

비정형데이터 수집을 통한 드라마 시청률 연관어 분석

강선경¹ · 이현창² · 신성윤^{3*}

Analysis of drama viewership related words through unstructured data collection

Sun-Kyoung Kang¹ · Hyun-Chang Lee² · Seong-Yoon Shin^{3*}

¹Department of Computer Software Engineering, Wonkwang University, Iksan, 54538, South Korea

²Department of Digital Contents Engineering, Wonkwang University, Iksan, 54538, South Korea

^{3*}School of Computer Information & Communication Engineering, Kunsan National University, Kunsan 54150, South Korea

요약

본 논문에서는 드라마의 시청률에 영향을 미치는 연관어 분석을 위해 정형화된 데이터와 비정형화된 데이터를 분석하는 내용이다. 정형화된 데이터 수집은 각 방송사의 드라마정보, 인물정보, 방송정보, 시청률정보라는 4가지 영역에서 총 19가지항목을 수집하였다. 비정형데이터는 각 방송사에서 드라마별로 운영되고 있는 게시판과 방영전 블로그와 방영후 블로그로부터 크롤링기법을 이용하여 수집하였다. 수집된 정형데이터로부터 각 방송사별 4가지 영역별에 따른 차이를 비교한 결과 방송사별 서로 유사한 결과 값을 보이고 있었다. 그리고 각 방송사의 드라마별 게시판과 블로그에서 수집된 비정형데이터로부터 출현빈도의 상관관계 분석을 통해 관련 연관어를 7개 도출하였다. 도출된 연관어는 신뢰성 분석을 통해 이루어졌다.

ABSTRACT

In this paper, we analyzed the stereotyped and non - stereotyped data in order to analyze the drama 's ratings. The formalized data collection collected 19 items from the four areas of drama information, person information, broadcasting information, and audience rating information of each broadcasting company. Atypical data were collected from bulletin boards, pre - broadcast blogs and post - broadcast blogs operated by each broadcasting company using a crawling technique. As a result of comparing the differences according to the four areas for each broadcaster from the collected regular data, the results were similar to each other. And we derived seven related words by analyzing the correlation of occurrence frequencies from unstructured data collected from bulletin boards and blogs of each broadcasting company. The derived associations were obtained through reliability analysis.

키워드 : 정형데이터, 비정형데이터, 연관어, 크롤링, 텍스트 마이닝

Key word : Structured data, Unstructured data, related words , crawling, Text mining

Received 08 June 2017, Revised 18 June 2017, Accepted 05 July 2017

* Corresponding Author Seong-Yoon Shin(E-mail:s3397220@kunsan.ac.kr , Tel:+82-63-469-4860)

School of Computer Information & Communication Engineering, Kunsan National University, Kunsan 54150, South Korea

Open Access <https://doi.org/10.6109/jkiice.2017.21.8.1567>

print ISSN: 2234-4772 online ISSN: 2288-4165

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/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. 서론

최근 들어 빅데이터를 활용하여 기업이나 정부 그리고 민간들간의 어떠한 방향제시나 융합지식을 도출할 수 있는 방법을 찾기 위한 예측기술들이 활발히 이루어지고 있다. 또한 바쁜 일상에서 사람들이 선호하는 트렌드나 좋아하는 연예인, 노래, 영화, 드라마, 등을 미리 분석하여 예측하려고 하는 노력도 다양한 분야에서 이루어지고 있다[1]. 이러한 현상은 모바일의 확산과 더불어 정형화된 자료뿐만 아니라 다양한 형태의 비정형화된 자료로부터 새로운 정보가 생성되고, 많은 정보가 전달 및 공유되는 활동이 활발히 이루어지고 있는 현상으로 우리 일상생활에 많은 부분을 차지하고 있기 때문이다. 최근에는 다양한 SNS 매체들로부터 정보가 생산되고 배포되는 현상으로 인하여 많은 자료나 정보들로부터 유의미한 정보를 추출하는 기술들이 연구되고, 이를 사용하여 개인화되어가는 현실생활에 좀 더 나은 방향을 제시해주는 방법으로 활용해 나가려는 추세이다. SNS로부터 수집된 방대한 다양한 자료들을 이용하여 대중의 집단지성에 표출된 일반적인 감정이나 선호도를 분석하여 다양한 분야에 활용하려는 현상이 많다 [2, 3]. SNS의 데이터에는 많은 정보와 다양한 생각들을 담고 있어 이에 대한 텍스트(Text mining), 감성분석에 관한 연구도 꾸준히 이루어지고 있다. 텍스트마이닝은 텍스트 형태의 비정형 데이터에 마이닝 기법을 적용한 것으로 텍스트에 포함된 단어들을 분석하여 단어 간의 관계를 파악하여 의미 있는 정보를 찾아내는 기법이라 할 수 있다. 뉴스, 트위터, 블로그, 게시판, facebook 등 텍스트로 구성된 비정형 데이터를 분석하는 방법들 중의 하나인 감성분석은 오피니언마이닝(Opinion mining)의 한 종류로 언론사에서 제공하는 뉴스, 사람들 사이에 공유되는 트위터, 블로그, 인터넷 포털에 게시된 글 등의 감성을 분석해 텍스트가 글의 주제에 대해 가지고 있는 감성이 긍정인지 부정적인지에 대해 분석해 내는 기법이다[4, 5]. 이렇게 분석된 연관어는 많은 분야에 활용되어지고 있다. 그 중 인기 있는 드라마에서 사용되는 언어나 패션 스타일, 가전제품, 출연배우 등이 시청자의 정서나 생활, 그리고 드라마 상영 중간접 광고로 인한 수익창출 및 시청자들의 소비패턴 등에도 많은 영향력을 끼치는 핵심요소로 자리 잡고 있다. 그로인해 이 드라마의 성공 요인이 사회적, 문화적, 경

제적 변화 요인으로까지 자리 잡을 수 있다. 본 논문에서는 드라마의 성공을 판단할 수 있는 시청률에 영향을 미치는 핵심 요소인 연관어를 찾아내어 분석하고자 한다[1, 2].

II. 관련연구

본 논문에서는 드라마의 성공을 판단할 수 있는 시청률에 영향을 미치는 연관어분석을 위해 텍스트 마이닝 기법을 사용한다. [4]

2.1. 데이터 서술 분석 기법

서술 분석 방법은 주어진 데이터로부터 현재의 상황을 파악하고 이를 적절하게 이해할 수 있도록 돕기 위한 분석방법론이다. 과거 또는 현재의 주어진 상황에서 발생하는 이벤트에 대한 설명을 제공하거나, 지금 일어나는 현상에 대한 요인 또는 일어날 징후를 찾아서 보여주는 것을 목적으로 하고 있으며, 서술 분석을 위해서는 전통적인 통계와 데이터 마이닝 기법들이 많이 사용되고 있다. 이는 특정 대상을 분석하기 위함보다는 데이터에 내재되어 있는 대상들의 관계를 발견하거나 이들 서로간의 형태를 설명할 수 있는 규칙을 도출하고자 할 때 많이 사용하는 방법이다.

서술 분석의 종류에는 연관성분석(Association Analysis), 집단 그룹핑(Grouping)을 위한 클러스터링(Clustering), 범주분류(Classification of Category) 등의 기법들이 있다.

2.2. 데이터 예측 분석 기법

예측 분석이란 앞으로 일어날 사건을 예측하기 위해 기존에 일어났던 현상이나 현재의 사실들을 분석하는 다양한 종류의 통계, 모델링, 기계학습(Machine learning) 및 데이터마이닝 기법 등을 모두 일컫는 분석 방법론이다. 서술 분석방법과의 다른 부분은 예측분석은 명확한 예측할 대상에 대한 변수를 규정하고 이에 대한 설명 변수를 결정하며, 예측대상 변수와 설명변수 사이의 관계를 명확히 규명하는 것에 초점을 둔다는 것이다. 예측분석 기법은 기존의 통계적 기법과 기계 학습적 기법으로 구분되어진다. 또한 통계적 기법의 종류에는 선형회기분석(Linear Regression model), 로지스

틱분석(Logistic Regression), 시계열분석(Time series analysis) 등의 분석기법이 존재하며, 기계학습 기법에는 의사결정트리(Decision Tree), 신경망(Neural networks), 유전자 알고리즘(Genetic algorithms), 베이저안 빌리브 네트워크(BBN : Bayesian Belief Network) 등의 분석 기법이 있다.

2.3. 데이터 최적화 분석 기법

최적화 분석기법은 다른 표현으로 처방분석(Prescriptive Analytics)이라 하며, 보통예측분석을 포함하고 있어서 예측을 기반으로 하는 여러 대안 중 최적의 의사결정에 대한 제안까지 포괄적으로 적용 및 활용되어지고 있는 방법이다. 어떠한 사건이나 결과를 예측하기 위한 기업과 조직에서는 대부분 의사결정을 산출하기 위해 수행해야 될 방법 중의 하나로써 이를 이용하여 산출된 결과값은 조직의 의사결정을 수행할 때 고려해야 되는 중요한 변수와 제약조건으로 작용될 수 있으므로 궁극적으로는 이들 예측과 의사결정에 대한 최적화를 추구하고자하는 경우 많이 사용되는 방법이다.

III. 데이터 수집

3.1. 정형데이터 수집 방법

본 논문에서는 드라마 시청률과 관련된 연관어 분석을 위해 그림1과 같이 드라마 정보, 인물정보, 방송정보, 시청률 정보라는 큰 분류를 4가지 영역으로 나누었으며, 이 4가지 영역에 세부 항목으로 드라마명, 장르, 관람등급, 총횟수, 주인공_남, 주인공_여, 조연1, 조연2, 방송사, 방송기간, 첫회시청률, 방영요일, 방영기간, 방송사, 연출가, 극본, 방영개월 수 등 16가지 정형데이터에서 획득될 수 있는 자료 내용목록과 게시판데이터, 방영전블로그, 방영 후 블로그 의 데이터관련 비정형 데이터 수집 목록을 정해서 데이터를 수집하였다.

정형데이터를 수집하기 위해서 대표방송 KBS, MBC, SBS 3사에서 2015년부터 2016년까지 방영되었던 드라마에 관련된 내용을 닐슨코리아와 TNMS에서 제공하는 드라마 정보를 하나하나 수작업을 통해 데이터 목록에 입력하여 의미 있는 정보화 작업을 수행하였다. 정형데이터 확보는 시간은 많이 걸렸지만, 정확하

고 신뢰성 있는 데이터를 확보하고 이를 분석해서 의미 있는 정보로 만들어 내는데는 그리 어렵지 않다는 장점을 가지고 있다. 그것이 바로 정형데이터의 장점이라 하겠다. 하지만 이런 정형데이터만으로는 좀 더 깊이 있는 정보를 획득하고 미래를 예측해 내기에는 부족한 부분이 있다. 따라서 대부분의 빅데이터 분석에서는 비정형화 된 데이터를 어떻게 수집하고 어떻게 분석해 내느냐가 가장 큰 해결해야 될 과제라 하겠다[6-8].

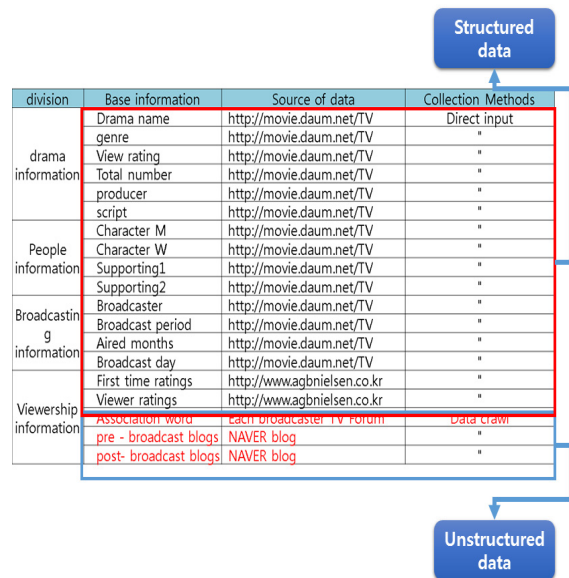


Fig. 1 Collection of drama information list

3.2. 비정형데이터 수집 방법

비정형 데이터 수집은 대표방송 KBS, MBC, SBS 3사에서 2015년부터 2016년까지 방영되고 있었던 드라마 홈페이지에서 운영되는 게시판의 시청자 의견 내용과 드라마가 시작하기 전 블로그의 내용 및 개수 그리고 드라마 시작 후 블로그의 내용 및 개수에 대해 수집하였다. 연관어 분석을 위해 첫 번째 Twitter의 데이터를 수집하였는데, 이는 지금 대부분의 사람들이 많이 사용하는 SNS 중 Facebook은 사용자 본인과 친구 관계에서 내용 검색이 가능하여 자료 수집의 제한이 있었기 때문이다. Twitter의 수집 방법으로는 Twitter에 별도 개발 계정을 등록하고 R package인 twitterR를 사용하여 Twitter검색 API(searchTwitter)를 통해 검색 및 데이터 내용을 저장하였다. 두 번째 NAVER 블로그 크롤링

방법으로 데이터를 수집하였다. NAVER 블로그 크롤링을 위한 주소 쿼리는 아래 그림 2와 같다.

```
http://cafeblog.search.naver.com/search.naver?where=post&query=드라마+2015-10-10&ie=utf8&st=sim&sm=tab_opt&date_from=http://www.imbc.com/broad/tv/drama/danger/board/index.html&date_to=NA&date_option=6&srchby=all&dup_remove=1&post_blogurl=&post_blogurl_without=&mson=0&start=1
```

Fig. 2 NAVER Address query for crawling blogs

NAVER 블로그 내용과 댓글 크롤링을 통해 단어 빈도 분석을 위한 정보 수집도 실시하여 데이터화 작업 후 저장하였다. 또한 각 드라마 방영전 1주일과 방영 후 1주일에 만들어진 블로그의 숫자도 데이터 작업 후 저장하였다. 셋째 각 방송사별 상영 드라마 게시판의 내용을 크롤링하는 방법을 사용하였다. 인코딩 문제에 따른 OS별 라이브러리를 사용하였다. htr패키지는 웹페이지 파싱을 통해 내용을 수집할 수 있는 라이브러리로 html()함수를 이용하여야 한다. 하지만 html()함수는 MAC, LINUX에서는 사용가능하나 Window에서는 작동이 되지 않는다는 문제점을 가지고 있다. 따라서 본 논문에서는 Window에 적합한 함수인 content(GET()) 함수를 사용하여 시청자 게시판의 내용을 크롤링하였다. 그림 3에서는 시청자 게시판 크롤링 쿼리를 보여주고 있다.

```
txt <-
content(GET(sprintf("http://imbbs.imbc.com/view.mbc?list_id=%s&page=1&bid=%s_board",list_id,drama_name)), "text")
html<-htmlTreeParse(txt,useInternal=TRUE,encoding="UTF-8")
list<-
xpathSApply(html,"//table//tr//td//table//tr//span",xmlValue)
```

Fig. 3 Address Query for viewers board Crawl

IV. 데이터 분석

4.1. 카이제곱검정

본 논문에서는 카이제곱검정을 사용하여 각 방송 3사별 드라마 정보 분류에 따른 데이터 결과 값이 평균 대비 어느 정도 유사도를 가지고 있는지 검증하는 방법으로 사용하였다.

카이제곱검정은 카이제곱 분포에 기초한 통계적 방법으로, 관찰된 빈도가 기대되는 빈도와 의미 있게 다른지의 여부를 검증하기 위해 사용되는 검증방법이다. 수집한 자료의 데이터가 주어졌을 때, 특히 명목척도 자료의 분석에 자주 이용되는 방법 중의 하나이다. 아래 식1과 같은 방법으로 계산한다. 검증 유형으로는 변인의 분포가 이항분포나 정규분포와 동일하다라는 가설을 설정하여 어떤 모집단의 표본이 그 모집단을 대표하고 있는지를 검증하는 데 사용하는 동질성 검증과 변인이 두 개 이상일 때 사용되는 기대빈도가 두 변인이 서로 상관이 없고 독립적이라고 기대하고 관찰빈도와 의 차이를 통해 기대빈도의 진위여부를 밝히는 독립성 검증이 있다. 본 논문에서는 동질성 검증 방법을 사용하여 비교분석을 실시하였다.

$$X^2 = \sum \frac{(Observation\ value - Expected\ value)^2}{Expected\ value} \quad (1)$$

본 논문에서는 각 방송시간 특성을 비교하기 위해 수집된 데이터로부터 그림4와 같이 방송사별 드라마 방영 시간대, 방영시작시기, 장르, 방영요일이라는 4개의 중분류를 두고 각 항목에 아침, 저녁, 심야, 봄, 여름, 가을, 겨울, 멜로, 가정, 사극, 기타, 일일, 월화, 수목, 주말이라는 소분류로 나누어 데이터를 정렬해 보았다. 이렇게 정렬된 데이터 값을 가지고 3개 방송사 간 특성을 카이제곱검정 방법으로 비교 분석한 결과 그림5와 같이 서로 유사한 결과값을 나타냄을 확인할 수 있었다.

Category	totality	Broadcaster			Chi-squared	
		KBS	MBC	SBS		
totality	169	57	53	59		
	100.0%	33.7%	31.4%	34.9%		
Broadcasting time zone	morning	9.5%	10.5%	9.4%	8.5%	0.843
	evening	21.9%	22.8%	24.5%	18.6%	
	Late night	68.6%	66.7%	66.0%	72.9%	
Broadcast start time	spring	25.4%	19.3%	26.4%	30.5%	4.282
	summer	23.7%	24.6%	24.5%	22.0%	
	autumn	29.6%	28.1%	34.0%	27.1%	
	winter	21.3%	28.1%	15.1%	20.3%	
genre	mellow	39.1%	40.4%	35.8%	40.7%	4.856
	Family	24.3%	26.3%	26.4%	20.3%	
	Historical drama	11.8%	14.0%	15.1%	6.8%	
	Etc	24.9%	19.3%	22.6%	32.2%	
Broadcast day	Daily	19.5%	22.8%	20.8%	15.3%	7.179
	Mon-Tuesday	24.9%	31.6%	17.0%	25.4%	
	Wed Thursday	29.0%	29.8%	30.2%	27.1%	
	Weekend	26.6%	15.8%	32.1%	32.2%	

Fig. 4 Broadcasters comparison between properties

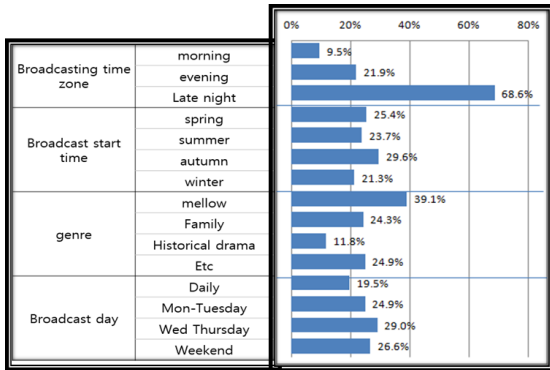


Fig. 5 Analysis using chi-square

4.2. 드라마 시청률과 블로그 관계 분석

본 논문에서는 드라마 시청률에 영향을 미치는 요소가 무엇인지 알아보기 위해 관련 연관어를 분석하는 과정에서 시청률과 블로그와의 관계 분석을 시행하였다. 이는 블로그에 나오는 관련 단어를 조사하는 과정에서 시청률이 좋은 드라마의 블로그의 개수 증가가 어떻게 변화되는지가 드라마의 승패를 예측할 수 있는 요소가 되는 것을 알게 되었다. 수집된 데이터로부터 첫 회 시청률, 드라마 시청률, 방영 전 1주일 블로그 수, 첫 방영 후 1주일간 블로그 수, 방영개월 수로 구분하여 각 방송사별로 평균과 표준편차를 구해 알아보았다. 그림 6에서 보는 것과 같이 시청률은 첫 회 시청률 보다 3사 모두 약 5% 상승하였고, 블로그 수는 방영전 블로그 수의 약 2배가 되는 것으로 나타남을 알 수 있다.

Item	Broadcaster	N	Average	Standard Deviation	F
First time Viewership	KBS	57	10.9	5.2	4.45*
	MBC	53	10.3	2.8	
	SBS	59	9.0	2.1	
	total	169	10.1	3.7	
Drama Viewership	KBS	57	15.8	10.6	1.99
	MBC	53	15.3	7.7	
	SBS	59	12.9	5.8	
	total	169	14.6	8.3	
Before broadcasting One week Blog number	KBS	57	230	185	2.58
	MBC	53	167	125	
	SBS	59	239	214	
	total	169	213	182	
After the first airing One week Blog number	KBS	57	414	340	4.34*
	MBC	53	330	260	
	SBS	59	552	540	
	total	169	436	411	
Aired months	KBS	57	3.8	2.3	2.75
	MBC	53	4.2	1.9	
	SBS	59	3.3	1.7	
	total	169	3.8	2.0	

Fig. 6 First time viewership and blog relation

이는 드라마의 시청률에 블로그의 내용이 영향을 미치며, 이에 따라 시청률 변화에 따라서 블로그의 수가 달라짐을 알 수 있다.

3개 방송사 모두 방송전 블로그를 가지고 있는 드라마의 시청률이 방송 후 시청률 보다 대부분 상승하였으며, 시청률이 상승하면서 블로그의 개수와 댓글의 개수도 증가하고 있는 것을 확인할 수 있었다. 이는 SNS의 영향력이 얼마나 그 드라마의 시청률과 연관이 있는지를 알려주는 결과라 할 수 있다. 그림7에서 보는 것과 같이 KBS, MBC, SBS 방송사의 드라마 시청률과 블로그와의 관계를 나타내 주고 있다. 각각의 방송사별 시청률의 %는 값은 다르지만, 그래프의 기울기 차이를 보면 거의 유사한 형태로 변화되고 있는 것을 알 수 있다.

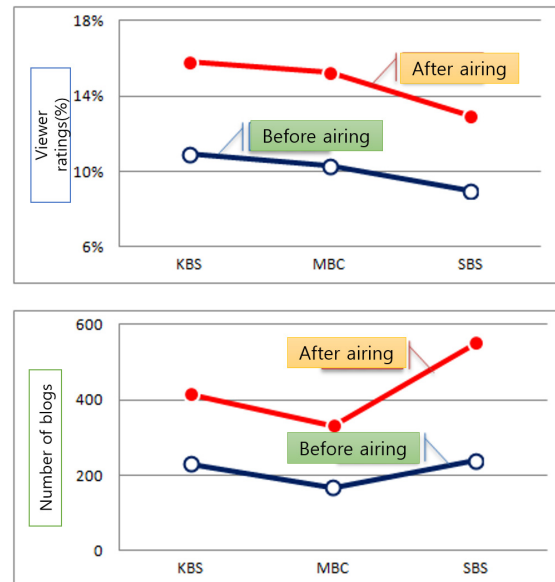


Fig. 7 Relationship between drama viewership and blog number

4.3. 비정형데이터 텍스트 마이닝

방송사마다 드라마 게시판의 글과 SNS의 대표인 Twitter의 글, 해당 드라마의 블로그의 내용과 댓글 등으로부터 수집된 비정형데이터로부터 의미 있는 정보를 발견하기 위해 본 논문에서는 텍스트 마이닝 기법을 사용하였다. 텍스트 마이닝(Text Mining)은 빅데이터 분석에서 텍스트 분석과 비슷하게 일컬어지는데, 인간

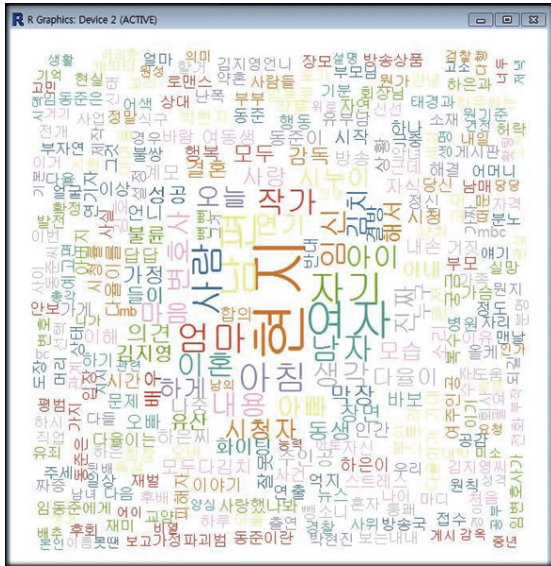


Fig. 8 Word cloud result image

의 언어로 쓰인 비정형 텍스트에서 자연어처리 기술을 이용하여 유용한 정보를 추출하거나, 연계성 파악, 분류 또는 군집화, 요약 등 빅데이터 안에 숨겨져 있는 의미 있는 정보를 발견할 수 있도록 도와주는 도구라 할 수 있다[9,10].

KBS, MBC, SBS 3사에서 2015년부터 2016년까지 방영되었던 드라마에 관련된 내용을 포함하고 있는 비정형데이터 수집 자료로부터 R을 이용하여 텍스트 마이닝을 실시한 후 문서의 키워드, 개념등을 직관적으로 파악할 수 있도록 핵심 단어를 시각적으로 돋보이게 하

는 워드 클라우드를 실시하여 그림 8과 같이 이미지화 된 결과를 획득하였다. 그림 8에서 보이는 것과 같이 엄마, 여자, 작가, 이혼, 사랑, 아침, 내용 등과 같은 단어가 많이 언급되어 있음을 알 수 있다.

또한 각 드라마의 어휘 출현빈도를 분석하고 모든 드라마에서 공통으로 출현되는 어휘만을 추출하여 어휘 출현빈도와 시청률과의 상관관계를 분석한 결과 그림 9와 같은 결과를 도출할 수 있었다. 연관어를 도출하기 위해 연관어들의 군집화와 요인화 작업을 실시하여 결과를 도출하였으며, 그로 인해 나온 결과 값으로는 방송, 시청자, 내용, 아이, 부탁, 여자, 사랑 이라는 총 7개의 연관어가 드라마의 시청률과 관련 있는 단어라는 결과를 얻을 수 있었다. 대부분의 시청자들이 좋아하는 그리고 관심 있는 드라마 관련 단어들은 3개의 방송사에서 방영되는 드라마 속에 7개의 단어가 공통적으로 가장 많이 들어가 있다라는 결론이다.

V. 결론

본 논문에서는 드라마의 시청률에 영향을 미치는 연관어를 분석하기 위해 정형화된 데이터와 비정형화된 데이터를 수집하고 그 수집된 데이터를 분석하고자 하였다. 데이터 수집은 KBS, MBC, SBS 3사에서 2015년부터 2016년까지 방영되었던 드라마에 관련된 내용으로 정형화된 데이터 수집은 각 방송사의 드라마정보, 인물정보, 방송정보, 시청률정보라는 4가지 영역에서 총 19가지항목을 수집하였으며, 비정형데이터는 각 방

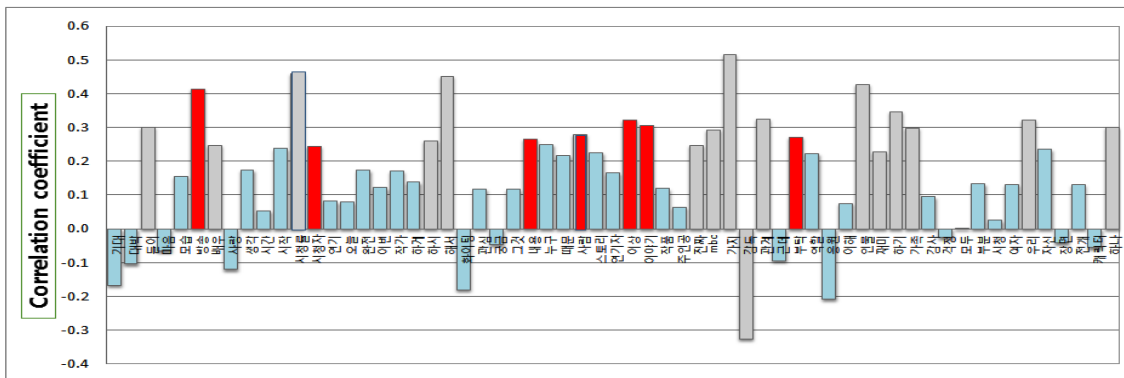


Fig. 9 Result of drama audience rating association

송사에서 드라마별로 운영되고 있는 게시판과 방영전 블로그와 방영후 블로그, 해당 드라마 Twitter의 댓글로부터 크롤링기법을 이용하여 수집하였다. 수집된 정형 데이터로부터 각 방송사별 4가지영역별에 따른 차이를 비교한 결과 방송사별 서로 유사한 결과 값을 보이고 있었다. 각 방송사의 드라마별 게시판과 블로그에서 수집된 비정형데이터로부터 출현빈도와 시청률과의 상관관계 분석을 통해 관련 연관어를 최종 7개 도출하였다. 앞으로 이렇게 도출된 7개의 연관어는 드라마 시청률을 예측하기 위한 감성분석 요소로 사용할 수 있을 것으로 보이며, 이를 활용하여 좀 더 정확한 시청률 예측 분석이 가능할 것으로 기대된다.

ACKNOWLEDGMENTS

This paper was supported by Ministry of SMEs and Startups. [Innovative Technology Development Project]

REFERENCES

- [1] S. K Kang, H. C Lee, and S. Y Shin, "Analysis of related words of drama viewership through SNS unstructured data crawling," in *Proceedings of the 41th Annual Conference of Korea Institute of information and Communication*, pp. 105, June 2017.
- [2] S. H Yun, K. H Lee, H. S Lim, D.R Kim, and J. H Kim, "The Method of Digital Copyright Authentication for Contents of Collective Intelligence," *Journal of the Korea Convergence Society*, vol. 6, no. 6, pp. 185-193, June 2015.
- [3] Y. W. No, D. Y. Kim, J. E. Han, J. T. Lim, K. S. Boek, and J. S. Yoo, "Hot Topic Prediction Scheme Considering User Influences in Social Networks", *Journal of Korea Contents Association Research*, vol.15, no. 8, pp24-36, Dec. 2015.
- [4] J. B. Lee, C. K. Lee, and K. J. Cha, "An Analysis of IT Trends Using Tweet Data", *Journal of Intelligence and Information Systems*, vol. 21, no. 1, pp143-159, Sep. 2015.
- [5] Y. J Lee, J. H Seo, and J. T Choi, "Fashion Trend Marketing Prediction Analysis Based on Opinion Mining Applying SNS Text Contents," *Journal of KIIT*. vol. 12, no. 12, pp. 163-170, June 2014.
- [6] H. J Kim and J. Y Chang, "Discovering News Keyword Associations Using Association Rule Mining," *The Journal of the Institute of Internet Broadcasting and Communication*, vol. 11, no. 6, pp. 63-71, Dec. 2011.
- [7] B. W. Kim, "Trend Analysis and National Policy for Artificial Intelligence." *Informatization Policy*, vol. 23, no. 1, pp. 74-93, Mar. 2016.
- [8] J. Y. Seo and C. Koh, "Big Data Analysis by Sensitivity Analysis," *Journal of the Society of Convergence Knowledge*, vol. 2, no. 1, pp. 15-21, June 2014.
- [9] J. S. Sohn, S. W. Cho, K. L. Kwon, and I. J. Chung, "Improved Social Network Analysis Method in SNS," *Journal of Intelligence and Information Systems*, vol. 18, no. 4, pp. 117-127, Mar. 2012.
- [10] S. N. Kang, Y. S. Kim, S. H. Choi, "Study on the social issue sentiment classification using text mining." *Journal of the Korean data & information science society*, vol. 26, no. 5, pp. 1167-1173, June 2015.
- [11] J. Y. Jang, "Automatic Retrieval of SNS Opinion Document Using Machine Learning Technique.," *The Journal of the Institute of Internet, Broadcasting and Communication*, vol. 13, no. 5, pp. 27-35, Sep. 2013.



강선경(Sun Kyoung Kang)

2010년 원광대학교 컴퓨터공학과 박사
 2016년 ㈜중은정보기술 연구소장
 2015년~현재 IITP 기술사업화 전문위원
 2017년~현재 한국정보통신학회 국제학술이사
 2017년~현재 원광대학교 컴퓨터소프트웨어공학과 조교수
 ※관심분야 : 멀티미디어 시스템 및 응용, 빅데이터, 영상처리, 가상현실



이현창(Hyun-Chang Leeg)

홍익대학교 컴퓨터학과(공학박사)
원광대학교 디지털콘텐츠공학과 교수
한국정보통신학회 국제학술 이사
※관심분야 : 시멘틱웹, 영상처리, 유비쿼터스 컴퓨팅



신성윤(Seong-Yoon Shin)

군산대학교 컴퓨터정보공학과 박사
한국정보통신학회 국문지부회장
한국정보통신학회 국제학술부회장
군산대학교 컴퓨터정보통신공학부 교수
※관심분야 : 멀티미디어 시스템 및 응용, 가상현실, 텔레메틱스