

# 빅데이터를 활용한 영화 흥행에 따른 리뷰길이 변화

## Changes in Review Length Based on the Popularity of Movies Using Big Data

조용희\*, 박이슬\*, 김혜진\*\*  
연세대학교 정보대학원\*, 연세대학교 근대한국학연구소\*\*

Yonghee Cho(yongfee@yonsei.ac.kr)\*, Yiseul Park(dewpark@yonsei.ac.kr)\*,  
Hea-Jin Kim(erin.hj.kim@yonsei.ac.kr)\*\*

### 요약

본 연구에서는 영화 관람 후 높은 평점을 매긴 집단과 낮은 평점을 매긴 집단 중 어느 집단이 영화에 대해 더 많은 이야기를 하는지, 즉 온라인 리뷰를 길게 작성하는지에 대해 알아보고자 하였다. 이를 위해 네이버 영화 API에서 제공하는 영화 평점과 리뷰 데이터를 수집하였고, 한국영화진흥위원회에서 제공하는 영화 손익분기점 데이터를 이용하여 영화를 흥행성공, 흥행부진, 흥행실패로 구분하여 영화 평점과 리뷰길이 간의 상관관계, 영화 개봉 전과 후, 흥행여부에 따른 리뷰길이의 특성, 마지막으로 영화 평점이 리뷰길이에 영향을 미치는가에 대한 회귀분석을 실시하여 제시하였다.

■ 중심어 : 빅데이터 | 영화 흥행 | 온라인 리뷰 |

### Abstract

The study aims to determine which groups leave longer(more active) online reviews(comments) on the film by separating groups, one that satisfied with the movie while the other group dissatisfied with the movie. The data used were rating scores and reviews(comments) from Naver Movie API, and break-even point data provided by Korea Film Commission. We analyzed the relationship between movie rating and review length, before and after movie opening, the characteristics of review length according to the box office, and whether the movie rating affects the review length.

■ keyword : Big Data | Popularity of Movies | Online Review |

## 1. 서론

영화는 경험제적 속성을 갖고 있기에 개인이 직접 소비하기 이전에 전문가 비평이나 인터넷으로 쉽게 접근할 수 있는 온라인 영화 리뷰를 활용하는 것이 일반적이다. 온라인 영화 리뷰는 영화 관람객들이 자신의 생각이나 의견 및 경험들을 자발적으로 온라인에서 표현하는

행위로 영화를 이미 관람한 이용자들이 해당 영화에 대해 평점과 후기를 공개함으로써 영화를 아직 관람하지 못한 잠재적 이용자들에게 영향을 미치게 된다[1][2]. 최근에는 인터넷과 통신기술의 발달로 이용자들이 온라인 영화 리뷰를 과거보다 더 쉽게 접근할 수 있게 되어 그 영향력은 점점 더 커지고 있다고 할 수 있다[3].

온라인 리뷰에 대한 유용성은 잠재 고객들에게 무수

\* 이 논문은 2017년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2017S1A6A3A01079581).

접수일자 : 2018년 03월 26일

수정일자 : 2018년 04월 13일

심사완료일 : 2018년 04월 16일

교신저자 : 김혜진, e-mail : erin.hj.kim@yonsei.ac.kr

히 많은 정보 가운데 유용한 정보만을 선별할 수 있도록 함으로써 비용을 감소시키고, 의사결정에 대해 확신을 하도록 하는 역할을 한다[4]. 네티즌 리뷰가 전문가들의 평론이나 평가 정도의 전문성은 확보하지 못할 수 있지만, 온라인 영화 리뷰의 특성을 분석하는 것은 유사한 특성을 공유하는 소비자들, 특히 온라인 영화 리뷰 작성자들의 성향을 파악할 수 있는 도구적 가치가 있다고 할 수 있다[5].

영화 리뷰 작성행위가 ‘자발성’ 또는 ‘적극성’으로 이루어진다고 가정하였을 때, 리뷰의 길이는 작성자의 ‘자발성’ 또는 ‘적극성’이 반영된 결과라고 할 수 있다. 왜냐하면 어휘가 많고 길이가 긴 글이 짧은 글에 비해 주제에 관한 내용 지식이 풍부하며 하고 싶은 말이 많다고 볼 수 있기 때문이다[6-8]. 또한 글의 길이가 길다는 것은 글을 쓰려고 했던 의지가 크고, 하고 싶은 말을 많이 만들어냈거나 찾았다는 의미로 해석할 수 있다[7][9].

기존의 영화 리뷰에 대한 분석은 주로 영화 평점(감성)과 흥행간의 관계를 다룬 연구들이 다수를 이루고 있다[10-12]. 본 연구에서는 영화 평점과 흥행간의 관계를 고려할 때 140자라는 제한된 리뷰의 길이를 변수로 활용하여 영화 평점(감성)과 리뷰길이와의 관계, 그리고 영화 관람 후 긍정 또는 부정 중 어느 감성이 리뷰를 더 적극적으로 작성하는 지를 영화의 흥행성공, 부진, 실패의 측면에서 분석하였다.

본 연구의 연구 질문은 다음과 같다.

영화 리뷰 작성자의 감성 즉, 영화평점과 영화 리뷰 길이는 관계가 있는가?

영화의 개봉 전과 후를 기준으로 영화의 흥행여부(성공, 부진, 실패)에 따라 평점 수와 리뷰길이의 변화가 있는가?

## II. 선행연구

### 1. 온라인 영화 리뷰

영화 온라인 리뷰에 대한 선행연구들은 대부분 영화 흥행에 영향을 미치는 변수를 고찰하여 검증하는 연구들이다. 또한 과거 한국 영화의 흥행성과 요인 연구들

에서 ‘개봉스크린 규모’를 제외하면 유의미한 설명력이 검증된 요인들이 많지 않다.

박승현, 장정현[13]은 흥행 결정요인으로 네티즌의 온라인 평점을 추가하여 개봉 스크린규모와 온라인 평점, 흥행성과의 영향관계를 제시하였다. 박승현, 송현주[14]는 온라인 평가 항목을 온라인 평가 빈도와 평균 평점으로 구별하여 온라인 구전 정보와 영화의 흥행성과의 관계를 분석하였다. 분석결과에 따르면, 개봉 이전 기간에 나타난 온라인 구전은 전체 흥행성과뿐만 아니라 주별 흥행성과에도 유의미한 영향력을 보여주는 것으로 나타났다.

리뷰의 양은 구전의도[15], 리뷰의 감성[3] 등과 관련이 있는데, 특히 Khare[16]는 리뷰 수를 조절변수로 활용하여 온라인 리뷰의 메시지 특성이나 소비자의 특성에 따라 온라인 리뷰의 설득효과가 달라질 수 있음을 규명하였다. 또한 온라인 리뷰의 수가 많고, 긍정적인 내용일수록 제품 판매나 소비자의 구매가 증가한다고 하였다. 하지만 Duan[17]과 최지은[18]의 연구결과에 의하면 온라인 리뷰의 수적 증가가 반드시 소비자 행동에 긍정적 영향을 미치는 것은 아니며, 오히려 온라인 리뷰가 적을 때, 온라인 리뷰의 영향력이 증가할 수도 있다. 이는 부정적인 리뷰가 긍정의 감성을 가지고 있는 리뷰보다 제품 또는 서비스의 판매에 더 큰 영향을 미치기 때문이다[19].

### 2. 온라인 영화 리뷰와 빅데이터

온라인 리뷰의 빅데이터는 온라인 리뷰의 평점, 댓글, 텍스트마이닝 기법을 이용한 감성 스코어 등이 포함된다. 김유영과 송민[11]은 리뷰의 감성분석을 통해 리뷰와 영화 흥행간의 연관성을 분석하였으며, 강지훈[10]은 평점, 댓글 수와 같은 관객 반응을 고려한 변수들을 이용하여 영화의 수요를 예측하고 실제 영화 흥행 성적과 비교하였다. 허민희[12]는 네이버의 영화 평점에 대한 의미 분석을 통해 영화에 대한 관객의 긍정 및 부정 표현의 발현 정도와 영화 흥행 간에 어떠한 상관성이 있는지 분석하였다.

리뷰 콘텐츠 자체를 활용한 연구로 Liu[20]는 고객 리뷰 데이터와 사용자 평가점수를 활용하여 요약보고서

를 생성하는 기법을 제안하였다. 국내 연구로는 김근형 [21]이 고객 리뷰 내 핵심개체들 사이의 연관성을 도출하는 방식으로 고객 리뷰를 요약 및 분석하는 기법을 제안한 연구가 있다.

### III. 연구 방법

#### 1. 분석대상의 선정

본 연구는 영화진흥위원회 사이트 기준으로 2017년 상반기에 개봉한 한국 영화 20편을 대상으로 하였다. 분석 기준인 영화의 흥행을 측정하는 방식을 정의함에 있어서 미국의 경우 박스 오피스(box office)라는 명칭으로 매출액을 성과의 기준으로 하거나 관객 수(admission)를 성과의 기준으로 할 수 있는데[22], 본 연구에서는 두 가지를 모두 채택하였다.

분석 기간은 2017년 1월부터 5월까지로, 이 기간 한국 영화는 총 165편을 개봉하였다. 그러나 전국 기준으로 관객 수 10만 명을 넘긴 영화는 22편으로 13%에 불과하였고, 이 중 100만 명을 넘긴 영화는 7편으로 전체 4%에 그쳤다. 관객 수 10만 명 이하의 영화들은 대부분 단편영화, 다큐멘터리, 독립영화 등 저예산 영화였다. 이러한 영화들은 본 연구에서 분석하고자 하는 영화 관람 후 만족 또는 불만족을 표현한 리뷰의 수가 적고, 분석 결과의 신뢰도와 타당도 또한 문제가 있을 것으로 판단되어 분석에서 제외하였다. 따라서 본 연구에서는 총 165개의 영화 중 관객 수가 10만 명 이상이었던 상업 영화 20편을 대상으로 분석을 진행하였다.

구체적인 분석을 위해 본 연구에서는 한국영화진흥위원회에서 제공하는 20개 영화의 총 관객 수와 영화사에서 발표하는 손익분기점을 기준으로 영화 별 손익분기점 대비 달성률을 산출하여[표 1] 영화의 흥행여부를 흥행성공, 흥행부진, 흥행실패로 정의하였다. 분석대상인 20개 영화의 흥행구분은 다음과 같다.

흥행성공: 달성률 100%를 넘긴 영화는 공조, 더킹, 재심, 해빙, 프리즌, 보안관으로 총 6편이다.

흥행부진: 달성률 40~100%이내의 영화로 조작된 도시, 아빠는 딸, 특별시민, 임금님의 사건수첩, 불한당 등

총 5편이다.

흥행실패: 달성률 40% 미만의 영화 9편이다(여교사, 그레 가족, 싱글라이더, 루시드드림, 비정규직 특수요원, 보통사람, 원라인, 어느날, 석조저택 살인사건).

선정된 20개 영화의 평점과 리뷰의 분석을 위해 국내에서 가장 방대한 영화 데이터베이스인 네이버 영화(movie.naver.com)에서 개봉 전과 개봉 후 데이터를 API를 통해 수집하였다. 자바(Java) 기반의 크롤러를 작성하여 개봉 전 11,141건과 개봉 후 133,454건, 총 144,595건의 데이터가 수집되었다.

표 1. 손익분기점 대비 달성률 기준 영화 분류

구분	제목	개봉 일	총 관객수 (천명)	손익분기점 (천명)	달성률
흥행성공 (6편)	공조	1/18	7,817	2,800	279%
	더킹	1/18	5,316	3,500	152%
	재심	2/15	2,421	1,600	151%
	해빙	3/1	1,204	1,160	104%
	프리즌	3/22	2,931	2,000	147%
	보안관	5/3	2,583	2,000	129%
흥행부진 (5편)	조작된 도시	2/9	2,514	3,000	84%
	아빠는 딸	4/12	646	1,500	43%
	특별시민	4/26	1,361	3,000	45%
	임금님의 사건수첩	4/26	1,633	3,000	54%
	불한당	5/17	902	2,200	41%
흥행실패 (9편)	여교사	1/4	116	700	17%
	그레 가족	2/15	46	700	7%
	싱글라이더	2/22	351	1,500	23%
	루시드드림	2/22	102	1,700	6%
	비정규직 특수요원	3/16	149	1,000	15%
	보통사람	3/23	383	2,000	19%
	원라인	3/29	434	1,500	29%
	어느날	4/5	234	1,000	23%
	석조저택 살인사건	5/9	350	1,900	18%

표 2. 영화 데이터 분석 예시 - 코딩

평점	리뷰길이	리뷰	날짜
10	22	굿굿 아주 너무너무 깊게 영화 잘봤습니다	2017.01.18
9	15	유해진 생활연구 꿀잼 ○人○	2017.02.12
6	23	현빈액션외에는 스토리고 뭐고 기억에 남는게 없음	2017.02.03
1	22	와아~ 윤제균 제작이다, 아주 균이네 균	2017.01.19.

네이버 영화의 경우 리뷰어(관람객)가 평점을 부여할

때 140자 이내의 제한된 리뷰를 작성해야하는 조건이 있다. 단순히 좋고 싫음만을 남기는 경우가 많지만, 평점을 낮게 혹은 높게 준 논리적 이유를 들어가며 140자가 가까이 채우는 사람들도 다수 존재하였다. [표 2]는 영화 '공조'에 대하여 수집된 리뷰의 예시이다.

2. 분석 방법

본 연구에서는 앞에서 언급한 각 영화의 리뷰 수, 평균 리뷰길이, 그리고 개별 리뷰의 작성날짜와 리뷰길이를 본 연구의 모델구축을 위한 변수로 사용하였으며 각 변수에 대한 정의는 다음과 같다.

- 리뷰 수: 평점 별 리뷰 수를 나타낸 값
- 리뷰길이: 평점 별 리뷰의 글자 수(공백포함)를 나타낸 값
- 평균길이: 리뷰길이를 리뷰 수로 나눈 값

표 3. 영화 20개 데이터 분석 - 평점, 길이, 평균길이

구분	총 리뷰 (수)	총 리뷰길이 (자)	평균 리뷰길이 (자)
전	11,141	403,881	36,252
후	133,454	5,135,667	38,483
총	144,595	5,539,548	38,311

수집한 개별 리뷰에 대하여 리뷰길이를 산출하였고, 이것을 리뷰 수로 나누어 각 영화에 대한 평균길이를 구하였다. 또한 영화의 개봉 전과 후를 기준으로 영화의 흥행여부(성공, 부진, 실패)에 따라 리뷰 수와 리뷰 길이가 어떻게 변화하는지를 분석하기 위하여 영화 개봉 전과 후를 기준으로 총 리뷰길이를 리뷰 수로 나누어 평균길이를 구하였다. 전체 영화 20개에 대한 총 리뷰 수, 총 리뷰길이, 평균 리뷰길이는 [표 3]과 같다.

마지막으로 연구문제에 대한 결과를 도출하기 위하여 다음의 총 3단계의 분석을 진행하였다.

첫 번째는 영화 평점과 리뷰길이 간 상관관계를 알아보기 위하여 Pearson의 이변량 상관분석을 실시하였다.

두 번째는 20개 영화를 개봉 전, 후 및 흥행 여부 세 가지로 조합하여 각 그룹의 리뷰길이의 특성을 비교 분석하여 제시하였다.

마지막으로 영화 평점이 리뷰길이에 미치는 영향이 유의한지 검증하기 위하여 영화 평점을 독립변수로 설

정하여 단순회귀분석을 실시하였다.

IV. 분석 결과

1. 영화 평점과 리뷰길이와의 상관관계

평점과 리뷰길이 간 상관관계를 알아보기 위하여 Pearson의 이변량 상관분석을 실시하였다. 분석의 편의성을 위해 개봉 전 데이터는 'BF', 개봉 후 데이터는 'AF'로 정의하였다. [표 4]는 개봉 전(BF)과 개봉 후(AF) 영화의 평점과 길이에 대한 상관분석의 결과이다. 개봉 전에는 흥행성공, 흥행부진, 흥행실패 모두 통계적으로 유의한(p<.05) 상관관계를 발견하지 못하였다[표 4(a)-(c)]. 그러나 개봉 후의 결과는 흥행성공의 경우 평점과 리뷰길이의 상관관계가 r=-0.940, p=0.000으로 강한 부(-)의 상관관계를 보였다[표 4(d)].

표 4. 영화 평점과 리뷰길이와의 상관관계

(a) 흥행성공 BF 상관관계				(b) 흥행부진 BF 상관관계			
		평점	길이			평점	길이
평점	Pearson 상관		1	평점	Pearson 상관		1
	유의확률 (양측)				유의확률 (양측)		
	N	10			N	10	
길이	Pearson 상관	.385	1	길이	Pearson 상관	.243	1
	유의확률 (양측)	.272			유의확률 (양측)	.498	
	N	10	10		N	10	10
(c) 흥행실패 BF 상관관계				(d) 흥행성공 AF 상관관계			
		평점	길이			평점	길이
평점	Pearson 상관		1	평점	Pearson 상관		1
	유의확률 (양측)				유의확률 (양측)		
	N	10			N	10	
길이	Pearson 상관	-.071	1	길이	Pearson 상관	-.940*	1
	유의확률 (양측)	.846			유의확률 (양측)	.000	
	N	10	10		N	10	10
(e) 흥행부진 AF 상관관계				(f) 흥행실패 AF 상관관계			
		평점	길이			평점	길이
평점	Pearson 상관		1	평점	Pearson 상관		1
	유의확률 (양측)				유의확률 (양측)		
	N	10			N	10	
길이	Pearson 상관	-.779*	1	길이	Pearson 상관	-.575	1
	유의확률 (양측)	.008			유의확률 (양측)	.082	
	N	10	10		N	10	10

\*\* 상관계수가 0.01 수준에서 유의합니다(양측).

\*\* 상관계수가 0.01 수준에서 유의합니다(양측).

또한 흥행부진 영화의 경우도 상관관계가  $r=-0.779$ ,  $p=0.008$ 로 강한 부(-)의 상관관계를 보이고 있으나[표 4(e)] 흥행실패 영화의 경우  $r=-0.575$ ,  $p=0.082$ 로 통계적으로 유의한 상관관계가 없는 것으로 나타났다[표 4(f)].

## 2. 개봉 전 · 후 영화 리뷰길이의 특성

흥행성공, 흥행부진, 흥행실패 영화들의 개봉 전, 개봉 후의 평균길이의 변화는 [그림 1]과 같다.

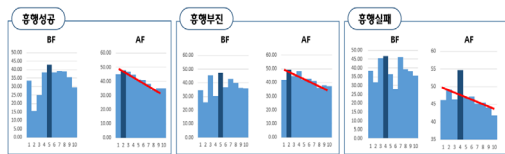


그림 1. 흥행별 개봉 전 · 후 리뷰길이

비교 분석 결과 영화 개봉 후에는 흥행성공과 흥행부진 영화에서는 2점에서 평균길이가 가장 많이 나타났으며, 강하게 우하향하는 형태의 그래프를 보인다. 이는 영화에 불만족한 사람들이 평균적으로 이야기를 더 많이 하고, 만족한 사람들은 상대적으로 적게 이야기했고 해석할 수 있으며 이 결과는 영화 평점과 리뷰길이의 상관관계 분석결과도 일치한다. 개봉 후 흥행실패에서도 약하게 우하향하는 형태가 나타났으나, 개봉 전

(BF)에서는 세 집단 모두 특별한 형태를 나타내지 않았다.

[그림 2]는 20개 영화의 BF, AF 데이터, 총 40개의 평균길이를 흥행성공, 흥행부진, 흥행실패로 산출한 결과이다. 흥행성공과 흥행부진의 AF 데이터를 살펴보면 평균길이가 대체로 낮은 평점에서 많고, 높은 평점에서 완만히 우하향하는 형태의 그래프 결과가 공통적으로 나타났다. 그러나 흥행실패 AF와 흥행성공, 부진, 실패 영화의 BF에서는 이러한 형태가 나타나지 않았다.

흥행성공 영화 6편의 BF와 AF 각각의 평균을 비교하여보면[그림 3], 흥행성공 BF에서 평균길이는 43글자로 평점 5점에서 가장 높고, 흥행성공 AF에서는 48.3글자로 평점 2점에서 가장 높다. AF의 그래프를 살펴보면 프리즌을 제외한 5개 영화와 평균 AF에서는 평점 3점 이하에서 가장 많은 평균길이가 있었고 우하향하는 흐름이 나타났다. BF에서는 ‘흥행부진’과 마찬가지로 평점과 그래프에서 공통적인 특성을 발견할 수 없었다.

흥행부진 영화 5편의 BF와 AF 각각의 평균을 비교하여보면[그림 4], 흥행부진 BF에서 평균길이는 48.7글자로 평점 5점에서 가장 높고, 흥행부진 AF에서는 49.7글자로 평점 2점에서 가장 높았다. AF의 그래프를 살펴보면 불한당을 제외한 4개 영화와 평균 AF에서는 평점 3점 이하에서 가장 많은 평균길이가 있었고 우하향하는 흐름이 나타났다. BF에서는 평점과 그래프에서 공통적인 특성을 발견할 수 없었다.

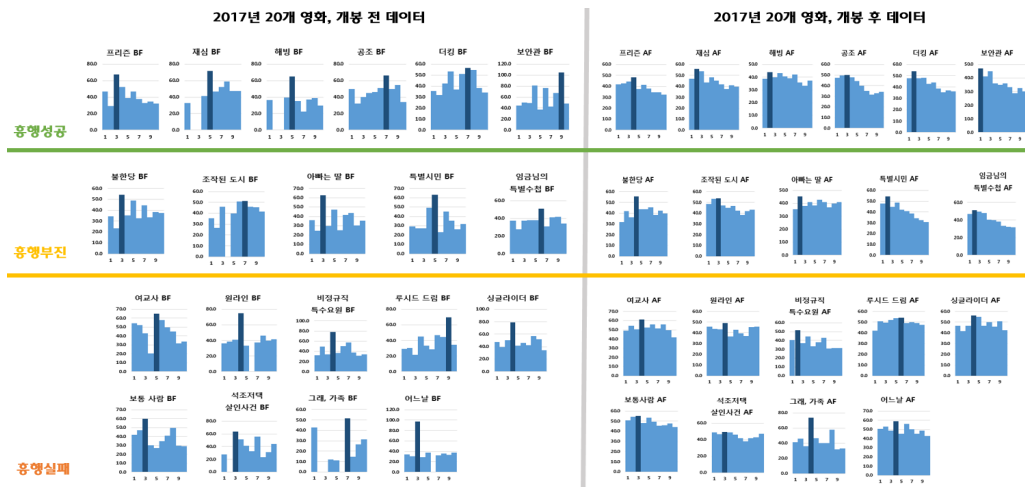


그림 2. 개봉 전 · 후 영화 리뷰길이 분석

마지막으로 흥행실패 영화 9편의 'BF', 'AF'에서도 공통적인 특성을 발견하지 못하였다.

### 3. 영화 평점이 리뷰길이에 미치는 영향

영화 평점이 길이에 미치는 영향이 유의한지 검증하기 위하여 영화 평점을 독립변수로 설정하여 단순회귀 분석을 실시하였다[표 5]. 개봉 전(BF)에는 흥행성공 영화의 경우  $R^2$ 는 0.148로 14.8%의 낮은 설명력을 지니며, p값은 0.272로 통계적으로 유의적인 영향을 미치지 못하는 것으로 나타났다[표 5(a)]. 흥행부진 영화와 흥행실패 영화의 경우에도  $R^2$ 는 0.059과 0.005로 5.9%, 0.5%의 매우 낮은 설명력을 지니며, p값 또한 0.498과

0.846으로 통계적으로 유의적인 영향을 미치지 못하는 것으로 나타났다[표 5(b)-(c)].

개봉 후(AF)에는 흥행성공 영화의 경우  $R^2$ 는 0.883으로 88.3%의 높은 설명력을 지니고 있으며, p값은 0.000으로  $p < 0.01$ 에서 통계적으로 유의적인 영향을 미치는 것으로 나타났다[표 5(d)]. 흥행부진 영화의 경우  $R^2$ 는 0.607로 60.7%의 설명력을 지니고 있으며, p값은 0.008로  $p < 0.01$ 에서 통계적으로 유의적인 영향을 미치는 것으로 나타났다[표 5(e)]. 한편, 흥행실패 영화의 경우  $R^2$ 는 0.331로 33.1%의 낮은 설명력을 지니고 있으며, p값은 0.082로 통계적으로 유의적인 영향을 미치지 못하는 것으로 나타났다[표 5(f)].

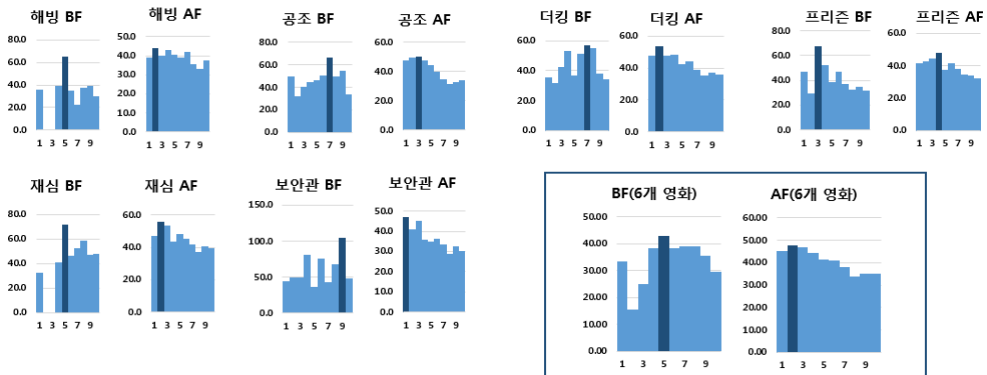


그림 3. 흥행성공 영화 평균 리뷰길이 분석

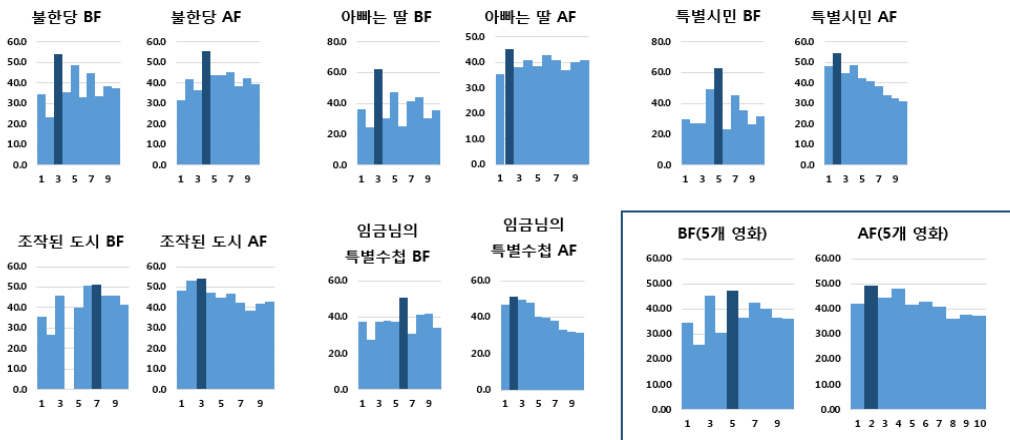


그림 4. 흥행부진 영화 평균 리뷰길이

표 5. 영화 평점과 리뷰길이 회귀분석 결과

(a) 흥행성공 BF 회귀분석

모형	R	R 제곱	수정된 R 제곱	추정값의 표준오차	Durbin-Watson
1	.385a	.148	.042	2,963	.337

(b) 흥행부진 BF 회귀분석

모형	R	R 제곱	수정된 R 제곱	추정값의 표준오차	Durbin-Watson
1	.243a	.059	-.058	3,115	.295

(c) 흥행실패 BF 회귀분석

모형	R	R 제곱	수정된 R 제곱	추정값의 표준오차	Durbin-Watson
1	.071a	.005	-.119	3,203	.119

(d) 흥행성공 AF 회귀분석

모형	R	R 제곱	수정된 R 제곱	추정값의 표준오차	Durbin-Watson
1	.940a	.883	.868	1,098	1,403

(e) 흥행성공 AF 회귀분석

모형	R	R 제곱	수정된 R 제곱	추정값의 표준오차	Durbin-Watson
1	.779a	.607	.557	2,014	1,514

(f) 흥행성공 AF 회귀분석

모형	R	R 제곱	수정된 R 제곱	추정값의 표준오차	Durbin-Watson
1	.575a	.331	.247	2,627	.831

a. 예측자: (상수), 길이  
 b. 종속변수: 평점

데이터 분석 결과와 통계적 검증 결과를 종합하여 보면, 영화 개봉 후 관람한 소비자들의 경우, 영화에 대해 불만족스러운 경우(낮은 평점)가 만족스러운 경우(높은 평점)보다 상대적으로 영화에 대해 더 많은 의견을 제시하는 것으로 나타났다. 이를 해석해보면, 영화를 재미있게 본 사람들은 영화에 대한 리뷰를 쓸 때, 좋은 평점을 남기고 왜 이 영화가 좋은지에 대해 길게 남기지 않는 반면, 영화에 대해 불만족한 사람들은 리뷰를 쓸 때, 이 영화가 왜 재미가 없는지에 대해 더 구체적으로 설명하며 비교적 긴 리뷰를 작성한다고 볼 수 있다.

## V. 결론 및 의의와 한계

본 연구에서는 영화에 만족하여 높은 평점을 매긴 집단 혹은 영화에 불만족하여 낮은 평점을 매긴 집단 중

어느 집단이 영화에 대해 더 많은 이야기를 하는지, 즉 리뷰를 길게 작성하는지에 대해 알아보려 하였다. 이를 위해 2017년 5월까지 개봉한 20개 영화의 총 123,014 건(개봉 전: 11,141건 / 개봉 후: 111,873건) 데이터를 수집하였고, 평점과 리뷰길이를 분석 데이터로 활용하였다. 영화 간 분류를 위해 영화의 손익분기점 달성률을 기준으로 흥행여부를 흥행성공, 흥행부진, 흥행실패 세 그룹으로 분류하였고 추가적으로 개봉 전과 후로 구분하여 총 여섯 개의 카테고리 영화를 그룹화하여 분석을 진행하였다.

데이터 분석 결과, 개봉 전 흥행성공과 흥행부진 영화의 평균 리뷰길이는 5점을 부여한 집단에서 가장 높게 집계되었다. 흥행실패 영화의 경우 4점을 부여한 집단에서 가장 높게 집계되었고 그래프 상에서 유의미한 결과를 찾지 못하였다. 또한 상관분석, 회귀분석에서도 통계적으로 유의미하지 않은 것으로 분석되었다.

개봉 후 흥행성공과 흥행부진 영화의 평균 리뷰길이는 2점을 부여한 집단에서 가장 높게 집계되었으며 낮은 평점에서 많고 높은 평점에서 적은 우하향 형태를 나타내었다. 하지만 흥행실패 영화의 경우 4점을 부여한 집단에서 가장 높게 집계되었고 약한 우하향 형태를 보였지만 상관분석과 회귀분석 결과 유의미하지 않은 것으로 분석되었다.

본 연구의 의의는 영화 리뷰길이에 대한 기존의 연구가 미비한 상황에서 영화 관람 후 소비자의 만족 혹은 불만족이 리뷰 작성에 어떠한 영향을 미치는가에 대해 데이터기반 분석을 시도하여 영화 개봉 후의 결과에서 평점과 리뷰길이와의 상관관계를 발견한 것이다.

본 연구는 2017년 1월부터 5월까지의 짧은 기간 동안 2017년 개봉영화 20편의 데이터를 수집·분석하였다. 따라서 제한된 표본 수로 인한 결론의 타당성 또는 일반화의 한계가 존재할 수 있다. 이를 극복하기 위하여 향후 연구에서는 더 많은 개봉영화를 대상으로 본 연구의 방법론을 적용하여 분석하여 봄으로써 연구결과의 일반화가 가능한지 가능해보는 시도가 필요할 것이다. 또한 후속 연구로 영화 평점 외의 추가적인 요인들, 예를 들면 리뷰 내용 속 출현 단어나 영화의 장르, 출연진, 제작진 등 소비자의 선호도 등의 요인을 추가하여 리뷰 길이

와 영화 흥행의 상관관계의 통제 변수로 활용하는 것도 유의미한 결과를 도출해 낼 수 있으리라 예상된다.

### 참고 문헌

- [1] 박지연, 전범수, “네티즌의 흥행 영화 리뷰에 포함된 감정 동사 이용 특성 연구,” 한국콘텐츠학회논문지, 제14권, 제5호, pp.85-94, 2014.
- [2] A. Chakravarty, Y. Liu, and T. Mazumdar, “The Differential Effects of Online Word-of-Mouth and Critics’ Reviews on Pre-Release Movie Evaluation,” *Journal of Interactive Marketing*, Vol.24, No.3, pp.185-197, 2009.
- [3] W. Duan, B. Gu, and A. B. Whinston, “Do online reviews matter?—An empirical investigation of panel data,” *Decision support systems*, Vol.45, No.4, pp.1007-1016, 2008.
- [4] S. W. Sussman and W. S. Siegal, “Informational Influence in Organizations: An Integrated Approach to Knowledge Adoption,” *Information Systems Research*, Vol.14, No.1, pp.47-65, 2003.
- [5] M. Clement, D. Proppe, and A. Rott, “Do critics make bestsellers? Opinion leaders and the success of books,” *Journal of Media Economics*, Vol.20, No.2, pp.77-105, 2007.
- [6] J. Chesky and E. H. Hiebert, “The effect of prior knowledge and audience on high school students’ writing,” *Journal of Educational Research*, Vol.80, No.5, pp.304-313, 1987.
- [7] 박현철, *정서 표현적 쓰기에서 글의 길이와 질(質)의 관계*, 고려대학교 대학원, 석사학위논문, 2009.
- [8] 서수현, “글쓰기에서의 내용 지식에 대한 개념 규정,” *국어교육*, 제121권, pp.107-128, 2006.
- [9] S. Brenner, *Motivation to write in grades three to six*, Unpublished Dissertation, University of Toronto, Canada, 2005.
- [10] 강지훈, 박찬희, 도형록, 김성범, “데이터마이닝 기법을 활용한 영화 흥행 실적 예측 기법,” 대한산업공학회 춘계학술대회 논문집, pp.142-156, 2014.
- [11] 김유영, 송민, “영화 리뷰 감성분석을 위한 텍스트 마이닝 기반 감성 분류기 구축,” *지능정보연구*, 제22권, 제3호, pp.71-89, 2016.
- [12] 허민희, 강필성, 조성준, “오피니언 마이닝을 이용한 영화 흥행의 예측,” *한국경영과학회 춘계공동학술대회 논문집*, Vol.2013, No.5, pp.487-500, 2013.
- [13] 박승현, 장정현, “온라인 영화 리뷰의 내용과 품질에 관한 탐색적 연구,” *언론과학연구*, 제12권, 제4호, pp.221-256, 2012.
- [14] 박승현, 송현주, “영화의 주별 흥행성과에 미치는 영향,” *한국언론학보*, 제56권 제4호, pp.210-235, 2012.
- [15] 여등승, 임규건, “온라인 영화관련 댓글의 양과 방향성이 구전 의도에 미치는 영향에 관한 연구,” *한국 IT 서비스학회 학술대회 논문집*, Vol.2014, pp.429-432, 2014.
- [16] K. Adwait, Lauren I. Labrecque, and Anthony K. Asare, “The Assimilative and Contrastive Effects of Word-of-Mouth Volume: An Experimental Examination of Online Consumer Ratings,” *Journal of Retailing*, Vol.87, No.1, pp.111-126, 2011.
- [17] D. Wenjing, B. Gu, and A. B. Whinston, “The Dynamics of Online Word-of-Mouth and Product Sales—An Empirical Investigation of the Movie Industry,” *Journal of Retailing*, Vol.84, No.2, pp.233-242, 2008.
- [18] 최지은, “항상 온라인 리뷰수가 많을수록 좋은가? : 영화 평점에 대한 소비자반응을 중심으로,” *광고학연구*, 제24권, 제7호, pp.87-103, 2013.
- [19] C. H. Wee, S. L. Lim, and M. Lewin, “Words-of-mouth communication in Singapore: With focus on effects of message-sidedness, source and user-type,” *Asia Pacific Journal of*



- Marketing and Logistics, Vol.1, No.2, pp.5-36, 1995.
- [20] B. Liu, "Sentiment analysis and opinion mining," Synthesis Lectures on Human Language Technologies, Vol.5, No.1, pp.1-167, 2012.
- [21] 김근형, 오성열, "온라인 고객리뷰 분석을 통한 시장세분화에 텍스트마이닝 기술을 적용하기 위한 방법론," 한국콘텐츠학회논문지, Vol.9, No.8, pp.272-284, 2009.
- [22] 장병희, *영화 흥행 요인*, 커뮤니케이션북스, 2015.

저 자 소 개

조 용 희(Yonghee Cho)

준회원



- 2016년 3월 ~ 현재 : 연세대학교 정보대학원 석사과정

<관심분야> : 빅데이터, 텍스트마이닝, 디지털콘텐츠

박 이 슬(Yiseul Park)

준회원



- 2017년 3월 ~ 현재 : 연세대학교 정보대학원 석사과정

<관심분야> : 빅데이터, 텍스트마이닝, 디지털콘텐츠

김 혜 진(Hea-Jin Kim)

정회원



- 2002년 8월 : 연세대학교 문헌정보학과(문헌정보학석사)
- 2015년 8월 : 연세대학교 문헌정보학과(문학박사)
- 2018년 2월 ~ 현재 : 연세대학교 근대한국학연구소 HK연구교수

<관심분야> : 빅데이터, 텍스트마이닝, 감성분석