http://dx.doi.org/10.14386/SIME.2018.26.4.57

빅데이터 확산에 대한 선행 데이터 탐색 및 국내 확산 과정의 시계열 분석

최진* · 김영준**

국문초록 : 빅데이터는 2010년 이후 다양한 산업 분야에서 빠르게 확산이 진행되었다. 본 연구에서는 빅데이터가 확산되는 초기 과정에 대한 시계열 분석을 통해 빅데이터의 범용 기술 특징을 분석하였고, 각 산업의 확산 특성 차이에 대해 조사하였다. 빅데이터를 키워드로 하여 논문, 특허, 뉴스 데이터, 구글트렌드를 분석하여 선행 지수에 해당하는 데이터를 탐색하였고, 논문과 특허보다 뉴스와 구글트렌드가 2년가량 선행하는 트렌드를 보임을 확인하였다. 구글트렌드를 이용하여 국내와 미국, 일본, 중국의 국가별 도입 시기와 확산 양산을 비교하였고, 뉴스 데이터를 통해 국내의 주요한 8가지 산업 분야에 대해 확산이 진행되는 과정을 정량적 그리고 사례를 바탕으로 분석하였다.

본 연구를 통해 빅데이터처럼 산업 전반에 걸쳐 영향을 주는 범용 기술이 어떻게 초기 확산이 이루어지는지에 대한 실증적 연구 방법을 제시하였고, 빅데이터가 국내에서 각 산업별 확산 속도 차이는 어디에서 비롯되는지 파악하였다. 본 논문에서 제시한 방법은 빅데이터 이외에 다른 기술의 확산 과정에도 분석할 수 있으며, 특정 국가내의 기술 키워드 확산에 해당하므로 개발도상국에서 외국으로부터 도입된 기술을 어떻게 받아들일지 분석하는데 사용 가능하다. 그리고, 기업 측면에서는 새로운 기술을 출시하고 이를 확산하고자 할 때 어떤 경로가 효과적인지 이해할 수 있다.

주제어: 빅데이터, 기술 도입, 혁신 확산, 뉴스 데이터

* 고려대학교 기술경영전문대학원 박사과정 (heisen@korea.ac.kr)
** 고려대학교 기술경영전문대학원 교수, 교신저자 (youngjkim@korea.ac.kr)

Exploring the leading indicator and time series analysis on the diffusion of big data in Korea

Jin Choi · YoungJun Kim

Abstract : Big Data has spread rapidly in various industries since 2010. We analyzed the general characteristics of big data through time series analysis on the initial process of spreading big data and investigated the difference of diffusion characteristics in each industry. By analyzing papers, patents, news data, and Google Trend using Big Data as a keyword, we searched for data corresponding to the leading indicator, and confirmed that trends in news and Google Trend preceded the papers and patents by two years. We used Google Trend to compare the introduction period of domestic, US, Japan, and China and quantify the process of spreading the eight main industries in Korea through news data. Through this study, we present an empirical research method on how the general technology spreads in several industry sectors and we have figured out where the spreading speed difference of big data originated in each industry in Korea. The method presented here can be used to analyze the technology introduced from foreign countries in developing countries because it can be analyzed in diffusion process of other technologies besides big data and corresponds to the diffusion of technology keywords in a specific country. And, on the corporate side, this approach shows what path is effective when it comes to launching and spreading new technologies.

Key Words : Big data, Technology introduction, Diffusion of innovation, New data

I. 서론

빠른 기술 발전과 제품의 변화 속도 증가에 의해 새로운 기술을 통한 기업의 혁신은 기업의 경쟁 우위 확보와 유지를 위해 중요한 요소에 해당한다 (Pateli 2015). 따라서 새로운 기술을 언제, 어떤 경로를 통해 받아들이고 적용하느냐는 기업의 성과를 위해 중요한 문제에 해당한다 (Greve 2015). 최근 Information and communication technology (ICT)은 많은 산업 분야와 융합되어 기업의 기술 혁신을 유발하고, 국가의 경제 발전에 중요한 요인으로 이해되고 있다 (Watanabe 2015, Pradhan 2015). 특히, 데이터의 양적인 축적과 분석 기술의 질적 향상에 의해 데이터의 중요성이 새롭게 부각되었다. 이러한 결과로 2010년 이후 빅데이터는 많은 기업들이 도입하여 활용하고 있고 (김방룡 2014) 데이터의 중요성에 대한 논의의 핵심적인 역할을 하고 있다 (McAfee 2012). 빅데이터는 Doug Laney (2001)에 의해 정의된 바와 같이 양(volume), 다양성 (variety), 속도 (velocity)의 3가지 새로운 속성을 가진 데이터와 이를 활용하는 기술에 해당한다. 빅데이터는 현재 마케팅, 제조, 금융, IT등 산업 전 분야에서 활용되고 가치를 창출하고 있다. 각 산업 분야에서 기업들은 빅데이터를 활용하여 제품과 서비스를 출시하고, 관련 리포트를 발표하며, 빅데이터와 관련된 새로운 비즈니스 모델을 창출하고 있다 (Bai 2016).

본 논문에서는 빅데이터가 도입되고, 초기에 확산되는 과정을 통해 빅데이터처럼 다양한 산업 분야에서 활용되는 범용 기술이 어떠한 확산 과정을 거치는지 살펴보고자 한다. 범용 기술 (general purpose technology)은 다양한 분야에 활용되는 기술로 모든 산업 분야에 영향을 주어 각 국가의 경제 발전에 중요한 역할을 담당하고, ICT 및 나노 기술 (nanotechnology)의 예제처럼 새로운 산업 영역을 제시하기 때문에 기업의 발전을 위해 포착과 수용이 중요한 기술에 해당한다. Liao (2016), Guerrieri (2007), Gossart (2015)에 의해 ICT 기술 관점에서 데이터의 범용 특성에 대해 전반적으로 분석된 경우는 존재하나, 최근에 논의되는 빅데이터에 초점을 맞추어 빅데이터의 범용 기술 특징과 이에 대한 확산 과정의 분석은 연구가 미흡한 상태이다. Hu (2017)는 빅데이터의 범용 특성과 확산 과정을 분석하기 위해 논문 데이터를 분야별로 분석하여 학제간 (interdisciplinary) 특성을 파악하였다. 하지만, Hu (2017)가 분석에 사용한 데이터는 논문 데이터에 국한되어 다양한 산업 분야에서 사용되는 빅데이터의 범용 기술 특성을 파악하기에 미흡하였다.

본 논문에서는 국내에 도입되어 산업 전반에 걸쳐 확산되는 초기 과정에 대한 시계열

분석을 통해 빅데이터의 범용 기술 특징을 분석하였고, 각 산업의 확산 특성 차이에 대해 조사하였다. 특히, 기존 문헌에서 다루지 못한 다양한 산업 분야에 대한 빅데이터의 영향력에 대해 논의하였다. 이는 빅데이터를 주제어 (keyword)로 하여 논문 데이터, 특허 데이터, 뉴스 데이터, 구글트렌드 (Google Trends)등 다양한 데이터를 통해 이루어졌다. 각 산업 분야의 확산 특성은 확산이 시작되는 초기 시점에 주로 차이를 보이는데, 빅데이터의 초기 확산을 보여주는 각 데이터의 특성을 통해 4가지 데이터 중 어떤 데이터가 빅데이터 키워드에 대한 선행 지수 (leading indicator)에 해당하는 지를 파악하였다. 그리고 본 데이터를 바탕으로 15가지 산업을 대상으로 빅데이터 현황을 파악하였고, 주요한 8가지 산업 분야에 대해 확산이 진행되는 과정을 정량적 사례를 바탕으로 분석하였다. 이를 통해 산업 전반에 걸쳐 영향을 주는 범용 기술이 어떻게 초기 확산이 이루어지는지에 대한 실증적 연구 방법을 제시하였고, 빅데이터의 경우 국내에서 각 산업별 확산 속도 차이는 어디에서 비롯되는지 파악하였다.

II장에서는 문헌 연구로 빅데이터 관련 선행 연구와 기술 확산에 대한 정량적 분석 방법에 대해 정리하였고 III장에서는 연구 방법으로 분석에 사용한 데이터 수집 방법을 설명하였다. IV장에서는 빅데이터 키워드 확산에 대한 선행 지수 탐색과 국내의 각 산업 분야에서 빅데이터의 확산이 발생하는 과정을 정리하였다. 그리고 V장에서는 분석 결과에 대한 논의와 의의를 요약하였다.

II. 문헌 연구

2.1 빅데이터 기술의 중요성

데이터의 중요성은 통계나 데이터 마이닝 (data mining)측면에서 지속적으로 활용되고 관심을 받아왔으나, 최근에 논의되는 빅데이터는 통상적인 데이터 보다 양(volume), 다양성 (variety), 속도 (velocity)측면에서 기존과 차이를 가지는 데이터의 분석을 통해 깊은 이해(intelligence)를 얻고 비즈니스 측면의 이점으로 활용될 수 있어 새로운 관심을 받고 있다 (Chen 2012). 최근에는 정확성 (Veracity), 가변성 (Variability), 가치 (Value)를 추가하여 6V로 빅데이터를 설명한다 (Demchenko 2014). McAfee (2012)에 따르면 경영자는 빅데이터를 통해 자신의 비즈니스를 정확히 측정하고 진단함으로써 의사결정과 기업의 성과를 향상시키는 데 활용 할 수 있다고 주장하였고, 많은 기업들의 응

용 사례를 보고하였다. 특히, IBM의 Jewell (2014)은 빅데이터를 데이터 분석을 통해 전체 산업 분야를 탈바꿈 시키는 연로로 표현하였다. 현재 빅데이터는 널리 확산되어 다양한 글로벌 기업에서 본 기술을 활용한 성공 스토리는 이미 정보 통신 분야뿐만 아니라 제조, 금융, 자동차 산업 등 많은 분야에서 널리 알려져 있다 (Taylor 2014, Wang 2015, 박세환 2014, 김태경 2014, 김정경 2016) 글로벌 기업들의 관심에 의해 빅데이터 처리에 필요한 분석 기법과 분산 처리 하드웨어 및 소프트웨어 시장을 위해 IT 업체들의 경쟁도 치열하게 전개되고 있다. IBM, EMC, HP등 주요 IT업체들은 빅데이터에 강점을 가진 업체들을 인수하는 전략을 통해 역량을 강화하였고 빅데이터에 특화된 제품들을 출시하였다. 기업 영역에서 빅데이터의 가치가 입증되면서 의료, 공공, 학술 연구 분야에서도 빅데이터의 도입이 진행되었다. 미국 정부는 2012년에 빅데이터 기술을 개발하고 응용프로그램 등을 활용하는 데이터 과학자의 다음 세대를 양성하기 위해 빅데이터 연구 개발 이니셔티브 (Big Data Research and Development Initiative)를 시작하였다. 그리고 2016년에는 빅데이터 연구 개발 전략 계획 (The Federal Big Data Research and Development Strategic Plan)을 공개하였다. 국내에서는 행정자치부 주도로 정부 3.0 핵심과제의 하나로 빅데이터 분석을 집중적으로 진행하였다. 이를 통해 산불위험예보, 개인정보보호, 의약품 부작용 분석, 시내버스 노선 도입 등 국민생활과 밀접한 분야의 가시적 성과를 보고하였다. 학문적으로는 빅데이터를 전문적으로 다루는 다양한 저널들이 발간되었다. (Big Data 저널, Journal of Big data 저널, Big data research 저널, Big data & Society 저널, Big Data and Information Analytics 저널, Big Data and Cognitive Computing저널 등 빅데이터에 대한 전문 저널들이 존재)

지난 10년간 마케팅, 제조, 금융, IT분야의 기업들은 빅데이터 기술을 이용하여 새로운 제품을 출시하고, 역량을 강화하기 위해 투자를 진행하였고, 관련 인력을 육성 및 신규 채용하는 등 빅데이터와 관련된 경영활동을 진행하였다. 또한 빅데이터와 관련된 새로운 비즈니스 모델이 탄생하였고 신규 스타트업 회사들이 생성되었으며 기술 확보를 위한 M&A가 활발하게 진행되었다. 학문적으로는 빅데이터를 주제 또는 keyword로 하는 수많은 논문과 특허가 발표되었고 빅데이터에 대한 학회나 저널이 새롭게 등장하였다. 이에 따라 빅데이터는 데이터 기술을 기반하고 있지만, 종래의 데이터 기술이나 ICT 기술과는 다른 새로운 기술 영역에 해당한다고 볼 수 있다.

2.2 빅데이터 기술의 정의

2000년도 이전의 문헌들을 통해 빅데이터의 기원에 해당하는 문헌들을 확인하면 빅데이터가 일관된 정의로 사용되지 않았음을 확인할 수 있다. 1980년에 사회학자인 Tilly (1984)는 역사에 대한 통계적 방법을 표현하기 위해 빅데이터라는 단어를 사용하였다. Bourgoin (1995)은 데이터 자산 (Data assets) 을 의미하기 위해 사용하였고, Michael Cox (1997)는 빅데이터에 대해 메인 메모리 (main memory), 로컬 디스크 (local disk), 외부 디스크 (remote disk)의 용량에 부담이 되는 대용량의 데이터라고 기술적인 방식으로 표현하였다. Douglas (2015)는 Bourgoin (1995)과 Cox (1997)의 정의에 대해 '비즈니스 분석 기법 (business analytics)에 대응하는 현재의 개념보다 단순히 데이터 수집에 해당하는 정의에 가깝다'라고 평가하였다. John Mashey (1997)는 빅데이터를 '정보 통신 기술에 대해 광범위한 의미를 내포하는 용어'로 사용하였다. John Mashey (1997)는 인터뷰를 통해 '지속적으로 발전하여 경계가 확장되는 IT기술을 표현하는 가장 단순하고 짧은 어구'로 빅데이터를 사용했다고 밝혔다. Gartner의 애널리스트인 Doug Laney (2001)는 3D Data Management 문헌에서 현재 빅데이터의 속성으로 받아들여지고 있는 Volume, Velocity and Variety에 대해 발표하였다. IBM (2014)은 '빅데이터는 존재하는 데이터보다 넓은 개념으로 새로운 비즈니스 모델을 가능하게 하거나 효과적으로 만드는 새로운 데이터의 수집, 데이터를 분석하는 새로운 방법'에 해당한다고 보고하였다. 즉, 빅데이터는 단어에 대해 엄격하게 정의가 이루어진 이후에 확산된 용어가 아니라 시간의 흐름에 따라 다양한 분야의 논의를 통해 개념이 형성된 용어에 해당한다. Diebold (2012)는 빅데이터에 대해 다음과 같은 의견을 제시하였다. '빅데이터는 …(중략) 아마도 1990년대 중반 Silicon Graphics사의 점심 식사 table에서 얘기하던 중 처음 등장한 것 같다.' 그리고 Gandomi (2015)는 '빅데이터는 정의가 빠르게 진화되고 있어 혼란들은 유발하고 있다'고 표현하였다. 이러한 정의의 모호함에 대해 Lealy (2013)는 '빅데이터는 초기에 단어의 뜻과 같이 통상적인 데이터보다 많은 양의 데이터를 의미했다. 하지만, 빅데이터는 다양한 종류의 데이터와 데이터를 활용한 분석 기법을 통합하는 많은 뜻을 가진 단어로 진화하게 되었다'고 해석하였다. 시간이 지남에 따라 비즈니스 영역에서 데이터의 양과 복잡도가 증가함에 따라 빅데이터라는 용어가 활발하게 사용하게 되었다. Sheng (2017)은 '빅데이터는 엄격한 정의를 내리기 어렵지만, 1990년대에 대용량 데이터베이스를 나타내기 위해 사용되어 온 방식과 현재는 전혀 다른 단어로 사용되고 있다. 아울러, 최근의 문헌들은 대용량의 데이터를 통해 비즈니스에 필요한 가치를 창출하는

특성이 두드러지게 나타난다'는 점을 지적하고 있다. 따라서 최근에 빅데이터가 주목을 받고 있는 이유는 데이터를 이용하여 비즈니스에 활용하는 경우, 고객의 수요, 현 비즈니스의 문제점, 향후 필요한 제품이나 서비스 등에 대한 깊은 이해 (intelligence)가 가능하기 때문이다. McAfee (2012)는 빅데이터를 이용하여 기업의 경영자들은 자신의 비즈니스를 정확히 계측하고, 진단할 수 있으며, 데이터를 통해 얻은 지식을 바탕으로 효과적인 의사 결정에 사용할 수 있다고 제시하였다.

이러한 문헌들을 종합하면 빅데이터는 초기에는 데이터에 관련된 기술을 의미하였으나 더 이상 대용량의 데이터 (large data volume) 또는 데이터 셋 (data set)으로 단순히 정의할 수 없음을 알 수 있다. Shollo (2016)이 제시한 바와 같이 빅데이터는 대용량 데이터에 대한 처리 기술에 국한되지 않고, 비즈니스 인텔리전스 (business intelligence) 측면으로 범위가 확장되고, 조직의 지식 (organizational knowledge)측면이 강조되고 있다. 결국, 빅데이터에 대한 정의는 시간이 지남에 따라 개념이 진화되어 대용량의 데이터에 관련된 기술뿐만 아니라, 각 산업에서 활용되는 목적에 따라 대용량 데이터를 어떻게 활용할 것인가가 빅데이터의 정의에 포함되었다. 즉, 빅데이터에서 활용되는 통계적 데이터 처리 기술, 데이터 관리 기술 등은 1990년대 이전에도 존재하였고 지속적으로 연구되어 온 분야에 해당한다. 하지만, 빅데이터는 앞에서 기술한 바와 같이 조직의 지식과 비즈니스 인텔리전스의 의미를 담고 있는 새로운 기술에 해당한다. 그리고 빅데이터라는 키워드로 논문, 특허, 뉴스, 인터넷 문헌들에서 사용되고 있다.

본 논문에서는 빅데이터를 '비즈니스에 대한 새로운 이해를 가능하게 하거나 필요한 가치를 창출하도록 만드는 데이터의 새로운 수집, 분석, 활용 방법'으로 해석하였다. 이에 따라 빅데이터를 이해하기 위해서는 다양한 산업에서 사용되는 특성에 대한 이해가 중요하고, 본 논문은 국내의 여러 산업에서 빅데이터를 받아들이고 확산되는 과정에 초점을 맞추고 있다. 빅데이터를 키워드로 하여 이 과정에 대한 분석을 진행하였다.

2.3 빅데이터 확산에 대한 선행 연구

다양한 분야에 응용되는 빅데이터의 범용성을 바탕으로 각 분야에서 확산 속도가 어떻게 차이를 보이는 지에 대해서는 소수의 문헌에서 논의되었다 (정보통신산업진흥원 2013). 2011년에 McKinsey는 여러 산업 분야의 빅데이터 확산 차이에 대한 분석 보고서를 발표하였다 (McKinsey 2011). 상품, 서비스업, 공공 부문의 3가지 카테고리에 대해

총 21개의 산업 분야를 대상으로 빅데이터의 잠재적 활용 가치를 분석하였다. (제조, 건설, 천연자원, 컴퓨터/전자제품, 부동산/렌탈/리스, 도매, 정보, 운송 및 창고보관업, 소매, 행정/지원, 숙박/음식업, 기타 서비스, 예술/엔터테인먼트/레저, 금융, 보험, 과학 및 기술 서비스, 기업 관리, 정부, 교육 서비스, 의료와 사회적 지원, 전력) 각 산업에 대해 지표 수립 편리성 (ease of capture index), 인력 (talent), IT 강도 (IT intensity), 데이터 지향 마인드 (data driven mind set), 데이터 가용성 (data availability)의 5가지 항목으로 분석한 결과 빅데이터의 활용은 모든 분야에서 유용하지만 산업별로 더 많은 이익을 얻을 수 있는 분야가 존재한다고 보고하였다. 국제적으로 거래되는 컴퓨터와 전자제품, 정보 산업 영역은 빅데이터 활용을 통해 실질적인 이익을 얻을 수 있는 대표적 분야로 선정하였다. 금융, 보험, 정부 영역은 규제 장벽 같은 제약 조건이 극복된다면 빅데이터를 통해 강력한 이익을 얻을 수 있는 조건을 구비하고 있다고 분석하였다. 모든 산업 분야에 빅데이터 활용의 장벽이 존재하지만 일부 영역은 구조적으로 높은 장벽이 존재한다고 보고하였다. 예를 들어, 교육을 포함한 공공 영역은 데이터 중심의 마인드와 이용 가능한 데이터가 부족하고, 의료 분야는 지금까지 상대적으로 IT투자가 낮다는 빅데이터 활용의 장벽이 존재한다. 소재, 제조, 전문 서비스 등의 영역은 IT투자 정도가 높고 이용 가능한 데이터가 풍부하며 데이터 활용 마인드도 정립되어 있기 때문에 타 분야에 비해 상대적으로 제약 조건의 수준이 낮다고 분석하였다. 소셜 미디어 업계는 프라이버시 이슈가 큰 문제이나 빅데이터를 이용하여 비즈니스화 할 수 있는 상품의 개발이 용이하다는 장점을 가지고 있다. 따라서 소득, 관심분야, 여행지, 가족관계, 쇼핑 내용을 데이터로 축적할 경우 온라인, 오프라인 비즈니스를 다양하게 전개할 수 있다고 예측하였다. 그리고 일반기업에서 빅데이터 활용의 가장 큰 과제는 전문 분석 능력을 갖춘 인력의 부족으로 분석하였다.

DELL, IBM, SAP, TERADATA 등 글로벌 IT업체들의 후원에 의해 진행된 EMA의 분석 보고서는 255명의 빅데이터 비즈니스 전문가를 대상으로 진행한 설문 조사 결과를 다루고 있다 (Devlin 2012). 본 보고서는 9가지 산업 분야에 대해 빅데이터의 도입 수준을 활용 단계, 계획 단계, 검토 단계로 구분하였다. 9가지 산업 분야를 활용 단계의 비중이 높은 순서에 따라 정렬하면 미디어 & 홍보, 소매, 산업, 금융, 제조, 건강관리, 공공시설 인프라, 레저, 공공 서비스의 순이다. 특히, 미디어 & 홍보와 소매 분야는 2012년 당시 응답자의 50%이상 기업에서 활용 단계에 있었던 반면, 공공 서비스 분야는 활용 단계가 25%이하였다. 미디어 및 홍보 분야에서 빅데이터의 활용 단계가 가장 높은 기업은 Facebook, Yahoo, Google이고, 소매 분야에서는 Walmart였다. 활용 단계와 계획 단계의

비중을 합쳐 9개 산업을 순서에 따라 정렬하면 미디어 & 홍보, 소매, 레저, 금융, 산업, 공공시설 인프라, 공공 서비스, 제조, 건강관리이다. 본 보고서는 계획 단계와 활용 단계의 합을 수용의 수준으로 해석하여 Rogers의 혁신 확산 모델을 바탕으로 각 산업이 혁신의 수용에서 어떤 단계에 있는지 분석하였다. 미디어 & 홍보, 소매, 레저는 Innovator 단계, 금융, 산업 분야는 Early adopters단계, 공공시설 인프라, 공공 서비스 분야는 Early majority, 제조, 건강관리 분야는 Late majority에 해당한다고 보고하였다.

Hu (2017)는 빅데이터의 범용성을 분석하기 위해 Web of Science를 이용하여 빅데이터에 대한 논문이 어떤 분야에 속하며 시간에 따라 논문 편수가 어떻게 변하는지 분석하였다. 본 논문에서는 Computer science와 Engineering분야가 빅데이터 연구에서 가장 높은 비중을 차지하고, 네트워크 분석을 통해 Computer science, Engineering, Social Sciences Business이 빅데이터 발전에 중요한 역할을 해왔다고 보고하였다.

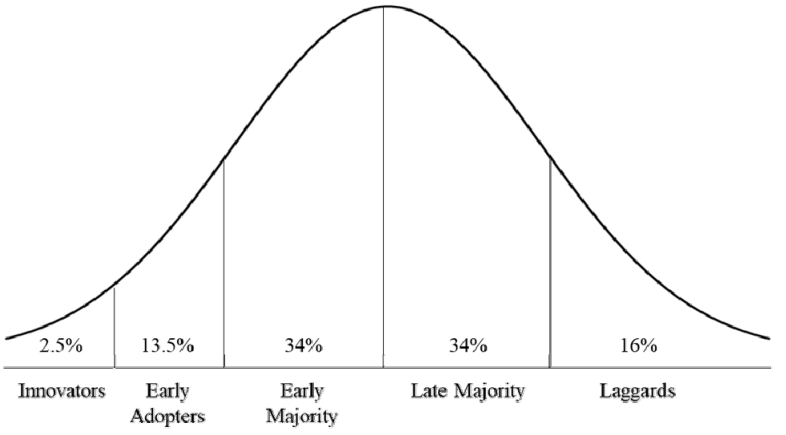
국내의 빅데이터 확산에 대한 연구는 개별 분야의 확산이 아니라 전체 산업을 대상으로 빅데이터 수용에 영향을 미치는 요인에 집중되었다 (박종현 2015, 김정선 2014, 김은영 2013, 이영주 2017). 국내 IT 중소기업을 대상으로 빅데이터 수용에 영향을 미치는 요인에 대한 회귀분석을 통해 기술적 가치, 기술적 준비, 비용수용, 조직적 지원, 기업 혁신성향, 경쟁압력, 외부 지원을 핵심 영향요인으로 선정하였다 (박종현 2015). 국내 빅데이터 실사용자 또는 잠재적 사용자를 대상으로 실증 분석하여 빅데이터의 인지된 유용성, 업무기술적합성, 개인 정보 침해 위험이 이용 의도에 유의미한 영향을 주고, 사용 용이성은 유의성이 없는 것으로 보고하였다 (이영주 2017). 그리고 빅데이터 기술의 확산 초기 단계에 조직들이 고려하는 요인은 조직의 혁신성 같은 기업의 성격보다는 신기술에 대한 불안함을 덜어줄 수 있는 조직의 여유자원과 정보시스템 인프라적 성숙도, 신기술의 인지된 혜택임을 보고하였다 (김은영 2013) 본 결과는 Rogers의 혁신확산이론에서 Innovators에서 Early adopters로 성장하기 위해 수용자 또는 수용조직이 초기 가격과 불확실성에 대한 부담을 기꺼이 수용할 수 있어야 한다는 내용과 일치한다고 보고하였다. (김은영 2013, Rogers 1993)

의료, 금융, 제조 등 특정 분야에서 빅데이터 기술의 확산과 수용에 대한 분석은 많은 시도가 있었으나 (Groves 2013, Fang 2016, Lee 2013) 동일한 기준을 바탕으로 빅데이터의 확산이 각 산업에서 어떻게 다른 지에 대해서는 연구가 부족한 현실이다. 앞서 정리한 McKinsey와 EMA의 선행 연구 역시 빅데이터에 대한 관심이 높아지던 시기인 2011년과 2012년 미국의 글로벌 기업을 대상으로 분석이 진행된 한계점을 가진다. 그리고 외국에서 도입된 빅데이터 기술이 국내의 산업별로 어떤 차이가 존재하는 지에 대한 국내

의 자체적인 연구도 부족하다. 본 연구는 빅데이터 기술의 확산이 수년간 진행된 현 시점에서 2000년부터 2016년의 데이터를 재분석하여 산업별로 빅데이터의 확산이 시작된 시점을 분석하고 이를 Rogers의 확산 모델을 통해 해석해보고자 한다.

2.4 Rogers의 기술 확산 모델

기술 확산은 Fourt (1960), Rogers (1962), Bass (1969)의 개척 이래로 1960년대부터 연구되어 다양한 분석과 결과가 존재한다. 기술의 확산에 대한 연구 방법에는 정량적 분석과 정성적 분석으로 구분이 가능하다. 이 중 정량적 분석은 모델을 통해 시간에 따라 제품의 사용자 수 또는 단위 시간당 수용 비율을 이용하고 있다. 이를 이용하면 확산의 패턴을 통해 확산의 단계를 정의하고, 특정 시점의 수용 수준, 수용자의 최종 수, 특정 시간의 수용 비율, 최대 수용 시점, 최대 수용 시점의 시기 등에 대해 분석이 가능하여 많은 관심을 받아왔다. 예를 들어, 혁신에 대한 정량적 모델링을 통해 Greve (2009)는 특정 컨테이너 (postPanamax container)의 확산을 설명하였고, Meade (2006)는 전화의 확산에 대해 설명하였다.



<그림 1> Rogers의 혁신 수용 주기 곡선

Rogers는 그림1과 같이 혁신을 수용하고 채택하는데 걸리는 시간과 혁신 성향에 따라 혁신 수용자를 5개 군으로 분류하였다. (김영준, 2017)

. 혁신 수용자 (innovators): 새로운 기술이 반영된 제품이나 서비스를 채택하는 것을 즐기는 사람에 해당한다. 이들은 신기술에 열광하고 신제품의 복잡한 사항을 익히기 마다하지 않으며 불확실성에 의해 나타나는 위험을 감수한다. 다른 사람의 구매 결정이 혁신 수용자에게 영향을 미치지 않는 가장 먼저 혁신을 수용하는 집단이다. 제품마다 다르지만 전체 수용자의 약 2.5%에 해당한다.

. 조기 수용자(Early Adopters): 이 집단은 남들보다 혁신을 앞서 채택하려는 경향이 있다. 다른 수용자의 구매 의사결정에 상당한 영향력을 미치며 초기 시장에서 대중화를 이끄는 소비집단이다. 오피니언 리더의 역할을 수행하기에 중요한 그룹이며 전체 수용자의 약 13.5%정도를 차지한다.

. 조기 다수 수용자 (Early Majority): 주류 시장으로 이동하면서 형성되는 이 부류는 혁신적인 기술이나 제품이 아닌 실제적인 문제에 신경을 쓰는 실용주의자이다. 이 집단은 혁신을 수용하는데 있어 안전을 추구하며 신중하다. 그리고 혁신 수용으로 인한 리스크를 줄이기 위해 Innovators와 Early Adopters에 의해 검증된 혁신 제품, 서비스, 프로세스를 선호한다.

. 후기 다수 수용자 (Later Majority): 위험 회피적이고 혁신에 대해 의심하는 경향이 있으며, 초기 대다수 소비집단에 영향을 받아 혁신을 채택하는 집단에 해당한다. 경쟁자에 뒤처지지 않고 따라잡기 위해 혁신을 수용하고, 가격에 민감한 특성을 지닌다.

. 지각 수용자 (Laggards): 전통적인 가치를 중시하고 현재의 상황을 유지하기를 바라는 집단이다. 혁신 수용에 부정적이고 사회에서 대중화 된 후에 마지못해 혁신을 수용하는 집단이다.

4.3장에서는 빅데이터의 국내 산업별 확산 특징을 파악하기 위해 연도별로 각 산업의 수용 단계를 분석하였다. Rogers의 기술 확산 모델을 바탕으로 국내 각 산업의 연도별 확산 단계를 정의하고, 이를 바탕으로 산업별 특징을 분석하였다.

2.5 범용 기술의 확산 요인

범용 기술은 다양한 분야의 활용되는 기술로 기존의 경제적, 사회적 구조에 근본적인 변화를 가져올 가능성이 높은 기술을 가리킨다. 기술의 활용 범위에 따른 분류 방식을 통해 범용 기술은 특정 기술에 반대되는 기술에 해당한다. (신인용 2014) Bresnahan (1995)는 범용 기술의 3가지 요건으로 (1) 기술적 파급성(technological pervasiveness)

(2) 기술적 개선효과(technological improvement) (3) 혁신 유발성(innovation spawning)을 제시하였다. 이러한 특징을 가진 범용기술에는 전기, 컴퓨터, 나노기술 등이 존재한다. 기존에 보고된 문헌들은 '범용 기술들이 상업적인 생존력을 가지기 위해 수십 년 간의 공적, 사적인 지원을 필요하다'고 보고하고 있다 (신인용 2014). 전기의 예를 보면, 삼상교류 전류의 송전에 성공한 것은 1890년의 일이지만, 사용량에 있어서 전기가 증기를 넘어선 것은 30년 후인 1920년에나 가능하게 되었으며, 전기의 송전 및 배전 시스템의 구축을 통하여 전 국민이 전기의 혜택을 누리게 된 것은 그 후로 도입 이후 거의 50년이 소요되었다. 구체적으로 영국에서 전기를 이용할 수 있게 된 가계는 1920년에 약 12%, 1950년에 86%로 확산되는데 오랜 시간이 소요되었다. (신인용 2014)

범용 기술의 확산에 대해서 다양한 문헌에서 연구가 진행되었는데, Helpman (1998)은 다음의 4가지 요소를 확산의 중요 요인으로 제시하였다. 1) 새로운 범용기술의 생산성 이점 (the productivity advantage of the new GPT), 2) 역사적인 충격 요인 (a 'historical stock' parameter), (3) 요구량 (a demand parameter), 그리고 (4) 연구 개발 요인 (R&D parameter)이 이에 해당한다. Thoma (2008)은 실리콘 벨리의 스타트업 기업인 Lonwork사의 제어기술을 사례로 범용 기술의 확산에 영향을 주는 4가지 변수를 제시하였다. (1) 새로운 기술을 통해 얻는 이득 (benefits received from the new technology), (2) 새로운 기술을 채택하는데 필요한 비용 (costs of adopting the new technology), (3) 수용자 입장의 네트워크 효과 (network effects on the demand side), 그리고 (4) 정보와 불확실성 (information and uncertainty)이 4가지 변수에 해당한다.

범용 기술뿐만 아니라 일반적인 기술 또는 혁신의 확산은 1962년 Rogers에 의해 연구가 시작된 이후 많은 연구자들에 의해 오랫동안 연구가 진행되었다. Greenhalgh (2004)은 사회학, 마케팅 등 다양한 분야에 대한 메타 연구를 통해 확산에 영향을 주는 총 11가지 요인을 제시하였다. (1) 상대적 장점 (relative advantage), (2) 수용성 (compatibility), (3) 낮은 복잡성 (low complexity), (4) 시도 가능성 (trial-ability), (5) 관측 가능성 (observability), (6) 재창조 가능성 (potential for reinvention), (7) 불명확한 경계 (fuzzy boundaries), (8) 위험도 (risk), (9) 이슈 사항 (task issues), (10) 필요한 지식의 속성 (무언/명확성) (nature of knowledge required (tacit/explicit)), 그리고 (11) 기술적 지원이 이에 해당한다.

III. 데이터 수집

빅데이터는 정의가 시간에 따라 진화된 특징을 가지며, 특정 기술이 아닌 기술의 복합체 및 기술의 사용 방법에 해당하기 때문에 일반적인 데이터로 빅데이터의 확산 과정을 분석하는데 한계가 있다. 따라서 본 논문에서는 기존 문헌에서 기술의 확산을 분석하기 위해 사용한 다양한 종류의 데이터들을 빅데이터 분석에 적용하여 각 데이터가 어떤 특징을 가지는지 확인하였고, 빅데이터 트렌드의 선행 지수에 해당하는 데이터를 탐색하였다.

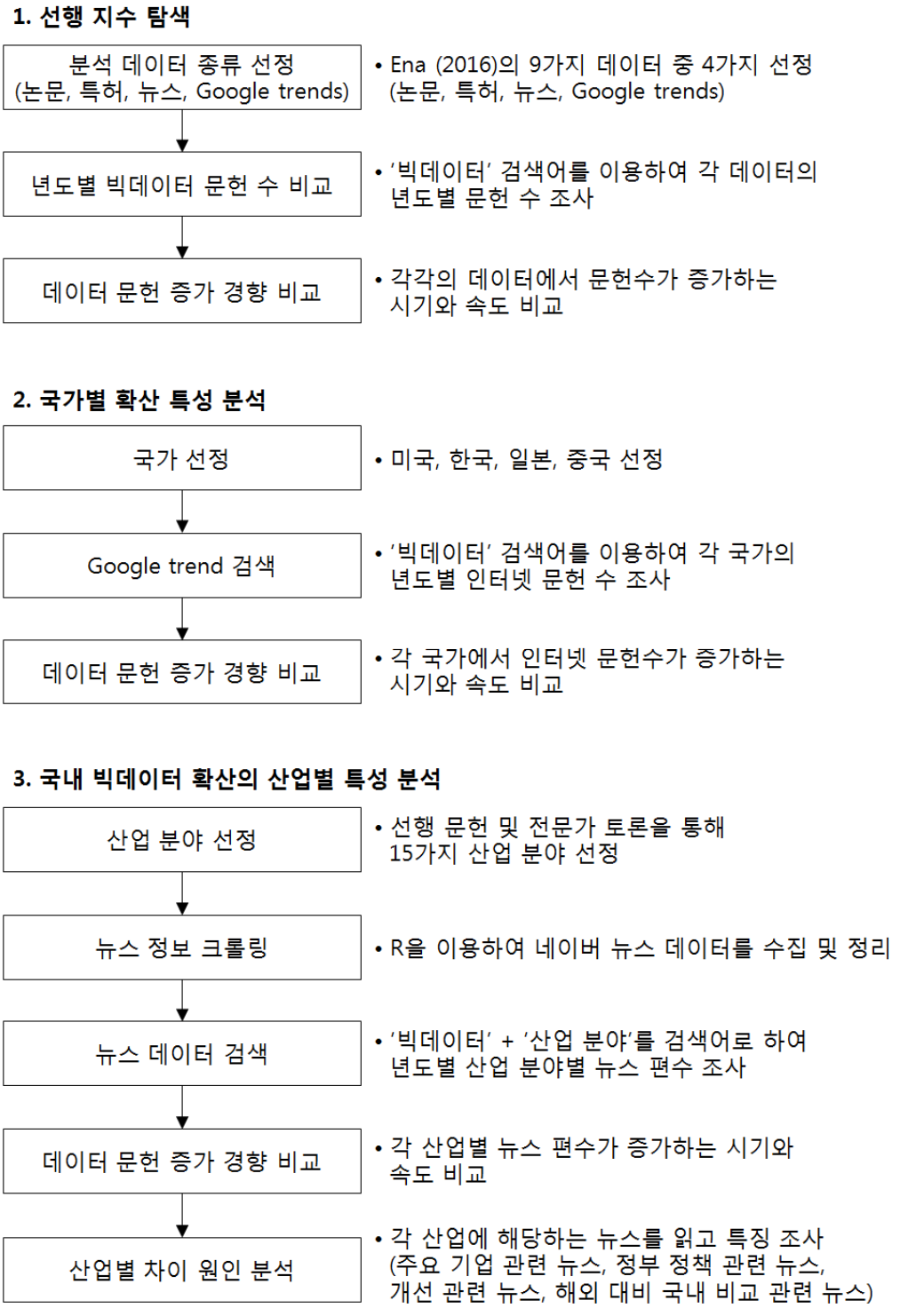
Ena (2016)는 사회와 기술 trend 분석을 위해 사용하는 데이터로 논문데이터, 특허 데이터, 뉴스 데이터, 기업 리포트, 학회 발표자료, 인터넷 검색 자료, 학위논문, 프리젠테이션 자료, 프로젝트 자료의 9가지 데이터를 제시하였고, 각각의 특징에 대해 분석하였다. Ena (2016)가 제시한 9가지 데이터 중, 기업 리포트, 학위논문, 프리젠테이션 자료, 프로젝트 자료는 시계열에 따른 확산 분석에 적합하지 않아 분석 과정에서 제외하였다. 그리고 논문 데이터와 학회 발표 자료의 경우 학술적인 연구 결과에 해당하는 공통된 특징이 있고, 본 연구에서 사용한 SCOPUS DB (database)에서 함께 검색이 가능하여 하나의 통합된 데이터로 분석하였다.

따라서 본 논문은 논문 데이터, 특허 데이터, 뉴스 데이터, 인터넷 검색 자료를 이용하여 빅데이터의 확산에 대해 분석하였다. 본 4가지 문헌들은 비즈니스 분석을 위해 다양한 문헌에서 사용되어왔다. Seyedghorban (2016)은 SCOPUS의 논문 데이터를 이용하여 1972년부터 2015년의 B2B 브랜딩 관련 문헌에 대해 분석하였다. Narin (1995)는 특허 데이터를 통해 특허가 기업의 성과를 평가하는 척도에 해당함을 제시하였다 (문진희 2017, 김준모 2014). Kubacki (2015)는 뉴스 데이터와 논문 데이터로 마케팅에 영향을 주는 요인을 분석하였다. 그리고 Fanelli (2013)은 뉴스 데이터를 이용하여 대중 매체에 의한 연구 결과물의 인용도 분석에 대해 보고하였다. Zhang (2017)은 인터넷 검색 자료의 빈도에 해당하는 구글트렌드 데이터를 이용하여 혁신적인 아이디어가 활발히 논의되기 이전 시기의 특징에 대해 분석하였다. 구글트렌드와 뉴스 데이터는 기존의 데이터로 분석이 어려운 문제점을 극복하기 위해 최근에 다양한 분석에 활용되고 있다 (Lie 2016, Preis 2013, Choi 2012). 특히, 뉴스 데이터는 정부, 대학, 기업의 최신 활동에 대한 다양한 정보를 담고 있다. 예를 들어, 기업의 신제품 출시, 투자, 타 기업과의 교류, 기업과 정부와의 교류, 인력 채용 등의 정보가 기업명, 날짜 등의 정보와 함께 자세하게 제시된다(박대민 2013, 김지혜 2010, 임수정 2018, 서윤교 2016). 하지만, 뉴스 데이터의 경우, 대중들의 유

행에 민감하게 작용하므로 특정 시점에 특정 주제의 데이터양이 급격하게 증가하는 문제가 발생할 수 있어, 본 논문에서는 분석을 통해 얻어진 결과를 각 분야의 전문가의 컨설팅과 패널 토론을 통해 데이터의 신뢰성에 대해 점검하였다.

논문 데이터는 SCOPUS 데이터베이스 (DB)를 사용하였다. SCOPUS DB는 연구 결과의 분석을 위해 빈번하게 사용되는 데이터베이스에 해당한다 (김현우 2015). SCOPUS에서 빅데이터에 대해 제목, 초록, 키워드를 2001년부터 2016년의 기간에 대해 검색하면, 총 8,986개 문헌이 검출된다. 특허는 Google patents 를 이용하였다. Google patents는 미국, 유럽, 아시아의 주요 국가들의 특허들에 대해 검색 서비스를 제공하고 있다. Google patents DB에서 빅데이터는 총 474개 문헌이 존재한다. 뉴스 데이터는 전 세계에서 발표되는 뉴스를 검색하기 위해 ProQuest 데이터 베이스를 이용하여, 총 104,578개의 데이터를 찾았다. 국내의 빅데이터 관련 뉴스는 네이버 뉴스를 통해 검색하였고, 총 31,261개의 데이터를 찾았다. 네이버 뉴스는 총 32가지의 국내 주요 언론사의 뉴스에 대해 검색이 가능하고, R 또는 Python같은 프로그램을 통해 뉴스를 크롤링하고 저장 및 분석이 가능한 장점을 가지고 있다.

수집된 데이터를 바탕으로 선행 지수 탐색, 국가별 확산 특성 분석, 국내 빅데이터 확산의 산업별 특성 분석을 진행하였고 각 단계별 절차는 그림 2에 정리하였다. 첫 번째, 선행 지수 탐색은 Ena (2016)의 9가지 데이터 중 논문, 특허, 뉴스, Google trends의 4가지 데이터를 선정하는 과정이다. 그리고 각 데이터에서 '빅데이터'를 검색어로 하여 연도별 문헌 수를 조사하였다. 시계열 데이터에서 각 데이터의 문헌수가 증가하는 시기와 속도를 비교하여 어떤 데이터가 선행하는지 분석하였다. 두 번째, 국가별 확산 특성 분석은 미국, 한국, 일본, 중국 4개 국가의 구글트렌드에 대해 '빅데이터'를 검색어로 하여 각 국가의 연도별 인터넷 문헌 수를 조사하고, 각 국가에서 인터넷 문헌수가 증가하는 시기와 속도 비교하는 과정에 해당한다. 세 번째, 국내 빅데이터 확산의 산업별 특성 분석은 5단계를 걸쳐 진행되었다. 우선, 선행 문헌 및 전문가 토론을 통해 15가지 산업 분야를 선정하였다. '빅데이터'와 '산업 분야'를 검색어로 하여 (예를 들어, '빅데이터'와 '은행'으로 검색) 연도별, 산업 분야별 뉴스 편수를 조사하였다. 각 산업별 뉴스 편수가 증가하는 시기와 속도 비교하였다. 특정 년도의 산업별 특징을 파악하기 위해 R 프로그래밍을 이용하여 2000년부터 2016년까지 네이버 뉴스 데이터를 수집하고 정리하였다.



<그림 2> 본 연구에서 빅데이터 확산의 특징을 분석하기 위해 사용한 절차

뉴스를 정리하는 과정에서 기업의 경영 활동 또는 정부의 빅데이터 관련 정책에 대해 검색하기 위해 표 1의 단어를 query로 사용하였다. Query는 EMA의 분석 보고서에서 (Devlin 2012) 정의한 도입 수준의 세 단계인 검토 단계, 계획 단계, 활용 단계에 대해 각각 검색이 가능하도록 구성하였다. 여기서 '가능성, 가치, 개척, 기대, 기회, 모색, 미래, 시대, 신규, 어려움, 주목'는 검토 단계와 연관성이 높다. '계획, 구축, 극대화, 마케팅, 비즈니스 모델, 시장, 사업, 산업, 성장, 주력, 집중, 초점, 추진, 투자'는 계획 단계와 연관성이 높고, '공개, 구현, 발표, 사례, 소개, 제공, 제시, 제품, 신제품, 출시, 활용, 고용, 교육, 육성, 인력, 채용, 경쟁, 경쟁력, 공략, 본격, 주도, 주도권, 전략, 진출, 간담회, 개최, 세미나, 컨퍼런스'는 활용 단계와 연관성이 높다. 하지만, query의 단계별 구분으로 도입 수준을 판단하지 않았고, 기사의 내용을 통해 도입 수준을 판단하였다. 그리고 범용기술 특성을 반영한 query는 사용하지 않았다. 본 논문에서는 다양한 산업에서 활용되는 현상을 파악하기 위하여 8개 산업에서 어떻게 활용되는 지 각 산업의 현황을 분석하였다.

특정 산업의 빅데이터 뉴스 중 본 query의 단어가 있는 기사를 읽고 전문가 토론을 진행하였다. 표 1의 query가 기업의 주요한 경영 활동을 검색하기에 부족할 수 있으나 31,261건의 뉴스데이터를 바탕으로 분석을 진행하였고, 검색된 뉴스에 대해 패널 토론으로 산업의 경영활동을 나타내는지 검증하였다.

<표 1> 빅데이터 기술 확산 단계 분석을 위한 기사 선정 query

경영 활동 검색을 위해 사용된 keyword들
가능성, 가치, 간담회, 개척, 개최, 경쟁, 경쟁력, 계획, 고용, 공개, 공략, 교육, 구축, 구현, 극대화, 기대, 기회, 마케팅, 모색, 미래, 발표, 본격, 분석, 비즈니스 모델, 사업, 산업, 세미나, 시대, 시장, 사례, 성장, 소개, 신규, 신제품, 어려움, 육성, 인력, 전략, 제공, 제시, 제품, 주도, 주도권, 주력, 주목, 진출, 집중, 채용, 초점, 추진, 출시, 컨퍼런스, 투자, 활용

전문가 토론은 다음과 같은 3가지 질문을 주고, 각 연도별, 산업별 특징되는 기사를 선정하였다.

1. 본 기사는 특정 산업에 해당하는 내용인가?
2. 빅데이터의 도입 수준이 활용 단계, 계획 단계, 검토 단계, 도입 전 단계 중 어느 단계에 속하는지 구분 가능한가?
3. 단계를 구분할 수 있는 구체적인 예제를 담고 있는가?

본 과정을 거쳐 선정된 기사를 표 5에서 표 8까지 연도별로 정리하였다.

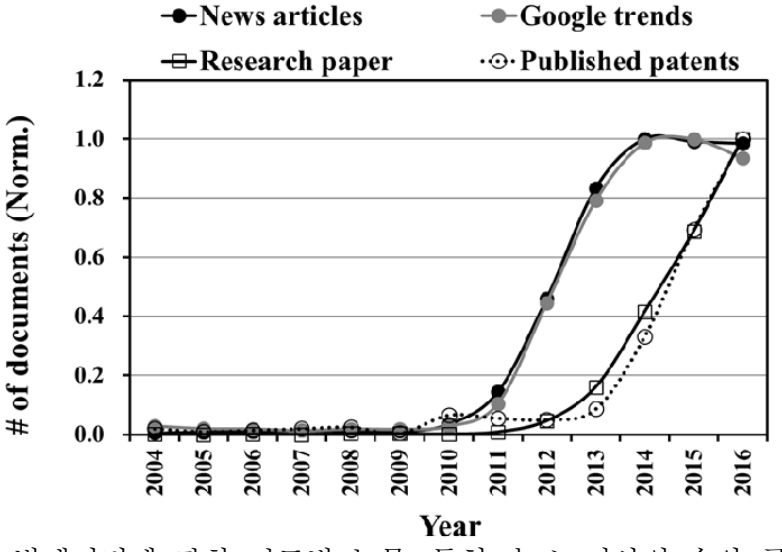
IV. 연구 결과

4.1 빅데이터 키워드 확산에 대한 선행 지수 데이터 탐색

빅데이터를 키워드로 하는 뉴스, 논문, 특허, 구글트렌드의 문헌 수를 시계열 그래프로 표현하면 그림 3과 같다. 본 4가지 데이터는 발표하는 목적에 따라 2가지로 유형을 분류할 수 있는데, 논문과 특허는 연구 결과를 발표하는 문헌에 해당하는 반면, 뉴스와 인터넷의 문헌은 기업이 자신의 관심사를 알리고 싶은 경우 혹은, 발표하는 기관에서 기업이나 대중이 관심 있어 할 내용인 경우 발표하는 특징을 지닌다. 따라서 앞의 2가지 문헌과 뒤의 2가지 문헌은 서로 다른 성격을 가진다. 그림 3은 총 4가지 문헌에서 빅데이터를 키워드로 하는 문헌 수가 시간에 따라 어떻게 변하는지 보여준다. 문헌 수가 시간에 따라 변동하는 것은 기술의 확산 속도 및 영향력과 비례하는 관계를 가진다 (이재윤 2015). 본 장에서는 위의 4가지 데이터에 대해 어떤 데이터가 빅데이터 키워드의 확산을 보여주는 선행 지수인지를 먼저 확인하고, 그 데이터를 이용하여 외국과 국내의 확산 속도 차이에 대해 논의하였다.

그림 3은 뉴스 데이터와 구글트렌드가 서로 유사한 트렌드를 가지고, 논문 데이터와 특허 데이터가 서로 유사한 트렌드를 가짐을 보여준다. 빅데이터와 관련하여 뉴스와 구글트렌드는 2011년부터 증가하여 2014년에 최대값을 가지고 2015년부터는 포화된 특성을 보인다. 한편, 등록된 특허 수와 논문 수는 뉴스 데이터에 비해 길게는 2년가량 느린 트렌드를 보여준다. 하나의 주제에 대해 4가지 데이터가 서로 다른 2가지 트렌드를 보이는 까닭은 각 데이터의 특징을 통해 이해될 수 있다. 논문과 특허는 연구를 통해 결과를 얻고 이를 등록하는데 시간이 소요되는 반면, 뉴스와 구글트렌드는 대중의 관심을 실시간으로 반영하는 특징을 지닌다. 만약 빅데이터를 키워드로 하는 주제가 학문적인 연구를 통해 확산이 시작되었다면, 논문이나 특허 발표 이후 대중적인 확산이 진행되기 때문에 뉴스나 구글트렌드가 시간적으로 느린 트렌드를 보일 것이다. 하지만, 그림 3과 같이 뉴스나 구글트렌드가 논문이나 특허보다 선행하는 현상은 빅데이터라는 키워드가 뉴스를 통해 대중에게 확산되고 본 확산이 연구자들을 촉진하여 빅데이터를 키워드로 하는 논문과 특허를 발표하도록 이끌었다고 해석할 수 있다. 따라서 뉴스와 구글트렌드 대비 논문 데이터와 특허 데이터의 시간차는 빅데이터 키워드와 관련된 정보를 입수한 후 연구를 시작하고 이에 대해 논문이나 특허를 발표하는데 걸린 시간으로 이해 가능하다. 즉, 빅데이터를 키워드로 하는 기술은 대중의 관심이 발생한 뒤 연구가 시작되는 주제로 해

석하는 경우 뉴스와 구글트렌드는 확산의 현상을 해석하는 선행 지수로 활용 가능하다. 일반적인 연구 결과는 논문이나 특허가 발표된 후 뉴스와 대중의 관심을 통해 확산이 진행되어 뉴스 데이터로 기술트렌드를 앞서서 파악하기 어려우나, 빅데이터를 키워드로 하는 문헌의 경우는 뉴스 데이터나 구글트렌드를 이용하여 변화의 흐름을 분석하는 것이 의미를 가진다.



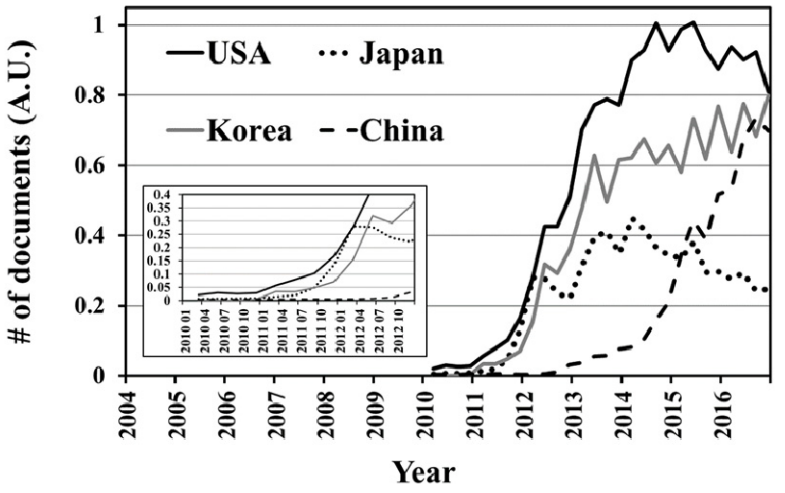
<그림 3> 빅데이터에 관한 년도별 논문, 특허, 뉴스 기사의 수와 구글트렌드

아울러, 특허나 논문은 기술에 대한 구체적 정보를 제공하는 장점을 지니지만, 기술 개발 내용에 국한되어 기업들 간의 비즈니스 관계에 대해서 정확히 파악하기 어렵고, 전체 산업 분야에서 발생하는 확산 과정을 분석하기에 접합하지 못한다. Hu (2017)의 문헌에 사용한 논문 데이터는 전체 논문의 63.0%가 컴퓨터 과학, 엔지니어링 기술, 통신 기술, 기타 과학 기술에 대한 기술 관련 내용을 차지하고 있어 산업에서 어떻게 사용되는 지에 대한 정보를 얻을 수 없다. 이에 비해 뉴스 데이터는 빅데이터에 대한 사회의 관심과 기업들의 다양한 경영활동 (투자, 제품 출시, 인력 고용, 마케팅 등)에 대한 정보를 제공한다. 따라서 검색어를 통해 어떤 산업에서 빅데이터에 대한 뉴스가 발표되었는지 확인 가능하다. 한편, 구글트렌드는 다양한 나라에서 빅데이터의 관심이 어떻게 다른지 분석하는데 활용 가능하다. 국내의 빅데이터 주제 확산을 분석하기에 앞서 국내의 확산 과정에 대한 전체적 트렌드를 파악하기 위해 미국, 한국, 일본, 중국에서 빅데이터의 관심이 어떻게 다른 지를 구글트렌드를 통해 분석하였다. 그리고 뉴스 데이터를 이용하여 국내에서 빅데이터라는 키워드가 각 산업에서 어떠한 확산 속도를 가지는지 비교 분석하였다.

4.2 선행 지수 데이터를 이용한 국내와 국외의 빅데이터 키워드 확산 특성 비교

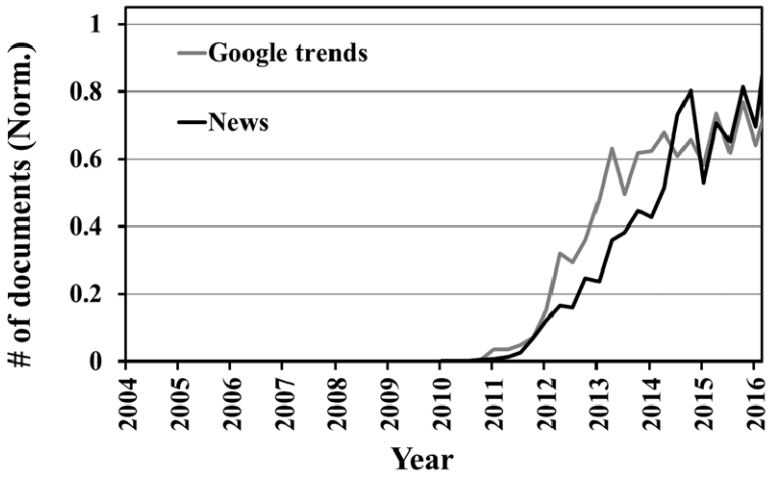
국내의 빅데이터 키워드 확산 특성을 미국, 일본, 한국, 중국의 구글트렌드 결과를 비교하여 시작 시기와 확산 속도를 분석하였다. 기존 문헌에 보고된 바와 같이 구글트렌드는 시간에 따른 인터넷 검색 량을 통해 대중의 관심을 보여준다. (신민수 2017) 그림 4는 2010년부터 2016년 사이 빅데이터에 대해 각 국가별로 명확히 다른 경향을 보임을 나타낸다. 4가지 국가 중 빅데이터는 2011년 미국에서 확산이 먼저 시작되었고, 이후 2011년에서 2014에 걸쳐 미국과 한국, 일본에서 빠르게 증가하는 추세를 보인다. 일본은 2014년 이후 감소하는 특징을 보이고, 중국은 2014년 이후 빠른 증가가 시작되는 특징을 보인다. 미국에서 2011년에 확산이 보이는 것은 미국 내의 기업과 정부의 노력으로 이해 가능하다. 2010년 12월 미국의 대통령 과학기술 자문위원회 (President's Council of Advisors on Science and Technology)는 'Designing a Digital Future' 보고서를 통해 모든 미국 연방 정부 기관은 빅데이터 전략이 필요하다고 강조하였다. (Holdren 2010) Gartner는 10대 전략 기술에서 2010년도에 advanced analytics를, 2011년에는 Next Generation Analytics와 Social Analytics를 선정하였다 (Gartner 2010, Gartner 2011). 2011년 2월에는 1400여명의 IT 전문가들이 참석하는 Making Data Work 주제의 데이터 관련 회의(Strata 행사)가 개최되었다. 특히, Strata 행사는 지금까지 IT를 지탱하는 하부 변수로만 취급되어 온 데이터의 위상을 높이고 데이터가 중요한 독립 분야임을 선언하였다. (박재득 , 2011) 아울러, 2012년 3월 미국 오바마 행정부는 빅데이터 연구 개발 계획(R&D Initiative)에 2억 달러를 투입한다고 발표하였다. (이응용 2012)

그림4에서 한국의 빅데이터 확산 트렌드는 미국 대비 6개월가량 시간차가 존재하고 2016년까지 지속적인 증가 경향이 존재함을 확인할 수 있다. 일본에 비해 빅데이터는 국내에서 2015년 이후에도 지속 성장하는 특징을 보이고, 중국에 비해 빠른 기술 도입 특성을 보인다. 따라서 빅데이터를 키워드로 하는 기술이 미국에서 시작하여(2.2장 '빅데이터의 정의' 참고) 각 국가에서 얼마나 빠른 속도로 확산되는지 알 수 있다. 본 논문에서는 2012년 이후 지속 성장해 온 빅데이터가 한국에서 각 산업별로 어떠한 과정을 거쳐 확산되었는지에 초점을 맞추고자 한다.



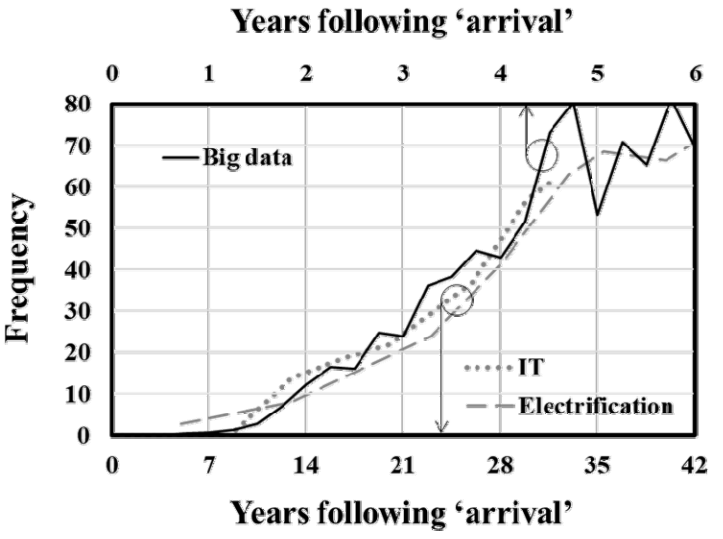
<그림 4> 미국, 일본, 한국, 중국의 빅데이터에 대한 구글트렌드

그림 5는 국내의 구글트렌드와 뉴스 데이터가 시간에 따라 어떻게 변화하는지를 보여준다. 국내의 뉴스 데이터 또한 구글트렌드와 기울기는 다르지만 2012년 이후 꾸준히 증가하는 경향을 보여준다. 특히, 2013년은 구글트렌드가 뉴스 데이터보다 높은 증가세를 보이는데, 이는 미국에서 빠르게 발전하는 빅데이터 키워드에 대해 관련 정보를 찾기 위해 많은 사람들이 인터넷을 이용한 증거에 해당하는 것으로 해석 가능하다. 뉴스의 경우 국내 정보를 전하는데 많은 비중을 차지하기 때문에 구글트렌드에 비해 기울기가 낮은 것으로 해석된다. 즉, 빅데이터를 키워드로 하는 기술이 국내에서 발생한 것이 아니라 외국에서 도입되었기 때문에 이러한 특징을 가지는 것으로 보인다. 국내의 빅데이터 뉴스는 2014년 중순 이후 높은 증가를 보이고 이후 지속적으로 증가하는 경향을 보여준다.



<그림 5> 국내 뉴스 데이터와 구글트렌드에서 빅데이터 관련 데이터 수

국내에서 진행된 빅데이터 키워드의 확산이 기존의 범용 기술 확산에 비해 어떤 속도를 가지는지 분석하였다. 대표적인 범용 기술로 전기의 확산이 있는데, 전기는 Bresnahan (2010)의 문헌을 통해 1890년대부터 1930년사이에 40년에 걸쳐 진행된 것으로 알려져 있다. 또한 컴퓨터의 확산은 1970년대부터 2000년에 걸쳐 발생하였는데, Bresnahan (2010) 문헌에서는 전기와 컴퓨터의 가격을 바탕으로 둘의 확산이 유사한 확산 속도를 가지는 것으로 보고하였다. 전기, 컴퓨터의 비용 변화와 빅데이터 뉴스 문헌수의 비교를 통해 빅데이터가 기존의 범용 기술에 비해 어떠한 확산 속도를 가지는 지 살펴보았다. 그림 6은 상단의 x축이 빅데이터가 도입 이후 년도에 해당하고, 하단의 x축은 전기나 컴퓨터의 도입 이후 년도에 해당한다. 전기와 IT는 유사한 기울기를 가지며 40년에 걸쳐 확산이 진행되는 과정을 보여주고 (Bresnahan (2010), 빅데이터의 뉴스 데이터는 6년에 걸쳐 확산이 진행되는 과정을 나타낸다. 이에 따라 빅데이터 뉴스는 7배 가량 빠른 확산 특성을 나타낸다. 빅데이터의 가격에 대한 데이터는 아직 정리되지 않았기 때문에 전기나 컴퓨터에 비해 실제 확산의 속도를 직접 비교하는 것은 어렵지만, 빅데이터 키워드가 빠른 확산 속도를 가지는 것과 뉴스데이터가 빅데이터 키워드의 확산 과정을 빠르게 대변하는 것은 분명하다.



<그림 6> 기술 도입 이후 년도에 따른 확산 정도

4.3 빅데이터의 국내 산업별 확산 특징 분석

2012년 이후 빠른 속도로 확산된 빅데이터가 국내의 산업별로 어떤 확산 차이를 가지는 지 뉴스데이터를 통해 분석하였다. 기술의 확산 과정을 분석하기 위해서는 각 산업에서 기업들이 빅데이터를 활용하는 현황 파악이 필요하다. 기술의 확산이나 메타 분석을 위해 기존 문헌에서 주로 사용된 논문이나 특허 데이터는 앞장에서 기술한 바와 같이 빅데이터의 분석에 적합하지 못하다. 구글트렌드나 뉴스 데이터가 빅데이터의 확산 경향을 보여주기에 적합한데, 이 중 뉴스 데이터는 각 산업내의 기업들이 빅데이터를 이용한 투자, 제품 출시, 기업간 교류, 인력 고용 등의 정보가 담겨있기 때문에 다양한 산업으로의 확산 특징을 분석하는데 용이하다.

본 장에서는 2000년부터 2016년까지 국내에서 발표된 총 32개의 주요 언론사의 31,261개의 뉴스를 이용하여 빅데이터와 산업의 주제어를 바탕으로 각 산업에서 어떻게 활용되는지 조사하였다. 빅데이터가 활용되는 분야 후보군을 선정하기 위해 표 2와 같이 빅데이터 업무를 경험한 적 있는 대기업, 중소기업, 정부기관 인력들과 인터뷰 및 패널 토의를 진행하였다.

<표 2> 빅데이터 활용 분석을 위한 전문가 인터뷰

대상	분야	인터뷰 방식
10명 (조건: 각 업계 10년 이상 종사자, 빅데이터 관련 업무 유경험자)	카드 회사, 의료 회사, 금융 회사, 마케팅 전문가, 정부 기관, 컨설팅 업체, 교육 업체, 국방 기관, 제조 업체, 전자 회사	대면 조사: 각 3회 이상 패널 토의: 2회

각 분야의 담당자는 각 업계에서 10년 이상 종사한 중역들을 대상으로 진행하였고, 2.3 선행 연구에서 기술한 McKinsey의 21가지 산업 분야와 EMA의 9가지 산업 분류를 참고하였고 패널 토의를 진행하여 총 15가지 분야를 선정하였다. McKinsey 분류 중 건설, 천연 자원, 부동산, 운송, 숙박, 예술, 전력 분야를 제외하였고, EMA 분류에서는 레저, 산업 분야를 제외하였다. 금융 분야에 대해선 은행, 카드, 주식을 별도로 구분하였다. 선정된 15가지 산업 분야와 각 산업 분야에서의 빅데이터 활용에 대한 참고 문헌은 표 3과 같다.

<표 3> 전문가 인터뷰를 통해 얻은 빅데이터의 활용 분야와 각 분야에 관련된 참고 문헌

산업분류	참고 문헌
SNS	Bello-Orgaz, Gema, Jason J. Jung, and David Camacho. "Social big data: Recent achievements and new challenges." *Information Fusion* 28 (2016): 45-59.
금융	Fang, Bin, and Peng Zhang. "Big data in finance" *Big Data Concepts, Theories, and Applications*. Springer, Cham, 2016. 391-412.
은행	Sun, N., et al. "iCARE: A framework for big data-based banking customer analytics." *IBM Journal of Research and Development*58.5/6 (2014): 4-1.
마케팅	Erevelles, Sunil, Nobuyuki Fukawa, and Linda Swayne. "Big Data consumer analytics and the transformation of marketing." *Journal of Business Research* 69.2 (2016): 897-904.
보안	Wang, Hongjian, et al. "Crime rate inference with big data." *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, 2016.
보험	Groves, Peter, et al. "The 'big data'revolution in US health care." *McKinsey Quarterly* 2 (2013): 3.

카드	Sobolevsky, Stanislav, et al. "Money on the move: Big data of bank card transactions as the new proxy for human mobility patterns and regional delineation. the case of residents and foreign visitors in spain." *Big Data (BigData Congress), 2014 IEEE International Congress on*. IEEE, 2014.
광고	Rabl, Tilmann, et al. "Solving big data challenges for enterprise application performance management." *Proceedings of the VLDB Endowment* 5.12 (2012): 1724-1735.
주식	Wu, Xindong, et al. "Data mining with big data." *IEEE transactions on knowledge and data engineering* 26.1 (2014): 97-107.
컨설팅	Minelli, Michael, Michele Chambers, and Ambiga Dhiraj. *Big data, big analytics: emerging business intelligence and analytic trends for today's businesses*. John Wiley & Sons, 2012.
교육	West, Darrell M. "Big data for education: Data mining, data analytics, and web dashboards." *Governance Studies at Brookings* 4 (2012): 1-0.
정치	Puschmann, Cornelius, and Jean Burgess. "The politics of Twitter data." (2013).
국방	Song, Xiao, et al. "Military simulation big data: background, state of the art, and challenges." *Mathematical Problems in Engineering* 2015 (2015)
의료	Groves, Peter, et al. "The 'big data' revolution in US health care." *McKinsey Quarterly* 2 (2013): 3.
제조업	Lee, Jay, et al. "Recent advances and trends in predictive manufacturing systems in big data environment." Manufacturing *Letters* 1.1 (2013): 38-41.

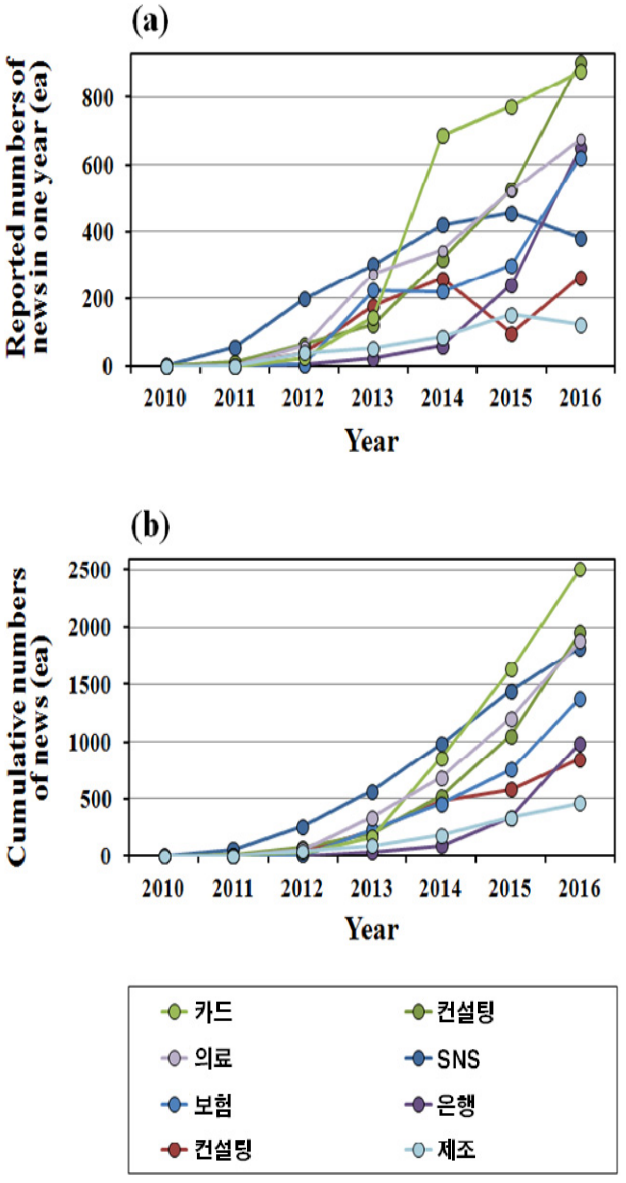
표 3에 선정된 15가지 산업 분야에 대해 국내 뉴스 데이터에서 빅데이터와 관련된 문헌의 수를 연도별로 정리하였다. 그 결과 총 15가지 산업 분야 중 국내 뉴스데이터에서 상위 8개 산업 분야에 해당하는 뉴스 데이터가 71.6%를 차지함을 확인하였다. 여기서 8가지 분야는 카드, 금융, 의료, 소셜 네트웍 서비스 (SNS), 보험, 은행, 컨설팅, 제조업에 해당한다. 8가지 산업 분야에서 빅데이터 관련 뉴스는 그림 7과 같이 시간에 따라 증가한다. 그림 7 (a)는 각 년도별 문헌 수에 해당하고, 그림 7 (b)는 총 누적 문헌 수에 해당한다. 따라서 그림 7 (a)는 해당 년도에 특정 산업에서 빅데이터와 관련된 뉴스의 양에 해당하므로 그 산업에서 빅데이터와 관련된 경영활동 (투자, 제품 출시, 인력 고용, 마케팅 등)이 얼마나 발생했는지와 연관지어 해석할 수 있다. 그리고 그림 7 (b)는 총 누적 문헌 수에 해당하기 때문에 각 산업에서 빅데이터에 대한 경영 활동의 누적량도 증가했다고 해석 가능하다.

표 4는 그림 7(a)에 해당하는 데이터를 백분율로 표시하여 해당년도에 발생된 뉴스데이터 중 특정 산업이 몇 퍼센트의 뉴스를 차지하고 있는지 보여준다.

<표 4> 빅데이터에 대한 산업계 경영활동 관련된 뉴스 데이터 중 특정 산업의 뉴스가 해당 년도에 차지하는 비율

	2010	2011	2012	2013	2014	2015	2016
카드	0.0%	0.0%	5.9%	10.9%	28.6%	25.2%	19.5%
금융	20.0%	15.9%	14.4%	9.4%	13.3%	17.2%	20.1%
의료	0.0%	3.7%	13.5%	20.7%	14.4%	17.0%	15.0%
SNS	80.0%	68.3%	44.9%	22.9%	17.5%	14.9%	8.5%
보험	0.0%	1.2%	1.4%	17.0%	9.2%	9.8%	13.8%
은행	0.0%	3.7%	1.4%	1.7%	2.5%	7.9%	14.4%
컨설팅	0.0%	6.1%	9.3%	13.4%	10.8%	3.2%	5.9%
제조	0.0%	1.2%	9.3%	4.0%	3.7%	5.0%	2.8%

그림 7과 표 4를 통해 빅데이터가 소개되고 도입된 초기에는 SNS와 관련된 뉴스가 2010년에는 80.0%, 2011년에는 68.3%, 2012년에는 44.9%로 높은 비중을 차지하고 있다. 2013년 이후에는 SNS관련된 뉴스는 비중이 점차 감소하고 있고, 특정 분야가 30% 이상 비중을 차지하지 않고 있다. 2014년에는 카드 분야의 빅데이터 뉴스가 3배 이상 증가하였다. 그리고 2016년에는 금융, 보험, 은행의 증가가 뚜렷하게 나타났다. 이를 통해 국내의 빅데이터는 2010년도에 SNS와 관련된 뉴스로 주로 소개되었고, 2013년 이후 다양한 산업으로 확산이 진행되었음을 확인할 수 있다.



<그림 7> 국내 8개 산업 분야에서 빅데이터 관련 (a) 년도별 뉴스 수와 (b) 년도별 누적 뉴스 수

해당 년도의 기사를 바탕으로 각 산업에서의 빅데이터 동향과 빅데이터 확산 과정을 분석하였고, 표 5에서 표 8까지 주요 뉴스 기사를 정리하였다.

<표 5> 2010년과 2011년 SNS와 금융 분야의 빅데이터 관련 경영 활동에 관한 뉴스 기사

년도	분야	뉴스 기사
2010년	SNS	소셜 네트워크를 통해 방대한 정보가 오고 가는 시대가 오면서 빅 데이터Big Data의 처리문제가 화두로 떠오르고 있다. (디지털데일리, 2010년 10월 25일자 기사)
2011년	SNS	온·오프라인을 통해 접수되는 고객 의견, 소셜네트워크(SNS) 상의 정보, 멀티미디어 콘텐츠, 기업 내부 정보 등을 어떻게 파악하고 경영과 마케팅에 접목시키는 지가 기업 생존의 관건으로 떠올랐다. (조선 비즈, 2011년 9월 15일자 기사)
2011년	금융	2012년 아태지역 은행 IT예산, 전년대비 7% 성장 새로운 기술 도입에 대해서는 빅데이터와 관련해 리스크 관리 및 금융사기 관리 분야에서 최초의 성공 사례들이 관측될 전망… (디지털 데일리, 2011년 12월 6일자 기사)

2010년 10월 25일자 기사에서 '소셜 네트워크를 통해 방대한 정보가 오고 가는 시대가 오면서 빅데이터의 처리 문제가 화두로 떠오르고 있다'고 발표하였다. 본 기사에서 빅데이터는 소셜 네트워크를 통해 발생하고 전달되는 정보의 량이 증가하면서 발생된 것으로 소개되고 있다. 2011년 SNS분야 기사에서는 SNS의 정보를 바탕으로 빅데이터 분석을 얼마나 잘하는 지를 기업의 중요한 문제로 제시하고 있다. 그리고 2011년 12월 6일자 기사에 따르면 금융 분야에서 2012년 IT관련 예산이 증가하고 있고 이는 빅데이터를 이용하여 리스크 관리 및 금융사기 관리에서 사용될 것으로 전망하고 있다. 표 3과 표 4를 통해 SNS분야와 금융 분야는 빅데이터가 처음 적용되는 분야로 Rogers의 기술 확산 모델에 따르면 빅데이터에 대해 처음 적용하고 이에 대한 효과를 확인하는 innovator역할로 해석 가능하다. Facebook과 Twitter같은 SNS는 국내에서 2011년도에 이용자수가 이미 1천만 명을 돌파하였기 때문에, SNS의 방대한 정보에 대해 빅데이터 분석을 적용할 수 있고 이를 활용하면 중요한 정보를 얻을 수 있다는 것을 누구나 쉽게 이해할 수 있기 때문에 SNS분야가 빅데이터 확산의 처음을 담당하고 있다. 2011년 금융 분야의 빅데이터 기사는 국내가 아닌 아시아 태평양 지역에 대한 전반적인 내용이고, 2012년에 수립된 예산에 대한 내용으로 국내의 구체적인 활동이 없음을 나타낸다.

<표 6> 2012년과 2013년 SNS, 금융, 의료, 보험, 카드분야의 빅데이터 관련 경영 활동에 관한 뉴스 기사

년도	분야	뉴스 기사
2012년	SNS	전 세계적으로 최근 수년간 페이스북 등의 소셜 네트워크 서비스(SNS)나 기기간 (M2M) 통신을 이용한 센서 네트워크, 그리고 기업의 IT 시스템에서 발생하는 대량의 데이터의 수집과 분석, 즉 이른바 빅 데이터의 활용이 활발해지고 있다. (머니투데이, 2012년 9월 25일자 기사)
2012년	금융	한 금융 서비스 기업은 상업 고객 대출 시 위험을 파악하고 줄이기 위해 Infosys의 빅 데이터 솔루션을 채택했다. 이 솔루션은 이 기업이 무수익 자산을 중심으로 많은 데이터를 처리하고, 실시간 통찰을 확보하며, 융자 과정을 능률화하도록 지원할 것이다. (연합뉴스, 2012년 10월 13일자 기사)
2012년	의료	제이슨 황은 파괴적 의료혁신이라는 개념이 등장한 이래 보건·의료에 미칠 가장 큰 기술 혁신으로 모바일헬스와 빅데이터로 꼽았다. (코리아헬스로드, 2012년 7월 13일자 기사)
2013년	SNS	빅데이터 흐름을 보면 여러 가지 데이터들이 쏟아져 나오는 것을 볼 수 있다... 그 중 소셜미디어 영역에서 급격하게 데이터가 증가하고 있다… 소셜네트워크분석, 소셜애널리틱스 등 이런 기술 요소들은 사실 헛물을 켜고 있는 것일 수도 있다. 최종적으로 남아 있는 기술들을 살펴보면 웹 분석, 예측 분석 등 이 기술들이 꼭 필요하다가 전망하고 있다. (IT데일리, 2013년 3월 19일자 기사)
2013년	금융	금융권 빅데이터사업 수주전 치열… 빅3 IT서비스기업 중 LG CNS 이외에 빅데이터 전담조직이 따로 없지만 빅데이터를 데이터분야의 연장선상에서 봤을 때 업체별로 기존 금융사업부에서 참여하는 형태로 파악된다. (디지털타임즈, 2013년 3월 12일자 기사)
2013년	의료	가톨릭의료원 산하 8개병원 '빅데이터' 주목…인간의 생명을 다루는 의학에도 이 빅데이터의 개념을 도입해 보는 것은 어떨까? 최근 한 의과대학이 그 시도를 하고 나서 귀추가 주목된다. (데일리메디, 2013년 4월 28일자 기사)
2013년	보험	건강보험 '빅데이터' 열린다…100만명 표본DB 공개… 건보공단은 22일 오후 이 같은 내용의 '국민건강정보 빅데이터 활용을 위한 표본 코호트DB 설명 및 공개방안 세미나'를 개최하고 DB 구축과 공개 범위, 향후 정보공개 방안 등에 대해 논의했다. (데이리팜, 2013년 3월 23일자 기사)
2013년	카드	라이프-로그란 위성위치확인시스템(GPS), 카메라, 카드 등의 디지털 기기를 통해 개인의 일상 생활을 기록하는 것을 말한다…. '라이프 로그가 개인의 '빅 데이터' 노릇을 하는 셈이다. (한겨레, 2013년 8월 28일자 기사)

2012년 9월 25일자 기사는 '최근 수년간 페이스북 등의 소셜 네트워크 서비스(SNS)에서 발생하는 대량의 데이터의 수집과 분석, 즉 이른바 빅데이터의 활용이 활발해지고 있다.'고 보도하고 있다. 즉, SNS분야에서는 빅데이터의 활용이 본격적으로 진행되고 있음을 나타낸다. 그리고 2013년 3월 19일 기사에서는 '소셜네트워크분석, 소셜애널리틱스 등 이런 기술 요소들은 사실 헛물을 켜고 있는 것일 수도 있다. 최종적으로 남아 있는 기술들을 살펴보면 웹 분석, 예측 분석 등 이 기술들이 꼭 필요하다'고 보도하였다. 이를 통해 SNS분야에서 진행되는 다양한 빅데이터 분석에 대해 검토하고 실제 유효한 방향을 고민하는 시기에 해당한다. 2012년 금융분야에서 빅데이터의 활용이 시작되어 한 기업에서 어떻게 사용되고 있는지 보여주고 있다. 그리고 2013년 3월 12일자 기사에서는 금융분야에서 3개의 대형 IT기업들이 금융권 빅데이터 사업을 수주하기 위해 경쟁하는 기사를 보도하고 있다. 그리고 의료분야에서는 2013년 4월 28일자 기사를 통해 특정 병원이 빅데이터를 도입하고 시도하는 현황에 대해 보고하였다. 이를 통해 금융 분야와 의료 분야는 2012년, 2013년이 Innovator 시기에 해당한다. 반면, 보험 분야의 경우 국민건강정보는 국가의 정책에 따라 관리되기 때문에 데이터 공개 범위를 논의하는 단계부터 시작하고 2013년 3월에 국민건강보험공단에서 이러한 세미나를 진행하였다.

<표 7> 2014년 SNS, 금융, 의료, 보험, 카드, 은행분야의 빅데이터 관련 경영 활동에 관한 뉴스 기사

년도	분야	뉴스 기사
2014년	SNS	공공치안을 책임지는 경찰, 소방, 국가안전 당국도 이제 소셜 데이터를 분석하고 업무에 적용할 수 있어야 한다. 싱가포르 경찰 출신으로 오라클 인더스트리 비즈니스 사업부에서 공공안전 부문 글로벌 총괄을 맡고 있는 홍잉 코 수석 이사는 … 각국 치안당국이 IT기술을 활용할 필요성을 강조했다. (ZDnet코리아, 2014년 2월 20일자 기사)
2014년	금융	금융권, 빅데이터 서비스·마케팅·상품개발 활발… 빅데이터를 도입한 기업들은 의사결정의 효율화, 업무 처리의 개선, 산업 경쟁력 강화 등의 효과를 꼽는다 금융사들도 그동안 보유한 고객들의 데이터를 이용한 상품개발이나 서비스에 활용하고 있다. (뉴스웨이, 2014년 6월 10일자 기사)
2014년	의료	보건의료분야 '빅데이터' 활성화 위한 첫걸음은 관련 '법률' 통합... 정부가 빅데이터에 주목하며 보건의료분야에도 이를 접목시켜 국가질병관리 DB, 식의약 DB, 심평원의 국가 의료정보 DB 등을 국가 중점 개방 데이터로 개방하고, 활용 기반을 조성하는 사업 등을 진행하고 있다. (메디컬투데이, 2014년 11월 19일자 기사)

2014년	보험	다중 지능으로 데이터 분석 해결한 캐글 … 캐글의 경연 대회를 통해 한 보험회사의 사고 예측 정확도가 예전보다 370% 나 높아지기도 했다. (한경 Business, 2014년 2월 10일자 기사)
2014년	카드	카드사들이 신성장동력으로 추진했던 빅데이터 사업에서 주도권 잡기 경쟁이 격화될 것으로 보인다. 빅데이터는 소비자의 구매 패턴 등을 분석해 마케팅에 활용할 수 있어 각광받고 있다. (이투데이, 2014년 4월 8일자 기사)
2014년	은행	전세계적으로 '빅데이터' 열풍이 불고 있지만, 국내 시중은행의 대부분은 빅데이터 활용에 소극적인 모습을 보이고 있다.... 13일 금융권에 따르면 우리·기업·외환·신한·국민은행 등 시중은행은 빅데이터 도입 검토 단계에 머무르고 있다. (아시아투데이, 2014년 10월 13일자 기사)

2014년 2월 20일자 기사는 SNS분야의 빅데이터 활용이 선도 사례로 소개되어 다른 분야의 적용을 촉진하는 역할을 하고 있음을 보여준다. 2014년 금융 분야는 다양한 금융사들이 빅데이터와 관련된 경영활동을 활발하게 진행하고 있음을 나타낸다. 한편 2014년의 카드산업은 2013년 금융 산업에서처럼 다양한 기업들이 빅데이터에 대한 초기 주도권 경쟁을 진행하는 단계에 해당한다. 반편, 은행업계에서는 빅데이터에 대해 소극적인 모습을 나타내며 아직 관련 경영활동이 시작되지 않음을 보여준다. 2014년은 각 산업별로 빅데이터 수용 단계에 큰 차이가 발생하는 시점으로 SNS분야에서는 충분한 활용을 넘어 다른 분야의 적용을 촉진하는 반면 은행분야에서는 시작도 진행되지 않았다.

<표 8> 2015년, 2016년 SNS, 금융, 의료, 보험, 카드, 은행, 제조업분야의 빅데이터 관련 경영 활동에 관한 뉴스 기사

년도	분야	뉴스 기사
2015년	SNS	SK텔레콤과 휴맵컨텐츠는 국내외 소셜네트워크서비스(SNS)·블로그·카페·게시판 등 온라인 채널에 표출되는 '빅데이터'를 분석해 활용하는 '스마트 아티스트 마케팅(SAM)'을 본격화한다고 10일 밝혔다. (데일리안, 2015년 2월 10일자 기사)
2015년	금융	유독 2015년의 혁신은 금융권에서 강하게 불고 있는 듯하다. '핀테크'와 '빅데이터'는 더 이상 낯선 IT용어가 아니라 경영자들 스스로 학습해서 익혀야할 정도의 상식적인 지식이 되었고, 실제 이 기술을 적용한 다양한 솔루션이 현실에 적용되고 있다. (대한금융신문, 2015년 12월 27일자 기사)

2015년	의료	국립암센터는 정밀의학 기반인 대규모 '암 정밀의학 코호트'를 구축한다. 전주기적 국가암관리정책 실현에 활용하는 것을 계획 중이다. 120만명 암 경험자와 생존자 데이터베이스(DB)를 구축해 암치료 후 생활습관 등 암 빅데이터 분석을 실시한다. 재발 예방, 치료 후 생존율 향상 등을 위한 전주기적 관리시스템을 개발한다. (전자신문, 2015년 10월 20일자 기사)
2015년	보험	"보험사 A부터 Z까지 미치는 빅데이터의 손" …. 8일 보험연구원의 '보험산업의 빅데이터 활용 가능성'에 따르면 독일을 대표하는 글로벌 금융사 뮌헨리(Munich RE)는 빅데이터가 요율 및 상품개발, 판매 및 마케팅, 언더라이팅, 보상에 이르는 보험산업 전 부문에 걸쳐 영향을 미칠 것으로 전망했다. (이코노믹 리뷰, 2015년 2월 9일자 기사)
2015년	카드	결제 패턴 등 빅데이터 분석을 통해 회원들에게 맞춤형 쿠폰을 제공하려는 카드회사들의 빅데이터 마케팅 경쟁이 가속화하고 있다. 삼성카드와 신한카드가 앞서 회원 개인별 맞춤형 쿠폰 서비스를 선보인 데 이어 롯데카드도 이달부터 720여만 회원 개개인에게 특화된 할인 쿠폰을 실시간으로 제공하는 서비스에 나선다. (한국경제, 2015년 8월 3일자 기사)
2015년	은행	금융위원회로부터 예비인가심사를 통과한 카카오뱅크와 K뱅크 컨소시엄이 이르면 내년 하반기부터 본격적인 금융 서비스에 나선다… 이들의 무기는 수년간 IT서비스를 통해 축적된 빅데이터다. 원금 상환능력이 떨어지는 저신용자가 주된 고객인 만큼 빅데이터 기반 신용평가모델을 소개한 것. 카카오뱅크는 KB국민은행·카카오톡·다음검색·샵검색 이용내역 등을 추가 데이터로 활용하겠다고 밝혔다. (아시아투데이, 2015년 11월 30일자 기사)
2015년	제조업	'스마트 제조업' 롤모델 개발·확산 시급... 전문가들은 국내 제조업 혁신을 위해 크게 4가지 키워드를 제시한다. 중소기업 지원책 마련, 인재 양성을 통한 생태계 확보, 사물인터넷과 빅데이터 활용, 기업이나 경영자의 인식 전환 등이다. (매일경제, 2015년 10월 2일자 기사)

2015년은 표 8과 같이 SNS, 금융, 의료, 보험, 카드, 은행 분야에서 빅데이터의 활용이 본격적으로 활용되는 시기에 해당한다. 이 시기에 각 산업의 기업들은 빅데이터와 관련된 새로운 제품이나 서비스를 출시하고, 고객에게 필요한 정보를 빅데이터 분석을 바탕으로 얻어 이를 활용하고 있다. 이 중 은행분야의 경우, 2014년까지 시작도 되지 않았으나 2015년에 카카오뱅크와 K뱅크 컨소시엄에서 본격적인 서비스가 등장하였다. 카카오뱅크는 카카오톡을 통해 SNS분야의 빅데이터 경험을 가진 기업에 해당하고, K뱅크는

KT, 우리은행, 노틸러스효성, GS리테일, 현대증권이 참여하는 컨소시엄으로 KT가 가진 빅데이터 활용 능력에 기반하고 있다. 즉, 다른 분야에서 빅데이터에 대한 역량을 보유한 기업이 은행분야로 진출하여 은행분야의 빅데이터 활용이 시작하게 된 경우에 해당한다. SNS 분야는 SNS데이터 외에 블로그, 카페, 게시판의 데이터와 통합되어 마케팅에 활용되었다. 그리고 2016년 1월 28일 조선 비즈 기사에 따르면 'LG CNS는 빅데이터 전문가와 일반인들이 손쉽게 포털, SNS, 언론기사 등에 등장하는 이슈의 흐름을 읽을 수 있는 빅데이터(방대한 양의 정보) 서비스 오디피아(www.odpia.org)를 개설했다고 27일 밝혔다'고 보고하였다. SNS분야에서 빅데이터는 기업이나 빅데이터 분석 전문가뿐만 아니라 일반 대중도 쉽게 사용할 수 있는 서비스를 제공하고 있다. 실제로 오디피아 사이트에서는 관심 있는 주제에 대해 단순 조작만으로 빅데이터 분석에 해당하는 그래프와 도표 등을 할 수 있도록 제공하고 있다. 즉, 2016년의 SNS분야는 일반 대중의 확산을 위해 SNS분야에서는 사용 방법이 표준화되고 이에 대해 쉽게 사용할 수 있는 서비스들이 존재하는 단계에 해당한다. 한편, 제조업분야는 2015년에 구체적인 활용이 되지 못하고 있으며, 이를 개선하기 위해 정부의 '중소기업 지원책 마련, 인재 양성을 통한 생태계 확보, 사물인터넷과 빅데이터 활용, 기업이나 경영자의 인식 전환'등이 필요하다고 보고하고 있다. 2015년의 제조업 분야는 2014년의 은행분야처럼 다른 산업분야 대비 속도가 늦은 상태이고, 이를 개선하기 위해 전문가들의 조언이 제시되는 단계에 해당한다.

본 분석 결과 국내의 빅데이터 확산에서 SNS분야는 2010년, 2011년부터 활용이 시작되므로 innovator에 해당한다. 반면, 제조업분야는 2015년에도 활용에 대한 구체적 방안이 수립되지 못하므로 Late Majority 또는 Laggard의 특징을 보여준다. Rogers의 분류를 적용하는 경우, 2011년, 2012년에 확산이 진행된 금융분야는 Early adaptor에 해당하고, 2013년, 2014년부터 활용이 시작된 의료, 보험, 카드분야는 Early majority로 볼 수 있다. 빅데이터는 현재도 발전과 진화가 계속 진행되고 있어 확산 모델에서 구분하는 Early majority, Late majority, Laggards에 대해선 추가 분석이 필요하다.

본 결과를 바탕으로 빅데이터의 국내 초기 확산 과정을 통해 산업별 빅데이터 도입 시기의 차이와 각 단계별 특징에 대해 다음과 같이 정리할 수 있다. 2000년 초부터 다양한 논의를 통해 개념이 정립된 빅데이터는 2010년, 2011년 미국의 정부 및 기업의 주도로 본격적인 발전이 시작되었고, 이에 대한 정보가 6개월가량의 시간차를 두고 국내에서 확산이 시작되었다. 2010년도에는 미국의 활용사례와 빅데이터의 정의가 소개되었고, 2011년 SNS분야 빅데이터가 활용되기 시작하였다. 이는 데이터의 방대한 양, 데이터 활용의 가치, 데이터 접근의 용이성에 의해 SNS분야에서 처음 도입되었다. 그 이후 2012

년 금융 분야에서 활용이 시작되었고, 2013년, 2014년부터는 의료, 보험, 카드 등의 분야로 확산이 시작되었다. 활용의 시작은 각 분야에서 이를 처음 도입하는 회사로부터 시작하고 대략 1년 후 여러 기업들이 도입하여 경쟁을 하는 특성을 보인다. 예를 들면, 금융 분야에서는 2012년에 '한 금융 서비스 기업이 빅데이터 솔루션을 채택하였다'는 기사가 있었고, 2013년에 '실시간 마케팅 빅데이터 사업에 삼성SDS와 LG CNS, SK C&C 빅3 IT서비스사가 입찰에 참여하여 금융권 빅데이터 사업 수주전이 치열하다'고 보고하고 있다. 그리고 2014년에는 '빅데이터 서비스·마케팅·상품개발 활발하다'고 보고하고 있다. 카드 분야에서는 2013년에는 '라이프-로그 기능을 소개'하는 기사였으나, 2014년에는 '카드사들이 신성장동력으로 추진했던 빅데이터 사업에서 주도권 잡기 경쟁이 격화될 것으로 보인다'고 보고하고 있다. 은행 분야에서는 2014년 '국내 시중은행의 대부분은 빅데이터 활용에 소극적인 모습을 보이고 있다'는 기사가 있으나, 2015년에는 '카카오뱅크와 K뱅크 컨소시엄이 이르면 내년 하반기부터 본격적인 금융 서비스에 나선다'고 보고하고 있다. 빅데이터를 활용하기 위해선 데이터의 준비가 필요하기 때문에 정부가 데이터를 관리하는 보험의 분야에서는 정부의 주도적인 역할이 중요하고, 개별 기업 내의 데이터를 활용하는 경우 기업 내에서 데이터를 분석하고 활용하는 역량에 의존하여 확산의 시기와 속도가 결정되었다. 특히, 은행의 경우 기존 기업들의 소극적인 도입에 따라 SNS 또는 통신 분야에서 빅데이터를 활용한 경험이 있는 기업들이 은행에 진출하는 양상을 보이고 있다.

빠르게 발전하는 기술에 따라 종래 기업들이 이를 어떻게 받아들이고 활용하는지는 기업의 혁신과 성과 창출을 위해 중요한 요인에 해당한다. 빅데이터가 미국에서 시작되어 국내로 확산되는 초기 과정을 통해 새로운 범용 기술을 받아들이는데 필요한 요소와 발전 양상을 예측하는데 적용 가능하다.

V. 결론

본 논문에서는 국내에 도입되어 산업 전반에 걸쳐 확산되는 초기 과정에 대한 시계열 분석을 통해 빅데이터의 범용 기술 특징을 분석하였고, 각 산업의 확산 특성 차이에 대해 조사하였다. 논문 분석을 통해 빅데이터의 특성 분석을 진행한 Hu (2017) 문헌의 한계를 극복하기 위해 논문 데이터, 특허 데이터, 뉴스 데이터, 구글트렌드를 복합적으로 이용하여 빅데이터라는 키워드에 대해 선행 지수에 해당하는 데이터가 무엇인지와 각 데이터의 트렌드 차이에 대해 분석하였다. 그 결과 빅데이터는 논문과 특허보다 뉴스와 구글트렌드가 2년가량 선행하는 트렌드를 보이는데, 이에 대해 뉴스를 통해 대중에게 확산되고 본 확산이 연구자들을 촉진하여 관련된 기술에 대해 논문과 특허를 발표하도록 이끈 결과로 해석하였다. 구글트렌드를 이용하여 각 국가별 도입 시기와 확산 양산을 비교하였고, 뉴스 데이터를 통해 국내의 주요한 8가지 산업 분야에 대해 확산이 진행되는 과정을 정량적 그리고 사례를 바탕으로 분석하였다.

국내에서 빅데이터는 2010년도 미국의 활용사례와 빅데이터가 무엇인지 소개된 이후, 2011년도에 SNS분야부터 활용되기 시작하였다. 이는 SNS분야의 방대한 데이터 양, 데이터 활용의 가치, 데이터 접근의 용이성에 비롯되었다. 그 이후 2012년 금융 분야에서 활용이 시작되었고, 2013년, 2014년부터는 의료, 보험, 카드 등의 분야로 확산이 시작되었다. 활용의 시작은 각 분야에서 이를 처음 도입하는 회사로부터 시작하고 대략 1년 후 여러 기업들이 도입하여 경쟁을 하는 특성을 보인다. 빅데이터를 활용하기 위해선 데이터의 준비가 필요하기 때문에 정부가 데이터를 관리하는 보험의 분야에서는 정부의 주도적인 역할이 중요하고, 개별 기업 내의 데이터를 활용하는 경우 기업 내에서 데이터를 분석하고 활용하는 역량에 의존하여 확산의 시기와 속도가 결정되었다. 특히, 은행의 경우 기존 기업들의 소극적인 도입에 따라 SNS 또는 통신 분야에서 빅데이터를 활용한 기업이 은행에 진출하는 양상을 보이고 있다. 본 분석은 뉴스 데이터를 통해 이루어졌는데, 뉴스 데이터는 기업의 경영활동과 관련된 정보를 제공하는 장점을 지니지만, 논문이나 특허에 비해 객관성 결여된 단점을 지니고 있다. 이러한 단점을 보완하기 위해 뉴스 데이터를 검색하기 위한 주제어 선정과 분석 결과에 대해 빅데이터 산업에 종사한 전문가 인터뷰를 통해 오류가 없는지 검토하였다.

본 논문은 산업 전반에 걸쳐 영향을 주는 범용 기술이 어떻게 초기 확산이 이루어지는지에 대한 실증적 연구 방법을 제시하였고, 빅데이터의 경우 국내에서 각 산업별 확산 속도 차이는 어디에서 비롯되는지 파악하였다. 본 논문에서 제시한 방법은 빅데이터 이

외에 다른 기술의 확산 과정에도 분석할 수 있으며, 특정 국가에서 기술 키워드의 확산에 해당하므로 개발도상국에서 외국으로부터 도입된 기술을 어떻게 받아들일지 분석하는데 사용될 수 있다. 그리고 기업 측면에서는 새로운 기술을 출시하고 이를 확산하고자 할 때 어떤 경로가 효과적인지 이해할 수 있다.

본 연구에는 다음과 같은 4가지 측면의 한계가 있다. 첫째, 뉴스 데이터를 이용하여 빅데이터 트렌드를 분석하고자 하였으나 뉴스데이터는 유행 등에 의해 민감하게 반응하므로 뉴스의 개수는 기술 확산 정도를 정확하게 대변하지 못한다. 아울러, 기사는 언론사에서 자체적인 기준에 의해 작성되므로 내용에 대해 검증되지 못하였다. 이러한 문제점을 보완하기 위해 32개 언론사에서 발표된 31,261개의 뉴스를 이용하여 분석하였고, 기사를 읽고 전문가 토론을 거쳤다. 하지만, 뉴스데이터라는 한계를 극복하기에는 부족함이 있다. 따라서 이러한 문제를 보완하기 위해 설문 조사 또는 기업의 데이터 활용 정보를 바탕으로 한 회귀분석 연구가 필요하다. 두 번째, 각 산업에 대한 확산 속도 차이를 비교하였으나 산업 내에는 다양한 기업들이 존재하고 빅데이터를 도입하고 활용할 수 있는 역량의 차이가 존재한다. 본 연구에서는 특정 산업의 빅데이터 활용 사례의 근거로 산업 내 몇몇 기업의 활용 사례를 예제로 들었다. 산업 내 기업들에서 몇 퍼센트 가량 활용하는지에 대한 후속 연구가 필요하다. 세 번째, 본 연구는 빅데이터를 키워드로 하여 데이터를 분석하였다. 빅데이터만 이용한 경우에도 3만 건 이상의 뉴스데이터가 존재하여 산업별 확산 트렌드 분석을 진행하였으나, 빅데이터를 구성하는 다양한 요소 및 연관 단어를 이용하여 확산 과정에 대한 더욱 정교한 분석이 가능하다. 따라서 후속 연구를 통해 다양한 연관 단어를 이용한 추가 분석이 필요하다. 네 번째, 빅데이터의 확산과 빅데이터 키워드의 확산은 차이가 존재한다. 분석에 사용한 논문, 특허, 뉴스, 구글 트렌드는 빅데이터를 키워드로 하여 확산을 분석하였기 때문에 빅데이터의 확산이 아닌 빅데이터 키워드의 확산에 해당한다. 따라서 본 논문의 4.1에서 설명한 뉴스와 구글 트렌드의 확산이 논문과 특허 등에 선행한다는 분석은 빅데이터 키워드의 확산에 해당하는 내용으로 빅데이터 자체의 확산에 대한 설명으로 부족하다. 따라서 이러한 부분을 보완하기 위해 단순히 빅데이터 키워드가 아니라 빅데이터를 구성하는 다양한 기술들을 이용하여 문헌 별 확산 특성에 대한 추가적인 분석이 필요하다.

참고문헌

(1) 국내문헌

김경태 (2014), 빅데이터 활용사례, 한국지능정보시스템학회, 한국지능정보시스템학회 학술대회논문집, pp.71-84.

김문수, 김호 (2003), 기술 및 수요속성에 따른 정보통신서비스 확산 패턴, 기술혁신연구, 제11권, 제2호, pp.71-89

김방룡, 홍재표, 고순주 (2014) 특허분석을 통한 빅데이터 기술개발 동향, ETRI 보고서

김은영, 이정훈, 서동욱 (2013), 빅데이터 시스템의 수용의도에 영향을 미치는 수용조직의 환경요인에 관한 연구, 한국데이타베이스학회지, 제20권, 제4호, pp.1-18

김정경 (2016), 국내·외 빅데이터 동향 및 성공사례, 대한산업공학회, ie 매거진, 제23권, 제1호, pp.47-52.

김정선, 송태민 (2014), 빅데이터 기술수용의 초기 특성 연구, 한국콘텐츠학회논문지, 제14권, 제9호, pp.538-555

김준모, 신준석 (2014), 특허 인용 네트워크와 동적 기술트리 분석을 활용한 기술 진화 경로 연구: 초고압 직류송전 시스템 사례, 기술혁신연구, 제22권, 제4호, pp.117-143.

김지웅, 허준, 김장일 (2013), 빅데이터의 금융기관 활용 사례, 전자공학회지 제40권, 제8호, pp.49-54.

김지혜, 장재영, 윤홍준, 김한준 (2010). 키워드 관련도를 이용한 뉴스기사의 연관검색 기법. 「한국정보과학회 학술발표논문집」, 제37권, 제1C호, pp.53-57.

김현우, 최윤정, 문영호 (2015). 과학계량학적 HCP 분석을 통한 연구개발 성과 평가. 「한국기술혁신학회 학술대회」, pp.226-240.

문진희, 권의준, 금영정 (2017), 특허 동시분류분석과 텍스트마이닝을 활용한 사물인터넷 기술융합분석. 「기술혁신연구」, 제25권, 제3호, pp.1-24.

박대민 (2013). 뉴스 기사의 빅데이터 분석 방법으로서 뉴스정보원연결망분석. 한국언론학보」, 제57권, 제6호, pp.234-262.

박세환 (2014), 빅데이터 기술 및 시장 동향, 정보통신기술진흥센터 주간기술동향, pp.15-24.

박재득, (2011) 빅 데이타 산업현황 및 대응방안, 「한국산업기술평가관리원」, KEIT PD Issue, 제11권, 제4호

박종현, 주창림, 신혁, 김문구 (2015), 기업의 빅데이터 수용의 영향요인, 한국경영학회 통합학술발표논문집 , pp.72-76

복경수, 유재수 (2014), 빅데이터 활성화 정책 및 응용 사례, 한국정보과학회, 정보과학회지 32(11), pp.46-57.

서윤교, 김시정 (2016). 뉴스 내용분석과 하이프 사이클을 활용한 기술기획의 탐색적 연구. 「기술혁신학회지」, 제19권, 제1호, pp.80-104.
서환주, 안정화 (2001), 정보통신기술의 확산과 결정요인, 기술혁신연구, 제9권, 제2호, pp.56-76
신민수, 박민규, 배성훈. (2017). 나노 인포매틱스 기반 구축을 위한 구글트렌드와 데이터 마이닝 기법을 활용한 나노 기술 트렌드 분석. 「산업경영시스템학회지」, 제40권, 제4호, pp.237-245.
신인용, 김현호 (2014), "일반범용기술과 경제성장. 「생산성논집」, 제28권, 제2호, pp.337-360.
이영주, 양현철 (2017), 활용 주체별 빅데이터 수용 인식 차이에 관한 연구: 활용 목적, 조직 규모, 업종 특성을 중심으로, 정보화정책, 제24권, 제1호, pp. 79-99
이응용, (2012) 미국정부의 빅데이터 R&D전략, 「인터넷&시큐리티 이슈, 한국인터넷진흥원」
이재윤, 최상희. (2015). 논문 인용 영향력 측정 지수의 편향성에 대한 연구. 「정보관리학회지」, 제32권, 제4호, pp.205-221.
임수정, 박덕근 (2018), 소셜미디어 분석을 활용한 재난안전산업 육성정책 수립방안, 「기술혁신연구」, 제26권, 제1호, pp.31-57.
정보통신산업진흥원 (2013), 빅데이터 확산에 따른 도전과 기회, 주간기술동향 1498

(2) 국외문헌

Accenture (2014). Big Success with Big Data.
Bai, J., Fan, J., & Tsay, R. (2016). Special Issue on Big Data, Journal of Business & Economic Statistics, Vol.34, pp.487-488
Bass, F. M. (1969). A new product growth for model consumer durables. Management science, Vol.15, No.5, pp.215-227.
Bourgoin, M. O., & Smith, S. J. (1995). Big Data--Better Returns, Leveraging Your Hidden Data Assets to Improve ROI. Artificial Intelligence in the Capital Markets, Probus Publishing Company.
Bresnahan, T. (2010). General purpose technologies. In Handbook of the Economics of Innovation (Vol. 2, pp. 761-791). North-Holland.
Bresnahan, T. F., & Trajtenberg, M. (1995). General purpose technologies 'Engines of growth'?. Journal of econometrics, Vol.65, No.1,pp. 83-108.
Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: from big data to big impact. MIS quarterly, Vol.36, No.4, pp.1165-1188.
Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. Economic

Record, Vol.88, No.s1, pp.2-9.

Cox, M., & Ellsworth, D. (1997, October). Application-controlled demand paging for out-of-core visualization. In Proceedings of the 8th conference on Visualization'97 (pp. 235-ff). IEEE Computer Society Press.

Demchenko, Y., Gruengard, E., & Klous, S. (2014). Instructional model for building effective Big Data curricula for online and campus education. In Cloud Computing Technology and Science (CloudCom), IEEE 6th International Conference (pp. 935-941)

Devlin, B., Rogers, S., Myers, J. (2012). EMA and 9sight Consulting Report, http://www.9sight.com/pdfs/Big_Data_Comes_of_Age.pdf

Diebold, F. (2012). A Personal Perspective on the Origin (s) and Development of'Big Data': The Phenomenon, the Term, and the Discipline, Second Version.

Douglas, K. (2015). Technologies for Data. Working paper of SAF05 project.

Ena, O., Mikova, N., Saritas, O., & Sokolova, A. (2016). A methodology for technology trend monitoring: the case of semantic technologies. Scientometrics, Vol.108, No.3, pp.1013-1041.

Fanelli, D. (2013). Any publicity is better than none: newspaper coverage increases citations, in the UK more than in Italy. Scientometrics, Vol.95, No.3, 1167-1177.

Fourt, L. A., & Woodlock, J. W. (1960). Early prediction of market success for new grocery products. The Journal of Marketing, Vol.25, No.2, pp.31-38.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, Vol.35, No.2, pp.137-144.

Gartner (2010) https://www.gartner.com/newsroom/id/1210613

Gartner (2011) https://www.gartner.com/newsroom/id/1454221

Gartner (2014). Survey analysis: Big Data Investment Grows but Deployments Remain Scarce in 2014.

Gossart, C. (2015). Rebound effects and ICT: a review of the literature. In ICT innovations for sustainability (pp. 435-448). Springer, Cham.

Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., & Kyriakidou, O. (2004). Diffusion of innovations in service organizations: systematic review and recommendations. The Milbank Quarterly, Vol.82, No.4, pp.581-629.

Greve, H. R. (2009). Bigger and safer: The diffusion of competitive advantage. Strategic Management Journal, Vol.30, No.1, pp.1-23.

Greve, H. R., & Seidel, M. D. L. (2015). The thin red line between success and failure: Path dependence in the diffusion of innovative production technologies. Strategic Management Journal, Vol.36, No.4, pp.475-496.

Guerrieri, P. A. O. L. O., & Padoan, P. C. (2007). Modelling ICT as a general purpose technology. Evaluation, https://pardee.du.edu/sites/default/files/ModelingICT.pdf

Helpman, E. (Ed.). (1998). General purpose technologies and economic growth. MIT press.

Hu, J., & Zhang, Y. (2017). Discovering the interdisciplinary nature of Big Data research through social network analysis and visualization. Scientometrics, Vol.112, No.1, pp.91-109.

IBM (2012). The real-world use of big data: How innovative enterprises extract value from uncertain data

Jewell, D., Barros, R. D., Diederichs, S., Duijvestijn, L. M., Hammersley, M., Hazra, A., ... & Portilla, I. (2014). Performance and capacity implications for big data. IBM Redbooks.

Kearns, M. J. (2010). Designing a digital future: Federally funded research and development in networking and information technology.

Kubacki, K., Rundle-Thiele, S., Pang, B., & Buyucek, N. (2015). Minimizing alcohol harm: A systematic social marketing review (2000 - 2014). Journal of Business Research, Vol.68, No.10, pp.2214-2222.

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. META Group Research Note, Vol.6, No.70.

Liao, H., Wang, B., Li, B., & Weyman-Jones, T. (2016). ICT as a general-purpose technology: The productivity of ICT in the United States revisited. Information Economics and Policy, Vol.36, pp.10-25.

Liu, N., An, H., Gao, X., Li, H., & Hao, X. (2016). Breaking news dissemination in the media via propagation behavior based on complex network theory. Physica A: Statistical Mechanics and its Applications, 453, pp.44-54.

McKinsey Global Institute (2011). Big data: The next frontier for innovation, competition, and productivity

Mashey, J. R. (1997, October). Big Data and the next wave of infraS-tress. In Computer Science Division Seminar, University of California, Berkeley.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. Harvard business review, Vol.90, No.10, 60-68.

Meade, N., & Islam, T. (2006). Modelling and forecasting the diffusion of innovation – A 25-year review. International Journal of forecasting, Vol.22, No.3, pp.519-545.

Narin, F. (1995). Patents as indicators for the evaluation of industrial research output. Scientometrics, Vol.34, No.3, pp.489-496.

National Information Agency (2014). Year of Information Society Static

O'Leary, D. E. (2013). BIG DATA', THE 'INTERNET OF THINGS'AND THE 'INTERNET OF SIGNS. Intelligent Systems in Accounting, Finance and Management, Vol.20, No.1, pp.53-65.

Pateli, A. G., & Giaglis, G. M. (2005). Technology innovation-induced business model change: a contingency approach. Journal of Organizational Change Management, Vol.18, No.2, pp.167-183.

Pradhan, R. P., Arvin, M. B., & Norman, N. R. (2015). The dynamics of information and communications technologies infrastructure, economic growth, and financial development: Evidence from Asian countries. Technology in Society, Vol.42, pp.135-149.

Preis, T., Moat, H. S., & Stanley, H. E. (2013). Quantifying trading behavior in financial markets using Google Trends. Scientific reports, 3, srep01684.

Rogers, E. M. (1976). New product adoption and diffusion. Journal of consumer Research, Vol.2, No.4, pp.290-301.

Seyedghorban, Z., Matanda, M. J., & LaPlaca, P. (2016). Advancing theory and knowledge in the business-to-business branding literature. Journal of Business Research, Vol.69, No.8, pp.2664-2677.

Sheng, J., Amankwah-Amoah, J., & Wang, X. (2017). A multidisciplinary perspective of big data in management research. International Journal of Production Economics, Vol.191, pp.97-112.

Shollo, A., & Galliers, R. D. (2016). Towards an understanding of the role of business intelligence systems in organisational knowing. Information Systems Journal, Vol.26, No.4, pp.339-367.

Taylor, L., Schroeder, R., & Meyer, E. (2014). Emerging practices and perspectives on Big Data analysis in economics: Bigger and better or more of the same?. Big Data & Society, Vol.1, No.2,

Thoma, G. (2008). Striving for a large market: evidence from a general purpose technology in action. Industrial and Corporate Change, Vol.18, No.1, pp.107-138.

Tilly, C. (1984). The old new social history and the new old social history. Review (Fernand Braudel Center), Vol.7, No.3, pp.363-406.

Wang, L., & Alexander, C. A. (2015). Big data in design and manufacturing engineering. American Journal of Engineering and Applied Sciences, 8, No.2, pp.223.

Watanabe, C., Naveed, K., & Neittaanmäki, P. (2015). Dependency on un-captured GDP as a source of resilience beyond economic value in countries with advanced ICT infrastructure: Similarities and disparities between Finland and Singapore. Technology in Society, Vol.42, pp.104-122.

Zhang, L., Xu, K., & Zhao, J. (2017). Sleeping beauties in meme diffusion. Scientometrics, Vol.112, No.1, pp.383-402.