JKIICE Journal of the Korea Institute of Information and Communication Engineering

한국정보통신학회논문지 Vol. 22, No. 6: 835~841, Jun. 2018

클라우드 기반의 공개의료 빅데이터 분석을 통한 삶의 질에 영향을 미치는 요인분석

김민경[1] · 조영복[2*]

An Analysis of Factors Affecting Quality of Life through the Analysis of Public Health Big Data

Min-kyoung Kim[1] · Young-bok Cho[2*]

[1]Department of Research & Development Center, SONOUM Inc, Cheongju, 28501, Korea
[2*]Department of Computer & Information Security, Daejeon University, Daejeon 34520, Korea

요 약

본 연구에서 공개 의료 빅데이터 분석을 지역사회건강조사 2012~2014년 자료를 이용해 개인의 건강관련 삶의 질 차이와 삶의 질에 영향을 미치는 요인을 분석하였다. 제안논문에서는 공개의료 빅데이터 분석을 위해 Hadoop 기반의 Spack을 이용해 병렬처리 지원을 위한 클라우드 메니저를 구성하고 개인의 삶의 질에 영향을 미치는 요인을 하드웨어의 제약없이 빠르게 분석하였다. 건강관련 삶의 질에 미치는 영향을 개인적 특성과 지역사회 특성으로 구분하여 단계별 다수준 회귀분석(ANOVA, t-test)을 실시하였다. 연구결과 개인별 삶의 질에 영향을 미치는 요인으로는 남자 평균 73.8점, 여자 평균 70.0점으로 남자가 여자보다 건강관련 삶의 질이 높은 것으로 나타났다.

ABSTRACT

In this study, we analyzed public health data analysis using the hadoop-based spack in the cloud environment using the data of the Community Health Survey from 2012 to 2014, and the factors affecting the quality of life and quality of life. In the proposed paper, we constructed a cloud manager for parallel processing support using Hadoop - based Spack for open medical big data analysis. And we analyzed the factors affecting the "quality of life" of the individual among open medical big data quickly without restriction of hardware. The effects of public health data on health - related quality of life were classified into personal characteristics and community characteristics. And multiple-level regression analysis (ANOVA, t-test). As a result of the experiment, the factors affecting the quality of life were 73.8 points for men and 70.0 points for women, indicating that men had higher health - related quality of life than women.

키워드 : 공개의료 빅데이터, 삶의질, ANOVA분석, 다수준 회귀분석, 하둡 스파크

Key word : Public Medical Big Data, Quality of life, ANOVA Analysis, Multi-level Regression Analysis, Hadoop Spack

Received 3 April 2018, **Revised** 5 April 2018, **Accepted** 21 April 2018

*** Corresponding Author** Young-Bok Cho(E-mail:ybcho@dju.ac.kr, Tel:+82-42-280-2406)
Department of Computer & Information Security, Daejeon University, Daejeon 34520, Korea

Open Access http://doi.org/10.6109/jkiice.2018.22.6.835 pISSN:2234-4772

ⓒThis is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(http://creativecommons.org/li-censes/by-nc/3.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Copyright ⓒ The Korea Institute of Information and Communication Engineering.

Ⅰ. 서 론

빅 데이터란 기존 데이터베이스의 데이터 저장·관리·분석 능력을 초과하는 다양한 형식을 가진 대량의 데이터를 의미한다. 빅 데이터는 여러 다양한 분야에서 생성 및 활용되고 있으며 특히 주목받고 있는 분야는 보건의료 분야로 빅 데이터가 활용되고 있다. 특히 보건의료 분야에서의 빅 데이터 분석은 사회경제적으로 큰 주목을 받고 있다[1,2]. 본 연구에서는 보건의료분야에서 생성되는 지역사회 건강조사 자료를 클라우드 환경에서 하둡과 DB Machine을 구축하고 분석한다. 현재는 과거에 비하여 경제성장, 의료기술의 발달, 평균수명의 연장 등으로 인해 국민의 생활수준은 향상되었음에도 불구하고 삶의 질은 낮아지고 있고 의료 초점은 질병 치료에서 질병 예방 및 관리로 패러다임의 변화로 국민 스스로 자신의 건강에 대해서 느끼고 평가하는 것이 중요하기 때문에 건강관련 삶의 질이 부각 되고 있다[3].

건강관련 삶의 질이란 "질병정도나 건강 수준에 영향을 받는 삶의 질"을 일컫는 말로 일상생활이나 주관적 경험에 대한 신체적, 정신적, 감정적 행복을 모두 포괄한 다면적인 개념을 포함하고 있다[2,4]. 또한 WHO는 삶의 질을 가치 체계, 목표, 기대, 관심과 관련하여 개인의 인식으로 정의하였다. 이는 사람의 신체 건강, 심리적 상태, 사회적 관계, 개인적 신념 및 환경의 특징에 대해 복합적으로 영향을 받는 광범위한 개념이다[5,6,7]. 그 중 건강경제 실현을 위한 해법으로 우리나라가 보유한 우수한 데이터와 IT/BT/의료 분야의 융합을 통해 창출되는 새로운 보건의료(IT-Health) 기술이 필요하다. 이러한 미래형 보건의료 기술의 중심에는 보건의료의 패러다임 변화와 더불어 요구되는 개인 맞춤형(personalized) 건강관리 서비스가 있다. 일상에서 생성되는 건강행태 데이터나 유전적으로 가지고 있는 유전정보나 진단 및 처방정보, 일생생활에서 생성되는 건강습관이나 행태 데이터 등 모든 정보를 종합적으로 분석해 개개인에 맞춤형 건강관리서비스를 제공할 수 있다면 질병예방, 약물 부작용 및 치료효과의 증대 등을 통해 의료비 감소는 물론 건강증진을 모두 만족할 수 있을 것으로 기대된다. [6,7].

이와 같이 변화하는 빅데이터 분석기술과 IT-Health 기기와 연동된 웨어러블 제품을 통한 개인 맞춤형 건강관리 서비스를 위해서는 필수가 되고 있다.

건강정보 빅데이터를 위해 현재 가장 많이 사용되고 있는 핏빗(Fitbit)이나 구글 글래스(Google glass), 헬스킷(Health kit) 그리고 구글핏(Google fit)등이 다양한 헬스케어 앱(application)과 연동되어 일상 생활 속에 활용되고 있다. 특히나 스마트 폰과 함께 일상생활 속에서 널리 사용되면서 개인별 운동량 추적, 식단 추천, 체중감량 조언, 신체상태 별 운동지도, 수면패턴 분석/관리, 혈압/스트레스 조절 등 맞춤형 건강관리 서비스를 제공하고 이를 통해 질병 예방 뿐 아니라 만성질환 관리도 가능하게 되고 있다. [8,9].

이와 같은 개인별 맞춤형 IT-Health 서비스 제공을 위한 데이터 소스, 즉, 다양한 디바이스로부터 생성된 정형/비정형 데이터들은 개별적으로도 그 가치가 매우 높지만, 다른 데이터와 상호 연계될 수 있다면, 그 잠재력은 기하급수적으로 커질 것이다. 따라서 본 논문에서는 공개 건강 빅데이터를 활용해 개인이 살아가는데 느끼는 건강관련 삶의 질을 분석한다. 삶의 질을 분석하기 위해 빅데이터 분석 플랫폼인 아파치 스파크 플랫폼을 이용하고 삶의 질 측정도구로는 이미 널리 사용되고 있는 Standard Gamble(SG), Time Trad-Off(TTO), visual Analogue Scale(VAS), SF-12, European EuroQoL questionnaire(EQ-5D)을 사용한다. 그 중 EQ-VAS는 눈금자 위에 오늘 자신의 건강상태가 얼마나 좋고 나쁜지를 표시하도록 하여 건강관련 삶의 질을 평가한 것이다.

건강관련 삶의 질은 객관적, 주관적 영역의 상호작용에 의해서도 영향을 받는다[9,10,11]. 그러므로 건강관련 삶의 질을 측정하기 위해서는 객관적, 주관적 지표가 모두 활용하는 것 정확하다. 대부분의 건강관련 삶의 질 연구는 개인수준에서 건강관련 삶의 질을 평가하고, 자료를 EQ-VAS 측정도구로 다수준 분석 방법을 사용한 분석은 거의 없는 실정이다. 따라서 본 연구는 공개의료 빅 데이터를 이용한 개인적 특성과 지역사회 특성에 따른 건강관련 삶의 질의 관련요인을 파악하여 지역사회 정책 개발의 기초자료로 제공하고자 한다.

Ⅱ. 관련연구

2.1. 빅데이터 분석 환경

빅데이터의 분석 기법들은 통계학과, 기계학습 그리고 데이터 마이닝 분야에서 이미 사용되고 있던 기법으

로 알고리즘을 대규모 데이터 처리에 맞도록 개선하여 빅데이터 처리에 적용시키고 있다. 클라우드는 다양한 클라이언트 장치에서 동작하는 기술로 네트워크 상에서 거대한 데이터 스토리지를 두고 필요할 때마다 별도의 디바이스를 이용해 사용 가능한 웹 기반 소프트웨어 서비스와 같아 보이지만, 실시간으로 클라우드 상에서 데이터 처리가 가능하고, 기업에서는 주로 가상화 기반으로 필요한 클라우드 서비스를 제공받고 있다. 빅데이터 분석 환경으로 오픈소스 환경인 클라우드 플랫폼을 많이 활용한다. 아파치 스파크는 빠른 범용 클러스터 컴퓨팅을 위한 플랫폼이다. 클라우드의 장점은 실시간으로 처리가 가능하며 동시에 대용량을 운영할 수 있어 이전에 버려졌던 데이터까지도 분석이 가능하도록 지원하고 있다.

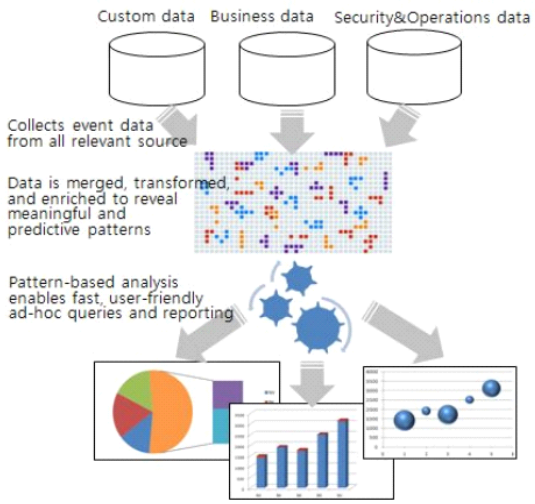

Fig. 1 Big-data analytics methodology

그림 1은 빅데이터 분석을 위한 클라우드 서비스를 도식화 한 것이다. 빅데이터 분석을 위해 클라우드 서비스 중 특히 고속의 처리를 위해 스파크는 맵리듀스보다 효율적으로 동작한다. 스파크는 기존 맵리듀스의 반복적인 알고리즘 수행 문제점과 비효율적이던 프레임워크 문제를 해결해 대화형 쿼리 형태의 작업에서의 매우 뛰어난 성능을 보여주고 있다. 특히, 하둡에 비해 평균 5배가량 고속의 결과를 보여 주고 있으며, 스파크는 분산처리를 요구하는 반복알고리즘이나 스트리밍 서비스 그리고 상호작용(interaction)등에서 파이프라인을 통해 범용성을 제공하는 특징을 갖는다. 즉 파이프라인을 통해 제공되는 분산처리는 다양한 작업을 편리하게 조합할 수 있어 주변 기기들과 유지보수 측면에서 부담을 줄인다는 특징을 갖는다.

또한 스파크는 함수형 프로그래밍 언어인 스칼라(Scala)를 자바가상머신(Java Virtual Machine; JVM) 상에서 사용하여 구현되었다. 그러나 스칼라 이외에도 파이썬(Python), 자바(JAVA), SQL, R 등 다양한 언어로 구성된 API와 풍부한 내부 라이브러리를 통해 쉽게 접근할 수 있다. 또한, 다른 빅 데이터 도구들과 유기적으로 연결된다. 즉 스파크는 HBase등 하둡클러스터에서 실행되고 다양한 하둡데이터 소스를 사용 할 수 있다. 스파크의 데이터 엔진은 메모리내 특정 상황 즉 스테이지와 스테이지 사이에는 디스트에 상태를 다시 써야하는 다중 스테이지 작업과 상호 비교하는 경우 하둡의 맵리듀스보다 최대 100배 정도 더 빠르다는 특징을 갖으며, 또한 데이터가 커서 메모리 안에 완전히 들어갈수 없는 아파치 스파크 작업의 경우라 할지라도 맵리듀스보다 최대 10배 이상 빠르다[11,12].

또한 스파크는 함수형 언어로 스칼라에 기반하기 때문에 스파크 API는 스칼라 컬랙션 라이브러리 API와 거의 동일하다. 즉 스파크 API는 상당히 간단하기 때문에 누구나 쉽게 사용할 수 있다는 장점을 갖는다.

2.2. 삶의 질 측정 도구 환경

수집된 자료는 빅 데이터 플랫폼을 통해 분류하고 통계적 처리는 SAS 9.4 프로그램을 이용하였다. 공개의료 빅데이터(지역사회건강조사) 자료를 기반으로 전체 인구를 대표할 수 있는 표본의 대표성을 위해 가중치를 부여하고 분석한다. 제안 모델에서 분석타입은 첫 번째 지역 간 변이를 확인하기 위한 독립변수를 모두 추가하지 않고 절편만 포함하는 기초모형을 분석한다. 두 번째는 지역 수준 고정 효과를 제외한 개인 수준의 변수만을 추가하여 개인요인이 건강관련 삶의 질에 미치는 영향을 분석하였다. 세 번째 모형에서는 두 번째 모형에 지역사회요인 변수를 추가하여 지역사회요인이 건강관련 삶의 질에 미치는 영향을 분석하였다.

제안 논문에서는 연구모델을 개인수준 변수와 지역수준 변수를 각각 투입하는 타입과 모두 투입하는 타입이다. 연구모델에서의 설명력은 기초타입의 각 수준 분산에서 연구모델의 변수 투입으로 설명되는 각 수준 분산의 비율로 구할 수 있다. 분석에는 SAS의 Glimmix

procedure를 이용하였다. 본 연구에서는 다수준 분석을 위해 3가지 타입을 구축하였다. 기초분석타입은 지역 간의 변이를 추정하기 위해 개인수준과 지역수준 변수 모두를 포함하지 않은 타입을 유의수준 0.05를 기준으로 구축하였다.

타입 1 : 개인수준

성별, 연령, 교육수준, 월 가구 소득, 아침결식, 중등도 이상 신체활동, 질환개수, 필요의료서비스 미 치료, 사회적 연결망 접촉, 사회활동 참여로 타입 구축.

타입 2 : 지역수준

읍면동별 박탈지수로 타입 구축.

타입 3 : 개인수준

지역수준 & 지역 수준 포함 타입 구축.

2.3. 연구 도구와 환경

본 논문에서 제안하는 시스템은 하둡기반 빅 데이터 질의처리 엔진이 탑재된 클라우드 환경에서 분석 서비스를 사용하였다. 연구 도구로 건강관련 삶의 질에 대해 살펴보기 위해 삶의 질은 건강 관련 삶의 질과 비 건강 관련 삶의 질로 구분할 수 있는데, 보건의료분야에서는 “건강 관련 삶의 질”이라는 표현을 주로 사용한다. 삶의 질에 대해서는 많은 정의가 있겠으나 개인의 경험, 신념, 기대나 인지 수준에 따른 신체적, 정신적, 그리고 사회적인 측면의 건강 수준이라고 정의한다[10,11,13]. 질보정수명(Quality Adjusted Life Years, QALYs)을 구하는데 필요한 효용 가중치(utility weight)를 구할 수 있는 EQ5D는 EuroQol 그룹에서 개발한 HRQoL 측정도구로 SF-36 등과 더불어 가장 많이 사용되고 있는 일반적 측정도구이다. 본 연구에서는 2012년부터 2014년 EQ-VAS에 응답한 19세 이상의 성인을 대상으로 분석타입을 정의하고 EQ-5D는 위의 문항을 초로 문항마다 상이한 가치점수를 부여하여 점수를 산출할 있다. 본 연구는 일반적으로 사용되는 시간 교환법(Time Trade-Off; TTO) 방식을 이용해 가치점수를 추정한다. 식 1은 가치 추정점수 TTO를 나타낸 것이다[12,13].

$$
\begin{aligned}
Y = 1 - (&0.050 + 0.096 \times M2 + 0.418 \times M3 \\
&+ 0.046 \times SC2 + 0.136 \times SC3 \\
&+ 0.051 \times UA2 + 0.208 \times UA3 \\
&+ 0.037 \times PD2 + 0.151 \times PD3 \\
&+ 0.043 \times AD2 + 0.158 \times AD3 \\
&+ 0.050 \times N3)
\end{aligned} \qquad (1)
$$

Ⅲ. 클라우드 환경에서 빅데이터 분석을 통한 삶의 질 측정

3.1. 클라우드 빅 데이터 분석 환경

의료정보 빅 데이터 분석을 위해 제안 논문에서는 오픈 소스 플랫폼인 클라우드 플랫폼을 활용한다. 클라우드 플랫폼에서 빅 데이터 분석을 위한 아파치 스파크 클러스터 구조를 그림 2와 같이 구성한다.

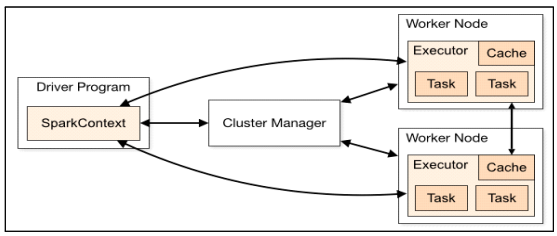

Fig. 2 Apache Spark cluster structure

그림 2의 Driver Program은 여러 개의 병렬적 작업으로 Worker Node에 있는 Executor에서 실행을 담당하고 SparkCotext는 메인 시작 지점으로 스파크 API를 활용하기 위해 필요하고 클러스터의 연결을 보여주고 RDD를 만드는데 사용된다. Cluster Manager는 Standalone, YARN, Mesos 등 클러스터의 자원을 관리하는 관리자 역할을 수행하고, Worker Node는 하드웨어 즉 서버로 하나의 물리적 장치에 여러 개 사용이 가능하다. 마지막으로 Executor는 프로세스 즉 하나의 워커 노드로 여러 개 사용이 가능하다.

제안 논문의 스파크를 활용한 데이터의 분산 처리는 클라우드 그리드 형태의 분산 인덱스 정보를 생성하고 RDD상에 저장한 다음 해당 인덱스 정보를 재사용해 삶의 질에 영향을 미치는 요인 분석을 수행한다.

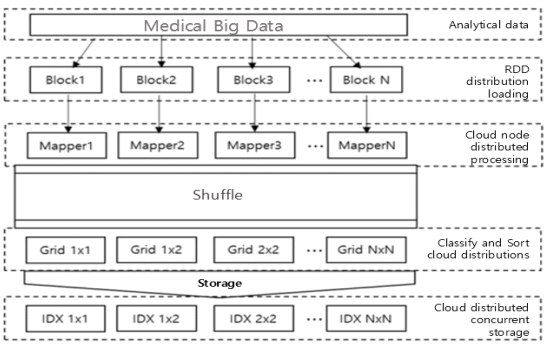

Fig. 3 RDD distribution structure in spark

그림 3은 제안 논문에서 빅 데이터 분석의 효율을 위해 탄력적 분산 데이터 처리를 스파크의 RDD(Resilient Distributed Data Set)을 활용한 구조를 나타낸 것이다. 또한 제안 논문에서는 몽고 DB를 구축하고 스파크 코어로 구성한 후 그림 3의 분산처리 결과를 저장한다. 또한 Spark은 인-메모리에서 처리하는 데이터의 손실을 방지하기 위해서 RDD를 활용하고 모든 데이터는 RDD로 만들어져서 처리에 사용된다. 다음은 RDD에서 데이터 셋을 2012년~2014년 지역사회건강조사 원시 자료를 대상으로 EQ-VAS(visual analogue scale, 0~100 scores)에 응답한 만19세 이상 성인의 자료와 통계청 시군구별 2차 자료를 추출해 분석데이터 셋을 구축한 후 daily를 이용해 키를 줄여 156 RDD에서 43 RDD로 축소한 것을 보여준 코드이다.

```
tally = dict()
for line in daily_show:
  year = line[0]
  if (year-1) in tally.keys():
    tally[year-1] = tally[year-1] + 1
  else:
    tally[year-1] = 1
tally = daily_show.map(lambda x: (x[0], 1))
  .reduceByKey (lambda x,y: x+y)
print(tally)
```

이와 같이 제안 논문은 분산실행을 위해 다수의 이그제큐터 프로세스로 드라이버 코어 프로세스를 결합하고 필요에 따라 확장과 축소를 탄력적 사용한다.

Ⅳ. 실험 및 결과

4.1. 성능평가

제안 논문에서는 하드웨어적 자원을 통한 분석성능평가를 실행하기 위해 각 노드에서 CPU자원의 사용량은 1Core로 한정하고 HDFS의 Replication 수는 2로 조정하였다. 클러스터 구성은 Master Node 1대와 Slave Node의 수를 각각 2, 4, 6, 8대 구성하고 메모리를 2GB부터 총 8GB까지 늘려가며 메모리 사용량을 평가하였다.

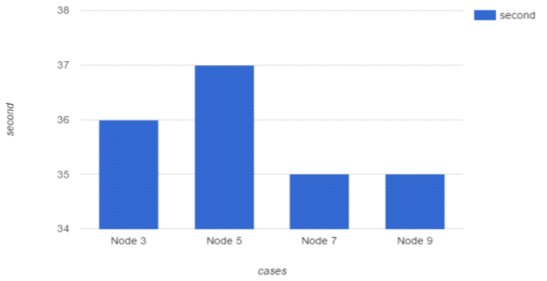

Fig. 4 Performance evaluation based on memory usage

그림 4는 Slave Node의 수가 6대, 8대인 경우에도 위와 같은 결과가 나타나는 것으로 보아 Hadoop Processing 수행시간에 RAM 자원의 영향은 적게 받는 것을 알 수 있다. 그림 5는 우리는 분석데이터의 확장성을 확인하기 위해 분석시 평균 Processing delay time을 측정한 결과 40개의 노드 확장까지는 큰 차이가 없음을 확인하였다.

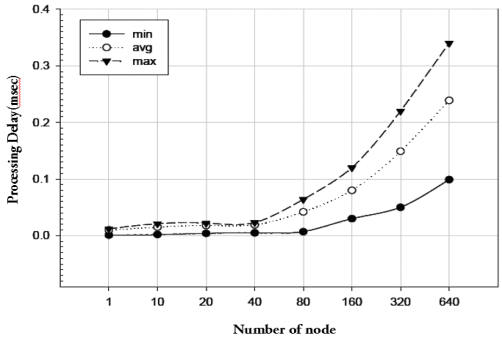

Fig. 5 Big data analysis delay in spark

4.2. 개인요인에 따른 건강관련 삶의 질

개인적 특성과 건강관련 삶의 질에 따른 결과는 APPENDEX Table 1과 같다. 성별에 따른 건강관련 삶의 질을 100점 기준 남자 평균 73.8, 여자 평균 70.0으로 남자가 여자보다 높았다. 개인적 특성에 따른 건강관련 삶의 질 수준은 남자와 여자 모두 연령이 낮을수록, 도시에 사는 경우, 교육수준이 높을수록, 월 가구소득이 높을수록, 세대유형은 2세대, 3세대, 1세대 순으로, 배우자가 있는 경우, 만성질환이 없는 경우, 스트레스가 적을수록 건강관련 삶의 질이 높았다. 또한 각 개인의 건강관련 삶의 질에 영향을 미치는 요인을 알아보기 위한 다수준 분석 결과로 각 타입에서 성별에 따른 차이를 보였으며 교육수준 및 다양한 항목에서 삶의 질과 유의성

을 보였다. 각 개인의 삶의 질에 영향을 미치는 건강관련 요인을 성별로 구분하고 단계별 다수준 회귀분석의 결과를 나타낸 것이다.

Ⅴ. 결론

본 논문에서는 클라우드 환경에서 공개의료 빅데이터 분석을 위해 하둡기반의 스파크를 이용해 개인의 삶의 질에 영향을 미치는 요인을 하드웨어의 제약 없이 빠르게 분석하였다. 분석데이터는 2012년부터 2014년 지역사회건강조사 자료 중 EQ－VAS에 응답한 19세 이상의 성인을 대상으로 일반적 특성, 건강행태, 사회 자본을 빅 데이터 분석을 통해 분석하고 분석결과를 삶의 질에 영향을 분석하였다. 분석결과 성별, 나이, 도시농촌 구분, 교육수준, 월 가구소득, 세대유형, 배우자 유무, 만성질환 유무, 스트레스수준이 건강관련 삶의 질에 영향을 미치는 것으로 나타났다. 다수준 분석에서 이를 보정한 상태에서는 고령인구비율, 복지예산비중, 제정자립도 등 지역사회특징이 건강관련 삶의 유의한 영향을 주어 건강관련 삶의 질은 개인적 특성뿐만 아니라 지역사회 특성에도 영향을 미치는 것으로 나타났다. 이와 같이 많은 변수적 속성을 지닌 의료 건강데이터 분석을 클라우드 환경에서 메모리 등 하드웨어의 종속 없이 활용할 수 있음을 증명하였다. 향후 의료 빅데이터 분석을 위해 클라우드와 오픈소스를 활용할 수 있는 하드웨어에 독립적인 플랫폼 제공을 위한 연구가 지속되어야 할 것이다.

REFERENCES

[1] K. Y. Kim, B. Y. Chun, S. Kam, S. W. Lee, K. S. Park, and S. C. Chae, “Development of Measurement Scale for the Quality of Life in Hypertensive Patients,” *International Journal of preventive medicine and public health*, vol.38, no.1, pp.61-70, February 2005.

[2] M. J. Calvert, and N. Freemantle, “Use of Health-related Quality of life in Prescrib in Research. Part1:Why Evaluated Health-related Quality of Life ?”, *Journal of clinical pharmacy and therapeutics*, vol. 28, no. 6, pp. 513-521. December 2003.

[4] A. Bowing, Z. Gabriel, J. Dykes, O. Evans, A. Fleissing, D. Banister, and S. Sutton, “ Let’s ask them: A national survey of definition of quality of life and its enhancement among people aged 65 and over,” *International Journal of Aging and Human Developent*, vol.56, no.4, pp.269-306, June 2003.

[5] Y. B. Cho S. H. Woo and S. H. Lee, “The Big Data Analysis and Medical Quality Management for Wellness”, *Journal of the Korea Society of Computer & Information*, vol.19, no. 12, pp. 101-109, December 2014.

[6] K. H. Ko, Y. H. An and K. W. Kim, “Measuring and enhancing ways of social capital in local government,” Research Report Seoul Korea Research Institute for Local Administration. 2012

[7] S,Y Kang, S.J Youn and Y.S. Shin, “A Basic Study on Environmental Health Service Utilizing Big Data,” Korea Environment Institute :KEI Working Paper, vol 7, 2015

[8] T.M Song, “Tend Analysis of Health and Welfare Policy Based on Social Big Data,” *Journal of the Health-welfare Policy Forum*, vol. 7, pp. 101-113, July 2017.

[9] C.M. An, “Big Data Platform Status and Issue Analysis,” ETRI Insight Report 2017-33, 2017

[10] D.M Seo, S.K Song, S.W. Lee and I. C. Yun,“Government Big Data Analysis Technology Application Trend.” Weekly Technology Trend Data of KOSEN Trend Report 1556-01, 2014

[11] S. S. Kim, M. Y. Chung, T. W. Lee and J. H. Won, “Implementation of Medical Data-Based Big Data Analytics Service,” *Journal of the Korea Information Science Society*, vol. no. pp.157-159, December 2015.

[12] A. Floratou, U. F. Minhas, and F.Ozcan, “SQL-on-Hadoop: Full Circle Back to Shared-nothing Database Architectures,” *In Proceedings of the VLDB Endowment*, vol. 7, no. 12, pp. 1295 – 1306, Hangzhou, China, 2014.

[13] J. H. Lee, M. K. Jea, M. J. Jo and H. S. Son, “ Big Data Utilization Trend in Healthcare,” *Journal of the Korea Institute of Information and Communication Engineering*, vol. 32, no.1, pp.63-75, January 2015.

APPENDIX.

Table. 1 Socio-demographic characteristics of the subjects

Factor		Men		Women	
		N(%)	M±SD	N(%)	M±SD
Age	19-29	22484(11.2)	77.9±15.6	26343(10.7)	75.4±16.0
	30-39	31855(15.9)	75.4±15.0	37228(15.2)	74.5±15.5
	40-49	40631(20.3)	75.9±15.3	45460(18.5)	75.2±15.7
	50-59	40917(20.4)	75.3±16.5	48102(19.6)	72.6±17.4
	60-69	31852(15.9)	72.8±17.7	38664(15.8)	67.0±19.2
	70+	32595(16.3)	66.1±20.2	49553(20.2)	58.8±21.4
Region	urban	113795(56.8)	75.0±16.1	138758(56.6)	72.2±17.6
	rural	86539(43.2)	72.3±18.5	106592(43.4)	67.2±20.2
Education	Ineducation	9754(4.9)	62.9±21.6	43967(17.9)	58.4±21.7
	Elementary school	27710(13.9)	68.1±19.8	42575(17.4)	65.8±19.5
	Middle school	23886(11.9)	71.9±18.0	25756(10.5)	71.3±17.8
	High school	75335(37.7)	75.2±16.2	74679(30.5)	74.3±16.3
	Over college	63374(31.7)	77.1±14.5	57977(23.7)	75.8±15.0
Monthly Family Income (unit 1, 000 won)	Under 1,000won	41201(21.5)	66.7±20.5	65067(27.8)	61.9±21.2
	1,010-2,000won	37877(19.8)	73.4±17.1	43767(18.7)	70.2±18.3
	2,010-3,000won	39544(20.6)	75.7±15.5	43098(18.4)	73.0±16.9
	3,010-4,000won	23961(12.5)	76.4±15.0	26684(11.4)	74.2±16.1
	Over 4,010won	48996(25.6)	77.5±14.6	55368(23.7)	75.4±15.9
Type of Household	One generation	75552(37.7)	71.3±18.7	95883(39.1)	66.1±20.3
	Two generation	106069(52.9)	75.5±15.9	123247(50.2)	73.0±17.3
	Three generation	18705(9.3)	74.4±16.9	26201(10.7)	70.2±18.8
Spouse	Yes	162853(81.3)	74.3±16.9	178354(72.7)	71.7±17.9
	No	37473(18.7)	72.0±18.4	66977(27.3)	65.4±20.7
Chronic Disease	Yes	68794(34.3)	70.2±18.7	98678(40.2)	64.0±20.5
	No	131540(65.7)	75.7±16.0	146672(59.8)	74.1±16.6
Perceived stress level	Not at all	43484(21.7)	77.2±17.0	48282(19.7)	73.6±19.0
	Not very	107073(53.5)	75.0±15.7	131064(53.5)	72.2±17.1
	Very	43185(21.6)	69.3±18.3	57294(23.4)	64.2±19.9
	Extremely	6495(3.2)	62.7±22.8	8485(3.5)	56.1±23.2

김민경(Min-Kyoung Kim)

2006: 연세대학교 보건학 석사
2016: 충북대학교 의학과 박사과정 수료
현 재: ㈜소노엠 기업부설연구소 선임연구원 및 협성대학교 보건관리학과 겸임교수

※ 관심분야: 빅데이터, 건강증진, 원격의료

조영복(Young-Bok Cho)

2005: 충북대학교 전자계산학과 공학석사
2012: 충북대학교 전자계산학과 공학박사
2016: 충북대학교 의학과 박사과정수료
2012-2018: 충북대학교 소프트웨어학과 초빙교수
현 재 : 대전대학교 정보보안학과 조교수

※관심분야: 의료영상처리, 정보보안, 의료정보보호, 모바일보안