

# 도교육청의 학생 지원을 위한 빅데이터 기술 활용 방안

이현진\*, 박석천\*\*, 김종현\*\*\*

\*가천대학교 일반대학원 모바일소프트웨어학과

\*\*가천대학교 컴퓨터공학과 정교수(교신저자)

\*\*\*위세아이텍 대표이사

e-mail : asrai7@naver.com

## Utilization Method of BigData Technology for Student Support in Education Office In-Local

Hyun-Jin Lee\*, Seok-Cheon Park\*\*, Jung-Hyun Kim\*\*\*

\*Dept of Mobile Software, Gachon University

\*\*Dept of Computer Engineering, Gachon University(Corresponding Author)

\*\*\*Representative Director, WISEITECH co., ltd

### 요 약

오늘날 학교는 교육수요자 중심으로 운영되며 정규 교육과정 이외의 교육활동이 많아지고 있다. 이러한 사교육의 증가는 사회적으로 이슈화 되고 있으며 정부나 교육청 같은 공공기관에서 방과후 활동 등 새로운 정책이나 방안을 제시하고 있지만 문제가 해결되지 않고 있다. 이러한 문제를 빅데이터 기술을 활용하여 교육에 앞서 학생들의 관심, 생각, 문제 등을 정보를 추출 및 분석하여 교육의 방향을 제시함으로써 사교육의 감소를 도모하고자 한다. 본 논문에서는 이러한 문제점을 해결하고자 빅데이터를 활용함으로써 교육청에게 학생들의 정보를 통계적으로 제공하여 교육 제도의 올바른 방향으로 정책 할 수 있도록 방안을 제시한다.

### I. 서 론

오늘날 사회의 변화는 여러 환경의 변화를 수반하고 있다. 이에 교육청에서는 사교육의 증가가 사회적으로 이슈화됨에 따라 사교육비 경감을 위해 실제적으로 방과후학교를 도입한다.

방과후학교를 도입 이후 학생들의 사교육비를 경감은 초등학교와 중학교에 대해서는 효과가 있었으나, 일반계 고등학교에 대해서는 명확한 효과가 없다고 한다. 교육청에서는 사교육의 감소와 교육의 질을 높이고자, 변화하는 사회요구에 능동적으로 대처하는 공교육의 신뢰회복을 꾀한다.

하지만, 교육청의 입장에서는 공교육이 학생들에게 미치는 영향 혹은 학생들의 관심점을 파악하기가 쉽지 않다.

따라서, 학생들이 생각하는 문제점 혹은 관심점을 파악하기 위하여 빅데이터를 활용하여 교육청에서 원하는 데이터를 추출하여 핵심단어를 제시하여 그 단어에 맞게 데이터를 추출함으로써 교육 제도의 올바른 방향으로 이끌 수 있도록 한다.

본 논문에서는 수없이 많은 빅데이터의 데이터중에 학생들에게 많이 거론되는 핵심단어를 제시하여 교육청에서

원하는 정보만을 추출하여 정보의 가치를 높이고 신뢰성 있는 데이터를 교육청에서 활용할수 있도록 빅데이터 기술 활용 방안을 제안하고자 한다.

### II. 관련 연구

#### 2.1 교육청

교육의 전문성과 지방교육의 특수성을 살리기 위하여 교육법에 규정된 당해 지방자치단체의 교육·학예에 관한 사무를 관장한다. 시·도의 교육청은 교육감의 감독 아래 국가행정사무 중, 시·도에 위임된 교육·학예에 관한 업무를 진행하며, 심의·의결기관인 시·도교육위원회와 함께 지방교육자치의 핵심을 이룬다[8].

또 1개 또는 2개 이상의 군·구를 관할구역으로 하는 교육청은 시·도교육청의 하급 교육행정기관으로, 장학관으로 임명하는 교육장의 감독아래 시·도의 교육·학예에 관한 사무를 일부 분담하는데, 그 사무는 공·사립의 초등학교·중학교·기술학교·유치원 등의 운영·관리에 관한 지도·감독 등이다[8].

서울특별시 교육청의 경우 1956년 10월 서울특별시 교육위원회로 발족하여 1991년 3월 교육청으로 개칭하였으며, 지속적인 직제개편을 거쳐 1999년 1월에 1실 2국 5담당관

9과로 정비하였다. 21세기를 선도하는 자율적·창의적·도덕적인 인간교육을 지표로 서울교육 새물결 운동의 지속적인 추진과 교단지원 중심의 열린 교육행정의 구현을 기본방향으로 하고 있다[8].

## 2.2 교육청 방과후학교 참여와 사교육비

방과후학교 정책의 주요 목적이었던 사교육비 경감을 위해 실제로 방과후학교 도입이후 학생들의 사교육비가 경감했는지에 대해 학술적, 정책적으로 많은 관심이 있었다. 주요 연구 결과들을 살펴보면, 방과후학교의 사교육비 경감 효과가 있다는 연구와 없거나 증명되지 않는다는 연구로 크게 나누어 볼 수 있다[7].

우선 방과후학교 참여가 사교육비 경감에 효과가 있다는 연구들을 살펴보면 김현철 외(2007)는 방과후학교 참여후 초등학교는 36.69%, 중학교는 19.7%, 고등학교는 25.38%가 사교육이 감소하였다고 하였으며, 소득이 낮을수록 사교육 감소효과가 크다고 하였다[7].

변종임 외(2009)는 방과후학교에 지속적으로 참여하는 학생들은 방과후학교에 전혀 참여하지 않는 학생과 간헐적으로 참여하는 학생에 비하여 사교육비 경감에서 효과를 보고 있는 것으로 보고 하였다. 우명숙·이수정(2010)은 중학교 영어 사교육비 경감에 방과후학교 참여가 효과가 있는 것으로 보고하고 있다[7].

김희삼(2012)은 방과후학교참여 학생의 3개월간 사교육비 지출액은 미참여 학생보다 13만5600원(연간 54만 2400원) 적은 것으로 나타났으며 시간·비용 대비 성적 향상 효과 또한 다른 유형의 사교육에 비해 1.4~6.5배 높은 것으로 나타났다고 밝혔다[7].

반면 방과후학교 참여가 사교육비 경감에 효과가 없다는 연구들을 살펴보면, 박소영(2008)은 방과후학교가 사교육비 경감에 효과가 없는 것으로 인식되었고, 실제 경감 효과도 없었다고 결론지었다. 이종구 외 (2009)의 연구에서도 방과후학교 사업은 사교육비 지출과 음의 관계를 갖지만, 사교육비 경감에 있어 본질적 효과가 있는 것은 아니라고 분석하고 있다[7].

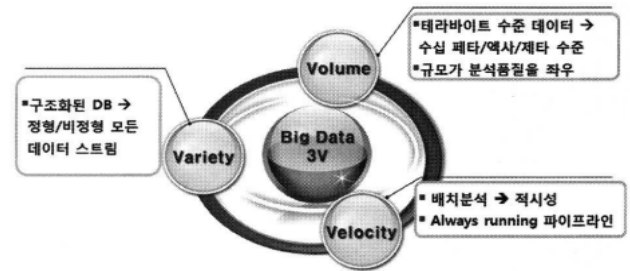
채재은 외(2009)의 연구에서는 방과후학교가 사교육비 및 학업성취도에 대해 유의한 정적인 관계를 지니고 있어 사교육비 경감에 효과적인 정책이 되지 못한다고 분석하고 있다. 한편 성낙일·홍성우(2008)는 초등학생과 중학생에 대해서는 효과가 있었으나, 일반계 고등학생에 대해서는 명확한 효과가 없다고 하였다[7].

## 2.3 빅데이터 개요 및 정의

빅데이터는 데이터의 규모가 방대하고(Volume),데이터의 종류가 다양하며(Variety), 데이터 처리 및 분석을 적시에 해결해야 하는(Velocity)특성을 가지고 있으며, 그 결과로 새로운 가치를 창출해 낼 수 있어야 한다[1].

3가지 속성을 가진 데이터가 ‘빅데이터’라는 게 대다수 전문가들의 공통된 의견이고 그림 2와 같이 가트너는 3V에 복잡성을 추가해 3V+C로 정의하기도 한다[2].

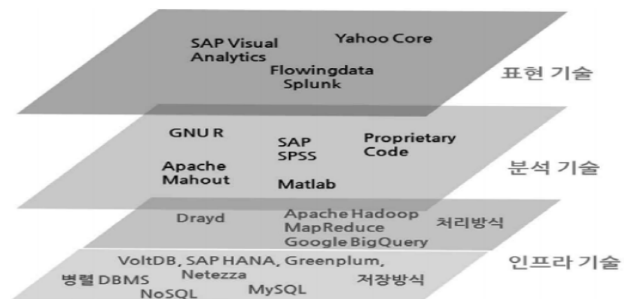
아래의 그림 1은 빅데이터 3대 특성이다. 빅데이터는 일반적인 데이터베이스로 저장, 관리, 분석 할 수 있는 한계를 넘어서며, 기업정보, 웹, 이미지/동영상,SNS, 센서 스트림 등 정형/비정형 데이터를 모두 포함하고, 분석과 예측에 있어서 실시간 처리 등 적시성을 요구한다[1].



(그림 1) 빅데이터 특성

## 2.4 빅데이터 기술 현황

빅데이터를 위한 주요 기술구성은 저장, 처리를 하는 인프라 기술과 분석 기술, 표현기술로 유형을 분류하며 그림 2와 같이 빅데이터 스택을 구성할 수 있다[3][4].



(그림 2) 빅데이터 스택 구성

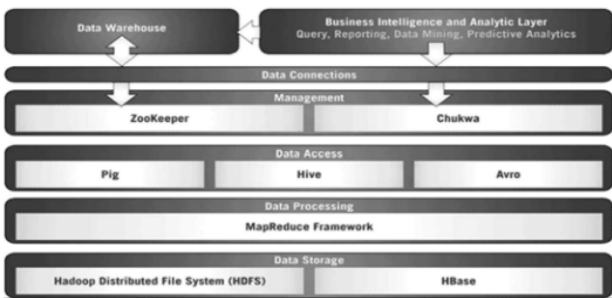
빅데이터가 가장 큰 이슈가 되고 있는 지금 빅데이터를 처리하기위한 많은 인프라들 또한 많은 관심을 받고 있다. 여러 가지들 중에 하둡(Hadoop), R, NoSQL 등이 많은 관심을 받고 있는데 특히 하둡은 많은 업체들이 개발에 나섰다, 기업들 역시 하둡 도입을 검토하고 있다[3].

하둡은 기업들이 비용과 복잡 그리고 처리문제로 인해 폐기했던 혹은 폐기 하려했던 데이터를 저장하고 처리 할 수 있게 해준다. 또한, 빅데이터라는 용어를 이끌어 낸 것이 하둡과 NoSQL의 성공에 결과라고 볼 수 있다[3][4].

하둡은 더크 커팅과 마이크 카파렐라가 구글의 GFS와 맵리듀스(MapReduce) 알고리즘을 구현한 것이다. 하둡은 빅데이터를 저장 분석하는데 사용되는 오픈소스 분산 처리 시스템이다[3].

특히 중요한 두 요소는 맵리듀스와 하둡 분산 파일 시스템(Hadoop Distributed File System, HDFS), Hbase가 있다. HDFS와 Hbase는 구글의 GFS와 BigTable의 영향을 받아 기본적으로 비용 면에서 효율적인 x86 서버로 가상화된 대형 스토리지를 구성하고, HDFS에 저장된 데이터셋을 간편하게 분산처리 할 수 있는 Java 기반의 맵리듀스 프레임워크를 제공한다[3].

하둡에는 HDFS와 맵리듀스 말고 다른 것들도 포함되어 있는데 그림3과 같이 피그(Pig), 하이브(Hive), H카탈로그(Hcatalog)등 하둡과 같이 사용 될 수 있는 여러 구성요소가 있다[3][5].



(그림 3) 하둡 시스템

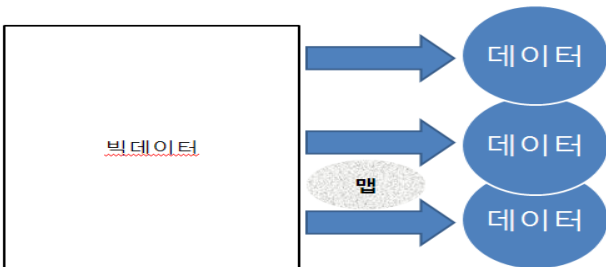
NoSQL은 전통적인 관계형 데이터베이스가 아닌 비관계형 데이터베이스로, 방대한 데이터를 유연하게 처리 할 수 있게 설계되어 있다. 하지만 NoSQL기술이 항상 채택되는 것은 아니다[3].

NoSQL기술은 네트워크 전송 중 일부 데이터 손실에도 시스템이 정상적으로 동작한다는 가용성에 중점을 두기 때문에 안정성이 중요한 시스템은 관계형 데이터베이스를 사용하는 것이 좋다[3][6].

### III. 교육청의 학생들을 위한 빅데이터 활용방안

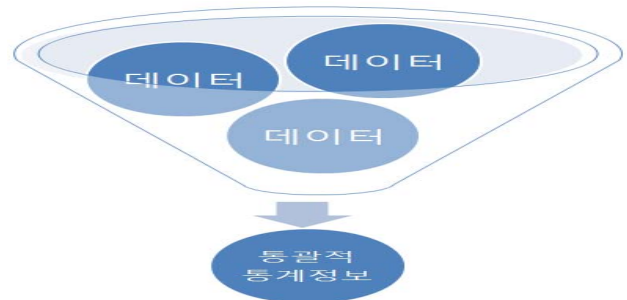
#### 3.1 빅데이터 수행절차

소셜네트워크 서비스내의 맵 과정을 통하여 흩어져있는 데이터를 매일 반복 수집하고, 맵을 이용하여 수집된 데이터를 한줄씩 나누어 각각의 단어를 추출한다. 아래의 그림 4는 소셜네트워크서비스에서 빅데이터기술을 활용하여 데이터를 분류하는 과정이다.



(그림 4) 빅데이터 분류 과정

아래의 그림 5는 그림 4를 통해서 분류된 데이터를 추출하는 과정이다.



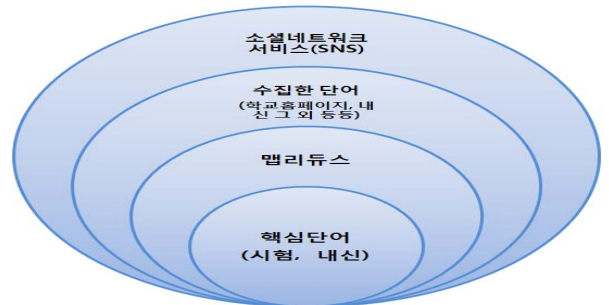
(그림5) 맵 데이터의 추출

이러한 과정들을 통하여 추출된 데이터는 통괄적인 통계 정보가 된다.

#### 3.2 학생 분석을 위한 빅데이터 프로파일링 방법

맵의 과정을 거친 문장들은 단어로 추출이 되고, 이러한 결과를 리듀스를 이용하여 데이터를 맵의 키 값에 따라 나누어진 데이터의 정렬과정을 거쳐 통괄적으로 추출된 정보를 통해서 핵심단어를 제공 할 수 있게 된다.

아래의 그림 6은 소셜네트워크 서비스에 빅데이터 기술을 적용하여 입력된 문장들을 단어별로 추출하고, 추출된 데이터를 정렬하여 단어의 개수를 카운트하여 카운트가 높은 단어순으로 핵심단어(시험, 내신, 등급)를 제공한다.



(그림 6) 소셜네트워크서비스 프로파일링

#### 3.3 교육청의 학생들을 위한 빅데이터 활용 방안

빅데이터에서 가장 중요한 것은 다양한 유형의 거대한 데이터를 수집할 수 있는 저장기술을 이용하여 분석할 수 있는 신뢰성이 높은 데이터 분석기술이라고 할 수 있다.

소셜네트워크서비스(SNS)를 통해서 정보를 수집하고, 데이터 분석을 통하여 모든 단어를 추출하고, 학생들이 많이 거론하는 핵심단어를 추출하여 교육청에게 학생들이 생각하는 문제점이나 관심사의 정보를 통계적으로 제공한다.

우리나라는 학원, 과외라는 사교육이 활성화 되어 학생들은 학교에서 배우는 교육보다는 학원에서 배우는 것을 더 신뢰하고 있다. 즉, 공교육의 신뢰를 잃고 있는 상황인 것

이다.

교육청에서는 사교육 방안으로 방과후 활동 등 여러 가지 방안을 제시하였다. 방과후 활동은 초등학교, 중학교에 효과가 있었으나 고등학생은 효과가 없었다.

따라서, 본 논문에서 제안한 빅데이터 기술 활용 방안을 이용하면 교육청은 학생들의 생각을 분석할 수 있으므로 정책에 더 효과적으로 활용할 수 있다.

의 방과후 학교참여가 사교육비 경감에 미치는 효과”, 2013  
[8] 네이버 지식백과, “교육청”

#### IV. 결 론

본 연구는 사교육활동을 감소하기 위해서 새로운 방안을 제시하기 위하여 빅데이터를 이용하여 10대들의 문제점 혹은 관심점을 분석함으로써 교육청에게 학생들의 정보를 통계적으로 제공하여 방과후에 관련된 새로운 정책을 제시하는데 도움이 되고자 한다.

최근 사교육 활성화로 인하여 공교육의 신뢰를 잃었고, 교육청에서는 사교육의 감소를 위해서 방과후 제도를 도입하였다.

그러나, 초등학생과 중학생에게는 효과가 있었으나 일반계 고등학생에 대해서는 효과가 없었다. 이로 인하여 방과후 정책은 신뢰를 잃고 있다.

따라서 본 논문에서는 학생들이 생각하는 문제점과 학생들의 관심점을 파악하고자 빅데이터 기술의 활용을 제안한다. 빅데이터를 활용하여 학생들이 생각하는 문제점, 관심점을 핵심 단어를 파악하고 교육제도의 올바른 방향 및 효과적인 정책을 할 수 있도록 제안한다.

향후, 본 논문에서 제안한 방법을 통해서 실제 시스템을 구현하여 얻어진 결과를 테스트하여 시스템을 보완 수정할 예정이다.

#### 사사의 글

본 연구는 2013년도 지식 경제부의 SW전문인력양성사업의 재원으로 정보통신산업진흥원의 고용계약형 SW석사과정 지원사업(HB301-13-1003)으로부터 지원받아 수행되었습니다.

#### 참고문헌

- [1] 안창원, 황승구 “빅 데이터 기술과 주요이슈”, 2012
- [2] 최 성, 우성구 “빅데이터 정의, 활용 및 동향”, 2012
- [3] 최대수, 김용빈 “빅데이터와 통합보안 2.0”, 2012
- [4] 최성우, 김호영, 김영국 “빅데이터 기술 및 분석 기법의 연구동향”, 2012
- [5] 강만모 외 2명, “빅 데이터의 분석과 활용”, 2012
- [6] ITWORLD, “빅 데이터의 이해”, IDG Tech Report, 2012
- [7] 심은석, 박균달, 김현진 “서울시 초·중·고등학교 학생