

토픽 모델링 분석을 통한 수학교육 연구 주제 분석

진 미 르 (아주대학교 대학원 학생)
고 호 경 (아주대학교 교수)[†]

본 연구는 최근 수학교육 연구 논문들의 연구 동향을 파악하기 위하여 2016년 이후의 수학교육 학술지 논문들을 대상으로 텍스트마이닝 기법 중 토픽 모델링과 트렌드 분석 기법을 활용하여 분석을 실시하였다. 분석 결과 빈도수가 높은 단어들을 조합하여 5개의 토픽을 추출하였으며 이를 통해 최근 활발히 이루어지고 있는 수학교육 연구 주제들을 파악할 수 있었다. 이는 다시 기 수행된 동향 연구들과 차이점과 유사한 점들을 도출할 수 있었는데, 이와 같은 동향 분석을 통해 최근 연구자들이 수학교육 연구에서 중요시 여기는 관점을 읽어 나감과 동시에 향후 주목하여야 할 연구 주제 및 방향에 대한 시사점을 제공한다.

I. 들어가는 말

최근 들어 수학교육에서는 시대의 흐름을 반영하여 영역 특수적인 측면에서 향후 교육의 방향을 설정하기 위한 연구들이 진행되고 있다. 예를 들면, 김영록(2017)은 미래세대가 필요로 하는 수학적 역량을 도출한 다음 이에 맞는 수학 기반 미래 인재상을 '수학적 사고력과 창의력을 갖춘 융합 인재'로 제시한 바 있고 수학을 교수하는 측면에서 강조해야 할 사고와 역량이 무엇인가에 대한 연구(최은미, 2017), 또는 보다 포괄적이고 일반적인 접근에서 미래 사회를 살아가기 위해 필요한 핵심 역량(key competency)을 도출하고 이를 수학교육의 목적과 방향에 대한 지평으로 삼고자 하고자 하는 시도(나귀수 외, 2018) 등 다양한 측면에서 시대를 반영하고자 하는 노력들이 이루어지고 있다.

시대의 변화와 흐름에 따라 수시로 변화하는 사회적 요구와 이에 따른 수학교육의 나아갈 방향을 모색하기 위해서는 현재까지 이루어진 수학교육학 관련 연구들이 시대의 흐름에 따라 어떻게 변화하여 왔는지 분석하고 현재의 동향을 분석하는 연구 역시 중요한 연구 중 하나이다(김래영 외, 2012). 이와 같은 필요성에 따라 수학교육의 연구 동향을 살펴본 연구들이 지속적으로 수행되고 있는데, 예를 들면, 1990년대 우리나라 수학교육 연구 동향 분석(최택영·송병근, 2001)처럼 국내 석사 논문을 분석한 연구나, 초등수학교육 연구동향(하수현 외, 2010)를 파악하기 위하여 국내 학술지에 게재된 논문들을 분석하여 국내 수학교육연구의 경향을 살펴본 연구들을 찾을 수 있다. 이렇듯 수학교육 연구 동향에 대한 연구는 지속적으로 이루어져 왔지만 대부분 특정 주제를 분류하는 연구이거나 연구 논문들의 분류 및 주제에 대한 빈도수 분석이 주를 이루어왔으며 이는 사람에 의해 주제를 분류하고 헤아리는 방법을 사용해 왔다. 그러나 이러한 기존의 방식으로는 시시각각 제공되어 방대해진 학술자료를 분석하기에는 많은 시간과 노력이 소요될 뿐만 아니라 이면의 정보들을 탐색하고 필요한 내용을 도출하는

* 접수일(2019년 7월 24일), 심사(수정)일(2019년 8월 2일), 게재확정일(2019년 8월 2일)

* ZDM분류 : C10

* MSC2000분류 : 97B10

* 주제어 : 텍스트마이닝, 토픽 모델링, 트렌드 분석

* 본 연구는 진미르의 석사학위 논문의 일부를 수정·보완하였음.

* 이 논문은 2016년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2016S1A5A2A01023238)

[†] 교신저자 : kohoh@ajou.ac.kr

것이 용이하지만은 않다.

인터넷 및 다양한 기술이 발달함에 따라 언제 어디서나 쉽게 정보를 검색하는 것이 가능해지고, 이러한 환경 속에서 정보는 무궁무진하게 쏟아져 나오며 따라 그 정보들은 이전과는 비교할 수 없을 정도로 많은 양의 데이터로 적재된다. 이러한 축적되는 데이터들 중에서 유의미한 정보를 찾기 위한 분석 방법은 기존의 분석 방법만으로는 어려움이 있기 때문에 또 다른 새로운 방법들을 시도해야한다(박참솔, 2017). 텍스트마이닝은 온라인상에 축적된 대량의 데이터를 트렌드 분석에 활용하는 객관적이고 실효성 높은 기법이다(Feldman & Dagan, 1995). 텍스트마이닝은 통계적인 접근을 통하여 개념 간의 연결성과 영향 관계를 파악하고 이를 시각화함으로써 데이터가 보여주는 의미를 도출하는 데에 유용하다(김선아 외, 2016).

이미 많은 영역에서 텍스트마이닝 분석 기법이 활발하게 활용되고 있으며 교육 분야에서 역시 이를 활용하는 연구가 늘어가고 있는 추세이다. 예를 들어 수학 담화에서 나타나는 교사의 감성적 언어를 분석하는데 텍스트마이닝을 활용하기도 하였다(손복은·고호경, 2018). 또한 연구 동향을 분석하는데 있어서도 텍스트마이닝을 이용한 연구가 수행된 바 있는데, 김선희·김수민(2018)은 2017년 국제 학술대회에서 발표된 논문 제목에 나타나는 단어의 빈도와 단어와 단어를 연결하는 의미 분석을 통해 최근 수학교육 연구에서 어떤 주제가 연구되고 있는지를 파악하는 연구를 실시하였다. 그러나 이 연구에서는 논문의 제목만을 가지고 분석함에 따라 연구의 트렌드를 충분히 파악하기는 어렵다는 한계점이 도출된 바 있다.

따라서 본 연구에서는 논문의 제목만 가지고 분석하는 것이 아닌 본문의 서론과 결론을 이용하여 분석을 실시하고자 하며, 2016년 이후 최근 3개년도의 수학교육 연구 학술지를 대상으로 다루고 있는 연구 논문에서 추출한 주요 주제들 간의 네트워크는 어떻게 나타나는지 도출하고자 토픽 모델링과 트렌드 분석을 실시하였다. 토픽 모델링(topic modeling)은 텍스트 데이터 내 단어들의 빈도수를 통계적으로 분석하여 전체 데이터에서 잠재적 주제인 토픽(topic)들을 자동으로 추출하여 분류한다는 점에서 주제 분석 시 유용하다. 또한 토픽 트렌드(topic trend)는 최근 다양한 분야에서 동향을 파악하기 위한 방법으로 활발하게 사용되고 있는 텍스트마이닝 기법이다.

이러한 분석을 통해 수학교육 연구 동향을 도출 해 낸다면, 이는 새로운 연구 방법을 통해 연구 동향 파악의 계보를 잇는다는 점 외에도 향후 주목하여야 할 연구 주제가 무엇인지를 판단할 수 있는 기초 자료를 제공한다는 측면에서 의의가 있을 것이다.

II. 수학교육 연구 동향 분석 선행 연구

연구 동향 분석을 진행하기에 앞서 선행연구의 고찰은 연구의 방향을 제시하고 분석 기준의 근간을 마련해 준다는 점에서 매우 중요하다. 2000년 이후에 수학교육의 연구 동향을 분석한 연구는 <표 II-1>과 같다.

<표 II-1> 2000년 이후 수학교육 연구 동향 분석 논문

저자명	연구 제목	연도
최택영, 송병근	1990년대 우리나라 수학교육 연구 동향 분석	2001
박경미	중등 수학교육 연구의 경향 분석	2003
이강섭	<수학교육>에 게재된 논문의 분류와 분석	2003
권정은, 최재호	초등수학교육 연구 동향 분석	2008
하수현 외	초등수학교육 연구동향	2010
박선영, 김원경	국내외 수학교육 연구 동향 비교 분석	2011
김동중 외	수학교육연구 및 혼합 연구방법 동향	2014
탁병주, 이경화	우리나라 통계교육 연구의 동향 분석	2017
김선희, 김수민	언어 네트워크 분석법을 이용한 최근 수학교육 연구 동향 탐색	2018

최택영, 송병근(2001)은 1990년부터 1999년까지 교육대학원 학위논문 3,214편을 연도별, 유목별, 지역별로 분류한 후 각각의 연구 분포가 고른지, 교육과정의 기대와 요구에 부합되게 진행됐는지 살펴보았다. 이 연구에서는 각 기준별 동향 뿐 아니라 지역에 따른 유목별 동향, 지역에 따른 연도별 동향, 유목에 따른 연도별 동향을 분석하는 등 분석 기준별 관계를 파악하려는 노력을 통해 다각적으로 자료를 해석하였다. 하지만 워낙 방대한 자료를 대상으로 하였고 때문에 논문의 전체적인 내용을 분석하지 않고, 제목과 초록만을 분석하였기에 지역별, 유목별, 연도별 분석 기준과 기준별 상호관계를 파악하는 분석 밖에 할 수 없는 단점이 있다. 따라서 본 연구에서는 논문의 전반적인 내용을 종합적으로 분석을 위하여 논문의 서론과 결론을 이용하여 분석하였다.

박경미(2003)는 2000년부터 2003년까지 <수학교육>과 <JRME>에 수록된 중등 수학교육 연구의 경향을 영역과 주제, 논문의 전개방식 및 구성 요소, 논문집의 구성 방식의 측면에서 살펴보았다. 주제 측면에서 <수학교육>에는 교과서의 개발과 적용에 관련된 연구가, <JRME>에는 기준에 기초한 교육과정의 효과 검증에 대한 연구가 많이 나타나지만 두 저널 모두 각국의 큰 관심사를 다룬다는 공통점이 있으므로 이 연구의 주제 분석을 통해 현안이 무엇인지 알 수 있다는 전제 하에 질적 접근을 통해 각 측면에서 포착되는 차이를 기술하는데 중점을 두었다.

이강섭(2003)은 1963년부터 2003년까지 <수학교육>에 수록된 논문 486편을 분석하고 연도와 MSC000기호, ZDM 분류 기준을 사용하여 분류하였다. 논문의 연구영역, 내용, 동향, 특징 등을 분석하여 우리나라 수학교육연구의 초기 상황에 대하여 정보를 제공하고 있다. 또한 영역별 연도별 분류와 더불어 많은 논문을 게재한 저자별 분석을 시도하여 저자별 편수와 관심분야를 간략하게 분석한 것이 특징이다.

권정은, 최재호(2008)는 1998년부터 2006년까지 초등수학교육에 관한 학술지 논문들의 연구 동향을 분석하여 분야별로 논문들의 연구가 얼마나 진행되었으며, 제7차 교육과정에 대한 요구와 적합하게 초등수학교육연구가 진행되고 있는지에 대한 분석을 실시하였다. 연구대상, 연구주제, 수학과 관련된 내용영역, 연구방법, 연구지역, 연구연도 등을 분석기준으로 하고, 분석대상을 '초등 수학'으로 검색하여 나온 학술지 및 논문 235편을 수집하고 내용분석을 실시하였다. 대상을 선정하는데 있어 모든 초등수학교육 연구와 관련된 논문이 포함되었다고 볼 수 없으므로 본 연구에서는 연구대상임에도 제외되는 경우가 없도록 학술지의 모든 논문들을 대상으로 분석을 실시하였다.

하수현 외(2010)는 한국학술진흥재단(한국연구재단)에 등재된 학술지 중 수학교육과 관련된 7개 학술지와 <초등수학교육> 학술지를 포함한 총 8개 학술지를 대상으로 2005년부터 2009년까지 5년간 발표된 논문 중 '초등수학교육'에 해당하는 논문 383편을 분석 대상으로 선정하였다. 주제, 연구방법, 연구대상, 수학 내용 영역별로 분류하여 초등수학교육 연구 동향에 관한 포괄적인 연구를 실시했다. <초등수학교육>의 연구 동향을 포괄적으로 파악했다는 점이 큰 특징이나, 하위 요소를 너무 세세하게 분류하여 전체적인 경향의 변화를 파악하기 힘들다. 따라서 본 연구에서는 하위요소에 대한 세세한 분석 보다는 전체적인 동향을 파악하는 것에 중점을 두고 분석을 실시하였다.

박선영, 김원경(2011)은 2005년부터 2009년까지 <수학교육>, <수학교육학연구>, <JRME>, <SM>에 수록된 논문 중 중등수학교육과 관련된 논문 587편을 대상으로 연구 분야별, 내용별, 학교급별, 방법별, 주제어별 연구 동향을 분석하였다. 국내외 국외의 학술지를 비교 분석하였다는 특징이 있으며, 비교적 많은 양의 논문을 대상으로 중등수학교육과 관련된 내용을 포괄적으로 다루었으나, 국내 학술지로는 2개만을 선정하여 국내의 동향을 대표한다고 보기 어렵다. 따라서 본 연구에서는 보다 다양한 학술지를 선정하여 분석을 진행하였다.

김동중 외(2014)는 2003년부터 2013년까지 국내 학술지에 등재 및 발표된 문헌 연구를 제외한 총 709편의 논문을 대상으로 하여 연구 동향 및 연구방법의 동향을 분석하였다. 그 중 혼합 연구방법을 사용한 논문을 가지고 연구 단계별로 혼합유형에 대해 조금 더 심층적으로 조사하였다. 수학교육논문에 연구방법이 어떠한 방법으로 사용되었으며 미래의 수학교육 연구에서의 연구방법의 방향과 다양성에 대한 시사점을 제시하고자 하였다.

탁병주, 이경화(2017)는 2000년부터 2016년까지 17년간 한국연구재단 등재후보지 이상의 주요 수학교육 학술지에 게재된 통계교육 연구논문 99편을 통해 국내 연구 동향을 분석하였다. 학술지별, 연구 대상별, 연구 방법별, 연구 주제별로 범주화하고 연도별로 그 분포를 확인하였고 국내 통계교육의 연구 동향을 2010년을 기준으로 전후로 분류하여 그 특징을 분석하고 향후 통계교육 연구의 과제와 발전 방향에 대한 시사점을 도출하였다.

김선희, 김수민(2018)은 2017년 국제 학술대회에서 발표된 연구의 제목을 언어 네트워크 분석법을 이용하여 분석하여 수학교육의 최근 연구 동향을 살펴보았다. 1,201편의 논문 제목에 언급된 단어의 빈도와 단어와 단어를 연결하는 의미 분석을 함께 살펴봄으로서 최근 수학교육 연구에서 어떤 주제가 연구되고 있는지를 구체적으로 파악하고, 그 동향에서 국내에서 미진한 주제를 찾아 국내 연구자들에게 정보를 제공하는 데 의의를 두었다. 그러나 위의 연구에도 논문의 제목만을 가지고 분석하였다는 한계점이 있다. 따라서 본 연구에서는 논문의 제목이 아닌 본문의 서론과 결론을 이용하여 분석을 실시하였다.

III. 연구 방법

1. 연구 분석 대상

기존의 연구 동향을 분석하는 논문들의 연구 대상 선정 방법을 살펴보면 크게 두 가지로 나타난다. 우선 학술지와 분석 기간을 선정한 후, 학술지에 게재된 논문들을 대상으로 연구 동향을 분석하는 방법이 있다(이수진 외, 2013). 다른 방법은 논문 검색사이트에서 키워드 검색을 통해 연구 대상을 선정하는 것이다(김혜미, 2016). 본 연구에서는 전반적인 맥락 및 연구 흐름을 분석하기 위해 전자와 같은 방법으로 다음과 같이 학술지와 분석 기간을 선정한 후 연구 동향을 분석하였다.

먼저, 분석 대상을 선정하기 위해 한국학술지인용색인(Korea Citation Index: KCI) 사이트를 통해 수학교육 관련 학술지를 검색하여 KCI에 등재(후보)된 학술지를 중심으로 학술지명, 발행 학회, 영향력지수 등을 비교하였다. 그러나 영향력지수는 학회지 별로 대동소이하여 이를 기준으로 삼아 선별하는 것은 무의미하다 판단하였다. 본 연구의 따라서 대상은 국내 수학교육관련 학술단체인 대한수학교육학회, 한국수학교육학회, 한국학교수학회에서 발행된 학술지를 모두 선정하였다. 위의 학술단체 학술지 외에도 다양한 학회의 학술지가 있고, 학위논문이나 연구보고서도 있다. 그러나 위의 학술지는 국내에서 수학교육학 연구를 목적으로 전국적으로 학술대회를 개최하는 학회들로서 국내의 수학교육 분야에서 권위를 인정받고 있기 때문에 수학교육에 대한 전반적인 연구 동향을 파악하는 데에 충분히 의미가 있을 것으로 여겨진다(박선영, 2011). 또한 초·중·고등학교급 모두에서의 전반적인 수학교육에 대한 트렌드를 분석하기 위해 초등수학에 관한 연구가 주를 이루고 있는 초등수학교육 관련 학술지는 제외하였다. 그 결과 본 연구는 한국수학교육학회에서 발행한 <수학교육>, <수학교육 논문집>, 대한수학교육학회에서 발행한 <학교수학>, <수학교육학연구>, 마지막으로 한국학교수학회에서 발행한 <한국학교수학회 논문집>을 분석대상으로 선정하였다.

분석 기간 선정에 앞서 선행연구를 살펴보면 분석 기간은 주로 5-10개년 주를 이루었는데, 이는 분석 기간이 길수록 시간의 흐름에 따른 트렌드 변화를 분석하기에 더 용이하기 때문이다. 하지만 논문의 서론과 결론을 이용하여 분석하는 본 연구의 특성상 기간이 길어질수록 텍스트의 양이 너무 방대해져서 분석하는데 한계점이 있었다. 따라서 본 연구는 기간에 따른 동향의 변화가 아닌 최근에 주로 이루어지고 있는 수학교육에 관한 연구의 트렌드를 분석하는 것을 목표로 하였다. 특히 최근에 이루어진 교육정책 중 연구 주제 선정에 가장 큰 영향을 미친다고 판단되는 '2015 개정 교육개정'이 발표된 이후인 2016년부터 1월부터 2018년 12월까지 총 3개년을 기간으로 선정하여 분석을 진행하였다.

<표 III-1> 연도별 학술지 논문 수(괄호는 영문으로 작성된 논문 수)

학회	학술지	2016년	2017년	2018년	합계
한국수학교육학회	수학교육	26	23	24(2)	73(2)
	수학교육 논문집	28(1)	30(2)	30(1)	88(4)
대한수학교육학회	수학교육학연구	41	42	29	112
	학교수학	44	40	38	122
한국학교수학회	한국학교수학회 논문집	25(2)	27(1)	23(1)	75(4)
합계		164(3)	162(3)	144(4)	470(10)

이렇게 선정된 기간별 학술지의 논문수를 살펴보면 <표 III-1>과 같다. 한국수학교육학회에서 발행한 <수학교육> 학술지에 게재된 논문은 2016년에 26편, 2017년에 23편, 2018년에 24편으로 총 73편의 논문이 게재되어 있고, 동 학회에서 발행한 <수학교육 논문집> 학술지에는 2016년에 28편, 2017년에 30편, 2018년에 30편으로 총 88편의 논문이 게재되어 있다. 대한수학교육학회에서 발행한 <수학교육학연구> 학술지의 경우 2016년에 29편, 2017년에 42편, 2018년에 41편으로 총 112편의 논문이 게재되어 있고, 마찬가지로 대한수학교육학회에서 발행한 <학교수학> 학술지의 경우 2016년에 38편, 2017년에 40편, 2018년에 44편이 게재되어 총 122편의 논문으로, 본 연구에서 선택한 학술지 중 가장 많은 논문이 게재되어 있었다. 마지막으로 한국학교수학회에서 발행한 <한국학교수학회 논문집> 학술지는 2016년에 25편, 2017년에 27편, 2018년에 23편이 게재되어 총 논문 수는 75편이었다. 선정된 학술지 별 논문수를 연도별로 살펴보면 2016년에 게재된 총 논문 수는 164편이었고, 2017년도에는 162편, 2018년도에는 144편이 게재되었다. 이렇게 선정된 논문 수는 총 470편이었는데 이 논문들 중 영문으로 작성된 논문 10편은 연구 대상에서 제외하였다. 이는 영문으로 된 텍스트를 포함을 시킨다고 하여도 한글 텍스트의 비율이 대부분이기 때문에 분석하는데 있어 미치는 영향이 미비하므로 영문으로만 작성된 논문은 제외하였다. 따라서 본 연구에서는 최종 연구 대상으로 2016년부터 2018년도까지 총 3개년도 동안 5개의 학술지에 게재된 총 460개의 논문을 선정하여 분석하였다.

2. 분석 방법

텍스트마이닝의 기능에는 텍스트 분류, 텍스트 군집화, 텍스트 요약, 그리고 텍스트 분할 등이 포함된다. 텍스트 분류는 텍스트의 내용에 의해 사용자가 미리 정의한 범주를 부여하는 과정을 말하는데, 이는 데이터마이닝의 패턴 분류에 대응된다. 텍스트 군집화에는 주어진 텍스트 집단을 내용의 유사성에 근거하여 이들을 군집화하여 여러 개의 서브그룹으로 나누는 과정을 말하며, 데이터마이닝의 패턴 군집화와 대응이 된다. 텍스트 요약은 텍스트로부터 전체 내용을 반영할 수 있는 문장 또는 문단 등 일부의 내용을 선택하는 과정을 말한다. 텍스트 분할은 내용이 긴 텍스트에 적용하는데, 1개의 텍스트를 내용에 따라 여러 개의 텍스트로 나누는 과정을 말한다. 이 경우에는 이를 구조화 시키는 과정이 필요하다(조태호, 2001).

텍스트마이닝의 기법 중 하나인 토픽 모델링이란 긴밀하게 연결되어 있는 연구들 안에서 독립적인 토픽들이 공존하거나 공유되는 것에 주목하여 그들 간의 연결 관계를 통하여 전체 집합 속 부분적인 토픽을 분석할 수 있는 기법으로 유의미한 데이터의 시간의 흐름에 따른 동향 파악에 효율적인 분석 방법 중 하나이다(김현희·이혜영, 2016). 텍스트마이닝 기법의 특징은 다음과 같다. 첫째, 인지능력을 사용하지 않아 객관적인 분석이 가능해 지고, 둘째, 분석결과를 시각화함으로써 연구의 관점을 쉽고 빠르게 전달할 수 있으며, 셋째, 도출된 결과를 연구

자의 관점에 따라 다각도로 조명해볼 수 있다는 것이다(이인선·나은영, 2018). 이처럼 텍스트마이닝 기법을 이용하여 다양한 연구들에 사용된 복잡하고 다양한 언어와 개념들을 비정형 데이터로 처리하고 통계적으로 분석을 함으로써 기존의 연구 동향에 대한 의미 있는 정보를 새롭게 조직할 수 있다.

IV. 텍스트마이닝을 활용한 분석 과정

1. 데이터 클리닝

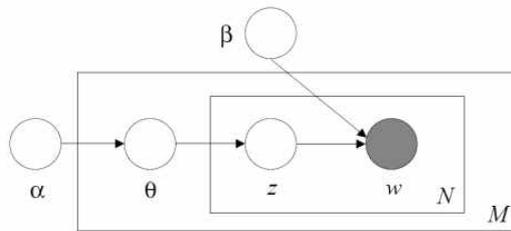
데이터 클리닝 단계는 자연어처리(NLP) 기법을 이용하여 앞서 수집한 텍스트 데이터를 분석이 용이한 형태로 변환하는 과정이다. 본 연구에서 데이터 클리닝은 R-Studio 프로그램을 이용하여 텍스트 데이터에 대한 전처리(preprocessing) 작업과 형태소 분석의 두 단계로 진행하였다. 전처리 단계에서는 수집된 텍스트 데이터 중 분석에 불필요한 단어 및 어구를 삭제하는 작업을 진행하였는데 텍스트에 포함되어 있는 각종 특수기호와 문장부호 등 분석 시 의미를 부여할 필요가 없는 불용어(stopword)를 제거한다. 이때 잘못 나타나는 단어들과 유사어지만 다르게 표현된 단어들을 통일하는 정규화(normalization)하는 작업 역시 반복적으로 실시해 주어야 한다. 다음 단계인 형태소 분석은 정보 추출이 가능하도록 텍스트를 자연어 형태로 변환하는 단계로 본 연구에서는 한국어 처리 패키지인 'KoNLP'와 'tm', 'stringr' 패키지를 활용하고자 한다. 전처리 작업을 거친 데이터를 품사분석을 실시하여 보통명사를 추출한다. 이때 명사를 추출한 후 추가적인 필터링을 해야 하는 단어들이 존재함을 발견할 수 있는데, 예를 들어 '가지'라는 단어의 경우 명사로 인식되어 추출되지만 의미를 갖는다고 판단하기 힘들다. 빈도수가 낮은 경우에는 빈도수 분석 결과에 크게 영향을 미치지 않지만, 좋은 분석 결과를 얻기 위해 불용어 제거와 함께 앞서 실시한 정규화 작업을 반복 시행하였다. 또한 명사 추출 시 나타나는 한 글자 단어들은 의미가 없는 경우가 많기 때문에 제거한 후 두 글자 이상으로 된 단어만 추출하였다(김영우, 2017).

2. 토픽 분석 모델 결정

1) 토픽 모델링(LDA 기반)

빈도수 분석은 어떠한 주제가 중요하게 나타나는지에 대해 효과적으로 보여주고 있지만, 단어의 빈도수만으로는 여러 쟁점을 보여주는 데 한계가 있다(박종희 외, 2015). 토픽 모델링(topic modeling)은 텍스트 데이터 내 단어들의 빈도수를 통계적으로 분석하여 전체 데이터에서 잠재적 주제인 토픽(topic)들을 자동으로 추출하여 분류한다는 점에서 주제 분석 시 유용하다. 토픽 모델링은 방대한 텍스트 자료로부터 특정 주제를 추출하는 알고리즘으로써 문서와 단어로 구성된 행렬(dtm)을 사용하여 문서에 잠재된 토픽의 등장 확률을 추정하는 통계적인 텍스트 처리기법이다(백영민, 2017). 또한 긴밀하게 연결되어 있는 연구들 안에서 독립적인 토픽들이 공존하거나 공유되는 것에 주목하여 그들 간의 연결 관계를 통하여 전체 집합 속 부분적인 토픽을 분석할 수 있는 기법으로, 유의미한 데이터의 시간의 흐름에 따른 동향 파악에 효율적인 분석 방법 중 하나이다(김현화·이혜영, 2016). 토픽 모델링의 종류는 잠재적 디리클레 할당 모형, 구조적 토픽모형, 상관 토픽 모형 기법이 있는데, 본 연구에서는 Blei et al.(2003)이 발표한 잠재적 디리클레 할당 모형(Latent Dirichlet Allocation: LDA)을 활용하고자 한다. 이는 LDA 기법이 다른 기법에 비해 결과 해석이 쉽고(Blei, 2012), 과적합(overfitting) 문제들을 해결하기 때문에 방대한 비정형 데이터로부터 다양한 토픽들을 도출하는 데 유리하기 때문이다(Griffiths & Steyvers, 2004). 특히 최근 주목하고 있는 다양한 토픽 모델링과 관련한 확장 모형 또한 LDA 기법에 기반하고 있기 때문에 향후 연구의 확장성 측면에서도 이 분석 기법이 유용하다(유예림, 2017).

LDA는 주어진 문서 내 토픽을 잠재적으로 가정하는 확률모델로 문서들이 잠재된 여러 개의 토픽들로 구성되어 있고, 다양한 단어들로 표현된다. 또한 이 토픽들의 구성 비율은 텍스트 자료 내 문서마다 다르며, 비율을 결정할 수 있는 확률분포로써 디리클레 분포가 존재한다고 본다. LDA는 문서 내에 토픽들의 확률분포 θ 와 각 토픽을 구성하는 단어들의 확률 분포 z 가 주어졌을 때, 문서를 구성하는 토픽을 확률적으로 선택하고 선택된 토픽에 존재하는 단어를 확률적으로 선택하는 샘플링과정을 반복함으로써 임의의 문서를 생성하는 모델이다(Blei et al, 2003). Blei et al.(2003)는 이러한 생성 과정을 [그림 IV-1]과 같이 시각화하였다.

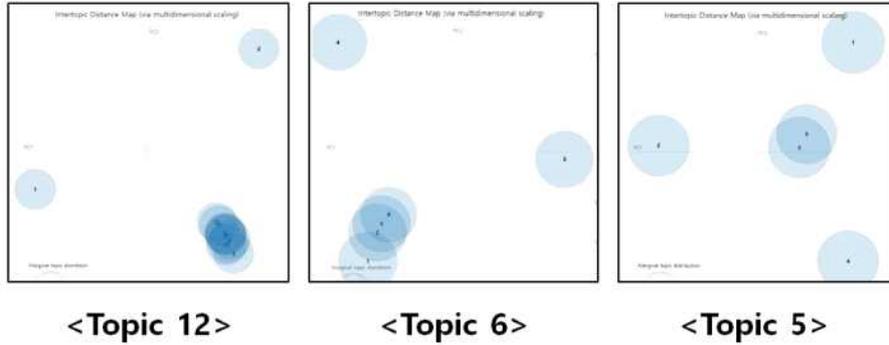


[그림 IV-1] LDA의 시각적 표현(Blei et al. 2003, p. 997)

[그림 IV-1]에서 M은 전체 문서의 집합을 나타내고 N은 특정 문서에 속한 단어의 집합이다. θ 는 문서들이 각 주제들에 속할 확률분포를 나타내며 디리클레분포의 매개변수인 α 에 의해 결정된다. z 는 문서 내의 단어들이 주제들에 속할 확률분포를 나타내며, 이는 θ 에 의한 다항분포로 선택된다. β 는 각 주제가 특정 단어를 생성할 확률을 나타내는 확률분포를 뜻한다. 결국 z 와 β 에 의해 실제 문서들의 단어분포인 w 가 결정된다. 이 모형에서 w 만이 실제로 문서들을 통해주어진 분포이고 나머지는 모두 잠재변수이다. LDA 알고리즘에서는 주어진 문서와 토픽들의 사전확률 분포인 α 와 토픽 내에서 단어의 사전확률분포인 β 의 파라미터 값을 활용해 반복적인 시뮬레이션을 통해 z 와 θ 를 추정한다(박상언·이병량, 2018).

LDA 분석 시에는 반복수행 횟수, 토픽의 개수를 사전에 설정해야 한다(유예림, 2017). 토픽 모형 생성의 반복수행 횟수는 분석 시간과 반비례 관계에 있기 때문에, 분석의 정확성과 효율성 모두를 충족하기 위해서는 적당한 수를 설정해 주어야 한다. 선행연구들에서 보통 1,000회에서 2,000회 사이의 값을 권고(안주영 외, 2016)하는 점을 참고하여, 본 연구에서는 반복횟수를 1,000회로 설정하였다. 다음으로 토픽의 개수를 설정하는 방법으로는 최적화한 확률모델에서 토픽을 비모수 통계로 추정하는 방법(Teh et al, 2012), 많은 토픽을 생성하여 LDA를 실시 후 유사한 토픽들끼리 결합함으로써 최종 결과를 생성하는 방법(Song, 2010 ; Yu, 2014), 여러 경우의 수를 고려하여 토픽 분석을 실시한 후 용어 분류가 유의미 하거나 정확성을 높게 가지는 결과를 선택하는 방법(Griffiths & Steyvers, 2004)이 있다.

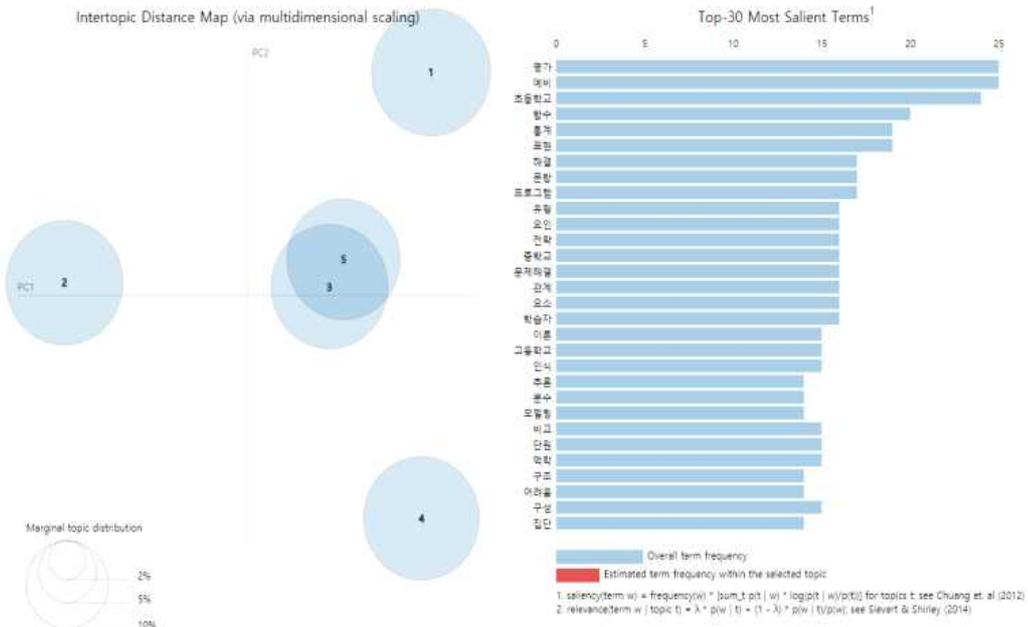
본 연구에서는 적절한 토픽수를 결정하기 위해 토픽수를 5개부터 12개까지 다양하게 변화시키면서 파일럿 LDA 분석을 실시하였고 시각적으로 토픽의 관계를 보여주는 IDM(Intertopic Distance Map)을 참고하여 토픽의 개수를 결정하였다([그림 IV-2]).



[그림 IV-2] Topic 개수 별 IDM

토픽수가 6개 이상일 경우에는 크게 3개의 그룹으로 분류된 뒤 두 그룹에 토픽이 하나씩 포함되고 나머지 한 그룹에 남은 모든 토픽이 들어가는 것을 확인할 수 있었다. 토픽수가 5개일 때는 4개의 그룹으로 분류된 뒤 하나의 그룹에서만 두 토픽이 나타나는 것으로 보아 가장 유의미하게 분류된다고 판단하였다. 이를 토대로 토픽의 개수를 5개로 설정하여 수학교육 연구의 잠재토픽을 추출하였다.

토픽을 추출한 후에는 각 토픽을 구성하고 있는 단어들을 비교 분석한 뒤 이론적 배경, 수학교육 연구의 맥락 등을 종합적으로 고려하여 토픽을 구성하는 단어들이 말하는 핵심 주제를 추론하였다. 마지막으로 IDM을 통해 토픽 모델링 결과를 시각화하여 토픽 간의 관계를 효과적으로 확인할 수 있었다. 본 연구에서 활용한 토픽의 단어들은 TF-IDF(Term Frequency-Inverse Document Frequency) 분석을 통해 추출한 단어들을 활용하였다.



[그림 IV-3] 전체 토픽에 대한 IDM

[그림 IV-3]은 전체 토픽에 대한 IDM(Intertopic Distance Map)으로 앞서 실시한 토픽 모델링의 결과를 시각적으로 표현해 준다. IDM을 살펴보면 우선 토픽들이 크게 4개의 그룹으로 분류된 것을 확인할 수 있다. 적정 모형을 찾아 나가는데 있어서 토픽의 교집합을 최소화하면서 찾아나가는데 예를 들어, 토픽 3과 토픽 5는 같은 그룹처럼 보이지만 단어의 구성을 보면 주제가 다른 영역임을 알 수 있다.

2) 토픽 트렌드

토픽 트렌드는 토픽 모델링을 실시한 후 이를 토대로 일정 기간 동안에 토픽들의 가중치의 변화를 나타낸 것이다. 학술지에 게재된 논문들에 대해 연도별 토픽 트렌드를 계산하기 위해 이전 연구에서 사용된 적이 있는 다음과 같은 공식을 적용하였다(Kim et al, 2015).

$$\varphi^{(t)} = \frac{\sum_{d=1}^{D_t} \theta^{(d)}}{D_t}$$

$\varphi^{(t)}$: 연도 t의 토픽 분포

D_t : 연도 t의 논문 수

$\theta^{(d)}$: 논문 d의 토픽 분포

위 공식은 결과적으로 해당 연도에 더 많은 논문에서 비중이 더 높게 다뤄진 토픽이 높은 가중치를 갖게 된다(박상언·이병량, 2018). 토픽 트렌드는 최근 다양한 분야에서 동향을 파악하기 위한 방법으로 활발하게 사용되고 있다. 본 연구에서는 위 공식을 사용하여 토픽 트렌드를 분석하였고 각 토픽들의 연도별 비율 변화를 분석하였다. 이를 통해 최근 수학교육 연구에서 나타나는 관심도가 높은 주제(hot topic)와 낮은 주제(cold topic)를 추론하여 수학교육의 연구 동향에 대한 심층적인 분석을 시도하였다.

V. 분석 결과

1. 토픽 모델링

<표 V-1> 토픽 모델링 분석 결과 추출된 5개의 토픽(Topic)

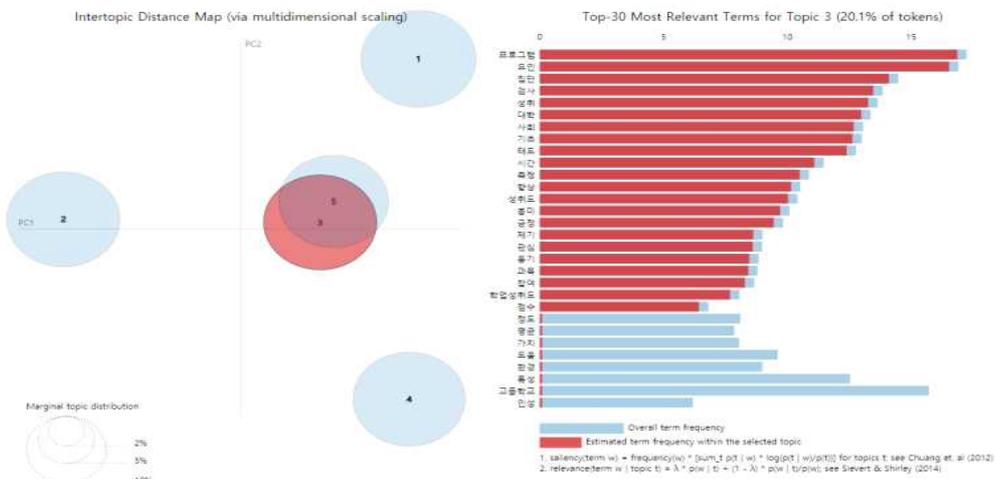
Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
요소	예비	프로그램	평가	통계
비교	중학교	요인	초등학교	표현
단원	학습자	집단	함수	해결
맥락	이론	검사	문항	유형
구성	고등학교	성취	문제해결	전략
역량	인식	대학	관계	추론
도입	일반	사회	모델링	분수
특징	기하	기초	도구	구조
탐구	기회	태도	계산	어려움
교실	특성	시간	사용	영제

토픽 2는 ‘예비’, ‘중학교’, ‘학습자’, ‘이론’, ‘고등학교’, ‘인식’, ‘일반’, ‘기하’, ‘기회’, ‘특성’이라는 단어들로 구성된다. 이 단어들의 의미와 맥락을 분석해보면 중·고등학교 학습자 특성 분석에 관한 단어들로 추론된다.

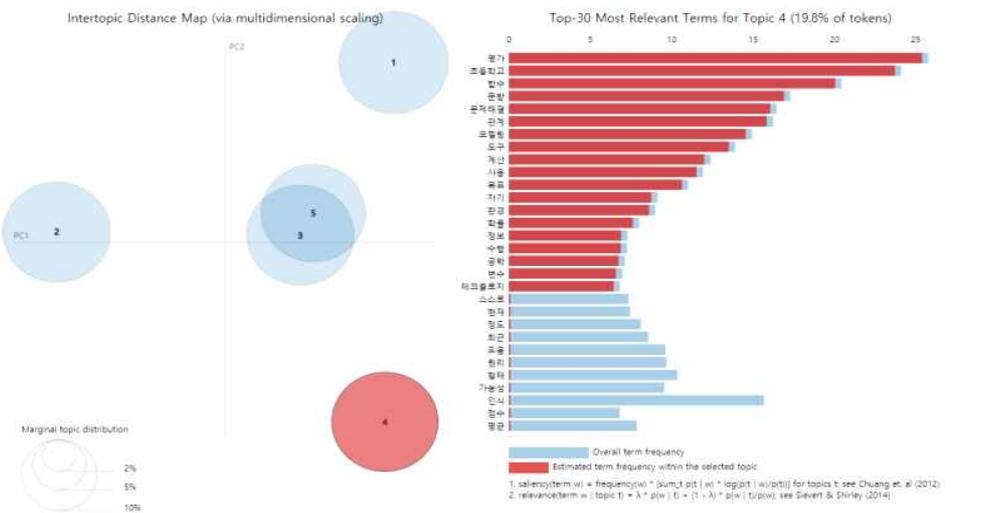
[그림 V-2]는 토픽 2를 시각화한 IDM으로써 오른쪽에 중요 단어가 나타나는데 이는 의미 있는 빈도 높은 주제들을 TF-IDF의 값을 이용하여 추출한 후 토픽 모델링한 결과이다.

토픽 3은 ‘프로그램’, ‘요인’, ‘집단’, ‘검사’, ‘성취’, ‘대학’, ‘사회’, ‘기초’, ‘태도’, ‘시간’이라는 단어들로 구성된다. 이 단어들의 의미와 맥락을 분석해보면 수학교육의 사회적 맥락, 학업성취도 검사에 관한 단어들로 추론된다.

[그림 V-3]은 토픽 3을 시각화한 IDM으로써 오른쪽에 중요 단어가 나타나는데 이는 의미 있는 빈도 높은 주제들을 TF-IDF의 값을 이용하여 추출한 후 토픽 모델링한 결과이다.



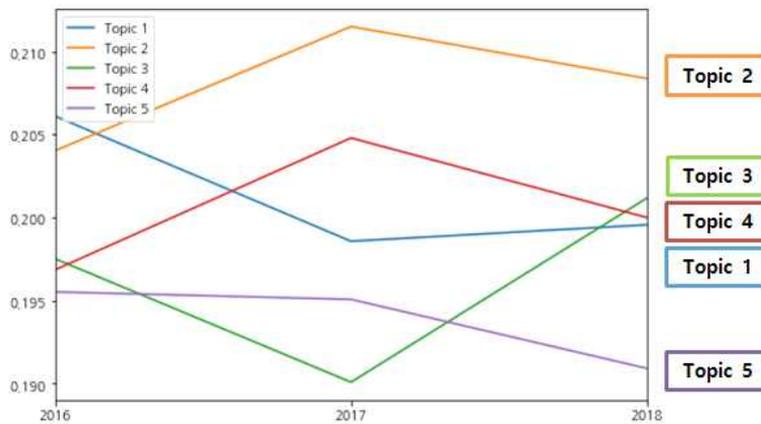
[그림 V-3] Topic 3에 대한 IDM



[그림 V-4] Topic 4에 대한 IDM

에는 토픽 1이 제일 많은 비중을 차지했고 토픽 2, 토픽3, 토픽 4, 토픽 5가 차례대로 비중이 적어졌다. 2017년의 경우 3개년도 중에 토픽별 차이가 가장 많이 났는데 토픽 2가 가장 많은 비중을 차지했고, 토픽 4, 토픽 1, 토픽 5, 토픽 3으로 비중이 점점 적어졌다. 2018년에는 2017년과 마찬가지로 토픽 2가 가장 많은 비중을 차지했고 다음으로 토픽 3, 토픽 4, 토픽 1이 비슷하게 나타났고 토픽 5가 가장 적은 비중을 보였다.

다음으로 연도별 토픽 비율을 시각화하여 토픽 트렌드를 살펴보면 [그림 V-6]와 같다. 이 경우 토픽의 변화 흐름을 분석하기 용이하다. 토픽 1의 경우 2016년에는 가장 많은 비중을 차지했지만 2017년에는 3번째, 2018년에는 4번째에 위치하면서 점점 비율이 낮아진 토픽으로 해석된다. 토픽 2의 경우 2016년에는 2번째로 많은 비중을 차지하고 2017년과 2018년에는 제일 많은 비중을 차지함으로써 5가지의 토픽 중 가장 관심이 많은 주제라고 판단된다. 토픽 3은 2016년에는 3번째에 위치했지만 2017년에는 제일 낮은 비중을 차지하였고, 2018년에는 2번째로 높은 비중을 차지함으로써 연도별로 차이가 큰 토픽이라고 할 수 있다. 토픽 4의 경우 2016년에는 4번째 비중을 차지했고 2017년에는 2번째로 많은 비중을, 2018년에는 3번째로 많은 비중을 차지했다. 마지막으로 토픽 5의 경우 2016년에는 제일 적은 비중을 차지했고 2017년에는 4번째, 2018년에는 다시 제일 낮은 비중을 차지함으로써 5개의 토픽 중에는 가장 낮은 비중으로 다루어진 것으로 나타났다.



[그림 V-6] 연도별 토픽 트렌드 그래프

VI. 나오는 말

본 연구는 텍스트마이닝 기법을 활용하여 수학교육에 대한 연구 동향을 살펴보고자 최근 3년간(2016년~2018년)의 논문, 총 460편의 논문 텍스트를 가지고 텍스트마이닝 기법을 이용하여 분석하였다. 그동안 수학교육 연구에서는 연구의 동향을 파악하기 위한 연구들이 거의 매년 실시되고 있는 추세이다. 본 연구는 수학교육 분야의 연구 동향을 파악하기 위한 계보로서의 중요성 외에도 텍스트마이닝 기법 중에서 동향을 추측하는데 보다 적합한 모델을 찾고 이를 통해 분석을 새롭게 시도했다는 점에서도 그 의의를 찾을 수 있다. 여기서 활용한 방법은 주제들 간의 관계가 유의미한 토픽끼리 묶인 집합들을 추출함으로써, 수학교육 연구 안에서 다루고 있는 주요 내용들은 무엇이며 이를 주제별로 묶으면 어떠한 주제들을 다루고 있는지를 통해 수학교육 연구의 흐름과 변화를 추정할 수 있었다.

본 연구 결과에서 나온 결과들은 먼저, TF-IDF 분석 결과를 토대로 한 토픽 모델링 기법 중 LDA를 활용하

여 토픽 분석 한 결과, 5개의 토픽을 추출할 수 있었다. 즉, 토픽을 구성하는 단어를 조합하여 최근 3개년 연구 논문들의 핵심 주제를 알 수 있었다. 이를 연도별 토픽 트렌드까지 같이 분석한 결과 가장 많이 등장한 주제는 ‘예비’, ‘중학교’, ‘학습자’, ‘이론’, ‘고등학교’, ‘인식’, ‘일반’, ‘기하’, ‘기회’, ‘특성’으로 구성된 토픽 2와 ‘평가’, ‘초등학교’, ‘함수’, ‘문항’, ‘문제해결’, ‘관계’, ‘모델링’ ‘도구’, ‘계산’, ‘사용’인 토픽 4로 나타났다.

2003년 동향 연구(이강섭, 2003)에서도 산출된 용어인 ‘문제해결’, ‘모델링’과 같은 주제는 수학교육에서는 지속적으로 수행되어오고 있는 연구라 할 수 있다. 단지 예전의 동향 분석 결과와 토픽 모델링 분석과 차별화되는 것은 ‘도구’와 ‘계산’과 같은 키워드가 함께 한다는 것이다. 다시 말하면 최근 연구들에서는 수학교실에서 계산량을 줄이고 도구를 활용한 문제해결과 수학 모델링 연구들이 수행되고 있다고 볼 수 있다. 흥미로운 사실은 이강섭(2003) 연구와는 별개로 진행되었던 연구(박경미, 2003)에서는 당시의 연구 동향에 ICT의 활용에 대한 연구가 활성화되고 있다는 분석 결과를 제시한 바 있다. 따라서 이 두 연구 결과를 조합해 보면 같은 토픽 모델링 분석이 유사한 결과를 도출하였다는 것을 알 수 있다.

2008년 수행된 동향분석(권정은, 최재호, 2008)에서 수업 설계 및 방법, 교육과정 및 교과서 분석 연구가 비중이 높게 나타났다고 보고한 바 있는데 본 분석에서도 교육과정 및 교과서 연구, 수업 설계 및 구성에 해당되는 토픽 1은 지속적으로 수행되고 있는 주제라 할 수 있다. 또한 학생들의 학습과 사고(오개념, 인지발달)에 관한 연구가 가장 주류를 이루었던 2009년 동향 분석 연구(김수연 외, 2009)와는 다르게 이러한 단어들 보다는 ‘역량’, ‘탐구’ 등이 키워드가 함께 동일 주제를 이루었다. 이는 역량 중심 교육과정의 변화와 발맞춘 트렌드라 해석된다.

또한 2010년 이후의 모든 수학교육 연구 동향 연구에서의 특징은 교육과정 연구가 주요 연구주제가 되었다는 것이다(하수연 외, 2010; 박선영·김원경, 2011; 김유경·방정숙, 2017). 이는 토픽 모델링 분석에서도 동일한 결과로써 교육과정 및 교과서를 분석한 연구 역시 비중이 높은 것으로 나타났다. 이는 2015 개정 교육과정과 개정 교과서가 학교에 적용하게 됨에 따라 이에 대한 연구도 같이 활성화 된 것으로 보인다.

반면 수와 연산, 도형에 대한 연구가 활발하지만 확률과 통계에 대한 연구가 미진하다(하수연 외, 2010)는 보고와는 다르게 본 연구에서는 토픽 5에 확률과 통계 연구가 주요 주제로 도출되었다. 또한 평가에 대한 연구가 매우 미진하다는 연구(김동중 외, 2014)와는 다르게 평가가 주요 연구 토픽 단어로는 도출되었고 2018년에는 그 빈도가 더욱 향상되기도 하였다. 그러나 평가를 세분화한 단어들이 눈에 띄지는 않은 것으로 미루어 평가에 대한 다양한 측면에서의 연구는 아직 부족한 실정이라 할 수 있다. 따라서 다양한 평가가 매우 중요해지는 시점에서 보다 구체화된 평가 연구가 활성화 될 필요가 있을 것이다.

또한 박경미(2003) 연구에서는 우리나라 연구에는 형평성과 사회적 정의와 수학교육이 연관된 논문이 부재하다는 결과를 제시하였는데, 토픽 모델링 분석에서도 토픽 3에 관련 단어가 등장하기는 하였으나 직접적인 단어가 도출되지 않은 것으로 미루어 그 이후에도 크게 달라진 바 없는 것으로 추측할 수 있다. 따라서 이러한 주제들은 수학교육에도 관심 가져야 할 주제 중 하나라 할 수 있다.

마지막으로, 토픽 모델링 분석을 통해 최근 중·고등학교 학습자 특성 분석에 관한 연구 매우 활발해지고 있다는 것을 알 수 있다. 이는 학생의 수준과 입장을 고려한 교수학습이 이루어져야 한다(예, 방숙희 강완, 2017; 최지선, 2017)는 최근의 주장에 따라 학습자의 특성을 분석하고 이를 고려한 교수·학습 연구가 많이 이루어지고 있음을 알 수 있다. 반면, 수학적 사고, 문제해결 전략에 관한 연구들의 경우 주요 5개 토픽 중에서는 다소 낮은 빈도로 나타나기는 하였으나 여전히 비중 있게 수행되고 있는 주제이기도 하다. 또한 이러한 주제는 수학교육 연구에서 기본적인 주제이고 지속적으로 수행되어야 할 연구라 할 수 있다(박선영·김원경, 2011). 따라서 향후에도 다양한 다른 주제들과 융합하여 다각도에서 수행될 필요와 가치가 있다.

참고문헌

- 고호경 · 최영우 · 박선정 (2014). 수학교육에서 빅데이터 활용 방안에 대한 소고, 한국수학교육학회지 시리즈E, **28(4)**, 573-588.
- Ko, H., Choi, Y. & Park, S. (2014). Study on Big Data Utilization Plans in Mathematics, *J. Korea Soc. Math. Ed. Ser. E: Communications of Mathematical Education*, **28(4)**, 573-588.
- 권정은 · 최재호 (2008). 우리나라 초등수학교육 연구의 동향 분석-국내 학술지를 중심으로, 한국초등수학교육학회지, **12(2)**, 149-163.
- Kwon, J. & Choi, J. (2008). An Analysis of Trends in Elementary Mathematics Education Research- Focussing on Mathematics Education Journals in Korea, *Journal of Elementary Mathematics Education in Korea*, **12(2)**, 149-163.
- 김대곤 · 원진영 (2013). 복합적 미래예측방법론 분석을 통한 미래재난예측기법 개발, 국립재난안전연구원, 25-35.
- Kim, D. & Won, J. (2013). *Development of Foresight Method for Future Disaster Through the Analysis of Complex Foresight Methodology*. National Disaster Management Institute, 25-35.
- 김동중 · 배성철 · 김원 · 이다희 · 최상호 (2014). 수학교육연구 및 혼합 연구방법 동향-최근 10년간 발표된 국내 학술지 논문을 중심으로, 수학교육 논문집, **28(3)**, 303-320.
- Kim, D., Bae, S., Kim, W., Lee, D. & Choi, S. (2014). Trends of Mathematics Education Research and Mixed Methods-Focusing on Domestic Mathematics Education Journals for the Last 10 years. *Communications of Mathematical Education*, **28(3)**, 303-320.
- 김래영 · 김구연 · 권나영 (2012). 연구 설계 및 연구 방법의 최근 동향: 초·중등 수학과 교육과정에 관한 연구를 중심으로, 학교수학, **14(3)**, 395-408.
- Kim, L., Kim, G. & Kwon, N. (2012). Trends in Research Design and Methods: Research on Elementary and Secondary Mathematics Curriculum, *School Mathematics*, **14(3)**, 395-408.
- 김선아 · 박진희 · 이현정 · 정유진 (2016). 텍스트마이닝 기법을 활용한 다문화 미술교육 연구 동향 분석 연구, 다문화교육연구, **9(2)**, 203-227.
- Kim, S., Park, J., Lee, H. & Chung, Y. (2016). A study on the research trends of multicultural art education using the text mining technique, *Multicultural Education Studies*, **9(2)**, 203-227.
- 김선희 · 김수민 (2018). 언어 네트워크 분석법을 이용한 최근 수학교육 연구 동향 탐색-2017년 국제 수학교육 학술대회 발표 논문을 중심으로, 학교수학, **20(4)**, 591-608.
- Kim, S. & Kim, S. (2018). The International Research Trends of Mathematical Education by Language Network Analysis Method-Based on the Titles of 2017 International Conference, *School Mathematics*, **20(4)**, 591-608.
- 김성근 · 조혁준 · 강주영 (2016). 학술연구에서의 텍스트 마이닝 활용 현황과 주요분석기법, 정보화연구, **13(2)**, 317-329.
- Kim, S., Cho, H. & Kang, J. (2016). The Status of Using Text Mining in Academic Research and Analysis Methods, *The Journal of Information Technology and Architecture*, **13(2)**, 317-329.
- 김수연 · 김영록 · 장재덕 (2009). 논문집 시리즈 A <수학교육>에 게재된 논문들의 분류와 분석: 2000년부터 2008년까지 게재된 논문들을 중심으로, 수학교육 논문집, **23(3)**, 638-705.
- Kim, S., Kim, Y. & Jang, J. (2009). Classifications and analysis of articles in Journal series A <the mathematics education> of Korean Society of Mathematics Education, *Communications of Mathematical Education*, **23(3)**, 638-705.
- 김영록 (2017). 미래 지능정보사회에 대비한 수학교육, 한국수학교육학회 학술발표논문집, **2017(1)**, 175-178.
- Kim, Y. (2017). Mathematics Education for Euture Intelligence Information Society. *The Korean Society of Mathematical*

- Education Academic Publications*, **2017(1)**, 175-178.
- 김영우 (2017). Do it! 쉽게 배우는 R 데이터 분석, 서울: 이지스퍼블리싱.
- Kim, Y. (2017). *Do it! R for Data Analysis*, Seoul: EasysPublishing.
- 김유경·방정숙(2017). 초등수학교육 연구동향: 최근 7년간 게재된 국내 학술지 논문을 중심으로, 초등수학교육, **20(1)**, 19-36.
- Kim, Y. & Pang, J. (2017). Research Trends in Elementary Mathematics Education: Focused on the Papers Published in Domestic Journals During the Resent Seven Years, *Education of Primary School Mathematics*, **20(1)**, 19-36.
- 김하진·송민 (2014). 동시출현단어 분석을 통한 국내외 정보학 학회지연구동향 파악. 정보관리학회지, **31(1)**, 99-118.
- Kim, H. & Song, M. (2014). A Study on the Research Trends in Domestic/International Information Science Articles by Co-word Analysis. *Journal of the Korean Society for Information Management*, **31(1)**, 99-118.
- 김현희·이혜영 (2016). 토픽 네트워크 분석을 활용한 데이터 마이닝 분야 연구 논문 분석, 한국컴퓨터정보학회 논문지, **21(5)**, 141-148.
- Kim, H. & Rhee, H. (2016). Trend Analysis of Data Mining Research Using Topic Network Analysis, *Journal of the Korea Society of Computer and Information*, **21(5)**, 141-148.
- 김혜미 (2016). 수학 문제해결에 관한 국내 연구 동향 분석. 학습자중심교과교육연구, **16(8)**, 831-850.
- Kim, H. (2016). An Analysis of the Research Trend in Korea Regarding Mathematical Problem Solving. *The Journal of Learner-Centered Curriculum and Instruction*, **16(8)**, 831-850.
- 나귀수·김연·이수진·박미미·김동원 (2018). 미래 시대의 수학교육 방향에 대한 연구, 수학교육학연구, **28(4)**, 437-478.
- Na, G., Kim, Y., Lee, S., Park, M. & Kim, D. (2018). Exploring the Direction of Mathematics Education in the Future Age, *The Journal of Educational Research in Mathematics*, **28(4)**, 437-478.
- 류지혜 (2018). 토픽 모델링을 이용한 보건학 분야 국내학술지 연구동향 분석, 고신대학교 대학원 석사학위 논문.
- Ryu, J. (2018). *Analysis of Research Trends in Korean Academic Journals on Health Sciences Using Topic Modeling*, Kosin University Graduate School.
- 민경아·유미현·고호경 (2011). 수학영재교육 관련 국내 연구 동향 분석, 한국학교수학회논문집, **14(3)**, 393-421.
- Min, K., Yoo, M. & Ko, H. (2011). An Analysis of Research Trend in Domestic Mathematics Gifted Education. *Journal of the Korean School Mathematics*, **14(3)**, 393-421.
- 박경미 (2003). 중등 수학교육 연구의 경향 분석-〈수학교육〉과 〈JRME〉에 수록된 논문 비교를 중심으로, 수학교육, **42(2)**, 211-230.
- Park, K. (2003). An Analysis of the Secondary Mathematics Education Research Trends in Korea by Comparing the Papers on <The Mathematical Education> and Those on <Journal for Research in Mathematical Education>, *The Mathematical Education*, **42(2)**, 211-230.
- 박상언·이병량 (2017). 텍스트 마이닝 기법을 활용한 한국 문화정책 연구 경향 분석, 한국거버넌스학회보, **24(3)**, 95-119.
- Park, S. & Lee, B. (2017). Trend Analysis of Korean Cultural Policy Studies Using Text Mining, *The Korean Governance Review*, **24(3)**, 95-119.
- 박선영 (2011). 국내외의 중등수학교육 연구 동향, 한국교원대학교 대학원 석사학위 논문.
- Park, S. (2011). *Trends in Researches on Secondary Mathematics Education in Korea and Other Countries*, Graduate School of Education Korea National University of Education.
- 박선영·김원경 (2011). 국내외 수학교육 연구 동향 비교 분석, 수학교육, **50(3)**, 285-308.
- Park, S. & Kim, W. (2011). A Comparative Analysis on Research Trends of Secondary Mathematics Education between

- Korea and Overseas, *The Mathematical Education*, **50(3)**, 285-308.
- 박종희 · 박은정 · 조동준 (2015). 북한 신년사(1946-2015)에 대한 자동화된 텍스트 분석, 한국정치학회보, **49(2)**, 27-61.
- Park, J., Park, E. & Jo, D. (2015). Automated Text Analysis of North Korean New Year Addresses, 1946-2015. *Korean Political Science Review*, **49(2)**, 27-61.
- 박참솔 (2017). 텍스트마이닝을 이용한 학술지 논문 동향 비교 분석, 호서대학교 대학원 석사학위 논문.
- Park, C. (2017). *Comparative Analysis of Journal Thesis Trends Using Text Mining*, The Graduate School Hoseo University.
- 방수희 · 강완 (2017). 학습자의 인지양식이 수학 과제 유형별 수학적 의사소통에 미치는 영향, 한국초등수학교육학회지, **21(4)**, 621-641.
- Bang, S. & Kang, W. (2017). Influences of Cognitive Styles on Students' Mathematical Communication by Types of Mathematical Tasks, *Journal of Elementary Mathematics Education in Korea*, **21(4)**, 621-641.
- 배규용 · 박주현 · 김정선 · 이영섭 (2013). 텍스트 마이닝 기법을 활용한 기후변화관련 식품분야 논문초록 분석, 한국데이터정보과학회지, **24(6)**, 1429-1437.
- Bae, K., Park, J., Kim, J. & Lee, Y. (2013). Analysis of the abstracts of research articles in food related to climate change using a text-mining algorithm, *Journal of the Korean Data And Information Science Society*, **24(6)**, 1429-1437.
- 백영민 (2017). R을 이용한 텍스트 마이닝, 파주: 한올아카데미.
- Baek, Y. (2017). *Text-Mining using R*. Paju: HanulAcademy.
- 서일원 · 전채남 (2013). 빅데이터 분석의 기술마케팅 활용에 관한 연구: 잠재 수요기업 발굴을 중심으로, 마케팅논집, **21(2)**, 181-203.
- Seo, I. & Jun, C. (2013). Analyzing the Bigdata for Practical Using into Technology Marketing: Focusing on the Potential Buyer Extraction, *Journal of Marketing Studies*, **21(2)**, 181-203.
- 손복은 · 고효경 (2018). 수학 담화에서 나타나는 교사의 감성적 언어 빈도 분석, 수학교육논문집, **32(4)**, 455-475.
- Son, B., & Ko, H. K. (2018). The Frequency Analysis of Teacher's Emotional Response in Mathematics Class, *Communications of Mathematical Education*, **32(4)**, 455-475.
- 송태민 · 송주영 (2016). R을 활용한 소셜 빅데이터 연구방법론, 서울: 한나래출판사.
- Song, T. & Song, J. (2016). *Social Big Data Research Methodology with R*, Seoul: Hannarae Publishing CO.
- 송혜지 · 박경수 · 정혜은 · 송민 (2013). 텍스트마이닝 기법을 활용한 한국의 경제연구 동향분석, 한국정보관리학회 학술대회 논문집, **20**, 47-50.
- Song, H., Park, K., Jung, H. & Song, M. (2013). Trend Analysis of Korean Economy in the Economic Literature by text mining techniques, *Korean Society for Information Society Conference Symposium*, **20**, 47-50.
- 신규식 · 최희련 · 이홍철 (2015). 신재생에너지 동향 파악을 위한 토픽모형 분석, 한국산학기술학회 논문지, **16(9)**, 6411-6418.
- Shin, K., Choi, H. & Lee, H. (2015). Topic Model Analysis of Research Trend on Renewable Energy, *Journal of Korea Academia-Industrial cooperation Society*, **16(9)**, 6411-6418.
- 안주영 · 안규빈 · 송민 (2016). 텍스트 마이닝을 이용한 매체별 에볼라 주제분석, 한국문헌정보학회지, **50(2)**, 289-307.
- An, J., Ahn, K. & Song, M. (2016). Text Mining Driven Content Analysis of Ebola on News Media and Scientific Publications. *Journal of the Korean Library and Information Science Society*, **50(2)**, 289-307.
- 유예림 (2017). 빅데이터 분석 기법을 활용한 2015 개정 교육과정 정책에 대한 언론보도 분석, 서울대학교 대학원 박사학위 논문.
- Yu, Y. (2017). *Analysis of media coverage on 2015 revised curriculum policy using Big Data Analysis*, The Graduate

- School Seoul National University.
- 유예림·백순근 (2016). 자동화된 텍스트 분석을 활용한 2015 개정 교육과정 정책에 대한 언론 보도의 쟁점 분석, 교육과정평가연구, **19(3)**, 127-156.
- Yu, Y. & Baek, S. (2016). Issue Analysis of the Related Mass Media's News Articles on the 2015 Revised National Curriculum Using Automated Text Analysis, *The Journal of Curriculum and Evaluation*, **19(3)**, 127-156.
- 육동인 (2017). 텍스트 마이닝을 활용한 직업학 연구동향 분석, 한국산학기술학회논문지, **18(3)**, 586-599.
- Yook, D. (2017). Text Mining-Based Analysis for Research Trends in Vocational Studies, *Journal of Korea Academia-Industrial cooperation Society*, **18(3)**, 586-599.
- 이강섭 (2003). <수학교육>에 게재된 논문의 분류와 분석- 통권 1호부터 통권 99호까지, 수학교육, **42(2)**, 227-237.
- Lee, K. (2003). A Classification and Analysis of the Articles in <the Mathematical Education>-From issue 1 to issue 99, *The Mathematical Education*, **42(2)**, 227-237.
- 이수진·이종학·김원경 (2013). 수학교육공학 연구 동향의 비교·분석, 교사교육연구, **52(2)**, 253-266.
- Lee, S., Lee, J. & Kim, W. (2013). A Study on Domestic and Foreign Research Trends of Mathematics Educational Technology. *Teacher Education Research*, **52(2)**, 253-266.
- 이인선·나은영 (2018). 텍스트마이닝을 이용한 인지분야 학회지 분석(2000-2017), 인문사회21, **9(3)**, 415-426.
- Yi, I. & Na, E. (2018). A Study on the Journal Analysis of Cognitive Field Using Text Mining (2000~2017), *The Journal of Humanities and Social science*, **9(3)**, 415-426.
- 임동훈 (2015). R을 이용한 빅데이터 분석, 과주: 자유아카데미.
- Lim, D. (2015). *Big Data Analysis using R*, Paju: FreeAcademy.
- 임화진 (2014). 빅데이터를 이용한 충남도 정책 키워드 분석, 충남발전연구원, 전략연구 2014-15.
- Lim, H. (2014). *Keyword Analysis of Chungnam Policy Using Big Data*, Chungnam Development Institute, Strategic Research 2014-15.
- 정근하 (2010). 텍스트 마이닝과 네트워크 분석을 활용한 미래예측 방법 연구, 한국과학기술기획 평가원.
- Jung, G. (2010). *A Study of foresight method based on textmining and complexity network analysis*, Korea Institute of S&T Evaluation and Planning.
- 조태호 (2001). 텍스트 마이닝의 개념과 응용, 지식정보인프라, **5**, 76-85.
- Jo, T. (2001). Concepts and Applications of Text Mining, *Journal of Scientific & Technological Knowledge Infrastructure*, **5**, 76-85.
- 채미현 (2018). 환경복지 정책 모형 설정에 관한 탐색적 연구-텍스트마이닝을 중심으로, 서울대학교 행정대학원 석사학위 논문.
- Chae, M. (2018). *An Exploratory Study on Establishing Environmental Welfare Policy Model-Based on Text Mining*, The Graduate School of Public Administration Seoul National University.
- 최은미 (2017). 4차 산업혁명에서의 STEAM과 기초수학 교육의 방향 연구, 한국교양교육학회 학술대회 자료집, 22-26.
- Choi, E. (2017). A Study on the Direction of STEAM and Basic Mathematics Education in the Fourth Industrial Revolution, *The Korean Association of General Education Conference Catalog*, 22-26
- 최지선 (2017). 수학 '확률과 통계' 영역에서의 중학교 2학년 학생의 성취 특성, 학습자중심교과교육연구, **17(21)**, 145-167.
- Choi, J. (2017). Characteristics of Korean Eighth Grade Students' Achievement in Probability and Statistics, *The Journal of Learner-Centered Curriculum and Instruction*, **17(21)**, 145-167.
- 최택영·송병근 (2001). 1990년도 우리나라 수학교육 연구 동향 분석: 석사학위 논문을 중심으로, 수학교육,

- 40(1), 77-92.
- Choi, T. & Song, B. (2001). Movement of Research in Mathematics Education in 1990's-focused on the master's theses in Korea. *The Mathematical Education*, 40(1), 77-92.
- 탁병주 · 이경화 (2017). 우리나라 통계교육 연구의 동향 분석-2000년 이후 발행된 국내 통계교육 연구논문을 중심으로, *수학교육학연구*, 27(2), 269-289.
- Tak, B. & Lee, K. (2017). An Analysis of Research Trends on Statistics Education in Korea from 2000 to 2016. *The Journal of Educational Research in Mathematics*, 27(2), 269-289.
- 하수현 · 방정숙 · 주미경 (2010). 초등수학교육 연구동향-최근 5년간 게재된 국내 학술지 논문을 중심으로, *한국수학교육학회지*, 49(1), 67-83.
- Ha, S., Pang, J. & Ju, M. (2010). Research Trends in Elementary Mathematics Education: Focused on the Papers Published in Domestic Journals During the Recent 5 Years, *The Mathematical Education*, 49(1), 67-83.
- 황명화 (2014). 여론 모니터링을 위한 비정형 빅데이터 시공간분석 방법론 연구, 국토연구원, 국토연 2014-10.
- Hwang, M. (2014). *Spatiotemporal Analysis of Unstructured Big Data for Public Opinion Monitoring*, Korea Research Institute for Human Settlements, 2014-10.
- Blei, D. M. (2012). Probabilistic topic models, *Communications of the ACM*, 55(4), 77-84.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 993-1022.
- Feldman, R., & Dagan, I. (1995). Knowledge discovery in textual databases(KDT), *KDD*, 95, 112-117.
- Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics, *Proceedings of the National academy of Sciences*, 101(suppl1), 5228-5235.
- Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 1-31.
- Ian, F. (2014). *wordcloud: Word Clouds. R package version 2.5* <https://CRAN.R-project.org/package=wordcloud>.
- Kim, W. J., Won, J. H., Park, S. U., & Kang, J. Y. (2015). Demand forecasting models for medicines through wireless sensor networks data and topic trend analysis. *International Journal of Distributed Sensor Networks*, 2015(10), 1-6.
- Song, Z. (2010). *Research on text categorization based on LDA*. Master's Degree Dissertation. Xi'an University of Technology, Xi'an, China.
- Teh, Y. W., Jordan, M. I., Beal, M. J., & Blei, D. M. (2012). Hierarchical dirichlet processes. *Journal of the American Statistical Association*, 101(476), 1566-1581.
- Turney, P. D., & Pantel, P. (2010). From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, 37, 141-188.
- Yu, S. Y. (2014). Exploratory study of developing a synchronization-based approach for multi-step discovery of knowledge structures. *Journal of Information Science Theory and Practice*, 2(2), 16-32.

Analysis of trends in mathematics education research using text mining

Jin, Mireu

Graduate School of Education
Ajou University
E-mail : mirfly19@ajou.ac.kr

Ko, Ho Kyoung[†]

Department of Mathematics Graduate School of Education
Ajou University
E-mail : kohoh@ajou.ac.kr

In order to understand the recent trends in mathematics education research papers, data mining method was applied to analyze journals of the mathematics education posterior to the year of 2016. Text mining method is useful in the sense that it utilizes statistical approach to understand the linkages and influencing relationship between concepts and deriving the meaning that data shows by visualizing the process. Therefore, this research analyzed the key words largely mentioned in the recent mathematics education journals. Also the correlation between the subjects of mathematics education was deduced by using topic modeling. By using the trend analysis tool it is possible to understand the vital point which researchers consider it as important in recent mathematics education area and at the same time we tried to use it as a fundamental data to decide the upcoming research topic that is worth noticing.

* ZDM Classification : C10

* 2000 Mathematics Subject Classification : 97B10

* Key words : text mining, topic modelling, trend analysis

* This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2016S1A5A2A01023238).

[†] corresponding author