



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

2021년 2월
석사학위 논문

인터넷 게임 방송 시청자의 채팅을 통한 후원행위 연구

조선대학교 대학원

경 영 학 과

황 요 셉

인터넷 게임 방송 시청자의 채팅을 통한 후원행위 연구

A Study on the Live Game Streaming Viewers'
Donation Behavior through Chatting

2021년 2월 25일

조선대학교 대학원

경 영 학 과

황 요 섭

인터넷 게임 방송 시청자의 채팅을 통한 후원행위 연구

지도교수 민 진 영

이 논문을 경영학 석사 학위 신청 논문으로 제출함

2020년 10월

조선대학교 대학원

경 영 학 과

황 요 섭

황요섭의 석사학위논문을 인준함

위원장 조선대학교 교 수

위 원 조선대학교 교 수

위 원 조선대학교 부교수

李 然 澤 (인)

홍영수 (인)

민진영 (인)

2020년 11월

조선대학교 대학원

【 목 차 】

ABSTRACT

제 1 장 서론	1
제 2 장 이론적 배경	4
제 1 절 인터넷 개인 방송	4
제 2 절 게임 방송	6
1. 게임 방송의 시청 동기	7
2. 게임 방송에서의 언어적 커뮤니케이션: 채팅	8
3. 게임 방송에서의 재정적 커뮤니케이션: 후원	10
제 3 장 연구 모형	13
제 4 장 방법론	16
제 1 절 연구 대상: Twitch	16
제 2 절 Twitch 내의 후원 파악	17
제 3 절 데이터 수집	18
제 4 절 분석 방법: Latent Dirichlet Allocation (LDA)을 이용한 토픽 모델링	21
제 5 절 데이터 전처리	23
제 5 장 분석 결과	27
제 6 장 결과 논의	38
제 7 장 결론	41
참고문헌	44

【 표 목차 】

[표 1] 결측 제거 후 수집 데이터	2 0
[표 2] 수집 데이터 내 후원 정보	2 1
[표 3] LDA 수행 예시	2 3
[표 4] 정수 인코딩 적용 예시	2 3
[표 5] 어간 추출, 표제어 추출 예시	2 5
[표 6] corpus 작성에 사용된 품사 태그	2 6
[표 7] Twitch 이모티콘과 의미	2 8
[표 8] 토픽 내 단어 해석	2 9
[표 9] 각 토픽 수 설정 별 결과	3 0
[표 10] 토픽 별 단어 구성	3 5
[표 11] 토픽별 채팅 비율	3 7

【 그림 목차 】

[그림 1] Twitch 게임 방송 예시(출처: 연두는말안드뤼)	7
[그림 2] 채팅 예시(출처: 좌: 연두는말안드뤼(Twitch), 우: 아크로 게임채널 (YouTube))	9
[그림 3] Twitch 비트 후원 예시	11
[그림 4] 연구 과정	15
[그림 5] 비트 입력 화면과 출력 예시	18
[그림 6] 비트 단위별 이모티콘	18
[그림 7] 수집된 채팅 데이터	20

ABSTRACT

A Study on the Live Game Streaming Viewers' Donation Behavior through Chatting

Hwang, Yo Seb

Advisor : Prof. Min, Jinyoung, Ph.D

Department of Business Administration

Graduate School of Chosun University

The live streaming contents market is continuously growing as they are free from the temporal and spatial limitation once the Internet is connected and accommodate various demands of the viewers by providing a variety of contents. Since the success of live-streaming service is on the viewer participation, researchers have studied the viewers' motivation to participate and even donate to streamers. However, few studies have investigated this phenomenon with the method that directly monitors the viewer's participation and donation behavior.

This study, therefore, delves into the viewers' chatting data from the video game live-streaming services and applies topic modeling to find the topics that appear and how the chats with the donation differ from the chats without the donation. The results show that four topics emerge in the chat data: 1) the game that streamers play, 2) the elements appeared in the game, 3) the conversations about streamers, and 4) the conversations unrelated to the streaming contents.

The fourth topic was the most frequently appeared in both general chats and donation chats, and it took more than a half in donation chats. The findings suggest that people who donate while watching video game live-streaming use a donation as a tool to show themselves and communicate with other users rather than talking about the game they are watching.

제 1 장 서론

인터넷 기술의 발전은 스마트폰의 보급과 함께 무선 통신 기기를 소유하고 있다면 누구든지 시간과 장소 상관없이 인터넷에 접속할 수 있게 하였다. 또한 콘텐츠의 용량 또는 기기의 데이터 저장 용량 제한마저 인터넷 스트리밍을 통해 해결하게 되어 미디어에서 제공하는 콘텐츠를 즉석에서 소비할 수 있게 되었다. 언제 어디서든 미디어 콘텐츠를 소비할 수 있는 환경은 사용자들의 콘텐츠 니즈를 충족시키기 위한 다양한 시장의 성장을 가져왔다.

스트리밍 서비스의 특징은 콘텐츠를 저장장치에 저장한 후 재생하지 않고 실시간으로 데이터를 수신하여 재생하는 것인데, 이를 통해 미리 제작이 완료된 콘텐츠를 송출하는 것이 아니라 콘텐츠 제작과 송출이 동시에 이루어지는 인터넷 방송이 가능하게 되었다. 이러한 스트리밍 서비스의 수요는 세계적으로 2012년 63억 달러에서 2017년 247억 달러로 연평균 31.4% 성장하였다(정민, 류승희, 2019). 현재 대중적으로 서비스되고 있는 스트리밍 서비스는 IPTV(Internet Protocol Television), OTT(Over The Top), 인터넷 방송(Internet Broadcasting) 세가지 유형으로 나뉘어지고 있다. 세 유형에서 가장 다른 운영 형태를 보여주는 것은 인터넷 방송이다. IPTV와 OTT에 참여하고 있는 콘텐츠 제공자는 지상파 또는 케이블의 채널 사업자 혹은 콘텐츠의 저작권을 보유한 법인이며, 플랫폼 사업자와 법적으로 계약을 맺고 서비스에 참여하고 있다. 반면 인터넷 방송의 콘텐츠 제공자는 개인이며, 특정 자격이 없어도 누구나 서비스에 참여할 수 있으며 서비스 참여와 이탈이 자유롭다. 따라서 인터넷 방송은 유명한 콘텐츠 제공자를 유치하고 이들을 유지하는 것이 콘텐츠 경쟁에서 중요한 전략으로 꼽히고 있다(윤혁진 *et al.*, 2019). 실제 사례를 들면, 유명 스트리머 대도서관이 AfreecaTV에서 YouTube로 이동하는 일이 있었다. 이

과정에 당시 기사들은 단순히 스트리머의 이동으로 끝나는 것이 아니라 시청자층도 함께 플랫폼을 이동할 것이며, 플랫폼의 수익이 감소할 수 있는 위기임을 언급했다(박세준, 2016; 채성오, 2016). 이 사례의 언론보도 이후 애플리케이션 정보 집계에서 AfreecaTV의 일일 활성 사용자 수치가 약 20만명가량 줄어든 것으로 관측하여 인기 스트리머의 이탈은 시청자의 이탈로 이어지는 것을 확인하였다(Ape, 2016). 스트리머의 플랫폼 선택은 방송 도구의 편의성 이외에 수익성이 중요한 기준으로 꼽힌다. 플랫폼에서 발생한 수익은 미리 정해진 비율에 따라 플랫폼과 스트리머가 나누어 가진다. 그래서 스트리머가 가져갈 수 있는 비율이 높을수록 신규 스트리머의 플랫폼 참여가 높아질 가능성을 가지고 있다. 이처럼 인터넷 방송은 단순히 개인이 만든 영상을 공유하고 시청하는 장소가 아니라 플랫폼간의 경쟁과 스트리머들이 보다 많은 시청자와 수익을 위해 서로 경쟁하고 있는 콘텐츠시장임을 알 수 있다.

방송을 통해 콘텐츠 제작자에게 돌아가는 수익은 영상 시청을 통해 발생하는 광고 수익과 시청자의 정기적인 구독, 홍보를 원하는 기업과의 제휴가 있다. 이는 여타 서비스와 유사하지만 인터넷 개인 방송이 가지는 특이한 요소는 이에 더해 시청자의 후원이라는 수익원이 존재한다는 것이다. 예를 들어 YouTube, Twitch, AfreecaTV 등에서 시청자는 공개된 방송을 시청할 때 별도의 금액을 지불할 의무는 없으나 자발적으로 플랫폼에서 판매하는 유료재화를 구입하여 출연자에게 선물하는 것이 가능하고, 이를 보통 후원이라 칭한다.

후원 행위는 통상적으로 널리 알려지고 사용되는 의미에서 후원과 다른 점이 있다. 대중적으로 알려진 후원은 재화를 받은 대상이 겪고 있는 어려움을 해소하기 위한 도움으로 인식되고 있으며 그에 따라 기부된 재화를 그 목적에 따라 사용하지 않으면 횡령으로 정의하고 있다(안영진, 2014). 그래서 후원과 관련한 연구는

아동복지와 공공시설, 공동체 의식과 관련된 사회학 연구들에서 활발히 이루어지고 있다(예: 권정미, 박태영, 2014; 김미숙, 김은정, 2005; 조윤희, 2019). 그러나 인터넷 방송의 후원은 어려움 해소보다는 스트리머의 수익원으로 인식되고 있으며 그 재화를 스트리머가 어떻게 사용하여도 책임을 묻는 일이 없다. 또한 기존 연구들에서는 재화를 선물하는 시청자의 후원 의도가 스트리머와의 관계 형성 또는 감사 표시의 성격을 또한 가지고 있음을 알아냈다(김성주, 안성아, 2020; 유재홍, 2016; 조성희, 양성병, 2020; 홍수봉, 김정규, 2017). 그래서 같은 단어이지만 그 성격이 다르기 때문에 구별하여 접근해야 하고, 다양한 관점에서 시청자의 후원 의도와 후원 행위에 대해 살펴볼 필요가 있다.

본 연구에서는 이러한 후원 행위가 인터넷 방송의 채팅 환경을 통해 발생하고 있다는 점에 주목하여 채팅 텍스트를 분석하여 후원 행위에 대해 살펴보고자 한다. 이 때 인터넷 방송이 다루는 콘텐츠 분야가 다양하므로 가장 큰 규모로 알려진 게임 방송을 대상으로 한다. 게임 방송의 영상 시청 중 발생한 채팅 데이터를 수집하여 채팅 텍스트를 토픽 모델링을 사용해 단어를 분류하고 시청자들이 어떤 주제들에 대해 대화하는지 살펴본다. 이에 더하여 각 주제의 비율을 후원이 발생하지 않은 일반 채팅과 비교하고 후원이 어떤 주제에서 많이 발생하는지, 후원 액수와 내용은 어떻게 나타나는지 살펴보고자 한다.

본 논문의 구성은 다음과 같다. 이론적 배경에서 인터넷 방송과 게임 방송의 정의에 대해서 설명하고 채팅과 후원 등 연구에서 다루고자 하는 개념을 상세히 설명한다. 연구모형에서는 문헌 연구를 통해 본 연구가 살펴보고자 하는 내용의 토대를 구축한다. 방법론에서는 해당 연구를 위해 사용한 기법 및 과정을 설명하며, 이후 데이터 분석 결과 및 시사점을 논의한다.

제 2 장 이론적 배경

제 1 절 인터넷 개인 방송

인터넷 방송은 멀티미디어 콘텐츠를 이용자의 수요가 있을 때마다 실시간으로 전송하는 서비스를 말한다(문주영, 2001). 현재 서비스 중인 인터넷 방송은 다양한 프로그램을 사전에 녹화하여 편집된 영상을 게시하는 VOD(Video On Demand)유형과 실시간으로 촬영되는 영상을 송출하는 스트리밍(Streaming) 유형으로 나뉘고 있다(최진웅, 2016). 시간과 장소에 제한 없이 콘텐츠를 즐길 수 있다는 장점이 있어 인터넷 방송으로 콘텐츠를 즐기는 시장은 스마트폰 보급과 함께 급속도로 성장하고 있다. 2020년 현재 인터넷 방송은 YouTube, Twitch 외에 AfreecaTV나 bilibili, 니코니코동화 등 여러 플랫폼이 각기 다른 기능과 콘텐츠를 시청자에게 제공하고 있다.

전세계에서 가장 많은 사용 시간을 보이고 있는 YouTube는 VOD와 스트리밍 서비스를 모두 제공하고 있고, 2016년 1월 모바일 앱 평균 체류 시간이 504.8분에서 2018년 1월 1031.4분으로 2 배 가량 증가한 것으로 나타났다(윤혁진 *et al.*, 2019). 스트리밍 서비스를 전문적으로 제공하는 Twitch는 2015년 150만 명의 스트리머, 월간 시청자 1억명의 규모에서 2018년 220만명의 스트리머, 일일시청자 1,500만명으로 성장하였고, 2020년 1분기에 평균 동시 시청자 최대 140만명을 기록하였다(IQBAL, 2020).

동일하게 인터넷 스트리밍 기술을 사용하는 IPTV, OTT서비스와 비교하여 인터넷 방송이 주목받는 이유는 개인이 자유롭게 창작자로서 플랫폼에 참여할 수 있기 때문이다. 개인이 자유롭게 플랫폼에 참여할 수 있다는 특징은 다양한 시청 수요를 만족시켜줄 다양한 콘텐츠를 인터넷 방송이 제공한다는 의미이기도 하다. 또한

인터넷이 연결되어 있고, 서비스에 접속가능한 단말기가 있다면 어디서든 시청이 가능하기 때문에 인터넷 방송은 콘텐츠에 대한 편리한 접근과 높은 선택의 다양성으로 지상파 TV나 케이블 같은 기존 미디어 산업을 대체할 가능성을 가지고 있다(정재민 *et al.*, 2005).

또한 인터넷 방송은 아마추어인 개인이 스스로 제작한 영상을 제공하는 서비스로 알려져 있지만 시청자가 증가하고 영상을 통해 수익을 창출할 수 있게 되면서 방송사와 신문사 등 기존 미디어 산업이 인터넷 방송 산업에 진출하게 되었다. 이러한 환경에서 수익과 저작권 등의 권리문제가 발생하였고, 해결하는 과정을 거치면서 아마추어의 창작활동 공간에서 준전문가의 콘텐츠 시장으로 변화하였다(오동일, 김효용, 2007).

인터넷 방송은 개인(1인 또는 소수의 팀)의 주도로 제작되는 방송이지만, 산업이 성장하면서 방송국, 신문사 등 기존 미디어가 플랫폼에 참여하고 있으므로 이와 구분하기 위하여 개인 방송이라 부르고 있다. 개인 방송은 기존 미디어와 달리 시간과 공간의 제약에서 벗어나 원하는 방식으로 다양한 형태로 제공되는 방송을 시청할 수 있다는 점과, 방송에서 실시간으로 이루어지는 스트리머와 시청자의 소통이 시청자들에게 인기를 끄는 요인으로 꼽히고 있다(이선미, 2015). 인터넷 방송의 수익은 광고와 시청자의 후원에서 발생한다. 광고 수익은 영상에 삽입된 광고를 시청하였을 때 발생하고 시청자의 후원은 영상을 시청하는 과정에서 시청자가 스트리머에게 자발적으로 건네는 재화 선물이기 때문에 영상을 시청하는 시청자가 많을수록 광고 수익이 증가하고 후원 발생의 기회가 늘어날 것이라 생각하는 인터넷 방송의 스트리머는 시청자와 조회수 증가를 목표로 시청자가 만족하는 영상을 만들기 위해 노력한다. 그러한 노력이 이어져 시청자의 영상 시청이 증가하면 할수록 주목도가 상승한 스트리머는 인플루언서로서 활동기회가 주어질 영상에 삽입된 광고

시청과 후원에 대한 수익 외에 제품 광고영상과 같은 추가 수익 활동 기회를 얻을 수 있다.

개인 방송은 시청자 수요에 따라 다양한 카테고리의 영상을 선보이고 있는데, 그 중에서 게임 카테고리가 가장 큰 규모를 가지고 있다. 2017년 기준 세계 비디오 게임 동영상 시청자는 6억6,500만명, 46억 달러의 시장 규모를 보였다(김주완, 2018). 또한 실시간 인터넷 방송 시간당 점유율 1위를 차지하고 있는 Twitch는 게임 카테고리를 전문으로 제공하는 플랫폼이기도 하다. YouTube, mixer 등 실시간 영상 공유 서비스를 제공하는 인터넷 방송 플랫폼 혹은 Facebook과 같은 소셜 네트워크 서비스에서도 게임 부문을 별도로 개설하여 제공하고 있다. 즉, 게임 카테고리의 개인 방송이 인터넷 개인 방송의 특징을 잘 보여주는 카테고리라 할 수 있겠다.

제 2 절 게임 방송

게임 방송이란 게임 플레이어가 자신의 게임 플레이 영상을 촬영 및 편집하고 해설과 감상을 덧붙여 제작한 2차 창작의 특징을 내포하는 콘텐츠를 포함한 인터넷 방송 콘텐츠를 말한다(한혜원, 김서연, 2014). 예를 들어 대도서관¹, 풍월량², 우왁굳³ 등은 방송을 통해 게임을 플레이하면서 겪는 다양한 상황에 대한 반응과 감상을 공유하고 있고, 김성희의 G식백과⁴, 중년게이머 김실장⁵과 같은 채널은 게임 산업 경험자들의 시각으로 게임과 관련한 이슈에 대해 해설하는 영상으로 인기를 끌고 있다.

¹ 대도서관: <https://www.youtube.com/user/BuzzBean11>

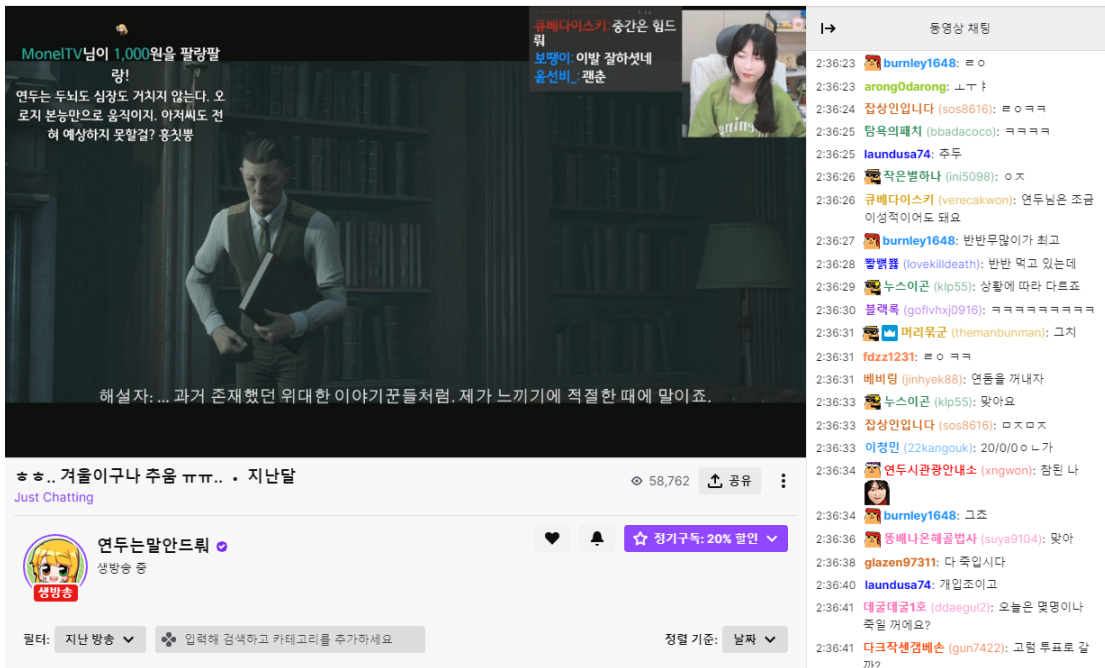
² 풍월량: <https://www.youtube.com/user/hanryang1125>

³ 우왁굳: <https://www.twitch.tv/woowakgood>

⁴ 김성희의 G식백과: https://www.youtube.com/channel/UCZ0bi2aVJngKLwFTU5g_fLQ

⁵ 중년게이머 김실장: https://www.youtube.com/channel/UCSgvQDeK_7CQj08fu7kp11w

[그림 1]은 Twitch에서 실제 게임 플레이를 공유하는 모습이다. 오른쪽 채팅창을 살펴보면 스트리머가 게임 플레이 경험을 실시간으로 공유하면서 시청자와 소통하는 것을 알 수 있다.



[그림 1] Twitch 게임 방송 예시(출처: 연두는말안드뤄)

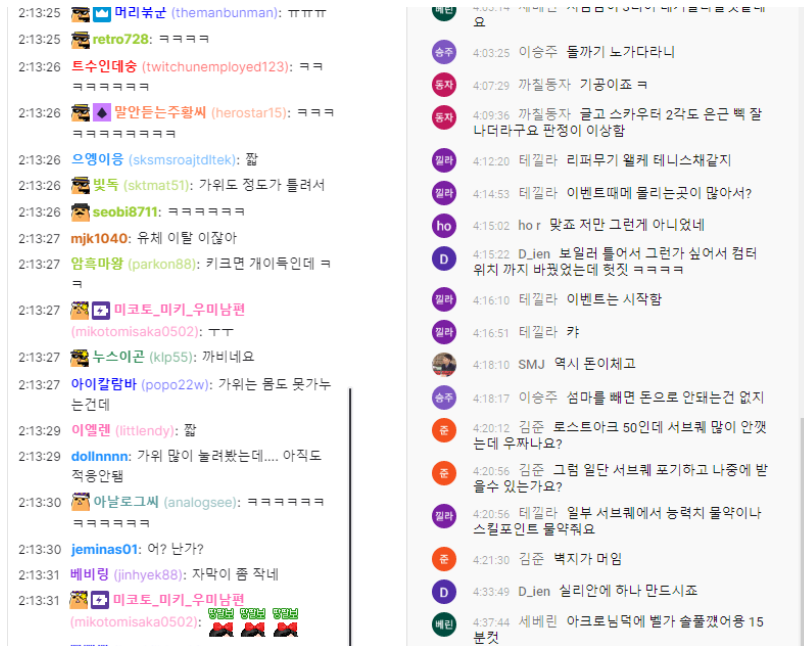
1. 게임 방송의 시청 동기

게임 방송의 특이한 점은 시청자들이 직접 게임을 플레이하면서 보낼 수 있는 시간을 플레이가 아닌 단순 관람에 들이고 있다는 점이다. 따라서 기존 연구들은 이렇게 플레이가 아니라 관람을 선택하는 시청자가 어떠한 사람들인지, 이들의 시청 동기나 채널 충성도와 관련된 시청 행위를 살펴보는 데 초점을 맞추고 있다. 김성주, 안성아 (2020)의 연구에서는 인터넷 방송을 시청하는 사람들은 각자의 시청목적이 재미인지, 정보 획득인지 교류 희망인지에 따라 스트리머에 대한

선호도가 달라지는데, 차별화된 기획을 보여주는 스트리머, 전문성을 가진 스트리머, 소통이 활발한 스트리머에 대해 각기 다른 시청 동기를 보여주는 것을 확인하였다. 게임 방송의 경우에는 시청자는 개인의 게임 숙련도와 무관하게 재미와 같은 경험적 동기를 중시하고 있으며, 여성보다는 남성이, 스마트 기기에 익숙한 20-30대가 시청 시간이 더 높은 것으로 나타났다(최민지 *et al.*, 2016). 조성휘, 양성병 (2020)은 중국의 인터넷 방송 플랫폼인 DouyuTV를 대상으로 게임 방송 시청자의 영상만족도와 후원의도에 미치는 요인을 파악하기 위해서 영상의 콘텐츠 품질을 시청가치와 관련정도, 제공하는 정보의 수준으로 정리하고 스트리머의 속성을 신뢰성과 전문성, 상호작용 정도로 나누어 영상 만족도에 미치는 영향을 확인하였다. 연구 결과 시청자는 시간을 들여 시청할 가치가 높고, 게임에 대한 전문성이 높거나 시청자와 소통이 활발한 스트리머의 방송일 수록 높은 시청만족도를 가질 수 있으며, 이러한 만족도는 후원 의도에 영향을 준다고 밝혔다.

2. 게임 방송에서의 언어적 커뮤니케이션: 채팅

인터넷 방송에서 채팅이란 스트리밍 방송에서 사용되는 도구로 실시간으로 송출되는 영상을 시청하는 시청자가 이용하는 대화수단이다. [그림 2]는 실제 YouTube와 Twitch의 게임 영상 화면의 오른쪽에 출력되는 채팅의 예시이다. 디자인의 차이는 있으나 둘 다 공통적으로 입력한 영상의 시간과 사용자, 텍스트를 표시하고 있다. 스트리머가 이렇게 실시간으로 작성된 채팅을 통해 시청자와 쌍방향으로 소통할 수 있다는 점은 시청 반응에 대한 수용이 즉각적으로 이루지고 따라서 시청자들에게 인터넷 방송이 인기 있는 요인 중 하나로 알려져 있다(윤혁진 *et al.*, 2019).



[그림 2] 채팅 예시(출처: 좌: 연두는말안드뤼(Twitch), 우: 아크로 게임채널(YouTube))

실시간으로 발생하는 채팅은 시청하는 방송의 시청자가 많고, 채팅 참여 의사가 높은 시청자가 많을수록 매우 많은 양의 채팅이 빠른 속도로 작성된다. 이와 관련해서 Ford *et al.* (2017)은 컴퓨터를 매개로 한 통신(Computer-mediated Communication)에 관한 이전 연구들을 통해 다수의 참여자가 존재하는 그룹에서 성공적인 채팅이 성립되기 위해 참여자 간의 관계와 참여자의 자기표현과 정체성이 중요하다고 가정하였다. 또한 Twitch 게임 방송 시청자들이 영상의 자극에 반응하여 이모티콘을 적극적으로 사용해 발생하는 대량의 채팅 속에서 일정한 주제가 나타나는 것을 발견하였는데, 게임 방송의 시청자들은 영상에 집중하여 영상의 흐름이 변할 때 대화의 주제를 빠르게 전환시키며 대규모 그룹의 대화가 성립되고 있음을 확인하였다. 또한 Luo *et al.* (2020)은 YouTube 스트리밍을 시청하는 채팅 참여자가 느끼는 감정이

스트리밍 종료 후 개별적으로 시청하는 감정보다 강하게 나타나는 것을 확인하였고 채팅 참여자에 의한 시청자의 감정 증폭이 있음을 밝혀낸 연구를 통해 시청자는 스트리머의 영상에만 반응하는 것이 아니라 다른 시청자의 행동에도 반응하는 것을 밝혔다.

이 밖에도 게임 방송의 채팅에 대한 연구들은 다수 존재한다. 예를 들어, 채팅의 규모(Poyane, 2018)와 방송 채널의 사회 인구 통계학적 특성 관계(Poyane, 2019)를 통해 시청 중 부정적인 내용의 채팅이 발생하는 원인들을 살펴보기도 하였다. 또한 채팅을 통해 영상 정보를 분석하려는 시도들도 있었는데, 채팅 텍스트의 언어 특성을 기반으로 영상의 하이라이트를 탐색하거나(최승혁 *et al.*, 2019), 채팅의 발생 빈도와 음성 데이터를 텍스트로 변환하여 토픽을 분석, 이에 따른 하이라이트 구간을 제공하는 방법을 제시한 연구(백종현 *et al.*, 2020)들이 그것이다. 더 나아가 채팅 텍스트의 영상 구간별 특징을 학습한 모델을 통해 시청 패턴을 탐색하기도 하였다(Zhang *et al.*, 2017). 이렇게 채팅은 게임 방송 자체의 특징이나, 이것의 시청자들의 시청 행태를 연구하는데 주요한 수단이 되고 있다.

3. 게임 방송에서의 재정적 커뮤니케이션: 후원

채팅과 더불어 인터넷 개인 방송의 특징적인 요소에는 후원이 있다. 후원이란 인터넷 개인 방송에서 시청자가 플랫폼이 제공하는 재화를 구입하여 스트리머에게 선물하는 것을 말한다. 방송 시청은 기본적으로 무료로 이루어지고 있고, 이러한 선물행위가 강제되지 않는 만큼 시청자의 자발적 선물행위가 현실의 기부 또는 후원과 비슷한 이미지를 가져 참여자들 간에 플랫폼에 구분되지 않고 통용되는 단어로 사용되고 있다. 각 플랫폼에서 제공하는 재화는 Twitch에서는 비트(Bits), YouTube에서는 슈퍼챗(Super Chat), AfreecaTV에서는 별풍선으로 지칭된다. 별풍선의 경우 단순히 재화를

전달하는 기능만을 가지고 있으며, 그 전달 과정에서 재화의 단위마다 구별되는 개성을 부여하기 위해 선물하는 재화의 단위에 따라 지정된 이모티콘이 달라지도록 하고 있다(유재흥, 2016). YouTube의 슈퍼챗은 채팅을 강조하는 방법으로 주로 쓰이는데, 스트리머가 금액의 단위에 따라 다르게 설정한 색상으로 채팅창의 메시지가 표시되고 일정시간 상단에 고정되어 누가 슈퍼챗을 사용하였는지 보여준다. Twitch의 비트는 금액에 따라 상이한 이모티콘 세트를 통해 일반 채팅보다 높은 주목도를 가진 채팅 메시지를 작성할 수 있다. [그림 3]은 Twitch에서 비트가 사용된 예시이다.



[그림 3] Twitch 비트 후원 예시⁶

이렇게 후원은 대체로 채팅을 통해서 이루어지고, 후원이 이루어질 경우 채팅 텍스트가 강조되거나 차별화되는 경우가 대부분이기 때문에 스트리머와 시청자, 시청자간 재정적 커뮤니케이션이면서 동시에 언어적 커뮤니케이션이라고 볼 수 있다.

⁶ Engadget: <https://www.engadget.com/2016-06-27-twitch-cheering-beta-bits-currency-tips.html>

AfreecaTV 시청자를 대상으로 시청자의 채팅과 후원의 관계에 대하여 알아본 연구(Yu *et al.*, 2018)에서는 채팅과 후원 발생은 별개로 이루어지며 채팅이 적음에도 후원이 발생하는 것은 스트리머와 시청자의 관계가 연예인과 팬의 관계처럼 팬덤 구조를 형성하기 때문이라고 밝히면서 채팅과 후원은 별개의 것으로 설명하고 있다.

그러나 AfreecaTV에서 사용하는 방식은 기 설정된 메시지가 입력된 이모티콘을 출력하는 방식인데 반해, Twitch와 YouTube에서는 후원이 채팅의 내용을 강조하는 유료기능처럼 사용되고 있으므로 채팅과 후원의 관계 자체를 관계없음으로 단정 짓기 보다는 위 연구가 이루어진 AfreecaTV와 다른 방식으로 후원이 이루어지는 플랫폼에서 채팅을 통해 시청자의 후원 의도를 살펴볼 필요가 있다.

제 3 장 연구 모형

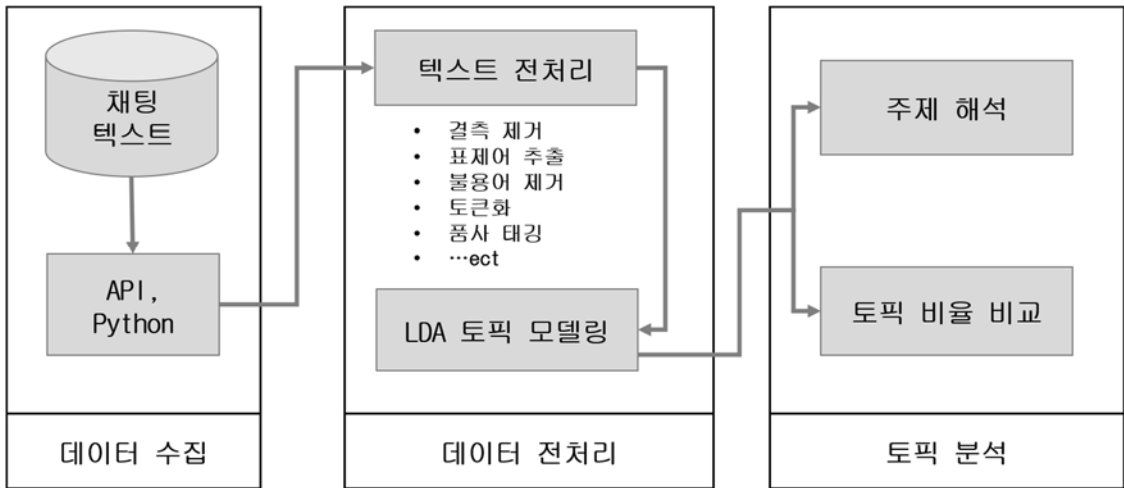
채팅을 통한 시청자의 후원 의도를 살펴보기 위하여 다음과 같은 가정을 하였다. 먼저, 실시간으로 발생하는 채팅은 얼핏 보면 각자 제각기 하고 싶은 말을 적을 뿐인 파편화된 메시지의 집합으로 보인다. 그러나 채팅에 참여한 다른 시청자에 의해 채팅 속에서 감정의 강화가 발생한다거나(Luo *et al.*, 2020), 시청중인 영상의 자극에 따라 시청자들 사이에서 일정한 주제가 발현된다는 연구(Ford *et al.*, 2017)를 고려한다면, 인터넷 방송의 채팅은 제각기 파편화된 개별의 텍스트가 아니라 주제가 있는 시청자 집단의 반응 집합임을 알 수 있다. 따라서 채팅 텍스트는 몇 가지 일정한 주제를 형성하고 있으며 인터넷 방송의 채팅 텍스트에서 주제를 살펴볼 수 있다면 후원 의도 또한 살펴볼 수 있을 것이다.

또한 후원 의도와 관련하여 리차드 바틀의 플레이어 4 유형(Bartle' s 4 Player Types)을 바탕으로 홍수봉,김정규 (2017)가 분류한 시청자 유형에 따르면 게임 자체의 요소 혹은 스트리머 관련 요소나 다른 시청자와 관련된 요소와 후원 의도가 무관하지는 않다는 것을 알 수 있다. 이 연구에서는 원하는 방송을 찾아 검색하는 성취형 게임 시청자, 특이한 게임을 다루는 방송을 찾아다니는 탐험형 게임 시청자, 다른 시청자 또는 스트리머와 대화하기를 원하는 사교형 게임 시청자, 의도적으로 방송 진행에 방해되는 행위나 부정적인 언사를 일삼아 자신의 존재를 드러내고 싶어하는 킬러형 게임 시청자로 시청자를 구분하였는데 성취형 후원자가 후원 경험 정도가 가장 높았고, 나머지 시청자 유형에 대해서는 탐험형, 사교형, 킬러형의 순서로 후원 경험이 나타났다. 즉, 게임 정보 혹은 공략 등의 시청자가 알고 싶었던 내용 제공에 대한 감사 의도의 후원이 많이 발생한다는 것인데, 따라서 후원 의도와 스트리머 및 게임의 요소와 후원의 관계를 채팅 데이터를 통해 살펴볼 수 있을

것이다.

채팅 데이터를 분석하는 방법으로는 문서에서 사용된 단어 간의 관계에 따라 그룹을 분류하는 토픽 모델링(Topic Modeling)을 사용한다. 채팅에서 많이 사용되는 단어의 카테고리를 확인하고 이를 바탕으로 토픽의 주제를 해석하는 분석 방법이다. 게임 방송에서의 후원에 관한 기존 연구들에서는 게임 방송 시청태도가 게임태도와 게임채택에 미치는 영향을 통해 게임방송 시청이 게임 구매로 이어질 수 있음을 밝혀낸 연구(이지수 *et al.*, 2019)와 같이 대부분 설문조사 방법을 사용하였다. 인터넷 사용자를 대상으로 한 설문조사는 인터넷 설문조사기관을 통해 많이 이루어지고 있는데, 인터넷 조사는 질문이 어렵거나 복잡할수록 응답률이 낮아지거나 잘못 응답하는 경우가 있으며, 응답자의 거짓, 불성실한 응답에 대한 문제가 있어 개선하기위해 노력해야하는 단점으로 제기되고 있다(심규호, 2009). 따라서 시청자의 후원 의도를 파악하기 위해서 인터넷 설문 조사를 보조하거나 대신하는 방법으로 후원이 이루어진 그 시점에서 이루어진 시청자의 행위를 직접 살펴보는 방법이 시도되어야 한다. 게임 방송을 시청하는 과정에서 시청자의 행동은 채팅, 후원, 구독하는 채널 등 플랫폼에 모두 기록되고 있기 때문에 이러한 데이터를 살펴보고 해석할 수 있다면 위에 명시한 설문조사의 한계를 극복하고 게임 방송에서의 후원 의도에 대해 보다 많은 정보를 제공할 수 있을 것이다.

또한 토픽 모델링을 통해 해석한 주제의 단어들이 후원채팅에서 어떤 비율로 사용되는지 확인해보고자 한다. 만약 후원 채팅이 일반 채팅과 그 성격이 다르다면 서로 다른 단어를 사용할 가능성이 높을 것이다.



[그림 4] 연구 과정

이러한 가정들을 확인하기 위하여 본 연구는 [그림 4]와 같은 연구 과정을 거친다. 게임 방송에서 발생한 채팅 데이터를 수집하여 토픽 모델링 훈련에 사용할 텍스트를 전처리한다. 전처리된 텍스트를 사용하여 토픽 모델링을 수행, 분류된 토픽에서 나타나는 단어로 주제를 해석하여 게임 방송의 시청자들이 어떤 주제를 대화하는지 확인한다. 이후 일반 채팅과 후원 채팅의 토픽 비율을 비교하여 일반 채팅과 후원 채팅을 비교 분석한다.

제 4 장 방법론

제 1 절 연구 대상: Twitch

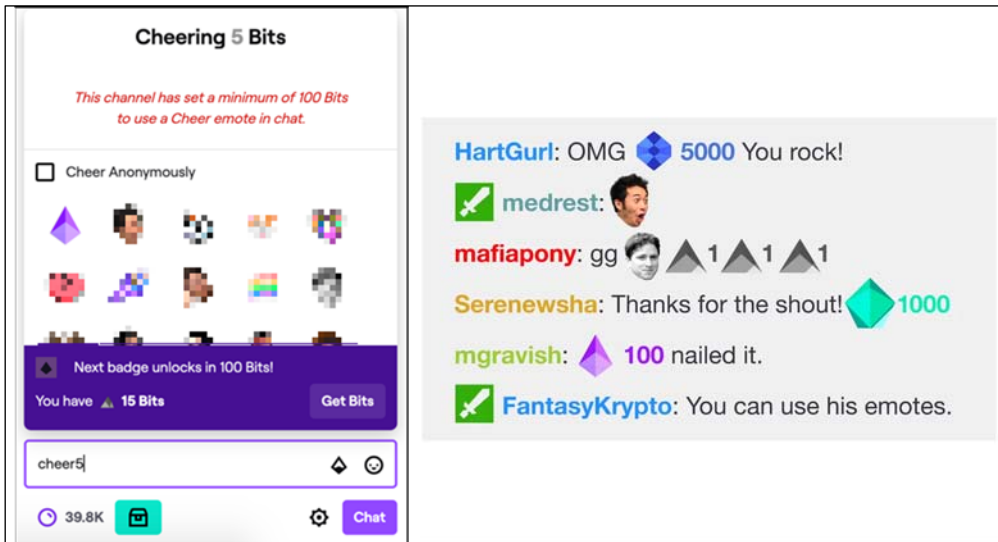
데이터 수집을 위해 게임 전문 인터넷 방송 플랫폼 Twitch를 대상으로 데이터를 수집하였다. Twitch는 2007년 서비스를 시작한 Justin.tv가 2011년 6월 게임 카테고리 방송만을 분할하여 만든 인터넷 방송 플랫폼이다. 2014년 9월 아마존에 인수되었으며, 2015년 2월 이후부터 대한민국에서 서비스되기 시작하였다⁷. Twitch는 게임 방송을 제공하는 플랫폼 중에서 가장 높은 시간당 시청율을 가지고 있는 플랫폼으로, 실시간 개인방송의 누적 시청 점유율만을 따지면 인터넷 방송 점유율 1위인 YouTube 보다도 높다(정미하, 2019). 게임 방송 제작 소프트웨어 지원 및 통계를 제공하는 Stream Elements의 2019년 12월 발표 자료에 따르면 시간당 시청률이 Twitch는 73%, YouTube Gaming이 21%, 그리고 Facebook gaming이 3%로 집계되어(Yosilewitz, Chase, 2019) 게임 방송 시청자들에게 있어서는 Twitch가 더 일반적인 게임 방송 플랫폼으로 받아들여지고 있다고 볼 수 있다. 따라서 게임 방송 시청자의 후원 행위를 연구하는 본 연구의 특성 상 가장 많은 이용자를 가진 Twitch의 게임 방송에서 발생하는 일반 채팅과 후원 채팅을 비교하여 본다면 게임 방송 시청자들의 후원 행위에 대해 이해하는데 도움이 될 것이다.

Twitch는 글로벌 서비스를 제공하고 있어 다양한 국가에서 서비스되고 있는데, 동일한 플랫폼 페이지에서 언어 선택을 통해 각 언어사용자의 게임 방송을 시청할 수 있으므로 접속 지역, 계정의 설정을 고려하지 않고 언어권 별로 데이터 수집이 용이하다. 본 연구에서는 텍스트 분석의 용이함을 고려하여 영어권 채널의 채팅 데이터를 대상으로 하였다.

⁷ Twitch Korea twitter 2015년 5월 2일: <https://twitter.com/TwitchKR/status/561787403673665537>

제 2 절 Twitch 내의 후원 파악

다음으로, 후원이 이루어졌는지 여부를 확인하기 위해서 Twitch의 스트리머 후원 서비스인 비트(Bits)를 살펴보았다. 비트는 Twitch에서 제시하는 조건을 충족(제휴)하였거나 계약관계(파트너)에 있는 채널에서 선물 형식으로 스트리머에게 보낼 수 있는 Twitch 내 재화로서 1 비트당 \$0.01의 가치를 가지고 있으며, 결제 시 수수료 40%가 부과된다(예시: 100 비트 = \$1.40). 스트리머는 받은 비트가 일정 단위 이상 누적되면 계좌로 현금화 가능하다. Twitch비트는 일반 채팅과 동일하게 채팅 데이터에 기록되며 [그림 5]의 좌측 이미지처럼 채팅 내용 작성 시 입력도구 옆의 마름모꼴 버튼을 사용하면 후원 채팅입력 창이 표시된다. 후원하고자 하는 단위를 입력하고 후원자 표시 여부 등을 설정하여 이용 가능하며 이용 시 선택적으로 이모티콘과 문구를 작성하게 된다. 실제 이용 화면에서 후원 채팅으로 작성된 채팅은 [그림 5]의 우측 이미지처럼 일반채팅과 함께 표시되는데, 일반 이모티콘과 차별화되는 애니메이션 또는 장식이 추가되어 있으며, 스트리머의 설정에 따라 해당 채팅에 대한 알림과 텍스트 내용을 읽어주는 효과가 방송 중에 실행된다. 후원 채팅 이모티콘은 [그림 6]과 같이 후원 단위에 따라 다른 이모티콘 효과를 제공하고 있으며, 스트리머가 채널에서 사용하는 이모티콘을 후원 단위에 따라 다르게 사용 가능 하도록 차별화할 수 있다.



[그림 5] 비트 입력 화면과 출력 예시



[그림 6] 비트 단위별 이모티콘

따라서 본 연구에서 수집 대상으로 삼는 채널은 제휴 또는 파트너 자격을 가지고 있어 비트의 사용이 가능한 영어권 채널을 대상으로 하며, 이 채널에서 발생하는 비트 정보를 포함한 채팅 데이터를 연구 대상으로 하였다.

제 3 절 데이터 수집

채팅 데이터를 수집하기 위해 Twitch에서 제공하는 API를 사용하였으며, 1) 채널

정보 수집, 2) 채널에 저장된 영상 정보 수집, 3) 영상에 저장된 채팅 정보 수집으로 3단계에 걸쳐서 이루어졌다.

데이터의 수집은 2020년 4월 3일부터 5월 7일까지 약 한달간 진행되었으며, 영어권 채널 중 시청자 1,000명이 넘는 채널 100개의 ID를 조회하여 채널 정보를 저장하였고, 이중에서 수집을 시작하기 전 3일을 기준으로 위에서 명시한 제휴 또는 파트너 자격을 가지고 있는 채널 50개를 무작위로 선택하였다.

Twitch 채널에서 실시간으로 방송이 종료될 경우, 별도로 저장하지 않기로 설정하지 않았다면 자동으로 방송 영상이 채널에 저장된다. 이러한 영상은 Twitch 계정의 등급에 따라 14일, 60일의 기간을 두고 보관된다. 따라서 수집 대상이 되는 영상은 이러한 각 채널에 저장된 60일 이내의 이전 게임 방송의 데이터이다. Twitch API를 통해 채널당 수집할 수 있는 영상은 최근영상을 기준으로 최대 100개로 제한되기 때문에 각 채널당 100개 이내의 영상 정보를 요청하여 확보하였다.

마지막으로 채널 별로 수집된 영상의 ID를 이용하여 각 영상의 실제 방송 당시 입력된 채팅 정보를 페이지 단위로 수집하였다. 수집된 채팅 정보는 채팅 데이터(채팅 텍스트, 후원 여부와 후원 액수), 데이터 식별을 위한 정보(채널ID, 비디오ID, 영상에 기록된 게임 태그, 방송시간(길이), 채팅 발생 구간, 작성자, 채팅ID, 채팅이 저장된 페이지, 채팅 입력 시간)로 구성되어 있다.

이렇게 수집된 데이터를 테이블 형태로 요약하면 [그림 7]과 같다.

In [122]: 1 pre_text[0].head()

Out [122]:

	channel	video	game_category	duration	time_stamp	comment_user	chat_id	comments	page	donation	spent	chat_time
0	2261919	97381002	Battlefield 1	400	23.342	swapt1_	81843c0c-a7d2-4f1c-8252-e2d11be9aef4	Lmfao	0	0	0	2016-10-16 12:20:27
1	2261919	97381002	Battlefield 1	400	27.848	hornsfan80	95d8dc7c-f85b-4fe3-95dd-4ee9763a4e2d	Imao Adam West really let himself go Keepo	0	0	0	2016-10-16 12:20:32
2	2261919	97381002	Battlefield 1	400	31.848	gex023	f898948e-80d4-4749-9fc1-460f6acfa99a	Lol	0	0	0	2016-10-16 12:20:36
3	2261919	97381002	Battlefield 1	400	32.106	thankphil	f8647964-7dba-4858-8cc7-daf06c4699f3	The hero we deserve	0	0	0	2016-10-16 12:20:36
4	2261919	97381002	Battlefield 1	400	35.179	jesse	605a52c8-d60b-4ee8-9f09-6c74bd857158	LUL	0	0	0	2016-10-16 12:20:39

[그림 7] 수집된 채팅 데이터

총 50개 채널에서 영상 2,848개에 대한 채팅 데이터를 수집하였으며, [표 1]에 내용을 요약하였다. 수집된 전체 채팅의 수는 28,655,636건이며, 연구에는 결측 데이터 494건이 제거된 28,655,142건의 채팅을 사용하였다.

[표 1] 결측 제거 후 수집 데이터

플레이 된 게임	172
전체 영상의 수	2,848
전체 영상 중 후원 발생 영상의 수	1,626
전체 채팅 수	28,655,142
후원 발생 채팅 수	28,890

수집된 데이터의 특성을 살펴보면, 채널 하나를 기준으로 영상이 가장 많은 경우는 100개, 적은 경우는 6개가 수집되었으며, 수집된 영상의 채널 별 평균은 56.96개, 중간 값은 56.5개였다. 수집 기준이 된 50개 채널 중 1개의 채널에서는 한 건의 후원도 발생하지 않았다. 나머지 49개의 채널을 살펴보면, 채널당 최소 한 건의 후원이라도 발생한 영상 수의 경우 많을 경우 62개, 적을 경우 1개의 영상에서 후원이 발생하였고 채널 당 평균 32.52개(중간 값은 33.5개)의 영상에서 후원이 발생하였다. 영상

한 건을 기준으로 발생한 후원을 집계하였더니 후원 채팅이 발생한 영상의 경우 최대 496건의 후원이 발생하였으며, 246개의 영상에서는 1건의 후원만이 발생하였다. 영상 당 평균 17.77건(중간 값은 6건)의 후원이 발생한다고 할 수 있다. 발생한 28,890건의 후원의 비트 총 합은 7,548,843 비트(\$ 75,488.43)로, 채널 기준으로는 하나의 채널에서 최대 1,197,003 비트(\$ 11,970.03), 최소 3비트(\$ 0.03), 평균 150,976.86 비트(\$1,509.77), 중간 값 70262.5(\$ 702.63~702.64)가 발생하였다. 영상을 기준으로 발생한 후원의 합을 비교한 결과 하나의 영상에서 최대 400,004비트(\$ 4,000.04), 최소 3비트(\$ 0.03), 평균 4,642.58 비트(\$ 46.43), 중간 값은 15,000 비트(\$ 150.00)가 발생한 것으로 나타났다. 후원 1건에서 발생한 비트의 최대값은 100,000(\$ 1,000.00), 중간 값은 100 비트(\$ 1.00)로 확인되었다. [표 2]에서 이 내용을 요약하였다.

[표 2] 수집 데이터 내 후원 정보

	최대	평균	중간	최소
채널당 후원 발생 영상 수	62	32.52	33.5	1
영상내 후원 발생 수	496	17.77	6.0	1
채널 당 발생한 비트 합계	1,197,003	150,976.86	70,262.5	3
영상 당 발생한 비트 합계	400,004	4,642.58	15,000.0	3
1회 후원 비트	100,000	-	100.0	1

제 4 절 분석 방법: Latent Dirichlet Allocation (LDA)을

이용한 토픽 모델링

본 연구에서는 실제 실시간 게임 방송의 시청자들에 의해 발생하는 채팅과 후원 텍스트를 이해하기 위해 LDA를 분석 방법으로 사용한다. LDA는 연구자가 설정한 토픽의 수에 맞추어 문서의 토픽 분포와 각 토픽 내의 단어 분포를 추정하고, 추정한

확률분포에 기반하여 토픽별로 단어를 분류하는 모델로, 각 문서에 존재하는 토픽을 파악하기 위해 사용하는 토픽모델링 기법의 하나이다. LDA모델은 연구자가 입력한 각 문서의 토픽 분포와 각 토픽 내의 단어 분포를 추정한다. LDA의 수행 과정은 다음과 같다⁸.

- 1) 연구자가 알고리즘에 지정한 토픽의 개수(k)를 지정하면 k 개의 토픽이 전체 문서(M)에 걸쳐 분포되어 있다고 가정한다.
- 2) 모든 단어를 임의의 토픽 하나(t)에 할당한다.
- 3) (이하 반복)단어 w 는 잘못된 토픽에 할당되어 있고, 다른 단어는 전부 올바른 토픽에 할당되어 있다고 가정. 이하를 비교하여 단어 w 의 토픽을 재할당한다.
 - 3-1) 문서 d 의 단어들 중 토픽 t 에 해당하는 단어들의 비율
 - 3-2) 단어 w 를 갖고 있는 모든 문서들 중 토픽 t 가 할당된 비율

반복이 종료되면 모든 단어들의 토픽 할당이 완료된 수렴상태가 된다. 이해를 돕기 위해 [표 3]에 간단한 수행 예시를 작성하였다. 이와 같이 LDA모델은 문서의 주제를 판별해주는 것이 아니라 확률분포에 기반하여 단어들을 토픽별로 분류하는 것이기 때문에 분류된 결과로부터 연구자가 어떤 내용에 대한 토픽인지 직접 보고 판단해야 한다. 본 연구에서와 같이 자유롭게 작성된 대량의 채팅에서 특정 주제를 눈으로 식별하기 어려운 경우 LDA를 통하여 공통된 주제들이 어떤 것이 있는지 살펴보는 데 용이하게 쓰일 수 있다.

⁸ LDA 개요(딥 러닝을 이용한 자연어 처리 입문): <https://wikidocs.net/30708>

[표 3] LDA 수행 예시

예시	문서1: 저는 사과랑 바나나를 먹어요 문서2: 우리는 귀여운 강아지가 좋아요 문서3: 저의 깜찍하고 귀여운 강아지가 바나나를 먹어요
각 문서의 토픽 분포	문서1: 토픽 A 100% 문서2: 토픽 B 100% 문서3: 토픽 B 60%, 토픽 A 40%
각 토픽의 단어 분포	토픽A: 사과 20%, 바나나 40%, 먹어요 40%, 귀여운 0%, 깜찍하고 0%, 좋아요 0% 토픽B: 사과 0%, 바나나 0%, 먹어요 0%, 귀여운 33%, 강아지 33%, 깜찍하고 16%, 좋아요 16%

제 5 절 데이터 전처리

LDA는 데이터 전처리 과정에서 작성된 단어집합을 사용하여 단어사전과 정수 인코딩이 적용된 말뭉치(corpus)를 만들어 모델 훈련에 사용한다. 말뭉치에 정수 인코딩을 적용하는 이유는 컴퓨터가 텍스트보다 숫자를 더 잘 처리할 수 있기 때문이다. 말뭉치에는 각 문서에 사용된 단어와 사용된 빈도가 기록된다. 이 말뭉치에 정수 인코딩을 적용한 경우, [표 4]과 같이 각 단어에 고유한 정수를 할당하여 훈련 모델이 데이터를 학습하기 용이한 형태, 즉 문자에서 숫자로 형태를 변화시킨다.

[표 4] 정수 인코딩 적용 예시

원문 텍스트	'vaughns picture reminds me of boogie2988s tin foil helmet picture Kappa'
전처리 결과	['vaughns', 'picture', 'reminds', 'boogie', 'tin', 'foil', 'helmet', 'picture', 'kappa']
말뭉치 (인코딩 적용 전)	[(kappa, 1), (vaughns, 1), (boogie2988s, 1), (foil, 1), (helmet, 1), (picture, 2), (reminds, 1), (tin, 1)]
말뭉치 (인코딩 적용 후)	[(12, 1.0), (44, 1.0), (92, 1.0), (93, 1.0), (94, 1.0), (95, 2.0), (96, 1.0), (97, 1.0)]

LDA를 이용한 토픽 모델링을 진행하기 위해서 먼저 수집된 채팅의 텍스트를 정제하는 과정을 진행하였다. 텍스트 전처리를 선행하는 이유는 말뭉치(corpus) 내에서 동일한 의미를 가졌으나 다른 용법으로 사용된 단어나 표현을 통일하여 중복된 의미를 갖는 단어의 수를 줄이고, 불필요한 텍스트를 제거하는 작업이 필요하기 때문이다. 이 작업을 통해 말뭉치 내의 단어 구조가 단순화되고, 훈련을 위한 단어 사전을 만드는 과정에서 텍스트 분석 모델의 성능을 저하시키는 노이즈(불용어, 특수문자, 숫자와 같이 해석에 큰 도움이 되지 않는 단어)를 줄일 수 있다.

이 과정에서 먼저 채팅 규정에 어긋나 삭제되었거나 요청단계에서 불러오지 못한 채팅 데이터를 제거하였다. 즉, 채팅 데이터의 텍스트가 N/A로 표기되어 있어 채팅 이력은 있으나 그 내용이 없는 것으로 확인되는 채팅을 제거했다.

다음으로, 모든 텍스트를 소문자로 변환한 후, 각 단어에 대한 표제어를 추출(Lemmatization)하였다. 표제어 추출은 단어의 형태를 통일시켜서 corpus 내의 단어 수를 줄이고 복잡성을 낮추기 위해 사용되는 정규화 기법 중 하나이다. 동일한 목적의 기법인 어간 추출(Stemming)이 알고리즘에서 정해진 규칙에 따라 단어의 어미를 자르는 것과 비교하여 표제어 추출은 기본 사전형 단어로 바꾸는 작업으로 [표 5]에서 각각이 적용된 결과의 차이를 볼 수 있다. 추출 속도가 어간 추출에 비해 느리나 정제되지 않은 언어가 많이 사용되는 채팅 텍스트에서 표제어 추출이 어간 추출보다 단어의 형태를 적절히 보존하는 양상을 보여주기 때문에, 토픽 모델의 해석에 편의성을 가지고 있어 본 연구에서는 표제어 추출을 사용하였다.

[표 5] 어간 추출, 표제어 추출 예시

예시	Stemming	Lemmatization
am	Am	be
the going	the go	the going
having	Hav	have

단어의 형태를 정리한 이후 I, we, be 등 문장에서 자주 사용되나 해석에 큰 영향이 없는 단어들이 불용어를 제거하였다. 이에 더해 ‘game’ 과 ‘cheer’ 또한 불용어로 보고 이들 또한 제거하였다. 왜냐하면 게임을 주제로 하는 텍스트에서 ‘game’ 이라는 단어는 특별한 의미를 갖지 않아 불용어로 볼 수 있기 때문이다. 또한 ‘cheer’ 는 후원 채팅을 입력할 때 자동으로 입력되는 이모티콘의 텍스트 이름으로서 텍스트의 내용과 거리가 있어 불용어로 처리하였다. 불용어 제거 후 해석이 어려운 2글자 이하의 단어를 추가로 삭제하였다. ly, im, gm과 같은 단어가 표제어 추출과 불용어 제거 등 텍스트를 가공하는 과정에서 발생하는데, 이를 해석하는 것이 문맥, 게임에 따라 게임 속에 등장한 사물의 명칭, 운영자에 대한 지칭, 브랜드의 약칭 등에 따라 달라지기 때문에 정확히 해석하기 어려워 2글자 이하 단어를 해석이 어려운 단어로 정리하여 제거하였다.

다음으로, 품사 태깅(Part-of-speech tagging)을 수행하였다. 동일한 단어라고 하더라도 문장의 위치에 따라 의미가 변하는 것을 파악하기 위한 작업을 품사 태깅이라고 한다. 토큰화 한 문장의 구조에 맞춰 각 단어의 품사를 표기하는 기능을 활용하여 토픽 모델링에 소요되는 시간을 단축할 수 있다. 본 연구에서는 토픽 해석에 용이하도록 각 채널 채팅에서 공통적으로 사용 빈도가 높게 나타나는 품사에 해당하는 단어를 골라 LDA 토픽 모델링에 사용할 단어집합을 작성하였다. 연구에 사용한 품사는 명사형 태그에는 NN(명사 단수형), NNS(명사 복수형), NNP(고유명사 단수형), 동사형 태그에는 VBD(동사 과거시제), VB(동사 기본형), VBP(비 3인칭 단수

현재형), VBN(동사 과거분사), VBG(동명사 또는 현재분사), VBZ(3인칭 단수 현재형), 형용사형 태그는 JJ(비교, 최상 표현 제외한 형용사), 부사형 태그는 RB(비교, 최상 표현 제외한 부사)를 사용하였다. 사용된 태그에 대해서 [표 6]에 정리하였다.

[표 6] corpus 작성에 사용된 품사 태그

품사	사용된 품사 태그
Noun	'NN', 'NNS', 'NNP'
Verb	'VBD', 'VB', 'VBP', 'VBN', 'VBG', 'VBZ'
Adjective	'JJ'
Adverb	'RB'

제 5 장 분석 결과

수집한 채팅 데이터에는 다수의 이모티콘 정보가 이모티콘의 이름으로 포함되어 있다. Twitch 에서 영어권 채팅은 사전에 시청자간 암묵적으로 형성된 채널의 룰에 따라 이모티콘을 사용하여 영상에 반응하는 경우가 빈번하다고 알려져 있다(Ford *et al.*, 2017). 예를 들어 [표 6]에서 첫번째 예로 들은 이모티콘은 과장된 분위기를 나타내는 상황에서 사용하는 이모티콘으로 그 이름이 “pogchamp” 이다. 따라서 채팅 데이터에서 이러한 이모티콘을 사용하였다면 “pogchamp” 라고 기록되게 된다. 이렇게 Twitch에서 자주 사용되는 이모티콘⁹에 대한 유래와 해석, 자주 사용되는 상황 등에 대한 설명은 KnowYourMeme, Urban Dictionary와 같은 사이트에서 찾아볼 수 있다.

[표 7]의 내용은 이러한 이모티콘을 정리한 예시이다. 예를 들어 흥분되는 상황 또는 과장된 분위기를 나타내는 상황에서 pogchamp를 사용할 수 있고, Lol(laugh out loud의 축약어)대신 LUL이 사용될 수 있으며, LUL의 이미지를 왜곡해서 더 큰 웃음을 표현하는 omegalul이 있다. 또한 시청자는 알고 있으나 스트리머는 알지 못하는 상황에서 사용하는 pepe laugh 등, 웃음과 즐거움을 표현하는 이모티콘이라도 방송의 상황과 시청자들이 느끼는 정도에 따라 다른 이모티콘이 사용됨을 확인하였다.

⁹ 이모티콘 해설 예시(Pogchamp): <https://knowyourmeme.com/memes/pogchamp>

[표 7] Twitch 이모티콘과 의미

이모티콘	이름	의미
	pogchamp	과장된 반응, 흥분
	LUL	웃음 (LOL과 같은 의미)
	omega lul	LUL의 변형 더 큰 웃음
	pepe laugh	크게 웃음 (미인지 상황에 대한)
	kebunS	긴장, 당황
	kebunLuv	호감 표시
	kebun10	플레이 또는 행동에 대한 긍정 평가
	zbtvEast	방향 표시

KnowYourMeme, Urban Dictionary 등에서 검색되지 않는 이모티콘의 경우에는 이 이모티콘이 사용된 영상에서 해당 이모티콘이 사용된 상황들을 직접 관찰하여 그 의미를 해석하였다.

이모티콘 외에도 일반적인 영단어가 아닌 단어들이 토픽 모델링 결과에 등장하는데 예를 들어 Chang과 같은 단어는 Grand Theft Auto V라는 게임을

스트리밍하는 스트리머들 사이에서 사용하는 단어로써 결성된 중국계 크루 또는 게임 내에서 등장하는 중국계 마피아 캐릭터를 지칭할 때 사용된다. 이름 또는 이미지에 Chang이 포함된 이모티콘도 그러한 내용으로 사용하는 것을 확인하였다.

‘lord_kebun’ 나 ‘ramee’ 는 스트리머의 이름이다. 본 연구에서 사용한 채팅 데이터로 토픽 모델링 분석 시 주제를 대표하는 단어들로 빈번하게 추출되는 단어들을 분류하여 정리하면 [표 8]과 같다.

[표 8] 토픽 내 단어 해석

카테고리	단어
이모티콘(감정표현)	lul, kebun10, kebunhype, kebunlul, kebunbbw, kebuncreep, kebunluv, kekw, sadkek, kappa, peepoclap, rameehuh
감정 표현	lol, lmao, pog, wtf, shit, good, fuck, nice...
게임 관련	chang, randy, chawa, wayne, zbtveast, kebuncg, kebungas, kebungw, kebunwu, kebunqwk, shotzcg, ratedecg ...
스트리머	lord_kebun, ramee, kebunlk(이모티콘)
일반단어	chat, clip, subscribed, month, channel, stream...

토픽 모델링을 실시할 때는 몇 개의 토픽으로 분류할 것인지를 정해주게 된다. 이 때 토픽의 수를 어떻게 정할지는 연구자가 살펴보고자 하는 내용에 따라 다를 수 있다. 긍정/부정 채팅에 대한 연구(Poyane, 2019)에서는 180으로 설정하여 분류된 토픽을 유사한 주제끼리 묶어 긍정적 표현이 사용된 채팅과 부정적 표현이 사용된 채팅으로 구분하였고, 시청 패턴과 채팅의 상관관계를 살펴본 연구(Zhang *et al.*, 2017)에서는 2부터 20개의 토픽 수를 사용하여 세가지 토픽 모델(EMTM, CTM, LDA)의 가능도 추정치(Held-out log likelihood)를 살펴보았다. 즉 연구의 초점에 맞춰 적절한 토픽 수를 설정해야 한다. 본 연구는 채팅 데이터의 주제를 확인하고 데이터 내에서 발생하는 후원이 각각의 주제마다 얼마나 발생했는지 살펴보는 것을 통해

후원 행위를 이해하고자 한다. 따라서 수집된 채팅 데이터의 주제를 파악하기 위한 토픽 수를 찾는 탐색적 과정으로 토픽 수를 2부터 시작하여 1씩 증가시키며 분류되는지 주제들을 살펴보았다.

토픽 수를 2부터 5까지 분류한 결과는 [표 9]와 같으며, 설정한 토픽 수에 따라 훈련된 LDA모델에서 각각의 토픽을 결정하는데 기여율이 높은 순으로 상위 10개의 단어를 카테고리 별로 정리한 것이다.

[표 9] 각 토픽 수 설정 별 결과

설정 토픽 수	토픽 번호	해당 토픽에 해당하는 단어
토픽수=2	토픽 1	<ul style="list-style-type: none"> • 감정표현 이모티콘: lul, • 감정표현: lol, lmao • 게임관련: chang, • 스트리머: lord_kebun, ramee • 일반단어: car, get, just, chat
	토픽 2	<ul style="list-style-type: none"> • 감정표현 이모티콘: kekw, kebunluv, kebunhype, kebunlul, sadkek • 감정표현: pog • 게임관련: kebuncg, kebungw, shotzcg • 일반단어: http
토픽수=3	토픽 1	<ul style="list-style-type: none"> • 감정표현 이모티콘: lul, kebun10, rameehuh, • 게임관련: kebuncg, shotzcg, zbtveast, ratedecg, kebungas • 스트리머 이모티콘: kebunlk • 일반단어: taco,
	토픽 2	<ul style="list-style-type: none"> • 감정표현 이모티콘: kebunluv, kebunhype, sadkek, kebunlul, • 감정표현: pog, lmao, • 게임관련: kebungw, • 일반단어: http, chat, big
	토픽 3	<ul style="list-style-type: none"> • 감정표현 이모티콘: kekw, • 감정표현: lol, shit • 게임관련: chang, • 스트리머: lord_kebun, ramee, • 일반단어: car, get, just, got,

토픽수=4	토픽 1	<ul style="list-style-type: none"> • 감정표현 이모티콘: lul, kebunluv, kebunhype, kebunbbw, kebuncreep, • 게임관련: kebungw, chang, • 스트리머: kebunlk, • 일반단어: got, time
	토픽 2	<ul style="list-style-type: none"> • 감정표현 이모티콘: kebunlul, kebun10, • 감정표현: lmao, • 게임관련: kebuncg, shotzcg, zbtveast, ratedecg, kebungas • 일반단어: http, subscribed
	토픽 3	<ul style="list-style-type: none"> • 감정표현 이모티콘: sadkek, lulw, • 감정표현: lol, pog, wtf, yes, • 스트리머: lord_kebun, • 일반단어: car, cop, guy
	토픽 4	<ul style="list-style-type: none"> • 감정표현 이모티콘: kekw, • 감정표현: shit, • 게임관련: randy • 스트리머: ramee, • 일반단어: get, just, chat, big, taco, n't,
토픽수=5	토픽 1	<ul style="list-style-type: none"> • 감정표현 이모티콘: kebunlul, kebunbbw, lulw, kebuncreep, • 감정표현: lmao, • 게임관련: chang, randy, wayne, • 스트리머(이모티콘): kebunlk, • 일반단어: gang
	토픽 2	<ul style="list-style-type: none"> • 감정표현 이모티콘: kebunluv, kebun10, ratedelul • 감정표현: pog, nice, • 게임관련 이모티콘: kebungw, kebungas, kebunsbso, • 일반단어: http, big,
	토픽 3	<ul style="list-style-type: none"> • 감정표현 이모티콘: kappa, • 감정표현: lol, • 게임관련: kebuncg, shotzcg, ratedecg, kebunwu, kebungwk, • 스트리머: lord_kebun, • 일반단어: subscribed, month
	토픽 4	<ul style="list-style-type: none"> • 감정표현 이모티콘: kekw, • 감정표현: shit, • 게임관련: zbtveast, • 스트리머: ramee, • 일반단어: get, just, chat, got, time, n't
	토픽 5	<ul style="list-style-type: none"> • 감정표현 이모티콘: lul, kebunhype, sadkek, • 감정표현: wtf, • 스트리머: lord_kebun, • 일반단어: car, taco, cop, look, truck

토픽 수를 2로 설정했을 경우, 토픽 1과 2의 차이를 살펴보면 토픽 2에서는 스트리머 관련 단어들이 순위에 나타나지 않는데 토픽 1에서는 이 단어들이 나타나는 것을 볼 수 있었다. 이 외에는 두 개의 토픽에서 큰 차이는 나타나지 않았고, 감정표현과 게임 관련 단어들이 두 토픽 모두에 나타나 두 개의 토픽을 별개의 토픽으로 뚜렷이 구별하기 어려웠다.

토픽 수를 3으로 설정한 경우부터 분류 내용에 대해 짐작할 수 있는 결과가 나타나기 시작하였다. 토픽 1에서 게임과 관련한 단어(kebuncg, shotzcg, zbtveast, ratedecg, kebungas)와 스트리머의 이름(kebunlk), 일반 단어(taco)를 확인하였고, 토픽 2는 일반 단어(http, chat, big)와 게임 관련 단어(kebungw)이외에 감정 표현 단어(kebunluv, kebunhype, sadkek, kebunlul, pog, lmao)가 주로 나타났다. 토픽 3에서는 게임 관련 단어(chang)와 스트리머의 이름(lord_kebun, ramee), 일반 단어(car, get, just, got)를 확인하였다. 각 토픽의 단어들을 살펴보면 토픽 1은 스트리머의 이름을 언급하는 형태가 텍스트가 아닌 이모티콘으로 이루어지고 있고, 게임 관련 단어가 다른 두 토픽과 비교해 가장 많이 나타났다. 그래서 스트리머가 플레이중인 게임의 내용을 언급하는 주제로 보이고, 토픽 2에서는 감정 표현이 주로 분류되어서 시청 중에 느낀 감정 표현 단어들이 모인 것으로 보인다. 토픽 3은 스트리머를 직접 언급하고 있고 일반 단어에서 car와 get, got이 언급되는 내용에서 스트리머의 행동을 언급하는 주제의 텍스트에서 많이 사용된 단어들이 분류되어 있는 것으로 보인다.

토픽 수를 4로 설정하여 분류된 내용을 살펴보면 다음과 같다. 토픽 1은 감정 표현 단어(lul, kebunluv, kebunhype, kebunbbw, kebuncreep)와 게임 관련 단어(kebungw, chang), 스트리머의 이름(kebunlk), 일반 단어(got, time, look)가 담겨있었다. 토픽 2는 감정표현(kebunlul, kebun10, lmao)과 게임관련 단어(kebuncg, shotzcg, zbtveast, ratedecg, kebungas), 일반 단어(http, subscribed)를 볼 수 있었고, 토픽 3은 감정표

현(sadkek, lulw, lol, pog, wtf, yes)과, 스트리머의 이름(lord_kebun), 일반 단어(car, cop, guy), 마지막 토픽 4는 감정표현(kek, shit)과, 게임 관련 단어(randy), 스트리머(ramee), 그리고 일반 단어(get, just, chat, big, taco, n't)를 확인하였다. 네 가지로 분류된 토픽이 내용을 비교하여 보았을 때, 토픽 1은 스트리머가 플레이 중인 게임에 대한 시청자의 반응을 연상하는 단어들이 담긴 내용이고, 토픽 2도 토픽 1과 동일한 내용으로 보이나 감정표현보다 게임내 등장하는 요소에 대한 반응에 집중한 단어들이 모여 있는 것으로 보인다. 토픽 3은 게임 관련 단어가 등장하지 않고, 감정표현과 스트리머의 이름이 등장하고 있어 스트리머의 활동에 대한 내용으로 보이고, 토픽 4는 토픽 1과 동일하게 스트리머를 언급하고 있는데, 둘의 차이를 비교하기 위해서 수집된 채널의 정보를 확인하였고, 토픽 1의 단어는 수집대상이 되었던 영상의 게임을 플레이하고 있는 스트리머였고, 토픽 4의 단어는 수집대상이 아니라 플레이 중에 시청자에 의해 언급된 스트리머(ramee)였다. 분류 상위에서 이러한 단어가 노출된 것은 해당 스트리머가 특정 스트리머와 함께 게임을 하거나 경쟁 관계에 있는 등 시청자들이 시청과정에서 이를 언급할 상황이 많았을 것으로 보인다. 이러한 차이와 다른 토픽 대비 감정 표현 이모티콘이 제일 적게 드러나는 것, 일반단어와 같이 직접 입력해야 하는 단어가 다른 토픽 대비 많이 등장하는 점을 바탕으로 토픽 4는 영상 시청 과정에서 발생한 시청자들의 영상과 때로 무관한 자유로운 대화에 사용된 단어들이 모여 있는 것으로 보인다.

마지막 토픽 수 5의 내용은 토픽 1에서 감정표현, 게임관련, 스트리머의 이름(kebun1k), 일반 단어(gang)를 토픽 2에서는 감정표현, 게임관련, 일반 단어(http, big), 토픽 3에선 감정표현(kappa, lol), 게임관련, 스트리머의 이름(lord_kebun), 일반 단어(subscribed, month)를 확인하여 세 토픽은 게임 플레이 관련 주제로 나누어진 것을 확인하였고, 토픽 4에서는 감정표현(kek, shit), 게임관련(zbtveast), 스트리머

의 이름(ramee), 일반 단어를 확인하였으며, 마지막 토픽 5는 감정표현(lul, kebunhype, sadkek, wtf), 스트리머의 이름(lord_kebun), 일반적으로 사용되는 단어(car, taco, cop, look, truck)를 통해 각각 영상과 무관한 시청자 의사표시와 스트리머 관련 주제로 이해하였다.

분류된 Topic의 내용을 살펴본 결과 토픽 수 3에서 5로 분류를 늘려갈 때마다 단어들이 특정 주제를 연상할 수 있도록 나뉘는 것을 확인할 수 있었다. 그러나 확인한 단어들의 종류인 감정표현과 게임관련 단어, 스트리머의 이름, 일반단어의 조합을 기반으로 토픽을 해석하는 과정에서 토픽 수가 5 부터 유사한 주제로 보이는 토픽의 차이를 구분하기 어려워져 세부기준(채널, 게임)에 대한 이해가 필요한 것으로 보였다. 따라서 여러 주제가 지나치게 유사한 주제로 분화되기 직전인 토픽 수 4가 각 토픽의 내용을 살펴보고 해석하기에 적절하다 판단하였고, 토픽의 수를 4로 지정 후 연구를 진행하였다. 훈련 결과 분류된 토픽 별 단어의 내용은 [표 10]과 같다.

[표 10] 토픽 별 단어 구성

토픽 번호	토픽 주제	토픽 해석
토픽 1	감정표현 이모티콘: lul(1), kebunluv(2), kebunhype(3), kebunbbw(8), kebuncreeper(10) 게임관련: kebungw(4), chang(5) 스트리머: kebunlk(9) 일반단어: got(6), time(7)	(스트리머의) 게임 플레이
토픽 2	감정표현 이모티콘: kebunlul(3), kebun10(7) 감정표현: lmao(4) 게임관련: kebuncg(1), shotzcg(5), zbtveast(6), ratedecg(8), kebungas(10) 일반단어: http(2), subscribed(9)	게임 요소
토픽 3	감정표현 이모티콘: sadkek(5), lulw(7) 감정표현: lol(1), pog(3), wtf(8), yes(9) 스트리머: lord_kebun(2) 일반단어: car(4), cop(6), guy(10)	스트리머
토픽 4	감정표현 이모티콘: kek(1) 감정표현: shit(6) 게임관련: randy(10) 스트리머: ramee(4) 일반단어: get(2), just(3), chat(5), big(7), taco(8), n't(9)	영상과 무관한 대화

토픽의 해석에서 토픽 2를 구성하는 단어에 채널 관련 단어(subscribed, month, stream, channel)가 포함되어 있는 것을 확인하였는데, 이 토픽에서 확인한 채널 관련 단어는 시청자의 방송 시청 중 정기 구독 결정 등으로 표시되는 공지 문자로 사용자가 아니라 시스템이 작성하는 채팅이다. 해당 텍스트의 발생에 관하여 염두에 두지 못해서 사전에 제거하지 못하였는데, LDA 모델이 토픽 2에 분류된 다른 단어와 구독 메시지의 단어를 유사한 주제로 파악했기 때문에 이 토픽에 함께 분류된 것으로 보인다. 해당 단어들이 스트리머가 언급하는 단어가 아니기 때문에 게임 요소 반응의 해석에서 제외하였다.

각 토픽의 해석은 적절한 토픽 수를 찾는 과정에서 미리 확인해둔 각 이모티콘과 단어의 의미를 기준으로 해석한 내용과 토픽내에서 나타나는 단어의 기여도를 통해

판단하였다. 토픽 1에서 감정표현과 게임관련 단어에 비해 스트리머의 이름은 9번째 기여도를 가지고 있어 스트리머와 게임이 아니라 스트리머의 게임 플레이로 해석하였다. 토픽 2는 감정표현과 게임관련 단어로 이루어졌고, 게임관련 단어는 시청자들이 영상의 게임에서 발생한 이벤트나 요소에 대해 반응하는 단어들이므로 게임 요소 반응으로 해석하였다. 토픽 3은 게임관련 단어가 등장하지 않고 스트리머가 2번째로 높은 기여도를 가지고 있어 시청자의 스트리머에 반응하여 작성한 채팅의 단어가 모인 것으로 보여 스트리머 관련이라고 해석하였다. 토픽 4는 앞서 토픽 수 비교에서 확인한 바와 같이 스트리머 이름의 특징과 다른 토픽과 비교하여 감정표현과 일반 단어의 비율을 통해 영상과 무관한 대화로 해석하였다.

분류된 토픽을 바탕으로 게임 방송의 채팅은 일반적으로 사용하는 단어와 감정표현 외에, 게임관련 단어, 스트리머와 같은 영상 시청과 관련된 단어를 사용하여 플레이 되고 있는 게임에 대한 주제와 스트리머에 대한 주제, 시청 중 채널의 룰이 아닌 개인의 생각을 주제로 대화하고 있는 것으로 보인다. 세부적으로 게임에 대한 주제는 스트리머가 플레이하고 있는 게임에 반응하는 주제와 시청 과정에 화면에서 등장하는 요소 혹은 이벤트에 반응하는 주제로 나뉘는 것으로 해석하였다.

다음으로 후원 채팅이 어떤 토픽에 많이 나타나는지 확인하기 위해 훈련된 LDA 모델을 사용하여 각 채팅의 토픽 확률을 구하였고, 최대 확률을 채팅의 토픽으로 간주하여 일반 채팅과 후원 채팅의 각 토픽 비율을 비교해 보았다.

수집된 전체 데이터에서 토픽 1부터 4까지의 비율은 순서대로 26.2 %, 13.29 %, 29.22 %, 31.29 %로 나타났다. 후원을 제외한 일반 채팅은 26.21 %, 13.29 %, 29.23 %, 31.27 %, 후원채팅은 16.15 %, 4.59 %, 20.15 %, 59.11 %로 각 비율이 확인되었다. 비율의 내용은 [표 11]에 정리하였다.

[표 11] 토픽별 채팅 비율

토픽 수	전체 채팅		일반 채팅		후원 채팅	
	채팅 수	비율	채팅 수	비율	채팅 수	비율
토픽 1 (게임 플레이)	7,508,025	26.20 %	7,503,359	26.21 %	4,666	16.15 %
토픽 2 (게임 요소)	3,806,923	13.29 %	3,805,597	13.29 %	1,326	4.59 %
토픽 3 (스트리머)	8,372,709	29.22 %	8,366,887	29.23 %	5,822	20.15 %
토픽 4 (영상과 무관한 대화)	8,967,485	31.29 %	8,950,409	31.27 %	17,076	59.11 %

일반 채팅은 수집된 전체 채팅의 대부분을 차지 하고 있어 각 토픽의 비율이 전체 채팅의 비율과 유사하게 나타나고 있다. 후원 채팅의 각 토픽이 후원 채팅 내에서 차지하는 비율의 순서는 작은 순부터 토픽 2(게임 요소), 토픽 1(게임 플레이), 토픽 3(스트리머), 토픽 4(영상과 무관한 대화)로 일반 채팅과 동일하지만, 이 토픽의 비율이 후원 채팅 주제 비율의 절반 이상인 59.11%를 차지하고 있는 것으로 나타났다.

제 6 장 결과 논의

본 연구를 통해서 Twitch 방송의 채팅 텍스트를 사용하여 1) 토픽들을 도출하고, 2) 도출된 토픽들의 단어로부터 그 토픽의 주제라고 할 수 있는 내용을 해석하여, 3) 일반 채팅과 후원 채팅이 그 토픽 면에서 차이가 있는지 살펴보았다.

수집된 데이터의 비트 후원 정보를 통해 실제 채팅 환경에서 시청자의 후원이 어떻게 발생하고 있는지 살펴본 결과, 후원 채팅의 횟수와 단위가 불규칙하게 발생하고 있지만 영상 데이터의 보존기간이 최대 60일임을 고려하였을 때, 60일 사이에 채널 평균 150,976.86 비트(\$ 1,509.77) 전후, 영상 평균 4,642.58(\$ 46.43)비트 수익이 발생하는 것을 제4장 제3절 데이터 수집에서 수집한 데이터의 정보를 통해 확인하였다. 조성휘,양성병 (2020)의 연구에서 시청만족도가 시청자의 후원의도에 미치는 영향을 살펴보면서 후원이 인터넷 방송의 주요 수익원으로서 주목받고 있음을 소개하였고, 채널의 영상에 대해 시청만족도가 높을수록 후원의도가 높아지는 것을 확인하였다. 이를 참고하여 수집된 데이터의 비트의 불규칙성은 시청만족도에 따라 발생한 것으로 보고 수익 창출 수단으로서 적합하다는 것을 알 수 있었다.

한 영상에서 최대 496회의 후원 채팅과 1회 최대 100,000 비트 후원이 발생하기도 하여 시청자들이 스트리머에게 때로 다수의 혹은 큰 액수의 후원을 할 수도 있음을 파악하였다. 시청만족도가 유료 후원 의도에 정의 영향을 미치므로(조성휘,양성병, 2020), 후원 채팅이 많이 발생한 영상을 분석하여 시청자의 시청만족도가 높게 나타난 영상을 구성하는 요인을 찾는 데 이용할 수 있을 것으로 보인다. 액수와 관련하여 실제 후원 채팅들의 텍스트를 살펴보았더니 6,000 비트(\$ 60) 이상의 후원에서 기본 후원 이모티콘 외에 응원 성격을 띄는 텍스트 또는 이모티콘을 사용하는 것을 확인하였고, 6,000 비트(\$ 60) 미만의 후원에서 액수가 낮아질 수록 응원 채팅 이외에

‘cheerwhal100(100 비트) u could paint his car yellow’ 나 ‘Kappa1500(1500 비트) If you drop soap on the floor, Is the floor clean or Is the soap dirty?’ 와 같이 영상과 무관하게 자유롭게 작성한 대화를 발견할 수 있었다. 이렇게 작성된 일정 액수 이하의 후원 채팅에서 일반 채팅에서나 작성할 법한 텍스트들이 발견되기 때문에 앞으로 Twitch 게임방송에서 채팅과 후원관계를 다룰 때 주목해야할 데이터로 보인다.

제 5 장 분석결과와 토픽별 채팅의 비율의 결과는 각 채팅이 담고있는 토픽들의 확률 중 최대인 토픽을 대표로 비율을 구한 것이다. 이를 통해서 일반 채팅의 토픽 크기가 영상과 무관한 대화, 스트리머, 게임플레이, 게임 요소 순으로 나타났고, 후원 채팅 또한 동일한 순서임을 확인하였다. 이들이 동일한 순서로 나타난 것은 4개의 토픽으로 분류한 단어들이 일반과 후원 채팅에 비슷한 비율로 사용되고 있다는 것을 의미하며, 후원 채팅에 작성된 텍스트가 일반 채팅에 작성된 텍스트와 크게 다르지 않은 것을 알 수 있다. 그러나 후원 채팅의 영상과 무관한 대화의 비율은 59.11%로 일반 채팅의 31.27%와 비교하여 후원채팅의 과반수 이상이 이러한 주제로 대화하고 있는 것으로 나타났다. 조성휘,양성병 (2020)의 연구에서 시청자의 충동구매 성향이 높을수록 유료 후원을 사용할 확률이 높다고 확인한 것을 통해 즉흥적으로 발생하는 채팅의 후원 채팅은 충동구매 성향이 높은 시청자가 이용할 가능성이 높다고 생각할 수 있다. 또한 한성희 (2012)의 연구에서 쾌락소비후 삶의 만족도에 영향을 미치는 요인 중 하나로 과시욕구가 간접적으로 삶의 만족도에 영향을 미치고 있음을 확인하였다. 재미를 목적으로 시청하는 시청자의 후원 채팅에는 스트리머와 다른 시청자가 자신이 비용을 지불하고 작성한 텍스트를 주목해주기를 바라는 과시욕구가 존재할 수 있다. 과시욕구에 기반한 채팅은 타인과 구별되고자 하므로 기존에 형성된 룰에 따르지 않고 개인적인 메시지를 직접 입력하게 된다. 이러한 채팅일수록 주목도 높은 후원 채팅을 사용하게 될 것이고, 그러한 배경에서 후원 채팅의 토픽 비율 중에 영상과

무관한 대화의 비율이 채널의 룰에 따라서 작성되는 다른 토픽보다 높게 나타나는 것으로 보이고, 일반 채팅보다 후원 채팅에서 더 많은 비율을 차지한 것으로 보인다. 이러한 과시욕구에서 작성한 채팅은 영상의 스트리머와 시청자들이 주목하는 것을 통해 커뮤니케이션하고 특정 관계를 형성하게 되므로 홍수봉,김정규 (2017)의 연구에서 플레이어 유형에 따라 정의한 시청자의 유형중 사교형 게임 시청자로 볼 수 있다. 그래서 홍수봉,김정규 (2017)의 연구에서 AfreecaTV 시청자의 후원 경험이 성취형에서 높게 나타나는 것에 비해 Twitch는 사교형이 더 높게 나타나는 것에서 두 플랫폼 시청자의 시청의도에 차이가 있는 것을 확인하였다.

제 7 장 결론

본 연구는 토픽 모델링을 사용하여 인터넷 게임 방송의 채팅 텍스트를 단어별 조합으로 분류하여 채팅의 주제를 해석하고 주제의 비율을 통해 후원과 채팅의 특징에 대해 살펴보는 연구를 진행하였다.

연구를 위해 Twitch의 영어권 스트리밍 채널 중 후원 활동이 가능한 채널 50개를 임의로 선정하여 영상에 기록된 채팅 데이터를 수집하였고, 분석결과 시청자들이 일반 단어 이외에 감정 표현과 게임 관련 단어, 스트리머의 이름을 채팅에서 언급하는 것을 확인했다. 시청자들이 이러한 단어를 사용하여 플레이중인 게임에 대한 주제, 게임 내에 등장한 요소에 대한 주제, 스트리머에 대한 주제, 전체 영상의 흐름과 무관한 내용의 주제로 대화하는 것을 발견하였다. 시청자의 대화 비율은 높은 것부터 영상과 무관한 대화, 스트리머에 대한 주제, 플레이중인 게임에 대한 주제, 게임 내에 등장한 요소에 대한 주제 순으로 나타났으며, 이러한 비율에서 후원 채팅은 전체 후원 채팅의 절반 이상이 영상과 무관한 대화 주제에서 발생하고 있음을 확인하였다. 더하여 후원의 단위가 소액일수록 이러한 의사표시 주제 발언이 많아 짐을 발견하였다.

이 연구는 게임 방송을 시청하는 시청자들을 이해하기 위한 연구에서 시청자들의 공통된 대화 주제와 후원 채팅에 관련한 정보를 제공하고 있다. 시청자는 영상 시청 과정에서 스트리머가 플레이하고 있는 게임에 대한 이야기와 스트리머 자체에 대한 이야기, 다른 시청자와 다른 자신의 생각을 채팅을 통해 표현하고 싶어하는 것을 확인하였다. 그 중에서 때로 영상과 무관한 시청자의 생각을 표현하는 대화가 분석을 통해 제일 많은 비율을 차지하는 것을 확인하였고, 이를 통해 Twitch 시청자는 게임이라는 공통된 주제로 교류하고 싶어하는 동기가 다른 시청 동기보다 많은 것을 발견

하였다. 후원 채팅의 비율에서 이 주제의 비율이 다른 주제보다 높은 것과 실제 후원 채팅의 텍스트를 살펴본 것을 통해 Twitch의 시청자는 소액의 후원 채팅을 사용해 일반 채팅에서 할 법한 대화를 한다는 것을 알 수 있었다. 이러한 연구 결과는 AfreecaTV의 채팅과 후원이 관계가 없다고 밝힌 Yu *et al.* (2018)의 연구결과에 반해 Twitch와 같이 채팅 강조 기능을 가지고 있는 후원 시스템을 사용하는 인터넷 방송은 후원 액수와 본 연구에서 다루지 못한 다른 요인(스트리머 유형, 콘텐츠 구성, 채널에서 다루는 게임이 소수인지, 다양한 게임을 다루고 있는지 등)에 의해 채팅과 후원 관계가 존재할 수 있음을 시사한다.

게임 방송의 스트리머에게 있어서 이 연구에서 채팅의 주제 분류 가능성은 방송 후 시청자의 반응을 확인하는 과정에서 주제별로 분류하여 채팅의 내용을 확인할 수 있다면 모든 채팅을 일일이 읽고 내용을 살펴봐야하는 수고를 절약할 수 있다는 편의를 제공할 수 있다.

플랫폼에 참여하는 스트리머를 늘리기 위해 노력하는 게임 방송 플랫폼에 있어서 방송에서 발생한 채팅의 주제를 파악하여 분류할 수 있다는 사실은 백종현 *et al.* (2020)연구에서 시도한 채팅 빈도를 통해 방송의 하이라이트 추출을 사용한 분석, 요약 플랫폼과 함께 채팅 주제를 사용한 지표 또는 편집 도구를 사용해서 스트리머에게 전략적인 영상제작 환경을 제공할 수 있고, 시청자의 채팅 트렌드를 통해 시청자가 관심을 보일만한 새로운 채널의 방송을 추천할 수 있을 것이다.

본 연구에서 채팅 데이터를 사용하여 토픽 모델링을 수행하여 채팅과 후원을 비교하고 채팅과 후원에 관계가 있을 수 있음을 발견하였으나, 해당 관계에 대해 설명할 수 있는 특성을 전부 고려하여 분석이 진행되지 못하였으므로 가능성을 보여준 것에 그치고 있다.

본 연구 결과를 통하여 시청자들이 작성하는 채팅의 텍스트를 주제에 따라 분류

할 수 있고, 따라서 후원 행위를 분석하는 과정에서도 이러한 채팅을 살펴볼 필요가 있다는 것을 제안한다. 또한 본 연구의 결과를 토대로 향후 연구에서는 주제 분류를 넘어 후원이 발생한 개별 채팅 데이터와 후원 데이터의 직접적 연결을 통해 채팅과 후원 사이의 구체적이고 상세한 관계를 살펴볼 수 있을 것이다.

참고문헌

- Ape, A. (2016, 11.04). 스타BJ들의 이민 속출! 유튜브 아닌 트위치가 웃음을 머금다? *Mobiinside*. Retrieved from <https://www.mobiinside.co.kr/2016/11/04/appape-afreecatv/> Accessed on 2020.11.30
- Ford, C., Gardner, D., Horgan, L. E., Liu, C., Tsaasan, A., Nardi, B., & Rickman, J. (2017). *Chat speed op pogchamp: Practices of coherence in massive twitch chat*. Paper presented at the Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems.
- IQBAL, M. (2020). *Twitch Revenue and Usage Statistics (2020)*. Retrieved from <https://www.businessofapps.com/data/twitch-statistics/> Accessed on 2020.11.30
- Luo, M., Hsu, T. W., Park, J. S., & Hancock, J. T. (2020). Emotional Amplification During Live-Streaming: Evidence from Comments During and After News Events. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW1), 1-19.
- Poyane, R. (2018). *Toxic communication during streams on Twitch. tv. The case of Dota 2*. Paper presented at the Proceedings of the 22nd International Academic Mindtrek Conference.
- Poyane, R. (2019). *Toxic Communication on Twitch. tv. Effect of a Streamer*. Paper presented at the International Conference on Digital Transformation and Global Society.
- Yosilewitz, A., & Chase. (2019). *State of the Stream 2019*. Retrieved from <https://blog.streamelements.com/tagged/state-of-the-stream> Accessed on

2020.11.30

- Yu, E., Jung, C., Kim, H., & Jung, J. (2018). Impact of viewer engagement on gift-giving in live video streaming. *Telematics and Informatics*, 35(5), 1450-1460.
- Zhang, C., Liu, J., Ma, M., Sun, L., & Li, B. (2017). *Seeker: Topic-aware viewing pattern prediction in crowdsourced interactive live streaming*. Paper presented at the Proceedings of the 27th Workshop on Network and Operating Systems Support for Digital Audio and Video.
- 권정미, & 박태영. (2014). 도농지역 공동체의식에 영향을 미치는 요인에 관한 연구: 대구광역시 달서구, 경상북도 의성군을 중심으로. *한국지역사회복지학*, 51, 127-154.
- 김미숙, & 김은정. (2005). 사회복지시설의 민간자원 동원에 영향을 주는 요인 연구: 후원을 중심으로. *한국사회복지학*, 57(2), 5-40.
- 김성주, & 안성아. (2020). 1 인 미디어 게임방송 유형과 시청동기가 후원의도에 미치는 영향. *문화경제연구*, 23, 85-106.
- 김주완. (2018, 5.7). 시청자 6억6500만명... 게임 동영상 중계 '황금알'. *한국경제*, p. A14. Retrieved from <https://www.hankyung.com/it/article/2018050770241>
Accessed on 2020.10.19
- 문주영. (2001). 인터넷 방송. *정보통신산업동향*, 2001(1), 34-50.
- 박세준. (2016, 11.04). 인기 BJ들 굿바이 아프리카TV. *주간동아*. Retrieved from <https://weekly.donga.com/3/all/11/774697/1> Accessed on 2020.11.30
- 백종현, 이규태, 이유혁, 홍성표, & 이영구. (2020). 스트리밍 콘텐츠 분석 및 요약 플랫폼. *한국정보과학회 학술발표논문집*, 1465-1467.

- 심규호. (2009). 인터넷 조사 개선 방안 연구. Retrieved from http://kostat.go.kr/sri/srikor/srikor_pbl/2/index.board?bmode=read&bSeq=&aSeq=369658&pageNo=1&rowNum=10&navCount=10&currPg=&searchInfo=srch&sTarget=title&sTxt=%EC%9D%B8%ED%84%B0%EB%84%B7+%EC%A1%B0%EC%82%AC Accessed on 2020.11.30
- 안영진. (2014). 사회복지법인에 대한 지방자치단체의 감독강화 방안. *지방자치법연구*, 14(1), 179-201.
- 오동일, & 김효용. (2007). 웹 2.0 시대의 한국 애니메이션 산업에 관한 고찰. *한국디지털컨텐츠학회지*, 3(1), 3-9.
- 유재흥. (2016). 개인방송, 그 수익의 근원: 누구를 위해 ‘별풍선’을 던지나. *한국통신학회지 (정보와통신)*, 33(4), 71-78.
- 윤혁진, 박찬술, & 이소중. (2019). 고속 성장 중인 1인 미디어 산업. Retrieved from https://www.kofia.or.kr/brd/m_186/view.do?seq=272&srchFr=&srchTo=&srchWord=&srchTp=&multi_itm_seq=0&itm_seq_1=0&itm_seq_2=0&company_cd=&company_nm=&page=2 Accessed on 2020.11.30
- 이선미. (2015). 마이너에서 주류 콘텐츠로 넘어온 개인 방송 서비스. *KT 경제경영연구* Issue & Trend, 1-10.
- 이지수, 권정은, 이현경, & 김상훈. (2019). 소비자의 게임 영상 시청과 게임 채택에 관한 연구. *소비자학연구*, 30(4), 239-259.
- 정미하. (2019, 10.26). 유튜브가 대세? 생방송 시장은 트위치가 더 대세! *조선일보*. Retrieved from https://www.chosun.com/site/data/html_dir/2019/10/25/2019102501815.html Accessed on 2020.10.14
- 정민, & 류승희. (2019). 콘텐츠 스트리밍 산업의 성장동력화가 시급하다:-ICT와의 결합으로 커지는 콘텐츠시장. *한국경제주평*, 829, 1-16.

- 정재민, 이화진, & 김영주. (2005). 미디어간 경쟁과 대체: 지상파방송, 케이블/위성 방송, 인터넷의 적소분석. *한국방송학보*, 19(4), 523-564.
- 조성휘, & 양성병. (2020). 게임 개인방송에서 시청자의 만족도 및 선물후원의도에 영향을 미치는 요인: DouYu TV 플랫폼을 중심으로. *인터넷전자상거래연구*, 20(3), 49-72.
- 조운호. (2019). 지역개발아동후원에 따른 후원아동과 비후원아동 간 격차 변화 분석. *국제개발협력연구*, 11, 81-99.
- 채성오. (2016, 10.15). '대도서관' 유튜브로...아프리카TV 이탈 가속화 될까. *한스경제*. Retrieved from <http://www.sporbiz.co.kr/news/articleView.html?idxno=48481> Accessed on 2020.11.30
- 최민지, 박정민, & 노기영. (2016). 1인 게임방송 시청에 영향을 미치는 요인에 관한 연구. *한국게임학회 논문지*, 16(6), 39-48.
- 최승혁, 박상준, 한재현, & 이현아. (2019). 게임 시청자 실시간 한글 채팅의 언어 자질에 기반한 하이라이트 탐지. *한국디지털콘텐츠학회 논문지*, 20(2), 335-342.
- 최진응. (2016). 인터넷 개인방송의 문제점과 개선방안. *이슈와 논점*(1187), 1-4.
- 한성희. (2012). 소비자 욕구 충족과 쾌락소비 후 감정형성이 삶의 만족도에 미치는 영향. *소비문화연구*, 15(1), 45-72.
- 한혜원, & 김서연. (2014). 온라인 개인 게임 방송의 스토리텔링 분석. *한국게임학회 논문지*, 14(2), 85-96.
- 홍수봉, & 김정규. (2017). 인터넷 개인 게임방송 시청자의 유형과 기부 행위 고찰: 리차드 바틀 (Richard Bartle)의 게임유형을 중심으로. *영상문화콘텐츠연구*, 13, 101-113.