

비정형 데이터 마이닝을 활용한 한국농수산대학 재학생의 학교생활 감성 분석(1)

An Analysis of School Life Sensibility of Students at Korea National College of Agriculture and Fisheries Using Unstructured Data Mining(1)

주진수

J. S. Joo
국립한국농수산대학¹

이소영

S. Y. Lee
국립한국농수산대학
농수산비즈니스학과²

김종숙

J. S. Kim
국립한국농수산대학
농수산비즈니스학과²

송천영

C. Y. Song
국립한국농수산대학
화훼학과³

신용광

Y. K. Shin
국립한국농수산대학
농수산비즈니스학과²

박노복*

N. B. Park
국립한국농수산대학
화훼학과³

Abstract

In this study we examined the preferences of eight college living factors for students at Korea National College of Agriculture and Fisheries(KNCAF). Analytical techniques of unstructured data used opinion mining and text mining techniques, and the analysis results of text mining were visualized as word cloud. The college life factors included eight topics that were closely related to students: 'my present', 'my 10 years later', 'friendship', 'college festival', 'student restaurant', 'college dormitory', 'KNCAF', and 'long-term field practice'. In the text submitted by the students, we have established a dictionary of positive words and negative words to evaluate the preference by classifying the emotions of positive and negative.

As a result, KNCAF students showed more than 85% positive emotions about the theme of 'student restaurant' and 'friendship'. But students' positive feelings about 'long-term field practice' and 'college dormitory' showed the lowest satisfaction rate of not exceeding 60%. The rest of the topics showed satisfaction of 69.3~74.2%. The gender differences showed that the positive emotions of male students were high in the topics of 'my present', 'my 10 years later', 'friendship', 'college dormitory' and 'long-term field practice'. And those of female were high in 'college festival', 'student restaurant' and 'KNCAF'.

In addition, using text mining technique, the main words of positive and negative words were extracted, and word cloud was created to visualize the results.

Key words : Unstructured data, Opinion mining, Word cloud, Visualization, 8 topics of college living

*교신저자: nobokpark@naver.com

1 Korea National College of Agriculture and Fisheries

2 Department of Agriculture and Fisheries Business, Korea National College of Agriculture and Fisheries

3 Department of Floriculture, Korea National College of Agriculture and Fisheries

I. 서론

빅 데이터 분석이란 대량의 데이터로부터 숨겨진 패턴과 알려지지 않은 정보를 찾아내기 위한 과정을 의미하며, 그 목적은 데이터 과학자들에 의해 분석된 정보를 토대로 각 분야의 의사결정을 수행하는 데 있다. 또한 빅 데이터 분석은 개인이나 기업 등에서 자료를 토대로 어떠한 의사결정을 할 때에 중요한 정보로써 사용되기 때문에 최선의 대안을 선택할 수 있도록 근거를 제시하는 중요한 역할을 하고 있다. 그래서 불확실성이 높고 의사결정이 초래하는 파급효과가 큰 의사결정일수록 실제 데이터 분석을 바탕으로 의사결정을 하는 경향이 더욱 높아지고 있는 상황이다.

그러나 빅 데이터 분석은 기존 데이터 분석과 달리 처리할 데이터 양이 방대하고 비정형 데이터의 비중이 높다는 점을 빼놓을 수 없다. 비정형 데이터란 그림이나 영상, 음성, 문서처럼 구조화되지 않은 데이터로서 책, 잡지, 문서, 의료기록과 같은 텍스트 정보, 음성 정보, 영상 정보와 같은 전통적인 데이터 외에 이메일, 트위터, 블로그 처럼 모바일 기기와 온라인에서 생성되는 데이터들을 의미한다.

이러한 비정형 데이터들에 대한 분석 기법은 정형 데이터 분석 기법인 데이터 마이닝(Data Mining)의 한계를 극복하기 위한 방편으로 꾸준히 발전해 오고 있으며, 텍스트 마이닝(Text Mining), 평판 분석(Sentiment Analysis) 혹은 감성 분석(Opinion Mining), 웹 마이닝(Web mining) 등의 분석 기법 유형이 있다.

본 연구에서는 빅 데이터 시대에 접어들며 데이터를 보다 효율적으로 정확하게 분석하고 비즈니스 등의 영역에 적용하려는 노력이 꾸준히 진행되고 있는 상황을 배경으로 한국농수산대학(이하 한농대) 학생들의 학교생활 만족도를 조사하기

위하여 설문조사와 빅 데이터 분석을 실시하였다. 학교생활 만족도의 조사 항목으로는 '나의 현재', '10년 후의 모습', '교우관계', 대학 축제인 '한농제', '후생관 식사', 기숙사의 '청학관', '한농대', '장기현장실습' 등 학생들에게 밀접한 8가지 주제를 선정하였다. 데이터 분석은 비정형 데이터 마이닝(Unstructured Data Mining)의 분석 기법인 감성 분석(Opinion Mining) 기법의 활용과 분석 결과의 이해를 돕기 위한 시각화 과정 등으로 이루어진다. 설문조사는 창업 설계를 하고 있는 3학년 학생을 대상으로 하였으며, 학교생활 요소에 대한 긍부정 감성을 한농대 농어업·농어촌연구소(이하 연구소)에서 제시한 단어/문장 표에서 자신의 감성에 부합하는 문구를 선택하여 직접 기입하도록 하였다.

II. 연구내용

1. 분석 도구 및 기법

빅 데이터 분석 도구로는 R 프로그래밍 언어, 빅쿼리, 프레스토 등이 활용되고 있는데, 본 연구에서는 R 프로그래밍 언어에 기반한 RStudio를 활용하였다. R 프로그래밍 언어는 오픈소스 프로젝트로 통계 계산 및 시각화를 위한 언어 및 개발 환경을 제공하며 이를 통해 기본적인 통계 기법부터 모델링, 최신 데이터 마이닝 기법까지 구현이 가능하며 통계적 컴퓨팅 언어로 다양한 통계 분석에 용이하다. 현재 R 프로그래밍 언어를 이용하여 다양한 빅 데이터 분석 및 예측 분석 등을 포함한 고급 분석 기술들이 연구 및 개발되고 있으며, 이 언어의 장점 중 하나는 사용자가 제작한 패키지를 추가하여 기능을 확장할 수 있다는 것이다.

빅 데이터 분석 기법으로는 ①주로 수치형 데이터에 대하여 확률을 기반으로 어떤 현상의 추정, 예측을 검정하는 통계적 분석 기법, ②대용량 데이터로부터 패턴인식, 인공지능, 고급 통계분석 기법 등을 이용하여 데이터 간의 상호 관련성 및 유용한 정보를 추출하는 데이터 마이닝(Data Mining) 기법, ③대용량 소셜 미디어를 언어분석 기반 정보 추출로 탐지와 이슈의 모니터링을 통하여 향후 추이를 분석하는 소셜 네트워크 분석 기법이 있다. ④마지막으로 텍스트 기반의 데이터로부터 정보 검색, 추출, 체계화, 분석을 포함하는 Text-processing 기술 및 처리 과정인 텍스

트 마이닝(Text Mining) 기법이 있으며 이와 관련된 분야에 평판 분석(Sentiment Analysis) 혹은 감성 분석(Opinion mining)이라고 불리는 기법 등이 있다.

본 연구에서는 텍스트 마이닝에서 발전된 감성 분석(Opinion mining) 기법을 활용하여 앞에서 설명한 8가지 학교생활 요소에 대한 감성을 평가하였다. 조사 대상은 재학생 가운데 장기현장실습 과정을 마치고 졸업을 준비하는 3학년 학생으로 하였으며, 조사는 2018년 12월 3일부터 2주간 실시하였다(Table 1).

Table 1. The composition of survey respondents by major N=279, (단위 : 명)

성별	전공	특용	식량	중소	채소	대가축	말산업	수산	화훼	과수
남학생		29	12	33	23	24	19	27	17	35
여학생		4		7	12	5	11	3	12	6
합계		33	12	40	35	29	30	30	29	41

학교생활 관련 8개 주제에 대한 감성표현의 문장은 학생이 직접 작성하는 게 아니라 Table 2와 같이 8개의 감성 영역(전체 386개 어구)으로 구성하여 제시된 단어/문장 표에서 골라 설문지에 기입하도록 하였다.

학생들은 ‘나의 현재’를 비롯하여 ‘10년 후의 모습’, ‘한농대’, ‘장기현장실습’ 등 8가지의 주제

에 대하여 각각 5개 이상의 어절이나 문장을 선택 기입해야 하며, 표에 없는 감성어는 학생이 작성하여 기입하였다. 본 연구에서 사용한 어구나 문장은 심리상담치료에서 사용하는 표현 단어^{10)~12)}를 기반으로 조사목적에 맞게 연구소에서 새롭게 작성하여 표로 제시하였다.

Table 2. The sensibility classification area and number of a phrases N=386, (단위 : 개)

영역	기쁨, 행복, 편안, 즐거움	사랑, 관심, 부러움, 배려	분노, 화, 성남, 적대적	미움, 증오, 부끄러움, 욕심, 죄책감	슬픔, 우울, 괴로움	두려움, 의심	자신감, 능력, 약함	피곤, 냉담
개수	62	43	61	41	54	48	61	16

2. Opinion Mining과 긍·부정 단어 사전

사람이 작성하는 텍스트 안에는 그 글의 주요 대상이 되는 주제(Topic)가 있을 것이고, 그 주제에 대한 글쓴이의 의견(Opinion)이 있게 마련이다. 글의 대상이 되는 주제를 찾아내는 작업을 토픽 모델링(Topic Modeling)이라고 하며, 주제에 대한 글쓴이의 의견을 파악하는 것을 감성 분석(Opinion Mining, 또는 감성 분류, sentiment classification)이라고 한다.

감성 분석은 소셜 미디어 등의 정형/비정형 텍스트의 긍정(Positive), 부정(Negative), 중립(Neutral)의 선호도를 판별하는 기술로서 특정 서비스 및 상품에 대한 시장규모 예측, 소비자의 반응, 입소문 분석 등에 활용되고 있다.

본 연구에서는 이 점에 착안하여 ‘나의 현재’로부터 ‘장기현장실습’까지 8가지 학교생활 주제를 한농대가 제공하는 서비스 및 상품으로 간주하고 이를 3년 동안 경험한 3학년 학생들의 주관적 평가, 의견, 정서적 태도를 탐색하여 보이지 않는

정보를 추출하는데 의미가 있다 하겠다.

감성 분석의 수행단계는 긍정, 부정, 중립을 표현하는 어휘 정보를 추출하는 단계, 세부 평가 요소와 그것이 가리키는 오피니언의 연결 관계를 포함한 문장을 인식하는 단계, 그리고 긍정, 부정, 중립 표현의 수 및 유용한 문장을 추출하여 리뷰 요약을 생성하면 선호도를 분류하는 세 단계로 되어있다.

본 연구를 위하여 첫 번째 어휘 정보를 추출하는 단계에서는 기존에 구축된 사전 등의 리소스를 이용해야 하지만 한국어에 대해서는 아직 활용할 만한 리소스가 존재하지 않아 연구소에서 수작업을 통하여 긍정 단어 사전과 부정 단어 사전을 구축하였다(Fig. 1). 이 사전을 이용하여 두 번째 단계인 세부 평가 요소와 긍정적인 오피니언인지 부정적인 오피니언인지 표현을 결정하였다. 마지막으로 긍정/부정 표현의 횟수 및 중요 문장의 추출과 워드 클라우드(word cloud)에 의한 시각화를 통하여 정보를 추출하였다.

	A	B
4	가슴 벅차다	
5	가슴이 넓다	
6	가슴이 뭉클하다	
7	가슴이 벅차다	
8	가치있다	
9	감격스럽다	
10	감미롭다	
11	감사하다	
12	감사함을 느끼다	
13	개운하다	
14	경탄할만 하다	
15	고맙다	
16	공손하다	
17	공평하다	
18	관대하다	
19	관심많다	

(a) positive word dictionary

	A	B
1	가슴이 답답하다	
2	가슴이 미어지다	
3	가슴이 아프다	
4	가슴저리다	
5	가슴이 저리다	
6	가슴이 찢어진다	
7	가없다	
8	가치없다	
9	가혹하다	
10	거부감이 들다	
11	거부감이 든다	
12	걱정스럽다	
13	겁이 나다	
14	겁이 많다	
15	겁에 질리다	
16	경멸감을 느끼다	

(b) negative word dictionary

Fig. 1. Dictionaries of emotional words in this study

3. 프로그래밍

본 연구의 감성 분석은 학생들이 제출한 설문지 텍스트의 내용을 긍정과 부정으로 분류하는 작업으로서 이를 위해서는 긍정과 부정의 감성 단어에 대한 모음집 사전이 필요하다. 이 과정의 알고리즘을 간략하게 설명하면 R의 패키지를 사용하여 비정형 데이터 마이닝 과정, 즉 비정형 데이터인 텍스트에서 불필요한 데이터를 제거하고 정형화하는 작업을 하는 것이다. 그다음으로 긍정 단어 사전과 부정 단어 사전을 참고해서 긍정과 부정으로 매핑되는 단어 수를 합산하여 문장의

긍정과 부정을 결정한다.

본 연구에서 활용한 RStudio의 프로그래밍에는 한글 자연어 분석 패키지 KoNLP, 데이터 전처리에 사용할 plyr 및 stringr 등의 패키지를 이용하였다. plyr 패키지는 두 개 이상의 데이터 프레임에 대상으로 key값을 이용하여 하나로 병합(merge)하거나, 집단 변수를 기준으로 데이터 프레임의 변수에 함수를 적용하여 요약 집계 결과를 제공하는 패키지이다. 또한 주요 단어의 시각화를 위하여 wordcloud 패키지와 글자 색깔을 표현하는 RColorBrewer 패키지를 사용하였다.

```

13 #read data
14 positive <- readLines("positive10.txt")
15 negative <- readLines("negative10.txt")
16 positive <- gsub(' ', '', positive)
17 negative <- gsub(' ', '', negative)
18
19 sentimental = function(sentences, positive, negative) {
20   scores = lapply(sentences, function(sentence, positive, negative) {
21     sentence = gsub('[:punct:]', '', sentence)
22     sentence = gsub('[:cntrl:]', '', sentence)
23     sentence = gsub('\\d+', '', sentence)
24     sentence = tolower(sentence)
25     word.list = str_split(sentence, '\\s+')
26     words = unlist(word.list)
27     # compare our words to the dictionaries of positive & negative terms
28     pos.matches = match(words, positive)
29     neg.matches = match(words, negative)
30     # match() returns the position of the matched term or NA
31     # we just want a TRUE/FALSE:
32     pos.matches = !is.na(pos.matches)
33     neg.matches = !is.na(neg.matches)
34     # and conveniently enough, TRUE/FALSE will be treated as 1/0 by sum():
35     score = sum(pos.matches) - sum(neg.matches)
36     val <- c(score=score, pos = sum(pos.matches), neg = sum(neg.matches))
37     # return(val)
38     return(score)
39   }, positive, negative)
40   # score.df = data.frame(score = scores, text = sentences)
41   score.df = data.frame(scores, text = sentences)
42   return(score.df)
43 }
44 #read input data
45 FileN <- "Da형재(전체).csv"
46 txt <- read.csv(FileN, header = T)
47 txt <- txt[,4]
48 txt <- gsub(' ', '', txt)
49 #class(txt)
50 result <- sentimental(txt, positive, negative)
51 #write.csv(result, "result.csv")
52 #print of negative words by frequency
53 neg <- subset(result, result[,1]in% c(-1))
54 wordcount1 <- table(neg[,2])
55 #head(sort(wordcount1, decreasing = T), 40)
56 wordcount10 <- head(sort(wordcount1,decreasing=T), 40)

```

Fig. 2. Programming with RStudio(Ver. 1.1.442)

본 연구의 프로그래밍 순서를 요약하면 다음과 같다. ①setwd 함수를 사용하여 분석 파일이 있는 경로 선택을 한 후 readLines 함수를 이용하여 긍정 단어 사전과 부정 단어 사전을 읽어 들였다. ②로드한 사전에서 긍정/부정 스코어 및 단어를 값으로 돌려주는 'sentimenatl'이라는 함수를 정의하고 ③학생들이 제출한 CSV 파일 형식의 텍스트를 전공별, 성별 등으로 구분하여 로드하였다. ④sentimenatl 함수를 이용하여 분석 결과를 추출한 후 ⑤빈도수를 출력하기 위하여 table 함수를 이용하여 단어별 도수분포표로 변환하고 결과를 CSV 파일 형식으로 출력하였다. ⑥마지막으로 감성표현에 많이 사용된 단어들을 wordcloud 함수를 사용하여 그림으로 시각화를 하였다.

Fig. 2에 '나의 현재'에 대한 CSV 파일 형식의 분석 프로그래밍 순서의 일부를 나타낸다. wordcloud 함수에는 최대 빈도 단어와 최소 빈

도 단어 폰트 사이의 크기 차이, 색상, 최소 및 최대 빈도수, 90도 회전된 단어의 출력 비율 등 다양한 옵션을 설정하여 화면상으로 출력을 조절할 수 있다.

III. 연구결과

1. 주제별 특성

한농대 학생들이 재학 3년 동안 경험한 8가지 평가요소에 대한 감성 분석(Opinion Mining) 결과는 다음과 같다.

먼저 Fig. 3은 8가지 주제에 대한 조사자 전체의 분석 결과로써 '현재'는 나의 현재, '10년 후'는 나의 10년 후의 모습, '후생관'은 후생관 식사, '장기현장'은 장기현장실습을 의미한다. Fig. 3을 보면 재학생들의 '교우관계'와 '후생관 식사'에

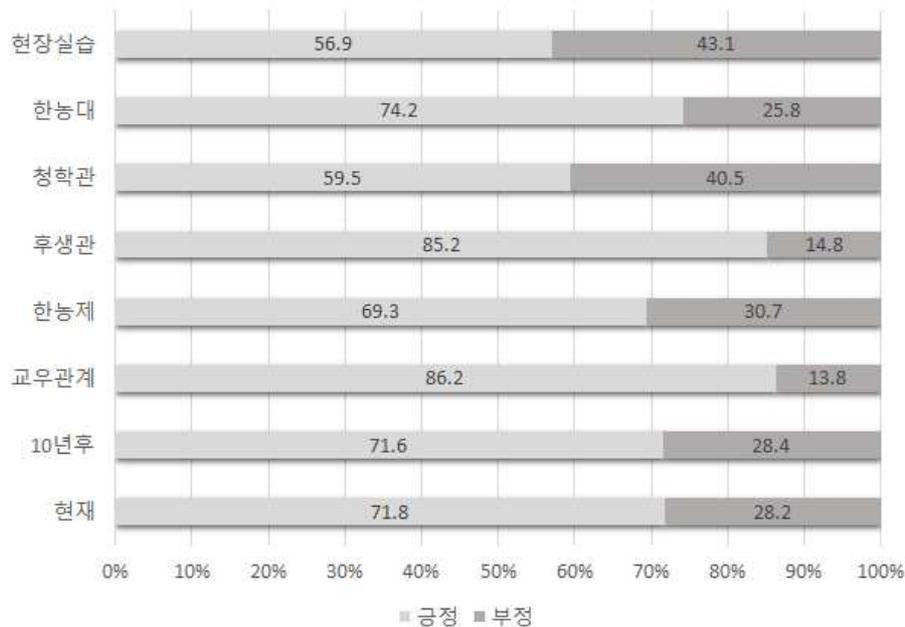


Fig. 3. The results of emotional evaluation on the eight topics of university life (light gray : positive, gray : negative)

대한 긍정적 감성은 85% 이상으로서 다른 주제들에 비하여 높은 호감도를 나타내는 것을 알 수 있다. 즉 재학생들은 학교 친구와는 친분을 충분히 쌓아 서로 친하게 지내는 인간관계를 형성하고 있음을 의미하며, 또한 아침, 점심, 저녁 3식을 제공하는 후생관에 대해서는 Fig. 5(e)의 '좋다', '감사함을 느끼다' 등의 감성어로 표현한 것과 같이 매우 높은 긍정의 만족도를 지니고 있음을 의미한다 할 수 있다.

이와 반대로 '장기현장실습'과 '청학관'에 대해서는 학생들의 긍정 감성이 60%를 넘지 않는 보통 수준의 만족도를 갖는 것으로 나타났다. 이는 Fig. 5(f)의 '골치 아프다', '가혹하다' 등으로 표현되는 기숙사 생활과 Fig. 5(h)의 '피곤하다', '괴롭다' 등으로 표현되는 장기현장실습에 대한 감성어에서 알 수 있듯이 젊은 학생들에게 기숙사 생활이 의무라는 것과 노동이 수반되는 장기간의 현장실습을 하는 것 자체에 대한 학생들의 솔직한 감성표현이라 할 수 있다.

다음으로 '나의 현재', '10년 후 모습', '한농제' 및 '한농대' 등의 주제에 대해서는 학생들의 긍정 감성은 69.3~74.2% 정도로서 약간 만족하는 것으로 나타났다.

2. 성별 특성

Table 3은 앞에서 살펴본 8가지 주제에 대하여 남녀 학생의 성별 차이를 나타낸 결과이며, 표 안의 ○은 여학생의 긍정 비율, ●은 여학생의 부정 비율이 남학생보다 높은 것을 나타낸다.

먼저 '현재', '10년 후', '교우관계', '청학관' 및 '장기현장실습' 등의 5개 주제에서 남학생, '한농제', '후생관' 및 '한농대' 등의 나머지 주제에 대해서는 여학생의 긍정적 감성이 높은 결과를 보이고 있다. 이는 바꾸어 말하면 전자의 5개 주제에서는 여학생, 후자의 3개 주제에서는 남학생의 부정적 감성이 높다는 것을 의미한다. 특히 '장기현장실습' 주제에서 여학생들의 부정적 감성이

Table 3. The results of emotional evaluation on the eight topics of university life by sex

topic	pos./neg.	male	female	note
my present	positive	73.8	64.2	
	negative	26.2	35.8	●
10 years later	positive	74.1	62.3	
	negative	25.9	37.7	●
a network of friends	positive	87.4	81.9	
	negative	12.6	18.1	●
Hannongje(college festival)	positive	67.5	75.9	○
	negative	32.5	24.1	
Husaengkwan(student restaurant)	positive	84.5	87.8	○
	negative	15.5	12.2	
Chunghakkwan(college dormitory)	positive	60.3	56.5	
	negative	39.7	43.5	●
KNCAF	positive	74.1	74.6	○
	negative	25.9	25.4	
Long-term field practice	positive	60.4	44.6	
	negative	39.6	55.4	●

○ : Girls have a higher percentage of positive words than boys

● : Girls have a higher percentage of negative words than boys

55.4%로서 8개 주제 가운데 유일하게 부정적 감성이 절반을 초과하는 것으로 나타났다. 이는 여학생들의 신체적 조건 및 현장실습장 환경 등의 영향이 다른 주제보다 부정적 이미지를 많이 갖게 한 것으로 여겨진다.

여학생들은 남학생들에 비하여 대학 축제, 후생관 식사 및 한농대에 대하여 긍정적 감성을 지니고 있으나 남학생에 비하여 자신의 현재나 미래에 대한 불안감, 재학 중 의무적인 기숙사 생활과 친구와의 인간관계 그리고 장기간에 걸쳐 이루어지는 현장실습 등에서는 불편함과 어려움이 높은 것으로 나타났다.

3. 전공별 특성

한농대는 작물계열, 원예계열, 산림계열, 축산계열, 수산계열 등 다양한 전공의 학생들이 재학 중에 있어 전공별 특성이 매우 다르다 할 수 있다. 따라서 8가지 연구 주제에 대하여 각 전공별로 긍부정 감성 특성을 비교한 결과를 Fig. 4에 나타낸다. Fig. 4의 세로축 상단의 '전체'는 각 주제에 대한 전체 응답자에 대한 응답, 세로의 굵은 점선은 '전체'의 긍부정 경계값과 각 전공별 결과와의 비교를 위한 연장선을 의미한다. 여기서는 응답자 수가 12명으로 적었던 식량작물학과의 분석 결과는 그림에 나타내지 않았다.

먼저 '나의 현재'에 대한 주제에서는 대가축학과, 중소가축학과, 말산업학과 학생들의 긍정 감성 비율은 다른 전공보다 높으며, 과수학과, 화훼학과, 채소학과, 특용작물학과의 긍정 비율은 전체보다 낮은 결과를 보이고 있다. 즉 축산계열 학생들은 다른 계열의 학생들에 비하여 현재의 자신에 만족하고 호감도가 높은 결과를 나타냈다(Fig. 4(a)).

또한 축산계열 학생들은 '10년 후' 미래의 자기 자신에 대해서도 자신감이 높게 나타났으며, 나머지 다른 계열 학생들은 현재에 비하여 미래를 희

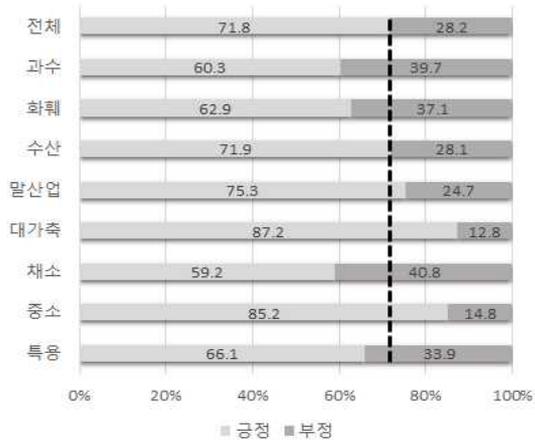
망적으로 생각하는 긍정적 사고 비율이 높아지는 결과를 나타낸다. 그러나 화훼학과 학생들은 현재에 대한 긍정 감성보다 미래에 대한 긍정 감성이 낮게 나타나 미래에 대한 불안감이 더 크게 나타나는 결과를 보인다. (Fig. 4(b)).

'교우관계'에 대한 주제에서는 화훼학과를 제외한 모든 전공에서 80% 이상의 긍정적 감성을 나타내고 있으며 특히 중소가축학과 학생들은 93%를 초과하는 매우 긍정적 사고를 하는 것으로 나타났다(Fig. 4(c)).

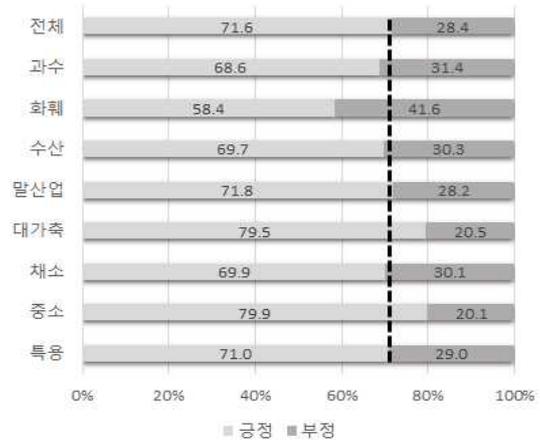
대학 축제인 한농제에 대한 이미지를 보면 과수학과, 수산양식학과 및 특용작물학과 학생들은 나머지 전공의 학생들보다 부정적 이미지가 큰 것으로 나타났다(Fig. 4(d)). 특히 수산양식학과 학생들은 대학 축제에 대한 부정적 감성 비율이 47.3%로서 거의 절반의 학생이 한농제에 대하여 호의적이지 않는 감정을 나타냈다.

한농대 학생들은 기숙사 생활이 의무이므로 후생관에서 하루 세끼 식사를 해결해야 하는데 후생관 식사에 대한 전체 학생들의 만족도는 85% 이상으로 매우 높게 나타났으나, 수산양식학과 학생들의 만족도는 66.5%로 그다지 높지 않은 결과를 보이고 있다(Fig. 4(e)). 수산양식학과 학생들의 만족도를 성별로 분석한 결과를 보면 여학생은 약 85%, 남학생은 약 65%로서 남학생들의 만족도가 낮게 나타났다. 후생관은 영양과 위생에 중점을 두고 운영하고 있으며, 영리를 목적으로 하지 않는 직영방식으로 운영하고 있어, 학생들의 건강한 식생활을 위해 친환경 식재료 및 천연조미료를 사용하는 등 평생 건강 기틀을 마련을 위한 급식을 제공하고 있는 게 긍정적 평가를 받는 이유로 여겨진다.

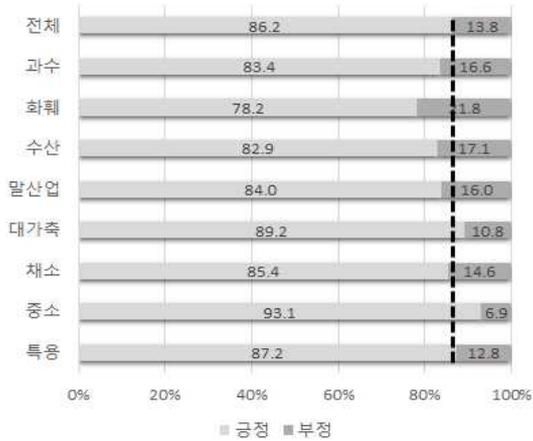
청학관 기숙사 생활에 대한 전체 학생들의 호감도는 59.5%로 낮게 나타났으나 이는 대학생들에게 기숙사 입사를 의무화하여 규칙적인 공동생활을 하도록 규정한 것이 원인이라 생각할 수 있다. 특히 과수학과와 수산양식학과 학생들의 만족도



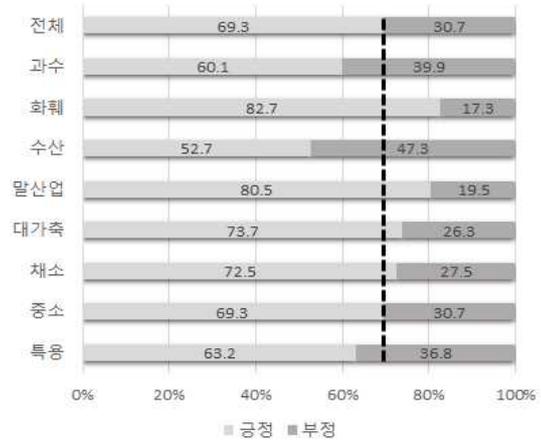
(a) my present



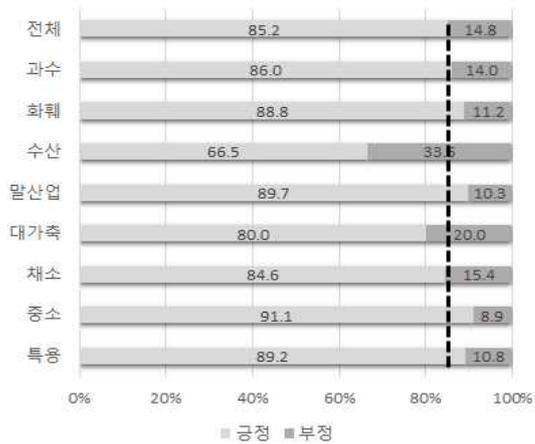
(b) 10 Years Later



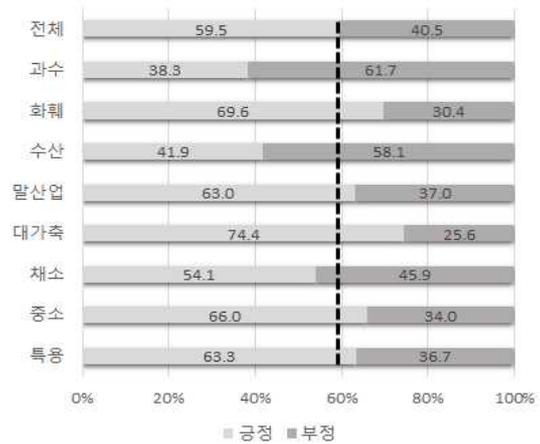
(c) friendship



(d) Hannongje(college festival)



(e) Husaengkwan(student restaurant)



(f) Chungchakkwan(college dormitory)

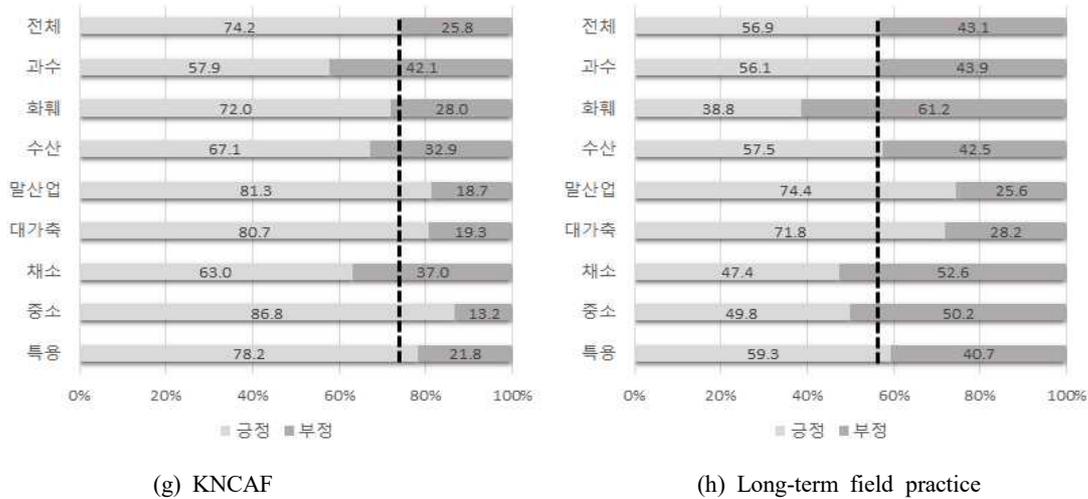


Fig. 4. The results of emotional evaluation by major on the 8 topics of university life (..... : baseline of the whole response)

Table 4. Rankings of positive response by major on 8 topics

주제	전공	특용	중소	채소	대가축	말산업	수산	화훼	과수
나의 현재		66.1	85.2	59.2	87.2	75.3	71.9	62.9	60.3
10년 후 모습		71.0	79.9	69.9	79.5	71.8	69.7	58.4	68.6
교우관계		87.2	93.1	85.4	89.2	84.0	82.9	78.2	83.4
한농제		63.2	69.3	72.5	73.7	80.5	52.7	82.7	60.1
후생관식사		89.2	91.1	84.6	80.0	89.7	66.5	88.8	86.0
청학관		63.3	66.0	54.1	74.4	63.0	41.9	69.6	38.3
한농대		78.2	86.8	63.0	80.7	81.3	67.1	72.0	57.9
장기현장실습		59.3	49.8	47.4	71.8	74.4	57.5	38.8	56.1
평균		72.2	77.6	67.0	79.6	77.5	63.8	68.9	63.8

청학관 기숙사 생활에 대한 전체 학생들의 호감도는 59.5%로 낮게 나타났으나 이는 대학생들에게 기숙사 입사를 의무화하여 규칙적인 공동생활을 하도록 규정한 것이 원인이라 생각할 수 있다. 특히 과수학과와 수산양식학과 학생들의 만족도는 각각 38.3%와 41.9%로 나타나 매우 부정적인 이미지를 보였다(Fig. 4(f)).

한농대에 대한 이미지는 74.2%의 긍정적 평가를 하고 있으나 과수학과(57.9%), 채소학과(63%)

그리고 수산양식학과(67.1%) 학생들 긍정적 만족도는 낮게 나타났다(Fig. 4(g)).

마지막으로 재학생들이 2학년 교과과정으로 의무적으로 수행하는 장기현장실습에 대한 긍정적 호감도는 56.9%로 8가지 주제 가운데 가장 낮은 만족도를 나타냈다. 특히 화훼학과(38.8%), 채소학과(47.4%) 및 중소가축학과(49.8%) 학생들은 매우 낮은 만족도를 보이고 있으나 말산업학과(74.4%)와 대가축학과(71.8%) 학생들은 다른 전

공의 학생들에 비하여 장기현장실습에 대한 긍정적 평가도가 높은 결과를 나타냈다(Fig. 4(h)).

다음 Table 4는 Fig. 3에 설명한 전공별 8가지 주제에 대한 긍정어 응답률과 그 평균을 나타낸다. 긍정어 응답률이 높을수록 학생들의 호감도 혹은 만족도가 높다는 것을 의미한다. 긍정어 응답률의 평균은 대가축학과, 중소가축학과, 말산업학과 등의 축산계열 학생들이 높게 나타났으며, 과수학과, 수산양식학과, 채소학과 및 화훼학과의 학생들의 긍정어 응답률은 70%를 넘지 않는 결과를 나타냈다.

4. 시각화

빅 데이터 시각화는 빅 데이터 분석 결과를 쉽게 이해할 수 있도록 시각적으로 표현하여 전달하는 과정을 말하는 것으로 분석에서 나타난 정보를 효과적으로 전달하는 것이다. 본 연구에서는 앞에서 설명한 8가지 주제에 대한 분석 결과의 시각화를 위하여 텍스트 마이닝 기법으로 구조화되지 않은 텍스트에서 주요 단어를 추출하고 단어들의 word cloud를 작성하여 학생들의 감성을 시각화하였다(Fig. 5). Fig. 5는 8가지 주제에 대하여 긍정어와 부정어에 대한 각각의 word cloud로서 출현 빈도수가 높은 문구가 크게 표현된다.

먼저 '나의 현재'에 대한 긍정 감성어로는 '개운하다', '기분 좋다', '감격스럽다', '기쁘다', '가슴 벅차다' 등의 어구가 많이 나타났으며, 부정 감성어는 '걱정스럽다'가 가장 많았으며 '기분이 묘하다', '막막하다', '골치 아프다' 등의 현실을 부정하는 표현이 많이 나타났다(Fig. 5(a)).

미래의 자신의 모습을 상상하는 '10년 후 모습'에 대한 긍정 감성어는 '능력 있다', '가슴 벅차다', '멋있다', '당당하다', '멋지다', '훌륭하다' 등의 자신감이 넘치는 어구가 많이 나타났으며, 부정어에는 '나의 현재'와 마찬가지로 '걱정스럽다'

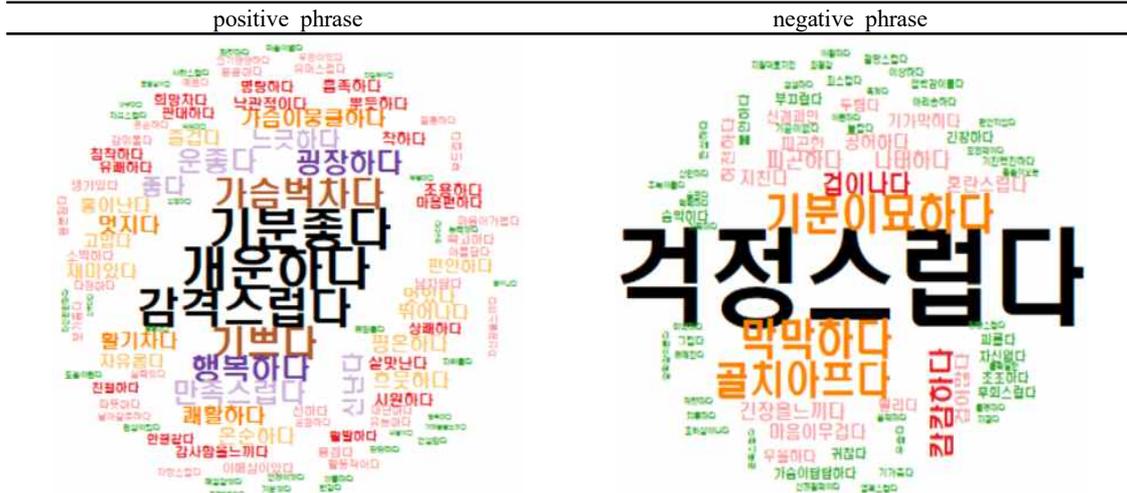
가 가장 많았으며 '겁이 난다', '막막하다', '기분이 묘하다', '혼란스럽다', '골치 아프다' 등의 자신감이 결여된 표현이 나타났다(Fig. 5(b)).

8가지 주제 가운데 학생들이 가장 높은 호감도를 보인 '교우관계'에 있어서는 '고맙다', '감사함을 느끼다', '우정이 있다' 등의 깊은 우정에 대한 감사함을 표현하는 감성어구가 많이 나타났으며, 부정 감성어는 '걱정스럽다', '미안하다', '골치 아프다', '밉다' 등의 어구가 많았다(Fig. 5(c)).

추수감사제 성격의 대학 축제 '한농제'에 대한 긍정 표현에는 '신난다', '흥겹다', '재미있다', '즐겁다', '흥이 난다' 등 축제의 즐거움과 흥겨움을 나타내는 어구가 많았으며, 부정 표현에는 '골치 아프다', '지루하다', '피곤하다', '귀찮다' 등의 어구가 많이 나타났다(Fig. 5(d)).

한편 '교우관계'와 더불어 매우 높은 호감도를 나타낸 '후생관 식사'에 대해서는 '좋다', '감사함을 느끼다', '흡족하다', '고맙다', '감사하다' 등의 감사와 만족감을 표현하는 어구가 많이 사용되었으며, '성에 안찬다', '지겹다', '당황스럽다' 등 약간의 부정적 감성어가 사용되었다(Fig. 5(e)). 여기서 음식에 대한 맛을 평가하는 '맛있다' 어구는 빈도 순위가 20위로 나타났으나 남학생(22위)과 여학생(10위)의 성별 간에는 분명한 차이를 보였다. 이는 혀의 표면에는 맛을 느끼는 미뢰가 여자가 남자보다 많아 남자에 비해 미각이 훨씬 예민하고 공감능력이 뛰어난 게 원인으로 여겨진다.

기숙사 생활을 평가한 '청학관'에 대한 긍정 감성어는 '편안하다', '재미있다', '좋다', '활기차다', '신난다' 등이 많이 나타났으며, 부정 감성어로는 '골치 아프다', '가혹하다', '걱정스럽다', '숨 막히다', '귀찮다' 등의 표현이 많이 나타났다(Fig. 5(f)). 기숙사 입사가 의무라면 한농대 학생뿐만 아니라 모든 학생들에게 긍정적 이미지를 기대하기 어려울 것이라 여겨진다. 따라서 장기현장실습과 함께 긍정적 이미지가 60%를 넘지 않는 주제로 판명되었다(Fig. 3).



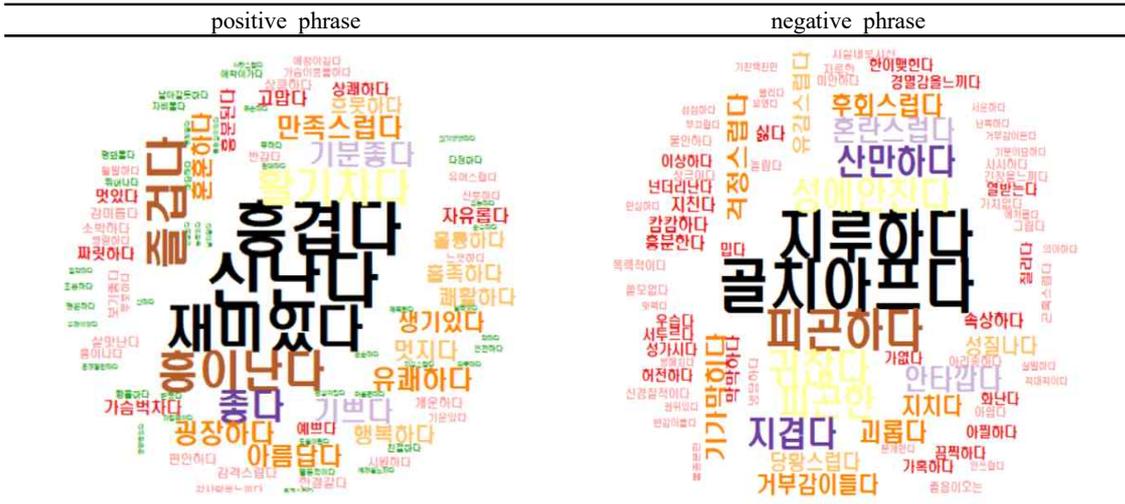
(a) my present



(b) my 10 years later



(c) friendship



(d) Hannongje(college festival)



(e) Husaengkwan(student restaurant)



(f) Chunghakkwan(college dormitory)

표현이 많이 나타났다(Fig. 5(h)).

IV. 적요

본 논문은 빅 데이터 분석기법을 이용하여 한국농수산대학 학생들의 대학생활 요소에 대한 선호도를 연구하기 위하여 비정형 데이터 분석기법으로서 감성 분석(opinion mining) 기법과 텍스트 마이닝 기법을 활용하였다. 분석도구로는 RStudio를 이용하였으며, 긍정과 부정의 감성을 분류하고 선호도를 평가하기 위한 긍정어 사전과 부정어 사전을 새롭게 작성하여 프로그래밍하였다. 비정형 텍스트에 대한 분석 결과는 도표와 워드 클라우드를 이용한 시각화 자료로 나타내어 정보를 추출하였다. 학교생활 요소로는 '나의 현재', '10년 후 모습', '교우관계', '한농제(대학 축제)', '후생관(식사)', '청학관(기숙사)', '한농대', '장기현장실습' 등 학생들에게 밀접한 8가지 주제를 대상으로 하였다.

분석 결과 한농대 학생들은 '후생관 식사'와 '교우관계'의 주제에 대해서 85% 이상의 긍정적 감성을 나타냈으나 '장기현장실습'과 '청학관(기숙사)'에 대해서는 긍정적 감성이 60%를 넘지 않는 만족도를 갖는 것으로 나타났다. 그리고 '나의 현재', '10년 후 모습', '한농제(대학 축제)' 및 'KNCAF' 등의 주제에 대해서는 69.3~74.2% 정도의 긍정적 감성을 나타냈다.

남녀 학생별 차이를 보면 '나의 현재', '10년 후 모습', '교우관계', '청학관(기숙사)' 및 '장기현장실습' 주제에서는 남학생의 긍정적 감성이 높게 나타났으며, '한농제(대학 축제)', '후생관' 및 '한농대' 주제에서는 여학생의 긍정적 감성이 높게 나타났다.

전공별 특징을 살펴보면 학생들은 '현재'나 '10년 후'의 자신의 모습에 대하여 71% 이상 긍정적 자신감을 지니고 있는 것으로 나타났다. 특히 축

산계열 학생들의 긍정적 감성이 높게 나타났으며, 화훼학과 학생들은 다른 전공의 학생들에 비하여 긍정적 감성이 낮게 나타나 자신감이 부족한 결과를 보였다.

'교우관계'에 대해서는 화훼학과를 제외하고 80% 이상의 긍정적 감성을 나타냈으며, 중소가축학과 학생들은 93%를 초과하는 적극적인 교우관계를 맺고 있는 것으로 나타났다. 대학 축제인 '한농제'에 대하여 전체 학생들의 긍정적 감성은 약 70% 정도이나 과수학과와 수산양식학과 학생들의 호감도는 60% 미만으로 축제에 대한 부정적 이미지가 높게 나타났다. '후생관 식사'에 대한 전체 학생들의 긍정적 감성은 85%를 넘어 매우 높은 만족도를 나타냈으나 수산양식학과 학생(남학생)들의 만족도는 매우 낮게 나타났다.

모든 학생들이 공동생활을 하는 '청학관'에 대한 학생들의 호감도는 59.5%로 낮게 나타났으며, 과수학과와 수산양식학과 학생들의 만족도는 약 42% 미만으로 더욱 부정적인 감성을 나타냈다. 또한 자신들이 3년간 학업을 재학한 한농대에 대해서는 74% 이상이 긍정적인 평가를 하는 것으로 나타났다. 특히 학생들의 호감도가 가장 낮게 나타난 장기현장실습에 대한 화훼학과, 채소학과, 중소가축학과 학생들의 호감도는 50%를 넘지 않는 매우 부정적 감성을 나타냈다.

빅 데이터 분석 결과를 쉽게 이해할 수 있도록 시각적으로 표현하기 위하여 텍스트 마이닝 기법으로 구조화되지 않은 텍스트에서 주요 단어를 긍정어와 부정어로 나누어 추출하고 그 단어들의 word cloud를 작성하여 학생들의 감성을 시각화하였다.

한농대는 학생들에게 지금보다 더욱 긍정적인 감성을 가지고 밝고 환한 말, 힘이 되고 용기를 주는 말, 사람을 기쁘게 하는 말을 많이 할 수 있도록 여건을 제공함으로써 학생들은 삶의 활기가 넘치고 성공적인 인생을 살아가는 행복을 만들 수 있으리라 여겨진다.

V. 참고문헌

1. 김경태, 안정국, 김동현. (2018). 빅 데이터 활용서 (I). 시대인.
2. 김영우. (2017). 쉽게 배우는 R 데이터 분석, 이지스퍼블리싱.
3. 나종화. (2017). R 데이터마이닝, 자유아카데미.
4. 남길임, 조은영. (2017). 한국어 텍스트 감성 분석, 커뮤니케이션북스.
5. 조민호. (2019). 데이터 분석 전문가를 위한 R 데이터 분석. 정보문화사.
6. 주진수 외 3인. (2018). 한국농수산대학 졸업생 영농정착 성공 사례집의 Text Mining. 현장농수산연구지 Vol. 20, No.2: 57-72.
7. 황인욱 외 3인. (2017). 청년농업인의 영농정착 성공 특성 및 영농수준에 관한 연구. 현장농수산연구지 Vol. 20, No.1: 145-162.
8. <http://100.daum.net/encyclopedia/view/125XX53100024>
9. <http://antilibary.org/490>
10. <http://blog.naver.com/PostView.nhn?blogId=sunilee0324&logNo=221312513268>
11. <http://cafe.daum.net/busanopenheart/GQTH/615?q=%EA%B0%90%EC%A0%95%ED%91%9C%ED%98%84%20%EB%8B%A8%EC%96%B4>
12. <http://cafe.daum.net/spyouth/6q2Q/248?q=%EA%B0%90%EC%A0%95%ED%91%9C%ED%98%84%20%EB%8B%A8%EC%96%B4>
13. <http://kutar37.tistory.com/27>
14. <https://mrkevinna.github.io/R-%EA%B8%B0%EC%B4%88-4/>

논문접수일 : 2019년 3월 15일

논문수정일 : 2019년 5월 10일

게재확정일 : 2019년 5월 20일