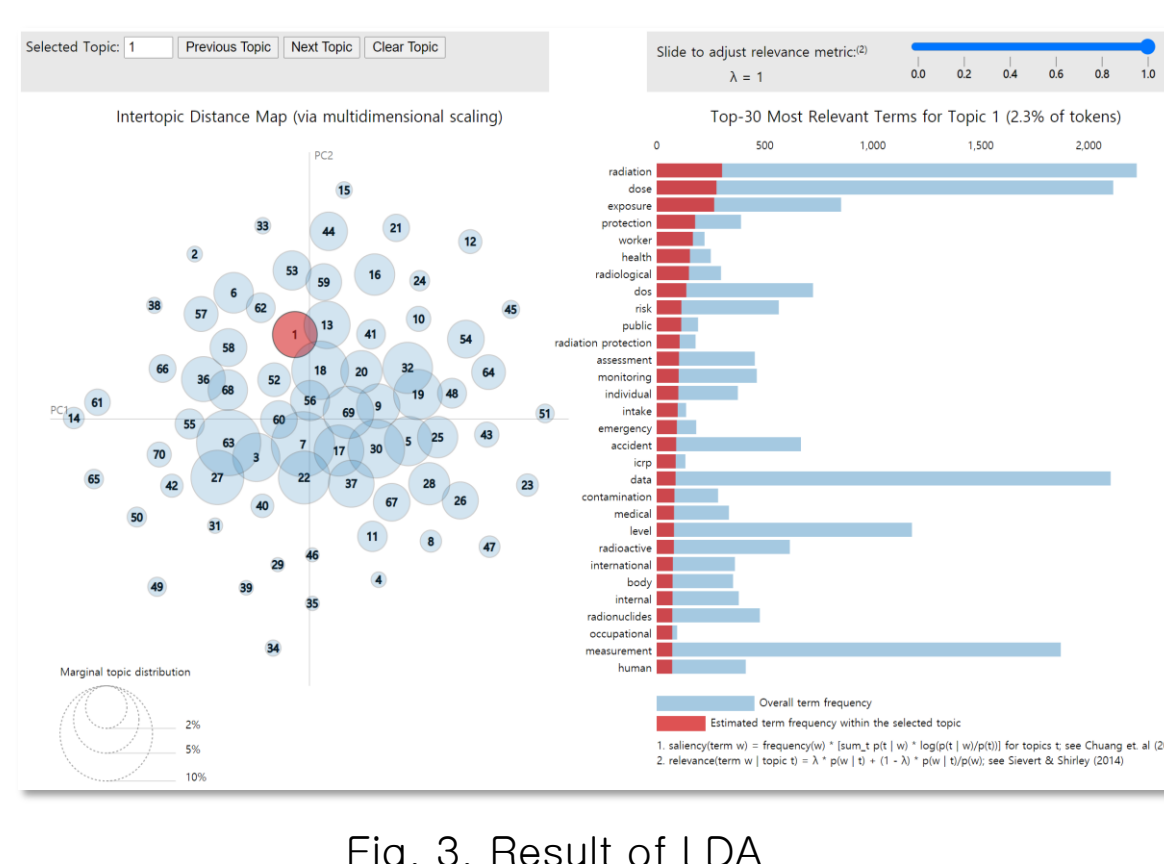
- We grouped similar subjects and gave the same names to them. Table 1 shows the classified topic names and some keywords of each topic that influenced the determination of the topic name.

|	Topic Name	Main keywords
1	Physics	'capture cross section', 'neutron diffusion', 'prompt gamma ray', 'gamma ray neutron', 'transport equation', 'neutron activation analysis', 'radioactive ion'
2	Radiation Application	'spectrometry', 'detector', 'target', 'scanner', 'ct', 'gamma ray dose', 'glow curve', 'photomultiplier', 'boron neutron capture therapy', 'resistive', 'plate chamber'
3	Radiation Protection	'activity concentration', 'annual effective', 'indoor', 'worker', 'concentration radon', 'dose assessment', 'tritium', 'dose equivalent'
4	I&C	'data acquisition', 'control', 'real time', 'signal', 'programmable', 'software', 'module', 'event', 'implementation', 'processing', 'system', 'algorithm'
5	Advanced Reactor	'inertial confinement fusion', 'tokamak', 'plasma', 'magnet', 'high temperature gas', 'blanket', 'tritium breeding', 'vacuum vessel'
6	Nuclear Fuel	'fuel assembly', 'spent fuel', 'cycle', 'burnup', 'repository', 'silicon carbide', 'pellet', 'thermal conductivity', 'mox', 'fission gas', 'enriched uranium'
7	Structural & Material	'strength', 'stress', 'mechanical', bismuth', 'stainless steel', 'tensile test', 'toughness', 'hardening', 'ductile', 'welding', 'tungstate', 'indium'
8	Policy & Economics	'energy consumption', 'renewable energy', 'cost', 'policy', 'expertise', 'sustainability', 'sustainability', 'thermal efficiency', 'solar', 'energy storage'
9	Chemical	'LiCl', 'cyclic voltammetry', 'aqueous solution', 'molten salt', 'electrochemical', 'electrolyte', 'aqueous solution', 'aqueous solution', 'ph value'
10	Safety	'probabilistic safety assessment', 'Fukushima', 'severe accident', 'decay heat', 'failure probability', 'loss coolant accident', 'accident management'
11	Thermal-Hydraulics	'heat transfer', 'fluid mechanics', 'critical heat flux', 'flow rate', 'decay heat', 'fluid dynamic', 'phase flow', 'turbulent', 'tube', 'steam'

Table 1 : Categorized Topic Names
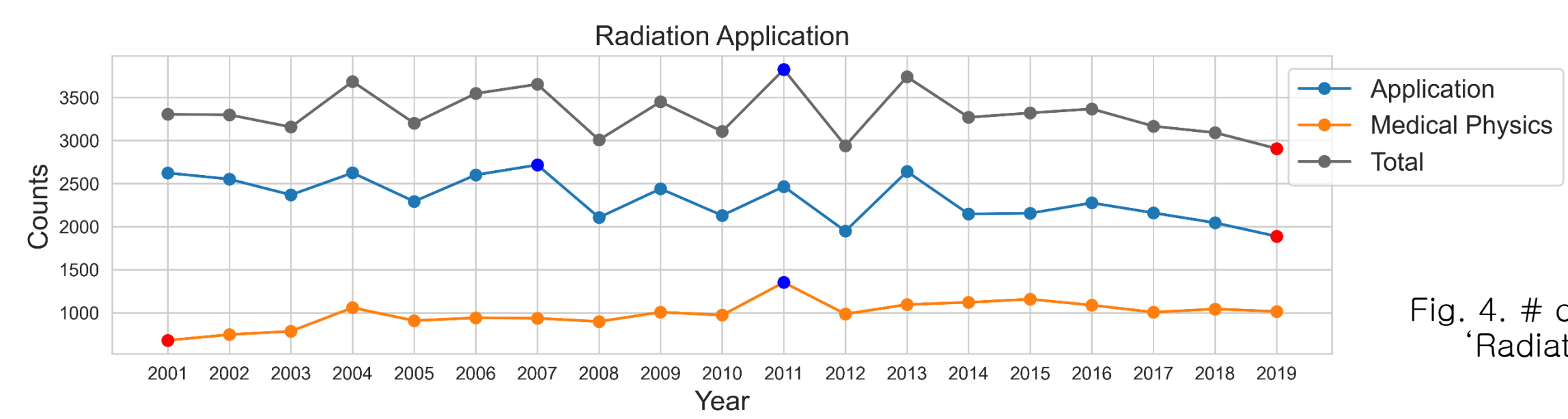
Fig. 4. # of papers in Topic 2 'Radiation Application'

  Shown in Fig.4, **Topic 2 'Radiation Application'** contains keywords for spectrometers, detectors, and other related equipment. And it also covers radiotherapy equipment and medical scanners(CT and MRI). Since radiation technology is used in various fields, Topic 2 covers at least 30% of all periods. Topic 2 contains keywords for spectrometers, detectors, and other related equipment.
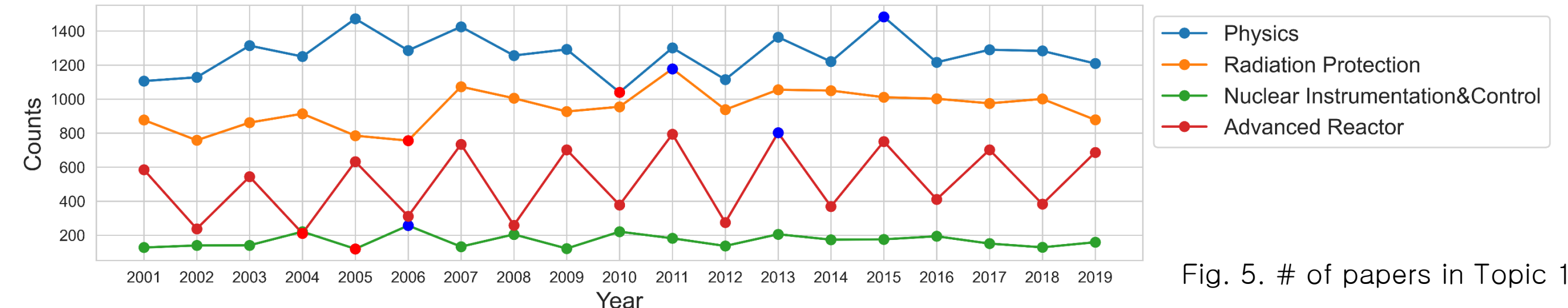

Fig. 5. # of papers in Topic 1,3,4,5

  Topic 1 'Physics' contains the study of neutron behavior, atomic reactions, neutron calculations, and kinetic physics. **Topic 3 'Radiation Protection'** discusses the absorption dose of workers or the protection of people and the environment from radiation. **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.5 shows the annual trend of Topic 1, 3, 4, 5.
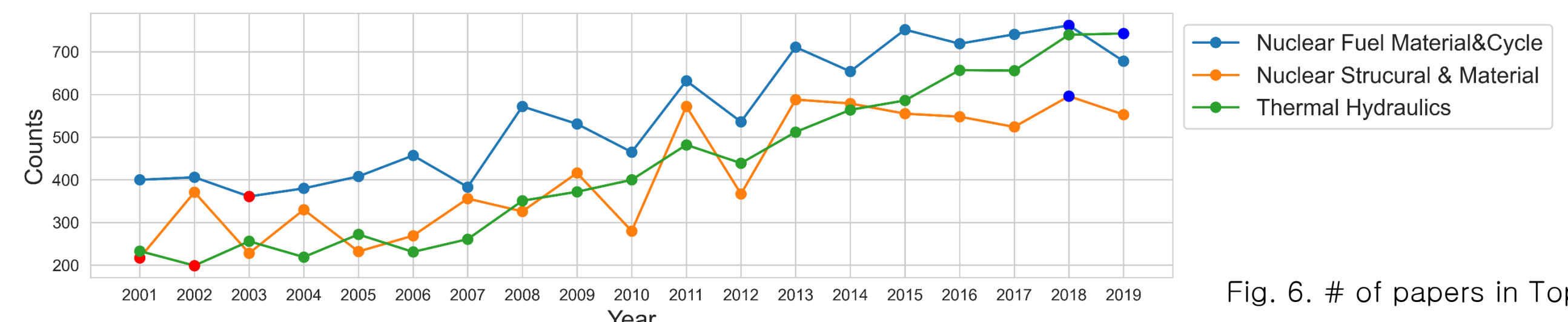

Fig. 6. # of papers in Topic 6,7,11

  The number of papers in Topic 6, 7, 11 is shown in Fig. 6. **Topic 6 'Nuclear Fuel'** 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 solution to many problems, so we understand the upward trends in topic 6. **Topic 7 'Structure and Materials'** consists of the words about structure analysis and material. After the 2011 Fukushima accident, the number of papers has increased. The increased research trend indicates efforts to maintain plant structure. **Topic 11 'Thermal Hydraulics'** 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.
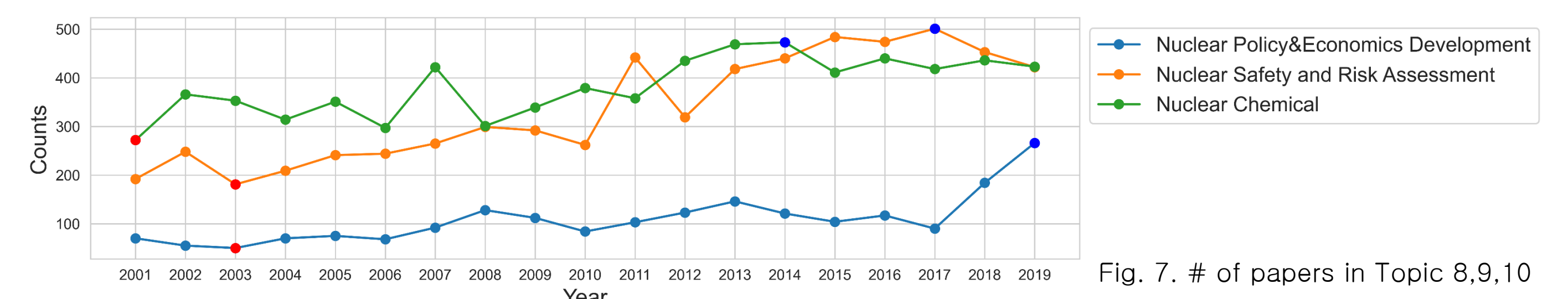

Fig. 7. # of papers in Topic 8,9,10

  Fig. 7 shows the trend of Topic 8 'Policy & Economics', Topic 9 'Chemical', Topic 10 'Safety'. All three of these topics are on the rise. **Topic 8 'Policy & Economics'** contains keywords that refer to discussions about the efficiency and economics of various energies. 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.

- 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, and observed the changes in each topic over time.

- Topic of radiation applications and nuclear medicine had the highest number of papers during the entire period, but their proportion gradually decreased. Research in the fields of fuel and structural analysis and thermal-hydraulics has increased rapidly. Most other topics are also being studied steadily.

- Interestingly, there are significant changes in trends in 2011 when the Fukushima nuclear accident occurred, and in 2012, the following year.

- 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. Also, we interpret that results show interest in safer and more economical use of nuclear power.

- This study is meaningful because it has never analyzed research trends in the entire paper on nuclear technology. Many web sites already support analysis of trends by subject or word. But to group sub-themes by topics of papers, researchers' knowledge must be added.

- [1] Benjamin Bengfort, Rebecca Bilbro, Tony Ojeda, Applied Text Analysis with Python, O'Reilly Media, 2018.

- [2] David M.Blei, Andrew Y.Ng, Michael I.Jordan Latent Dirichlet Allocation, Journal of Machine Learning Research 3, p. 993-1022, 2003.