

커뮤니케이션 전공학생과 소셜미디어/빅데이터 교육*

김 호 동

(아주대학교 미디어학과 부교수)

소셜미디어와 빅데이터에 대한 관심이 높아지면서 이와 관련된 사회현상에 대한 연구가 커뮤니케이션 학제에서도 활발히 진행되고 있다. 반면, 이런 추세에 맞추어 소셜미디어와 빅데이터에 어떻게 접근하고 활용할 것인지에 대한 교육적 커리큘럼을 전공학생들에게 제시하지 못하고 있는 것 또한 현실이다. 본 논문은 커뮤니케이션 전공학생이 소셜미디어와 빅데이터에 대한 관심을 갖는다면 어떤 종류의 지식과 경험, 그리고 기술을 갖춰야 하는가에 대한 논의를 함으로써 역동적으로 창출되고 변화하는 미디어산업 환경에서 커뮤니케이션 전공자들의 역할을 재고하려는 목적을 갖는다. 연구자는 새로운 현상을 조직적으로 살펴볼 수 있도록 돕는 전통적인 커뮤니케이션 연구와 새로운 현상 자체에 대한 연구를 계속 진행하는 것 외에, (1) 방법론적인 측면에서 새로운 미디어 현상에 유연하게 적응할 수 있도록 하는 교육과; (2) 커뮤니케이션 전공학생의 수준에 맞는 전산학적인 지식에 대한 교육이 필요하다는 주장을 한다. 이 두 가지에 대한 보강은 커뮤니케이션 연구가 더욱더 실제 현상과 밀접하게 진행이 될 수 있도록 하여 줄뿐만 아니라, 전공학생들은 광고/PR회사 리서치 회사, 미디어관련 회사에서 현 추세에 맞는 업무를 수행할 수 있도록 도울 것이다.

주제어: 소셜미디어, 빅데이터, 커뮤니케이션 연구, 커뮤니케이션 교육

* 이 논문은 2011년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 연구되었음
(NRF-2011-327-B00951)

1. 들어가는 글

최근 소셜미디어와 빅데이터에 대한 관심이 큰 화두를 이루어 그 뜨거움이 식을 줄 모르고 있다(윤홍근, 2013; 최성 & 우성구, 2012). 커뮤니케이션 학제에서는 두 단어가 공학적인 내용을 포함하는지라 각 기술에 대한 직접적인 연구보다는, 소셜미디어와 빅데이터를 활용하는 사회적인 현상에 대한 인문, 사회학적인 연구를 활발히 진행하고 있다. 개인 정보 보호와 관련된 연구(Lee, Lim & Zo, 2012; 김상현, 2010; 최대선, 김석현, 조진만 & 진승현, 2013); 영리 비영리 목적의 기관과 기업의 활용과 그 효과에 대한 연구(김유정 & 김혜영, 2010; 김정화, 2011; 문달주, 2011; 조민숙 & 이태희, 2013; 최영근, 김선홍 & 송용택, 2012); 저널리즘적 효과에 관한 연구(임종섭, 2013; 장병희 & 남상현, 2012; 최진호 & 한동섭, 2011) 그리고 경제적인 파급 효과에 관한 연구(김원호, 2013; 이규철 & 원희선, 2012) 등 다양한 분야에서 주목할 만한 연구결과들 보여주고 있다.

그러나 소셜미디어와 빅데이터가 일부 경영학 분야를 제외한 사회과학을 하는 학생들, 특히 커뮤니케이션 전공과 어떻게 연계가 되고, 관련 직종이나 직업에 진출하기 위해서 무엇을 해야 하는가에 대한 고민은 많지 않은 현실이다. 외국의 경우를 보면, 빅데이터와 소셜미디어는 공학, 수학, 사회학이 어우러져 시너지 효과를 내는 복합적인 학문이라고 하며(Schutt, 2012), 공학계열의 학습에 사회과학, 경영, 정치경제 등을 접목하려는 시도가 보인다(Dumbill, 2012; Manovich, 2012; S. Miller, 2012). 특히 미국의 NSF(National Science Foundation)가 펜실베이니아 주립대학교의 정치외교학과에서 시도하는 빅데이터와 관련된 큰 규모의 융합교육 대학원과정을 지원하기로 결정한 것은 빅데이터와 사회과학 사이에 간극이 크지 않음을 보여주는 것이다(“Big Data Social Science”, 2013).

신문방송학과나 광고홍보학과보다 경영학과 졸업생이 마케팅 분야에서 더 두각을 나타낸다는 인식, 의료, 공학 쪽 전공의 높은 취업률에 대비되는 커뮤

니케이션 전공자의 저조한 취업률, 커뮤니케이션 전공과 같은 사회과학에서는 일반적인 전공지식을 습득할 뿐이라는 인식 등이 만재해 있다고 한다(서동균, 2013; 정성민, 2012; 교육정책네트워크, 2013). 이 글은 커뮤니케이션과 사회과학 전공지식이 세상의 시선과 구미를 끄는 빅데이터와 소셜미디어 분야와 어떤 연관이 있고, 어떻게 실질적인 지식을 제공하는 학문 분야가 될 수 있는지에 대해 알아봄으로써 커뮤니케이션 전공이 궁극적으로는 다이내믹한 사회에 일조하는 지식과 인력을 창출하는 창구가 될 수 있다는 것을 주장하려고 한다.

이 논문은 커뮤니케이션을 전공한 한 사람으로서 IT와 미디어 산업 발전과 전개에 대한 개인적인 이해를 토대로 교육적 설계에 대한 지적과 논의를 담고 있으므로, 일반 논문의 형식을 갖추기 보다는 전문적이고 체계적인 전문가들의 관심을 불러일으키고, 새롭게 전개되는 미디어, IT 산업의 경향에 커뮤니케이션 전공 학생들이 발맞춰 갈 수 있는 교육적인 설계를 호소하고자 하는 세부적인 목적을 갖는다.

연구자는 커뮤니케이션 전공을 하는 학생이 소셜미디어와 빅데이터에 관심을 가진다면 어떤 종류의 지식과 경험, 그리고 기술을 갖추어야 하는지를 정리하려고 한다. 이를 위해서 우선 연구자는 소셜미디어와 빅데이터에 대한 간략한 소개와 설명을 하고, 데이터 사이언스와 커뮤니케이션 전공 간의 역동적인 관계에 대한 언급을 한다. 이 후 관련된 커뮤니케이션 지식이 소셜미디어나 빅데이터의 분석에 어떻게 쓰이는지에 대한 몇 가지 예를 들어봄으로써 혜안을 가진 전문가들의 주목에 호소하고자 한다.

2. 소셜미디어와 빅데이터

데이터 사이언스와 커뮤니케이션 전공

소셜미디어는 웹2.0의 기술적 철학적 의미를 공유하는 것으로서, 콘텐츠를 생성, 가공, 배포하는 것과 이런 행위 중에 사용자들이 맺는 관계, 그리고 이런 활동과 관계를 원활하게 해주는 테크놀로지를 총괄하는 것이라고 하겠다

(Ahlgqvist et al., 2008). 그리고 이 서비스를 이용하면서 사용자가 축적하는 데이터가 소비자, 투표권자, 영향력자 등으로 일컬어지는 사용자들의 심리와 행위에 대한 이해에 무한한 도움이 된다는 아이디어가 정리되면서 빅데이터에 대한 관심이 급증하게 되었다.

빅데이터라는 용어는 2011년 매킨지 글로벌 연구소가 “빅데이터: 차기 혁신과 경쟁, 생산성에 관한 보고서”라는 보고서를 발표한 이래 널리 알려졌다. 보고서 내용은 당시의 데이터가 폭발적으로 증가하는 현상에 대한 지적과 이런 현상이 보다 편리하고 다양한 소비활동을 도울 것이라는 주장이었다. 또한 기존의 데이터를 다루던 프로그램과 어플리케이션의 한계에 대한 지적과 함께, 빅데이터 분석의 중요성, 그리고 이를 처리할 수 있는 데이터 사이언티스트(data scientist)의 필요성에 대한 역설을 담고 있었다(Maniyka et al., 2011).

흔히 빅데이터는 볼륨(volume), 속도(velocity), 다양(variety)의 세가지 특징으로 정의 된다. 현대의 소셜미디어를 활용하면서 축적되는 데이터가 양적으로 너무 크거나(volume), 빠르게 옮겨 다니거나(velocity), 기존 데이터베이스 구조에 맞질 않아서(variety) 다른 방법으로 처리해야 하는 것들이기에 3V로 축약되어 설명이 되는 것이다(Dumbill, 2012; Weiss, 1998).

초대용량의 규모(volume)를 가진 빅데이터를 처리하기 위해서 HDFS(Hadoop Distributed File System)나 맵리듀스(Map Reduce) 기술 등이 개발되어 이를 효과적으로 처리할 수 있는 하둡 시스템에서 사용된다. 또한 서비스나 상품의 제조사는 적시에 소비자의 의견을 수집해서 분석을 하고, 이를 제조나 상품개선에 사용하는 것이 중요할 텐데, 인터넷과 모바일을 이용하여 쏟아져 나오는 소셜미디어적인 데이터는 그 속도와 볼륨이 어마어마하여 전통적인 방법으로 정리 정돈하여 데이터를 활용하기가 어렵다. 이런 특징에 발 맞춰 데이터를 수집, 정리하고 이를 아웃풋으로 다시 활용하여 회사의 상품 생산이나 개선과 관련된 의사결정 등을 해야 하는 점에서 빅데이터의 특징으로 속도(velocity)가 거론된다. 또한 서비스의 기획이나 상품의 개발에는 데이터가 활용되기 마련인데, 인터넷과 모바일 테크놀로지를 이용하여 생성되는 콘텐츠로서의 데이터

들은 대개 계획된 테이블과 테이블 간의 관계로 저장되기 보다는 유동적(flexible)이고 원형적(raw)인 비정형 데이터 형태로 저장되어 사용된다. 예를 들면, 트위터나 페이스북, 블로그, 뉴스 등의 테크놀로지에서 통용되는 텍스트, 사진, 동영상 등의 정보에 대한 분석은 어느 한 가지로 굳어진 정형적인 데이터를 의미하는 것이 아닌 다양한 형태의 데이터를 말하는 것이다. 이렇게 유동적이며 다양한 모습을 가진 데이터(variety)를 다루기 위해서 NoSQL과 같은 기술적인 장치가 고안되어 사용된다.

이런 추세를 정리하자면, 인터넷과 모바일 테크놀로지(PC, 스마트 TV, 태블릿, 스마트폰 등등을 포함하는)로 소셜미디어를 사용하면서 생성되는 데이터에 대한 중요성이 급부상하고 있으며, 이와 함께 빅데이터와 관련된 기술과 현상을 다루기 위해서 필요한 전문가적인 인재상인 데이터 사이언티스의 양성이 중요하다는 의견이 널리 퍼져 있다고 하겠다(김우용, 2013; 송태민, 2012).

데이터 사이언티스트는 (1) 빅데이터를 저장, 활용할 수 있는 시스템을 구축할 수 있는 엔지니어링적인 기술; (2) 사용자들이 이롭게 사용할 수 있는 소셜미디어를 구축할 수 있고, 필요한 데이터를 효율적으로 추출하여 목적에 맞게 사용될 수 있도록 준비하는 프로그래밍의 지식과 기술; (3) 추출된 데이터를 수학적 특히 통계학적으로 다룰 줄 아는, 따라서 복잡한 것처럼 보이는 현상을 일목요연하게 정리할 수 있는 지식과 기술; (4) 그리고 사용자의 미디어 사용행태 분석을 통해서 사용자들과 사용자들이 속한 사회 문화적인 시스템을 이해할 수 있는 사회(심리)학적인 지식과 기술 등을 갖춘 인재로 요약될 수 있다(Chang, Kauffman & Kwon, in press; Snijders, Matzat & Reips, 2012). IBM의 밤브리(Anjul Bhambhani)가 위의 인재상이 종합적임을 강조하면서 “거의 르네상스 인간형”에 가깝다고까지 표현을 한 이유도 공학, 통계학, 사회과학 등의 다방면에 관한 지식이 중요하기 때문이다(Zikopoulos, Eaton, deRoos, Deutsch & Lapis, 2012).

그렇다면, 소셜미디어와 빅데이터로 대표되는 최근의 경향과 전통적인 커뮤니케이션(혹은 사회학관련) 전공을 공부하는 학생들과의 관계는 어떠한가? 즉

소셜미디어의 이용과 그 결과 얻어지는 빅데이터와 관련하여 커뮤니케이션 전공학생들이 무엇을 배우고 준비해야 할 것인가(Groenfeldt, 2013)?

우선 명백한 것은 커뮤니케이션 전공학생들이 엔지니어링 관련 지식을 습득할 필요는 없다는 점이다. 기술적 관점에서 접근하는 빅데이터와 소셜미디어와 관련된 분야로는 빅데이터의 플랫폼이라고 할 수 있는 하둡, 하둡의 생태계, NoSQL, 빅데이터 엔지니어링, 등과 관련된 지식이다. 반면 커뮤니케이션 학생들이 관심을 두어야 할 분야는 데이터와 관련된 분야이다. 마케팅전략광고, PR등의 수립과 실시, 새로운 정책이나 서비스의 기획과 실시 등이 커뮤니케이션 전공자들이 취업 후 광고회사나 PR회사, 리서치 회사, 그리고 일반 기업의 홍보, 커뮤니케이션 전공자로서 담당해야 할 업무라고 할 수 있는데, 이런 업무들이 “빅데이터의 분석”에 기반을 두고 이루어지게 되기 때문이다 (Roberts, Kayande, & Stremersch, in press; Ruppert, Law, & Savage, 2013).

“빅데이터 분석에 기반”이 의미하는 것은 다양한 형태의 데이터 분석 방법을 이야기하는 것인데, 이런 종류의 것으로는 텍스트마이닝(text mining), 웹마이닝(web mining), 오피니언마이닝 혹은 감성마이닝(opinion mining 혹은 sentiment analysis), 소셜네트워크분석(analysis of social networks), 클러스터분석(cluster analysis) 등이 있다(He, Zha & Li, 2013). 이런 분석을 통해서 사회적 이슈나 문제점 파악(텍스트마이닝), 웹 사이트들 간의 구조적인 분석(웹마이닝), 상품에 대한 평판, 특정 관심이나 행동 패턴에 따른 사회적 그룹(analysis of social networks) 이나 소비 혹은 구입 패턴이나 특징에 따른 군집(cluster)에 대한 분석 등이 가능하다 Burnap et al., in press).

가령 예를 들자면, 미래의 선거전략 방향은 소셜미디어인 트위터의 트윗 메시지 중 후보자의 이름이 거론되는 문장의 메시지를 분석함으로써 가늠할 수 있게 된다고 주장이 되는데, 이는 텍스트마이닝을 통해서 가능하다. 새롭게 출시한 자동차 모델 A에 대한 사람들의 평가가 전통적인 서베이 기법을 이용하는 리서치 회사로부터 나오는 것이 아니라, 광고회사의 데이터사이언스 부서에서 이루어지는 오피니언마이닝으로부터 얻어지게 된다. 모델 A 출시 한달

후에, 모델 A 단어가 들어있는 mc2day의 문장들을 수집하고 이를 감성단어를 기반으로 분석하여 소비자의 감성을 파악한다.

또한 웹사이트 간의 하이퍼링크 분석을 통해서 구조적 우위에 있는 웹사이트들을 파악하고 이를 기반으로 마케팅계획을 짜나갈 수 있게 된다(Barabasi & Frangos, 2002). 혹은 특정 상품을 구매한 기록을 바탕으로 구매자 간의 상관관계를 실시간으로 분석하여 비슷한 관심사를 가진 소셜네트워크를 추출해 내고, 이런 그룹에 특화된 마케팅이나 서비스 등을 기획하여 제공할 수도 있다. 반대로 사용자들이 선택한 노래들 간의 상관관계를 통하여 비슷한 성향의 노래들을 군집화한 후 이를 (다음 곡을 제공하는) 추천서비스에 이용할 수도 있다. 아마존닷컴의 “이 책과 함께 아래의 책들이 많이 팔렸습니다”라는 추천서비스도 이에 해당한다.

요점은 소셜미디어 사용으로 인해 빅데이터가 양산이 되면서 사회과학 전공자가 진지하게 도전해 볼만한 분야가 더욱 늘어난다는 것이다. 단, 이를 이해하는 데는 공학적인 분야가 맞물려 있어서 사회과학 혹은 인문학 전공자들이 접근하기에는 어려운 내용이 산재해 있지 않을까 하는 막연한 느낌으로 인해 효과적으로 접근하지 못하고 있는 상황이다. 위의 분석방법이나 내용에 대한 지식을 갖추게 되면 커뮤니케이션 전공자들은 리서치 회사의 여론조사나 시청률조사, 광고대행사의 통합적 마케팅 기획이나 미디어 기획, 대기업 홍보팀에서의 홍보전략 등의 분야에서 커뮤니케이션에 중점을 두는 사업에서 활약을 보일 수 있을 것이다.

그렇다면 이를 위해서 무엇을 배워야 하는가? 다음 절에서는 두 가지를 강조하는데, 첫째가 방법론적인 지식의 강화이다. 방법론은 각 전문분야마다 정도의 차이는 있지만 커뮤니케이션 전공자라면 누구나 익혀야 할 지식이다. 다만 현재의 방법론 지식이 학술적인 가설테스트와 접목되어 습득이 되면서, 유연하게 사고하고 생각하는데 사용되지 못한다는 점을 지적하려고 한다. 두 번째로는 전산학적인 지식이다. 전산학이라면 커뮤니케이션 전공자의 전공분야가 아니므로 생소하게 느껴질 수도 있겠지만, 연구자는 데이터사이언스 교육

학제를 준비하는 공학계열의 학자들 또한, 인문/사회과학(+경영)의 중요성을 인식하고 전공자로 하여금 관련 지식을 쌓도록 권하고 있는 형편이고, 무엇보다도 앞으로의 학문의 추세가 “융합”과 “통섭”인데, 이에 대해 커뮤니케이션 전공자가 소극적으로 대처할 필요는 없다고 본다. 따라서 다음 절에서는 이 두 가지에 대해서 논하도록 하겠다.

3. 무엇을 학습해야 하는가

위의 예들이 던지는 가장 중요한 점은 소셜미디어와 빅데이터 개념이 우리가 생활하는 사회 현실과 직접적으로 관련이 있고 영향을 주는 현상이라는 점이다. 전통적인 의미에서의 의제설정에 관한 연구도 이제는 종이신문이 아닌 포털 서비스의 뉴스서비스와 연관되어 진행되어야 하며 포털 서비스 사용에서 나타나는 사용자 소셜데이터의 수집과 분석이 기존의 이론적 관점을 확인, 검증, 혹은 부정하는데 중요하게 작용할 것이라는 점이다¹⁾. 다시 말하면, 전통적인 커뮤니케이션 연구를 하더라도 소셜미디어 현상과 빅데이터에 대한 지식이 없이는 과거의 이론적인 프레임에만 매달려 달라진 현상을 연구하는 오류를 범할 수 있다는 점이다. 따라서 빅데이터와 소셜미디어 현상 자체가 커뮤니케이션 혹은 미디어 관련 전공학생의 관심사가 되어야 한다.

또 한 가지 짚고 넘어가야 할 것은 빅데이터 연구에 있어서 부딪히는 어려운 점들이 사회 혹은 커뮤니케이션 이론을 가이드로 살펴보면 종종 해결되다는 점이다(Snijders et al., 2012). 따라서 소셜미디어의 사용과 빅데이터의 분석이라는 사회적 현상을 이론적인 관점에서 접근하는 전통적인 연구 또한 중요하다고 하겠다. 위의 두 가지를 전제로 하고, 빅데이터의 출현으로 중요해진 분야는 방법론적인 지식과 전산학적 지식이다. 다음 절에서 이에 대한 논의를

1) 실제로 커뮤니케이션 분야에서는 이와 관련하여 많은 연구를 이루어 왔다 (김경모, 2012; 김사승, 2006; 김유정, 2013; 안차수, 2013; 이희은, 2009; 임종섭, 2013; 장병희 & 남상현, 2012; 최진호 & 한동섭, 2011).

하겠다.

1) 방법론적 지식

우선 방법론(통계학)적인 지식이 중요하다. 커뮤니케이션 전공자로서 대개는 조사방법론과 비슷한 일련의 수업을 통해서 통계학적인 지식을 다지곤 한다. 이런 지식이 빅데이터의 분석에 직접 활용이 될 수 있어야 한다. 예를 들면 전공자들에게 상관관계분석은 아주 생소한 것은 아니다. 그러나 이 개념이 추천서비스에 어떻게 적용되느냐의 질문에는 정확한 답을 제시하지 못하는 것이 보통이다.

<표 1>은 7인의 사용자가 6개의 영화에 대해서 평가한 결과를 표로 정리한 것이다(공란은 평가가 되지 않은 상태를 의미). 이런 종류의 데이터는 흔히 영화나 음악과 같은 문화상품, 테크놀로지에 대한 리뷰, 음식이나 숙박과 같은 서비스에 대한 평가가 이루어지는 웹서비스에서 많이 나타나게 되는데, 사용자들의 영화에 대한 평가에 기초하여 각 영화에 대한 개인적인 점수와 추천이 가능하게 된다. 이에 대한 예를 보자면, <표2>는 사용자 i 가 관람하지 않은 영화 A, C, F에 대한 추정점수를 구하는 과정을 정리한 것이다. 커뮤니케이션 전공자는 사용자 i 가 본 영화의 점수와 다른 사람들의 영화에 대한 점수를 모두 고려하여 평가예측을 하려고 한다. 우선 가장 간단한 방법은 영화 A, C, F의 평가값을 모두 더하여 가장 높은 점수를 받은 영화를 추천해 주는 것이다. 그러나 i 와 (영화평가에 있어서) 상관관계가 높은 사용자들에게 가중치를 주어 계산을 하면 더 정확한 예측을 할 수 있을 것이다. 따라서 i 와 다른 사용자들과의 상관관계수(r)를 구한 후, 이를 각 사용자의 평가치와 곱한 값을 개인의 평가점수로 사용한다(각각 $r * A$, $r * C$, $r * F$). 이 점수를 모두 더한 값을 각 영화에 대한 평가점수라고 생각할 수 있는데, 이 방법의 한 가지 단점은 평가한 사람이 많은 영화가 높은 점수를 받을 수 있다는 점이다. 이를 상쇄하기 위해서 각 계산에 사용된 r 값들을 모두 더한 후, sum행의 값을 이 값으로 나누어 준다(wsum 행의 값). 결과 값을 보면, 영화 F가 가장 높은 점수를 받아서 추천

1순위가 되며 A, C가 다음 순위가 된다.

<표 1> 소셜미디어 사용자의 영화에 대한 평가 데이터. (Russell, 2011)에서 응용

	Film A	Film B	Film C	Film D	Film E	Film F
a	5	7	6	7	5	8
b	6	7	3	10	7	6
c	5	6		7		8
d		5	6	8	4	8
e	8	7	3	6	4	5
f	6	8		10	7	6
i		9		8	2	

<표 2> 상관관계(r)를 이용한 소셜미디어에서의 영화추천 서비스²⁾

	similarity (r)	Film A	r * A	Film C	r * C	Film F	r * F
a	0.991	5	4.955	6	5.946	8	7.928
b	0.381	6	2.286	3	1.143	6	2.286
d	0.592			6	3.552	8	4.736
e	0.980	8	7.840	3	2.940	5	4.900
f	0.663	6	3.978			6	3.978
sum			19.059		13.581		23.828
wsum			3.015		2.944		3.607
			6.321		4.613		6.606

대개 커뮤니케이션 전공자들은 통계의 결과값을 가설검증과 연관 짓기 때문에 이렇게 상관관계 값을 추천시스템을 만드는데 사용하는 예는 생소하기 마련이다. 그러나 간단한 상관관계 값이라도 이렇게 소셜미디어 추천시스템에 사용된다는 것을 이해하고 나면 좀 더 유연한 사고방식을 갖는 전공자가 될 수 있다. 덧붙이자면, 상관관계 값이 아닌, 유클리드 거리(Euclidean distance)나 맨하탄 거리(Manhattan distance) 등의 다른 척도를 이용한 추천시스템을 생각해 볼 수 있다.

2) <표 2>에서 사용자 i와 유사한 취향의 사용자들만을 고려하기 위해서 상관관계계수가 0 혹은 그 이하인 경우의 사용자는 제외하였다.

후에 언급하겠지만 위에서와 같이 아이템(영화)과 아이템(영화) 간의 관계를 측정할 수 있는 기법이나 모델링이 가능하게 되면 (통계학적인 방법론에 기초한 측정이나 모델링을 말한다), 전공학생은 프로그래밍 기술을 이용하여 각 아이터과 가장 유사한 아이터들을 자동으로 추출해 낼 수 있는 평선이나 객체를 개발하고 이를 통해서 사용자가 추천 결과를 직접 경험해 볼 수 있도록 하는 서비스를 만들어 제공할 수 있게 된다. 즉, (통계) 방법론적인 지식과 코딩에 대한 지식을 결합하여, 실질적인 서비스 제공의 기초를 그려볼 수 있게 되는 것이다³⁾. 설혹 커뮤니케이션 전공학생이 실제로 프로그래밍을 구현하지는 않더라도, 그 원리를 이해하고 있다면, 프로그래밍 파트를 맡은 동료와 효율적인 작업을 수행할 수 있는 자질을 갖추게 되는 이점이 있다. 사실, 프로그래밍에 능통한 전문가들이 사회학자들에게 원하는 것은 데이터가 제공해 주는 사실(facts)을 들어내는 것이 아닌, 사회학적인 통찰력(insights)과 직관력에 기초한 데이터의 분석과 이를 수학적인 모델로 완성해주는 실력이다(Mann, 2012).

위의 예는 간단한 상관관계 계수를 이용한 영화 간의 평가관계를 수치화하는 것이었지만, 좀 더 복잡한 통계방법을 이용하여 영화(혹은 분석의 대상이 되는 그 어떤 것)를 그룹화할 수도 있다. 클러스터 분석(cluster analysis)이나 MDS(multi-dimensional scaling), 사회연결망분석(social network analysis) 등이 그 예인데, 지면 관계상 각각의 방법에 대한 구체적인 설명을 모두 할 수는 없지만, 한가지 예를 더 들어서 이런 방법론이 소셜미디어와 빅데이터와 연관 있음을 설명하겠다.

<표 3>은 a, b, c . . . 의 사용자가 구입한 도서들의 데이터를 매트릭스 형태로 기록한 데이터이다(G1, 8x9 크기의 데이터). 사용자 a가 해당 열의 도서를 구입했다면 1을, 그렇지 않으면 0을 주었다. 예를 들면, a는 커뮤니케이션 이론(comTheo), PR(pr), 광고(adv), 그리고 뉴미디어(newMedia)를 구입하였다. 이 데이

3) 가령, python의 for 문을 이용하여 각 아이터에 루프를 걸어 유사한 아이터의 순서를 정하는 프로그램 예시로 들어볼 수 있기는 하지만, 지면을 할애해야 하고, 또한 기술적인 면에 치중하게 되므로 생략하였다.

터만을 가지고도 사용자서비스를 기획해 볼 수 있지만 (가령, 행의 합은 사용자의 도서 구입량을 의미하므로, 자주 구매하는 사용자를 추려서 이벤트성 토큰을 제공할 수도 있겠다), 이 데이터를 가공함으로써 소셜미디어에 어울리는 서비스를 기획해볼 수도 있다.

<표 3> G1: 온라인 서점에서 사용자가 구입한 서적에 관한 데이터 (1 = 구입; 0 = 비구입)

	comTheo	pr	adv	broadc	internet	cam Shoot	edit	newMedia	CMC
a	1	1	1	0	0	0	0	1	0
b	1	0	0	1	0	1	1	0	0
c	0	0	0	1	1	1	1	0	0
d	1	0	0	0	1	0	0	1	1
e	1	0	0	1	0	1	1	0	0
f	1	1	1	0	0	1	0	0	0
g	0	1	1	0	0	1	1	0	0
h	1	0	0	0	1	0	0	1	1

<표 4>는 <표 3>의 G1 과 G1의 치환 매트릭스인 G1' 를 곱한 매트릭스를 나타내며, <표 5>는 G1' 와 G1을 순서대로 곱한 결과인데 각각을 M1, M2로 부르기로 한다. 결과로 얻는 M1은 사용자 간의 관계를 나타내는 매트릭스이고, M2 판매된 도서 간의 관계 매트릭스이다⁴⁾. 우선 <표 4>를 보면, 대각선 값(c_{ij} , $i=j$ 일 때)이 의미하는 것은 개인의 도서구입량이다. 가령 사용자 b의 경우에는 4권의 책을 구입하였다 ($C_{2,2}=4$). 그 외의 셀 값들은 해당 행과 열의 사

4) 두 매트릭스 Y (g x h)와 W (h x k)가 있다고 하면, $Z = Y * W$ 는 $Z = [z_{ij}]$ 처럼 표시할 수 있고 Z의 각 셀값은 아래와 같이 정해진다.

$$z_{ij} = \sum_{l=1}^h y_{il}w_{lj}$$

이 때, 매트릭스 Z의 크기는 g x k 가 된다. 위의 예를 보면, 8 x 9 크기의 매트릭스 데이터 G1과 이를 치환한 G1' (9 x 8)을 곱하면 M1과 같은 (8 x 8) 크기의 매트릭스를 구할 수 있다. 반면에 그 순서를 바꾸어서 G1'와 G1을 곱하면 M2와 같은 (9 x 9) 크기의 매트릭스를 구하게 된다.

용자 간의 관계를 나타내는 데이터이다. 가령, $C_{2,5}$ (사용자 b와 e 에 해당하는 셀의 값)의 가치는 4 인데 이것이 의미하는 것은 b와 e 가 구입한 동일한 서적의 숫자가 4임을 의미한다. 원래 데이터를 살펴보면, b는 커뮤니케이션이론(comTheo), 방송학(broadc), 카메라촬영(camShoot), 편집(edit)의 책을 구입하였고, e 또한 동일한 4권의 도서를 구입하였다. 따라서, b와 e는 (적어도 서적에 관한 관심사만큼은) 관심사가 유사한 사용자라는 것을 알 수 있으며, e가 향후 구입하는 도서가 b에게도 중요할 수 있다는 판단을 할 수 있다(혹은 그 반대도 마찬가지이다). 나아가서는 (프라이버시 문제가 해결된다면) 동일한 관심사를 가진 사용자들을 묶어서 특별한 이벤트를 제공해 줄 수도 있을 것이다.

<표 4> M1: 사용자x사용자 관계 데이터 (G1 x G1'의 결과)

	a	b	c	d	e	f	g	h
a	4	1	0	2	1	3	2	2
b	1	4	3	1	4	2	2	1
c	0	3	4	1	3	1	2	1
d	2	1	1	4	1	1	0	4
e	1	4	3	1	4	2	2	1
f	3	2	1	1	2	4	3	1
g	2	2	2	0	2	3	4	0
h	2	1	1	4	1	1	0	4

M2 또한 M1의 경우와 같이 해석할 수 있는데 판매된 도서 간의 관계를 나타내준다(<표 5> 참조). M2의 대각선 값은 해당 도서의 판매량을 의미하며, 대각선에 해당하지 않는 셀의 값들은 책들 간의 관계를 보여준다. 예를 들면, 카메라촬영(camShot)책은 총 5권이 팔렸으며, 이 중 4권이 편집(edit)책과 함께 판매되었다는 것을 알 수 있다. 따라서 서비스 기획자는 웹사이트를 방문한 무명의 사용자가 카메라촬영 책을 살펴보고 있을 때, 편집 책이 이와 동시에 가장 많이 팔리는 책이라는 추천 서비스를 제공할 수 있게 된다. 더 나아가서는 두 책을 묶어서 판매를 하는 전략을 세울 수도 있겠다.

〈표 5〉 M2: 도서x 도서 관계 데이터 (G1' x G1의 결과)

	comTheo	pr	adv	broadc	internet	camShoot	edit	newMedia	CMC
comTheo	6	2	2	2	2	3	2	3	2
pr	2	3	3	0	0	2	1	1	0
adv	2	3	3	0	0	2	1	1	0
broadc	2	0	0	3	1	3	3	0	0
internet	2	0	0	1	3	1	1	2	2
camShoot	3	2	2	3	1	5	4	0	0
edit	2	1	1	3	1	4	4	0	0
newMedia	3	1	1	0	2	0	0	3	2
CMC	2	0	0	0	2	0	0	2	2

위의 예는 인터넷 사용자의 도서구입 내역 데이터를 이용하여 사용자 간의 관계와 판매된 도서 간의 관계를 파악하고 이를 바탕으로 합리적인 서비스를 기획하는 예를 들었던 것인데, 사회연결망분석(social network analysis)이 이런 류의 분석의 기초가 되는 통계적 방법론이다. 비교적 간단한 예를 들었지만, 사회연결망 분석이 제공하는 여러 가지 중심성(centrality) 측정이나 군집분석방법(N-clique, K-plex clique, MDS, CONCOR(Convergence of CORelations) 등등)을 이용할 수도 있다. 전통적으로 통계방법론은 커뮤니케이션 전공자들이 접하는 분야의 하나다. 여기에 더하여 이런 방법론들이 소셜미디어나 빅데이터와 관계 있음을 이해하고 이를 활용할 수 있는 훈련이 중요해지고 있다.

이와 더불어 아래에서 언급할 전산학적인 지식(프로그래밍 기술)이 더해지면 위에서 언급하는 사회학적인 데이터의 분석이 기술, 기계적으로 활용 되어 사용자들이 유용하게 사용할 수 있는 서비스로 발전될 수 있게 된다. 사실, 커뮤니케이션 전공자가 이런 코딩작업을 전산학을 전공한 인력과 협력하여 할 경우가 더 많겠지만, 이런 경우에도 프로그래밍에 대한 이해가 있는 전공자라면, 창의적이고 적극적인 서비스를 기획하고 전산전공 인력과도 효율적으로 커뮤니케이션을 하는 인력이 될 수 있다.

2) 전산학적 지식

전산학적(computer programming) 지식은 아마도 가장 논란을 부르는 분야일 것이다. 프로그래밍에 대한 지식을 얻고 싶으면, 당연히 전산학을 전공해야 한다고 주장할 수도 있겠지만, 소셜미디어와 빅데이터의 시대에서는 이런 주장이 꼭 합당한 것만은 아니다. 네이버의 한종호 이사가 언급하는 것처럼 “앞으로는 어떤 전공을 막론하고 제2 외국어는 영어, 제1 외국어는 프로그래밍이 됩니다”라는 의견이 설득력을 가지고 회자되는 편이다(한종호, 2013). 실제로 요사이의 대학전산 입문과정 수업은 대개 타 전공학생을 위한 반이 따로 개설되고 있는 형편이다.

그렇다면 사회과학을 하는 전공자는 어느 정도의 전산학적 지식이 필요한가? 가장 최소한의 지식은 “커뮤니케이션을 위한 지식”이라고 하겠다. 즉, 리서치회사나, PR회사, 광고회사, 기업의 마케팅관련 부서, 콘텐츠 서비스 기업 등등 커뮤니케이션이나 문화산업과 관련된 분야에서 새로운 상품이나 서비스를 기획할 때에 소셜미디어와 빅데이터가 어떤 것이며 어떤 원리로 움직이는가를 이해하고, 전산학 전문인력과 같이 일할 때 문제없이 이들과 커뮤니케이션을 하고, 창의적인 서비스 기획을 주도하는 사람이 필요한 것이다. 따라서 프로그래밍 지식이 중요하다는 의견의 단초는 커뮤니케이션이라고 하겠다.

이는 생각보다 중요하다. 최근의 빅데이터에 대한 늘어나는 관심으로 지적되는 낭비 중의 하나는 어디에 어떻게 쓸지 모르는 상태로 막대한 비용을 들여서 빅데이터를 수집할 수 있는 장치와 기술을 먼저 들이려고 하는 행위이다. 이는 기술에 대한 이해가 모자라는 상태에서 기술이 무엇이든지 해결해 줄 것이라는 기술유토피아적인 태도에서 나타나는 현상이다 (Dumbill, 2012; 함유근 & 채승병, 2012). 좀 더 현명한 인력이라면 전산학적인 지식과 사회학적인 이해를 무기로 어떤 종류의 데이터를 생산해 내는 사회적 미디어가 필요하다는 것을 창의적으로 파악하고, 이를 줄기로 하여 세부적인 기술을 구현해 가도록 지도할 것이다. 이는 “전산학을 아는 사회과학도 혹은 인문학도”가 두각을 나타낼 수 있는 분야인 것이다 (최광선, 2012; 최성 & 우성구, 2012).

그렇다면 “커뮤니케이션을 위해서 필요한 최소한”의 지식은 어떤 것인가? 이에 대한 의견은 분분하겠지만, 가장 기초가 되는 것으로는 우선 웹 관련 지식과 프로그래밍 관련 지식이 있겠다. 전자는 인터넷의 웹 페이지가 세워지고 돌아가는 원리에 대한 것으로 이런 류의 지식은 웹에서 생성되고 발전되며 소멸되기도 하는 많은 종류의 서비스들을 이해하는 데 도움이 될 뿐 더러 이런 서비스를 기획하고 실현하는 데에도 중요한 역할을 한다.

HTML, CSS3 등에 대한 지식은 월드와이드웹 페이지가 어떻게 만들어지는가, 웹 페이지의 내용과 형식(디자인)이 왜, 어떻게 구분되는가, 웹 생태계란 결국 무엇을 의미하며 이런 환경에서 살아남기 위한 웹 서비스의 특징이 무엇인가 등에 대한 고민과 고찰이 가능하도록 돕는다. PHP, JavaScript와 MySQL에 대한 기초적인 지식은 월드와이드웹에서의 상호작용이 어떻게 구현되는가를 이해할 수 있도록 해준다. 또한 사회적 상호작용의 결과로 생성되는 데이터들이 어떤 형식으로 어떻게 저장되는가를 파악할 수 있도록 해 줌으로써 새로운 상품기획이나 서비스의 개발에 어떻게 적용될 것인가에 대한 전력적인 구상을 짤 수 있도록 해준다.

또한 최근의 HTML5의 등장은 멀티미디어 콘텐츠가 월드와이드웹 서비스에 어떻게 접목되는지 등에 대한 안목을 가질 수 있도록 한다. HTML이 진화한 HTML5는 스마트 기기의 서비스 제공에 크게 활용되고 있으므로 HTML5의 이해는 스마트 폰이나 스마트 TV의 (흔히 앱이라고 부르는) 서비스의 기획과 구현에 큰 도움이 되며, 이렇게 해서 얻을 수 있는 빅데이터는 다시 소비자나 사용자를 이해하는 기반으로 활용한다.

서비스기획이나 개발과 같은 실무적인 측면에서 도움이 될 뿐만 아니라, 커뮤니케이션 연구 혹은 사회과학적 안목에도 많은 도움이 된다. HTML과 CSS로 만들어지고 표현이 되는 인터넷에서 상호작용이란 무엇인가에 대한 이해를 통해서 HCI에 대한 공학적 접근의 연구가 가능해진다. 또한 상호작용에서의 보안에 관한 문제에 대한 이해를 통해서 개인정보 이슈에 대한 심도 있는 토론을 할 수 있게 되며, 근 10년 동안 논란이 되어 왔던 액티브X가 어떻게 생겨났

으며, 왜 현재까지 사용되고 있는지에 대한 이해와 다른 기술과의 충돌에 대한 문제점들을 연구의 주제로 다룰 수 있게 된다.

이와 함께 필요한 지식은 코딩(프로그래밍)과 관련된 지식이다. 전산학을 전공하는 학생이라면 기초가 되는 알고리즘, 이산수학, 자료구조 등의 기초지식을 토대로 프로그램의 원리에 대한 공부를 하게 되고 이 과정에서 C와 C++ 등의 프로그래밍 랭귀지 기술과 지식을 습득하게 되며 이를 통해서 다른 언어를 접하더라도 지습할 수 있는 능력을 기르게 된다. 커뮤니케이션 전공자들은 이런 과정을 모두 습득하는 것을 지양하고, JAVA와 같은 프로그램 랭귀지에 대한 기초적인 소개를 통해서 프로그래밍과 유저인터페이스, 상호작용 등에 대한 안목을 얻어야 한다. 이 부분에 있어서는 많은 논란의 여지가 있겠지만, 이런 종류의 지식은 결국 이전에 언급한 사회통계학적인 지식을 활용하는 작업, 데이터를 시각화하는 작업 등과 긴밀한 연관을 가지므로 무척 중요하다고 하겠다 (Whittaker, 2011).

예를 들면, 앞서 언급한 통계를 수행하는 방법으로 최근에 주목을 받는 프로그램이 R이다. R은 통계 계산과 이를 축약 표현하는 그래픽을 위한 프로그래밍 언어이자 소프트웨어 환경을 말한다. 각 통계방법이 모듈형식으로 구성되어 있고, 각 모듈은 오픈소스 방식으로 자발적 전문 참여자들에 의해서 관리가 되므로 R에서는 다양한 통계방법을 사용해 볼 수 있다. 또한 R의 강점은 데이터의 시각화를 정교하게 할 수 있다는 점이다. R의 단점은 인터페이스가 콘솔의 형태를 띠는 점과 R 자체가 프로그래밍 환경이라는 점이다. 이 점은 전문가에게 이점으로 작용할 수도 있겠지만, 통계를 처음 다루는 초보자들에게는 커다란 산이 되어 버린다. 그러므로 프로그래밍에 대한 기초적인 지식은 통계에 대한 지식과 기술을 습득하는데도 많은 도움을 준다.

또한 프로그래밍에 대한 지식은 커뮤니케이션과 사회현상 연구의 도구가 된다는 점에서 무척 유용하다. 소셜미디어의 가장 큰 특징 중의 하나는 소셜미디어의 사용에서 생성되는 데이터의 활용이다. Facebook⁵⁾, Twitter⁶⁾ ,

5) <https://developers.facebook.com/docs/reference/api/>

Youtube⁷⁾, Flickr⁸⁾, Foursquare⁹⁾, Google+¹⁰⁾ 등 소셜미디어로 유명한 서비스들은 사용과 관련된 데이터의 일부를 이용할 있도록 API(Application Programming Interface) 방법을 제공하고 있는데, 이를 이용하여 원하는 데이터나 자료를 열람하여 볼 수 있다. 프로그래밍 지식이 있는 사용자들은 API로 데이터를 얻어 이를 다시 창의적인 서비스를 제공하는데 활용하기도 한다. 커뮤니케이션 연구에 있어서 앞서 언급된 소셜미디어 사이트들의 사용자들의 사용행태나 심리, 그리고 사회적인 행위의 교환 등이 중요한 연구의 관심사인데, 이와 관련된 데이터를 구해서 사회과학의 이론에 접목시켜 살펴볼 수 있다면 커뮤니케이션 연구에 많은 도움이 될 것이다.

이와 관련된 예를 들어보면 아래는 지난 대선 후보를 리트윗 했던 트위터 아이디를 추적한 자료 중 박근혜 후보를 중심으로 정리한 데이터의 일부이다 (<표 6> 참조). 첫 번째 열은 리트윗을 당한 아이디(retweetee)를 의미하고 두 번째 것은 리트윗을 한 아이디를 말하며(retweeter), 마지막 열은 그 횟수(freq)를 말한다. 첫 번째 행을 이해하자면 at_pgh 라는 아이디는 박근혜 후보의 트윗 메시지를 21번 리트윗 하였다는 뜻이다. 연구자는 트위터의 공개된 API(Application Programming Interface)를 이용하여 retweeter로 밝혀진 아이디들을 다시 retweetee로 파악을 한 후, 같은 작업을 수행하였고, 이를 3단계까지 실시하였다. 결국, 이렇게 얻은 데이터는 대선 후보의 트위터 메시지가 리트윗이라는 메시지 전파방법을 통해서 어떻게 확산(diffusion)되는가를 나타내 주는 데이터라고 하겠다. 이는 프로그램 코딩에 대한 간단한 지식을 가진 학생이라면 소셜미디어에서 제공하는 공개적 API를 이용하여 데이터를 얻을 수 있고, 이를 사회현상을 분석하는데 활용할 수 있는 통찰력을 길러줄 수 있다는 예라고 하겠다.

6) <https://dev.twitter.com/docs/api/1.1/overview>

7) https://developers.google.com/youtube/getting_started

8) <http://www.flickr.com/services/developer/>

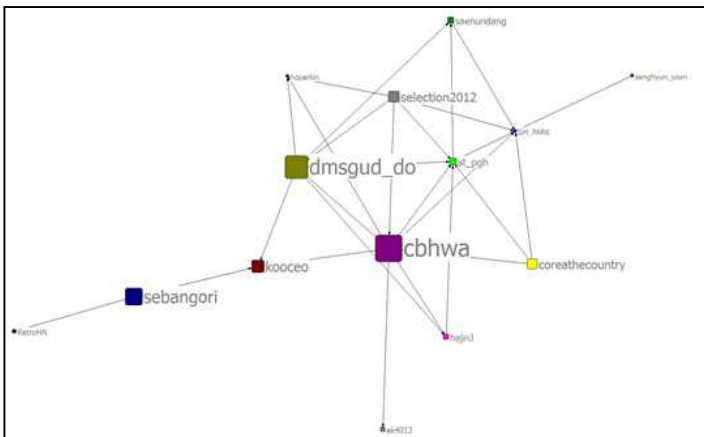
9) <https://developer.foursquare.com/overview/>

10) <https://developers.google.com/+api/>

〈표 6〉 대선 후보를 중심으로 한 리트윗 관계를 정리한 데이터

Retweetee	Retweeter	freq
GH_PARK	at_pgh	21
GH_PARK	saunakim	1
GH_PARK	saenuridang	1
at_pgh	GH_PARK	26
at_pgh	saunakim	14
at_pgh	cbhwa	11
...

연구자는 나아가 위의 데이터를 이용하여 트위터 사용자x사용자 (m x m 크기의) 매트릭스를 구하였는데, 이는 박근혜 대통령 후보를 중심으로 누가 누구를 리트윗 하는가에 대한 관계를 나타내는 것이었다. 이에 따라서 연구자는 매트릭스에서 추출한 매개중앙성(betweenness centrality)을 토대로 아래와 같은 시각화를 시도하였다. <그림 1>은 박근혜 후보의 리트윗에 참가하는 사용자들 중 가교역할을 하여 다른 사람들에게 메시지를 효과적으로 전달하는 사용자들로 dmsgud_do 와 cbhwa가 파악이 되고 있다는 것을 나타낸다.



〈그림 1〉 리트윗 네트워크에서 영향력이 높은 사용자들의 추출

〈표 7〉 리트윗 사용자들에 대한 정보

	er (팔로워 숫자)	ing (친구 숫자)	tweets (트윗 숫자)	regday (등록 경과일)	Desc (소개글)
saenuridang	29311	29295	7789	851	대한민국 새누리당 공식 트위터 계정입니다!
sebangori	3440	10	6	188	나라를 망하게 하는 것은 外侵이 아니라, 공 직자의 부정부패에 의한 民心의 離反이다.' '술마시기를 좋아하는 나라는 망하고, 차를 마 시기 좋아하는 나라는 흥한다.' (다산 정약 용, 1762~1836)
holhye	22350	44	1	715	...
...

또한 연구자는 위에서 파악된 트위터 사용자들의 아이디를 이용하여 사용자의 팔로워숫자, 친구숫자(팔로워하는 사람숫자), 트윗메시지를 만든 숫자, 트위터의 경험(트위터아이디 등록부터 연구시점까지의 날짜수) 등과 자신을 소개하는 소개글을 수집한 후 이를 함께 정리하였다 (<표 7> 참조). 연구자는 통계 프로그램인 R과 R의 통계모듈들을 이용하여 소개글에서 형태소를 제외한 단어만을 추출하고, 이를 다시 소셜네트워크 형식의 데이터로 변형한 후 (사용자 x 단어), R을 다시 이용하여 단어 구름(world cloud)를 만들 수 있다(<그림 2> 참조). 연구자는 이를 통해서 박근혜 후보 중심의 리트윗 네트워크의 소개글에서 발견되는 중요한 개념으로 대한민국, 국민, 행복, 박근혜, 사랑 등이 있다는 것과 중복, 자유, 정의로 대표되는 좌파계열에 대한 적개심이 박근혜 리트윗 네트워크 참여자들의 주된 테마였다고 주장하는 근거로 데이터를 이용한다.

사회과학을 하는 학생의 입장에서 보면, 이런 분석은 전혀 새로운 것이 아니라 학습과정에서 익혀왔던 방법론적인 지식과 사회학적인 통찰력을 활용한 것이라고 볼 수 있는데, 여기에 개방된 API를 이용할 수 있는 기술 혹은 지식이 반침이 되어 필요한 데이터를 얻어 낸 것이 기존의 사회학적인 연구와 다른 점이라고 하겠다.

창의적이고 적극적인 소셜미디어와 빅데이터 전공인을 배출하는 것이 필요한 것이다.

<표 8>은 빅데이터 시대에 발맞추어 일부 대학교와 교육기관에서 운영하는 빅데이터 관련 교육과정을 정리한 것이다. 우선 언급되어야 할 것은 전산학(Computer Science)과 경영학이 관련된 빅데이터 교육 프로그램은, 특히 미국의 경우에는 여기서 논의할 수 없을 정도로 많아지고 있는 추세라는 점이다 (Henschen, 2013; Mages, 2013). 따라서 <표 8>에서는 사회과학이나 커뮤니케이션 학제와 어느 정도 관련이 있다고 생각되는 교육 프로그램만을 나열하였다.

<표 8> 일부 대학교의 빅데이터 관련 교육 프로그램

대학교	소속	프로그램 이름
경희대학교 ¹¹⁾		소셜네트워크과학 학과
경희대학교 ¹²⁾	경영대학원	빅데이터분석전문가과정
국민대학교 ¹³⁾		데이터사이언스 학과
연세대학교 ¹⁴⁾	정보대학원	빅데이터석사과정
영남대학교 ¹⁵⁾		디지털융합비즈니스학과
충북대학교 ¹⁶⁾		비즈니스데이터융합학과
한양대학교 ¹⁷⁾		정보사회학과
아주대학교 ¹⁸⁾	미디어학과	소셜미디어전공
서울대학교 ¹⁹⁾		빅데이터센터
Columbia Univ ²⁰⁾	Journalism	Dual Degree: Journalism and Computer Science
Pen State Univ ²¹⁾		Big Data Social Science

11) <http://www.khu.ac.kr/upload/etc/131008a.pdf>

12) http://khmba.khu.ac.kr/html_2013/

13) <http://gds.kookmin.ac.kr/major/human/business>

14) http://cmsdv.yonsei.ac.kr/gsi/sub03/sub0311/sub03_11.asp

15) <http://graduate.yu.ac.kr/>

16) <http://bigdata.chungbuk.ac.kr/>

경희 대학교의 경우에는 소셜네트워크과학이라는 대학원 교육과정을 준비하고, 2014년 신입생을 모집하는 것으로 알려져 있다. 그러나 구체적인 커리큘럼 등은 확정되어 있지 않은 형편이다. 동시에 경영학대학원에서는 빅데이터 분석전문과정을 개설하여 회계, 경영 등의 학부 전공을 가진 학생들을 IT경영 전문가로 키우려고 하고 있다. 국민대학교, 연세대학교, 충북대학교, 그리고 영남대학교 또한 발 빠르게 움직여 빅데이터 관련 대학원 학제를 개설하고 신입생을 모집하고 있으며 관련 산업체와 연계하여 산업적인 협력관계도 모색하고 있는 것으로 알려져 있다.

한양대학교의 정보사회학과는 기존의 정보사회학과 빅데이터 이슈를 결합하여 커리큘럼을 강화하여 학부과정의 프로그램을 제공하고 있다. 아주대학교의 경우에는 미디어학과 내에 소셜미디어라는 전공을 두고, 학부생들로 하여금 소셜미디어 서비스를 기획, 제작해보고, 빅데이터를 다루어 볼 수 있도록 하고 있다. 한양대학교와 아주대학교의 경우에는 학부과정의 커리큘럼에 중점을 둔다는 특징이 있다.

서울대학교의 경우, 관련 학과가 존재하는 것은 아니지만, 빅데이터센터를 두고 9개의 분과를 구성하여 운영하고 있고, 관련 교수와 학자들을 센터의 연구 참여자로 끌어들여 연구의 방향을 확대, 정리해가고 있다. 예를 들면, 빅데이터센터는 기반기술, 데이터과학 및 분석기술 등의 공학적인 접근의 분과는 물론이려니와, 법/정책, 보건의료, 생명/환경, 미래산업경제, 사회복지, 방송/문화/스포츠, 그리고 인력양성 분과를 각각 두어, 관련 학제와 학과의 연구자들

17) <http://www.hanyang.ac.kr/user/structureDirect.action?structureSeq=333>

18) <http://it.ajou.ac.kr/media/major/major02.jsp>

19) <http://bigdata.snu.ac.kr/bx/>

20) <http://journalism.columbia.edu/page/276-dual-degree-journalism-computer-science/279>

21) <http://bdss.psu.edu/>

을 연구원으로 끌어들이는 한편, 관련 산업에서 나타날 수 있는 빅데이터나 소셜미디어 현상을 다각도에서 관찰, 연구할 수 있는 환경을 구축하고 있다. 서울대학교의 경우에는 특정 학제를 선정하여 집중을 하는 것보다, 위와 같이 연구센터를 두고, 조금이라도 빅데이터와 관련이 있는 학제의 연구원들이 능동적으로 참여할 수 있도록 하여, 전산학, 통계학, 경영학은 물론, 법학, 커뮤니케이션학, 의학, 스포츠학, 생물학 등등 다양한 분야에서의 연구가 가능한 환경을 구축한 것이 특징이다.

앞서 언급한 것처럼 미국의 경우에는 경영, 회계 관련 교육과정에서 이미 다양한 빅데이터관련 교육 프로그램을 운영하고 있는데(Mages, 2013), 그 중에서 <표 8>에서 소개하는 두 프로그램은 사회과학 학제가 주가 되어 운영되고 있는 프로그램이다. 우선 콜롬비아 대학의 저널리즘 학과에서는 저널리즘과 전산학을 동시에 전공하도록 하는 복수 전공 제도를 만들어 제공하고 있고, 이를 통해서 대규모의 신문, 방송사의 IT 관련 에디터나, 데이터마이닝 기술을 갖춘 전문 사회부기자(investigative reporter)를 전문적으로 양성하고, IT관련 미디어 기업 창시자, 뉴스전문매체에 특화된 웹사이트 디자이너/관리자를 배출하는 것을 목표로 제시하고 있다. 펜실베니아 주립대학교는 정치학파가 주도가 되어 NSF로부터 펀드를 제공받아 다학제 융합 대학원과 전문교육 과정(Integrative Graduate Education and Research Traineeship)을 개설하여 운영하고 있다. 이 과정을 자세히 살펴보면, 빅소셜데이터와 관련된 이슈와 접근방법, 소셜데이터분석과 관련된 이슈와 접근방법 등 두 과목을 전공필수로 제시하고 있고, 학부과정의 전공과 상관없이 데이터마이닝, 머신러닝, 비주얼분석 등 3 과목 중 하나를 선택하여 수강하도록 하고 있다. 또한 사회과학 배경을 가진 학생의 경우 네트워크학, 비전(vision)트래킹, 대규모데이터의 규칙성, 패턴인식, 정보추출과 조직, 웹애널리틱스 등과 통계학 등을 선택하여 수강하도록 하여, 전산학과 통계학 관련 지식을 점진적으로 쌓도록 하고 있으며, 전산관련 배경의 학생은 데이터모델링, 정치관련데이터와 예측, 사회연결망분석방법, 원인추정, 다차원모델링 등등의 사회과학과 관련된 모델링, 방법론적인 지식을

쌓도록 하고 있다. 마지막으로 이 프로그램은 빅데이터와 관련된 윤리와 법제에 대한 수업을 필수로 두고 있다. 위의 두 프로그램을 보면, 융합적인 접근을 위해서 전산학과 사회과학, 그리고 통계학 분야의 지식과 기술을 단기간 내에 쌓도록 디자인 되어 있으며, 앞으로도 이와 관련된 교육 프로그램이 사회학과 관련된 학제에서도 더 늘어날 것으로 보인다 (Mann, 2012).

연구자는 이 논문을 통하여 커뮤니케이션 학제에 맞는 전공인을 배출하기 위해서 어떤 종류의 공학적인(그리고 사회학적, 통계학적인) 지식이 필요한가를 언급하였지만, 구체적인 교육과정과 내용은 사회과학이라는 큰 학제에 속한 여러 전문가들이 머리를 맞대고 구상해봐야 할 것이다. 때 맞춰 국내의 한 기업에서는 소프트웨어 아카데미를 개설하고 인문 사회과학 전공자들을 교육한 후, 인재로 채용하는 프로그램을 구상, 실천하고 있는데, 이 또한 잘 쌓여진 사회과학적 소양이 현 추세에 맞는 공학적 지식과 잘 어우러지면 상당한 시너지 효과를 낼 것이라는 비전에 근거한 것이라고 하겠다(이동수 & 서한영, 2013).

연구자는 또한 새로운 IT산업과 미디어 산업에 대한 준비에 초점을 맞추어 이 글을 전개하였지만, 빅데이터, 소셜미디어와 관련하여 나타나는 정치, 경제, 사회학적인 현상에 대한 전통적인 연구에도 이런 융합적인 교육 디자인이 필요하다는 것을 언급하고 싶다. 개인정보의 보호와 관련법안에 관한 연구, 의제 설정과 제삼자효과이론, 침묵의 나선이론 등등의 전통적인 커뮤니케이션 이론 연구 또한 소셜미디어와 빅데이터에 대한 이해에 바탕을 두고 재조명할 수 있도록 해야 한다.

마지막으로 이 논문은 방법론의 적용과 이론적 모델을 설정하는데 한계점을 갖는다. 특히 이 논문은 커뮤니케이션 학제의 다 방면에 걸친 전문가들이 머리를 맞대고 무엇을 할 것인가에 대한 논의를 시작하자는 단초를 제공하려는 목적을 가질 뿐이지, 이론적으로 구체적이고 명확한 함의와 방향을 제시하려는 목적을 갖지는 않는다는 한계점을 지적하고 싶다. 다양한 경험과 지식을 가진 전문가들의 토론, 토의를 통한 결론의 도출이 한 개인이 가진 단편적인

지식과 기술에 앞선다는 것이다.

또한 이 논문에서 거론된 예시들은 연구자만의 지식에 기반을 둔 것으로서 그 한계점이 있다는 것을 언급하고자 한다. 연구자의 전공인 사회연결망 분석 방법의 예시를 주로 들었는데, 사회과학에서 주력하는 모델링 기법이나, 인지/행동 심리학적인 이론과 접근 방법 등등 또한 유용한 예로서 활용될 수 있다 (A. Miller, 2011). 구글의 라그하반(Prabhakar Raghavan)이 경제학의 가격이론(Pricing Theory), 게임이론, 거래비용이론(Transaction Cost Theory) 등이 이미 옥션이나 온라인 광고 시스템에 적극적으로 적용되어 왔다는 점을 들어 사회과학자들의 좀 더 활발한 참여가 필요하다는 점을 이야기한 것도 사회과학이 빅데이터와 밀접한 관련이 있다는 것을 말해주는 예라고 하겠다 (Mann, 2012). 커뮤니케이션 학제에서도 이런 추세에 발맞추어 급격한 IT 산업계의 변화에 능동적으로 대처하는 인력을 양성할 수 있도록 힘을 썼으면 하는 바람이다.

■ 참고문헌 ■

- 김경모(2012). 온라인 뉴스 확산과 여론 형성. 『언론과학연구』, 12(4), 35~72.
- 김시승(2006). 뉴스 블로그의 성격에 관한 분석. 『언론과학연구』, 6(2), 113~148.
- 김상현(2010). 소셜미디어와 프라이버시. 『정보과학회지』, 28(3), 41~46.
- 김우용(2013, August 23). SAS코리아, '데이터 과학자' 주제 대학생 특강. ZDNet Korea.
- 김원호(2013). 빅데이터와 인포그래픽스의 섹스. 『마케팅』, 47(5), 9~14.
- 김유정(2013). 소셜네트워크서비스 이용에 대한 비교 연구. 『언론과학연구』, 13(1), 5~32.
- 김유정 · 김혜영(2010). 소셜미디어 마케팅에서 정의와 신뢰가 온라인 구권의도에 미치는 영향. 『인터넷전자상거래연구』, 10(3), 131~155.
- 김정화(2011). 소셜미디어가 뮤지엄 경영에 끼친 영향과 활용 방안 연구. 『문화정책논총』, 25(2), 35~62.
- 문달주(2011). 소셜 미디어시대의 가치마케팅 전략. 『문화산업연구』, 11(2),

161~178.

서동균(2013). 취업에 있어서 사회과학전공자들의 빛과 어둠.

<http://blog.naver.com/PostView.nhn?blogId=skcareers&logNo=138322881>

송태민(2012). 보건복지 빅 데이터 효율적 활용방안. 『보건복지포럼』, (193), 68~76.

안차수(2013). 유명인의 정치참여와 대학생의 정치태도 및 투표행위에 관한 연구. 『언론과학연구』, 13(2), 193~226.

윤홍근(2013). 문화산업에서 빅데이터의 활용방안에 관한 연구.

『글로벌문화콘텐츠』 (10), 157~179.

이규철 · 원희선(2012). 新기술(빅데이터) 등장에 따른 경제적 파급효과 및 법(규제) 연구. 『한국통신학회지(정보와통신)』, 29(11), 48~54.

이동수 · 서한영. (2013). 삼성그룹 SCSA, 인문학도의 꿈을 채용합니다!

<http://blog.samsung.com/3390>

이희은(2009). 위키피디어 정보의 기술문화적 함의. 『언론과학연구』, 9(2), 461~497.

임종섭(2013). 소셜 미디어 특성을 다룬 신문기사와 공중파 TV뉴스의 프레임 비교분석. 『언론과학연구』, 13(3), 527~555.

장병희 · 남상현. (2012). 소셜미디어의 발전과 네트워크 저널리즘 전망.

『언론과학연구』, 12(4), 457~496.

정성민(2012). 취업률이 경쟁력, 취업 우수 대학을 주목하라: 4년제 대학 취업률 집중 분석. 『대학저널』.

<http://www.dhnews.co.kr/news/articleView.html?idxno=18531>

조민숙 · 이태희. (2013). 소셜 미디어를 통한 마케팅 커뮤니케이션 활동이 브랜드가치, 고객만족, 브랜드판단에 미치는 영향 연구.

『관광레저연구』, 25(4), 61~80.

최광선(2012). 『소셜 빅데이터 분석 서비스』.

최대선, et al. (2013). 빅데이터 개인정보 위협 분석 기술. 『정보보호학회지』, 23(3), 56~60.

- 최성 · 우성구(2012). 빅데이터 정의, 활용 및 동향. 『정보처리학회지』, 19(2), 10~19.
- 최영근 · 김선홍 · 송용택(2012). 소셜커머스 서비스 만족도 분석을 통한 마케팅 전략에 관한 연구. 『대한경영학회지』, 25(5), 2255~2269.
- 최진호 · 한동섭. (2011). 정치인 트위터와 신문 · 방송뉴스의 의제 상관성에 관한 연구. 『언론과학연구』, 11(2), 501~532.
- 교육정책네트워크(2013). [프랑스] 인문학, 사회과학 후원계획.
http://edpolicy.kedi.re.kr/EpnicGlobal/Epnic/EpnicGlobal01Viv.php?Ac_Group=1&Ac_Num0=15549
- 한중호 (2013, September 10). 프로그래밍의 중요성에 관한 대화.
- 함유근 · 채승병 외(2012). 『빅데이터 경영을 바꾸다』. 삼성경제연구소.
- Ahqvist, T., et al. (2008). *Social media roadmaps*. Helsinki: Edita Prima Oy.
- Barabasi, A.-L., & Frangos, J. (2002). *Linked: The New Science Of Networks Science Of Networks*. Basic Books.
- Big Data Social Science. (2013). <http://bdss.psu.edu/>
- Burnap, P., et al. (in press). Detecting tension in online communities with computational Twitter analysis. *Technological Forecasting and Social Change*. doi: <http://dx.doi.org/10.1016/j.techfore.2013.04.013>
- Groenfeldt, T. (2013). For Data Scientists, Math Skills Are Not Enough. Retrieved September 2
<http://www.forbes.com/sites/tomgroenfeldt/2013/08/23/for-data-scientists-math-skills-are-not-enough/>
- He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management*, 33(3), 464-472. doi: <http://dx.doi.org/10.1016/j.ijinfomgt.2013.01.001>
- Henschen, D. (2013). *Big Data Analytics Master's Degrees: 20 Top Programs*. Retrieved from

- <http://www.informationweek.com/big-data/big-data-analytics/big-data-analytics-masters-degrees-20-top-programs/d/d-id/1108042?>
- Lee, H., Lim, D., & Zo, H. (2012). 빅데이터 시대의 개인정보 과잉 인식이 사용자 저항에 미치는 영향. 『한국지능정보시스템 학회』.
<http://www.dbpia.co.kr/Article/3093624>
- Mages, E. (2013, September 16). As demand for big data analysts grows, schools rush to graduate students with necessary skills, *Washington Post*. Retrieved from http://www.washingtonpost.com/business/capitalbusiness/as-demand-for-big-data-analysts-grows-schools-rush-to-graduate-students-with-necessary-skills/2013/09/13/a1fbaf3e-1a66-11e3-82ef-a059e54c49d0_story.html
- Mann, R. (2012). Five minutes with Prabhakar Raghavan: Big data and social science at Google. Retrieved from <http://blogs.lse.ac.uk/impactofsocialsciences/2012/09/19/five-minutes-with-prabhakar-raghavan/>
- Manovich, L. (2012). Trending: The Promises and the Challenges of Big Social Data. In M. K. Gold (Ed.), *Ddata in the Digital Humanities* (pp. 460-475): Univ Of Minnesota Press.
- Manyika, J., et al. (2011). Big data: The next frontier for innovation, competition, and productivity.
- Miller, A. (2011). Five minutes with Andrew Miller MP: "It's important that people handle information in an intelligent way, and social science has a huge role in this.". Retrieved from <http://blogs.lse.ac.uk/impactofsocialsciences/2011/12/09/five-minutes-with-andrew-miller-mp-%E2%80%9Cimportant-that-people-handle-information-in-an-intelligent-way-and-social-science-has-a-huge-role-in-this%E2%80%9D/>
- Miller, S. (2012). Social Science Pushing Data Frontiers. Retrieved from <http://www.information-management.com/blogs/social-science-pushing-data-frontiers-10023673-1.html>

- Roberts, J. H., Kayande, U., & Stremersch, S. (in press). From Academic Research to Marketing Practice: Exploring the Marketing Science Value Chain. *International Journal of Research in Marketing* doi: <http://dx.doi.org/10.1016/j.ijresmar.2013.07.006>
- Ruppert, E., Law, J., & Savage, M. (2013). Reassembling Social Science Methods: The Challenge of Digital Devices. *Theory, Culture & Society*, 30(4), 22-46. doi: 10.1177/0263276413484941
- Russell, M. A. (2011). *Mining the social web Analyzing data from Facebook, Twitter, LinkedIn, and other social media sites*. O'Reilly.
- Schutt, R. (2012). *Course Description: Introduction to Data Science*, Columbia University. Retrieved from <http://columbiadatascience.com/2012/08/29/course-description/>
- Snijders, C., Matzat, U., & Reips, U.-D. (2012). Big data: Big gaps of knowledge in the field of Internet science. *International Journal of Internet Science*, 7(1), 1-5.
- Weiss, S. M. (1998). *Predictive data mining: A practical guide* Morgan Kaufmann.
- Whittaker, Z. (2011). Why every school student should be taught to code. ZDNet.
- Zikopoulos, P., et al. (2012). Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data Retrieved from <http://www-01.ibm.com/software/data/infosphere/data-scientist/>

최초 투고일 2013년 10월 09일

논문 수정일 2013년 11월 25일

게재 확정일 2013년 11월 30일

Social Media and Big Data Education for Communication Students

Kim, Hyo D.

Associate Professor, Department of Media
Ajou University

As social media and big data got intensive attention from various academic fields, the social phenomena related to these terms and ideas are actively investigated in the communication discipline. Nevertheless, the discipline has neglected to provide its students with visions of how these trends can be accessed and utilized in their major. The study aims at revamping the provisions of the roles of communication students in their career fields by providing insights on what are the proper knowledge, experiences, and skills when students got interested and involved in social media and big data. In addition to keeping up with traditional communication research works, and taking a full attention to the new phenomena, I propose that (1) revised statistical and methodological education that may help students adopt to rapidly changing social media world; and (2) newly and properly designed programing education are needed. I argue that reinforcement on these two aspects will not only make communication research more closely related to real worlds, but also help communication major graduates to perform better tasks and jobs in their career fields.

Key Words: Social Media, Big Data, Communication Research,
Communication Education