

## 시리즈물 리뷰의 긍정 유사도가 흥행에 미치는 영향을 분석하기 위한 텍스트 분류

\*김수진 조형민

성균관대학교 소프트웨어대학

\*tnwsl6744@naver.com hyungmin.cho@skku.edu

Text Classification to Analyze the Effect of Positive Similarity  
in Series Reviews on the Box Office Performance

\*Kim Sujin Cho Hyungmin

Sungkyunkwan Computer Science and Engineering University

## 요약

오늘날 인터넷이 보편화되었고, 최근에는 최근에는 코로나19 유행으로 사람들이 집에 머무르는 시간이 많아지면서 여러 온라인 플랫폼을 통해 영화, 드라마 등의 프로그램을 시청하는 것에 관심이 많아지고 있다. 또한, 그러한 시대적 흐름에 따라 시즌제 형식의 시리즈물을 통해 보다 퀄리티 높은 콘텐츠를 보고자 하는 소비자 니즈도 증가하고 있다. 시리즈물은 전편과 속편이 유기적으로 연결되기 때문에 전편의 리뷰를 분석하여 관객의 니즈를 파악하고 그것을 속편에 반영하는 것이 중요해 보인다. 따라서 본 연구에서는 텍스트 분류를 통해 시리즈물의 전편과 속편 리뷰의 긍정 유사도를 비교하고, 나아가 긍정 유사도가 흥행 성적에 유의미한 영향을 미치는지 알아보려고 한다.

## 1. 서론

오늘날 인터넷이 보편화 되었고, 최근에는 코로나19 유행으로 사람들이 이전보다 집에 머무르는 시간이 많아지면서 영화나 드라마 등의 프로그램을 시청하는 것에 관심이 많아지고 있다. 이와 함께 이야기가 연속되어 더 깊이 있는 콘텐츠를 담아낼 수 있는 시리즈물에 대한 소비자 니즈도 증가하고 있으며, 실제로 많은 시리즈물이 제작되고 있다.

하지만 시리즈물의 전편에 대한 반응이 좋고 흥행에 성공했다고 하더라도, 그 속편에 대한 반응이 항상 좋고 흥행하는 것은 아니다. 오히려 전편의 긍정적인 부분을 담아내지 못했다는 평가 함께 흥행에 실패하는 경우도 많다. 시리즈물의 경우, 전편과 속편의 유기적 연결성이 강하기 때문에 전편의 리뷰를 분석하여 관객의 니즈를 파악하고 그것을 속편에 반영하는 것이 중요해 보인다.

따라서 본 연구에서는 텍스트 분류를 통해 시리즈물의 전편과 속편에 대한 긍정적인 리뷰를 분류해내고, 분류해낸 리뷰에서 유의미한 키워드를 추출하여 유사도를 비교한다. 이후 유사도가 흥행 성적에 유의미한 영향을 미치는지 알아보려고 한다. 긍정 유사도가 높을수록 즉, 관객이 느꼈던 전편의 긍정적인 점이 속편에 유지될수록 흥행에 성공하는지 분석해볼 것이다. 유의미한 분석 결과가 나온다면, 앞으로 시리즈물 제작자들이 어떤 방향으로 제작하면 좋을지에 대한 의견도 제시할 수 있을 것이다.

본 연구의 Overview는 다음과 같다. 총 113편의 영화(42개의 시리즈)를 선정한다. 이후 네이버 리뷰 사이트에서 시리즈물의 각 시즌에 대한 리뷰를 스크래핑한 후, 감성분석을 통해 긍정적인 리뷰를 분류해낸다. 분류된 리뷰 속 빈도수가 높으면서 유의미한 키워드를 추출한다. 추출된 키워드 목록의 유사도를 구하여 흥행 성적과 비교한다.

## 2. 관련 연구

영화 흥행과 관련하여 빅데이터를 활용하는 연구들이 여러 방향으로 진행 중이지만, 특히 관객이 남긴 영화 리뷰를 분석하는 연구들이 많았다.

첫째로 '네티즌의 흥행 영화 리뷰에 포함된 감정 동사 이용 특성 연구'[1] 논문에서는 흥행한 영화의 네티즌 리뷰에 '재미있다'의 감정 동사가 많이 사용된 것으로 보아 네티즌이 영화를 판단할 때 '재미'를 가장 큰 요인으로 생각하고 있다고 기술했다. 또한, 감정 동사를 군집 분석한 결과, 재미를 기준으로 '몰입'이나 '감동' 등 감정적인 동사를 기준으로 군집되었다고 한다. 이는 관객의 심리적, 정신적인 내부 세계의 '희, 노, 애, 락'에 대한 주관적인 생각이 영화 흥행의 중요한 요소라는 것과 리뷰에 적힌 키워드를 분석하는 것이 흥행 여부를 파악하는 것에 있어 유의미한 일임을 보여준다.

둘째로 'Research on the Movie Reviews Regarded as Unsuccessful in Box Office Outcomes in Korea: Based on Big Data Posted on Naver Movie Portal'[2]에서는 영화 평론을 바탕으로 어떤 키워드 및 주제가 영화 등급에 긍정적인 영향과 부정적인 영향을 미치는지 분석했다. 분석 결과, 평론에서 추출된 특정 영화 주제가 영화 평점에 부정적인 영향을 미쳤다고 한다. 이를 통해 평론을 분석하는 것이 영화의 작품성과 흥행을 판단하는데 중요한 역할을 할 수 있음을 알 수 있다.

'평점 및 리뷰 분석을 통한 유료 모바일 게임 애플리케이션의 흥행 요인 연구'[3]는 유료 모바일 게임 애플리케이션 리뷰를 분석함으로써 모바일 게임 애플리케이션의 흥행 요인을 파악할 수 있다고 기술했다. 모바일 애플리케이션은 직접 사용해보기 전에는 품질을 알 수 없다는 특

성 때문에 소비자 불확실성이 높아 새로운 소비자들이 기존의 소비자들이 작성한 평점이나 리뷰를 참고하여 최종 사용 여부를 결정한다고 하는데, 이는 영화도 마찬가지인 것으로 생각된다. 따라서 영화 리뷰를 분석함으로써 영화의 흥행 요인도 파악 가능하다고 유추할 수 있다.

영화 리뷰를 분석하는 방법을 연구한 논문으로는 '영화리뷰의 분류 성과 향상에 관한 연구'[4]와 'Movie Review Classification Based on a Multiple Classifier'[5], 'sentimental Analysis of Movie Reviews using Twitter'[6] 등이 있다. 세 연구에서는 텍스트마이닝을 이용하여 영화리뷰의 감성정보를 자동으로 분류하고 예측 성과를 보다 향상시킬 수 있도록 다양한 변수 선택법을 적용했다.

### 3. 설계 및 구현

본 연구에서는 113편의 영화(42개의 시리즈) 리뷰를 분석하였다. 다양한 분석을 위해 외국 시리즈물(아이언맨, 트와일라잇, 미션임파서블, 해리포터, 데드풀 등)과 국내 시리즈물(조선명탐정, 가문의 영광, 국가대표, 신과함께 등) 그리고 애니메이션 시리즈물(토이스토리, 보스 베이비, 겨울왕국 등)이라는 카테고리로 나누어 분석했다.

그 중 엽기적인 그녀, 킹스맨, 슈퍼배드, 스파이더맨을 예시로 분석 과정을 살펴보겠다. '엽기적인 그녀'와 '킹스맨'은 속편 흥행 실패작으로 볼 수 있으며, '슈퍼배드'와 '스파이더맨'은 속편 흥행 성공작으로 볼 수 있다. 흥행의 성공 및 실패를 구분하는 기준은 관객 수로 하였다. 속편의 관객 수가 전편의 95% 이상인 경우, 흥행 성공으로 판단하였다. 속편의 관객 수가 전편의 관객 수보다 많다면 흥행에 성공했다고 볼 수 있지만, 개봉 당시의 상황에 따라 관객 수는 조금씩 영향을 받을 수 있기 때문에 5%의 오차범위를 두었다.

엽기적인 그녀는 1편의 관객 수가 480만 명이었으나, 2편은 7.7만 명이다. 이러한 관객 수의 감소는 1편에 비해 2편 흥행이 실패했음을 알 수 있다. 킹스맨 또한 마찬가지다. 킹스맨의 1편, 2편, 3편의 관객 수는 각각 102만 명, 494만 명, 102만 명으로 2편은 흥행에 성공했으나 3편은 흥행에 실패했다고 볼 수 있다. 다음 표 1은 각 시리즈물의 관객 수와 흥행 성공 및 실패 여부를 정리한 것이다.

표 1 시리즈물의 관객수 및 흥행 여부

	1편	2편	3편
엽기적인 그녀	480만	7.7만(실패)	
킹스맨	102만	494만(성공)	102만(실패)
슈퍼배드	102만	96만(성공)	332만(성공)
스파이더맨	725만	802만(성공)	755만(성공)

분석 단계는 1) 전편과 속편의 리뷰를 스크래핑, 2) 감성분석을 통해 긍정적인 리뷰만 분류, 3) 분류된 리뷰 속 빈도수가 높고 유의미한 키워드 추출, 4) 추출된 키워드 목록의 유사도 비교다. 우선 긍정적인 리뷰를 스크래핑하기 위해 네이버 영화 리뷰 사이트(movie.naver.com)에서 시리즈물의 각 시즌에 대한 리뷰를 추출한다. 웹 페이지의 정보를 스크랩하기 위해 BeautifulSoup 패키지를 사용하고 HTTP 요청을 보낼 수 있도록 Requests 패키지를 사용한다. 이후 감성분석을 통해 긍정적인 리뷰만 분류해내 텍스트 파일로 저장한다. 저장된 긍정적인 리뷰에서

빈도수가 높으면서 유의미한 명사를 추출하여 텍스트 파일로 저장하고 시리즈물의 전편과 속편 유사도를 비교한다.

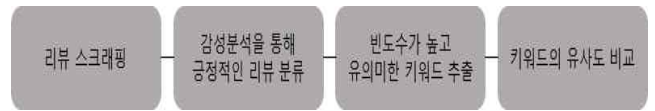


그림 1 텍스트 분류 단계

이때 사용한 유사도 공식은 자카드 유사도(Jaccard Similarity)이다. 키워드를 통해 두 집합이 얼마나 비슷한지를 측정하는 방법으로 공식은 다음과 같다.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

자카드 유사도 공식을 사용하여 서로 다른 영화의 긍정 유사도를 비교한 결과, 평균적으로 0.28의 유사도를 보였다. 따라서 키워드 유사도의 범위가 0.2~0.3인 경우, 유사도가 낮은 편이라고 판단하였다.

표 2 다른 영화의 유사도 비교

영화 1	영화 2	유사도
트와일라잇 1편	미션임파서블 6편	0.31
트와일라잇 2편	미션임파서블 6편	0.26
어벤저스 1편	토이스토리 4편	0.30
드래곤길들이기 1편	조폭마누라 1편	0.19
엽기적인 그녀 1편	킹스맨 1편	0.20
엽기적인 그녀 1편	해리포터 1편	0.17
스파이더맨 1편	해리포터 1편	0.29
스파이더맨 1편	킹스맨 1편	0.35
킹스맨 1편	해리포터 1편	0.25
⋮	⋮	⋮

### 4. 결과 및 분석

앞서 엽기적인 그녀 시리즈는 1편에 비해 2편의 관객 수가 현저히 적어 흥행에 실패했다고 판단했다. 다음 표들은 엽기적인 그녀 1편과 2편의 리뷰에서 추출한 긍정 리뷰와 키워드다. 추출한 긍정 키워드의 유사도를 비교한 결과, 0.23으로 유사도가 낮은 범위(0.2~0.3)에 속하므로 전편과 속편의 유사도가 낮다고 판단할 수 있다.

표 3 엽기적인 그녀 1편의 긍정 리뷰

긍정적인 리뷰
이때의 전지현은 정말 아름다웠었지
코믹이지만 한편으로는 가슴을 아릿하게 만드는 영화지.
역시 한국 로맨틱 코미디 영화중에 최고다
원조 '국민 첫사랑'은 따로 있었다.
지금봐도재밌다 경우는너무순수하고 전지현은너무이쁘다 ㅠㅠ
⋮

표 4 엽기적인 그녀 1편의 긍정 키워드

키워드
전지현, 로맨틱, 멜로, 시절, 운명, 감동, 차태현, 명작, 한국, 사랑, 그녀, 코미디 ...

표 5 엽기적인 그녀 2편의 긍정 리뷰

긍정적인 리뷰
와 이것도 개재밌다. 이거 추천합니다
영화를 봤습니다!!
차태현 너무좋아요!!
견우를추억하며 ..차태현은 좋아요
어제 봤는데, 기대이상.
:

표 6 엽기적인 그녀 2편의 긍정 키워드

키워드
차태현, 빅토리아, 연기, 견우, 전작, 감동, 재미, 견우, 엽기, 스토리 ...

같은 방법으로 각 시리즈물의 긍정 리뷰와 키워드를 추출하여 유사도를 비교한 결과와 흥행 성공 여부는 표 7과 같다. 유사도가 낮다고 판단되는 범위인 0.2~0.3인 경우 흥행에 실패하였으며, 유사도가 0.5 이상인 경우 흥행에 성공하였다.

위와 같은 방식으로 총 69개의 모든 케이스를 분석해본 결과, 1) 유사도가 높으며 흥행에 성공한 경우는 39개였으며, 2) 유사도가 낮으며 흥행에 실패한 경우는 22개였다. 총 69개의 케이스 중 61개의 대부분의 시리즈물이 1), 2) 에 해당되었기 때문에 전편과 속편의 긍정 유사도가 높을 때 흥행에 성공할 확률이 높고, 전편과 속편의 긍정 유사도가 낮을 때 흥행에 실패할 확률이 높다고 판단했다. 즉, 관객이 전편에서 느꼈던 긍정적인 측면이 속편에 반영되는 경우 흥행에 성공할 확률이 높다고 할 수 있다.

표 7 각 시리즈물 전편과 속편의 긍정 유사도 및 흥행 성공 여부

	1편과 2편	2편과 3편
엽기적인 그녀	0.23(실패)	
킹스맨	0.54(성공)	0.32(실패)
슈퍼배드	0.5(성공)	0.52(성공)
스파이더맨	0.56(성공)	0.5(성공)

하지만 해리포터와 같은 장기적인 시리즈나 마이펫의 이중생활 같은 애니메이션 시리즈는 다른 결과가 나오기도 한다. 해리포터 시리즈는 총 8편이며, 각 시즌의 관객 수는 (33만, 13만, 186만, 349만, 370만, 295만, 285만, 442만)이다. 이를 통해 3편(186만)이 2편(13만)에 비해 흥행에 성공했으며 6편(295만)이 5편(370만)에 비해 흥행에 실패했다고 판단할 수 있다. 그러나 2편과 3편의 긍정 유사도는 0.51이었으며, 5편과 6편의 긍정 유사도는 0.56이었다. 두 경우 모두 긍정 유사도가 0.5를 초과했다. 이는 전편과 속편의 긍정 유사도가 높을 때, 흥행에 성공한다는 앞선 결과와 반대되는 결과다.

해리포터뿐만 아니라 다른 4편 이상의 장기시리즈물 중 '트와일라

잇 3-4편, 미션임파서블 4-5편'도 유사도가 높지만 흥행에 실패했음을 보였다. 또한 '마이펫의 이중생활 1-2편, 쿵푸팬더 2-3편, 보스베이비 1-2편, 드래곤 길들이기 2-3편' 등의 애니메이션 시리즈도 마찬가지로의 결과를 보여 총 7개의 케이스가 긍정 유사도가 높지만 흥행에 실패함을 보였다.

장기 시리즈나 애니메이션 시리즈는 골수팬들이 많고, 그들이 전편의 리뷰와 속편의 리뷰를 비슷하게 작성하는 경향이 있어 흥행 여부와 상관없이 리뷰의 긍정 유사도가 높은 것으로 판단된다.

긍정 유사도가 낮지만 흥행에 성공한 영화도 있다. 데드풀의 경우, 속편의 긍정 리뷰를 분석한 결과, 전편의 긍정 리뷰에는 존재하지 않았던 새로운 인물과 신선한 각본에 대한 내용이 많았다. 그에 따라 유사도는 낮았지만, 흥행에는 성공했다.

### 5. 결론 및 향후 연구

본 연구에서는 텍스트 분류를 통해 시리즈물의 전편과 속편의 긍정적인 리뷰를 분류해냈고, 리뷰 속 빈도수가 높고 유의미한 키워드를 비교하여 긍정 유사도를 구했다. 나아가 그러한 긍정 유사도가 흥행 성적에 유의미한 영향을 미치는지 알아보았다. 이에 따라 도출 가능한 결론은 다음과 같다.

전편의 리뷰에서 추출한 긍정적인 키워드를 후속편에 반영하여 긍정 유사도를 높인다면, 흥행에 성공할 확률이 높아진다. 하지만 장기 및 애니메이션 시리즈의 경우, 골수팬이 존재하고 그들이 유사한 리뷰를 쓸 가능성이 높아 흥행 여부와 상관없이 유사도가 높게 나올 수 있다.

또한 새로운 인물이나 신선한 각본, 더 높은 완성도 등으로 전편과 속편의 긍정 유사도가 낮더라도 흥행하는 경우가 있었다. 따라서 긍정적인 점을 반영하고자 노력하는 것뿐만 아니라 관객들에게 신선하고 새롭게 느껴질 수 있는 요소들을 추가하는 것도 시리즈를 흥행시킬 수 있는 방법이 될 수 있다.

소포모어 징크스(Sophomore jinx)는 성공적인 첫 작품 혹은 활동에 비해 그에 이은 작품 혹은 활동이 부진한 경우를 가리키는 용어다. 따라서 시리즈물 제작자는 소포모어 징크스를 깨고자 노력하는데, 속편 제작 예정 영화의 제작 기획 단계와 마케팅 단계에서 텍스트 분류를 통해 추출해낸 전편의 긍정 키워드를 적극 활용한다면 흥행을 유도할 수 있을 것으로 기대된다.

### 6. 참고문헌

[1] 박지연, 전범수, "네티즌의 흥행 영화 리뷰에 포함된 감정 동사 이용 특성 연구", 한국콘텐츠학회논문지, 2014.  
 [2] Ho-Seong Jeon, "Research on the Movie Reviews Regarded as Unsuccessful in Box Office Outcomes in Korea: Based on Big Data Posted on Naver Movie Portal", Kangwon National University Business and Economics Research Institute, September 2021.  
 [3] 나가은(Na, Ga Eun), 신민수(Shin, Min Soo), "평점 및 리뷰 분석을 통한 유료 모바일 게임 애플리케이션의 흥행 요인 연구", 한국IT서비스학회, 2018.  
 [4] 이재화(Jae Hwa Lee) , 김양석(Yang Sok Kim) , 이충권(Choong Kwon Lee), "영화리뷰의 분류성과 향상에 관한 연구", 한국IT서비스학

회, 2018.

[5] Tsutsumi, Kimitaka, Shimada, Kazutaka, Endo, Tsutomu, "Movie Review Classification Based on a Multiple Classifier", KISTI, 2007.

[6] Monisha V. N, Madhumitha N, Nimishamba B, Poornima N, Dr. Shashikumar D. R, "Sentimental Analysis of Movie Reviews using Twitter", IJERT, August 2020.