

<https://doi.org/10.7236/JIIBC.2020.20.3.163>

JIIBC 2020-3-23

빅데이터 기반 환자 간병 방법 분석 연구

A Study on Big Data Based Method of Patient Care Analysis

박지훈*, 황승연**, 윤범식***, 최수길****, 이돈희*****, 김정준*****, 문진용*****, 박경원*****

Ji-Hun Park*, Seung-Yeon Hwang**, Bum-Sik Yun***, Su-Gil Choe****,
Don-Hee Lee*****, Jeong-Joon Kim*****, Jin-Yong Moon*****, Kyung-won Park*****

요약 정보통신기술의 발전과 함께 데이터의 생산량이 기하급수적으로 증가하면서 빅데이터에 대한 관심이 높아지고 있다. 빅데이터 관련 기술들도 발전함에 따라 여러 분야에서 빅데이터가 수집, 저장, 처리, 분석, 활용되고 있다. 특히 보건 의료 분야에서의 빅데이터 분석은 사회경제적으로도 큰 영향력을 발휘할 수 있기 때문에 큰 주목을 받고 있다. 빅데이터 기술을 환자 진단 데이터 분석에 활용하여 간단한 병원 진료에 투여되는 막대한 비용을 절감할 수 있을 것으로 전망된다. 따라서 본 논문에서는 환자 데이터를 분석하여 병원에 가기 어려운 환자나 의학적인 전문 지식이 없는 간병인들에게 의사의 진단과 가까운 간병 가이드 정보를 제시하고자 한다. 먼저 수집된 환자 데이터를 HDFS에 저장하고, 하둡 환경에서 빅데이터 처리 및 분석 도구인 R을 이용하여 데이터를 처리한 후 분류분석을 한다. R의 다양한 기능들을 웹에 구현하기 위해 활용되는 R Shiny를 이용하여 웹 서버에 시각화를 한다.

Abstract With the development of information and communication technologies, the growing volume of data is increasing exponentially, raising interest in big data. As technologies related to big data have developed, big data is being collected, stored, processed, analyzed, and utilized in many fields. Big data analytics in the health care sector, in particular, is receiving much attention because they can also have a huge social and economic impact. It is predicted that it will be able to use Big Data technology to analyze patients' diagnostic data and reduce the amount of money that is spent on simple hospital care. Therefore, in this thesis, patient data is analyzed to present to patients who are unable to go to the hospital or caregivers who do not have medical expertise with close care guidelines. First, the collected patient data is stored in HDFS and the data is processed and classified using R, a big data processing and analysis tool, in the Hadoop environment. Visualize to a web server using R Shiny, which is used to implement various functions of R on the web.

Key Words : Big Data, Hadoop, R, Classification analysis, R Shiny

*준회원, 한국산업기술대학교 컴퓨터공학과 학부생

**준회원, 안양대학교 ICT 융합공학부 대학원생

***준회원, 한국산업기술대학교 컴퓨터공학과 학부생

****준회원, 한국산업기술대학교 컴퓨터공학과 학부생

*****정회원, SK 주식회사 수석

*****정회원, 안양대학교 ICT 융합공학부 조교수

*****정회원, 강동대학교 방송영상미디어과 교수

*****정회원, 한국산업기술대학교 컴퓨터공학부 조교수

접수일자 2020년 3월 12일, 수정완료 2020년 5월 2일

게재확정일자 2020년 6월 5일

Received: 12 March, 2020 / Revised: 2 May, 2020 /

Accepted: 5 June, 2020

*****Corresponding Author: chrisndanny@kpu.ac.kr

Dept. of Computer Engineering, Korea Polytechnic University, Korea.

I. 서 론

최근 정보통신기술의 발전과 함께 디지털 세상이 열리면서 빅데이터와 관련된 기술들이 발전하게 되었고, 제조, 유통, 통신 등 다양한 분야에서 생산되는 데이터들을 수집한 후 저장하고 처리 및 분석을 할 수 있게 되었다. 보건의료 분야에도 데이터가 폭발적으로 증가하고 있다^[1]. 국내의 경우, 건강보험심사평가원은 연간 4.600만 명, 200억 건의 의료 빅데이터를 보유하고 있다. 국가적으로 공공의료 서비스의 개선을 위해 빅데이터를 활용하여 분석하는 것은 중요하다. 이와 같은 방대한 양의 보건의료 데이터를 잘 활용하면 헬스 케어 분야에 투여되는 막대한 비용을 절감할 수 있을 것으로 전망된다^[2].

또한, 정보통신기술과 의료정보기술을 결합한 질병예방, 진단, 치료와 사후관리 서비스를 연계한 디지털 헬스케어 산업과 다양한 스마트인공지능이 전문의보다 더 높은 정확도로 질병을 진단하는 시대가 다가오는 만큼 헬스케어 분야에서 빅데이터와 딥 러닝의 활용은 이제 선택이 아닌 필수로 자리 잡고 있다^[3]. 우리나라의 의료복지 수준을 개선하고 의료 서비스의 질을 향상시키기 위해 헬스케어 빅데이터와 딥 러닝에 많은 관심과 투자 및 연구가 이뤄져야 한다^[4]. 따라서, 본 논문에서는 환자 진단 데이터를 분석하고, 간병인에게 간병 방법을 제공하여 병원 진료비용과 시간 등을 줄일 수 있는 빅데이터 기반 분석 시스템을 개발한다.

본 논문의 구성은 다음과 같다. 2장에서는 본 연구에서 사용되는 기술들에 대하여 설명한다. 3장에서는 빅데이터 환자 진단 데이터 분석 시스템의 설계와 구현에 대하여 설명한다. 4장에서는 본 논문의 내용을 정리하며, 기대 효과에 대하여 설명하고 마친다.

II. 관련 기술

1. 빅데이터(Bigdata)

빅데이터는 기존 데이터베이스 관리 도구의 능력을 넘어서는 대량의 정형 데이터뿐만 아니라 비정형 데이터까지 포함한 데이터를 의미하고, 이러한 데이터로부터 가치를 추출하고 결과를 분석하는 기술이다. 빅데이터는 정치, 사회, 경제, 문화 등 다양한 분야에서 발생하고 있다. 그리고 인류에게 가치 있는 정보를 제공할 가능성을 제시하며 그 중요성이 강조되고 있다^[5].

2. Apache Tomcat

아파치 소프트웨어 재단에서 개발한 서블릿 컨테이너만 있는 웹 애플리케이션 서버로써, 웹 서버와 웹 컨테이너의 결합으로 다양한 기능을 컨테이너에 구현하여 다양한 역할을 수행할 수 있는 서버를 의미한다. 웹 서버가 정적인 데이터를 처리하는 서버라면, 톰캣은 데이터 조작이 필요하거나 동적인 데이터 처리가 필요한 경우 사용한다^[6].

3. JSP(Java Server pages)

JSP는 자바 언어를 기반으로 웹사이트를 쉽게 구축할 수 있도록 만들어주는 서버사이드 스크립트 언어로써, 자바에서 제공하는 API나 데이터베이스와의 연동이 가능하도록 만들어주는 JDBC 등과 같은 자바 기능을 그대로 사용할 수 있으며, 운영체제에 상관없이 플랫폼에 독립적인 장점이 있다^[7].

4. 분류분석

분류분석은 데이터의 실체가 어떤 집단에 속하는지 예측하는데 사용하는 데이터 마이닝 기법이다. 클래스로 나눈다는 점에서 클러스터링과 유사하지만, 각 클래스가 어떻게 정의되는지 미리 알아야 한다는 차이가 있다. 따라서 분류분석의 데이터는 훈련 데이터와 시험 데이터로 나눌 수 있으며, 훈련 데이터를 통해 모형을 만들고 시험 데이터를 통해 모형의 타당성을 확인할 수 있다. 대표적인 알고리즘으로 K-NN(K-Nearest Neighborhood), Decision Tree, Bayesian Theorem 등이 있다^[8].

5. R Shiny

R Shiny는 데이터 분석 시각화 도구인 R을 사용해서 웹 애플리케이션을 제작할 수 있게 해주는 패키지이다. R Shiny는 UI, Server, 입출력, render와 배포로 구성되어 있다. UI는 실제로 사용자가 보는 화면을 의미하고, Server는 실제 서버가 브라우저와 통신하는 과정 전체를 간단하게 만들어주는 역할을 한다. 입출력은 input과 output 함수를 통해 데이터를 UI와 Server가 교환하며 output 함수는 render 함수로 선언된다. Shiny는 Shiny-server를 통해서 동작하며 shiny app이 동작할 수 있는 서버를 의미한다^[9].

III. 설계 및 구현

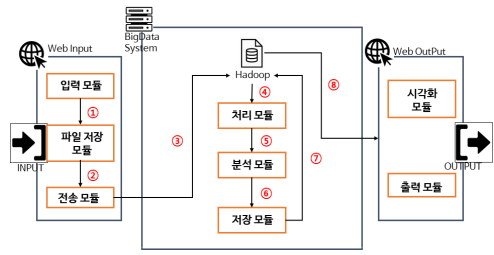


그림 1. 빅데이터 기반 환자 간병 방법 분석 시스템 구성도
Fig. 1. Big Data Based Method of Patient Care Analysis System Composition diagram

빅데이터 기반 환자 간병 방법 분석 시스템은 그림 1과 같이 Web Input Process, Big Data System Process, Web output Process 3가지로 구성하였다.

Web Input Process는 간병 방법이 필요한 환자의 특성을 웹 서버에 입력하는 모듈, 입력한 환자 데이터를 CSV 형식의 파일로 저장하는 모듈, 환자 데이터가 저장된 CSV 파일 빅데이터 분산 저장 시스템에 전송하는 모듈로 구성된다.

Big Data System Process는 전송 모듈로부터 전송되는 데이터를 가공하는 처리하는 모듈, 기존 데이터를 학습하여 정제된 데이터를 분석하는 모듈, 분석된 데이터를 기존 데이터 셋에 추가하여 하나의 파일로 구성한 후 빅데이터 저장 솔루션을 이용하여 분산 저장하는 모듈로 구성된다.

Web output Process는 처리 및 분석이 완료된 결과물을 시각화하는 모듈, 시각화한 도표를 웹 서버로 출력해주는 모듈로 구성된다.

본 논문에서는 기존 환자 데이터를 학습시키고 대응 방법을 분석하여 의학적 전문 지식이 없는 간병인에게 환자 간병 방법을 제공하기 위한 시스템을 개발하고자 한다.

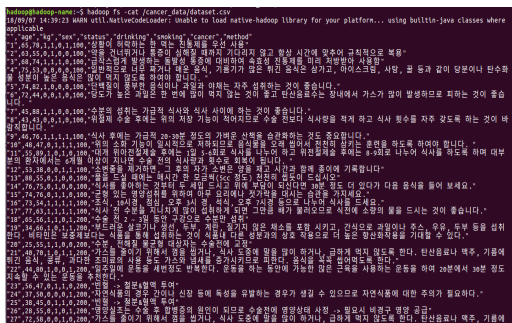


그림 2. HDFS에 저장된 기존 환자 데이터
Fig. 2. Existing Patient Data stored in HDFS

학습을 위한 기존 환자 데이터는 실제 의료 데이터 수집에 제약이 있어 실제 간병인을 대상으로 한 온라인 설문조사, 의료 홈페이지와 공공데이터를 활용해 데이터를 수집하였다. 수집한 데이터는 Hadoop 클러스터의 4개 노드에 분산 저장한다. 그림 2는 HDFS에 저장된 기존 환자 데이터이다.

간병 방법 예측을 위한 테스트 데이터는 입력 모듈인 웹 서버를 통해 수집하였다. 그림 3은 Tomcat과 JSP를 사용하여 구축한 웹 서버 화면이다.

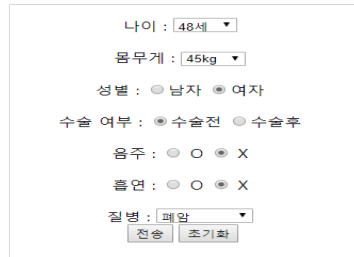


그림 3. 테스트 데이터 수집을 위한 웹 서버
Fig. 3. Web Servers for Collecting Test Data

간병인은 그림 3에서 간병 방법이 필요한 환자의 특성인 나이, 몸무게, 성별, 수술 여부, 흡연, 질병을 입력하여 테스트 데이터를 생성한다. 입력된 데이터는 파일 저장 모듈을 통해 CSV 파일로 저장된다. 저장되는 과정에서 나이, 몸무게는 입력 수치 값 그대로 저장되지만 성별, 수술 여부, 흡연, 질병은 지정된 수치 코드로 변환되어 저장된다. 그림 4를 통해 수치 코드로 변환되어 저장된 CSV 파일의 데이터를 확인할 수 있다.

	A	B	C	D	E	F	G	H
1	age	kg	sex	status	drinking	smoking	cancer	method
2		48	45	0	1	0	400	

그림 4. 입력된 데이터가 저장된 CSV 파일
Fig. 4. CSV File with entered stored Data

그림 4를 보면 간병 방법(method)를 제외한 컬럼들이 채워있는 것을 확인할 수 있다. 수치 코드로 변환된 컬럼을 보면 sex의 0은 여성, status의 1은 수술 전, drinking의 0은 X, smoking의 0은 X, cancer의 400은 폐암을 의미한다. 간병 방법은 분류 분석을 통해 예측할 것이다.

CSV 파일은 HDFS의 Name 노드와 공유된 폴더에 저장되어 전송 모듈을 통해 Hadoop으로 전송된다. 그림 5는 Local에 저장된 CSV 파일을 HDFS에 전송하는 명령어이다.

```
mv /mnt/hgfs/VMShare/input_data.csv /home/hadoop/R/Workspace/Shiny/
```

그림 5. 로컬에 저장된 데이터를 HDFS에 전송하는 명령어
Fig. 5. Commands for storing data stored in Local on HDFS

데이터 처리 모듈은 그림 2의 학습 데이터와 그림 4의 테스트 데이터를 바탕으로 간병 방법을 분류한다. 분류 분석을 위해 R 프로그래밍을 사용하여 K-NN 알고리즘을 적용하였다. R 프로그래밍은 hadoop 환경에서 구동하였다.

Step 1	data_set<-read.csv("dataset.csv", sep=";", header=T, stringsAsFactors = FALSE, fileEncoding = "UTF-8", encoding = "UTF-8")
Step 2	input_data<-read.csv("input_data.csv", sep=";", header=T, stringsAsFactors = FALSE, fileEncoding = "UTF-8", encoding = "UTF-8")
Step 3	i <- nrow(data_set)

그림 6. 데이터 처리를 위한 데이터 불러오기
Fig. 6. Load Data for Data Processing

그림 6의 Step 1은 분류 분석에 필요한 학습 데이터를 불러오기 위한 소스코드이다. 그림 6의 Step 2는 분류 분석에 적용할 테스트 데이터를 불러오기 위한 소스코드이다. 그림 6의 Step 3은 학습 데이터 행의 수를 저장하기 위한 소스코드이다.

Step 1	data_set <- data_set[-c(1)]
Step 2	rbind_data<-rbind(data_set, input_data)

그림 7. 데이터 가공
Fig. 7. Data Processing

그림 7의 Step 1은 불러온 학습 데이터 셋에서 불필요한 열인 index를 제거하기 위한 소스코드이다. 그림 7의 Step 2는 분석을 위해 학습 데이터와 테스트 데이터의 프레임에 결합하기 위한 소스코드이다.

Step 1	rbind_data\$age<-as.numeric(rbind_data\$age) rbind_data\$kg<-as.numeric(rbind_data\$kg) rbind_data\$sex<-as.numeric(rbind_data\$sex) rbind_data\$status<-as.numeric(rbind_data\$status) rbind_data\$drinking<-as.numeric(rbind_data\$drinking) rbind_data\$smoking<-as.numeric(rbind_data\$smoking) rbind_data\$cancer<-as.numeric(rbind_data\$cancer) rbind_data\$method<-as.factor(rbind_data\$method)
--------	---

그림 8. 데이터 정제
Fig. 8. Data Refinement

그림 8의 Step 1은 가공한 데이터를 분석을 위해 정제하는 소스코드이다. 분류 분석은 거리를 기반으로 데이터를 분류하기 때문에 데이터의 자료형은 숫자로 구성되어야 한다. as.numeric 함수를 사용하여 가공된 데이터의 각 컬럼을 숫자형으로 변환하였다.

Step 1	train_data<-rbind_data[1:i,1:7] test_data<-rbind_data[i+1,1:7]
Step 2	train_labels<-rbind_data[1:i,8] test_labels<-rbind_data[i+1,8]
Step 3	install.packages("class") library(class)
Step 4	test_pred <- knn(train = train_data, test = test_data, cl=train_labels, k=6)
Step 5	rbind_data[i+1,8]<-test_pred write.csv(rbind_data,"dataset.csv")

그림 9. K-NN 알고리즘을 활용한 분류 분석
Fig. 9. Classification Analysis Using the K-NN Algorithm

그림 9의 Step 1은 정제된 데이터의 분류 분석을 위해 학습 데이터와 테스트 데이터를 분류하기 위한 소스코드이다. 그림 9의 Step 2는 데이터를 학습시킨 후 예측된 테스트 데이터의 결과를 저장할 위치를 설정하는 소스코드이다. 그림 8의 Step 9은 K-NN에 필요한 관련 패키지를 설치 및 적용하는 소스코드이다. 그림 9의 Step 4는 K-NN 알고리즘을 활용하여 분류 분석을 실행하는 소스코드이다. knn 함수에 그림 9의 Step 1에서 분류한 학습 데이터와 테스트 데이터를 적용하고 k의 값은 knn 함수 수행에서 가장 좋은 평가척도 값을 나타낸 6으로 설정하였다. 그림 9의 Step 5는 학습된 데이터로 분류한 테스트 데이터의 결과를 그림 7의 Step 2에서 결합한 프레임에 저장한 후 데이터 셋을 갱신하는 소스코드이다. 그림 10을 보면 그림 4에서 비어있던 간병 방법의 값이 분류된 것을 확인할 수 있다.

405	48	45	0	1	0	0	400	체력이 많이 소진되는 질병이다. 질병은 단백질식품, 비타민 무..
-----	----	----	---	---	---	---	-----	--------------------------------------

그림 10. 분류된 테스트 데이터
Fig. 10. Categorized Test Data

그림 11은 학습 데이터의 일부를 보여준다.

36	45	80	1	1	1	1	400	기름기가 많은 음식이나 인공적인 조미료를 사용한 음식..
37	48	45	0	1	0	0	400	체력이 많이 소진되는 질병이다. 질병은 단백질식품, 비타민 무..
38	38	70	1	0	1	1	400	주치의를 항상 찾아가고 통증을 느낄 때 바로 의사에게 처방받..

그림 11. 학습 데이터
Fig. 11. Learning Data

분류된 테스트 데이터인 그림 10과 학습 데이터의 일부분인 그림 11을 비교해보면 간병 방법의 값이 정확하게 분류된 것을 확인할 수 있다.

분류 분석이 완료된 데이터를 웹 서버에 시각화하기 위해 R shiny를 사용하였다. 먼저 해당 질병에 해당하는 모든 간병 방법의 키워드를 WordCloud로 시각화하기 위해 갱신된 데이터 셋을 가공한다.

Step 1	<pre>ncancer <- rbind_data[i+1,] ncancer_data_cancer <- rbind_data [%]% filter(cancer == ncancer)</pre>
Step 2	<pre>ncancer_method <- ncancer_data_cancer\$method write.table(ncancer_method, "ncancer_method.txt", row.names = FALSE, quote = FALSE, append = FALSE)</pre>

그림 12. 데이터 셋 가공
 Fig. 12. Data Set Processing

그림 12의 Step 1은 데이터 셋에서 가장 최근 추가된 테스트 데이터의 질병 값에 해당하는 행을 필터링하기 위한 소스코드이다. 그림 12의 Step 2는 필터링 된 행에 해당하는 간병 방법을 추출한 후 WordCloud로 출력하기 위해 txt 파일로 생성한다.

Step 1	<pre>cancers <- list("Cancer" = "cancer") getTermMatrix <- memoise(function(cancer) { if (!(cancer %in% cancers)) stop("Unknown book") text <- readLines("ncancer_method.txt") word_data <- sapply(text, extractNoun, USE.NAMES = F) data_unlist <- unlist(word_data) data_unlist <- Filter(function(x) {nchar(x)}=2), data_unlist) wordcount <- table(data_unlist) wordcount_top <- head(sort(wordcount, decreasing = T), 100) m = as.matrix(wordcount_top) sort(rowSums(m), decreasing = TRUE) })</pre>
--------	---

그림 13. R shiny wordcloud를 위한 요소 설정
 Fig. 13. Element Settings for R shiny Wordcloud

그림 13의 Step 1은 R shiny에서 wordcloud를 사용하기 위해 요소를 설정하는 소스코드이다. List 설정부터 사용 데이터, 키워드 필터, 빈도 순 정렬 등을 설정할 수 있다.

Step 1	<pre>MAN <- nrow(ncancer_data[ncancer_data\$sex==1,]) WOMAN <-nrow(ncancer_data[ncancer_data\$sex==0,]) yesDrinking <-nrow(ncancer_data[ncancer_data\$drinking==1,]) noDrinking <-nrow(ncancer_data[ncancer_data\$drinking==0,]) yesSmoking <-nrow(ncancer_data[ncancer_data\$smoking==1,]) noSmoking <-nrow(ncancer_data[ncancer_data\$smoking==0,])</pre>
Step 2	<pre>sex <- c(MAN, WOMAN) drinking <- c(yesDrinking, noDrinking) smoking <- c(yesSmoking, noSmoking) ncancer_data <- data.frame(sex, drinking, smoking)</pre>

그림 14. R shiny barplot을 위한 요소 설정
 Fig. 14. Element Settings for R shiny Barplot

Step 1	<pre>function(input, output, session) { data_set<-read.csv("dataset.csv", stringsAsFactors = FALSE, fileEncoding = "UTF-8", encoding = "UTF-8") i <- nrow(data_set) pred<-data_set[i,9] pred output\$mente <- renderText({ paste("The way that is right for me : ") }) output\$method <- renderText({ paste(pred) })</pre>
Step 2	<pre>terms <- reactive({ input\$update isolate({ withProgress({ setProgress(message = "Processing corpus...") getTermMatrix(input\$selection) }) }) }) wordcloud_rep <- repeatable(wordcloud) output\$plot <- renderPlot({ v <- terms() wordcloud_rep(names(v), v, scale=c(5,0.1), min.freq = input\$freq, max.words=input\$max, rot.per=0.1, random.order = FALSE, random.color = TRUE, colors = brewer.pal(8, "Dark2"), family = "font") })</pre>
Step 3	<pre>output\$factorPlot <- renderPlot({ barplot(ncancer_data[,input\$factor], main=input\$factor, ylab="Number", xlab="Factor") }) }</pre>

그림 15. R shiny 웹 구현
 Fig. 15. R shiny Web Implementation

그림 14의 Step 1은 R shiny에서 barplot를 사용하기 위해 요소를 설정하는 소스코드이다. 수치코드로 변환된 요소들이 어떤 의미를 가지고 있는지 정의를 해준다. 그림 14의 Step 2는 수치코드로 변환된 요소들의 범주를 정의해주기 위한 소스코드이다.

그림 15는 R shiny를 사용하여 wordcloud와 barplot

를 웹 서버에 구현하기 위한 Server의 소스코드이다. 그림 15의 Step 1은 wordcloud와 함께 출력할 간병 방법 문구를 저장하는 소스코드이다. 그림 15의 Step 2는 그림 12에서 추출한 특정 질병에 해당하는 간병 방법들을 wordcloud로 구현하기 위한 소스코드이다. 그림 15의 Step 3은 특정 질병을 앓고 있는 환자들의 평균 특성을 barplot으로 구현하기 위한 소스코드이다.

그림 16은 R shiny를 사용하여 구현한 데이터를 웹 서버에 출력하기 위한 UI의 소스코드이다. 그림 16의 Step 1은 shiny 웹 서버에 사용되는 tag들을 디자인하고 출력하기 위한 소스코드이다. 그림 16의 Step 2는 wordcloud의 Frequency와 Number of Words 조건을 사용자가 직접 조절할 수 있도록 wordcloud와 함께 사용될 sidebar을 설정하기 위한 소스코드이다. 그림 16의 Step 3은 그림 15의 Step 2에서 구현한 wordcloud를 shiny 웹 서버에 출력하기 위한 소스코드이다. 그림 16의 Step 4는 barplot의 factor를 사용자가 직접 조절할 수 있도록 barplot와 함께 사용될 sidebar을 설정하기 위한 소스코드이다. 그림 16의 Step 5는 그림 15의 Step 3에서 구현한 barplot을 shiny 웹 서버에 출력하기 위한 소스코드이다. 그림 17을 통해 최종적으로 구현된 shiny 웹 서버를 확인할 수 있다.

Step 1	<pre>fluidPage(tags\$head(tags\$style(HTML(" h1, h3{ font-family: 'Monospace'; font-weight: bold; } h1{ color: #045FB4; })))), h1("Response method & Word Cloud"), h3(textOutput("mente")), h4(textOutput("method")), hr()</pre>
Step 2	<pre>sidebarLayout(sidebarPanel(selectInput("selection", "memu:", choices = cancers), actionButton("update", "Change"), hr(), sliderInput("freq", "Minimum Frequency:", min = 1, max = 50, value = 1), sliderInput("max", "Maximum Number of Words:", min = 1, max = 300, value = 100)), mainPanel(plotOutput("plot"))), hr()</pre>
Step 3	<pre>h1("Ratio of cancer patients"), sidebarLayout(sidebarPanel(selectInput("factor", "Factor:", choices=colnames(cancer_data), hr(), helpText("Number 1: MAN, Number 2: WOMAN"), helpText("Number 1: YES, Number 2: NO")), mainPanel(plotOutput("factorPlot"))))</pre>

그림 16. R shiny 웹 UI
Fig. 16. R shiny Web User Interface



그림 17. R shiny 웹 서버 구현 결과
Fig. 17. Results of R shiny web server implementation

그림 17의 1번 tag는 그림 15의 Step 1과 그림 16의 Step 1을 구현한 결과이다. 테스트 데이터인 그림 4를 분석하여 분류된 간병 방법 결과를 출력한 화면이다. 그림 17의 2번 sidebar는 그림 13과 그림 16의 Step 2의 구현 결과이다. 사용자는 sidebar의 Frequency, Number of Words 설정을 조정하면서 wordcloud 결과를 확인할 수 있다. 그림 17의 3번 wordcloud는 그림 15의 Step 2와 그림 16의 3의 구현 결과이다. 그림 17의 2번 sidebar에서 적용된 설정에 맞게 출력된 wordcloud 화면이다. 그림 17의 4번 sidebar는 그림

14와 그림 16의 Step 4의 구현 결과이다. 사용자는 sidebar에서 원하는 factor를 선택하여 barplot 결과를 확인할 수 있다. 그림 17의 5번 barplot은 그림 15의 Step 3과 그림 16의 Step 5의 구현 결과이다. 환자 질병에 걸린 환자들의 평균 특성을 그림 17의 4번 sidebar에서 선택한 factord에 맞게 출력된 barplot 화면이다. 그림 17을 통해 의학 지식이 없는 간병인도 웹 서버에 환자의 특성만 입력하면 간병 방법을 제공받을 수 있다. 또한, 환자와 같은 질병을 앓고 있는 환자들이 받은 진단 결과들을 wordcloud로 시각화하여 보여줌으로써 간병에 필요한 주요 키워드를 통해 환자에게 다양한 간병 방법을 제시할 수 있다. barplot을 통해서 같은 질병을 앓고 있는 환자들의 평균 특성을 보여줌으로써 나이, 담배, 술과 같은 질병의 원인을 제시할 수 있다. 따라서 환자 진단 데이터 분류 분석을 통해 병원에 가기 어려운 환자나 의학적인 전문 지식이 없는 간병인들도 환자에게 의사의 진단과 가까운 간병 방법을 제시할 수 있다.

IV. 결 론

본 논문에서는 환자 진단 데이터의 환자 특성에 따른 간병 방법을 분류 분석하여 진단 시간과 비용 등을 줄일 수 있도록 환자 혹은 간병인에게 간병 방법을 제공하기 위한 시스템을 구현하였다. 빅데이터 기반의 환자 진단 데이터 분석을 통해 병원에 가기 어려운 환자나 의학적인 전문 지식이 없는 간병인들에게 의사의 진단과 가까운 간병 방법을 제시할 수 있다. 또한 이와 같은 다양한 질병의 환자 진단 데이터를 활용하면 병원 진료에 투입되는 막대한 의료비용을 감소시킬 수 있을 것으로 전망된다.

References

- [1] Ji-Hye Lee, Mi-Kyung Je, Myung-Ji Cho, Hyun -Suk Son, "Trends in the Use of Big Data in Health Care," The Journal of The Korean Institute of Communication Sciences, Vol. 32, No. 1, pp. 63-75, 2014.12.
- [2] Sung-Soo Kim, Moon-young Chung, Tae-whi Lee, Jong-ho Won, "Implementation of Medical Data-Based Big Data Analytics Service," Journal of the Korea Information Science Society, pp. 157-159, 2015.12.

- [3] Gil-Ha Park, "A Trends of Big Data Processing for Digital Healthcare," Journal of Korea Creative Content Association, Vol. 15, No. 1, pp. 35-37, 2017.3.
- [4] Ye-Lim Kim, Chun-Ki Kwon, Yong-Hae Kong, "Real-time Web System Development for Effective Nursing & Care Integration Services," Journal of the Korea Academy Industrial Cooperation Society, Vol. 17, No. 12, pp. 41-52, 2016.12.
DOI: <https://doi.org/10.5762/kais.2016.17.12.41>
- [5] Uh-Soo Kyun, Sung-Hoon Cho, Jeong-Joon Kim, Young-Gon Kim, "A Study on Perception for Public Safety of Seoul Citizens using Multiple Regression Analysis," Journal of The Institute of Internet, Broadcasting and Communication, Vol. 18, No. 1, pp. 195-201, 2018.2.
DOI: <https://doi.org/10.7236/IIBC.2018.18.1.195>
- [6] Yong-Tae Kim, Yoon-Su Jeong, Gil-Cheol Park, "A Design of Web Server Architecture Environment for Reliability Enhancement and Secure Web Services," Journal of the Korea Institute of Information and Communication Engineering, Vol. 14, No. 2, pp. 343-350, 2010.2.
DOI: <https://doi.org/10.6109/jkiice.2010.14.2.343>
- [7] Hyun-il Sung, Sang Chul Kim, "Construction Of Astronomical Database Based On MySQL and JSP", Publications of The Korean Astronomical Society, Vol. 19, No. 1, pp. 109-119, 2014.
DOI: <https://doi.org/10.5303/pkas.2004.19.1.109>
- [8] Kan-gwon Shin, Sang-woo Shim, "Expressway Travel Time Prediction Using K-Nearest Neighborhood", Journal of the Korean Society of Civil Engineers, Vol. 34, No. 6, pp. 1873-1879, 2014.
DOI: <http://dx.doi.org/10.12652/Ksce.2014.34.6.1873>
- [9] Jongh-wa Na, Eun-ji Hwang, "Visualization analysis using R Shiny", Journal of the Korean Data And Information Science Society, Vol. 28, No. 6, pp. 1279-1290, 2017.
DOI: <http://dx.doi.org/10.7465/jkdi.2017.28.6.1279>

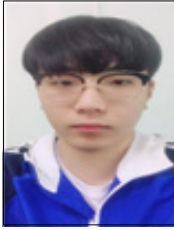
저 자 소 개

박 지 훈(준회원)



• Ji-Hun Park is currently studying at Korea Polytechnic University for a bachelor's degree in computer engineering in 2018. His research interests include the Database systems, Big Data, Data Analysis, etc.

황 승 연(준회원)



- Seung-Yeon Hwang is received his BS in Department of Computer Science at Korea Polytechnic University in 2019. He is currently studying MS in Department of Computer Science at Anyang University. His research interests include Database System, Big Data, Data Analysis, Machine Learning, etc.

윤 범 식(준회원)



- Bum-Sik Yun is currently studying at Korea Polytechnic University for a bachelor's degree in computer engineering in 2018. His research interests include the Network Security, Big Data, Security Analyzation, etc.

최 수 길(준회원)



- Su-Gil Choe is currently studying at Korea Polytechnic University for a bachelor's degree in computer engineering in 2018. His research interests include the Big Data, Autonomic Computing, etc.

이 돈 희(정회원)



- Don Hee Lee received his M.S degree in Computer Engineering from Yonsei University, Korea, in 2005, and a Ph.D. in Computer and Information Communication Engineering from Konkuk University, Korea, in 2016. He has been working for SK holdings Ltd, Bundang, Korea, from 2002 to present. His research interests include databases, big data, spatio-temporal index in wireless communications, location-based services, and information system audit.

김 정 준(정회원)



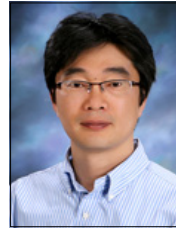
- Jeong-Joon Kim received his BS and MS in Computer Science at Konkuk University in 2003 and 2005, respectively. In 2010, he received his PhD in at Konkuk University. He is currently a professor at the department of ICT Convergence Engineering at Anyang University, His research interests include Database Systems, Big Data, Semantic Web, Geographic Information Systems (GIS) and Ubiquitous Sensor Network (USN), etc.

문 진 용(정회원)



- Jin Yong Moon received his MS in Computer Science at Konkuk University in 1998. Then he received PhD from Suwon University in 2001. He is currently a professor in the department of Visual Broadcasting Media at Gangdong University. His research interests include Database Systems, Web Science, Geographic Information Systems (GIS) and Multimedia Systems, etc.

박 경 원(정회원)



- Kyungwon Park received his BS and MS in Mathematics at Chung-Ang University in 1995 and 1998, respectively. In 2004, he received his Ph. D at University of South Carolina . He is currently a professor at the department of Computer Science at Korea Polytechnic University. His research interests include Nonlinear Approximation Theory, Image Analysis, Machine Learning, etc.