

공연 포스터의 이미지 특성을 활용한 딥러닝 기반 관객예측

조유정, 강경표*, Yao hui**, 권오병***
경희대학교, *경희대학교, **경희대학교, ***경희대학교

yujung251@khu.ac.kr, *kpkang0646@naver.com,
zgyh97y@khu.ac.kr, *obkwon@khu.ac.kr

Deep Learning-Based Audience Prediction Using Poster Image Characteristics

Yu Jung Cho, Kyung Pyo Kang *, Yao Hui**, Oh byung Kwon***
Kyung Hee Univ., * Kyung Hee Univ., ** Kyung Hee Univ., *** Kyung Hee Univ.

요약

공연예술 기관에서의 공연에 대한 흥행 예측은 공연예술 산업 및 기관에서 매우 흥미롭고도 중요한 문제이다. 이를 위해 출연진, 공연장소, 가격 등 정형화된 데이터를 활용한 전통적인 예측방법론, 데이터 마이닝 방법론이 제시되어 왔다. 그런데 관객들은 공연안내 포스터에 의하여 관람 의도가 소구되는 경향도 있음이 분명함에도 불구하고, 포스터 이미지 분석에 의한 흥행 예측은 거의 시도되지 않았다. 최근에는 이미지를 통해 판별하는 CNN계열의 딥러닝 방법이 개발되면서 포스터 분석의 가능성이 열렸다. 이에 본 연구의 목적은 공연관련 포스터 이미지를 통해 흥행을 예측할 수 있는 딥러닝 방법을 제안하는 것이다. 이를 위해 KOPIS에 공개된 포스터 이미지를 학습데이터로 하여 Pure CNN, VGG-16, Inception-V3, ResNet50 알고리즘을 통해 예측을 수행하였다. 그 결과 흥행예측 정확도 80% 이상의 높은 성과를 보였다. 또한 공연 관련 정형데이터를 활용한 전통적 방법론과의 앙상블을 시도하였다. 본 연구는 공연예술 분야에서 이미지 정보를 활용하여 흥행을 예측하는 첫 시도이며, 본 연구에서 제안한 방법은 연극 외에 영화, 기관 홍보, 기업 제품 광고 등 포스터 기반의 광고를 하는 영역으로도 적용이 가능할 것이다.

I. 서론

공연예술 기관에서의 콘텐츠에 대한 흥행 예측은 공연예술 산업 활성화에 중요한 이슈이다. 과거 흥행 예측 방법으로는 전통적인 예측 기법인 다중회귀분석 [1][2]과 Bass모형[2]외에 Random Forest, KNN, 인공신경망[3] 등 데이터 마이닝 알고리즘 등이 제안되어 왔다.

지금까지의 흥행 예측 연구는 대부분 출연진[4]이나, 공연장소[5][6], 일반인 평가[4], 전문가 평가[7] 등 정형적인 공연 특성에 의한 것이었다. 그러나 최근 이미지 기반의 판별 문제가 CNN 등 딥러닝 알고리즘에 의하여 해결 가능해짐으로써 공연 포스터와 같은 이미

지로도 흥행을 예측하는 시도가 가능 해졌다.

공연예술 분야의 흥행 관련 연구는 성과 예측 모형을 직접 제시하는 것보다는 연극의 장르(유희적, 교육적, 교훈적 연극), 원작의 유무, 연극의 속성[8]등 연극의 특성이 흥행에 미치는 연구가 대부분이다[9]. 예측 모형 연구가 활발하지 않은 이유는 영화만큼 예측 관련 데이터가 풍부하지 않고[8], 설문조사나 면접 방법 등을 통해 자료를 확보하기 때문에 성능이 좋은 예측 모형을 만들기가 용이하지 않기 때문이다.

이에 본 연구에서는 비정형 데이터인 공연관련 포스터 이미지를 통해 흥행을 예측해보고, 정형화된 공연 특성 데이터만으로 예측한 것과 성능을 비교해 봄으로

써 비정형 데이터인 연극 공연 포스터로 연극의 흥행 여부를 판단할 수 있는지와 정형화된 공연 관련 데이터와 포스터 이미지 특성을 복합적으로 활용하여 흥행을 예측하였을 때 어느 것이 더 우수한 성능을 나타내는지 알아보고자 한다.

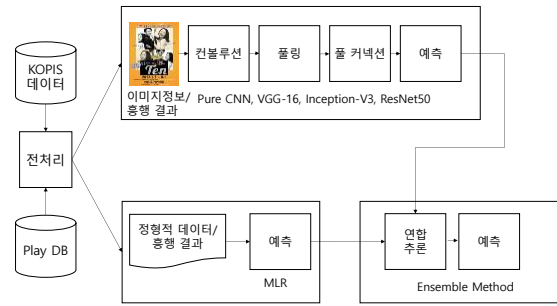
II. 본론

본 연구에서 공연 포스터 이미지를 기반으로 흥행을 예측하기 위해 <그림1>과 같이 예측을 수행하고자 한다. 먼저 복수의 오픈 데이터를 수집하고 전처리를 한 후에 포스터 이미지 데이터에 대해서는 복수의 CNN 계열 알고리즘들로 흥행을 예측하고, 정형화된 데이터는 회귀분석을 통해 흥행을 예측한다. 그리고 두 가지의 흥행 예측 결과를 바탕으로 앙상블 기법을 수행한다. 연극공연에 대한 정보는 공연예술통합전산망인 KOPIS(<http://www.kopis.or.kr>)에서 API key를 발급받아 연극정보, 연극상세정보, 연극포스터 등을 확보하였다.

학습에 사용되는 모델은 CNN을 기반으로 한 모델이다. 수집된 데이터는 추정 관객수를 기준으로 관객수가 500명 미만일 경우 흥행실패, 500명 이상일 경우 흥행성공으로 분류하였다. 추정 관객수는 공연횟수와 1회 당 최대 유치 관객수, 즉 공연장 규모의 곱으로 구했다. 따라서 CNN 모델은 흥행 실패와 흥행성공 두 개의 클래스로 분류되도록 모델을 설계하였다. CNN 모델은 각각 3개의 Convolution layer와 maxpooling 층으로 이루어진 Pure CNN과 Python의 패키지 Keras의 내장 사전 훈련 모델로 있는 VGG16, Inception-v3, Resnet50 모델을 사용하였다.

III. 결론

공연예술 분야는 인공지능, 특히 딥러닝이 적용될 수 있는 유력한 분야임에도 불구하고, 그동안 많은 연구가 진행되지 못했다. 이에 본 연구는 연극 분야가 보유하고 있는 이미지를 포함한 오픈 빅데이터를 토대로 연극의 활성화를 위한 예측과 의사결정에 도움이 되는 딥러닝 방법을 제안했다는 데 의미가 있다.



<그림 1> 전체적인 프로세스

ACKNOWLEDGMENT

이 논문 또는 저서는 2020년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2020S1A3A2A02093277)

참고 문헌

- [1] 김상호, 한진만. (2014). “한국 영화의 흥행성과 결정요인 분석.” 사회과학연구, 53(1), pp.191-214.
- [2] 김보경, & 임창원. (2018). “GLS 와 Bass 모형을 결합한 하이브리드 모형을 이용한 영화 관객 수 예측.” 응용통계연구, 31(4), pp.447-461.
- [3] 정찬미. (2020). “영화 관객 수 예측을 위한 머신러닝 기법의 성능 평가 연구.” 석사학위논문, 이화여자대학교
- [4] 김연형, 홍정한 (2011). “영화 흥행 결정 요인과 흥행 성과 예측 연구.” 한국통계학회논문집, 18(6), pp.859-869.
- [5] 권선주. (2014). “영화 흥행성과의 분석과 예측: 뉴스와 웹사이트 데이터 이용.” 문화경제연구, 17(1), pp.35-55.
- [6] 우종필, & 이용환. (2018). “빅데이터 분석을 통한 천만 관객 영화 예측 모델.” 한국빅데이터학회지, 3(1), pp.63-71.
- [7] 전범수, & 이준영. (2019). “한국 개봉 흥행 영화 평점이 성과에 미치는 영향: 일반 이용자와 전문가 평점의 비교.” 언론과학연구, 19(4), pp.227-253.
- [8] 송은아. (2013). “장기공연 연극의 특성에 관한 연구.” 한국엔터테인먼트산업학회 학술대회 논문집, pp.132-136.
- [9] 김태희, & 신형덕. (2013). “어린이공연의 내용, 장르, 원작유무가 공연 흥행에 미치는 영향: 등교기간의 조절효과를 중심으로.” 한국산학기술학회 논문지, 14(10), pp.4762-4768.