

# 홍콩 영화에 관한 고객 리뷰의 텍스트 마이닝 기반 분석: 관객 선호도의 진화 발견

손화양\* · 이정승\*\*

## Text Mining-Based Analysis of Customer Reviews in Hong Kong Cinema: Uncovering the Evolution of Audience Preferences

Huayang Sun\* · Jung Seung Lee\*\*

### Abstract

This study conducted sentiment analysis on Hong Kong cinema from two distinct eras, pre-2000 and post-2000, examining audience preferences by comparing keywords from movie reviews. Before 2000, positive keywords like 'actors,' 'performance,' and 'atmosphere' revealed the importance of actors' popularity and their performances, while negative keywords such as 'forced' and 'violence' pointed out narrative issues. In contrast, post-2000 cinema emphasized keywords like 'scale,' 'drama,' and 'Yang Yang,' highlighting production scale and engaging narratives as key factors. Negative keywords included 'story,' 'cheesy,' 'acting,' and 'budget,' indicating challenges in storytelling and content quality. Word2Vec analysis further highlighted differences in acting quality and emotional engagement. Pre-2000 cinema focused on 'elegance' and 'excellence' in acting, while post-2000 cinema leaned towards 'tediousness' and 'awkwardness.' In summary, this research underscores the importance of actors, storytelling, and audience empathy in Hong Kong cinema's success. The industry has evolved, with a shift from actors to production quality. These findings have implications for the broader Chinese film industry, emphasizing the need for engaging narratives and quality acting to thrive in evolving cinematic landscapes.

Keywords : Hong Kong Cinema, Text Mining, Sentiment Analysis, Customer Reviews, Audience Preference

Received : 2023. 02. 10.    Revised : 2023. 08. 23.    Final Acceptance : 2023. 08. 26.

\* First Author, Department of Health Management, Jinzhou Medical University, China, e-mail : 1207282652@qq.com

\*\* Corresponding Author, School of Business, Hoseo University, 12, Hoseodae-gil, Dongnam-gu, Cheonan-si, Chungcheongnam-do, 31066, Korea, 82-41-560-8361, e-mail : jslee@hoseo.edu

## 1. 서 론

중국 영화 산업은 최근 급속한 발전을 이루어내며 중국 영화업계가 대중의 시선을 사로잡고 있으며, 더 많은 사람이 중국 영화를 즐기며 자신의 자유시간을 보내는 추세를 보인다[Ahn, 2021]. Chen et al.[2022]의 조사 결과에 따르면 2016년부터 2021년까지 중국 영화 총매출은 455.1억 원에서 472억 원으로 증가하며, 중국 영화 산업의 성장을 확인할 수 있다. 이러한 성장과 함께 인터넷의 발전은 사용자들에게 공공 플랫폼에서의 의견, 태도, 견해, 감정을 표현할 기회를 제공하고, 이는 영화 분야의 대중 플랫폼에서도 영화 감상과 관련된 리뷰 및 평가의 활발한 교환을 끌어냈다.

1997년 홍콩의 중국 반환과 2000년의 중국 영화 관련 제도의 개혁과 함께 홍콩 영화는 대중 시장에 주입되며 중국 영화 산업의 일부로 자리매김하게 되었다. 그러나 홍콩 영화업계는 불경기, 혁신 정신의 결여, 스타의 중국 진출 등으로 인해 어려움을 겪고 있으며, 이러한 어려움을 극복하려는 홍콩 영화인이 중국 본토로 진출하여 새로운 발전 기회를 모색하고 있다[Lardy, 2014]. 따라서 홍콩 영화의 변화와 중국 영화 산업의 발전 간의 상관관계를 이해하고자 하는 필요성이 존재한다.

본 연구의 목적은 텍스트마이닝 및 감성분석 기술을 활용하여 1997년 중국 반환을 기준으로 3년간의 홍콩 영화와 2000년 이후의 홍콩 영화에 대한 감정 분석을 수행하여 두 시기의 영화 감상자들의 태도를 이해하고, 이를 통해 영화 제작진에게 영화 제작 방법에 대한 참고 자료를 제공함으로써 홍콩 영화와 중국 영화 산업의 발전을 지원하는 것이다.

본 연구의 범위는 1997년 중국 반환 이후부터 2000년 이후 3년 동안의 홍콩 영화를 대상으로 한다. 이를 위해 네이버 영화 사이트에서 많은 감상평이 작성된 홍콩 영화 중 공감순으로 상위 10개의 영화를 조사 대상으로 선정하였다. 따라서 본 연구는 두 시기의 홍콩 영화에 대한 감정 분석을 수행하고자 하며, 연구 결과를 통해 중국 영화 산업의 발전을 지원하는 인사이트를 제공하고자 한다.

## 2. 이론적 배경

### 2.1 텍스트 마이닝

텍스트 마이닝(text mining)은 비정형 데이터에 대

한 과학적 분석 절차로, 데이터로부터 통계적으로 의미 있는 개념이나 특성을 추출하고 이들 간의 패턴 및 추세와 같은 고품질의 정보를 추출하는 과정을 말한다. 텍스트 마이닝은 데이터의 형태에 따라 정형데이터와 비정형 데이터로 나눌 수 있으며, 주로 정형데이터를 다루는 데이터 마이닝과 비정형 데이터를 다루는 텍스트 마이닝으로 나눌 수 있다[Fay, 2018]. 텍스트 마이닝은 주로 텍스트 범주화, 텍스트 클러스터링, 개념 및 특성 추출, 개념 간 및 특성 간의 관계 예측 등의 작업을 수행하며, 이를 통해 비정형 데이터에 포함된 중요한 정보를 발견하고 분석한다. 이러한 기술은 다양한 분야에서 응용되며, 특히 텍스트 데이터의 부가적 가치를 끌어내기 위해 효과적으로 사용된다.

TF-IDF(Term Frequency-Inverse Document Frequency) 계산법은 정보검색 및 텍스트 분석 분야에서 널리 사용되는 방법의 하나이다[Qaiser and Ali, 2018]. 이 계산법의 목표는 단어가 특정 문서 내에서 얼마나 중요한지를 측정하는 것이다. TF-IDF는 두 가지 주요 구성 요소로 이루어져 있다. 단어의 사용 빈도(TF)는 어떤 특정한 단어가 문장 또는 문서 내에서 얼마나 자주 나타나는지를 나타낸다. 이것은 해당 단어의 상대적인 중요성을 나타낸다. 역문서 빈도(IDF)는 해당 단어가 다른 문서에서 얼마나 희귀하게 나타나는지를 나타낸다. 이것은 단어가 문서 집합 내에서 얼마나 중요한지를 나타낸다. 이러한 계산법을 통해 각 문서 내에서 특정 단어의 중요성을 측정하고, 이를 기반으로 검색 및 분류 작업을 수행할 수 있다. TF-IDF는 정보검색, 문서 분류, 텍스트 요약 등 다양한 텍스트 분석 작업에 활용된다.

Word2vec은 자연어 처리(Natural Language Processing, NLP) 분야에서 널리 사용되는 기술 중 하나로, 대량의 텍스트 데이터로부터 단어 간의 의미적 관계를 학습하고 단어를 벡터로 표현하는 데 사용된다[Church, 2017]. Word2vec은 신경망 모델을 기반으로 하며, 단어의 의미를 수치화하여 단어 간 유사성을 계산할 수 있게 해준다. Word2vec 모델은 단어 간의 연관성을 파악하고, 단어를 벡터 형태로 표현한다. 이를 통해 단어 간의 의미 유사성을 계산할 수 있으며, 벡터 연산을 통해 의미상으로 유사한 단어를 찾거나 단어 간의 의미적 관계를 분석할 수 있다. 이러한 기술은 문서 분류, 문서 군집화, 텍스트 생성, 기계 번역 등 다양한 자연어 처리 작업에 활용된다.

## 2.2 영화 리뷰 분석

영화 리뷰 분석은 텍스트마이닝을 통해 수행되며, 다양한 연구자들이 이를 다양한 관점에서 탐구하고 있다 [Xu et al., 2020]. Yessenov and Misailovic (2009)은 TF-IDF와 LDA 토픽 모델링을 활용하여 영화 리뷰에서 긍정적인 측면인 연기력, 연출력, 소재의 가치를 강조하고 부정적인 측면으로는 내용표절을 지적하였다. 또한, Rahman and Hossen(2019)은 TF-IDF, Word2vec, 그리고 LDA 토픽 모델링을 사용하여 특수효과와 배우의 감정 및 추억을 긍정적으로 평가하고, 결말 및 악역 파에 대한 불만을 부정적으로 언급하였다. 또한, Mcauliffe and Blei(2007)의 연구에서는 LDA 토픽 모델링, TF-IDF, Word2vec을 활용하여 주제 적극성과 연기를 긍정적으로 평가하고, 줄거리의 심심함과 진실성 부족을 부정적으로 언급하였다. 더불어, Lin and He(2009)은 LDA 토픽 모델링을 통해 연기력이 뛰어나며 감동적인 스토리를 긍정적으로 평가하고, 소재가 낡고 혁신적인 측면이 부족하다는 부정적인 의견을 나타냈다.

이와 같은 연구들은 텍스트마이닝 기법을 활용하여 영화 리뷰 도메인에서 다양한 측면을 분석하고 있으며, 영화 리뷰에 대한 이해를 높이기 위한 중요한 연구들로 평가된다. 이러한 연구들은 영화 산업 및 영화 관련 의사결정에 도움을 주는 데 이바지하며, 자연어 처리와 텍스트 분석 분야에서의 중요한 응용 분야 중 하나이다.

## 3. 분석 방법론

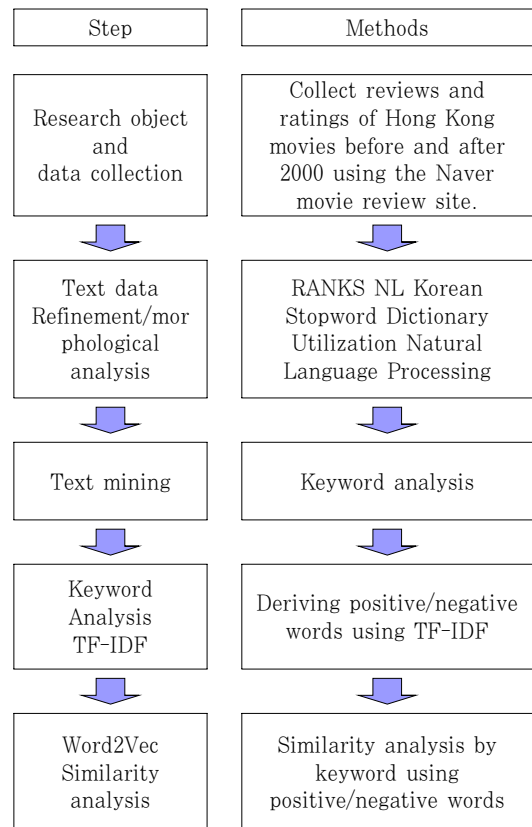
### 3.1 분석 절차

본 연구는 주로 네이버 웹사이트에서 두 시기의 홍콩 영화에 대한 리뷰 내용을 추출하여 TF-IDF 알고리즘과 Word2vec 알고리즘을 통해 리뷰 텍스트 키워드 분석과 키워드의 유사도 분석을 진행하여 관객들의 감정적 태도를 찾아내고, 마지막으로 두 시기의 홍콩 영화의 차이를 분석하였다.

〈Table 1〉과 같이 데이터 수집을 위해 두 시기의 홍콩 영화 리뷰를 대상으로 네이버에 접속해 네이버 영화리뷰 사이트(<https://movie.naver.com>)에 접속해 2000년 이전과 이후의 홍콩 영화 리뷰 데이터를 수집

했으며, 1시기의 홍콩 영화 리뷰는 1976년부터 1997년까지 총 4만 3,000건을 대상으로 했다. 2시기의 홍콩 영화는 2001년부터 2020년까지 리뷰 수가 총 38,000건이었다. 이 중 두 시기의 영화를 각각 만 편씩 선정하여 분석 텍스트로 삼았다. 리뷰 내용, 평점을 합산해 최저 1점부터 최고 10점까지 점수를 매겼다.

〈Table 1〉 Research methods and composition



데이터 전처리를 위해 데이터 획득 후 중복 평가, 무의미한 리뷰, 빈 값으로 평가된 리뷰는 분석에 의미가 없으므로 삭제했다. 마지막으로, RANKS NL에서 제공해주는 한국어 불용어 사전을 이용하여 감정 분석에 불필요한 불용어를 제거하며, 텍스트에는 (나, 의, 너, 및...) 빈도가 높지만, 실제 의미가 별로 없는 단어로 주로 어조 조사, 부사, 전치사 등의 불용어 단어 표와 보충 어휘를 사용하여 이러한 고빈도 무의미한 어휘를 제거했다.

### 3.2 텍스트 마이닝 방법론

우선 두 시기의 홍콩 영화 리뷰 데이터를 TF-IDF에 도입하고 10점 별점을 활용하여, 10점 점수제를 5점 평점제로 변환하여 기술지표 1~5점은 비추천, 6-10점은 추천으로 한다[Jeon, 2022] 6-10점은 1점으로 1~5점은 0점으로 변환하여 1점의 경우 긍정, 0점의 경우 부정으로 정하고, 빈도 분석 기법인 TFIDF를 적용하여 빈도수와 긍/부정 평가를 매칭하여 분석했다.

TF-IDF에서 분석한 키워드 결과를 Word2vec 처리기술을 이용해 분석하여 TF-IDF에서 긍정 및 부정 관련된 키워드와 유사도를 도출하고, 이에 따라 사용자 코멘트의 감정 주제를 분석하여 두 시기 주제의 차이를 비교했다.

## 4. 분석 결과

### 4.1 TF-IDF 분석 결과

2000년 이전에 조사된 10,000편의 영화리뷰 중 1점부터 5점까지 0점의 경우 부정 총 1,080개로 전체의 10.8%, 6점부터 10점까지 1점의 경우 긍정 총 8,920개로 전체의 89.2%를 차지했다. 2000년 이후에 조사된 만 개의 영화리뷰 중 1점부터 5점까지 총 3,387개로 전체의 33.87%, 6점부터 10점까지 총 6,913개로 전체의 69.13%를 차지했다.

TF-IDF를 이용하여 빈도 분석을 진행했으며, 분석에 불필요한 명사를 제거하고, okt 형태소 분석기를 이용하여 형태소를 분석하였으며, 불용어 사전을 만들어 불용어를 제거하였고, 단어의 빈도와 역문서 빈도를 사용하여 단어마다 중요한 정도를 가중치에 부여하여 단어의 빈도를 계산하였다.

2000년 이전의 홍콩 영화 리뷰에서 긍정 키워드와 부정 키워드는 크게 몇 가지 주제로 나눌 수 있다. 다음 <Table 2>와 같이 긍정 키워드에서 양조위, 주윤발 등 명배우(배우), 액션 코미디 중 명연기(연기), 홍콩 영화 등 사회적 지위(결작)와 긴장감 넘치는 분위기(분위기) 등으로 관객들의 사랑을 받고 있지만 부정 리뷰에서는 주로 영화의 구성 억지(구성), 내용 설정과 공감(공감) 부족, 배우에 대한 미화 폭력 감이

(폭력) 많은 등 관객의 부정을 받고 있다.

<Table 2> 20 Keywords Extracted and Positive and Negative Words before 2000

Order	Positive words (similarity)	Negative words (similarity)
1	주윤발(1.58)	실망(3.4)
2	양조위(1.49)	별로(3.2)
3	연기(1.49)	감시자(3.2)
4	분위기(1.31)	최악(2.35)
5	코믹(1.24)	억지(2.07)
6	액션(1.20)	설정(2.05)
7	형님(1.15)	대가(1.63)
8	결작(1.14)	테러(1.53)
9	성룡(1.04)	공감(1.43)
10	최고(0.99)	가오(1.46)
11	스토리(0.99)	적당(1.40)
12	배우(0.96)	기대(1.38)
13	명작(0.91)	캐릭터(1.34)
14	부활(0.91)	미화(1.34)
15	형제(0.86)	쓰레기(1.17)
16	김장(0.86)	악역(1.12)
17	영웅본색(0.85)	과거(1.11)
18	전설(0.82)	주연(1.09)
19	홍콩 영화(0.81)	알바(1.08)
20	대사(0.81)	수가(1.07)

2000년 이후의 영화 평론에서 긍정적인 상황은 다음 <Table 3>과 같다. 주로 영화의 대규모 할리우드(규모), 2000년 이전에 영화들 보다 청춘 요소들이 첨가되어 드라마처럼 청춘, 첫사랑 시절 등 좋은 추억(드라마), 그리고 미모가 뛰어난 배우(양양) 등으로 관객들의 호평을 받고 있으나 부정적 리뷰에서는 대부분 스토리 표절, 스토리 지루함 애니메이션처럼 유치함(유치), 연기 대사 능력 부족(연기), 제작비용 과다(제작비) 등으로 관객의 부정을 받고 있다.

〈Table 3〉 20 Keywords Extracted and Positive and Negative Words after 2000

Order	Positive words (similarity)	Negative words (similarity)
1	인생(2.31)	애니메이션(3.26)
2	스케일(1.17)	짜통(2.90)
3	여운(2.10)	노잼(2.58)
4	감독관(2.08)	알바(2.35)
5	시절(1.95)	개연(2.36)
6	첫사랑(1.95)	발연기(2.31)
7	최고(1.86)	쓰레기(2.21)
8	청춘(1.85)	유치(1.92)
9	재미나(1.83)	졸작(1.46)
10	할리우드(1.77)	망작(1.72)
11	미소(1.75)	별로(1.67)
12	재미(1.750)	억지(1.50)
13	누구(1.68)	졸작(1.46)
14	기대(1.68)	중뿡(1.46)
15	양양(1.61)	제작비(1.45)
16	눈물(1.60)	수준(1.42)
17	공감(1.51)	자본(1.22)
18	오락(1.47)	스토리(1.17)
19	대박(1.45)	자고(1.12)
20	생각(1.41)	지루함(1.12)

## 4.2 Word2vec 분석 결과

### 4.2.1 공통 단어 분석 결과

두 시기 키워드를 TF-IDF 출력 키워드로 선택한 결과, 같은 키워드로 연기력과 공감이 나타났다.

먼저 공통 키워드 중 연기력을 살펴보았다. 다음 〈Table 4〉에 보는 바와 같이 2000년 이전 홍콩 영화 배우들의 연기에 대한 키워드는 멋지다, 배우, 장면, 슬프다, 성룡, 좋다, 가다, 대단하다, 죽다, 감독 등이다. 분석 결과를 통해 관객들은 1기 홍콩 영화에 대해 명장면이 많고 배우의 연기에 정서가 풍부하며 연출력이 뛰어나다는 평을 들을 정도로 1기 홍콩 영화의 연기에 긍정적인 반응을 보였다. 2000년 이후 홍콩 영화 배우들의 연기에 대한 키워드는 연출, 연기력, 몰입, 아쉽다, 어색하다, 자체, 지루하다, 배우, 내용, 죽다 등이다. 분석 결과를 통해 2000년 이후 홍콩 영화에 대한 관객들의 평가는 부족, 지루함, 아쉬움 등 부

정적인 표현들이 대부분 차지하였다.

〈Table 4〉 Word2vec Analysis of Positive Words (acting) in TF-IDF

2000-words	Similarity	2000+words	Similarity
멋지다	0.999809	연출	0.999901
배우	0.999800	연기력	0.999880
장면	0.999797	몰입	0.999878
슬프다	0.999796	아쉽다	0.999875
성룡	0.999790	어색하다	0.999875
좋다	0.999782	자체	0.999873
가다	0.999776	지루하다	0.999871
대단하다	0.999770	배우	0.999870
죽다	0.999769	내용	0.999867
감독	0.999768	죽다	0.999866

다음으로 공통 키워드 중 공감을 살펴보았다. 다음 〈Table 5〉에 보는 바와 같이 2000년 이전 홍콩 영화에 대한 공감성 키워드는 이야기, 버리다, 끄다, 소재, 때문, 가슴, 가다, 부족하다, 표현, 생각, 버려라 등이다. 부족하다 등의 단어가 첫 시기의 홍콩 영화는 공감대 형성에 부족하다는 결과를 도출해냈다. 2000년 이후 홍콩 영화의 공감에 대한 키워드는 영화로, 처럼, 이네, 같이, 있다, 청춘, 홍콩 영화, 지금, 사실, 추억 등이다. 결과물을 통해 관객들은 2기 홍콩 영화들이 공감대를 더 강하게 형성할 수 있음과 동시에 자신의 과거를 회상할 수 있다는 것을 알 수 있었다.

〈Table 5〉 Word2vec Analysis of Positive Words (empathy) in TF-IDF

2000-words	Similarity	2000+words	Similarity
이야기	0.999809	영화로	0.999900
버리다	0.999800	처럼	0.999887
끄다	0.999797	이네	0.999878
소재	0.999796	같이	0.999875
때문	0.999790	있다	0.999874
가슴	0.999782	청춘	0.999873
가다	0.999776	홍콩 영화	0.999871
부족하다	0.999770	지금	0.999864
표현	0.999769	사실	0.999861
생각	0.999768	추억	0.999860

#### 4.2.2 비공통 단어 분석 결과

비공통 단어 비교에서는 시기별 대표주제어를 몇 개 선택했는데, 우선 2000년 이전의 영화리뷰 키워드 중 두 개를 고른다.

먼저 비공통 단어 중 억지를 살펴보았다. 다음 <Table 6>에 보는 바와 같이 억지와 유사한 키워드는 맞다, 자체, 죽다, 대단하다, 구성, 보이다, 이 영화, 전혀, 긴장감, 내용 등이다. 관객들은 2000년 이전 홍콩 영화를 보면서 영화의 장르적 구성, 영화 속 죽음, 긴장감 등에 대해 다소 억지스럽다고 생각하고 있음을 알 수 있다.

<Table 6> Comparison of Word2vec Uncommon Words (Forced) (before 2000)

Similar Words	Similarity
자체	0.999629
죽다	0.999624
대단하다	0.999622
구성	0.999622
보이다	0.999621
이영화	0.999618
전혀	0.999614
긴장감	0.999609
맞다	0.999608
내용	0.999606

다음으로 비공통 단어 중 분위기를 살펴보았다. 다음 <Table 7>에 보는 바와 같이 분위기와 유사한 키워드는 스토리, 연출, 배우, 시나리오, 뛰어나다, 음악, 전개, 괜찮다, 좋아하다, 캐릭터 등이다. 2000년 이후 홍콩 영화 키워드 결과물을 보면, 2000년 이전에 홍콩 영화 속 이야기, 줄거리, 배우의 연기, 배경음악의 설정 등이 관객들이 느끼는 영화의 전반적인 분위기가 긍정적임을 알 수 있다.

비공통어 비교에서는 시기별 대표주제어를 몇 개 선택했는데, 2000년 이후의 영화리뷰 키워드 중 두 개를 고른다.

<Table 7> Word2vec comparison of non-common words (mood) (before 2000)

Similar Words	Similarity
스토리	0.999707
연출	0.999700
배우	0.999635
시나리오	0.999605
뛰어나다	0.999604
음악	0.999598
전개	0.999590
괜찮다	0.999575
좋아하다	0.999572
캐릭터	0.999572

먼저 비공통 단어 중 유치하대를 살펴보았다. 다음 <Table 8>에 보는 바와 같이 유치하다와 유사한 키워드는 치다, 역사, 목성, 감성, 끄다, 들어가다, 기술, 현실 등 있다. 주제가 출력의 키워드 결과를 통해 관객들은 2000년 이후 홍콩 영화에서 역사 소재의 촬영, 우주 등 기발한 소재의 구상, 영화 기술까지 다루는 것이 다소 유치하고 현실성이 떨어진다고 느꼈다고 밝혔다.

<Table 8> Word2vec Comparison Non-common Words (after 2000) (lingering)

Similar Words	Similarity
청춘	0.999527
주제	0.999508
첫사랑	0.999499
찍다	0.999480
드라마	0.999479
보여주다	0.999479
재미	0.999478
중화시장	0.999472
결말	0.999470
추억	0.999471

다음으로 비공통 단어 중 여운을 살펴보았다. 다음 <Table 9>에 보는 바와 같이 여운과 유사한 키워드는 청춘, 주제, 첫사랑, 찍다, 드라마, 보여주다, 재미, 중화시장, 결말, 추억 등이다. 키워드 출력의 키워드 결

과를 통해 관객들이 2000년 이후 홍콩 영화를 보고 여운이 남았다고 느낀다는 것을 알 수 있었다. 청춘의 테마가 드라마처럼 재미있고, 청춘의 추억을 불러일으킬 수 있으며, 첫사랑의 시간을 떠올릴 수 있는 때 등 시청자들은 극에 몰입해 자신의 과거를 그리워하는 것을 알 수 있다.

〈Table 9〉 Word2vec Comparison Non-common Words (after 2000) (childish)

Similar Words	Similarity
거슬리다	0.999902
치다	0.999901
역사	0.999897
목성	0.999896
감성	0.999892
끄다	0.999891
감성	0.999891
들어간다	0.999890
기술	0.999888
현실	0.999888

Word2vec을 통해 도출된 결과, 여운, 분위기라는 단어에 대해 시청자들은 긍정적인 반응을 보였지만 억지스러움, 유치함이라는 단어에 대해서는 부정적인 반응을 보였다.

## 5. 결론 및 향후 연구 방향

### 5.1 결론 및 시사점

본 연구에서는 홍콩 영화 두 시기를 대상으로 고객의 리뷰를 감정 분석하였다. 두 시기의 영화리뷰를 모아 주요 키워드를 출력한 후 비교하며 홍콩 영화 두 시기의 특징의 차이를 알아보았다.

첫째, 우선 TF-IDF 분석 결과에서 알 수 있듯이 2000년 이전 영화리뷰에서 긍정적인 키워드가 주로 배우, 연기, 결작, 분위기로 나왔다. 점에서 배우의 인기와 연기, 홍콩 영화의 사회적 지위, 영화의 분위기 등이 2000년 이전 홍콩 영화가 사랑받는 주요 요인임을 알 수 있다. 또한 부정적인 키워드 구성, 공감, 폭력이 부정적인 주요 키워드임을 알 수 있다. 그리고 영

화 구성 부족, 공감 부재, 영화 장면 폭력 등이 관객들의 첫 홍콩 영화에 대한 부정적 평가의 원인이라는 것을 알 수 있다.

2000년 이후 TF-IDF 분석 결과를 보면 긍정 리뷰 키워드는 규모, 드라마, 양양 등이었다. 그만큼 2000년 이후 홍콩 영화의 제작 규모와 드라마 같은 전개는 여운을 남겼으며, 양양처럼 미모가 뛰어난 배우들이 2000년 이후 홍콩 영화에 대한 관객들의 호감을 샀던 요인도 있다. 부정적인 평가의 키워드는 스토리, 유치, 연기, 제작비 등이었다. 이처럼 영화의 스토리텔링 능력 부족과 콘텐츠의 유치함, 연기력 부족, 제작비 과다 등이 홍콩 영화에 대한 부정적 평가의 주요 원인이 되고 있음을 알 수 있다.

둘째, Word2vec을 통해 두 시기의 TF-IDF 결과에서 나온 같은 단어를 분석한 결과, 먼저 '연기'라는 단어를 예로 들며 첫 시기의 키워드는 멋, 슬픔, 우수, 결출함 등을 꼽았다. 2000년 이후 나온 키워드는 지루함, 어색함, 재미 등이었다. 연기력에서 두 시기가 확실하게 대비되는 다음, '공감'이라는 단어의 첫 번째 시기에 나온 키워드는 버리기, 끄기, 부족하기 등이다. 두 번째 시기에 나온 키워드는 함께, 청춘, 추억, 현실 등이다. '공감'이라는 단어에도 두 시기의 확실한 대조를 보였다.

두 시기의 비공통어 분석을 통해 첫째 시기의 억지스러움과 분위기가 홍콩 영화의 두드러진 점이고, 둘째 시기의 홍콩 영화는 유치하지만, 추억을 불러일으키는 내용을 담고 있으며, 두 시기의 전형적인 특징은 확실하게 대비되는 각기 다른 특색을 보인다.

TF-IDF가 출력하는 키워드와 Word2vec에서 도출한 주제를 보면 관객들에게 사랑받는 영화, 배우의 연기와 영화의 내용이 곧 영화 성공의 열쇠가 된다. 좋은 연기와 좋은 스토리텔링이 관객들의 사랑을 받게 되는 것이다. Word2vec을 통해 두 시기 홍콩 영화의 TF-IDF에 대해 같은 키워드를 출력한 결과, 영화가 관객의 공감을 얻을 수 있는지도 중요한 요소이다. 두 시기 영화는 공감 측면에서 확실한 대조를 이룬다. 그만큼 관객이 영화에 몰입할 수 있도록 하여 공감을 유발할 수 있는 영화가 사랑받고 있음을 알 수 있다. Word2vec을 통해 TF-IDF가 출력한 두 시기 각각의 특색 있는 키워드를 분석한 결과, 한 편의 영화가 좋은 분위기와 함께 관객들의 여운을 자아내는 것도

중요하다. 좋은 분위기와 추억을 불러일으키는 영화도 관객의 사랑을 받는 요인 것이다.

셋째, 분석 결과를 종합할 때 2000년 이전 홍콩 영화는 배우들의 연기력은 뛰어난데 영화 내용은 억지스럽다. 2000년 이후 홍콩 영화는 제작 규모는 웅대하지만, 내용과 연기가 부족했다. 그만큼 관객들이 영화 속 배우들의 연기와 영화 스토리에 가장 많이 신경을 쓰고 있다는 것을 알 수 있게 하는 대목이다. 연기와 내용은 한 영화를 평가하는 중요한 요소다. 두 시기의 홍콩 영화 변화의 핵심은 주로 배우들의 연기였고, 영화 내용에 변화가 있었음을 알 수 있다. 동시에 관객들이 영화 한 편을 가장 중요하게 여기는 요소가 무엇인지 보여준다. 이번 연구를 통해 두 시기 홍콩 영화의 변화를 분석한 결과 향후 홍콩 영화는 물론 중국 영화의 제작 방향에도 시사하는 바가 크다. 좋은 소재, 좋은 스토리, 좋은 연기가 있어야 진정한 대중적 사랑을 받을 수 있고, 영화의 길이 넓어진다.

## 5.2 향후 연구 방향

홍콩 영화 장르별로 영화를 분류해 수집하지 않으면 영화별로 비교하는 것보다 분석 결과가 정확하지 않을 수 있다. 또한 새로운 관람 고객이 새로운 리뷰를 쓰기도 하는데, 본인이 받은 리뷰는 최신이 아닐 수도 있으며, 수집이 전면적이지 못할 수도 있다. 더불어 나이별 영화리뷰를 수집하지 않고, 연령대별로 영화를 보는 시각이 달라서 분석 결과가 구체적이지 않을 수도 있다.

이번 연구를 통해 홍콩 영화 감정 분석에 그치지 않고 앞으로는 중국 영화에 대한 보다 상세한 분석을 통해 시간이 지남에 따라 사람들이 좋아하는 영화의 장르가 변화하고, 연령대별 구체적인 분류연구, 다른 영화의 종류별 분류연구, 영화 수집량 증가, 영화의 리뷰와 감정 분석 등을 통해 관객이 더 나은 영화를 선택하는데 도움이 되는 방안, 영화 제작자들이 관객이 좋아하는 영화를 어떻게 제작해야 하는지 심도 있게 연구할 예정이다.

## References

- [1] Ahn, J. A., "China's Post '80s and '90s Generation and Selective Reception of Foreign Popular Culture: The Case of Beijing", *Journal of the Korea Contents Association*, Vol. 14, No. 6, 2014, pp. 34-43.
- [2] Chen, J., Chen, X., Chao, X. and Wang, H., "How electronic word of mouth dynamically influences product sales and supplies: An evidence from China film industry", *Economic research-Ekonom-ska istraživanja*, Vol. 35, No. 1, 2022, pp. 637-656.
- [3] Church, K. W., "Word2Vec", *Natural Language Engineering*, Vol. 23, No. 1, 2017, pp. 155-162.
- [4] Fay, C., "Text Mining with R: A Tidy Approach", *Journal of Statistical Software*, Vol. 83, 2018, pp. 1-3.
- [5] Jeon, H. S., "Research on Movie Reviews and Ratings Through Big Data - Focused on Relationship Between Major Features of Movie Reviews and the Box Office Outcomes", *Consumer Culture Research*, Vol. 25, No. 1, 2022, pp. 1-26.
- [6] Lardy, N. R., *Markets over Mao: The rise of private business in China*, Columbia University Press, 2014.
- [7] Lin, C. and He, Y., "Joint sentiment/topic model for sentiment analysis", In *Proceedings of the 18th ACM conference on Information and knowledge management*, November 2009, pp. 375-384.
- [8] Mcauliffe, J. and Blei, D., "Supervised topic models", *Advances in Neural Information Processing Systems*, 2007, p. 20.
- [9] Qaiser, S. and Ali, R., "Text mining: use of TF-IDF to examine the relevance of words to documents", *International Journal of Computer Applications*, Vol. 181, No. 1, 2018, pp. 25-29.

[1] Ahn, J. A., "China's Post '80s and '90s



- [10] Rahman, A. and Hossen, M. S., "Sentiment analysis on movie review data using machine learning approach", The Proceedings of 2019 international conference on bangla speech and language processing (ICBSLP), September, 2019, pp. 1-4.
- [11] Xu, A., Qi, T., and Dong, X., "Analysis of the Douban online review of the MCU: based on LDA topic model", Journal of Physics: Conference Series, Vol. 1437, No. 1, 2020, p. 012102.
- [12] Yessenov, K. and Misailovic, S., "Sentiment analysis of movie review comments", Methodology, Vol. 17, 2009, pp. 1-7.

## ■ 저자소개



### 손 화 양

현재 중국 랴오닝성 진저우의과대학 건강관리학원 교사로 재직 중이다. 호서대학교 국제통상학과에서 학사, 동 대학 경영학과에서 석사학위를 취득하였다. 주요 관심 분야는 경영 데이터를 활용한 텍스트 마이닝, 딥러닝 분석 등이다.

닝, 딥러닝 분석 등이다.



### 이 정 승

현재 호서대학교 경영학부 부교수로 재직 중이며, 한국데이터전략학회 회장을 맡고 있다. KAIST 경영과학과에서 학사 및 석사, 동 대학원 경영대학에서 경영공학전공으로 박사학위를 취득하였으며, SNS 사이트

'올드보이(oldboy.co.kr)', 재무컨설팅사이트 '베스트머니(bestmoney.co.kr)', 블록체인 기반 개인이력인증 솔루션 (Career Ledger) 기업 '쥬스마트아이'를 창업하였다. 주요 관심 분야는 스마트 팩토리, 스마트 그리드, 지속가능 공급망이다.