

텍스트 마이닝 기법을 이용한 유튜브 추천 알고리즘의 필터버블 현상 분석

An Analysis of Filter Bubble Phenomenon on YouTube Recommendation Algorithm Using Text Mining

신유진*, 이상우**
LGU+, 연세대 정보대학원**

Yoo Jin Shin(syj3344@naver.com)*, Sang Woo Lee(leesw726@yonsei.ac.kr)**

요약

이 연구에서는 필터버블 현상의 주요 요인인 추천 알고리즘의 정치적 편향성(추천 알고리즘이 이용자가 선호하는 정치 성향의 영상을 제한적으로 제공하는 것)과, 이용자들의 선택적 노출(이용자가 자신이 선호하는 정치 성향의 영상을 자발적으로 선택하는 것)을 실증적으로 검증하고자 하였다. 이를 위해 새로운 유튜브 계정 2개를 개설하여 각각의 계정을 보수/진보 계정으로 일주일 동안 훈련시켰고, 각 계정에서 추천받은 영상들은 이를 간격으로 수집하였다. 텍스트 마이닝(Text Mining) 방법을 통해 보수 계정에서는 보수 성향의 영상이 더욱 추천되는지, 진보 계정에서는 진보 성향의 영상이 더욱 추천되는지를 알아보았다. 또한 각각의 계정에서 정치적으로 편향된 주제들이 다뤄지고 있는지를 관찰하였다. 설문조사를 통해 유튜브로 정치 및 뉴스 영상을 소비하는 이용자들에게 보수/진보 계정에서 6일째에 추천된 영상 리스트를 제공하여 이용자들이 선택적 노출을 보이는지를 알아보았다. 연구결과, 시간이 지날수록 보수 계정에서는 보수 성향의 영상과 채널이 더욱 추천되고, 진보 계정에서는 진보 성향의 영상과 채널이 더욱 추천되었으며, 보수 계정과 진보 계정에서 추천된 영상들은 대부분 정치적으로 편향된 주제를 다루고 있는 것으로 나타났다. 응답자들의 약 77%는 자신이 선호하는 정치 성향의 영상에 선택적으로 노출되어 보이는 것으로 나타났다.

■ 중심어 : | 유튜브 | 추천 알고리즘 | 필터버블 | 선택적 노출 | 뉴스 이용 | 정치적 편향성 |

Abstract

This study empirically confirmed 'the political bias of the YouTube recommendation algorithm' and 'the selective exposure of user' to verify the Filter Bubble phenomenon of YouTube. For the experiment, two new YouTube accounts were opened and each account was trained simultaneously in a conservative and a liberal account for a week, and the "Recommended" videos were collected from each account every two days. Subsequently, through the text mining method, the goal of the research was to investigate whether conservative videos are more recommended in a righties account or lefties videos are more recommended in a lefties account. And then, this study examined if users who consumed political news videos via YouTube showed "selective exposure" received selected information according to their political orientation through a survey. As a result of the Text Mining, conservative videos are more recommended in the righties account, and liberal videos are more recommended in the lefties account. Additionally, most of the videos recommended in the righties/lefties account dealt with politically biased topics, and the topics covered in each account showed markedly definitive differences. And about 77% of the respondents showed selective exposure.

■ keyword : | YouTube | Recommendation Algorithm | Filter Bubble | News | Political Bias | Selective Exposure |

I. 서론

최근 언론진흥재단에서 발표한 <디지털 뉴스 리포트 2019>에 따르면 2019년에 들어서 45-54세 연령대의 디지털 뉴스 의존도(50%)는 텔레비전 뉴스 의존도(47%)를 앞지르기 시작했다. 뉴스 소비를 위해 활용하는 디지털 매체 중, 우리나라에서 가장 많이 사용하는 매체는 네이버(66%)와 유튜브(40%)로 나타났다[1]. 특히 유튜브를 활용한 뉴스 소비는 지난해 대비 이용자 비율이 급상승하였으며, 50대 이상의 연령층에서 급속도로 증가하고 있는 양상을 보였다[2].

이에 기존 언론들도 유튜브를 통해 콘텐츠를 유통하고, 정당 및 정치인들도 유튜브를 통해 대중과 소통하기 시작하였다. 유튜브 저널리즘이라는 용어가 등장할 정도로 유튜브의 영향력이 확대되면서 학계에서는 유튜브를 통한 뉴스 소비에 대해 다양한 논의가 벌어지고 있다. 그 중 유튜브의 추천 알고리즘이 이용자의 이념적 성향을 강화하여 이념의 양극화를 초래할 수 있다는 '필터 버블(Filter Bubble)' 현상에 대한 우려가 대표적이다. 필터버블(Filter Bubble)이란 추천 알고리즘에 의해 생기는 정보편식 현상으로, 추천 알고리즘이 이용자가 좋아할만한 정보만 제한적으로 제공하면서 이용자는 자신이 선호하는 정보를 더욱 용이하게 선택하게 되고 그 결과 스스로의 문화적 또는 이념적 거품(bubble)에 갇히는 현상을 말한다[3].

선행 연구에 따르면, 유튜브를 통해 정치 및 뉴스 소비하는 이용자들에게 필터 버블 현상이 일어나는 주된 요인은 두 가지로 압축할 수 있다[4][5]. 하나는 이용자의 정치적 성향에 맞춰 제한적으로 정보를 제공하는 추천 알고리즘의 정치적 편향성이고, 다른 하나는 이용자가 자신이 선호하는 정치 성향의 영상을 자발적으로 선택하는 선택적 노출이다. 하지만 필터버블 현상에 대한 선행연구는 대부분 구글(Google) 추천 알고리즘을 대상으로 하며, 이용자의 선택적 노출(selective exposure) 요인은 배제한 채, 추천 알고리즘이 이용자의 이념적 성향에 따라 편향적으로 정보를 제공하는가를 검증하거나[6][7][8], 추천 알고리즘의 작동 방식을 밝히는 데에 주로 초점이 맞춰져 있었다[9][10][11].

본 논문은 실제로 유튜브 이용자의 정치 성향에

따라 유튜브가 어떤 정치 뉴스를 추천하는지를 살펴보고, 유튜브 이용자들은 자신의 정치성향에 따라 특정 정치 성향의 영상을 선택적으로 시청하는지를 살펴보고자 하였다.

이에 본 논문에서는 필터버블 현상의 주요 요인인 추천 알고리즘의 정치적 편향성(추천 알고리즘이 이용자가 선호하는 정치 성향의 영상을 제한적으로 제공하는 것)과, 이용자들의 선택적 노출(이용자가 자신이 선호하는 정치 성향의 영상을 더욱 선택하는 것)을 종합적으로 알아보고자 하였다.

이를 위해 새로운 유튜브 계정 2개를 개설하여 각각의 계정을 보수/진보 계정으로 일주일 동안 훈련시켰고, 각 계정에서 추천받은 영상들을 이를 간격(2일, 4일, 6일)으로 수집하였다. 수집한 영상들을 텍스트 마이닝(Text Mining) 방법을 통해 보수 계정에서는 보수 성향의 영상이 추천되는지, 진보 계정에서는 보수 성향의 영상이 추천되는지를 살펴보고, 각각의 계정에서 어떤 토픽이 다뤄지고 있는지를 관찰 하였다. 다음으로 유튜브를 통해 정치 및 뉴스 영상을 소비하는 이용자들에게 보수/진보 계정에서 6일째에 추천된 영상 리스트를 제공하여 이용자들이 자신의 정치 성향에 따라 선택적 노출(자신이 선호하는 정치 성향의 영상을 자발적으로 선택하는 태도)을 보이는지를 알아보았다.

본 논문의 의의는 여러 전문가들이 우려하고 있는 [4][5][12][13] 유튜브 추천 알고리즘으로 인한 필터 버블 현상을 실증적으로 검증하는 것이다. 본 연구는 추천 알고리즘의 정치적 편향성(추천 알고리즘이 이용자의 정치 성향에 따라 제한적으로 영상을 제공하는 것)과 이용자들의 선택적 노출(이용자들이 자신이 선호하는 정치 성향의 영상을 자발적 선택하는 것)을 텍스트 마이닝과 설문조사를 활용하여 종합적으로 검증하는데에 의의가 있다.

II. 기존 연구 및 연구문제

1. 유튜브의 추천 서비스

유튜브는 이용자의 관심사와 선호도를 파악하여 이용자가 원할만한 정보를 선별적으로 제공한다. 유튜브

에 접속하여 로그인을 하면 메인화면에 개인 맞춤형 추천 영상 리스트가 제공되고, 영상을 하나 재생하면 ‘다음 동영상’으로 표시된 목록에도 추천 영상이 제공된다.

유튜브의 최고 상품 담당자(CPO)인 닐 모한(Neal Mohan)의 인터뷰에 따르면 유튜브 이용자들의 시청 시간 중 70%가 추천 알고리즘에 의해 이뤄진다고 한다 [11]. 하지만 대부분의 기업이 그렇듯 유튜브 추천 알고리즘은 차별화된 서비스를 제공할 수 있는 자원이기 때문에 외부에 공개되지 않는다. 더욱이 유튜브 추천 알고리즘은 영상을 통해 광고 수익을 얻는 크리에이터와 광고주들 간의 이해관계가 얽혀있기 때문에 공개되기 어렵다. 다만 유튜브 추천 알고리즘은 구글의 인공지능(AI) 알고리즘을 기반으로 구축되었기 때문에 구글 연구자들이 발표한 논문을 통해 일부분을 파악할 수 있다.

구글 연구자들이 발표한 논문에 따르면, 유튜브 추천 영상 목록은 이용자가 이전에 어떤 영상을 시청했는지, 얼마동안 시청했는지, 어떤 키워드로 검색했는지 등과 같은 이용자의 이용 기록과 더불어 이용자의 인구통계학적 정보를 알고리즘에 반영하여, 현재 재생하는 영상과 비슷한 주제의 영상, 함께 재생된 적이 많은 영상으로 이뤄진다. 또한 추천 영상 순서는 이용자가 특정 영상을 클릭한 후 얼마나 시청하였는지, 영상에 ‘좋아요’를 눌렀는지 등과 같은 이용자의 행동을 알고리즘에 반영하여 정해진다. 나아가 이용자가 추천 목록의 상단에 있는 영상을 개인적 선호가 아닌 단순히 추천 목록의 맨 위에 있어서 클릭한 경우, 알고리즘은 이용자의 선호를 학습하기보다 알고리즘의 행위를 학습하게 되기 때문에 높은 순위의 영상은 가중치를 떨어뜨리고, 추천 목록의 하단의 영상은 가중치를 높이는 방식으로 작동한다[4][9][11].

2. 유튜브와 필터버블 현상

필터 버블(Filter Bubble)이란 추천 알고리즘에 의해 생기는 정보편식 현상으로, 추천 알고리즘이 이용자가 좋아할만한 정보만 제한적으로 제공하면서 이용자는 자신이 선호하는 정보를 더욱 용이하게 선택하게 되고 그 결과 스스로의 문화적 또는 이념적 거품(bubble)에 갇히는 현상을 말한다[3]. 뉴스 분야는 사람들에게 다양

한 정보를 전달하여 여론 형성을 돕는 역할을 수행할 필요가 있기 때문에 필터버블 현상에 대한 우려가 더욱 큰 분야이다. 추천 알고리즘이 뉴스를 이용자의 선호에 맞게 선별하여 전달한다면, 이용자는 다양한 관점을 접할 수 없고 개인의 이념 또는 세계관을 넓히기는 어려워지며 건전한 담론을 형성하기도 어려워진다[4].

유튜브의 경우, 추천 알고리즘을 밝히지 않고 있으며 일부 알려진 바에 따르면 필터 버블을 고려하기보다 이용자의 영상 시청 시간을 늘리고 참여도를 높이기 위한 목적으로만 추천 알고리즘을 개선한다[4]. 이에 따라 유튜브를 통해 정치 및 뉴스 영상을 소비하는 사람이 증가하는 현 시점에서, 유튜브 추천 알고리즘으로 인한 필터버블(Filter bubble) 현상에 대한 우려는 더욱 깊어지고 있다.

3. 연구문제

유튜브 추천 알고리즘의 정치적 편향성 문제는 지속적으로 제기되고 있다. 하지만 실제로 유튜브에서 어떤 정치 및 뉴스 영상들이 얼마나 추천되며, 서로 다른 정치 성향을 지닌 사용자 계정에 따라 정치적으로 편향된 영상들이 추천되는지, 또한 시간이 지날수록 편향성이 더욱 심화되는지 등에 대하여 실증적으로 분석한 경우는 많지 않다. 특히 영어권에 비해, 한국의 정치 및 뉴스 장르의 추천 영상에 대한 분석은 매우 부족한 편이다.

방송과 신문의 뉴스보도를 대상으로 정치적 편향성을 측정하는 연구를 오래전부터 진행되었다. 시퍼(Schiffer)는 뉴스보도의 정치적 편향성을 측정하는 기존 연구들을 분석한 결과, 뉴스보도의 정치적 편향성을 크게 두 가지 방법으로 측정한다고 주장하였다[14].

하나는 뉴스보도의 물리적인 양이 어느 한 정당에 편중되어 있는지에 대한 정도를 측정하는 것이고, 다른 하나는 뉴스보도가 갖는 논조가 어느 한 정당에 편향되어 있는지에 대한 정도를 측정하는 것이다. 본 논문에서도 기존 연구의 방법론을 참고하여 정치적 편향성의 측정 기준을 물리적 편향성과 주제의 편향성으로 구분하였다.

물리적 편향성을 측정하기 위해 <연구문제 1>을 도출하여 보수 성향으로 훈련된 계정에서는 보수 성향의

상 및 채널이, 진보 성향으로 훈련된 계정에서는 진보 성향의 영상 및 채널이 더욱 추천되는지를 알아보고, 주제의 편향성을 측정하기 위해 <연구문제 2>를 도출하여 각 계정에서 다루는 주제의 차이를 비교하여 분석하였다.

연구문제 1. 보수 계정에서는 보수 성향의 영상 및 채널이, 진보 계정에서는 진보 성향의 영상 및 채널이 더욱 추천되는가?

연구문제 2. 보수 계정과 진보 계정에서 추천되는 정치 및 뉴스 영상의 주제들은 정치적으로 편향된 차이를 보이는가?

추천 알고리즘이 어떤 정보를 추천하든 결국 정보를 선택하는 것은 개인의 자발적인 의지이기 때문에 추천 알고리즘의 작동 방식뿐만 아니라 자신이 선호하는 정보만을 선택하는 개인의 의지도 필터 버블을 일으키는 주된 요인이다. 실제로 페이스북 이용자들을 대상으로 이용자들이 뉴스피드를 통해 접하는 뉴스의 이념적 다양성을 연구한 사이언스지의 발표에 따르면, 정치적 성향을 프로필에 표시한 이용자들이 다양성이 더 낮은 뉴스피드를 보고 있는 것은 맞으나, 이는 추천 알고리즘에 의한 결과가 아니라 이용자가 스스로 자신의 이념과 비슷한 콘텐츠를 선택하였기 때문이라는 결론을 제시하였다[15].

하지만 추천 알고리즘의 필터 버블 현상을 분석한 대부분의 연구들은 이용자가 자신이 선호하는 정치 성향의 영상을 자발적으로 선택하는 것에 대한 요인은 배제한 채, 추천 알고리즘이 이용자의 이념적 성향에 따라 선별적으로 정보를 제공하는가를 검증하는 데만 초점이 맞춰져 있거나[6][7][8] 추천 알고리즘의 작동 방식을 밝히는 데에 주로 초점이 맞춰져 있다[9][10][16]. 오세욱의 연구[4]에서 유튜브 추천 알고리즘에 대한 이용자의 인식과 이용 행태 그리고 추천 콘텐츠 선택 시 중요하게 생각하는 요인 등을 조사하였지만, 이용자가 자신이 선호하는 정치 성향의 영상을 선택적으로 소비하는지에 대한 조사는 이루어지지 않았다. 이에 본 논문에서는 특정 정치 성향을 지닌 유튜브 이용자가 실제

로 자신의 정치성향에 따라 영상을 선택적으로 소비하는지 혹은 자신과 다른 성향의 정치 영상도 함께 소비하는지를 확인해보고자 하였다. 이를 위해 <연구문제 3>에서는 이용자들이 유튜브를 통해 추천 받은 영상을 그대로 이용하는지 아니면 자신이 선호하는 정치 성향의 영상을 선택적으로 이용하는지를 알아보았다.

연구문제 3. 유튜브 이용자들은 자신이 선호하는 정치 성향의 영상을 선택적으로 소비하는가?

III. 텍스트 마이닝 분석

1. 텍스트 마이닝(Text Mining)

텍스트 마이닝은 문자로 이루어진 비정형 텍스트 데이터를 수집하여 의미 있는 정보를 찾아내는 분석 방법을 의미한다. 텍스트 마이닝의 기법으로는 주제에 따라 문서를 나누는 문서 분류(Document Classification)와 군집 분석(Clustering Analysis), 원하는 정보를 가진 문서를 찾는 정보 검색(Information retrieval), 필요한 정보를 찾는 정보 추출(Information Extraction), 문서 내에 단어들의 출현 빈도를 분석하는 빈도 분석(Frequency Analysis) 등이 있는데[17], 본 논문에서는 빈도분석과 군집분석 방법을 사용하였다.

2. 유튜브 추천 영상 수집

텍스트 마이닝 분석을 진행하기 위해서는 분석하고자 하는 텍스트 데이터를 수집하는 과정이 필요한데, 본 연구에서는 추천된 유튜브 영상의 “채널명”과 “제목”을 사용하였고 데이터 수집을 위해 순수한 상태의 계정 2개를 개설하였다. 계정을 만들 때 설정하는 개인 정보는 사용자의 이름, 생일, 성별이 있다. 보수 계정의 이름은 김 보수, 진보 계정의 이름은 김 진보라고 설정하였고, 생일은 2000년 11월 25일, 성별은 공개 안함으로 두 계정 모두 동일하게 설정하였다. 그다음 김 보수의 계정은 보수 성향의 채널 5개를 구독하고, 김 진보의 계정은 진보 성향의 채널 5개를 구독하였다. 각각의 계정이 구독한 5개의 채널은 국내 유튜브 정치 카테고리

고리에서 구독자 수와 누적조회 수를 기준으로 보수, 진보 성향의 상위 5개 채널을 선정하였다[18]. 이에 따라 보수 계정은 신의한수(60만), 펜앤드마이크 정규재 TV(41.2만), 황장수의 뉴스브리핑(37.2만), 고성국 TV(28.1만), TV홍카콜라(26.4만), 진보 계정은 사람사는세상 노무현재단(74.2만), 판지방송국(71.4만), 미디어 몽구(25.9만), 서울의 소리(47.8만), 유재일(12.8만) 채널을 구독하였다.

이후 첫날을 제외한 2019년 11월 26일부터 12월 1일까지 6일 동안 매일 각 계정이 구독한 5개 채널의 영상을 2시간씩 시청하였으며, 이 과정을 통해 의도적으로 보수 계정은 보수 성향의 계정으로, 진보 계정은 진보 성향의 계정으로 훈련을 시켰다. 또한 훈련을 시키는 6일 동안 이를 단위(2일째, 4일째, 6일째)로 각 계정에서 제공하는 추천 영상의 “채널명”과 “제목”을 파이썬(Python 3.6.4) 셀레늄(selenium)을 활용하여 웹스크래핑(Web Scrapping) 방법으로 수집하였다. 이를 간격으로 추천 영상의 “채널명”과 “제목”을 수집한 이유는 각 계정의 시청 활동이 추천 알고리즘에 충분히 반영되는 시간을 고려하였기 때문이다.

데이터 수집을 완료한 후, <연구문제 1>을 알아보기 위해 보수/진보 각 계정에 추천된 영상들의 “채널명”을 빈도 분석(Frequency Analysis)하였으며, [연구문제 2]를 알아보기 위해서는 K-평균 군집분석(Clustering Analysis)을 진행하였다.

3. 빈도분석(Frequency Analysis)

빈도분석을 위해 보수/진보 각 계정에서 2일째, 4일째, 6일째에 수집한 영상들의 채널을 조사하고, 각 채널의 정치적 성향을 구분하였으며 추천된 빈도를 계산하였다. 그리고 각각의 날짜마다 추천된 전체 영상 중 보수 성향의 영상 및 채널과 진보 성향의 영상 및 채널의 비율을 계산하였다.

추천된 영상의 정치 성향의 구분방법은 최선규·유수정·양성은의 연구에서 사용했던 방법론을[19] 본 논문 에 맞게 수정하여 적용하였다. 또한 채널의 정치 성향은 영상을 제공하는 각 채널을 개설한 크리에이터의 당적(黨的)에 따라 진보 또는 보수 성향의 채널로 구분하였다. 예를 들어 1) 크리에이터가 보수 성향의 정당에

속해 있거나, 2) 보수 성향의 정당에 소속된 적이 있었던 경우 해당 채널의 크리에이터를 보수 성향의 크리에이터로 분류하였다. 또한 3) 크리에이터의 당적이 모호한 경우, 최근 치러진 총선(제21대 국회의원 선거)과 관련하여 어느 정당의 홍보영상을 채널에 게시하였는지를 관찰하여 정치적 성향을 구분하였다. 이에 더해 언론사의 정치적 편향성을 검증한 선행연구[19]에 따라 4) 한겨레, 경향신문은 진보 언론사로 조선, 중앙, 동아일보는 보수 언론사로 보고, 진보 언론사에서 만든 채널은 진보 성향의 채널, 보수 언론사에서 만든 채널은 보수 성향의 채널이라고 구분하였다.

4. 군집분석(Clustering Analysis)

<연구문제 2>를 알아보기 위해서는 K-평균 군집분석(Clustering Analysis)을 진행하였다. K-평균 군집분석은 데이터 공간에서 개체 간의 거리를 계산하여 K 개의 군집으로 개체를 구분하는 분석 방법이다[20][21]. K-평균 군집분석을 위한 데이터는 보수/진보로 훈련된 두 개의 계정에서 2일째, 4일째, 6일째에 추천된 영상들의 “제목”을 사용하였다. 예를 들어, 보수 계정에서 2일째에 504개의 영상이 추천되었으므로, 수집한 제목의 개수는 504개이다. 각각의 제목은 여러 개의 단어로 이루어져 있는데 이를 형태소 단위로 나누고 명사만 추출한 뒤, 벡터화(vectorization)하는 작업을 진행하였다. 이 과정에서 각 제목마다 ‘단어 행렬(DTM)’이 형성되고, 이 행렬을 통해 2일차의 제목들에는 어떤 단어가 가장 높은 빈도로 등장하며, 주로 어떤 단어들이 함께 등장하는지를 알 수 있다. 그리고 벡터화 된 단어들을 모두 합쳐서 말뭉치(corpus)를 형성하고, 말뭉치(corpus)를 k개의 군집(토픽)으로 분류하기 위해 k값을 지정한다. 이를 통해 데이터 공간 안에 k개의 중심값이 임의로 지정되고 k개의 군집이 형성된다. 또한 중심값은 같은 군집에 속한 개체(단어)들의 평균 유클리디안 거리(Euclidean distance) 값을 계산하여 새로운 중심 값으로 이동하는 것을 반복한다.

본 논문에서는 보수/진보 계정에서 수집한 약 4000여개의 영상 제목에 포함되어 있는 모든 단어들이 군집 대상 개체이기 때문에 군집 분석의 여러 가지 방법 중, K-평균 알고리즘 기반의 군집 분석 방법을 활용하였

다. 최적의 k값을 정하기 위한 방법으로는 실루엣(Silhouette) 기법을 사용하였다. 실루엣 기법은 군집의 품질을 정량적으로 계산하는 방식으로, 실루엣 계수는 아래의 수식으로 계산된다[수식 1].

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

수식 1.

a(i)는 군집 내 데이터들의 평균 거리를 계산한 응집도(cohesion)이며, b(i)는 군집 간의 분리도(separation)를 나타내는 값으로, b(i)의 값은 크고 a(i)의 값은 작을수록 군집의 개수가 최적화 되어 있다는 것을 의미한다. 이에 따라 s(i) 값이 1에 가까울수록 군집의 개수가 최적화 되었다는 의미이고, 0에 가까울수록 군집으로 분리하는 것이 무의미하다는 의미이다. 본 논문에서는 실루엣 계수를 통해 k값을 6으로 지정하였다.

이상의 과정을 통해 본 논문에서는 보수/진보 각각의 계정에서 2일째, 4일째, 6일째 추천된 영상들의 제목을 토대로 각 날짜마다 6개의 군집으로 군집화(Clustering) 하여, 총 48개의 군집을 추출하였다. 그리고 각 날짜마다 추출된 6개의 군집은 날짜별, 계정별(보수/진보)로 비교 분석하였다[표 1].

IV. 이용자의 선택적 노출 측정

대상자는 유튜브를 통해 정치 및 뉴스를 소비하는 성인인 진보성향의 이용자 42명, 보수성향의 이용자 36명이었으며, 기간은 2020년 6월 19일부터 2020년 6월 21일까지 약 3일간 진행하였다.

실험을 위해 먼저 보수/진보 성향으로 훈련시킨 각각의 계정에서 6일째에 수집한 추천 영상들 중 15개씩 영상을 추출하여 실험 처치물을 만들었다. 즉, 보수 계정에서 수집한 추천 리스트 중 15개의 영상을 추출해 보수 성향의 추천 리스트를 하나 만들고, 진보 계정에서 수집한 추천 리스트 중 15개의 영상을 추출해 진보 성

향의 추천 리스트를 만들었다. 15개의 영상은 모두 정치 및 뉴스 장르의 영상이며, 실제 추천 리스트의 진보 또는 보수 성향의 채널 비율을 적용하여 구성하였다.

예를 들어 보수 계정에서 6일째에 추천된 정치 및 뉴스 장르의 영상이 총 300개고, 이 중 보수 성향의 채널이 240개(80%), 진보 성향의 채널이 60개(20%)라면, 보수 계정에서 추출한 15개의 영상 중 12개(80%)는 나머지 3개(20%)를 진보 성향의 영상으로 구성한 것이다. 진보 계정에서 추출한 15개의 영상도 마찬가지로 정치 및 뉴스 영상 중 보수 또는 진보 성향의 채널 비율을 적용하여 구성하였다. 15개의 영상을 모두 정치 및 뉴스 장르로 구성한 이유는 해당 실험의 대상이 유튜브를 통해 정치 및 뉴스 영상을 소비하는 이용자들이며, 이들이 자신이 선호하는 정치 성향의 영상을 자발적으로 선택하는지를 알아보고자 했기 때문이다. 참여자들에게 보수/진보 계정에서 15개씩 추출하여 구성한 2개의 추천 리스트를 제공하여, 시청하고 싶은 영상을 선택하도록 하였다. 그리고 실험 대상자의 선택적 노출 정도를 '자신의 정치성향과 같은 성향의 정보를 이용하는 정도'에서 '자신의 정치성향과 반대되는 성향의 정보를 이용하는 정도'의 차이를 통하여 측정하였다. 이에 따라 선택적 노출 값이 양수(+)라면 선택적 노출을 보인다고 판단하였으며, 0 또는 음수(-)라면 선택적 노출을 보이지 않는다고 판단하였다.

V. 연구결과

1. 물리적 편향성 검증 결과

〈연구문제 1〉을 알아보기 위해 유튜브 추천 알고리즘의 물리적 편향성을 분석한 결과, 시간이 지날수록 보수 계정에서는 보수 성향의 영상이 더욱 추천되고 진보 계정에서는 진보 성향의 영상이 더욱 추천되는 것을 확인하였다. 구체적으로 보수 계정에서 계정을 개설한 첫째 날 추천된 504개의 영상 중, 정치 및 뉴스 장르 영상은 43개(8.5%)였고 이중 보수 성향의 영상은 6개(14%), 진보 성향의 영상도 6개(14%)였다. 2일째에는 추천된 528개의 영상 중, 정치 및 뉴스 장르의 영상이 161개(30.5%)로 늘어났고, 이중 보수 성향의 영상이

130개(80.7%), 진보 성향의 영상이 8개(5%)로 나타나 정치 및 뉴스 장르의 영상과 함께 보수 성향의 영상이 급증한 모습을 보였다. 4일째에는 추천된 506개의 영상 중, 정치 및 뉴스 장르의 영상이 164개(32.4%)였고, 이중 보수 성향의 영상이 133개(81.1%), 진보 성향의 영상이 3개(1.8%)였다. 6일째는 추천된 520개의 영상 중, 정치 및 뉴스 장르의 영상이 202개(38.8%)였고 이중 보수 성향의 영상이 174개(86.1%), 진보 성향의 영상이 3개(1.5%)를 차지하여 시간이 지날수록 정치 및 뉴스 장르의 영상과 함께 보수 성향의 영상이 꾸준히 증가하는 모습을 보였다.

한편 보수 계정에서 기타 장르의 영상은 첫째 날은 전체 영상 중 91.5%를 차지하였고, 2일째에는 69.5%, 4일째에는 67.6%, 6일째에는 61.2%를 차지하며 시간이 지날수록 감소하는 추이를 보였지만 해당 계정으로 정치 및 뉴스 장르의 영상만 보았음에도 여전히 높은 비율로 추천되는 모습을 보였다[표 1].

표 1. 보수 계정에서 추천된 영상 빈도분석

	전체 영상 수	기타 장르 영상	정치 및 뉴스 장르 영상			
			전체	보수 성향	진보 성향	성향 모름
보수 계정	504개	461개 (92%)	43개 (9%)	6개 (14%)	6 (14%)	31 (72%)
2일째	528개	367개 (70%)	161개 (31%)	130 (81%)	8 (5%)	23 (14%)
4일째	506개	342개 (68%)	164개 (32%)	133 (81%)	3 (2%)	28 (17%)
6일째	520개	318개 (61%)	202개 (39%)	174 (86%)	3 (2%)	25 (12%)

진보 계정에서는 계정을 개설한 첫째 날은 추천된 510개의 영상 중, 정치 및 뉴스 장르의 영상이 50개 (98.6%)였고 이중 보수 성향의 영상이 6개(12%), 진보 성향의 영상이 4개(8%)였다. 2일째에는 추천된 510개의 영상 중, 정치 및 뉴스 장르의 영상이 229개(44.9%)로 늘어났고 이중 보수 성향의 영상은 8개(3.5%), 진보 성향의 영상은 179개(78.2%)로 정치 및 뉴스 장르의 영상과 함께 진보 성향의 영상이 급증한 모습을 보였다. 4일째에는 추천된 510개의 영상 중, 정치 및 뉴스

장르의 영상이 230개(45.1%)였고 이중 보수 성향의 영상이 8개(3.5%), 진보 성향의 영상이 182개(79.1%)였다. 6일째에는 추천된 510개의 영상 중, 정치 및 뉴스 장르의 영상이 238개(46.6%)였고 이중 보수 성향의 영상이 9개(3.8%), 진보 성향의 영상이 199개(83.6%)를 차지하여 시간이 지날수록 정치 및 뉴스 장르의 영상과 함께 진보 성향의 영상이 꾸준히 증가하는 모습을 보였다.

한편 진보 계정에서 기타 장르의 영상도 보수 계정과 마찬가지로 첫째 날은 전체 영상 중 90.2%를 차지하였고, 2일째에는 55.1%, 4일째에는 54.9%, 6일째에는 53.3%를 차지하며 시간이 지날수록 감소하는 추이를 보였지만 해당 계정으로 정치 및 뉴스 장르의 영상만 보았음에도 50%가 넘는 비율로 추천되는 모습을 보였다[표 2].

표 2. 진보 계정에서 추천된 영상 빈도분석

	전체 영상 수	기타 장르 영상	정치 및 뉴스 장르 영상			
			전체	보수 성향	진보 성향	성향 모름
진보 계정	510개	460개 (90%)	50개 (10%)	6개 (12%)	4 (8%)	40 (80%)
2일째	510개	281개 (55%)	229개 (45%)	8 (4%)	179 (78%)	42 (18%)
4일째	510개	280개 (55%)	230개 (45%)	8 (4%)	182 (79%)	40 (17%)
6일째	510개	272개 (53%)	238개 (53%)	9 (4%)	199 (84%)	30 (12%)

또한 보수 계정에서 추천된 보수 성향의 채널과 진보 계정에서 추천된 진보 성향의 채널을 빈도 분석한 결과, 시간이 지날수록 보수 계정에서는 보수 성향의 채널이 더욱 증가하였으며 진보 계정에서는 진보 성향의 채널이 더욱 증가하는 것을 확인하였다. 이는 보수 계정에서는 보수 성향의 영상이, 진보 계정에서는 진보 성향의 영상이 더욱 추천되는 현상이 하나의 채널에서 파생된 영상 수가 늘어난 것뿐만 아니라, 보수 계정에서는 보수 성향의 채널이, 진보 계정에서는 진보 성향의 채널이 더욱 늘어난 결과라고 볼 수 있다.

2. 주제의 편향성 검증 결과

〈연구문제 2〉를 위해 주제의 편향성을 분석한 결과, 보수 계정과 진보 계정에서 추천된 영상들은 대부분 정치적으로 편향된 주제를 다루고 있었으며 각 계정에서 다루는 주제가 확연히 다른 모습을 보였다.

구체적으로 2일째 보수 계정에서는 ① 문재인 정권 퇴진을 요구하는 광화문 집회, ② (당시) 자유한국당 황교안 대표의 단식 투쟁 ③ 문재인 정권의 부정선거 의혹 등의 주제를 다루고 있다. 반면 진보 계정에서는 ① 청와대 앞 투쟁에 대한 비판 ② 문재인 대통령에 대한 우호적인 평가 ③ 문재인 대통령의 한·아세안 양자회담 등의 주제를 다루고 있다. 이를 통해 2일차의 보수/진보 각 계정에서 다루는 주제가 정치적으로 매우 편향되어 있으며, 서로 다른 주제를 다루고 있다는 것을 알 수 있으며, 또한 청와대 앞 집회에 대해서는 두 계정 모두 다루고 있지만 각 계정의 관점이 매우 다르다는 것을 알 수 있다.

4일째 보수 계정에서는 ① 조국에 대한 검찰 수사 ② 문재인 대통령 부정선거 의혹 ③ (당시) 황교안 자유한국당 대표의 단식 투쟁과 필리버스터 ④ 문재인 정권을 규탄하는 광화문 집회 등에 대한 주제를 다루고 있다. 반면 진보 계정에서는 ① 과거 박근혜 전 대통령에 대한 탄핵 요구 영상 ② 공수처 설치 및 검찰 개혁 주장 ③ 검찰의 수사를 받는 조국을 두둔하는 내용 ④ 홍준표, 박사모, 황교안 등 보수 성향의 정치인 또는 단체에 대한 비판 등에 대한 주제를 다루고 있다. 4일째 또한 2일째와 마찬가지로 보수/진보 각 계정에서 다루는 주제가 정치적으로 편향되어 있으며, 서로 다른 주제를 다루고 있다. 또한 조국 사건은 두 계정 모두 다루고 있지만, 각 계정의 관점이 매우 다르다는 것을 알 수 있다.

6일째 보수 계정에서는 ① 문재인 정권의 언론장악 의혹 ② 문재인 탄핵을 요구하는 광화문 집회 ③ 박근혜 전 대통령 탄핵 재판에 대한 의혹 ④ 여당의 검찰개혁 주장에 대한 비판 ⑤ 문재인 정부의 공수처 설립 주장에 대한 비판 등이었다. 반면 진보 계정에서는 ① 검찰 개혁을 요구하는 촛불집회 ② 자유한국당의 단식 투쟁과 필리버스터에 대한 비판 ③ 검찰의 수사를 받는 조국을 두둔하는 내용 ④ 홍준표, 지만원, 박사모, 황교

안, 나경원 등 보수 성향의 정치인 또는 단체에 대한 비판 ⑤ 선거제도 개혁에 대한 의지 등의 주제를 다루고 있다.

6일째 또한 2일째, 4일째와 마찬가지로 보수/진보 각 계정에서 다루는 주제가 정치적으로 편향되어 있으며, 서로 다른 주제를 다루고 있다는 것을 알 수 있다. 또한 광화문 집회에 대하여 두 계정 모두 다루고 있지만, 보수 계정에서는 문재인 탄핵을 요구하는 집회에 대한 영상이 추천되었으며 진보 계정에서는 검찰개혁을 요구하는 집회에 대한 영상이 추천되었다.

3. 이용자들의 선택적 노출 검증 결과

〈연구문제 3〉을 알아보기 위해 실험을 진행한 결과 전체 보수 및 진보 성향을 지닌 78명의 응답자 중, 60명(76.9%)의 선택적 노출 값이 1 이상으로 선택적 노출을 보였으며, 18명(23%)은 선택적 노출 값이 0 이하로 선택적 노출을 보이지 않았다. 이에 따라 본 논문에서는 응답자의 약 77%가 선택적 노출을 보인다는 것을 확인하였다. 또한 진보 성향을 지닌 응답자의 평균 선택적 노출은 5.33(SD=4.83), 보수 성향을 지닌 응답자의 평균 선택적 노출은 3.38(SD=4.74)로 나타났다. 정치 성향별 선택적 노출 정도의 차이가 유의미한지를 알아보기 위하여 신뢰구간 95% 수준에서 독립표본 T 검증을 하였으나, 유의미한 차이가 없는 것으로 나타났다.

VI. 결론

본 논문에서는 유튜브를 통한 정치 및 뉴스 영상 소비가 급증하고 있는 상황에서 많은 전문가들이 우려하고 있는 유튜브 추천 알고리즘으로 인한 필터버블(Filter Bubble) 현상을 실증적으로 검증하고자 하였다. 이를 위해 필터버블 현상의 주된 요인을 추천 알고리즘의 정치적 편향성과, 이용자의 선택적 노출을 종합적으로 알아보았다.

그 결과 시간이 지날수록 보수 계정에서는 보수 성향의 영상이 더욱 추천되고 진보 계정에서는 진보 성향의 영상이 더욱 추천되는 것을 확인하였다. 또한 보수 계정과 진보 계정에서 추천된 영상들은 대부분 정치적인

로 편향된 주제를 다루고 있었으며 각 계정에서 다루는 주제가 확연히 다른 모습을 보인다는 것을 확인하였다. 마지막으로 실험 참여자 78명 중 60명(76.9%)이 선택적 노출을 보인다는 것을 확인하였다.

Ⅶ. 한계 및 시사점

본 논문은 다음과 같은 한계를 가진다. 첫째, 본 논문에서 유튜브 추천 알고리즘의 정치적 편향성을 알아보는 실험은 일주일간 진행한 것으로, 단기간의 정치적 편향성은 확인하였지만 장기적인 관점에서의 정치적 편향성은 확인하지 못하였다. 여러 선행 연구에서 밝혔듯이 유튜브는 이용자가 특정 영상을 얼마나 시청하였으며, 어떤 채널을 구독하고, 어떤 영상에 '좋아요'를 눌렀는지 등과 같이 이용자의 복합적인 행동을 지속적으로 반영하여 추천영상 리스트를 업데이트한다. 따라서 장기적인 관점에서 추천 영상의 편향성 정도는 달라질 수 있다. 후속연구에서는 실험 기간을 더욱 길게 잡아 장기적인 관점에서 이용자의 정치 성향에 따라 추천 영상의 정치적 편향성이 나타나는지를 살펴보는 것이 필요하다.

둘째, 본 논문에서는 유튜브 추천 알고리즘의 주제의 편향성을 알아보기 위해 k-평균 군집분석을 실시하였다. 그러나 하나의 군집에 포함된 단어들이 반드시 하나의 주제를 대표하는 것이라고 볼 수는 없었으며, 하나의 군집에 여러 개의 주제가 포함되어 있기도 하였다. 또한 군집에 속한 단어들이 정확히 어떤 맥락에서 사용되었는지를 파악하는 것도 한계가 있었다. 본 논문에서는 이를 보완하기 위해 내용분석을 실시하였지만 후속연구에서는 감성 분석, 이념 분류기 등의 방법을 통해 군집에 속한 단어들의 문맥적 의미를 더욱 효율적으로 분석할 필요가 있다. 그러나 본 논문은 유튜브 추천 알고리즘으로 인한 필터버블 현상에 대한 논쟁이 뜨거운 가운데, 유튜브를 통해 정치 및 뉴스 영상을 소비하는 이용자들이 추천 알고리즘으로 인해 필터버블(Filter bubble) 현상에 갇힐 가능성이 높다는 것을 종합적으로 검증하였다는 의의가 있다. 또한 본 논문은 보수 및 진보 성향의 계정에 따라 유튜브에서 추천되는

영상이 정치적 편향성을 보인다는 것을 실증적으로 검증하여, 유튜브 추천 알고리즘으로 인한 필터버블 현상에 대한 우려가 현실이 될 수 있다는 점을 밝혔다.

* 본 논문은 신유진의 연세대학교 정보대학원(해당년도) 석사학위논문을 수정, 보완하여 작성함.

참고문헌

- [1] 김선호, 김위근, *디지털 뉴스 리포트 2019 한국*, 한국언론진흥재단 - 로이더저널리듬연구소, pp.13-28, 2019.
- [2] http://www.koreanclick.com/insights/newsletter_view.html?code=topic&id=552&page=1, 2019. 11.26.
- [3] E. Pariser, *The filter bubble: How the news personalized web is changing what we read and how we think*, Penguin Books, 2012.
- [4] 오세욱, 송해엽, "유튜브 추천 알고리즘과 저널리즘," 한국언론진흥재단, 정보통신정책연구, 제19권, 제2호, pp.69-92, 2019.
- [5] 정철운, "가짜뉴스 유튜브 극우보수와 저널리즘," 인물과사상, 제247호, pp.168-183, 2018.
- [6] 전준영, 황소윤, 윤영미, "개인화 알고리즘으로 필터버블이 형성되는 과정에 대한 검증," 멀티미디어학회 논문지, 제21권, 제3호, pp.369-381, 2018.
- [7] M. Haim, A. Graefe, and H. B. Brosius, "Burst of the filter bubble? Effects of personalization on the diversity of Google News," *Digital Journalism*, Vol.6, No.3, pp.330-343, 2018.
- [8] E. Nechushtai and S. C. Lewis, "What kind of news gatekeepers do we want machines to be? Filter bubbles, fragmentation, and the normative dimensions of algorithmic recommendations," *Computers in Human Behavior*, Vol.90, pp.298-307, 2019.
- [9] P. Covington, J. Adams, and E. Sargin, "Deep neural networks for youtube recommendations," *Proceedings of the 10th ACM Conference on Recommender Systems*,

pp.191-198, 2016.

[10] <https://www.tubefilter.com/2016/06/23/reverse-engineering-youtube-algorithm/>, 2016.6.23.

[11] <https://www.nytimes.com/2019/03/29/technology/youtube-online-extremism.html>, 2019.3.29.

[12] 강준만, “SNS 모바일 유튜브 시대의 언론,” 인물과사상, 제251호, pp.47-84, 2019.

[13] 양선희, “유튜브 저널리즘의 시대, 전통적 저널리즘의 대응현황과 과제,” 사회과학연구, 제31권, 제1호, pp.245-262, 2020.

[14] A. J. Schiffer, “Assessing partisan bias in political news: The case(s) of local senate election coverage,” Political Communication, Vol.23, No.1, pp.23-39, 2006.

[15] E. Bakshy, S. Messing, and L. Adamic, “Exposure to ideologically diverse news and opinion on Facebook,” Science, Vol.348, No.6239, pp.1130-1132, 2015.

[16] Z. Zhao, L. Hong, L. Wei, J. Chen, A. Nath, S. Andrews, A. Kumthekar, M. Sathiamoorthy, X. Yi, and E. Chi, “Recommending what video to watch next: a multitask ranking system,” In Proceedings of the 13th ACM Conference on Recommender Systems, pp.43-51, 2019.

[17] S. M. Weiss, M. Indurkha, and T. Zhang, “Fundamentals of predictive text mining,” Springer, p.41, 2010.

[18] URL:<https://zdnet.co.kr/view/?no=20190326173948>, 2019.03.27.

[19] 최선규, 유수정, 양성은, “뉴스 시장의 경쟁과 미디어 편향성: 취재원 인용을 중심으로,” 정보통신정책연구, 제19권, 제2호, pp.69-92, 2012.

[20] S. W. Lee, “Comparison of Initial Seeds Methods for K-Means Clustering,” Journal of Korean Society for Internet Information, Vol.13, No.6, pp.1-8, 2012.

[21] A. K. Jain, “Data Clustering: 50 years beyond K-means,” Pattern Recognition Letters, Vol.31, No.8, pp.651-666, 2010.

저 자 소 개

신 유 진(Yoo Jin Shin)

정회원



- 2015년 2월 : 숙명여자대학교 국문학/언론정보학(학사)
- 2020년 2월 : 연세대학교 맞춤형융합-데이터 사이언스 (석사)
- 2020년 9월 ~ 현재 : LGU+ 기술부문 빅데이터전략팀

<관심분야>: 빅데이터, 콘텐츠

이 상 우(Sang Woo Lee)

정회원



- 1992년 2월 : 연세대학교 화학과(학사)
- 1994년 2월 : 연세대학교 화학과(석사)
- 1994년 5월 : Telecommunication, Michigan State University(석사)
- 2002년 6월 : Mass Communication, Indiana University(박사)
- 2002년 7월 ~ 2008년 8월 : 정보통신정책연구원 연구위원
- 2008년 ~ 현재 : 연세대학교 정보대학원 교수

<관심분야> : 미디어, 콘텐츠, 빅데이터