

# 소셜 빅데이터로 본 경기도 행복지도, 2013~2017

The Happiness Map of Gyeonggi Province

Created through Big Data Analysis of Social Media, 2013~2017

이 병 호\* · 김 가 연\*\*

Lee, Byung-Ho · Kim, Ga-Yeon

## ■ 목 차 ■

- I. 서론
- II. 선행연구
- III. 연구설계
- IV. 분석결과
- V. 요약 및 결론

이번 연구는 경기도 31개 기초자치단체를 대상으로 2013년부터 2017년까지 행복과 결부된 주요 연관어 및 긍정적, 부정적 감성어 언급 빈도 추세를 소셜 빅데이터를 통해 분석하였다. 이를 위해 우선 지역행복에 영향을 주는 다양한 요인을 10개 영역으로 구분하고 영역별 주제어 사전을 구축하였다. 그리고 소셜 빅데이터 분석 방법의 하나인 웹 크롤링 분석을 통해 시기별 행복 관련 명사형 주제어 사용 추세를 분석하였다. 나아가 지역별 긍정적, 부정적 감성형용사 출현 빈도를 확인하고 이를 경기도 행복지도로 제시하였다. 분석결과, 지역의 행복과 상관성이 높은 요인 중에서도 특히 주거문제는 시기를 막론하고 두드러진 연관 주제로 나타났다. 아울러 최근에는 개인의 사생활 특히 휴식과 여가활동에 대한 언급 빈도가 나날이 높아지는 추세이다. 한편 감정의 공간적 파급효과는 행복할 때에 비해 불행할 때 더 두드러진다. 이러한 소셜 빅데이터 분석결과는 행정통계나 설문조사로는 확인할 수 없던 내용을 포함한다. 따라서 이번 연구는 기존 연구의 한계를 보완함으로써 지역의 행복 수준에 대한 총체적인 이해를 도모할 뿐만 아니라 삶의 질을 높이는 지역정책 개발의 새로운 단초를 제공한다.

\* 제1저자, 교신저자, 아주대학교 사회학과 조교수

\*\* 제2저자, 경기연구원 공존사회연구실 연구원

논문 접수일: 2018. 8. 10, 심사기간: 2018. 8. 10~2018. 9. 13, 게재확정일: 2018. 9. 13.

□ 주제어: 행복지도, 소셜 빅데이터, 웹 크롤링, 파급효과, 경기도

This research wants to identify the factors associated with the degree of happiness at local level and test the evidence of spatial ripple effects of emotional well-being. It focuses on 31 administrative subunits in Gyeonggi Province and analyzes social big data regarding these areas from 2013 to 2017. To do so, we first established a keyword dictionary including thematic nouns and emotive adjectives. We then analyzed social big data employing a web crawling method. We also presented the happiness map of Gyeonggi Province for visualization. As a result, we found several patterns of happiness. The question of housing has been the major factor affecting happiness. Further, it appears that the degree of regional happiness has been increasingly influenced by personal conducts particularly rest and leisure rather than societal concerns. We also found that, as the happiness map of Gyeonggi Province indicates, negative emotions are more spatially contagious than positive emotions. Our findings suggest that social big data analysis as a novel approach can not only broaden our understanding of happiness and emotional well-being, but also contribute to the development of regional policy whose goal is to enhance the quality of life.

□ Keywords: Happiness Map, Social Big Data, Web Crawling, Ripple Effect, Gyeonggi Province

## I. 서론

각 개인마다 삶의 양태는 다르지만, 시공간에 관계없이 공통적으로 추구하는 중요한 가치는 개인의 삶의 질 향상에 있다. 그리고 삶의 질을 높이는데 있어 물리적 환경 구축은 기본적인 욕구 충족이란 측면에서 중요한 요소로 인식되어 왔다(이승철, 2011; 윤은기, 2012; 이윤진·김남숙·기영화, 2016). 따라서 지금까지 물리적 발전은 개인의 삶의 질과 지역사회의 질적 향상에 있어 중요한 목표로 작용하였다. 기반시설 구축, 인구 증가, 산업화 그리고 경제적 성장을 이뤄감에 따라 사회 전체의 효용 증대와 더불어 개인 역시 많은 혜택을 누렸음은 부정할 수 없는 사실이다(Burton, 2003; Yang, 2008; Dempsey et al., 2009). 하지만 압축 고도 성장 시대가 저물고 장기적인 저성장 시대에 접어든 지금 부의 불평등, 사회갈등, 이념대립 그리고 초저출산과 인구고령화 등 지속가능한 발전을 저해하는 다양한 문제들이 대두되는 상황이다.

이런 현실에서 최근 정부와 학계는 양적 성장을 넘어 삶의 질과 사회의 질을 제고하는 방안을 모색하기 시작했다. 이런 관점은 사회를 이루는 구성원 개인의 행복을 중요한 요소로 인식하고 있으며, 성장의 척도에 행복 수준(degree of happiness)을 중시하는 인식의 전환을 수반한다. 뒤에 상론하겠지만 OECD는 질적 요소를 측정하기 위해 ‘더 나은 삶 지수’를 만들어 주기적으로 그 결과를 공표하고 있으며, UN에서도 2012년부터 ‘세계행복보고서’를 발표하고 있다.

지금까지 개발된 행복에 대한 여러 지표는 대개 거시 행정통계와 설문조사를 통해 얻어진 정형화된 자료를 활용해 구축되었다. 하지만 이러한 통계자료는 수집에 많은 비용이 들고 공개까지 장시간 소요되기에 급변하는 상황에 기민하게 대응하기 어렵다는 한계를 가진다. 이런 문제를 극복하기 위해 최근에는 비정형 자료가 주목받는 상황이다. 특히 정보통신기술의 발달로 사람들이 자신의 생각을 자유롭게 표현할 수 있으며 그 양은 무제한으로 증가하는 상황이라는 점에서 통상 빅데이터로 불리는 방대한 정보를 해석하고 활용하는 것은 의미가 있다(Najafabadi et al., 2015).

그 동안 빅데이터는 주로 실생활에 필요한 다양한 정보 제공을 목적으로 활용되어 왔지만, 이를 활용하여 지역 차원의 행복과 같이 감정을 분석한 연구는 아직까지 상대적으로 미흡한 편이다. 하지만 분명 개인이나 지역 차원의 기분(mood)도 빅데이터를 활용한다면 측정 가능한 수준으로 자료화할 수 있다(Mayer-Schönberger & Cukier, 2013: 93). 가령 소셜 빅데이터 자료를 수집하여 문화권별로 사람들이 느끼는 감정을 분석한 연구(Golder & Macy, 2011)가 있다. 따라서 이번 연구는 사회발전의 질적 측면에서 핵심이라 할 지역사회의 행복

수준을 탐색하기 위해 소셜 빅데이터 분석을 시행하고 나아가 매핑(mapping)을 통해 시각화한 행복지도를 제시함으로써 현상을 보다 직관적으로 파악하고자 한다.

이를 위하여 우선 분석 대상을 설정한다. 지역은 경기도 31개 기초자치단체, 시기는 2013년부터 2017년까지로 한다. 그리고 행복의 측정과 관련하여 개념에 대한 조작적 정의를 내린 다음, 각종 포털사이트의 개인 홈페이지, 블로그, SNS를 아우르는 비정형 자료를 수집한다. 분석방법은 소셜 빅데이터 분석 그 중에서도 웹 크롤링(web crawling) 분석을 시행한다. 주제어 사전을 기준으로 웹 크롤링을 통해 도출된 단어들의 빈도분석과 언급추이를 살펴본다. 나아가 지역별 긍정적, 부정적 감성형용사 출현 빈도를 확인하고 그 결과를 커뮤니티 매핑을 통해 시각화한다.

이번 연구는 소셜 미디어를 통해 얻어진 방대한 빅데이터를 통해 개인의 행복에 미치는 다양한 요인을 체계적으로 측정하였다는 점에서 의의가 있다. 더 나아가 시각화를 통해 분석지역을 시각적, 직관적으로 이해할 수 있도록 제시한다는 점에서 이전 연구들과 차별성을 가진다. 특히 정보의 양이 방대한 현대 사회에서 불특정 다수의 견해를 쉽게 파악할 수 있도록 하며, 설문조사나 인터뷰를 통해서 알 수 없는 정보를 추가하여 기존 연구들을 보완, 발전시킬 수 있다는 측면에서 향후 지역 차원의 행복 관련 연구에도 단초를 제공할 수도 있다.

## II. 선행연구

그간 국가의 발전 정도를 측정하는 기준은 국내총생산(GDP)이나 구매력평가(PPP)와 같은 양적인 기준으로 측정해온 것이 일반적이었지만, 최근 들어 각국 정부는 삶의 질이나 행복 같은 질적 성장에 주목하는 추세이다(변미리, 2013, 2014). 이런 움직임은 특히 프랑스 정부의 요청으로 시작된 사회발전과 경제성과를 측정하기 위한 연구가 나온 이후 세계적인 주목을 받기 시작했다(Stiglitz, Sen, & Fitoussi, 2009). 국제기구 가령 UN은 2012년부터 주기적으로 세계행복보고서(World Happiness Report)를 발행하여 회원국의 시민이 체감하는 행복 수준을 조사하고 있다. 아울러 OECD 역시 웰빙(well-being) 수준을 측정하기 위해 '더 나은 삶 지수'(Better Life Index; 이하 BLI)를 개발해 2011년부터 그 결과를 공개하고 있다(Durand, 2015; OECD, 2017). 아래에서 상론하겠지만, 이번 연구는 여러 삶의 질 혹은 행복 관련 지표 가운데 BLI 분류체계를 따른다.

한편 대한민국 정부 역시 국정기조를 국민의 행복과 삶의 질 향상으로 설정했으며, 이는 정책 수립과 시행에 있어 중요한 목표로 자리매김하는 상황이다(우창빈, 2013). 주지하듯이 한

국은 단기간에 눈부신 경제성장을 이루었지만 개인이 실제 느끼는 행복이나 이에 영향을 주는 삶의 질 수준은 다른 선진국에 비해 높지 않다(김운태, 2010). 가령 최근 공개된 2018 세계행복보고서 결과에서 조사대상 156개국 중 57위(Helliwell, Layard, & Sachs, 2018), OECD의 BLI에서는 38개국 가운데 28위에 그치는 실정이다(OECD, 2016). 이와 같이 낮은 수준에 머무르고 있는 국민의 행복 수준을 높일 정책을 만들고 집행하기 위해서는 무엇보다 이에 대한 정확한 측정과 함께 한국적 맥락에서 개인의 행복에 영향을 주는 요인을 실증적으로 밝혀야 한다.

지난 10여 년간 정부와 학계에서 삶의 질 혹은 행복에 대한 측정 문제는 활발히 논의되고 있다(김승권·장영식·조홍식·차명숙, 2008; 황명진·심수진, 2008; 이희길·심수진, 2010; 한준 외, 2011; 변미리, 2013; 김선기·박승규, 2014). 통계청은 ‘국민 삶의 질 종합지수’(Korean Quality of Life)를 통해 삶의 질과 관련한 각종 자료를 제공하고 있으며, 변미리(2014)는 서울형 행복지표 체계 구축을 통해 지역주민의 행복 수준을 측정하고 행복지수에 영향을 주는 요인을 분석하였다. 최근 경기도 지역을 대상으로 한 연구에서도 단순히 GDP로 측정되는 경제적 지표나 물질적인 조건뿐만 아니라 개인들의 심리적인 만족과 행복 그리고 삶의 질적 수준까지 포괄적으로 고려하고 있다(이병호·손웅비, 2017).

요컨대 장기적인 저성장 국면에 접어든 상황에서 발전국가 시기의 급속한 양적 성장에서 질적 성장으로 관심이 옮겨지는 추세이며 삶의 질 수준을 높이기 위한 정책을 펼치기 위해 행복에 대한 연구는 국내외에서 갈수록 주목받고 있다.

그동안 행복 및 삶의 질 수준을 측정하고 분석하기 위한 자료로는 주로 개인 단위 미시 서베이 자료와 거시 행정통계 등이 활용되어왔다(김선기·박승규, 2014). 아울러 최근 데이터 시각화의 중요성이 커지면서, 국가 혹은 지역의 수준을 비교하기 위해 자료에 속성치(attribution value)를 부여하고 이를 그룹화시킨 매핑 방식이 주목받고 있다. 원래 매핑 방식은 보건, 환경, 도시경관 등의 영역에서 다양하게 활용되어왔는데, 근래에는 사람들의 행복을 지역적으로 분류하고 그 특징을 비교하는 연구에도 적용되는 추세이다.

행복을 주제로 매핑한 연구는 2009년 영국 레스터 대학의 심리학자 White 교수가 처음으로 작성한 2006년 세계행복지도를 꼽을 수 있는데 여기서는 국가별 개인의 행복 수준을 제시하였다(White, 2009). 아울러 근래에는 설문조사 자료뿐만 아니라 빅데이터를 활용한 매핑 분석도 활용되는 추세이다. 예를 들어 미국 전역에서 사람들이 트위터에 올린 텍스트에 대한 웹 크롤링을 시행하여 행복과 함께 사용된 연관어를 수집하고 이를 매핑하여 지역별 행복과 상관성이 높은 핵심어를 비교분석한 연구가 있다(Schwartz et al., 2013).

나아가 빅데이터, 머신러닝 등을 활용한 지역별 행복지도는 도시주민의 행복 수준을 시각화하는 방법으로 국내외에서 각광받고 있다. 예컨대 런던시는 UCL(University College London)의

CASA(Centre for Advanced Spatial Analysis) 센터를 통해 ‘alpha’ 프로토타입의 지역 대시보드(dashboard)를 2012년에 선보였다. 당시 영국의 연방 정부에서는 소셜 시뮬레이션(social simulation)을 위한 웹 기반 플랫폼(web-based platform)인 e-Infrastructure를 지원하였다. 이 사업을 통해 시간이나 상황에 따라 가변적인 다양한 정보, 예컨대 날씨, 대기의 질, 교통체증, 지역 뉴스 등을 시각적으로 쉽게 사람들에게 전달할 수 있게 되었다. 더 나아가 런던에서는 소셜 미디어와 행복지수(Happiness Index)를 활용한 지역 사회지도를 제작하였고 이후 이 프로젝트는 글래스고(Glasgow), 브라이튼(Brighton) 지역으로 확대되었다. 한편 Kitchin, Lauriault, & Wilson(2017)은 2014 Dublin Dash Board 프로젝트를 수행하면서 행복지수 적용을 포함한 지역의 다양한 사회현상을 지도로 구현하였다.

국내에서는 도시환경, 교통, 범죄 등 다양한 분야에서 매핑을 활용한 연구가 다수 진행되었다. 그러나 직접측정이 어려운 삶의 질, 행복 수준을 매핑 방식으로 연구하는 것은 아직 소수에 불과하다. 최근에서야 다양한 시도가 이루어지고 있으며, 여기에는 서울시 행복지수를 분석하고 예측하기 위해 가용한 빅데이터로 행복지표를 도출하고 그 결과를 서울시 행복지도를 통해 시각화한 신동윤·송유미(2017)의 연구를 들 수 있다.

### Ⅲ. 연구설계

#### 1. 주제어 사전 구축

소셜 빅데이터 분석 방법 중 하나인 웹 크롤링 방식으로 자료를 확보하기 위해서는 몇 가지 작업이 선행되어야 한다. 이 가운데 특히 주제 분석을 위해 필요한 용어 사전을 자료를 확보하기 전에 미리 구축해야 한다. 따라서 연구진은 행복이란 개념과 연관된 어휘를 체계적으로 분류한 주제어 사전을 구축하였다. 기존의 범용사전을 사용하는 대신 이번 연구의 목적에 맞춰 연구자가 사전에 설정한 주제어 사전을 활용하였으며 이를 만드는 세부적인 과정은 다음과 같다.

먼저 앞 장에서 살펴본 행복을 측정하기 위한 다양한 지표 가운데 공신력을 가진 분류방식을 선택할 필요가 있다. 이번 연구는 기존 행복지표 연구(변미리, 2014)에서 일반적으로 사용되는 OECD가 개발한 BLI를 사용하였다. <표 1>은 BLI를 구성하는 세부 영역 및 지표를 보여준다.

〈표 1〉 OECD의 BLI 영역별 지표

영역	지표명칭
주거(Housing)	주거 관련 지출(Housing expenditure)
	기본 시설이 있는 가구(Dwellings with basic facilities)
	개인당 방수(Number of rooms per person)
소득(Income)	가계 금융자산(Household financial wealth)
	가계 순 가처분소득(Household net adjusted disposable income)
직업(Job)	직업안정성(Job security)
	개인소득(Personal earning)
	고용률(Employment rate)
	장기실업률(Long-term unemployment rate)
공동체(Community)	지원관계망의 질(Quality of support network)
교육(Education)	기대교육기간(Years in education)
	학생의 역량(Student skills)
	교육성취(Educational attainment)
환경(Environment)	수질(Water quality)
	대기오염(Air pollution)
시민참여 (Civic engagement)	규칙 제정에의 참여(Stakeholder engagement for developing regulations)
	투표참여율(Voter turnout)
건강(Health)	자기보고 건강상태(Self-reported health)
	기대수명(Life expectancy)
삶의 만족 (Life satisfaction)	삶에 대한 만족도(Life satisfaction)
안전(Safety)	살인율(Homicide rate)
	야간보행 안전체감도(Feeling safe walking alone at night)
일과 삶의 균형 (Work-Life balance)	여가와 개인적 돌봄에 쓴 시간(Time devoted to leisure and personal care)
	장시간 근로자(Employees working very long hours)

다음으로 BLI 영역 가운데 ‘삶의 만족’ 영역을 제외한 10개 영역에서 행복과 연관이 있는 어휘를 도출하였다. 즉 각 영역에 해당하는 명사와 이에 대한 인식과 태도를 보여주는 감성형 용사를 함께 도출하여 사전을 구축하였다. 행복에 대한 단어사전은 〈표 2〉와 같이 정리할 수 있다. 특히 아래에서 살펴볼 감성분석 결과는 표에서 제시한 감성형용사 사전을 통해 긍정과 부정으로 구별해 분석하였다. 이를 위해 긍정과 부정을 나타내는 형용사를 분류해 텍스트 분석을 실시한 다수의 선행연구를 참고하였다(장효진, 2001; 박경미·박호진·김형곤·고희동, 2011; 강범모, 2014; 양승준·이보연·김희웅, 2015; 임좌상·김진만, 2015). 예컨대 교육 영

역에서 ‘행복’이라는 단어와 입학시험에 붙거나 떨어지는 것이 함께 사용되었을 때 합리적인 경우는 행복에 긍정적인 것으로 불합격한 경우는 부정적인 것으로 간주하였다. 한편 ‘행복주택’과 같이 행복이 포함된 고유명사는 불용어(不用語)로 처리해 분석에서 제외하였다.

〈표 2〉 행복과 연관된 단어사전

영역	명사형 주제어	감성형용사	
주거	전세, 분양, 역세권, 재개발, 신혼부부, 재건축, 청년주택, 사회주택, 임대주택, 복지, 쪽방, 아파트, 전원주택, 마당, 대출, 단독주택, 빌라, 고시원, 원룸, 임대수익, 연립, 부동산시장, 월세, 매매, 주택담보대출, 다세대주택, 주택청약, 주택연금, 장기임대, 조합, 신축, 정비구역, 신도시, 입주물량, 관리비, 규제, 투자, 투기, 금리, 도시형	긍정	안락한, 쾌적한, 기쁜, 편안한, 기대하는, 편리한, 좋은, 활기찬, 충족하는, 추천하는, 부듯한
		부정	힘든, 고통스러운, 불안한, 싫은, 혐오스러운, 나쁜, 슬픈, 우울한, 암담한, 전망이 어두운, 심각한, 낙후된, 위험한, 불편한, 짜증나는, 괴로운
소득	가치분소득, 최저임금, 임금인상, 실질소득, 세금, 종합소득세, 상속, 공제, 퇴직금, 로또, 상금, 근로소득, 양도소득, 연말정산, 맞벌이, 외벌이, 저축, 투자, 투기, 대출(가계부채), 이자, 물가, 재산, 수당, 금수저	긍정	여유로운, 풍족한, 충분한, 즐기는, 좋은, 기쁜, 안정적인
		부정	화나는, 부족한, 힘든, 분노하는, 절망하는, 반발하는, 위기감의, 짜증나는, 슬픈, 소외감, 격차가 벌어지는, 우울한, 대책없는
직업	대기업, 자영업, 중소기업, 공무원, 공기업, 인턴, 사업, 이직, 계약직, 정규직, 아르바이트, 프리랜서, 관리직, 고위직, 기술직, 전문직, 단순노무직, 무직, 연구직, 직장인	긍정	인정받는, 승진하는, 좋은, 보람된, 즐거운, 의욕있는, 시작하는, 쉬운, 즐기는, 잘하는, 기쁜, 좋은, 안정적인
		부정	그만두는, 못하는, 질책하는, 힘든, 고된, 짜증나는, 괴로운, 찢리는, 보람없는, 하기싫은, 의욕없는, 피곤한, 싫은, 슬픈, 불안한, 귀찮은
공동체	반상회, 부녀회, 주민센터, 자치, (지역 혹은 특정집단) 온라인 카페, SNS, 단지, 참여, 소속	긍정	권장하는, 화해하는, 활발한, 즐거운, 협조적인, 좋은, 신나는, 기쁜, 공손한, 흐뭇한, 이해하는
		부정	나쁜, 조장하는, 못된, 짜증나는, 귀찮은, 슬픈, 이기적인, 비난하는, 싫은, 무례한, 오해하는
교육	학군, 명문대, 진학, 학원, 대학교, 진학률, 유학, 학교폭력, 수준, 성취도, 시험, 수능, 학생부, 실적, 고등학교, 중학교, 입시, 상위권, 중·하위권, 교사, 따돌림(왕따), 학부모, 성적, 인성	긍정	합격한, 좋은, 붙는, 잘하는, 기쁜, 성공한, 보람된, 의욕적인, 보장되는, 가치있는, 책임있는, 사랑하는, 인정받는, 우수한, 자신있는
		부정	피곤한, 실패한, 떨어지는, 스트레스받는, 고달픈, 힘든, 짜증나는, 화나는, 우울한, 부담되는, 이기적인, 신뢰감없는, 상실한, 외로운, 외면하는, 미워하는, 증오하는, 모욕적인, 괴로운



영역	명사형 주제어	감성형용사	
환경	미세먼지, 소음, 공해, 광해, 수질오염, 매연, 배기가스, 대기상태, 먼지, 민원	긍정	상쾌한, 맑은, 좋은, 깨끗한, 향긋한, 편안한, 기분좋은, 청량한, 향기로운, 개선되는
		부정	심각한, 나쁜, 더러운, 썩은, 악취나는, 악독한, 불편한, 악화되는, 구질구질한, 불쾌한, 싫은, 괴로운, 참을 수 없는, 짜증나는, 화나는
시민 참여	선거, 청원, 시위, 집회, 서명, 출마, 국회의원, 시민단체, 탄핵, 고소·고발	긍정	신뢰하는, 잘하는, 좋은, 존경하는, 인정하는, 지지하는
		부정	무관심의, 믿을 수 없는, 화나는, 분노하는, 싫은, 짜증나는, 언짢은, 탐탁치않은, 열받는, 의심스러운
건강	암, 질병, 죽음, 운동, 정신, 장수	긍정	좋은, 회복하는, 즐기는, 즐거운, 자신있는, 유지하는, 활기찬, 건강한, 생기넘치는
		부정	두려운, 암울한, 불안한, 회복가능성이 없는, 나쁜, 악화되는, 힘든, 아픈, 괴로운, 자신없는, 고통스러운, 어려운
안전	피난대피시설, 소방시설, 보험, 범죄(살인, 절도, 성범죄 등), 침수, 홍수, 지진, 산사태, 전쟁	긍정	안심되는, 안전한, 충분한, 적절한, 신뢰할 수 있는
		부정	불안한, 두려운, 충분하지 않은, 위험한, 신뢰할 수 없는
일과 삶의 균형	여가, 레저, 운동, 여행, 공부, 업무, 일, 사색, 독서, 퇴근, 출근, 개인, 장시간	긍정	여유로운, 즐기는, 균형있는, 적절한, 보장된, 유연한
		부정	피곤한, 과중한, 여유없는, 힘든, 고된, 부족한, 경직된

## 2. 자료 및 분석방법

웹 크롤링의 대상은 앞서 논의한 주제어 사전에 제시된 단어를 기준으로 한다. 특히 이번 연구는 사전에 정의한 주제어들을 인터넷을 통해 수집하는 소셜 빅데이터 분석 방식을 취한다. 여기서 말하는 소셜 빅데이터는 한국의 주요 언론사 기사뿐만 아니라 트위터(twitter), 페이스북 facebook)과 같은 SNS, 블로그, 그리고 포털 사이트의 그룹이나 카페에 올라온 게시물 등을 아우른다. 빅데이터 수집을 위해 구글 웹 크롤러(Google Web crawler)에서 제공하는 소스코드를 활용하였다.

이러한 웹 크롤링 방식은 불특정 다수의 일반적인 의견을 수렴하기 위해 주로 사용된다. 비용이 많이 드는 실태조사나 인터뷰를 하지 않아도 온라인에서 빈번하게 사용되는 단어에 대

한 분석을 통해 사람들의 의식이나 태도를 파악할 수 있다는 측면에서 자료수집을 위한 조사 비용이 들지 않는다는 장점이 있다. 다만 빅데이터의 속성상 불필요하거나 연관이 없거나 결과를 왜곡시킬 여지가 있는 정보를 연구 설계단계에서 철저히 걸러내어야만 신뢰할만한 분석 결과를 얻을 수 있다. 따라서 연구 목적에 부합하는 분석대상을 정하기 위한 코딩과정이 필요하며, 그 구체적인 과정은 다음과 같다.

일단 시공간적 범위는 2013년부터 2017년까지 5년 동안 경기도 전역에서 쓰인 행복과 관련한 자료로 한정하며, 경기도 지역 내의 다양성을 고려하기 위해 시·군 단위 31개 기초자치단체를 분석단위로 설정하였다. 경기도 지역에서 행복이 동시에 포함된 자료를 모으기 위해, 여기서는 구글 웹 크롤러에서 제공하는 html 소스를 활용하여 이를 Python 2.7코드로 변환시켜 노출되는 웹 페이지에서 '경기도 31개 행정구역 명칭'과 '행복'이 포함된 문장(line)을 csv 파일 형태로 저장하였다. 구체적으로 웹 크롤링을 수행할 때 주로 쓰이는 Python 플러그인(Plug-in) 가운데 하나인 Scrapy 1.6 프로그램을 활용하였으며 웹 크롤링이 허용되는 범위 안에서 코드작업(encoding)을 통해 언급된 단어와 빈도값을 도출하였다.

분석결과를 제시하기 위해 경기도 31개 기초자치단체에 대한 지역 명칭과 행복과 관련한 주제어 사전에 등장하는 어휘를 입력해 확인된 용례의 빈도와 상호 연계성 수준을 우선 확인하였다. 이를 위해 2013년부터 2017년까지 연도별로 BLI의 10개 영역에 대한 단어 사용 건수의 추이를 살펴보았으며, 그 결과는 Tagxedo 웹페이지([www.tagxedo.com](http://www.tagxedo.com))에서 제공하는 프로그램을 이용하여 단어구름 형태로 시각화하였다. 지역에 따라 특정 시기에 영역별 언급 횟수의 변화를 확인한다면 행복과 밀접하게 영향을 주고받는 주요 현안이 무엇인지 식별할 수 있다. 가령 뒤에서 상술하겠지만 주로 안전 영역에 해당하는 세월호 사건은 경기도에서도 안산시 및 그 인근 지역에 심대한 영향을 주었다. 마찬가지로 건강 영역에 속하는 메르스 감염 문제는 2015년 평택지역에서 처음 발견되었고 당시 경기도 상당수 지역의 행복에 부정적인 영향을 주었다.

다음으로 지역 간 행복 수준을 비교하기 위해 ArcGis 10.3 프로그램을 이용하여 경기도 행복지도를 제작하였다. 지역별 Shp 파일을 바탕으로 지난 5년간 행복 및 이와 관련한 10개 영역에 대해 언급된 단어의 빈도 값을 입력한 속성테이블(attribute table)을 만들고, 이를 단계구분도 기법을 활용해 연도별로 매핑하였다. 이런 과정을 통해 만들어진 경기도 행복지도는 시기별, 지역별 행복 수준을 직관적으로 파악할 수 있다는 장점을 가진다.

## IV. 분석결과

### 1. 명사형 주제어 빈도분석

여기서는 ‘경기도’와 ‘행복’을 포함하는 문장을 웹 크롤링 방식을 통해 추출하고, 이를 앞선 표에서 제시한 10개 영역별 주제어 사전에 포함되는 명사와 함께 언급된 건수를 제시한다. 연도별 그리고 영역별 빈도분석 결과는 <표 3>과 같다.

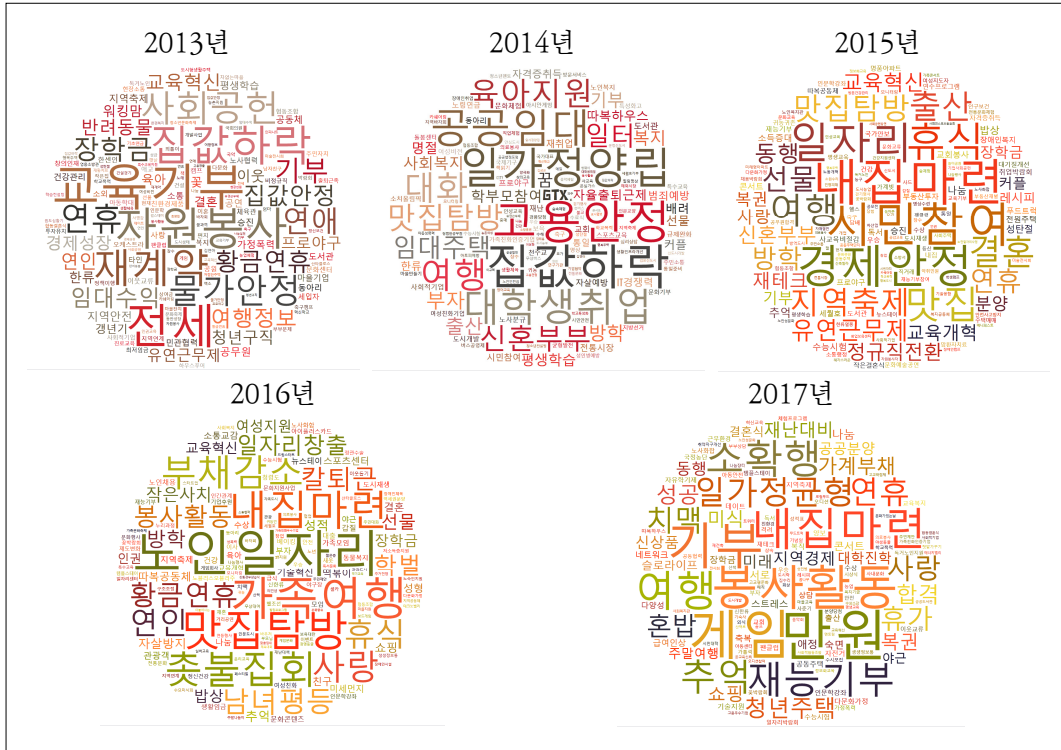
<표 3> 영역별 명사형 주제어 언급 빈도, 2013~2017

영역	2013년	2014년	2015년	2016년	2017년	합계
주거	112,485	92,654	136,772	145,670	125,462	613,043
소득	73,879	79,549	82,148	91,244	95,698	422,518
직업	65,841	71,553	80,266	87,640	92,124	397,424
공동체	28,694	39,567	50,724	58,831	71,105	248,921
교육	41,503	44,028	47,564	51,742	58,127	242,964
환경	17,942	24,598	38,068	49,229	57,982	187,819
시민참여	5,194	6,074	7,569	20,720	42,898	82,455
건강	10,574	12,721	13,691	12,549	20,024	69,559
안전	6,418	26,889	10,415	9,217	16,439	69,378
일과 삶의 균형	10,524	9,927	10,124	18,742	24,128	73,445
합계	373,054	407,560	477,341	545,584	603,987	2,407,526

경기도 지역에서 2013년부터 2017년까지 행복과 함께 언급된 명사형 주제어의 사용빈도는 240여만 건에 달한다. 시기별로 보면 2013년 37.3만 건에서 2017년 60.4만 건으로 꾸준히 증가하는 추세를 보인다. 그렇지만 사용빈도 증감추세는 세부 영역별로 차이를 나타낸다. 가령 시민참여 영역은 2015년까지 완만한 상승세를 보이다 2016년부터 급증하였으며, 안전 영역은 2014년 세월호 사건으로 인해 크게 부각되었으나 이후 2016년까지 줄어들었다.

영역별 특징에 대해 논의하기에 앞서, 시기에 따라 행복과 관련이 있는 명사형 주제어 사용빈도를 <그림 1>과 같이 시각화하였다.

〈그림 1〉 행복과 관련된 명사형 주제어 빈도, 2013~2017



그림에서 보듯 주거와 일자리 그리고 사회참여는 모든 시기에서 행복과 매우 긴밀한 관계를 가진다. 2013년과 2014년에는 집값과 관련한 언급이 두드러지며, 2015년부터는 내 집 마련 즉 자가 소유에 대한 관심이 높아지는 추세이다. 자원봉사나 각종 기부활동 같은 사회참여 역시 행복과의 상관성이 높는데, 여기서 특기할만한 것은 2016년 촛불집회는 행복감과도 높은 관련성을 가진다는 점이다. 아래에서는 세부 영역별 특징을 보다 상세히 논의한다.

주거 영역은 10개 세부 영역 가운데 가장 많이 언급되고 있으며 5년간 61.3만 건으로 전체 언급 빈도의 25%를 차지한다. 이는 경기도와 같은 수도권 지역에서 주거문제는 사람들의 행복감에 심대한 영향을 주고 있음을 시사한다. 구체적으로 노인, 신혼부부, 청년, 대학생 등 주거 취약계층을 위한 공공임대 주택 제도를 통한 주거비 부담 완화라는 기대감이 큰 영향을 주었음을 알 수 있다. 나아가 제2기 신도시 이후 주거지 입주 물량 확대를 목적으로 한 신도시(하남신도시, 양주 운정신도시 등)들이 경기도 지역에 조성되면서 신규 분양 주택 및 거주 환경에 대한 기대감이 반영되었다고 볼 수 있다. 특히 입주 물량이 큰 폭으로 증가한 2014년과 2015년에는 언급된 건수도 급등하였다. 한편 소득과 직업은 경기도 지역에서 주거 다음으로

중요한 요소이며, 지난 5년간 두 영역 모두 언급 빈도가 점증하고 있다. 특히 작금의 취업난을 반영하듯 청년취업과 노인일자리에 대한 언급 빈도가 높게 나타났다.

지난 5년간 공동체와 교육 영역에 대한 언급 횟수는 공히 24만여 건인데, 영역별 증가속도는 차이가 존재한다. 교육은 완만하게 증가하는데 반해, 공동체는 최근으로 올수록 급등하는 추세이다. 이에 따라 2015년부터는 교육에 비해 공동체 영역에 대한 언급 빈도가 높아지고 있으며, 그 격차는 갈수록 커지는 상황이다. 공동체에 대해 높아지는 관심을 감안할 때, 지역을 기반으로 한 다양한 활동이나 이웃과의 교류가 사람들의 행복에 점점 더 유의미한 연관을 가지고 있음을 시사한다.

한편 환경 문제는 해가 갈수록 언급 횟수가 높아지는데, 특히 대기 질 문제인 미세먼지에 대한 언급이 높은 상황이다. 그리고 시민참여는 주로 정치 이슈에 대한 영역인데 <그림 1>에서 보듯이 2015년 핵심 연관어로 시민참여가 부각된 것으로 확인할 수 있다. 사회참여에 대한 열기는 2016년 촛불집회와 2017년 대통령 선거 등을 통해 급속도로 고조되는 추세를 보인다. 이는 적극적인 사회참여를 통해 자신의 행복을 스스로 찾아간다는 인식이 높아지고 있음을 알 수 있다.

마지막으로 건강, 안전, 그리고 일과 삶의 균형 영역은 다른 주제에 비해 행복과 관련해 언급되는 빈도는 낮은 편이지만 여전히 행복과 밀접한 관련을 가진다. 특히 개인의 여가활동과 연관이 있는 일과 삶의 균형의 경우 2016년부터 관심이 크게 증가하고 있으며 소소한 일상의 보장이나 사소한 것에서 느끼는 행복감이 강조되었다. 구체적으로 <그림 1>에서 보듯이 게임, 맛 집 탐방, 휴식, 소확행(小確幸, 작지만 확실한 행복) 등의 주제어가 최근 두드러진다.

경기도 31개 시·군의 행복과 관련하여 2013~2017년 사이 도출된 명사형 주제어를 분석한 결과, 2014년까지는 주로 주거문제, 고용 및 취업문제, 일가정양립과 같은 사회정책 현안이 개인의 행복에 미치는 영향이 높은 것으로 조사되었다. 그런데 2015년부터는 개인의 활동 영역에서 느끼는 행복이 강조되고 있음을 알 수 있다. 이 시기에 여행, 맛 집 탐방, 게임, 소확행 등 개인 여가에 대한 언급 비중이 증가하는 것을 볼 수 있다. 결론적으로 경기도 지역에서 행복에 영향을 미치는 요인은 사회정책과 지역 현안에서 개인의 실생활이나 활동 영역으로 점차 변화하는 추세이다.

## 2. 경기도 행복지도

경기도 지역 행복 수준과 감정의 지역별 파급효과를 동시에 살펴보기 위해 명사형 주제어와 함께 도출된 감성형용사를 별도로 분류해 정리하였다. ‘행복’ 그리고 경기도 31개 시·군의 ‘지역명칭’과 함께 도출된 감성형용사를 다시 긍정 표현과 부정 표현으로 구분하여 집계하였다.

마지막으로 이 결과 값을 기반으로 지도에 매핑하는 작업을 통해 '경기도 행복지도'를 구현하였다.

이를 위하여 우선 경기도 31개 기초자치단체를 분석단위로 설정하였다. 여기서 각 지역의 명칭과 행복에 대한 웹 크롤링 검색결과를 앞선 <표 2>에 제시한 10개 영역별 감성형용사 단어사전을 기준으로 재분류하였다. 지역별 그리고 시기별 긍정적, 부정적 형용사 어휘에 대한 언급 횟수는 아래 <표 4>와 같다.

분석결과를 논의하기에 앞서, 웹 크롤링 대상인 소셜 미디어 사용자들이 만든 빅데이터 자료에 대한 신뢰도를 검증하였다. 여기서는 텍스트 분석의 신뢰도 검증을 위해 TF-IDF 가중치 값을 활용하여 행렬을 구축하고 Mantel 검정(Mantel test)도 실시하였다(박승리·김진화·백승익, 2017). 그리고 경기도 31개 지역별 감성형용사 웹 크롤링 결과의 신뢰성을 검증하기 위해 대조군을 설정하여 비교하였다. 즉 소셜 미디어 사용자들에 의해 경기도 지역명칭, 행복 그리고 감성형용사가 함께 언급된 자료의 신뢰성을 검증하고자 공식적인 언론 기사를 동일한 방법으로 분석하였다. 이때 소셜 미디어 이용자들이 언급한 비속어는 분석에서 제외하였다. 나아가 두 행렬 간 상관관계의 통계적 유의성을 확인하기 위해 Mantel 검정을 실시하였다. 여기서는 연도별 긍정 표현과 부정 표현을 구분해 분석했는데, 전체 언론 노출 빈도와 소셜 미디어에서 도출된 감성형용사의 언급 횟수에 대한 Mantel 상관계수  $r$ 값은 <표 4>에서 보듯이 0.63~0.86 수준으로 전반적으로 상관성이 뚜렷했으며, 상관계수 10개 모두 95% 신뢰수준에서 통계적 유의성을 가진다.

경기도 전체를 대상으로 한 행복 관련 긍정 및 부정 표현의 언급 횟수 추세의 특징은 다음과 같다. 2013년의 행복에 관한 긍정 표현은 14,103건으로 부정 표현 11,928건 보다 높게 나타났으나, 세월호 사고가 발생한 2014년에는 긍정 표현은 13,139건으로 전년 대비 소폭 하락한 반면 부정 표현은 20,060건으로 급증했다. 2015년도 긍정 표현은 13,637건으로 2014년과 대동소이하나, 부정 표현은 세월호 사건에 대한 여파와 함께 경기도 북부 접경지역에서 발생한 안보위기 상황 등으로 인해 21,549건으로 증가하였다. 하지만 2016년부터 다시 긍정적인 표현이 부정적인 표현에 비해 언급 횟수가 많아지는 추세이다. 2016년에는 긍정 표현과 부정 표현은 각각 22,097건과 19,941건으로 나타났으며, 대통령 선거가 치러진 2017년에는 긍정 표현 29,603건 부정 표현 18,301건으로 전년에 비해 격차가 커졌다.

〈표 4〉 지역별 감성형용사 언급 빈도, 2013~2017

지역	2013년		2014년		2015년		2016년		2017년	
	긍정	부정	긍정	부정	긍정	부정	긍정	부정	긍정	부정
수원시	898	391	675	1,074	520	1,279	1,793	968	2,114	995
성남시	1,043	194	692	503	863	560	1,400	332	2,067	414
의정부시	254	634	288	668	266	684	472	629	592	640
안양시	759	451	231	804	248	855	497	758	618	778
부천시	371	671	171	998	189	1,012	410	997	637	962
광명시	713	220	649	647	628	697	829	625	1,149	600
평택시	569	399	353	715	381	728	575	644	714	589
동두천시	96	128	184	299	201	683	388	381	495	407
안산시	154	592	220	1,528	224	1,856	431	2,021	462	1,772
고양시	428	325	810	650	861	666	1,048	632	1,769	565
과천시	746	266	697	624	701	608	914	575	1,324	512
구리시	552	187	661	571	676	535	907	510	1,077	479
남양주시	567	161	817	337	821	331	1,036	321	1,652	296
오산시	392	432	159	626	175	670	505	681	891	642
시흥시	186	523	111	735	122	741	312	675	509	645
군포시	583	120	543	431	537	413	847	400	946	388
의왕시	444	200	424	511	429	500	676	461	857	442
하남시	850	314	800	417	719	404	977	354	1,641	320
용인시	701	297	866	449	872	442	1,097	392	1,976	341
파주시	465	464	512	673	510	698	691	665	737	635
이천시	518	673	408	770	427	747	655	686	725	601
안성시	392	224	422	576	500	601	743	587	829	533
김포시	267	618	231	995	246	988	488	878	636	674
화성시	220	417	157	534	200	555	418	563	559	516
광주시	434	129	445	323	511	343	774	317	836	300
양주시	177	646	179	799	183	821	399	834	472	749
포천시	169	660	187	744	202	873	415	888	469	787
여주시	178	276	190	444	231	490	493	516	611	486
연천군	112	751	128	850	244	975	439	958	485	735
가평군	513	142	557	213	534	224	798	218	921	145
양평군	352	423	372	552	416	570	670	475	833	353
합계	14,103	11,928	13,139	20,060	13,637	21,549	22,097	19,941	29,603	18,301
Mantel statistic r	0.867	0.712	0.742	0.631	0.728	0.695	0.751	0.722	0.795	0.794

다음으로 지역별 추이 변화를 직관적으로 확인하기 위해 <표 4>의 정보를 시각화하였다. 경기도 31개 시·군별 행복과 관련된 긍정 및 부정 형용사 언급 빈도를 단계구분도(choropleth map)로 표현하면 각각 <그림 2>, <그림 3>과 같다. 단계구분도로 시각화하기 위해 먼저 <표 4>의 언급 빈도를 지역별 인구규모를 고려한 수치로 바꾸었다. 아울러 지도를 통해 공간적 파급효과(ripple effects)를 확인하기 위해 공간적 차원의 가중치도 부여하였는데 여기서는 공간적 인접성(contiguity)을 기준으로 하는 Queen 방식으로 공간가중행렬(spatial weighted matrix)을 구축하였다. 마지막으로 그림에 나타나는 단계구분을 위해 계급분류의 한 방법인 표준편차분류법을 적용하였다.

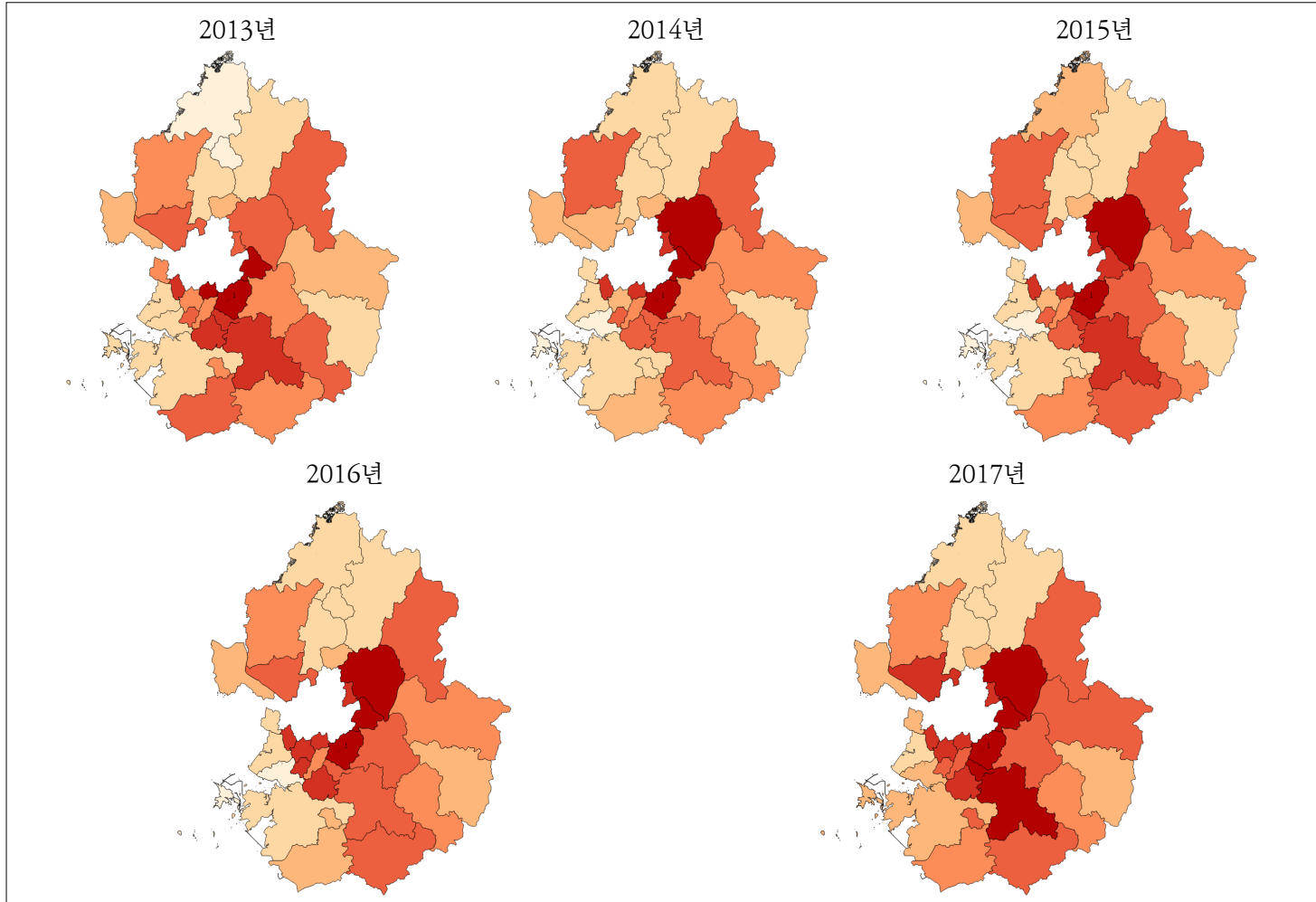
그림을 통해 알 수 있듯이, 긍정 및 부정 표현의 언급 빈도는 시공간적으로 상이하다. 먼저 <그림 2>에 제시한 긍정 표현의 경우 과천시, 광명시, 하남시, 성남시, 남양주시 같이 지역 기반시설이 잘 구축이 되어 있는 지역 또는 개발(혹은 재생)에 활기를 띠는 지역일수록 대체로 긍정적인 표현을 언급한 빈도가 높게 나타났다. 연도별 변화를 살펴보면 서울과 인접한 상당수 지역에서 농도가 점점 진해지고 있는데, 특히 2017년에는 이전에 비해 그 주변 지역까지 행복도가 상승하였다. 하지만 양주시, 의정부시 등은 근래 신도시 개발(가령 양주신도시)이나 도시환경정비사업 등으로 인해 지역이 활기를 띠고 지역주민의 삶의 질이 나아질 여지가 있음에도 긍정 표현이 뚜렷하게 나타나지 않는다. 또한 남양주시와 포천시 경우처럼 인접 지역 이터라도 상이한 결과를 보이는 곳도 존재한다. 따라서 공간적 인접성 그 자체(spatial contiguity *per se*)는 지역의 행복에 대한 긍정 표현을 결정하는 요인으로 단정하기 어렵다.

<그림 3>에서 제시한 부정 표현은 대체로 긍정 표현과 반대로 나타났다. 여기서 주목할 것은 2014년부터 수년간 세월호 사고와 관련이 있는 서남부 지역에서 부정적인 표현이 급증했다는 사실이다. 특히 2014년에는 비단 안산시뿐만 아니라 부천시, 시흥시, 수원시, 광명시, 안양시 등 인근 지역까지 행복 관련 부정 표현은 두드러지게 증가했다. 이듬해인 2015년에도 안산시와 인접한 상당수 지역에서는 부정 표현이 높은 수준을 유지했다. 이런 결과는 특정 지역의 불행의 감정이 다른 지역으로 확산될 수 있음을 경험적으로 보여준다. 다시 말해 경기도 지역에서 행복의 공간적 파급효과는 미약하거나 존재하지 않는데 반해, 불행의 파급효과는 뚜렷하다.

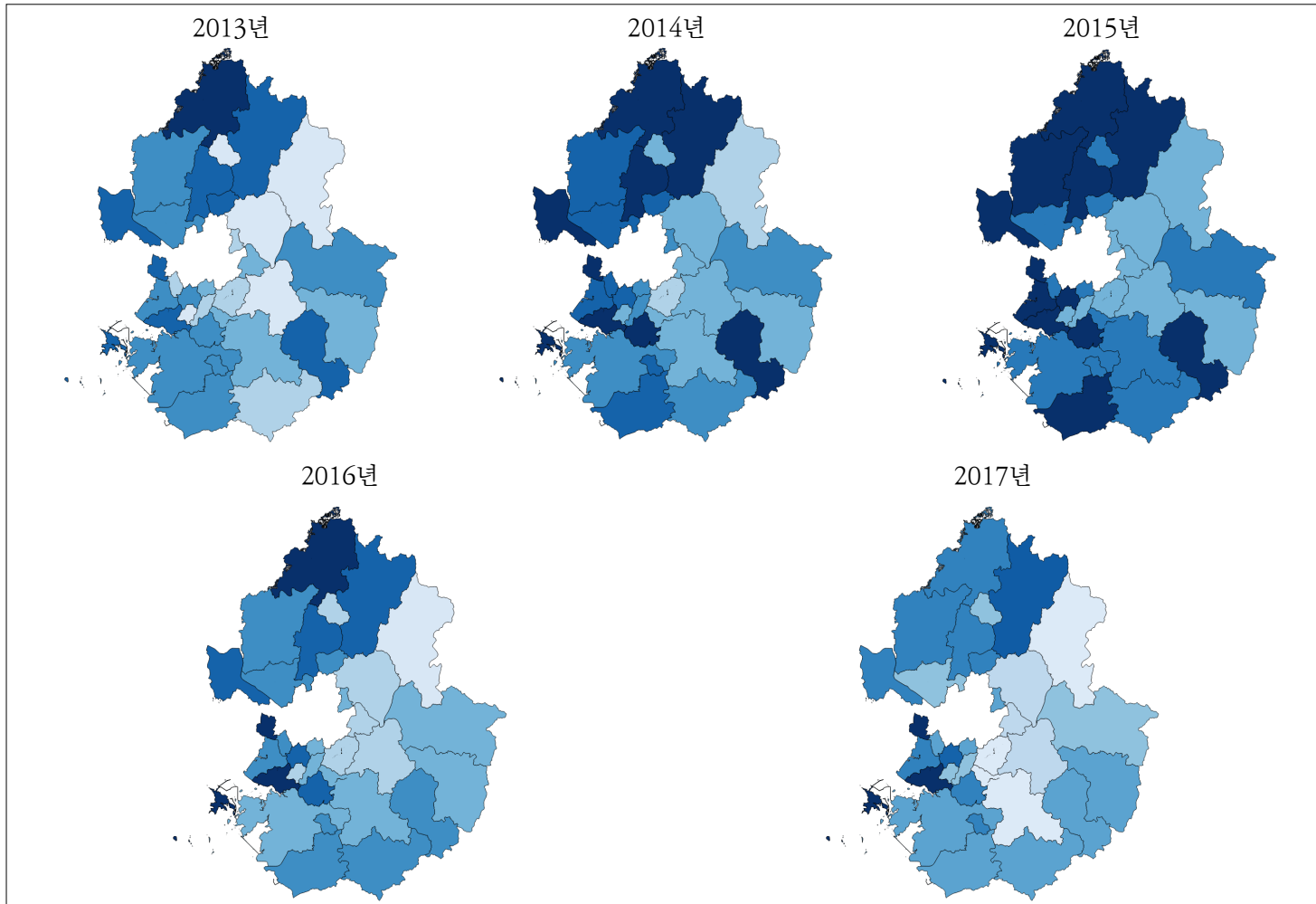
나아가 접경지역이나 그린벨트 등의 이유로 지역개발에서 상대적으로 소외된 경기도 북부 지역에서 부정적인 표현을 사용하는 빈도가 높다. 이 가운데 연천 지역은 지속적으로 부정 표현이 빈번했는데, 이는 해당 시기 냉랭했던 남북관계가 영향을 주었을 개연성이 존재한다. 특히 2015년에는 경기도 북부 대부분 지역에서 부정 표현이 매우 뚜렷한데, 여기에는 그해 8월 파주 지역 DMZ 목함지뢰 도발과 연이은 연천 지역 대북확성기 포격 사건이 상당한 영향을 주었을 개연성이 높다. 이는 지역주민이 체감하는 위협인식이 행복감에 유의미하게 부정적인 영향을 준다는 기존의 연구결과와도 일맥상통한다(이지은·강민성·이승중, 2012).



〈그림 2〉 긍정 표현에 대한 행복지도, 2013~2017



〈그림 3〉 부정 표현에 대한 행복지도, 2013~2017



## V. 요약 및 결론

이번 연구는 경기도 31개 기초자치단체를 대상으로 2013년부터 2017년까지 행복과 결부된 주요 연관어 및 긍정적, 부정적 감성어 언급 빈도 추세를 소셜 빅데이터를 통해 분석하였다. 이를 위해 우선 지역행복에 영향을 주는 여러 요인을 10개 영역으로 구분한 다음 영역별 주제어 사전을 구축하고, 이를 기준으로 빅데이터 분석(웹 크롤링)을 시행하였다. 경기도 지역에서 일상적으로 느끼는 행복에 대한 감정을 파악하기 위하여 불특정 개인이 소셜 미디어를 통해 언급한 비정형 자료를 분석하였다. 주요 분석결과는 다음과 같다. 지역의 행복과 상관성이 높은 요인 가운데 주거의 안정성과 함께 취업, 구직활동 등 일자리와 관련된 언급이 높은 비중을 차지하며 특히 집값, 내 집 마련과 같은 주거문제는 시기를 막론하고 두드러진 연관 주제로 나타났다.

한편 소셜 빅데이터 자료에서 나타나는 행복 관련 핵심 주제는 시기별로 차이를 보인다. 2013년과 2014년은 고용이나 집값 문제와 함께 자원봉사나 기부와 같은 사회참여에 관한 언급 빈도가 높았다. 하지만 2015년부터는 주로 맛집탐방, 여행, 게임, 휴식 등 개인의 일상적 행위에 대한 언급 횟수가 높아지는 추세를 보인다. 이런 결과는 사람들이 추구하는 가치와 행복을 얻는 요인이 변하고 있음을 시사한다. 다시 말해 지난 5년 동안 행복과 밀접하게 관련된 핵심 주제는 정부의 정책이나 사회참여 활동에서 사생활 특히 휴식과 여가활동으로 옮겨가는 추세라고 해석할 수 있다. 최근 회자되는 ‘소확행’이라는 신조어는 바로 이런 사회적 분위기를 반영하는 것으로 간주할 수 있는데, 이번 연구에서는 유의미한 행복 연관어로 2017년 처음 등장했다.

또한 감정 상태를 통해 지역의 행복 수준을 살펴보면, 사회적 파장을 야기하는 부정적인 사건이 지역 전체의 감정적 웰빙에 주는 영향은 두드러지는 것으로 나타났다. 즉 감정의 공간적 파급력은 행복할 때에 비해 불행할 때 더 크게 나타난다. 예를 들면 세월호 사건의 중심 지역인 안산시를 보면 부정적 표현은 해당 지역이 위치한 경기도 서남부 인접 지역까지 파급됨을 확인할 수 있다. 이와 유사하게 2015년 여름 경기도 접경지역에서 발생한 목함지뢰 도발 및 대북확성기 포격 사건은 경기도 북부 대부분 지역에 영향을 주었다. 요컨대 지역민이 느끼는 불안과 위협 인식은 행복감에 부정적인 영향을 주며 이는 인근 지역으로 파급된다.

이번 연구의 의의는 다음과 같다. 그동안 빅데이터와 관련된 연구는 주로 교통, 시설입지, 상권분석 등 지역 행동패턴을 분석하는 것에 초점을 맞춰왔다. 그리고 이러한 연구는 미래의 결과를 예측하거나 물리적 환경 개선을 위한 자료로 활용되고 있다. 하지만 이번 연구는 물리적 환경이 아닌 실제 사람들이 느끼고 생각하는 행복이란 정서적 주제를 소셜 빅데이터를 활

용해 분석했으며, 이를 통해 지방행정의 궁극적 목적인 주민의 복리증진을 달성하기 위한 지역정책 개발에 쓰일 수 있는 단초를 제공하였다.

나아가 이 연구는 웹 크롤링 방식으로 수집한 소셜 빅데이터를 통해 지역주민의 행복 수준을 측정함으로써 그동안 거시 행정통계와 서베이 자료 중심의 연구가 가지는 문제점을 극복하고자 했다. 행정통계는 주민의 주관적인 체감도를 반영하지 못하는 생태학적 오류(ecological fallacy)의 가능성을 내포하며, 주민 대상 서베이는 낮은 응답률 그리고 신뢰할 만한 표본 크기를 확보하기에 비용이 많이 들고 오랜 시간이 걸린다는 한계가 존재한다. 물론 소셜 빅데이터로 수집된 정보는 본질적으로 소셜 미디어 활동을 적극적으로 하는 사람들의 의견을 과도하게 반영한다는 표본의 대표성 문제가 발생할 여지가 존재한다. 이런 한계에도 불구하고, 소셜 빅데이터는 신뢰할만한 서베이 자료가 부재할 경우 선택할 수 있는 대안이다. 만일 대규모 서베이 자료가 있다면, 텍스트 중심의 비정형 소셜 빅데이터와 숫자로 코딩된 서베이 자료는 상호보완적 관계를 가지며 그 분석결과는 지역의 행복 수준에 대한 총체적인 이해를 가능하게 한다.

## 【참고문헌】

- 강범모. (2014). 텍스트 맥락과 단어 의미: 잠재의미분석. 「언어학」, 68: 3-34.
- 김선기·박승규. (2014). 「지역생활권 삶의 질 분석 및 지역격차 해소방안」. 한국지방행정연구원.
- 김승권·장영식·조홍식·차명숙. (2008). 「한국인의 행복결정요인과 행복지수에 관한 연구」. 한국보건사회연구원.
- 김윤태. (2010). 행복지수와 사회문화적 분석. 「문화경제연구」, 13(1): 23-45.
- 박경미·박호진·김형곤·고희동. (2011). SNS에서 오피니언마이닝 연구. 「정보과학회지」, 29(11): 54-60.
- 박승리·김진화·백승익. (2017). 텍스트 마이닝을 이용한 빅 데이터의 신뢰도 검증에 대한 연구. 「경영정보관련 춘계통합학술대회 논문집」, 2017: 524-529.
- 변미리. (2013). 「도시발전의 질적 성장 모니터링: 행복지수와 메가시티인덱스 구축 연구」. 서울연구원.
- 변미리. (2014). 「메가시티 삶의 질과 서울형 행복지표」. 서울연구원.
- 신동윤·송유미. (2017). 빅데이터를 이용한 서울시 행복지구 분석 및 예측을 위한 실험 및 고찰. 「KIBIM Magazine」, 7(1): 1-8.
- 양승준·이보연·김희웅. (2015). 행복·불행 관련 요인에 대한 토픽모델링 분석 연구. 「한국경영정보학회 춘계통합학술대회 논문집」, 2015: 577-591.
- 우창빈. (2013). 행복을 추구하는 행복정책은 가능한가?: 행복을 목표로 하는 공공정책의 이론, 논란, 가능성과 조건들. 「행정논총」, 51(2): 283-381.
- 윤은기. (2012). AHP (Analytic Hierarchy Process) 기법을 이용한 삶의 질 평가요인들의 상대적 중요성 분석. 「한국행정학보」, 46(2): 395-419.
- 이병호·손웅비. (2017). 삶의 질 측정을 위한 종단 조사 설계 연구: 경기도민 삶의 질 조사의 사례를 중심으로. 「GRI연구논총」, 19(3): 1-23.
- 이승철. (2011). 도시유형별 삶의 질 수준과 영향요인. 「지방정부연구」, 15(2): 97-119.
- 이윤진·김남숙·기영화. (2016). 도시 공동주택 기반 주민 공동체 학습과정 사례연구. 「평생학습사회」, 12(1): 195-221.
- 이지은·강민성·이승중. (2012). 도시위험인식이 행복감에 미치는 영향: 서울시 주민을 중심으로. 「지방정부연구」, 18(2): 559-588.
- 이희걸·심수진. (2010). 사회지표 개편 방향 탐색. 「한국사회」, 11(1): 47-77.
- 임좌상·김진만. (2015). 한국어 트위터 감정의 핫스팟 분석. 「멀티미디어학회 논문지」, 18(2): 233-243.

- 장효진. (2001). 감정동사 및 감정형용사 분류에 관한 연구. 「한국정보관리학회 학술대회 논문집」, 2001(8):29-34.
- 한준 외. (2011). 「국민 삶의 질 측정을 위한 분석틀」. 한국사회학회.
- 황명진·심수진. (2008). 한국의 행복지수 개발. 「조사연구」, 9(3): 93-117.
- Burton, E. (2003). Housing for an Urban Renaissance: Implications for Social Equity. *Housing Studies*, 18: 537-562.
- Dempsey, N. et al. (2009). The Social Dimension of Sustainable Development: Defining Urban Social Sustainability. *Sustainable Development*, 19: 289-300.
- Durand, M. (2015). The OECD Better Life Initiative: *How's Life?* and the Measurement of Well-being. *Review of Income and Wealth*, 61(1): 4-17.
- Golder, S. A., & Macy, M. W. (2011). Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength across Diverse Cultures. *Science*, 333(6051): 1878-1881.
- Helliwell, J., Layard, R., & Sachs, J. (2018). *World Happiness Report 2018*. New York: Sustainable Development Solutions Network.
- Kitchin, R., Lauriault, T. P., & Wilson, M. W. (Eds.). (2017). *Understanding Spatial Media*. Sage.
- Mayer-Schönberger, V. & Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Think and Work*. London: John Murray.
- Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep Learning Applications and Challenges in Big Data Analytics. *Journal of Big Data*, (2015)2: 1.
- OECD. (2016). *Society at a Glance 2016: OECD Social Indicators*. Paris: OECD Publishing.
- OECD. (2017). *How's Life? 2017: Measuring Well-being*. Paris: OECD Publishing.
- Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Lucas, R. E. et al. (2013). Characterizing Geographic Variation in Well-Being Using Tweets. *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media (ICWSM)*, pp. 583-591.
- Stiglitz, J. E., Sen, A., & Fitoussi, J. (2009). *Report by the Commission on the Measurement of Economic Performance and Social Progress*. Paris: Commission on the Measurement of Economic Performance and Social Progress.

- White, A. (2009). A Global Projection of Subjective Well-being. The First Published Map of World Happiness.  
[http://news.bbc.co.uk/1/shared/bsp/hi/pdfs/28\\_07\\_06\\_happiness\\_map.pdf](http://news.bbc.co.uk/1/shared/bsp/hi/pdfs/28_07_06_happiness_map.pdf)
- Yang, Y. (2008). A Tale of Two Cities: Physical Form and Neighborhood Satisfaction in Metropolitan Portland and Charlotte. *Journal of the American Planning Association*, 74: 307-323.

---

**이 병 호:** 미국 University of Michigan at Ann Arbor에서 2011년 사회학 박사학위를 취득하였고 현재 아주대학교 사회학과 조교수로 재직하고 있다. 주요 관심분야는 삶의 질, 인구, 가족, 정체성, 중국사회 등이다. 주요 논문으로는 “삶의 질 측정을 위한 종단 조사 설계 연구”(2017), “노동시장 공간의 이중구조화와 불평등”(2017), “다자녀 출산의 결정요인 연구”(2017) 등이 있다. (byungholee@ajou.ac.kr)

**김 가 연:** 이화여자대학교 및 동 대학원에서 행정학 석사학위(2013, 논문: 서울시 주택바우처의 주거지 수요요인 분석)를 취득하였고 현재 경기연구원 연구원으로 재직하고 있다. 주요 관심분야는 도시의 지속가능한 성장, 도시 삶의 질, 공간계량 및 매핑, 머신러닝을 활용한 빅데이터 분석, 소셜 네트워크 분석을 이용한 지역 현상 분석 등이다. (soleil419@gmail.com)

