

텍스트 마이닝 분석을 통한 수학교육 연구 동향 분석

A Text Mining Analysis for Research Trend about the Mathematics Education

진미르 · 고희경¹⁾

ABSTRACT. In this paper we used text mining method to analyze journals of mathematics education posterior to the year of 2016. To figure out trends of mathematics education research, we analyzed the key words largely mentioned in the recent mathematics education journals by Term Frequency and Term Frequency-Inverse Document Frequency method. We also looked at how these keywords match up with the key words that appear of education to prepare for future society. This result can infer the characteristics of mathematics education research in the aspect upcoming research topics.

I. 들어가는 말

우리 사회는 이미 제4차 산업혁명기에 진입했으며 이로 인해 우리의 삶에도 많은 변화가 예상될 것이라고 한다. 이러한 사회적 변화에 맞춰 교육 분야에서도 이를 대비하고자 하는 노력을 기울여야 한다는 당위성과 함께 제4차 산업혁명시대로 대변되는 미래사회를 대비한 교육이란 무엇인가에 대한 논의 역시 활발히 일고 있다. 예를 들어, 제4차 산업혁명이 만드는 사회는 소프트웨어 중심사회로 변화하고 있으며 사회의 구성원에게는 복합 문제 해결능력, 비판적 사고 능력, 창의력, 인적 자원 관리 능력, 협업능력, 감성 능력, 판단 및 의사 결정 능력, 서비스 지향성, 협상능력, 인지적

Received July 28, 2019; Revised August 20, 2019; Accepted August 29, 2019.

2010 Mathematics Subject Classification: 97D30

Key words: research trend, mathematics education, keyword, text mining

진미르 석사 논문 중 일부를 발췌하여 수정·보완하였음.

본 연구는 2018학년도 아주대학교 일반연구비 지원에 의하여 연구되었음.

1) Corresponding author.

유연력 등이 더욱 요구되고 있기 때문에 교육에서의 목표와 내용은 이와 함께 가야 한다는 것이다(류태호, 2017).

수학교육계에서 역시 다른 어떤 교과보다 4차 산업혁명시대 대비라는 거대 담론 하에 다양한 측면에서의 변화를 예측하고 방향을 가늠하는 노력을 기울이고 있다. 예를 들면, 수학이나 수학교육이 추구해야 하는 것이 무엇이며 학생들은 어떠한 역량을 함양해야 하는가에 대한 합의점을 도출하고 이를 적용하기 위한 방법들을 구안해나가고자 하고 있다(예, 최은미, 2017; 나귀수 외, 2018).

이와 더불어 수학교육연구 측면에서는 수학교육에 필요한 연구들이 얼마나 방향성 있게 진행되고 있는지를 판단하고 향후 연구의 방향을 설정하기 위한 연구들이 지속적으로 수행되어오고 있다. 그 중 대표적인 것이 선행된 연구들의 분석을 통해 동향을 분석하고 향후 연구의 방향을 제시하는 것이다(이수진 외, 2013; 김혜미, 2016). 이러한 연구는 학문이 발전하기 위한 여러 단계 중 하나로 학문영역에서 축적된 연구성과들에 대한 분석을 통하여 더 나은 연구진행을 촉구하는 방법으로써, 연구들의 대상, 주제, 방법 등 전반적인 연구동향에 대한 정기적 분석은 향후 나아가 할 방향을 제시하는 매우 의미 있는 작업이라 할 수 있다(육동인, 2017). 따라서 본 연구는 수학교육에서 수행된 연구들의 분석을 통해 수학교육 연구의 동향을 파악하고자 한다.

텍스트 마이닝은 온라인상에 축적된 대량의 데이터를 트렌드 분석에 활용하는 객관적이고 실효성 높은 기법이다(Feldman & Dagan, 1995). 텍스트 마이닝은 통계적인 접근을 통하여 개념 간의 연결성과 영향 관계를 파악하고 이를 시각화함으로써 데이터가 보여주는 의미를 도출하는 데에 유용하다(김선아 외, 2016). 이에 본 연구는 최근 수학교육 연구 논문들의 텍스트를 분석하기 위한 방법으로 텍스트 마이닝 기법을 활용하여 최근의 수학교육 연구에 높은 빈도로 등장한 키워드를 추출하고 이를 시각적으로 제시하고자 한다. 이를 위하여 2016년부터 2018년도까지 3개년도의 수학교육 학술지의 텍스트들을 분석하였고 이를 통하여 연구에서 주로 다루어준 관심사는 무엇인지 유추하고자 하였다. 이와 더불어 수학교육 연구 영역이 4차 산업혁명에 따른 교육의 변화를 얼마나 반영하고 있는지 키워드 분석을 통해 가늠하고자 하였다. 이는 수학교육 연구 분야에서의 미래사회 대비라는 키워드 측면에서 향후 주목하여야 할 연구 주제가 무엇인지를 판단하고 시사점을 제시함으로써 연구 분야의 지평을 넓히는데 기여하고자 한다.

II. 미래 사회 대비 수학교육에서의 키워드

1. 미래 사회 대비 교육에서의 키워드

OECD는 2018년에 발표한 ‘OECD 2030 학습 틀(The OECD Learning Framework 2030)’에서 ‘변환적 역량(Transformative competencies)’ 즉, ‘새로운 가치 창출(Creating new value)’, ‘긴장과 딜레마 조정(Reconciling tensions and dilemmas)’, ‘책

임지기(Taking responsibility)' 등 3가지의 역량을 제시하였다. 이 중 새로운 가치 창출 역량을 뒷받침하는 구성요소에는 적응력, 창의력, 호기심 및 개방성이 키워드로 들어간다. 긴장과 딜레마 조정 역량은 예를 들어, 평등과 자유, 자율성과 공동체, 혁신과 연속성, 그리고 효율성과 민주적 과정의 조화 등과 같은 키워드가 포함되며, 책임지기 역량의 핵심은 자기 제어(Self-control), 자기 효능감, 책임, 문제해결 및 적응력을 포함하는 자기 통제(self-regulation)를 미래 사회 대비 교육의 핵심 키워드로 제시하였다(OECD, 2018).

United Nations Educational, Scientific and Cultural Organization(UNESCO)에서는 미래 사회 대비 학생들의 역량 함양 키워드로 먼저, '알기 위한 학습 역량(Learning to know)'으로 이를 위해서는 체계적 사고, 분석적 질문 및 비판적 사고, 문제해결력, 미래지향적 사고, 변화수용능력, 학문간 접근법에 대한 이해 등이다. 다음은 사회·정서적 영역 측면에서 가치와 태도로 그리고 '존재하기 위한 학습 역량(Learning to be)'과 '더불어 살아가기 위한 학습 역량(Learning to live together)'이라 할 수 있다(UNESCO, 2015).

또한 OECD 프레임워크의 역량(OECD, 2018)에서는 ICT·디지털·데이터 리터러시, 협력, 비판적 사고, 문제해결력, 자기 관리(Self-regulation/ Self-control), 공감(Empathy), 존중(Respect), 끈기/회복력(Persistence/ Resilience), 학생 주도성(Student Agency), 책임감 가지기(Taking responsibility), 긴장과 딜레마 해소하기(Reconciling tensions and dilemmas), 예측(Anticipation), 컴퓨팅 사고력/프로그래밍/코딩(Computational thinking/ Programming/Coding), 그리고 마지막으로 금융 소양(Financial literacy), 기업가 정신(entrepreneurship/enterprising) 등을 키워드로 제시하였다.

국내 연구 중 류태호(2018)에서는 미래사회 대비 교육에서의 주요 역량을 복합 문제 해결능력, 협업능력, 비판적 사고 능력, 감성 능력, 인적 자원 관리 능력, 서비스 지향성, 판단 및 의사 결정 능력, 인지적 유연력, 협상능력, 창의력으로 설정하였으며 이영희 외(2018) 연구에서는 문제해결력, 의사소통능력, 협업능력, 비판적 사고력, 소프트웨어, 인성·감성교육, 다문화, 직업진로, 글로벌시민의식, 창의력 등을 키워드로 미래 교육의 방향성을 상정하였다.

2. 미래 사회 대비 수학교육에서의 키워드

이러한 교육 전반적인 분야에서 미래 사회 대비 주요 키워드들은 수학교육 분야에도 그 정신을 이어받아 영역 특수적인 측면을 고려하여 제시되고 있다. 예를 들어, 김광진, 한재성(2017)에 따르면, 4차 산업시대에서는 새로운 소통의 미디어 생태계가 생겨나면서 글로벌 연결로 다양성이 극대화되는 사회라는 것이다. 이러한 새로운 사회에서는 창의력과 문제의 조직적 해결능력을 갖춘 인재가 필요하고, 이러한 시대의 요구는 수학교육에서 과거 공급자 중심의 일방적 경험 전달보다는 교육 수요자인 학

생에 맞춰 진행되어야 한다. 또한 최근 기술을 기반으로 미래 사회 대비 교육에서의 주요 키워드는 클라우드와 인공지능 비판적 사고, 소통, 협동능력, 창의력. 수학교육에서 개인 맞춤형, 개별화 등을 주장하고 있다.

최은미(2017)는 4차 산업혁명 시대 수학 영역에서 관심을 가져야 할 부분이 STEAM이고 수학적 모델링을 통한 문제해결능력이 핵심이라 보았다. 또한 사회적인 여러 문제를 제대로 인식하기 위한 ‘비평적 사고’의 중요성 및 ‘창의성’을 강조하였다.

최지웅(2017)은 4차 산업혁명의 핵심 화두는 빅데이터, IoT 사물인터넷, 인공지능기술, 증강현실, 가상현실 등이며 이에 교육의 키워드를 ‘빅데이터, 사물인터넷, 인공지능, 증강·가상현실, 3D프린터’ 소프트웨어 교육과 코딩교육을 중심으로 미래사회 대비 교육의 방향성을 주장한 바 있다.

박진해(2017)는 과거와는 전혀 다른 제 4차 산업혁명의 시대에는 비판적이고 창의적인 사고력과 빠르게 변화하는 과학과 기술에 적응할 수 있는 능력이 필수적이라고 한다. 따라서 학생들에게 이미 알려진 사실을 잘 이해만 시키는 교육이 아닌 기본 원리를 잘 이해하여 새로운 아이디어를 창출할 줄 아는 창의적인 사고력 교육과 변화에 적응할 수 있도록 해야 하며 이에 따라 수학 교육의 목표를 다음과 같이 제시하였다. 먼저, 수와 식을 이용한 수학의 개념을 정확히 이해하여 사회와 자연현상을 수학적으로 기술, 분석하고 추론하여 의사소통하고 창의적이고 비판적인 사고력을 함양할 수 있어야 한다는 것이다. 또한 빠른 속도로 변해가는 시대에 적응하고 미래 사회의 구성원에게 필수적인 역량을 제공하여 국제화된 사회의 시민으로서 합리적인 의사결정능력과 민주적 소통능력을 함양한다.

마지막으로 나귀수 외(2018)에서는 4차 산업혁명 시대는 다양한 분야의 기술들의 융합, 즉, 인공지능이 적용된 자동화·지능화, 사람·사물·공간 등 모든 것이 연결되고 상호 작용하는 초연결(Hyper-connect) 사회를 말하며 따라서 주요 내용은 ‘지능정보사회’를 살아가기 위한 역량에 초점을 맞추었다.

위의 각 연구에서 제시한 수학교육의 주요 키워드를 정리하면 <표 1>과 같다.

교육부 (2015)		김광진·한재웅 (2017)	최은미 (2017)	최지웅 (2017)	박진해 (2017)	나귀수 외 (2018)	
문제해결			문제해결			문제해결	복합적 문제해결
문제 이해	문제해결 접근법					문제발견	수학적 아이디어
통제 및 반성	협력적 문제해결						
문제 만들기	수학적 모델링				수학적 모델링	모델링	
추론	관찰과 추측				분석, 추론	추론	
논리적 절차 수행	수학적 사실 분석				개념 이해	논리적 추론	수학적 지식
정당화	추론 과정의						

	반성						
창의융합		창의력	창의성	창의융합형 사고	창의성	창의융합	창의성
의사소통		소통			의사소통	의사소통	
수학적 표현	협력과 존중	협동능 (협업)				수학적 표현	협력
정보처리						정보처리	빅데이터
정보처리, 문제해결, 창의성, 발표능력	공학전공, 연구, 활용			코딩, 컴퓨팅 사고		코딩	프로그래밍
수학적 태도 및 실천						수학적 태도	
가치 인식	학습태도	생각방법				자율성	
자기 관리		자기주도 학습				자기주도성	능동적 학습
합리적 의사결정	시민의식				의사결정능력	사회 정서적 역량	
		비판적 사고	비평적 사고		비판적 사고	비판적 사고	
				인성감성, 디지털교과서		공감	정서 지능
						미래시대	인지적 역량
				활동, 탐구, 문제해결, 협업		타인의 존중	인지적 유연성
				통계교육		STEAM	MINT
		개별화		맞춤형 교육		시각화	공간

<표 1> 미래사회 대비 수학교육에서 제시하는 핵심 키워드

III. 텍스트 마이닝(Text Mining)

1. 텍스트 마이닝의 개념

텍스트 마이닝은 글 혹은 문서를 뜻하는 ‘Text’와 채굴, 채광을 뜻 하는 ‘Mining’이 결합된 단어로, 글 혹은 문서에서 의미 있는 정보를 추출하여 활용하기 위한 분석 도구이다. 즉, 인터넷상에 있는 복합문서(워드프로세서, 이메일, PDF 등)와 인터넷 페이지 등의 비정형 텍스트 형태로 구성되어 있는 정보들을 어떻게 활용하는가에 대한 방법을 텍스트 마이닝(Text Mining)이라고 한다(김대곤 외, 2013). 데이터마이닝과 텍스트 마이닝의 차이는 대상이 텍스트 데이터와 수치 데이터라는 점에서 구분되고 텍스트 마이닝은 데이터마이닝과 달리 이를 구조화시키는 과정이 필요하다. 과거 정보 시스템은 주로 수치 데이터를 저장하였으나 최근에는 수치 데이터보다는 텍스트 즉 문서형태의 데이터를 주로 저장한다. 수치 데이터인 경우 구조화 되어 있어 컴퓨터가 직접적인 처리가 가능하나, 텍스트 데이터의 경우에는 이를 구조화 시키는 과정이 필

요하다. 텍스트 마이닝은 텍스트 데이터를 대상으로 하여 그들 간의 암묵적인 정보를 추출하는 과정으로 정의할 수 있다(조태호, 2001).

텍스트 마이닝은 텍스트 데이터를 대상으로 유용한 정보를 추출해 내는 방법론으로서 자연어로 구성되어 있는 비정형 데이터에서 일정한 패턴과 관계를 추출하여 유의미한 정보를 찾아내는 기법이다. 자연어 처리(Natural Language Processing: NLP)에 기반을 둔 기술로써 사람들이 말하는 언어를 컴퓨터 혹은 프로그램이 이해할 수 있도록 변환시켜준다(임동훈, 2015). 또한 비정형 텍스트의 연계성을 파악함으로써 분류 혹은 군집화하거나 요약하는 등 빅데이터 속에 숨겨진 의미 있는 정보를 발견하기 위해 주로 사용된다(송태민·송주영, 2016).

텍스트 마이닝은 텍스트 데이터로부터 의미 있는 정보 등을 찾아낼 수 있다는 ‘통계적 의미론적 가설(statistic semantic hypothesis)’에 근거하고 있다. 이 가설은 “사람들의 말이나 글 등에서 나타나는 여러 단어 사용의 통계적인 규칙으로부터 그 사람들의 의도를 찾아낼 수 있다”는 전제를 바탕으로 한다(Turney & Pantel, 2010). 또한 텍스트 마이닝은 대용량의 데이터로부터 사용자가 관심을 가지고 있는 정보를 단어 수준이 아닌 맥락(context) 수준의 의미까지 찾아내 주는 프로세스를 의미한다. 오늘날 정보 및 지식의 폭발적인 증가로 상당한 부분을 자동적으로 처리할 수 있도록 하는 방법을 필요로 하고 있으며 방대한 데이터 속에서 숨겨진 패턴양식을 발견하여 특정 주제와 관련한 데이터를 검색하는 방법으로 발전하고 있다. 텍스트 마이닝의 일반적인 프로세스는 비정형 정보수준 → 정보처리 → 정보추출 → 정보분석 등으로 진행되며 객관적인 수학적 모델이나 알고리즘을 통해 유용한 정보를 도출하는 방법이다(정근하, 2010).

텍스트 마이닝은 많은 것을 제공하지만, 잘못된 해석을 하지 않기 위해 주의할 사항이 있다. Grimmer & Stewart (2013)는 텍스트 마이닝의 사용에서 주의할 4가지 사항을 제시했다. 첫째로 텍스트 데이터가 어떻게 생성되었는지에 대해 정확한 원인을 알 수 없으므로 텍스트 분석의 정량적 모델을 구축할 때 해석에 신중을 기해야 한다. 둘째, 텍스트에 대한 연구자의 해석이 필요하며 수치 계산은 연구자의 판단을 돕는 것으로만 활용되어야 한다. 셋째, 텍스트 분석의 절대적인 방법은 아직 발견되지 않았으므로 텍스트 마이닝은 항상 분석 목적과 일치해야 한다. 넷째, 유효성 확인이 필요하며 이를 수행하는 데는 여러 가지 방법이 있다. 이러한 주의사항을 종합해보면 텍스트 마이닝은 연구자의 판단을 위한 최소한의 분석요인을 제공하는 것으로서 연구자가 설정한 연구목적과 부합될 때 잘 활용되었다고 할 수 있다.

2. 텍스트 마이닝을 활용한 분석 방법

텍스트 마이닝의 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. 텍스

트 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).

텍스트 마이닝의 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).

텍스트 마이닝 기법의 특징은 다음과 같다. 첫째, 인지능력을 사용하지 않아 객관적인 분석이 가능해지고, 둘째, 분석결과를 시각화함으로써 연구의 관점을 쉽고 빠르게 전달할 수 있으며, 셋째, 도출된 결과를 연구자의 관점에 따라 다각도로 조명해볼 수 있다는 것이다(이인선·나은영, 2018). 이처럼 텍스트 마이닝 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. 연구 분석 과정

1. 분석 대상 선정

학술지에 게재된 논문들의 데이터를 수집하기 위한 학술연구 제공서비스로는 국회도서관의 전자문서 검색서비스, 한국교육학술정보원의 RISS, (주)한국학술정보의 KISS, 누리미디어의 디비피아(DBpia)가 대표적이다. 이 중 국내에서 가장 많은 논문을 보유하고 있는 학술연구 제공 서비스는 디비피아(DBpia)로 2019년 6월 7일 기준으로 2,851,248건의 논문을 보유하고 있다. 학술지를 선정하는데 참고했던 한국학술지 인용색인(KCI)사이트와 더불어 가장 많은 논문을 보유하고 있는 디비피아(DBpia)사이트를 통해서 각 학술지에 게재된 논문들을 수집하였다. 논문의 제목 및 초록은 학술연구 제공 사이트에서 간단히 텍스트로 정리되어 받을 수 있고, 일반적으로 여러 동향 분석 연구에서도 논문 초록을 이용하여 분석대상으로 삼지만(김하진·송민, 2014; 송혜지 외, 2013; 신규식 외, 2015; 배규용 외, 2013) 본 연구의 경우 텍스트의 양이 많을수록 유의미한 결과를 기대할 수 있기 때문에 논문의 제목 및 초록만으로는 텍스트의 양이 부족하다고 판단하였다. 따라서 수집한 논문의 PDF 및 한글(*.hwp)문서에서 서론과 결론의 텍스트를 추출하여 데이터화 하였고 엑셀(*.csv)문서로 정리하였

다. 수집한 논문들 중 텍스트가 추출되지 않는 논문들이 존재하였는데 이러한 경우에는 해당 논문의 국문 초록을 이용하여 분석하였다. 이때 주의해야 할 것은 텍스트 데이터의 양이 논문마다 모두 제각각이기 때문에 형태소 추출 시 빈도수에 영향을 줄 수 있는 점에 유의하여 동일한 수준으로 통제하여야 한다(채미현, 2018).

본 연구에서는 이와 같은 과정을 거쳐 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.

전체 분석 기간(2016년-2018년)동안 학술지를 통해 발표된 수학교육 연구 논문은 <수학교육>에서 73편, <수학교육 논문집>에서 88편, <수학교육학연구>에서 112편, <학교수학>에서 122편 <학교수학회 논문집> 75편으로 총 470편이 게재되었다. 이 중 영문으로 작성된 10편의 논문을 제외한 460편의 논문을 분석 대상으로 선정하였다.

2. 빈도 분석 및 워드 클라우드 생성

텍스트 마이닝은 일반적인 텍스트 데이터를 전처리 작업과 형태소 분석을 통해 단어를 정제하고 텍스트 속에서 유용한 단어를 추출하고 그 빈도수를 도출한다. 여러 문서에서 동시에 사용되는 단어의 빈도수를 통해 문장의 의미를 전달하는 해당 단어의 중요도를 파악할 수 있다(서일원·전채남, 2013). 본 연구에서는 연도별 학술지에 게재된 논문들이 어떠한 부분에 주목했는지 일차적인 탐색을 위해 빈도수 분석을 활용하였다. ‘빈도수 분석’은 특정 문서들 내에서 자주 사용되는 단어를 추출하여 언급되는 빈도수에 따라 중요도를 분석하는 방법(황명화, 2014)으로 발생빈도를 파악하여 문서의 주제 또는 문서에 대한 태도나 감성을 추론 할 수 있다.

빈도수를 분석하기 위해서는 대부분의 경우 문장을 개별단어로 분리하는 토큰화 과정(tokenization)이 필요한데 한국어의 경우 문장을 개별단어로 분리하기 위한 방법으로 형태소 분석을 실시한다(김성근 외, 2016). 본 연구에서도 데이터 클리닝 과정을 거쳐 추출된 단어들을 이용하여 빈도수 분석을 실시하였고 가장 높은 빈도수를 차지하는 단어부터 내림차순으로 정렬한 후, 상위 빈도수 20개의 단어를 추출하였다. 이렇게 추출된 단어들이 갖는 의미를 분석하였고, 이를 통해 최근 연구들의 동향을 추론해 보았다. 또한 연도별 상위 빈도수 단어들을 비교하여 공통적으로 중요하게 나타나는 단어와 각 연도별로 특별한 의미를 갖는 단어들을 분석하였다. 단어들에 대한 분석 후에는 word cloud 패키지(Ian, 2014)를 이용하여 빈도수 분석 결과를 시각적으로 나타내었다. 워드 클라우드(word cloud)란 언급된 빈도수에 비례하여 단어의 크기와 색을 다르게 하여 시각적으로 표현한 것으로 이를 통해 어떠한 단어가 빈번하게 언급되었는지 효과적으로 확인할 수 있다.

3. TF-IDF(Term Frequency-Inverse Document Frequency) 분석

앞서 분석하였던 빈도수 분석은 특정 단어가 문서들에서 얼마나 등장하는 지를 나타내는 ‘단어 빈도(Term Frequency: TF)’로 쉽게 결정할 수 있다. 여기서 TF 값이 큰 단어일수록 문서에서 중요한 단어라고 판단할 수 있으나, 만약 모든 문서에서 자주 등장한다면 보편적인 단어라는 것을 의미할 수도 있다(육동인, 2017). 이는 ‘문서 빈도(Document Frequency: DF)’의 값이 큰 단어라고 할 수 있는데, DF 값은 특정 단어가 문서 집합 내에서 얼마나 많은 문서에서 나타나는가를 의미한다. 예를 들어 ‘수학’이라는 단어는 수학교육 연구 논문에서는 빠짐없이 등장하여 높은 TF 값을 가질 것이다. 하지만 모든 문서에서 자주 나타나기 때문에 DF 값 또한 커질 것이고 ‘수학’이라는 단어의 중요도는 낮아진다. 이런 경우와 같이 모든 문서에서 자주 나타나는 보편적인 단어를 필터링 하기 위해 ‘단어 빈도-역문서 빈도(Term Frequency-Inverse Document Frequency: TF-IDF)’가 도입 되었는데, 이는 단순한 빈도처리가 아닌 단어의 빈도수 확률을 기준으로 출현 빈도를 재가공한 것이다(임화진, 2014). ‘역문서 빈도(Inverse Document Frequency: IDF)’는 특정 단어가 문서들 사이에서 얼마나 공통적으로 나타나는지를 나타내는 값으로, 전체 문서의 빈도수를 특정 단어를 포함하고 있는 문서의 빈도수로 나눈 후 로그를 취하여 계산한다. TF-IDF 값은 문서에서의 단어 빈도수를 나타내는 TF와 특정단어가 출현하고 있는 문서의 빈도수 DF의 역수인 IDF를 곱한 값이다. 즉, 특정 문서 내에서 단어 빈도가 높을수록, 전체문서에서 그 단어를 포함한 문서가 적을수록 TF-IDF 값이 커지며 구체적인 계산식은 다음과 같다(육동인, 2017).

$$TF-IDF = TF \times \log\left(\frac{N}{DF}\right)$$

TF : 문서 내 특정단어 빈도수

$$IDF : \log\left(\frac{N}{DF}\right)$$

DF : 특정 단어가 포함된 문서의 개수

N : 전체 문서의 개수

본 연구에서는 위의 식을 활용하여 분석을 하였으며 TF-IDF 분석 결과 높은 값을 나타내는 상위 20개의 단어를 추출하였고, 이를 통해 핵심적인 의미를 가지는 단어를 분석하였다. 또한 연도별 상위 빈도수 단어들을 비교하여 공통적으로 중요하게 나타나는 단어와 각 연도별로 특별한 의미를 갖는 단어들을 분석하였고 앞서 실시하였던 상위 빈도(TF) 단어와의 비교를 통해 최근 연구 논문의 핵심 키워드를 추출하였다.

V. 연구 분석 결과

1. 빈도수 분석 및 워드 클라우드 분석 결과

전체 분석 기간인 2016년부터 2018년 동안의 수학교육 연구 논문에서 데이터 클리닝 과정을 거친 후 나타난 단어는 총 16,149 단어였다. 이들 중 높은 빈도로 출현한 상위 단어 20개를 빈도수(TF) 기준으로 정리하면 <표 2>와 같다.

순위	주제어	빈도수(TF)	순위	주제어	빈도수(TF)
1	수학	7736	11	과정	1436
2	학생	4418	12	교과서	1331
3	연구	3849	13	교육과정	1282
4	교사	2847	14	해결	1279
5	학습	2747	15	활용	1265
6	문제	2376	16	내용	1223
7	수업	2312	17	개념	1182
8	분석	1672	18	이해	1171
9	교육	1565	19	개발	950
10	결과	1506	20	지식	948

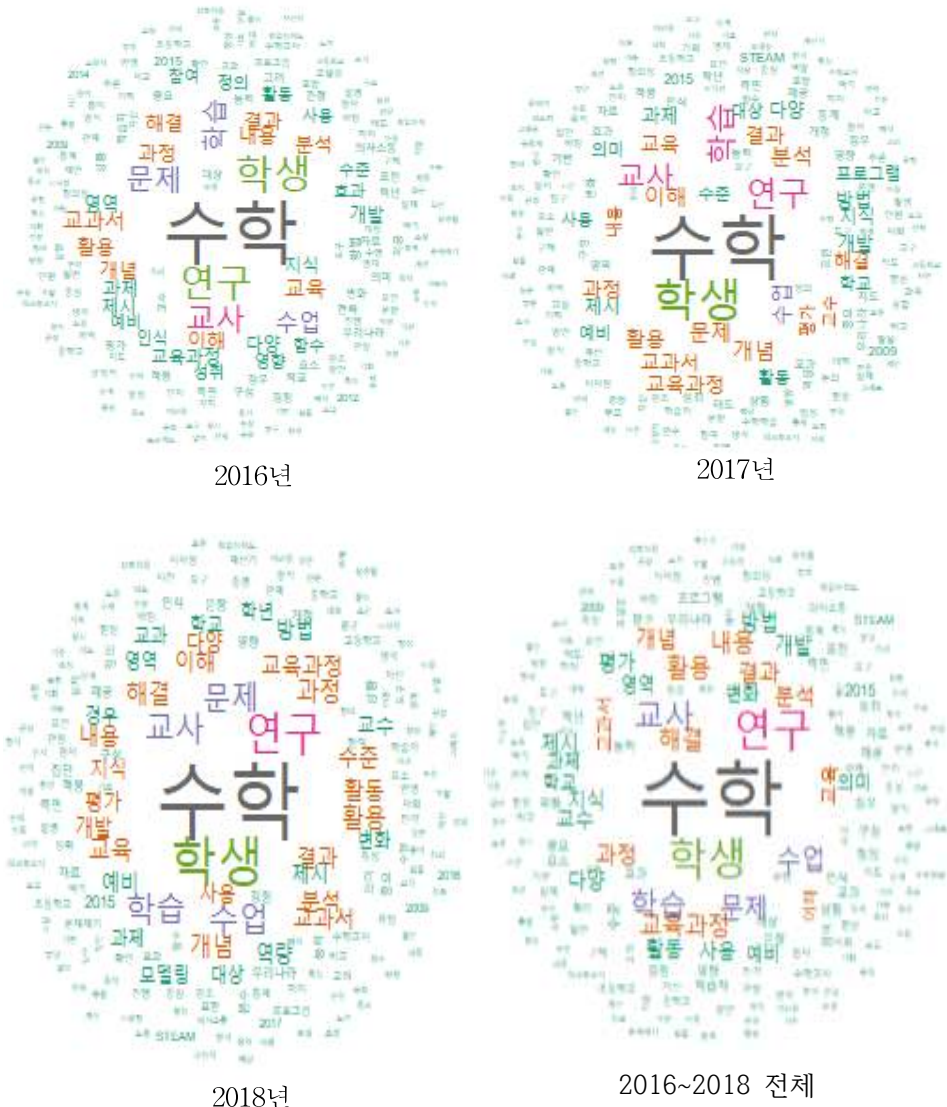
<표 2> 전체 분석 기간 학술지 논문에서 추출된 상위 빈도수 단어(TF 기준) 가장 많이 나타난 단어는 ‘수학’으로 총 7746번 나타났고, ‘학생’(4184번), ‘연구’(3829번), ‘교사’(2847번), ‘학습’(2747번), ‘문제’(2376번), ‘수업’(2312번), ‘분석’(1672번), ‘교육’(1565번), ‘결과’(1506번)가 많은 빈도수를 가지는 단어로 나타났다. ‘수학’, ‘분석’, ‘결과’의 단어들은 수학교육 연구 논문에 많이 사용되는 보편적인 단어로서 연구 주제에 대해 직접적인 의미를 갖고 있다고 해석하기 힘들다. 수학교육 연구에 있어서 ‘학생’과 ‘학습’, ‘교사’와 ‘수업’이라는 단어를 분석해 보면 교사중심으로만 생각하는 것이 아니라 학생의 수준과 입장을 고려한 연구가 많이 진행되고 있다는 사실을 알 수 있다. 수학교과의 성격상 수학적 개념을 이해하고 적용하기 위해 다양한 문제를 다루는 만큼 ‘문제’에 대한 연구 또한 많이 진행되고 있다는 사실을 알 수 있다.

전체 분석 기간 학술지 논문의 빈도수 분석을 한 후, 이를 시각적으로 표현하기 위해 워드 클라우드를 구현한 결과는 [그림 1]과 같다. 앞서 빈도수 분석을 했을 때 나타난 상위 빈도수 단어들에 워드 클라우드 중앙으로 모여 있는 것을 한눈에 확인할 수 있고, 주변으로 상위 빈도수 20개의 단어들까지 표현하였다.

연도별 분석 과정을 살펴보면 다음과 같다. 먼저, 2016년 학술지를 통해 발표된 수학교육 연구 논문은 <수학교육>에서 26편, <수학교육 논문집>에서 28편, <수학교육학연구>에서 41편, <학교수학>에서 44편 <학교수학회 논문집> 25편으로 총 164편이 게재되었다. 이 중 영문으로 작성된 3편의 논문을 제외한 161편의 논문을 분석하였다.

2016년 수학교육 연구 논문에서 데이터 클리닝 과정을 거친 후 나타난 단어는 총 7532 단어였다. 이들 중 높은 빈도로 출현한 상위 단어 20개를 빈도수(TF) 기준으로

정리하면 <표 3>과 같다. 가장 많이 나타난 단어는 ‘수학’으로 총 2631번 나타났고, ‘연구’(1421번), ‘학생’(1367번), ‘교사’(1027번), ‘학습’(898번), ‘문제’(846번), ‘수업’(802번), ‘교육’(599번), ‘분석’(564번), ‘결과’(512번)가 많은 빈도수를 가지는 단어로 나타났다.



[그림 1] 2016~2018년 빈도수 기준 워드 클라우드

2016년 학술지 논문의 빈도수 분석을 한 후, 이를 시각적으로 표현하기 위해 워드

클라우드를 구현한 결과는 [그림 1]과 같다. 앞서 빈도수 분석을 했을 때 나타난 상위 빈도수 단어들이 워드 클라우드 중앙으로 모여 있는 것을 한눈에 확인할 수 있고, 주변으로 상위 빈도수 200개의 단어들까지 표현하였다. 2017년 학술지를 통해 발표된 수학교육 연구 논문은 <수학교육>에서 23편, <수학교육 논문집>에서 30편, <수학교육학연구>에서 42편, <학교수학>에서 40편 <학교수학회 논문집> 27편으로 총 162편이 게재되었다. 이 중 영문으로 작성된 3편의 논문을 제외한 159편의 논문을 분석하였다.

2017년 수학교육 연구 논문에서 데이터 클리닝 과정을 거친 후 나타난 단어는 총 7305 단어였다. 이들 중 높은 빈도로 출현한 상위 단어 20개를 빈도수(TF) 기준으로 정리하면 <표 3>과 같다. 가장 많이 나타난 단어는 ‘수학’으로 총 2392번 나타났고, ‘학생’(1440번), ‘연구’(1123번), ‘학습’(946번), ‘교사’(899번), ‘수업’(617번), ‘문제’(559번), ‘분석’(539번), ‘교과서’(499번), ‘교육’(497번)이 많은 빈도수를 가지는 단어로 나타났다.

2017년 학술지 논문의 빈도수 분석을 한 후, 이를 시각적으로 표현하기 위해 워드 클라우드를 구현한 결과는 [그림 1]과 같다. 앞서 빈도수 분석을 했을 때 나타난 상위 빈도수 단어들이 워드 클라우드 중앙으로 모여 있는 것을 한눈에 확인할 수 있고, 주변으로 상위 빈도수 200개의 단어들까지 표현하였다.

2018년 학술지를 통해 발표된 수학교육 연구 논문은 <수학교육>에서 24편, <수학교육 논문집>에서 30편, <수학교육학연구>에서 29편, <학교수학>에서 38편 <학교수학회 논문집> 23편으로 총 144편이 게재되었다. 이 중 영문으로 작성된 2편의 논문을 제외한 142편의 논문을 분석하였다.

2018년 수학교육 연구 논문에서 데이터 클리닝 과정을 거친 후 나타난 단어는 총 7283 단어였다. 이들 중 높은 빈도로 출현한 상위 단어 20개를 빈도수(TF) 기준으로 정리하면 <표 3>과 같다. 가장 많이 나타난 단어는 ‘수학’으로 총 2713번 나타났고, ‘학생’(1611번), ‘연구’(1305번), ‘문제’(971번), ‘교사’(921번), ‘학습’(903번), ‘수업’(893번), ‘분석’(569번), ‘결과’(529번), ‘교육과정’(507번)이 많은 빈도수를 가지는 단어로 나타났다.

2018년 학술지 논문의 빈도수 분석을 한 후, 이를 시각적으로 표현하기 위해 워드 클라우드를 구현한 결과는 [그림 1]과 같다. 앞서 빈도수 분석을 했을 때 나타난 상위 빈도수 단어들이 워드 클라우드 중앙으로 모여 있는 것을 한눈에 확인할 수 있고, 주변으로 상위 빈도수 200개의 단어들까지 표현하였다.

연도별 빈도수 분석 결과를 서로 비교해보면, 먼저 7개의 상위 단어는 ‘수학’, ‘학생’, ‘연구’, ‘교사’, ‘학습’, ‘문제’, ‘수업’으로 순위의 차이는 있지만 모든 연도에서 동일하게 나타났다. 따라서 이 7개의 단어를 최근 3개년 수학교육 연구의 키워드라고 할 수 있다. 다음으로 나타나는 ‘분석’, ‘교육’, ‘결과’, ‘과정’, ‘교과서’, ‘해결’, ‘활용’, ‘내용’, ‘개념’, ‘이해’라는 단어들은 모든 연도에서 공통적으로 나타났다. 전체적으로 봤을 때 각 연도에 특별히 두드러지는 단어는 나타나지 않고 다소 고르게 나타났음을 확인할 수 있다. 이는 최근 3개년 동안 이루어진 연구들의 키워드가 전반적으로 비슷하게 나

타났다고 볼 수 있고 연구의 트렌드가 일정하다고 판단된다.

순위	2016	2017	2018	전체
1	수학	수학	수학	수학
2	연구	학생	학생	학생
3	학생	연구	연구	연구
4	교사	학습	문제	교사
5	학습	교사	교사	학습
6	문제	수업	학습	문제
7	수업	문제	수업	수업
8	교육	분석	분석	분석
9	분석	교과서	결과	교육
10	결과	교육	교육과정	결과
11	과정	교육과정	해결	과정
12	해결	결과	과정	교과서
13	교과서	활용	교육	교육과정
14	내용	과정	활용	해결
15	개념	내용	개념	활용
16	활용	이해	이해	내용
17	이해	평가	활동	개념
18	영역	교수	내용	이해
19	방법	개념	교과서	개발
20	과제	해결	수준	지식

<표 3> 연도별 학술지 논문에서 추출된 상위 빈도 단어(TF 기준)

연도별로 다르게 나타난 단어들을 살펴보면 2016년에는 ‘영역’, ‘방법’, ‘과제’가 나타났고 2017년에는 ‘교육과정’, ‘평가’, ‘교수’, 2018년에는 ‘교육과정’, ‘개발’, ‘지식’이라는 단어가 나타났다. 특히 2016년 이후에 ‘교육과정’ 단어의 빈도수가 많아진 것으로 보아 2015 개정 교육과정이 학술지 논문 주제에 영향을 크게 미친 것으로 추론할 수 있다. 또한 2017년에 높게 나타난 ‘평가’, ‘교수’ 단어의 빈도수는 2018년도에는 현저하게 줄었으며 오히려 ‘활동’, ‘수준’ 단어의 빈도수가 향상 되었다. 이는 최근 학생들의 활동과 수준에 대한 관심도가 높아졌음을 알 수 있다. 연도별 빈도수 분석 결과에는 나타나지 않았지만 전체적으로 빈도수를 분석했을 때 ‘개발’, ‘지식’이 나타나는 것으로 보아 이 단어들과 관련된 연구들이 많이 수행되었음을 알 수 있다(예를 들면, 교수·학습 자료 개발, 예비교사 지식 등).

2. TF-IDF 분석 결과

전체 분석 기간(2016년-2018년)동안 학술지를 통해 발표된 수학교육 연구 논문에서 TF-IDF 기준으로 값이 큰 상위 빈도 단어 20개를 추출하였다(표 4).

최근 3개년 연구에서 가장 TF-IDF 값이 큰 단어는 ‘평가’로 나타났고, 다음으로는

‘초등학교’, ‘예비’, ‘함수’, ‘통계’순으로 나타났다. TF-IDF 분석 결과 ‘평가’라는 단어가 가장 중요한 단어로 나타났는데 이는 평가 방법에 관한 연구가 가장 핵심적으로 나타난다고 볼 수 있다. 학교급으로는 초등학교, 중학교 순으로 중요하게 고려되었으며, 교과서 단원으로는 함수와 통계 단원이 중요하게 고려된다는 것을 알 수 있다. ‘예비’라는 단어는 예비교사, 예비 학생 등 의미 있게 쓰이기 때문에 높은 값을 갖는 것으로 추론된다. 그 뒤로는 ‘표현’, ‘요소’, ‘문항’, ‘중학교’, ‘비교’순으로 나타났다. 이 단어들을 앞서 분석하였던 전체 분석 기간 상위 빈도수(TF) 단어와 비교하여 보면 ‘해결’이라는 1개의 단어만 공통으로 나타났고 나머지 단어는 모두 새롭게 나타났다.

연도별로 분석한 결과를 살펴보면 다음과 같다. 2016년에 학술지를 통해 발표된 수학교육 연구 논문에서 TF-IDF 기준으로 값이 큰 상위 빈도 단어 20개를 추출하면 값이 가장 큰 단어는 ‘교과서’로 나타났고, 그 다음으로는 ‘과제’, ‘예비’, ‘활동’, ‘분석’순으로 나타났다. 이를 통해 2016년 연구에서 가장 중요하게 고려된 단어는 ‘교과서’인 것을 확인할 수 있었는데, 2015 개정 교육과정으로 인해 바뀌는 교과서에 관한 연구와 더불어 ‘과제’라는 단어 역시 중요하게 나타났다. 또한 2015 개정 교육과정에서 다양한 활동을 통한 6가지 핵심역량 함양을 강조하고 있기 때문에 2016년 연구에서 ‘활동’이라는 단어가 중요하게 고려되었다고 판단할 수 있다. 그 뒤로는 ‘평가’, ‘지식’, ‘우리나라’, ‘영역’, ‘학년’순으로 나타났다. 이 단어들을 앞서 분석하였던 2016년 상위 빈도수(TF) 단어와 비교하여 보면 ‘교과서’, ‘과제’, ‘분석’, ‘영역’, ‘이해’ 5개의 단어만 공통으로 나타났고 나머지 단어는 새롭게 나타났다.

다음으로 2017년에 학술지를 통해 발표된 수학교육 연구 논문에서 TF-IDF 기준으로 값이 큰 상위 빈도 단어 20개를 추출하였고 그 중 가장 비중이 큰 단어는 2016년과 같이 ‘교과서’로 나타났고, 그 다음으로는 ‘교육과정’, ‘평가’, ‘개념’, ‘과제’ 순으로 나타났다. 2017년 연구에서도 2016년과 같이 가장 중요하게 고려된 단어는 ‘교과서’인 것을 확인할 수 있었고, 두 번째로 중요하게 고려된 단어는 ‘교육과정’이었다. 이 단어는 2016년에는 나타나지 않았지만 2017년에는 중요하게 고려된 단어인 것으로 나타났는데 이는 앞서 실시한 빈도수 분석에서도 마찬가지로 2016년에는 포함되지 않고 2017년부터 포함되었던 단어로 의미가 있다고 판단된다. 다음으로 나타난 ‘평가’의 경우 2016년에도 상위 20개의 단어에 포함되었는데 ‘개념’, ‘과제’ 단어의 경우 2016년에는 나타나지 않고 2017년에 새롭게 포함되었다. 그 뒤로는 ‘통계’, ‘지식’, ‘예비’, ‘활동’, ‘지도’순으로 나타났다. 이 단어들을 앞서 분석하였던 2017년 상위 빈도수(TF) 단어와 비교하여 보면 ‘교과서’, ‘교육과정’, ‘평가’, ‘개념’, ‘활용’ 5개의 단어만 공통으로 나타났고 나머지 단어는 새롭게 나타났다.

2018년에 학술지를 통해 발표된 수학교육 연구 논문에서 TF-IDF 기준으로 값이 큰 상위 빈도 단어 20개를 추출하였고 그 중 TF-IDF 값이 가장 큰 단어는 2016년, 2017년도와 같이 ‘교과서’로 나타났고, 그 다음으로는 ‘개념’, ‘평가’, ‘교육과정’, ‘예비’순으로 나타났다. 2018년 연구에서도 전년도와 같이 가장 중요하게 고려된 단어는 ‘교과서’인 것을 확인할 수 있었고, 두 번째로 중요하게 고려된 단어는 ‘개념’이었다.

이 단어는 2016년에는 나타나지 않았지만 2017년에도 4번째로 중요하게 고려된 단어인 것으로 나타났다. 다음으로 나타난 ‘평가’, ‘교육과정’, ‘예비’의 경우 전년도에도 상위 20개의 단어에 포함되어 중요한 단어라는 것을 알 수 있었다. 그 뒤로는 ‘모델링’, ‘역량’, ‘활동’, ‘학년’, ‘과제’순으로 나타났다. 이 단어들을 앞서 분석하였던 2017년 상위 빈도수(TF) 단어와 비교하여 보면 ‘교과서’, ‘개념’, ‘교육과정’, ‘활동’, ‘수준’ 5개의 단어만 공통으로 나타났고 나머지 단어는 새롭게 나타났다.

지금까지 연도별로 분석한 결과를 비교하여 보면, 빈도수 분석의 경우 단어가 나타난 단순 빈도수(TF)만을 가지고 분석했기 때문에 상위의 단어들이 연도별로 공통적으로 나타났다. 하지만 TF-IDF 분석의 경우 많은 문서에 공통적으로 나타나는 단어들의 점수가 낮아지기 때문에 연도별로 다양한 단어들이 나타났다. 공통적으로 나타나는 단어는 ‘교과서’, ‘과제’, ‘예비’, ‘활동’, ‘평가’, ‘지식’, ‘학년’, ‘수준’으로 이 단어들이 최근 3개년 동안 연구에서 중요하게 고려된 단어들이라고 할 수 있다. 연도별로 다르게 나타난 단어들을 살펴보면 ‘교육과정’의 경우 빈도수(TF) 분석과 마찬가지로 2016년에는 나타나지 않았지만 2017년과 2018년에는 공통으로 나타났다. 또한 2015 개정 교육과정에서 강조하는 ‘역량’이라는 단어가 2018년 연구에서 상위 단어로 나타났고 전체 분석기간에서는 핵심역량 중 하나인 ‘문제해결’이라는 단어가 나타났다. 이를 통해 2015 개정 교육과정이 중요하게 고려되었다는 것을 추론할 수 있었다. 마지막으로 ‘모델링’이라는 단어는 2016년에는 나타나지 않았지만 2017년에는 19번째, 2018년에는 6번째 상위 빈도(TF-IDF) 단어로 나타나면서 점차 관심도가 높아지고 있는 단어라고 추정된다.

순위	2016	2017	2018	전체
1	교과서	교과서	교과서	평가
2	과제	교육과정	개념	예비
3	예비	평가	교육과정	초등학교
4	활동	개념	평가	함수
5	분석	과제	예비	통계
6	평가	통계	모델링	표현
7	지식	지식	역량	요소
8	우리나라	예비	학년	문항
9	영역	활동	활동	중학교
10	학년	지도	과제	해결
11	수준	방법	지식	프로그램
12	함수	그래프	수준	문제해결
13	정의	프로그램	문항	비교
14	학습자	자료	교과	단원
15	전략	활용	정의	학습자
16	능력	학년	통계	요인
17	사고	문항	문제해결	관계

18	상황	수준	집단	맥락
19	성취	모델링	학교	이론
20	이해	개정	개정	구성

<표 4> 연도별 학술지 논문에서 추출된 상위 빈도 단어(TF-IDF 기준)

V. 결론 및 시사점

교육부(2015)에서는 미래인재 양성을 4차 산업혁명시대와 연결 짓고 빅데이터 분석, 인공지능(AI), 가상(VR) 증강현실(AR) 등 지능정보기술을 활용하여 개인의 흥미, 수준에 따른 맞춤형 교육 서비스 제공, 컴퓨팅 사고, 창의·융합형 사고, 인성, 감성 등 미래사회에서 요구되는 역량 교육 강화를 위한 교육으로 그 방향을 설정하고 있다. 수학교육 선행연구들이 최근의 경향을 얼마나 반영하고 있는지 지속적으로 점검하면서 이를 반영해 미래에 필요한 내용들을 도출하는 것은 연구의 발전적인 측면뿐만 아니라 더 나은 수학 교육을 위해서도 필요한 일이라 할 수 있다(김래영 외, 2012). 본 연구는 텍스트 마이닝 기법을 활용하여 수학교육연구에서 의미 있게 등장하는 키워드들을 통해 중요하게 다뤄지고 있는 연구 주제는 무엇인지를 살펴보고자 하였다.

이를 위하여 국내의 수학교육학 대표 학술지인 <수학교육>, <수학교육 논문집>, <수학교육학연구>, <학교수학>, <한국학교수학회 논문집>에 게재된 논문들의 최근 동향을 살펴보기 위해 최근 3년간(2016년~2018년)의 논문, 총 460편을 텍스트 마이닝 기법을 이용하여 분석하였다.

본 연구의 분석 과정은 3단계의 과정을 거쳐 진행되었다. 첫 번째 단계는 데이터 수집 및 클리닝의 단계로, 연구대상을 선정한 후 논문 데이터를 수집하여 서론과 결론 부분의 텍스트를 추출, 전처리 과정을 거쳐 형태소 분석을 실시하였다. 두 번째 단계에서는 키워드 분석을 하였다. 추출한 명사를 토대로 빈도수 분석을 실시하고 워드 클라우드를 구현하였다. 또한 빈도수(TF)만으로는 핵심 키워드를 판별하기 힘들기 때문에 TF-IDF 분석도 실시하였다.

본 연구의 결과를 통해 얻은 결론을 정리하면 다음과 같다.

첫째, 빈도수(TF) 분석을 통해 최근 3개년 수학교육 학술지에 게재된 논문에서 가장 많이 나타나는 단어들을 추출할 수 있었다. 상위 빈도수 단어는 ‘수학’, ‘학생’, ‘연구’, ‘교사’, ‘학습’, ‘문제’, ‘수업’ 순으로 나타났고 그 외 주요단어로는 ‘교과서’와 ‘교육과정’이 있었다. 이 연구 결과에서 얻을 수 있는 시사점은 최근 수학교육 연구에서 학생과 교사에 관한 연구가 많이 이루어지고 있다는 것이다. 교사와 학생은 꾸준히 관심도가 높았던 주제로 학술지 논문의 전반적인 핵심 키워드로 나타났다. 학생의 수준과 입장을 고려한 연구가 많이 진행되고 있고, 교사의 수업 설계 및 교과서 분석에 관한 연구도 많이 이루어지고 있는 것으로 추정된다. 2015 개정 교육과정에 관한 연구 역시 점차 관심도가 높아지는 것으로 나타났다. 2015 개정 교육과정이 현재 도입

되어 시행되고 있으므로 앞으로도 점점 더 많은 연구가 진행될 것으로 예상된다. 하지만 2015 개정 교육과정에서 제시하는 수학과 핵심역량에 관한 연구는 아직 미비한 것으로 나타났다. 따라서 수학과 핵심역량, 인재상, 정의적 영역 등 아직까지 많이 나타나지 않은 주제와 더불어 참신한 아이디어를 바탕으로 보다 다양한 수학교육 연구가 이루어질 필요가 있다.

둘째, TF-IDF 분석을 통해 최근 3개년 동안의 연구에서 중요하게 고려된 단어를 찾을 수 있었다. ‘교과서’, ‘과제’, ‘예비’, ‘활동’, ‘평가’, ‘지식’, ‘학년’, ‘수준’이 중요하게 고려되는 단어였고 그 외에 중요하게 고려되는 단어로는 ‘교육과정’과 ‘역량’이라는 단어가 나타났다. 이 연구 결과에서 얻을 수 있는 시사점은 다음과 같다. ‘교과서’라는 단어는 빈도수(TF) 분석에서도 핵심 단어로 나타났는데 여러 문서에서 동시에 나타나진 않았지만 교과서를 중요하게 다루었던 것으로 분석된다. 이는 교과서 비교 분석 및 개발에 관한 연구가 핵심적으로 이루어지고 있다고 추정된다. 또한 ‘예비’라는 단어가 중요하게 고려되고 있음을 알 수 있었고, 예비 교사 및 예비 학생들에 관한 연구에서 중점적으로 나타났다고 추정된다. 특히 빈도수(TF) 분석과 마찬가지로 ‘교육과정’이 핵심 단어로 추출되었고 추가로 ‘역량’이라는 단어가 나타났다. ‘역량’이라는 단어는 빈도수(TF) 분석에는 나타나지 않았는데 적은 문서에서 핵심적으로 나타났다고 추론할 수 있다. 마지막으로 빈도수(TF) 분석과 비교하여 봤을 때 공통적으로 나타나는 단어는 ‘교과서’, ‘교육과정’, ‘과제’, ‘활동’, ‘평가’, ‘지식’이었고 이를 통해 최근 3개년 동안의 연구 논문들의 핵심 키워드를 확인할 수 있었다.

이렇게 나타나는 키워드들을 미래 사회를 대비하고자하는 연구들에서 나타나는 핵심 키워드들과 연관 지어 본다면 다음과 같다.

먼저, TF-IDF를 통해 추출된 키워드 중, <표 1>에서 제시한 핵심 키워드와 일치하는 것은 ‘문제해결’, ‘모델링’, ‘활동’, ‘분석’, ‘지식’, ‘해결’, ‘통계’, ‘상황’, ‘전략’ 등이다. 즉, 이러한 연구들은 최근의 수학교육 연구와 미래사회 대비 핵심 내용들과 일치한다고 추측할 수 있다.

반면, TF-IDF를 통해 추출된 키워드 중에는 ‘협업 능력’, ‘자기관리’, ‘비판적 사고’, ‘코딩’, ‘감성 능력’, ‘글로벌 역량’, ‘창의력’ 등은 나타나지 않았음을 주목할 필요가 있다. 특히 협업능력과 같은 것은 최근 연구들에서 미래사회 대비 교육에서의 주요 역량이라 주장한다(예, OECD, 2018; OECD, 2018; UNESCO, 2015; 류태호, 2018; 이영희, 2018). 또한 <표 1>에서 제시한 핵심 키워드들 중 TF-IDF를 통해 추출된 키워드에는 나타나지 않는 것은 ‘창의·융합(창의성)’, ‘정보처리’, ‘코딩’, ‘자기 관리’, ‘의사결정’, ‘비판적 사고’ 등이다. 특히 ‘창의·융합(창의성)’은 많은 연구에서 공통적으로 미래사회 대비 교육에서 주요한 역량 또는 교육의 목표임을 주장하고 있다(예, 김광, 한재성, 2017; 최은미, 2017; 최지웅, 2017; 박진해, 2017; 나귀수 외, 2018). 이는 최근 3개년의 수학교육 논문들이 이러한 키워드를 포함한 연구들이 많이 이루어지고 있지 않은 결과라고도 추측할 수 있다. 따라서 향후 수학교육연구가 미래사회를 대비하는 방향과 보다 밀접해지기 위해서는 이러한 키워드를 고려한 연구가 활성화 될 필요가

있다.

참고문헌

- [1]교육부(2015). **수학과 교육과정**. 서울: 교육부.
- [2]김광진, 한재성(2017). 4차 산업혁명과 수학교육. **한국멀티미디어학회지**, 21(4), 33-38.
- [3]김대곤, 원진영(2013). **복합적 미래예측방법론 분석을 통한 미래재난예측기법 개발**. 국립재난안전연구원, 25-35.
- [4]김래영, 김구연, 권나영(2012). 연구 설계 및 연구 방법의 최근 동향: 초·중등 수학과 교육과정에 관한 연구를 중심으로. **학교수학**, 14(3), 395-408.
- [5]김선아, 박진희, 이현정, 정유진(2016). 텍스트 마이닝 기법을 활용한 다문화 미술교육 연구 동향 분석 연구. **다문화교육연구**, 9(2), 203-227.
- [6]김성근, 조혁준, 강주영(2016). 학술연구에서의 텍스트 마이닝 활용 현황과 주요분석기법. **정보화연구**, 13(2), 317-329.
- [7]김하진, 송민(2014). 동시출현단어 분석을 통한 국내외 정보학 학회지연구동향 파악. **정보관리학회**, 31(1), 99-118.
- [8]김현희, 이해영(2016). 토픽 네트워크 분석을 활용한 데이터 마이닝 분야 연구 논문 분석. **한국컴퓨터정보학회논문지**, 21(5), 141-148.
- [9]김혜미(2016). 수학 문제해결에 관한 국내 연구 동향 분석. **학습자중심교과교육연구**, 16(8), 831-850.
- [10]나귀수, 박미미, 김동원, 김연, 이수진(2018). 미래 시대의 수학교육 방향에 대한 연구. **수학교육학연구**, 28(4), 437-478.
- [11]류태호(2017). **4차 산업혁명 교육이 희망이다**. 서울: 경희대학교 출판문화원.
- [12]박진해(2017). 제 4차 산업혁명에 대비한 교양기초교육으로서의 수학교육의 방향. **한국교양교육학회 학술대회 자료집**, 27-29.
- [13]배규용, 박주현, 김정선, 이영섭(2013). 텍스트 마이닝 기법을 활용한 기후변화관련 식품분야 논문초록 분석. **한국데이터정보과학회지**, 24(6), 1429-1437.
- [14]서일원, 전체남(2013). 빅데이터 분석의 기술마케팅 활용에 관한 연구: 잠재 수요기업 발굴을 중심으로. **마케팅논집**, 21(2), 181-203.
- [15]송태민, 송주영(2016). **R을 활용한 소셜 빅데이터 연구방법론**. 서울: 한나래출판사.
- [16]송혜지, 박경수, 정혜은, 송민(2013). 텍스트 마이닝 기법을 활용한 한국의 경

- 제연구 동향분석. **한국정보관리학회 학술대회 논문집**, 20, 47-50.
- [17]신규식, 최회련, 이흥철(2015). 신재생에너지 동향 파악을 위한 토픽모형 분석. **한국산학기술학회 논문지**, 16(9), 6411-6418.
- [18]육동인(2017). 텍스트 마이닝을 활용한 직업학 연구동향 분석. **한국산학기술학회논문지**, 18(3), 586-599.
- [19]이수진, 이종학, 김원경(2013). 수학교육공학 연구 동향의 비교·분석. **교사교육연구**, 52(2), 253-266.
- [20]이영희, 윤지현, 홍섭근, 임재일, 백병부(2018). 미래교육 관련 연구 메타분석을 통한 미래교육의 방향. **교육문화연구**, 24(5), 127-153.
- [21]이인선, 나은영(2018). 텍스트 마이닝을 이용한 인지분야 학회지 분석(2000-2017). **인문사회**21, 9(3), 415-426.
- [22]임동훈(2015). **R을 이용한 빅데이터 분석**. 파주: 자유아카데미.
- [23]임화진(2014). **빅데이터를 이용한 충남도 정책 키워드 분석**. 충남발전연구원, 전략연구 2014-15.
- [24]정근하(2010). **텍스트 마이닝과 네트워크 분석을 활용한 미래예측 방법 연구**. 한국과학기술기획 평가원.
- [25]조태호(2001). 텍스트 마이닝의 개념과 응용. **지식정보인프라**, 5, 76-85.
- [26]채미현(2018). 환경복지 정책 모형 설정에 관한 탐색적 연구-텍스트 마이닝을 중심으로-. 서울대학교 행정대학원 석사학위 논문.
- [27]최은미(2017). 4차 산업혁명에서의 STEAM과 기초수학 교육의 방향 연구. **한국교양교육학회 학술대회 자료집**, 22-26.
- [28]최지웅(2017). 4차 산업혁명이 다가온 미래 사회의 수학교육 동향 고찰. **교과교육연구**, 10(1), 59-72.
- [29]황명화(2014). **여론 모니터링을 위한 비정형 빅데이터 시공간분석 방법론 연구**. 국토연구원, 국토연 2014-10.
- [30]Feldman, R., & Dagan, I. (1995). Knowledge discovery in textual databases(KDT). *KDD*, 95, 112-117.
- [31]Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21(3), 267-297.
- [32]Ian, F. (2014). wordcloud: Word Clouds. R package version 2.5. <https://CRAN.R-project.org/package=wordcloud>.
- [33]OECD. (2018). The future of education and skills education 2030. OECD. Retrieved from <http://www.oecd.org/education/2030>.
- [34]Turney, P. D., & Pantel, P. (2010). From frequency to meaning: Vector

space models of semantics. *Journal of Artificial Intelligence Research*, 37(1), 141-188.

- [35]UNESCO. (2015). Global citizenship education: Topics and learning objectives. United Nations Educational, Scientific and Cultural Organization.

Jin Mireu

Graduate School of Education Korea University

Suwon, 16499 Korea

E-mail address: mirfly19@ajou.ac.kr

Ko, Ho Kyoung

Ajou University

Suwon, 16499 Korea

E-mail address: kohoh@ajou.ac.kr