

상관토픽모델을 활용한 유튜브 크리에이터 소비자의 온라인 구전 특성에 관한 연구*

오혜라 한국외국어대학교 미디어커뮤니케이션학과 석사과정**

정윤재 한국외국어대학교 미디어커뮤니케이션학부 교수***

본 연구는 소비자의 능동적인 콘텐츠 확산에 대한 이해를 높이기 위해 소비자가 어떤 이야기를 어떻게, 왜 말하는지 등의 온라인 구전 특성을 살펴보고자 한다. 이를 위해, 유튜브 크리에이터로 활동 중인 소비자 147명으로부터 개방형 응답으로 온라인 구전 특성을 수집하고 상관토픽 모델링을 수행했으며 토픽 간 관계를 나타낸 네트워크 구조는 상관계수 10% 이상의 토픽만 간선으로 표현했다. 분석 결과 271개의 키워드에서 '지식', '소통', '추억', '참여', '기록', '일상', '경험', '제작', '취미', '활동' 등의 10개 토픽이 도출되었다. 연관성에서는 토픽을 구성하는 키워드 중 '일상'이 가장 높게 나타났으며 토픽3(추억)과 8(제작), 토픽1(지식)과 2(소통), 토픽4(참여), 5(기록), 6(일상), 7(경험), 9(취미), 10(활동)이 유사한 것으로 파악됐다. 네트워크 구조로 도출한 상관관계에서는 토픽9(취미)와 토픽5(기록)가 .20의 상관계수 값으로 가장 높게 나타났다. 그동안 논의가 체계적으로 이루어지지 않았던 유튜브 크리에이터로서 소비자를 CTM을 통해 미시적인 관점에서 탐색한 본 연구는 다음과 같은 학술적·실무적 의의를 지닌다. 첫째, 한정된 마케팅 예산을 해결할 수 있는 소비자의 온라인 구전에 내재한 가치를 영상 제작 생산·공유하는 소비자의 관점에서 살펴본 방법론적인 측면에 학술적 함의를 제공한다. 둘째, 분석 결과는 소셜미디어뿐 아니라 마케팅, 광고, 커뮤니케이션 등 다양한 분야의 연구 토대와 이론적 근거를 마련해줄 것이다. 특히 기업이 위기 커뮤니케이션 상황을 접했을 때 소비자로부터 완충작용을 하는 커뮤니케이션 역할을 기업이 수행하는 데 있어 실무적 의의를 제공할 것이다.

KEY WORDS 개인 미디어 • 영상 콘텐츠 • 유튜브 • 온라인 구전 • 상관 토픽 모델

* 이 연구는 한국외국어대학교 교원연구지원사업 지원에 의하여 이루어진 것임. 본 논문은 2020년도 한국광고학회의 온라인 설문조사 지원을 받아 진행되었음.

** hyeraira@hufs.ac.kr, 제1저자

*** yjcheong@gmail.com, 교신저자

1. 연구배경 및 목적

유튜브(YouTube)는 뒷광고 이슈로 콘텐츠의 투명성을 촉진해야 하는 상황에 놓여있다. 소셜 네트워크 서비스(Social Networking Service, 이하 SNS) 중 유튜브가 광고성 콘텐츠가 많다는 인식이 가장 높았으며, 이에 대한 인식은 이전보다 더 증가하는 추세(오픈서베이, 2021)이기 때문이다. 또한, 기업들이 자사 제품의 온라인 리뷰를 임의로 작성하거나, 경쟁제품에 대한 부정적 리뷰를 작성하는 등 온라인 구전 정보를 조작하는 사건도(김성훈, 2018) 그 원인으로 작용한다. SNS 이용자를 대상으로 거부감이 크거나 신뢰감이 낮은 게시물 형태를 조사한 결과 ‘광고가 아닌 것처럼 보이도록 만든 광고 게시물’과 ‘실제 후기를 가장한 홍보 게시물’로 나타났다(오픈서베이, 2021). 설문 결과에서 나타난 부정적인 인식을 형성하는 위의 두 특징을 드러내고 있는 유튜브에 대한 문제 해결의 필요성이 요구된다. 또한, 소비자의 동영상 이용 행태에서도 기업이 유튜브를 어떻게 활용해야 할지에 대한 새로운 판단의 필요성이 나타난다. DMC(2021)의 ‘2020 인터넷 동영상 시청행태 및 광고 태도 분석’ 보고서에 따르면 동영상 콘텐츠의 이용량은 점점 늘어나는데 오히려 동영상 광고 클릭 경험자는 줄어든 것으로 조사됐다. 이는 타겟에게 광고 클릭을 유도하기가 더 어려워졌으며, 타겟에게 그들이 관심이 있는 상품의 광고를 노출하는 것이 매우 중요함을 보여준다.

코로나 19 팬데믹은 원시시대 이후 사회를 형성해 살아온 인간에게 분리를 강요했다. 하지만 사람들은 ‘연결’의 가치에 더욱 집중했으며, 함께 대화하고 행동하는 기업, 브랜드, 유튜버가 많은 구독자를 얻고 사랑받고 있다(김경달 · 씨로켓리서치랩, 2020). 영상 콘텐츠를 통해 소통하고 소비하는 방식도 다차원적으로 나타났다. 코로나 19로 동영상 이용 시간이 급증하자 그동안 배송 경쟁에 집중하던 유통업계는 ‘미디어 커머스(Media Commerce)’로 방향을 전환했다. 미디어(콘텐츠)와 커머스(전자상거래)를 접목한 서비스인 미디어 커머스는 유튜브 등 SNS, 쿠팡, 네이버 및 넷플릭스(Netflix) 등에서 주로 영상 콘텐츠를 활용해 상품을 마케팅한다(김경달 등, 2020). 소셜미디어에서 영향력을 갖는 인물을 칭하는 ‘일반인 인플루언서’(최지윤 · 정윤재, 2017)를 대상으로 구매 전환이 잘 일어나는 플랫폼을 조사한 결과 유튜브(68.5%)가 1위로 꼽혔다(안선혜, 2020). 이상의 맥락은 기업이 유튜브 크리에이터로 활동 중인 소비자를 토대로 기존 마케팅 전략 수립의 한계점을 해결할 필요성을 시사한다.

실제 기업에서는 일반인 인플루언서를 활용한 마케팅이 점점 증가하는 추세다(최지운 등, 2017). 영상 콘텐츠 창작 측면에서도 ‘크리에이터를 넘어 인플루언서의 시대’로 진화(한국콘텐츠진흥원, 2019)하며 유튜브 크리에이터로 활동 중인 소비자에 대한 이해의 필요성을 강조한다. 다양한 영역에서 유튜브를 활용한 소통이 일어나는 모습은 사회 각층의 소셜미디어 크리에이터의 참여와 영상 콘텐츠 공급의 활성화를 이끌 것이다.

따라서 본 연구는 팔로워 수가 크고 적음에 상관없이 자발적으로 영상을 직접 제작·공유하는 ‘유튜브 크리에이터로서 소비자(Consumer as YouTube Creator)’의 능동적인 콘텐츠 확산에 대한 이해를 높이는 데 목적을 둔다. 유튜브에 대한 부정적인 인식 속에서도 콘텐츠 수요와 공급은 기업과 소비자의 높은 상호협정을 통해 부정적인 인식을 낮추고 콘텐츠 생산·소비에 혁신을 유도할 수 있기 때문이다. 또한, 기업이 원하는 소비자와의 커뮤니케이션 형성은 유튜브 크리에이터로서 소비자가 어떤 이야기를 어떻게, 왜 말하는지 등의 특성을 파악하는 것이 선행되어야 가능하다. 이에 본 연구는 유튜브에서 자신의 채널을 운영하며 콘텐츠를 직접 제작·공유하는 소비자로 연구대상 범위를 선정하였다. 소비자가 자발적으로 애착을 갖고 제작·공유한 영상의 특징에 대해 객관적으로 분석한 연구 결과를 토대로 소비자의 주체적인 온라인 구전을 확산할 요소를 탐색하고자 본 연구는 유튜브 크리에이터로서 소비자의 입장에서 온라인 구전 특성을 도출한다. 이를 통해 유튜브 크리에이터로서 소비자로부터 실현 가능한 온라인 구전과 자발적인 온라인 구전 확장을 위한 전략 방안을 제시한다. 콘텐츠를 소비하는 유튜브 이용자로 연구가 편중되어 있던 상황에서 본 연구는 연구대상의 한계를 극복하고 새로운 관점을 구축한다는 점에서 의의가 있다.

이상의 분석을 위해 소비자가 직접 작성한 데이터에 기초하여 문체탐색과 의사결정을 하는 빅데이터 과학의 일환인 코퍼스 언어학 기반의 내용분석수법(Corpus Linguistic Based Content Analysis)을 적용한다. 구체적으로 유튜브 크리에이터로서 소비자에게 개방형 형태의 질문(Open-ended question)으로 응답을 얻어 온라인 구전 특성을 객관적으로 탐색·기술하고자 한다. 또한, 일상 행동 간 연결성을 보이는 소비자의 특성에 따라 도출된 온라인 구전 차원 간의 상관성을 함께 분석하고자 한다. 서로 다른 소비자로부터 도출된 행동 특성일지라도 독립적이고 명확히 구분되는 것은 아니며, 형성된 개개인의 특성은 커뮤니케이션을 통해 지속해서 변화할 수 있기 때문이다.

온라인 구전을 직접 수행하는 대상에 주목하는 것은 향후 소비자의 구전 방향을 예측하는 실무적 측면의 이론적 근거로서 작용할 것이다. 소비자가 자발적으로 애착을 갖고 제작한 영상의 특징을 마케터 혹은 기업이 객관적이고 신속하게 파악하는 것은 온라인 구전의 시너지효과를 모색하는데 기여할 수 있다. 또한, 본 연구 결과는 유튜브 크리에이터로서 소비자에게서 생산되고 소비되는 영상 콘텐츠를 기반으로 한 마케팅 전략 수립에 도움이 될 것이다.

2. 이론적 고찰

1) 개인 미디어 영상 콘텐츠

오늘날 디지털 미디어에서 생산된 콘텐츠는 양적은 물론 질적으로도 확장되는 과정에 있다. 이는 소셜미디어를 중심으로 변화하는 디지털 미디어 환경 속에서 사회 주체이자 미디어 사용자인 ‘개인’의 의미 변화와 연결된다. 개인은 근대에서 자유롭고 자기 스스로 결정력을 지닌 인간 고유의 특성을 의미한다(이희은, 2019). 이러한 특성은 콘텐츠 기획과 제작, 그리고 유통이나 송출에 이르기까지 개인이 혼자서 운영하는 형태인 개인 미디어(문화체육관광부, 2018)가 콘텐츠 생산자의 다양한 욕구와 취향을 반영할 수 있음을 시사한다.

개인 미디어의 장점은 개인이 인터넷상에서 네트워크의 주체적 구성원으로 콘텐츠를 생산하고 유통할 수 있다는 것이다. 즉 개인은 미디어에 대해 단순히 소비자가 아니라 생산자로서 혹은 생산적으로 참여하는 것이 가능해졌다(배기형, 2016). 개인 미디어 콘텐츠 생산자들은 커뮤니티를 중심으로 이용자와의 긴밀한 상호작용을 형성하기 때문에 인플루언서(Influencer)로서 브랜드 및 다양한 기관과의 협업도 가능하다. 더불어 콘텐츠와 커머스를 결합한 비즈니스를 수행한다는 차원에서 산업 효과도 크다.

기존 방송 산업이 방송사와 제작사, 그리고 소수 스타 연예인만 경제적 이익을 취할 수 있는 구조라면, 개인 미디어 콘텐츠는 생산자들이 오리지널 콘텐츠(Original Contents)를 만들기 때문에 팬덤(Fandom)과 지식 재산(Intellectual Property; IP)을 기반으로 새로운 소구 수단을 찾고 있는 브랜드와 협업하여 부가가치를 올릴 수 있다

(문화체육관광부, 2018). 특히 개인 미디어는 그동안 제도화된 미디어에 담기지 못했던 많은 표현이 드러난다는 점에서 중요하다(이희은, 2018). 기존 1인 미디어(One-person Media) 콘텐츠가 게임, 뷰티, 먹방, 키즈 등의 분야에 편중되어 있던 과거와 달리 중장년층의 1인 미디어 콘텐츠 이용이 급격히 늘어나면서 정치, 북튜버, 재테크, 부동산, 자기계발, 마음공부, 농사, 요리 등 신규 채널이 활성화되었다(김해원, 2018).

학술적·실무적 측면에서 살펴본 개인 미디어의 특성은 콘텐츠 생산자로부터 자발적으로 생산된 콘텐츠의 특성을 탐색하는 본 연구의 중요성을 제시한다. 개인 미디어 콘텐츠 생산 과정 차이는 곧 콘텐츠 생산자의 가치 차이를 의미하며, 콘텐츠 생산자의 가치는 콘텐츠 생산과 확산의 출발점으로 작용하기 때문이다. 개인 미디어 콘텐츠 생산자의 행동에 대한 미시적 차원의 접근은 디지털 미디어 환경, 콘텐츠 및 콘텐츠 생산자의 속성 등 종합적인 맥락에 대한 이해가 동반되어야 가능하다. 개인 미디어 콘텐츠 생산자에 대한 새로운 학문적 접근을 시도하는 본 연구는 콘텐츠 생산 주체로서의 소비자의 가치를 분석하는 차원을 넘어 소비자 관점에서 광고 전략을 수립하는 데 있어 기반으로 작용될 수 있을 것이다.

구체적으로 콘텐츠 크리에이터의 주력 플랫폼(Platform)이 유튜브라는 특징(문화체육관광부, 2018)에 따라 본 연구에서는 유튜브 크리에이터로서 소비자가 직접 기술한 데이터를 바탕으로 분석하고자 한다. 소비자는 유튜브를 통해서 영상을 시청하는데 그치는 것이 아닌 본인의 글·사진·영상·댓글 업로드를 적극적으로 하는 것으로 나타났다(오픈서베이, 2021). 유튜브는 매월 920억 페이지뷰를 기록하고 있으며 이는 미국의 3대 네트워크 텔레비전 방송국이 지난 60년간 제작한 영상 콘텐츠와 같은 양의 콘텐츠가 불과 두 달 만에 유튜브에서 만들어지고 있음을 의미한다(안대천·김상훈, 2012). 따라서, 유튜브 채널(YouTube Channel)을 직접 운영하는 소비자로부터 영상 콘텐츠의 특성을 이해하고 자발적인 콘텐츠 생산 가치를 추적할 필요가 있다. 이에, 개인에 의해 자발적으로 생산된 영상 콘텐츠의 특성에 초점을 두는 본 연구는 유튜브 크리에이터로서 소비자의 특성을 파악하는 데 있어 방향성을 제시할 것이다.

개인 미디어 영상 콘텐츠는 그동안 광고학, 인류학, 심리학, 교육학 등 다양한 학문에서 행동, 태도, 동기 등과 연결돼 논의가 이루어져 왔다. 이효성(2020)은 유튜브 속 영상 콘텐츠 이용과 이용 동기 요인들에 관해서 탐구하였다. 윤지영 등(2020)은 유튜브 브이로그(vlog) 이용 경험이 있는 20~30대 일반인을 대상으로 브이로그 이용 동기

와 이용자 특성이 브이로그 이용 만족 및 지속적인 이용 의도에 미치는 영향을 분석했다. 정선임 등(2021)은 유튜브 플랫폼과 1인 미디어에 초점을 두고 의학 정보 유튜브 콘텐츠 크리에이터의 평판과 콘텐츠 속성이 시청자의 건강 행동 의도에 어떠한 영향을 미치는지 분석했다. 김자경과 이정기(2018)는 자율 감각 쾌락 반응(Autonomous Sensory Meridian Response; ASMR)과 같은 무자극 영상 콘텐츠에 대한 20대 대학생들의 지속적 이용 의도 결정요인을 탐색했다. 이처럼 미디어, 이용자, 영상 콘텐츠 특성, 심리적 요인 등 개인 미디어 영상 콘텐츠에 대해 다루는 학술적 논의가 다양하게 이루어지고 있으나, 콘텐츠를 생산하는 구전 발신자로서 소비자의 관점을 분석한 결과는 찾아보기 어렵다는 점에서 여전히 한계를 갖는다. 이러한 맥락에서 본 연구는 유튜브 크리에이터로서 소비자를 대상으로 생산한 영상 콘텐츠의 특성을 탐색하고 이를 바탕으로 자발적이고 능동적인 온라인 구전과의 논리적 연결고리를 규명하고자 한다.

2) 유튜브와 온라인 구전

확장된 상호작용과 더불어 인터넷의 발달이 가져온 시간과 공간의 자율성은 자발적으로 정보를 생산하고, 공유하는 구전 경향을 증가시켰다(신병철, 2003). 전통적인 미디어와는 달리 1인 미디어가 디지털 환경의 속성을 상당히 반영하고 있다는 점을 고려하면 향후 1인 미디어가 개별 이용자에게 미치는 파급력은 더욱 높아질 것이다(유현주·김현, 2020). 그중 유튜브(YouTube)는 웹 기반의 영상 콘텐츠 서비스 확산의 촉매제 작용을 하고 있다(송진·이영주, 2016). 소비자의 콘텐츠 생산율은 2020년 조사 대상자의 17.4%에서 2021년 33.2%로 15.8% 증가했으며(오픈서베이, 2021) 자발적이고 능동적인 유튜브 채널 운영은 크게 확장 중이다. 최근에는 자신만의 채널을 만들어 스트리밍을 통해 라이브 방송을 진행하거나 직접 제작한 동영상을 주문형 비디오(Video on Demand; VOD) 형태로 업로드하는 ‘유튜브 크리에이터’로서의 소비자의 활동이 증대되었다. 유튜브 크리에이터는 동영상을 직접 제작하여 유튜브에 업로드하는 사람으로 유튜브 이용자, 광고주와 함께 유튜브 생태계를 구성하는 핵심 요소를 말한다(노희운, 2019). 소셜미디어 플랫폼상에서 다수의 팔로워(follower)를 보유하고 영향력이 높으며(이신형·김셋별, 2019) 미디어 플랫폼을 기반으로 인지도를 쌓은(박선경·김상훈, 2020) 인플루언서와 달리 크리에이터는 무언가를 새롭게 만들어 내는 사람이다(이용

관, 2019). 즉, 1인 미디어 크리에이터는 개인이 다양한 콘텐츠를 직접 생산하고 공유할 수 있는 소통 플랫폼(채널)에서 콘텐츠를 제작·유통하는 자를 말한다(이용관, 2019). 본 연구에서는 불특정 다수의 소비자가 정보제공, 관심사 탐색, 즐거움, 사회적 상호작용을 위해 유튜브 영상을 제작·공유하는 활동을 설명하고자 1인 미디어 크리에이터에 대한 기존의 문헌에 따라 유튜브 크리에이터로서 소비자를 ‘개인의 가치, 관심사, 욕구 등에 따라 유튜브에서 영상 콘텐츠를 제작·공유하는 모든 소비자’로 정의한다.

구전(Word-of-Mouth, WOM)은 오늘날 웹을 통해 온라인 구전 중심으로 나타나고 있다. 온라인 구전(electronic Word-of-Mouth, eWOM)은 콘텐츠를 빠르게 전파하고 접근성이 높으며 눈으로 직접 관찰할 수 있고 신뢰를 제공한다(Cheung & Thandani, 2012). 특히 소비자는 개인적인 중요성이나 관심도를 의미하는 관여도(안광호 등, 2014)가 높은 것을 공유하는 모습을 보인다(Greenwald & Leavitt, 1984). 콘텐츠 이용자에게 친숙하거나 흥미 있는 분야(Laaksonen, 1994)가 광고처럼 주의력을 이끌 수 있다(Greenwald et al., 1984)는 점은 기업을 대신한 유튜브 크리에이터로서 소비자가 구현하는 마케팅 효과의 근거로 작용할 수 있다. 즉, 유튜브 크리에이터로서 소비자는 관심 있는 정보를 직접 생산하고 공유하는 활동을 수행하는 구전 정보원의 역할의 모습을 보인다.

일반적으로 온라인 구전에서는 정보 발신자의 익명성을 전제로 한다(Tseng & Fogg, 1999). 하지만 유튜브를 통한 온라인 구전 시 발화자의 신원(Identity)이 명확히 나타나며 익명성이 보장되지 않는다(이혜린 등, 2019). 즉, 유튜브를 통한 온라인 구전은 크리에이터와 크리에이터가 생산한 영상 콘텐츠가 신뢰성을 결정토록 하며 영상 콘텐츠에는 크리에이터가 중요하게 생각하는 속성들이 반영된다. 따라서 자발적이고 능동적으로 생산된 유튜브 영상 콘텐츠의 특성을 올바르게 이해하는 것은 유튜브 크리에이터로서 소비자의 온라인 구전 특성을 파악하는 것이라 할 수 있을 것이다.

이상의 이론적 고찰을 바탕으로 본 연구는 유튜브 크리에이터로서 소비자의 영상 콘텐츠에서 나타나는 특성을 이해하고 실질적으로 내재 되어 있는 온라인 구전의 특성을 파악하고자 한다. 청과 리(Cheung & Lee, 2012)는 온라인 구전을 장려하는 가장 중요한 요소로 ‘소유(ownership)’, ‘승리(winning)’, ‘돕고 싶어 하는 의지(willingness to help)’를 제시하였다. 앤더슨(Anderson, 1998)은 발신자 관점에서 ‘열의(Enthusiasm)’, ‘구체성(detail)’, ‘호의성(Praise or Favorableness)’ 등의 세 가지 구전 특성을 밝혀냈

다. 열의는 구전의 빈도와 접촉대상의 숫자, 구체성은 구전 정보의 양, 호의성은 제품이나 서비스에 대한 구전 정보의 긍정·부정성이다. 하지만 국내 연구에서는 온라인 구전 주체를 중심으로 탐색한 연구가 미비하다. 온라인 구전에 관한 기존 연구들은 크게 온라인 구전의 선행요인을 규명하기 위한 연구(김성훈, 2003; 김해룡 등, 2004; 이은영, 2004; 이은영·이태민, 2005; 이현선·리대용, 2004; Chatterjee, 2001; Henning-Thurau & Walsh, 2003; Henning Thurau, 2004)와 온라인 구전의 효과를 규명하기 위한 연구(Chevalier & Mayzlin, 2003; Godes & Mayzlin, 2004), 그리고 온라인 구전 정보의 양, 방향성(부정/긍정), 구전 정보의 평가내용 등 소비자의 온라인 구전 정보를 탐색한 연구(박철·정수연, 2006)로 구분된다(박철·이태민, 2006). 가령 노단 등(2020)은 온라인 매체의 구전 양이 드라마의 성과에 어떠한 영향을 미치는지 방송사의 지상파 여부에 따라 살펴보았다. 하지만 드라마와 관련된 온라인 구전의 양적 속성을 통한 인지 효과만을 검증하며 구전 발신자의 입장에서 구전 의도 등을 탐색하는 데 한계를 보였다. 김인규와 차성수(2020)는 모바일 게임 앱 내 소비자의 사용 후기와 댓글을 특징과 시기별로 수집하여 구전의 다이내믹스(Dynamics) 및 게임 앱 성과 간의 상관성을 분석했다. 게임 앱에서 구전 발신자가 남긴 리뷰(Review)를 내용 분석하며 구전에 대한 양적·질적 방법론을 보완했는지라도 구전하는 소비자의 입장에서 왜 구전을 하는지, 그리고 그에 따라 나타난 구전의 특성은 무엇인지 발신자 입장에서의 구전의 중요한 원천을 살펴보지 못한 한계를 지적할 수 있다. 기존 연구에서 미비하였던 유튜브 크리에이터로서 소비자를 탐색하는 본 연구는 이들에게 내재한 온라인 구전 특성을 파악하는 차원을 넘어 소비자의 관점에서 도출된 자발적으로 생산된 온라인 구전의 의미를 제시할 수 있을 것이다.

3) 텍스트마이닝 기반 상관 토픽 모델(Correlated Topic Model)

설문조사는 조사자가 설계한 설문 틀에 따라 응답하는 방법으로 응답자의 관점에서 분석대상에 대한 의견과 상관없는 특성을 답하도록 하는 한계점을 보일 수 있다 (Tapachai & Waryszak, 2000). 이는 조사자 관점에 따라 도출된 분석 결과가 소비자에 내재한 의사구조 등의 특성을 파악하는 데 제한적일 수 있음을 시사한다. 또한, 현대 사회에서 기하급수적으로 증가하는 정보량에서 원하는 정보를 단번에 찾는 것은 점

점 어려워지고 있으며 대량의 데이터 내에서 주어진 데이터가 어떠한 정보와 가치를 포함하고 있는지 파악하지 못할 때도 있다. 빅데이터 플랫폼 성격을 지닌 유튜브 속 온라인 구전의 객관적인 성격을 탐색하는 방법론적 관점을 확장함으로써 이에 대한 이론적 근거를 마련하고, 실무적인 가치로 기능하기 위해서는 구전 주제인 유튜브 크리에이터로서 소비자 관점에서의 종합적인 접근이 필요하다.

텍스트 마이닝(Text Mining) 중 수많은 데이터를 탐색해 정보 추출을 도와주고 숨겨진 패턴을 정리해주는 분석법인 토픽 모델링(Topic Modeling)(안정국 등, 2015)은 연구대상에 내포된 의미를 탐색하는 데 유용하다. 서비스 또는 제품에 대한 솔직한 경험과 의견에 대해 이용자들이 남기는 온라인 리뷰(Pang & Lee, 2008)를 활용한 기존 연구는 텍스트 마이닝을 통해 서비스나 제품 사용의 주요 요인을 파악했다(황해정 등, 2016; 이소현 등, 2019). 선행연구에 근거하여 본 연구도 비정형 데이터(Unstructured Data)에서 자연어 처리(Natural Language Processing) 기술에 기반을 두어 의미 있고 유용한 정보를 추출하는 과정(이소현 등, 2020; Sebastiani, 2002; Hearst, 1999)인 텍스트 마이닝을 분석 방법에 도입하고자 한다. 텍스트 마이닝에서 다루는 다양한 방법론 중 하나인 토픽 모델링은 방대한 텍스트로 구성된 문헌 집단에서 핵심 주제들을 찾아내기 위한 자연어 처리 알고리즘(Algorithm)으로 각각의 문서 속에 잠재된 토픽(Topic)을 찾아내고자 개발됐다(Blei et al., 2003). ‘어휘들의 의미는 어휘 자체에 내재하기보다는 어휘가 사용되는 맥락 혹은 함께 사용되는 어휘들과의 관계에 기초한다.’는 전제에서 만들어진 토픽 모델링(남춘호, 2016)은 여러 문서가 임의의 토픽 집합으로 구성된 것으로, 그 문서를 구성하는 토픽과 단어의 중요도를 확률로 제시한다(Steyvers & Griffiths, 2007).

그중 상관 토픽 모델(Correlated Topic Model, 이하 CTM)은 블레이와 래퍼티(Blei & Lafferty, 2007)가 개발한 것으로 로지스틱 정규분포(logistic normal distribution)를 따르며 추출된 토픽 간의 상관성을 모형화한 토픽 모델링이다. CTM은 토픽을 생성하는 과정에서 단어 간 연관성에 따라 가중치로 세분화하며 소비자의 외적으로 나타나는 행동과 성격 등의 내적 요소를 상호 연결한 토픽 구조를 체계적으로 구현할 수 있다. 유튜브 크리에이터로서 소비자로부터 자발적이고 능동적으로 생산된 영상 콘텐츠의 특성을 탐색하는 본 연구는 양적 연구뿐 아니라 CTM을 통해 질적 분석을 통합적으로 활용한다는 측면에서 기존의 내용분석이 갖는 단점을 보완하고 심층적인 통찰을 얻

을 수 있어 유용하다. 따라서 본 연구는 텍스트 마이닝을 기반으로 한 CTM을 통하여 분석을 수행하고자 한다. 아래 <그림 1>은 CTM을 시각화한 것이며(Blei et al., 2007), <표 1>은 <그림 1>의 각 기호가 의미하는 바를 나타낸다(Konstantin et al, 2009). <그림 1>에서 각 원은 랜덤 변수(random variable) 또는 모델 매개 변수(model parameter)를 나타내며, 간선(edge)은 변수 사이의 확률적 의존성(probabilistic dependencies) 또는 조건부 독립성(conditional independencies)을 지정한다(Konstantin et al, 2009).

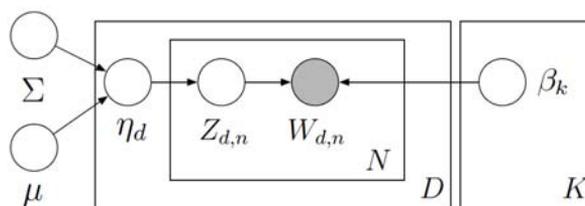


그림 1. Graphical model representation for the CTM

표 1. CTM 기호 의미

기호	의미
D	전체 문서에서 나타난 말뭉치 개수
d	1, 2, ..., D , 말뭉치에 있는 개별 문서의 인덱스
N	d 번째 문서의 단어 개수
n	1, 2, ..., N , 문서 d 속 단어 출현 인덱스
K	전체 토픽 수(하이퍼 파라미터)
κ	1, 2, ..., K , 잠재된 변수의 인덱스 (“주제”)
$w_{d,n}$ (or w_n)	d 문서에서 n 번째 단어 (W 는 단어)
$z_{d,n}$ (or z_d)	d 번째 문서에서 n 번째 단어에 할당된 토픽 (Z 는 토픽)
β_k	k 번째 토픽의 단어 분포
$\eta_d \sim \mathcal{N}(\mu, \Sigma)$ (or η)	각 문서에 대한 토픽 분포를 지정하는 K -차원 벡터
μ and Σ	다변량 가우시안 프로세스(multivariate Gaussian process)의 매개변수 (평균 및 공분산 행렬)

Note. Konstantin et al.(2009)을 바탕으로 재구성

본 연구에서는 각 토픽을 단일하고 독립적인 개체로 고려한 잠재 디리클레 할당 (Latent Dirichlet Allocation, 이하 LDA)이 아닌 토픽 간 상관관계를 통한 온라인 구

전의 양상을 파악할 수 있도록 CTM으로 연구를 진행하였다. 실생활의 다양한 문서들은 서로 상관관계가 존재하는 것이 일반적인 현상이기 때문에, 토픽 간의 상관관계 모델링(Modeling)이 어려운 LDA에 비해 CTM이 본 연구에 적합하다(김규하·박철용, 2015). CTM은 유튜브 크리에이터로서 소비자들의 응답에서 자발적인 온라인 구전의 전반적인 특성을 도출하고 상관관계를 밝힘으로써 구전 발신자의 관점에서 논의한 온라인 구전 특성 연구를 보완할 것이다. 결과적으로 텍스트 마이닝을 통해 살펴본 온라인 구전에 대한 분석은 관련한 향후 연구의 타당성을 높여주는 근거로 작용하고 관련된 국내연구의 부재를 보완할 수 있을 것이다.

도출된 토픽에 대한 네트워크 분석(Network Analysis)은 텍스트(text)에 출현하는 키워드 간의 관계를 간선으로 표시해 구축되는 구조(structure)의 특징을 해석하는 방법으로 특정 연구나 현상을 이해하는 데 탁월하다는 장점이 있다(남현동·남태우, 2020). 이에 따라 본 연구에서는 CTM을 통해 도출한 토픽 간 관계를 토픽 간 거리 지도(Intertopic Distance Map, 이하 IDM)와 네트워크 구조로 시각화한다. IDM은 토픽의 비중과 토픽간 거리를 나타내는 그림으로 각 토픽이 다른 토픽과 어느 정도의 연관성을 지니는지와 토픽 간의 유사성을 파악할 수 있게 한다(조정원·우영운, 2019). 네트워크 구조는 토픽을 구성하는 단어들의 관계뿐 아니라 토픽 간 관계를 통해 토픽의 중요도를 나타낸다. CTM과 IDM, 네트워크 구조로 자발적인 온라인 구전의 특성을 밝히고 특성 간 관련성을 실증적으로 규명하고자 본 연구는 다음과 같은 연구문제를 설정하였다.

RQ1. 유튜브 채널을 운영하는 소비자가 관심과 공유의 목적을 위해 생산한 영상 콘텐츠로부터 추출된 토픽의 특성은 어떠한가?

- 1-1. 토픽을 구성하는 키워드는 무엇인가?
- 1-2. 토픽을 구성하는 키워드의 빈도수는 어떠한가?

RQ2. 유튜브 채널을 운영하는 소비자가 관심과 공유의 목적을 위해 생산한 영상 콘텐츠로부터 추출된 토픽은 어떠한 연관성을 가지는가?

- 2-1. 추출된 토픽을 구성하는 키워드 간 관계는 어떠한가?
- 2-2. 추출된 토픽 분포도는 어떠한가?

RQ3. 유튜브 채널을 운영하는 소비자가 관심과 공유의 목적을 위해 생산한 영상 콘텐츠로부터 추출된 토픽의 구조적 관계는 어떠한가?

3. 연구방법

온라인 구전 분석의 세부 절차는 <그림 2>와 같다. 추출된 데이터로부터 CTM을 통한 토픽 모델링을 수행하여 키워드의 내재적 특성을 탐색하였으며, 토픽의 분포도와 상관성을 파악하기 위해 토픽 분포도 및 네트워크 구조를 분석하였다.

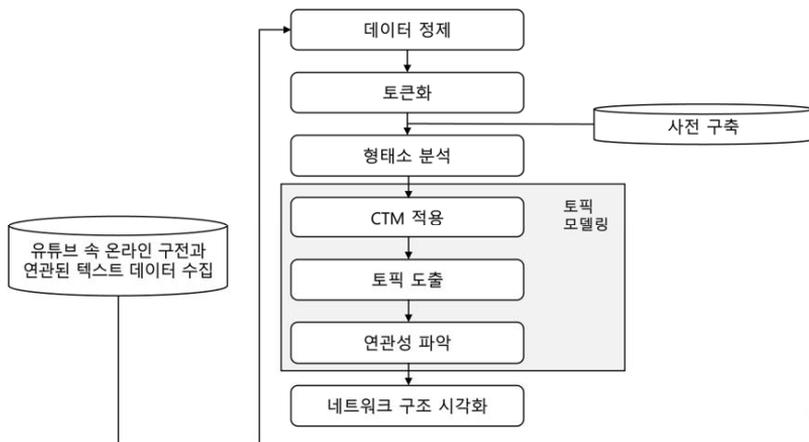


그림 2. CTM을 통한 텍스트마이닝 진행 과정

1) 데이터 수집

본 연구의 연구대상 범위는 유튜브 채널을 직접 운영하는 유튜브 크리에이터 소비자로 한정하였다. 유튜브는 소셜미디어 중 이용률이 가장 높으며 정보 탐색 이용에 있어 1년 전 대비 67.3% 증가하였기 때문이다. 또한, 유튜브에서는 흥미 위주나 유용한 콘텐츠를 얻고자 이용하는 비중이 두드러지게 나타났다(오픈서베이, 2021). 직접 영상 콘텐츠를 제작·공유하는 정보 확산 차원에서 소비자의 온라인 구전의 내재적 의미를 조명하는 본 연구목적에 따라 기저가 될 데이터의 출처는 유튜브 크리에이터 소비자로 설정

하였다. 데이터는 ‘귀하의 영상 중 높은 조회 수/좋아요/공유 수와 관계없이 귀하가 애착을 갖고 제작했던 영상 또는 라이브 방송의 특징은 무엇인지 자유롭게 기재하여 주십시오.’를 질의어로 한 개방형 형태의 온라인 서베이(online survey)를 전문 리서치 기관 마크로밀엠브레인(Macromill Embrain)을 통해 수집하였다. 2020년 9월 16일부터 10월 6일까지 총 160부의 설문지가 회수되었으며 응답이 불성실하거나 결측치가 있는 응답지를 배제하고 실제 분석에는 147부의 설문지를 활용했다.

2) 데이터 분석

데이터 분석 시 프로그래밍 언어는 파이썬(Python), 통합 개발 환경은 주피터 노트북(Jupyter Notebook), 라이브러리(Library)는 토모토파이(tomotopy)를 사용하였다. tomotopy로 추출한 10개의 토픽은 상관관계 분석 후 pyLDAvis를 통한 IDM과 Barnes-Hut 알고리즘을 통한 계층적 네트워크 구조로 시각화하였다.

(1) 데이터 클리닝

소비자의 개방형 응답으로부터 추출한 데이터는 분석에 적합한 형태로 재가공하는 전처리(preprocessing)와 형태소 분석(morphological analysis)이 필요하다. 전처리 단계의 정확성 여부가 정보검색의 방향을 결정한다고 해도 과언이 아니므로 수집한 결과를 잘 분류하여 분석이 가능한 형태로 만드는 것이 매우 중요한 기술이 된다(조영임, 2013). 인간이 생활 속에서 사용하는 언어인 비정형 형태의 자연어(natural language)를 컴퓨터 언어로 정형화 후 연구목적에 적합한 정보를 효율적으로 추출하고자 텍스트 형태의 데이터를 전처리하는 것이다. 본 연구의 데이터셋(dataset)은 CTM을 위해 다음과 같은 정제 기준을 적용하였다.

‘~것’, ‘~때’와 같이 내용을 파악하는 데 보조적인 역할을 하는 기능어와 불필요하게 길거나 수식어가 있는 단어, ‘...’, ‘:’ 등의 특수문자 및 부호는 삭제하였다. 또한, ‘관련’과 같이 연구목적에서 벗어난 토픽은 제거하고 형태소 분석에 사용되는 범용 사전에 의해 연구목적과 다르게 분할되는 것을 방지하고자 사용자 사전을 구축하였다. 가령 ‘떡방’, ‘영상미’, ‘구독자 수’와 같이 내용상 한 단어로 쓰여도 좋은 토픽은 띄어쓰기 없이 분석에 사용하고자 단어 그대로 사용자 사전에 등록하였다. ‘vlog’, ‘blog’ 등 한글-영어, 한글-

한자 동의어의 경우 한글로 통일하였다. 문장을 의미의 최소 단위로 바꾸는 작업인 형태소 분석은 한국어 데이터 처리에 유용한 코엔엘파이(KoNLPy)를 사용하였다.

(2) CTM을 사용한 토픽 추출과 연관성 분석

토픽 모델링의 주요 파라미터(parameter)는 토픽 개수(K)를 선정하는 것이다(강현아 · 임희석, 2020). 토픽의 개수를 결정하는 방법에 대한 선행연구를 살펴보면 다음과 같이 정리할 수 있다(홍영희, 2019). 첫 번째로 연구자가 해석 가능한 해당 분야에서의 적합성 등을 고려하여 적당한 수를 지정한 토픽 모형에 적용하는 것이다(강범일 등, 2013; 배정환 등, 2013; Grant et al., 2013). 두 번째로 실험 데이터의 특징에 따라 연구자가 주관적으로 정하거나(홍정립 등, 2019), 일관성 점수(coherence score) 등을 적용할 수 있다. 최적의 토픽 수에 대한 통계적 해법은 없으나(남춘호, 2016), 다만 전체 토픽을 대상으로 얼마나 많은 수의 의미 있고 분석적으로 유용한 토픽을 산출해주는가에 따라 토픽 수를 선택하는 것이 적합하다(남춘호, 2016; Dimaggio et al., 2013).

본 연구는 유튜브 크리에이터 소비자로부터 설문을 진행함에 따라 다수의 표본 수 확보에 어려움이 있었다. 또한, 유튜브 크리에이터로부터 수집한 자발적인 온라인 구전 데이터를 코퍼스(corpus)로 구축한 선행연구가 부재하였기 때문에, 키워드의 적합한 정의의 기준이 되는 참조 코퍼스(reference corpus)를 바탕으로 분석 대상이 되는 목표 코퍼스(target corpus)를 구축하는 것은 적절하지 않았다. 따라서 본 연구는 토픽 수(K)의 결정에 있어 토픽 모델링을 반복적으로 수행하여 도출한 키워드의 의미 분석에 무리가 없는 값 10으로 토픽 수를 설정하였다.

또한, 정제화 과정을 거친 데이터에서 임의의 두 키워드 쌍에 대한 단어 간 연관성을 나타내는 척도 점별 상호 정보량(Pointwise Mutual Information, 이하 PMI)으로 토픽 응집성(Topic Coherence)을 측정했다. PMI를 이용하면 사람이 판단하는 것과 매우 흡사하게 토픽의 일관성을 측정할 수 있다(이정빈 · 인호, 2017; Newman et al., 2010). 데이터에서 특정 단어와 다른 단어 간 함께 사용된 정도에 중요도를 두는 PMI는 <식 1>과 같이 정의된다. <식 1>에서 $p(X)$ 는 키워드 X 가 전체 말뭉치(corpus)에 등장할 확률이며, $p(X, Y)$ 는 키워드 X 와 Y 가 한 토픽에 동시에 등장할 확률이다.

식 1. 점별 상호 정보량(Pointwise Mutual Information)

$$PMI(X; Y) = \log \frac{p(X, Y)}{p(X)p(Y)} = \log \frac{p(X|Y)}{p(X)} = \log \frac{p(Y|X)}{p(Y)}$$

추출할 토픽 개수(k)를 10으로 설정한 본 연구는 각 토픽의 임의의 두 키워드 쌍에 대한 키워드 간 연관성을 나타내는 척도 PMI가 커지도록 λ (람다) 값을 0.0~1.0 중 0.2로 설정 후 키워드들을 도출함으로써 각 토픽의 특징이 명확히 구분되어 추출될 수 있도록 하였다. 초깃값은 베타 샘플 수 10, μ = 평균 벡터, Σ = 공분산 행렬에서 추출하도록 하였다. 토픽과 각 토픽에 속한 단어를 확인하며 필터링 되지 않은 불용어와 본 연구의 정제 기준에 포함된 단어는 불용어 사전에 추가하여 필터링하고 어휘 사전에 재구축하는 등 생성과 확인 과정을 반복했다.

4. 연구 결과

1) 표본의 특성

147명의 유튜브 크리에이터로서 소비자를 대상으로 온라인 서베이를 한 결과 분석대상 중 남성은 68명(46.3%), 여성은 79명(53.7%)이며 연령대는 만19~29세($n=48$, 32.7%), 만30~39세($n=40$, 27.7%) 순으로 높게 나타났다. 유튜브 크리에이터로 활동한 기간은 6개월 미만($n=73$)이 49.7%로 가장 높았다. 4년 이상 활동한 응답자도 4.1%($n=6$)로 확인되었다. 직접 제작한 영상 콘텐츠 업로드의 평균 주기는 주 1회($n=121$, 82.3%), 주 2~3회($n=16$, 10.9%) 순으로 높게 나타났다. 보유한 유튜브 채널 구독자 수를 살펴보면 1천 명 미만이 126명(85.7%)으로 가장 높았고, 100만 명 이상의 구독자를 보유하고 있는 유튜브 크리에이터도 2명(1.4%)으로 조사되었다. 제작·공유한 영상 장르는 중복 응답 기준으로 라이프($n=45$, 13.9%)가 가장 높았으며, 게임($n=33$, 10.2%), 푸드($n=27$, 8.3%), 동물펫($n=23$, 7.1%) 순으로 그 뒤를 따랐다.

유튜브 크리에이터가 보유한 구독자 수에 따른 온라인 구전 특성을 살펴보기 위해 제작·공유한 영상 장르에 대해 교차분석을 하였다. 결과를 살펴보면 1천 명 미만의 구독자 수를 보유한 유튜브 크리에이터 중 41명은 라이프 콘텐츠를 가장 많이 제작·공유했으며 해당 장르를 제작한 크리에이터의 91.1%를 차지했다. 또한, 본 표본에서 의료·건강($n=5$)과 IT·기술·과학($n=6$)은 1천 명 미만의 구독자 수를 보유한 유튜브 크리에이터만 제작·공유한 것으로 나타났다. 1천 명 이상 1만 명 미만의 구독자 수 크리

에이터는 푸드($n=8$, 53.3%)에 대한 콘텐츠 제작·공유를 가장 많이 했으며, 장르별로 분류했을 때는 자동차($n=6$)가 37.5%로 가장 높은 비중을 보였다. 100만 명 이상의 구독자 수를 보유한 크리에이터는 공통으로 동물펫($n=2$) 콘텐츠를 다룬 점이 특징이었으며, 게임($n=1$), 푸드($n=1$), 뷰티패션스타일($n=1$)의 장르도 제작·공유한 것으로 확인됐다. 1만 명 이상 5만 명 미만 채널은 뉴스·정치·이슈와 연예인, 5만 명 이상 10만 명 미만 채널은 경제·금융·재테크, 자동차, 뉴스·정치·이슈, 10만 명 이상 100만 명 미만 채널은 취미 장르에서 각각 한 명씩 나타났다. 1천 명 미만, 1천 명 이상 1만 명 미만, 100만 명 이상의 채널보다 제작·공유한 제작의 특성에서 선별성이 명확히 드러나지는 않았다.

2) 생산된 영상 콘텐츠로부터 추출된 토픽의 특성: RQ1

응답 수는 최소 1개 이상 조건으로 하며 단어 가중치를 PMI(Pointwise Mutual Information)로 계산한 결과 아래 <표 2>와 같이 총 147개의 응답에서 토픽을 구성하는 271개의 키워드가 추출됐다(RQ1-1). 271개의 키워드가 10개의 토픽에서 나타내는 총 빈도수는 466이었다(RQ1-2).

표 2. 10개의 토픽을 구성하는 271개의 키워드

영상, 브이로그, 일상, 게임, 편집, 재미있다, 좋아하다, 만들다, 사람, 정보, 주제, 여행, 자유, 소통, 공유, 많이, 모습, 강아지, 대하다, 가장, 활동, 시청자, 도움, 관심, 패션, 생활, 콘텐츠, 업로드, 라이브, 보이다, 추억, 운동, 취미, 교육, 귀엽다, 분야, 관련, 처음, 방송, 실시간, 제작, 즐거움, 음악, 내용, 먹방, 촬영, 자연, 영상미, 고양이, 수업, 아이, 직장, 반려견, 즐기다, 함께, 노력, 직접, 연주, 노래, 지식, 전달, 기존, 리뷰, 위하다, 사회, 경험, 기록, 어리다, 공감, 나오다, 가족, 요리, 배경음악, 신경, 요소, 제품, 비교, 올리다, 구도, 사용, 다이어트, 새롭다, 메뉴, 딱히, 구독자 수, 얼마이다, 저장, 건강, 차량, 시승, 전달력, 실력, 다양, 접근성, 용이, 남기다, 온라인, 헬스장, 다른, 동료, 동호회, 구입, 골프, 연습, 방법, 환경, 정성, 들어가다, 자녀, 역사, 컨셉, 생각, 그저, 그대로, 산책, 해외, 소개, 패션쇼, 유명한, 밴드, 신규, 접하다, 겨울왕국, 라이, 특징, 후원, 채팅, 식물, 키우다, 실무, 동기, 동창, 리액션, 너무, 공감대, 다시, 지금, 학교, 과제, 시연, 스피드, 페인팅, 제공, 주식, 토크, 경제, 유튜브, 육아, 베이비, 특정, 자신, 특별, 매니아층, 토크, 편하다, 과외, 영화, 채널, 알차다, 생방송, 활기차다, 인상, 멋지다, 풍경, 행복감, 지인, 베이킹, 관심거리, 남자, 어르신, 손쉽다, 캠핑, 로또, 밀집, 동물, 평소, 하울, 시간, 들이다, 바탕, 전자제품, 개봉, 창단, 쿠키코, 언덕, 정업, 이별, 여름, 눈싸움, 음원, 커버, 부르다, 인천, 동네, 레고, 나라, 아직, 블로그, 흥미, 제일, 플레이, 최대한, 아기, 음식, 장난감, 언박싱, 힘들다, 개인, 테이블, 수집, 관하다, 다루다, 기억, 느끼다, 집중, 스타일, 검토, 소비자, 현명, 소비, 드리다, 이제, 시작, 긴장감, 설레다, 공략, 편안, 인기, 검색어, 레고프렌즈, 실제, 재건축, 현장, 마지막, 철거, 이루어지다, 마음, 아프다, 가지다, 세일, 의미, 생일날, 기념, 조금, 재미, 매끄럽다, 흘러가다, 모든, 애착, 홍보, 문화, 언어, 구독자, 진짜, 성공, 부자, 중요, 행복, 주장, 대학원, 지내다, 반경, 범위, 기법, 같이, 전반적, 주의, 기을 이다, 작업, 반응, 장소

10개의 토픽을 구성하는 키워드는 아래 <표 3>과 같다. <표 3>의 키워드는 토픽 내에서 발현될 확률이 높은, 즉 토픽에 강한 영향을 미치는 상위 10개의 핵심 키워드이다. 토픽별 괄호 안의 숫자는 해당 토픽을 구성하는 키워드들의 빈도수이다. 토픽 구성 키워드는 토픽 모델링 결과로 도출된 것이며, 토픽 라벨은 해당 토픽 확률이 높게 측정된 문서를 확인하고 토픽별 추출된 키워드의 의미적 연관성을 반영하여 연구자가 직접 설정하였다. 키워드는 해당 토픽의 주제와 연관성이 높은 키워드라 할 수 있다.

표 3. 토픽별 상위 8개의 핵심 키워드 (k=10)

Keyword	Topic 1 (65)	Topic 2 (60)	Topic 3 (56)	Topic 4 (50)	Topic 5 (48)
	지식	소통	추억	참여	기록
1	일상	게임	영상	주제	브이로그
2	정보	소통	강아지	분야	좋아하다
3	교육	시청자	음악	대하다	편집
4	공유	가장	추억	먹방	라이브
5	패션	업로드	모습	연주	제작
6	많이	실시간	귀엽다	직접	수업
7	생활	자유	관련	노래	베이킹
8	관심	리뷰	반려견	사용	개인

Keyword	Topic 6 (42)	Topic 7 (46)	Topic 8 (43)	Topic 9 (31)	Topic 10 (25)
	일상	경험	제작	취미	활동
1	운동	사람	만들다	재미있다	콘텐츠
2	즐거움	도움	내용	여행	건강
3	가족	처음	보이다	활동	활기차다
4	위하다	기록	영상	취미	생방송
5	요리	비교	방송	블로그	다이어트
6	요소	올리다	자연	커버	베이비
7	함께	노력	촬영	음원	육아
8	구도	다시	영상미	창단	동창

Note. 토픽 번호는 중요도/비중과 무관하며, 토픽 모델링 과정에서 임의로 부여된 것

유튜브 크리에이터로서 소비자가 자발적이며 애착을 갖고 제작한 영상의 특징에 대한 주요 주제는 ‘지식’, ‘소통’, ‘추억’, ‘참여’, ‘기록’, ‘일상’, ‘경험’, ‘제작’, ‘취미’, ‘활동’으로 나타났다. 토픽1(지식)에서는 ‘정보’, ‘교육’, ‘패션’뿐 아니라 ‘공유’, ‘생활’, ‘관심’ 등의 키워드가 함께 제시되어 있다. 이는 지식 노하우를 가진 사람이라면 누구나

다양한 주제와 형태의 콘텐츠를 만들고 소통하는 지식 공유 플랫폼으로 유튜브가 활용된다고 해석할 수 있다. ‘소통’으로 명명한 토픽2는 ‘게임’, ‘시청자’, ‘실시간’, ‘자유’, ‘리뷰’ 등으로 구성되어 있다. 토픽2를 구성하는 키워드로는 ‘게임’의 비중이 가장 크지만, ‘게임’의 콘텐츠 특성과 토픽을 구성하는 키워드의 전반적인 맥락을 살펴봤을 때 본 토픽은 ‘소통’과 관련되어 있다고 해석할 수 있다. 토픽3(추억)은 토픽 내에서 ‘추억’ 키워드가 출현했을 뿐 아니라 ‘강아지’, ‘음악’, ‘반려견’ 등의 키워드로 자신의 삶을 영상을 통해 남기는 특성을 명확히 보여주고 있다. 토픽4를 구성하는 주요 키워드인 ‘떡방’, ‘연주’, ‘직접’, ‘사용’ 등은 다양한 ‘주제’와 ‘분야’에서 유튜브 크리에이터로서 소비자의 새로운 경험을 공유하는 모습을 보여줌에 따라 ‘참여’로 토픽을 명명하였다. 토픽5는 ‘브이로그’가 주요 키워드로 구성되어 있으며, ‘라이브’, ‘수업’ 등 크리에이터가 유튜브에서 ‘좋아하는’ 콘텐츠로 시청자와의 소통을 남기는 모습을 유추할 수 있다. 이에 따라 토픽5는 ‘기록’으로 명명하였다. 토픽6은 ‘운동’, ‘가족’, ‘요리’, ‘함께’ 등과 같이 일상적인 소재와 연관된 콘텐츠를 촬영하고 공유하는 키워드들로 추출되어 ‘일상’으로 명명하였다. 소비자가 자발적으로 직접 생산하고 공유함에 따라 정해진, 인위적인 내용이 아닌 그들의 일상 모습을 다채로운 내용으로 솔직하게 공유하고 있음을 알 수 있다. 토픽7(경험)의 경우 ‘사람’, ‘처음’, ‘기록’, ‘비교’와 같이 삶에서 자신의 이야기를 공유하고 소통하는 키워드로 구성되어 있다. 구성 키워드들은 특징적인 주제를 내포하기보다 유튜브 플랫폼으로 개인의 경험이 어떻게 이루어지는지 나타내고 있다. 토픽8은 다른 토픽과 달리 ‘만들다’, ‘영상’, ‘보이다’, ‘내용’, ‘영상미’ 등과 같이 콘텐츠를 촬영하고 공유하는 제작과정을 다루는 키워드들로 추출되어 ‘제작’으로 명명하였다. 토픽9를 구성하는 키워드 결과에서 ‘재미있다’, ‘여행’, ‘활동’ 등은 크리에이터의 관심사와 경험을 유튜브를 통해 간접적으로 제공한다는 것을 유추할 수 있다. 이상의 키워드에 근거하여 토픽9는 ‘취미’로 설정하였다. 마지막으로 토픽10은 ‘건강’, ‘활기차다’, ‘다이어트’ 등 실내외에서 할 수 있는 콘텐츠가 공통으로 나타날 뿐 아니라 ‘베이비’, ‘육아’ 등 아이를 키우는 주제를 다루는 특성이 두드러진다. 이를 반영하여 토픽10은 ‘활동’으로 토픽 명을 선정했다. 토픽의 빈도수는 25~65로 나타났다. 특히 토픽1(지식)을 구성하는 키워드가 10개의 토픽 중 가장 빈번하게 발생했다. 이를 통해 유튜브 크리에이터로서 소비자의 온라인 구전에서 정보, 지식과 관련된 특성이 가장 두드러지게 나타나고 있음을 알 수 있다.

3) 생산된 영상 콘텐츠로부터 추출된 토픽간 연관성: RQ2

연구문제 2에서는 앞서 추출된 토픽 간의 연관성을 내재적 의미를 바탕으로 분석했다. 271개로 압축된 키워드를 연관성이 높은 확률로 살펴보고자 아래 <표 4>에서는 다른 키워드와 함께 나타나는 경우가 많은 순으로 키워드를 추출했다. 연관성은 단어의 가중치와 로지스틱 정규분포를 따르는 CTM의 특성에 따라 공분산 행렬에서 상관 계수 (correlation coefficient)를 계산했다. <표 4>에서 플롯(float)은 전체 데이터 내 해당 키워드가 차지하는 비중이다.

표 4. 연관성이 높은 순으로 추출한 키워드

No	Keyword	float	No	Keyword	float
1	일상	.084979	11	아이	.030438
2	정보	.065695	12	눈싸움	.024221
3	교육	.056885	13	편안	.024221
4	공유	.056099	14	장난감	.021493
5	패션	.054250	15	스타일	.021493
6	많이	.052190	16	제공	.021493
7	생활	.050897	17	연박싱	.021493
8	관심	.050643	18	알차다	.021493
8	공감	.042946	19	자유	.021215
10	직장	.033166	20	정성	.019897

분석 결과를 살펴보면 ‘일상’, ‘정보’, ‘교육’, ‘생활’, ‘직장’ 등 전반적으로 유튜브 크리에이터로서 소비자의 일상 소재를 다룬 키워드가 연관성에서 강하게 나타났다 (RQ2-1). 그 중 ‘일상’ 키워드가 .084979로 키워드 간 연관성에서 가장 높은 것으로 나타났다. ‘일상’과 ‘생활’ 키워드를 포함한 실제 응답을 살펴보면 ‘딱히 구독자 수와 뷰에 얽매이지 않고 일상 저장용으로 업로드 함’, ‘아이의 일상 중 많이 웃는’, ‘귀여운 강아지의 일상’, ‘가족 일상’, ‘브이로그 일상’, ‘대학원 생활을 하며 지내는 일상 모습을 공유’, ‘직장 생활 브이로그’, ‘생활에 밀접한 내용과 사회적 관심도가 높은 내용’ 등과 같이 가족, 동물, 직장 등 자신의 삶과 관심사를 솔직하게 풀어낸 온라인 구전이 활발히 나타나는 모습을 보였다. ‘영상’ 키워드에 이어 ‘정보’(.065695), ‘교육’(.056885), ‘공유’(.056099), ‘패션’(.054250) 순으로 높은 연관성을 보였다.

본 연구에서는 토픽의 분포, 토픽의 주요 키워드, 토픽 간 연관성을 살펴보기 위해 pyLDavis 라이브러리(library) 활용했다. 분석 결과 <그림 3>과 같이 토픽의 크기와 토픽 간 거리를 나타내는 IDM(Intertopic Distance Map), 토픽을 구성하는 키워드의 발현 정도인 현저성(saliency)과 토픽 간 연관성(relevance)이 확인됐다. IDM은 각 토픽이 다른 토픽과 어느 정도 연관성을 지니는지와 토픽 간의 유사도를 제시한다(신명선·조경원, 2019). <그림 3>의 좌측은 토픽 K개의 차원을 제1 주성분(Principal Component 1, 이하 PC1)과 제2 주성분(Principal Component 2, 이하 PC2)의 2차원 임베딩 벡터(embedding vector)로 축약해 나타낸 것이며, <그림 3>의 우측은 전체 코퍼스 내에서 가장 중요한 키워드 30개를 출력한 것으로 ‘브이로그’, ‘영상’, ‘게임’, ‘재미있다’ 등을 살펴볼 수 있다.

본 연구에서는 10개의 토픽에 대한 토픽별 중요한 단어 30개를 출력하도록 설정하였는데, 토픽을 구성하는 단어의 출현 조건을 결정하는 λ 값은 단어 빈도를 고려하여 0.0~1.0 중 0.2로 조절하여 각 토픽을 구성하는 단어를 뚜렷이 구분하고자 하였다. 원은 본 연구에서 설정한 10개의 토픽이며 원의 크기는 전체 키워드의 빈도수 중 해당 토픽에 배분된 단어의 빈도수로 원이 클수록 해당 토픽이 응답자로부터 많이 다뤄진 주제임을 의미한다. 파란색 막대그래프는 단어의 전체 빈도수로 각 단어와 해당 토픽

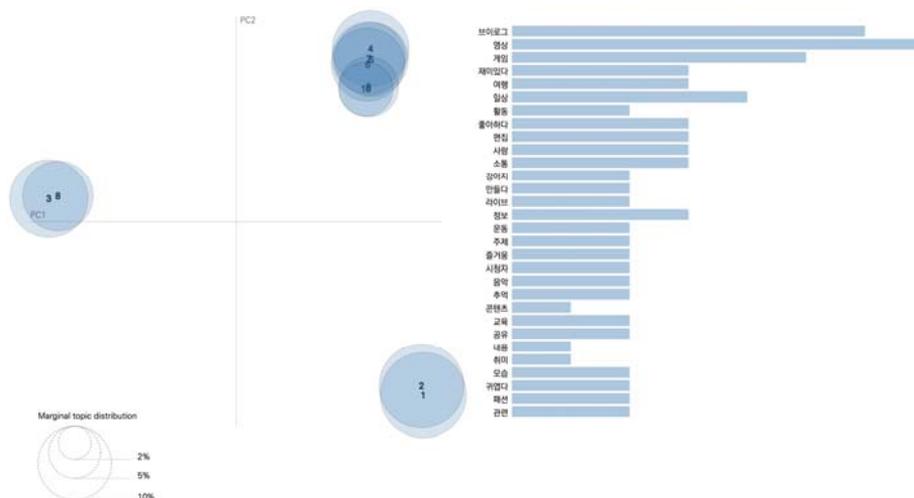


그림 3. (좌)Intertopic Distance Map(IDM) Analysis | (우) 전체 데이터에서 현저성이 높은 30개 키워드

1. 현저성(term w) = $\text{frequency}(w) * [\sum_t p(t | w) * \log(p(t | w)/p(t))]$ for topics t (Chuang et al, 2012)
2. 연관성(term w | topic t) = $\lambda * p(w | t) + (1-\lambda) * p(w | t)/p(w)$ (Sievert & Shirley, 2014)

과의 연관성을 나타낸다.

분석 결과 <그림 3>과 같이 토픽1(지식)의 원이 가장 큰 것으로 확인되어 분석한 데이터에서 토픽1의 상대적 중요도(relative prevalence)가 가장 높은 것을 알 수 있다. 또한, 10개의 클러스터(cluster)는 주제에 따라 분포 비율에 차이를 나타내는 것으로 볼 수 있다(RQ2-2). 그중 가장 큰 원인 토픽1(지식)은 10개 토픽 중 13.7% 비율을 차지했다. 그 뒤를 이어 토픽2(소통) 13.3%와 토픽3(추억) 11.3%, 토픽4(참여) 10.9%, 토픽5(기록) 10%, 토픽6(일상) 9.9%, 토픽7(경험) 9.8%, 토픽8(제작) 9.2%, 토픽9(취미) 6.6%, 토픽10(활동) 5.4% 순으로 토픽 비율을 차지하였다. 토픽10(활동)은 가장 작은 원을 보이며 ‘건강’, ‘생방송’ 등을 다룬 토픽10(활동)이 유튜브 크리에이터로서 소비자들 사이에서 상대적으로 적게 나타나는 온라인 구전 특성임을 시사한다.

같은 사분면에서 비슷한 위치에 존재하는 토픽들은 서로 유사한 문맥을 지니며 의미적 연관성을 보인다. 반면 토픽 간의 거리가 먼 경우 판별 타당도가 높고 주제가 뚜렷하게 구분된다. 토픽들의 겹치는 영역을 살펴본 결과 ①토픽1(지식)과 2(소통), ②토픽3(추억)과 8(제작), 그리고 ③토픽4(참여), 5(기록), 6(일상), 7(경험), 9(취미), 10(활동) 등 세 집합이 나타났다. 토픽들이 한 사분면에 집중되어 있으며 겹치는 영역이 크다는 것은 토픽 간 높은 유사성과 비슷한 문맥을 지녔다는 점을 말해준다. 토픽 간 유사성이 높다는 것은 토픽이 구성하는 키워드를 많이 공유하고 있음을 나타낸다. 즉, 겹친 토픽은 유튜브 크리에이터로서 소비자의 유사한 온라인 구전 특성을 나타낸다. 즉, 도출된 결과에 따라 지식과 소통, 추억과 제작, 그리고 참여, 기록, 일상, 경험, 취미, 활동을 다룬 세 개의 주제 묶음에서 콘텐츠 간 높은 연관성과 유사성을 보임을 알 수 있다. 토픽 분포를 바탕으로 살펴본 토픽 간 연관성, 유사성에 대한 이상의 분석을 정리하면 토픽1(지식)과 토픽2(소통), 토픽3(추억)이 영상 콘텐츠 주제이자 특성으로 많이 다뤄지고 있으며 10개의 주제는 크게 3개로 나뉘어 비슷한 특성의 결과를 보여주고 있다.

4) 생산된 영상 콘텐츠로부터 추출된 토픽의 구조적 관계: RQ3

연구문제 3에서는 토픽 간 상관관계를 좀 더 명확히 밝히기 위하여 연구문제 1과 2에서 도출한 토픽, 키워드와 토픽 간 연관성을 바탕으로 상관 계수(correlation coefficient)

와 네트워크 구조를 도출하였다. pyLDAvis 라이브러리는 하나의 기준으로 다차원의 벡터(vector)를 2차원으로 축소하여 표현한 결과이므로 IDM에 표현된 토픽의 거리로만 토픽의 관계를 분석하는 것에는 한계가 있다. 토픽 간의 거리에 대하여 더욱 깊이 있는 이해를 위해서는 다각도로 벡터 공간을 형상화하여 살펴볼 필요가 있다(전승봉·고태경, 2020). 따라서 연구문제 3에서는 토픽 간 0.1 이상의 상관 계수만 간선으로 나타냄으로써 토픽 간의 상관관계를 네트워크 구조로 파악하였다. 본 연구에서는 pyvis 라이브러리와 Barnes-Hut 알고리즘을 사용해 10개의 토픽을 네트워크 구조로 시각화했다. 토픽 간 상관 계수는 0.1 이상인 경우에만 간선을 나타내도록 하였으며, 토픽 간 상관계수의 크기가 클수록 노드 간의 링크 두께는 굵어지도록 설정하였다. 라벨은 단어 간 연관성에 따라 추출된 271개의 단어 중 토픽을 구성하는 상위 3개를 나타내도록 하였다. 아래 <그림 4>는 유튜브 크리에이터로서 소비자의 온라인 구전 특성을 나타내는 10개의 토픽을 네트워크 구조로 제시한 것이다.



그림 4. 토픽의 상관관계에 따른 네트워크 구조

<그림 4>와 같이 10개 중 6개의 토픽은 .15~.20의 상관계수 값을 나타냈다. 관계가 깊은 소비자의 온라인 구전 특성으로는 토픽5(기록: 브이로그, 좋아하다, 편집)와 토픽9(취미: 재미있다, 여행, 활동)가 .20으로 가장 높은 상관계수 값을 보였다. 상관성이 가장 높게 나타난 토픽5(기록)와 토픽9(취미)는 소비자가 좋아하는 일상이나 취미 등의 콘텐츠에 대해 직접 제작하고 리뷰 하고자 하는 소비자의 욕구를 실질적으로 확인할 수 있는 결과라고 해석할 수 있다. 또한, 소비자가 좋아하는 콘텐츠를 독립적이고 자발적으로 기록하고 싶어 하는 것을 시사하며 소비자의 관심사와 직접 참여가 가능한

소재에 대한 조사가 필요하다고 정리할 수 있다. 그에 이어 토픽2(소통: 게임, 소통, 시청자)는 토픽7(경험: 사람, 도움, 처음)과 .16의 상관성을 나타냈으며 토픽9(취미)와도 .15의 상관관계를 갖는 것으로 확인됐다. 또한, 토픽8(제작: 만들다, 내용, 보이다)과 토픽3(추억: 영상, 강아지, 음악)이 .16의 상관계수 값을 보이며 유튜브 크리에이터로서 소비자 개인의 사적 영역에서 영상을 통한 정보 공유 활동을 적극적으로 참여시킬 수 있음을 제시한다. 그 외 토픽1(지식: 일상, 정보, 교육), 토픽4(참여: 주제, 분야, 대하다), 토픽6(일상: 운동, 즐거움, 가족), 토픽10(활동: 콘텐츠, 건강, 활기차다)은 토픽 간 상관관계를 지니지 않는 것으로 확인됐다.

5. 결론

1) 논의 및 시사점

2059세대에서 인터넷 동영상 광고는 관심 있는 상품 정보가 나와 클릭하는 경우가 가장 높게 나타났지만, 실수로 잘못 클릭한 경우가 2위로 나타났다(DMC, 2020). 더욱이 뒤 광고나 허위광고 등에서 비롯된 문제를 한정된 기업의 마케팅 예산에서 해결할 새로운 전개가 필요한 상황이다. 이러한 상황에서 개인 미디어의 발달과 확장은 기업의 새로운 커뮤니케이션 방향을 제시했다. 개인 미디어 콘텐츠는 성장이 정체된 국내 영상 산업에 새로운 기회를 제공해 주고 있다. 방송 콘텐츠 제작과 유통 플랫폼이 제한된 국내 영상 산업에서 개인 미디어 콘텐츠는 방송 콘텐츠와 차별화된 포맷과 소재로 콘텐츠의 영역을 확대하고 있으며, 이를 기반으로 광고 수익 외에 새로운 수익 모델을 만들어가고 있다(문화체육관광부, 2018). 이상의 맥락에 따라 본 연구는 유튜브 크리에이터로서 소비자를 대상으로 자발적인 콘텐츠 제작·공유 특성을 탐색함으로써 소비자의 온라인 구전에 대한 이해를 높이는 데 목적을 두었다. 분석을 위해 다수의 정보생산과 공유 활동이 이루어지는 유튜브에서 채널을 직접 운영하는 유튜브 크리에이터인 소비자가 자발적이며 주체적으로 제작·공유한 영상 콘텐츠의 내재적 특성을 개방형 응답으로 수집했으며, 특성 간 상관관계를 탐색했다. 맥락과 관련된 단서들을 이용하여 유사한 의미를 가진 단어들을 클러스터링하는 방식으로 주제를 추론하는 기법(Steyvers

et al., 2007; Blei, 2012; 강범일 등, 2013)인 토픽 모델링 중 CTM을 적용하여, 수집된 데이터로부터 주제를 추출했다. 구체적으로, 총 147개의 데이터를 통해 ① tomotopy 라이브러리를 활용한 CTM을 바탕으로 271개의 키워드와 10개의 토픽이 도출되었고, ② pyLDavis를 활용한 토픽 분포도를 통해 토픽 간 유사성이 분석되었다. 또한, ③ pyvis와 Barnes-Hut으로 네트워크 구조 분석을 하여 데이터에 함축된 상관관계를 파악했다.

개인의 온라인 구전 특성을 구성하는 키워드와 키워드의 빈도수를 확인한 연구문제1에서는 PMI로 단어의 연관 정도에 따라 가중치를 부여해 총 10개의 토픽을 구성하는 271개의 키워드가 추출됐다. 271개의 키워드는 발현 확률이 높고 자발적인 온라인 구전에 따라 생성된 영상의 주요 특성을 의미한다. ‘지식’, ‘소통’, ‘추억’, ‘참여’, ‘기록’, ‘일상’, ‘경험’, ‘제작’, ‘취미’, ‘활동’ 등 10개 토픽의 구성 차원을 살펴봤을 때 31에서 65까지 토픽별로 키워드의 빈도수가 다양하게 나타나고 있었다. 토픽별로 토픽에 가장 큰 영향을 미치는 1순위 키워드를 확인한 결과 토픽1(지식)은 ‘일상’, 토픽2(소통)은 ‘게임’, 토픽3(추억)은 ‘영상’, 토픽4(참여)는 ‘주제’, 토픽 5(기록)는 ‘브이로그’, 토픽6(일상)은 ‘운동’, 토픽7(경험)은 ‘사람’, 토픽8(제작)은 ‘만들다’, 토픽9(취미)는 ‘재미있다’, 토픽10(활동)은 ‘콘텐츠’로 나타났다. 추출된 키워드를 세부적으로 살펴보면, 수익을 올리는 것이나 돈벌이 수단으로 유튜브를 시작하는 것이 아닌 유튜브를 통해 자신이 좋아하는 것을 누군가와 공유하고 기록을 남기고 싶다는 바람으로 시작해야 한다고 주장한 백종원 대표(강일용, 2020)의 말처럼 유튜브 크리에이터로서 소비자는 ‘일상’, ‘정보’, ‘교육’을 ‘공유’하고 ‘관심’ 있으며 ‘공감’ 가는 콘텐츠를 제작하는 모습이 두드러지게 나타나고 있다. 또한, 유튜브가 우리에게 준 가장 강력한 혜택인 ‘자유’(박현우, 2017)를 증명하듯이 토픽2(소통)에서는 유튜브가 크리에이터에게 제공하는 커뮤니케이션 방법인 ‘소통’, ‘시청자’, ‘자유’, ‘리뷰’ 등의 키워드가 추출되었다. 특히, 2030을 중심으로는 브이로그를 통해 타인과 공감대를 형성하고 소통하는 모습이 두드러지게 나타나고 있다. 브이로그를 하는 사람인 ‘브이로거(Vlogger)’는 반려견의 일상, 직장 생활, 여행, 맛집 탐방 등 기억하고 싶은 순간이나 타인에게 공유하고 싶은 것을 영상으로 남겨 자신의 블로그나 SNS 등에 공유한다(김수완, 2020). 토픽을 구성하는 주요 단어인 ‘브이로그’, ‘베이킹’, ‘육아’, ‘강아지’, ‘추억’, ‘운동’, ‘동창’ 등을 통해서도 소비자가 공유하는 정보에 자신의 기호와 객체를 연계하여 정보 생산을 이루고 있

음을 유추해볼 수 있다. 개인 미디어는 콘텐츠 기획과 제작, 그리고 유통이나 송출에 이르기까지 개인이 혼자서 운영하는 특성을 보인다. 크리에이터는 자신이 만든 콘텐츠를 통해 수용자와 지속적인 상호작용을 하면서 공감대를 형성하고 이들과 커뮤니티를 형성하기도 한다(문화체육관광부, 2018). 이러한 측면을 반영하여 볼 때, 유튜브 크리에이터인 소비자로부터 ‘제작’과 관련한 키워드를 중심으로 온라인 구전 특성이 나타났다고 해석해볼 수 있다.

추출한 키워드 간 연관성을 분석한 연구문제2에서는 ‘일상’이 .084979로 가장 높게 나타났으며 이어서 ‘정보’(.065695), ‘교육’(.056885), ‘공유’(.056099), ‘패션’(.054250)의 순으로 나타났다. 토픽을 구성하는 단어의 출현 조건을 결정하는 λ 값을 0.2로 설정한 IDM 분석을 통해서 토픽1(지식)이 13.7%의 비율로 가장 높은 분포 비율을 지닐 수 있었다. 또한, 2차원(PC1, PC2)으로 축약하여 사분면에 속한 토픽들의 의미적 연관성을 살펴봤을 때 세 개의 클러스터인 ① 토픽1(지식)과 2(소통), ② 토픽3(추억)과 8(제작), 그리고 ③ 토픽4(참여), 5(기록), 6(일상), 7(경험), 9(취미), 10(활동)이 도출되었다. 이를 통해 각 클러스터에 속한 토픽끼리 높은 유사성을 지녔다는 것을 추적할 수 있었다. 유튜브가 정보 유익성 측면에서 다른 SNS에 비해 상대적으로 높은 긍정적인 인식이 형성된 모습은(오픈서베이, 2021) 가장 높은 키워드 빈도수를 보인 토픽1(지식)의 결과와 연결 지어 해석할 수 있다. 또한, 이론적 고찰에서 살펴봤듯이 온라인 구전을 장려하는 가장 중요한 요소인 ‘소유 의식’, ‘승리’, ‘돕고 싶어 하는 의지’(Cheung et al., 2012)가 토픽5(기록), 토픽2(소통)와 ‘도움’, ‘정보’, ‘직접’, ‘수업’ 등의 키워드를 통해 실제 국내 유튜브 크리에이터로서 소비자의 온라인 구전 특성으로 나타남을 알 수 있다.

연구문제1과 2의 결과를 바탕으로 유튜브 크리에이터로서 소비자의 관심과 공유에 따라 생산된 영상 콘텐츠 토픽의 구조적 관계를 살펴본 연구문제3은 추출된 토픽 간 상관계수를 통해 연관성을 분석하고, 토픽 간의 연결망을 네트워크 구조로 표현했다. 상관계수 10% 이상의 토픽만 간선으로 연결한 결과 토픽3(추억)과 토픽8(제작), 토픽9(취미)와 토픽2(소통), 토픽5(기록)와 토픽9(취미), 토픽2(소통)와 토픽7(경험)에서 상관성을 보여주었다. 토픽 간 상관관계를 분석하는 것은 소비자의 자발적인 온라인 구전 특성을 개별적이고 독립적인 것이 아닌 소비자 간 연결된 행동임을 시사한다. 즉, 유튜브 콘텐츠처럼 소비자의 행동도 실제 네트워크처럼 연결을 이루고 있으며 소비자

간 영향을 미칠 수 있다고 해석할 수 있다.

소비자들은 서로 간의 가치를 표현하고, 반영하고자 실제적인 연결 여부와 상관없이 유튜브라는 가상공간에서 영상 콘텐츠를 생산하고 공유하는 행위를 통해 내·외적의 상호작용을 실현한다. 이러한 점에서 본 연구는 다음과 같은 학술적·실무적 의의가 있다. 첫째, 의사결정 과정의 본질적인 주체인 유튜브 크리에이터인 소비자로부터 자유롭게 작성된 응답에서 출발하여 내재한 정보생산 및 확산 가치를 살펴본다, 소비자의 온라인 구전 특성에 대해 조사하는 방법론적인 측면에서 학술적인 함의가 있다. 특히, 구전 정보 특성 또는 소비자에 대한 탐색이 주로 거시적인 분야에서 이루어졌던 기존 연구와 달리 본 연구는 애착을 갖고 제작한 영상의 특성을 유튜브 채널을 운영하는 소비자가 직접 개방형 질문에 답하여 의의가 있다. 즉, 소비자의 인식과 의견이 함축된 데이터로부터 온라인 구전 특성에 대한 핵심 토픽과 키워드를 도출하고 이를 현상학적으로 이해하며 그 함의를 제시함으로써 본 연구의 유용성을 제시한다.

둘째, 그동안 논의가 체계적으로 이루어지지 않았던 유튜브 크리에이터의 온라인 구전 특성에 대해 크리에이터의 입장에서 도출된 데이터로 탐색함으로써 소셜미디어뿐 아니라 마케팅, 광고, 커뮤니케이션 등 다양한 분야의 연구 토대와 이론적 근거를 마련해줄 것이다. 온라인 구전의 주체에 미치는 영향 또는 주체의 특성에 대한 논의는 구전이나 유튜브에 대한 연구물들의 축적에 비해 현재까지도 체계적으로 이루어지지 않고 있다. 디지털 미디어 환경에서 소비자의 주체성, 크리에이터의 주체성, 온라인 구전이 의미하는 바를 단순히 이론적으로만 고찰했던 것을 고려할 때, 본 연구는 소비자가 어떻게 자신의 이야기를 표현하고 공유하는지 가상공간에서의 소비자 세계를 실증적으로 분석하는 데 의의가 있다.

셋째, 소비자가 직접 느끼는 감정, 행동 등에 대한 토픽을 직접 온라인 구전을 한 경험이 있는 소비자의 응답을 바탕으로 도출하는 과정은 소비자의 온라인 구전을 높이는 밑바탕으로 작용할 수 있다. 본 연구에서 도출된 토픽들은 절대적이고 독립적이며 개별적인 의미를 지닌 것이 아닌 소비자들의 다양한 현상을 연결한다는 점에서 상황에 적합한 전략을 순환적으로 구축할 수 있는 실무적 함의를 제공할 것이다. 예컨대 본 연구결과 중 토픽10(활동)에서 나타난 바와 같이 ‘건강’에 대한 콘텐츠를 제작·공유할 경우 ‘다이어트’ 콘텐츠를 함께 녹이고 ‘생방송’으로 진행하는 것이 구전 발신자 입장에서는 자발적인 온라인 구전 기회를, 구전 수신자 입장에서는 해당 콘텐츠에 대한 소비

의도를 높일 수 있을 것이다. 따라서 본 연구에서 제시한 토픽과 키워드는 구전 수신자 및 발신자인 소비자가 자발적인 온라인 구전에 긍정적인 반응을 보일 수 있도록 무엇을, 언제, 어디서, 어떻게, 왜 이야기할지 기획 의도와 제작 방향을 간접적으로 제시한다. 이를 통해 바이럴 마케팅에서 소비자의 자발적이고 지속적인 온라인 구전 증진을 유도하는 인사이트와 방안 수립에 이바지하였다는 측면에서 시사점을 지닌다. 본 맥락과 같은 연구를 향후 기업에서 진행하는 것은 온라인 구전에 대한 다양한 소비자 관점을 형성할 수 있을 것이다.

넷째, 공통키워드에서 핵심어를 선별할 때 특정 문서에서만 등장하는 중요한 키워드들이 공통키워드 선정단계에서 배제되어 개별 문서들이 갖는 고유한 정보가 누락될(정민지 등, 2019) 가능성을 평균(μ)과 공분산 행렬(Σ)의 두 개의 모수로 상관성을 고려하는 CTM을 적용해 보완하였다. 토픽 간 독립성을 가정하는 디리클레 분포를 적용한 LDA가 문서 전체를 탐색해서 모든 토픽을 도출할 수 있을 때까지 키워드를 예측하기 어려운 것과 달리 로지스틱 정규분포를 적용한 CTM은 토픽 간 상관성을 바탕으로 관련된 키워드를 추론한다(Blei et al., 2006). 즉, 상관성으로 LDA 보다 많은 토픽과 토픽을 구성하는 키워드를 탐색할 수 있는 CTM으로 개체별, 그리고 개체 간 특성을 파악했다는 점에서 본 연구는 함의를 지닌다. 구체적으로 유튜브가 이용자의 자발적인 콘텐츠 생산과 소비의 구조로 운영된다는 점에서 소비자가 내부적으로 자유롭게 의견을 표출하고 주도적으로 정보 확산을 이끄는 특성을 실제로 일어나는 현상 속에서 텍스트 마이닝을 통해 도출한 것은 크리에이터의 관점에서 가치, 동기 등을 고찰하는 데 방향성을 제시할 것으로 기대한다.

다섯째, 기업이 위기 커뮤니케이션 상황에 대응하는 관점에서 소비자의 온라인 구전 특성에 함축된 토픽을 미리 파악하는 것은 소비자로부터 완충작용을 하는 커뮤니케이터 역할을 하는 데 도움을 주며 본 연구방법의 실무적인 가치를 증가시킨다. 자사에 적합한 온라인 구전 토픽을 영상 콘텐츠 제작과 마케팅에 반영하고 그 개념에 대해 명확히 이해하며 소비자와의 관계망을 구축함으로써 효율적으로 대처할 수 있기 때문이다. 유튜브 크리에이터로서 소비자의 개별적인 문맥을 고려해 추출한 키워드는 소비자가 지닌 차별적인 특징을 마케터, 기업 등에 객관적이고 신속하게 전달하는 데 도움을 줄 것으로 기대한다. 특히 CTM은 특정 이슈에 대해 소비자가 지닌 인식은 무엇이며, 인식 간 어떠한 연관성이 내포되어 있는지 시의적절하게 파악하는 데 효과적일 것이

다. 또한, 소비자가 추구하는 차별적인 온라인 구전 특성과 그 연관성을 바탕으로 자사 브랜드 및 제품에 적합한 크리에이터를 타겟화하여 바이럴 마케팅 효과를 극대화하는 등 광고홍보 분야에서 실무적인 유용성을 높일 수 있을 것이다.

여섯째, 디지털 컨버전스(Digital Convergence)가 진행되며 커뮤니케이션 또한 온·오프라인이 결합한 양상으로 나타나고, 일상생활에서 미디어가 필수적인 요소로 자리 잡으며 미디어 이용자의 사용 이력과 선호도 등 대량의 데이터에서 기반 된 콘텐츠가 중점적으로 생산된다. 이용자 중심의 검색 및 추천 서비스를 구현하기 위해서는 이용자의 관점에서 정보를 추출하고 연관성을 분석하는 기법이 필수적임에 따라 본 연구의 연구 방법과 결과는 긴밀한 상호작용을 보이는 산업과 소비자 행동의 논리적인 연결고리를 규명하는 측정 도구로 적용 가능하다는 의의를 보인다.

2) 연구의 한계점 및 향후 연구과제

디지털 미디어 환경에서 소비자가 콘텐츠를 생산하고 소비하면서 그 역할이 점차 주체적으로 변화함에 따라 온라인 구전 특성을 나타내는 토픽은 더 다양하거나 오히려 줄어들 수도 있다. 즉, 본 연구에서 도출한 토픽은 비지속적일 수 있다. 디지털 미디어에서 사용하는 우리의 데이터는 불에 구워서 나온 도기가 아닌 계속해서 증가하고 변형하는 인간을 담고 있는 유기체이다. 따라서 도출된 구조 내 알고리즘의 편향성을 줄이는 차원에서의 지속적인 후속 연구가 요구된다. 본 연구는 정보를 생산하고 공유하는 크리에이터의 입장에서 온라인 구전의 내재적 특성을 개방형 질문으로 탐색하는 출발 점임에 따라 이론에 입각한 접근의 한계가 있었다. 향후 연구에서는 소비자의 온라인 구전 특성을 설명할 수 있는 이론적 배경의 부족함을 채우고 소비자에게 도출 가능한 데이터를 추가로 탐색해 분석 결과에 대한 타당성과 소비자의 자발적이고 능동적인 온라인 구전 특성의 일반화 적용 가능성을 높일 필요가 있겠다. 또한 구독자 수, 조회 수, 좋아요 수, 공유 수, 장르 등 유튜브 크리에이터별 특성을 구분함으로써 이들이 지닌 영향력에 따라 나타나는 온라인 구전 특성을 구분해 살펴볼 필요가 있다. 표본 수를 증가시킨 향후 연구에서 유튜브 크리에이터로서 소비자를 새로운 기준으로 분리하는 등 소비자의 내재적 특성을 이해하는 데 다양한 방법을 구현한다면 다각적인 논의가 이루어질 수 있고, 실무적인 활용도도 높일 수 있을 것으로 기대된다.

참고문헌

- 강범일·송민·조화순 (2013). 토픽 모델링을 이용한 신문 자료의 오피니언 마이닝에 대한 연구. *한국문헌정보학회지*, 47(4), 315~334.
- 강일용 (2020, 1월 31일). 백종원 “유튜브는 취미생활로 접근해야...돈 벌이되면 고통”. *아주경제*, <https://www.ajunews.com/view/20200131145352146>
- 강현아·임희석 (2020). 토픽모델링과 주성분 분석을 활용한 온라인 쇼핑 검색 질의 유형 분류. *한국정보처리학회 2020 온라인 추계학술발표대회 논문집*, 765~768.
- 김경달·씨로켓서치랩 (2020). *유튜브 트렌드 2021-연결역량이 중요한 시대!* 서울: 이은북.
- 김규하·박철용 (2015). 토픽 모형 및 사회연결망 분석을 이용한 한국데이터정보과학회지 영문초록 분석. *한국데이터정보과학회지*, 26(1), 151~159.
- 김성훈 (2003). 제품관여도 및 제품지식에 따른 온라인 구전정보 활용 연구. *광고학연구*, 14(1), 257~280.
- 김성훈 (2018, 7월 6일). 매크로를 이용한 영화 평점 조작의 실체. *씨네21*, http://www.cine21.com/news/view/?mag_id=90596
- 김수완 (2020, 4월 4일). “제 일상 어떤가요?” 브이로그에 빠진 2030. *아시아경제*, <https://www.asiae.co.kr/article/2020040320323342835>
- 김인규·차성수 (2020). 온라인 구전과 마케팅 성과의 다이내믹스 연구: 모바일 게임 앱 리뷰를 중심으로. *한국콘텐츠학회논문지*, 20(12), 36~48.
- 김자경·이정기 (2018). 20대 대학생들의 무자극 영상 콘텐츠 지속적 이용의도 결정요인 연구-무자극 영상 콘텐츠 이용동기, 스트레스, 외로움, 확장된 계획행동이론 변인을 중심으로. *사회과학연구*, 34(2), 29~52.
- 김해룡·최현국·이문규 (2004). 인터넷 구전 수용의 영향요인. *한국마케팅학회 추계학술대회 발표논문집*, 245~255.
- 김해원 (2018). 2019 1인 미디어 산업 동향 분석과 2020 전망. *미디어 산업 보고서 시즌3-1인 미디어 산업 동향*, 3, 3~21.
- 남춘호 (2016). 일기자료 연구에서 토픽모델링 기법의 활용가능성 검토. *비교문화연구*, 22(1), 89~135.
- 남현동·남태우 (2020). 한국 플랫폼 정부의 방향성 모색: 공공기관 연구보고서에 대한 토픽 모델링과 네트워크 분석. *디지털융복합연구*, 18(2), 139~149.

- 노단·김성철·백현미 (2020). 온라인 구전이 드라마 성과에 미치는 영향: 지상파와 비지상파 차이를 중심으로. *한국방송학보*, 34(1), 70~113.
- 노희윤 (2019). 국내 OTT 서비스 이용 현황 분석-유튜브(Youtube)를 중심으로. *정보통신 방송정책*, 31(4), 30~40.
- 문화체육관광부 (2018). *개인미디어 콘텐츠 육성 방안 연구*. 세종: 문화체육관광부.
- 박선경·김상훈 (2020). 인스타그램에 나타난 인플루언서의 자아표현에 관한 연구. *OOH광고학연구*, 17(2), 35~58.
- 박철·이태민 (2006). 온라인 구전효과에 영향을 미치는 요인에 관한 비교문화적 실증연구. *경영학연구*, 35(6), 1617~1647.
- 박철·정수연 (2006). 온라인 구전에 관한 내용분석: 사이트 유형과 제품유형에 따른 차이를 중심으로. *광고연구*, 70, 91~118.
- 박현우 (2017, 7월). 브랜드, 유튜브를 다시 생각하다. *Think with Google*, <https://www.thinkwithgoogle.com/intl/ko-kr/marketing-strategies/video/rediscovery-youtube-brand-awareness/>
- 배기형 (2016, 4월 25일). 1인 방송과 크리에이터. *Naver 지식백과*, <https://terms.naver.com/entry.nhn?docId=3543409&cid=42171&categoryId=58478>
- 배정환·손지은·송민 (2013). 텍스트 마이닝을 활용한 2012년 한국대선 관련 트위터 분석. *지능정보연구*, 19(3), 141~156.
- 송진·이영주 (2016). 웹 기반 동영상 콘텐츠 이용에 관한 탐색적 연구. *사이버커뮤니케이션학보*, 33(2), 43~85.
- 신명선·조경원 (2019). 텍스트마이닝을 활용한 한국어언어치료학회지의 토픽 모델링 및 트렌드 분석: (2002~2018). *언어치료연구*, 28(3), 81~91.
- 신병철 (2003). *브랜드 확장에서의 시너지 효과측정 및 매개과정에 관한 연구*. 석사학위논문, 고려대학교 대학원, 서울.
- 안광호·하영원·박홍수 (2014). *마케팅원론*. 경기: 학현사.
- 안대천·김상훈 (2012). SNS유형별 광고속성 평가 및 태도에 관한 연구: 블로그, 트위터, 페이스북, 유튜브의 비교. *광고학연구*, 23(3), 53~84.
- 안정국·이규현·김희웅 (2015). 정보시스템 연구 트렌드 변화 분석: 토픽 델링과 네트워크 분석. *2015 한국경영정보학회 추계통합학술대회*, 561~570.
- 오픈서베이 (2021, 3월). *소셜미디어·검색 포털 트렌드 리포트 2021*. 오픈서베이.
- 유현주·김현 (2020). 1인 미디어 콘텐츠의 인플루언서 요인이 상품 구매 의도에 미치는 영

- 향: 브랜드 진정성의 매개효과를 중심으로. *한국콘텐츠학회논문지*, 20(1), 45~59.
- 윤지영·유지운·이장석 (2020). 유튜브 브이로그 이용 동기 및 이용자 특성이 이용 만족 및 지속이용의도에 미치는 영향. *한국콘텐츠학회논문지*, 20(4), 189~201.
- 이소현·김진솔·윤상혁·김희웅 (2020). 텍스트마이닝 기법을 이용한 모바일 피트니스 애플리케이션 주요 요인 분석. *한국IT서비스학회지*, 19, 117~137.
- 이소현·손새아·김희웅 (2019). 텍스트마이닝을 이용한 미용성형 주요 요인에 관한 연구. *지식경영연구*, 20(1), 45~75.
- 이신형 (2019). 인플루언서 마케팅의 부메랑 효과: 인플루언서 유형과 사회적 거리의 상호작용이 부정적 구전 의도에 미치는 영향. *대한경영학회지*, 31(11), 2005~2028.
- 이용관 (2019). 1인 미디어 크리에이터 활동 영향요인 분석: 소득창출 및 경력개발 가능성에 대한 인식수준을 중심으로. *미디어 경제와 문화*, 17(3), 43~69.
- 이은영 (2004). 커뮤니티 특성과 사이트 특성이 온라인 구전의 수용과 확산에 미치는 영향에 관한 연구. *한국마케팅학회 추계학술대회 발표논문집*, 305~325.
- 이은영·이태민 (2005). 온라인 환경에서의 정보특성이 구전효과에 미치는 영향에 관한 연구: 소비자 제품지식의 조절효과를 중심으로. *광고학연구*, 16(2), 7~33.
- 이정빈·인호 (2017). 토픽모델링 성능 향상을 위한 불용어 자동 생성 기법. *2017년 한국정보처리학회 춘계학술발표대회 논문집*, 24(1), 869~872.
- 이현선·리대용 (2004). 구전으로서 온라인 사용후기 효과에 관한연구. *홍보학연구*, 8(2), 234~268.
- 이혜린·양미현·최지희·김다은·장수연 (2019). 1인 미디어의 공신력 판단기준에 대한 세부 요소 분석: 뷰티유튜버를 중심으로. *소비자학연구*, 30(6), 75~95.
- 이효성 (2020). 유튜브의 이용동기에 대한 탐색적 고찰. *커뮤니케이션학연구: 일반*, 28(3), 77~99.
- 이희은 (2018). 디지털 미디어 환경에서의 '개인'의 의미에 대한 탐색적 연구. *한국언론정보학보*, 93, 39~68.
- 전승봉·고태경 (2020). 도로 위의 군비경쟁: LDA 토픽모델을 활용한 SUV의 인기 요인 탐구. *디지털융복합연구*, 18(10), 239~252.
- 정민지·신상민·최용석 (2019). 텍스트 데이터 분석을 위한 근접성 데이터의 생성과 군집화. *응용통계연구*, 32(3), 451~462.
- 정선임·신기하·유효정·조재희 (2021). 의학 정보 유튜브 크리에이터 평판 및 콘텐츠 속성이 시청자의 건강행동 의도에 미치는 영향: 정보 신뢰도의 매개 효과를 중심으로. *홍보*

- 학연구, 25(1), 108~149.
- 조경원·우영은 (2019). 텍스트 마이닝을 이용한 4차 산업 연구 동향 토픽 모델링. *한국정보통신학회논문지*, 23(7), 764~770.
- 조영임 (2013). 빅데이터의 이해와 주요 이슈들. *한국지역정보학회지*, 16(3), 43~65.
- 최지윤·정운재 (2017). 뷰티 인플루언서 마케팅 활용 전략: 매스미디어와 소셜미디어의 비교를 중심으로. *광고학연구*, 28(4), 47~72.
- 한국콘텐츠진흥원 (2019). 세상을 잇는 방송영상 콘텐츠의 현재와 미래-2019 BCWW NewCon 취재기. *방송 트렌드&인사이트*, 20, 45~57.
- 홍영희 (2019). 소프트웨어(SW)교육 정책에 대한 토픽 분석: 부산지역 종합지를 중심으로. *통계연구*, 24(2), 52~77.
- 홍정림·유미림·최보름 (2019). 토픽 모델링을 활용한 모바일 증강현실 앱 사용자 리뷰 분석. *한국디지털콘텐츠학회논문지*, 20(7), 1417~1427.
- 황의록·김창호 (1995). 구전(口碑)커뮤니케이션에 관한 문헌연구. *광고연구*, 26, 55~84.
- 황해정·심혜린·최준호 (2016). 빅데이터 분석을 활용한 사용자 경험 평가 방법론 탐색: 아마존 에코에 대한 온라인 리뷰 분석을 중심으로. *한국콘텐츠학회논문지*, 16(8), 517~528.
- DMC (2020, 7월 2일). 2020 인터넷 동영상 시청행태 및 광고 태도 분석. https://www.dmcreport.co.kr/report/surveyReport/premiumView?reportcode=DMCSR2020039&drtpdeth=RPT_TYPE_3&keyword_type=REPORT_KEYWORD_1
- Anderson, E. W. (1998). Consumer satisfaction and word of mouth. *Journal of Service Research*, 1(1), 5~17.
- Blei, D. M. (2012) Probabilistic topic models. *Communications of the ACM*, 55, 77~84.
- Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of science. *The Annals of Applied Statistics*, 1(1), 17~35.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 993~1022.
- Chatterjee, P. (2001). Online reviews: Do consumers use them? *Advances in Consumer Research*, 28, 129~133.
- Cheung, C. M. K., & Lee, M. K. O. (2012). What drives consumers to spread electronic word of mouth in online consumer-opinion platforms. *Decision*

- Support Systems*, 53(1), 218~225.
- Cheung, C. M., & Thandani, D. R. (2012). The impact of electronic word-of-mouth communication: A literature analysis and integrative model. *Decision Support Systems*, 54(1), 461~470.
- Chevalier, J. A., & Mayzline, D. (2003). The effects of word of mouth on sales: Online boor reviews. *Working Paper #10148*, National Bureau of Economic Research. Retrieved from <http://www.nber.org/papers/w10148>
- Day, G. S. (1971). Attitude change, media and word of mouth. *Journal of Advertising Research*, 11(6), 31~40.
- Dimaggio, P., Nag, N., & Blei, D. M. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture. *Poetics*, 41(6), 570~606.
- Godes, D., & Mayzlin, D. (2004). *Using online conversations to study word of mouth communication*. Working Paper, Harvard University Business School.
- Grant, S., Cordy, J. R., & Skillicorn, D. B. (2013). Using heuristic to estimate an appropriate number of latent topics in source code analysis. *Science of Computer Programming*, 78(9), 1663~1678.
- Greenwald, A. G., & Leavitt, C. (1984). Audience involvement in advertising: Four levels. *Journal of Consumer Research*, 11(1), 581~592.
- Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(1), 5228~5235.
- Hearst, M. A. (1999). Untangling text data mining. *In Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics*, 3~10.
- Henning-Thurau, T. (2004). Motive des lesens von kundenartikulationen im internet: Theoretische und empirische analyse. *Konsumentenverhalten im Internet*. Gabler Verlag, 171~193.
- Henning-Thurau, T., & Walsh, G. (2003). Electronic word-of-mouth: Motives for and consequences of reading customer articulations on the Internet. *International Journal of Electronic Commerce*, 8(2), 51~74.
- Laaksonen, P. (1994). *Consumer involvement: Concepts and research*. London: Routledge.

- Newman, D., Llau, J. H., Grieser, K., & Baldwin, T. (2010). Automatic evaluation of topic coherence. *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 100~108.
- Pang, B., & Lee, L. (2008). *Opinion mining and sentiment analysis*. MA: Now Publishers, Inc.
- Salomatin, K., Yang, Y., & Lad, A. (2009). Multi-field correlated topic modeling. *In Proceedings of the 2009 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics*, 628~637.
- Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys*, 34, 1~47.
- Steyvers, M., & Griffiths, T. (2007). Probabilistic topic models. *Handbook of Latent Semantic Analysis*, 427(7), 424~440.
- Tapachai, N., & Waryszak, R. (2000). An examination of the role of beneficial image in tourist destination selection. *Journal of Travel Research*, 39(1), 37~44.
- Tseng, S., & Fogg, B. J. (1999). Credibility and computing technology. *Communication of the ACM*, 42, 39~44.
- Xiao, M., Wang, R., & Chan-Olmsted, S. (2018). Factors affecting YouTube influencer marketing credibility: A heuristic-systematic model. *Journal of Media Business Studies*, 15(3), 1~26.

논문투고일: 2021년 04월 20일

논문심사일: 2021년 04월 30일

게재확정일: 2021년 05월 10일

Abstract

A Study on the Characteristics of eWOM of Consumers as YouTube Creators Using Correlated Topic Model*

Hye-Ra Oh**

M.A. student, Dept. of Media Communication, Hankuk Univ. of Foreign Studies

Yunjae Cheong***

Professor, Division of Media Communication, Hankuk Univ. of Foreign Studies

This study aims to examine eWOM characteristics such as what, how, and why consumers tell stories in order to increase understanding of the consumers actively spreading content. To this end, eWOM characteristics were collected and CTM was performed with open responses from 147 consumers working as YouTube creators, and network structures representing relations between topics with a correlation coefficient of 10% more were represented as lines. As a result of the analysis, 10 topics such as 'knowledge', 'communication', 'memory', 'participation', 'recording', 'daily', 'experience', 'production', 'hobby' and 'activity' were derived from 271 keywords. In the association, 'daily' was the highest among the keywords that make up the topic, and Topic 3(memory) and 8(production), 1(knowledge) and 2(communication), 4(participation), 5(recording), 6(daily), 7(experience), 9(hobby), and 10(activity) were found to be similar. In the correlation derived from the network structure, Topic 9(hobby) and Topic 5(recording) showed the highest correlation coefficient values of .20. As a YouTube creator who has not been systematically discussed, this study, which explored consumers through CTM from a microscopic perspective, has the following academic and practical implications. First, it provides academic implications on the methodological aspects of looking at the value inherent in a consumer's eWOM that can address a limited marketing budget. Second, the analysis results will provide a foundation and theoretical basis for research in various fields such as marketing, advertising, and communication as well as social media. In particular, This study will provide practical

* This work was supported by Hankuk University of Foreign Studies Research Fund.

** First Author, hyeraira@hufs.ac.kr

*** Corresponding Author, yjcheong@gmail.com

significance for companies to act as a buffer from consumers when faced with crisis communication situations.

KEY WORDS Personal Media • Video Content • YouTube • eWOM • Correlated Topic Model