

토픽 모델링을 이용한 방송미디어 관련 소셜 미디어 콘텐츠 분석*

박 상 언**

Analysis of Social Media Contents about Broadcast Media through Topic Modeling*

Sangun Park**

■ Abstract ■

Numerous people share their TV experience with other viewers on social media such as personal blogs and Twitter. It means that broadcast media, especially TV, affects the responses on social media. Moreover, the responses affect broadcast media ratings back. Social TV tried to use the relationship in marketing activities such as advertisement by analyzing the TV related social behavior. However, most of them used just the quantities of social media responses. This study analyzes the subjects of the responses on social media about specific TV dramas through topic modeling, and the relationship between the changes of popular topics and viewer ratings of the drama over specified periods. Five representative Korean dramas of 2014 were selected and Blog contents including viewer ratings about the dramas were collected from naver.com which is the representative portal in South Korea. The proposed analysis framework consists of three steps which are Blogs crawling, topic modeling, and topic trend analysis. We found some implications from the results of the topic trend analysis. Firstly, there were specific topics on dramas in social media. Secondly, the topics had some meaningful relationships with viewer ratings. Lastly, there were differences between the topics of dramas with higher viewer ratings and those with lower viewer ratings.

Keyword : Social Media, Broadcast Media, Topic Modeling, Topic Trends

1. 서론

소셜 미디어는 최근 가장 인기 있는 연구주제 중 하나이다. 소셜 미디어에서의 반응을 이용하여 선거 결과를 예측(Kushin and Yamamoto, 2010)하거나 마케팅 캠페인의 수행에 활용하는 등(Michaelidou et al., 2011), 소셜 미디어에 대한 분석은 다양한 분야에서 활용되고 있다. 본 연구에서는 방송미디어의 시청률과 소셜 미디어에서의 반응 사이의 관계를 살펴보고자 한다. 사람들은 TV를 통해 다양한 프로그램을 접하고 그에 대한 경험이나 의견을 소셜 미디어를 통해 다른 사람들과 공유하게 된다. 이러한 소셜 미디어에서의 반응은 다시 TV 프로그램의 시청률에 영향을 미칠 수 있다. 이와 같이 방송 미디어와 소셜 미디어 사이의 연관성에 주목한 연구는 이전부터 수행되어 왔다. 예를 들어 소셜 TV는 TV와 관련된 소셜 행태에 대한 연구로서, 주로 TV를 보면서 시청자들 간에 의견을 교환할 수 있는 새로운 형태의 애플리케이션을 제공하거나(Cesar and Geerts, 2011a), 특정 TV 프로그램에 대해 소셜 미디어에서의 반응을 측정하였다(Cesar and Geerts, 2011b). 그러한 예들 중 하나로 Bluefin(Deb, 2012)은 사용자들의 반응을 두 개의 척도로 측정하여 TV 프로그램에 대한 시청자들의 참여에 대해 새로운 통찰력을 제공하고자 하였다. 소셜 TV는 한 때 많은 주목을 받았으나, 2014년에는 대부분의 소셜 TV 사업들이 실패한 것으로 나타났다(Roettgers, 2014). 이들 사업들은 주로 시청자들이 TV를 보면서 의견을 교환할 수 있는 새로운 형태의 SNS를 제공하고자 했는데, 주요 실패원인은 이미 SNS 시장에 페이스북과 트위터라는 공룡들이 자리 잡고 있었기 때문이다(Roettgers, 2014).

그러나 이러한 소셜 TV 서비스가 실패했다고 해서, 소셜 미디어로부터 얻을 수 있는 통찰력이 필요 없다는 것을 의미하지는 않는다. BBC 방송의 경우 시청정보를 분석하여 시청자 맞춤형 방송을 추천하고 있으며(Brown, 2013), 이러한 추천에서

소셜 미디어 기반의 정보는 유용하게 사용될 수 있다. 또한 넷플릭스(Netflix)가 빅데이터 분석을 기반으로 '하우스 오브 카드'라는 드라마를 제작한 것은 잘 알려진 사실이다. 이 드라마는 시청자 방송 콘텐츠 수요 및 선호도 분석과 소셜 데이터 분석 결과를 토대로 방송 콘텐츠 성공을 예측하였다. 빅데이터 분석을 통해 시청자들이 원하는 주연배우, 감독, 주제를 선정하여 제작하였는데, 2013년 1/4 분기에만 2백만 명의 신규 시청자를 확보하는데 성공하였다. 즉, 소셜 미디어를 이용한 시청자 반응 분석은 드라마와 같은 방송 콘텐츠의 제작에 중요한 정보들을 제공한다고 할 수 있다. 한국 문화관광연구원에 따르면 2014년 인기드라마였던 "별에서 온 그대"의 국내 생산유발효과를 101,105,200만 원으로 추산하고 있다(Jung, 2014). 따라서 우리나라에서도 드라마의 기획은 매우 중요한 산업이며, 방송 미디어와 소셜 미디어 간의 관계를 분석함으로써 드라마의 완성도를 높일 수 있다면, 충분히 시도해 볼 만한 일이라 할 수 있다(Park et al., 2015).

과거 소셜 TV를 비롯한 연구들에서 방송 미디어가 소셜 미디어에 미친 영향은 주로 검색어 순위 혹은 검색량과 같은 정량적 지표가 대상이었다. 그러나, 본 연구에서는 그와 같은 정량적인 지표보다 소셜 미디어에서 방송 콘텐츠에 대해 기술되는 토픽과 같이 의미적이고 정성적인 부분에 대한 분석 방법을 제안하고자 한다. 토픽 모델링(Blei et al., 2010)과 같이 잘 알려진 텍스트 마이닝 기법을 이용하여 드라마와 같은 방송 콘텐츠에 대해 시청자들이 소셜 미디어에서 주로 기술하는 토픽들을 분석한 후, 그 토픽들과 드라마의 시청률과의 관계를 분석함으로써 어떠한 토픽들이 시청률에 영향을 미치고, 또 반대로 시청률이 토픽들에 어떤 영향을 미치는지 둘 간의 관계를 알아보고자 한다. 이를 위해 먼저 드라마가 방영된 기간 동안의 소셜 미디어 내용을 수집하고, 토픽을 분석한 후에, 정해진 기간 동안의 토픽 트렌드를 분석함으로써 각 토픽들이 시청률과 함께 어떻게 변화했는지를 살펴보고자 한다. 이와 같이 토픽 트렌드와 시청률의 변화를 봄으로

써 각각의 토픽과 시청률의 관계를 직관적으로 파악해낼 수 있을 것으로 기대된다.

본 논문의 구성은 다음과 같다. 제 2장에서는 토픽 모델링에 대한 기존 연구를 알아보고자 한다. 제 3장에서는 본 연구에서 제시하는 방송 미디어와 소셜 미디어 간의 관계를 분석하기 위한 방법론에 대해 설명하고, 제 4장에서는 분석 결과와 시사점에 대해 기술하고자 한다. 마지막으로 제 5장에서는 결론과 향후 연구에 대해 설명하고자 한다.

2. 토픽 모델링과 토픽 트렌드

토픽 모델링은 텍스트 마이닝 분야에서 유용하게 사용되는 도구이다(Blei and Lafferty, 2009). 주 사용목적은 매우 큰 다큐먼트의 집합으로부터 유용한 패턴을 찾아내는데 있다. 이 패턴은 워드들로 구성된 벡터로서 토픽이라고 불린다. 즉 토픽은 문서들에 내재한 단어들의 확률분포이며, 연관성이 높은 단어들로 이루어진다. 토픽 모델링은 이러한 토픽들의 통계적인 집합을 문서들로부터 추출하기 위한 방법론이라고 할 수 있다. 본 연구에서는 토픽 모델링에서 가장 잘 알려진 알고리즘 중 하나인 LDA (Latent Dirichlet Allocation) (Blei et al., 2003)를 이용하였다. LDA에서 각 문서들은 토픽들의 특정한 집합으로 간주된다. 이 때 유일하게 우리가 관측할 수 있는 변수들은 문서 내에 있는 특정 단어들이다. LDA는 토픽 분포와 토픽을 구성하는 단어들의 분포에 잠재된 변수들을 추출하기 위해, 디리클레 다항분포에 기반한 통계적 추론방법을 사용한다.

토픽 트렌드는 시간에 따라 변화하는 토픽들의 추세를 제공한다. Matsubara et al.(2012)은 LDA가 직접적으로 토픽 트렌드를 제공하지는 않는 것으로 지적하고 있으나, 본 연구에서는 전체 기간에 대해 LDA를 수행하고 그 결과를 이용해 각 날짜에 따른 토픽 값을 구한 후 그 추세를 보는 알고리즘을 이용하였다.

토픽 모델링과 토픽 트렌드는 다양한 분야의 연구에서 활용되고 있다. Hyun et al.(2015)은 토픽

에서 발견된 이슈에 대해 클러스터링을 적용하여 상위 개념의 이슈를 도출하기 위한 이슈 클러스터링 방법을 제안하고 소비자 선호 이슈 관점에서 적용하고자 하였다. Bae et al.(2014)은 대표적인 소셜 네트워크 서비스인 트위터를 대상으로 하여 이슈 트래킹 시스템을 제안하였다. Bolelli et al.(2009)은 디지털 도서관의 문서들로부터 다양한 토픽들을 추출하는데 토픽 모델링 기법을 이용하였다. 이 연구에서는 문서들을 일정 시간 간격에 따라 분류하고 각각의 세그먼트에 대해 토픽 모델링을 수행한 후에, 시간에 따라 토픽들이 어떻게 진화해왔는지를 보이고자 했다. Martie et al.(2012)은 안드로이드 버그 리포트에 토픽 트렌드를 적용하고자 했다. 안드로이드의 오픈소스 프로젝트들을 분석하기 위해 버그에 대한 토론들이 올라온 공개 이슈들에 대해 토픽 분석을 실시함으로써 버그와 관련된 토픽들이 어떻게 변화하는지 살펴보고 이를 개발과정에 반영하고자 했다. 기존 연구들에 비해 본 연구는 소셜 미디어를 대상으로 하여 토픽을 들여다보는 것에 그치지 않고, 분석된 토픽들과 실제 방송 지표인 시청률과의 연관성을 보고자 했다는 점에서 차이가 있다.

3. 토픽 트렌드 분석을 이용한 방송 미디어와 소셜 미디어 분석방법

3.1 방송미디어 분석의 중요성

한국 문화관광연구원은 드라마 “별에서 온 그대”의 경제적 파급효과에 대한 분석을 제공하고 있다 (Jung, 2014). 분석에 따르면 2014년 7월 당시 수출 협상중인 국가의 예상판매액의 총액은 213,100만 원으로 잡고 있으며, 중국에 판매된 금액인 119,200만 원을 합산하면 332,400만 원 정도의 수출이 이루어질 것으로 추산하고 있다. 여기에 드라마로 인한 광고 총액은 685,700만 원 정도, 관련 상품의 매출은 5,495,500만 원으로, 관광 수입은 46,516,400만 원으로 추산하고 있다. 결론적으로 국내생산유발효과를

〈Table 1〉 Five Korean Dramas Selected

Title	Period	Number of episodes	Broad-casting days	Viewer Rating (Avg)
Wang Family	2013. 08. 31~2014. 02. 16	50	Weekent	33%
Empress Qi	2013. 10. 28~2014. 04. 29	51	Mon, Tues	21.9%
Dr. Stranger	2013. 05. 05~2014. 07. 08	20	Mon, Tues	11.7%
My Love from the Star	2013. 12. 18~2014. 02. 27	21	Wed, Thurs	24%
Misaeng(Incomplete Life)	2014. 10. 17~2014. 12. 20	20	Fri, Sat	5.4%

101,105,200만 원으로 추산하고 있는데 어느 정도 과장될 수 있다는 점을 감안하더라도 한 편의 드라마가 매우 높은 경제적 파급효과를 갖고 있다는 것을 알 수 있다.

넷플릭스(Netflix)가 제작한 “House of Cards”에 대한 분석과 배경을 제공하는 뉴욕타임즈의 기사에서는 “Giving Viewers What They Want”라는 제목에서 알 수 있듯이, 드라마를 제작하는데 있어, 빅데이터 분석을 이용해 시청자들의 취향과 요구사항을 분석하는 것이 매우 일반적인 경향이 되었음을 보여준다(Carr, 2013).

이상의 내용으로 볼 때 한 편의 드라마를 만드는 것은 매우 중요한 산업임을 알 수 있으며, 시청률과 소셜 미디어의 분석으로부터 시청자들의 관심이 어떻게 변화하고 어떻게 시청률에 영향을 미치는 지 알아낼 수 있다면 충분히 시도해볼 만한 일이라 할 수 있다.

3.2 드라마와 소셜 미디어의 선택

본 연구에서는 2014년도에 방영된 드라마들을 대상으로 하여 그 중 인기가 높았던 드라마들을 선정하였다. 이를 위해 뉴스핌에서 제공하는 2014년 지상파 드라마 주말, 일일, 수목극 시청률 순위(Lee, 2014)를 이용하여 각 부문별 시청률 1위 드라마와, 종합편성 채널에서의 시청률 1위 드라마를 <Table 1>과 같이 선정하였다. 또한 SBS의 드라마 중에서 기대는 높았으나 실제 시청률은 2위에 그친 드라마 -“닥터이방인”을 추가로 선정하였다. <Table 1>에서 보는 바와 같이 한국의 드

라마는 일일드라마를 제외하고 일주일에 방영되는 요일이 일정간격이 아니라는 단점이 있다. 예를 들어 수목드라마는 수요일과 목요일에 연달아서 방영되지만, 나머지 요일에는 방영되지 않는다. 이로 인해 소셜 미디어의 반응을 어느 요일을 기준으로 수집해야 하는 지에 대한 이슈가 발생한다.

다음 단계에서는 분석대상 소셜 미디어 서비스를 선택하였다. DMC 미디어의 “소셜 미디어 이용실태”에 따르면(Dmc_Marketing_Team, 2014), 2014년 한국에서 가장 많이 이용한 소셜 미디어는 페이스북으로 77.4%의 이용률을 보였다. 그러나 페이스북의 경우, 전체 사용자들이 게시하는 내용을 대상으로 검색하여 결과를 가져오기 위한 API나 방법이 제공되지 않고 있다. 두 번째로 가장 많이 사용된 소셜 미디어는 블로그로 29.8%의 이용률을 보였다. 본 연구에서는 비교적 크롤링이 용이한 블로그를 대상으로 선택하였으며, 그 중에서도 한국 블로그 마켓에서 2015년 5월을 기준으로 71.2%의 점유율을 보인 네이버 블로그를 선택하였다.¹⁾

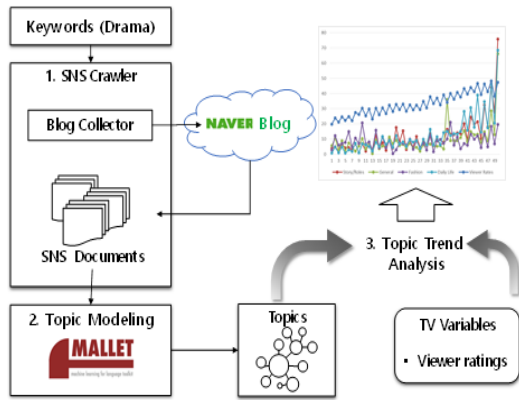
3.3 토픽 트렌드 분석 프레임워크

본 연구에서 제안하는 분석 프레임워크는 <Figure 1>과 같다. 첫 단계는 네이버 블로그로부터 대상 드라마와 관련된 블로그 글들을 수집하는 단계이다. 기간은 드라마가 방영된 날짜를 기준으로 하였으며, 향후 연구 보정이나 보완을 위해 드라마 방영기간 동안 방영되지 않은 날에 대해서도 날짜

1) 출처 : <http://www.blogchart.co.kr/>.

별로 문서들을 수집하였다. 둘째 단계는 문서들을 대상으로 토픽 모델링을 수행하는 단계이다. LDA 분석에 있어서 많이 알려지고 또 사용되는 말렛(Mallet)을 이용하였으며, 그 결과로는 전체 기간의 토픽들의 분포와 각 문서 별로 토픽들의 분포를 얻을 수 있다.

셋째 단계는 드라마가 방영된 날짜에 대비하여 각 토픽들의 날짜별 트렌드를 구하는 단계이다. 이 단계에서는 드라마의 시청률과 각 토픽의 트렌드를 비교하여 시각적으로 상호 연관관계를 확인해 볼 수 있다.



<Figure 1> Analysis Framework

4. 토픽 트렌드 분석 결과

4.1 블로그 문서 수집결과

네이버 블로그를 대상으로 한 드라마 관련 문서 수집에서는 자바로 직접 크롤링 프로그램을 작성하여 수행하였다. 블로그 검색을 위한 키워드로 드라마의 제목과 등장인물의 연기자 이름과 등장인물의 이름 등 다양한 방법을 시도하여 검사한 결과, 드라마의 제목만을 키워드로 이용하였다. 등장인물의 연기자 이름과 같이 다른 키워드를 사용한 경우에는, 드라마와 연관성이 있는 문서는 대부분 드라마 제목으로 한 결과와 중복되고, 중복되지 않은 경우에는 드라마와 연관성이 없는 결과가 많았

다. 따라서 문서의 연관성을 높이기 위해 제목만을 사용하기로 결정하였다. 블로그 문서의 본문을 가져온 후, 형태소 분석기인 KLT2000(KLT, 2014)을 이용하여 명사만을 추출하였다. 이 과정에서 먼저 필요 없는 단어는 불용어로 한 번 걸러지고, 뒤에 토픽 모델링 수행 단계에서 다시 걸러진다. 토픽 트렌드 분석과정에서 사용할 시청률은 AGB닐슨미디어리서치²⁾에서 제공하는 시청률을 바탕으로 네이버에서 각 드라마에 대해 편집하여 제공하는 기간별 시청률을 사용하였다.

<Table 2>는 수집된 문서들의 결과를 보여준다. 표에는 각 드라마 별로 수집된 블로그 문서의 수와 날짜 별 평균 문서 수가 나타나 있다. 주목할 점은, ‘왕가네 식구들(Wang Family)’의 경우, 시청률이 가장 높음에도 불구하고 일당 평균 문서 수는 가장 낮다는 점이다. 이는 주말드라마인 ‘왕가네 식구들’의 특성 상, 시청자들의 분포가 40~50대 주부와 같이 SNS를 많이 사용하지 않는 특정 층에 집중되었기 때문인 것으로 판단된다. 반면, ‘미생(Misaeng)’의 경우에는 낮은 시청률에도 불구하고 일당 평균 문서 수는 ‘별에서 온 그대(My Love from the Star)’에 이어 두 번째로 높은 526건을 기록하였다. 이러한 결과는 후에 시청률과 토픽 간의 관계를 분석한 결과와도 연관이 있을 것으로 생각된다.

<Table 2> Crawling Results

Title	Number of Documents	Avg. No of Documents per Day	Viewer Rating (Avg.)
Wang Family	4,818	64.2	33%
Empress Qi	10,130	131.6	21.9%
Dr. Stranger	5,249	175.0	11.7%
My Love from the Star	43,913	1372.3	24%
Misaeng (Incomplete Life)	15,787	526.2	5.4%

2) 출처 : <http://www.agbnielsen.co.kr/>.

4.2 드라마 관련 소셜 미디어 토픽 분석 결과

수집된 문서들에 대한 토픽 분석은 MALLET (Mimno, 2013)에서 제공하는 LDA(Latent Dirichlet Allocation)(Blei, 2003) 기반의 토픽 모델링 도구를 사용하여 수행하였다. 토픽 모델링은 문서가 다양한 토픽들의 집합체이고 각각의 토픽은 단어들의 확률적인 분포로 이루어졌다는 가정 아래, 주어진 문서들로부터 각 문서들의 토픽의 분포를 찾아내고자 하는 방법론이라고 할 수 있다.

즉, 이 단계에서 수행하는 토픽 모델링은 각 드라마가 방영된 기간 동안에 쓰여진 블로그 문서들이 어떠한 토픽들을 담고 있는 것인가를 알아내고자 하는 것이다. 특히 다음 단계에서 그러한 토픽들의 시간에 따른 트렌드를 보고자 하기 때문에, 한번 나타났다가 사라진 토픽보다는 가급적 방영 기간 동안 지속된 토픽들을 찾아내고자 하였다. 토픽 모델링에서 중요한 이슈 중 하나는 토픽 수의 결정이다. 토픽의 수가 지나치게 많은 경우 수많은 각각의 토픽들을 해석하기가 쉬지 않고, 토픽의 수가 적은 경우에는 토픽들이 지나치게 일반화되어 의미 있는 토픽을 찾기가 어렵기 때문이다. 따라서 본 연구에서는 적절한 토픽의 수를 결정하기 위해 5개부터 시작하여 20개까지 다양한 수의 토픽으로 분석을 실시하였다. 그 결과, MALLET으로 10개의 토픽을 만드는 분석을 실시하고, 그 중에서 다시 가중치(weight)가 0.1보다 높은 토픽들을 선정하는 방식으로 토픽 모델링을 실시하였다. MALLET은 두 종류의 결과를 제공하는데, 첫째로 전체 문서들에 대해 종합적으로 추출한 토픽(단어들의 집합)과 각 토픽의 가중치를 제공하고, 둘째로 각 문서들에 대해서 앞에 추출된 토픽들의 문서 단위의 가중치를 제공한다. 이 단계에서는 첫째 결과를 이용하여 토픽들을 선정하였으며, <Table 3>은 ‘별에서 온 그대’에 대한 결과를 토픽 모델링 결과를 보여주고 있다. 분석에서 제공되는 토픽은 단어들의 집합이기 때문에 토픽의 제목은 분석자의 주관적인 판단에 의해

많이 결정되는데, <Table 3>의 경우는 제목이 비교적 단어들을 잘 반영하고 있는 것으로 판단된다.

<Table 3> Example of Topics

Topic Title	Weight	Topic
Food	0.13184	Food, price, cheese, strawberry, pizza, ice coffee, sweet potato, cafe...
Fashion	0.12519	Overcoat, price, style, sunglasses, dress, jacket, fashion, skirt, shirt, ...
Story, Roles	0.14532	Songee, Dominjun, Semi, Jaekyong, alien, star, epilogue, kiss, supernatural power, ...
Drama-General talk	0.18174	Drama, viewer rating, casting, review, actress, The heirs, The thieves, acting ability...
Daily Life	0.20568	You, today, mobile, thought, photo, love, We, human, sister, time, one day, friend, now...

예를 들어 ‘음식(Food)’과 ‘패션(Fashion)’의 경우에 각기 직접적인 연관이 있는 단어들로 구성이 되었으며, ‘줄거리, 역할(Story, Roles)’의 경우에는 실제 주인공들의 극중 이름인 Songee, Dominjun, Semi, Jaekyong 외에, 줄거리와 연관이 높은 단어들로 이루어져 있다. 그 외에 주목할 토픽은 ‘드라마에 대한 일반적인 내용(Drama-General talk)’으로 드라마의 구체적 내용과는 관계없이 일반적인 드라마에 대한 내용을 담고 있다. 마지막으로 ‘일상생활(Daily Life)’에 대한 토픽은 드라마와 관계없이 일상적인 단어들을 담고 있으며, 이는 블로그가 일기의 성격을 갖고 있기 때문인 것으로 생각된다.

<Table 4>는 전체 드라마에 공통적으로 발견된 토픽들을 정리한 표이다. 가중치의 합이 높고 5개의 드라마에 공통으로 나타난 토픽들은 ‘일상생활(Daily life)’, ‘일반적인 드라마 관련(Drama-General talk)’, ‘줄거리, 역할(Story, Roles)’ 순이었다.

<Table 4> Common Topics over Dramas

Topics	Drama Title	# of Dramas	Weight Sum
Daily Life	Wang Family, Empress Qi, Dr. Stranger, My Love from the Star, misaeng	5	1.07
Drama-General talk	Wang Family, Empress Qi, Dr. Stranger, My Love from the Star, misaeng	5	1.04
Drama-story, Roles	Wang Family, Empress Qi, Dr. Stranger, My Love from the Star, misaeng	5	0.82
Food	Empress Qi, Dr. Stranger, My Love from the Star, Misaeng	4	0.50
Background, material (ex. History, alien)	Empress Qi, My Love from the Star	2	0.42
Fashion	Wang Family, Dr. Stranger, My Love from the Star	3	0.37
Philosophy	Misaeng	1	0.29

4.3 드라마 관련 소셜 미디어 토픽 트렌드 분석 결과

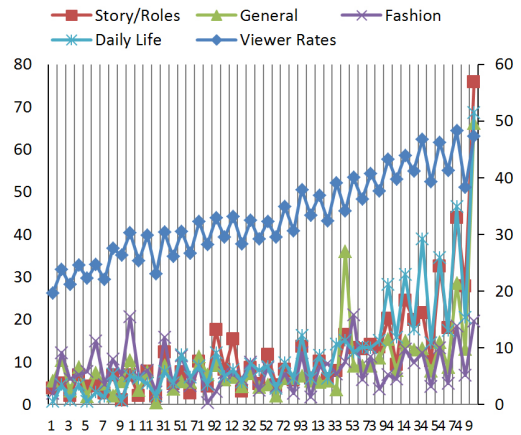
토픽 트렌드는 일정 기간 동안에 토픽들의 가중치의 변화를 나타낸 것이다. 드라마 별로 각 방영분에 대해 토픽 트렌드를 계산하기 위해 다음과 같은 공식을 적용하였다. 이전 연구에서는 당일 작성된 블로그 문서들을 대상으로 하여 각 토픽에 대해 평균값을 계산하여 토픽 트렌드 값을 구하였다. 본 연구에서 드라마에 대한 블로그 문서들을 대상으로 적용하여 토픽 트렌드를 구하기 위한 구체적인 공식은 아래와 같다(Kim et al., 2015).

$$\varphi^{(t)} = \frac{\sum_{d=1}^{D_t} \theta^{(d)}}{D_t}$$

- $\varphi^{(t)}$: 방영일 t의 토픽 트렌드 분포
- D_t : 방영일 t의 총 문서 수
- $\theta^{(d)}$: 문서 d의 토픽 분포

그러나, 위와 같이 계산된 식은 각 토픽들에 대한 상대적인 분포로서, 당일에 해당 토픽의 절대적인 인기는 반영하지 못한다. 상대적인 지표로 계산된 토픽 트렌드를 시청률과 연관시켜 트렌드를 보는 것은 바람직하지 않기 때문에, 이를 보정하기 위하여 본 연구에서는 위에서 구한 토픽 트렌드 값에 당일의 블로그 문서 수를 곱하였다.

<Figure 2>는 ‘왕가네 식구들’의 토픽 트렌드를 보여주고 있다. 좌측의 Y좌표는 토픽의 트렌드 값



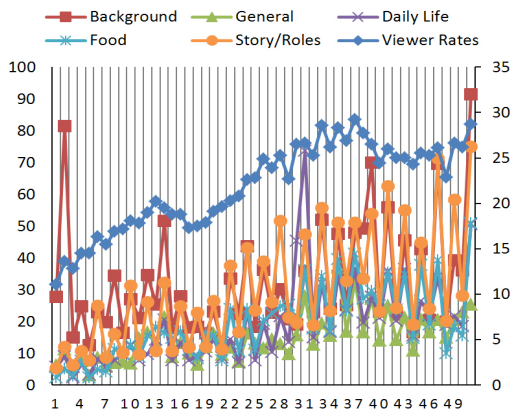
<Figure 2> Topic Trends : ‘Wang Family’

을 나타내고, 우측의 Y좌표는 시청률의 %값을 나타낸다. X축은 방영된 일자를 나타낸다. 이 그래프에서는 시청률이 꾸준히 상승하는 반면에 각 토픽의 값은 후반부에 가파르게 상승하는 것으로 나타난다. 특히 ‘일상생활’과 ‘즐거워, 역할’의 토픽값이 후반부에 상대적으로 더 상승하고 있다. 그러나, 토픽과 시청률 사이의 상관관계는 잘 보여지지 않고 있다.

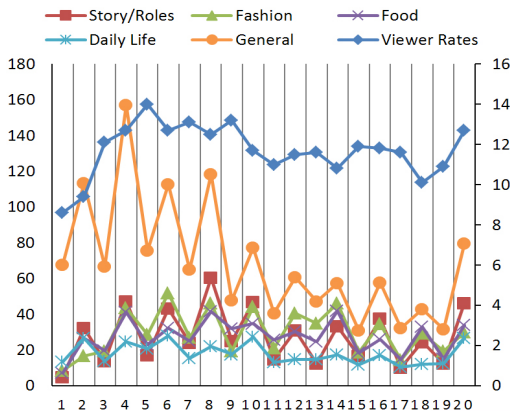
<Figure 3>은 ‘기황후’의 토픽 트렌드를 보여준다. ‘기황후’의 경우에는 토픽의 값들이 문서 수의 영향으로 인해 오르락내리락하는 경향이 있다. 즉 일주일 중 방영이 시작된 월요일과 화요일의 블로그 문서의 수가 차이가 크기 때문에 이로 인해 피크가 반복되는 경향이 있다. 전반적으로 토픽들이 시청률과 함께 움직이는 경향이 있음을 볼 수 있다.

<Figure 4>는 ‘닥터이방인’의 토픽 트렌드를 보여준다. ‘닥터이방인’은 기대만큼 흥행에 성공하지 못한 드라마로, 시청률 추이에서도 그러한 면이 보인다. 즉 초반 1/4시점까지는 시청률이 올라갔으나 그 이후로는 완만한 하락세를 보여준다. 이에 따라 토픽들도 전반적으로 하향의 추세를 보이고 있다. 다른 드라마와 뚜렷한 차이를 보이는 부분은 토픽들 중에서 ‘드라마 일반’에 대한 토픽이 초반부터 끝까지 가장 우세한 토픽이라는 점이다. 이는 드라마에 대해 다뤄지는 토픽들이 드라마 주변에서 맴돌고, 드라마의 세부적인 내용-‘줄거리, 역할’이나 ‘일상생활’과 연결되지 못한 것으로 해석된다.

<Figure 5>는 ‘별에서 온 그대’의 토픽 트렌드이다. ‘별에서 온 그대’는中间的 한 번의 피크를 제외하고는



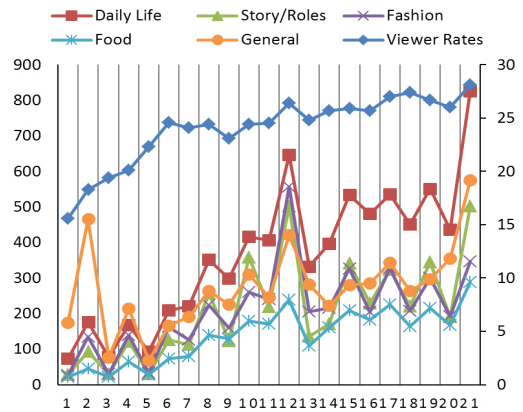
<Figure 3> Topic Trends : ‘Empress Qi’



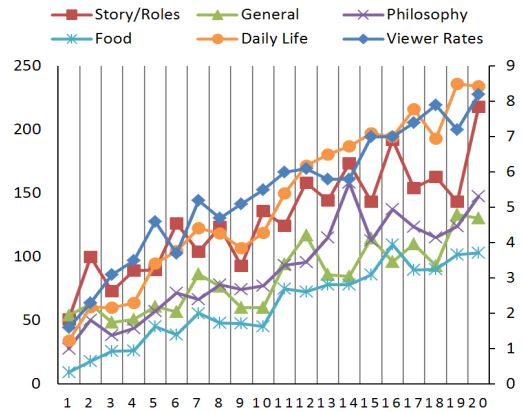
<Figure 4> Topic Trends : ‘Dr. Stranger’

시청률과 토픽 지수 모두 꾸준한 상승세를 보이고 있다.中间的의 피크는 다큐먼트 수의 영향으로, 다음 방영차수에서 갑작스런 하락세를 한번 보인다. 향후 분석에서는 이러한 특이점의 원인을 찾아보는 것도 흥미로울 것으로 생각된다. 그래프는 전반적으로 시청률과 토픽의 영향을 잘 보여주고 있는 것으로 판단된다.

<Figure 6>는 ‘미생’의 토픽 트렌드를 보여준다. ‘미생은’ 드라마들 중에서 가장 명확한 시청률과 토픽 간의 연관관계를 보여준다. 시청률과 다른 토픽들을 분간할 수 없을 만큼 거의 유사하게 움직이고 있다. 대부분의 토픽들이 그래프 상으로는 시청률과 함께 움직이는 것으로 보인다. 이는 시청률이 다른 드라마에 비해 낮은 반면 소셜 미디어에서의 참여는 높기 때문인 것으로 생각된다.



<Figure 5> Topic Trends : ‘My Love from the Star’



<Figure 6> Topic Trends : ‘Misaeng’

4.4 소셜 미디어 토픽과 시청률 사이의 연관성

이상과 같이 토픽 트렌드 기법을 이용하여 드라마가 방영된 기간 동안 드라마의 시청률과 드라마에 대한 소셜 미디어에서의 반응 사이의 연관관계를 살펴보았다. 그래프를 이용한 분석을 통해 다음과 같은 시사점들을 얻을 수 있었다. 첫째, 드라마에 대한 소셜 미디어에서의 반응이 낮으면 둘 간의 연관성을 찾기가 어렵다. 블로그 문서의 수가 다른 드라마에 비해 낮은 ‘왕가네 식구들’의 경우, 시청률이 매우 높음에도 불구하고 둘 간의 연관성을 찾기가 어려웠고, 이는 ‘기황후’에서도 마찬가지였다. 둘째, 시청률이 점차 떨어지는 드라마는 마찬가지로 소셜 미디어의 토픽도 살아나지 못하는 경향이 있다. 이는 소셜 미디어에서 다루지는 내용이 드라마 주변의 이야기에서 벗어나지 못하기 때문인 것으로 보이며, ‘닥터 이방인’을 그 예로 볼 수 있다. 셋째, 토픽과 시청률의 연관성이 높거나 소셜 미디어에서의 반응이 높은 드라마일수록 후반에 ‘일상생활’에 대한 토픽이 높게 나타났다. 이는 ‘미생’과 ‘별에서 온 그대’에서 동일하게 나타났다. 넷째, ‘음식’과 ‘패션’에 대한 토픽은 대부분의 드라마에서 다루지고 있으나 비중은 낮았다.

5. 결 론

본 연구에서는 방송 미디어와 소셜 미디어 간의 연관성을 알아보기 위해 드라마의 시청률과 그 드라마에 대한 블로그 문서 사이의 관계를 분석할 수 있는 방법을 제안하고자 하였다. 이를 위하여 먼저 2014년에 가장 시청률이 높았던 드라마 5개를 선정하고, 이들을 대상으로 네이버 블로그로부터 드라마가 방영된 날에 작성된 관련 문서들을 수집하여, 토픽 모델링을 이용해 문서의 내용에 대한 토픽을 분석하였다. 그리고 토픽 트렌드 분석을 이용하여 각 드라마에 대해 토픽들의 시간에 따른 추이를 그 드라마의 시청률과 함께

살펴보았다.

본 연구의 연구결과에 대한 의미를 요약해 보면, 첫째 소셜 미디어에서 드라마에 대한 토픽들이 실제로 존재하고 있음을 확인할 수 있었으며, 둘째 그러한 토픽들이 시청률과 상관관계를 갖고 있음을 확인할 수 있었다. 마지막으로 시청률이 높은 드라마와 그렇지 않은 드라마 사이에는 주로 다루지는 토픽에 차이가 있다는 점을 알 수 있었다. 이상과 같이 기존에는 짐작만 했던 내용을 실제 데이터를 통해 확인했다는 점에서 의의가 있다고 볼 수 있다. 활용성 측면에서는, 첫째 소셜 미디어 분석을 통해 시청률이 높은 드라마의 특징을 알아낼 수 있고, 이를 이용하여 드라마 기획 단계에서 어떤 점에 초점을 맞추어야 하는지를 알 수 있을 것으로 기대된다. 둘째 좋은 드라마를 만들기 위해 어떤 토픽들에 시청자들이 주목하도록 해야 하는지에 대한 시사점을 얻을 수 있다. 셋째 드라마 초반의 토픽들을 분석해 보면 드라마의 흥행가능성에 대한 정보를 얻을 수 있으며 동시에 어떤 점에 더 주력해야 하는 지에 대해서도 알 수 있을 것으로 기대된다.

본 연구에서는 드라마에 대한 블로그 문서들의 토픽들과 드라마의 시청률 간의 관계를, 시간에 따른 그래프로 시각화하고 그 속에서 의미를 찾으려고 하였다. 그러나, 제시된 관계가 통계적으로 어느 정도 유의한 지에 대해서는 제시하지 못하고 있다. 따라서 향후에는 이를 극복하기 위해 VAR(Vector AutoRegression)와 같은 시계열 분석 기법을 이용하여 보다 깊이 있는 분석을 수행하고자 한다. 또한, 대부분의 드라마가 일주일에 이를 연속 방영되기 때문에, 본 연구와 같이 방영 일만을 대상으로 분석할 경우 일주일 중 5일 동안의 소셜 미디어 반응은 분석에 포함되지 못하고 있다는 한계가 있다. 이와 같은 한계의 극복 외에도 향후 연구에서는 그래프에서의 특이점에 대해 원인을 분석함으로써 성공적인 드라마와 그렇지 않은 드라마 간의 차이를 밝혀낼 수 있을 것으로 기대된다.

References

- Bae, J.H., N.G. Han, and S. Min, "Twitter Issue Tracking System by Topic Modeling Techniques", *Journal of Intelligence and Information Systems*, Vol.20, No.2, 2014, 109-122.
- Blei, D., L. Carin, and D. Dunson, "Probabilistic Topic Models", *Signal Processing Magazine, IEEE*, Vol.27, No.6, 2010, 55-65.
- Blei, D.M. and J.D. Lafferty, "Topic Models", *Text mining : Classification, Clustering, and Applications*, Vol.10, No.71, 2009, 34.
- Blei, D.M., A.Y. Ng, and M.I. Jordan, "Latent Dirichlet Allocation", *the Journal of Machine Learning Research*, Vol.3, 2003, 993-1022.
- Bolelli, L., S. Ertekin, D. Zhou, and C.L. Giles, "Finding Topic Trends in Digital Libraries", *Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital libraries*, 2009.
- Brown, A., "Big Data, Big Opportunity?", available at <http://www.broadcastnow.co.uk/tech/facils/ibc-2013/big-data-big-opportunity/5060598.article>, 2013.
- Carr, D., "Giving Viewers What They Want", *The New York Times*, available at http://www.nytimes.com/2013/02/25/business/media/f-or-house-of-cards-using-big-data-to-guarantee-its-popularity.html?_r=0 2013.
- Cesar, P. and D. Geerts, "Past, Present, and Future of Social Tv : A Categorization", *Proceedings of Consumer Communications and Networking Conference (CCNC), 2011 IEEE*, 2011a.
- Cesar, P. and D. Geerts, "Understanding Social Tv : A Survey", *Proceedings of the Networked and Electronic Media Summit(NEM Summit 2011)*, Torino, Italy, 2011b, 94-99.
- Deb, R., "Social Tv", TVB, available at http://www.tvb.ca/pages/deb+roy+tv+day+2012_htm, 2012.
- Dmc_Marketing_Team, "2014 Sns Usage Report", DMC Media, available at <http://www.slideshare.net/BeckyKim/2014-sns-37219822>, 2014.
- Hyun, Y.J., N.G. Kim, and Y.H. Cho, A Multi-Dimensional Issue Clustering from the Perspective Consumers' Interests and R&D, *Journal of Information Technology Services*, Vol.14, No.1, 2015, 237-249.
- (현윤진, 김남규, 조윤희, "소비자 선호 이슈 및 R&D 관점에서의 다차원 이슈 클러스터링", *한국IT서비스학회지*, 제14권, 제1호, 2015, 237-249.)
- Jung, H.I., "Economic Ripple Effect of 'My Love from the Star'", *Cultural Tourism*, available at https://www.kcti.re.kr/webzine2/webzineView.action?issue_count=32&menu_seq=4&board_seq=1, 2014.
- (정현일, "드라마 '별에서 온 그대'의 경제적 파급효과", *문화관광*, No.7, 2014.)
- Kim, W.J., J.H. Won, S.U. Park, and J.Y. Kang, "Demand Forecasting Models for Medicines through Wireless Sensor Networks Data and Topic Trend Analysis", *International Journal of Distributed Sensor Networks*, Vol.2015, No.1, 2015, 1-6.
- KLT, "Klt : Korean Language Technology", available at <http://nlp.kookmin.ac.kr/HAM/kor/>, 2014.
- Kushin, M.J. and M. Yamamoto, "Did Social Media Really Matter? College Students' Use of Online Media and Political Decision Making in the 2008 Election", *Mass Communication and Society*, Vol.13, No.5, 2010, 608-630.
- Lee, H., 2014 Drama Rank, NEWSPIM, available

- at <http://www.newspim.com/view.jsp?newsId=20141222000007> 2014.
- Martie, L., V.K. Palepu, H. Sajnani, and C. Lopes, "Trendy Bugs : Topic Trends in the Android Bug Reports", *Proceedings of the 9th IEEE Working Conference on Mining Software Repositories*, 2012.
- Matsubara, Y., Y. Sakurai, C. Faloutsos, T. Iwata, and M. Yoshikawa, "Fast Mining and Forecasting of Complex Time-Stamped Events", *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2012.
- Michaelidou, N., N.T. Siamagka, and G. Christodoulides, "Usage, Barriers and Measurement of Social Media Marketing : An Exploratory Investigation of Small and Medium B2b Brands", *Industrial Marketing Management*, Vol.40, No.7, 2011, 1153-1159.
- Mimno, D., "Tutorial : Machine Learning with Mallet", available at <http://mallet.cs.umass.edu/mallet-tutorial.pdf>, 2013.
- Park, S.U., J.Y. Kang, Y.S. Bang, and K.S. Han, "Relationship Between Audience Engagement On Social Media And Broadcast Media-Ratings", *Proceedings of the 15th International Conference on Electronic Business*, 2015.
- Roettgers, J., "Let's Face It : Social Tv Is Dead", GIGAOM, available at <https://gigaom.com/2014/01/29/lets-face-it-social-tv-is-dead>, 2014.

◆ About the Authors ◆**Sangun Park (supark@kgu.ac.kr)**

Professor Sangun Park is currently an associate professor of MIS at Kyonggi University. He received his B.A. in Computer Science, M.S. and Ph. D. in Management Engineering from Korea Advanced Institute of Science and Technology. His current primary research interests are Big Data Analysis, Semantic Web, Cloud Computing, Intelligent Information Systems, Artificial Intelligence and Text Mining.