

영화의 흥행 성과와 리뷰 감정어휘와의 관계 분석

Relationship Analysis between the Box Office Performance and Sentimental Words in Movie Review

주저자

문성민 (Mun, Seong Min)

아주대학교 라이프미디어 협동과정 통합디자인연구실 연구원

공동저자

하효지 (Ha, Hyo Ji)

아주대학교 라이프미디어 협동과정 통합디자인연구실 연구원

교신저자

이경원 (Lee, Kyung Won)

아주대학교 미디어학과

kwlee@ajou.ac.kr

목차

1. 서론

- 1-1. 연구의 배경 및 필요성
- 1-2. 연구의 목적 및 방법

2. 관련 연구

- 2-1. 영화 흥행 관련 연구
- 2-2. 영화 감정 어휘 관련 연구

3. 데이터 수집 및 정제

- 3-1. 영화 리뷰 데이터 수집
- 3-2. 감정 어휘 사전 구축
- 3-3. 영화 리뷰 데이터 정제
- 3-4. 데이터 특성 파악

4. 시각화 기능 및 제안

- 4-1. Parallel coordinates의 개념
- 4-2. Parallel coordinates의 기능
- 4-3. Parallel coordinates를 활용한 분석

5. 결론 및 제안

참고문헌

(요약)

본 연구는 영화 리뷰 값이 1000개 이상인 673개의 영화를 대상으로 영화 장르별로 관객들이 느끼는 감정 어휘의 분포를 탐색하고 영화 흥행도와 감정 어휘의 관계를 파악하는데 목적이 있다. 분석을 위해 네이버 영화 리뷰를 크롤링하고 감정 어휘 사전을 활용하여 7개의 대표 감정 어휘, 영화 티켓 판매액, 영화 관람 관객 수, 상영 스크린 수, 한 스크린 당 영화 관람 관객 수, 영화의 장르, 영화의 영문 이름으로 구성된 데이터를 생성하였다. 연구 목적을 위한 분석은 상관관계 분석 방법과 Parallel coordinates 시각화 분석 방법을 사용하였다. 연구 결과로는 첫째, 장르에 따른 영화 흥행도를 분석하여 영화 흥행도가 가장 높은 장르는 코미디이고 가장 낮은 장르는 호러라는 결과를 확인하였다. 둘째, 모든 장르에서 'Happy'와 'Surprise'의 값이 높게 나오지만 다른 장르들에 비해 판타지 장르의 영화는 지루한 감정이 많이 느껴지고, SF장르의 영화는 화나는 감정이 많이 느껴진다는 결과를 확인 하였다. 셋째, 모든 장르의 감정 어휘 별 상관관계를 분석하여 'Disgust'의 값이 높아질 때 'Anger'의 값도 높아지고, 'Happy'의 값이 높아질 때 'Surprise'의 값은 낮아진다는 결과를 확인하였다. 넷째, 영화 흥행도에 따른 감정 어휘를 분석하여 영화 흥행도와 'Happy'는 선형 관계이지만 영화 흥행도와 'Fear'는 비선형 관계인 것을 확인 하였다.

(Abstract)

This study aims to understand distribution of the sentimental words on each genre and find relationship between box office performance and sentimental words in movie review using 673 movies that have more than 1,000 reviews. For the analysis, crawling movie reviews and made data was composed movie genre, movie name, sales, attendance, screen, normal attendance, 7 sentimental words. For analysis results, we used correlation analysis and Parallel coordinates. As a results, First, the highest box office value of the genre is comedy and the lowest box office value of the genre is horror through analyze box office on each genre. Secondly, Movie genre of fantasy feel a lot of boring emotion and Movie genre of SF feel a lot of anger emotion even if 'Happy' and 'Surprise' have highest sentiment value on every genre. Third, We found 'Anger' increase sentimental value when 'Disgust' increase sentimental value and 'Surprise' decrease sentimental value when 'Happy' increase sentimental value through analyze correlation relationship between sentimental words using total data. Fourth, We found 'Happy' have linear relationship between box office and 'Fear' have non-linear relationship between box office through analyze sentimental words according to box office performance.

(Keyword)

Box Office Performance, Movie genre, Parallel coordinates, Sentimental dictionary, Visual analysis

1. 서론

1-1. 연구의 배경 및 필요성

문화산업의 큰 부분을 차지하는 영화 산업은 매년 성장하고 있으며 최근 미디어 기술의 발달로 영화 시장의 규모가 더 커지고 있다. 영화진흥위원회가 2014년에 발표한 영화 산업 결산 원고에 따르면 2014년 극장 입장권 매출액은 역대 최대 액수인 1조 6,641억 원을 기록, 2013년 대비 7.3% 증가했고 관객 수도 2억 명을 돌파하며 2013년 대비 약 0.8% 증가한 2억 1,506만 명에 달하였다. 또한 인구 1인당 평균 관람 횟수는 2013년보다 증가하며 4.19회에 달하였다.¹⁾

영화 산업의 성장과 함께 영화 흥행에 영향을 미치는 요인을 도출하기 위한 다양한 연구가 진행되어 왔다. 한국 영화 산업에 대한 연구는 주로 영화의 흥행을 높이기 위해 영화 흥행에 영향을 미치는 요인을 도출하기 위한 방향으로 진행되어 왔다. 국내에서는 1997년 김휴종의 “한국 영화 스타의 스타파워 분석”이라는 연구를 시작으로 김영현 외 1명(2011), 박승현 외 2명(2011) 등의 연구가 진행되어 왔으며 이러한 연구를 통해 감독의 명성, 제작, 유통사의 역량, 평론가의 리뷰, 감독, 작가, 제작자 등 영화와 관련된 관계자들의 역할 및 상호간의 연결 등이 영화의 성공여부를 결정할 수 있는 요인으로 밝혀졌다.²⁾³⁾⁴⁾

그러나 이제까지 연구된 영화 흥행에 영향을 미치는 요인들만으로는 영화의 흥행을 설명하는데 한계점이 존재한다. 선행 연구에서 도출된 영화의 성공 여부를 결정할 수 있는 요인들을 충족하지만 영화 흥행에 실패한 영화의 예로 2014년 개봉한 영화 ‘익스펜더블3’가 있다. 이 영화의 경우 유명한 해외 스타들이 다수 등장하고 유명한 감독이 연출하고 해외에서 이미 흥행에 성공을 했지만 국내 시장에선 흥행(누적 관객 152,025명)에 성공 하지 못했다. ‘익스펜더블3’에 대한 영화 리뷰는 “1,2편과 스토리가 똑같고 지루하다.”, “팬서비스용 재연 액션, 그러니까 그들의 팬들에게만” 등이 있었고 리뷰를 통해 영화에서 지루한 감정을 느꼈다는 것을 확인 할 수 있었다. 반면에 영화 흥행에 영향을 미치는 요인들을 충족하지 못하지만 영화 흥행에 성공한 영화의 예로 2014년 개봉한 영화 ‘님아, 그 강을 건너지 마오’가 있다. 영화 ‘님아, 그 강을 건너지 마오’는 독립영화로써 영화 흥행에 영향을 미치는 요인들을 충족하지 못하지만 독립 영화로써 흥행 기록을 새로이 갱신하며 흥행(누적 관객 4,801,445명)에 성공한 영화이다.⁵⁾ ‘님아, 그 강을 건너지 마오’에 대한 영화 리뷰는 “노부부가 살아온 인생 여정이 너무나 가슴 아프고 감동스러웠습니다.”, “정말 가슴이 먹먹해집니다. 상영관 좀 늘려주세요~ 보고 싶어 하는 사람들이 많아요.” 등이 있었고 리뷰를 통해 영화에서 슬픈 감정을 느꼈다는 것을 확인 할 수 있었다.

영화 흥행에 영향을 미치는 요인은 객관적인 정보에 의한 요인과 주관적인 정보에 의한 요인으로 나눌 수 있다. 객관적인 정보에 의한 요인은 감독의 명성, 제작, 유통사의 역량, 영화 평론가의 리뷰, 감독, 작가, 제작자 등과 같은 요인이며, 주관적인 정보에 의한 요인은 구전의 의한 정보전달 측면에서 매우 중요하며 영화에 대한 관객들의 생각, 관객들이 영화를 통해 느낀 감정 등 관객에 의해 발생하는 요인들이라고 할 수 있다.⁶⁾

1) 2014년 한국 영화산업 결산, 영화진흥위원회 정책연구부, p.6, 2014.

2) 김휴종, “한국영화스타의 스타파워분석”, 삼성경제연구소 연구보고서, 1997.

3) 김영현, 홍정환, “영화 흥행 결정 요인과 흥행 성과 예측 연구”, 한국통계학회논문집, 제18권, 제6호, p.865, 2011.

4) 박승현, 송현주, 정안규, “한국영화의 흥행성과 결정 요인에 관한 연구”, 언론과학연구, 제11권, 제4호, p.240, 2011.

5) 영화진흥위원회, <http://www.kobis.or.kr/kobis/business/mast/mvie/searchMovieList.do>

6) 성영신, 박진영, 박은아, “온라인 구전 정보가 영화 관람 의도에 미치는 영향”, 광고연구, 제57권, p.39, 2002.

위의 사례를 통해 객관적인 정보에 의한 요인만으로 영화 흥행과의 관계를 규명하기에는 한계가 있다고 볼 수 있다. 따라서 주관적인 정보에 의한 요인과 영화 흥행과의 관계를 규명하는 연구 등 다양한 관점에서 연구가 진행되어야 한다.

1-2. 연구의 목적 및 방법

본 연구는 주관적인 정보에 의한 요인 중 하나인 영화를 보고 관객들이 느끼는 감정과 영화 흥행과의 관계를 규명하기 위해 각 영화의 장르 별로 관객들이 영화를 보고 느끼는 감정의 분포를 탐색하고 영화 흥행도와 감정 어휘의 관계를 파악하는데 연구 목적을 두고 있다. 이러한 연구 목적을 달성하기 위한 방법은 다음과 같다.

첫째, 영화를 통해 관객들이 느끼는 감정을 확인하기 위해 네이버 영화 평에서 리뷰의 개수가 1000개 이상인 673개의 영화를 선정하고 전체 리뷰 텍스트 데이터를 JAVA언어를 사용하여 크롤링(Crawling)하였다.

둘째, 표본으로 선정된 673개의 영화 리뷰 텍스트 데이터에서 감정 어휘를 추출하기 위해 선행연구를 기반으로 감정 어휘 사전을 구축하였다.

셋째, 감정 어휘 사전을 사용하여 영화 리뷰 텍스트 데이터를 정제하여 감정 어휘 값을 생성하고 'TF-IDF'공식을 활용하여 각 영화에 대한 감정어의 분포를 표준화 하였다. 또한 영화의 주요 장르, 누적 매출액, 누적 관객 수, 상영 스크린 수 등 영화에 대한 객관적인 정보를 나타내는 요인과 병합하였다.

넷째, 정제된 데이터를 분석하기 위해 Parallel coordinate 시각화를 제작하고 Parallel coordinate를 활용하여 각 영화의 장르별로 대표 감정 어휘의 분포를 탐색하고 영화 흥행도와 감정 어휘의 관계를 도출하였다.

다섯째, 분석 결과를 해석하고 연구의 시사점과 연구의 한계, 향후 연구 방향을 제시하였다.

2. 관련 연구

2-1. 영화 흥행 관련 연구

영화 산업 시장의 규모가 확대됨에 따라 영화의 흥행에 영향을 미치는 요인을 도출해 내기 위한 다양한 연구들이 진행 되어오고 있다.

국내에서는 1997년 김휴종의 '한국 영화 스타의 스타파워 분석'이라는 연구와 2011년 박승현, 송현주, 정완규의 '한국영화의 흥행성과 결정요인에 관한 연구' 등 다양한 연구가 있다. 김휴종의 연구에서는 1988년부터 1995년까지의 529편의 영화를 분석하여 배우와 감독의 스타 파워가 흥행에 미치는 영향력을 검증하였으며⁷⁾ 박승현 외 2명의 연구에서는 개봉 스크린 규모, 제작비, 전문가 평가, 온라인 평가 등이 흥행에 미치는 영향력을 검증하였다.⁸⁾

해외에서는 1983년 리트만(Litman)의 'Predicting Success of Theatrical Movies: An Empirical Study'와 1991년 와이아트(Wyatt)의 'High concept, product differentiation, and the contemporary U.S film industry'등의 연구가 있다. 리트만의 연구는 1970년대 개봉한 영화 155편

7) 김휴종.

8) 박승현, 송현주, 정완규.

에 대한 정보를 수집하고 회귀분석을 실시하여 제작비, 개봉스크린 규모, 전문가 평점, 배급사 파워, 아카데미 수상실적 등의 요소와 공상 과학/판타지, 코미디, 공포의 세 가지 장르 요소가 유의미한 영향력을 지닌다는 결과를 도출하였고⁹⁾ 와이어트의 연구는 제작비, 전문가 평점, 스타 배우의 파워, 아카데미 수상실적, 여름시즌 개봉의 요인이 영화 흥행에 영향을 미친다는 결과를 도출하였다.¹⁰⁾ 영화 흥행도 관련 선행 연구의 내용은 <표1>과 같다.

<표 1> 영화 흥행도 관련 논문

저자	논문	흥행에 영향을 미치는 요인
김휴중	한국영화스타의 스타파워분석(1997)	배우의 스타 파워, 감독의 스타 파워
박승현 외 2명	한국 영화의 흥행성과 결정 요인에 관한 연구(2011)	개봉 스크린 규모, 제작비, 전문가 평가, 온라인 평가
송현주 외 1명	영화의 흥행성과 제작비 규모와의 관계(2012)	온라인 평가, 영화 평점, 제작비, 코미디 장르
이윤정 외 1명	원작의 유무와 형태가 영화 흥행에 미치는 영향(2013)	소설 원작의 존재, 소설 원작의 형태, 소설 원작의 영향력, 영화 관련 외부 콘텐츠
Litman, B	Predicting Success of Theatrical Movies: An Empirical Study(1983)	제작비, 개봉스크린 규모, 전문가 평점, 배급사 파워, 아카데미 수상실적, 공상과학/판타지, 코미디, 공포(3가지 장르)
Wyatt, R. O	High concept, product differentiation, and the contemporary U.S film industry(1991)	제작비, 전문가 평점, 스타 배우의 파워, 아카데미 수상실적, 여름시즌 개봉

2-2. 영화 감정 어휘 관련 연구

주관적인 정보에 의한 요인은 영화를 관람한 관객에 의해 생성되며 영화에 대한 평가 리뷰 안에서 감정 어휘를 추출하여 분석하면 영화 흥행과 감정 어휘 사이의 관계를 파악 할 수 있다. 감정 어휘에 대한 연구는 언어에 따라 연구 방법이 상이한데 한글의 텍스트 기반 감정 언어에 대한 연구는 2008년 이준웅, 송현주, 나은경, 김현석의 연구가 있고 영화 리뷰를 활용하여 감정 어휘를 분석한 연구는 2014년 박지연, 전범수의 연구가 있다.

이준웅 외 3명의 연구(2008)에서는 유사성 분류 자료를 근거로 군집 분석을 수행하였고, 감정 어휘를 기본 수준에서 ‘기쁨’, ‘궁지’, ‘사랑’ 등 긍정적인 정서들과 ‘공포’, ‘분노’, ‘연민’, ‘수치’, ‘좌절’, ‘슬픔’ 등 부정적인 정서로 총 9개의 정서 범주로 나눌 수 있음을 제시하였다. 또한 ‘기쁨’의 경우 다시 ‘기쁨’과 ‘통쾌’로 나누어지며 나누어진 ‘기쁨’도 ‘재미’, ‘즐겁’, ‘유쾌’, ‘흥겹’, ‘신나’, ‘기쁘’, ‘열광’으로 나누는 등 대표 감정 어휘가 포함하는 세부 감정 어휘에 대한 온톨로지를 구축하였다.¹¹⁾

박지연 외 1명의 연구(2014)에서는 한국 및 외국 흥행 영화에 대한 네티즌 리뷰를 중심으로 네티즌 리뷰에 사용된 감정 동사와 흥행 영화와의 관계를 분석하였다. 연구 결과로는 네티즌들이 영화를 판단 할 때 재미를 가장 큰 요인으로 생각하고 있다는 점과 한국 영화의 경우 재미를 기준으로 몰입

9) Litman, B, “Predicting Success of Theatrical Movies: An Empirical Study”, Journal of Popular Culture, 16 (Spring), p.166, 1983.

10) Wyatt, R. O, “High concept, product differentiation, and the contemporary U.S film industry”, Current research in film Audiences, economics, and law, Vol. 5, p.93, 1991.

11) 이준웅, 송현주, 나은경, 김현석, “정서 단어 분류를 통한 정서의 구성 차원 및 위계적 범주에 관한 연구”, 한국 언론 학보, 제52(1)권, p.101, 2008.

이나 감동 등 감정적 동사를 기준으로 영화가 군집화 되고 외국 영화는 재미있는 영화와 재미없는 영화로 군집화 된다는 결과를 얻었다.¹²⁾

3. 데이터 수집 및 정제

3-1. 영화 리뷰 데이터 수집

본 연구에서는 영화의 장르 별로 감정 어휘의 분포를 탐색하고 영화 흥행도와 감정 어휘의 관계를 파악하기 위해 아래와 같이 두 단계의 데이터 수집 과정을 거쳤다.

첫째, 영화 리뷰의 수집을 자동화하기 위해 데이터를 수집 할 수 있게 JAVA를 사용하여 해당 사이트에 대한 웹 크롤러를 제작하였다. 크롤러는 네이버 영화 홈페이지에서 특정 영화의 관람객 댓글과 리뷰들을 정제되지 않은 데이터 형태로 수집하도록 설계되었다.

둘째, 수집된 영화 데이터 중에서도 리뷰의 개수가 1000개 이상인 영화들만 다시 필터링하였고 최종적으로는 2289개의 영화 중 673개의 영화에 대한 리뷰 데이터가 수집되었다.

3-2. 감정 어휘 사전 구축

본 연구에서는 감정 어휘 사전을 구축하기 위해 여러 방법을 거쳤으며 감정 어휘 사전 구축 과정은 다음과 같다.

첫째, 본 연구의 기반이 된 선행 연구 중 하효지 외 2명의 연구(2013)를 참고하여 68개의 감정 어휘를 선정하고 7개의 대표 감정 어휘 범주('Happy', 'Sad', 'Fear', 'Anger', 'Surprise', 'Disgust', 'Boring')로 분류 및 군집화 하는 작업을 시행하였다.¹³⁾ 해당 감정 어휘 사전을 바탕으로 리뷰 데이터를 분류하여 673개 영화 리뷰에 대한 1차 감정 어휘 데이터를 생성하였다.

둘째, 본 연구에서는 실제 영화 리뷰 데이터에서 영향력이 미미한 감정 어휘를 제거하여 조금 더 정확한 결과를 얻고자 하였으며 이를 위해 1차적으로 TF-IDF공식을 활용하여 감정 어휘 데이터를 표준화시켰다. TF-IDF는 각 감정어 집단의 단어 빈도수(tf : Term Frequency)와 역문서빈도(idf : Inverse Document Frequency)를 곱하여 감정 어휘를 표준화시킨 공식으로써 TF-IDF값에 관한 공식은 다음과 같다.

$$TF-IDF(t,d,D) = tf(t,d)*idf(t,D)$$

셋째, 감정 어휘 개수를 줄이기 위해 각 감정 어휘에서 나타날 수 있는 TF-IDF 스코어의 최대치를 구하였다. 예를 들어 '경악하다'의 경우 모든 영화에서 TF-IDF 스코어의 비율이 0.8% 이하인 반면에 '달콤하다'의 경우는 적어도 한 개의 영화에서는 TF-IDF 스코어의 비율이 42%에 달하는 것을 뜻한다. 따라서 TF-IDF스코어의 비율이 10%미만인 감정 어휘를 제거하고 최종적으로 36개의 감정 어휘를 선택하였다. 최종적으로 선택된 36개의 감정 어휘는 크게 'Happy', 'Surprise', 'Boring', 'Sad', 'Anger', 'Disgust', 'Fear'의 성격으로 나뉘게 되며, 감정 어휘에 관한 내용은 <표 2>와 같다.

12) 박지연, 전범수, “네트즌의 흥행 영화 리뷰에 포함된 감정 동사 이용 특성 연구”, 한국 콘텐츠 학회, 제14(5)권, p.88, 2014.

13) 하효지, 김기남, 이경원, “영화 리뷰의 감정 어휘 공간 및 영화 관람의 상황분석 연구”, 디자인 융복합 학회, 제 12(6)권, p.51, 2013.

〈표 2〉 최종 선정된 36개의 감정 어휘 표

대표 감정 어휘	세부 감정 어휘
Happy	Happy, Sweet, Funny, Exited, Pleasant, Fantastic, Gratified, Enjoyable, Energetic
Surprise	Surprised, Ecstatic, Awesome, Wonderful, Great, Touched, Impressed
Boring	Calm, Drowsy, Bored
Sad	Pitiful, Lonely, Mournful, Sad, Heartbroken, Unfortunate
Anger	Outraged, Furious
Disgust	Ominous, Cruel, Disgusted
Fear	Scared, Chilly, Horrified, Terrified, Creepy, Fearsome

3-3. 영화 리뷰 데이터 정제

본 연구는 영화의 장르 별로 관객들이 영화를 보고 느끼는 감정의 분포를 탐색하고 영화 흥행도와 감정 어휘의 관계를 파악하는데 연구 목적을 두고 있으며 이를 위해 앞의 과정을 통해 얻은 데이터 중 대표 감정 어휘와 각 영화의 객관적인 정보를 담은 데이터를 병합하여 최종적인 데이터를 생성하였다. 병합된 영화에 대한 객관적인 정보 데이터는 영화 티켓 판매액, 영화 관람 관객 수, 상영 스크린 수, 한 스크린 당 영화 관람 관객 수, 영화의 장르, 영화의 영문 이름 등이었으며 영화진흥위원회의 통계 결과를 바탕으로 데이터를 생성하였다.¹⁴⁾ 영화 흥행도를 나타내는 요인으로 1983년 리트만(Litman)의 연구에서는 누적 매출액을 영화 흥행도로 사용하였으며,¹⁵⁾ 최근 연구들 중 2012년 박승현 외 1명의 연구에서는 누적 관객 수를 영화 흥행도를 나타내는 요인으로 사용하였다.¹⁶⁾ 본 연구에서는 누적 관객 수를 상영 스크린 수로 나누어 한 스크린에서의 누적 관객 수를 영화 흥행도로 사용하였으며 위의 데이터와 병합하여 분석에 사용하였다.

3-4. 데이터의 특성 파악

3-4-1. 영화 장르 별 흥행도의 분포와 감정 어휘 분포 분석

673개의 영화에 대한 총 판매액의 평균은 10,170,432,626원이고 누적 관객 수의 평균은 1,470,000명, 평균 상영 스크린은 330개 상영관, 흥행도를 나타내는 한 스크린 당 누적 관객 수의 평균은 4,010명이었다. 또한 영화의 장르는 드라마(25.9%), 액션(19.2%), 코미디(13.8%), 멜로 로맨스(7.6%), SF(7.3%)순으로 드라마가 제일 많은 비중을 차지하였으며, 제작 국가는 한국(44.1%), 미국(38.2%), 일본(4%), 프랑스(1%), 미국&독일(0.7%) 순으로 한국과 미국이 대부분의 비중을 차지하였다. 마지막으로 673개의 영화에 대한 리뷰에서 감정어의 분포는 ‘Happy’(34.3%), ‘Surprise’(31.7%), ‘Sad’(13.3%), ‘Boring’(7.6%), ‘Fear’(5.7%), ‘Anger’(4.5%), ‘Disgust’(2.7%)순으로 ‘Happy’와 ‘Surprise’가 차지하는 비율이 높았다. 최종적으로 분석에 사용 될 데이터의 특징은 〈표 3〉의 내용과 같다.

14) 영화진흥위원회.

15) Litman, B.

16) 박승현, 송현주, “영화의 흥행성과와 제작비 규모와의 관계: 2011년 한국영화의 흥행결정요인 분석”, 사회과학연구, 제51집 1호, p.67, 2012.

〈표 3〉 영화 데이터의 특징

구분		빈도	비율(%)
장르	드라마	174	25.9
	액션	129	19.2
	코미디	93	13.8
	멜로 로맨스	51	7.6
	SF	49	7.3
제작국가	한국	297	44.1
	미국	257	38.2
	일본	27	4
	프랑스	7	1
	미국 & 독일	5	0.7
감정어휘	Happy	1143281	34.3
	Surprise	1054108	31.7
	Sad	443010	13.3
	Boring	253862	7.6
	Fear	189763	5.7
	Anger	150013	4.5
	Disgust	89827	2.7

4. 시각화 기능 및 제안

4-1. Parallel coordinates의 개념

본 연구에서 다변량으로 되어 있는 영화 데이터를 분석하기 위해 여러 시각화 분석 방법 중 Parallel coordinates(평행 좌표계)를 사용하였다. Parallel coordinates 시각화 분석 방법은 N차원 공간 안의 점들의 집합을 보여주기 위한 방법으로 일반적으로 수직의 형태이며 N개의 등간격 평행 라인으로 이루어져 있다. 또한 시계열 데이터 시각화에도 밀접한 관계가 있으며 데이터 내 변수간의 관계를 파악하는데 용이하다.¹⁷⁾ 이 방법은 1985년 Inselberg, A. 가 구체적으로 제안하였고 최근까지 다양한 학문 영역에서 사용되고 있다. Inselberg, A.의 연구에 따르면 Parallel coordinate는 각 변수의 대부분 라인이 평행일 때 두 차원 사이에 유사한 관계가 형성된다고 해석할 수 있으며, 대부분의 라인이 교차할 때는 상이한 관계가 형성된다고 해석한다.¹⁸⁾

또한 본 연구에서는 Parallel coordinates 시각화 방법에 선택된 데이터의 평균 값을 나타내는 기능, 영화의 장르를 선택하는 기능, 축을 변경하는 기능, 축을 제거하는 기능, 하나의 영화를 선택하여 데이터의 특징을 확인하는 기능, 영화의 제목 명으로 데이터를 검색하는 기능 등 분석에 필요한 다양한 인터랙션 기능을 추가하였다.

4-2. Parallel coordinates의 기능

본 연구에서는 정확한 시각화 분석을 시행하기 위해 기존의 Parallel coordinates 시각화 방법에 다양한 기능을 추가하였다. 연구에 사용된 Parallel coordinates 시각화의 기능은 다음과 같다.

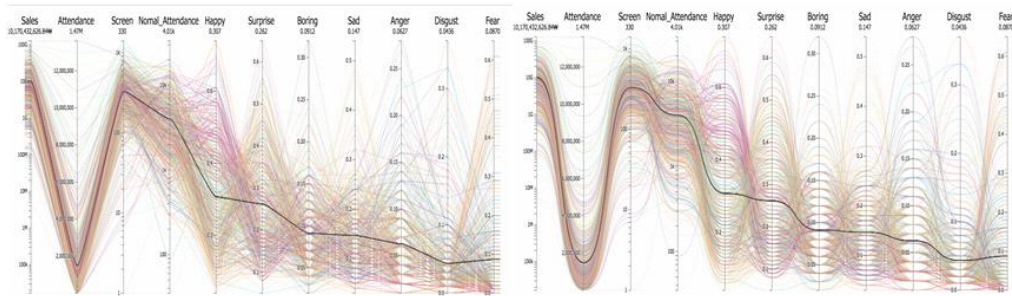
4-2-1. 번들링(Bundling)

Parallel coordinates 시각화는 일반적으로 축이 변경 될 때 데이터의 연결 표현을 직선으로 표현하고 있다. 하지만 직선으로 데이터 사이의 연결을 표현할 경우 데이터의 양이 많으면 축이 전혀 보이지 않고 데이터 사이의 패턴 또한 발견하기 어렵다. 본 연구에서는 〈그림 1〉의 오른쪽과 같이 데이터 사이의 연결 표현에 번들링 기능을 추가 하였으며 번들링 기능을 통해 데이터 사이의 연결을 표

17) 위키 백과사전, Parallel coordinate의 정의, <http://ko.wikipedia.org/>

18) Inselberg, A, The plane with Parallel coordinates, The Visual Computer, p. 79, 1985,

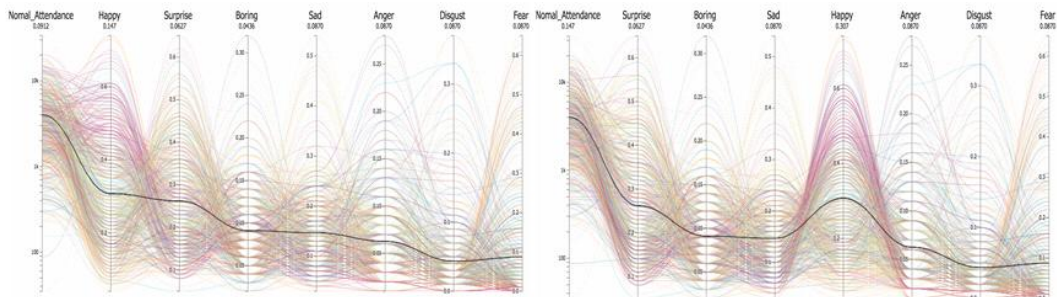
현하면 해당 축의 데이터들을 일정 수준으로 묶어서 표현하기 때문에 데이터들이 군집화 되는 경향을 쉽게 확인 할 수 있다.



〈그림 1〉 (왼쪽) 일반적인 Parallel coordinates (오른쪽) 변들링 기능을 추가한 Parallel coordinates

4-2-2. 축(Axes)

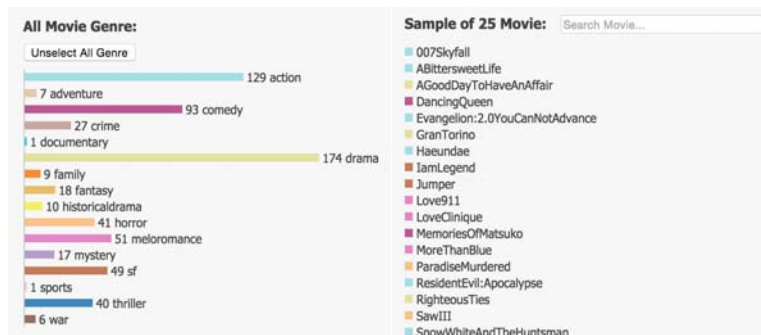
일반적인 Parallel coordinates 시각화는 데이터 변수를 축으로 설정하고 데이터 변수 사이에서 변화하는 데이터들의 패턴을 확인하는데 매우 용이하다. 본 연구에서는 Parallel coordinates에 〈그림 2〉와 같이 데이터 변수 축의 삭제 기능과 축의 순서를 이동시키는 기능을 추가하여 분석을 용이하게 하였다.



〈그림 2〉 (왼쪽) 기본적인 데이터 축의 순서 (오른쪽) Happy의 데이터 변수 축 순서 변경

4-2-3. 색상(Colour)

장르별로 비교 분석을 하기 위해서는 영화의 장르별로 구분 할 수 있는 기능이 필요한데, 본 연구에서는 〈그림 3〉과 같이 영화가 속하는 주요 장르에 따라 line 그래프의 색상을 다르게 지정하여 사용자가 영화의 장르별로 데이터를 구분 할 수 있게 하였다.



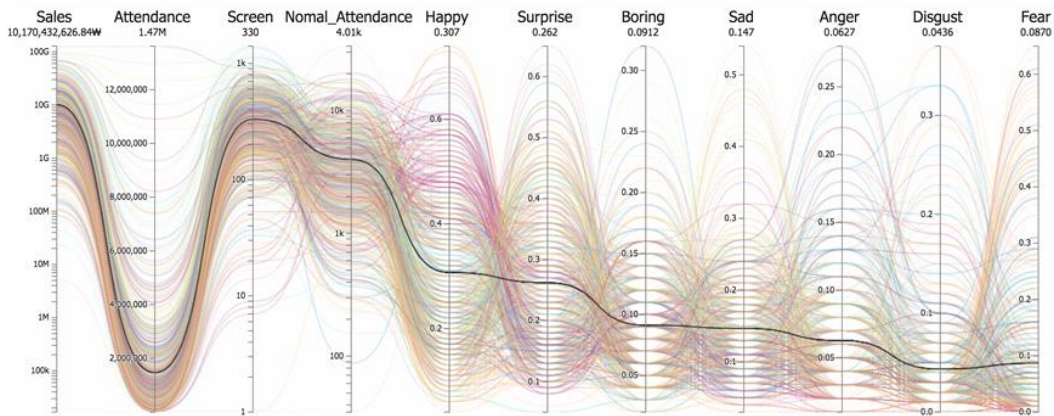
〈그림 3〉 영화의 장르별로 지정된 색상

4-2-4. 기술 통계(Descriptive statistic)

일반적으로 시각화 분석에서 사용되는 Parallel coordinates는 데이터 패턴, 데이터 변수 축에서 발생하는 군집화, 변수 축 사이의 직선 기울기 등 시각적으로 확인이 가능한 부분만으로 해석을 해야 한다. 본 연구에서는 다양한 관점으로 분석을 시행하기 위해 <그림 4>와 <그림 5>와 같이 선택된 데이터의 평균을 나타내는 평균선, 선택된 각각의 데이터 변수 축의 평균 값, 선택된 영화 수의 합계를 나타내는 기능을 추가하였다.



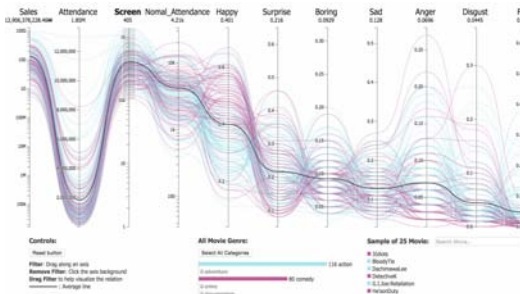
<그림 4> 선택된 데이터 변수들의 평균값과 영화 수의 합계



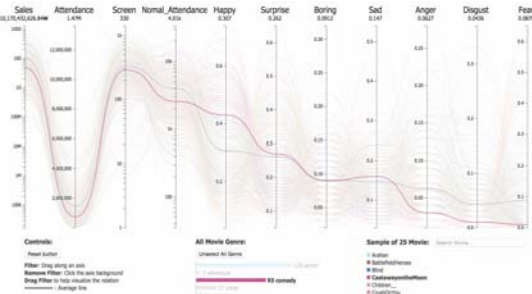
<그림 5> 선택된 데이터 변수들의 평균 값과 평균 선 (굵은 line 그래프)

4-2-5. 데이터 선택(Data Selection)

영화의 장르 별로 대표 감정 어휘의 분포를 탐색하고 영화 흥행과 감정 어휘 사이의 관계를 파악하기 위해서는 장르 별 데이터의 패턴 비교, 분포 확인, 조건에 따른 패턴 변화 등을 확인하여야 한다. 본 연구에서는 분석을 용이하기 위해 <그림 6>과 <그림 7>과 같이 장르 선택 기능, 영화 검색 기능, 조건에 따른 데이터 필터링 기능, 하이라이트 기능 등을 추가하여 분석을 용이하게 하였다.



<그림 6> 데이터 선택: 장르가 액션 & 코미디이고 상영 스크린 수가 100개 이상인 영화



<그림 7> 2009년 개봉작 김씨 표류기(Castaway on the Moon)에 대한 하이라이트 View

4-3. Parallel coordinates를 활용한 분석

본 연구는 영화 리뷰를 사용하여 영화의 장르 별로 대표 감정 어휘의 분포를 탐색하고 영화의 흥행도와 감정 어휘의 관계를 파악하는데 연구 목적 두고 있다. 분석을 위해 7개의 대표 감정 어휘, 영화 티켓 판매액, 영화 관람 관객 수, 상영 스크린 수, 한 스크린 당 영화 관람 관객 수, 영화의 장르, 영화의 영문 이름 등으로 이루어진 최종 데이터를 생성하였으며 연구 목적을 위한 분석은 상관관계 분석 방법과 Parallel coordinates 시각화 분석 방법을 사용하였다.

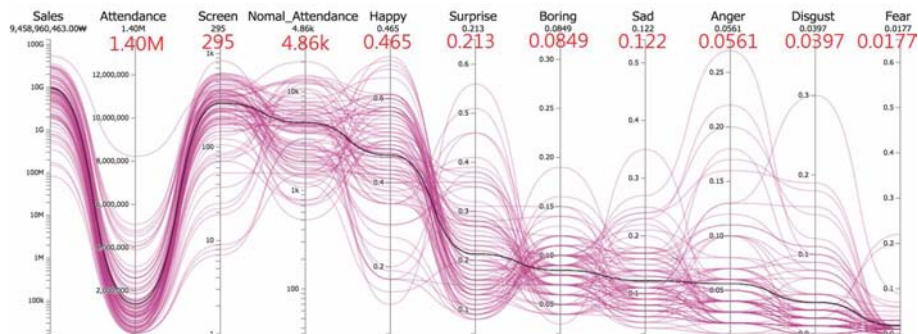
4-3-1. 영화 장르 별 흥행도의 분포와 대표 감정 어휘의 분포

본 연구에 사용된 673개의 영화의 대표 장르는 드라마(25.9%), 액션(19.2%), 코미디(13.8%), 멜로로맨스(7.6%), SF(7.3%)등의 순서이며 데이터의 개수가 15개 이상인 영화의 장르 별 흥행도와 감정 어휘 분포는 <표4>와 같다.

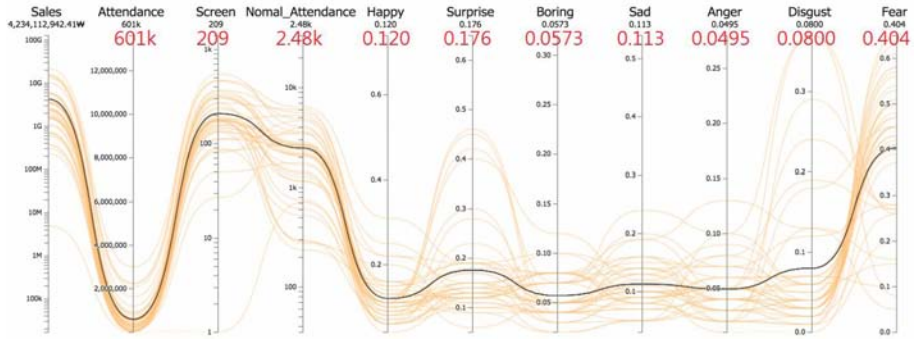
<표4> 데이터의 개수가 15개 이상인 영화의 장르 별 감정 어휘 분포(평균 값)

대표 장르	흥행도	Happy	Surprise	Boring	Sad	Anger	Disgust	Fear
드라마(174)	4,270	0.249	0.333	0.0935	0.186	0.054	0.035	0.0476
액션(129)	4,060	0.351	0.233	0.0926	0.13	0.0774	0.0455	0.0715
코미디(92)	4,860	0.465	0.213	0.0849	0.122	0.0561	0.0397	0.0177
멜로로맨스(52)	3,590	0.325	0.330	0.102	0.161	0.0438	0.021	0.019
SF(49)	4,130	0.274	0.227	0.102	0.163	0.0833	0.0504	0.102
호러(41)	2,480	0.120	0.176	0.0573	0.113	0.0495	0.080	0.404
스릴러(40)	3,520	0.207	0.253	0.0898	0.118	0.0828	0.0688	0.182
범죄(27)	3,680	0.283	0.250	0.101	0.121	0.0715	0.0722	0.103
판타지(18)	3,310	0.368	0.240	0.104	0.159	0.0728	0.0239	0.0339
미스터리(17)	2,480	0.225	0.217	0.0912	0.123	0.0506	0.0347	0.259

본 연구에서는 누적 관객 수를 상영스크린 수로 나누어 한 스크린에서의 누적 관객 수를 영화 흥행도로 사용하였으며 영화의 흥행도가 제일 높은 장르인 코미디의 경우, 한 스크린 당 평균 누적 관객 수는 4,860명 이었고 감정어의 분포는 <그림 8>과 같이 'Happy'(0.465), 'Surprise'(0.213), 'Sad'(0.122), 'Boring'(0.0849), 'Anger'(0.0561), 'Disgust'(0.0397), 'Fear'(0.0177)순이었으며 평균적으로 행복한 감정이 다른 감정에 비해 많이 느껴진다는 결과를 Parallel coordinates를 통해 확인 할 수 있다. 다음으로 영화의 흥행도가 제일 낮은 장르인 호러의 경우, 한 스크린 당 평균 누적 관객 수는 2,480명 이었고 감정어의 분포는 <그림 9>와 같이 'Fear'(0.404), 'Surprise'(0.176), 'Happy'(0.120), 'Sad'(0.113), 'Disgust'(0.0800), 'Boring'(0.0573), 'Anger'(0.0495)순이었으며 다른 감정에 비해 무서운 감정이 많이 느껴진다는 결과를 Parallel coordinates를 통해 확인 할 수 있다.



<그림 8> 영화의 대표 장르가 코미디인 영화에 대한 감정 어휘 분포



〈그림 9〉 영화의 대표 장르가 호러인 영화에 대한 감정 어휘 분포

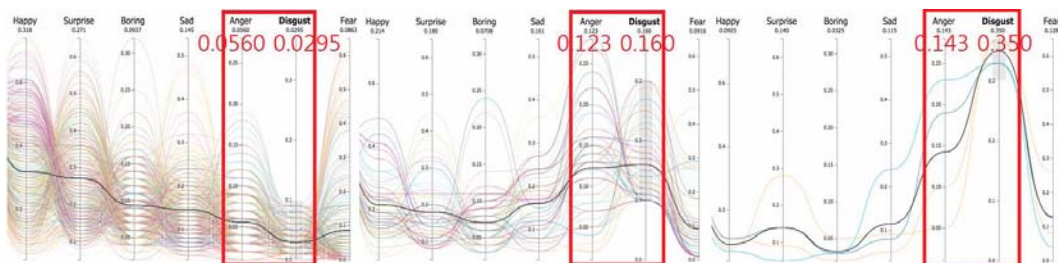
4-3-2. 영화의 대표 감정 어휘 사이의 상관관계

본 연구에서는 감정 어휘를 36개로 세분화 하였으며 이를 다시 'Happy', 'Surprise', 'Boring', 'Sad', 'Anger', 'Disgust', 'Fear'로 군집화 하였다. 장르 구분 없이 전체 영화를 대상으로 상관계수 (Correlation) 공식을 활용해 대표 감정 어휘 별 상관관계를 분석한 결과는 〈표5〉와 같으며 모든 변수들은 $P < 0.05$ 로 선형 연관성이 있었다.

〈표 5〉 전체 영화의 대표 감정 어휘 별 상관관계 표

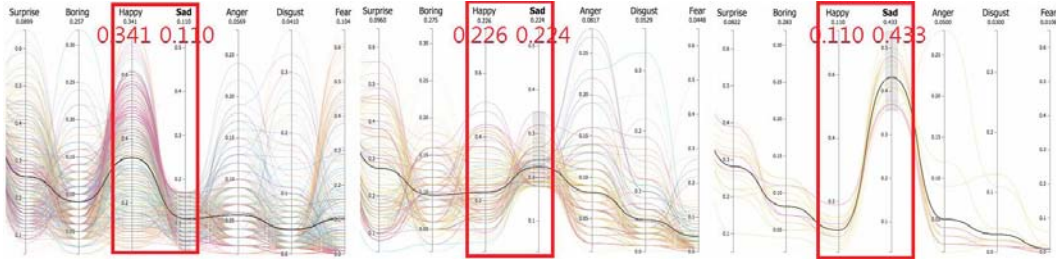
	Happy	Surprise	Boring	Sad	Anger	Disgust	Fear
Happy	1.00	-0.286	0.016	-0.444	-0.175	-0.300	-0.466
Surprise	-0.286	1.00	-0.136	0.042	-0.407	-0.338	-0.300
Boring	0.016	-0.136	1.00	0.052	0.068	-0.173	-0.205
Sad	-0.444	0.042	0.052	1.00	0.202	0.045	-0.215
Anger	-0.175	-0.407	0.068	0.202	1.00	0.514	-0.097
Disgust	-0.300	-0.338	-0.173	0.045	0.514	1.00	0.122
Fear	-0.466	-0.300	-0.205	-0.215	-0.097	0.122	1.00

전체 영화를 대상으로 대표 감정 어휘 별 상관관계를 분석한 결과 상관관계가 정(正)의 방향으로 가장 높은 감정 어휘는 'Disgust'와 'Anger'로 상관관계는 0.514이다. 이를 Parallel coordinates를 통해 확인 한 결과는 다음과 같다. 'Disgust'에 대한 구간 값이 0.1이하인 경우 'Disgust'와 'Anger'의 감정 어휘 값은 0.0295와 0.0560, 'Disgust'에 대한 구간 값이 0.1부터 0.3인 경우 'Disgust'와 'Anger'의 감정 어휘 값은 0.160와 0.123, 'Disgust'에 대한 구간 값이 0.3이상인 경우 'Disgust'와 'Anger'의 감정 어휘 값은 0.350와 0.143이었다. 이는 'Disgust'의 값이 높아질 때 'Anger'의 값도 같이 높아진다고 볼 수 있으며 분석 결과는 <그림 10>과 같다.



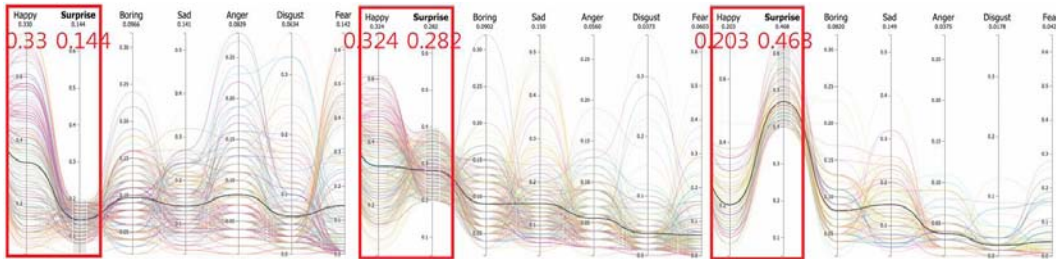
〈그림 10〉 Disgust값이 0.1이하(왼쪽), 0.1이상 0.3이하(가운데), 0.3이상(오른쪽)일 때의 분석 결과

다음으로 상관관계가 부(不)의 방향으로 가장 높은 감정 어휘는 'Happy'와 'Sad'로 상관관계는 -0.444이다. 이를 Parallel coordinates를 통해 확인 한 결과는 다음과 같다. 'Sad'에 대한 구간 값이 0.175이하인 경우 'Happy'와 'Sad'의 감정 어휘 값은 0.341와 0.110, 'Sad'에 대한 구간 값이 0.175부터 0.35인 경우 'Happy'와 'Sad'의 감정 어휘 값은 0.226와 0.224, 'Sad'에 대한 구간 값이 0.35 이상인 경우 'Happy'와 'Sad'의 감정 어휘 값은 0.110와 0.433이었다. 이는 'Sad'의 값이 높아질 때 'Happy'의 값은 낮아진다고 볼 수 있으며 분석 결과는 <그림 11>과 같다.



<그림 11> Sad값이 0.175이하(왼쪽), 0.175이상 0.35이하(가운데), 0.35이상(오른쪽)일 때의 분석 결과

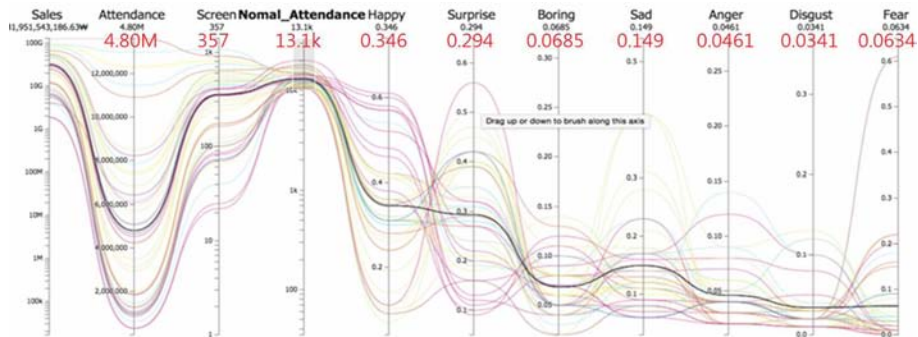
마지막으로 모든 영화 장르에서 값이 높게 나왔던 'Happy'와 'Surprise'의 상관관계를 분석한 결과 부(不)의 방향으로 상관관계(-0.286)가 있는 것을 확인하였다. 이를 Parallel coordinates를 통해 확인 한 결과는 다음과 같다. 'Surprise'에 대한 구간 값이 0.2이하인 경우 'Happy'와 'Surprise'의 감정 어휘 값은 0.330과 0.144, 'Surprise'에 대한 구간 값이 0.2부터 0.4인 경우 'Happy'와 'Surprise'의 감정 어휘 값은 0.324와 0.282, 'Surprise'에 대한 구간 값이 0.4이상인 경우 'Happy'와 'Surprise'의 감정 어휘 값은 0.203와 0.468이었다. 이는 'Surprise'의 값이 높아질 때 'Happy'의 값은 낮아진다고 볼 수 있으며 분석 결과는 <그림 12>와 같다.



<그림 12> Surprise값이 0.2이하(왼쪽), 0.2이상 0.4이하(가운데), 0.4이상(오른쪽)일 때의 분석 결과

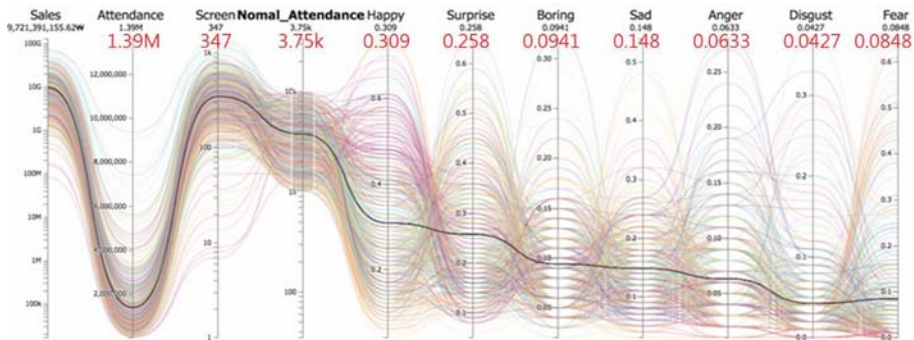
4-3-3. 영화 흥행도에 영향을 미치는 대표 감정 어휘 분석

본 연구에서는 누적 관객 수를 상영 스크린 수로 나누어 한 스크린에서의 누적 관객 수를 영화 흥행도로 사용하였으며 모든 장르의 영화를 대상으로 흥행도 값에 조건을 주어 영화의 흥행도와 감정 어휘의 관계를 분석한 결과는 다음과 같다.



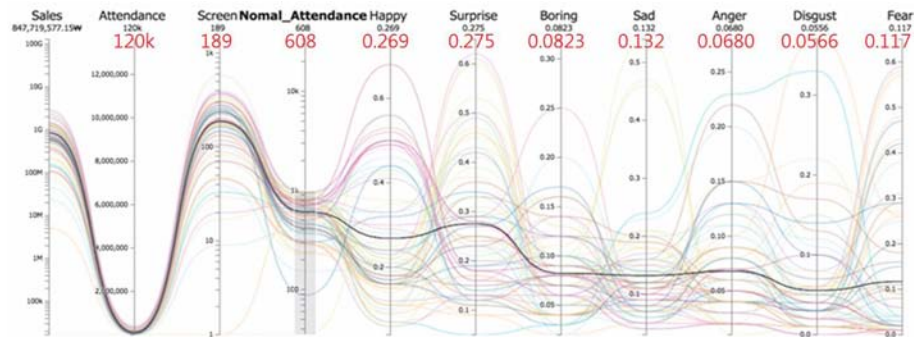
<그림 13> 영화 흥행도가 10,000명 이상인 경우 분석 결과

영화 흥행도가 10,000명 이상인 경우, <그림 13>과 같이 조건에 해당하는 영화는 41개의 영화였으며 대표 감정 어휘의 분포는 'Happy'(0.346), 'Surprise'(0.294), 'Sad'(0.149), 'Boring'(0.0685), 'Fear'(0.0634), 'Anger'(0.0461), 'Disgust'(0.0341)의 순서로 영화의 흥행도가 높을 때 값이 높게 분포하는 대표 감정 어휘는 행복한 감정과 놀라운 감정이라는 결과를 도출하였다.



<그림 14> 영화 흥행도가 10,000명 이하이고 1,000명 이상인 경우 분석 결과

영화 흥행도가 10,000명 이하이고 1,000명 이상인 경우, <그림 14>과 같이 조건에 해당하는 영화는 556개의 영화였으며 대표 감정 어휘의 분포는 'Happy'(0.309), 'Surprise'(0.258), 'Sad'(0.148), 'Boring'(0.0941), 'Fear'(0.0848), 'Anger'(0.0633), 'Disgust'(0.0427)의 순서로 값이 높게 분포하는 대표 감정 어휘는 조건이 흥행도가 10,000명 이상인 경우일 때와 같다는 결과를 도출하였다.



<그림 15> 영화 흥행도가 1,000명 이하인 경우 분석 결과

영화 흥행도가 1,000명 이하인 경우, <그림 15>와 같이 조건에 해당하는 영화는 75개의 영화였으며 해당 대표 감정 어휘의 분포는 'Surprise'(0.275), 'Happy'(0.269), 'Sad'(0.132), 'Fear'(0.117), 'Boring'(0.0823), 'Anger'(0.0680), 'Disgust'(0.0566)의 순서로 영화의 흥행도가 낮을 때 값이 높게 분포하는 대표 감정 어휘는 놀라운 감정과 행복한 감정이었으며 영화의 흥행도가 낮은 경우 무서운 감정의 값이 다른 조건들에 비해 높다는 것을 확인하였다.

5. 결론 및 제언

영화 산업의 성장과 함께 영화 흥행에 영향을 미치는 요인을 도출하기 위한 다양한 연구가 진행되어 왔으며 이제까지의 연구들을 통해 감독의 명성, 제작, 유통사의 역량, 영화 평론가의 리뷰, 감독, 작가, 제작자 등 영화와 관련된 관계자들의 역할 및 상호간의 연결 등이 영화의 성공 여부를 결정할 수 있는 요인으로 밝혀졌다. 하지만 객관적인 정보에 의한 요인만으로 영화 흥행과의 관계를 규명하기에는 한계가 있으며 주관적인 정보에 의한 요인과 흥행 사이의 관계를 규명하는 연구의 필요성이 제기되었다.

본 연구는 이러한 상황에서 주관적인 정보에 의한 요인과 영화 흥행과의 관계를 규명하고자 하였으며 특히 영화의 장르별로 관객들이 느끼는 감정 어휘의 분포를 탐색하고 영화의 흥행도와 감정 어휘 사이의 관계를 파악하는데 연구 목적을 두고 진행하였다.

최종적으로 7개의 대표 감정 어휘, 영화 티켓 판매액, 영화 관람 관객 수, 상영 스크린 수, 한 스크린 당 영화 관람 관객 수, 영화의 장르, 영화의 영문 이름 등으로 이루어진 데이터를 분석에 활용하였으며 분석 도구로는 Parallel coordinates 시각화 분석 방법을 사용하였다. 본 연구를 통해 도출된 결과는 다음과 같다.

첫째, 영화의 흥행도 값이 가장 높은 장르와 가장 낮은 장르를 확인 한 결과, 영화의 흥행도가 제일 높은 장르는 코미디(4,860명)였고, 대표 감정어의 분포는 'Happy'(0.465), 'Surprise'(0.213), 'Sad'(0.122), 'Boring'(0.0849), 'Anger'(0.0561), 'Disgust'(0.0397), 'Fear'(0.0177)순이었다. 다음으로 영화의 흥행도가 제일 낮은 장르는 호러(2,480명)였고, 대표 감정어의 분포는 'Fear'(0.404), 'Surprise'(0.176), 'Happy'(0.120), 'Sad'(0.113), 'Disgust'(0.0800), 'Boring'(0.0573), 'Anger'(0.0495)순이었다. 코미디 영화 중 하나인 '스윙걸즈'의 리뷰는 "너무나 깔끔하게 유쾌한 영화", "정말 재미있고 풋풋하네요."등 행복한 감정을 나타내는 리뷰들이 있었고, 호러 영화 중 하나인 '새벽의 저주'의 경우 "좀비영화 중 이보다 무서운 영화는 없는 듯 단연 최고", "무서워서 지릴뻔했음"등 무서운 감정을 나타내는 리뷰들이 있었다.

둘째, 대부분의 장르에서 'Happy'와 'Surprise'의 값이 가장 높았으나 각각의 대표 감정어휘 값이 가장 큰 장르는 다음과 같았다. 'Happy'(0.465)의 값이 가장 높은 장르는 코미디, 'Surprise'(0.333)의 값이 가장 높은 장르는 드라마, 'Boring'(0.0849)의 값이 가장 높은 장르는 판타지, 'Sad'(0.122)의 값이 가장 높은 장르는 드라마, 'Anger'(0.0561)의 값이 가장 높은 장르는 SF, 'Disgust'(0.0397)의 값이 가장 높은 장르는 범죄, 'Fear'(0.0177)의 값이 가장 높은 장르는 호러인 것을 확인 하였다. 이를 통해 대부분의 장르에서 'Happy'와 'Surprise'의 값이 높게 나오지만 다른 장르들에 비해 판타지 장르의 영화는 지루한 감정이 많이 느껴지고, SF장르의 영화는 화나는 감정이 많이 느껴진다는 결과를 도출하였다. 판타지 영화 중 하나인 '해리포터와 불사조 기사단'의 리뷰는 "시리즈 중 가장 지루한 듯함", "재미도 없는데 지루하기까지 함"등 지루한 감정을 나타내는 리뷰들이 있었고, SF 영화 중 하나인 '엘리시움'의 경우 "배우들이 아깝다. 쓰레기 영화.", "예고편에 속은 듯 욕 나온다."등 화나는 감정을 나타내는 리뷰들이 있었다.

셋째, 영화의 대표 감정 어휘 사이의 상관관계를 분석한 결과, 정(正)의 상관관계 값이 가장 높은 감정 어휘는 'Disgust'와 'Anger'(0.514)였고, 'Disgust'의 값이 높아짐에 따라 'Anger'의 값도 높아지는 것을 확인하였다. 다음으로 부(不)의 상관관계 값이 가장 높은 감정 어휘는 'Happy'와 'Sad'(-0.444)였고, 'Sad'이 값이 높아질 때 'Happy'의 값은 낮아지는 것을 확인하였다. 또한 'Happy'와 'Surprise'(-0.286)도 부(不)의 상관관계를 가졌으며 이를 통해 대부분의 장르에서 'Happy'와 'Surprise'의 감정이 많이 분포 하지만 'Happy'의 값이 높아질 때 'Surprise'의 값은 낮아진다는 결과를 확인 할 수 있었다.

넷째, 모든 장르를 대상으로 영화의 흥행도와 감정 어휘의 관계를 분석하여 확인한 결과, 영화 흥행도가 10,000명 이상인 경우, 대표 감정어휘의 분포는 'Happy'(0.346)와 'Surprise'(0.294)가 가장 높았다. 다음으로 영화 흥행도가 10,000명 이하이고 1,000명 이상인 경우에도 마찬가지로 'Happy'(0.309), 'Surprise'(0.258)가 가장 높았다. 마지막으로 영화 흥행도가 1,000명 이하인 경우, 위의 경우와 마찬가지로 'Surprise'(0.275), 'Happy'(0.269)의 값이 가장 높았으며 'Fear'(0.117)의 경우 다른 조건보다 높은 값을 가졌다는 점에서 두 조건과 상이하였다. 이를 통해 영화 흥행도의 값 높아질 때 'Happy'값도 같이 높아졌고, 영화 흥행도 값이 낮아질 때 'Fear'값은 높아진다는 결과를

확인하였다.

기존의 선행 연구들은 객관적인 정보에 의한 요인을 중심으로 흥행도에 영향을 미치는 요인을 탐색하는 방향으로 이루어 졌지만 본 연구는 주관적인 정보에 의한 요인 중 하나인 영화를 보고 관객들이 느끼는 감정과 영화 흥행과의 관계를 규명하기 위해 영화 장르별로 감정 어휘의 분포를 탐색하고 영화의 흥행도와 감정 어휘의 관계를 파악하였다는데 시사점이 있다. 하지만 감정 어휘 값을 생성 할 때 영화 리뷰 안의 명사 표현만으로 데이터를 생성하였다는데 한계점이 있다. 따라서 향후에는 관용 표현 처리 방법, 논항 구조 파악 방법, 강도에 따른 인식 방법에 대한 형태소 분석 연구를 진행하여 본 연구에 도입하여야한다.

참고문헌

논문

- Adel Ahmed, Vladimir Batagelj, Xiaoyan Fu, Seok-Hee Hong, Damian Merrick, Andrej Mrvar, "Visualisation and Analysis of the Internet Movie Database", Asia-Pacific Symposium on Visualisation 2007 IEEE, 2007.
- Litman, B, "Predicting Success of Theatrical Movies: An Empirical Study", Journal of Popular Culture, 16 (Spring), pp.159-175, 1983.
- Hyoji Ha, Wonjoo Hwang, Hanmin Choi, Gi-nam Kim, Hansung Kang, Kyungwon Lee, "CosMovis: Semantic Network Visualization on Sentiment Words for Movie Recommendation System", VIS 2014 Computer society IEEE, 2014.
- Inselberg, A, "The plane with Parallel coordinates", The Visual Computer, pp. 69-91, 1985.
- Parrott, W, "Emotions in Social Psychology", Philadelphia: Psychology Press, 2001.
- Rolf A. Zwaan, Mark C. Langston and Arthur C. Graesser, "The Construction of Situation Models in Narrative Comprehension: An Event-Indexing Model", Psychological Science, Vol. 6, No. 5, pp.292-297, 1995.
- 김연형, 홍정한, "영화 흥행 결정 요인과 흥행 성과 예측 연구", 한국통계학회논문집, 제18권, 제6호, pp.859-869, 2011.
- 박승현, 송현주, 정완규, "한국영화의 흥행성과 결정 요인에 관한 연구", 언론과학연구, 제11권, 제4호, pp.231-258, 2011.
- 성영신, 박진영, 박은아, "온라인 구전 정보가 영화 관람 의도에 미치는 영향", 광고연구, 제57권, pp.31~52, 2002.
- 김휴종, "한국영화스타의 스타파워분석", 삼성경제연구소 연구보고서, 1997.
- 이준웅, 송현주, 나은경, 김현석, "정서 단어 분류를 통한 정서의 구성 차원 및 위계적 범주에 관한 연구", 한국 언론 학보, 제52(1)권, pp.85-116, 2008.
- 김정호, 김명규, 차명훈, 인주호, 채수환, "한국어 특성을 고려한 감성 분류", 감성 과학, 제13(3)권, pp.449-458, 2010.
- 하효지, 김기남, 이경원, "영화 리뷰의 감정 어휘 공간 및 영화 관람의 상황분석 연구", 디자인 융복합 학회, 제12(6)권, pp.43-59, 2013.
- 한덕웅, 강혜자, "한국어 정서 용어들의 적절성과 경험 빈도", 한국 심리학회지, 제19권, pp.78-98, 2000.
- 박지연, 전범수, "네티즌의 흥행 영화 리뷰에 포함된 감정 동사 이용 특성 연구", 한국 콘텐츠 학회, 제14(5)권, pp.85-94, 2014.
- 박승현, 송현주, "영화의 흥행성과 제작비 규모와의 관계 : 2011년 한국영화의 흥행 결정 요인 분석", 사회 과학 연구 제51권 1호 , pp. 45~79, 2012.

인터넷 사이트

- 위키 백과사전, Parallel coordinate의 정의, <http://ko.wikipedia.org/>
- 영화진흥위원회 , <http://www.kobis.or.kr/kobis/business/mast/mvie/searchMovieList.do>