韓國IT서비스學會誌
第15卷 第1號
2016年 3月, pp.97-111

Journal of Information Technology Services
http://dx.doi.org/10.9716/KITS.2016.15.1.097

빅데이터를 활용한 정책분석의 방법론적 함의 : 기회형 창업 관련 소셜 빅데이터 분석 사례를 중심으로

이영주* · 김도훈**

Methodological Implications of Employing Social Bigdata Analysis for Policy-Making : A Case of Social Media Buzz on the Startup Business

Young-Joo Lee* · Dhohoon Kim**

■ Abstract ■

In the creative economy paradigm, motivation of the opportunity based startup is a continuous concern to policy-makers. Recently, bigdata anlalytics challenge traditional methods by providing efficient ways to identify social trend and hidden issues in the public sector. In this study the authors introduce a case study using social bigdata analytics for conducting policy analysis. A semantic network analysis was employed using textual data from social media including online news, blog, and private bulletin board which create buzz on the startup business. Results indicates that each media has been forming different discourses regarding government's policy on the startup business. Furthermore, semantic network structures from private bulletin board reveal unexpected social burden that hiders opening a startup, which has not been found in the traditional survey nor experts interview. Based on these results, the authors found the feasibility of using social bigdata analysis for policy-making. Methodological and practical implications are discussed.

Keyword : Social Bigdata, Sementic Network Analysis, Startup, Policy Decsionmaking

Submitted : October 26, 2015 1st Revision : January 5, 2016 Accepted : January 7, 2016

* 한국정보화진흥원 수석연구원, 정보시스템학 박사, 교신저자
** (주)아르스프락시아 대표

1. 서 론

1.1 창조경제와 기회형 창업

박근혜 정부는 창조경제를 “국민의 창의성과 과학기술, 정보통신기술(ICT)의 융합을 통해 산업과 산업이 융합하고 산업과 문화가 융합해 새로운 부가가치를 창출함으로써 새로운 성장 동력과 일자리를 만들어내는 경제”로 정의하고 있다. 창조경제 정책이 특히 집중적으로 육성하고자 하는 분야는 기술기반의 기회형 창업 또는 혁신형 창업이다. 전 세계적으로 저성장의 기조가 자리 잡은 가운데 기술기반의 혁신형 창업은 저성장의 돌파구로 주목받으면서 창조경제 추진의 중요한 수단으로 자리매김 되어 있다.

기회형(혁신형) 창업은 창업자가 새로운 분야의 사업을 펼치기 위해 자발적으로 창업하는 것으로, 창업에 대한 부담이 적은 나라일수록 기회추구형 창업이 높은 편이다. 하지만 우리나라는 OECD 국가 중 ‘기회추구형 창업’ 비중이 가장 낮고, 창업소기업의 부가가치 창출 비중도 부진한 편이다. OECD가 파악한 우리나라 소기업의 부가가치 창출 비중은 2011년 기준 13.7%로 일본(3.98%), 스위스(13.62%)에 이어 세 번째로 낮은 수준이며, 소기업의 생산성도 근로자 1인당 3만 5,100달러로 아일랜드(2만 9,500달러)에 이어 두 번째로 낮은 수준이다.

한편 우리나라에서도 이미 기술기반 창업의 경제적 잠재력이 확인된 상태이다. 중소기업청과 창업진흥원의 「2013년 기준 창업기업실태조사」에 의하면, 2011년 기준 창업기업 수는 총 164만개로 2009년 156만개 대비 3년간 2.4% 증가하여, 전체 중소기업의 50.7%를 차지하고 있다. 또한 창업기업에 근무하는 “종사자 수”는 492만 명으로 전체 중소기업 종사자(‘11년 1,263만 명) 중 38.9%로 국내 일자리 창출 보고(寶庫) 역할을 하고 있다. 업종별로는 제조업이 14.2만 개(8.7%), 지식서비스업 23.9만 개(14.6%)로 기술기반 창업이 전체 창업기업의 23.3%(38.2만 개)에 달하고 있으며, 나머지 도·소매업은 47.8만 개(29.2%), 숙박·음식점업은 44.4만 개(27.1%) 수준이다. 특히 평균 종사자 수는 제조업 6.4명, 지식서비스업 4.6명으로 기술기반 창업이 도·소매업, 숙박·음식점업에 비해 고용창출 효과가 큰 것으로 나타났다. 매출액 또한 2011년 기준 제조업 매출액은 123조 원, 지식서비스업은 144조 원으로 기술기반 창업이 전체 창업기업 매출액의 29.0% 이상을 차지하는 등 기술기반 창업은 자영업 창업에 비해 부가가치 창출효과가 큰 것으로 나타났다.

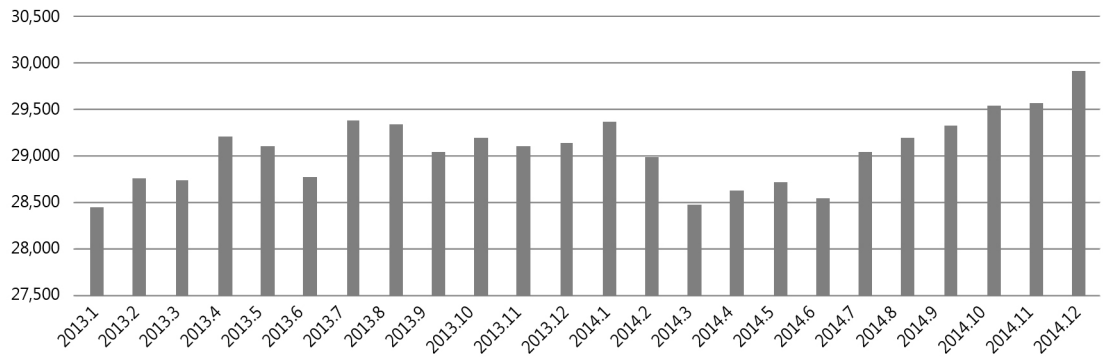
1.2 우리나라의 창업 지원정책 현안

우리나라의 기회형 창업에 대한 잠재적 이슈는 창업에 대한 최근 통계를 산출해 보아도 알 수 있다. 기회형 창업 현황에 대한 거의 유일한 통계인 벤처기업수의 변화를 보면 2014년 상반기에 급격하게 감소하였으나 하반기에 들어 다시 상승하고 있는데, 이러한 변화에 대한 추적과 정책적 시사점 분석이 필요한 상황임을 알 수 있다.[1)]

2015년에 들어서도 창조경제의 핵심수단인 기회형 창업의 활성화가 기대되고 있는 가운데 박근혜 정부에서는 2015년에도 창업 활성화를 위한 정책 지원을 아끼지 않고 있다. 2015년 정부의 벤처·창업지원 예산은 6개 부처, 32개 세부사업으로 2014년 기준 2조 1,661억 원, 2015년 예산안 기준으로는 1조 9,115억 원 수준[2)]에 달한다. 하지만 사업 유형별로 보면, 예산규모로는 자금지원(금융) 사업의 비중이 2015년 예산안 기준 약 79%로 가장 높으며, 자금지원 사업을 제외한 인프라(교육, 공간 등), 인식조성 사업의 예산비중은 낮은 편으로, 직접적 자금지원에 우선순위가 매우 높은 편이다.

창업에 대한 직접적인 자금지원 뿐 아니라 창업에 대한 긍정적인 사회적 여건 조성은 이전부터 중요한 요인으로 지적되어 왔다. 과학기술혁신 관련 기존 연구에서도 벤처형 중소기업이나 기술창업 지원시 선별적이고 직접적인 지원 보다는 잠재력이

1) 한국벤처기업협회 벤처통계시스템 통계자료 재가공.
2) 국회예산정책처, 벤처·창업 지원 정책의 주요 쟁점과 개선과제, 2014. 11.



〈Figure 1〉 No. of Monthly Venture Capital Startups

높은 신규진입자(Potential Entrants)를 포함하여 동기를 부여할 수 있도록 정보제공, 창업인큐베이팅 프로그램 등 간접적 지원을 통한 환경조성의 중요함을 지적해 왔다(Shin and Song, 1998; Kim, 2007).

따라서 기회형 창업 진흥을 위한 대규모 예산을 투자함에 있어 투자할 예산의 비중이나 투입 방법 등 포트폴리오 기반 정책 수립을 위해서는 현재 기회형 창업 여건에 대한 정확한 현황 분석이 필요하다. 그러나 현재까지는 기회형 창업의 사회적 여건에 대해 파악하거나 예측할 수 있는 통계나 지표가 부재하고, 관련 정책을 수립하는 데 있어 현장의 이슈를 파악하기는 데에는 대규모 설문조사나 전문가 의견수렴의 방법이 거의 유일한 것이 현실이다.

1.3 빅데이터 분석의 기회

빅데이터는 그 양과 다양한 형태, 처리의 복잡성으로 인해 기존의 정량적 분석 방법으로는 수집, 저장, 분석, 시각화가 어려운 특성을 가지고 있다(Mckinsey, 2011). 빅데이터 처리를 위한 컴퓨팅 기술이 비약적으로 발전하면서, 민간부문에서는 대용량 정형데이터 분석 뿐 아니라 뉴스 등 소셜미디어에 대한 비정형데이터분석을 통해 마케팅, 평판관리 등 경영 의사결정에 적극적으로 활용하고 있다(Song, 2013; Kim et al., 2014; Kim et al., 2012).

특히 트위터, 페이스북 등 소셜 네트워크 서비스(SNS)에 드러나는 이슈 분석에는 네트워크에 참여하는 개별 메시지에 드러나는 단어 간의 관계성을 분석하는 의미연결망 분석(semantic network analysis) 기법이 사용되고 있다(Choi and Kweon, 2014). 의미연결망 분석은 사회의 시스템 구조를 파악할 때 행위자 사이의 관계를 중시하는 사회연결망 분석(social network analysis)을 커뮤니케이션 메시지에 적용시킨 방법이다(Wang and Rada, 1998; Carley, 1993). 사회연결망 분석이 사회현상을 만들어내는 원동력을 각각의 개체가 아니라 개별 행위자가 연결되어 있는 관계성에 주목한다면 이를 원용한 의미연결망 분석은 행위자 대신 메시지의 구성요소인 단어들의 구조적 관계를 통해 의미를 분석한다(Kim and Kim, 2015; Sudhahar et al., 2015; Wang and Rada, 1998; Wasserman and Faust, 1994).

사회 구성원들 간의 담론을 파악하는 전통적 방법이 노동집약적이고 연구자의 주관적 해석을 배제할 수 없다는 한계가 있다면. 의미연결망 분석은 대용량의 비정형데이터를 컴퓨터 프로그래밍 기반으로 분석하기 때문에, 이러한 단점을 보완할 수 있다. 나아가 개별 단어의 고정된 속성이나 Park and Leydesdorff(2004), Jang(2001) 개념이 아니라 상호 작용하는 관계의 맥락 속에서 역동적인 의미를 포착할 수 있는 장점이 있다(Kim, 2011).

한편 공공정책에 빅데이터를 적용한 국내연구는 아직까지 대부분 빅데이터 분석 시스템 구축방안이나 거버넌스에 관한 연구가 대부분으로, 공공분야의 정책의사결정에 빅데이터 분석을 적용한 연구는 초기 수준이다(Kim et al., 2015). 따라서 본 연구에서는 빅데이터를 활용한 정책분석의 방법론

적 타당성과 함의를 탐색해보는 사례연구의 일환으로 창업에 대한 소셜 빅데이터를 대상으로 의미연결망 분석을 통해 기회형 창업에 대한 사회적 여건을 분석해보고 시사점을 도출해 보고자 한다.

2. 분석방법 및 절차

2.1 분석 대상 자료 수집

본 연구의 분석 대상이 되는 비정형데이터는 소셜 미디어에 나타난 기회형 창업에 대한 온라인 버즈(Buzz)이다. 수집 범위로는 주요 언론의 온라인 기사, 개방형 블로그, 그리고 민간의 대표적 창업문의·상담사이트를 선정하였다. 민간의 사이트를 선정하기 위해 DC인사이드의 창업갤러리, 창업 114 등 3개 사이트를 후보로 선정하고, 게시글의 양과 내용이 충분히 연구 취지에 맞는지를 검토하여 최종적으로 DC인사이드의 창업갤러리를 결정하였다. 실제로 창업갤러리 게시판에서는 사용자들이 현장에서 마주치는 문제에 대해 상호 소통하며 실제적인 담론을 형성하고 있는 것으로 파악되었다. 따라서 소셜미디어 분석에서 주로 사용되는 소셜 네트워크 서비스(SNS)인 페이스북(Facebook)과 트위터(Tweeter)보다 대표성이 높을 것으로 판단하였다.

분석 기간은 최근 3년(2012. 1월~2014. 10월)으로 설정하였고, 웹크롤링 방식으로 각 미디어에서 창업 관련 원문을 추출하기 위한 검색어를 설정하여 수집하였다.

수집을 위한 검색어는 창업의 유형을 분류하는 일반적인 분류기준인 1인 창업, 벤처창업, 프랜차이즈, 소상공업을 1차 검색키워드로 사용하였고, 수집된 멘션들의 판별타당성을 확보하기 위해 43개 세부 검색어와 제외를 사용하였다. 데이터 수집 및 정제에는 약 1개월(2014. 11. 1.~11. 28)이 소요되었으며 사전 필터를 통해 상업적 광고 또는 단순 제품홍보 관련 멘션은 제거하였고, 사후 노이즈 제거를 위해 1차 추출 후 샘플링 기반 원문검토를 통해 분석목적과 무관한 멘션을 제거하였다.

수집 및 정제 결과 최근 3년 창업 관련 기사와 블로그 포스팅 본문 129,200건과 DC인사이드 게시물 1,437건을 확정하였고, 수집한 텍스트데이터는 원문과 함께 메타데이터로 생성일, 유형(기사/블로그/DC인사이드), 댓글수, 제목 등을 생성하였다.

2.2 분석 방법 및 절차

분석에 사용된 주요 방법은 각 미디어 별 버즈량 기초 통계(빈도, 점유율 등) 분석과 의미연결망 분석(Semantic Network Analysis), 감성분석(Sentiment Analysis) 등이다. 각 분석 결과는 정량화 후 기존 통계지표와의 추세 비교 등을 포함한다.

특히 소셜미디어에서 표상(represent)되는 창업에 대한 사회적 프레임의 특징을 분석하기 위해 의미망 분석을 활용하였다. 사회적 프레임은 특정 사회적 카테고리에 속한 행위자들의 논리적 사고에 선행하는 전형적으로 구조화된 믿음체계(Lakoff and Johnson, 1980) 또는 해석의 스키마(Goffman, 1974)로서, 당사자들이 이슈, 상황, 관계, 상호작용 등을 정의하고 해석하며 표상하는 방식이다(Dewulf et al., 2009). 다시 말해, 각 행위자들이 가지는 특정 이슈에 관한 판단과 사고가 바탕하고 있는 전제나 근거를 일컫는다고 할 수 있다. 이와 같은 프레임을 통해 개인과 집단은 “문제의 핵심이 무엇이며 어떻게 해결되어야 하는가”를 규정할 수 있다.[3)]

프레임 구조의 구체적인 특징을 살펴보기 위해서는 의미망 분석의 방법이 매우 효과적이다. 의미망 분석법은 텍스트에 대한 기존의 질적 분석과 양적 분석인 텍스트 마이닝 기법을 혼합한 것으로

3) 이러한 프레임은 다음의 세 가지 전형적인 속성(Baden, 2010)을 갖는다. (1) 특정 대상 및 이슈에 관한 상(image)의 형성에 있어 프레임은 준거집단이 사용하는 언어적 컨셉과 선택적 친화성을 갖는다. (2) 프레임에는 이런 언어적 컨셉들을 매개하고 통제하는 중심적 컨셉이 존재한다. 이러한 중심적 컨셉을 추종하는 의미상의 일관성이 핵심적인 의미체계를 형성한다. (3) 이렇게 형성된 특정 집단의 의미체계는 사회적 이슈에 대한 논쟁에 있어 전형적인 인과성(causality), 가치판단 및 대안의 준거를 제공한다.

자동화 텍스트 분석과 의미망 분석 기술을 순차적으로 적용한 것이다. 의미망분석은 “문제의 핵심이 무엇이며 어떻게 해결되어야 하는가?”에 관한 각 발화자의 인식의 프레임을 포착하며, 이를 통하여 쟁점이 되는 사안들에 대한 인식 차의 원인과 현상, 전망을 드러낼 수 있도록 하는데 유용한 분석방법이다(Jo et al., 2015). 특히 각 행위주체들이 언급한 키워드들의 연결망 속에서 핵심적인 이슈 프레임을 발굴하기 위해 의미망 분석 방법을 활용하여 기존의 정성(예 : 인터뷰), 혹은 정량(예 : 객관식 설문) 한쪽에 치우친 텍스트 분석 기법에 비해 심화된 수준의 통합적 분석을 가능케 해준다.

의미연결망 분석은 특정 주제를 가정하지 않은 상태에서 단어들의 빈도와 한 문장 안에서 동시에 사용되는 단어들의 관계를 통해 메시지들의 공통된 의미화 패턴을 추출하는 작업이다. 이때, 분석의 단위가 되는 단어를 사회연결망 상의 노드(Node)로 설정하고 노드 간 연결 상태를 링크(link)로 나타내어 링크로 드러나는 단어들의 공동출현(co-occurrence) 관계를 통해 의미들의 구조를 해석한다(Wang and Rada, 1998). 단어들의 사용빈도와 관계는 전체 버즈에서 에서 강조되는 상징성을 보여줄 수 있고 단어들의 결합을 통해 특정한 의미를 만들어내는 경향을 파악할 수 있다(Park and Leydesdorff, 2004).

의미망 분석의 전 단계에서 진행한 자동화 텍스트 분석은 원자료를 분석 가능한 자료로 변환하기 위한 작업이라 할 수 있다. 이는 빅데이터 분석에 활용되는 자연어 처리(Natural Language Processing : NLP) 기술을 활용한 것으로서, 한국어의 품사, 문법구조, 언어습관을 인지하는 데이터베이스 및 인공지능 알고리즘 설계에 의한 자연어 처리 기술을 통해 대용량의 텍스트 데이터에서 주요 형태소 혹은 단어를 추출하고 그 빈도수를 분석한다. 이후에 진행된 의미망 분석은 텍스트 내에서 동시에, 그리고 선후에 출현하는 단어 중 통계적으로 유의미한 관계를 가지는 단어들을 추출해 낸다.

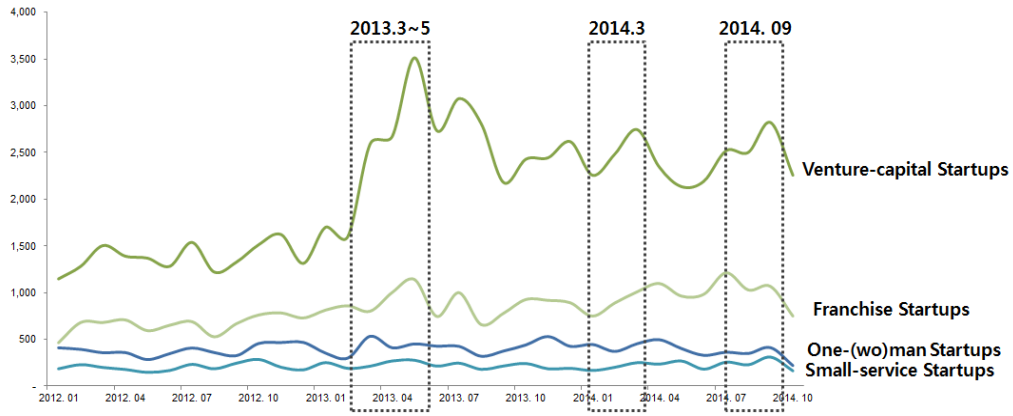
본 연구에서는 의미망분석이 창업과 관련된 온라인 미디어에 드러나는 버즈에 나타나는 주요 키워드들의 사용빈도와 주요 키워드들 간의 의미적 연관구조를 파악함으로써, 사회적으로 형성된 창업 여건과 이슈를 발굴하는 데 사용되었다.

아울러 감성분석(Sentiment Analysis)은 자연어에 대해 기계학습을 바탕으로 문장의 긍정, 부정과 같은 문맥 정보를 추출하거나 분류하는 기법이다. 본 연구에서는 기존 연구에서 주로 사용하는 감성 사전을 기반으로 분석을 수행하였다(Kim et al., 2012; Kim et al., 2014; Jo et al., 2015). 실제로 수집된 문장들에 대해 사전에 정의된 감성사전을 바탕으로 비교를 통해 긍/부정의 극성을 정의하였고, 개별 문장의 극성을 미디어별, 주제별로 합산하여 비율을 계산하였다.

자료수집 및 빅데이터 분석에 필요한 알고리즘 및 분석인프라는 빅데이터 기반 컨설팅 기업인 아르스프락시아(구 트리움)의 검색엔진 및 의미망 분석도구 옵티마인드(Optimind v2.0)가 사용되었다.

옵티마인드는 대용량의 텍스트 데이터에 대한 의미연결망 분석과 감성분석을 자동화한 아르스프락시아의 상업용 빅데이터 분석 시스템이다(www.arspraxia.com). 카네기멜론 대학에서 개발된 의미망 분석 툴인 오토맵(CASOS, 2007) 등 학술용 분석 도구가 비교적 적은 양의 데이터에 대해 반자동화(semi-automated) 된 코딩 기법을 활용하여 의미망을 추출한다면, 옵티마인드는 테라바이트 단위의 대용량 텍스트에 대해 사람의 개입 없이 완전히 자동화된 방식으로 코딩과 분석을 가능케 해준다.

본 연구에서 사용한 텍스트 데이터의 의미망 및 감성 분석의 절차는 다음과 같다. 첫째, 블로그 등 특정 매체의 데이터를 크롤링하여 포스팅 수를 산출한다. 둘째, 자연어처리를 통해 추출된 동시출현하는 명사어 중 통계적으로 유의미한 관계들을 연결하여 복잡계 형태의 인지지도를 만들고, 핵심적인 의미망을 그래픽으로 표현한다. 셋째, 추출된 핵심 단어들을 주제별로 카테고리화 하여, 연도별로 사회의 주요 관심과 이슈가 어떻게 변해 왔는지를 조망한다. 넷째, 주제별로 선별된 텍스트 샘플에 대



〈Figure 2〉 Trend in Online News Articles Regarding Startups

한 감성분석을 통해, 긍·부정 감성의 비율을 계산하고 외부 데이터와의 통계적 상관성을 검증한다.

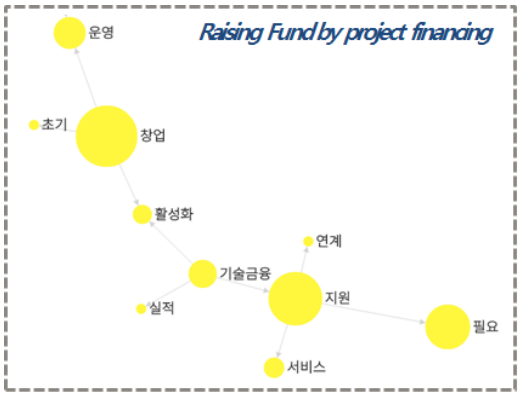
3. 분석 결과

3.1 온라인 뉴스에 나타난 기회형 창업에 대한 여론 동향

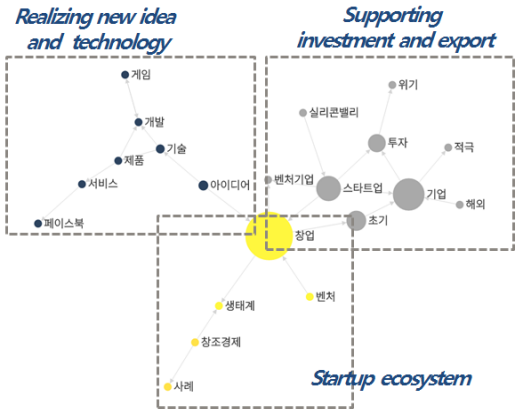
창업에 대한 온라인뉴스의 보도량은 박근혜 정부 출범 후 눈에 띄게 증가하였다. 특히 새정부가 출범한 후 창조경제 실현계획과 차년도 실현계획이 발표되는 시점에 전체적인 언급량이 증가하고 있었다. 창업유형별 언급량은 벤처창업 > 프랜차이즈 > 1인 창조기업 > 소상공업 순으로, 전반적으로는 1인 창업보다는 벤처창업에 대한 언론의 관심이 높은 것으로 파악된다.

뉴스기사의 의미연결망 분석을 통해 나타난 기회형 창업에 대한 주요 키워드와 의미구조는 다음과 같다.

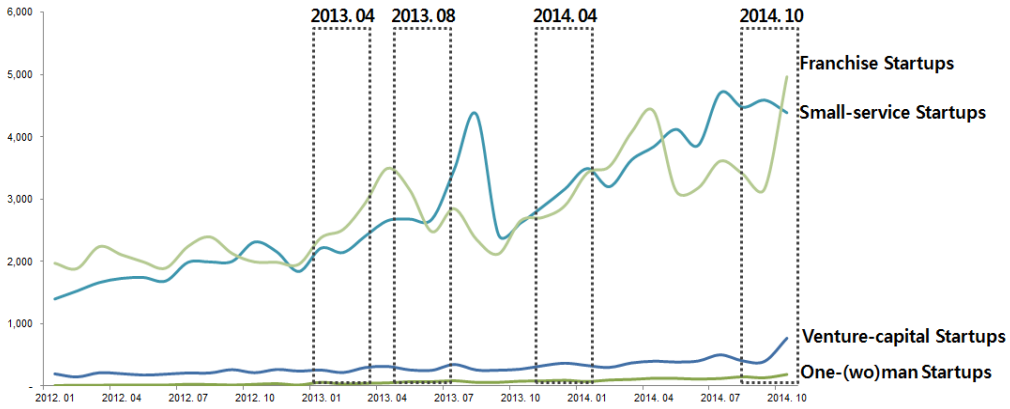
먼저 1인 창업에 대한 주요 정책키워드로는 기술금융, 연계, 실적, 지원, 서비스 등으로, 다양한 기관과 연계한 기술금융을 통한 자금지원이 창업 초기의 안정적 운영과 창업 활성화의 중요한 수단으로 자주 언급되고 있으며, 박근혜 정부 출범 초기에는 사회적 기반 형성과 지원정책의 실적 부족에 대한 우려도 제기되고 있었다.



〈Figure 3〉 Semantic Networks in Online News Articles Regarding One-(wo)man Startups



〈Figure 4〉 Semantic Networks in Online News Articles Regarding Venture-Capital Startups



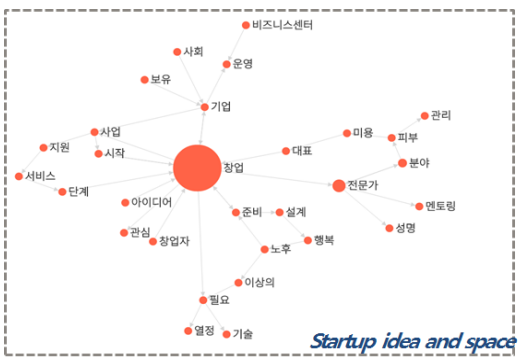
〈Figure 5〉 Trend in Blog Posts Regarding Startups

다음으로 벤처창업과 관련한 주요 정책키워드로 아이디어, 기술, 서비스, 해외, 실리콘벨리, 생태계, 사례 등이 도출되었으며, 아이디어와 기술의 제품화/서비스화 방법, 투자 및 해외진출에 대한 사례와 지원방안 등 벤처창업의 활성화에 대한 정부의 지원정책과 창업생태 조성을 통한 창조경제 실현 등 벤처창업의 긍정적 가치에 대하여 보도하고 있었다.

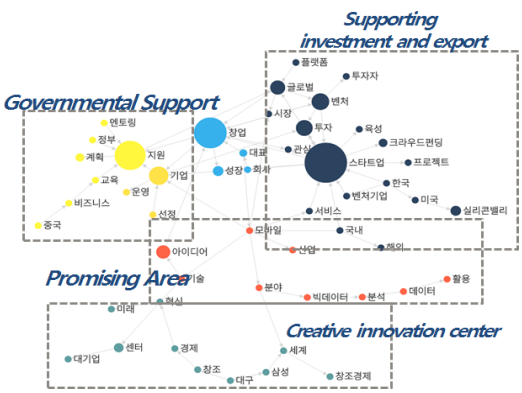
3.2 개방형 블로그에 나타난 기회형 창업에 대한 포스팅 동향

블로그는 미디어의 특성상 창업에 대한 정보제공에 관한 포스트가 다수를 이루었다. 전반적인 포트스량 증가 추세는 인터넷 뉴스와 마찬가지로 2013년 이후 눈에 띄게 증가하는 추세를 보이고 있다. 다만 유형별로는 프랜차이즈 > 소상공업 > 벤처창업 > 1인 창조기업 순으로 생계형 창업에 대한 정보제공성 포스트 비중이 높다.

블로그에 나타난 1인 창업 관련 담론은 프랜차이즈 및 소상공창업과 크게 다르지 않았다. 각 유형 모두 다양한 창업아이템과 정부의 지원 내용에 대한 소개가 주를 이루고 있었다. 벤처창업에 대해서는 유망분야에 대한 다양한 소개와 더불어, 투자유치 및 해외시장 진출에 관한 다양한 사례, 정부지원 및 멘토링 서비스에 대한 소개, 각지의 창조경제혁신센터에 대한 소개 등 최전 정부정책에 부응하는 내용이 주를 이루고 있었다. 개방형

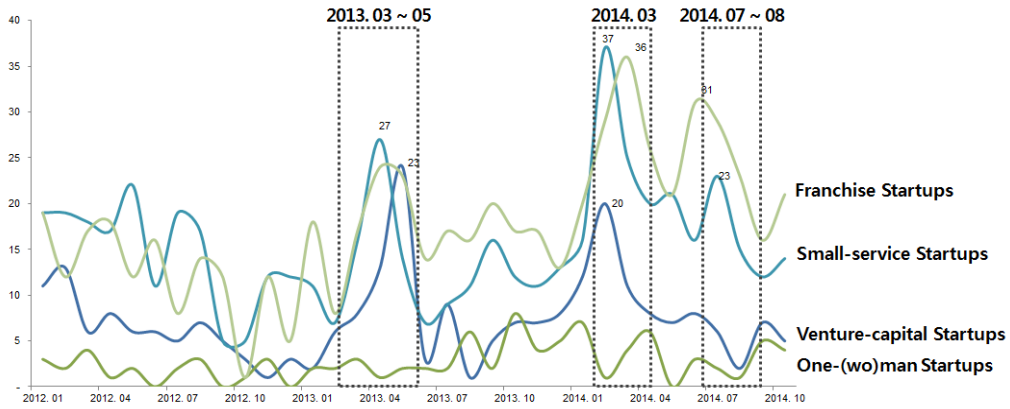


〈Figure 6〉 Semantic Networks in Blogs Regarding One-(wo)man Startups



〈Figure 7〉 Semantic Networks in Blogs Regarding Venture-Capital Startups

블로그의 특성상 정부 정책의 홍보와 확산의 통로 역할을 수행하고 있음을 알 수 있다.



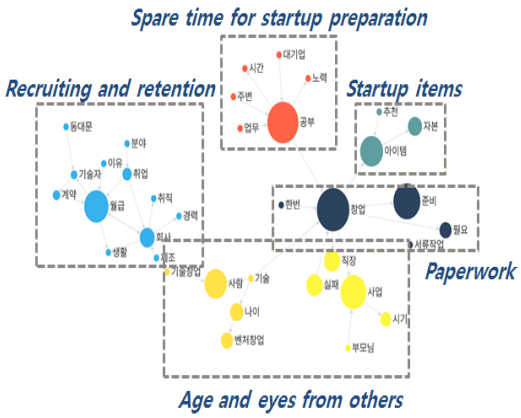
〈Figure 8〉 Trend in Startup-Bulletin Board

3.3 DC인사이드에 나타난 창업에 대한 포스팅 동향

DC인사이드의 창업갤러리 게시판에서 게시되었던 창업 관련 상담 및 문의 내용에 대한 분석 결과는 다른 미디어와는 확연히 다른 특징을 보이고 있다. 특히 온라인 뉴스나 블로그에 비해 정부정책과 연관된 내용 보다는 실제 현장에서 발생하는 다양한 문제와 애로사항, 불만사항 등을 파악할 수 있었다. 창업 유형별 언급량을 보면 전체적으로 변동폭이 매우 불안하지만 정부의 특정 정책 이벤트와는 무관하게 점유율 순위는 비교적 일관된 결과를 보이고 있다. 언급 비중도 역시 미디어 특성상 프랜차이즈 > 소상공업 > 1인 창조기업 > 벤처창업 순으로 기회형 창업에 대한 언급 비중이 현저히 낮다. 실제 민생경제 현장에서는 정부가 적극적으로 펼치고 있는 기회형 창업에 대한 저변이 확산되어 있지 않다고 판단할 수 있는 일부 실마리를 제공하고 있다.

DC인사이드에 게시된 기회형 창업 관련 상담·문의글에 대한 키워드 및 의미망 분석 결과는 다음과 같다.

먼저 1인 창업에 대해서는 전반적으로 창업 아이템에 관한 다양한 의견과 함께 창업에 대한 고충의 유형을 파악할 수 있었다. 주요 정책키워드 그룹으로 파악한 고충의 유형으로는 기술자, 월급, 계약, 경력 등으로 구성되는 '구인 및 유지의 어려움', 나이, 실패, 시기, 부모님 등으로 구성되는 '나이와 주변 시선에 대한 부담감', 준비, 필요, 서류작업, 한번 등으로 구성되는 '창업에 필요한 서류작업에 대한 부담감', 시간, 업무, 노력, 공부 등으로 구성되는 '창업 준비과정의 부담감' 등이 파악되었다. 특히 창업을 준비하는 직장인들이 대부분 일과 창업준비를 병행하고 있는데 이에 따른 어려움과 주변의 시선, 그리고 정부의 규제나 법적 절차가 창업에 주는 부담감이 여전히 상당한 수준임을 알 수 있었다.



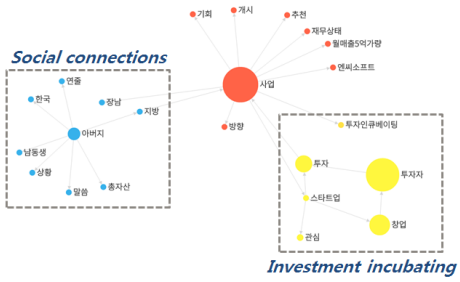
〈Figure 9〉 Semantic Networks in Startup-Bulletin Board Regarding One-(wo)man Startups

다음으로 벤처창업에 대한 상담·문의글에 대한 키워드 그룹과 의미망 분석 결과로는 먼저 스타트업, 투자자, 인큐베이팅, 관심, 창업 등으로 구성되는 의미망으로 기존의 벤처라는 용어가 '스타트업'으로 전환되면서 성공한 벤처기업에 대한 다양한 논의와 더불어 스타트업 투자를 확보하는 방안에 대한 관심이 형성되었고, 다음으로 연줄, 장남, 아버지, 총자산 등으로 구성되는 의미구조는 연줄에 의지하는 우리사회의 한계에 대한 비판적인 내용이 언급되고 있었다. 1인 창업과 벤처창업에 대한 고충과 관련된 게시글 중 대표적인 사례는 다음과 같다.

(창업준비 관련) "아버지 일 물려받던가, 노하우 배워 이쪽일대에 법인체 한번 내보려 현재 업무와 공부를 병행하고 있다… (중략)…일단 니가 하고픈 분야에 취업을 해라. 꼭 하고픈 분야가 아니더라도 취업을 하던가 뭐가 됐든 고등 기술이나 교육을 들어봐."

(연줄에 의지) "체인점창업이라면 상관없지만 저희아버지나 저 같은 경우는 연줄이 많이 필요합니다. 연줄을 만들기 위해선…없네요 뭐. 인맥을 최대한 활용하는 게 중요."

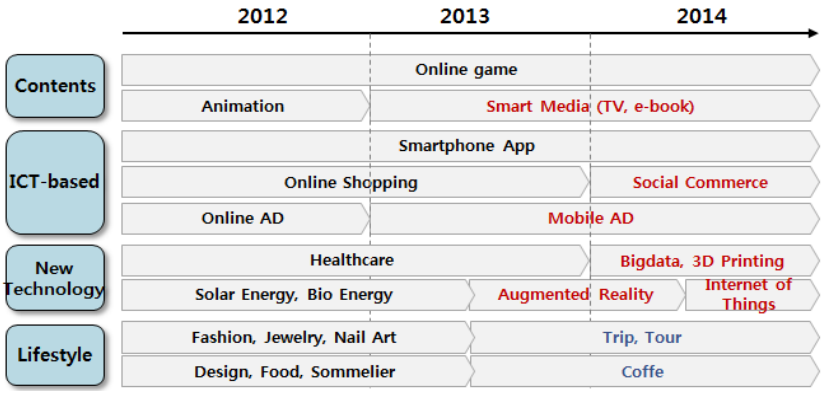
(학연) "기술창업은 말할 것도 없고 우리나라에서 사업으로 성공 했다는 사람들은 죄다 sky야 … (중략)… 하다못해 백종원처럼 그냥 쌩요식업으로 상공한 사람도 연세대라는거,… 결국 사업도 학벌 따라 가는 경우가 많더라…"



〈Figure 10〉 Semantic Networks in Startup-bulletin Board Regarding Venture-Capital Startups

3.4 기회형 창업에 대한 관심 분야의 연도별 추이

창업 유망분야 도출을 위해 수집한 인터넷기사, 블로그, DC인사이드 창업갤러리 게시글에 기회형 창업유형인 1인 창업 및 벤처창업의 검색키워드와 유의미한 빈도와 관계를 보여주는 연관키워드를 연도별, 유형별로 각각 100개씩 추출하였다. 도출된 키워드를 대상으로 창업 분야를 콘텐츠창업, ICT 기반창업, 신기술기반창업, 문화생활창업 등 크게 4개 분야로 대분류를 생성하였고, 다시 대분류 안에서 유의미한 유형을 보이는 중분류 카테고리를 생성하였다. 분류분석 결과 <Figure 11>과 같이 관심을 보이는 창업 분야가 연도별로 변화하는 양상을 파악할 수 있었다.



〈Figure 11〉 Trend in Social Media Keywords of Opportunity-based Startups

먼저 콘텐츠 분야는 게임 관련 창업에 대한 지속적인 관심과 함께, SNS에 사용하는 캐릭터나 애니메이션 관련 창업에서 TV나 전자책 등 스마트 미디어플랫폼을 기반으로 콘텐츠를 제작/유통하는 창업에 대한 관심으로 이동 중에 있음을 확인하였다. 콘텐츠 유통을 위한 디바이스간 융합과 차별화가 각각 발전하면서도 상호 영향을 주고받을 것으로 보이지만, 이의 기반에는 스마트미디어 플랫폼이 공통기반으로 확산될 것으로 예상된다.

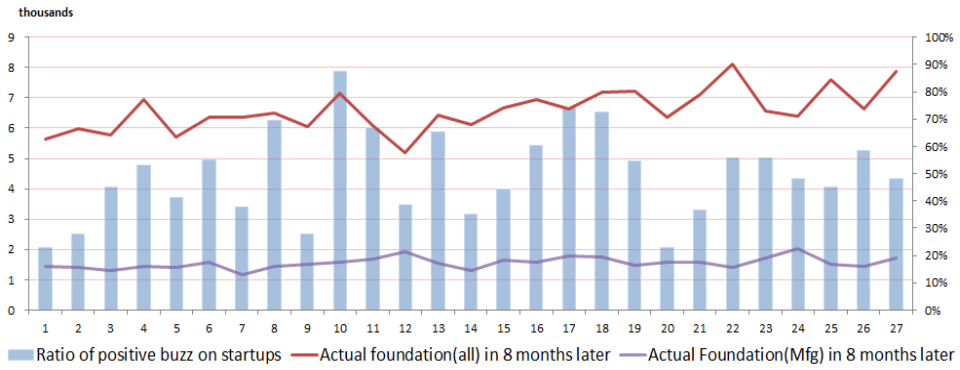
ICT 기반 창업은 인터넷과 모바일을 통한 비즈니스 모델을 개발하고 연계하는 데 관심이 높았다. 스마트폰 앱 개발에 관한 관심은 지속적으로 유지되는 한편, 온라인 쇼핑몰은 SNS와 융합된 소셜커머스로 변화 중이며, 인터넷 광고에 대한 관심 또한 온라인에서 모바일로 옮겨가고 있음을 확인하였다. 모바일 전자상거래 관련 기술이 진화하는 추세에 추어 모바일 기반 창업 또한 유망할 것으로 예상된다.

신기술 기반 창업 분야는 주로 벤처창업 형태로 2012년에는 헬스케어, 태양광, 바이오등에 관심이 높았다가 증강현실, 사물인터넷 등 새로운 유망기술로 관심이 이동하였다. 최근의 유망기술은 ICT와 과학기술의 융합이 일어나는 영역으로 기술 간 융합과 이로 인한 새로운 비즈니스모델 발굴은 지속적으로 확산될 것으로 보인다.

생활 분야는 문화콘텐츠에 대한 최근의 트렌드에 부응하여 Kpop이나 여행/관광과 접목한 창업에 대한 관심이 높아졌다. 이는 여가활동에 대한 욕구가 증가하는 사회적 여건과 그 맥락을 같이 하고 있다.

3.5 사회적 창업여건 변화와 실제 창업과의 관계

소셜미디어에 나타난 창업에 대한 긍정적 또는 부정적 인식이 실제 창업과 연관이 있는지를 알아보기 위해 감성분석 결과와 통계치를 연결하여 분석해 보았다. 먼저 수집한 창업관련 기사/블로그포스트/DC인사이드 게시물에 대한 감성분석 결과에 따라 긍정적 감성비율을 산출하고, 각 미디어 유형별 버즈량에 따른 가중평균값을 계산하였다. 이에 따라 2012년 1월부터 2014년 3월까지 총 27개월 동안의 월별 긍정적 감성비율(%)을 산출하였다. 다음으로 통계청에서 제공하는 신설법인 수 통계자료를 활용하여 월별 신설법인수를 얻었다. 최근 중소기업청과 창업진흥원에서 실시한 2013년 기준 '창업기업 실태조사' 결과에 따르면 창업에 대한 준비를 시작하고 실제 창업에 이르는 기간이 평균 8.7개월로 나타났는데, 본 분석에서는 이를 반영하여 분석 기간 내의 특정 월의 긍정적 감성비율(%) 값과 8개월 후의 실제 신설법인수를 매핑하였다. 따라서 최종적으로 2012년 1월~2014년 3월까지의 소셜미디어의 긍정감성 비율과 2012년 9월~2014년 11월의 신설법인 수가 매핑된 27개의 월별 통계치를 구하였다.



〈Figure 12〉 Comparison of Trends between Media Sentiments and Actual Foundations in 8 Months

분석 결과는 <Figure 12>와 같이 창업 관련 긍정적인 여론 비율(%)의 증감은 8개월 후 실제 창업(통계청 신설법인 수)의 증감과 대략적으로 연동됨을 확인할 수 있었다. 그러나 표본의 수가 작고 피어슨 상관계수 또한 0.13으로 상대적으로 낮아 통계적 유의성은 확인할 수 없었다.

4. 논 의

4.1 정책적 함의

본 연구에서는 창조경제로의 패러다임 전환 시기에 창조경제 달성의 가장 중요한 원동력인 창업, 특히 기회형 창업의 사회적 여건에 초점을 맞추었다. 소셜 빅데이터 분석 결과 기회형 창업 중 벤처 창업에 대해서는 온라인 기사와 블로그에서 비중있게 관심을 가지고 있으나, 1인 창조기업에 대한 관심은 상대적으로 적은 것으로 나타났다. 벤처창업보다 1인 창조기업에 대한 인터넷기사의 비중이 낮은 것은 정부 정책의 우선순위가 벤처창업에 더 쏠려 있음을 대변하고 있어, 일반 국민의 고충을 해소하기 위한 더 많은 정책적 관심이 요구된다. 이러한 결과는 여전히 대기업과 중소기업의 동반성장을 할 수 있는 사회적 여건이 미흡하다는 기존 연구 결과와 유사하다(Lee et al., 2013).

블로그와 DC인사이드 창업갤러리의 분석 결과를 보면, 일반 국민들은 전반적인 경기침체와 일자리부족 등으로 기회형 창업보다는 소상공업이나 프랜차이즈 등 생계형 창업으로만 관심이 쏠려 있음을 확인할 수 있었다. 창업 준비생들은 1인 창업이나 벤처창업 등 기회형 창업은 준비하는 과정이 어렵고 인맥 등 선천적으로 주어진 여건이 있는 사람들만 하는 것으로 인식하여, 상대적으로 쉽게 추진할 수 있는 생계형 창업에 더 관심이 높은 것으로 보인다. 창업 여건에 대한 사회적 여건 조성을 위해서는 단순히 제도적 지원 환경 뿐 아니라 이면에 깔려있는 사회적 인식에 대한 면밀한 검토가 필요함을 알 수 있다. 누구나 창업에 지원하고, 직장에서의 경험이 창업 도전으로 연계될 수 있는 지원 프로그램의 개발 등이 정책적 대안으로 검토되어야 할 것이다.

4.2 이론적 함의

본 연구의 분석을 통해 발견한 이론적 시사점은 다음과 같다.

첫째로 기존의 창업 정책 연구에서는 창업 여건 또는 창업 준비자에 대한 설문조사 중심의 조사연구가 다수를 이루고 있었다(Lee et al., 2013; Yook and Jeon, 2014). 그러나 빅데이터 분석 기법과 분석 기반기술의 발전에 따라 소셜미디어에서 형성되는 사회 구성원들의 인식과 행위에 대한 비정형분석이 새로운 방법론으로 활용될 수 있다. 본 연구에서도 창업 지망생의 고충을 파악하는 데 있어 기존 문헌과 이론에 바탕을 둔 사전 설문 설계를 통한 분석이 갖는 한계를 극복할 수 있음을 시사한다. 중소기업청에서 실시한 유사한 설문조사 결과와 소셜미디어 빅데이터 분석에 의한 결과가 <Table 1>에 정리되어 있다. 결과를 보면 기존의 설문조사에서 발견된 고충도 유사하게 나타난 한편, 미처 발견하지 못했던 고충 또한 발견되었다. 기존에는 대규모 설문의 한계를 극복하기 위해 참여연구나 포커스그룹 인터뷰를 사용해 왔으나, 관찰 대상의 대표성을 확보하는 데 어려움이 있었다. 하지만 빅데이터 분석 기법을 잘 활용한다면 기존의 조사 방법론을 보완할 수 있을 뿐 아니라 대표성을 확보하면서도 비용효율적인 방법으로 새로운 분석 결과를 얻을 수 있다.

두 번째로 본 연구에서는 빅데이터 분석을 통해 대중의 관심 변화에 대한 거시적 트렌드 분석의 가능성을 확인하였다. 사회적 관심에 대한 트렌드와 ICT 기술 발전 양상, 거시적 사회변화를 지속적으로 추적한다면 미래의 창업 유망분야에 대한 예측 또한 가능할 것으로 보인다. 이러한 분석 결과는 향후 정부의 창업 지원 정책을 수립하는 데 충분히 활용될 수 있을 것이다.

〈Table 1〉 Comparison of Keywords ranks between Social Media and Surveys

Rank	Social Media		Survey (as a whole)
	Opportunity based startup	Livelihood based startup	
1	Recruiting	Securing initial fund	Securing initial fund(52.7%)
2	Business idea	Profitability	Fear of failure(32.0%)
3	Securing initial fund	Recruiting	Maintaining livelihood(28.0%)
4	Fear of failure	Business idea	Lack of information, skill and experience(18.8%)
5	Spare time for preparation	Stable operation	Negative social norms(2.1%)
6	Social connections	-	Negative peer opinion(1.1%)
7	paperwork	-	Limiting previous job security(0.3%)

세 번째로 빅데이터의 정성적 분석과 기존 정량적 통계자료를 결합한 분석을 통한 예측모형의 초기 가능성을 확인하였다. 이는 주가예측(Kim et al., 2014; Kim et al., 2012) 등 민간에서 적극적으로 시도되고 있는 빅데이터 기반 미래 예측이 공공부문에도 가능함을 시사한다(Song and Song, 2014). 본 분석에서는 시계열 자료의 부족으로 통계적 유의성에는 한계를 가지고 있지만 향후 더 많은 시계열 데이터를 확보하고, 정량화 과정의 정교화를 통해 보완한다면 소셜 빅데이터와 경제지표를 결합한 창업준비지수와 창업예측지수 등 사회적 창업 여건 예측모형의 개발 가능성은 충분하다고 볼 수 있다. 또한 동일한 방법론을 적용하여 가계부채, 개인파산 등 다른 분야의 경제지수에 대한 분석 · 예측에도 적용 가능할 것으로 판단된다. 다만 정성적 데이터의 수집과 해석에 있어서는 검색방법이나 키워드 설정에 있어 연구자의 주관성을 완전히 배제할 수 없는 것이 현실이므로, 향후 지속적으로 방법론의 보완이 필요하고, 정량화에 있어서도 신중을 기해야 할 것이다.

5. 결 론

21세기 새로운 원유라고 불리는 빅데이터는 그 특징을 이해하고 분석 목적을 명확히 설정하여 충분히 합리적인 해석 과정을 거치면 아주 귀중한 정보를 얻을 수 있다. 2012년 미국 대선에서 오바마캠프가 소셜 빅데이터 분석을 통해 의사결정의 근거로 활용한 것을 시작으로 소셜 빅데이터는 공공정책 입안자들에게 주목의 대상이 되어 왔다. 하지만 빅데이터를 이용한 공공부문의 정책분석 및 이에 근거한 의사결정 수립은 아직 본격적인 활성화 단계는 아니며 방법론적으로도 정리되어 있지 않은 상황이다.

본 연구는 기회형 창업에 대한 소셜 빅데이터 분석을 통해 정책적 현안에 대한 심층분석과 의사결정 지원을 위한 빅데이터 분석 사례를 제시하였다. 소셜미디어에 나타나는 온라인 버즈(Buzz)는 향후 정부에서 정책 수요를 파악하고 지원 방향을 설정하는 중요한 도구가 될 수 있음을 시사한다. 소셜미디어에서 형성되고 파급되는 여론의 방향은 실제 정치적, 경제적 활동으로 이어지게 마련인데, 빅데이터 분석 기법을 적용하여 정량화한 데이터를 통해 가까운 미래의 상황을 예측할 수 있는 모델을 지속적으로 개발하여 지능적이고 선제적인 정책 대응체계를 마련할 것을 제안한다.

References

Baden, C., Communication, Contextualization and Cognition : Patterns and Processes of Frames' Influence on People's Interpretation of the EU Constitution, Ph.D thesis, Amsterdam School of Communication Research, 2010.

CASOS, *Automap User's Guide*, Carnegie Mellon University Press, Pittsburgh, 2007.

Choi, Y.J. and S.H. Kweon, "A Semantic Network Analysis of the Newspaper Articles

on Big data", *Journal of Cybercommunication Academic Society,* Vol.31, No.1, 2014, 241-286.

(최윤정, 권상희, "'빅데이터' 관련 신문기사의 의미연결망 분석", *사이버커뮤니케이션학보*, 제31권, 제1호, 2014, 241-286.)

Carley, K., "Coding Choices for Textual Analysis : A Comparison of Content Analysis and Map Analysis," *Sociological Methodology*, Vol.23, No.1, 1993, 75-126.

Dewulf, A., B. Gray, L. Putnam, R. Lewicki, N. Aarts, R. Bouwen, and C. Van Woerkum, "Disentangling Approaches to Framing in Conflict and Negotiation Research : A Meta-Paradigmatic Perspective", *Human relations*, Vol.62, No.2, 2009, 155-193.

Goffman, E., *Frame analysis : An Essay on the Organization of Experience,* New York : Harper and Row, 1974.

Jang, H.Y., "Exploring the Structure of Media Criticism : A Network Analysis of the Problems in Korean Newspaper Reporting", *Korean Journal of Communication and Information*, Vol.16, 2001, 108-133.

(장하용, "언론보도와 비평의 구조 : 신문보도의 비평에 대한 네트워크 분석", *한국언론정보학보*, 16권, 108-133, 2001.)

Jo, H.J., J.H. Seo, and J.T. Choi, "OAR Algorithm Technology Based on Opinion Mining Utilizing Stock News Contents", *Journal of Korean Institute of Information Technology*, Vol.13, No.3, 2015, 111-119.

(조혜진, 서지훈, 최진탁, "주식 뉴스 콘텐츠를 활용한 오피니언마이닝 기반의 OAR 감성사전 알고리즘 기법", *한국정보기술학회논문지*, 제13권, 제3호, 2015, 111-119.)

Kim, L. and N. Kim, "Connecting Opinion, Belief and Value : Semantic Network Analysis of a UK Public Survey on Embryonic Stem Cell Research", *Journal of Science Communication,* Vol.14, No.1, 2015, 1-24.

Kim, D.S., M.S. Lim, W.X.S. Wong, C. Liu, N.G. Kim, J.H. Park, and W.Y. Kil, "A Methodology for Analyzing Public Opinion about Science and Technology Issues Using Text Analysis", *Journal of Information Technology Services,* Vol.14, No.3, 2015, 33-48.

(김다솜, 임명수, William Xiu Shun Wong, 윤한술, 류신, 김남규, 박준형, 길우영, "텍스트 분석을 활용한 과학기술이슈 여론 분석 방법론", *한국IT서비스학회지,* 제14권, 제3호, 2015, 33-48.)

Kim, S., "Middle Power's Diplomatic Strategies in the Perspective of Networks : Applying Theories of Structural Holes and Positional Power", *International Politics Review*, Vol. 51, No.3, 2011, 51-77.

(김상배, "네트워크로 보는 중견국 외교전략 : 구조적 공백과 위치권력 이론의 원용", *국제정치논총,* 제51권, 제3호, 2011, 51-77.)

Kim, Y.M., S.J. Jeong, and S.J Lee, "A Study on the Stock Market Prediction Based on Sentiment Analysis of Social Media", *Entrue Journal of Information Technology*, Vol.13, No.3, 2014, 59-70.

(김영민, 정석재, 이석준, "소셜 미디어 감성분석을 통한 주가 등락 예측에 관한 연구", *엔트루 저널*, 제13권, 제3호, 2014, 59-70.)

Kim, Y.S., N.K. Kim, and S.R. Jung, "Stock-Index Invest Model Using News Big Data Opinion Mining", *Journal of Intelligence and Information Systems*, Vol.18, No.2, 2012, 143-150.

(김유신, 김남규, 정승렬, "뉴스와 주가 : 빅데이터 감성분석을 통한지능형 투자의사결정모형", *지능정보연구*, Vol.18, No.2, 2012, 143-150.)

Kim, J.M., "Policy Directions of Facilitating Tech-

nology Innovation SME Startups", *Science and Technology Policy*, STEPI, Jul · Aug 2007.

(김주미, "기술혁신형 중소기업 창업 활성화를 위한 정책방향", 과학기술정책, 2007.)

Lakoff, G. and M. Johnson, *Metaphors we live by*. Chicago : University of Chicago Press, 1980.

Lee, Y.J., K.C. Jung, B.Y. Jang, and C.H. Nah, "The Success Factors of Technology-based Startups and Supporting Policy", *Science and Technology Policy,* 2013.

(이윤준, 정기철, 장병열, 나청호, "기술창업의 성공조건과 지원정책", *과학기술정책연구원 정책연구*, 2013.)

McKinsey, Big Data : The Next Frontier for Innovation, Competition, and Productivity, 2011.

Park, H.W. and L. Leydesdorff, "Understanding the KrKwic : A Computer Program for the Analysis of Korean Text", *Journal of the Korean Data Analysis Society,* Vol.6, No.5, 2004, 1377-1387.

(박한우, Leydesdorff, "한국어의 내용분석을 위한 KrKwic 프로그램의 이해와 적용", *한국자료분석학회,* Vol.6, No.5, 2004, 1377-1387.)

Shin, T.Y. and W.J. Song, "Determinants of Success of New Technology-Based Firms and Innovation Policy", *Science and Technology Policy,* STEPI, Vol.98, No.14, 1998.

(신태영, 송위진, "신기술창업기업의 성공요인 분석과 정책과제", *과학기술정책관리연구소 정책연구,* Vol.98, No.14, 1998.)

Song, T.M. and J.Y. Song, "Risk Prediction of Internet Addiction Disorder by Using Social Big Data", *Health and Social Welfare Review*, Vol.34, No.3, 2014, 106-134.

(송태민, 송주영, "소셜 빅데이터를 활용한 인터넷 중독 위험예측 모형", *보건사회연구*, Vol.34, No.3, 2014, 106-134.)

Song, E.J., "A Study on the Case Analysis of Customer Reputation based on Big Data", *Journal of the Korea Institute of Information and Communication Engineering*, Vol.17, No.10, 2013, 2439-2446.

(송은지, "빅데이터를 이용한 고객평판 사례분석에 관한 연구", *한국정보통신학회논문지*, 제17권, 재10호, 2013, 2439-2446.)

Sudhahar, S., G. Veltri, and N. Christianini, "Automated analysis of the US presidential elections using Big Data and network analysis", *Big Data and Society*, Vol.2, No.1, 2015.

Wang, W. and R. Rada, "Structured Hypertext with Domain Semantics", *ACM Trans. Inform. Syst.*, Vol.16, 1998, 372-412.

Wasserman, S. and K. Faust, *Social Network Analysis : Methods and Applications,* Cambridge : Cambridge University Press, 1994.

Yook, C.H. and I.O. Jeon, "Effect of Entrepreneur's Characteristics on the Intention of Entrepreneruship : Focusing on the Moderating Effect of Entrepreneurship Mentoring in Korean", *Journal of the Korea Contents Association,* Vol.14 No.10, 2014, 661-676.

(육창환, 전인오, "창업가의 특성요인이 창업의지에 미치는 영향 : 국내 창업 멘토링의 조절효과를 중심으로", *한국콘텐츠학회논문지,* 제14권, 제10호, 2014, 661-676.)

Jo, E.J., J.S. Park, and B.R. Jin, "A study on the Communication Frame among Tripartite Commission Members", Economic and Social Development Commission of Korea, 2015.

(조은주, 박종식, 진보래, "한국 노사정의 사회적 대화 프레임 분석을 통한 소통 활성화 방안 연구 : 통상임금, 근로시간 단축, 정년연장의 쟁점을 중심으로", 경제사회발전노사정위원회, 2015.)

About the Authors

Young-Joo Lee (lyj@nia.or.kr)

Young-Joo Lee is a Principal Researcher at the National Information Society Agency (NIA). He received his Ph.D in Information Systems from Graduate School of Yonsei University, M.S. in Business Administration from Korea Advanced Institute of Science and Technology (KAIST), and B.S. in Socioloy at Yonsei University. His research interests include future study, big data analytics, management of technology, enterprise architecture, information systems, and other governmental and social issues regarding information technology.

Leo Dhohoon Kim (leo_kim@arspraxia.com)

Leo Kim is the CEO of Ars Praxia, a consulting company that specializes in big data analysis and innovation consulting. His career interests lie at the governance and discourses of science & technology and innovation in educational and industrial fields. Since his Ph.D study in Science Policy Research Unit (Univ. of Sussex), Leo has also been engaged in academic works researching interaction of actors, strategies, narratives, and knowledge in the public sphere of science, and developed the method of semantic network analysis and big data processing for pragmatic applications. He has published articles related to these topics in several international journals.