

의료 빅데이터를 활용한 강원도 사망 원인 및 특성 분석

Analysis of Mortality Cause and Properties using Medical Big Data in Gangwon

정대현*, 권오영*, 구영덕**
강원연구원*, 한국과학기술정보연구원**

Dae-hyun Jeong(gregori79@rig.re.kr)*, O-young Kwon(bayside5050@rig.re.kr)*,
Young-duk Koo(ydkoo@kisti.re.kr)**

요약

의료정보의 급속한 발달로 인하여 막대한 의료 빅데이터 정보가 축적되고 있다. 이러한 의료 빅데이터는 고령화와 의료비의 가파른 상승을 해결하는데 중요한 자료로 활용될 가능성이 큰 분야이다. 특히 한국의 경우 전 국민대상 주민등록번호와 전산화된 이용 자료들을 기반으로 빅데이터를 활용할 수 있는 의료 인프라 측면에서는 다른 국가에 비해 우위를 점할 수 있다. 본 연구에서는 국민건강보험공단 빅데이터 자료와 통계청 사망자 데이터를 활용하여 강원도의 사망자 분석 및 사망률에 영향을 미치는 요인을 분석하기 위해 상관분석을 수행한 후 시사점을 도출하였다. 분석 결과, 강원도 사망률과 관계된 주요 변수로는 자기지역 병원 인프라 이용률, 소득수준, 고령화 및 인구 밀도 등으로 나타났다. 따라서 소득격차에 따른 불평등과 지역 의료 인프라 부족 등은 지역의 사망률에 부정적 영향을 미치고 있는 것으로 나타나고 있어, 지역 병원 인프라 개선 및 소득 수준 향상을 위한 지역별 맞춤형 정책지원이 필요하다. 본 연구의 결과는 의료 빅데이터를 활용하여 강원도민 사망자를 분석하고, 다양한 사회 지표와 상관분석을 통해 그 원인을 파악해 보았다는 점에서 연구의 의의가 있다.

■ 중심어 : | 의료 빅데이터 | 사망률 | 상관분석 |

Abstract

Due to the rapid development of medical information, vast amounts of medical data are accumulating, and such medical data is highly likely to be used as an important data for solving the aging population and the rapid rise in medical cost. Especially in Korea, there are resident registration numbers and computerized usage data for all citizens, so it can be superior to other countries in terms of medical infrastructure that can utilize big data. The purpose of this study was to analyze the factors affecting the mortality and death rate of Gangwon using the Big Data and the National Statistical Office data centered on Kangwon province. As a result of analysis, major variables related to the mortality rate of Gangwon were hospital infrastructure utilization rate, income level, aging population and population density. Therefore, inequalities due to income disparities and insufficient local medical infrastructures were affecting the local mortality rate, and policy support was needed to improve the local hospital infrastructure and income level. The results of this study were meaningful in that medical big data were used to analyze the deaths of people in Gangwon, and the causes of the deaths were analyzed through various social indicators and correlation analysis.

■ keyword : | Medical Big Data | Mortality Rate | Correlation Analysis |

* 본 연구는 한국과학기술정보연구원 연구과제로 수행되었습니다.

접수일자 : 2018년 08월 06일

수정일자 : 2018년 08월 28일

심사완료일 : 2018년 08월 28일

교신저자 : 구영덕, e-mail : ydkoo@kisti.re.kr

I. 서론

의료정보의 급속한 발달로 인하여 막대한 양의 정보가 쌓이고 있다. 이러한 정보를 이용하여 임상연구를 하고자하는 수요가 증가하고 있으며, 고령화와 의료비의 가파른 상승을 해결해야 하는 문제에 직면해 있다. 이에 대한 대안으로 빅데이터의 활용에 대한 목소리가 높다[1]. 빅데이터(big data)는 기존의 데이터베이스 관리도구가 수집, 저장, 관리, 분석하는 범위를 넘어서는 규모의 데이터로 정의되지만, 이의 활용은 양적인 개념에 국한되지 않고 데이터의 분석과 활용의 전 과정을 포괄한다[2]. 빅데이터는 방대한 자료의 분석을 통해 미래를 예측하는 기반을 마련해 주기 때문에 국가의 미래 경쟁력을 좌우할 수 있는 핵심요인으로 평가 받고 있다. 세계경제포럼은 2012년 가장 주목해야 할 과학기술로 빅데이터 분야를 선정하였다[3]. 또한 여러 연구에서 공공분야에 빅데이터를 활용 시 미국의 추가 생산성은 1% 향상되며, 영국정부의 총 예산 중 2.5-4.5%가 절감될 수 있다고 예측하였다[4]. 빅데이터에 대한 연구는 미디어 분석, 문화콘텐츠 분야 이외에도 재난관리 등 다양한 분야에서 빅데이터를 활용한 연구가 진행 중이다[5-7]. 보건 의료 분야에서도 이러한 현상은 예외가 아니다. 매킨지는 '보건의료 분야의 빅데이터 혁명'이라는 보고서에서 빅데이터를 통해 근거중심 의로서 비스 제공 및 의료진반에 걸친 패러다임 변화가 가능하며, 혁명의 핵심은 자료원이며 변화를 위한 전제조건은 자료원간 연계라고 제시하였다[8]. 특히, 영국 보건부는 2013년에 Personalized Health and Care 2020을 발표하여 분산된 의료 데이터들을 수집, 저장, 연계, 분석하는 독립조직을 설립하여 의료 빅데이터를 활용하고 있다. 미국의 경우에는 민간 보험사들을 중심으로 다양한 정보 통합 플랫폼이 활성화되도록 간접적인 역할을 수행하여 데이터 상호운용성을 높이기 위한 정보제공을 진흥하고 있다[9][10].

우리나라는 전 국민대상 주민등록번호와 전산화된 의료 이용자료들이 있기에 빅데이터를 활용할 수 있는 인프라 측면에서는 선진국과의 경쟁에서 충분히 비교우위를 점할 수 있다. 이를 이용하여 질병의 원인적 연

관성, 치료효과 파악 및 보건의료 분야 정책 입안에 필요한 근거들을 효율적으로 생성할 수 있다. 실제 국민건강보험공단, 국민심사평가원에서는 전 국민의 의료 정보를 구축하고 이를 서비스하고 있다. 특히 국민건강보험공단에서는 영양기관 정보, 검진기관 정보 및 장기요양 기관 정보를 통합 구축하여 서비스 하고 있으며, 통계청 사망자 데이터와 연계도 가능하다. 그러나 고령화, 소득수준, 사망원인 등은 지역별로 다르기 때문에, 미시적으로 빅데이터를 활용한 분석이 이루어져야 정확한 정책 도출이 이루어짐에도 불구하고 대부분 전국을 대상으로 분석이 이루어지고 있다.

본 연구에서는 강원도를 중심으로 국민건강보험공단 의료 빅데이터와 통계청에서 제공하는 사망자 분석 정보를 활용하여 사망자 분석 및 사망률에 영향을 미치는 주요 요인들에 대해 살펴보고, 그 시사점을 도출하였다.

강원도는 저출산·고령화가 심화되고 있으며, 연도별 사망자도 2009년부터 크게 증가하고 있다. 또한 전국 17개 광역시·도의 소득 상위 20%와 하위 20% 집단 간 기대수명격차를 살펴보면 강원도는 전남과 더불어 격차가 7.6년이 차이 나고 있어 최하위를 기록하고 있다 [6].

따라서 고령화와 의료비의 가파른 상승, 소득격차에 따른 건강불평등 문제 해결을 위해 의료 빅데이터를 활용하여 그 원인을 파악하고자 하였다.

II. 방법론

본 연구에서는 강원도 주요 사망자 데이터를 분석하고, 각 변수간 상관분석을 통해 강원도 사망률에 영향을 미치는 주요 변수들이 어떠한 것인지를 파악하였다. 이를 위해, 국민건강보험공단에서 제공하는 맞춤형 DB를 신청하여 국민건강보험공단 내 빅데이터 서버에서 SAS 프로그램을 활용하여 2002년부터 2015년까지 강원도민 전수데이터 약 670GB 데이터를 구축하였다. 그 이후 구축된 데이터를 통계청 2015~2016년 강원도 주소로 된 사망자 21,826명의 데이터와 연계하여 분석을 수행하였다. 국민건강보험공단 빅데이터에서 제공하는

정보는 강원도에 주소가 등록된 환자의 보험료(20분위) 정보, 읍·면·동 단위의 주소 및 이용병원의 주소 정보가 기록되어 있으며, 통계청에서 제공하는 사망자 정보는 직업, 교육상태, 혼인상태, 사망원인 등이 기록되어 있다.

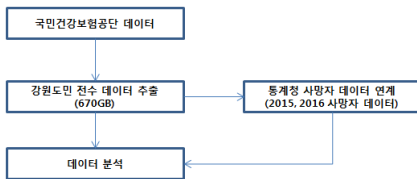


그림 1. 국민건강공단, 통계청 데이터 추출

주요 분석사항으로는 지역별 인구 10만명 당 사망자 수, 성별, 연령, 소득수준, 인프라 분석, 직업, 혼인상태, 교육정도, 사망장소, 사망원인 분석 등을 수행하였다. 지역별 인구 10만명 당 사망자 수 분석에 있어서 인구 정보는 통계청에서 제공하는 주민등록상 인구 수를 활용하여 분석을 수행하였다.

또한, 분석된 결과를 바탕으로 기타 통계청에서 제공하는 GRDP, 인구밀도, 지역별 고령화 지수 등과 같이 상관분석을 수행하여 지역 내 사망자와의 관계에 대해 분석을 수행하였다.

III. 분석결과

국민건강보험공단과 통계청 사망자 데이터를 연계하여 분석한 결과, 2015~2016년 강원도 내 사망자 성별 비중은 남성이 55%, 여성이 45%를 차지하여 상대적으로 남성의 비율이 높은 것으로 나타났다.

(단위: 명, %)

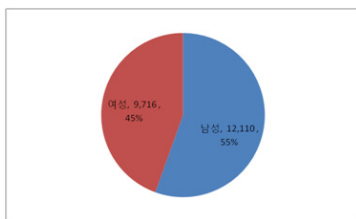


그림 2. 강원도 사망자 성별 비중

강원도 사망 연령비중을 살펴보면, 71세 이상의 노인층이 14961명으로 약 68.5%를 차지하여 가장 많은 비중을 차지하고 있으며, 고령화에 따라 사망률이 증가하고 있는 것으로 나타났다.

(단위: 명, %)

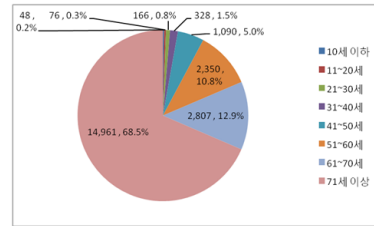


그림 3. 강원도 사망자 연령 비중

사망자의 국민건강보험공단 20분위를 분석하면, 소득이 전혀 없거나 균인으로 분류되는 0분위가 3,923명으로 전체 18%, 차상위 계층인 1분위의 경우 1,231명으로 5%, 2~10분위에 해당하는 사람이 6,260명으로 29%를 차지하여 전반적으로 소득 수준이 낮은 것으로 나타났다.

(단위: 명, %)

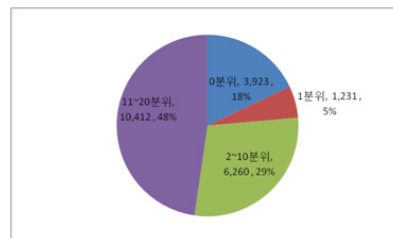


그림 4. 강원도 사망자 소득수준

강원도 소득격차가 심화되는 상황에서 강원도 사망자 소득수준이 가장 낮은 차상위 계층 비율을 지역별로 살펴보면, 정선군과 고성군이 각각 8.7%, 8.6%로 나타나 상대적으로 사망자 중 차상위계층의 비율이 높은 것으로 나타났다.

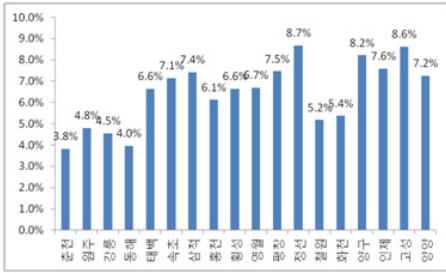


그림 5. 강원도 사망자 지역별 사망자 차상위계층 비율

강원도 전체 사망자의 직업 분포를 살펴보면, 학생, 가사 및 무직의 비율이 15,636명으로 전체 72%를 차지하여 대부분이며 농업, 임업 및 어업 속련 종사자가 2,973명으로 13%로 그 뒤를 잇고 있다.

<단위: 명, %>

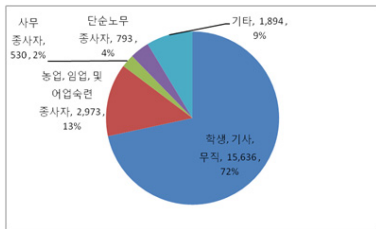


그림 6. 강원도 사망자 직업 분포

강원도 혼인 상태를 살펴보면, 배우자가 있거나 사별한 경우가 각각 10,249명, 8,571명으로 47%, 39%를 차지하여 대부분을 차지하고 있다.

<단위: 명, %>

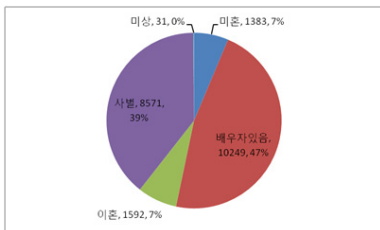


그림 7. 강원도 사망자 혼인상태 분포

강원도 교육상태를 살펴보면, 불취학에서 고등학교 이하의 비중이 19,787명으로 전체 91% 대부분인 것으로 나타났다.

<단위: 명, %>

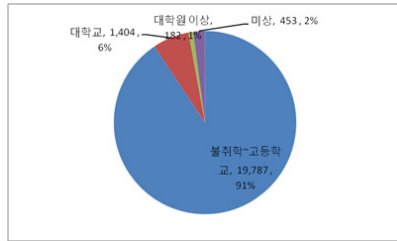


그림 8. 강원도 사망자 교육상태 분포

강원도 사망장소를 살펴보면, 의료기관에서 사망한 사람이 16,075명으로 74%를 차지하여 큰 비중을 차지하고 있으며, 주택에서 사망한 비율도 4,304명으로 전체 20%를 차지하고 있다.

<단위: 명, %>

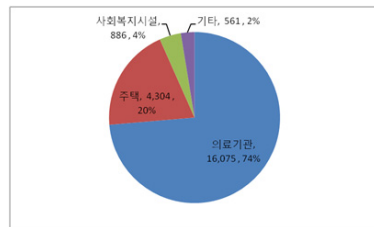


그림 9. 강원도 사망자 장소 분포

강원도 사망자 사망원인을 살펴보면, 악성신생물(암질병)이 6,030명으로 전체 28%를 차지하고 있다. 특히 폐암이 암질병으로 인한 사망자 중 1,402명으로 가장 많은 사망자를 나타내고 있다.

<단위: 명, %>

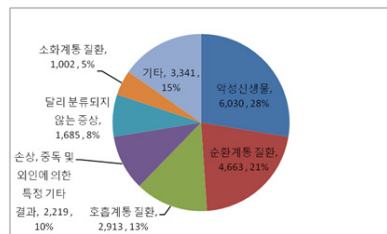


그림 10. 강원도 사망자 사망원인 분포

(단위: 명)

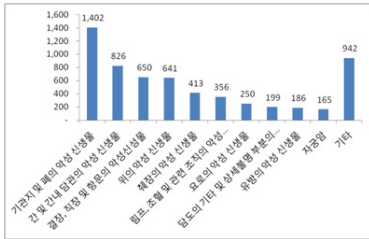


그림 11. 강원도 사망자 사망원인 주요 암질병 사망자

강원도 사망자 인구 10만명당 사망자 수를 살펴보면, 양양군이 인구 10만명당 사망자가 1,009.2명으로 가장 높은 사망률을 기록하고 있으며, 영월군, 정선군이 그 뒤를 잇고 있다.

(단위: 명)

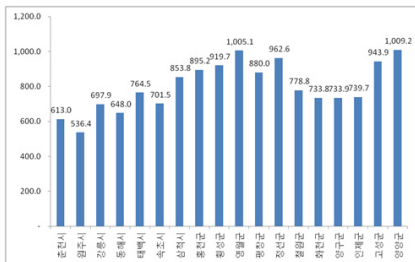


그림 12. 강원도 사망자 지역별 인구 10만명당 사망자 수

강원도 자살자 인구 10만명당 사망자 수를 살펴보면, 평창군이 인구 10만명당 사망자가 45.7명으로 가장 높은 사망률을 기록하고 있으며, 양구군, 정선군이 그 뒤를 잇고 있다.

(단위: 명)

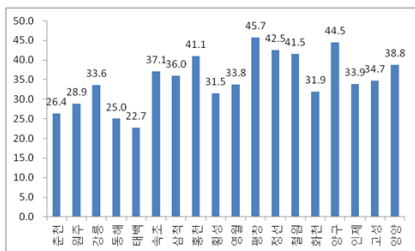


그림 13. 강원도 사망자 지역별 인구 10만명당 자살자 수

국민건강보험공단 데이터 상, 수진자의 주소와 해당 수진자가 이용한 병원 주소를 매칭시켜 분석하면, 그림에서와 같이 춘천권, 속초권, 원주권 및 강릉권을 중심으로 병원 인프라를 이용하고 있는 것으로 나타났다. 상대적으로 춘천, 원주 및 강릉 지역에 거주하는 주민들은 살고 있는 지역의 병원을 이용하는 비율이 높았으며, 강원 남부 지역의 경우 자기 지역의 병원을 이용하는 비율이 매우 낮은 것으로 분석되었다.

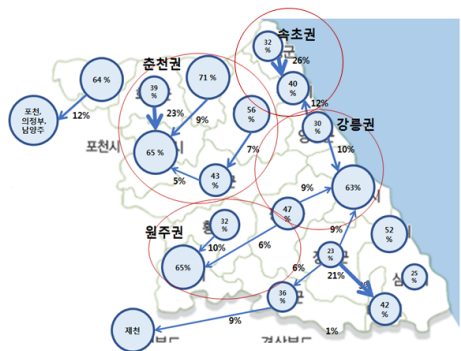


그림 14. 강원도 데이터 상 수진자 주소와 병원 이용 주소 간 관계망

강원도 주민의 국민건강보험공단 데이터와 통계청 사망자 데이터를 연계하여 강원도 18개 시·군별로 분석한 결과와 통계청에서 발표한 18개 시·군별 지역별 고령화 비율, GRDP, 인구밀도 데이터 사이에 상관분석을

표 1. 강원도 주요 통계분석 결과간 상관분석 표

| | 전체 사망률 | 자살률 | 병원인프라 이용 비율 | 고령화 | GRDP | 인구밀도 | 사망자차상위 비율 |
|-------------|---------|--------|-------------|---------|---------|--------|-----------|
| 전체 사망률 | 1 | | | | | | |
| 자살률 | 0.413 | 1 | | | | | |
| 병원인프라 이용 비율 | -.766** | -0.091 | 1 | | | | |
| 고령화 | .977** | 0.317 | -.750** | 1 | | | |
| GRDP | -.678** | -0.332 | .575* | -.629** | 1 | | |
| 인구밀도 | -.599** | -0.325 | 0.207 | -.623** | 0.377 | 1 | |
| 사망자차상위 비율 | .653** | .525* | -.530* | .563* | -.679** | -0.393 | 1 |

* p<0.05, ** p<0.01

수행한 결과, 지역 소득수준이 낮은 지역일수록, 인구밀도가 낮은 지역일수록, 자기 지역의 병원을 이용하는 비율이 낮을수록 인구 10만명 당 사망자 수가 높은 것으로 나타났다. 또한 사망자 중 차상위 비율이 높은 지역일수록, 고령화가 높은 지역일수록 인구 10만명 당 사망자의 수가 높은 것으로 나타났다.

IV. 결론

지금까지 국민건강보험공단 데이터와 통계청 사망자 데이터를 연계하여 강원도민 2015년, 2016년 사망자 데이터를 분석하였다. 분석결과, 강원도 사망자는 61세 이상의 고령층이 큰 비중을 차지하고 있으며 저소득 계층을 중심으로 사망자 비중이 높은 것으로 나타났다. 특히 차상위 계층이 많은 지역일수록 지역에 사망자가 많은 것으로 분석되었다. 사망자의 교육수준은 전반적으로 낮은 것으로 나타났으며, 의료기관에서 사망하는 사람이 대부분이나 주택에서 사망하는 비중도 상대적으로 높은 것으로 나타났다. 주요 사망원인으로는 암질병에 의한 사망자가 가장 많았다. 특히 폐암, 간암 등에 의한 사망자가 가장 큰 비중을 차지하고 있는 것으로 나타났다. 지역별로 사망률을 살펴보면, 인구 10만명 당 사망률은 폐광지역이 상대적으로 높으나, 차상위 계층 사망률은 접경지역이 대체로 높은 것으로 나타났다. 또한 주요 통계분석 간 상관분석결과, 고령층이 많은 지역일수록 인구밀도가 낮고 병원 인프라 이용 비율이 낮아 고령층에 대한 사망률이 큰 것으로 분석된다.

이러한 분석결과를 바탕으로 본 연구에서는 다음과 같은 시사점을 도출하였다. 첫 번째, 강원도 건강 형평성은 전국 최하위권을 기록하는 상황에서 사망자의 소득수준 및 지역 경기 침체는 실제 강원도 사망률에 큰 영향을 미치는 요인이 되고 있다. 이는 강원도의 고령화가 심화되는 상황에서 소득이 낮은 사람들을 중심으로 건강 상태 점검 및 지속적인 관리가 필요하다. 두 번째, 강원도 주요 통계분석 결과 간 상관분석 결과에서도 나타난 바와 같이 지역 내 병원 인프라 수준은 해당 지역의 사망률에 큰 영향을 미치는 요인 중 하나이다.

따라서 지역 내 병원 인프라 개선 또는 확충을 통해 지역 내 사망률을 낮추기 위한 정책 수립이 필요하다. 마지막으로 자살율에 영향을 미치는 요인 중 하나가 경제적 수준으로 차상위 계층을 중심으로 자살예방 정책이 필요하다.

이를 위해서는 지역 내 인프라 확충 및 개선을 위해 강원도 내 지역거점 공공병원 추진 등의 정책이 필요하다. 지역거점 공공병원은 의료 취약지에 건립되어 지역 거점 의료기관의 역할을 수행하는 병원으로 지역 내 거점 의료기관에 대한 강원도 주요 사망원인을 중심으로 특성화 병원으로 나아가는 방향이 필요하다. 또한 지역의 의료인력(의사, 간호사) 구인난을 해결하기 위한 지역 대학들과의 연계, 협력을 통한 교육체계 개선 등의 전략 수립이 필요하다. 또한 지역 거점 병원을 가지 못하는 지역을 대상으로 원격의료 시스템 구축, 앰블런스 상설화 시스템 구축 등 상대적으로 빈곤층인 노인을 대상으로 지속적인 모니터링 시스템 구축이 필요하다.

본 연구의 결과는 의료 빅데이터를 활용하여 강원도민 사망자를 분석하고, 다양한 사회 지표와 상관분석을 통해 그 원인을 파악해 보았다는 점에서 연구의 의의가 있다. 국민건강보험공단과 통계청에서 제공하는 데이터는 전수 데이터이기 때문에 의료 상태를 파악하는데 있어 정확한 분석이 가능하다. 이러한 의료 빅데이터는 다양한 사회 지표와 같이 연계하여 분석을 하는 경우 현실적인 정책 대안을 수립이 가능하기 때문에 연구의 가치가 있다. 그러나 본 연구는 강원도 사망자 데이터를 활용한 원인분석에 있어 의료 빅데이터를 활용하여 요약 통계량 및 상관분석 정도에 그쳤다는 한계가 있다. 향후 연구에서는 요인분석 등을 통해 변수 간 선후 관계를 파악하여 의료 빅데이터를 활용한 건강관리 시스템 구축 등에 대한 연구가 추가로 필요하다.

참고 문헌

- [1] 김성수, “의료기관의 빅데이터 활용 방안에 대한 연구,” 한국디지털정책학회, Vol.12, No.2, pp.397-407, 2014.

[2] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, and C. Roxburgh, *Big data: the next frontier for innovation, competition, and productivity*, Lexington: McKinsey; 2011.

[3] World Economic Forum. Big data, big impact: new possibilities for international development. Geneva: World Economic Forum; 2012.

[4] S. J. Ko and Y. H. Chung, "Health risk prediction using big health data," Health welf Policy Forum, Vol.193, pp.43-52, 2012.

[5] 이병엽, 임종태, 유재수, "빅 데이터를 이용한 소셜 미디어 분석 기법의 활용," 한국콘텐츠학회논문지, Vol.13, No.2, pp.211-219, 2013.

[6] 이오준, 박승보, 정다울, 유은순, "소셜 빅데이터를 이용한 영화 흥행 요인 분석," 한국콘텐츠학회논문지, Vol.14, No.10, pp.527-538, 2014.

[7] 신동희, 김용문, "국내 재난관리 분야의 빅 데이터 활용 정책방안," 한국콘텐츠학회논문지, Vol.15, No.2, pp.377-392, 2015.

[8] P. Groves, B. Kayyali, D. Knott, and S. V. Kuiken, *The 'big data' revolution in health care: accelerating value and innovation*, Lexington: McKinsey Global Institute; 2013.

[9] HSCIC. www.hscic.gov.uk/article/2656/

[10] ONC, HHS, Federal Health IT Strategic Plan 2015-2020, 2014.

권 오 영(O-young Kwon)

정회원



- 2010년 2월 : 연세대학교 행정학과(행정학 석사)
- 2013년 8월 : 연세대학교 행정학과(행정학 박사)
- 2015년 11월 ~ 현재 : 강원연구원 부연구위원

<관심분야> : 공공정책, 규제정책, 지방행정, 갈등관리 등

구 영 덕(Young-duk Koo)

정회원



- 1988년 2월 : 인천대학교 기계공학과(공학사)
- 1990년 2월 : 인천대학교 기계공학과(공학석사)
- 1998년 2월 : 인천대학교 기계공학과(공학박사)

▪ 1991년 4월 ~ 현재 : 한국과학기술정보연구원 책임연구원

<관심분야> : 빅데이터 정보분석

저 자 소 개

정 대 현(Dae-hyun Jeong)

정회원



- 2009년 8월 : 한양대학교 신소재공학과(공학석사)
- 2015년 8월 : 성균관대학교 기술경영학과(공학박사)
- 2015년 11월 ~ 현재 : 강원연구원 부연구위원

<관심분야> : 빅데이터, 과학계량학, 네트워크 분석, 기술혁신