

# 텍스트 네트워크 분석과 임베딩 모델을 이용한 일본 과학기술정책 의제 변화 분석

김민호\*

김영준\*\*

## 〈目 次〉

- |           |          |
|-----------|----------|
| I. 서론     | IV. 분석결과 |
| II. 선행연구  | V. 결론    |
| III. 연구설계 |          |

## 〈요 약〉

일본 과학기술기본계획은 일본 과학기술정책의 주요 키워드 및 정책의 변화를 거시적인 관점에서 파악할 수 있는 정책 텍스트이다. 본 논문의 목적은 일본 '과학기술기본계획'과 '과학기술·이노베이션기본계획'을 대상으로 텍스트 네트워크 분석과 임베딩 모델을 이용하여 텍스트 네트워크 구조를 밝히고, 정책의제의 변화를 분석하는 것이다. 본 논문에서는 자연어로 작성되어있는 일본 과학기술정책 텍스트를 텍스트 네트워크로 구축하고, 중심성 및 코사인 유사도를 계산한 뒤 분석한다.

분석결과, 중심성이 높은 주요의제가 시기별로 변화하고 있음을 확인할 수 있다. 이는 각 기간마다 정책의 주요 의제가 다양하게 나타나고 있음을 시사한다. 그리고 정보통신기술과 관련된 용어들은 최근 기본계획에서 '사회'와의 코사인 유사도가 높아진 것으로 나타났다. 텍스트 분석을 통해 일본의 최근 과학기술기본계획이 '사회'와 정보통신기술의 조화를 강조하며 'Society 5.0' 개념을 추구하고 있음을 확인할 수 있었다.

【주제어: 일본 과학기술정책, 텍스트 네트워크 분석, 텍스트 임베딩 모델, 과학기술기본계획, 자연어처리】

\* 제1저자, 고려대학교 과학기술협동과정 박사과정(max22justice@gmail.com)

\*\* 교신저자, 고려대학교 기술경영전문대학원 교수(youngjikim@korea.ac.kr)

논문접수일(2023.11.1), 수정일(2024.1.10), 게재확정일(2024.1.30)

## I. 서론

일본의 과학기술기본법(科学技術基本法)은 1995년 11월 8일 참의원 본회의에서 전회 일치로 가결, 성립되었다. 과학기술기본법은 향후 일본 과학기술정책의 기본틀을 제시하고 있으며, 일본 정부가 과학기술의 진흥을 강력히 추진해 나가는데 근간이 되는 성격을 갖는다(水間英城, 1996). 과학기술·이노베이션기본법<sup>1)</sup> 제 1조에서는 “과학기술·이노베이션 창출의 진흥에 관한 시책을 종합적이고 계획적으로 추진”이라는 목적을 정하고 있으며, 제 12조에서는 “정부는 과학기술·이노베이션 창출의 진흥에 관한 시책의 종합적이고 계획적인 추진을 도모하기 위해 과학기술·이노베이션 창출의 진흥에 관한 기본적인 계획을 책정해야 한다.”고 규정하고 있다. 이에 따라 일본 정부는 과학기술기본계획에 의거하여 과학기술정책을 추진하고 있다. 일본의 과학기술기본계획은 연구개발 추진에 관한 종합적인 방침, 연구시설의 정비, 연구개발과 관련된 정보화 촉진 및 기타 연구개발 추진을 위한 환경정비에 관해 정부가 종합적이고 계획적으로 강구해야 할 시책, 기타 과학기술 진흥에 관해 필요한 사항을 정하고 있으며(赤池伸一, 2019), 일본 과학기술정책의 주요 키워드 및 정책의 변화를 거시적인 관점에서 파악할 수 있는 과학기술정책 텍스트이다.

텍스트 분석은 데이터베이스의 대규모화가 진행된 오늘날 계속해서 증가하고 있는 정책 텍스트에서 중요 정책의제를 도출하는 데 있어 매우 효과적인 분석 도구라 할 수 있다. 특히 해외 과학기술정책에 대한 텍스트 분석은 글로벌적인 과학기술정책 동향과 의미를 파악하는데 효과적인 분석방법이다. 이 분야에서 텍스트 분석을 수행하면 다양한 국가 및 기관에서 발표된 과학기술정책 텍스트를 분석하여 해외 과학기술정책 동향을 파악할 수 있다. 이를 통해 새로운 기술의 등장, 정책 변화 등을 파악하여 다양한 국가에서 어떠한 정책 분야에 주목하고 있는지를 이해할 수 있다. 그리고 정책 텍스트에서 추출된 정보를 기반으로 정부 및 기관의 과학기술정책 방향성을 이해할 수 있다.<sup>2)</sup> 이는 향후 한국의 과학기술정책에 대한 방향을 설정하는 데 도움이 될 수 있다. 그러나 일본어로 작성된 과학기술정책 텍스트는 언어적인 장벽으로 인해 텍스트 네트워크 분석을 적용하여 분석하는데 있어 어려움이 있다. 따라서 일본어로 작성된 과학기술정책 텍스트를 대상으로 텍스트 네트워크 분석을 진행하기 위해서는 분석가능 데이터로의 변환 가능성 검토와 텍스트 네트워크 특징 및 구조에 주

1) 2021년 4월 과학기술기본법등의일부를개정하는법률에 의해 과학기술기본법은 과학기술·이노베이션 기본법으로 변경되었다.

2) 미국 특허 문서 초록(박주섭 외, 2018), 항공우주산업 관련 저널(김희수 외, 2021), 중국 과학기술 성과 100선(임강희·강경환, 2023), 중국공산당 보고서(임강희 외, 2023) 등 해외 과학기술관련 텍스트를 대상으로 텍스트 분석기법을 적용하여 과학기술 발전방향을 분석하는 연구가 수행되었다.

목할 필요가 있다.

본 논문의 목적은 일본 ‘과학기술기본계획’과 ‘과학기술·이노베이션기본계획’을 대상으로 텍스트 네트워크 분석과 임베딩 모델을 이용하여 텍스트 네트워크 구조를 밝히고, 정책 의제의 변화를 분석하는 것이다. 본 논문에서는 자연어로 작성되어있는 일본 과학기술정책 텍스트를 텍스트 네트워크로 구축하고, 중심성 및 코사인 유사도를 계산한 뒤 분석한다. 이것은 정책의제 간의 관계성을 파악함으로써 과학기술기본계획 전체를 부감하고, 일본 과학기술정책의 동향을 파악할 수 있다.

본 논문의 구성은 다음과 같다. 2장에서는 일본 과학기술기본계획, 텍스트 분석기법을 적용한 정책 텍스트 분석과 관련된 선행연구를 검토했다. 3장은 연구문제, 분석의 틀, 데이터 수집 및 전처리 과정, 그리고 분석방법을 설명하는 연구설계이다. 4장은 실증분석으로 텍스트 네트워크 특징과 구조를 분석하기 위해 텍스트 네트워크 분석과 임베딩 모델을 이용하여 분석하였다. 5장은 본 연구의 결과와 연구결과에서 도출할 수 있는 정책적 시사점, 그리고 본 연구의 한계점을 제시한다.

## II. 선행연구

### 1. 일본 과학기술기본계획 및 과학기술·이노베이션기본계획의 이해

일본은 과학기술기본법 제9조<sup>3)</sup>에 따라 과학기술기본계획을 수립하였다. 과학기술기본 계획 책정에 있어서는 종합과학기술회의(현재는 종합과학기술·이노베이션회의)를 거치게 되어있다.<sup>4)</sup> 과학기술기본계획에서는 연구개발 추진에 관한 종합적인 방침, 연구시설의 정비, 연구개발과 관련된 정보화 촉진 및 기타 연구개발 추진을 위한 환경정비에 관해 정부가 종합적이고 계획적으로 강구해야 할 시책, 기타 과학기술 진흥에 관해 필요한 사항을 정하게 되어 있다(赤池伸一, 2019). 일본 정부는 지금까지 제1기(1996년~2000년), 제2기(2001년~2005년), 제3기(2006년~2010년), 제4기(2011년~2015년), 제5기(2016년~2020년) 과학기술기본계획을 각의결정하였다. 그리고 2021년 4월 과학기술기본법등의일부를개정하는법률에 의해 과학기술기본법은 과학기술·이노베이션기본법(科学技術・イノベーション

3) 정부는 과학기술 진흥에 관한 시책의 종합적이고 계획적인 추진을 도모하기 위해 과학기술 진흥에 관한 기본적인 계획을 책정하여야 한다.

4) 과학기술기본법(科学技術基本法) 제9조, 내각부설치법(内閣府設置法) 제26조.

基本法)으로 변경함에 따라, 2021년부터 향후 5년간을 대상으로 하는 제6기 과학기술·이노베이션기본계획을 각의결정하였다. 이에 따라 일본 정부는 책정된 기본계획에 따라 과학기술정책을 추진해 오고 있다.

일본의 종합과학기술·이노베이션회의는 과학기술정책을 담당하는 조직이다. 과학기술정책은 문부과학성, 경제산업성 등 많은 부처와 관련된 사업이 많기 때문에, 일본의 종합과학기술 이노베이션회의는 범부처(省庁横断) 조직이라 할 수 있다. 과학기술 기본계획의 내용은 종합과학기술·이노베이션회의의 전문조사회에서 논의를 거쳐 차기 계획이 검토된다(小林信一 外, 2019). 또 문부과학성 과학기술학술심의회, 경제산업성 산업구조심의회 등에서 각 성의 소관사항을 중심으로 한 기본계획 책정을 위한 적극적인 검토가 이루어져 보고서가 작성된다(小林信一 外, 2019).

과학기술기본계획은 각기 마다 내용구성은 변화하고 있지만, 정부연구개발투자목표의 설정, 중점연구개발분야와 해결해야 할 중점사회과제의 설정, 연구개발시스템의 개혁, 과학기술이노베이션 정책의 추진체제 등의 사항이 대체로 포함되어있다(小林信一 外, 2019; 研究開発戦略センター, 2022).

## 2. 텍스트 분석기법을 적용한 정책 텍스트 분석

복잡한 관계에 내재된 구조적 유사성을 탐색적으로 찾아내기에 유용한 기법이라 할 수 있는 네트워크 분석은(고길근, 2007), 구성요인들 간의 관계와 추세를 파악할 수 있다는 장점이 있다(윤호열 외, 2020). 네트워크 분석은 데이터베이스의 대규모화가 진행된 오늘날 계속해서 증가하고 있는 정책 텍스트에서 중요 정책의제를 도출하는 데 있어 매우 효과적인 분석기법이라 할 수 있다.

내용분석은 일련의 절차를 사용하여 텍스트에서 타당한 추론을 하는 연구 방법이다(Weber, 1990). 내용분석은 모든 유형의 텍스트에 적용할 수 있는데(Bengtsson, 2016), 대부분의 내용분석은 인쇄된 자료(신문, 잡지, 책, 정부 간행물 등)를 분석대상으로 하지만, 인쇄된 자료 이외의 자료(영화, 라디오 방송 등)도 분석대상으로 적절하다(Saraisky, 2015). 텍스트 분석기법을 활용한 내용분석은 다양한 텍스트 자료를 대상으로 분석을 수행하여 의미 있는 정보를 찾고자 하는 분석방법으로서, 트위터 계정을 대상으로 하는 텍스트 분석을 통해 주요 주제 클러스터를 식별(Herrero-Solana & Faba-Pérez, 2023)하거나 다양한 행위자들 간의 상호작용(황한찬·엄석진, 2019)을 분석할 수 있었으며, 역대 대통령의 주요 연설문(박치성·신나리, 2021)이나 정부 시정연설(고기동·이은미, 2022)을 대상으로 텍스트 분석 기법을 적용하여 정책변화의 동인을 분석할 수 있었다. 또한, 국민이 작성한 의견을 분석하

여 정책이슈를 도출할 수 있는데, 정책제안 플랫폼에서 수집된 데이터를 통해 정책이슈를 도출하였으며(김희철·한수미, 2020), 국민청원글을 대상으로 하여 정책이슈를 분석할 수 있다(김찬우, 2019; 황한찬·엄석진, 2021).

텍스트 분석기법을 적용한 내용분석 방법은 정책 텍스트 내 관련 주제의 본질적인 사실과 관련 경향을 명확하게 분석할 수 있어 정책분석에 활용해 왔다(Peng & Liu, 2016; Liao, 2018). 주요 정책 텍스트에 대한 내용분석을 통해, 정책목표와 거버넌스 메커니즘을 분석하였으며(Alves & Lee, 2022), 다년간의 정책내용 변화를 분석하여 데이터 내용의 주제 및 특징을 파악하거나(이은미 외, 2016; 김민호 외, 2021; 신승윤 외, 2021), 정책하위체제 구조를 분석하여 정책변동과정을 심도 있게 분석하거나(이수정·박광국, 2022), 정책방향과 우선순위를 확인하는 연구가 수행되었다(Guo et al., 2023). 또한, 특정 국가나 대상에 국한되는 것이 아니라, 여러 국가의 정책 텍스트를 종합하여 비교 분석하는 연구도 수행되었다(Saheb & Saheb, 2023). 텍스트 마이닝을 통해 단어를 추출한 뒤 단어 간의 관계를 통해 정책변동 및 정책방향을 탐색하는 연구는 경제정책(고기동, 2019; 연지현·김성원, 2022), 산업정책(황한찬·엄석진, 2018), 교육정책(김옥일, 2008; 임봉조, 2022), 북한정책(이가영, 2020; 남성욱·백연주, 2022), 여성정책(김학실, 2012), 지방정책(박주섭·홍순구, 2016), 과학기술정책(박주섭 외, 2017; 이주영·정효정, 2017; 김민호 외, 2021; 김희수 외, 2021) 등 다양한 정책연구 분야에서 활용되고 있다.

### III. 연구설계

#### 1. 연구문제

본 연구는 자연어로 작성된 일본의 ‘과학기술기본계획’과 ‘과학기술·이노베이션기본계획’을 대상으로 문자 인코딩과 데이터 전처리 방법론을 적용하여 분석가능한 데이터로 변환한 후, 텍스트 네트워크 분석과 임베딩 모델을 적용하여 텍스트 네트워크의 특징과 주요 정책의제가 시기별로 변화하고 있음을 파악하고자 한다. 본 연구의 연구문제는 다음과 같다.

연구문제 1: 텍스트 네트워크 분석을 적용하여 일본 과학기술정책 텍스트의 텍스트 네트워크의 특징과 정책의제가 시기별로 변화하고 있음을 확인할 수 있는가

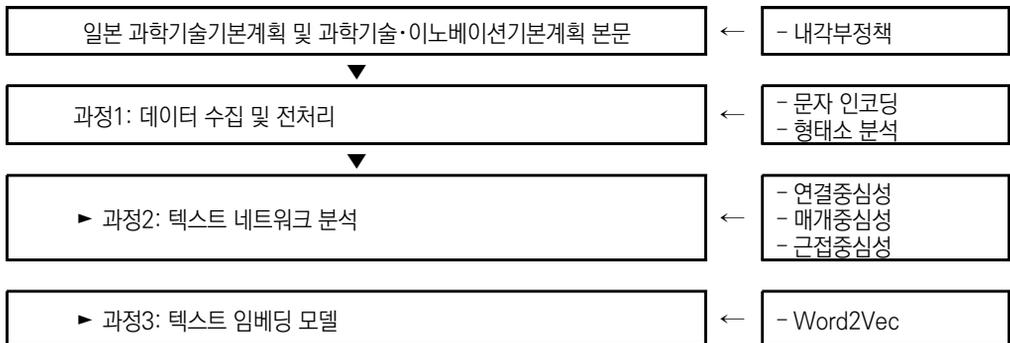
연구문제 2: 텍스트 임베딩 모델을 적용하여 일본 과학기술정책 텍스트의 텍스트 네트워크의 특징과 정책의제가 시기별로 변화하고 있음을 확인할 수 있는가

## 2. 분석의 틀

일본 과학기술기본계획에 대한 텍스트 네트워크의 특징과 정책의제가 시기별로 변화하고 있음을 파악하기 위한 전체 프로세스를 소개한다.

본 논문은 세 단계로 구성된다. 첫째, 분석대상 자료는 내각부정책(内閣府の政策) 과학기술정책 데이터베이스에서 수집하였다. 분석대상의 범위는 제1기 과학기술기본계획부터 제6기 과학기술·이노베이션기본계획까지이다. 수집된 정책 텍스트는 분석프로그램이 인식할 수 있도록 데이터 전처리 과정을 수행하였다. 둘째, 텍스트 네트워크 분석 모델을 적용하여 텍스트 네트워크 구조를 구성하였으며, 정책의제 및 정책의제 간의 연결 정도를 파악하기 위해 연결중심성, 매개중심성, 근접중심성 척도를 계산하였다. 또한 정책의제의 중심성 및 연결 정도를 시각화해 직관성을 높였다. 셋째, 단어 간 의미적 유사성을 포착하는데 유용한 텍스트 임베딩 모델을 적용하여 텍스트 네트워크 구조의 변화를 확인하고자 하였다.

〈그림 1〉 논문의 전반적인 프로세스



## 3. 데이터 수집 및 전처리

논문의 분석대상인 일본 과학기술기본계획 및 과학기술·이노베이션기본계획 본문은 내각부정책 과학기술·이노베이션 데이터베이스에서 수집하였다. 과학기술기본계획의 근거가 되는 일본의 과학기술기본법은 1995년 11월 8일 참의원 본회의에서 가결, 성립되었다. 과학기술·이노베이션기본법 제 12조에서는 “정부는 과학기술·이노베이션 창출의 진흥에 관한 시책의 종합적이고 계획적인 추진을 도모하기 위해 과학기술·이노베이션 창출의 진흥에 관한 기본적인 계획을 책정해야 한다.”라고 규정하고 있다. 이에 따라 일본 정부는 지금까지 제1기(1996년~2000년), 제2기(2001년~2005년), 제3기(2006년~2010년), 제4기(2011년~

2015년), 제5기(2016년~2020년) 과학기술기본계획을 각의결정하였으며, 2021년도부터 향후 5년간을 대상으로 하는 제6기 과학기술·이노베이션기본계획을 각의결정하였다. 이에 따라 일본 정부는 책정된 기본계획에 의거하여 과학기술정책을 추진해 오고 있다.

본 논문의 분석대상인 일본의 ‘과학기술기본계획’과 ‘과학기술·이노베이션기본계획’ 본문은 정리되지 않은 비정형(unstructured) 텍스트이다. 우선 텍스트 분석을 수행하기 위해 분석 가능한 데이터로 변환하는 과정을 수행하였다. 컴퓨터가 각기 다른 연도에 작성된 문자를 동일하게 인식할 수 있도록, 모든 기본계획의 문자 인코딩(character encoding) 방식을 통일시켰다. 그리고 일본어 형태소 분석 알고리즘(MeCab)을 이용하여 텍스트 분석에서 의미가 없거나 관련이 없는 단어, 특수기호 등을 제거한 뒤, 단어들을 추출하였다(김민호 외, 2021).

분석대상 자료인 ‘과학기술기본계획’과 ‘과학기술·이노베이션기본계획’에 포함된 글자 수와 데이터 전처리과정을 거친 글자 수를 <표 1>에 정리하였다. 데이터 전처리과정을 거친 글자 수는 전처리과정 이전 글자 수에 비해 절반 이하로 감소했다. 또한, 기본계획이 갱신되면서 글자 수가 증가하고 있다.

<표 1> 분석대상 자료에 포함된 글자 수(데이터 전처리 전/후)

연도	분석대상 자료	전처리 전 글자 수	전처리 후 글자 수
1996~2000년	제1기 과학기술기본계획	26,359	12,700
2001~2005년	제2기 과학기술기본계획	46,274	21,730
2006~2010년	제3기 과학기술기본계획	55,560	25,531
2011~2015년	제4기 과학기술기본계획	60,477	29,647
2016~2020년	제5기 과학기술기본계획	68,466	33,715
2021~2025년	제6기 과학기술·이노베이션기본계획	112,701	53,973

## 4. 분석방법

### 1) 텍스트 네트워크 분석

데이터베이스의 대규모화가 진행된 오늘날 계속해서 증가하고 있는 정책 텍스트에서 텍스트 네트워크 분석은 중요 정책의제를 도출하는 데 있어 매우 효과적인 분석 도구라 할 수 있다. 일본 ‘과학기술기본계획’과 ‘과학기술·이노베이션기본계획’의 텍스트 네트워크 특징 및 구조를 분석하기 위해 텍스트 네트워크 분석을 적용하였다. 우선 전처리과정을 거친 단

어들을 텍스트 네트워크 분석을 할 수 있도록 각각의 단어를 노드(Node)로 하였으며, 인접 단어 간의 관계를 링크(Link)로 설정하였다. 그리고 인접단어 간의 관계가 많을수록 링크의 가중치가 높아지도록 설정하였다.

연결중심성(Degree Centrality)은 분석대상인 노드와 연결된 주변노드의 총합으로 측정되며, 그 관계를 맺은 인접한 노드의 수를 기준으로 측정되므로 중심단어가 주변단어와 어느 정도 관련되어 있는지를 나타낸다. 연결성은 노드가 갖는 링크의 수로 정의된다. 따라서 연결중심성( $C_D$ )은 중심으로 삼을 단어 노드( $i$ )와 인접한 단어 노드( $j$ )의 최단거리( $L_{ij}$ )를 모두 더해서 구할 수 있다. 연결중심성은 다음 공식(1)과 같이 표현할 수 있다.

$$C_D(i) = \sum_j^n L_{ij} \quad (1)$$

매개중심성(Betweenness Centrality)은 두 노드를 연결하는 최단 링크 간에 분석대상 노드를 포함하는 링크의 수로 계산된다. 즉 중개자의 기능을 하는 단어를 파악할 수 있는 지표가 된다. 매개중심성( $C_B$ )은 노드( $s$ )와 노드( $t$ )로 가는 최단 경로의 개수( $\sigma_{st}$ )중에서 측정되는 노드( $i$ )를 거쳐가는 경로의 개수( $\sigma_{st}(i)$ )를 더해서 구한다. 매개중심성은 공식(2)로 표현할 수 있다.

$$C_B(i) = \sum_i^n \frac{\sigma_{st}(i)}{\sigma_{st}}, (s \neq t \neq i) \quad (2)$$

근접중심성(Closeness Centrality)은 분석대상인 노드가 다른 노드들과 얼마나 가까운지를 파악할 수 있다. 근접중심성이 높은 노드는 텍스트 네트워크 구조 속에서 다른 노드에 도달하기 위한 연결거리가 짧고 구조적으로 중심에 있음을 의미한다. 근접성은 거리의 역수로 정의되었다(Bavelas, 1950). 따라서 근접중심성은 노드와 다른 모든 노드 사이의 최단 경로 길이의 합의 역수로 구할 수 있다. 근접중심성( $C_C$ )은 기준 노드( $i$ )에서 노드( $i$ )를 제외한 다른 노드들( $j$ )까지의 최단 경로의 길이( $L_{i,j}$ )에 대해 평균을 내고( $n-1$ ), 그 값을 역수로 해서 구할 수 있다. 근접중심성은 다음 공식(3)과 같이 표현할 수 있다.

$$Cc(i) = \frac{1}{\frac{1}{n-1} \sum_{i \neq j} L_{i,j}} \tag{3}$$

또한, 노드의 중요도와 다른 노드 간의 관계에 대한 직관을 높이기 위해 시각화 작업이 수행되었다. 노드의 연결성 정도에 따라 노드의 크기를 차등하였으며, 노드와 인접노드의 연결 빈도에 따라 링크의 굵기를 차등하여 설정하였다. 그리고 각각의 ‘과학기술기본계획’과 ‘과학기술·이노베이션기본계획’은 글자 수에 차이가 있어, 시각화에 사용할 단어의 숫자를 다르게 조정하였다.<sup>5)</sup>

## 2) 텍스트 임베딩 모델(Word2Vec), 코사인 유사도, 상관분석

2013년에 발표된 "Efficient Estimation of Word Representations in Vector Space" 논문(Mikolov et al., 2013)에서는 텍스트 임베딩 모델이 소개되었다. 이 논문은 자연어처리 분야에서 단어를 벡터로 표현하는 방법으로, 텍스트 임베딩 모델의 주요 구조를 제시했다. 텍스트 임베딩 모델은 단어 간 의미적 유사성을 포착하여 자연어처리 및 관련 분야에서 광범위하게 활용되고 있다.

본 논문에서는 텍스트 임베딩 모델 중에서 Skip-gram 알고리즘을 사용하였다. Skip-gram은 자연어 처리에서 사용되는 단어 임베딩 기법 중 하나로 중심단어를 중심으로 주변단어를 예측하는 알고리즘으로 단어를 벡터로 표현할 수 있다(Mikolov et al., 2013). 또한, 알고리즘의 연산속도를 증가시키기 위해 negative sampling을 적용하였다. negative sampling은 전체 말뭉치 중에서 빈도수가 높은 단어가 선택되도록 하는 샘플링기법이다. 분석을 위한 Skip-gram 알고리즘과 negative sampling의 파라미터는 기본값으로 설정하였다.

데이터 전처리 과정을 거쳐 추출한 단어들을 대상으로 텍스트 임베딩 모델을 적용하여 다차원 벡터공간에 임베딩한 뒤, 두 벡터의 유사도인 코사인 유사도를 측정하였다. 코사인 유사도는 두 벡터가 가리키는 방향이 얼마나 유사한가를 의미한다.<sup>6)</sup> 코사인 유사도는 다음 공식(4)와 같이 표현할 수 있다.

5) 제1기 과학기술기본계획은 연결성 10개 이상. 제2기, 제3기, 제4기, 제5기 과학기술기본계획은 연결성 15개 이상. 제6기 과학기술·이노베이션기본계획은 연결성 20개 이상.

6) 두 벡터의 방향이 완전히 동일할 경우에는 1의 값을 가지며, 반대방향이면 -1이다.

$$\text{sim}(x, y) = \frac{x \cdot y}{\|x\| \times \|y\|} = \frac{\sum_{i=1}^n x_i \times y_i}{\sqrt{\sum_{i=1}^n x_i^2} \times \sqrt{\sum_{i=1}^n y_i^2}} \quad (4)$$

과학기술기본계획 간의 관계를 이해하기 위해 두 벡터의 유사도인 코사인 유사도를 대상으로 피어슨(Pearson) 상관분석을 수행하였다. 피어슨 상관계수는 두 변수(X, Y) 사이의 선형 상관관계의 강도와 방향을 설명할 수 있다. 피어슨 상관계수는 두 변수의 공분산을 변수 X의 표준편차와 변수 Y의 표준편차를 곱한 값으로 나누어 구할 수 있다. 피어슨 상관계수는 다음 공식(5)과 같이 표현될 수 있다.

$$r_{XY} = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_i (X_i - \bar{X})^2} \sqrt{\sum_i (Y_i - \bar{Y})^2}} \quad (5)$$

## IV. 분석결과

### 1. 텍스트 네트워크 분석

전처리 작업을 수행한 후, 텍스트 네트워크 분석 모델을 적용하여 텍스트 네트워크 구조를 구성하였다. 그리고 정책의제의 중심성 및 연결 정도를 시각적으로 파악하기 위해 <표 8>과 같이 중심성 분석결과를 가시화하고 정리하였다.

다음으로 정책의제 및 정책의제 간의 연결 정도를 파악하기 위해 연결중심성, 매개중심성, 근접중심성을 측정하였다. 그리고 중심성 값을 분석하여 의미 있는 정보를 찾고자 했다.

#### 1) 제1기 과학기술기본계획의 중심성

제1기 과학기술기본계획에 대한 텍스트 네트워크 분석 결과를 <표 2>로 정리하였다. 제1기 중심성의 경우 다른 기본계획과는 다르게 '확충'의 중심성(연결 61, 매개 28862, 근접 0.000473)이 높게 나타났다. '확충'과 관련된 단어를 살펴본 결과 '확충-자금'의 연결정도는

13, '확충-연구'의 연결정도는 8, '확충-대학'의 연결정도는 4였다. 시각화 그림을 보면, 제1기에서는 '연구'가 가장 연결이 많은 중심노드가 되고, '연구'를 중심으로 인접노드들이 연결된다. 제1기의 경우 '연구'와 '개발'의 연결정도가 높다는 점을 제외하면 중심성이 큰 노드는 없다. 제1기에서는 그 후의 기본계획에 비해 '과학'과 '기술' 간의 연결정도가 많지 않으며, 연결노드 또한 많지 않다는 특징을 보인다.

〈표 2〉 제1기 과학기술기본계획의 중심성

순위	연결중심성		매개중심성		근접중심성	
	연결중심성	연결중심성	매개중심성	매개중심성	근접중심성	근접중심성
1	연구 研究	269	연구 研究	286694	연구 研究	0.000601
2	과학 科学	89	과학 科学	55638	정비 整備	0.000504
3	추진 推進	89	정비 整備	51140	추진 推進	0.000500
4	정비 整備	88	추진 推進	45343	과학 科学	0.000492
5	개발 開発	70	기술 技術	40046	개발 開發	0.000492
6	기술 技術	65	평가 評価	35126	충실 充實	0.000478
7	사회 社会	62	사회 社会	32437	사회 社会	0.000474
8	확충 扩充	61	일본 我が国	31781	확충 扩充	0.000473
9	기관 機関	61	개발 開發	31139	기술 技術	0.000473
10	평가 評価	61	기관 機関	29131	평가 評価	0.000473
11	활동 活動	60	확충 扩充	28862	기관 機関	0.000471
12	활용 活用	56	대학 大学	27450	지원 支援	0.000471
13	충실 充實	56	필요 必要	24404	정보 情報	0.000471
14	일본 我が国	55	활용 活用	23774	국제 国際	0.000470
15	대학 大学	54	활동 活動	23770	활용 活用	0.000470

## 2) 제2기 과학기술기본계획의 중심성

제2기 과학기술기본계획에 대한 텍스트 네트워크 분석 결과를 〈표 3〉에 정리하였다. 분석결과 제2기에서는 제1기와 마찬가지로 '정비'의 중심성이 높게 나타났다(연결 12위, 매개 15위, 근접 10위). 이는 초기 과학기술기본계획에서는 그 후의 기본계획과는 다른 주요 정책의제로서 '확충'과 '정비'에 관한 문언이 담겨 있을 것으로 해석된다. 시각화 그림을 보면, 제2기에서는 제1기와 마찬가지로 '연구'가 가장 연결이 많은 중심노드가 되고, '연구'를 중심으로 인접노드들이 연결된다. 제1기의 경우 '연구'와 '개발'의 연결정도가 높다는 점을 제외하면 중심성이 큰 노드는 없다. 한편, 제2기의 경우 '연구' 이외에도 '기술' 노드가 중심성이 크고, 인접노드가 많이 연결되어있다. 제2기 과학기술기본계획은 제1기 과학기술기본

계획보다 ‘기술’과 ‘과학’의 중심성이 높다는 것을 알 수 있다. 보다 구체적으로 제1기에서는 과학기술과 연결된 단어는 ‘진흥’, ‘육성’뿐이지만, 제2기에서는 ‘진흥’, ‘육성’, ‘혁신’, ‘통신’, ‘시스템’, ‘이전’ 등의 단어가 과학기술과 연결된다. 과학기술 관련 구체적인 정책의제가 형성되고 있음을 알 수 있다.

〈표 3〉 제2기 과학기술기본계획의 중심성

순위	연결중심성		매개중심성		근접중심성	
1	연구 研究	339	연구 研究	422214	연구 研究	0.000397
2	기술 技術	218	기술 技術	261357	기술 技術	0.000370
3	과학 科学	153	과학 科学	109434	과학 科学	0.000351
4	사회 社会	132	평가 評價	98081	필요 必要	0.000344
5	평가 評價	129	사회 社会	92431	사회 社会	0.000338
6	필요 必要	126	필요 必要	82473	개발 開發	0.000337
7	일본 我が国	109	일본 我が国	68728	평가 評價	0.000337
8	기관 機關	106	국제 國際	58897	추진 推進	0.000335
9	개발 開發	96	개발 開發	58762	일본 我が国	0.000333
10	국제 國際	96	기관 機關	53086	정비 整備	0.000333
11	추진 推進	93	정보 情報	47765	산업 産業	0.000331
12	정비 整備	90	환경 環境	46584	국제 國際	0.000329
13	산업 産業	86	대학 大学	42012	활용 活用	0.000329
14	환경 環境	82	촉진 促進	41819	성과 成果	0.000328
15	대학 大学	81	정비 整備	41367	환경 環境	0.000326

### 3) 제3기 과학기술기본계획의 중심성

제3기 과학기술기본계획에 대한 텍스트 네트워크 분석 결과를 〈표 4〉로 정리하였다. 제3기 중심성의 경우 다른 기본계획에 비해 ‘경쟁’의 중심성(연결 104, 매개 40199, 근접 0.000321)이 높았다. 제3기 과학기술기본계획에서 ‘경쟁-자금’의 연결정도는 58, ‘경쟁-국제’의 연결정도는 24, ‘경쟁-강화’의 연결정도는 15인 점을 고려하면, ‘경쟁’은 주요 정책의제이다. 시각화 그림을 보면, 제3기에서는 다른 기본계획에서는 나타나지 않았던 ‘연구-여성’의 연결성이 나타났다. 제3기에서의 ‘여성’의 중심성(연결 19, 매개 4190, 근접 0.000273)은 다른 기본계획보다 중심성 순위가 높았다. ‘연구-여성’ 연결정도의 경우 제1기에서는 1, 제2기에서는 3이었으나, 제3기에서는 16으로 크게 증가했다. 제3기에서는 다른 기본계획에 비해 ‘여성’과 관련된 정책의제의 비중이 커졌음을 추론해볼 수 있다.

〈표 4〉 제3기 과학기술기본계획의 중심성

순위	연결중심성		매개중심성		근접중심성	
	연어	중심성	연어	중심성	연어	중심성
1	연구 研究	321	연구 研究	393635	연구 研究	0.000383
2	기술 技術	175	기술 技術	143910	과학 科学	0.000345
3	과학 科学	166	과학 科学	132331	기술 技術	0.000342
4	대학 大学	149	대학 大学	109381	대학 大学	0.000334
5	사회 社会	115	국제 国际	84221	국제 国际	0.000328
6	국제 国际	107	사회 社会	83767	사회 社会	0.000328
7	경쟁 競争	104	정비 整備	57297	필요 必要	0.000326
8	필요 必要	101	일본 我が国	52450	일본 我が国	0.000326
9	평가 評価	100	필요 必要	51252	기관 機関	0.000325
10	일본 我が国	98	분야 分野	48994	강화 強化	0.000325
11	기관 機関	97	평가 評価	45021	추진 推進	0.000323
12	정비 整備	96	제도 制度	44917	중요 重要	0.000321
13	인재 人材	92	인재 人材	43522	경쟁 競争	0.000321
14	활동 活動	90	활동 活動	40779	평가 評価	0.000321
15	추진 推進	89	경쟁 競争	40199	활동 活動	0.000321

#### 4) 제4기 과학기술기본계획의 중심성

제4기 과학기술기본계획에 대한 텍스트 네트워크 분석 결과를 〈표 5〉로 정리하였다. ‘이노베이션’의 연결중심성이 제1기 0, 제2기 0, 제3기 44, 제4기 72로, 그전의 기본계획에 비해 높게 측정되었다. ‘이노베이션’과 관련된 정책의제는 제3기에서 등장한 이후로 제4기부터 비중이 커졌음을 추론해볼 수 있다. 시각화 그림에서는 2011년의 ‘동일본대지진(東日本大震災)’과 관련하는 ‘동일본’, ‘대지진’, ‘복흥(復興)’, ‘자연’, ‘재해’ 등의 단어가 주변노드로서 등장했다. ‘동일본대지진’은 ‘연구’, ‘과학’, ‘기술’ 등 중심성이 높은 노드와의 관련성은 높지 않지만, 제4기 과학기술기본계획에서 영향력이 있는 노드임을 알 수 있다.

〈표 5〉 제4기 과학기술기본계획의 중심성

순위	연결중심성		매개중심성		근접중심성	
	연어	중심성	연어	중심성	연어	중심성
1	연구 研究	322	연구 研究	424965	연구 研究	0.000347
2	기술 技術	189	기술 技術	194895	추진 推進	0.000321
3	추진 推進	176	과학 科学	156763	기술 技術	0.000319
4	과학 科学	174	추진 推進	141123	과학 科学	0.000314

5	일본 我が国	149	일본 我が国	112375	일본 我が国	0.000304
6	개발 開発	125	개발 開発	103009	개발 開発	0.000304
7	대처 取組	123	국제 国際	70195	대처 取組	0.000303
8	국제 国際	110	사회 社会	69672	촉진 促進	0.000300
9	정책 政策	108	대처 取組	63689	국제 国際	0.000299
10	대학 大学	101	세계 世界	59506	사회 社会	0.000299
11	사회 社会	99	촉진 促進	56993	강화 強化	0.000295
12	촉진 促進	98	지원 支援	56883	전략 戦略	0.000291
13	세계 世界	98	정책 政策	54427	정책 政策	0.000290
14	강화 強化	91	연휴 (제휴) 連携	52401	활용 活用	0.000290
15	지원 支援	90	평가 評価	51991	이노베이션 イノベーション	0.000290

5) 제5기 과학기술기본계획의 중심성

제5기 과학기술기본계획에 대한 텍스트 네트워크 분석 결과를 <표 6>에 정리하였다. 제5기의 중심성의 경우 다른 기본계획에 비해 ‘기업’의 중심성(연결 121, 매개 79961, 근접 0.000274)이 높았다. 그리고 ‘이노베이션’의 경우, 제4기보다 중심성의 순위가 올라갔다. 시각화 그림을 보면, 제5기에서는 다른 기본계획에서는 등장하지 않았던 ‘기업’과 인접노드(‘벤처’, ‘중소’)가 등장한다. 제5기 과학기술기본계획에서는 다른 기본계획과 다른 주요 정책의제로 ‘기업’, ‘중소’, ‘벤처’ 등이 꼽혔으며, ‘이노베이션’과 관련된 정책의제의 비중이 커졌음을 추론해볼 수 있을 것이다.

제5기의 중심성의 경우 이전의 기본계획에 비해 ‘사회’의 중심성이 증가하였으며, 인접노드(‘스마트’, ‘변혁’, ‘실장’)가 등장하였다. 일본 정부는 2016년 제5기 과학기술기본계획을 통해 일본이 목표로 삼을 미래 사회로 ‘Society 5.0’을 제시했기 때문으로 판단된다.

<표 6> 제5기 과학기술기본계획의 중심성

순위	연결중심성		매개중심성		근접중심성	
1	연구 研究	281	연구 研究	367472	연구 研究	0.000307
2	기술 技術	248	기술 技術	352353	기술 技術	0.000305
3	과학 科学	174	과학 科学	158613	일본 我が国	0.000290
4	사회 社会	167	사회 社会	146979	사회 社会	0.000288
5	추진 推進	158	일본 我が国	119935	과학 科学	0.000287
6	일본 我が国	155	추진 推進	108586	대처 取組	0.000284
7	대처 取組	135	대처 取組	81227	추진 推進	0.000282
8	중요 重要	127	기업 企業	79961	중요 重要	0.000282

9	기업 企業	121	중요 重要	73964	활용 活用	0.000277
10	인재 人材	119	세계 世界	73844	강화 強化	0.000277
11	활용 活用	116	인재 人材	73621	인재 人材	0.000277
12	대학 大学	113	활용 活用	68479	이노베이션 イノベーション	0.000275
13	강화 強化	113	강화 強化	67776	기업 企業	0.000274
14	이노베이션 イノベーション	109	대학 大学	64167	창출 創出	0.000273
15	세계 世界	103	이노베이션 イノベーション	59072	필요 必要	0.000273

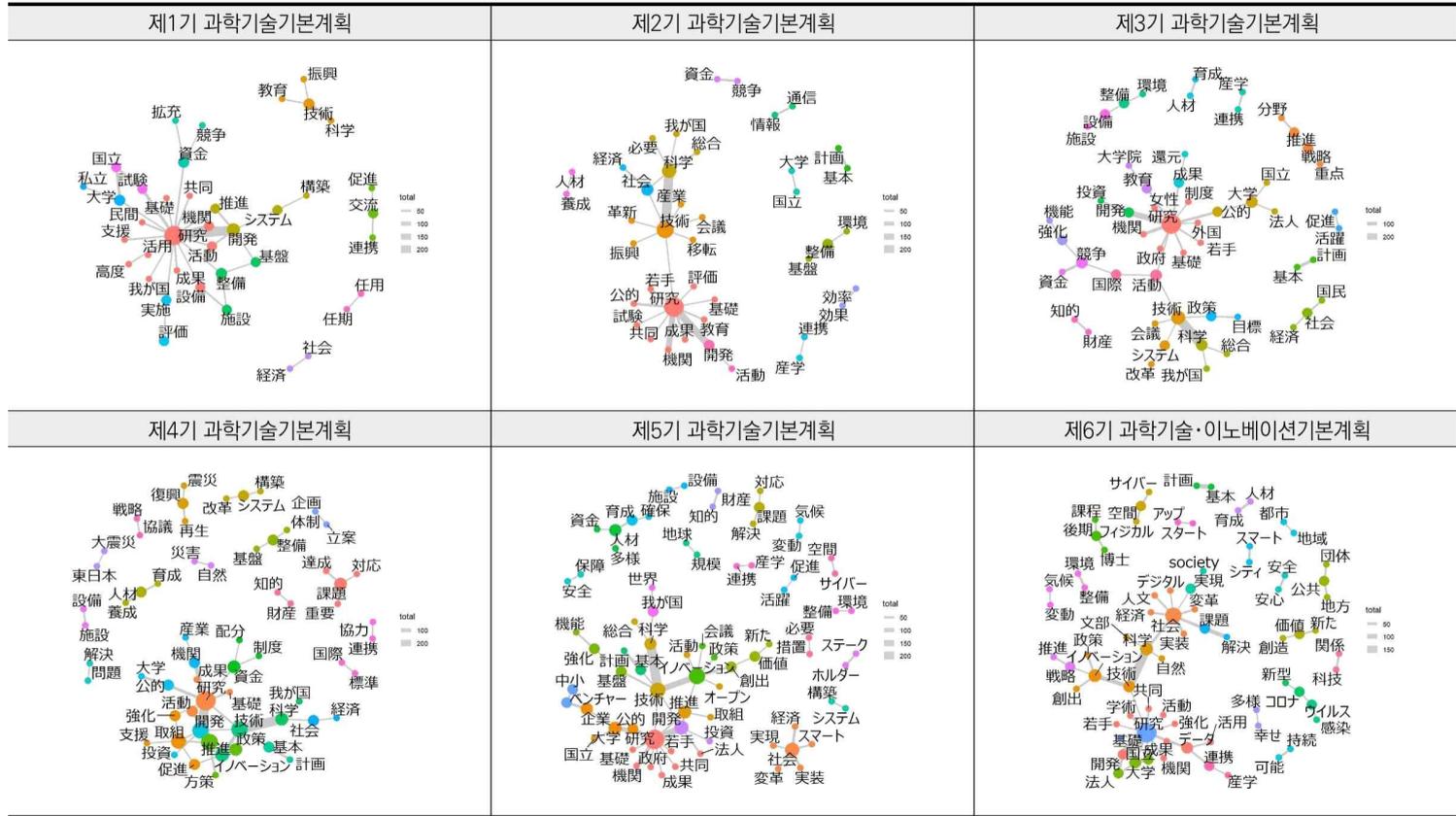
### 6) 제6기 과학기술·이노베이션기본계획의 중심성

제6기 과학기술·이노베이션기본계획에 대한 텍스트 네트워크 분석 결과를 <표 7>로 정리하였다. 제6기 과학기술·이노베이션기본계획에서 ‘사회’의 중심성은 매우 높았다(연결, 매개, 근접중심성 2위). 다음으로 ‘데이터(データ)’와 ‘제휴(連携)’도 중심성이 높게 나타났다. 시각화 그림을 보면, 제6기 과학기술·이노베이션기본계획에서는 ‘사회’가 중요한 위치를 점하고 있다. ‘사회’의 인접노드로서 ‘문제-해결’, ‘경제’, ‘실현’, ‘개혁’, ‘인문’이 나타났다는 것을 고려하면, 과학기술·이노베이션기본계획에서 사회문제해결은 매우 중요한 정책 의제임을 알 수 있다. 이 밖에도 ‘신형’, ‘코로나’, ‘바이러스’, ‘감염’ 등 ‘코로나바이러스감염증’ 같은 사회적 이슈도 주변 노드로 등장했다.

<표 7> 제6기 과학기술·이노베이션기본계획의 중심성

순위	연결중심성		매개중심성		근접중심성	
1	연구 研究	405	연구 研究	855984	연구 研究	0.000204
2	사회 社会	294	사회 社会	495486	사회 社会	0.000197
3	대학 大学	223	대학 大学	343221	활용 活用	0.000190
4	활용 活用	207	활용 活用	263802	대학 大学	0.000187
5	일본 我が国	184	일본 我が国	224846	추진 推進	0.000186
6	추진 推進	177	실현 実現	206070	일본 我が国	0.000185
7	실현 実現	174	기술 技術	200945	실현 実現	0.000185
8	데이터 データ	171	세계 世界	199315	대처 取組	0.000183
9	기술 技術	167	추진 推進	191330	데이터 データ	0.000182
10	세계 世界	161	데이터 データ	188885	제휴 連携	0.000182
11	대처 取組	159	대처 取組	154230	기술 技術	0.000182
12	과학 科学	145	과학 科学	149021	세계 世界	0.000181
13	국제 國際	139	국제 國際	142487	과학 科学	0.000180
14	연휴 (제휴) 連携	137	기업 企業	137813	정비 整備	0.000180
15	전략 戦略	132	정보 情報	132969	촉진 促進	0.000178

〈표 8〉 텍스트 네트워크 분석 시각화



## 2. 텍스트 임베딩 모델

본 연구에서는 데이터 전처리를 통해 추출된 단어들을 대상으로 텍스트 임베딩 모델을 적용하여 다차원 벡터 공간에 표현하였다. 텍스트 임베딩 모델 중에 Skip-gram 알고리즘은 중심단어를 기반으로 주변단어를 예측하는 메커니즘을 갖추고 있어 텍스트의 복잡한 의미 구조를 분석하는데 유용하다.

일본 과학기술기본계획 간의 상호관계를 파악하기 위해 코사인 유사도를 활용한 피어슨 상관분석을 수행하였다. 피어슨 상관계수와 유의확률을 <표 9>에 정리하였다. 분석결과, 모든 과학기술기본계획 간에는 통계적으로 유의미한 양(+)의 상관관계가 나타났다. 특히, 초기 계획인 제1기에서 제4기까지의 과학기술기본계획 간에는 다소 높은 상관관계가 확인되었다. 그러나 제4기에서 제5기, 그리고 제5기에서 제6기로 갱신될 때 상관계수는 각각 0.393, 0.257로 감소하였다. 뿐만 아니라, 제6기 과학기술·이노베이션기본계획은 이전 기본계획과 낮은 상관관계를 나타내어, 기본계획 간의 구조적 변화를 추적하는 데에 도움이 될 수 있음을 시사한다.

<표 9> 피어슨 상관계수

ITEM	제1기	제2기	제3기	제4기	제5기	제6기
제1기 과학기술기본계획	1					
제2기 과학기술기본계획	.436***	1				
제3기 과학기술기본계획	.405***	.506***	1			
제4기 과학기술기본계획	.327***	.438***	.457***	1		
제5기 과학기술기본계획	.294***	.378***	.367***	.393***	1	
제6기 과학기술·이노베이션기본계획	.175***	.176***	.198***	.222***	.257***	1

\*\*\*: p<0.001 \*\*: p<0.01 \*: p<0.05

제4기에서 제5기 과학기술기본계획으로 갱신될 때 상관계수가 다소 감소한 원인과 6기 과학기술·이노베이션기본계획이 이전 기본계획과 낮은 상관관계를 나타내는 원인을 고찰하면, 텍스트 네트워크 분석에서 확인한 것과 유사하게, 2016년 제5기 과학기술기본계획에서 도입된 ‘Society 5.0’이라는 미래 사회 모델이 영향을 끼친 것으로 추론된다.

텍스트 임베딩 모델을 활용하여 다차원 벡터 공간에 단어를 임베딩한 후, 코사인 유사도를 통해 두 벡터 간의 유사도를 계산하였다. 이에 대한 결과를 <표10>에 정리하였다. 분석결과, ‘사회’와 관련된 텍스트 네트워크 구조가 제5기 과학기술기본계획 이전의 계획과는 상당한 차이를 보임을 확인할 수 있었다. 특히, ‘AI’, ‘IOT’, ‘ICT’, ‘사이버’와 같은 정보통신기

술 관련 단어들은 제1기부터 제4기까지의 계획에서는 ‘사회’와는 두 벡터가 가리키는 방향이 유사하지 않게 나타났으나, 제5기 이후로는 코사인 유사도가 높아진 것으로 나타났다. 특히, 제6기에서는 전반적인 코사인 유사도 값이 상승하였다. 이를 통해 일본의 최근 과학 기술기본계획은 ‘사회’와 정보통신기술의 조화를 강조하는 경향이 있음을 추론할 수 있다. 또한, ‘society’, ‘라이프사이클’, ‘라이프스타일’과 같이 사회생활과 관련된 단어들도 제1기부터 제4기까지의 계획에서는 ‘사회’와의 유사도가 없었으나, 제5기와 제6기에서는 유사도가 상승한 것으로 나타났다. 이러한 결과는 일본 정부가 제시한 ‘Society 5.0’이 가상공간과 현실사회를 고도로 융합시키는 시스템으로 경제발전과 사회적 과제해결을 양립하는 인간 중심 사회를 추구하고 있음을 반영한다.

〈표 10〉 코사인 유사도

단어	제1기	제2기	제3기	제4기	제5기	제6기
society	-	-	-	-	0.797	0.825
延長線(연장선)	-	-	-	-	0.241	0.597
浸透(침투)	-	-	-	-	0.081	0.417
ライフサイクル(라이프사이클)	-	-	-	-	0.125	0.394
サイバー(사이버)	-	-	-	-	0.234	0.376
進化(진화)	-	-	-	-	0.113	0.337
共感(공감)	-	-	-	-	0.022	0.334
適応(적응)	-	-	-	-	0.220	0.231
ビジネス(사업)	-	-	-	-	0.062	0.210
AI	-	-	-	-	0.079	0.210
IOT	-	-	-	-	0.298	0.209
概念(개념)	-	-	-	-	0.044	0.208
試行(시행)	-	-	-	-	0.219	0.204
ICT	-	-	-	-	0.163	0.204
アジェンダ(아젠다)	-	-	-	-	0.131	0.171
取り込み(캡처)	-	-	-	-	0.044	0.153
実装(실장)	-	-	-	-	0.374	0.147
規範(규범)	-	-	-	-	0.164	0.105
スピード(스피드)	-	-	-	-	0.339	0.098
ライフスタイル(라이프스타일)	-	-	-	-	0.410	0.093

## V. 결론

본 논문의 목적은 일본의 ‘과학기술기본계획’과 ‘과학기술·이노베이션기본계획’에 기록된 텍스트를 대상으로 데이터 전처리 방법을 이용하여 분석 가능한 형태의 데이터로 변환한 후, 텍스트 네트워크 분석과 임베딩 모델을 활용하여 텍스트 네트워크의 특징과 주요 정책 의제가 시기별로 변화하고 있음을 체계적으로 파악하는 것이다. 본 논문에서 사용된 분석 절차를 통해 일본 과학기술정책 텍스트에 대한 텍스트 분석을 심층적으로 수행하였으며, 이 과정에서 의미 있는 내용이 도출되었다.

첫째, 일본 과학기술정책의 텍스트 네트워크 구조와 정책의제 변화를 고찰하면, 중심성이 높은 주요의제가 시기별로 변화하고 있음을 확인할 수 있다(〈표 11〉).

〈표 11〉 일본 과학기술정책의 정책의제 키워드 변화

연도	분석대상 자료	주요 정책의제 키워드
1996~2000년	제1기 과학기술기본계획	연구, 개발
2001~2005년	제2기 과학기술기본계획	과학, 기술
2006~2010년	제3기 과학기술기본계획	여성, 경쟁
2011~2015년	제4기 과학기술기본계획	이노베이션, 동일본대지진
2016~2020년	제5기 과학기술기본계획	사회, 중소기업, 이노베이션
2021~2025년	제6기 과학기술·이노베이션기본계획	사회, 코로나바이러스감염증

텍스트 네트워크 분석 결과를 각 과학기술기본계획별로 세분화하여 살펴보면, 해당 계획들이 채택한 주요 의제를 파악할 수 있다. 제1기에서는 ‘연구’와 ‘개발’ 간의 강한 연결성이 확인되지만, 중심성을 갖춘 특정 노드는 찾아보기 어렵다. 초기 계획에서는 주로 일반적인 과학기술 연구 및 개발에 주력했음을 시사한다. 제2기에서는 ‘연구’에 더해 ‘기술’이 강한 중심성을 보이며, 이와 연관된 다수의 인접노드가 관찰된다. 이러한 패턴은 제2기 이후로 과학과 기술에 대한 구체적인 정책이 강조되었다는 것을 나타내며, 연구와 기술의 통합적인 발전을 시사한다. 제3기에서는 ‘경쟁’과 ‘여성’이 주요 정책의제로 부각된다. 특히 ‘여성’과 관련된 노드가 돋보이며, 해당 주제의 중심성이 높게 나타났다. 이는 과학기술분야에서 성평등에 대한 관심이 높아지고 있으며, 특히 여성 참여를 강조하는 정책의 중요성이 부각되고 있다는 것을 의미한다. 제4기에서는 ‘이노베이션’이 강한 중심성을 보인다. 이는 제3기 이후로 ‘이노베이션’과 관련된 정책의 중요성이 증가하고 있음을 시사한다. 또한, ‘동일본대지진’이 주요 정책의제로 등장하는데, 이는 지역적·사회적 이슈가 일본 과학기술기본계획

방향성에 미치는 영향을 반영한다.

제5기에서는 ‘이노베이션’의 중심성 순위가 제4기에 비해 상승하였고, ‘기업’, ‘중소’, ‘벤처’ 등 ‘중소기업’과 관련된 내용이 주요 정책의제로 강조되었다. 특히 ‘기업’의 중심성이 높아지면서 ‘중소기업’과 관련된 내용이 주요 정책의제로 부상했다. 제5기의 중심성에서는 이전 기본계획에 비해 ‘사회’의 중심성이 상승하였으며, 인접노드로는 ‘스마트’, ‘변혁’, ‘실장’ 등이 등장하였다. 이러한 변화는 제5기 과학기술기본계획을 통해 제시된 ‘Society 5.0’의 영향으로 해석된다. 제5기에서 처음으로 제시된 ‘Society 5.0’를 계승한 제6기 과학기술·이노베이션기본계획에서도 ‘사회’의 중심성은 매우 높게 나타났다(연결, 매개, 근접중심성 2위). 또한, ‘사회’의 인접노드로는 ‘문제-해결’, ‘경제’, ‘실현’, ‘개혁’ 등이 등장하였으며, 이를 통해 사회문제해결이 매우 중요한 정책과제임을 확인할 수 있다. 시각화 그림에서는 ‘신형’, ‘코로나’, ‘바이러스’, ‘감염’ 등과 같은 사회적 이슈도 주변노드로 나타났다. 이러한 결과를 통해 향후 일본 정부는 현재 시대의 사회적 상황에 대응하고자 과학기술정책에서 ‘사회’와 관련된 국가전략을 마련하여 집중적으로 추진할 것을 예상할 수 있다.

둘째, 텍스트 임베딩 모델을 사용하여 분석을 진행했다. 이 분석에서는 단어들을 다차원 벡터로 임베딩하고, 이를 기반으로 코사인 유사도를 계산하여 단어 간의 유사성을 측정했다. 이 결과를 정리한 <표 11>에서는 ‘사회’와의 관련성이 시간에 따라 어떻게 변하는지를 확인했다. 분석결과, ‘AI’, ‘IOT’, ‘ICT’, ‘사이버’ 등 정보통신기술 관련 용어들은 최근 과학기술기본계획에서는 ‘사회’와의 유사도가 높아진 것으로 나타났다. 특히, 제5기 이후에서는 정보통신기술 용어들과 ‘사회’ 간의 유사성이 상승하는 경향을 보였다. 또한, ‘society’, ‘라이프사이클’, ‘라이프스타일’과 같이 사회생활과 관련된 단어들도 최근 기본계획에서는 ‘사회’와의 유사성이 높아진 것으로 확인되었다. 이러한 결과를 통해 일본의 최근 과학기술기본계획이 ‘사회’와 정보통신기술의 조화를 강조하고, ‘Society 5.0’ 개념을 추구하고 있음을 정량적으로 확인할 수 있었다. 텍스트 임베딩 모델을 활용한 분석은 과학기술기본계획의 변화를 자세하게 이해하는 데 도움이 되었다.

본 연구는 일본 ‘과학기술기본계획’과 ‘과학기술·이노베이션기본계획’에 텍스트 네트워크 분석과 임베딩 모델을 적용하여 분석한 것이다. 이러한 유형의 분석에 수반되는 한계점과 향후 추가로 고려해야 할 연구과제는 다음과 같다.

첫째, 본 연구는 일본 ‘과학기술기본계획’과 ‘과학기술·이노베이션기본계획’의 본문만을 대상으로 하고 있다. ‘과학기술·이노베이션기본계획’에는 본문 이외에도 ‘지금까지의 주요경위’, ‘과학기술·이노베이션기본계획주요지표·참고지표데이터집’, ‘종합과학기술·이노베이션회의의사요지’ 등의 자료들이 있다. 따라서 본문뿐만 아니라 관련 자료도 고려하

여 분석을 하면, 보다 심층적인 분석이 가능할 것이다.

둘째, 본 연구는 연결중심성, 매개중심성, 근접중심성에 초점을 맞추어 진행되었다. 중심성 척도에는 연결중심성, 매개중심성, 근접중심성 이외에도, 고유벡터중심성, 페이지랭크 중심성(PageRank) 등 다양한 중심성 척도가 있다. 본 연구는 노드의 연결 정도와 근접 정도에 초점을 맞췄기 때문에 연결중심성, 매개중심성, 근접중심성을 중점으로 분석하였으나, 다른 관점에서 중심성을 측정할 수 있을 것이다.

## 참고문헌

- 고기동. (2019). 아이디어와 정책선택에 관한 경험적 연구: 경제정책의 변화를 중심으로. 『행정논총』, 57(3): 89-115.
- 고기동·이은미. (2022). 코로나19 위기와 정책아이디어 전환: 정책아이디어 변화 궤적과 탈경로에 관한 비교 분석. 『행정논총』, 60(1): 1-25.
- 고길근. (2007). 정책네트워크 연구의 유용성과 사회연결망 이론 활용 방법의 고찰. 『행정논총』, 45(1): 138-164.
- 김민호·윤호열·최상욱. (2021). 인공지능망 기반 자연어처리를 적용한 연도별 정책내용 변화 분석에 관한 연구: 일본 IT신전략 (2018-2020)을 대상으로. 『정보통신정책연구』, 28(1), 1-27.
- 김옥일. (2008). 교육행정정보화 정책네트워크 구조 변화에 관한 연구: 사회연결망 분석의 적용. 『행정논총』, 46(2): 255-279.
- 김찬우. (2019). 국민 청원 데이터를 통해 본 주요 개혁 이슈. 『예술인문사회 융합 멀티미디어 논문지』, 9(2): 823-832.
- 김학실. (2012). 여성정책변동과 연구경향 분석: 네트워크 텍스트 분석을 중심으로. 『한국행정학보』, 46(2): 241-264.
- 김희수·김재현·황인서·이동훈·최원석. (2021). 키워드 네트워크 분석을 통한 항공우주산업 분야 과학기술의 연구 동향 분석. 『한국항공경영학회지』, 19(4): 27-45.
- 김희철·한수미. (2020). 온라인 클라우드소싱 플랫폼 데이터의 텍스트 마이닝을 통한 도시계획 정책 이슈 도출. 『기술혁신학회지』, 23(6): 1159-1182.
- 남성욱·백연주. (2022). 북한의 ICT 발전전략의 미래경향성에 관한 연구: 텍스트마이닝 기법을 중심으로. 『평화학연구』, 23(3): 67-93.
- 박주섭·홍순구. (2016). 텍스트마이닝 기법을 활용한 부산시 지역혁신정책 동향분석. 『지방정부연구』, 20(1): 1-20
- 박주섭·홍순구·김종원. (2017). 토픽모델링을 활용한 과학기술동향 및 예측에 관한 연구. 『한국산업

- 정보학회논문지», 22(4): 19-28.
- 박주섭·김나량·한은정. (2018). 키워드 네트워크 분석을 활용한 과학기술동향 분석. 한국산업정보학회논문지», 23(2): 63-73.
- 박치성·신나리. (2021). 토픽모델링을 활용한 역대 대통령의 정책변화 분석. 「한국정책과학학회보」, 25(4): 1-33.
- 신승윤·최한별·김동욱. (2021). 지방정부의 공공데이터 내용 분석: 텍스트 마이닝을 통한 분석결과를 중심으로. 「한국행정연구」, 30(2): 129-171.
- 연지현·김성원. (2022). 텍스트 마이닝을 활용한 경제정책기록서비스 연구: 경제정책방향을 중심으로. 「한국기록관리학회지」, 22(2): 117-133.
- 윤호열·김민호·최상욱·박근우. (2020). NTIS 국가연구개발사업 정보를 활용한 지역 과학기술 연구동향 분석: 충남 지역 NTIS 데이터를 중심으로. 「기술경영」, 5(4): 21-44.
- 이가영. (2020). 김정은 시기 북한 경제정책의 변화경향성 연구: 텍스트마이닝 분석을 중심으로. 「동북아경제연구」, 32(1): 27-54.
- 이수정·박광국. (2022). 음악공연권 정책변동과정에 대한 시론적 연구: 옹호연합모형(ACF)과 사회연결망분석(SNA)의 적용. 「행정논총」, 60(1): 133-166.
- 이은미·김동욱·고기동. (2016). 비정형 데이터 분석의 제도변화 적용에 관한 연구: 개인정보보호 제도를 중심으로. 「한국정책과학학회보」, 20(2): 217-240.
- 이주영·정효정. (2017). 한국 과학기술계 기술혁신 논의의 흐름과 변화: 한국과학기술단체총연합회의 『과학과 기술』을 중심으로, 1968-2017. 「기술혁신학회지」, 20(4): 1015-1035.
- 임강희·강경환. (2023). 텍스트마이닝을 활용한 중국의 핵심 과학기술 확보 동향 연구. 「한국산학기술학회논문지」, 24(4): 228-233.
- 임강희·최현주·강경환. (2023). 중국공산당 주요 보고서 분석을 통한 중국 과학기술 발전방향 연구: 텍스트마이닝 기법을 활용하여. 「한국산학기술학회논문지」, 24(7): 317-324.
- 임봉조. (2022). 텍스트 네트워크 분석(TNA) 방법을 활용한 초등돌봄교실 연구동향 분석. 「종교교육학연구」, 70: 47-67.
- 황한찬·엄석진. (2018). 한국의 고도성장기의 산업정책 네트워크 분석: 수출진흥확대회의와 월간경제동향보고회의를 중심으로. 「한국행정학보」, 52(4): 557-588.
- 황한찬·엄석진. (2019). 소셜미디어 활용을 통한 지방행정의 반응성 개선: 성남시의 트위터 활용 사례를 중심으로. 「한국행정학보」, 53(4): 243-273.
- 황한찬·엄석진. (2021). 지방정부에서 전자청원 시스템의 효과성: 서울시 전자청원과 서면 청원의 비교를 중심으로. 「한국행정학보」, 55(3): 321-350.
- Alves, Ana C., & Lee, Su-Hyun. (2022). China's BRI developmental agency in its own words: A content analysis of key policy documents. *World Development*, 150: 1-13.
- Bavelas, Alex. (1950). Communication Patterns in Task-Oriented Groups. *The Journal of the Acoustical Society of America*, 22(6): 725-730.

- Bengtsson, Mariette. (2016). How to plan and perform a qualitative study using content analysis. *NursingPlus Open*, 2: 8-14.
- Guo, Meng., Li, Qingyang., Wu, Chenyang., Vine, Scott L., & Ren, Gang. (2023). Content analysis of Chinese cities' Five-Year Plan transport policy documents. *Case Studies on Transport Policy*, 13: 1-13.
- Herrero-Solana, Víctor., & Faba-Pérez, Cristina. (2023). European Patent Office (EPO) en Twitter: A content analysis. *World Patent Information*, 74: 1-5.
- Liao, Zhongju. (2018). Content analysis of China's environmental policy instruments on promoting firms' environmental innovation. *Environmental Science & Policy*, 88: 46-51.
- Peng, Huatao., & Liu, Yang. (2016). A comprehensive analysis of cleaner production policies in China. *Journal of Cleaner Production*, 135(1): 1138-1149.
- Saheb, Tahereh., & Saheb, Tayebbeh. (2023). Topical review of artificial intelligence national policies: A mixed method analysis. *Technology in Society*, 74: 1-14.
- Saraisky, Nancy G. (2015). Analyzing Public Discourse: Using Media Content Analysis to Understand the Policy Process. *Current Issues in Comparative Education*, 18(1): 26-41.
- Mikolov, Tomas., Chen, Kai., Corrado, Greg., & Dean, Jeffrey. (2013). Efficient Estimation of Word Representations in Vector Space, arXiv preprint arXiv:1301.3781, 1-12.
- Weber, Robert P. (1990). *Basic Content Analysis*, 2nd Ed. CA: SAGE Publications.
- 赤池伸一. (2019). 科学技術政策の歴史と今後の課題: - 次期科学技術基本計画の策定に向けて. *情報の科学と技術*, 69(8): 358-363.
- 研究開発戦略センター. (2022). *日本の科学技術・イノベーション政策*. 東京: 研究開発戦略センター.
- 小林信一・赤池伸一・林隆之・富澤宏之・調麻佐志・宮林正恭. (2019). 科学技術基本計画の変遷と次期の展望. *研究技術計画*, 34(3): 190-215.
- 内閣府. (1996). 第1期科学技術基本計画 本文.
- 内閣府. (2001). 第2期科学技術基本計画 本文.
- 内閣府. (2006). 第3期科学技術基本計画 本文.
- 内閣府. (2011). 第4期科学技術基本計画 本文.
- 内閣府. (2016). 第5期科学技術基本計画 本文.
- 内閣府. (2021). 第6期科学技術・イノベーション基本計画 本文.
- 水間英城. (1996). 科学技術基本法について. *情報管理*, 38(12): 1080-1093.
- 科学技術基本法. (공포 1995.11.5., 법률 제130호).
- 科学技術基本法等の一部を改正する法律. (공포 2020.6.24., 법률 제63호).
- 科学技術・イノベーション基本法. (개정 2020.6.24., 법률 제130호).
- 内閣府設置法. (공포 1999.7.16., 법률 제89호).
- 일본 내각부정책 과학기술·이노베이션(<https://www8.cao.go.jp/cstp/>).

## ABSTRACT

### Analysis of Changes in Japan's Science and Technology Policy Agenda Using Text Network Analysis and Text Embedding

Minho Kim & YoungJun Kim

The Science and Technology Basic Law in Japan establishes the fundamental framework for science and technology policy. Japan formulates Science and Technology Basic Plan in accordance with this law. The Science and Technology Basic Plan is a policy text that allows for a macroscopic understanding of the major keywords and changes in Japan's science and technology policy. The purpose of this paper is to explain the text network structure of Japan's Science and Technology Basic Plan using text network analysis and text embedding models, as well as to analyze changes in the policy agenda. Japan's science and technology policy documents written in natural language are constructed as text networks in this paper, and centrality and cosine similarity are calculated.

It can be observed that important policy agendas are changing over time when analyzing the text network structure in Japan's science and technology policy. This indicates the diversity of key policy issues in each period. Moreover, terms related to information and communication technology in recent basic plans have shown increased cosine similarity to 'society.' This result confirms that Japan's recent Science and Technology Basic Plan emphasizes the harmony between 'society' and information and communication technology, pursuing the concept of 'Society 5.0.'

【Keywords: Japan's science and technology policy, text network analysis, science and technology basic plan, text embedding model, natural language processing】