

토픽 모델링을 활용한 국내 초등 정보교육 연구동향 분석

심재권

고려대학교 영재교육원

요약

초등학생을 대상으로 인공지능교육에 대한 관심이 증대되면서 기존에 수행된 초등 정보교육 연구를 거시적인 관점에서 분석하여 현재의 상황을 이해하고 후속연구의 시사점을 제공하기 위한 노력이 필요한 시점이라 할 수 있다. 본 연구는 우리나라 초등 정보교육의 연구동향을 조망하고자 하는 목적으로 정보교육학회논문지를 분석하였다. 분석을 위한 데이터는 정보교육학회논문지의 창간호에서 2020년까지 출간된 논문을 모두 선정하였고, 토픽 모델링하여 연구주제 11개를 도출하였다. 연구결과, 가장 높은 비중인 토픽 T1은 약 38%를 차지하는 것으로 분석되었고, 토픽 T1에 기여도 순에 따라 교육, 연구, 분석, 초등, 정보의 키워드가 도출었다. 토픽들의 연도별 회귀분석 결과, 연구의 트렌드가 컴퓨팅사고력, 소프트웨어교육, 인공지능교육 등으로 변화하고 있는 것으로 나타났다. 본 연구의 의의는 초등 정보교육과 관련된 텍스트 데이터를 객관적으로 클러스터링하였다는 점에서 의미가 있다고 할 수 있다.

키워드 : 연구동향분석, 토픽모델링, 잠재디리클레할당, 초등정보교육

Analysis of Research Trends in Elementary Information Education in Korea using Topic Modeling

Jaekwoun Shim

Korea University Center for Gifted Education

Abstract

As interest in artificial intelligence education for elementary school students has recently increased, it is necessary to analyze the existing elementary information education research from a macroscopic point of view to understand the current situation and to provide implications for subsequent research. This study analyzed Journal of The Korean Association of Information Education for the purpose of looking at the research trend of elementary information education in Korea. For the data of the study, all papers published until 2020 in the first issue of the journal were selected, and 11 research topics were derived by modeling topics. As a result of the study, topic T1, the highest proportion, was analyzed to account for about 38%, and keywords such as education, research, analysis, elementary school, and information were derived according to the order of contribution to topic T1. As a result of regression analysis according to the year of the topic, it was found that the research trend is changing to computing thinking, software education, and artificial intelligence education. The significance of this study is that text data related to elementary information education is objectively clustered.

Keywords : Research Trend Analysis, Topic Modeling, LDA, Elementary Information Education

이 논문은 2020년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(NRF-2020R111A1A01058353).

논문투고 : 2021-01-18

논문심사 : 2021-01-18

심사완료 : 2021-01-30

1. 서론

새로운 기술의 파급력과 정도에 따라 교육에 미치는 정도는 상이하지만 기술의 변천은 자연스럽게 교육에 영향을 미치고 있다. 최근에는 인공지능 기술의 발전이 산업, 사회, 문화에 영향을 미치게 됨에 따라 초중등 교육에서도 인공지능을 소양의 차원에서 진로와 직업, 학업에 이르기까지 폭넓게 변화가 요구되고 있다[9]. 교육에 영향을 끼친 기술을 살펴보면, 인공지능 이전에는 소프트웨어가 미치는 파급력이 상당하였고, 거슬러 올라가자면 스마트, 디지털교과서, ICT리터러시, 교육정보화 등 기술의 발전이 다양하게 교육에 영향을 주고 있다는 것을 확인할 수 있다[10]. 현재의 컴퓨터과학 혹은 컴퓨터공학이라고 일컬어지는 학문은 컴퓨터라고 일컬어지는 대표적인 자동화 기계를 활용하여 다양한 문제를 효과적으로 해결하는 방법에 대한 체계적인 지식체계로 사회와 산업에 미치는 중요성이 인정되어 초중등 교육에서 컴퓨터교육 혹은 정보교육으로 중요한 자리를 차지하고 있다.

컴퓨터와 관련된 대표적인 교육은 컴퓨터를 사용하여 문제를 해결할 수 있도록 프로그램을 구현하는 프로그래밍 교육이라고 할 수 있다[16]. 초등학생을 위한 프로그래밍 교육은 시모어 페퍼트가 1967년에 처음으로 디자인하여 1970년대에 등장한 LOGO가 시초라고 할 수 있고, 1990년대에 엘런 케이가 개발한 이토이를 거쳐 미치 레즈닉의 스크래치에 이르기까지 프로그래밍을 처음 접하는 학생을 위한 교육환경과 교육내용을 다양하게 연구개발하고 있다[14]. 컴퓨터의 성능이 발전함에 따라 업무의 생산성과 효율성이 증가하게 되었고, 컴퓨터를 능숙하게 사용하여 학생이 과제를 수행하는 등 컴퓨터 활용 능력을 키워주기 위한 교육의 중요성이 대두되었다[8]. 또한, 정보사회가 고도화됨에 따라 발생하는 다양한 역기능에 대처하기 위한 목적으로 학생에게 올바른 가치관을 심어주고 행동양식을 길러주기 위한 정보윤리 교육이 강조되었다[4].

이처럼 컴퓨터와 관련된 기술과 학문의 변화는 초등학생을 위한 정보교육에서 정보통신기술활용교육, 정보소양, 디지털교과서, 스마트교육, 소프트웨어교육의 화두를 거치면서 다양한 연구주제와 교육내용이 등장하게 되었다[11]. 또 다른 변화가 예고되는 인공지능 기술의

등장이라는 시점을 고려한다면, 기존에 수행된 정보교육 연구를 메타적으로 분석하고, 초등 정보교육 연구의 성장과 함께 후속연구에 시사점을 제공하기 위한 노력이 필요한 시점이라 할 수 있다.

본 연구의 목적은 우리나라 초등 정보교육의 거시적인 변화를 조망하고자 하는 목적으로 정보교육학회논문지의 연구주제를 분석하고자 한다. 정보교육학회논문지는 우리나라에서 초등학생을 위한 정보교육 연구가 가장 활발히 이루어진 논문지로 정보교육학회논문지에 출판된 논문들의 흐름이 우리나라 정보교육의 변천사라고 할 수 있다.

연구주제를 도출하는 방법은 연구자가 직접 분류하는 방법은 주관적이거나 확인할 수 있는 텍스트의 양에 한계가 있어 컴퓨팅 파워를 활용하여 자동으로 논문 내에서 사용된 단어의 패턴을 군집화(clustering)하는 토픽 모델링을 활용하였다[12]. 분석논문은 정보교육학회논문지의 창간호에서 2020년까지 출간된 논문을 모두 선정하였고, 국문 제목, 국문 초록, 국문 키워드 텍스트 데이터로 분석하여 연구주제를 도출하였다.

2. 이론적 배경

2.1. 토픽 모델링

연구자가 사전에 가정할 조건이 없이 대용량의 텍스트 자료에서 토픽을 모델링하는 기법 중에 하나인 잠재 디리클레 할당(Latent Dirichlet Allocation, LDA)은 이산자료에 대한 확률적 생성 모형으로 문서 집합 내에 잠재된 토픽들을 분석하여 도출할 수 있다[6]. 확률적 생성 모형이란 확률 분포와 파라미터에 대한 정보를 알고 있을 때, 샘플링을 통해 잠재 변수(latent variable)를 발견할 수 있다는 관점을 반영한 모형이다[2]. 즉, LDA는 문서 속의 텍스트로 관찰된 변수(observed variable)를 통해 숨겨진 변수(hidden variable)를 추론하는 것을 목적으로 하여 문서 집합을 구성하는 토픽 비율과 각 토픽별 단어 분포를 추론한다.

LDA 알고리즘은 관찰할 수 있는 단어인 변수 w , 관찰된 변수인 단어를 기반으로 보이지 않는 변수인 문서별 토픽 분포(θ), 토픽 할당(z), 토픽별 단어분포(β)를

추론하고, 디리클레 파라미터인 α 와 η 에 의해 각 문서를 구성하는 토픽 분포와 토픽을 구성하는 단어의 분포를 추정한다.

2.2. 토픽 모델링의 교육분야 활용

토픽 모델링 기법을 활용한 연구는 다음과 같다. 이종형(2020)은 컴퓨터교육과 관련된 논문의 연구동향을 분석하기 위한 목적으로 2000년부터 2019년까지 학술연구정보서비스(RISS)에 게재된 논문 422편을 분석하였다[7]. 연구방법은 토픽 모델링과 함께 키워드 분석, 중심성 분석, 사회연결망 분석 및 시계열 회귀분석을 수행하였다. 연구결과, 추출한 주제어 중 ‘컴퓨터교육’ ‘프로그래밍교육’ ‘소프트웨어교육’ ‘컴퓨팅사고력’의 순서로 출현빈도가 높게 나타났고, 핵심 주제어 간 연관성은 ‘소프트웨어교육-컴퓨팅사고력’ 사이에서 가장 강하게 연결되었다. 마지막으로 ‘컴퓨터교육’에 관한 토픽 중 ‘컴퓨팅사고력’ ‘소프트웨어교육’ ‘프로그래밍교육’은 년도별 상승추세를 갖는 토픽으로, ‘유아컴퓨터교육’은 하강하는 추세의 토픽으로 분석되었다. 다만, 분석에 활용된 422편의 논문을 선택하는 과정에서 연구자의 주관이 개입할 수 있다는 점에서 컴퓨터교육의 전반을 살필 수 있는 방법이 필요하다.

신동조(2020)는 수학교육과 관련된 논문의 국내외 연구동향을 비교하기 위한 목적으로 2000년부터 2019년까지 7종의 KCI 등재지에 게재된 3,114편의 수학교육 논문과 5종의 SSCI 등재지에 게재된 1,636편의 수학교육 논문을 수집하여 토픽 모델링하였다[13]. 연구결과, 국내외 수학교육 연구는 16개의 유사한 주제와 7개의 상이한 주제로 분류할 수 있는 것으로 분석되었다.

우창우(2020)는 국가의 정책적 방향성을 수립하는데 지원하기 위한 목적으로 국가연구개발사업을 통해 수행되고 있는 ICT분야의 연구과제를 토픽 모델링 하였다[15]. 연구는 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 분석하였다. 연구결과, ICT분야 연구과제에 대한 연구토픽은 인공지능, 빅데이터, 사물인터넷과 같은 지능정보기술로 확인되었고 연구동향에는

초실감미디어에 관한 연구가 활발히 진행되고 있음을 확인하였다.

마지막으로 윤빛나리(2020)은 교육분야에서 정권에 따른 마이스터고 정책이슈와 특성을 분석하기 위한 목적으로 교육부가 발표한 234건의 마이스터고 관련 보도자료를 토픽 모델링하였다[17]. 연구결과, 9개의 토픽이 관찰되었으며, 그 가운데 IP Meister Program, 산관학 협력, 선취업 후진학, 행사·홍보, 글로벌 현장학습 등 학생 지원 제도 및 성과 홍보에 대한 토픽이 다수 등장한 것으로 분석되었다. 추가로 정권별 보도자료 분석에서는 이명박 정부와 박근혜 정부의 마이스터고 정책이슈는 서로 유사한 특성을 보였으나, 문재인 정부는 상대적으로 독자적인 정책 이슈를 형성하는 것으로 나타났다.

기존 연구의 시사점을 정리하면 다음과 같다. 첫째, 토픽 모델링 기법을 활용하여 해당 연구분야에서 연구동향을 거시적인 관점에서 분석이 충분히 가능하다. 둘째, 토픽 모델링 뿐 아니라 키워드 분석, 시계열 분석 등의 추가적인 분석을 통해 연구동향에 대한 추가적인 정보를 제공할 수 있다. 마지막으로 과거의 결과를 바탕으로 향후 연구의 방향을 수립하는데 실증적인 근거로 활용가능하여 초등 정보교육에 대한 거시적인 연구의 흐름을 파악할 수 있을 뿐 아니라 다양한 시사점을 도출할 수 있다.

3. 연구방법

3.1. 연구대상

본 연구에서는 한국정보교육학회(The Journal of Korea Association of Information Education) 논문지의 제 1권 1호(1997년)부터 제 24권 6호(2020년)까지 총 24년간의 논문 총 1,149편을 분석 대상으로 하였다.

3.2. 데이터 수집

데이터의 수집은 정보교육학회논문지에 수록된 논문의 제목, 국문 초록, 국문 키워드를 분석하였다. 대부분의 논문지에서 제목, 초록, 키워드는 공통적으로 구성되어 있는 형식으로, 연구자가 연구의 핵심을 압축적으로

요약하여 해당 논문의 내용을 가장 적절하게 나타낸 부분이라고 할 수 있다. 본 연구에서는 논문의 전반적인 내용을 분석하고 경향성을 파악하기 위해서 제목, 키워드 및 국문 초록이 적합하다고 판단하였고, LDA 알고리즘 기반 토픽 모델링 분석을 수행 하였다. 데이터 수집을 위해 RISS의 서지반출 기능을 이용하여 논문의 국문 제목, 국문 초록, 국문 키워드를 추출하였다. 정확하게 데이터를 수집하기 위해서, 추출된 데이터를 2인이 교차하여 논문과 수집된 데이터를 확인하였고, 누락된 부분은 직접 크롤링하여 총 1,149편의 논문 데이터를 수집하였다. 정보교육학회논문지의 초창기 일부 논문에서 데이터가 일부 누락된 경우에는 국문 제목, 국문 초록, 국문 키워드를 최대한 수집하여 분석하였다.

3.3. 데이터 전처리

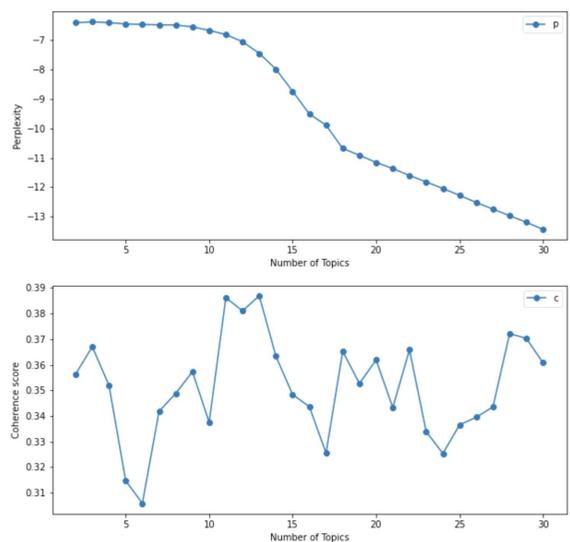
전처리는 수집한 텍스트 데이터에 포함된 특수문자와 문장부호 등을 제거하는 정제와 분석에 활용할 단어를 선정해내는 선별 과정을 수행하였다. 정제는 카카오에서 개발한 형태소 분석기인 khaiii를 사용하여 데이터 정제를 실시하였다. 선별은 개념을 나타내는 품사인 명사와, 수량이나 순서를 나타내는 품사인 수사를 분석에 활용할 대상으로 선정하여 khaiii 분석기의 보통명사(NNG), 고유명사(NNP), 의존명사(NNB), 대명사(NP), 수사(NR)를 추출하였다. khaiii 분석기는 기계학습 기반의 알고리즘을 이용하여 형태소 태깅을 수행하므로, 학습 데이터에 포함되어 있지 않은 단어가 존재할 경우 분석이 올바르게 분석되지 않을 수 있다. 이를 해결하기 위해서, 전처리에서 형태소가 잘못 태깅된 데이터와 단어는 수작업으로 수정 및 치환을 수행하였다. 또한 명사 중에 ‘위’, ‘예’, ‘속’ 등과 같은 한 글자 단어들은 의미 파악이 어렵고, 주제 분석에 큰 영향을 주지 않을 것으로 판단하여 분석 대상에서 제외하였다. 마지막으로, 영문으로 표기된 단어 가운데에도 ‘SW’, ‘ICT’와 같이 주제 분석에 있어 유의미한 정보를 제공할 수 있어 영문도 대상에 포함하였고, 저자가 작성한 원래의 데이터를 가공하지 않고 최대한 그대로의 것으로 분석하기 위해 ‘AI’와 ‘인공지능’ 텍스트를 구분하여 분석하였다. 위의 전처리 과정을 통해 최종적으로 총 5,086개 단어를 분석에 활용하였다.

3.4. 토픽 개수 설정

LDA 기반 토픽 모델링을 수행하기 위해서는 적합한 토픽의 개수를 결정하는 것이 중요하다. 전체 논문과 단어에 비해 토픽 수를 적게 설정하는 경우 개별 토픽이 지나치게 광범위한 내용을 포함하는 주제로 설정될 수 있고, 토픽 수를 많이 설정하는 경우 하나의 토픽으로 분류되어야 할 토픽이 다수의 각기 다른 토픽으로 나뉘게 되어 결과가 부정확할 수 있다.

일반적으로는 토픽의 개수를 설정하기 위해 참조하는 지수로는 복잡도(perplexity)와 응집도(coherence)를 사용하고 있다[1]. 복잡도는 설정된 토픽 모델이 실제 결과를 얼마나 정확하게 예측하는지를 나타내는 척도로, 값이 낮을수록 결과를 더 정확하게 예측함을 의미한다. 응집도 지수는 값이 클수록 설정된 각 토픽에 속한 단어들이 서로 유사한 정도를 의미한다[3].

본 연구에서는 토픽의 개수를 2에서 30까지 1씩 증가시킨 모델에 대해 복잡도와 응집도 지수를 (Fig. 1)과 같이 모두 도출하였고, 복잡도 지수가 급격히 감소하기 이전에 응집도 지수가 가장 높은 지점을 선정하여 토픽의 개수를 최종적으로 11개로 설정하여 LDA 토픽 모델링을 수행하였다.



(Fig. 1) Result of Perplexity and Coherence

3.5. 토픽 모델링

LDA 토픽 모델링은 python의 Gensim 패키지를 활용하였고, 디리클레 모수인 α 와 η 는 'auto'로 설정하였고, 1000회 반복하여 훈련하여 결과를 도출하였다.

4. 연구결과

4.1. 키워드 분석

정보교육학회논문의 창간호인 1997년 1권 1호에서 2020년 24권 6호까지의 논문을 대상으로 주요 키워드를 TF-IDF로 분석한 결과는 <Table 1>과 같다. 상위빈도 키워드를 살펴보면 학습, 프로그래밍, 정보, 로봇, 컴퓨터, 시스템, 교육, SW, 평가, 교사, 초등, 과정, 수업, 컴퓨팅, 활용, 학습자, 문제, 교수, 개발, 소프트웨어 순으로 분석되어 정보교육과 관련된 키워드가 도출된 것으로 확인되었다.

논문이 출간된 년도별 주요 키워드를 TF-IDF로 분석한 결과는 <Table 2>와 같다. 분석결과, 1순위로 분석된 키워드로 로봇 5회, 학습 4회, 정보 3회, 교육 2회, 프로그래밍 2회, SW 2회, 평가, 프로젝트, AI, 인공지능이 각 1회씩 나타났다.

<Table 1> Top 20 Keywords based on TF-IDF

Rank	Keyword	TF-IDF	Rank	Keyword	TF-IDF
1	학습	38.82	11	초등	23.09
2	프로그래밍	36.43	12	과정	22.88
3	정보	36.23	13	수업	22.68
4	로봇	33.03	14	컴퓨팅	22.6
5	컴퓨터	31.51	15	활용	22.12
6	시스템	28.13	16	학습자	21.99
7	교육	26.34	17	문제	21.95
8	SW	25.67	18	교수	21.38
9	평가	25.39	19	개발	20.97
10	교사	23.56	20	소프트웨어	20.7

4.2. 토픽 모델 분석

정보교육학회논문의 창간호인 1권 1호에서부터 2020년 24권 6호까지 전체 논문의 국문 제목, 국문 초록, 국문 키워드를 토픽 모델링한 결과 총 11개의 토픽으로 분류되었고, 해당 토픽에 대한 기여도가 높은 순으로 15개씩 <Table 3>과 같이 나열하였다.

정보교육학회논문지에서 도출된 토픽의 년도별 변화를 시각화한 결과는 (Fig. 2)와 같다. 토픽의 년도별 전체 평균은 T1(37.7%), T2(18.5%), T3(13.9%), T4(8.1%), T5(4.7%), T6(4.5%), T7(3.5%), T8(2.6%),

<Table 2> Top 10 Keywords based on TF-IDF by Year

Year	Paper	1	2	3	4	5	6	7	8	9	10
1997	10	교육	연구	CAI	자료	WEB	개별	멀티미디어	원격	편집기	하이퍼미디어
1998	34	교육	정보	에이전트	시스템	설계	연구	컴퓨터	소프트웨어	초등	교수
1999	25	정보	교육	학습	검색	시스템	컴퓨터	인터넷	문화	심층	목표
2000	22	학습	평가	그룹	코스웨어	학습자	컴퓨터	정보	활동	설계	문제
2001	33	평가	활용	정보	프로그래밍	학습자	컴퓨터	학습	게임	인터넷	시스템
2002	33	프로젝트	학습	학습자	주도	문제	모형	중심	교육	시스템	정보
2003	34	학습	상담	환경	평가	교수	학습자	시스템	ICT	정보	프로그래밍
2004	50	학습	컴퓨터	통신	정보	상담	학습자	ICT	활용	평가	윤리
2005	58	영재	평가	학습	컴퓨터	정보	과학	교육	ICT	활용	시스템
2006	40	로봇	컴퓨터	정보	수업	학습	교과서	ICT	평가	교육	능력
2007	49	학습	시스템	컴퓨터	정보	게임	로봇	토론	프로그래밍	학교	중독
2008	45	프로그래밍	컴퓨터	평가	정보	시스템	초등	학습	로봇	윤리	PBL
2009	50	사이버	로봇	학습	프로그래밍	정보	수업	정확습	교육	컴퓨터	전략
2010	62	로봇	프로그래밍	학습	정보	문제	시스템	컴퓨터	프로그램	게임	교육
2011	63	프로그래밍	로봇	학습	정보	영재	과학	스크래치	저작	프로그램	교육
2012	49	정보	로봇	프로그래밍	스마트	학습	현실	알고리즘	증강	학습자	영재
2013	48	로봇	STEAM	학습	프로그래밍	컴퓨터	스마트	정보	프로그램	개발	스마트폰
2014	63	로봇	프로그래밍	정보	컴퓨터	과학	학습	게임	내용	스마트폰	교육
2015	53	SW	프로그래밍	로봇	소프트웨어	디지털교과서	학습	수업	정보	알고리즘	과정
2016	64	정보	SW	프로그래밍	컴퓨팅	교사	활동	과정	컴퓨터	알고리즘	수업
2017	68	SW	프로그래밍	ICT	정보	컴퓨팅	학습	소프트웨어	로봇	개발	교사
2018	69	로봇	SW	교사	컴퓨팅	학습	교수	가상현실	예비	프로그래밍	정보
2019	66	인공지능	컴퓨팅	SW	사고력	메이커	프로그램	학습	교사	언플러그드	소프트웨어
2020	61	AI	인공지능	SW	로봇	사고력	평가	컴퓨팅	초등	프로그래밍	빅데이터

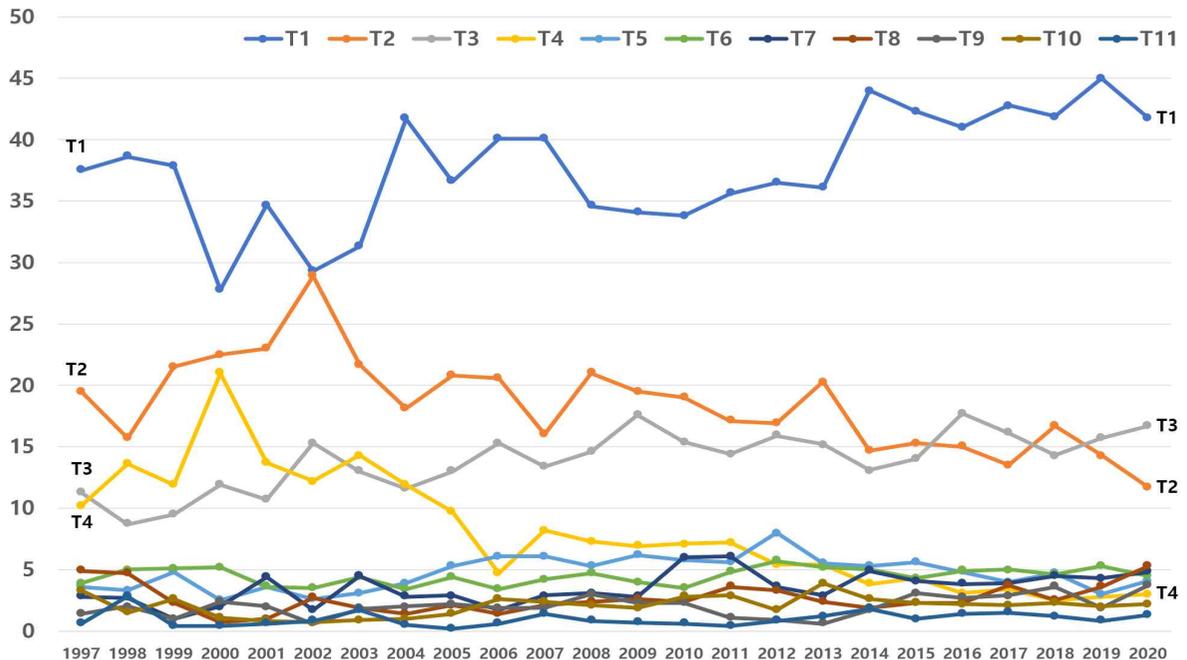
<Table 3> Result of Topic Modeling

Topic	Keywords
T1	교육, 연구, 분석, 초등, 정보, 과정, 교사, 활용, 컴퓨터, 소프트웨어, 학생, 결과, 내용, 개발, AI
T2	학습, 활용, 활동, 학습자, 개발, 교수, 능력, 단계, 적용, 평가, 이해, 자기주도, 문제해결력, 연구, 설계
T3	수업, 효과, 문제, 연구, 적용, 결과, 학습, 사고력, 학생, 학년, 향상, 차이, 해결, 모형, 중심
T4	시스템, 스크래치, 사고, 설계, 구현, 온라인, 관리, 서비스, 상호, 이용, 제공, 자동, 구체, 작용, 사용자
T5	영향, 온라인, 분석, 긍정, 연수, 사용, 몰입, 인터넷, 유의, 효능감, 중독, 디지털교과서, 개정, 만족도, 사회
T6	컴퓨팅, 집단, 확인, 기준, 요소, 스마트, 핵심, 피지컬, 대학교, 시각, 모바일, 기기, 도구, 측정, CS
T7	프로그래밍, 사고력, 프로그램, 게임, 언어, 영재, 코딩, 엔트리, 논리, 블록, 개발, 결과, 추상, EPL, 초등
T8	인공지능, 데이터, 알고리즘, 디지털, 학업, 성취, 환경, 모델, 제안, 탐색, 절차, 방법, 놀이, 예측, 데이터과학
T9	SW, 평가, 유형, 빅데이터, 윤리, 수행, 소프트웨어교육, 이러닝, 학점, IPTV, 자녀, SWEET, 객관, 준거, 년도
T10	로봇, 검사, 사후, STEAM, 지능, 표본, 문항, 진로, 선행, 진단, 강사, 보조, 직업, 여학생, 센서
T11	역량, 리터러시, ICT, 스마트폰, 정도, 경로, 단어, 측정, 습관, 문서, 타당도, 비버챗런지, 위험, 토픽, 안전

T9(2.0%), T10(2.0%), T11(1.0%)의 순으로 분석되었다. 연도별 변화를 살펴보면, T1은 창간호에서 2020년까지 가장 높은 비중을 나타내고 있는 것으로 분석되었고, 최근에 와서 T2는 감소하고, T3는 증가하는 추세인 것으로 나타났다. 마지막으로 T4의 경우 초창기에는 10% 이상의 비중을 차지하다가 2005년 이후 감소하여 최근의 비중은 5% 미만인 것으로 분석되었다. 전체적으로 살펴보면, 상위 4개의 토픽이 전체의 75%이상의 비중을 차지하고 있는 것으로 나타났다.

각 토픽의 연도별 비율 값으로 회귀계수를 도출하여 양의 값이면 정보교육학회논문지 내에서 연구의 관심이 증가되는 핫이슈(hot-issue)으로 분류하였고, 회귀계수가 음의 값이면 콜드이슈(cold-issue)으로 분류한 결과는 <Table 4>와 같다[5].

토픽 T1, T3, T7, T9은 유의수준 .05에서 상승하는 것으로 분석되었고, 토픽 T2, T4는 하락하는 것으로 분석되었다. 정리하면, T1과 T3의 비중은 증가하고 T2와 T4의 비중은 감소하고 있다고 해석할 수 있다.



(Fig. 2) Rate of Change of Topics by Year(%)

<Table 4> Regression Analysis by Topic

Topic	Regression coefficient (standard error)	p-value	Trend
T1	0.380(0.111)	.002	Rise
T2	-0.367(0.084)	.000	Fall
T3	0.252(0.048)	.000	Rise
T4	-0.577(0.072)	.000	Fall
T5	0.061(0.039)	.128	-
T6	0.032(0.019)	.111	-
T7	0.105(0.032)	.004	Rise
T8	0.037(0.034)	.297	-
T9	0.055(0.023)	.027	Rise
T10	0.038(0.023)	.112	-
T11	0.013(0.018)	.454	-

5. 결론

본 연구는 초등 정보교육의 거시적인 변화를 조망하고자 하는 목적으로 정보교육학회논문지의 창간호에서 2020년에 발간된 24권 6호까지의 논문을 대상으로 키워드 분석과 토픽 모델링을 수행하였다.

본 연구의 시사점은 다음과 같다. 첫째, 정보교육학회 논문지에 가장 많은 비중(약 38%)을 차지하는 토픽 T1은 기여도 순으로 교육, 연구, 분석, 초등, 정보의 키워드가 도출되었고, 통계적으로 유의한 수준에서 년도에 따라 비중이 상승하는 것으로 나타나 논문의 초등 정보교육을 대표하는 특성이 강화되고 있는 것으로 해석할 수 있다.

둘째, 상대적으로 많은 연구비중을 차지하고 있는 토픽 T2(학습, 활용, 활동, 학습자, 개발 등)와 T4(시스템, 스크래치, 사고, 설계, 구현 등)가 감소하고, 토픽 T7(프로그래밍, 사고력, 프로그램, 게임, 언어 등)과 T9(SW, 평가, 유형, 빅데이터, 윤리 등)는 증가하고 있어 연구의 트렌드가 컴퓨팅사고력, 소프트웨어교육, 인공지능교육 등으로 변화하고 있는 것으로 나타났다.

마지막으로 연구자가 임의로 초등 정보교육의 분류체계를 제안하기 보다는 수집된 대량의 텍스트 데이터를 객관적으로 클러스터링하였다는 점에서 의미가 있다고 할 수 있다.

본 연구의 한계점은 첫째, 한국어 형태소 분석기의 성능에 따라 연구의 결과가 영향을 받을 수 있다. 둘째, 분석에 활용된 데이터는 논문의 국문 제목, 국문 초록,

국문 키워드로 논문 모든 텍스트를 다루지 못하였고, 텍스트를 전처리하여 활용하는 과정에서 논문의 정보가 유실될 수 있다.

향후 연구로는 단어간 상관을 고려한 발전된 토픽 모델링 기법인 상관 토픽 모델(Correlated Topic Model)을 활용할 필요가 있고, 분류된 토픽의 이름을 자동으로 라벨링해주는 방법에 대한 고려가 필요하다.

참고문헌

- [1] 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.
- [2] David M. Blei, Andrew Y. Ng, Michael I. Jordan. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993-1022.
- [3] Jonathan Chang, Jordan Boyd-Graber, Sean Gerrish, Chong Wang, David M. Blei. (2009). Reading Tea Leaves: How Humans Interpret Topic Models. *Advances in Neural Information Processing Systems*, 22, 288-296.
- [4] Kim, K.M., Kim, S.S., and Kim, S.S.(2014). Development of Assessment Tool about Professionalism of the Information Ethics Teachers. *Journal of The Korean Association of information Education*, 18(1), 1-11.
- [5] Kim, S.Y.(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.
- [6] Kim, T.K., Choi, H.R., and Lee H.C.(2016). A Study on the Research Trends in Fintech using Topic Modeling. *Journal of the Korea Academia-Industrial cooperation Society*, 17(11), 670-681.
- [7] Lee, C.H., Kim, U.J.(2020). A Research Trend

- Analysis of Computer Education Using Topic Modeling. *The Journal of Korean Association of Computer Education*, 23(6), 15-23.
- [8] Ma, D.S et al(2008). A Study on the Computer Application Ability Gap of the Elementary Student. *Journal of The Korean Association of information Education*, 12(2), 163.171.
- [9] Ministry of Government(2020). *Education Policy Direction and Core Tasks in the Age of Artificial Intelligence*.
- [10] Moon, W.S.(2017). A Study on the Trend of papers published by Korean Association of Information Education. *Journal of The Korean Association of information Education*, 22(6), 681-687.
- [11] Park, S.J.(2017). Analysis of Information Education Related Theses Using R Program. *Journal of The Korean Association of information Education*, 21(1), 57-66.
- [12] Shin, M.S., Cho, K.W.(2019). Analysis on Topic Modeling and Trend of Journal of Speech-Language & Hearing Disorders using Text Mining: (2002~2018). *Journal of speech-language & hearing disorders*, 28(3), 81-91.
- [13] Shin. D.J.(2020). A comparative study of domestic and international research trends of mathematics education through topic modeling. *The Mathematical Education*, 59(1), 63-80.
- [14] T. McNerney(2004). From turtles to tangible programming bricks: Explorations in physical language design. *Personal and Ubiquitous Computing*, 8(5), 326-337.
- [15] Woo. C.W., Lee J.Y.(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.
- [16] Yang, C.M(2014). Meta-Analysis on the Effects of Programming Education using Educational Programming Languages. *Journal of The Korean Association of information Education*, 18(2), 317-324.
- [17] 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, 프로그래밍교육, 온라인교육
 e-mail: jaekwoun.shim@gmail.com