

온라인 동영상 플랫폼의 알고리즘은 어떤 연관 비디오를 추천하는가: 유튜브의 K POP 뮤직비디오를 중심으로

What Do The Algorithms of The Online Video Platform Recommend: Focusing on Youtube K-pop Music Video

이영주*, 이창환**

서울과학기술대 IT정책전문대학원*, 서울과학기술대 데이터사이언스학과 일반대학원**

Yeong-Ju Lee(roselee@seoultech.ac.kr)*, Chang-Hwan Lee(lch1181@ds.seoultech.ac.kr)**

요약

본 연구는 온라인 동영상 플랫폼에 적용되는 추천 알고리즘을 이해하고자 유튜브에서 K-pop 뮤직비디오의 콘텐츠 특성과 재생 시 추천되는 연관 비디오(related video)의 관계를 규명하고 네트워크 분석을 통해 어떤 비디오가 연관 비디오로 추천되는지 살펴보았다. 분석 결과, K-pop 재생 시 비디오의 좋아요 수가 추천 순위에 영향을 주었으며 대부분 같은 채널에 속하거나 동일한 기획사에서 제작한 비디오가 연관 비디오로 추천되었다. 그리고 연관 비디오의 네트워크 분석 결과, K-pop 뮤직비디오의 네트워크가 강하게 형성되어 있으며 연관 비디오의 네트워크 분석에서 BTS의 뮤직비디오가 중심성이 높게 나타났다. 이러한 연구결과는 K-pop 간의 네트워크가 강하기 때문에 K-pop을 검색 쿼리로 입력해서 비디오를 시청할 때는 연속적으로 K-pop을 즐길 수 있지만, 반대로 다른 장르의 비디오를 시청할 때는 K-pop이 연관 비디오로 추천되지 못할 수 있음을 의미한다.

■ 중심어 : | 콘텐츠 추천 시스템 | 알고리즘 | 온라인 동영상 플랫폼 | 연관 비디오 |

Abstract

In order to understand the recommendation algorithm applied to the online video platform, this study examines the relationship between the content characteristics of K-pop music videos and related videos recommended for playback on YouTube, and analyses which videos are recommended as related videos through network analysis. As a result, the more liked videos, the higher recommendation ranking and most of the videos belonging to the same channel or produced by the same agency were recommended as related videos. As a result of the network analysis of the related video, the network of K-pop music video is strongly formed, and the BTS music video is highly centralized in the network analysis of the related video. These results suggest that the network between K-pops is strong, so when you enter K-pop as a search query and watch videos, you can enjoy K-pop continuously. But when watching other genres of video, K-pop may not be recommended as a related video.

■ keyword : | Contents Recommendation System | Algorithm | Online Video Platform | Related Video |

* 이 연구는 서울과학기술대학교 교내연구비의 지원으로 수행되었습니다.

접수일자 : 2020년 02월 14일

수정일자 : 2020년 03월 23일

심사완료일 : 2020년 03월 31일

교신저자 : 이영주, e-mail : roselee@seoultech.ac.kr

I. 서론

최근 IT 기업들의 데이터 수집 및 분석 역량이 증가함에 따라 가입자 기반이 큰 온라인 플랫폼의 영향력이 더욱 강화되고 있다. 온라인 플랫폼들은 서비스의 이용 과정에서 발생하는 데이터를 목표 수용자의 니즈나 선호도에 맞춰 비즈니스 모델을 정교화시키면서[1] 소비자 만족을 높이고 있다. 플랫폼이 개인화된 서비스를 제공함으로써 이용자들이 정보 검색에 소요되는 시간을 줄이고 적합한 정보를 쉽게 찾아낼 수 있는 것도 어느 정도 사실이다[2]. 그런데 서비스 만족도가 커질수록 이용자의 고착화는 커지고[3] 플랫폼의 수익도 더불어 증가한다.

선호 장르의 추종(genre-following) 경향이 두드러진 미디어 산업에서도[4], 데이터 기반의 콘텐츠 추천 엔진은 핵심적인 역할을 한다. 막대한 양의 콘텐츠가 있는 온라인 동영상 플랫폼에서 이용자들은 검색 경험을 최적화해야 하는 어려움에 직면하게 된다. 하지만 이용자들은 자신이 원하는 비디오를 찾기 위해 많은 시간을 소비하는 대신 특정 채널을 구독하거나 선호도를 표시하며 자기만의 히스토리를 만들어간다[5]. 그리고 동영상 플랫폼은 이용자의 선호도와 콘텐츠 이용 패턴을 분석하여 추천 시스템에 적용한다. 특히 넷플릭스와 유튜브는 알고리즘을 통해 콘텐츠 큐레이션을 고도화시키고 있다[5][6]. 넷플릭스는 이용자들의 선호 장르, 스토리, 시청 패턴 등 복합적 요인에 기반한 추천 알고리즘을 개발하여 이용자들의 콘텐츠 이용을 늘림과 동시에[6] 신작 구매 비용을 줄여 수익구조를 개선하고 있다[7]. 유튜브 역시 머신 러닝과 딥러닝을 통해 이용자들에게 연관 동영상(related video)을 추천하고 있다[5]. 유튜브에서 연관 동영상을 추천한 이후 이용자들의 시청시간은 20배 증가했고, 그 중 70%는 추천된 동영상 중에서 소비되고 있다[8]. 양면 시장의 속성을 지니는 유튜브는 이용자가 동영상을 많이 볼수록 광고에 더 노출되면서 수익도 증가하고 있다[9].

이렇듯 추천 알고리즘이 콘텐츠 이용에 막대한 영향을 미치기 때문에 온라인 동영상 플랫폼의 추천 알고리즘에 대한 이해는 선택이 아닌 필수가 되고 있다. 학술적/실용적 차원에서 넷플릭스나 유튜브의 추천 알고리

즘의 원리를 추정해보려는 시도가 이루어지고 있다. 그러나 주로 공학 분야에서 알고리즘의 최적화를 위한 요인을 찾는 연구가 이루어졌으며, 추천되는 연관 동영상의 특성과 비디오간 네트워크에 대한 고찰은 상대적으로 미흡한 상황이다. 그나마 알고리즘이 주기적으로 바뀌고 있어 연구결과를 일반화하여 설명하거나 예측하기 어려운 한계가 있다.

이에 본 연구에서는 온라인에서 가장 많은 동영상이 유통되는 유튜브에서 어떤 연관 동영상이 추천되는지 K-pop의 뮤직비디오를 중심으로 알아보고자 한다. K-pop은 유튜브에서 채널 구독자와 조회 수가 가장 높은 콘텐츠 유형으로서 K-Pop을 시청할 때 어떤 연관 동영상이 추천되는가에 따라 K-pop의 확산에 중요한 영향을 미칠 수 있다. 이에 K-pop 아티스트 상위 15개 그룹을 중심으로 이용자의 선호도가 반영된 콘텐츠 특성이 연관 동영상의 추천 순위에 어떤 영향을 주는지 알아보고자 한다. 그동안 유튜브에서 K-pop의 확산에 대한 연구는 많이 이루어졌지만, 알고리즘 추천으로 이루어지는 연관 비디오에 주목한 연구는 거의 없었다. 유튜브는 알고리즘을 주기적으로 바꾸고 있어서 시기와 연구대상에 따라 그 결과도 달라지고 있다. 이에 본 연구는 머신러닝과 딥러닝을 토대로 하는 알고리즘이 K-pop 뮤직비디오의 연관비디오 추천에 미치는 영향을 추정함으로써 콘텐츠 창작자와 이를 전달하는 서비스 제공자에게 실용적인 함의를 제공할 수 있을 것으로 기대한다. 동시에 알고리즘에 의해 추천되는 동영상이 무엇인지 실증적으로 제시함으로써 정책적 관점에서 온라인 플랫폼에서 사용되는 추천 알고리즘에 대한 논의를 시작할 수 있는 단초를 제공하고자 한다.

II. 관련 문헌 검토

1. 온라인 동영상 플랫폼과 K-Pop

1.1 온라인 동영상 플랫폼 생태계

인터넷이 도입되면서 글로벌 음악 산업은 대전환을 맞게 된다. CD와 같은 앨범 대신 온라인으로 음원을 다운로드 받거나 스트리밍 방식으로 음악을 소비하는 행태가 급증하면서 스포티파이, 애플뮤직, 아마존뮤직,

텐센트뮤직 등의 구독형 플랫폼이 음원 유통을 대체하고 있다. 2030년까지 음악 스트리밍 매출이 500% 증가하고 유료 가입자 계정도 5배 증가할 것으로 전망되고 있다[10].

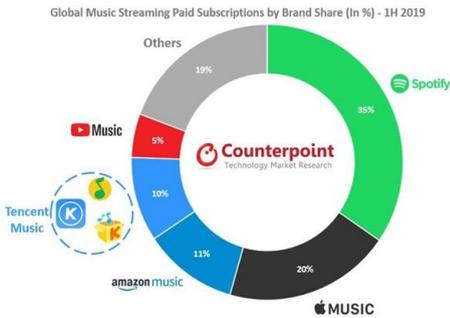


그림 1. 구독형 음악 스트리밍 서비스의 점유율

(출처: Strabase, "2018년 음악 산업에 영향을 미칠 3가지 요인", NewsBrief, 2018)

유튜브는 이러한 음악의 생산과 수용의 변화를 가져오는 데 가장 결정적인 역할을 하고 있다. 2006년 구글에 의해 1조 5천억 원으로 인수된 이후, 유튜브는 동영상, 검색, 쇼핑, 저널리즘 등 모든 장르를 망라하는 동영상 플랫폼으로 올라섰다. 국내 온라인 동영상 플랫폼 이용률이 유튜브가 82.5%로 압도적으로 앞선 가운데, 네이버TV 42%, 페이스북 26%, 인스타그램 17.2%로 추격하고 있지만[11], 유튜브에의 쏠림현상은 고착화되고 있다.

2018년 국내 이용자들이 선호하는 장르는 예능(55.7%), 음악/뮤직비디오(40%) 등 오락 장르가 압도적으로 많다[11]. 그리고 비실시간 영상이 많지만, 2017년부터 라이브 서비스도 제공하면서 스포츠 중계도 네이버TV에 이어 유튜브 이용률이 많아지고 있다.

유튜브가 성장하게 된 것은 무엇보다도 스마트폰의 보급에 따른 미디어 소비 패턴의 변화와 소셜 미디어를 통한 동영상 공유의 확대, 그리고 이용자가 창작한 콘텐츠를 플랫폼에 자유롭게 업로드하고 다수의 크리에이터들이 제작한 콘텐츠를 자유롭게 제공할 수 있는 생태계를 구축했기 때문이다[12]. 이용자들이 음악을 쉽게 접근할 수 있게 되면서 매스업과 샘플링을 통한 창작도 용이해지고 있다[13]. 현재 유튜브는 이용자가 제작한 콘텐츠뿐만 아니라 레거시 미디어의 프로그램도

유통되는 등[14] 새롭고 다양한 포맷과 내용의 콘텐츠 뿐만 아니라 전문 미디어 그룹이 제작하는 고품질의 콘텐츠까지 제공되고 있다[15].

1.2 K-Pop의 글로벌 유통

K-pop은 무료로 이용할 수 있는 동영상 플랫폼 이외에 아이튠즈나 스포티파이와 같은 유료 구독 플랫폼에서도 접근 가능하다[16]. 온라인으로 음원이 소비되는 환경으로 변하면서 한국 엔터테인먼트 기업은 위기가 기회로 맞게 된다. 아티스트의 앨범 판매 수입이나 콘서트 수익은 감소하는 대신 글로벌 시장으로 수용자 접점이 확대되는 기회가 생기고 있다[17]. 특히 다양한 형태의 소셜 미디어를 통해 K-pop 아티스트와 전세계에 퍼진 팬들이 연결되고, 팬들도 커버 영상이나 리액션 영상을 업로드하면서 팬덤을 표출한다[17].

비즈니스 차원에서도 많은 변화가 있는데, 음악을 유통시키는 거래비용은 낮아지고 상품판매 수익, 로얄티, 그리고 온라인 플랫폼에서 분배되는 광고수익 등 새로운 B2B 수익이 증가하였다[18]. 국내 엔터테인먼트 기업은 조회 수에 비례해서 광고수익을 받을 수 있는 유튜브와 네이버와 같은 동영상 플랫폼을 적극적으로 활용하기 시작했고[19], 플랫폼의 특성에 맞게 콘텐츠를 차별화하며 프로모션하였다[20]. 뮤직비디오는 아티스트를 홍보함과 동시에 브랜딩(branding)을 효과적으로 할 수 있어서 적극적으로 활용되었다[21][22]. 방탄소년단의 경우 블로그에는 멤버의 일기 영상을, 유튜브에는 뮤직비디오나 댄스 연습 동영상을, 브이 라이브(V-live)에는 자체 제작 예능물을, 인스타그램에는 멤버의 사진을, 트위터엔 근황을, 페이스북엔 공식 일정과 국내외 언론 자료 등을 게재한다[23].

한편, K-pop은 글로벌 수용자에게 소구력을 높이기 위해 두세 개의 장르가 혼합된 장르의 음악이 많으며[24], 국적, 민족, 문화가 뒤섞인 초국가적 이미지를 내세우며 마케팅을 하고 있다[25].

2. 온라인 동영상플랫폼의 콘텐츠 추천시스템

2.1 알고리즘 기반 추천 서비스의 특성과 유형

대부분의 온라인 동영상 플랫폼은 데이터를 기반으로 수용자의 이용 행태를 분석할 수 있는 알고리즘을

개발하여 지시된 바를 수행하는 인공지능 시스템을 갖추고 있다. 인공지능은 지식을 축적하고 패턴을 인식하고 의사결정을 내리는 처리 속도가 매우 빠르며, 이용자 개인의 정치적 성향, 소득, 구매 습관 등을 예측하는 통찰력도 갖춰나가고 있다[26].

온라인 동영상 플랫폼이 도입하고 있는 추천 시스템은 사용자가 선호할 만한 아이템을 추천함으로써 사용자에게 적합한 콘텐츠를 제공한다. 콘텐츠 추천 시스템은 협업 필터링과 콘텐츠 기반 필터링을 기반으로 한다[27].

먼저 협업 필터링(Collaborative filtering)이란 대규모의 사용자 행동 정보를 분석하여 비슷한 성향의 사용자들이 기존에 좋아했던 항목을 추천하는 기술로 연관된 콘텐츠의 속성을 분석하는 대신, 이용자가 두 콘텐츠를 같이 이용했다는 기록을 바탕으로 추천한다[27]. 협업 필터링은 기존 자료를 바탕으로 이루어지기 때문에 새로운 항목은 추천되지 못하며 이용자들의 관심이 상대적으로 적은 항목은 제공되지 못하는 경우가 많아서 슢립 현상이 나타날 수 있다.

반면 콘텐츠에 기반한 필터링은 항목 자체를 분석하여 추천하는데, 이용자가 특정 음악을 검색해서 듣고 있다면 그 음악과 유사한 음악을 추천하는 방식이다. 이에 각 음악 항목을 분석한 프로파일과 사용자의 선호도를 계산한 프로파일을 추출하여 유사성을 계산한다[28].

한편, 모델 기반 협력 필터링은 이 두 가지 추천 방식을 보완한 것인데 항목 간 유사성을 단순 비교하는 데 그치지 않고 자료 안에 내재한 패턴을 파악하여 추천한다. 예를 들어 어느 사용자가 A라는 드라마를 좋아하는 경우, 드라마와 관련된 주위 정보를 이용해 선호 이유를 유추하여 다른 드라마를 추천하는 방식이다[27].

이처럼 알고리즘은 미리 정해진 규칙, 절차, 명령에 따라 빅데이터를 일련의 과정을 거쳐 결과로 도출하는데, 알고리즘을 설계할 때 어떤 요인에 가중치를 두느냐에 따라 추천 결과가 달라질 수 있다.

2.2 유튜브 알고리즘에 관한 기존 연구 검토

광고수익을 수익원으로 하는 대다수 플랫폼과 마찬가지로 유튜브도 이용자들의 시청시간이 길어지도록

알고리즘이 설계되어 있다[29]. 시청시간은 특정 영상의 조회 수, 평균 시청시간, 세션 시작, 세션 시간, 세션 종료, 업로드빈도, 그리고 영상길이 등의 요인으로 구성되며 각 요인의 가중치는 명확하게 파악하기 어렵다.

유튜브에서는 이용자가 검색 쿼리를 입력해서 비디오를 시청하면 이와 관련된 다른 비디오들을 추천해준다. 초기에 동일한 태그가 붙은 비디오들을 추천하다가 이후 머신러닝을 통해 복잡하고 정교하게 진화되고 있다[30]. 이용자가 보는 비디오의 가중치는 개별 이용자의 동영상 소비 활동에 따라 달라지며, 이용자가 시청하는 비디오 정보뿐만 아니라, 활동 내용, 프로필, 초기에 입력한 쿼리 정보를 기반으로 연관 비디오의 추천 내용이 달라진다[31][32].

유튜브가 추천 시스템을 적용한 이후 추천받은 비디오가 추천받지 않은 비디오보다 더 많이 시청되고 있다. 연관 비디오 20위 내에서 이용자들이 클릭하는 비율은 41.6% 정도이며, 1, 2 순위에 있는 비디오의 클릭 비율이 가장 높다[33].

한편, 유튜브는 개별 이용자가 업로드한 비디오보다 많은 구독자를 확보하고 있는 상업 채널에서 올린 비디오를 우선적으로 노출하도록 설계한다. 많은 구독자가 속한 채널을 노출할수록 이용자들이 유튜브에 오래 머무를 수 있기 때문이다. 채널 구독자가 많지 않거나 기존에 업로드한 영상의 조회 수가 적으면 그 채널의 동영상이 연관 비디오로 추천되는 가능성이 낮아진다[34]. 따라서 이용자들도 새롭게 업로드된 동영상보다 이미 조회 수가 많이 누적된 비디오를 추천받을 가능성이 크다.

2.3 유튜브에서의 콘텐츠 확산에 대한 기존 연구

유튜브에서 콘텐츠 이용이 증가할수록 콘텐츠가 어떻게 확산되고 조회 수가 높아지는가에 관심이 증가하고 있다[35]. 유튜브에서의 콘텐츠 확산에 대한 연구들은 콘텐츠의 일반적인 요인이나 친구 맺기나 구독을 중심으로 한 인적 네트워크 요인에 주로 초점을 두고 있다.

먼저 비디오의 댓글 수, 구독자 수, 평점, 콘텐츠 경과 시간 등은 콘텐츠 조회 수에 영향을 미치며 각 요인이 대체로 상관관계를 갖는다[36]. 수용자들은 콘텐츠를 선택할 때 콘텐츠 내용, 댓글 수나 좋아요 수와 같은 다

른 사람들의 평가를 고려한다[37]. 댓글 수와 좋아요 수는 비디오의 조회에 긍정적인 영향을 주는 것으로 나타났다[38].

다음으로 비디오 간 연관성을 기반으로 네트워크를 분석한 연구에서는 비디오의 매개 중심성이 높을 때 비디오의 인기가 높아진다는 결과를 도출하였다. 비디오를 더 많이 연결하고 매개할수록 비디오가 많이 조회되는데, 싸이의 '강남 스타일'이 가장 대표적인 예이며 [38] 콘텐츠 유형 중에서는 소셜 인플루언서나 이용자가 제작한 영상의 매개 중심성이 높게 나타났다[39].

또한 콘텐츠 등록 초기에는 구독자 네트워크[40] 연결 정도가 영향을 미치지만, 이후에는 친구 간의 네트워크나 매개 중심성 그리고 클러스터링 계수 등이 영향을 미치는 것으로 분석되었다[36]. 한편, 다른 연구에서는 구독자 수, 등록 이후 경과 시간, 매개 중심성, 댓글 수, 근접 중심성, 클러스터링 계수, 평균 평점 순으로 영향력이 크지만[41] 네트워크 구조적 요인, 즉 연결 정도 중심성과 아이겐벡터 중심성은 조회 수에 부적 영향을 주는 것으로 밝혀졌다. 이러한 결과가 나타난 것은 너무 많은 콘텐츠와 연결성이 높을 경우 유사 콘텐츠들이 존재하여 조회 수가 분산될 수 있기 때문에 불리하다고 해석되고 있다[41].

또한 연관 비디오는 36% 정도 또 다른 비디오의 연관 비디오로 추천되는데, 조회 수가 많은 상위 5개 비디오는 연관 비디오 그래프에서 서로 연결되어 있는 것으로 밝혀졌다[42].

III. 연구문제 및 연구방법

1. 연구문제

본 연구에서는 온라인 동영상 플랫폼에서 적용하는 추천 알고리즘에 따라 콘텐츠 노출이 달라질 수 있다는 점에 주목하고, 연관 비디오(related video)로 어떤 콘텐츠가 추천되는지를 알아보고자 하는 연구 목적을 지니고 있다. 이를 위해 K-pop 뮤직비디오의 콘텐츠 특성이 콘텐츠 조회 수와 연관 비디오의 추천 순위에 미치는 영향을 알아보고, 연관 비디오 간 형성되는 네트워크 특성을 살펴보고자 한다. 본 연구에서 설정한 연

구문제는 다음과 같다.

연구문제 1. K pop 뮤직비디오의 콘텐츠 특성이 비디오의 조회 수에 미치는 영향은 어떠한가?

연구문제 2-1. K pop 뮤직비디오의 콘텐츠 특성의 차이에 따라 연관 비디오로 추천되는 순위가 어떻게 달라지는가?

연구문제 2-2. K pop 뮤직비디오의 콘텐츠 특성이 연관 비디오의 추천 순위에 어떠한 영향을 미치는가?

연구문제 3. 원곡 비디오와 연관 비디오 간 네트워크 구조는 어떠한 특성을 지니는가?

2. 연구방법

2.1 주요 개념의 조작적 정의

본 연구에서 수집한 비디오 메타데이터에는 비디오의 제목, 아티스트, 비디오 유형, 업로드 시기, 조회 수, 댓글 수, 좋아요 수, 음악 장르가 포함되어 있다. 채널 구독자 수란 비디오가 업로드된 채널을 구독하는 팔로워들의 수를 말한다. 연관 비디오란 특정 비디오에 대해 유튜브 추천 시스템에 의해 제시되는 비디오를 의미하며 상위 1~20위까지 추천된 동영상으로 제한했다. 조회 수란 해당 비디오의 인기가 확산 정도를 측정하는 지표로 해당 비디오의 조회한 숫자로 표시된다. 좋아요 수는 해당 비디오를 시청한 이후 이용자들이 좋아요를 누른 수를 의미하며, 댓글 수는 해당 비디오에 달린 댓글 수로 이용자들의 참여나 공감 정도를 파악할 수 있는 지표이다[41]. 콘텐츠 등록 기간은 해당 비디오가 유튜브에 등록된 이후 경과된 기간으로 데이터를 크롤링한 날부터 비디오를 최초 등록한 일까지 계산하였다.

표 1. 주요 개념의 조작적 정의

요인	조작적 정의
채널 구독자 수	해당 채널을 구독하는 팔로워들의 수
연관 비디오	특정 비디오에 대해 유튜브 추천 시스템에 의해 제시되는 비디오
조회 수	해당 비디오의 조회 수
댓글 수	해당 비디오에 달린 댓글 수
좋아요 수	해당 비디오에 이용자들이 좋아요를 누른 수

콘텐츠 등록 기간	해당 비디오가 유튜브에 등록된 이후 경과된 일수
연관 비디오 순위	해당 비디오가 연관 비디오로 추천된 순위
연관 비디오의 채널 다양성	특정 비디오를 시청할 때 추천되는 연관 비디오에 포함된 채널 수로 1~20로 측정
연결 중심성	한 노드에 연결된 노드 수를 의미하며 이 값이 클수록 다른 노드에 미치는 영향이 크다고 볼 수 있음.
근접 중심성	네트워크의 중심성을 측정하는 것으로 모든 노드와의 거리값의 역수를 말함. 즉, 중앙에 있을수록 모든 노드에 가깝기 때문에 이 값이 크게 나옴.
매개 중심성	특정 노드가 다른 노드의 연결을 어느 정도 매개하는가를 의미
아이겐벡터 중심성	특정 노드에 연결된 노드들의 중요성을 반영하여 이 값이 클수록 영향력이 큰 노드들이 많이 연결되었음을 의미

한편, 네트워크 분석은 각 대상의 속성을 비롯하여 대상이 연결됨으로써 발생하는 다양한 관계적 특성을 분석할 수 있는 장점이 있다. 중심성이란 한 행위자가 전체 네트워크에서 중심에 위치하는 정도를 표현한 것으로 전체 네트워크에서 각각의 객체가 콘텐츠 확산에 어떠한 역할을 하는지 규명할 수 있다[39]. 먼저 연결 중심성은 한 노드가 얼마나 많은 연결 관계를 갖는가를 의미하며, 노드에 연결된 연결 수로 산출된다. 한 노드에 연결된 링크가 많을수록 연결 중심성이 높으며 네트워크상에서 큰 영향력을 갖는다. 아이겐벡터 중심성은 중심성이 높은 노드와의 연결이 많을수록 영향력이 높다고 본다. 매개 중심성은 네트워크상에서 특정 노드가 다른 노드의 연결을 어느 정도 매개하고 있는가를 나타낸다[39].

2.2 분석방법

표 2. 분석대상 K-pop 아티스트의 채널 정보

아티스트	채널 구독자 수	채널 총 조회 수	채널 다양성	N
블랙핑크	31100000	7839247199	8.39	56
빅히트	30000000	10518199541	10.46	28
방탄소년단	23100000	3466560992	9.19	42
씨아이	13000000	7309388725	13.05	91
빅뱅	11700000	5546225788	9.04	98
아이콘	6170000	1435673826	10.00	70
트와이스	5900000	707844336	8.92	84
2NE1	4580000	1675528186	7.60	77
갯세븐	4320000	505932491	8.86	63
세븐틴	4300000	1071133608	9.38	63

엑소	4000000	69462646	9.66	35
위너	3000000	862653631	8.17	35
Mamamoo	2000000	340474487	8.64	42
소녀시대	1970000	51891646	7.53	49
레드벨벳	1730000	86410512	7.31	70

※ 2019년 11월 11일 기준

먼저 유튜브 채널 구독자 수 기준으로 상위 15개 K-pop 채널을 선별하여 각 채널에 업로드된 뮤직비디오 중 10개 미만씩 총 129곡을 추출하였다. 그런 다음 129곡의 연관 비디오를 상위 20개까지 유튜브 API에 요청하였다. 수집된 연관 비디오들의 ID를 호출하여 상세 정보를 정리하였다. 데이터 수집은 2019년 11월 11일부터 17일까지 1주일간 매일 오후 5시에 추출하였다. 크롤링을 통해 추출된 연관 비디오는 18,060(129 x 20곡 x 7일)이 아닌 17,939를 얻었는데, 일부 동영상의 경우 연관 비디오가 20개가 아닌 경우가 있었기 때문이다.

[표 2]에 제시된 바와 같이 구독자 수가 가장 많은 채널은 2019년 11월 기준으로 3,110만 구독자를 확보한 블랙핑크이며, 빅히트 3,000만, 방탄소년단 2,310만, 씨아이 1,300만, 빅뱅 1,170만 순으로 나타났다. BTS의 경우 기획사 채널(빅히트)과 아티스트 채널(방탄소년단)에 모두 뮤직비디오가 업로드되어 있으며 두 채널의 구독자 수를 합하면 5,300만이다. 채널에 업로드된 동영상의 총 조회 수는 빅히트 채널이 105억 이상으로 가장 많은 것으로 나타났다.

또한 아티스트별로 연관 비디오로 추천되는 채널 다양성을 살펴보았는데, PSY 13.05, 빅히트 10.46, 아이콘 10 순으로 PSY가 다양한 채널에서 연관 비디오로 추천되고 있으며, 소녀시대 등 걸그룹의 채널 다양성이 상대적으로 낮은 것으로 밝혀졌다.

다음은 연관 비디오 목록에 나타난 비디오의 조회 수, 댓글 수, 좋아요 수, 곡길이, 경과일 간에 상관관계를 분석해보았다. 조회 수, 댓글 수, 좋아요 수 간에 정적 상관관계가 있는 것으로 나타났다. 그러나 경과일과 댓글 수나 좋아요 수는 부적 상관관계가 발견되었는데 경과일이 클수록 댓글 수와 좋아요 수는 감소하였다. 상관계수와 독립변인 간의 상관관계의 정도를 의미하는 VIF값이 10지 넘지 않아 다중공선성에는 문제가 없

는 것으로 나타났다.

표 3. 주요 변인간 상관관계

구독자		조회수	댓글 수	좋아요수	곡길이	경과일
전체	조회 수	1	.561**	.859**	-.028**	.070**
	댓글 수		1	.811**	-.029**	-.110**
	좋아요 수			1	-.052**	-.107**
	곡길이				1	-.030**
	경과일					1
500만 이하	조회 수	1	.595**	.840**	-.047**	.019
	댓글 수		1	.831**	-.039**	-.123**
	좋아요 수			1	-.068**	-.186**
	곡길이				1	-.030**
	경과일					1
500만 ~1000만	조회 수	1	.690**	.941**	-.030*	.128**
	댓글 수		1	.800**	-.042**	-.017
	좋아요 수			1	-.055**	.007
	곡길이				1	-.008
	경과일					1
1000만 이상	조회 수	1	.611**	.827**	-.090**	.193**
	댓글 수		1	.856**	-.050*	-.092**
	좋아요 수			1	-.104**	-.006
	곡길이				1	-.076**
	경과일					1

그러나 15개의 그룹을 구독자 규모에 따라 1000만 이상, 500~1000만, 500만 이하의 세 그룹으로 구분하여 변인 간 상관관계를 다시 살펴본 결과 500~1000만 구독자 수가 있는 그룹에서 조회 수와 좋아요 수의 상관관계가 가장 높은 것으로 분석되었다(.941). 그러나 1000만 이상의 구독자 수를 갖는 그룹 3의 경우 좋아요 수와 댓글 수간의 상관관계가 가장 높게 나타났다(.856).

다음으로 네트워크 분석을 위해 원곡 비디오 데이터와 연관 채널 또는 연관 비디오 데이터를 2차원 형태인 2-mode network로 만들었다. 각 원곡 비디오에서 연관 동영상이 나왔으면 1, 아니면 0으로 코딩하고, 연관 채널의 경우 Original과 Channel을 이용하여 2-mode network를 만들며 연관 비디오는 Original과 Video를 이용하여 2-mode network를 만들었다. 2-mode network를 만든 뒤 행렬 곱으로 Original x Original과 같은 1-mode network 행렬을 만들고 다음 단계로 근접 중심성, 매개 중심성, 연결 중심성, 아이

겐벡터 중심성을 분석하였다. 네트워크 분석 및 시각화는 Gephi 소프트웨어를 활용했다.

IV. 연구결과

1. 분석 비디오와 연관 비디오의 기본 통계

표 4. 분석 비디오와 연관 비디오의 기본 통계

	세부항목	합계	평균
원곡 비디오 (N=129)	채널 구독자수	1190190000	9226279.07
	조회수	26625549900	206399611.63
	좋아요수	290430741	2268990.16
	댓글 수	44272349	343196.50
연관 비디오 (N=17937)	채널 구독자수	160047630000	8922764.68
	조회수	2982431634482	166272600.46
	좋아요수	28109715466	1567135.83
	댓글 수	2989680946	166676.75
	경과일	24395752	1360.08

분석대상 비디오 129곡의 평균 채널 구독자 수는 922만이며, 비디오 당 평균 조회 수는 2억 633만, 평균 좋아요 수는 226만, 평균 댓글 수는 34만 정도로 팬덤이 높은 곡이라는 것을 알 수 있다.

한편 1주일간 추출한 연관 비디오 17,937곡에 대한 평균 채널 구독자 수는 892만으로 원곡 비디오보다 다소 적다. 비디오당 평균 조회 수는 1억 6627만, 평균 좋아요 수는 156만, 평균 댓글 수는 16만 6천 개로 조사되었다. 연관 비디오의 댓글 수가 원곡 비디오의 댓글 수의 1/2 수준이다. 연관 비디오가 유튜브에 업로드된 경과일은 평균 1360일, 즉 45개월 정도로 나타났다.

2. K-pop의 조회 수에 영향을 미치는 요인

분석 대상인 129곡의 뮤직비디오의 조회 수에 영향을 미치는 요인을 살펴본 결과, 좋아요 수와 채널 구독자 수가 통계적으로 유의미한 변인이며 댓글 수는 영향을 주지 않는 것으로 밝혀졌다. 그 중 좋아요 수는 표준화 계수(β) .743으로 매우 영향력이 큰 변인이지만, 채널 구독자 수는 음의 방향으로 조회 수에 영향을 주는 것으로 나타나(β =-.188) 기존 연구와 상충되는 결과를 보였다. 독립변인의 종속 변인에 대한 설명력을 나타내

는 R 제곱값은 .659으로 나타났다.

표 5. K-pop 뮤직비디오의 조회 수에 영향을 미치는 요인

독립변인	표준화 계수 (β)	t	유의확률
(상수)		1.180	.240
채널 구독자 수	-.188	-2.553*	.012
좋아요 수	.743	4.481**	.000
댓글 수	.179	1.242	.217
F		80.619**	

3. K-pop 콘텐츠 특성의 차이에 따른 연관 비디오의 추천 순위

표 6. K-pop 콘텐츠특성의 차이에 따른 연관 비디오의 추천 순위

		N	M	SD	F
조회 수	5천만 이하	8697	10.33	5.595	29.673**
	5천만~1억	2637	10.84	5.705	
	1억~3억	4117	10.49	5.902	
	3억 이상	2486	10.28	5.928	
좋아요 수	100만 이하	10683	10.33	5.588	6.388**
	100~200만	3156	10.76	5.881	
	200~300만	1568	10.43	6.047	
	300만~400만	922	10.99	6.124	
	400만이상	1608	10.18	5.797	
댓글 수	10만 이하	12006	10.43	5.604	2.497*
	10만~20만	2429	10.21	5.871	
	20만~30만	1198	10.78	6.201	
	30만~40만	497	10.79	5.788	
	40만 이상	1807	10.42	6.029	

다음으로 유튜브에 업로드된 K-pop 뮤직비디오의 콘텐츠 특성, 즉 조회 수, 좋아요 수와 댓글 수의 차이에 따라 연관 비디오로 추천되는 순위가 어떻게 달라지는지 살펴보았다. 분석 결과, 3억 이상으로 조회 수가 가장 많은 비디오의 연관 순위는 10.28이었으며, 좋아요 수도 400만 이상일 때 10.18로 가장 낮은 것으로 나타났다. 그러나 댓글 수가 많다고 해서 연관 비디오 추천 순위가 가장 낮은 것은 아니었다.

4. K-pop의 연관 비디오의 추천 순위에 영향을 미치는 요인

표 7. K-pop의 연관 비디오의 추천 순위에 영향을 미치는 요인

독립변인	표준화계수 (β)	t	유의확률
(상수)		116.381	.000
조회 수	-.099	-5.699**	.000
댓글 수	-.026	-1.790	.073
좋아요 수	.094	3.876**	.000
경과일	-.010	-1.275	.202
F		11.833	

연관 비디오의 추천 순위에 영향을 미치는 요인을 회귀분석을 통해 살펴본 결과, 각 비디오의 좋아요 수와 조회 수가 통계적으로 유의미한 영향을 주는 요인으로 나타났다. 단, 좋아요 수가 많을수록 연관 순위가 높게 나타났지만($t=3.876$), 조회 수는 음의 관계로 영향을 주는 것으로 나타났다($t=-5.699$). 한편 댓글 수와 경과일은 유의미한 영향을 주지 않은 것으로 밝혀졌다. 이러한 결과는 비디오에 대한 수용자의 긍정적 반응을 나타내는 지표가 높을수록 연관 비디오로 추천됨으로써 비디오의 확산에 영향을 주는 것으로 해석할 수 있다.

5. 원곡비디오와 연관비디오의 네트워크분석결과

본 연구에서는 유튜브의 원곡 비디오와 연관 비디오의 목록을 바탕으로 129개 비디오 x 20개 연관 비디오의 네트워크를 분석하였다.

5.1 원곡 비디오의 네트워크 분석

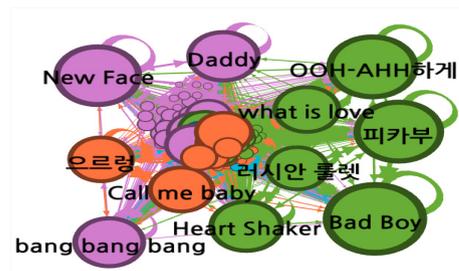


그림 2. 원곡 비디오의 매개 중심성

원곡 비디오 129곡의 네트워크는 크게 세 클러스터로 구분된다. 즉 걸그룹 노래(우아하게/what is love/Heart Shaker, Bad Boy, 러시안 룰렛, 피카부와 EXO(으르렁, Call me baby)나 그리고 빅뱅(뱅뱅뱅)같은 5년 이상 된 그룹, 그리고 PSY(Daddy)로 구성된다.

이들 노래들은 다른 노드들을 최단 경로로 연결하는 데 중요한 역할을 하고 있다. 반면 갓세븐 if, 싸이의 hangover, we are the one, 오빤 딱 내 스타일과 같이 2015년 이전 발매된 곡이거나 Black pink의 kill this love dance practice와 같이 뮤직비디오가 아닌 영상이 매개 중심성이 낮은 것으로 분석되었다.



그림 3. 원곡 비디오의 근접 중심성

다음으로 다른 노드들과의 거리가 가까운 근접 중심성을 나타내는 분석은 매개 중심성과 유사한 결과를 보이고 있다. 걸그룹 노래와 PSY, EXO, 빅뱅 등의 노래가 근접 중심성이 높게 나타나면서 다른 뮤직비디오와 네트워크 형성 시 허브 역할을 하고 있다. 근접 중심성이 낮은 뮤직비디오 역시 if, Hangover, 오빤 딱 내 스타일, we are the one, gentleman으로 매개 중심성이 낮은 뮤직비디오와 중복된다.



그림 4. 원곡 비디오의 아이젠벡터 중심성

아이젠벡터 중심성이 상위에 있는 뮤직비디오는 매개 중심성과 비슷한 결과를 보이는데, 특이하게 세븐틴 울고 싶지 않아/갓세븐 니가하면/엑소 Tempo가 높게 랭크되고 있다. 아이젠벡터 값이 높은 노드들은 중심성이 큰 영향력 있는 노드들과 연결되어 있음을 의미한다. 반면 아이젠벡터 중심성이 낮은 노래는 hangover,

if, 오빤 딱 내 스타일, we are the one, gentleman 등 매개 중심성이 낮은 노래와 중복된다.

5.2 연관 비디오의 네트워크 분석

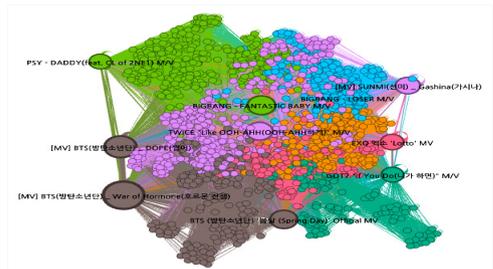


그림 5. 연관 비디오의 매개 중심성

다음으로 연관 비디오의 네트워크 특성을 분석한 결과 BTS, PSY, GOY7, BIGBANG, 걸그룹의 5개 그룹으로 구분된다. 특히 BTS 호르몬 전쟁, 봄날, DOPE과 PSY Daddy와 빅뱅의 Fantastic baby, Loser/트와이스의 우아하게가 매개 중심성 값이 높다. 반면 매개 중심성이 낮은 영상은 개인이 올린 영상과 GOT7 공식 채널, 외국 노래인 것으로 분석되었다.

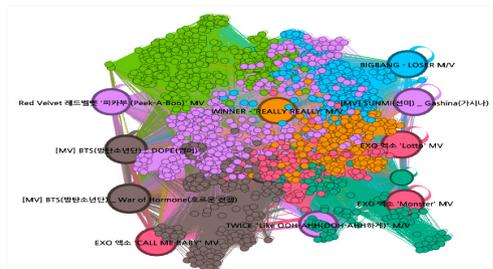


그림 6. 연관 비디오의 근접 중심성

연관 비디오의 근접 중심성 역시 BTS의 노래가 가장 높으며 레드벨벳의 피카부, 트와이스의 우아하게, 엑소의 Monster와 Lotte, Call me baby가 상위에 랭크되었다. 근접 중심성이 낮은 노래는 조회 수가 400만 이하로 GOT7 노래가 포함되었다.

연관 비디오 중 아이젠벡터 중심성이 높은 비디오는 BTS의 호르몬전쟁, DOPE이 압도적으로 높으며, 레드벨벳의 피카부와 BTS의 노래가 연결되어 있었다. 엑소의 노래가 아이젠벡터 중심성이 높게 나타났고, 선미의

뮤직비디오가 내포하는 사회문화적 차원의 심층적인 분석과 동영상이 소비되는 맥락에 대한 분석이 요구되는 지점이다. 그럼에도 K pop의 재생시 추천되는 17,000여 개의 연관 비디오 데이터를 실증적으로 분석했다는 데 의의를 두고자 한다.

참 고 문 헌

- [1] G. Sivinski, A. Okuliar, and L. Kjolbye, "Is big data a big deal? A Competition law approach to big data," *European Competition Journal*, Vol.13, No.2-3, pp.199-277, 2017.
- [2] 김도균, "과학기술정보 서비스 플랫폼에서의 빅데이터 분석을 통한 개인화추천 서비스 설계," *한국비블리아학회지*, 제28권, 제4호, pp.501-518, 2017.
- [3] M. L. Katz, "Multisided Platforms, Big Data, and a Little Antitrust Policy," *Review of Industrial Organization*, Vol.54, pp.695-716, 2019.
- [4] 조성기, 이영주, "유료방송 가입자의 장르 선호도와 VOD 구매의 관계에 관한 연구: IPTV 영화 이용을 중심으로," *한국콘텐츠학회논문지*, 제16권, 제11호, pp.91-102, 2016.
- [5] M. Airoidi, D. Beraldo, and A. L. Gandini, "Follow the algorithm : An exploratory investigation of music on Youtube," *Poetics*, Vol.57, pp.1-13, 2016.
- [6] B. Hallinan and T. Striphas, "Recommended for you : The Netflix Prize and the production of algorithmic culture," *New Media & Society*, Vol.18, No.1, pp.117-137, 2016.
- [7] 김조한, *플랫폼 전쟁*, 메디치, 2017.
- [8] Strabase, "음악 산업에서도 인공지능과 데이터 활용의 중요성 증대, 열성팬 찾기 등에 활용," *NewsBrief*, 2017.
- [9] <https://www.nytimes.com/2019/06/14/opinion/youtube-algorithm.html?searchResultPosition=1>, 2020.2.1.
- [10] Strabase, "2018년 음악 산업에 영향을 미칠 3가지 요인," *NewsBrief*, 2018.
- [11] 나스미디어, "미디어 이슈 리포트," 2018.
- [12] 김경달, *유튜브 트렌드 2020*, 이은북, 2019.
- [13] C. Cayari, "The YouTube effect : how YouTube has provided new ways to consume, create, and share music," *International Journal of Education & the Arts*, Vol.12, No.6, pp.1-28, 2011.
- [14] 강미선, "유튜브는 어떻게 갓튜브가 됐나 : 동영상으로 1020 사로잡고 수익배분으로 콘텐츠 확보," *신문과 방송*, 제570호, 2018.
- [15] 금준경, "국내 언론사의 유튜브 생존 전략 : 오리지널 콘텐츠 라이브, 서브 채널에 승부 걸다," *신문과 방송*, 제570호, pp.11-15, 2018.
- [16] http://www.hani.co.kr/arti/culture/culture_general/893488.html#csidx6e800d7bfeda60aa382d9417d326097, 2020.2.1.
- [17] Jin Dal Yong and Yoon Kyong, "The social mediascape of transnational Korean pop culture : Hallue 2.0 as spreadable media practice," *New media & society*, Vol.18, No.7, pp.1277-1292, 2016.
- [18] I. Oh and Gil-sung Park, "From B2C to B2B : Selling Korean Pop Music in the Age of New Social media," *Korea Observer*, Vol.43, No.3, pp.365-397, 2012.
- [19] 조병철, 심희철, "K-POP 한류의 성공요인분석과 한류 지속화 방안 연구 : 스마트 미디어 기반 실감 콘텐츠 활용을 중심으로," *한국콘텐츠학회논문지*, 제13권, 제5호, pp.90-102, 2013.
- [20] I. Oh and H. G. Lee, "Mass media technologies and popular music genres : K-pop and Youtube," *Korea Journal*, Vol.53, No.4, pp.34-58, 2013.
- [21] J. van Dijck, "Users like you? Theorizing agency in user-generated content," *Media, Culture & Society*, Vol.31, No.1, pp.41-58, 2009.
- [22] J. H. Ahn, S. Oh, and H. Kim, "Korean pop takes off! Social media strategy of Korean entertainment industry," In : *Proceedings on the 10th international conference on service systems and service management*, New York : IEEE, pp.774-777, 2013.
- [23] http://www.hani.co.kr/arti/culture/culture_general/893488.html#csidx663d8a4d6db0df38cf3

- 974c11248671, 2020.2.1.
- [24] D. Y. Jin, *Globalization and Media in the Digital Platform Age*, Routledge, p.126, 2020.
- [25] U. Siriyuvasak and S. Hyunjoon, "Asianizing K-pop: production, consumption and identification patterns among Thai youth," *Inter Asia Cultural Studies*, Vol.8, No.1, pp.109-136, 2007.
- [26] 정두희, *AI 초격차 시대가 온다*, 청림출판, 2019.
- [27] 서봉원, "콘텐츠 추천 알고리즘의 진화," *방송 트렌드 & 인사이트*, 제5권, pp.19-24, 2016.
- [28] M. Kaminskas and F. Ricch, "Contextual music information retrieval and recommendations : State of the art and challenges," *Computer Science Review*, Vol.6, pp.89-119, 2012.
- [29] 오세욱, "알고리즘으로 본 유튜브의 미디어 지향," *관훈저널*, 제61권, 제1호, pp.11-17, 2019.
- [30] M. Bendersky, L. Garcia-Pueyo, J. Harmsen, V. Josifovsky, and D. Lepikhin, *Up Next : Retrieval methods for large scale related video suggestion categories and subject descriptors*, KDD, 2014.
- [31] 유소엽, 정옥란, "사용자의 소셜 카테고리를 이용한 유튜브 동영상 추천 알고리즘," *정보과학회논문지*, 제42권, 제5호, pp.664-670, 2015.
- [32] Q. Huang, B. Chen, J. Wang, and T. Mei, "Personalized Video Recommendation through graph propagation," *ACM Transaction on Multimedia Computing, Communications, and Applications*, Vol.10, No.4, p.32, 2014.
- [33] R. Zhou, S. Khemmarat, and L. Gao, "The impact of YouTube recommendation system on video views," In M. Allman (Ed.), *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement (IMC'10)*, New York, NY: ACM Press, pp.404-410, 2010.
- [34] M. Gielen and J. Rosen, "Reverse Engineering The YouTube Algorithm," *tubefilter*, 2016.
- [35] 이선정, 이수범, "유튜브 상에서의 K-뷰티 한류 콘텐츠 확산 전략 연구," *GRI 연구논총*, 제20권, 제3호, pp.231-259, 2018.
- [36] H. Yoganasimhan, "Impact of social network structure on content propagation : A Study using Youtube data," *Quantitative Marketing and Economics*, Vol.10, pp.111-150, 2012.
- [37] 박정아, 임지은, 황장선, "유튜브 브랜드 채널 콘텐츠의 커뮤니케이션 전략," *한국광고홍보학보*, 제20권, 제2호, pp.95-151, 2018.
- [38] 임병학, "유튜브의 방문횟수에 대한 온라인 비디오 네트워크의 영향 : 강남 스타일의 패러디 비디오를 중심으로," *유라시아연구*, 제10권, 제1호, pp.1-15, 2013.
- [39] W. W. Xu, J. Y. Park, J. Y. Kim, and H. W. Park, "Networked cultural diffusion and creation on Youtube : An analysis of Youtube memes," *Journal of Broadcasting & Electronic Media*, Vol.60, No.1, pp.104-122, 2016.
- [40] A. Susarla, J. H. Oh, and Y. Tan, "Social Networks and the diffusion of user-generated content : Evidence from Youtube," *Information Systems Research*, Vol.23, No.1, pp.23-41, 2012.
- [41] 박병언, 임규건, "일반영향요인과 댓글 기반 콘텐츠 네트워크 분석을 통한 유튜브상의 콘텐츠 확산 영향요인 연구," *지능정보연구*, 제21권, 제3호, pp.19-36, 2016.
- [42] G. Sivinski, A. Okuliar, and L. Kjolbye, "Is big data a big deal? A Competition law approach to big data," *European Competition Journal*, Vol.13, No.2-3, pp.199-277, 2017.

저 자 소 개

이 영 주(Yeong-Ju Lee)

정회원



- 1991년 2월 : 이화여자대학교 영어영문학(문학사)
- 1998년 2월 : 서강대학교 신문방송학(석사)
- 2005년 8월 : 이화여자대학교 신문방송학(박사)
- 2009년 3월 ~ 현재 : 서울과학기술대학교 IT정책전문대학원 교수

