

# 국민청원글의 토픽 모델링을 통한 교육이슈 분석

심재권

고려대학교 영재교육원

## 요약

교육과 관련된 이슈는 다양한 집단과 상황이 서로 복잡하게 연계된 사회문제로 교육과 관련된 현상을 분석하여 이슈와 문제를 구체적으로 발견하는 것은 쉽지 않은 일이다. 한국어 기반 텍스트 분석은 정량적인 형태로 분석이 가능하고, 텍스트 분석기법의 발전에 따라 연구적인 성과를 내고 있어 교육과 관련된 이슈를 한국어 텍스트로 된 데이터에서 도출하는데 충분히 활용할 수 있다. 본 연구는 청와대 국민청원 홈페이지 게시판의 육아/교육 분야의 청원글을 수집하고 텍스트 분석방법을 활용하여 교육계의 이슈와 문제를 도출하고자 하였다. 분석은 토픽 모델링 기법 중 잠재 디리클레 할당(LDA)을 통해 6개 토픽을 도출하였고, 주요 키워드의 연관규칙을 분석하여 그래프로 시각화하였다. 기존의 설문을 통한 교육의 이슈를 도출하는 방법 이외에 추가로 텍스트 기반의 분석방법을 통해 이슈를 충분히 발견할 수 있다는 점에서 향후 연구의 방향과 정책에 시사점을 제공할 수 있다.

키워드 : 국민청원, 토픽 모델링, 잠재 디리클레 할당, 연관규칙, 교육이슈

## Analysis of Educational Issues through Topic Modeling of National Petitions Text

Jaekwoun Shim

Korea University Center for Gifted Education

## Abstract

Education related issues are social problems in which various groups and situations are intricately linked to each other. It is difficult to find issues by analyzing social phenomena related to education. Korean based text analysis can be analyzed in a quantitative. With the development of text analysis techniques, research results have been recently achieved, and it can be fully utilized to derive educational issues from text data in Korean. In this study, petition articles in the field of childcare/education were collected on the online-board of the Blue House National Petition website, and text analysis was used to derive issues in the education world. The analysis derived 6 topics through Latent Dirichlet Allocation(LDA) among topic modeling techniques. The association rules of major keywords were analyzed and visualized as graphs. In addition to deriving educational issues through the existing questionnaire, it can provide implications for future research directions and policies in that issues can be sufficiently discovered through text-based analysis methods.

Keywords : National Petitions, Topic Modeling, Latent Dirichlet Allocation, Association Rule, Education Issue

## 1. 서론

교육은 유아에서 대학생과 성인에 이르기까지 폭넓은 대상과 지식의 습득, 진로와 직업, 취미와 여가 등 다양한 내용으로 다면적인 사회적 맥락 속에서 진행되고 있다. ‘교육열’로 대표되는 교육의 문제는 사회적으로 중요한 이슈이지만, 다양한 집단과 상황이 서로 복잡하게 연계되어 있어 교육과 관련된 현상을 분석하고 문제를 발견하는 것은 어려운 일이다[3]. 하지만, 미래세대를 위해 교육과 관련된 이슈는 중요한 과제이고 문제를 해결하기 위해서는 현상을 구체적으로 분석할 필요가 있다.

사회적 이슈는 언론사와 여론조사를 통해 생산 및 확산되지만, 최근에는 IT기술을 활용하여 개인 미디어를 통해 이슈를 생산하고 유통하여 소비하는 방식으로 변화하였다[15]. 개인의 의견과 생각이 네트워크를 통해 생산과 재생산의 과정을 거쳐 확산되는 소셜 네트워크 사회의 도래는 사회적 이슈에 대한 복잡성이 기존보다 증대되었지만, 개인의 의견과 생각이 구체적인 데이터로 저장되어 컴퓨팅 환경에서 분석이 가능해졌다[5].

국민청원 홈페이지는 전 국민의 의견을 듣기 위한 창구로 2017년 8월에 청와대 홈페이지를 개편하면서 신설되었다[1]. 국민청원 홈페이지는 공개된 게시판에 청원 글을 작성하고 청원에 동의한 참여자 수에 따라 해당 청원과 관련된 부서에서 답변하는 의사소통 창구로 운영되고 있다. 국민청원은 개인이 구체적인 사안을 자유롭게 정부에 직접 요구할 수 있고, 누구나 참여하여 동의할 수 있다는 점에서 우리나라의 폭넓은 사회적인 이슈의 장이라고 할 수 있다. 즉, 언론사를 통해서 작성되는 뉴스에 비하여 국민 누구나 작성할 수 있는 국민청원글은 국민 개개인의 의견을 대변한다고 할 수 있다. 따라서, 우리나라 교육과 관련된 이슈를 도출하고 분석하기에는 국민청원 홈페이지의 육아와 교육관련 분야에 작성된 청원글이 현실 적합성이 높다고 할 수 있다.

국민청원 홈페이지의 청원글은 청원 제목과 내용이 한국어로 작성된 텍스트로 구성되어 이를 분석하기 위해서는 텍스트 처리를 통해 이슈를 도출하는 방안이 연구될 필요가 있다. 수집된 텍스트 문서들에서 등장하는 단어를 분석하여 의미를 발견하고 토픽을 도출하는 방법인 토픽 모델링(Topic Modeling)은 사회적 이슈를 도출하기에 적합한 방법이라고 할 수 있다[6][9]. 추가적으

로, 연관분석을 통해 텍스트 문서에 등장하는 키워드의 연관성을 분석하고 규칙을 시각화할 수 있다.

본 연구는 교육계의 쟁점과 이슈를 파악하고, 해당 이슈의 내부적인 연계 구조를 분석하여 연구의 방향과 교육정책의 시사점을 제공하고자 하였다. 이를 위해, 토픽 모델링 중에서 잠재 디리클레 할당(Latent Dirichlet Allocation, LDA)을 사용하여 국민청원 홈페이지의 육아와 교육 분야에 작성된 청원글을 분석하여 토픽을 도출하였고, 시각화하여 우리나라 교육의 핵심적인 이슈를 분석하였다.

## 2. 이론적 배경

### 2.1. 토픽 모델링

대용량의 텍스트에서 토픽을 모델링 하는 방법 중에 하나인 잠재 디리클레 할당(LDA)은 문서 집합 내에 잠재된 토픽을 도출하는 확률적 생성 모형이다[6][12]. 확률적 생성 모형은 확률 분포와 파라미터를 통해 잠재 변수(latent variable)를 발견하는 모형이다[6]. 즉, LDA는 관찰된 변수(observed variable)를 통해 숨겨진 변수(hidden variable)를 추론할 수 있고, 이를 문서 집합에 적용하여 토픽 비율과 각 토픽별 단어 분포를 예측할 수 있다.

### 2.2. 교육분야의 토픽 모델링 활용

교육분야에서 토픽 모델링을 적용한 연구를 살펴보면 다음과 같다. 강승지(2020)는 유아 사교육의 현상을 분석하기 위한 목적으로 국내 신문 기사, 블로그의 게시글과 댓글을 수집하여 토픽 모델링하였다[10]. 국내 신문 기사는 7개 토픽, 블로그의 게시글과 댓글에서는 6개의 토픽으로 구분되었고, 우리나라 교육정책의 핵심적인 토픽과 유아 사교육에서 엄마의 역할에 대한 정책적인 시사점을 도출하였다. 심재권(2020)은 교육관련 이슈도출을 목적으로 국민청원 홈페이지 청원글의 제목을 수집하여 토픽 모델링하였다[7]. 토픽 모델링을 통해 교육관련 법령, 대학입시, 교육 관련 범죄, 교육환경, 유치등교육, 교원처우, 교육정책, 고등교육, 중등교육 등의 9개 토픽을 도출하였다. 손태권(2021)는 초등수학교육 연구

의 방향을 분석하기 위한 목적으로 2001년에서 2020년까지 초등수학교육과 관련된 논문을 토픽 모델링하였다 [13]. 초등수학교육의 토픽은 분수, 통계교육, 수와 연산, 도형, 문제해결, 추론, 교과서 분석, 교육과정 분석, 교수 관행, 예비교사 교육, 영재교육, 오개념, 학생의 인지정 의적 요인, 평가, 효과성 검증의 15개로 분류되었다. 15 개의 토픽 중에 연구비중이 증가하는 토픽은 통계교육, 교과서 분석, 교육과정 분석이고, 감소하는 토픽은 문제 해결, 효과성 검증인 것으로 분석되어 향후 국내 초등수 학교육의 발전 방향과 후속 연구에 대한 시사점을 도출 하였다. 윤빛나리(2020)은 교육분야 중에서 마이스터고 에 따른 정책이슈와 특성을 분석하기 위한 목적으로 교 육부가 발표한 234건의 마이스터고 관련 보도자료를 토픽 모델링하였다[16]. 마이스터고 관련하여 9개의 토픽 을 도출하였고, 그 중에서 학생 지원 제도 및 성과 홍보 에 대한 토픽이 다수 등장한 것으로 분석되었다.

ICT 국가전략을 수립에 정책적 방향성을 지원하기 위한 목적으로 우창우(2020)는 국가연구개발사업의 ICT 분야의 연구과제를 토픽 모델링하였다[2]. 분석자료는 NTIS(National Science and Technology Information Service)로부터 최근 5년간 국가 연구개발사업의 전체 연구과제 정보를 다운로드 받아 정보통신기획평가원 (IITP)의 EZone 시스템과 매칭하여 ICT분야 연구과제 5,200건을 대상으로 하였다. ICT분야 연구과제에 대한 연구토픽은 인공지능, 빅데이터, 사물인터넷 등의 지능 정보기술에 대한 고도화 및 응용연구를 수행하는 것으 로 확인하였다.

토픽 모델링을 활용하는 교육분야의 연구를 종합해보 면, 교육분야와 관련된 현상을 파악하고 분석하기 위한 목적으로 텍스트 데이터를 수집하고, 토픽 모델링을 통 해 얻어진 토픽의 내재적 의미와 시계열적 변화를 분석 하여 연구의 방향과 정책적 시사점을 도출하는데 활용 할 수 있는 것으로 나타났다. 즉, 거시적인 관점에서 분 석이 충분히 가능하고, 키워드 분석, 시계열 분석 등의 추가적인 분석을 통해 현상에 대한 추가적인 정보를 제 공할 수 있다.

### 3. 연구방법

#### 3.1. 연구대상

본 연구는 한국교육의 핵심적인 이슈를 도출하고자 국민청원 홈페이지에서 육아와 교육 분야에 작성된 청 원글의 제목과 내용을 연구대상으로 선정하였다. 분석 은 데이터 수집, 데이터 전처리, 최적의 토픽 수 결정, 결정된 토픽 수에 따른 토픽 모델링 수행 순으로 진행 되었다. 추가로, 데이터 전처리를 통해 도출된 키워들 간의 연관성을 기반으로 네트워크 분석을 진행하였다.

#### 3.2. 데이터 수집

국민청원 홈페이지 게시판에 글을 작성시 입력하는 항목은 제목, 분야, 내용, 링크, 검색태그이고, 작성된 청 원글을 통해서 제목, 동의인원, 분야, 작성일, 마감일, 청 원인, 내용을 확인할 수 있다. 데이터의 수집은 육아와 교육 분야 게시판에 2021년 2월 14일까지 작성된 모든 청원글에 대해 웹크롤러를 사용하여 제목, 내용을 수집 하였다. 수집결과, 28,482개의 청원글을 수집하였고, 년 도별 청원글의 수는 <Table 1> 과 같다.

<Table 1> Research Subject

Year	2017	2018	2019	2020	2021	Total
Document	7,275	17,015	3,030	1,026	136	28,482

#### 3.3. 데이터 전처리

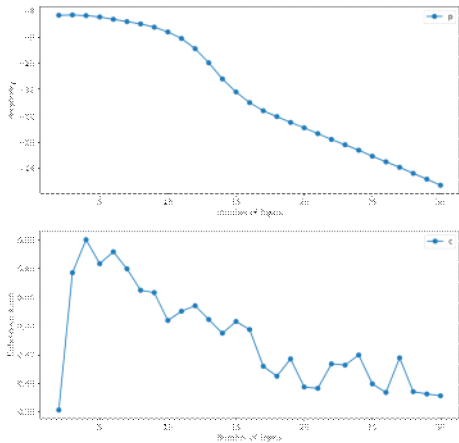
전처리는 수집한 텍스트 데이터에 포함된 특수문자와 문장부호 등을 제거하고, 분석에 사용할 단어를 선정하 기 위해서 카카오의 형태소 분석기 khaiii를 사용하여 보통명사, 고유명사, 의존명사, 대명사, 수사를 추출하여 최종적으로 총 91,609개 단어를 분석에 활용하였다.

#### 3.4. 최적의 토픽수 결정

토픽 모델링에서 토픽의 수를 적게 설정하면 하나의 토픽이 여러 개의 토픽을 포함하게 되고, 반대로 토픽의 수를 많이 설정하면 하나의 토픽이면 충분할 토픽이 다

수의 토픽으로 나뉘질 수 있어 적합한 토픽의 수를 결정하는 것은 중요하다. 일반적으로는 토픽의 수를 결정하기 위해 혼란도(perplexity)와 응집도(coherence) 지수를 참고하여 결정할 수 있다[4]. 혼란도는 도출된 토픽 모델이 결과를 정확하게 예측하는지를 나타내는 지수로, 값이 낮을수록 학습이 잘 되었다는 의미이다. 응집도 지수는 토픽의 일관성에 관한 값으로 토픽에 포함된 단어들 의미론적으로 유사한 정도를 의미한다[8][14].

본 연구에서는 토픽의 수를 2에서 30까지 1씩 증가시킨 모델에 대해 혼란도와 응집도 지수를 (Fig. 1)와 같이 모두 도출하였고, 토픽의 수를 최종적으로 6개로 설정하여 LDA 토픽 모델링을 수행하였다.



(Fig. 1) Result of Perplexity and Coherence

### 3.5. 토픽모델링

LDA 토픽 모델링은 python의 Gensim 패키지를 사용하였고, 디리클레 모수인  $\alpha$  는 'auto'로 설정하였고, 1000회 반복하여 훈련하여 결과를 도출하였다.

### 3.6. 연관 키워드 분석

문서에 함께 나타나는 키워드 분석은 연관규칙을 활용하기 위해 python의 apyori 패키지를 사용하였다. 분석은 모든 단어를 쌍으로 만들어 연관규칙을 만들어 지지도(0.01), 신뢰도(0.2), 향상도(3)를 설정하여 연관성이 높은 100개 규칙을 선정하여 그래프로 표현하였다.

## 4. 연구결과

### 4.1. 키워드 분석

국민청원 홈페이지에서 육아와 교육 분야에 작성된 청원글의 제목과 내용의 주요 키워드를 도출하기 위한 목적으로 TF-IDF 분석을 수행한 결과는 <Table 2>와 같다. 상위빈도 키워드를 살펴보면 아이, 학생, 교육, 학교, 교사, 폐지, 청소년, 어린이집, 유치원, 생각, 시간, 대학, 수능, 부모, 선생, 사람, 시험, 문제, 보육, 처벌 순으로 분석되어 교육과 육아와 교육에 관련된 키워드가 도출된 것으로 확인되었다.

<Table 2> Top 20 Keywords based on TF-IDF

Rank	Keyword	TF-IDF	Rank	Keyword	TF-IDF
1	child	756.04	11	time	391.07
2	student	635.46	12	university	383.77
3	education	619.34	13	SAT	364.27
4	school	592.25	14	parents	354.41
5	teacher	586.41	15	teach	349.42
6	abolition	558.49	16	person	320.35
7	teenager	531.82	17	exam	312.8
8	Daycare Center	525.67	18	problem	307.63
9	kindergarden	513.47	19	nursery	300.22
10	think	413.46	20	punishment	299.78

청원글이 작성된 년도별 주요 키워드를 TF-IDF로 분석한 결과는 <Table 3>와 같다. 분석결과, 1순위로 분석된 키워드는 아이 4회, 청소년 1회로 나타났다.

<Table 3> Top 10 Keywords based on TF-IDF by year

Rank	2017	2018	2019	2020	2021
1	teenager	child	child	child	child
2	abolition	student	student	starting school	teacher
3	law	teacher	school	student	abuse
4	child	education	education	class	notify
5	education	Daycare Center	kinder garten	Attending school	Jeong-In
6	Adolescents Protection	school	teacher	school	Child abuse
7	school	kinder garten	Daycare Center	teacher	children
8	student	time	university	Daycare Center	education

9	SAT	university	teaching	on-line	school
10	punishment	think	time	nursery	student

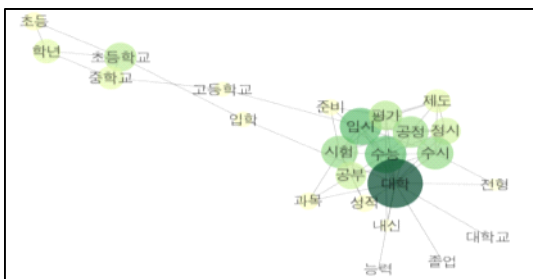
년도에 따른 주요 키워드의 TF-IDF 결과, 상위 키워드인 아이, 교육, 학생, 학교, 생각, 교사는 2017년에서 2021년의 분석기간 동안 높은 비율을 유지한 것으로 분석되었다.

#### 4.2. 연관 키워드 분석

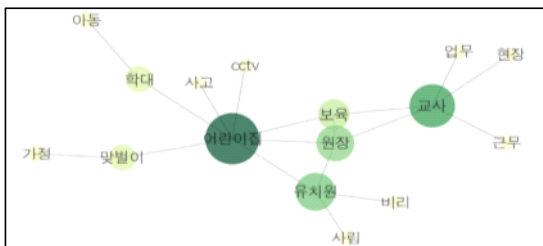
청원글에서 함께 등장하는 키워드 간의 연관성을 분석하기 위한 목적으로 연관규칙을 분석하여 그래프로 시각화하였다. 시각화는 키워드를 노드로 표현하고, 연관성이 높은 상위 100개의 규칙을 선으로 연결하였다. 연결된 선이 많을수록 진한 색으로 시각화하였고, 키워드가 10개 이상 연관된 그래프는 (Fig 2A, 2B)와 같다.

첫 번째 그래프(Fig. 2A)는 대학을 중심으로 수능, 입시, 시험, 공정, 수시, 공부, 평가 등의 대학입시와 관련된 키워드가 연관된 것으로 분석되었다.

두 번째 그래프(Fig. 2B)는 어린이집을 중심으로 유치원, 교사, 원장, 보육, 맞벌이 등 어린이집을 대상의 보육과 관련된 키워드가 등장하는 것으로 분석되었다.



(Fig. 2A) Related Keyword Analysis (University Entrance)



(Fig. 2B) Related Keyword Analysis (Kindergarten)

정리해보면, 교육의 주요이슈는 대학입시와 어린이집에 관한 이슈가 주요 키워드로 함께 등장하여 현재 우리나라의 교육계가 당면한 가장 중요한 이슈는 대학입시와 유아보육 문제인 것으로 해석할 수 있다.

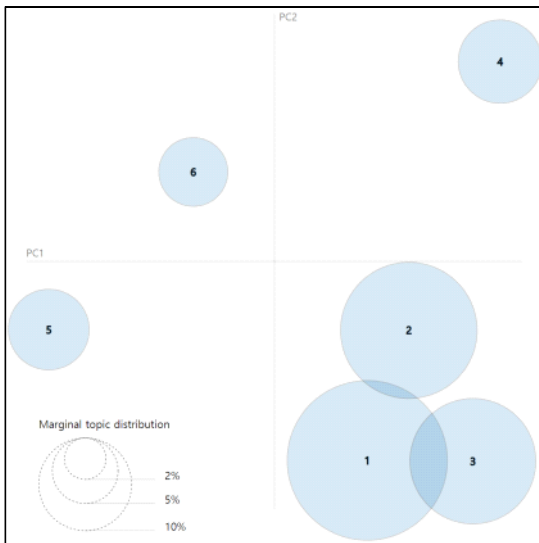
#### 4.3. 토픽 모델링

국민청원 홈페이지에서 육아와 교육 분야에 작성된 청원글의 제목과 내용을 대상으로 토픽 모델링하여 T1 (유아교육), T2(교육정책), T3(인사정책), T4(입시/임용), T5(청소년법), T6(학교폭력)로 분류되었고, 해당 토픽에 대한 기여도가 높은 순으로 키워드를 <Table 4> 과 같이 나열하였다.

<Table 4> Top 20 Keywords by Topic

Topic	Keywords
T1 (Early Childhood Education)	child, think, daycare center, parent, childcare, support, kindergarten, teacher, mom, time, us, president, family, favor, person, private, reality, maternity, childcare fee, mind
T2 (Education Policy)	abolition, student, education, thinking, school, class, society, studying, english, problem, person, boy, need, grade, time, president, high school, academy, confidence, private education
T3 (HR Policy)	teacher, education, fixed-term teacher, lecturer, regular, policy, full-time, transition, field, ministry of education, government, opposition, operation, work, full-time teacher, demand, national, increase, period
T4 (SAT/ Appointment Teacher)	teacher appointment, exam, SAT, university, evaluation, student, fair, institution, preparation, entrance examination, selection, college of education, pass, selection, examination, opportunity, study, problem, period, school record
T5 (Adolescents Protection Law)	adolescent, protection law, punishment, youth law, case, crime, protection, assault, victim, perpetrator, petition, person, adult, revision, middle school girls, abolition of law, violence, action, murder
T6 (School Violence)	school, youth law, teacher, violence, homeroom teacher, friend, student, school board, victim, perpetrator, classroom, blue house, respect, dispatch, victim student, female student, report, human rights, counseling, vice-principal

토픽의 년도별 전체 평균은 T1(유아교육, 33.3%), T2(교육정책, 24.2%), T3(인사정책, 17.0%), T5(청소년법, 12.2%), T4(입시/임용, 8.1%), T6(학교폭력, 4.8%) 순으로 분석되었다. 상위 3개의 토픽이 전체의 약 75%를 차지하는 것으로 나타났고, 각 토픽 간의 관련성을 시각화한 결과는 (Fig. 3)와 같다.



(Fig. 3) Visualize Relevance between Topics

시간의 흐름에 따라 토픽 분포를 <Table 5>는 연단위로, (Fig. 4)는 월단위로 시계열 분석하였다.

<Table 5> Percentage of Topic Change over the Year(%)

Topic	2017	2018	2019	2020	2021	Total
T1 (Early Childhood Education)	28.3	34.9	35.2	37.3	32.8	33.3
T2 (Education Policy)	23.0	24.9	24.5	21.1	13.1	24.2
T3 (HR policy)	14.9	17.5	17.9	20.8	15.9	17.0
T4 (SAT/ Appointment Teacher)	7.4	8.8	6.6	7.0	6.6	8.1
T5 (Edolescents Protection Law)	21.4	8.9	9.3	6.9	22.8	12.2
T6 (School Violence)	4.5	4.6	6.0	6.4	8.4	4.8

T1(유아교육)은 30%대의 높은 비율을 차지하는 것으로 분석되어 교육계에서 유아교육과 관련된 이슈가 일정부분 크게 차지하는 것으로 나타났다. T2(교육정책)과 T3(인사정책)은 매년 10~20%의 비율인 것으로 나타났고, T4(입시/임용)과 T6(학교폭력)는 10%대 미만의 분포인 것으로 분석되었다. T5(청소년법)은 이슈가 크게 되는 사건이 발생하는 경우 20%대의 높은 비율로 이슈가 발생하는 것으로 분석되었다.

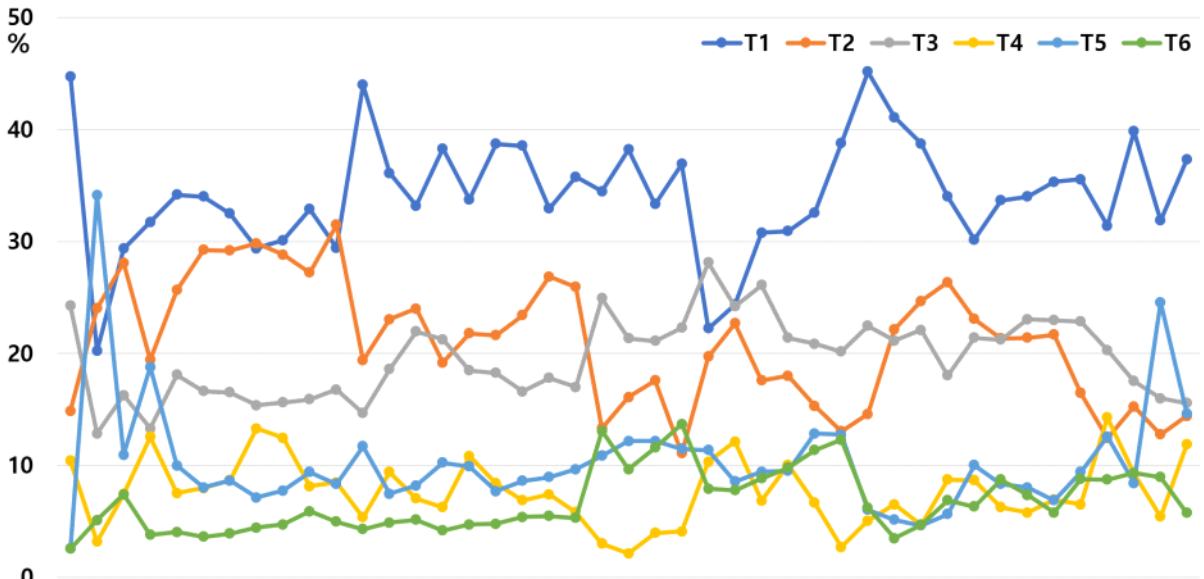
<Table 6> Regression Analysis by Topic

Topic	Regression coefficient	p	Trend
T1 (Early Childhood Education)	.197	.205	-
T2 (Education Policy)	-.465	.002	Fall
T3 (HR policy)	.340	.026	Rise
T4 (SAT/Appointment Teacher)	-.159	.309	-
T5 (Edolescents Protection Law)	-.078	.618	-
T6 (School Violence)	.449	.003	Rise

각 토픽의 월별 비율 값으로 회귀계수를 도출하여 양의 값이면 국민적인 관심이 증가되는 핫이슈(hot-issue)으로 분류하였고, 회귀계수가 음의 값이면 콜드이슈(cold-issue)으로 분류하여 <Table 6>과 같이 분석하였다[11]. 토픽 T3(인사정책)과 T6(학교폭력)은 유의수준 .05에서 상승하는 것으로 분석되었고, 토픽 T2(교육정책)은 하락하는 것으로 분석되었다. 즉, T3(인사정책)과 T6(학교폭력)의 비중은 증가하는 핫이슈이고 T2(교육정책)은 감소하는 콜드이슈라고 해석할 수 있다.

### 5. 결론

본 연구는 교육계의 쟁점과 이슈를 도출하고자 하는 목적으로 국민청원 홈페이지에 게시된 육아와 교육 분야의 청원글을 수집하여 잠재 디리클레 할당(LDA)로 분석하였다. 분석결과, 토픽은 T1(유아교육), T2(교육정책), T3(인사정책), T4(입시/임용), T5(청소년법), T6(학교폭력)로 구분되었고, T3(인사정책)과 T6(학교폭력)은



(Fig. 4) Percentage of Topic Change over the Month(%)

찾이슈이고 T2(교육정책)은 콜드이슈로 분석되었다. 주요 키워드들 간의 연관성을 분석한 결과, 우리나라 교육은 대학입시를 중심으로 구성된 네트워크와 유치원을 중심으로 유아교육과 관련된 네트워크가 형성되어 기존의 연구와 언론을 통해 얻은 여론과 유사한 것을 확인할 수 있었다.

연구를 통해 도출한 시사점은 다음과 같다. 첫째, 교육과 관련된 이슈는 교육계 내부의 독립된 관심사를 넘어서서 보다 폭넓은 사회의 여론에 크게 영향을 받는 것으로 분석되었다. 교육의 핵심 키워드라 할 수 있는 교육과정, 목표, 평가, 내용, 환경에 관한 키워드 보다는 해당 시점에 발생한 유아 혹은 청소년과 관련된 사회적 사건과 맞물려 이슈로 크게 부각된 것으로 분석되었다. 둘째, 육아와 교육분야의 이슈는 교육과 관련된 다양한 키워드가 복잡하게 연관된 것으로 분석되었다. 따라서 교육 문제는 교사, 학생, 학부모의 다양한 계층과 집단의 이해관계와 제도, 기관 등의 교육시스템을 고려한 다면적인 접근이 필요하다. 즉, 해당 사안에 대한 심도 있는 이해가 수반되어야 올바른 방향으로 해결책을 수립할 수 있다. 마지막으로 기존에 교육계에서 가지고 있었던 유아교육, 사교육, 대학입시, 교원임용, 학교폭력 등의 기존에 잘 알려진 문제 뿐 아니라 코로나19로 인해 발생한 온라인교육, 개학과 등교의 새로운 이슈가 발생하고 있어 이에 대한 대비가 필요함을 알 수 있다.

이러한 결과를 토대로 볼 때, 국가정책을 도출하거나 여론을 분석하기 위해서 충분한 데이터가 수집이 된다면, 본 연구에서 제안한 토픽 모델링 방식이 데이터가 충분하다면 하나의 방법이 될 수 있음을 보여주었다.

본 연구의 한계점은 첫째, 한국어 형태소 분석기 성능에 따라 토픽을 도출하는 결과가 영향을 받을 수 있다. 둘째, 분석에 사용한 데이터는 국민청원 홈페이지에 게시된 청원글로 우리나라 교육의 모든 내용을 포괄하지 못할 수도 있다는 한계가 있다. 향후 연구로는 잠재 디리클레 할당(LDA) 보다 발전되어 키워드 간의 상관을 고려한 토픽 모델링 기법인 상관 토픽 모델(Correlated Topic Model)을 활용할 필요가 있고, 분류된 토픽에 이름을 자동으로 라벨링하는 방식을 활용할 필요가 있다.

**참고문헌**

[1] Blue House National Petition Homepage [Internet]. <https://www1.president.go.kr/petitions/FAQ>

[2] C. W. Woo and J. Y. Lee (2020). Investigation of Research Topic and Trends of National ICT Research-Development Using the LDA Model. *Journal of the Korea Convergence Society, 11(7)*, 9-18.

- [3] C.D. Kang (2008). A Social Historical Study on the Korean Paranoid Educational Fervor and Status Desire. *The Korea Educational Review*, 14(2), 5-32.
- [4] D. Newman, J. H. Lau, K. Grieser, and T. Baldwin (2010). Automatic Evaluation of Topic Coherence. Human Language Technologies. *The 2010 Annual Conference of the North American Chapter of the ACL*, 100-108.
- [5] D.W. Yoon and H.J. Choe (2019). Analysis of the Core Concepts of Middle School Informatics Textbook Using Big Data Analysis Techniques. *Journal of Creative Information Culture*, 5(2), 157-164.
- [6] David M. Blei, Andrew Y. Ng and Michael I. Jordan (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993-1022.
- [7] J. W. Min, and J. K. Shim (2020). A Study on Analysis of National Petition Data for Deriving Current Issues in Education. *Journal of Creative Information Culture*, 6(2), 57-64.
- [8] Jonathan Chang, Jordan Boyd-Graber, Sean Gerrish, Chong Wang and David M. Blei (2009). Reading Tea Leaves: How Humans Interpret Topic Models. *Advances in Neural Information Processing Systems*, 22, 288-296.
- [9] M. Steyvers, and T. Griffiths (2007). *Probabilistic topic models*, Handbook of latent semantic analysis, Lawrence Erlbaum Associates Publishers.
- [10] S. J. Kang and Y. J. Shon (2020). Phenomenon of Early Childhood Private Education through Topic Modeling Analysis: Focusing on Domestic Newspaper Articles and Blogs. *Journal of Future Early Childhood Education*, 27(1), 177-199.
- [11] S. Y. Kim (2020). Analysis of Research Trends in Journal of the Korean Society for Industrial and Applied Mathematics Using Topic Modeling and Implications for Industrial Mathematics Education. *Secondary Education Research*, 68(2), 267-293.
- [12] T. K. Kim, H. R. Choi and H. C. Lee (2016). A Study on the Research Trends in Fintech using Topic Modeling. *Journal of the Korea Academia-Industrial cooperation Society*, 17(11), 670-681.
- [13] T. K. Son, and S. H. Hwang (2020). Analysis of the Research Trends of Domestic Elementary Mathematics Education Using Topic Modeling. *Journal of Elementary Mathematics Education in Korea*, 25(1), 61-80.
- [14] W.J. Choi, J.W. Seol, H.S. Jeong, and H.M. Yoon (2018). Comparison and Analysis of Subject Classification for Domestic Research Data. *The Journal of the Korea Contents Association*, 18(8), 178-186.
- [15] Y. Kim, N.K. Kim and S.J. Lee (2021). A Study on the Influence of Communication Characteristics of One-Person Media on Intention to Contents Acceptance: Focusing on the Mediating Effect of Parasocial Interaction. *The Korean Journal of Advertising*, 32(2), 163-188.
- [16] Yon, B.N (2020). A Topic Modeling Analysis on the Policy Issues of Meister High School. *Journal of Vocational Education & Training*, 23(1), 39-67.

### 저자소개

#### 심재권



2007 경인교육대학교 컴퓨터교육과 (교육학사)  
 2012 고려대학교 컴퓨터교육학과 (이학석사)  
 2017 고려대학교 컴퓨터교육학과 (이학박사)  
 2017~현재 고려대학교 연구교수  
 관심분야: 컴퓨터교육, 프로그래밍 교육, 온라인교육  
 e-mail: jaekwoun.shim@gmail.com