

YouTube 동영상 의견분석을 통한 사용과 충족 이론 측정 : 트로트 가수 조명섭 동영상을 중심으로

Analyzing Comments of YouTube Video to Measure Use and Gratification Theory Using Videos of Trot Singer, Cho Myung-sub

홍한국*, 임병학**, 김삼문***

동의대학교 경영정보학부*, 부산외국어대학교 경영학부**, 동의대학교 소프트웨어공학과***

Han-Kook Hong(honghk@deu.ac.kr)*, Byung-hak Leem(bhleem@pufs.ac.kr)**,
Sam-Moon Kim(sammoon@deu.ac.kr)***

요약

본 연구의 목적은 소셜미디어 중 하나인 YouTube 동영상 사용자들이 남긴 의견을 추출하여 분석하는 질적 연구방법을 제시한다. 이를 위해서 YouTube 동영상 사용자의견을 사용하여 사용과 충족 이론의 쾌락적 충족, 사회적 충족, 그리고 실용적 충족을 빈도분석과 토픽모델링을 통해 측정하였다. 측정결과, YouTube KBS 한국방송 채널 중 트로트 가수 조명섭 동영상을 사용자들이 시청하는 이유는 첫 번째로 높은 빈도를 보이는 것이 쾌락적 충족을 위해서였다. 다음 순으로 사회적 충족과 실용적 충족으로 나타났다. 단어-문서 네트워크 분석에서 연결정도중심성은 '응원', '감사', '화이팅', '최고' 등이 높게 나타났고, 매개중심은 '감사', '응원', '화이팅' 등의 단어가 높게 나타나 연결정도 중심성과 유사함을 보였다. 아이겐벡터중심성은 '사랑', '마음', '감사' 등의 단어가 높게 나타나 사용자들의 의견들에 가장 영향력이 높은 단어들을 알 수 있다. 이는 YouTube의 트로트 가수 조명섭 동영상 시청자들 중 대다수가 동영상에 대해 사랑과 감사의 마음을 보이고 있음을 알 수 있다. 위의 세 가지 중심성 분석결과는 동영상을 시청하는 동기로 사용충족 이론의 쾌락적 충족과 사회적 충족 관련 단어들인 높은 값을 보이고 있다. 본 연구는 설문조사 기반의 구조방정식 모형을 따르지 않고, 질적분석 연구를 자동화한 텍스트마이닝 기법을 사용하여 YouTube동영상을 사용하는 동기를 사용 및 충족 이론에 의해 밝혀냈다는 것에서 연구 함의를 찾을 수 있다.

■ 중심어 : | 의견분석 | 사용충족이론 | 소셜미디어분석 | 토픽모델링 | 텍스트네트워크분석 |

Abstract

The purpose of this study is to present a qualitative research method for extracting and analyzing the comments written by YouTube video users. To do this, we used YouTube users' feedback to measure the hedonic, social, and utilitarian gratification of use and gratification theory(UGT) through by using analysis and topic modeling. The result of the measurement found that the first reason why users watch the trot singer, Cho Myung-sub's video in the KBS Korean broadcasting channel is to achieve hedonic gratification with high frequency. In word-document network analysis, the degree of centrality was high in words, such as 'cheering', 'thank you', 'fighting', and 'best'. Betweenness centrality is similar to the degree of centrality. Eigenvector centrality also shows that words such as 'love', 'heart', and 'thank you' are the most influential words of users' opinions. The results of the centrality analysis present that the majority of video users show their 'love', 'heart' and 'thank you' for the video. it indicates that the high words in centrality analysis is consistent with the high frequency words of hedonic and social gratification dimension of the UGT. The study has research methodological implication that shed light on the motivations for watching YouTube videos with UGT using text mining techniques that automate qualitative analysis, rather than following a survey-based structural equation model.

■ keyword : | Use and Gratification Theory | Social Media Analytics | Topic Modeling | Text Network Analysis |

* 이 논문은 2020학년도 동의대학교 교내연구비에 의해 연구되었음(202003310001).

접수일자 : 2020년 06월 29일

심사완료일 : 2020년 08월 24일

수정일자 : 2020년 08월 24일

교신저자 : 홍한국, e-mail : honghk@deu.ac.kr

I. 서론

웹 트래픽을 분석하는 Amazon 자회사인 Alexa.com 에 따르면 YouTube는 세계에서 가장 인기 있는 소셜 미디어 사이트이다. YouTube 이용자 수는 Facebook 또는 Wikipedia와 같은 웹 거인의 사용자 수를 초과했다. 지난 12년 동안 YouTube는 사용자가 다양한 장르의 동영상을 찾고 볼 수 있는 다양한 플랫폼이 되었다.

YouTube의 시청자에 대한 다양한 연구 목적은 동영상 시청 행위, 동영상에 묘사된 문제 및 사용과 충족에 대해 가치 있는 정보 자원을 만드는 데 있다. 지금까지 YouTube는 중요한 토픽[1], 의사소통가치[2], 동영상의 소주제별 내용[3] 및 시스템 자체[4] 등의 주제를 가지고 연구되어 왔다. 또한 전통적인 내용분석의 일환으로 YouTube에 대한 토픽분석 및 감성분석도 연구되어 왔다[5]. 최근에는 텍스트마이닝 기술의 발달로 내용분석의 자동화가 가능해졌다. 오늘날 혁신적인 기업들은 자동화된 텍스트마이닝 기반 소셜미디어 분석도구를 사용하여 YouTube의 데이터를 자동으로 수집, 결합 및 분석하여 시청자, 고객, 그리고 경쟁자를 이해하고 고객관리 및 경쟁전략을 수립하는 새로운 방법을 찾고 있다.

본 연구는 소셜미디어 중 하나인 YouTube 동영상 사용자들이 남긴 의견을 추출하여 분석하는 질적 연구를 통해 지속적으로 가수의 인기관리를 위한 팬관리 방안을 제시한다. 본 연구의 목적은 텍스트분석과 토픽모델링을 사용하여 YouTube 동영상 사용자의견이 사용과 충족 이론을 반영하고 있는지를 밝히고자 한다. 따라서 다음 장은 사용충족이론, 소셜미디어 분석과 토픽모델링에 대한 이론적 배경과 선행문헌 연구를 기술하였다. 3장은 유튜브 동영상 사용자 의견추출, 분석방법 및 절차에 대해서 기술하고, 이에 따른 분석결과를 4장에서 기술하였다. 마지막 5장은 연구결과 요약, 한계점, 향후 연구방향을 포함하는 결론을 기술하였다.

II. 이론적 배경

1. 사용 및 충족 이론 (Use and Gratification Theory)

UGT는 사용자의 심리적, 사회적 요구를 충족시키는 서로 다른 미디어와 다양한 콘텐츠에 사용자를 끌어들이고 유지하는 충족 또는 혜택에 대한 연구를 말한다[6]. UGT의 주요 목표는 미디어와 상호작용할 때 사용자가 사회 및 개인 충족에 대한 이해를 높이고 사용자의 동기를 설명하기 위해 사용자가 특정 유형의 매체를 선택하는 원인을 밝히는 것이다[7]. 지금까지 UGT는 사용자들이 신문, 전화, 라디오와 같은 매스미디어를 사용하는 동기를 측정했다. 최근에는 SNS[8][9], 모바일 SNS[10], 온라인게임[11] 등의 소셜 미디어를 사용하는 동기를 UGT를 활용하여 밝히고 있다. 따라서 UGT는 사용자의 주어진 요구를 충족시키고 지식, 휴식, 사회적 상호작용, 기분전환 또는 탈출을 향상시킬 수 있는 미디어와 콘텐츠를 선택하는 방법에 대해 설명한다. 오늘날 사용자는 미디어의 수동적 소비자가 아니라, 오히려 미디어 소비를 지배하고 미디어를 자신의 삶에 해석하고 통합하는 데 적극적인 역할을 한다[12].

SNS 사용자는 자주 정보추구, 사회적 상호작용, 자신의 표현, 즐거움, 사회적 존재 및 소속감과 같은 개인 및 사회적 욕구를 충족시키고자 한다[1][13][14]. Dhir and Tsai[15]는 즐거움, 노출, 탈출, 정보추구 및 사회적 영향의 충족감에 따라 Facebook 사용의 동기가 다르다고 주장했다. Hsiao et al.[9]는 모바일 소셜앱의 지속적인 사용은 실용주의적 동기, 쾌락적 동기 및 사회적 동기로 측정될 수 있음을 입증하였다. Balakrishnan과 Griffiths[16]는 YouTube 중독의 원인이 유튜브 콘텐츠 제작 및 방문의도에 있다고 가정하고 제작 및 방문 동기가 콘텐츠 충족(공유와 자기표현), 사회적 충족, 프로세스 충족(즐거움, 자기표현, 시간보내기) 그리고 기술적 충족(미디어 매력과 편리성)임을 실증 분석하였다. Hossain[7]도 또한 페이스북 사용의도는 즐거움, 시간보내기, 정보탐색, 자기표현, 사회적 존재 및 사회적 사회작용에 의해서 영향을 받고 있음을 보였다. Hossain, et al.,[7]은 또 Facebook의 좋아요(Like) 행위 동기를 쾌락적 충족(즐거움과 시간보내기), 실용주의적 충족(정보탐색과 자기표현), 사회적 충족(사회적 상호작용과 사회적 존재)으로 나누어 실증 분석하였다 그 결과 쾌락적 충족엔 즐거움이, 실용주의적 충족엔 정보탐색이, 사회적 충족엔 사회적 상

호작용이 '좋아요'를 선택하는 행위에 영향을 미치고 있는 것으로 나타났다. 이들의 연구들은 모두 설문조사에 의해 구조방정식모형을 사용하여 실증 분석하였다. 이 연구방법은 미디어 사용의도의 동기를 밝히는 데 많은 기여를 해왔다. 하지만 이 방법은 설문응답자의 과거 기억에 의존해야 하고 시간 소모적이란 단점을 가지고 있다. 컴퓨터와 텍스트마이닝 기술의 발달로 사용자들이 사용후기로 남긴 의견의 텍스트데이터를 분석하여 UGT 기반 미디어 사용동기를 밝힐 수 있다면 획기적인 새로운 연구방법으로 여겨질 수 있다.

따라서 본 연구는 Balakrishnan과 Griffiths[2]의 연구와 Hossain[7]의 연구를 기반으로 UGT을 YouTube동영상 사용자 의견에 얼마나 반영하고 있는지를 텍스트 분석을 통해 입증하고자 한다.

2. 소셜미디어 분석

최근 빅데이터 분석의 부상으로 기업조직에서 실무자와 연구자 모두에게 중요한 연구 분야로 떠오르는 것은 소셜미디어 분석이다. Wedel & Kannan[17]에 의하면, 소셜미디어 분석은 사용자들이 생성하여 남긴 데이터(footprint)를 추출하여 미래의 추세나 행동을 예측하는 것으로 소셜미디어 플랫폼에서 데이터를 수집하고 분석하여 의사 결정자가 특정 문제를 해결하도록 도움을 주는 것을 말한다. 소셜미디어 분석의 명성은 인기 있는 소셜미디어 플랫폼을 통해 기업이 엄청난 양의 고객데이터에 접근할 수 있게 되면서 급증했다. 소셜미디어 플랫폼은 콘텐츠제작자와 콘텐츠소비자 특유의 그룹에 중점을 두고 있다. 소셜미디어 분석은 소셜미디어 데이터를 분석하여 소비자의 혁신적인 아이디어를 얻고 고객 관계를 강화할 수 있다. Fortune 500대 기업들은 경쟁우위를 확보하기 위해 소셜미디어 분석을 사용하고 있다. 많은 유형의 조직 중에서 기업은 소셜미디어 분석에 가장 활발한 사용자이다. 기업에 소셜미디어 데이터를 분석하여 고객이 왜 제품이나 서비스를 구매하는지를 잘 이해하는 것은 경쟁우위를 유지하는 데 중요한 역할을 한다[18]. 기업은 소셜미디어 플랫폼에서 수집한 정보와 기존 고객정보를 결합하여 고객행동을 보다 잘 이해할 수 있다[19]. marketandmarkets.com에 따르면 전 세계 소셜미디어 분석시장은 27.6%의 연간

복리성장률로 2015년 16억 달러에서 2020년 54억 달러로 성장할 것으로 예측하고 있다. 북미 지역은 제조, 보건, 운송 및 물류와 같은 여러 산업에서 소셜미디어 분석이 널리 보급되어 소셜미디어 분석공급업체들의 최고 수익창출지역이고, 유럽지역이 그 뒤를 잇고 있다. 따라서 혁신적인 기업의 관리자는 소셜미디어 분석도구를 사용하여 소셜미디어에서 데이터를 자동으로 수집, 결합 및 분석하여 고객을 이해하고 고객관계를 관리하며 새로운 제품을 디자인하는 새로운 방법을 찾고 있다.

기업이 소셜미디어 분석을 적용할 수 있는 영역은 시장 지향성과 적시성에 따라서 실시간과 비실시간, 경쟁자와 고객에 따라 4가지의 소셜미디어 분석범주가 있을 수 있다[20]. 다음 [표 1]은 소셜 미디어를 통한 실시간 고객분석, 비실시간 고객분석, 실시간 경쟁분석 및 비실시간 경쟁분석에 대해 요약하여 보여 주고 있다.

표 1. 기업의 소셜미디어분석 유형

		적시성	
		실시간	비실시간
시장지향	고객	실시간 고객소셜미디어분석 대응적 마케팅: 온라인 리뷰(의견)분석-키워드 분석, 위치분석, 담론분석, 불만분석, 경고분석	비실시간 고객 소셜미디어분석 적극적 마케팅-수익성 고객 식별, 소셜네트워크 분석, 영향력 분석, 웹분석, 감성분석
	경쟁자	실시간 경쟁자 소셜미디어분석 운영정보-가격과 판촉 모니터링, 뉴스 분석, 신제품 발표, 인수 합병	비실시간 경쟁자 소셜미디어분석 전략 및 전술 정보-경쟁자 가격의 주기적 추세 분석, 신제품 생산 및 개발, 기술개발, 고객불만, 종업원 의견

자료: Lee [20]의 논문기업 소셜미디어 분석 유형을 정리함.

다양한 소셜미디어 플랫폼 중 본 연구가 지향하고 있는 YouTube기반 내용분석에 대한 선행연구들을 [표 2]에 요약하여 보여주고 있다. Grant, et al.[21]와 Rashield & Zeeshan[22]는 전통적인 내용분석에 의해 YouTube동영상을 분석하였으며, Thelwall, M.[4]와 Parabhoi & Saha[23]은 텍스트마이닝 방법에 의한 분석을 하였다. 본 연구는 UGT이론을 기반으로 YouTube 동영상 시청하는 동기를 밝힌다는 점에서 차이점을 찾을 수 있다.

표 2. 소비자 리뷰의 내용분석 연구들의 예

저자	대상	분석 기법
Grant et al.[21]	YouTube Videos & Comments	내용분석
Rashied & Zeeshan[22]	YouTube Videos & Comments	넷노그래피
Theilwall, M.[23]	YouTube Comments	토픽모델링과 감성분석
Parabhoi & Saha[24]	YouTube Comments	감성분석

3. 토픽모델링

소셜미디어의 고객 의견을 분석하려면 먼저 의견 데이터에 텍스트마이닝을 적용해야 한다. 텍스트 마이닝은 출현하는 단어의 빈도와 비율을 사용하여 통계적으로 추정하고 텍스트데이터를 요약하고 분류하는 데 사용된다[24]. 이는 대량의 문서에서 의미 있는 정보를 추출하는데 목적을 두고, 비정형의 텍스트데이터에서 유용한 모델, 추세, 패턴 또는 규칙을 찾는 데 목을 둔다[25]. 지난 10년 동안 비지도 기계학습방법인 확률적 토픽모델링은 사회과학분야에서 텍스트마이닝 도구로 점점 더 관심을 지속적어 받아왔다[25]. 토픽모델링의 기본개념은 모든 문서가 사전에 알려지지 않은 여러 토픽을 표현할 수 있다는 것이다[26]. 따라서 토픽모델링은 텍스트문서가 단어로 구성되어 있고 여러 문서에서 말하는 토픽이 강하게 연결된 단어의 조합으로 표현될 수 있기 때문에 텍스트문서에서 숨겨진 토픽을 추출할 수 있게 해준다[27]. 이를 위해 토픽모델링의 기본가정은 각 문서는 각 토픽을 다양한 정도로 표현될 수 있고, 각 문서는 여러 토픽에 속할 수 있다[28].

다양한 토픽모델링 알고리즘 중에서, Blei et al.[29]가 개발한 LDA(Latent Dirichlet Allocation)가 가장 널리 사용되는 모델링방법이다[29; 30]. LDA의 장점은 대규모 문서를 처리하고 식별된 잠재 토픽을 해석할 때 여러 토픽모델링 알고리즘 중 가장 높은 성능을 가지는 것으로 알려져 있다 [28]. 문서에 단어의 동시 출현으로 단어와 토픽의 의미를 정의하는 토픽모델링은 관계적 근법을 따르고 있다[26]. 다시 말해서, LDA는 주어진 문서에 잠재된 토픽을 가정하는 확률적 모델이며, 문서 내에 토픽의 확률적 분포와 각 토픽을 구성하는 단어의 확률분포가 주어지면, LDA는 문서를 구성하고 선택된 토픽의 단어를 반복적으로 샘플링하여 확률적으로 토픽을 선택한다[27]. 즉 LDA는 각 문서가 토픽의 혼합

으로 표현되고 토픽은 단어의 혼합으로 표현되는 구조를 갖는다. LDA에 의한 토픽모델링의 근본 원리는 관측 가능한 문서-단어 행렬은 토픽-단어 행렬과 문서-토픽 행렬로 분해 될 수 있다는 것이다. Calheiros et al.[29]에 따르면 LDA는 각 토픽에 속하는 선정된 리뷰의 확률을 결정하고, 고려된 각 단어의 근접성에 따라 리뷰를 그룹화 할 뿐만 아니라 어떤 토픽이 더 많은 관심을 끌고 있는지 식별하는 데 도움이 된다.

토픽모델링에 사용되는 대부분 데이터는 웹사이트 및 온라인광고[31], 소셜미디어게시물[32], 온라인제품 리뷰[33]와 같은 다양한 형태의 비구조화된 텍스트데이터이다. 물론 이미지 [14], 구매기록, 모바일앱 사용기록[34] 및 인터넷검색 행동의 흔적[35][36]과 같은 다른 형태의 데이터도 있다. Reisenbichler & Reutterer [37]에 따르면, 마케팅 분야에서 토픽모델링은 온라인 텍스트소비자 리뷰 및 서비스 연구, 영업/소매업, 소셜 미디어 분석, 이미지데이터 분석, 온라인광고 분석, 그리고 마케팅 문헌 연구 등에 적용되고 있다.

III. 연구방법

1. 동영상 사용 동기 요인 정의

YouTube동영상 이용은 쾌락적 충족, 사회적 충족 그리고 실용주의적 충족에서 동기부여 됨을 2장에서 알 수 있었다[16][38]. 쾌락적 충족은 즐거움과 시간 보내기, 사회적 충족은 사회적 존재와 사회적 상호작용, 그리고 실용주의적 충족은 정보공유 및 자신의 표현에 의해 측정된다.

YouTube는 상호작용을 통해 동영상방문자의 경험과 사용자간 즐거움을 새롭게 해 준다. 따라서 효과적인 즐거움, 시간소모 드리고 자신의 표현을 할 수 있는 YouTube의 역할을 통해 쾌락적 충족을 만족시켜 준다[38]. 따라서 쾌락적 충족은 YouTube동영상을 통해 즐거움과 재미를 얻고 행복과 마음의 쉼을 느끼는 정도를 말한다[38]. YouTube의 주요 활동은 사용자가 비디오 공유를 통해 사회적 상호작용에 참여할 수 있도록 하는 것이다. YouTube의 의사소통 구조와 온라인에 게시된 다양한 비디오는 다양한 사회적 상호작용을 가

능하게 해 주고 있다[39]. 마지막으로 사회적 충족은 이러한 사회적 상호작용을 말한다. 사회적 상호작용은 사용자간 소셜미디어 상에서 인적 네트워크를 형성하고 새로운 만남을 통해 가치를 공유하는 정도를 말한다 [16]. YouTube는 동영상과 의견을 사용자에게 정보 출처로 표현할 수 있게 하여 사용자들간 정보를 공유하고 사용자 자신을 표현할 수 있게 해준다. 이는 사용자의 실용적 충족을 만족시킬 수 있도록 해 주고 있다. 정보공유는 유용하고 도움이 될 수 있는 새로운 정보를 제공하고 공유하는 정도를 말한다. [표 3]은 쾌락적, 사회적, 그리고 실용적 충족에 대해 정의와 관련 키워드를 보여 주고 있다.

표 3. 유튜브 동영상 사용 동기 정의 및 관련 키워드

사용동기	정의	관련 키워드
쾌락적 충족	즐거움과 재미를 얻고 행복과 마음의 평화를 느끼는 정도	즐거움, 기쁨, 행복, 힐링, 마음, 감동
사회적 충족	사용자간 YouTube상에서 인적 네트워크를 형성하고 새로운 만남을 통해 가치를 공유하는 정도	최고, 마음, 모습, 파이팅
실용적 충족	유용하고 도움이 될 수 있는 새로운 정보를 제공하고 공유하는 정도	목소리, 가요, 트로트, 무대, 조명

2. 연구절차

토픽모델링은 [그림 1]에서 보는 바와 같이 5단계를 거쳐 수행된다[40]. 토픽모델링 프로세스는 분석하고자 하는 텍스트의 원천을 정의하는 것에서 시작된다. 게시된 텍스트는 동적이고, 대량이며, 다양한 언어로 되어 있으면서 노이즈가 있다. 따라서 토픽모델링을 위해 올바른 원천을 찾는 것은 유용한 비즈니스 인사이트를 추출하는데 중요하다. 다음 단계는 기계학습에 가장 기본이 되는 자연어처리기술을 사용하여 텍스트를 분해, 정제, 여과를 하는 전처리이다. 텍스트에서 의미를 추출하기 위한 컴퓨터와 알고리즘을 사용하여 문장구조와 품사의 결정, 추출된 객체의 명명, 그리고 불용어를 제거한다. 이 단계의 대부분은 자동적으로 수행되지만 상황에 따라 연구자의 개입이 요구된다. 예를 들면, 여과단계에서 불필요하거나 원치 않는 용어를 제거하기 위해서는 수동적으로 정제를 해야 한다. 세 번째 단계는 텍스트에 분석알고리즘을 적용하기 전에, 텍스트는 분석을 위해 컴퓨터가 이해할 수 있는 형태로 변환되어야 한다. 따라서 앞 단계에서 정제된 텍스트는 벡터공간모

델과 같은 선형대수기반 기법을 사용하여 수치화된 표현으로 변환되어야 한다. 즉 텍스트는 문서-단어 행렬, 토픽-단어 행렬, 문서-토픽 행렬로 변환된다. 이때 텍스트에 나타난 단어빈도와 단어동시 출현에 의한 단어 행렬이 생성된다. 이 단계가 완료되어야 비로소 다음 단계의 토픽분석을 수행할 수 있다. 이 단계는 단어의 군집화를 통해 텍스트에서 토픽을 추출하고 해석하는 것이다. 따라서 토픽모델링은 이러한 복잡한 알고리즘을 사용하여 인간이 하는 방법으로 텍스트에서 의미를 추출하는 것이다.

Schmiedel et al.[25]에 따르면, 데이터수집 단계에서 데이터가 토픽모델링을 사용하여 특정 연구 질문에 대답하는 데 도움이 되어야 하고, 결과의 일반화가 가능해야 하며, 토픽모델링 알고리즘 이후의 통계는 정확하고 의미 있는 결과를 생성하기에 충분한 양의 텍스트가 필요로 하기 때문에, 데이터의 관련성, 대표성 및 규모가 고려되어야 한다.

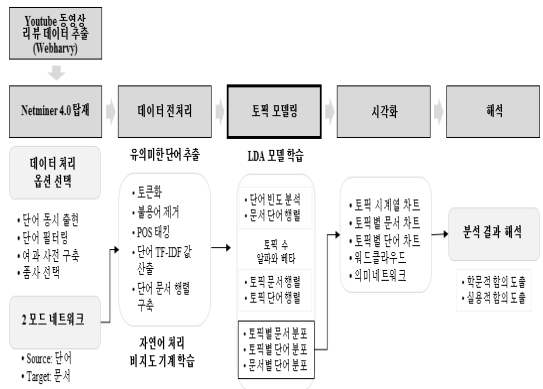


그림 1. 연구 절차

YouTube의 거대한 사용자의견 텍스트데이터를 추출하기 위하여, 웹 크롤러인 Webharvy를 사용하였다. 이 크롤러 도구는 YouTube동영상에 대해 개별적인 사용자의견의 수천 개까지 추출할 수 있도록 지원한다. 사용자의견 분석을 위하여 KBS한국방송채널에서 본 연구와 관련된 동영상에 대한 의견만 추출하였다. 본 연구는 사용자의 의견이 독립적인 기여라는 것을 확보하기 위해 의견에 대한 답글은 제외하였다.

본 연구는 2019년 말부터 2020년 중반까지 트로트

계에 인기를 끌고 있는 조명섭 가수의 YouTube 동영상의 댓글을 대상으로 하였다. 조명섭 트로트가수와 관련된 인기 있는 상위 5개 채널은 KBS 한국방송, KBS Entertainment, 튜브 54, 미스트로트TV, 조명섭 다음팬클럽이다. 4채널들 중에서 KBS의 동영상이 다른 채널들과 대부분이 중복되는 동영상들이기에 KBS한국방송 채널에서 제공한 YouTube동영상 시청자의견을 추출하였다. KBS한국방송채널에서 “트로트가 좋아”1차경연 하이라이트 동영상을 2019년 11월 28일 업로드 되어 2020년 2월1일 기준 2,303,373 방문자를 기록하고 있고 1,808개의 시청자의 의견을 추출하였다. 다음 [그림 2]는 YouTube 상에서 추출한 시청자의견 부분을 보여 주고 있다.



그림 2. 조명섭 가수 YouTube 동영상 시청자의견

분석을 위한 최종 데이터세트는 시청자가 남긴 1,808개의 의견 중 1,188개의 유용한 의견을 대상으로 하였다. 전처리 후 토픽모델링을 위해 두 글자 이상의 단어와 TF-IDF값 0.5 이상 되는 단어를 추출하여 1,318개 단어와 1,116개 문서를 생성하였다. 문서로 된 의견은 짧은 한 개 내지 두 개 정도의 문장으로 이루어져 있다. 의견의 길이는 그 자체의 가치에 영향을 미치지 않는 것으로 보이지만, 정보의 풍부함을 어느 정도 반영한다[41]. 또한 스마트폰 및 태블릿과 같은 소형 휴대용 기기가 널리 채택됨에 따라 리뷰가 점점 짧아지면서 연구자는 토픽모델링을 사용하여 짧은 텍스트를

분석하는 데 여러 문제들을 고려해야 한다[42].

본 연구는 Netminer 4.0 소프트웨어 패키지의 토픽 모델링 도구를 사용하여 말뭉치의 토픽을 찾아냈다. Netminer 4.0은 SNA (Social Network Analysis)를 기반으로 대규모 네트워크 및 텍스트 데이터의 탐색적 분석 및 시각화를 위한 기계학습 및 데이터 마이닝 도구이다. Schmiedel et al.[25]에 따르면, 토픽 발굴은 토픽의 노이즈(소음)를 줄이기 위해 표준 및 연구자가 정의한 모든 불용어 제거를 필요로 한다. 본 연구에서 불용어는 일반적인 단어, 감정을 반영하는 단어 및 연구 맥락에서 의미 없는 단어 등을 포함한다.

텍스트데이터에 LDA를 적용할 때, 연구자는 생성될 토픽의 수를 선택해야 한다. Westerlund et al. [43]의 개념에 따라, 본 연구는 중첩을 피하고 토픽의 해석 가능성을 보장하는 시행착오를 토대로 토픽의 수를 선택했다. Calheiros et al.[29]에 따르면 소수의 토픽은 토픽간 중복을 피하고, 차례로 구조타당성을 확보하는 데 도움이 된다[25]. 또한, Lei and Law[44]는 소수의 토픽은 일반적으로 소비자 리뷰에서 핵심속성을 포함하고 있음을 제시하고 있다. 이 개념은 Maier et al.[45]의 주장과 잘 부합되며, 가능한 토픽의 수는 이론적인 개념, 연구맥락 및 연구의 목적과 일치하도록 세분화하는 것이 실용적이다. 본 연구는 최적토픽수를 시행착오를 통해 토픽이 서로 배타적으로 나누어지고, 상위 10까지 등장하는 단어들이 서로 겹쳐지지 않을 때를 선택했다. 또한 토픽-단어의 2 모드 네트워크로 표현하여 스타형일 때 가장 설명력이 높은 것으로 판단하여, 본 연구는 토픽 수를 6개로 선택하였다.

토픽 수 K 와 더불어 문서의 토픽분포 생성을 위한 디리클레분포 파라미터 α 와 토픽의 단어분포 생성을 위한 디리클레분포 파라미터 β 도 결정되어야 잠재적 토픽이 문서들에서 결정된다. 알파값을 높게 잡으면 문서의 토픽연관확률이 동일하게 나타나며, 이는 문서에 모든 토픽이 동일한 확률로 분포함을 의미한다. 베타값을 높게 잡을 시는 토픽의 단어 연관확률이 동일하며, 토픽에 대한 모든 단어의 연관확률이 동일함을 의미한다. 따라서 이들 값의 최적 값은 없으나 본 연구는 기존 연구자의 결과를 기반으로 설정하였다. Griffiths and Steyvers[46]에 따르면 α 와 β 의 선택은 LDA 결과에

중요한 영향을 줄 수 있다. 그들은 $\alpha=50/k$ (k : 토픽의 수)와 $\beta=0.01$ 을 사용했다. β 에 작은 값을 사용한 이유는 연구의 특정 영역을 나타내는데 더 많은 토픽을 만들기 위해서였다. Steyvers와 Griffiths[47]는 파라미터 α 와 β 가 토픽의 수와 어휘 크기에 따라 다르다는 것을 입증했다. 또 그들은 $\beta=0.01$ 이외의 값을 그리고 $\alpha=50/k$ 값을 사용할 것을 권고하고 있다. 최근에 Zhao, et al.[49]은 LDA의 파라미터 α 값은 0.01과 0.1 사이, β 값은 0.01일 때 가장 설명력이 높은 토픽모델을 만들 수 있음을 제안했다.

표 4. LDA의 파라미터 관련 연구

저자	응용분야	테스트 코퍼스	α	β
Steyvers & Griffiths [47]	토픽 분류	TASA	50/K	0.01
Lu et. al., [48]	토픽 분류 및 정보검색	TDT2	[0.1, 0.5] or [0.5, 2]	0.01
Zhao et al. [49]	토픽 식별	Reuters-21578	[0.01, 0.1]	0.01
Naili, et al. [50]	토픽 식별	Al-Watan	0.5	0.01

본 연구에서 $\alpha=50/k$ (k : 토픽의 수)에 의해 8.33으로 하였으며, $\beta=0.01$ 을 사용했다[47]. 관찰된 문서들 통해 잠재된 토픽을 추출하는데 기반이 되는 문서-단어 행렬의 2 모드 네트워크 (source: 단어, target: 문서 ID)가 추출되었다. 이 행렬은 문서(열) 내의 단어(행)의 분포(빈도)를 나타낸다. 실제 텍스트마이닝에서 단어-문서 행렬은 토픽모델링의 LDA 알고리즘에 대한 입력으로 사용되는 벡터-공간 표현의 2차원 매트릭스 구조를 생성한다. 행렬의 각 셀은 문서에 단어가 등장하는 빈도를 표현한다. 단어-문서 행렬은 빈도를 통해 문서와 문서 간의 관계를 특징짓는다[51].

LDA를 이용한 토픽모델링에서 비지도학습을 위해 가장 널리 사용하는 샘플링방법이 깃스샘플링(Gibbs sampling)이다. 깃스샘플링은 Monte Carlo Markov-chain (MCMC) 알고리즘으로, 각 변수의 조건부분포를 효율적으로 계산할 수 있는 경우 결합분포에서 샘플을 생성하는 통계적 추론의 강력한 기법이다. 본 연구는 이 깃스샘플링 방법을 사용하여 1000회 반복을 통해 학습하였다.

IV. 연구결과

1. 빈도 분석

YouTube의 KBS한국방송채널에서 “트로트가 좋아” 1차 경연 하이라이트 동영상의 시청자의견 1,116개 의견(문서)을 비지도학습 기반의 언어처리에 의한 데이터 전처리한 결과, 1,318개의 단어가 수집되었다. 트로트가 좋아 1차 경연대회 하이라이트에 대한 시청자의견에 등장한 단어들을 빈도수에 의해 정렬하였고, 그 중 상위 100개의 단어를 추출하여 [표 5]와 같이 정리하였다. 1위는 쾌락적 충족을 나타내는 단어 ‘노래’로 423번 언급되었으며, 2위와 3위는 각각 실용적 충족을 나타내는 단어 ‘가수’, ‘목소리’로 각각 212번, 172번 언급되었다. 4위는 사회적 충족을 나타내는 단어 ‘최고’로 111회 언급되었다. 각 동영상 사용동기별 키워드들의 빈도와 비율을 다음 [표 5]에서 보여 주고 있다.

빈도분석으로 의미를 찾기는 쉽지 않다 그래서 단어가 많이 사용은 되었지만 중요하지 않은 단어들도 있을 것이다. 이러한 것을 위해서 TF-IDF값을 계산하여 일정 임계값보다 큰 것을 대상으로 하여 토픽분석을 진행한다. TF-IDF값이 0.5이하를 갖는 단어들은 유의미한 단어가 아닌 것으로 판정되어 0.5 임계값을 기준으로 그 이상 되는 단어를 대상으로 하였다. 다음 [그림 3]은 IF-IDF가 0.5 이상의 값을 가지는 단어들로 구성된 워드클라우드를 보여 주고 있다. 큰 값이 TF-IDF값이 큰 단어를 의미한다.

흥미로운 사실은 ‘노래’, ‘가수’, ‘목소리’는 빈도수는 높지만 TF-IDF 값이 작아 워드클라우드에 큰 글씨로 보여주고 있지 않다. 이는 보통 일반적으로 YouTube 청취자들이 많이 사용하는 단어로 중요성이 높지 않음을 알 수 있다. [그림 3]의 워드클라우드를 볼 때 시청자들이 조명섭 트로트 가수를 최고의 칭찬과 사랑, 파이팅의 응원, 그리고 조명섭 가수의 트로트 노래에 감동, 감사, 행복한 마음, 건강의 단어가 주류를 이루고 있다. 조명섭 트로트 가수는 노래를 통해 사람들에게 감동과 행복을 주어 마음에 감성을 불러 일으키는 천상의 목리소를 가지고 있으며, 이로 인해 이런 가수에 대한 보호해지 지키기 위한 칭찬, 사랑, 응원이 나타나고 있다.

이러한 빈도분석은 직관적 의미를 파악하는데 도움을 주지만 전체를 요약하거나 단어와 단어간 연계성, 단어와 문서간 연계성 등 깊이 있는 분석은 어렵다. 이

표 5. 빈도와 TF-IDF값에 의한 상위 100개 단어

		%		%		%		%	
	423	12.68							
가	212	6.35	32	0.96	18	0.54	13	0.39	
	172	5.16	32	0.96	17	0.51	13	0.39	
	111	3.33	31	0.93	17	0.51	13	0.39	
	93	2.79	30	0.9	17	0.51	13	0.39	
	92	2.76	30	0.9	16	0.48	13	0.39	
	82	2.46	28	0.84	16	0.48	12	0.36	
	75	2.25	27	0.81	16	0.48	12	0.36	
가	69	2.07	26	0.78	16	0.48	12	0.36	
	60	1.8	24	0.72	16	0.48	12	0.36	
	60	1.8	24	0.72	16	0.48	12	0.36	
	59	1.77	23	0.69	가	15	0.45	12	0.36
	57	1.71	22	0.66		15	0.45	11	0.33
	56	1.68	22	0.66		15	0.45	11	0.33
	54	1.62	22	0.66		15	0.45	11	0.33
	48	1.44	21	0.63		15	0.45	11	0.33
	46	1.38	21	0.63	가	15	0.45	11	0.33
	46	1.38	21	0.63		14	0.42	11	0.33
	45	1.35	21	0.63		14	0.42	11	0.33
	40	1.2	20	0.6		14	0.42	11	0.33
	36	1.08	20	0.6		14	0.42	10	0.3
	35	1.05	20	0.6		14	0.42	10	0.3
	33	0.99	19	0.57		14	0.42	10	0.3
	33	0.99	18	0.54		14	0.42	10	0.3
	33	0.99	18	0.54		14	0.42	10	0.3



그림 3. 단어의 IF-IDF 값에 의한 워드클라우드

러한 부분을 보완하기 위한 방법이 토픽 모델링과 텍스트네트워크 분석이다.

2. 토픽모델링

본 연구의 토픽모델링 분석은 토픽간 충분히 독특하고, 해석하기 쉽고, 의미 있는 결과를 제공하는 6개 토픽을 선정하였다. LDA의 확률적 접근 방식은 토픽에서 단어가 중복되는 결과를 고려하여 본 연구는 상위단어에 중복을 최소화하여 토픽이 동영상 시청자의견에서

실제로 서로 다른 속성을 반영하도록 보장하는 것을 목표로 했다. 3장의 토픽모델링 방법에 따라 Netminer 4.0을 사용하여 토픽에 대한 다음과 같은 결과를 얻었다. 선행연구에서 검토한 바와 같이 YouTube를 통해 비디오를 시청하는 동기는 쾌락적 충족, 사회적 충족 그리고 실용적 충족을 위해서 입을 알 수 있었다. 첫 번째는 감사, 응원, 사랑, 행복, 눈물과 힐링 등으로 이루어진 즐거움을 나타내는 쾌락적 충족이다. 다음은 할머니, 파이팅, 대한민국 등 사회적 작용, 최고, 선행, 건강, 마음, 모습 등의 사회적 충족을 나타내고 있다. 마지막으로 실용적 충족을 나타내는 단어는 트롯트, 가수, 무대, 조명, 목소리, 노래, 천상, 행복, 소리 등이다. 실용적 충족은 토픽 1과 토픽 6으로 각각 185회, 195회 문서에 나타났으며, 사회적 충족은 토픽 2와 4로 각각 154회, 179회가 문서에 등장했다. 마지막 쾌락적 충족은 토픽 3과 5로 각각 190회, 213회가 사용자들이 남긴 문서(의견)에 등장했다. 사용자들이 남긴 의견에서 가장 많이 등장하는 토픽은 쾌락적 충족으로 403회 등장했다. 다음으로는 사회적 충족을 나타내는 토픽이 380회, 뒤이어, 실용적 충족을 나타내는 토픽이 333회

로 나타났다. 따라서 사용자들이 YouTube트로트 동영상을 시청하는 가장 큰 동기는 쾌락적 충족임을 알 수 있다. 다음 [표 6]은 토픽별 핵심키워드와 토픽이 등장하는 문서 수를 보여 주고 있다.

표 6. 각 토픽별 핵심키워드와 사용된 문서수

토픽	핵심 키워드 (상위 5개)	사용된 문서 (의견) 수
1 (실용적 충족-1)	가수, 트로트, 사람, 무대, 조명	185
2 (사회적 충족-1)	파이팅, 대한민국, 군대, 할머니, 처음	154
3 (쾌락적 충족-1)	감사, 응원, 사랑, 영상, 행복	190
4 (사회적 충족-2)	최고, 선생, 건강, 마음, 모습	179
5 (쾌락적 충족-2)	노래, 가요, 마음, 눈물, 힐링	213
6 (실용적 충족-2)	목소리, 노래, 천상, 행복, 소리	195

표 7. 토픽별 단어 출현 확률분포

	1		2		3		4		5		6	
	가	가	가	가	가	가	가	가	가	가	가	
1	가	0.2144	0.0632	0.0854	0.1147	0.3085	0.1750					
2		0.0627	0.0619	0.0951	0.0498	가	0.0703	0.1246				
3		0.0375	0.0467	0.0520	0.0452	0.0160	0.0329					
4		0.0211	0.0422	0.0408	0.0408	0.0124	0.0238					
5		0.0162	0.0231	0.0374	0.0338	0.0092	0.0155					
6		0.0134	0.0217	0.0198	0.0290	0.0090	0.0140					
7		0.0089	0.0201	0.0172	0.0231	0.0082	0.0104					
8		0.0078	0.0174	0.0153	0.0207	0.0077	0.0098					
9		0.0077	0.0164	0.0126	0.0149	0.0077	0.0090					
10	가	0.0070	가	0.0158	0.0124	0.0139	0.0077	0.0082				

표 8. 토픽별 문서(의견) 출현 확률분포

	1		2		3		4		5		6	
	Doc	Doc	Doc	Doc	Doc	Doc	Doc	Doc	Doc	Doc	Doc	
1	Doc302	0.2277	Doc860	0.3754	Doc33	0.2325	Doc3	0.2435	Doc246	0.2536	Doc343	0.2230
2	Doc594	0.2233	Doc164	0.3203	Doc1110	0.2318	Doc740	0.2428	Doc598	0.2321	Doc568	0.2150
3	Doc1014	0.2213	Doc983	0.2858	Doc933	0.2282	Doc197	0.2297	Doc160	0.2286	Doc677	0.2124
4	Doc711	0.2207	Doc872	0.2494	Doc68	0.2270	Doc867	0.2292	Doc173	0.2208	Doc274	0.2074
5	Doc260	0.2187	Doc400	0.2421	Doc367	0.2178	Doc310	0.2265	Doc429	0.2091	Doc954	0.2040
6	Doc566	0.2126	Doc116	0.2226	Doc1112	0.2138	Doc102	0.2259	Doc473	0.2072	Doc588	0.2036
7	Doc455	0.2099	Doc245	0.2166	Doc986	0.2133	Doc1	0.2256	Doc824	0.2032	Doc234	0.2024
8	Doc350	0.2089	Doc1086	0.2164	Doc354	0.2131	Doc380	0.2245	Doc1027	0.1979	Doc437	0.2014
9	Doc704	0.2076	Doc277	0.2086	Doc948	0.2126	Doc987	0.2197	Doc135	0.1974	Doc71	0.2014
10	Doc942	0.2045	Doc1007	0.2064	Doc742	0.2094	Doc149	0.2178	Doc233	0.1970	Doc1024	0.1999

LDA 알고리즘에 의한 토픽 모델링은 모든 단어는 한 토픽에 할당되고 모든 토픽은 한 문서에 할당된다. 따라서 모든 토픽별 단어의 확률의 합은 1이 되어야 하

고, 문서별 토픽의 확률의 합은 1이 되어야 한다. [표 7]은 토픽별 단어가 출현하는 확률분포, [표 8]은 문서에 토픽이 나타나는 확률분포를 보여주고 있다. 이 표들은 각 토픽별 핵심키워드와 등장 문서수 상위 10위에 대해 자세히 보여주고 있다.

3. 텍스트 네트워크 분석

단어의 공동 출현에 의한 빈도를 사용하여 단어 문서 간의 네트워크를 분석하였다. 단어문서네트워크는 TF-IDF 값을 계산하여 이를 가중치로 사용하여 연결정도중심성, 매개중심성, 아이젠벡터중심성을 분석한 결과를 [표 9][표 10]에 보여주고 있다.

연결정도중심성은 단어가 얼마나 많은 연결관계를 가지고 있고 중심이 되느냐를 나타낸 정도를 말하는데, 많은 연결을 가질수록 다른 단어에 미치는 영향정도가 커지기 때문에 상대적으로 우위에 있다고 볼 수 있다 [52]. 주요 단어 연결정도중심성을 확인한 결과, '응원', '감사', '화이팅', '최고' 등의 단어가 높음을 확인하였다. 이러한 단어들이 동영상 사용자의견들 중에서 대다수의 사용자들이 사용하여 사용자들간 많은 연결이 되어 있음을 알 수 있다. 대다수의 사용자들이 조명섭 트로트 가수의 노래에 대해 감사하고, 응원하고, 파이팅을 하고 있음을 보이고 있다.

아이젠벡터중심성은 고유벡터중심성 또는 위세중심성이라 하며, 연결된 단어의 개수뿐만 아니라, 연결된 관계가 얼마나 중요하지도 함께 고려함으로써 연결중심성의 개념을 확장한 개념이다. 그러므로 아이젠벡터는 네트워크 내에서 가장 영향력 있는 중심노드를 찾는 데 유용한 지표가 된다[53]. 가장 영향력 있는 중심노드를 찾기 위해 주요 단어의 아이젠벡터중심성을 계산한 결과, '사랑', '마음', '감사' 등의 단어가 높음을 보였다. 이러한 단어들이 동영상 사용자 의견들 중에서 높은 영향력을 보이고 있음을 알 수 있다. 대다수의 사용자들이 조명섭 트로트가수의 노래에 대해 사랑과 감사의 마음을 보이고 있음을 알 수 있다.

마지막으로 단어와 단어를 연결 해주는 정도를 측정하는 매개중심성 측정결과는 '감사', '응원', '화이팅' 등이 단어가 높은 값을 보이고 있다. 이 단어들은 연결정도중심성과 유사하게 나타났다. 이 단어들은 사용자들

이 자주 의견에 사용한 단어로 전체 네트워크 속에서 단어와 단어를 가장 많이 매개해 주는 역할을 하는 것을 나타낸다. 따라서 연결정도중심성은 네트워크의 지역적인 중심성을 의미하며, 매개중심성은 네트워크의 글로벌중심성을 의미함을 알 수 있다.

표 9. 단어 네트워크의 중심성

1		0.0797		0.0683	0.2195	
2		0.0744		0.0672	0.2144	
3		0.0717		0.0607	0.2019	
4		0.0717		0.0582	가	0.1881
5		0.0529		0.0580		0.1872
6		0.0466		0.0524		0.1865
7	가	0.0466		0.0510		0.1807
8		0.0457	가	0.0436		0.1663
9		0.0457		0.0415		0.1482
10		0.0430		0.0404		0.1478

단어 문서 네트워크는 TF-IDF 값을 계산하여 이를 가중치로 사용하여 문서네트워크의 중심성 분석의 결과를 [표 10]에서 보여주고 있다. 문서 연결정도중심성은 문서가 얼마나 많은 연결관계를 가지고 있고 중심이 되느냐를 나타낸 정도를 말하는데, 많은 연결을 가지면 가질수록 다른 문서에 미치는 영향정도가 커지기 때문에 상대적으로 우위에 있다고 볼 수 있다[52]. 주요 문서 연결정도중심성을 확인한 결과, '문서991', '문서1', '문서33' 등의 문서가 높음을 확인하였다. 이러한 문서들이 동영상 사용자의견들 중에서 대다수의 사용자들이 사용하여 사용자들간 많은 연결이 되어 있음을 알 수 있다. '문서991'은 '김경숙'이란 사용자가 '멋지다 명섭군, 최고다 명섭군, 여러분 명섭군 많이 응원하구 홍보합시다, 보물입니다 보물'이란 의견이다. '문서33'은 '임수연'이란 사용자가 의견을 남긴 "넌넌 이쁜사람 조명섭 가수 (내 마음의 막내아들 삼음) 할 수 있으니깐 도전하는 것입니다 ..(중략)..이젠 대학생도 되었으니 더 발전 있는 삶이 될 것을 기대해보면서 값진 나날이 되길 기도하고 바랄게요. 막내아들 사랑해요 파이팅"이다. 이 사용자의 의견이 문서 네트워크에서 다른 사용자들의 의견을 선도하고 있음을 의미한다.

아이젠벡터중심성은 연결정도중심성의 개념을 확장한 개념으로 네트워크 내에서 가장 영향력 있는 중심노드를 찾는데 유용한 지표가 된다[53]. 가장 영향력 있는

중심노드를 찾기 위해 주요 문서의 아이젠벡터중심성을 계산한 결과, '문서1', '문서33', '문서824' 등의 문서가 높음을 보였다. 이러한 문서들은 연결정도중심성과 유사하게 보이고 있다. '문서1'은 '김용배'란 사용자가 남긴 의견은 "Kbs 아들 조명섭군을 응원합니다. 어릴 때부터 몸의 아픔, 가난의 고통을 현인 선생님의 노래로 위로 받고 견디고 이겨낸 조명섭군. 우울증까지 이기고 가수의 꿈을 ..(중략)..kbs 최고 파이팅 조명섭군 최고 파이팅"이다. 이 의견이 전체 문서네트워크에서 다른 사용자들의 의견에 영향력을 가장 높게 행사하고 있음을 알 수 있다.

마지막으로 문서와 문서를 연결 해주는 정도를 측정하는 매개중심성 측정결과는 '문서1', '문서991', '문서81' 등의 문서가 높은 값을 보이고 있다. 이 문서들은 연결정도중심성과 유사하게 나타났다. 이 문서의 사용자들이 다른 사용자들 의견에 많은 영향을 주는 문서들이면서 전체 네트워크 속에서 문서와 문서를 가장 많이 매개해 주는 역할을 하는 것을 나타낸다.

표 10. 문서네트워크의 중심성

	()ID	()ID	()ID	()ID	()ID	
1	Document991	0.0281	Document1	0.0552	Document1	0.2800
2	Document1	0.0273	Document991	0.0467	Document33	0.2633
3	Document33	0.0266	Document81	0.0334	Document824	0.2297
4	Document81	0.0258	Document824	0.0331	Document991	0.2258
5	Document824	0.0250	Document33	0.0307	Document81	0.2171
6	Document987	0.0228	Document986	0.0249	Document986	0.1875
7	Document986	0.0205	Document987	0.0247	Document860	0.1413
8	Document860	0.0182	Document1033	0.0192	Document1068	0.1325
9	Document546	0.0167	Document546	0.0188	Document987	0.1297
10	Document550	0.0152	Document860	0.0184	Document983	0.1269

V. 결론 및 한계점

본 연구는 텍스트분석과 토픽모델링을 사용하여 YouTube 동영상 사용자의견이 사용과 충족 이론을 쾌락적 충족, 사회적 충족, 그리고 실용적 충족에 대해서 빈도분석과 토픽모델링을 통해 측정하였다. YouTube KBS 한국방송 채널 중 트로트 가수 조명섭 동영상을 사용자들이 시청하는 이유는 첫 번째로 높은

빈도를 보이는 것이 쾌락적 충족을 위해서였다. 다음 순으로 사회적 충족과 실용적 충족으로 나타났다.

단어-문서 네트워크 분석에서 연결정도중심성은 '응원', '감사', '화이팅', '최고' 등이 높게 나타났고, 매개중심성은 '감사', '응원', '화이팅' 등의 단어가 높게 나타나 연결정도중심성과 유사함을 보였다. 아이겐벡터중심성은 '사랑', '마음', '감사' 등의 단어가 높게 나타나 사용자들의 의견들에 가장 영향력이 높은 단어들을 알 수 있다. 이는 YouTube의 트로트 가수 조명섭 동영상 시청자들 중 대다수가 동영상에 대해 사랑과 감사의 마음을 보이고 있음을 알 수 있다. 위 세 중심성 분석결과는 동영상을 시청하는 동기로 사용충족 이론의 쾌락적 충족과 사회적 충족관련 단어들이 높은 값을 보이고 있다.

본 연구는 지금까지의 설문조사 기반 구조방정식 모형에 의한 실증분석을 탈피하여 질적분석을 자동화한 텍스트마이닝을 기반으로 한 빈도분석과 토픽모델링을 사용하여 YouTube 동영상을 사용하는 동기를 사용과 충족 이론에 의해 밝혀냈다는 데 연구방법론적 함의를 보였다. 두 번째 학문적 함의는 빈도분석이나 토픽모델링으로 설명할 수 없는 단어와 문서간 관계를 텍스트네트워크 분석에 의해 보완을 해 줌으로써 질적방법을 좀 더 견고하게 하였다. 실무적 함의는 트로트 가수 조명섭의 YouTube 동영상 방문자 수를 높이고 소셜미디어 상 홍보를 극대화하기 위해서 영향력 있는 단어들에 해쉬태그 사용하는 것이다.

이와 같이 연구결과에 대한 시사점을 제공해 줌에도 불구하고 본 연구의 몇 가지 극복해야 할 점이 있다. 우선 사용자의 동영상에 대한 거짓 의견이나 의도적인 악성 의견을 본 연구로는 찾아내기가 어렵다. 좀 더 정교한 텍스트마이닝 기술의 발달로 가능성을 기대해 본다. 둘째로는 사용충족 이론의 사용동기 차원인 쾌락적 충족, 사회적 충족, 실용적 충족과 관련된 핵심단어 선정 시 전문가의 판단을 필요로 한다는 것이다. 즉 연구자의 주관적 판단이 개입할 수밖에 없다는 것이다. 마지막으로 텍스트 전처리 중 특수문자나 이모티를 제거함으로써 의견 작성자의 의도를 해칠 수 있다. 이런 부분까지 반영할 수 있는 텍스트마이닝을 기대해 본다.

참고 문헌

- [1] R. Briones, X. Nan, K. Madden, and L. Waks, "When vaccines go viral," *Health Com*, Vol.27, pp.478-485, 2012.
- [2] S. P. Lewis, N. L. Heath, M. J. Sornberger, and A. E. Arbuthnott, "Helpful or harmful? An examination of viewers' responses to non-suicidal self-injury videos on YouTube," *Journal of Adolescent Health*, Vol.51, pp.380-385, 2012.
- [3] R. Jaspal, A. Turner, and B. Nerlich, "Fracking on YouTube: Exploring risks, benefits and human values," *Environmental Values*, Vol.23, pp.501-527, 2014.
- [4] M. Thelwall, P. Sud, and F. Vis, "Commenting on YouTube videos: From guatemalan rock to El Big Bang," *Journal of the American Society for Information Science and Technology*, Vol.63, pp.616-629, 2012.
- [5] T. Desai, A. Shariff, V. Dhingra, D. Minhas, M. Eure, and M. Kats, "Is content really king? An analysis of the public's response to medical videos on YouTube," *PLoS One*, Vol.8, No.12, e82469, 2013.
- [6] A. Dunne, M. Lawlor, and J. Rowley, "Young people's use of social networking sites a uses and gratifications perspective," *Journal of Research in Interactive Marketing*, Vol.4, No.1, pp.46-58, 2010.
- [7] M. Hossain, "Effects of uses and gratifications on social media use," *Research Review*, Vol.3, No.1, pp.16-28, 2019.
- [8] I. M. Al-Jabri, M. S. Sohail, and N. O. Ndubisi, "Understanding the usage of global social networking sites by Arabs the lens of uses and gratifications theory," *Journal of Service Mgt*, Vol.26, pp.662-680, 2015.
- [9] C. H. Hsiao, J. J. Chang, and K. Y. Tang, "Exploring the influential factors in continuance usage of mobile social apps: customer value perspectives," *Informatics*,

- Vol.33, No.2, pp.342-355, 2015.
- [10] C. Gan, "Understanding WeChat users' liking behavior: An empirical study in China," *Computer Human Behavior*, Vol.68, pp.30-39, 2017.
- [11] X. Li, Y. Wang, A. Zhang, C. Li, J. Chi, and J. Ouyang, "Filtering out the noise in short text topic modeling," *Information Sciences*, Vol.456, pp.83-96, 2018.
- [12] Katz, Elihu, Jay G. Blumler, and Michael Gurevitch, "Uses and Gratifications Res," *Public Opinion Quarterly* 4th, Vol.37, pp.509-523, JSTOR. Web. 14 Oct. 2011.
- [13] C. M. K. Cheung and M. K. O. Lee, "A theoretical model of intentional social action in online social networks," *Decision Support Systems*, Vol.449, pp.24-30, 2010.
- [14] D. Cao, R. Ji, D. Lin, and S. Li, "Visual sentiment topic model based image sentiment analysis," *Multimed Tools Application*, Vol.2016, No.75, pp.8955-8968, 2014.
- [15] A. Dhir and C. C. Tsai, "Understanding the relationship between intensity and gratifications of Facebook use among adolescents and young adults," *Telematics Inf.* Vol.34, No.4, pp.350-364, 2017.
- [16] J. Balakrishnan and M. Griffiths, "Social media addiction: What is the role of content in YouTube?," *Journal of Behavioral Addictions*, Vol.6, No.3, pp.364-377, 2017.
- [17] Wedel & Kannan, "Marketing Analytics For Data-Rich Environments," *Journal Of Marketing: AMA/MSI Special Issue*, Vol.80, pp.97-121, 2016.
- [18] G. Brooks, A. Heffner, and D. Henderson, "A SWOTanalysis of competitive knowledge from social media for a small start-up business," *Review of Business Information Systems*, Vol.8, No.1, pp.23-34, 2014.
- [19] M. Sigala and K. Chalkiti, "Knowledge management social media, and employee creativity," *International Journal of Hospitality Mgt*, Vol.45, pp.44-58, 2015.
- [20] I. Lee, "Social media analytics for enterprises: Typology, and processes," *Business*, Vol.61, pp.199-210, 2018.
- [21] P. Grant, E. Botha, and J. Kietzmann, "Branded mobs: Moving toward a deeper understanding of consumers responses to video advertising," *Journal of Advertising*, Vol.15, No.1, pp.28-42, 2015.
- [22] Y. Rashied and M. Zeeshan, "Customer Attitude towards Online Ads of Smartphone Brands," *J of Mgt Sciences*, Vol.5, No.2, pp.40-64, 2018.
- [23] M. Thelwall, "Social media analytics for YouTube comments: and limitations," *International Journal of Social Res Methodology*, Vol.21, No.3, pp.303-316, 2018.
- [24] L. Parabhoi and P. Saha, "Sentiment Analysis of YouTube Comments on Koha Open Source Software Videos," *International Journal of Library and Information Studies*, Vol.8, No.4, pp.96-102, 2018.
- [25] T. Schmiedel, Oliver Muller, and Jan vom Brocke, "Topic Modeling as a Strategy of Inquiry in Organizational Research: A Tutorial With an Application Example on Organizational Culture," *Organizational Research Methods*, Vol.22, No.4, pp.941-968, 2018.
- [26] A. Bittermann and A. Fischer, "How to Identify Hot Topics in Psychology Using Topic Modeling," *Zeitschrift für Psychologie*, Vol.226, No.1, pp.3-13, 2018.
- [27] E. Katz, J. G. Blumler, and M. Gurevitch, "The Uses of Mass Communications: Current Perspectives on Gratifications Research," Newcastle, UK, pp.19-32, 1974.
- [28] B. Jeong, J. Yoon, and L. M. Lee, "Social media mining for product planning: A product opportunity mining approach based on modeling and sentiment analysis," *International Journal of Information Mgt*, 2019.

- [29] A. C. ros, S. Moro, and P. Rita, "Sentiment classification of consumer-generated online reviews using topic modeling," *Journal of Hospital Market Management*, Vol.26, No.7, pp.675-693, 2017.
- [30] Leem, "Using Topic Modeling to Analyze on YouTube," *Logos Mgt Review*, Vol.17, No.4, pp.67-82, 2019.
- [31] D. T. Le, C. T. Nguyen, Q. T. H. Coltech, X. H. Phan, and S. Horiguchi, Matching and ranking with hidden topics towards online contextual advertising. 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2008.
- [32] A. Schieber, A. Hilbert, S. Sommer, and K. Heinrich, "Analyzing customer sentiments in microblogs based approach for Twitter datasets. In: Proceedings of the Seventeenth Americas Conference on Information Systems," Detroit, Michigan August 4th-7th, 2011.
- [33] S. Tirunillai and G. J. Tellis, "Mining marketing meaning from online chatter: Strategic brand analysis of latent Dirichlet allocation," *Journal of Marketing Research*, Vol.51, No.4, pp.463-479, 2014.
- [34] T. Do and D. Gatica-Perez, "By their apps you shall understand them: mining large-scale patterns of mobile phone usage," In: Proceedings of the 9th international conference on MUM'10, 1-3. 2010.
- [35] N. Schroder, A. Falke, H. Hruschka, and T. Reutterer, "Analyzing browsing and purchasing across multiple websites based on latent Dirichlet allocation," ALLDATA, ISBN: 978-1-61208-552-4. 2017.
- [36] M. Trusov, L. Ma, and Z. Jamal, "Crumbs of the cookie: User profiling in customer-base analysis and behavioral targeting," *Marketing Science*, Vol.35, No.3, pp.405-426, 2016.
- [37] M. Reisenbichler and T. Reutterer, "Topic modeling in marketing: advances and research opportunities," *Journal of Business*, Vol.89, pp.327-356, 2019.
- [38] M. Hossain, M. Kim, and N. Jahan, "Can "Liking" Behavior Lead to Usage Intention on Facebook?," *Uses and Gratification Theory Perspective*, Vol.1166, pp.1-13, 2019.
- [39] A. Susarla, A. Barua, and A. B. Whinston, "Understanding the service component of application service Empirical analysis of satisfaction with ASP services," *MIS Quarterly*, Vol.27, No.1, pp.91-123, 2003.
- [40] Khan, *Creating Value with Social Media Analytics: managing, and Mining Social Media Text, Networks, Action, Locations. Apps, Hyperlinks*, 2018.
- [41] R. Robinson, T. T. Goh, and R. Zhang, "Textual factors in online product reviews: a foundation for a more influential approach to opinion mining," *Electronic Commerce Research*, Vol.12, No.3, pp.301-330, 2012.
- [42] S. Xiong, K. Wang, D. Ji, and B. Wang, "A short text sentiment-topic model for product reviews," *Neurocomputing*, Vol.297, pp.94-102, 2018.
- [43] M. Westerlund, S. Leminen, and M. Rajahonka, "A Topic Modelling Analysis of Living Res," *Technology Innovation Mgt Review*, Vol.8, No.7, pp.40-51, 2018.
- [44] S. Lei and R. Law, "Content Analysis of TripAdvisor Reviews on Restaurants: A Case Study of Macau," *Journal of Tourism*, Vol.16, No.1, pp.17-28, 2015.
- [45] D. Maier, A. Waldherr, P. Miltner, G. Wiedemann, A. Niekler, A. Keinert, B. Pfetsch, G. Heyer, U. Reber, and T. Häussler, *Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology*, *Journal of Business*, 2018.
- [46] Griffiths and Steyvers, Finding Scientific Topics, Colloquium of the National Academy of Sciences, Mapping Knowledge, held May 9-11, 2003, at Beckman Center of the National Academies of Sciences and Engineering in Irvine, CA, 2004.
- [47] M. Steyvers and T. Griffiths, "Probabilistic

topic models. In Handbook of Latent Semantic Analysis; Lawrence Associates," Inc.: Mahwah, NJ, USA, 2007.

- [48] Y. Lu, M. Qiaozhu, M. ChengXiang, "Investigating task performance of probabilistic topic models," *Inf. Retrieval* Vol.14, No.2, pp.178-203, 2011.
- [49] W. Zhao, J. Chen, and W. Zen, *Best practices in building topic models with LDA for mining regulatory textual documents*, CDER 9TH NOV. 2015.
- [50] M. Naili, A. Chaibi, and H. Ghezala, "Arabic topic identification based on empirical studies of topic models. *Revue Africaine de la Recherche en Informatique*," *ARIMA Journal*, Vol.27, pp.45-59, 2017.
- [51] D. Delen and M. D. Crossland, "Seeding the survey and analysis of research with text mining," *Expert Systems with Applications*, Vol.34, No.3, pp.1707-1720, 2008.
- [52] H. J. Ban and J. K. Jun, "A study on the semantic network analysis of luxury hotel and business hotel through the big data," *Culinary Science & Hospitality Research*, Vol.25, No.1, pp.18-28, 2019.
- [53] P. Bonacich, "Some unique properties of eigenvector centrality," *Social Networks*, Vol.29, pp.555-564, 2007.

저 자 소 개

홍 한 국(Han-Kuk Hong)

정회원



- 1988년 : 고려대 통계학과(경제학사)
- 1990년 : KAIST 산업공학과(공학 석사)
- 2000년 : KAIST 경영공학과(공학 박사)
- 1990년 ~ 1996년 : 삼성경제연구소 선임연구원
- 1997년 ~ 1998년 : 삼성화재 과장
- 2000년 ~ 현재 : 동의대학교 경영정보학과 교수
 <관심분야> AI, 빅데이터, SNA분석

임 병 학(Byunghak Leem)

정회원



- 1987년 : 고려대 산업공학(공학사)
- 1996년 : KAIST 경영공학과(공학 석사)
- 2002년 : University of Texas at Arlington(공학박사)
- 1989년 3월 ~ 1997년 : 삼성전자 과장
- 2002년 ~ 현재 : 부산외국어대학교 경영학부 교수
 <관심분야> 생산관리, 공급망관리, 성과관리

김 삼 문(Sam-Moon Kim)

정회원



- 1990년 : 동명대 경영정보학과(학사)
- 2011년 : 동의대 경영정보학과(석사)
- 2014년 : 동의대 경영정보학과(박사)
- 1999년 ~ 2015년 : (주)bullsbroadband, SK브로드밴드
- 2015년 ~ 현재 : 동의대학교 응용 S/W공학과 교수
- <관심분야> : DEA, 데이터마이닝, CRM, SCM