

머신러닝 기반의 디지털 방송 프로그램 유형 분류 및 활용 방안 연구¹

A Study of the Classification and Application of Digital Broadcast Program Type based on Machine Learning

윤상혁 (Sang-Hyeak Yoon) 연세대학교 정보대학원/스마트미디어랩²
이소현 (So-Hyun Lee) 연세대학교 정보대학원³
김희웅 (Hee-Woong Kim) 연세대학교 정보대학원⁴

ABSTRACT

With the recent spread of digital content, more people have been watching the digital content of TV programs on their PCs or mobile devices, rather than on TVs. With the change in such media use pattern, genres(types) of broadcast programs change in the flow of the times and viewers' trends. The programs that were broadcast on TVs have been released in digital content, and thereby people watching such content change their perception. For this reason, it is necessary to newly and differently classify genres(types) of broadcast programs on the basis of digital content, from the conventional classification of program genres(types) in broadcasting companies or relevant industries. Therefore, this study suggests a plan for newly classifying broadcast programs through using machine learning with the log data of people watching the programs in online media and for applying the new classification. This study is academically meaningful in the point that it analyzes and classifies program types on the basis of digital content. In addition, it is meaningful in the point that it makes use of the program classification algorithm developed in relevant industries, and especially suggests the strategy and plan for applying it.

Keywords: Digital content, Broadcast program classification, Machine learning, Clustering analysis, Spectral clustering algorithm

1) 논문접수일: 2019년 7월 11일; 1차 수정: 2019년 8월 21일; 게재확정일: 2019년 8월 22일

2) 제 1저자 (scottyoon@yonsei.ac.kr)

3) 교신저자 (sohyun1010@yonsei.ac.kr)

4) 제 3저자 (kimhw@yonsei.ac.kr)

1. 서론

디지털이 일반화되면서 TV, 라디오, 신문, 잡지 등의 전통적인 4대 매체가 주를 이루던 시대를 지나 이제는 본격적으로 온라인과 모바일을 기반으로 한 디지털 콘텐츠가 확산되고 있다. 2017년 글로벌 소비자들은 하루 평균 56분을 온라인 동영상 시청하였고, 2018년에는 67분을 시청하였다(Mobvista 2018). 특히, 온라인 동영상 중에서 방송 프로그램을 온라인에서 시청하는 형태가 크게 증가하였다(Mikos 2016; Vanattenhoven and Geerts 2015). 방송 프로그램은 방송 전반에 걸쳐 가장 중요한 요소 중의 하나로써 한 사회 내에서 방송의 이념, 문화, 제도 등이 구체화되는 영역이다(주영호·황성연 2004). 방송 프로그램의 장르(유형) 구분이 중요한 이유는 각각의 프로그램이 어느 장르(유형)에 속하는지에 따라서 제작 방식, 내용과 형식, 그리고 시청자가 얻는 즐거움이나 정보가 다르기 때문이다(주창운 2004). 또한, 방송사나 산업이 프로그램에 투자하는 방식도 장르(유형)에 따라 달라지며(주창운 2004), 방송 프로그램의 특성에 따라 광고 효과가 다르게 발생되며 광고 상품 설계의 기초가 된다(Snapp 2019). 특히, 온라인 동영상을 통한 방송 프로그램 시청 행태가 증가하면서 동영상을 기반으로 한 온라인 광고도 증가하였다(공감신문 2019). 다시 말해, 디지털 방송 프로그램 유형 분류가 함께 제공되는 온라인 광고의 효과에 영향을 미칠 수 있다는 것이다. 그러므로, 방송 프로그램의 장르(유형)는 단지 프로그램들을 분류하는 것을 넘어 광고와 같은 관련 분야에까지 영향을 끼친다.

방송 프로그램의 장르(유형)는 시대의 흐름과 시청자의 트렌드에 따라 끊임없이 변화한다. 최근 미디어 사용 패턴의 변화와 함께 기존에 일반화되어 사용해 왔던 프로그램 장르(유형)의 성격이 변하는 장르(유형) 혼합 현상이 나타나고 있다(주창운 2004). 이는 드라마와 오락, 다큐멘터리와 예능 등 프로그램 형식과 내용의 혼

합된 형태가 확대되고 있는 것을 나타낸다. 그리고, 하나의 프로그램 장르(유형) 내에서도 두 가지 이상의 장르(유형) 혼합 현상이 나타나기도 한다. 무엇보다도 기존에 TV를 중심으로 방영되었던 프로그램들이 동영상의 디지털 콘텐츠로 변화되면서 이를 시청하는 시청자의 인식도 변하고 있는 것이다. 그러므로, 미디어 사용의 변화와 함께 최근 TV 뿐만 아니라 온라인에서 다양한 디지털 콘텐츠를 통한 방송 프로그램의 시청은 프로그램 유형 구분의 새로운 변화를 요구한다고 할 수 있다(Narducci et al. 2018). 이와 관련하여 방송 프로그램의 장르(유형) 분류를 위한 기준 및 분류 시스템에 대한 연구들이 수행되었지만(Pan et al. 2014; 주창운 2004), 이들은 전형적인 TV 기반의 프로그램 장르(유형)를 재 분류하는 것에만 초점을 두었다. 전통적인 장르(유형) 분류는 대부분 공급자인 방송사와 관련 산업의 주관적인 판단으로 시청자의 인식과 동떨어진 분류를 할 수 있다. 즉, 디지털 콘텐츠 기반의 방송 프로그램 시청 환경에서 공급자 관점의 주관적인 판단을 배제하고 객관적이고 과학적인 방법으로 TV프로그램 유형을 분류하고 이를 활용할 수 있는 구체적인 방안을 제시한 연구가 없었다. 따라서, 기존에 방송사 또는 관련 산업 등에서 분류한 프로그램의 장르(유형)와는 차별적으로 온라인 환경에서 디지털 콘텐츠를 시청하는 시청자들의 시청 패턴을 기반으로 프로그램의 유형을 새롭게 분류할 필요가 있다.

이에, 본 연구에서는 온라인 환경에서 디지털 콘텐츠를 기반으로 방송 프로그램을 시청하는 시청자들의 시청 패턴을 나타내는 시청 로그를 기반으로 프로그램 유형을 새롭게 구분하고 이를 활용하는 방안을 제시하고자 한다. 구체적으로 디지털 콘텐츠 기반 시청 로그 데이터를 이용하여 방송 프로그램의 유형을 새롭게 분류하고 이를 활용하는 방안을 제시하고자 한다. 방송 프로그램의 유형을 분류하는 방법으로 머신러닝(Machine Learning) 및 빅데이터(Big Data) 분석을

사용하여 수용자 인식 기반의 온라인 전용 방송 프로그램을 분류한다. 사용자 로그 기반의 빅데이터 분석을 통해 프로그램의 특성을 정확하게 파악하여 효과적인 마케팅 전략을 수립하는데 도움을 줄 수 있다. 본 연구는 미디어 사용의 변화로 시청자들의 방송 프로그램 시청 패턴이 달라진 것을 고려하여 디지털 콘텐츠 시청 로그를 기반으로 프로그램을 분류한 후 활용 방안을 제시한 것에 학술적 의의가 있다. 더불어, 실무적으로는 디지털 콘텐츠 기반 시청 로그로 프로그램을 분류하고, 시청자 인식 기반의 TV프로그램 분류를 통해 시청자의 요구에 맞는 구체적인 전략 및 방안을 제시했다는 의의가 있다.

2. 개념적 배경

2.1 방송 프로그램 장르(유형) 분류

장르(유형)는 일반적으로 내용, 대상, 기능, 형식에 따라 분류할 수 있으며, 대체로 방송 프로그램의 장르(유형)를 분류할 때는 기능에 의한 분류의 기준을 사용한다(주영호·황성연 2004). 기능에 의한 분류에서 대표적으로 NHK의 프로그램 장르(유형) 분류를 들 수 있는데 이는 1) 보도, 2) 드라마, 3) 오락, 4) 음악, 5) 스포츠, 6) 교육, 7) 생활정보, 8) 기타 등의 기능에 초점을 두고 프로그램을 분류한 것이다. 하지만, 이러한 기능적 분류 체계는 프로그램을 분류하는 명확한 기준이 없고, 최근에는 분류하기 어려운 프로그램들이 점점 증가하면서 명확한 분류가 더 어려워졌다. 그래서, 지금까지 방송사와 시청자, 그리고 정책당국자가 각자 편리한 프로그램 분류 방식을 사용해 왔다(주영호·황성연 2004).

장르(유형) 관습은 어느 정도 고정되어 있지만, 산업과 시청자의 취향에 따라서 변화한다. 특히, 방송 프로그램은 시청률에 직접적으로 영향을 받기 때문에 다른 문학이나 영화 장르(유형)에 비해 변화의 속도가 빠

른 편이다(주창윤 2004). 긴 시간을 통해 형성된 장르(유형)는 특정 하위 장르(유형)들이 모여서 장르(유형) 관습을 형성하고, 둘 사이의 상호작용은 장르(유형) 구성체를 형성하고 또 다른 장르(유형) 구성체와 상호작용하면서 장르(유형) 네트워크를 만들어간다(원용진 1997; 주창윤 2004). 하나의 장르(유형) 내에서 변형과 혼합이 이루어지는 것처럼 장르(유형) 간 혼합 현상도 나타난다. 방송 프로그램 장르(유형)는 형식과 내용에 의한 전달방식으로 구분되는데, 이 때 다양한 결합이 이루어진다는 것이다.

대표적으로 최근에는 하나의 방송 프로그램 내에서 서로 다른 장르(유형)들이 결합되는 경우가 증가하고 있다. 예를 들어, 기존에 방송사가 지정해 놓은 방송 프로그램의 장르(유형)로는 ‘오락’이지만, 오락 프로그램 내에서 정보나 지식을 포함함으로써 형식은 오락 또는 예능을 사용하고 내용은 생활정보나 교육 등을 담고 있는 것을 말한다. 이는 방송사의 입장에서 기존의 장르(유형) 관습에서 벗어나 새로운 장르(유형)를 만드는 방법 중의 하나로 시청자들의 관심과 시청률을 높일 수 있다. 하지만, 이러한 장르(유형) 혼합 현상은 프로그램의 장르(유형)를 정의하는데 어려움이 생긴다. 또한, 방송 프로그램의 장르(유형)를 나타내는데 고려해야 할 사항 중의 하나로 시청자의 기대지평이 있다(김희경 2004). 시청자는 특정 프로그램이 어떻게 전개되는지 알고 있으며, 장르(유형)마다 그들이 얻는 즐거움이 다르다. 즉, 장르(유형) 관습이 만들어내는 방식에 따라 시청자의 기대지평은 달라지게 된다. 따라서 장르(유형)를 분류하는데 시청자가 프로그램에 갖는 기대지평, 즉 시청자의 프로그램에 대한 인식을 포함시키는 것이 필요하다(김희경 2004; 주창윤 2004).

특히, 최근에는 스마트 디바이스의 보급 증가로 콘텐츠 소비 방식에도 변화가 일어나고 있다. 기존에는 주로 TV를 통해서만 방송 프로그램을 시청하였다면 이제는 PC, 태블릿 PC, 스마트폰 등으로 방송 프로그램을 동

영상으로 시청하며 방송 프로그램 시청에 대한 인식도 변화하고 있다(김종무 2015). 즉, 온라인 환경에서 동영상과 같은 디지털 콘텐츠를 통한 방송 프로그램 시청은 시청자들에게 방송 프로그램에 대한 다른 인식을 가지게 할 수 있다. 이것은 시청자의 입장에서 TV가 아닌 디지털 콘텐츠 기반의 방송 프로그램 시청에 대한 선호도나 인식이 다를 수 있다는 것이다. 이러한 시청자들의 방송 프로그램에 대한 인식 변화는 그들에게 다른 프로그램을 추천하거나 동영상과 같은 디지털 콘텐츠의 방송 프로그램에 광고를 함께 제공하는 광고주와 같은 관련 산업의 전략 변화에도 큰 영향을 끼칠 것이다. 그리하여, 본 연구에서는 이러한 콘텐츠 소비 방식이 변하고 있는 현 시점에서 TV 외 온라인을 통해 방송 프로그램을 시청하는 시청자들의 로그 데이터를 이용하여 방송 프로그램 장르(유형)의 새로운 분류를 시도하고자 한다.

2.2 선행연구

방송 프로그램의 다양성과 프로그램 장르(유형) 분류에 대한 이슈들이 제기되면서 이와 관련된 연구들도 진행되고 있다. 또한, 기존에 TV를 통해 주로 시청하던 방송 프로그램을 최근 다양한 플랫폼과 콘텐츠로 시청하면서 매체와 장르 별 프로그램 시청에 대한 관심도 증가하고 있다. 이와 관련하여 기존 연구들은 매체별 프로그램 장르(유형)의 선호도(김종무 2015; 전범수 2011), 프로그램 추천을 위한 프레임워크 또는 그룹 분류(Hara et al. 2004; Kurapati and Gutta 2002), 프로그램 장르(유형) 구분을 위한 기준 및 시스템 제안(Pan et al. 2014; 주창윤 2004) 등의 연구들이 있었다(아래 <표 1> 참조). 그 밖에, 프로그램 장르(유형) 분류 체계와 시청패턴의 일치성 분석(주영호·황성연 2004), 프로그램 장르(유형)의 다양성과 시장 상황과의 관계(박소라 2003), 소셜 iTV에서 프로그램 장르(유형)의

<표 1> 선행 연구

구분	저자	연구 목적	연구 결과
매체별 프로그램 장르(유형)의 선호도	김종무(2015)	플랫폼의 종류에 따른 시청자들의 선호 장르에 대해 조사하고자 함	TV로 시청할 때 선호하는 장르는 버라이어티쇼, 드라마 순이었으며, 스마트폰으로 시청할 때는 드라마, 애니메이션 순이었음
	전범수(2011)	장르 선호도가 시청 정도에 끼치는 영향을 조사하고자 함	지상파 방송 채널들은 뉴스/시사 분야와 드라마 장르를 중심으로 경쟁이 구축됨
프로그램 추천을 위한 프레임워크 또는 그룹 분류	Hara et al.(2004)	시청자가 과거에 시청하는 프로그램들이 해당 시청자의 프로그램 선호도를 나타낸다고 가정하고 시청자별 특성을 조사하고자 함	8개의 시청자 그룹을 나타내었고, 이는 시청자의 TV 접촉, 그들이 선택한 프로그램을 시청한 동기, 그리고 TV 외 그들의 관심사를 설명함
	Kurapati and Gutta(2002)	시청자의 기존 시청 프로그램 패턴을 반영한 프로필을 기반으로 프레임워크를 제안하고자 함	시청자들에게 즉각적인 개인화 서비스를 제공하는 프로그램 추천 및 프로필 기반 프레임워크를 개발함
프로그램 장르(유형) 구분을 위한 기준 및 시스템 제안	Pan et al.(2014)	방송 프로그램의 새로운 분류 시스템을 제시하고자 함	새로운 방송 프로그램 분류 시스템을 제안하고, 제안한 시스템이 기존의 전통적인 분류 시스템 보다 더 이점이 있는 것으로 밝힘
	주창윤(2004)	방송 프로그램 장르 분류와 장르별 유사성 인식의 기준에 대해 조사하고자 함	방송 프로그램의 장르 구분을 위한 새로운 구분 기준을 설정하고, 프로그램 장르와 각각의 하위 장르의 구분 기준을 제시함

역할(Geerts et al. 2008) 등의 다양한 연구들이 진행되었다.

구체적으로 주영호·황성연(2004)은 기존 텔레비전 프로그램 유형 분류 체계와 시청행태와의 일치성 분석을 통하여 기존의 분류체계가 시청자의 시청 다양성을 유지 및 관리하는 기준으로서의 역할을 하는지를 실증적으로 분석하였다. 방송 프로그램의 시청행태와 프로그램 유형의 일치도를 검증하기 위하여 일주일 동안의 텔레비전 시청률 데이터를 이용하였다. 연구 결과, 기능적 프로그램 분류는 시청자의 시청행태와 동일성을 보이지 못했으며, 기존 분류 내에서의 프로그램 시청유사성도 유형별로 다른 요인으로 구분되는 시청유사성을 보였다. 김종무(2015)는 다양하게 변화하고 있는 플랫폼의 종류에 따라 시청자들이 선호하는 장르가 어떠한지를 실험 연구를 통해 조사하였다. 연구 결과, 방송 프로그램을 시청할 때 가장 선호하는 플랫폼은 스마트폰과 TV로 나타났다. 장르에 대한 선호도는 버라이어티쇼, 영화, 드라마 순으로 나타났다. TV로 시청할 때 선호하는 장르는 버라이어티쇼와 드라마 순이었고, 스마트폰으로 시청할 때는 드라마와 애니메이션 순으로 나타났다. 마지막으로 실험자가 선호하는 프로그램 장르와 플랫폼을 통해 시청하는 장르 간에는 유사한 선호도를 보였으나, 스마트폰을 통해 시청하는 경우에는 다른 선호도 형태를 보였다.

전범수(2011)는 매체별 장르 선호도가 각각의 채널별 시청 정도에 끼치는 영향을 조사하였다. 조사를 위해 성인 500명을 대상으로 개별 면접 조사를 통한 설문조사를 실시하였다. 연구 결과로 먼저 지상파와 케이블TV의 장르별 선호도를 응답자 특성과 매체별로 구분하여 나타내었다. 더불어, 매체별 장르별 선호도가 채널 별 시청 정도에 끼치는 영향을 살펴본 결과, 지상파 방송 채널들은 뉴스/시사 분야와 드라마 장르를 중심으로 경쟁 시장이 구축되었다. 이는 지상파 모든 채널들 간에 드라마와 뉴스/시사 프로그램 편성 경쟁이

치열하다는 것을 보여준다. 박소라(2003)는 방송 프로그램 장르 다양성의 여러 차원 가운데 장르의 다양성과 시장 상황 간의 관계를 조사하였다. 전체적으로 방송 프로그램의 다양성 추세를 살펴보고, 경쟁의 정도에 따라 다양성 지수는 어떻게 변하는지를 조사하였다. 조사에서 방송사의 매년 한 주간 방송되었던 프로그램들을 분석 대상으로 하여 그 프로그램들을 15가지 장르로 구분하여 다양성의 변화 추이를 살펴보았다. 연구 결과, 방송 프로그램 장르의 다양성은 다소 줄거나 유지되었고, 경쟁이 심한 시대에서는 더욱 다양성이 감소하는 경향을 보였다. 즉, 경쟁의 도입은 일시적으로 다양성을 감소시키고 있었다.

Kurapati and Gutta(2002)는 프로그램의 즉각적인 개인화 서비스를 제공하기 위하여 시청자가 기존에 시청한 프로그램 시청 패턴을 반영한 프로필을 기반으로 개인화된 프레임워크를 제안하였다. 여기서 프로그램 시청 패턴은 보통 클러스터의 관점에서 각 클러스터가 특정 패턴을 나타내는 방송 프로그램의 특정 세그먼트로 구성된다. 프로그램 시청 패턴을 클러스터링 하는데 7명의 시청자들의 실제 프로그램 시청 기록을 사용하였다. 연구 결과, 시청자들에게 즉각적인 개인화 서비스를 제공하는 프로그램 추천 및 프로필 기반 프레임워크를 개발하였다. Hara et al.(2004)는 시청자가 과거에 시청한 프로그램들이 해당 시청자의 프로그램 선호도를 가장 잘 나타낸다고 가정하고 이를 조사하였다. 이를 검증하기 위하여 1600명을 대상으로 설문조사를 실시하고 본 데이터를 분석에 사용하였다. 연구 결과, 8개의 시청자 그룹을 나타내었고 이는 시청자의 텔레비전 접촉, 그들이 선택한 프로그램을 시청한 동기, 그리고 텔레비전 이외의 그들의 관심사를 설명한다. Geerts et al.(2008)는 방송 프로그램 장르가 소셜 iTV의 사용에 어떠한 역할을 하는지를 조사하였다. 연구 결과, 뉴스, 퀴즈 및 스포츠 프로그램들이 시청할 때 가장 많이 이야기하는 장르로 동시 발생하는 소셜 iTV 시스템

에 적합한 것을 나타내었다. 동시에 발생하지 않는 소셜 iTV의 경우에는 영화, 뉴스, 다큐멘터리 및 음악이 잠재적으로 인기있는 장르로 나타났다. 이를 통하여, 소셜 iTV 설계 및 평가에 어떻게 영향을 끼칠지에 대해 논의 되었다.

대부분의 방송 프로그램 관련 연구들은 매체 구분에 따른 장르(유형) 선호도(김종무 2015; 전범수 2011), 프로그램의 장르(유형) 구분을 위한 제안(Pan et al. 2014; 주창윤 2004), 그리고 장르(유형) 분류와 시청패턴, 시장상황, 매체와의 관계(Geerts et al. 2008; 박소라 2003; 주영호·황성연 2004)를 다루고 있었다. 하지만, 최근 디지털 콘텐츠 기반의 프로그램 시청 증가와 관련하여 이를 기반으로 프로그램의 장르(유형)를 이해하고 새롭게 분류한 연구는 없었다. 이와 유사하게 프로그램 장르(유형) 구분을 위한 기준 및 시스템을 제안한 연구로 주창윤(2004)은 그 동안 모호하게 분류되어 왔던 방송 프로그램 장르를 명확히 구분하기 위하여 기존 방송 프로그램의 장르 분류와 장르 별 유사성 인식에 대한 기준이 무엇인지를 조사하였다. 조사를 위해 정성적/ 정량적 연구 방법을 사용되었다. 먼저, 문헌 조사와 방송 관련 종사자의 심층 집단 토론을 통하여 기존 프로그램 장르를 구분하였다. 이를 통해 도출된 결과는 방송 관련 종사자와 시청자를 대상으로 설문조사를 실시하여 실증적으로 분석하여 방송 프로그램의 장르 구분을 위한 새로운 구분 기준을 설정하고, 프로그램 장르와 각각의 하위 장르의 구분 기준을 제시하였다. Pan et al.(2014)은 기존의 방송 프로그램을 분류하는 분류 시스템이 가지는 한계점을 극복하기 위하여 새로운 프로그램 분류 시스템을 제안하였다. 새로운 분류 시스템에서는 방송 프로그램의 태그들을 기반으로 프로그램들을 분류하는 것이다. 해당 연구에서 다차원 텔레비전 프로그램 태깅을 구현하는 정규 표현 매칭, 웹사이트 인터페이스 연결, 웹 크롤러 세 가지 방법을 제안하고, 이를 통하여 방송 프로그램 태그들이 다

차원적으로 조합되어 분류될 수 있게 하였다. 연구 결과, 기존의 전통적인 분류 시스템과 비교하여 본 연구를 통해 새롭게 제안한 시스템이 더 이점이 있는 것으로 밝혀졌다. 하지만, 기존 연구들은 TV 기반 방송 프로그램 장르(유형)를 기존 분류 방식과 다르게 재 분류하는 것에만 초점을 둔 것으로 최근 증가하고 있는 디지털 콘텐츠 기반의 프로그램 시청과 관련하여 실제 시청자들의 시청 기록을 이용하여 프로그램의 유형을 재 분류하고 이를 해석한 연구는 없었다. 그리하여, 본 연구에서는 온라인 매체를 통해 방송 프로그램을 시청하는 시청자들의 실제 데이터를 기반으로 프로그램 유형을 새롭게 분류하고, 이를 통해 분류된 프로그램 유형들의 해석 및 활용 방안을 제시한다. 특히, 디지털 콘텐츠를 기반으로 방송 프로그램을 새롭게 분류하기 위해 실제 시청자들의 로그 데이터를 이용하여 프로그램 유형을 분류했다는데 의의가 있다.

3. 연구방법론

3.1 클러스터링 기법

클러스터링 기법은 데이터 집합이 가지고 있는 구조를 발견하는 것으로 다차원 공간상에서 유사한 객체 집단을 식별하는 다변량 통계 기법에 속하는 방법이다(Xu and Kim 2008). 이 기법은 유사도가 높은 객체 끼리 군집을 K개로 분류하고 각 군집에 속한 객체 간의 유사성과 다른 군집의 객체 간의 차이를 보여준다(Tombros et al. 2002). 대상들을 분류하기 위한 기준이 모호할 경우, 다양한 특징을 지닌 객체를 군집화할 수 있다는 장점이 있다(강남준 등 2008). 클러스터링 기법은 마케팅, 경영, MIS 분야 등 다양한 분야에서 활용되어 왔다. 일반적으로 클러스터링 기법은 비계층적 기법과 계층적 기법으로 분류할 수 있다. 비계층적 기법은 연구자가 정한 군집의 수에 따라, seed값을 시

작으로 각 객체간의 거리를 최소화 할때까지 분류하는 방식으로 계산 복잡도가 낮다는 장점이 있다(이상훈·한치근 2015; 한승희 2009). 계층적 기법은 트리 구조로 유사성이 높은 객체끼리 계속 클러스터를 확대해 나가는 기법이다. 비계층적 기법에 비해 결과가 안정적이라는 장점이 있다(Tombros et al. 2002; 한승희 2009). 하지만, 해당 기법들은 속성들의 형태가 다르거나 같은 형태의 속성이라도 값의 범위가 다양할 경우 거리 측정 기준을 설정하는데 어려움이 따른다. 즉, 선형으로 분류할 수 없는 데이터(non-convex data)를 변환하여 처리할 수가 없다는 한계점이 있다. 이에 대한 대안으로 선형으로 분류할 수 없는 데이터(non-convex data)를 변환하여 처리 할 수 있는 스펙트럴 클러스터링(Spectral clustering)과 같은 머신러닝 기법들이 학계와 실무에서 각광을 받고있다(허경용 등 2008).

스펙트럴 클러스터링은 머신러닝 알고리즘 기반의 클러스터링 기법 중의 하나로 라플라시아(Laplacian) 행렬을 이용한 그래프 분할 접근법을 활용해서 군집을 분류한다. 스펙트럴 클러스터링은 주어진 데이터 포인트들 간의 유사도 매트릭스(Affinity Matrix)를 구성하고, 이 매트릭스의 고유값(eigenvalue)과 고유벡터(eigenvectors)를 이용해 원본 데이터를 분류한다(Ng et al. 2002). 특히, 선형으로 분류할 수 없는 데이터를 선형 분류가 가능한 데이터로 치환하여 처리할 수 있다는 장점으로 학계와 업계에서 주목을 받고 있다. 이런 장점으로 기존 데이터 마이닝 기법으로 분류가 어려웠던 이미지 인식 분야나 복잡한 사용자 이용 패턴 분류에 탁월한 결과를 내고 있다(허경용 등 2008).

본 연구에서는 디지털 콘텐츠를 기반으로 방송 프로그램을 시청하는 시청자들의 시청 로그를 기반으로 방송 프로그램을 분류하기 위하여 클러스터링 기법을 사

용하고자 한다. 디지털 콘텐츠 기반에서 시청자들이 인식하는 방송 프로그램 분류가 명확하지 않은 상황에서 시청자 로그 데이터를 이용한 클러스터링 분류는 유용한 방법론이 될 수 있다. 여기서 시청 로그 데이터는 시청자의 관심과 선호를 기반으로 나타난 것으로 유사한 시청 패턴을 가진 방송 프로그램 집단은 고유한 유형을 형성하게 된다(이경렬·김형중 2017). 즉, 본 연구에서는 시청 로그를 활용한 클러스터링 기법을 통하여 시청자 인식 기반의 디지털 전용 방송 프로그램의 유형을 새롭게 분류하고자 한다.

3.2 데이터 수집

본 연구를 위해 온라인 미디어랩사인 '스마트미디어랩(SMR)'⁶의 협조를 받아 온라인 방송 프로그램 시청 로그 데이터를 사용하여 분석을 진행하였다. 스마트미디어랩사에서는 (1) 라이브(Live) 형태의 온에어, (2) 비디오 형태의 VoD, (3) 5분 내외의 짧은 하이라이트 형태의 TV 클립(Clip)의 다양한 디지털(동영상) 콘텐츠로 방송 프로그램을 제공하고 있으므로 본 연구의 데이터 수집을 위하여 활용하기에 적합하다고 판단하였다. 특히, 스마트미디어랩(SMR)에서는 월 평균 50억 건 이상의 동영상과 광고 로그 데이터를 수집하고 있으며, 해당 로그 데이터는 시청자들이 사용하는 플랫폼과 콘텐츠 유형(Live/VoD/Clip)에 따라 분류되어 저장하고 있다. 이는 시청자들이 온라인에서 방송 콘텐츠를 시청할 때마다 비식별화 처리된 개인 식별자와 함께 생성된다. 개인 식별자 코드를 통해 각 사용자의 동영상 시청 기록(Watching History)과 성·연령 정보를 확인할 수 있다. 본 연구에서는 스마트미디어랩이 2019년 5월 25일부터 6월 21일까지 수집한 동영상 로그 데이터 가운데, 무작위 기법을 통해 샘플 식별자 1,573,209명의

5) 스마트미디어랩은 2014년 SBS와 MBC가 공동 투자해 설립한 온라인 미디어랩사로 국내 주요 방송사인 SBS, MBC, KBS, CJ E&M, JTBC 등과 파트너십을 맺고 방송 콘텐츠에 온라인 광고 판매를 대행하고 있음. 현재, 네이버/카카오, POOQ, CJ헬로비전/HCN 등과 같은 다양한 디지털 케이블 서비스들과 협력하여 방송 콘텐츠를 제공하고 있으며, 월 평균 100억원 이상의 취급고를 달성하며 국내 대표적인 미디어랩사로 성장 중임

<표 2> 분석에 활용된 TV프로그램 목록 (2019년 5월 25일 ~ 2019년 6월 21일)

번호	프로그램명	장르	방송사	샘플 조회수
1	봄밤	드라마	MBC	2,797,613
2	프로듀스 X 101	예능	CJ ENM	1,886,285
3	바람이 분다	드라마	JTBC	1,851,010
4	검색어를 입력하세요 WWW	드라마	CJ ENM	1,815,692
5	슈퍼맨이 돌아왔다	예능	KBS	1,688,425
6	나 혼자 산다	예능	MBC	1,647,196
7	아스달 연대기	드라마	CJ ENM	1,588,411
8	단 하나의 사랑	드라마	KBS	1,252,547
9	강식당2	예능	CJ ENM	875,839
10	동상이몽2_너는 내 운명	예능	SBS	826,656
11	황금어장-라디오스타	예능	MBC	801,712
12	백종원의 골목식당	예능	SBS	755,531
13	아는 형님	예능	JTBC	687,185
14	슈퍼밴드	예능	JTBC	685,404
15	미운 우리 새끼	예능	SBS	578,289
(중략)				
78	플레이어	드라마	CJ ENM	1,309
79	Show Me The Money 8	예능	CJ ENM	1,183
80	니가 알던 내가 아냐 V2	예능	CJ ENM	615
81	더 콜2	예능	CJ ENM	486
82	닥터탐정	드라마	SBS	18

27,655,017개 데이터를 활용하였다. 로그 데이터 1개당 콘텐츠 유형에 따라 별도의 가중치를 주지 않고 재생 수 1회로 처리하였다. 로그데이터를 통해 1회 이상 재생된 프로그램은 총 82개이며(<표 2> 참조), 여기에는 예고편 영상을 통해 제공되고 있는 방송 예정 프로그램도 포함되어 있다.

3.3 연구설계 및 분석 절차

본 연구의 데이터 분석 절차는 아래 <그림 1>과 같이 5단계로 진행되었다. 사용자 로그 데이터를 바탕으로 (1) 데이터 탐색, (2) 변수 선택, (3) 유사도 매트릭스 생성, (4) 군집 알고리즘 및 군집 K의 결정 그리고 마지막으로 (5) 군집 시각화 및 군집 명 결정의 순서로 진

행된다. (1) ‘데이터 수집’ 단계에서는 로그 데이터를 분석에 적합하도록 매트릭스 구조를 변형 시키고 데이터 탐색을 하였다. 로그 데이터의 남녀 비율을 확인해 본 결과, 여성(54.8%)의 비율이 남성(45.2%)보다 다소 높게 나타났다. 연령 비율은 30대(29.8%)가 가장 높았으며, 그 다음으로 40대(26.3%), 20대(23.4%), 50대이상(14.3%), 10대(6.3%) 순이었다. 콘텐츠 유형 비율은 TV클립(72.5%)이 가장 많았으며, 그 다음으로 VoD(15.6%)와 라이브(12.0%) 순이었다.

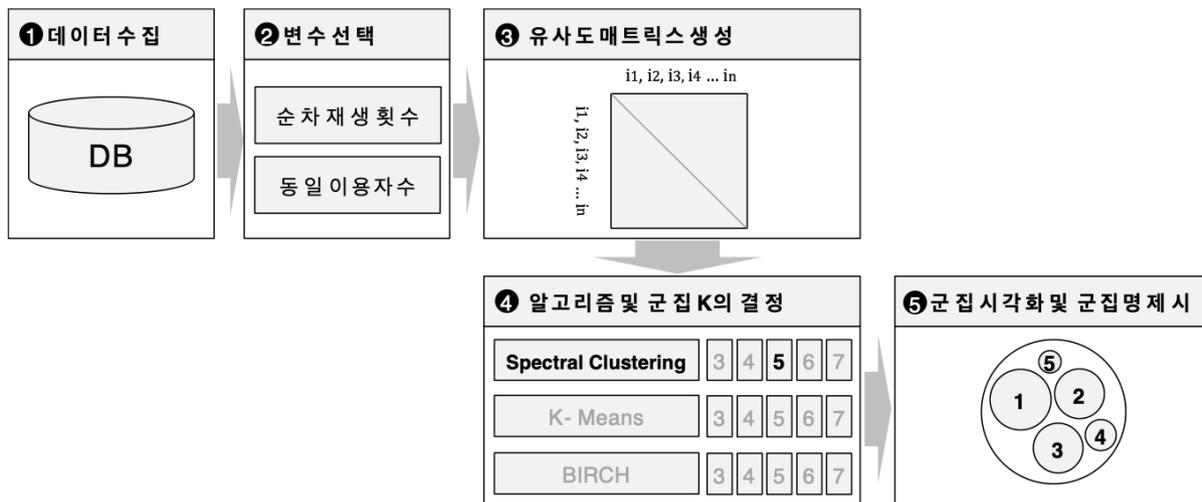
(2) ‘변수 선택’ 단계에서는 앞 단계에서 수집한 데이터 가운데 클러스터링 분석에 활용할 변수를 선택하였다. 본 연구에서는 사용자들의 온라인에서 방송 프로그램 소비 패턴에 따른 최종 군집의 특징 해석을 용

이 할 수 있도록 순차재생횟수와 동일 이용자수를 채택하였다. 순차재생횟수란 개인 식별자 코드별로 시청한 프로그램 중 전/후로 시청한 프로그램간의 재생횟수의 합으로 하였다. 프로그램 간의 전/후 관계가 없는 경우에는 0으로 처리하였다. 동일 이용자수는 프로그램 별로 이용자 동일한 이용자가 동시 발견된 수(Co-occurrence)로 처리하였다. 단, 두 변수 모두 이용자 세션 별로 분리하여 집계했다. 변수 선택 후에 (3) '유사도 매트릭스(Affinity Matrix)'를 생성하였다. 유사도 매트릭스란 각 아이템간의 유사도를 측정된 행렬 구조를 의미한다. 스펙트럴 클러스터링 분석을 위해서는 유사도 매트릭스를 생성하는 것이 필수이다. 앞 단계에서 선택된 변수의 속성이 모두 연속형이므로 코사인 유사도를 이용하여 유사도 매트릭스를 생성했다.

그리고 본 연구에서는 (4) '군집 알고리즘 및 군집 K의 결정'을 위해서는 실루엣 평가를 실시하였다. 실루엣 평가는 군집된 데이터 사이의 거리와 다른 군집간의 거리를 고려하여 군집이 명확하게 되었는지를 실루엣 값으로 비교하는 방법이다. 실루엣 값은 -1에서 1사이의 값이 나오며, 1에 가까울수록 클러스터링 분석이 잘 된 것이다(박성진 등 2017). 클러스터링 알고리즘 선택을 위해 비계층적 기법의 대표적인 알고리즘인 K-means,

계층적 기법의 대표적인 알고리즘인 Birch, 마지막으로 머신러닝 알고리즘인 스펙트럴 클러스터링 기법을 비교하였다. 또한 클러스터링 기법은 군집의 개수 K값을 어떻게 정하느냐에 따라 군집의 결과가 달라지므로 K값에 따른 실루엣 값도 함께 비교하였다. 초기값에 따라 결과가 달라질 수 있어, 최소 10회 이상의 클러스터링을 반복 수행하고 가장 좋은 결과를 선택하였다. 분석은 파이썬(Python) 3.5의 기본 패키지와 분석 라이브러리인 scikit-learn을 이용하였다. 그 결과 스펙트럴 클러스터링 기법으로 군집의 수가 5개 일 때 가장 우수한 결과(0.504)가 나왔다. 이 같은 결과는 전통적인 통계 기법에 비해 머신러닝 알고리즘이 보다 정확한 분류가 가능하다는 것을 의미한다.

마지막으로 (5) '군집 시각화 및 군집명 제시'의 단계에서는 분석된 군집들을 시각화 톨로 표현하고 각 군집에 적합한 이름을 정하였다. 클러스터링 시각화는 군집간의 거리는 코사인 유사도를 활용하고 군집의 크기는 재생수로 표현하였다. 군집 명의 선정은 각 TV프로그램들의 핵심키워드와 그들 간의 관계를 고려하여 연구자들 간의 토의 과정을 걸쳐 결정하였다. 이 과정에서 단어 빈도 분석과 키워드 추출 분석을 통해 도출된 키워드를 빈도수와 가중치를 활용하였다. 특히, 키워드 추



<그림 1> 데이터 분석 절차

출을 통해 5위까지의 주요 키워드들을 도출하여 해당 키워드들이 각 프로그램의 핵심 내용 및 컨셉을 나타내는지를 확인하였다(이소현 등 2019).

4. 분석 결과

클러스터링 기법을 통해 개발한 알고리즘을 통해 디지털 콘텐츠를 기반으로 한 방송 프로그램의 유형을 새롭게 분류한 결과와 이를 시각화하여 아래의 <표 3>와 <그림 2>로 나타내었다. 클러스터링을 반복해서 수행한 결과, 스펙트럴 클러스터링 기법으로 군집의 수가 다섯 개였을 때 가장 우수한 결과가 도출되어 총 프로그램을 다섯 개의 군집으로 분류하고, 각 군집에서 세부 주제들을 추가 분류하였다. 분류된 다섯 개의 군집들의 군집명은 각 군집이 포함하는 프로그램들을 소

개하는 시놉시스의 핵심 키워드들과 그들 간의 관계를 고려하여 선정하였다. 이 과정에서 핵심 키워드는 단어 빈도 분석과 키워드 추출 분석을 통해 빈도수와 가중치가 높은 키워드를 도출한 후 연구자들 간의 토의를 통해 최종 군집명과 주제어를 결정하였다.

첫 번째 군집의 ‘생활’은 핵심 키워드로 생활, 며느리, 결혼, 집안일, 대모, 여성, 반찬, 시장, 전수 등이 도출되었고, 이를 기반으로 ‘살림’과 ‘음식’의 두 가지 주제로 다시 분류할 수 있다. 즉, 핵심 키워드 생활, 며느리, 결혼, 집안일, 대모, 여성 등을 통하여 ‘살림’의 주제어를 도출하였고, 여기에 해당하는 프로그램은 살림하는 남자들, 이상한 나라의 며느리, 생활의 달인 등이 있었다. 주제어 ‘살림’에서는 생활에 필요한 정보와 상식을 제공하거나 집 안에서 일어나는 이야기를 콘텐츠로 제공하는 것을 나타낸다. 다음으로 반찬, 시장, 전수 등의 키워드들을 통하여 ‘음식’의 주제어를 도출하였고, 여기에

<표 3> 프로그램 별 관련 키워드 기반의 군집명 제시

구분	군집명	주제어	키워드	관련 방송 프로그램
군집1	생활	살림	생활, 며느리, 결혼, 집안일, 대모, 여성	살림하는 남자들, 이상한 나라의 며느리, 생활의 달인
		음식	반찬, 시장, 전수	수미네 반찬, 수요미식회
군집2	음악	음악	음악, 뮤직, 무대, 노래, 아이돌, 장르, 음원, 인기가요	쇼 음악중심, 뮤직뱅크, SBS 인기가요, M Countdown
군집3	소통	가족	엄마, 아빠, 아들, 자식, 부부, 가족, 육아	슈퍼맨이 돌아왔다, 세상에서 제일 예쁜 내 딸, 동상이몽-너는 내 운명
		토크	감동, 웃음, 사연, 토크, 공감, 고민	미운오리새끼, 황금어장-라디오스타, 아는형님, 대국민 토크쇼 안녕하세요, 마이 리틀 텔레비전, 전지적 참견 시점
군집4	경쟁	능력 (재능)	아이돌, 손흥민, 능력, 검사, 천재, 영재, 대통령	프로듀스X101, 영재발굴단, 보좌관-세상을 움직이는 사람들, 손세이셔널-그를 만든 시간, 호텔 델루나, 60일, 지정 생존자, 으라차차 만수로
		프로젝트 (게임)	대결, 반전, 축구, 작전, 리얼, 게임	복면가왕, 강식당, 뭉쳐야 찬다, 슈퍼히어로, 비긴어게인, 니가 알던 내가 아냐
군집5	회상	사랑	사랑, 봄밤, 로맨스, 운명, 마음, 믿음	봄밤, 단 하나의 사랑
		청춘	사고뭉치, 예측불가, 사랑, 좌절, 실패	불타는 청춘
		역사	혁명, 연대기, 동학, 영웅, 태고	아스달 연대기, 녹두꽃



<그림 2> 디지털 콘텐츠 기반 방송 프로그램 클러스터링 시각화

는 수미네 반찬, 수요미식회 등의 프로그램들이 있었다. 주제어 ‘음식’에서는 음식을 만들거나 먹는 내용의 콘텐츠를 제공하는 것으로 이와 관련해서는 1980년대 요리 전문가가 요리법을 알려주는 정통 요리 프로그램으로 시작하여 현재는 맛집 탐방 및 오락 콘텐츠를 포함하여 요리법을 알려주는 프로그램들이 제공되고 있다 (충청매일 2019).

둘째, 군집 ‘음악’은 핵심 키워드로 뮤직, 무대, 노래, 아이돌, 장르, 음원, 인기가요 등이 도출되었고 여기에 해당하는 프로그램은 쇼 음악중심, 뮤직뱅크, SBS 인기가요, M Countdown 등이 있다. 일반적으로 음악 방송은 가수와 음악가들이 출연하여 자신들의 곡을 선보이는 방송을 나타낸다(나무위키 2019). 하지만, 최근에는 음악을 주제로 한 다양한 장르(유형)의 프로그램들

이 증가하면서 단지 노래와 무대를 제공하는 것 이상으로 음악을 소재로 한 다양한 프로그램들이 방영되고 있다. 셋째, 군집 ‘소통’은 핵심 키워드로 엄마, 아빠, 아들, 자식, 부부, 육아, 감동, 웃음, 사연, 공감, 고민 등이 도출되었고, 이를 기반으로 ‘가족’과 ‘토크’의 두 가지 주제로 다시 분류할 수 있다. 핵심 키워드 엄마, 아빠, 아들, 자식, 부부, 육아 등을 통하여 ‘가족’의 주제어를 도출하였고, 여기에 해당하는 프로그램은 슈퍼맨이 돌아왔다, 세상에서 제일 예쁜 내 딸, 동상이몽-너는 내 운명 등이 있다. 주제어 ‘가족’에서는 육아, 부부 또는 부모와 자식의 관계를 보여주는 가족의 이야기를 콘텐츠로 제공하는 것을 나타낸다. 예를 들어, ‘동상이몽-너는 내 운명’의 프로그램에서는 다양한 분야의 커플들이 살아가는 모습을 보여주고, 두 사람이 함께 사는 것

의 가치를 살펴보는 프로그램으로 부부라는 가족의 이야기를 담고 있다. 다음으로 감동, 웃음, 사연, 공감, 고민 등의 키워드들을 통하여 ‘토크’의 주제어를 도출하였고, 여기에는 미운오리새끼, 황금어장-라디오 스타, 대국민 토크쇼 안녕하세요 마이 리틀 텔레비전 등의 프로그램들이 있었다. 여기에서는 특정 주제, 사연, 고민, 상황 등에 대하여 호스트와 게스트가 자유롭게 이야기함으로써 감동과 웃음을 이끌어내는 것을 나타낸다. 이와 관련한 프로그램은 ‘황금어장-라디오 스타’ 처럼 게스트들의 개인적인 이야기를 끌어내는 독보적 토크 프로그램도 있지만, ‘전지적 참견 시점’이나 ‘미운오리새끼’ 처럼 리얼 일상을 공유하고 이에 대해 이야기하는 다양한 컨셉들이 존재한다.

넷째, 군집 ‘경쟁’은 핵심 키워드로 손흥민, 아이돌, 능력, 검사, 천재, 영재, 대결, 반전, 축구, 작전, 리얼, 게임 등이 도출되었고, 이를 기반으로 ‘능력(재능)’과 ‘프로젝트(게임)’의 두 가지 주제로 다시 분류할 수 있었다. 핵심 키워드 손흥민, 아이돌, 능력, 검사, 천재, 영재 등을 통하여 ‘능력(재능)’의 주제어를 도출하였고, 여기에 해당하는 프로그램은 영재발굴단, 보좌관-세상을 움직이는 사람들, 손세이셔널-그를 만든 시간 등이 있었다. 여기에서는 특정 능력이나 재능을 기반으로 이야기를 전개해 나가거나 실제 다른 능력이나 재능을 가진 사람들의 생활을 내용으로 스토리를 제공하는 것을 나타낸다. 예를 들어, 프로그램 ‘보좌관-세상을 움직이는 사람들’에서는 정치 플레이어들의 이야기를 드라마로 보여주고 있으며, 프로그램 ‘영재 발굴단’과 ‘손세이셔널-그를 만든 시간’은 주인공들의 재능이나 성공의 리얼 스토리를 보여준다. 다음으로 대결, 반전, 축구, 작전, 게임 등의 키워드들을 통하여 ‘프로젝트(게임)’의 주제어를 도출하였고, 여기에는 프로듀서 X 100, 복면가왕, 뭉쳐야 찬다, 니가 알던 내가 아냐 등의 프로그램들이 있었다. 여기에서는 퀴즈, 대결 등과 같은 컨셉으로 프로젝트 또는 게임을 콘텐츠로 제공하는 프로그램들을 나

타낸다. 이와 관련한 프로그램으로 ‘프로듀서 X 100’은 경쟁을 기반으로 한 글로벌 아이돌을 육성하는 것이고, ‘비긴어게인’은 국내 최정상의 뮤지션들이 해외의 낯선 도시에서 버스킹에 도전하는 것이며 ‘니가 알던 내가 아냐’는 주인공의 최 측근과 비 관계자의 퀴즈 대결을 보여준다.

마지막으로 군집 ‘회상’은 핵심 키워드로 사랑, 봄밤, 로맨스, 운명, 마음, 예측불가, 좌절, 실패, 사고뭉치, 혁명, 연대기, 동학, 영웅 등이 도출되었고, 이를 기반으로 ‘사랑’, ‘청춘’ 그리고 ‘역사’의 세 가지 주제로 다시 분류할 수 있었다. 핵심 키워드 사랑, 봄밤, 로맨스, 운명, 믿음 등을 통하여 ‘사랑’의 주제어를 도출하였고, 여기에 해당하는 프로그램은 봄밤, 단 하나의 사랑 등이 있었다. 또한, 핵심 키워드 사고뭉치, 예측불가, 좌절, 실패 등의 키워드들을 통하여 ‘청춘’의 주제어를 도출하였고, 여기에는 불타는 청춘 프로그램이 있었다. 즉, 여기서는 사랑이나 청춘을 주제로 연인 간의 로맨스 이야기나 젊은 시절의 추억을 회상하는 컨셉으로 스토리를 전개해 나가는 것을 나타낸다. ‘사랑’을 주제로 한 프로그램은 드라마가 가장 많지만, 최근에는 실제 커플들의 이야기 뿐 아니라 처음 만나는 남녀의 데이트를 보여주는 관찰 프로그램도 인기를 얻고 있다. 그리고 주제어 ‘청춘’과 관련된 프로그램의 ‘불타는 청춘’에서는 중견 스타들이 자연스럽게 친구가 되는 과정 속에서 자신들의 과거의 추억을 회상하고 이야기하는 내용을 담고 있다. 마지막으로 혁명, 연대기, 동학, 영웅 등의 키워드들을 통하여 ‘역사’의 주제어를 도출하였고, 여기에는 아스달 연대기, 녹두꽃 등의 프로그램들이 있었다. 주제어 ‘역사’에서는 과거의 역사와 관련된 정보를 제공하거나 역사를 기반으로 한 이야기를 제공하는 것을 나타낸다.

5. 결과 토의 및 시사점

5.1 연구결과 토의 및 활용 방안

본 연구에서는 변화된 미디어 환경에 따른 시청자들의 디지털 방송 프로그램 시청 패턴을 분석한 후 사용자 인식 기반의 디지털 방송 프로그램을 분류를 하였다. 사용자 시청 패턴 분석은 다양하고 빠르게 변화되는 시청자의 시야를 정확하게 이해할 수 있었다. 또한, 머신러닝 분석은 기존의 전통적인 데이터 분석에 비해 군집 분류 성능이 뛰어나 보다 객관적이고 과학적인 분류가 가능함을 증명함으로써 머신러닝 분석 방법의 유용성을 제시하였다. 추가로, 본 연구에서는 머신러닝 분석 결과를 토대로 플랫폼사/콘텐츠사 차원/광고회사 차원에서의 구체적인 방안 및 전략을 아래 <표 4>과 같이 제시한다.

첫째, 플랫폼사 차원에서는 온라인 전용 방송 프로그램 분류를 활용한 ‘온라인 맞춤형 채널’ 구축을 제안한다. 최근 OTT 서비스에서 다양한 온라인 맞춤형 채널

을 제공하고 있다. 일 예로 콘텐츠연합플랫폼의 poog에서는 MBC ‘무한도전 채널’, KBS ‘1박 2일 채널’과 같이 인기있는 방송 프로그램을 24시간 제공하는 온라인 전용 채널을 제공하고 있다. 온라인 전용 채널은 사용자가 좋아하는 프로그램을 계속 시청할 수 있다는 장점이 있지만, 프로그램 다양성이 부족해 일부 매니아를 제외한 일반 시청자를 동인 할 요소가 다소 부족하다. 본 연구에서 새롭게 분류한 방송 프로그램에 따라 해당 채널을 편성하는 방안은 시청자들이 관심에 따라 다양한 프로그램을 지속적으로 소비하게 하는 마케팅 전략이 될 수 있다. 또한, 시청자의 관심사(주제어)를 기반으로 해당 주제어가 속하는 군집의 다른 주제어의 프로그램들을 함께 제공할 수 있다. 예를 들어, ‘슈퍼맨이 돌아왔다’와 같은 주제어 ‘가족’의 프로그램들 주로 시청하는 시청자에게 해당 프로그램이 속한 군집 ‘소통’의 다른 주제어인 ‘토크’에 해당하는 프로그램들을 함께 제공해 줄 수 있다. 그 밖에 모바일을 통해 짧은 동영상 소비가 늘어난 시청 변화에 맞춰 5분 내외의

<표 4> 플랫폼사/콘텐츠사/광고회사 차원의 활용 방안 제안

구분	사용자 로그 기반의 디지털 방송 프로그램 유형 분류 활용 방안	
	활용 방안	내용
플랫폼사 차원	① 온라인 맞춤형 채널(Live) 구축	<ul style="list-style-type: none"> • 각 군집에 속한 방송 프로그램이 계속 재생되는 온라인 전용 채널 구축 • 시청자의 관심사(주제어)를 기반으로 한 맞춤형 채널 제공
	② 추천 알고리즘 개발	<ul style="list-style-type: none"> • 동영상 추천 알고리즘을 활용하여, 시청자들에게 최적의 방송 프로그램 추천 • 시청 기록이 없는(Cold-Start issue) 비로그인 또는 초기 시청자를 대상으로 한 주제별 방송 프로그램 제안
콘텐츠사 차원	③ 유통 판매 상품 기획	<ul style="list-style-type: none"> • VoD 및 OTT 서비스에 유통 시, 온라인 전용 패키지 상품 기획에 활용 • 해외 유통 판매 시 번들링(Bundling) 전략을 통한 방송 프로그램 가치의 극대화
광고회사 차원	④ 그룹 타게팅 상품 개발	<ul style="list-style-type: none"> • 디지털을 통한 방송 프로그램 광고 집행 시, 그룹 타게팅 상품(방송 프로그램) 개발에 활용 • 브랜드 또는 광고 소재에 적합한 광고 상품(방송 프로그램) 제안
	⑤ 광고 효과 극대화를 위한 최적의 매체 전략 수립	<ul style="list-style-type: none"> • 클러스터 내에 있는 방송프로그램 구매 시 정밀화된 타겟에서 광고 반복 노출 • 클러스터 별로 방송 프로그램을 판매해서 광고 도달을 극대화

하이라이트 영상(TV클립) 단위 채널을 편성하여 변화된 시청자의 방송 시청 의도를 높일 수 있다(조석현·이현지 2016). 다음으로 플랫폼사에서 추천 알고리즘에 본 연구 결과를 활용하는 것을 제안한다. 동영상 플랫폼사들은 사용자들의 체류시간을 높이고자 추천 알고리즘을 적극적으로 사용하고 있다. 하지만 전통적인 추천 시스템은 이전의 시청 기록이나 선호도 기록이 없는 비로그인 또는 초기 시청자에게 적절한 추천을 제공할 수 없는 문제를 가진다. 이러한 문제를 추천시스템 연구에서는 cold-start 이슈라고 정의한다(김동현 등 2016). 이러한 cold-start 이슈를 해결하는 방안으로 본 연구의 분류 방법이 활용될 수 있다. 시청자가 최초로 시청한 프로그램을 중심으로 클러스터링 내에 있는 프로그램들을 추천하여 시청자가 유사성이 높다고 느낄만한 프로그램을 계속 제안할 수 있다. 이후 시청자의 시청 기록이 충분히 쌓인 이후에는 개인별 맞춤형 추천으로 자동으로 변경한다면 추천 시스템의 성능을 향상시킬 수 있다.

둘째, 콘텐츠사(방송사) 측면에서 유통 판매 상품 기획에 활용될 수 있다. 정보통신정책연구원의 2018년도 방송시장 경쟁상황 평가서에 따르면 2011년도부터 2017년도까지 지상파방송사 및 PP 방송프로그램 판매 매출액의 연평균 성장률은 9.1%로 성장 중이다(정보통신정책연구원 2018). 이 같은 결과는 VoD 서비스 및 OTT 사업자 증가와 해외 유통 판매의 증가에 기인하였다. 방송 프로그램 가치를 극대화 하기 위해 방송 프로그램 번들링(Bundling) 전략⁶에 본 연구의 결과를 이용할 수 있다. 콘텐츠사에서 본 연구의 새로운 프로그램 분류체계에 따라 온라인 전용 패키지 단위 판매 기획에 이용할 수 있다. 본 연구에서 제시하는 프로그램 분류체계는 시청자 인식을 기반으로 하였으므로, 해당 분류체계를 통한 그룹 상품 판매는 기존 분류체계와 비교하여 그 만족도가 더 높을 것이다. 해외 유통 판

매 시에도 개별 프로그램 단위의 단위가 아니라 묶음 단위로 판매하여 방송 프로그램 전체의 가치를 극대화할 수 있다.

셋째, 광고회사 측면에서 타게팅 광고 상품 판매에 활용될 수 있다. 최근 방송 프로그램을 온라인으로 시청하는 시청자가 늘면서 방송 프로그램을 활용한 온라인 동영상 광고 시장이 크게 성장하고 있다. 온라인 광고 시장에서 광고주에 적합한 타게팅 상품 개발이 중요한데, 본 연구 결과를 바탕으로 그룹(클러스터링) 타게팅 상품 개발이 가능하다. 기존 연구에 따르면, 해당 브랜드나 광고 소재가 콘텐츠와 관련성이 높을 경우 광고 효과에 긍정적인 영향을 끼친다고 하였다 (Kononova and Yuan 2015). 그러므로, 본 연구 결과를 활용하여 광고 상품을 최적의 광고주에게 제안할 수 있다. 예를 들어 군집 1인 '생활'과 관련한 타게팅 상품을 개발하여 가전용품, 생활용품, 식품 등의 광고주에게 제안할 수 있다. 이는 TV 기반의 프로그램 장르(유형) 분류에서는 드라마, 예능, 시사 등과 같이 프로그램의 성격으로 장르(유형)를 분류했다면, 본 연구에서는 프로그램 콘텐츠 자체로 분류하기 때문에 해당 프로그램과 광고 소재를 연결하는 것이 수월하고 그 효율성이 클 것으로 기대한다. 마지막으로 광고효과 극대화하기 위한 광고 콘텐츠(방송 프로그램) 구매 전략 수립에 이용할 수 있다. 광고 콘텐츠 구매 시 군집 내에 있는 프로그램들을 구매하게 되면 시청 패턴이 유사한 시청자들에게 타겟 할 수 있어서 광고 빈도를 높이는데 도움을 줄 수 있다. 반대로 군집 별로 프로그램들을 구매하게 되면 시청 패턴이 다른 시청자들에게 도달할 수 있어 도달율(Reach)을 높일 수 있다.

5.2 연구의 한계 및 향후 연구방향

본 연구의 한계점과 이를 보완한 향후 연구 방향은 다음과 같다. 첫째, 본 연구에서는 분석 단위가 방송 프

6) 다른 종류의 콘텐츠를 하나로 결합하여 가격을 차별화하는 마케팅 전략

로그로 한정했다는 한계가 있다. 미디어의 발전과 함께 방송프로그램 소비를 라이브, VoD와 같은 방송 프로그램 단위 뿐 아니라 분절화 된 TV클립 소비로 확대되었다. 향후 연구에서 클러스터링 분석을 할 때 클립 단위 분석을 통해 세분화된 시청자 시청 패턴을 분석이 가능할 것이다. 또한 콘텐츠 차원이 아닌 이용자 차원에서 시청 로그 데이터 기반의 이용자 클러스터링 분석도 가능하다. 이용자 기반 클러스터링 분석을 해서 각 군집에 맞는 편성 또는 마케팅 전략에 관한 연구가 향후 진행되길 기대한다. 기존 연구에서는 디지털 라디오 로그 데이터를 활용한 라디오 청취자 유형분석(이경렬·김형중 2017) 연구가 있었으나 디지털 방송프로그램에 관한 시청자 유형 분석 연구는 부족한 상황이다.

둘째, 본 연구에서 제한된 데이터를 활용한 점이 확장된 연구 결과를 기대할 수 있다. 본 연구에서 수집한 로그 데이터 기간이 4주로 비교적 짧다는 한계를 가진다. 방송 프로그램의 방영 기간은 평균 8주 이상이기 때문에 로그 데이터 기간 확대를 통해 추가적인 인사이트를 도출할 수 있다. 또한 군집 명 제시 단계에서 활용한 시놉시스(synopsis) 글자 수는 프로그램 당 300자 내외로 고도화된 텍스트마이닝을 하기에는 글자 수가 부족하였다. 본 연구에서 각 군집에 있는 프로그램 소개 글자를 모두 활용했음에도 키워드 추출이나 단어 빈도 분석과 같은 기초적인 분석에 그쳤다는 아쉬움이 있다. 향후 연구에서는 프로그램 소개 뿐 아니라 관련 SNS 포스트, 프로그램 관련 댓글, 기사와 같은 관련 텍스트를 추가해서 분석해, LDA 토픽 모델링, 네트워크 분석과 같은 추가 분석을 통해 군집 명의 타당성을 높이길 기대해 본다.

셋째, 향후 연구에서는 본 연구에서 제시된 분류 체계와 기존 방송사에서 분류 체계를 비교 및 분석할 수 있다. 방송사(제공자)가 인식하는 분류체계와 시청자(수용자)가 인식하는 분류 체계 간의 차이를 비교해 보고 상호 보완적으로 활용할 방안 에 관한 연구가 필요하

다. 이와 같은 연구는 제공자와 수용자 간의 인식의 차이를 좁히고 확장된 분류 체계를 수립하는데 도움이 될 것이다. 덧붙여 본 연구의 분류체계를 통한 플랫폼사/콘텐츠사/광고회사 차원의 활용 방안의 검증도 필요하다. 실제 비즈니스에서 본 연구의 분류체계를 활용해서 상품이나 전략에 이용했을 때 효과가 있는지 검토가 필요하다. 특히, 본 연구에서 제안한 그룹 타게팅 상품이 개발된다면 광고 효과 측정은 로그 단위로 수집되기 때문에 정량적인 검증을 할 수 있다. 검증 결과는 본 연구의 의의를 높이고 구체적인 활용 방안 설정에 도움을 줄 수 있을 것이다.

5.3 연구의 학술적/실무적 의의

본 연구의 학술적 시사점은 다음과 같다. 첫째, 본 연구는 최근 증가하고 있는 디지털 콘텐츠 기반의 방송 프로그램 시청과 관련하여 실제 시청자들의 시청 기록을 이용하여 프로그램을 분류하고 이를 해석하였다. 기존 관련 연구에서는 TV 기반 방송 프로그램의 장르(유형)를 재 분류하는 것에 중점을 두었으며(Pan et al. 2014; 주창운 2004), 디지털 콘텐츠 기반의 사용자 로그 데이터를 이용하여 프로그램을 분류한 연구는 없었다. 특히, TV 기반 방송 프로그램의 장르(유형) 분류는 방송사와 같은 공급자가 직접 결정하므로 실제 프로그램을 시청하는 시청자의 장르(유형) 인식과 차이가 있을 수 있다. 그리하여, 본 연구에서는 공급자 관점의 주관적 판단을 제외하고 동영상과 같은 디지털 콘텐츠로 방송 프로그램을 시청하는 실제 시청자들의 로그 데이터를 이용함으로써 시청자 인식을 기반으로 프로그램을 분류하였다. 특히 시청자 로그 데이터는 시청자의 관심과 선호를 기반으로 나타낸 것으로 유사한 시청 패턴을 가진 방송 프로그램 집단은 시청자 인식 기반의 고유한 유형을 형성할 수 있다.

둘째, 본 연구에서는 방송 프로그램을 분류하기 위하여 머신러닝과 빅데이터 기반의 클러스터링 기법을 이

용하였다. 이를 통해 데이터 집합이 가지고 있는 구조를 발견한 것으로 유사도가 높은 객체끼리 군집을 분류하였다(Tombros et al. 2002). 이와 관련한 기존 연구에서도 효율적인 마케팅 전략을 세우기 위하여 다량의 고객 데이터를 이용하여 고객을 분류하는데 클러스터링 기법을 이용하였다(이영찬 2018). 특히, 최근 머신러닝 기법들이 학계와 실무에서 활용되고 있어서 본 연구에서는 프로그램 분류를 위해 머신러닝 알고리즘 기반의 스펙트럴 클러스터링 기법을 사용하였다. 본 기법은 선형으로 분류할 수 없는 데이터를 선형 분류가 가능한 데이터로 치환하여 처리할 수 있어서 기존 데이터 마이닝 기법으로 분류가 어려웠던 사용자 이용 패턴 분류에 적합한 기법으로 적용될 수 있었다(허경용 등 2008).

셋째, 개발한 프로그램 분류 알고리즘을 통해 방송 프로그램 분류 후 그 결과를 시각화하고, 단어 빈도 분석과 키워드 추출을 기반으로 각 군집의 군집명을 제시하였다. 즉, 각 군집에서 포함하고 있는 방송 프로그램들을 소개(시놉시스)하는 내용에서 가장 많이 언급되거나 핵심이 되는 키워드들을 도출하여 연구자들 간의 토의 과정을 통해 군집명을 결정하였다. 그리고, 각 군집의 크기와 거리를 고려하여 분류된 군집들을 시각화하여 가시성을 높였다.

실무적 시사점으로는 본 연구에서 개발한 디지털 콘텐츠 기반의 방송 프로그램 분류 알고리즘의 사용과 본 연구에서 도출한 결과를 기반으로 플랫폼사, 콘텐츠사, 광고 회사 차원의 구체적인 전략 및 방안을 제시하였다(<표 3> 참조). 즉, 본 연구에서는 디지털 콘텐츠를 기반으로 방송 프로그램을 새롭게 분류했을 뿐만 아니라 관련 산업 분야에서 이를 어떻게 활용할 수 있는지에 대한 구체적인 방안 및 전략을 제시하였다. 이는 변화하고 있는 미디어 사용 패턴에 맞추어 방송 프로그램을 새롭게 분류하고 활용함으로써 관련 산업 분야의 효과적인 마케팅 전략 수립 및 가치를 극대화할 수

있다. 더불어 이를 기반으로 지식경제 사회를 주도하는 핵심 산업으로 언급되고 있는 디지털 콘텐츠 사업에 장르(유형) 간 융합을 일으킴으로써 새로운 시너지 효과를 일으킬 수 있다(신용재·이동현 2016).

참고 문헌

[국내 문헌]

1. 강남준, 이종영, 이혜미 2008. “군집분석 방법을 사용한 미디어 레퍼토리 유형분석,” *한국방송학보* (22:2), pp. 7-46.
2. 김동현, 박찬영, 오진오, 유환조 2016. “추천시스템에서의 Cold-Start 문제에 관한 연구 동향,” *정보과학회지* (34:6), pp. 16-21.
3. 김종무 2015. “TV 시청 플랫폼 차이에 따른 장르 선호도 분석: TV와 스마트폰 플랫폼 중심으로,” *한국디자인문화학회지* (21:2), pp. 151-160.
4. 김희경 2004. “텔레비전 다큐멘터리 수용자의 기대지평에 따른 텍스트 해석의 차이에 관한 연구,” *한국언론학보* (48:6), pp. 166-195.
5. 박성진, 강범수, 김성수, 김상균. 2017. “데이터 클러스터링을 활용한 게이미피케이션 환경에서의 플레이어 유형 분석,” *한국게임학회 논문지* (17:6), pp. 77-88.
6. 박소라 2003. “경쟁 도입이 텔레비전 프로그램 장르 다양성에 미치는 영향에 대한 연구: 1989년 이후 지상파 방송 편성표 분석을 통하여,” *한국언론학보* (47:5), pp. 222-250.
7. 신용재, 이동현 2016. “디지털 문화 콘텐츠 산업이 지식경제사회에 미치는 파급효과 분석,” *지식경영연구* (17:1), pp. 73-89.
8. 원용진 1997. “장르 변화로 읽는 사회: 인기 드라마

- < 모래성> 과< 애인> 을 중심으로,” *언론과 사회*, pp. 100-133.
9. 이경렬, 김형중 2017. “스마트폰 앱 로그데이터를 활용한 라디오 청취자 유형분석: SBS 고릴라(Gorealra) 를 중심으로,” *방송과 커뮤니케이션* (18:4), pp. 5-49.
 10. 이상훈, 한치근 2015. “스펙트럴 클러스터링 기반 커뮤니티 통합이 그래프의 중심성 지표에 미치는 영향 분석,” *한국인터넷정보학회 학술발표대회 논문집*, pp. 281-282.
 11. 이소현, 손새아, 김희웅 2019. “텍스트마이닝을 이용한 미용성형 주요 요인에 관한 연구,” *지식경영연구* (20:1), pp. 45-75.
 12. 이영찬 2018. “군집분석과 연관규칙을 활용한 고객 분류 및 장바구니 분석: 소매 유통 빅데이터를 중심으로,” *지식경영연구* (19:4), pp. 59-76.
 13. 전범수 2011. “매체별 장르 선호도가 지상파방송 채널 시청 정도에 끼치는 영향,” *방송통신연구*, pp. 148-170.
 14. 정보통신정책연구원 2018. *2018년도 방송시장 경쟁상황 평가*, 방송통신위원회.
 15. 조석현, 이현지 2016. “방송클립동영상 이용이 방송 시청의도에 미치는 영향,” *한국콘텐츠학회논문지* (16:10), pp. 645-655.
 16. 주영호, 황성연 2004. “시청행태에 근거한 프로그램 유형분류의 유효성에 관한 연구: 시청행태와 유형분류의 일치성을 중심으로,” *한국방송학보* (18:1), pp. 217-244.
 17. 주창윤 2004. “텔레비전 프로그램 장르 분류기준에 관한 연구,” *방송통신연구*, pp. 105-136.
 18. 한승희 2009. “연관 태그의 군집화를 위한 클러스터링 기법 비교 연구,” *한국문헌정보학회지* (43:3), pp. 399-416.
 19. 허경용, 김광백, 우영운 2008. “스펙트럼 군집화에서 블록 대각 형태의 유사도 행렬 구성,” *멀티미디어학회논문지* (11:9), pp. 1302-1309.

[국외 문헌]

1. Geerts, D., Cesar, P., and Bulterman, D. 2008. “The Implications of Program Genres for the Design of Social Television Systems,” in *Proceedings of the 1st International Conference on Designing Interactive User Experiences for TV and Video*, ACM, pp. 71-80.
2. Hara, Y., Tomomune, Y., and Shigemori, M. 2004. “Categorization of Japanese TV Viewers Based on Program Genres They Watch,” *User Modeling and User-Adapted Interaction* (14:1), pp. 87-117.
3. Kononova, A., and Yuan, S. 2015. “Double-Dipping Effect? How Combining YouTube Environmental PSAs with Thematically Congruent Advertisements in Different Formats Affects Memory and Attitudes,” *Journal of Interactive Advertising* (15:1), pp. 2-15.
4. Kurapati, K., and Gutta, S. 2002. “Instant Personalization via Clustering TV Viewing Patterns,” *IASTED’s ASC*.
5. Narducci, F., Musto, C., de Gemmis, M., Lops, P., and Semeraro, G. 2018. “TV-Program Retrieval and Classification: A Comparison of Approaches Based on Machine Learning,” *Information Systems Frontiers* (20:6), pp. 1157-1171.

6. Ng, A. Y., Jordan, M. I., and Weiss, Y. 2002. "On Spectral Clustering: Analysis and an Algorithm," in *Advances in Neural Information Processing Systems*, pp. 849-856.
7. Pan, X., Yin, F., and Chai, J. 2014. "Delaying Tagging of Television Programs and Association Rule Mining," in *2014 IEEE 17th International Conference on Computational Science and Engineering*, IEEE, pp. 192-197.
8. Tombros, A., Villa, R., and Van Rijsbergen, C. J. 2002. "The Effectiveness of Query-Specific Hierarchic Clustering in Information Retrieval," *Information Processing & Management* (38:4), pp. 559-582.
9. Xu, Y. C., and Kim, H.-W. 2008. "Order Effect and Vendor Inspection in Online Comparison Shopping," *Journal of Retailing* (84:4), pp. 477-486.
5. Snapp. 2019. "패션·유통업계, '서바이벌 프로그램'에 빠졌네~," http://zine.istyle24.com/Fashion/FashionView.aspx?Idx=44350&Menu=12&C_=23069

[URL]

1. 공감신문. 2019. "제이와이네트워크 "유튜브 광고" 짧은 시간이라도 고품질 영상 제작 중요" - 공감신문," <http://www.gokorea.kr/news/articleView.html?idxno=253758>
2. 나무위키. 2019. "음악 방송 - 나무위키," <https://namu.wiki/w/%EC%9D%8C%EC%95%85%20%EB%B0%A9%EC%86%A1>
3. 충청매일. 2019. "[오늘의 칼럼]음식 프로그램의 변화 - 충청매일," (<http://www.ccdn.co.kr/news/articleView.html?idxno=568786#09SX>)
4. Mobvista. 2018. "2018년 디지털 광고 시장의 주요 트렌드 결산과 2019년의 방향," <https://www.mobvista.com/kr/blog/2018%EB%85%84-%EB%94%94%EC%A7%80%ED%84%B8-%E%A%B4%91%EA%B3%A0-%EC%8B%9C%EC%9E%A5%EC%9D%98-%E%CA3%BC%EC%9A%94-%ED%8A%B8%EB%A0%8C%EB%93%9C-%EA%B2%B0%EC%82%B0%EA%B3%BC-2019%EB%85%84%EC%9D%98-%EB%B0%A9/>

저 자 소 개



윤상혁 (Sang-Hyeak Yoon)

연세대학교 정보대학원에서 정보시스템 박사과정이다. 현재 스마트미디어랩(SMR)에서 데이터 분석 업무를 맡고 있다. 주요 연구분야는 머신러닝, 디지털 마케팅, 비즈니스애널리틱스 등이다. 관련 연구들은 Electronic Commerce Research, IVEY Publishing 등에 논문이 게재되었다.



이소현 (So-Hyun Lee)

연세대학교 정보대학원에서 정보시스템 박사학위를 취득한 후, 현재 연세대학교 정보대학원 연구교수로 재직 중이다. 주요 연구분야는 디지털 비즈니스, 소셜미디어 등이다. 관련 연구들은 Information Systems Research, Communications of the ACM, Information & Management, Behaviour & Information Technology, International Journal of Information Management 등에 논문이 게재되었다.



김희웅 (Hee-Woong Kim)

National University of Singapore 정보시스템학과에서 근무한 후, 현재 연세대학교 정보대학원 교수로 재직 중이다. 주요 연구분야는 디지털 비즈니스, 정보시스템 관리 및 활용 등이다. 관련 연구들은 MIS Quarterly, Information Systems Research, Journal of Management Information Systems 등에 50여 편의 논문이 게재되었다. JAIS, IEEE TEM의 편집위원으로 활동했고, KrAIS 회장을 역임했다.