

수학교육에서 빅데이터 활용 방안에 대한 소고¹⁾

고 호 경 (아주대학교)

최 영 우 (아주대학교)

박 선 정 (국가수리과학연구소)[†]

최근 사회의 각 분야에서 언급되고 있는 빅데이터 열풍을 교육분야에서는 어떻게 수용 혹은 활용할 것인가? 본 연구는 수학교육에서의 빅데이터 활용 방안을 모색하기 위해 먼저 빅데이터의 개념과 활용 예시 등에 대해 살펴본 후, 두 가지로 향후 연구 방향을 모색하였다. 먼저, 기술과 환경의 변화에 따라 데이터 표현과 수용방식의 변화가 요구되고 있다. 다시 말해 수많은 정보들을 시각적으로 묘사하거나 필요한 정보를 '분석하고 추론'하여 효율적이고 명확하게 제공할 수 있도록 데이터를 다루는 학습 내용 및 방법의 변화가 요구되고 있고, 수학교육계는 21세기 학습자 역량강화와 연계하여 이를 수용하기 위한 교육과정 변화를 주도할 필요가 있다. 두 번째, 교수·학습 측면 뿐 아니라 교육분야에서 보다 적극적으로 필요한 데이터들을 수집하고 이러한 데이터 속에 기존에 알지 못했던 새로운 정보들을 파악함으로써, 이를 활용한 효율적인 수학교수·학습 변화를 추구하고 수학 교육에 대한 정책적 결정 등에 활용할 수 있도록 많은 관심과 연구를 촉진할 필요가 있다.

I. 서론

최근 정보통신 하드웨어와 소프트웨어의 발달과 더불어 기하급수적으로 증가되는 정보량에 따라 이를 통해 새로운 가치를 창출하고자 하는 빅데이터에 관심이 모아지고 있다(Manyika, J. et al., 2011). 최근 급증하는 디지털 정보량은 2011년도에 이미 전 세계적으로 생산, 유통되는 디지털 정보량이 1.8 제타바이트(ZB), 약 2조 기가바이트에 달하고 있어서 전 세계 인구 1인당 평균적으로 280기가바이트에 해당하는 엄청난 정보량이 축적되었다고 보고되고 있다(김지숙, 2013). 이에 맞춰 IT를 활용한 다양한 산업분야의 전자화나 자동화 역시 크게 진전됨으로써 기술의 발전과 더불어 처리하고 활용할 수 있는 데이터 역시 늘어나고 있는 추세를 반영하여, 미국 과학기술정책자문위원회가 2010년 모든 연방정부는 빅데이터 전략 수립이 필요하다고 판단함에 따라 오바마 행정부는 빅데이터 기술연구개발에 2억 달러 이상을 투자하는 '빅데이터 연구개발 이니셔티브(Big Data Research and Development Initiative)'를 발표(2012.03)하였으며, 일본 역시 빅데이터 활용 특별부회에서 빅데이터 활용을 위한 정책 기본방향을 정리한 '빅데이터 활용 기본전략(2012.05)'을 발표하였다(한국정보화진흥원, 2013). 우리 정부 역시 빅데이터가 민간 및 공공부문의 새로운 가치를 창출하는 미래 국가경쟁력의 핵심으로 대두됨에 따라 '스마트국가 구현을 위한 빅데이터 마스터 플랜'을 발표(2012.11)하고 빅데이터를 구축 및 활용하고자 하는 국정과제를 수립하였다(김윤정 외, 2013).

디지털 시대에는 실생활 속에서 축적되는 다양한 유형의 데이터가 증가할수록 데이터의 활용가치는 무한히

1) 본 논문은 2013학년도 아주대학교 일반연구비 지원에 의하여 연구되었음

* 접수일(2014년 11월 9일), 심사(수정)일(2014년 11월 16일), 게재 확정일(2014년 11월 26일)

* ZDM 분류 : A5, B1

* MSC2000 분류 : 9702, 97C99

* 주제어 : 빅데이터, 수학교육, 수학교육연구

† 교신저자 : seonjeong1124@nims.re.kr

상승한다는 판단 하에, 쌓여 있는 데이터 속에서 유용하고 가치 있는 정보를 찾기 위한 노력을 끊임없이 기울이고 있으며, 이러한 빅데이터 처리의 특징을 만족시키기 위한 다양한 스토리지, 컴퓨팅 기술 및 분석기법들을 개발하고 있다(한국정보화진흥원, 2013). 예를 들어, 빅데이터를 분석하는 기법 중 하나인 마이닝 기법은 기업의 의사결정, 마케팅 외에도 금융, 행정, 의학, 교육 등의 분야에서 혁신적으로 적용되고 있거나 그 가능성을 이야기하고 있다(한국정보화진흥원, 2013). 많은 기업에서는 이미 빅데이터 분석과 활용을 통해 경제적 이익을 창출하는데 관심을 두고 있는데 가령, 구글은 통계적 기계 번역(statistical machine translation)이라고 부르는 자동번역 시스템 기술을 개발하여 컴퓨터에게 문법을 가르치지 않고 사람이 이미 번역한 수억 개의 문서에서 패턴을 조사해서 언어 간 번역 규칙을 스스로 발견하도록 하는 방식을 도입하고 있다(김지현 외, 2014).

그러나 다양한 분야에서 빅데이터 기술을 활용한 성과 사례가 나오고 있는 것에 반해, 교육 분야에서 빅데이터의 활용 단계는 아직 시작 단계에 불과하다. 현재 학습 평가와 학습 분석 등에 대한 연구가 상당히 진전되어 일부 적용단계 수준까지는 와 있기는 한데, 이 분야 역시 아직 국내에서는 빅데이터 기반으로서의 데이터 수집, 검색, 분석, 그리고 시각화 관련 기술이 구현 가능한 상황은 아니다(임현정, 2014). 교육 분야에서 빅데이터를 활용하는 아이디어가 미진하거나 구체적인 사례를 제시하지 못하는 데는, 공학자나 데이터 과학자들은 교육에서 무엇이 필요한 데이터이고 어떻게 활용해야 하는지에 대한 설계가 분명치 않아 아직은 초기적인 데이터 표집이나 분석 방안 정도를 제안하고 있는 상황이라 할 수 있다.

수학교육 분야 역시 다른 교육 분야와 상황은 크게 다르지 않아서 빅데이터의 활용 방안에 대한 연구가 거의 전무한 실정이다. 그러나 향후 빅데이터를 활용한다면 그동안 얻지 못했던 많은 정보들을 얻을 수 있을 가능성이 농후한데, 예를 들어 수학교육에 관한 내용 영역 지식(contents domain knowledge)을 기반으로 행동영역이나 상황에 대한 연결망분석(actions-context network analysis)을 활용하여 빅데이터화한 자료를 축약(reduction) 처리하고, 이를 시각화(visualization) 함으로써 보다 종합적인 자료를 활용함으로써 개인별 지식영역에 효율적인 학습방법이나 학습 효과 등을 제안할 수 있을 것이다.

따라서 본 연구는 수학교육 분야에서 구체적인 빅데이터 연구를 진행하기 전에, 수학교육에서 빅데이터를 활용한 자료 분석과 활용에 대한 필요성을 제안하고 빅데이터를 활용하여 무엇을 할 것인가에 대한 질문을 생성하는데 도움을 제공함으로써 빅데이터를 활용한 연구의 활성화에 기여하고자 한다.

II. 빅데이터의 개념

1. 빅데이터의 정의

빅데이터에 대한 정확한 학술적 정의는 없다. 학문적 특성에 따라 혹은 이를 활용하는 영역에 따라 조금씩 다르게 정의하고 분류하고 있는 실정이다. 빅데이터에 대한 논의들을 살펴보면 빅데이터의 특징을 크게 데이터 측면과 데이터 저장, 관리, 처리의 측면에서 제시하는 것으로 보이며, 2011년 McKinsey Global Institute에서는 3V 라고 해서, 크기(volume)·속도(velocity)·다양성(variety) 측면에서 기존 데이터와 차별화시키는 것으로 회자되고 있다.

미래사회의 특성 대비 빅데이터의 역할을 살펴보면(<표 1>), 빅데이터는 미래사회의 불확실성과 리스크를 통찰력을 바탕으로 예측하고 대응하는데 기여할 것이며, 점점 스마트해지고 융합되어 가는 사회에서의 경쟁력을 강화시키는 수단이 될 것이라고 보고 있다(한국정보화진흥원, 2013).

<표 1> 미래사회의 특성과 빅데이터의 역할(한국정보화진흥원, 2013)

| 미래사회 특성 | 빅데이터의 역할 | |
|---------|----------|--|
| 불확실성 | → 통찰력 | - 사회현상, 현실세계의 데이터를 기반으로 한 패턴분석과 미래 전망 - 여러 가지 가능성에 대한 시나리오 시뮬레이션 - 다각적인 상황이 고려된 통찰력을 제시 - 다수의 시나리오로 상황 변화에 유연하게 대처 |
| 리스크 | → 대응력 | - 환경, 소셜, 모니터링 정보의 패턴 분석을 통한 위험징후, 이상 신호 포착 - 이슈를 사전에 인지·분석하고, 빠른 의사결정과 실시간 대응 지원 - 기업과 국가 경영의 투명성 제고 및 낭비요소 절감 |
| 스마트 | → 경쟁력 | - 대규모 데이터 분석을 통한 상황인지, 인공지능 서비스 등 가능 - 개인화, 지능화 서비스 제공 확대 - 소셜(니즈)분석, 평가, 신용, 평판 분석을 통해 최적의 선택 지원 - 트렌드 변화 분석을 통한 제품 경쟁력 확보 |
| 융합 | → 창조력 | - 타분야와의 결합을 통한 새로운 가치창출(의료정보, 자동차정보, 건물정보, 환경정보 등) - 인과관계, 상관관계가 복잡한 컨버전스 분야의 데이터 분석으로 안전성 향상, 시행착오 최소화 - 방대한 데이터 활용을 통한 새로운 융합시장 창출 |

2. 빅데이터 처리 기술과 분석 기법

빅데이터 분석 기술은 크게 데이터 처리를 위한 인프라 기술과 데이터를 분석하는 분석기법으로 나누어 볼 수 있다.

가. 빅데이터 처리 기술

근래에 빅데이터 분석을 수행할 수 있는 인프라 기술은 다양하게 개발되고 있지만 일반적으로 많이 활용되는 것 중에는 대용량 데이터 처리 능력을 위한 분산처리 기술인 하둡, 의미 분석 기술과 진보된 알고리즘 및 데이터 마이닝 기술을 보여주는 R, 비정형 데이터를 처리하기 위한 기술인 NoSQL 등이 있다(김지숙, 2013).

이에 대해 그 특징들 위주로 간단히 살펴보면, 우선 하둡은 다수의 서버를 묶어 분산 처리하는 플랫폼으로서, 분산처리를 하는 'Map' 단계와 결과를 취합하는 'Reduce' 단계로 이루어진 것이며, 구글의 맵리듀스(MapReduce) 모델을 본떠 만든 것으로 Apache Group에 의해 제작되고 배포되었다(김지숙, 2013). 맵리듀스 프레임워크는 저렴한 컴퓨터를 수백, 수천 대를 서로 연결해서 만든 클러스터에서 병렬 분산 프로그램을 처리할 수 있는 방법을

제시해준다. 구글(Google)의 맵리듀스 또는 오픈 소스 진영에서 만든 하둡(Hadoop)은 그러한 병렬 분산프로그램을 개발하기 위한 맵리듀스 프레임워크를 제공하는 효율적인 툴이다(심규석 외, 2014). 하둡은 아파치 소프트웨어 라이선스에 따라 전체 프레임 워크가 오픈 소스이기 때문에 기본 소프트웨어에 대한 라이선스 비용이 없을 뿐만 아니라 고가의 하드웨어나 고성능 프로세서가 필요치 않고 저가의 상용 서버를 사용할 수 있기 때문에 비용 절감 효과가 있는 것으로 평가되고 있다(양성은, 2014). 빅데이터 처리 성능 향상을 위한 병렬 처리 성능 향상 기법 및 다양한 분석을 위한 병렬 질의 처리 기법에 대한 연구는 앞으로도 지속될 것이며 이로 인해 데이터 처리 능력은 꾸준히 향상될 것으로 전망된다(박준호 외, 2014)

두 번째 오픈 소스 프로젝트 R은 통계 계산 및 시각화를 위한 언어 및 개발환경을 제공하며, 기본적인 통계 기법부터 모델링, 최신 데이터 마이닝 기법 등이 구현 가능하다. 이렇게 구현한 결과는 그래프 등으로 시각화할 수 있기 때문에, Java나 C, Python 등의 다른 프로그래밍 언어와의 연결도 용이하다. 또 R의 경우에는 무료로 사용할 수 있고, 다양하게 활용할 수 있는 패키지들이 빠르게 추가되고 있어 최근 들어 더욱 주목을 받고 있다. 위의 장점들로 인하여 R은 통계 분석 분야에서 활용도를 높이고 있는데, 하둡 환경상에서 분산처리를 지원하는 라이브러리 덕분에 구글, 페이스북, 아마존 등 빅데이터 분석이 필요한 기업에서 대용량 데이터 통계분석 및 데이터 마이닝을 위해 널리 사용되고 있다(김지숙, 2013).

세 번째, NoSQL은 데이터의 양이 지속적으로 증가하고 해석해야 할 데이터의 요구가 일관적이지 않은 웹 시장의 요구를 수용하고, 기존의 관계형 데이터베이스의 수평적 확장성의 한계를 해결하고자 대두되었다. 관계형 데이터베이스의 경우, 일관성(Consistency: 모든 노드는 같은 시간에 같은 데이터를 보여줘야 한다)과 가용성(Availability: 일부 노드가 다운되어도 다른 노드에 영향을 주지 말아야 한다)에 중점을 두고 있는 반면, NoSQL 기술은 네트워크 전송 중 일부 데이터를 손실하더라도 시스템은 정상 동작이 되는 분할지속성(Partition Tolerance)에 중점을 두고 일관성과 가용성은 보장하지 않는다(김지숙, 2013).

나. 빅데이터 분석 기법

빅데이터 시대에는 데이터가 가장 관심이 주목되는 대상이 됨으로써, IT의 주 관심사가 인프라 구축, 기술, SW에서 데이터로 전이되었다고 한다. 최근 빅데이터가 이슈가 되자 데이터분석의 중요성에 대한 인식과 다양한 데이터 마이닝 기법이 새롭게 조명되고 있다(한국정보화진흥원, 2013). 다양한 마이닝 중 몇 가지 대표적인 것들을 살펴보면, 우선 데이터 마이닝(Data Mining)을 들 수 있는데, 데이터 마이닝은 대용량의 데이터, 데이터베이스 등에서 감춰진 지식, 기대하지 못했던 경향, 새로운 규칙 등의 유용한 정보를 발견하는 과정 데이터 마이닝을 통해 정보의 연관성(순차 패턴, 유사성 등)을 파악함으로써 가치 있는 정보를 의사결정에 활용하는 기술이다. 다시 말해, 기업이 보유하고 있는 각종 생산품 관련 자료, 거래자료, 마케팅 활동의 피드백 자료와 기타 외부자료를 포함하여 사용 가능한 데이터를 기반으로 숨겨진 정보, 기대하지 못했던 패턴, 새로운 법칙과 관계 등을 찾아내어 이를 기업 경영을 위한 정보로 활용하는 것이다(김지숙, 2013).

두 번째, 텍스트 마이닝(Text Mining) 또는 텍스트 기반의 지식 발견(Knowledge Discovery in Texts)이라는 것은, 텍스트 기반 데이터베이스로부터 기존에 알려지지 않은 새로운 정보를 자동적으로 추출하는 기법을 의미하는 것으로, 비/반정형 텍스트 데이터에서 자연어처리(Natural Language Processing)기술을 기반으로 유용한 정보를 추출, 가공하는 것을 목적으로 한다(구주나, 2014). 텍스트 마이닝 기술을 통해 방대한 텍스트 문치에서 의미 있는 정보를 추출해내고, 다른 정보와의 연계성을 파악하며, 텍스트가 가진 카테고리를 찾아내거나 단순한 정보 검색 그 이상의 결과를 얻어낼 수 있다. 컴퓨터가 인간이 사용하는 언어(자연어)를 분석하고 그 안에 숨겨진 정보를 발굴해 내기 위해 대용량 언어자원과 통계적, 규칙적 알고리즘이 사용되고 있다(김지숙, 2013).

세 번째, 최근에 새로운 여론 분석기술로 주목받고 있는 오피니언 마이닝(Opinion Mining)을 들 수 있다. 이는 소셜미디어와 웹사이트 등에 나타난 여론과 의견을 분석하여 유용한 정보로 재가공하는 기술로써, 오피니언

마이닝 혹은 감정분석을 활용하면 네티즌이 웹 상에 올려 놓은 각종 글들을 분류하여, 더 객관적이고 정확한 정보로써 활용하고자 하는데 사용한다. 이러한 활용성의 오피니언 마이닝 기법에서 발생할 수 있는 오류를 피하기 위하여 사용되는 연구의 3단계를 김정숙(2012)은 다음과 같이 제시한 바 있다;

1단계: '주관성 분석'으로 주어진 텍스트가 주관적인지 객관적인지 결정하는 것으로, 주어진 텍스트에 나타난 저자의 태도를 판단하는 단계이다. 2단계: '극성 분석'으로 텍스트가 주관적인 의견을 갖고 있을 경우 긍정인지 부정인지를 분류하는 단계이다. 3단계: '극성의 정도 분석'으로 주관적인 텍스트에 대하여 긍정적인 정도와 부정적인 정도를 측정하는 단계이다. 오피니언 마이닝은 특정 서비스 및 상품에 대한 시장규모 예측, 소비자의 반응, 입소문 분석 등에 활용되고 있으며, 공공분야의 경우 민원의 원인이나 문제점 등을 파악하는 것이 용이해 서비스를 개선할 수 있다. 또한, 기업은 특정 제품에 대한 고객의 반응을 빠르게 파악하고 선호도를 역으로 추론하는데 효과적으로 활용할 수 있다. 정확한 오피니언 마이닝을 위해서는 전문가에 의한 선호도를 나타내는 표현과 사전(dictionary) 구축 역시 필요하다

네 번째, 웹 마이닝(Web Mining)이라 하여 인터넷상에서 수집한 정보를 데이터 마이닝 방법으로 분석으로 통합하는 기법이다. 웹 마이닝은 콘텐츠 마이닝(웹 검색, 수집 데이터), 구조 마이닝(웹 사이트 구조), 활용 마이닝(사용자 이용형태) 등으로 세분될 수 있다(한국정보화진흥원, 2013).

다섯 번째, 소셜 마이닝(Social Mining)이라는 것으로 소셜 네트워크의 분석(Social Network Analysis)기법을 사용한다. 이는 집단 내 개체의 연결 상태 및 연결 구조의 특성을 수량적으로 파악하고 시각적으로 표현하는 기법으로, 수학의 그래프 이론에 뿌리를 두고 있으며, 소셜 네트워크 연결구조 및 연결 강도 등을 바탕으로 사용자의 명성 및 영향력을 측정하여, 소셜 네트워크상에서 입소문의 중심이나 허브 역할을 하는 사용자를 찾는데 활용되고 있을 뿐 아니라, 전통적인 네트워크 이론에 기반한 다양한 중심성 척도들이 편리하게 계산되기 때문에 점차 그 활용 분야가 확대될 것으로 내다본다(조인동, 2011).

여섯 번째, 현실 마이닝(Reality Mining)이란 사람들의 행동패턴을 예측하기 위해 사회적 행동과 관련된 정보를 휴대폰 등 모바일 기기를 통해 얻고 분석하는 기술로서, 이러한 기기들을 통해 현실에서 발생하는 정보를 기반으로 인간관계와 행동 양태 등을 추론하는데 활용된다(한국정보화진흥원, 2013)

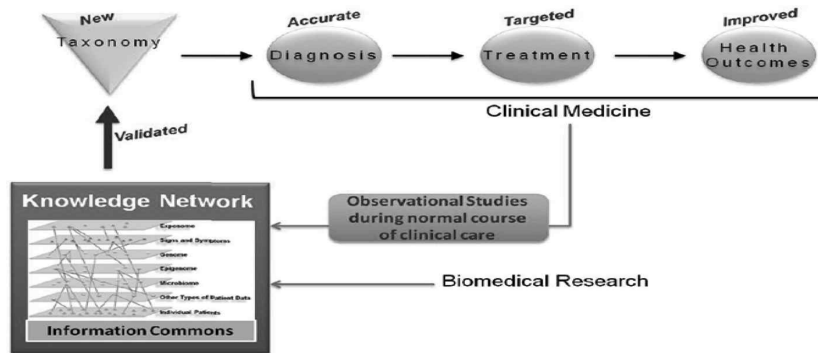
마지막으로 군집분석(Cluster Analysis)은 비슷한 특성이 있는 개체를 합쳐가면서 최종적으로 유사 특성의 군을 발굴하는데 사용되는 기법이다. 군집분석을 위하여 가장 흔히 사용하는 자료는 간격척도 혹은 비율척도로 측정된 거리 값(Distance measures)의 경우에 따라 서열척도로 측정된 값들로 군집분석을 실시한다(김지숙, 2013). 계층적 군집 분석은 계층적 방법에 의해 군집화를 한 다음 그 결과로부터 가장 적절한 수의 군집 수를 결정하여 다시 비 계층적 방법에 의해 분석하면서 그 수를 지정하는 방법을 사용하는 기법이다(김정숙, 2012).

III. 빅데이터 활용 사례

빅데이터 활용에 대한 예는 다양한 분야에서 찾아볼 수 있는데, 먼저 빅데이터 기술의 활용을 의학 분야에 사용하고자 하는 연구들을 예로 들 수 있다. 이는 의료기관의 원활한 운영과 수준 높은 연구에 빅데이터를 활용하고자 하는 시도들로서, 의료계는 타산업에 비해 비정형데이터가 차지하는 비중이 높고 이용할 자료는 존재하지만 효과적인 이용을 위한 분석시스템이나 체계가 부족하기 때문에 빅데이터를 활용하여 자료 활용 폭을 넓힐 수 있다고 하였다(김성수, 2014). 또한 김지현 외(2014)는 개인의 생체정보나 유전자형 등을 분석하여 맞춤형으로 약물을 처방한다는 맞춤형요법은 개인의 유전정보 수집을 전제로 하는데, 빅데이터들이 좀 더 용이하게 처리될 수 있게 될 것이기 때문에 맞춤형요법 적용이 보다 더 용이해질 수 있다고 보았다. 실제로 미국의 National Research Council(NRC)에서 발행된 'Toward Precision Medicine: Building a Knowledge Network for

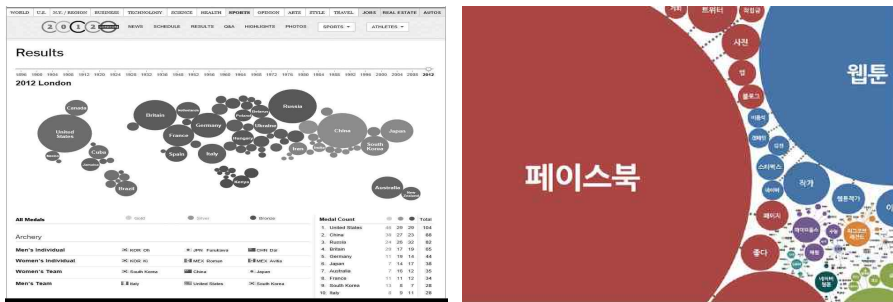
Biomedical Research and New Taxonomy of Disease'에서는 정보 활용을 통한 질병 치료와 맞춤형물요법에 대해 단순히 신약개발과정 속에서의 피드백이 아니라, 다른 분야의 축적된 정보로부터 약에 대한 지식을 창출해내는 과정에 대해서도 설명하고 있다고 소개한다(김지현 외, 2014).

이미 익히 알려진 바와 같이, 빅데이터를 국가현안이나 공공분야, 지방 행정에도 그 효용 가능성이 제기 되면서 보다 적극적으로 활용하고자 하는 연구가 활발히 진행되고 있다. 예를 들어, 가치가 높은 정부·공공기관 데이터 공유와 과학기술 분야의 빅데이터 활용체계 구축을 통해 국정 현안 사전대응, 기회 포착 등이 가능할 것이라는 것이다(김윤정 외, 2013). 공공분야인 교통시스템, 수자원시스템, 방범시스템, 탈세방지시스템과 같은 분야에서 역시 빅데이터를 통한 시스템 효율화가 가속화 될 것이라 보았으며, 이는 이미 브라질, 싱가포르, 미국 등과 같은 나라에서도 시도되고 있다고 보고하고 있다(이성훈, 2013). 또한 노규성(2014)은 역시 빅데이터를 활용하여 사용자의 요구분석, 주요성공요소 기반 추진, 시범사업, 성과평가, 성과에 따른 인센티브, 공동기반 구축 등의 정책과 행정 서비스에서 예측 및 새로운 정책 개발 등이 가능하다 하였다.



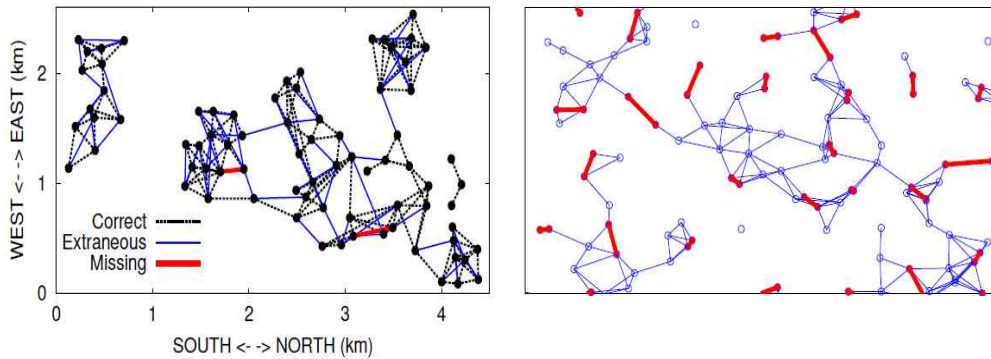
<그림 1> 빅데이터를 활용한 의학 및 맞춤형물 연구(NRC, 2011; 김지현 외 2014 재인용)

이렇듯 빅데이터가 가장 큰 관심을 받고 있는 이유 중 하나는 데이터의 시각화이다. 뉴욕타임스는 정보 데이터를 시각화한 인터랙티브 인포그래픽과 모션 인포그래픽 등 다양한 인포그래픽을 선보이고 있는데, 이러한 인포그래픽과 사용자들이 함께 참여하는 인터랙티브한 인포그래픽은 사용자들에게 높은 호응과 관심을 이끌어 낼 수 있었다(김한수·김준교, 2014).



<그림 2> 정보 데이터 시각화 예시(좌: 올림픽 결과에 대한 인포그래픽으로서의 올림픽 메달지도, 김한수·김준교, 2014, p.256, 우: 소셜 매트릭스 탐색어 맵을 활용한 연결형 인포그래픽, 이상훈, 2013, p.85)

그 외에도 너무나도 다양한 분야에서 다양한 종류의 빅데이터들이 창출되고 이를 활용하려는 연구들이 시행되고 있는데, 그 중 또 다른 한 예가 공간 빅데이터의 활용이라 볼 수 있다. 이는 예를 들어, 근접성 및 규모, 위도, 경도, 고도 등 위치참조를 위한 공간적 속성을 나타내는 정보를 의미하는 공간정보가 데이터의 관리·처리·분석에 있어서 데이터의 양(Volume), 데이터 생성속도(Velocity), 형태의 다양성(Variety)이 현재의 시스템으로 수용이 어려운 상태인 빅데이터로 다루어지고 있다(안중욱 외, 2013). 위성영상 정보, 항공사진 정보 등과 같은 래스터(Raster) 정보, 점·선·면으로 이루어진 벡터(Vector) 정보, 도로망, node, edge, path 등의 그래프(Graph) 정보로 분류할 수 있고, 질병, 재난, 재해, 범죄 등의 발생 분포 분석, 핫스팟 및 핫스팟 가능 지역 탐지, 공간적 상관관계 분석 등에 활용될 수도 있고 Navteq의 Temporally detailed roadmaps 등 도로망, 전력망, 공급망 등과 같은 공간적 그래프를 포함하며, 시간에 따른 접근성 분석, 최적 출발시간 산정, 시간대에 따른 최적 경로 분석 등에 활용이 가능하다(안중욱 외, 2013). 예를 들어 Zhou(2013)의 연구에서는 빅데이터가 실제에서 새로운 네트워크 디자인을 구현하는데 사용될 수 있는가를 보여주고 있는데, <그림 3>에서와 같이 대량의 숫자로부터 정확하게 간섭 조건을 특성화해 나간 그래프 예시(그림 3 왼쪽)이며, 충돌 조건을 설명하는 샘플 네트워크 영역을 빅데이터 분석기법으로 제시한 예시이다.



<그림 3> 빅데이터를 활용하여 교통 네트워크 디자인을 위한 기초자료를 제시한 예시(Zhou, 2013)

IV. 교육에서의 빅데이터 활용 연구

교육 분야의 데이터는 노출된 데이터를 구하는데 여러 가지 제약조건이 있을 뿐 아니라 비정형 데이터가 주류를 이루고 있다. 따라서 교육관련 빅데이터는 이를 어떻게 처리해서 학습 시스템을 개선할 것인가가 중요한 관건인데 이러한 기술에 대한 연구 분야는 학습 분석(learning analytics)와 교육 데이터 마이닝(educational data mining) 분야로 나뉘어 볼 수 있다. 교육 데이터 마이닝은 데이터에서 새로운 패턴을 찾아 새로운 알고리즘과 모델 개발 방향의 기술이고, 학습 분석은 교육 체제에서 알려진 예측 모델을 적용하기 위한 기술 방향이다(임현정, 2014).

<표 2> 학습 분석과 교육 데이터 마이닝 기술 특징(임현정, 2014, p.53)

| 학습 분석 | 교육 데이터 마이닝 |
|------------------------------------|-----------------------------|
| - 정보과학, 사회학, 심리학, 통계학, 컴퓨터 과학, 교육학 | - 통계학, 기계학습, 데이터마이닝 |
| - 전체를 이해하고자 함 | - 학습을 작은 부분들로 분해 |
| - 기존의 방법과 모델을 적용하여 학습과 관련된 문제를 분석 | - 데이터를 분류하고 관계를 찾아 예측모델을 개발 |
| - 자동화된 처리 대신에 교사의 결정을 지원 | - 학생에 대한 자동화된 처리 (적응학습 시스템) |

교육분야에서 빅데이터를 활용하기 위해서는 데이터를 구축하기 쉬운 스마트 환경에서의 설계가 일반적인 접근이라 볼 수 있는데, 송지숙(2014)은 빅데이터 기반의 스마트교육 활용방안을 통해 학생에게 데이터에 기반한 개별적인 맞춤형 교육을 제공할 것을 제안한 바 있다. 이 연구에 따르면, 학생들의 학습 이력데이터 즉, 학생의 응답정확도, 오답의 횟수, 문제풀이 시간 데이터, 학생의 학습활동 등의 데이터를 하둠 분산 파일 시스템에 저장하여 그 데이터를 맵리듀스로 처리하여서 학생의 온라인 학습데이터와 온라인 학습으로 발생하는 학생의 데이터를 분석하여 이를 학업 성취도 향상에 활용할 수 있을 것이라 하였다.

최근에는 스마트 학습과 유사한 환경이라 할 수 있는 클라우드 환경과 디지털교과서에 빅데이터를 적용하고자 하는 시도도 일고 있다. 스마트 교육의 일환으로 도입하고 있는 디지털 교과서 개발 동향과 향후 방향은 “학습자의 학습기록 분석을 통한 맞춤형 적응형 교육 서비스 제공이 가능한 플랫폼이며, 연결주의와 생각공유 기반의 협업학습 기능을 중심으로 다양한 교수-학습 활동이 이루어지는 플랫폼(성정숙·김현철, 2013 p.56)”이라고 한다. 디지털교과서는 시/공간에 구애받지 않는 멀티미디어 학습 자료로서 학습자의 학습 수준에 따른 능력별 학습과 양방향 학습을 지원하는 교수-학습지원 시스템적 특징(박춘자, 2011)을 강화시키기 위한 방안으로 빅데이터를 활용하여 학습자의 학습 기록 분석 등을 통해 맞춤형 학습과 자기주도적 학습을 실현하고자 하는 노력을 기울이고 있다(성정숙·김현철, 2013).

그 외 교육용 자료를 구축하기 위한 시도들도 있는데, 예를 들어 전원표(2014)는 많은 인력과 시간이 소비되는 개체명 클래스를 빅데이터 기법을 활용하여 능동학습 방법을 이용한 개체명 사전 반자동 구축 방법을 제안한 바 있다. 또한 손정은(2013)은 빅데이터에서 데이터 종류도 매우 다양해짐에 따라 그에 따른 처리 기술이 ICT(Information & Communication Technology)분야의 중요한 이슈가 되고 있다고 하였으며, 분석 자체가 정보 검색 기술을 기반으로 이루어져야 하지만 이를 통하여 빅데이터로부터 더 많은 분석 정보를 도출할 수 있을 것이라 제안한 바 있다(손정은, 2013).

빅데이터를 교과교육에 활용하고자 하는 시도도 있는데, 예를 들어 임현정(2014)은 음악 교육 분야에 대한 빅데이터의 활용은 개별 맞춤형 교육에 한정되어 있기는 하나 예술적 자료에 대한 빅데이터 작업을 시작한 단계에 와있으며 앞으로도 음악 교육 특수성을 살리기 위한 빅데이터 적용 노력을 꾸준히 기울일 필요가 있다고 제안한 바 있다.

V. 수학교육에서 빅데이터 활용 방안 예시

1. 스마트 교육환경에서의 빅데이터 활용

정보화시대에 이어 스마트시대에 접어들면서 데이터가 자원으로 활용되면, 이미 클라우드에 기반한 학습분석(learning analytics) 기법을 디지털교과서에 적용함으로써 학습자의 평가에 이용하거나 맞춤형 상호작용적 서비스를 제공하는 데 이용하려는 시도가 있다. 그 예로, 성정숙·김현철(2013)의 조사에 따르면 Kno사의 Kno Me의 경우에는 학습자의 행동, 상호작용 수준, 시간관리, 진도 등에 대한 통계 정보를 실시간으로 제공하는 등 교수자에게 학습자를 평가하거나 학습지도를 하는데 유용하게 사용될 수 있는 정보를 관리해 나간다는 것이다. 그러나 아직까지 스마트환경이 학습에 충분히 도입된 시점이 아니며 학습분석에 대한 데이터 처리나 분석 기술에 시작 단계에 불과한 수준이라 할 수 있다. 그러나 이러한 학습자에 대한 다양하고 정확한 분석은 NMC Horizon Report 2013 Higher Education Edition에서 향후 2-3년 내에 교육 공학기술 분야의 핵심 발전안으로 선정될 만큼 주목할 만한 기술이다(Johnson, et al., 2013). 따라서 이러한 맞춤형 적응형 교육 서비스 제공이 가능하기 위해서는 빅데이터를 교육에 전적으로 도입하기 위한 연구는 필수불가결한 일이라 할 것이다.

이러한 주장에 근거해서 수학교육에서 빅데이터를 활용하게 될 장면을 떠올려보기로 하자. 우선 전통적인 교육방법에서는 학생의 평가나 학습활동과 관련된 대부분의 데이터가 인위적으로 기록하지 않고는 기록된 내용을 보존하거나 활용하는 것은 거의 불가능하다. 그러나 스마트 환경을 활용한 스마트 교육에서 이루어지는 학습활동은 거의 모든 내용이 디지털 데이터로 저장 가능하다. 예를 들어 한 학생이 학습에서 활용하는 과정에서 생성되는 빅데이터와 관련된 학습활동 자료의 예를 보면 다음과 같은 상황을 생각해 볼 수 있을 것이다;

- 학생이 스마트 기기를 사용하여 온라인 접속한 뒤 온라인 학습을 진행함
- 필요한 특정 수학 개념에 대한 동영상 강의나 자바 애플릿 환경에서 제공하는 내용을 학습하거나 반복 학습함
- 스마트 기기가 판단하여 제공한 문제를 풀거나 필요한 문제를 선택하여 해결함(특정 문제는 빠른 시간에 풀지만 특정 문제는 오랜 시간이 걸림)
- 과제를 제출하거나 온라인 퀴즈를 풀고 온라인 설문조사에 응함
- 필요한 콘텐츠를 다운로드하여 이용하고 다른 스마트 기기에 콘텐츠를 동기화하여 다른 학생과 콘텐츠를 공유함
- 학습과 관련된 검색을 하거나 자신의 풀이과정 등을 남김
- 대화방에서 다른 학생과 대화를 하거나 SNS를 이용하여 글을 게시하거나 질문을 남김
- 각 학생의 학습 목표, 학습 성향, 기술 및 능력, 총 학습 시간, 각 문제에 대한 풀이 시간, 정답률, 성취도 등의 학습자 정보가 저장됨.

스마트 환경을 백분 활용하기 위해서는 이러한 각 학습 과정에서 발생하는 모든 데이터를 빅데이터로 처리, 분석하여 활용할 수 있다. 예를 들어 스마트 기기가 학생에게 맞는 문제를 제공할 때, 각 학생들이 학업 시 발생한 데이터를 활용하여 '학생을 예측해서 문항을 제공'하는 데 빅데이터 처리기법을 활용할 수 있으며, 개별 학생들의 장단점이나 학습에 필요한 내용 등에 대한 '시각화된 결과물 제공'에서도 빅데이터는 가치 있게 활용될 수 있을 것이다.

이와 같은 주장에 대한 근거 중 하나로, 최근 기사 중에 "포토매쓰"라는 어플리케이션이 소개된 바 있다(<http://media.daum.net/digital/others/newsview?newsid=20141027181907285>). 여기서는 스마트폰으로 문제를 인식시키고 그 풀이과정을 제시해주는 어플리케이션에 대한 소개였는데, 현재의 수준으로도 충분히 구현할 수 있는 어플리케이션이다(Wolfram Alpha를 이용하면 카메라 인식은 되지 않지만, 현재 보다 더 어려운 문제의 풀이도 볼 수 있음). 따라서 미래에는 학생들의 답안을 지필로 쓴 것이든, 타블릿 PC에 쓴 내용이든 그 풀이과정을 컴퓨터가 인식해서 채점을 하거나, 스마트교육에 있어서 학생의 풀이과정을 실시간 트래킹함으로써 학생들의 답안을 살펴봄으로써 새로운 해법을 찾을 수도 있을 것이고, 현재 학생이 어느 부분에서 틀렸는지를 추적하고 올바른 해법과 함께 다음에 풀 문제를 자동으로 제시하는 것이 가능할 수 있다는 것이다. 또한 교수학적으로도 어떤 형태

의 오답이 많은지가 필요에 따라 자동으로 산출되고 이에 근거해 필요한 교수학습 내용이 개별 학생에 맞게 제공될 수 있을 것이다.

2. 수학교육 연구에서의 빅데이터 활용

빅데이터는 각종 수학교육 관련 연구에서 활용 가능하다. 예를 들어, 빅데이터는 기존의 자료 분석의 방법이나 활용 방안과는 차별화 된다는 특성(장영재, 2013)을 활용하여, 리더십 개발 연구에도 빅데이터를 활용하려는 움직임이 일고 있다. 그동안 설문지 중심의 횡단연구에 대한 문제점 등이 제기되면서 현대적 리더십 이론을 연구하기 시작한 초기에 구축된 이론으로 리더의 신체적 특징, 지능, 성격, 태도, 사회적 성향 등이 리더십의 본질이라는 중단적 관점의 특성이론을 활용하려 하는 것이다. 이때 빅데이터를 활용하여 리더십개발의 다양한 측면에 대해서 보다 광범위한 입체적 분석이 가능해질 수 있다는 주장이다(백기복·김정훈, 2013). 이와 마찬가지로 정보화 시대의 사회적 요구에 부응하기 위해서는 수학교육학적 성찰을 바탕으로 새로운 연구방법의 가능성을 추구해야 한다. 정보화에 대한 적극적인 수용은 연구 자료의 정보화만을 의미하는 것이 아니라 정보화 기술에 대한 폭 넓은 이해를 바탕으로 새롭게 수학교육학 연구 방법을 정립하는 것도 포함한다. 빅데이터는 수학교육학의 방법론과 범주에 큰 기여를 할 수 있다. 우선 방법론적인 면에서 수학교육학의 연구에 보다 많은 사례들을 다양하게 제시할 수 있다. 빅데이터 시스템을 활용하게 된다면 많은 유사 사례들을 분석할 수 있고, 수학교육학 연구에 대한 다층적인 접근과 해석을 시도해 볼 수 있게 됨에 따라 연구의 객관성을 좀 더 높일 수 있을 것이다. 또한 범주의 측면에서 빅데이터는 수학교육학의 연구 범위를 확대시킬 수 있다.

그 중 한 예로 ‘근거 이론(grounded theory)’과 같은 연구방법론의 응용을 들 수 있다. 기존의 근거 이론 연구 방법은 어떤 현상에 대한 경험적 자료들을 수집한 후 개념들의 수정·통합을 통해 귀납적인 과정을 거쳐 자료들을 분석하고 이를 통해 어떠한 현상들에 관한 기본적인 이론을 설정해 나가는 방식이다(Strauss & Corbin, 1998). 이는 연구 참여자가 의미 있게 받아들이고 있는 주요 사건이나 문제점을 참여자의 관점에서 파악하고자 하는 것으로서, 특정 현상에 대한 변인을 찾고 구체화 되지 않은 현상에 대한 기본적인 이론을 발전시키고자 할 때 활용할 수 있다(Charmaz, 2013). 그러나 이러한 과정에서 자료를 개념화하여 도출시켜 범주화 시키는 과정에서 판단된 텍스트를 중심으로 여러 차례 반복되는 내용이나 의미 그리고 주제에 이름을 붙이는 개념화 과정을 진행해 나가야 하는데(고호경 외, 2013), 여기에는 많은 인력과 시간이 소요된다. 따라서 이렇게 사람에게 의해 일일이 판단하고 분류하며 진행해야 하는 작업에 빅데이터 분석을 접목함으로써 연구의 수월성을 높일 수 있을 것이다.

이러한 주장에 대한 간략한 예시를 만들어 보이기 위하여 실제 빅데이터라 할 수는 없으나 유사한 데이터를 가지고 실험한 결과를 한 가지 보이고자 한다. 본 연구자들은 우선 대도시, 중소도시, 읍면지역에 소재한 학교들이 가급적 고루 포함될 수 있도록 중·고등학교 학생 총 2036명을 대상으로 수학을 잘할 수 있는 요인(목표 달성 원인)을 조사하였다. 조사한 의견에 대한 결과를 얻기 위하여, 수학교사 집단(한 집단에 2~3명 포함)을 6개의 집단으로 편성하여 학생들의 답변을 임의 표본 추출한 후 각 반응들을 분류해 보도록 하였다. 이 분류는 ‘1) 수학에 대한 흥미와 호기심’, ‘2) 수학에 대한 자신감’, ‘3) 수학에 대한 불안’, ‘4) 수학의 유용성 인식’, ‘5) 자기 조절력³⁾’, ‘6) 창의적 사고’, ‘7) 수학 수업에 참여’, ‘8) 학습 방법적 측면(선수학습, 개념의 복잡성, 난이도, 복습 등)’, ‘9) 외적 요인 조절(컴퓨터, 휴대폰 등)’, ‘10) 기타’로 범주화하여 학생들의 반응을 분류해 본 결과는 다음

3) 자기 조절력은 ‘수학 공부를 열심히 하려 노력한다.’, ‘수학 시간에 배운 내용을 확실히 알려고 노력한다.’, ‘수학 문제를 풀 때, 답을 구할 때까지 중단하지 않고 열심히 풀려고 노력한다.’, ‘수학 공부를 잘하기 위해 계획을 세우고 스스로 노력한다.’ 등의 노력에 원인을 기술한 항목이다.

<표 3>과 같이 나왔다.

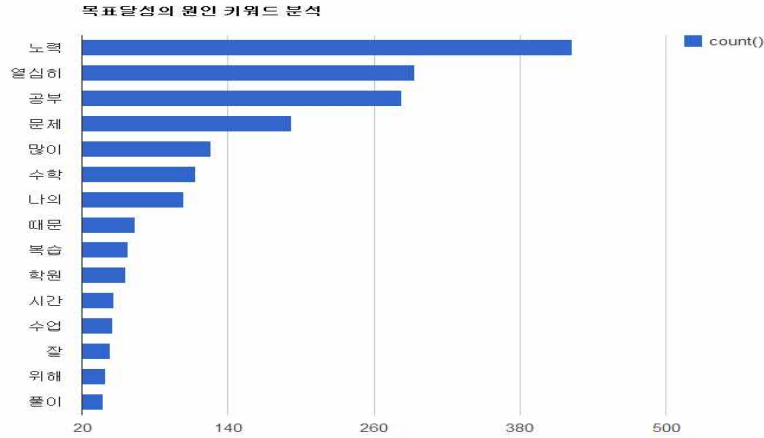
<표 3> 중·고등학교의 수학을 잘할 수 있는 요인에 대한 의견 조사 결과

| 요인 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 임의 표집수 |
|-----|-----|-----|-----|-----|------|------|-----|------|-----|------|--------|
| 집단1 | 3 | 3 | 4 | 0.5 | 56 | 2 | 2 | 15 | 11 | 0.1 | 2036 |
| 집단2 | 3 | 9 | 3 | 0.7 | 63 | 0.86 | 2 | 6 | 10 | 1 | 694 |
| 집단3 | 4 | 3 | 3 | 1 | 60 | 1 | 3 | 10 | 9 | 6 | 1992 |
| 집단4 | 4 | 3 | 2 | 1 | 59 | 2 | 3 | 11 | 11 | 4 | 1218 |
| 집단5 | 3 | 2 | 5 | 1 | 50 | 1 | 3 | 10 | 10 | 15 | 946 |
| 집단6 | 4 | 5 | 3 | 0.5 | 65 | 0.5 | 1 | 15 | 7 | 5 | 350 |
| 총합 | 21 | 25 | 20 | 4.7 | 353 | 7.36 | 14 | 67 | 58 | 31.1 | 7,236 |
| 평균 | 3.5 | 4.2 | 3.3 | 0.8 | 58.8 | 1.2 | 2.3 | 11.2 | 9.7 | 5.2 | |

이번에는 학생들의 반응을 사람이 일일이 분류하지 않고, Google Fusion Table에서 집계한 결과를 비교해 보고자 한다. 이를 위해 엑셀의 스플릿(split) 함수를 이용해서 각 문장을 단어 단위로 나눈 다음에, 나누어진 모든 단어를 하나의 컬럼(column)에 나열하여 분석한 결과는 <그림 4>와 같이 노력, 열심히, 공부 등의 키워드로 정리가 되어 노력이란 단어에 가장 중심이 실림으로써, 사람이 분류하여 정리한 결과와 맥락이 유사함⁴⁾을 확인할 수 있었다⁵⁾.

이와 같은 아주 작은 하나의 예시를 제시하기는 하였으나 이러한 예시는 빙산의 일각에 불과하고, 앞으로 수학교육 연구에서 질적 연구자는 빅데이터 분석을 적극 활용할 필요가 있다. 빅데이터 분석은 질적 연구자에게 시간과 인력 절감 측면에서 큰 도움이 될 수 있을 뿐만 아니라, 질적 연구자의 영역 지식이 빅데이터의 전 처리 과정에 결정적인 기여를 할 수 있다. 따라서 질적 연구자가 프로그램 기획에 적극 참여하여 텍스트와 그 분석방법에 대한 깊이 있는 영역 지식을 제공한다면 휴리스틱 측면을 대폭 개선시켜줄 수 있을 것이다. 빅데이터 분석은 표집이나 코딩을 최소화하고, 답론 구조를 부분이 아닌 전체로서 보여주는 큰 장점이 있다. 물론 빅데이터 분석은 자료의 특성과 연구목적에 따라 프로그램이 설계·수정되기 때문에 프로그래밍 작업이나 사전 구축 작업이 불가피하다. 그러나 질적 연구자의 영역은 비정형 데이터 분석을 통해 근거 이론을 창출하는 측면에 있다고 할 때 빅데이터 분석은 인문사회와 과학기술 간 융합연구 성격이 강하며 그 효과도 기대할 수 있다.

4) 목표달성 원인으로 가장 많이 언급된 단어는 ‘노력’, ‘열심히’, ‘공부’, ‘문제’, ‘많이’, ‘수학’, 등이 나온 것으로 문장을 유추해 보면 ‘열심히 노력하고, 수학 문제를 많이 풀어본 경우에 목표를 달성했다.’등과 같이 형성해 볼 수 있다.
 5) 본 고에서 제시한 것은 샘플 예를 만들기 위하여 시도한 것에 불과하며 보다 정확한 분석을 위해서는 학생들이 언급한 단어들을 비슷한 분류로 묶는 작업 등도 필요하고 이에 대한 분석을 위해 주관식 응답을 Matrix 형태로 바꿔서(해당 단어가 있으면 1, 없으면 0) 데이터 처리를 하거나 주관식 답변만으로 clustering 분석 등의 다양한 시도를 할 필요가 있을 것이다.



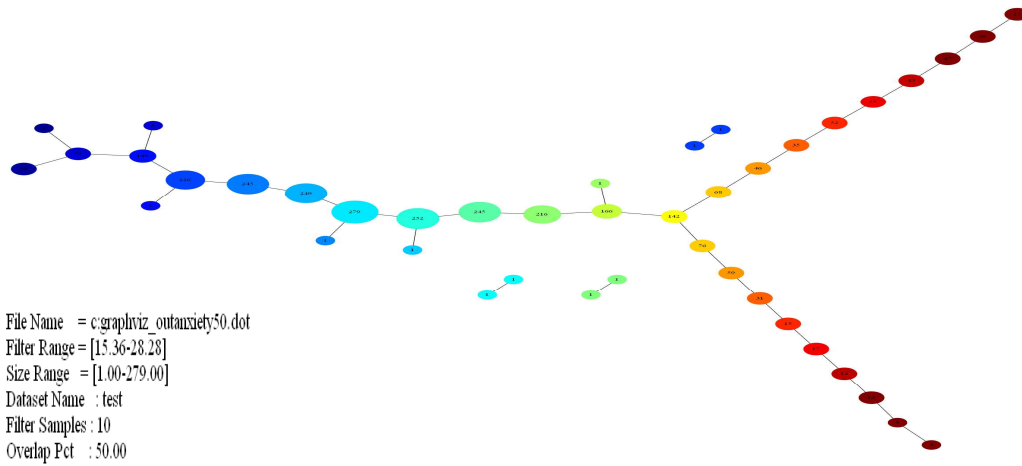
<그림 4> 수학을 잘할 수 있는 요인에 대한 Google Fusion 분석 결과

빅데이터의 오피니언 마이닝 기법 등도 질적 연구자들이 적극 활용해 볼 수 있는데, 이러한 빅데이터 분석을 통해 수학 학습에 관한 수요자나 제공자의 반응을 분석하여 선호도나 활용도 등을 추측하여 교육과정이나 정책적 제언에 활용할 수 있을 것이다.

김성수(2014)는 의료계에서 일어나는 다양한 임상과 검사 결과를 DB로 구축하고 이를 효과적으로 분석하기 위한 방안으로 빅데이터를 활용한 분석을 제안하였다. 수학교육에서도 최근 수학을 인지적인 측면뿐만 아니라 정의적인 측면까지 고려하면서 각종 검사와 상담과 치유와 같은 임상적 활동을 폭넓게 수행하고 있다. 그러나 이러한 데이터가 어딘가에 고여 있거나 묻혀버리지 않고 이를 분석하여 활용할 수 있는 방안으로서 빅데이터 사용을 제안할 수 있다.

이러한 데이터 분석을 제안하기 위한 예시로서 <그림 5>는 학생들이 보이는 수학 불안에 대한 원인을 TDA(Topological Data Analysis) 기법의 하나인 맵퍼(MAPPER, Singh, G. et al (2007)를 이용하여 분석한 것이다. 이것은 양상 또는 중요도를 나타낼 수 있는 필터함수를 사용하여 비슷한 필터값을 가지는 데이터들로 나누고, 같은 필터값을 가지지만 본질적인 특성이 다른 영역들을 군집(clustering)하여, 각각의 클러스터들의 연결 관계를 그래프로 제시한 것이다.

이러한 방법을 활용하면 서로 연관이 높은 요인들끼리의 관련성 등을 분석할 수 있게 되므로 실제로 수학불안을 치료하기 위한 요인별 처리를 할 때 활용해 볼 수 있을 것이다. 또한 이러한 기법은 수학 불안의 원인 등을 유사 항목별로 나타내는 것에 그치는 것이 아니라 학생들의 학업 상담이나 진로 지도 시 유사 성향이나 귀인 등을 묶어서 분석하고 처치에 활용해 볼 수 있을 것이다.



<그림 5> 수학불안 원인에 대한 데이터를 TDA 기법으로 그래프로 표현한 예시

VI. 나가는 말

학생의 학습 과정은 그 자체가 빅데이터가 될 수 있다. 이는 수학교육 연구에서 연구대상이 변화했고, 그에 따라 방법론과 이론도 변화할 수 있다는 점을 시사한다. 따라서 본 고에서 주장하고자 하는 바는 먼저, 수학교육 연구자는 빅데이터 자료에 대한 분석에 관심을 두고 이에 대한 연구를 다각도에서 진행해 볼 필요성이 있다는 것이다. 예를 들어, 앞에서 제시한 바와 같이 빅데이터 자료를 축약하고 분석하는 방법으로서, 학생과 교사의 신념이나 가치 등을 분석하기 위한 텍스트 마이닝을 시도해 볼 수 있을 것이다. 이는 사람들의 성향이나 지지 등을 알고자 하기 위해 저널리즘 등에서 활용될 수 있는 방법인데 이를 응용하여, 먼저 크롤링 기법을 활용한 자료를 수집하고 비정형자료를 정형화한 자료로 변환한 후, 자료 분석 및 시각화 등으로 나누어 자료를 분석하여 적절히 활용해 볼 수 있을 것이다.

그러나 빅데이터 분석이 표집이나 코딩을 최소화하고, 담론 구조를 부분이 아닌 전체로서 보여주는 큰 장점이 있기는 하나 도출된 결과를 해석하는 데 어려움을 겪을 수 있다. 따라서 연구자의 영역 지식은 비정형 빅데이터 분석의 기본적인 요건이 되며, 이런 면에서 빅데이터 분석은 인문사회와 과학기술 간 융합연구 성격이 강한 것으로 규정하기도 한다(조완섭, 2014). 빅데이터를 활용하여 자료를 모으고 분석한다 하더라도 그 최종 해석은 사람이 해야 하는데, 이러한 연구가 공학자나 데이터 과학자만이 할 수는 없는 일이기 때문이다. 따라서 수학교육 관련에서 나오는 많은 데이터들을 수학, 공학 분야에서만 할 수는 없고 수학교육 학자들과 협업하여 학제간 연구로 진행함으로써, 학문들 사이의 의사소통을 통해 학제간 협력과 공동 연구를 높여야 한다.

두 번째, 진출한 바를 통해 도출된 시사점은 수학 교수·학습 측면에서의 제언이다. 빅데이터 시대를 살아가는 우리가 빅데이터를 위한 기술개발과 연구 진행 못지않게 이를 교육적인 측면으로 이끌어 교육현장을 어떻게 적합하게 바꾸는가에 대한 문제가 남아있다. 근래 빅데이터 시대를 이끌어갈 전문가로 데이터를 수집, 정리, 조사, 분석, 가시화할 수 있는 전문가인 ‘데이터 과학자(data scientist)’에 관한 관심과 양성에 관심이 모아지고 있다(조완섭, 2014). 이러한 인재를 양성하기 위해서는 학교교육에서는 데이터를 ‘분석·추론’하는 능력이 요구되고, 따라서 데이터 과학자의 역량 및 마이닝 기법들의 기본 능력인 패턴 분석, 규칙 찾기, 분석통합, 상관관계 찾

기, 유추·추론, 통찰 등의 요소들을 교육을 통해 적극적으로 길러질 수 있도록(최지은, 2014) 수학교육의 변화 역시 요구되고 있다.

그 외에도 빅데이터 관련하여 변화를 고려해야하는 측면을 제시해보면, 빅데이터 분석은 때에 따라 프로그래밍 작업이 수반되어야 하는데, 이는 자료의 특성과 연구목적에 따라 프로그램이 설계·수정되어야 하기 때문이다. 따라서 수학교육에서의 소프트웨어 교육이나 코딩 교육에서도 빅데이터와의 연관성은 피할 수 없는 상황이다. 따라서 수학교육에서 학문의 답을 조금 허물면서 융합형태의 교육으로 진행한다거나 데이터 분석을 기반으로 과학적인 의사결정을 하는 교육 등 다각도의 발전 방향을 모색해 보아야 할 것이다.

종합적으로, 방대한 데이터 속에서 의미를 발굴하고 그것을 새로운 가치로 창출해내는 빅데이터 시대에는 창조적, 통합적인 사고와 직관력을 필요로 하며 수많은 데이터 속에서 목적에 부합하는 데이터를 찾고, 효과적인 질문과 분석 결과를 공급하는 창의적 분석가의 양성이 필요하다(한국정보화진흥원, 2013). 따라서 이러한 시대에 걸맞은 우수한 데이터 과학자와 데이터를 활용할 수 있는 인재들을 양성해 나가기 위하여서는, 이에 필요한 기본적인 소양 및 능력 함양에 목적을 둔 수학 교수·학습의 질적, 양적 변화가 동시에 이루어져야 할 것이다.

참 고 문 헌

- 고호경 · 김동원 · 이환철 · 최태영 (2013). 근거이론적 접근에 따른 수학학습 상담 발문 유형에 대한 연구. 한국학 교수학회논문집, **17(1)**, 73-92.
- 구주나 (2014). 한국어 텍스트 마이닝의 특징 고찰 및 실제 빅데이터에의 적용. 숙명여자대학교대학원 석사학위논문.
- 김정숙 (2012). 빅 데이터 활용과 관련기술 고찰, 한국콘텐츠학회지, **10(1)**, 9-116.
- 김지숙 (2013). 빅데이터 활용과 분석기법 고찰. 고려대학교대학원 석사학위논문.
- 김지현 · 김 홍 · 손광은 · 송윤수·윤지혜 · 임희창 · 정상호 · 전주홍 (2014). 빅데이터의 의학적 활용. 정보과학회지, 특집호 **3**, 18-26.
- 김성수 (2014). 의료기관의 빅데이터 활용방안에 대한 연구. Journal of Digital Convergence, **12(2)**, 397-407.
- 김윤정 · 최희석 · 한희준 · 김재수 (2013). 국가 현안 대응을 위한 빅데이터 활용 방안 : 국가과학기술지식정보서비스를 중심으로. 2013년 한국컴퓨터종합학술대회 논문집, 558-560.
- 김한수 · 김준교 (2014). 스포츠 인포그래픽스(Infographics)의 효과적인 정보 표현 방법에 관한 연구: 인터넷 신문을 중심으로. 디지털디자인학연구, **14(1)**, 252-260.
- 노규성 (2014). 사례분석을 통한 지방행정의 빅데이터 활용 전략. Journal of Digital Convergence, **12(1)**, 87-97.
- 박준호 · 북경수 · 유재수 (2014). 빅데이터 병렬 처리 기술 동향. 정보과학회지, 특집호 **1**, 18-26.
- 박춘자 (2011). 스마트 러닝환경에서의 디지털) 교과서 개발 모델. 숭실대학교 박사학위논문.
- 백기복 · 김정훈 (2013). 리더십개발 연구: 현황과 과제. 대한경영학회지, **26(7)**, 1929-1946.
- 성정숙 · 김현철 (2013). 국내외 디지털교과서 특성 및 동향 분석 : 콘텐츠 플랫폼 중심으로. 한국컴퓨터교육학회 하계 학술발표논문지, **17(2)**, 51-56.
- 손정은 (2013). 교육 이슈 빅데이터에 대한 자동 분류 처리 기술에 관한 연구. 동의대학교 대학원 석사학위논문.
- 송지숙 (2014). 스마트교육의 현황 및 빅데이터 기반 스마트교육의 활용방안. 숙명여자대학교 교육대학원 석사학위논문.
- 심규석 · 김영훈 · 이정훈 · 김진현 · 박윤재 (2014). 빅데이터 분석을 위한 맵리듀스 알고리즘의 최근 연구 동향.

정보과학회지, 특집호 1. 27-32.

- 안종욱·이미숙·신동빈 (2013). 공간빅데이터 개념 및 체계 구축방안 연구. *Journal of Korea Spatial Information Society*, 21(5), 43-51.
- 양성은 (2014). 하둡 기반의 빅데이터 영상 처리를 통한 차량 이동경로 추적 시스템의 설계 및 구현. 강원대학교 대학원 석사학위논문.
- 이상훈 (2013). SNS상에서 빅 데이터를 활용한 인포그래픽스 연구. 경북대학교 대학원 석사학위논문.
- 이성훈 (2013). 공공서비스 지원을 위한 빅데이터. 2013 한국정보기술학회 하계학술대회 논문집, 121-123.
- 임현정 (2014). 스마트 음악 교육을 위한 ICT 기반 기술 동향 조사 연구. 공주대학교 대학원 석사학위논문.
- 장영재 (2013). 교통 물류의 빅데이터 응용사례. 한국경영과학회/대한산업공학회 춘계공동학술대회 논문집, 5, 889-907.
- 전원표 (2014). 능동학습법을 이용한 개체명 사진 자동 구축. 강원대학교대학원 석사학위논문.
- 조완섭 (2014). 빅데이터 활용과 데이터 과학자. 정보과학회지, 1, 59-65.
- 조인동 (2011). 데이터 마이닝과 소셜 네트워크 분석을 통한 학술 키워드 추천 방법론. 국민대학교대학원 석사학위논문.
- 최지은 (2014). 창의적 문제해결력 증진을 위한 퍼즐활용 교육의 효과. 한국교원대학교 석사학위 논문.
- 한국정보화진흥원 (2013). 새로운 미래를 여는 빅데이터 시대. 서울: 한국정보화진흥원.
- Charmaz, K. (2013). 근거이론의 구성 [*Constructing Grounded Theory: A Practical Guided through Qualitative Analysis*] (박현선·이상균·이채원 역). 서울: 학지사 (원전은 2006에 출판).
- Johnson, L., Adams Becker, S., Cummins, M., Estrada, V., Freeman, A., and Ludgate, H. (2013). *NMC Horizon Report: 2013 Higher Education Edition*. The New Media Consortium.
- Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: technique and procedures for developing grounded theory*. New York: SAGE Publications, Inc.
- Manyika, J. et al. (2011). *Big data: The next frontier for innovation, competition and productivity*. McKinsey Global Institute.
- Zhou, X. (2013). *Network Design in the Big-Data Age*. UNIVERSITY OF CALIFORNIA doctoral Dissertation.

Study on Big Data Utilization Plans in Mathematics Education

Ko, Ho Kyoung

Ajou University
kohoh@ajou.ac.kr

Choi, Youngwoo

Ajou University
youngwoo@ajou.ac.kr

Park, Seonjeong[†]

National Institute for Mathematical Sciences
seonjeong1124@nims.re.kr

How will the field of education react to the big data craze that has recently seeped into every aspect of society? To search for ways to use big data in mathematics education, this study first examined the concept of big data and examples of its application, and then pursued directions for future research in two ways. First, changes in the representation and acceptance of data are required because of changes in technology and the environment. In other words, the learning content and methodology of data treatment need to be changed by describing a myriad amount of data visually or by 'analyzing and inferring' data to provide data efficiently and clearly. Additionally, the mathematics education field needs to foster changes in curricula to facilitate the improvement of students' learning capacity in the 21st century. Second, it is necessary to more actively collect data on general education and not merely on teaching or learning to identify new information, pursue positive changes in the teaching and learning of mathematics, and stimulate interest and research in the field so that it can be used to make policy decisions regarding mathematics education.

* ZDM classification : A5, B1

* 2000 Mathematics Classification : 9702, 97C99

* key words : big data, mathematics education

† Corresponding author