JKSCI

한국컴퓨터정보학회논문지
Journal of The Korea Society of Computer and Information
Vol. 31 No. 3, pp. 139-147, March 2026
https://doi.org/10.9708/jksci.2026.31.03.139

A Korean Movie Genre Prediction Model Using Multi-Representation Learning

Jong-Hyun Kim*

*Associate Professor, College of Software and Convergence (Dept. of Design Technology), Inha University, Incheon, Korea

[Abstract]

In this study, we propose a multi-representation learning model that integrates heterogeneous information—movie poster images, audience reviews, and textual phrases extracted from posters—to predict the genres of Korean films. While visual features are learned from posters using a CNN, relying solely on posters may fail to fully capture genre context, especially when discrepancies exist between the director's intended genre and the audience's perceived genre. To address this limitation, we additionally construct a Word2Vec-based review analysis model and an LSTM-based text genre prediction model using OCR-extracted poster phrases. By integrating the genre probabilities from the three models, the proposed approach achieves approximately 75% overall stable performance, effectively compensating for genre-specific errors compared to the single poster-based CNN model (approximately 78%). The proposed multimodal framework enhances the interpretability and reliability of genre prediction, and future work will focus on expanding genre categories and refining the weighting of OCR-based textual features to further improve performance.

▸ **Key words**: Movie genre prediction, Multi-representation learning, Convolutional Neural Network, Natural Language Processing

[요 약]

본 연구에서는 영화 포스터 이미지, 관람객 리뷰, 포스터 내 문구 등 서로 다른 표현 정보를 통합해 한국 영화 장르를 예측하는 다중 표현 학습 모델을 제안한다. 포스터는 CNN으로 시각적 특징을 학습하되, 포스터만으로 장르 맥락을 반영하기 어렵고 감독 의도와 관객 경험이 달라질 수 있다는 한계를 고려해 리뷰(Word2Vec)와 OCR 문구(LSTM) 기반 텍스트 모델을 추가로 구성했다. 세 모델의 장르 확률을 통합한 결과, 단일 포스터 CNN(약 78%) 대비 특정 장르에서 오류를 보완하며 전체적으로 약 75%의 안정적 성능을 보였다. 제안 모델은 멀티모달 결합을 통해 예측의 해석성과 신뢰도를 높였으며, 향후 장르 확장과 OCR 문구 가중치 조정으로 성능 개선을 목표로 한다.

▸ **주제어**: 영화 장르 예측, 다중 표현 학습, 합성곱 신경망, 자연어 처리

• First Author: Jong-Hyun Kim, Corresponding Author: Jong-Hyun Kim
*Jong-Hyun Kim (jonghyunkim@inha.ac.kr), College of Software and Convergence (Dept. of Design Technology), Inha University
• Received: 2026. 01. 02, Revised: 2026. 02. 10, Accepted: 2026. 02. 20.

http://www.ksci.re.kr pISSN:1598-849X | eISSN:2383-9945

I. Introduction

영화 장르는 관람객이 영화를 선택하는 과정에서 가장 핵심적인 판단 요소 중 하나이며, 영화 산업 전반에서 마케팅 전략 수립, 관객 타깃팅, 추천 시스템 구축 등에 중요한 역할을 한다. 그러나 한 영화는 복수의 장르를 동시에 포함할 수 있으며, 장르의 경계가 명확하지 않은 경우가 많아 자동화된 장르 분류는 여전히 어려운 문제로 남아 있다. 특히 한국 영화의 경우, 감독이 의도한 메시지와 관객이 체감하는 분위기 사이에 차이가 발생하기도 하여 장르 예측의 모호성이 더욱 커지는 경향이 있다. 이러한 이유로 영화 포스터나 리뷰와 같은 다양한 표현 요소를 활용한 장르 분석 기법에 대한 연구가 지속적으로 이루어지고 있다[1-3].

영화 포스터는 제한된 공간에 주요 인물, 색감, 촬영 구도, 문구 등을 압축적으로 전달하는 시각적 매체로서 장르적 특징을 내포하고 있으며, 이미지 기반 장르 분류 연구에서 중요한 분석 단서로 활용되어 왔다. 합성곱 신경망(CNN)의 발전으로 포스터 이미지의 시각적 패턴을 학습하는 연구는 높은 성과를 보여 왔으나[1,4], 이미지 정보만으로는 영화의 서사적 맥락, 분위기, 감독 의도 등을 충분히 반영하기 어렵다는 근본적 한계가 존재한다. 특히 스릴러-드라마, 코미디-로맨스처럼 경계가 유동적인 장르에서는 포스터 이미지에만 의존할 경우 오분류 가능성이 높다.

이러한 한계를 보완하기 위해 자연어 기반 장르 분석 연구가 주목받기 시작했다. 관람객 리뷰는 영화에 대한 주관적 경험과 내용을 반영하고 있으며, 장르적 분위기나 영화적 특징을 간접적으로 드러내는 언어적 단서를 제공한다. 기존 연구에서는 시놉시스나 리뷰를 이용하여 장르를 분류하는 다양한 접근이 제안되었으며[5,6], 텍스트 기반 분석이 서사 정보와 감정 정보를 반영하는 데 효과적임이 보고되었다. 그러나 이러한 방법은 포스터에서 전달되는 시각적 분위기나 마케팅 의도를 활용하지 못한다는 한계를 가진다.

최근에는 포스터 이미지와 텍스트, 메타데이터 등을 함께 활용하는 멀티모달 기반 장르 분류 연구가 제안되고 있으며, 서로 다른 표현 정보를 결합할 경우 단일 모달 기반 방법보다 성능이 향상될 수 있음이 보고되었다[7,8], 그러나 기존 연구의 상당수는 대규모 영어권 데이터셋이나 예고편 영상 등 특정 형태의 데이터에 초점을 맞추고 있으며, 실제 관람객의 인식과 밀접하게 연결된 리뷰와 포스터 문구를 동시에 활용하여 한국 영화 장르를 분석한 연구는 상대적으로 부족한 상황이다.

또한 영화 장르는 제작사나 플랫폼에서 제공하는 공식 장르 정보에 의해 구분되는 경우가 많지만, 실제 관람객이 인지하는 분위기나 정서적 경험은 이러한 분류와 차이를 보이는 경우가 존재한다. 소비자 관점에서의 장르 예측은 이러한 인지적 차이를 반영함으로써 추천 시스템에서의 사용자 만족도 향상, 콘텐츠 탐색 과정에서의 직관적인 분류 지원, 그리고 관객 인식 기반의 콘텐츠 분석과 같은 응용 분야에서 보완적인 정보를 제공할 수 있다.

본 연구는 이러한 문제점을 해결하기 위해 영화 포스터 이미지, 관람객 리뷰, 포스터 내 문구라는 세 가지 이질적 표현 정보를 통합적으로 활용하는 다중 표현 학습(Multi-representation Learning) 기반 장르 예측 모델을 제안한다. 포스터 이미지는 CNN을 통해 시각적 특징을 학습하고, 리뷰 데이터는 Word2Vec 기반 자연어 처리 모델로 장르 관련 단어 분포를 분석하며, 포스터 문구는 OCR로 추출한 후 LSTM 기반 텍스트 모델을 활용하여 장르적 단서를 해석한다. 이후 세 모델에서 산출된 장르별 확률값을 결합하여 최종 장르를 예측함으로써 단일 모달리티 기반 모델의 한계를 보완하였다.

본 연구의 차별성과 의의는 다음과 같다. 첫째, 한국 영화 데이터를 대상으로 포스터 이미지, 관람객 리뷰, 포스터 문구라는 서로 다른 표현 정보를 동시에 활용하여 장르 예측을 수행하였다. 둘째, 소비자 반응 기반 정보와 시각 정보를 결합함으로써 공식 장르 분류에서 발생할 수 있는 인지적 차이를 보완하는 접근을 제시하였다. 셋째, 각 모달리티의 역할과 한계를 실험적으로 비교하고, 통합 모델이 오분류를 보완하는 과정을 분석함으로써 멀티 표현 기반 장르 예측의 실질적인 효과를 정량적·정성적으로 확인하였다. 이러한 점에서 본 연구는 한국 영화 환경에서의 멀티 표현 기반 장르 분석의 가능성을 제시한다는 의의를 가진다.

영화 장르는 제작사나 플랫폼에서 제공하는 공식 장르 정보에 의해 구분되는 경우가 많지만, 실제 관람객이 인지하는 분위기나 정서적 경험은 이러한 분류와 차이를 보이는 경우가 존재한다. 소비자 관점에서의 장르 예측은 이러한 인지적 차이를 반영함으로써 추천 시스템에서의 사용자 만족도 향상, 콘텐츠 탐색 과정에서의 직관적인 분류 지원, 그리고 관객 인식 기반의 콘텐츠 분석과 같은 응용 분야에서 보완적인 정보를 제공할 수 있다. 따라서 관람객의 반응과 언어적 표현을 기반으로 한 장르 추정은 기존 장르 체계를 대체하기보다는 이를 보완하는 유용한 접근 방식이라 할 수 있다.

II. Related Work

1. Movie Poster Image-Based Genre Classification

영화 포스터는 장르를 직관적으로 전달하는 시각 매체이기 때문에, CNN을 활용한 이미지 기반 장르 분류 연구가 활발히 진행되어 왔다. Chu와 Guo는 대규모 영화 포스터 데이터셋을 구축하고, 시각적 외형 및 객체 정보를 함께 학습하는 심층 신경망을 설계하여 멀티레이블 장르 분류 문제를 다루었다[1]. 국내에서도 CNN을 이용해 영화 포스터 이미지만으로 장르를 분류하는 연구가 보고되었으며, 한국 영화 포스터를 대상으로 한 멀티레이블 장르 예측 모델이 제안된 바 있다[2].

Sung 등은 포스터 이미지를 입력으로 하는 CNN 기반 분류기를 구현하여, 포스터에 내재된 시각적 패턴이 장르 분류에 유효함을 보였다[3]. Korai 등은 3만 장 이상의 RGB 포스터 이미지를 수집하여 28개 장르에 대한 다중 장르 분류를 수행함으로써, 대규모 데이터셋에서의 포스터 기반 학습 가능성을 검증하였다[4]. 또한, 영화 포스터의 색채 특성을 장르와 연관지어 분석한 연구도 다수 제안되었는데, 한국 영화 포스터를 대상으로 장르별 색채 경향을 비교한 연구[3]와 포스터 색상 분포를 이용해 박스오피스 성과를 예측하는 연구가 대표적이다[5].

그러나 이러한 연구들은 대부분 포스터 이미지 단일 모달리티에 의존한다는 한계를 가진다. 포스터는 장르적 분위기를 압축적으로 전달할 수 있지만, 이야깃거리나 서사 구조·감정선과 같은 정보는 제한적으로만 표현되기 때문에, 장르 간 경계가 모호한 경우에는 오분류가 빈번하게 발생할 수 있다. 이를 보완하기 위해 포스터 이미지에서 장르 간 상관관계를 고려하는 인터-레이블 상관 구조, 또는 포스터 내 얼굴·객체 영역을 별도로 추출하여 활용하는 구조 등이 제안되었으나, 여전히 이미지 기반 정보에 국한된다는 점에서 한계가 지적된다[6].

2. Text-Based Movie Genre Classification

텍스트 정보는 영화의 내용과 분위기를 보다 직접적으로 반영하기 때문에, 시놉시스·리뷰·스크립트 등을 활용한 텍스트 기반 장르 분류 연구 역시 활발히 진행되어 왔다. Battu 등은 영화 시놉시스만을 입력으로 하여 장르와 관람 등급을 동시에 예측하는 다중 출력 딥러닝 모델을 제안하였고, 장르를 9개 클래스로 재정의하여 멀티클래스 분류 문제로 다룬 바 있다[7]. Van der Lee는 자막 코퍼스를 대상으로 텍스트·구문·내용 특성을 복합적으로 추출하여 비디오 장르를 분류하는 연구를 수행하였다[8]. Cuomo는 영화 대본 텍스트를 활용해 장르를 분류하는 방법을 제안하며, 시놉시스나 리뷰보다 더 풍부한 서사 정보가 장르 분류 성능 향상에 기여할 수 있음을 보였다[9].

최근에는 트랜스포머(transformer)를 활용한 다중 장르 분류가 활발히 연구되고 있다. Shaukat 등은 플롯 텍스트를 입력으로 하는 멀티라벨 장르 분류 문제를 대상으로, 다중 트랜스포머와 계층적 어텐션을 결합한 장르 주의(Genre Attention) 기반 모델을 제안하여 장르 간 중첩을 효과적으로 처리하였다[10]. Zhang 등은 시놉시스를 중심으로 언어적 특징을 강화하는 언어 증강(language augmentation) 전략과 쇼트 샘플링을 결합하여, 비디오 기반 장르 분류에 텍스트 정보를 적극적으로 활용하는 프레임워크를 제안하였다[11].

이러한 텍스트 기반 연구들은 서사 정보와 감성 정보를 반영할 수 있다는 강점을 가지지만, 포스터에서 전달되는 시각적 분위기나 마케팅 의도와 같은 정보를 활용하지 못한다는 점에서 한계가 있다. 또한, 리뷰의 경우 관람객의 주관적 평가가 강하게 반영되어 감독의 의도와 장르 레이블 간 괴리가 발생할 수 있으며, 다국어·비정형 텍스트가 혼재하는 현실 환경에서 전처리 부담도 크다[7].

3. Multimodal/Multi-representation Movie Genre Classification

최근에는 포스터 이미지, 시놉시스, 예고편 영상, 메타데이터 등을 함께 활용하는 멀티모달(Multimodal) 영화 장르 분류가 주목받고 있다. Li 등은 포스터·플롯·예고편·메타데이터 등 다양한 모달리티를 통합하는 멀티모달 장르 분류에서, 도메인 지식 그래프를 결합한 자기지도(self-supervised) 어텐션 및 대조학습(contrastive learning) 프레임워크를 제안하여, 장르 중심 임베딩 공간을 형성함으로써 장르 간 구분성을 향상시켰다[12]. Vishwakarma 등은 영화 예고편의 시각·음향·메타데이터(줄거리 설명 등)를 함께 사용하여 인지적·정서적 특징을 추출하는 심층 네트워크를 설계하고, 트레일러 기반 다중 모달 장르 분류의 가능성을 제시하였다[13].

Sulun 등은 다양한 사전학습 모델에서 추출한 고수준 특징(장면, 객체, 텍스트, 음성, 음악, 효과음 등)을 경량 분류기와 결합하는 방식으로 예고편 기반 멀티모달 장르 분류를 수행하여, 모달리티 결합이 단일 모달 대비 유의미한 성능 향상을 가져올 수 있음을 보였다[14]. 한편, 포스터와 시놉시스를 함께 사용하는 멀티모달 장르 분류 모델도 제안되어, 포스터 이미지와 줄거리 요약을 동시에 입력으로 받는 심층 네트워크 구조가 구현되었다[15].

최근에는 다양한 모달리티의 정보를 결합하여 영화 장르를 분류하려는 연구가 활발히 이루어지고 있다. 특히 사전학습된 특징을 활용하여 시각, 음향, 텍스트 정보를 통합하는 접근이 제안되었으며, 이러한 방법은 단일 모달리티 기반 접근보다 향상된 장르 분류 성능을 보이는 것으로 보고되었다. 예를 들어 Sulun 등은 예고편 영상에서 추출한 시각, 음향, 텍스트 특징을 결합하여 장르를 분류하는 멀티모달 프레임워크를 제안하였으며, 사전학습된 특징을 활용하는 방식이 효율적인 장르 분류에 효과적임을 보였다[15].

또한 최근 연구에서는 대규모 사전학습 모델과 멀티모달 표현 학습을 결합하여 장르 중심의 임베딩 공간을 형성하고, 장르 간 구분성을 향상시키는 방법도 제안되었다. Li 등은 도메인 지식 그래프와 자기지도 학습을 결합한 멀티모달 장르 분류 프레임워크를 제안하여 다양한 영화 데이터에서 성능 향상을 보고하였다[13]. 이러한 연구들은 서로 다른 형태의 정보를 통합하는 접근이 영화 장르 분류 문제에서 중요한 연구 방향임을 보여준다.

본 연구 역시 이러한 흐름을 반영하여 영화 포스터 이미지, 관람객 리뷰, 포스터 문구라는 서로 다른 표현 정보를 결합하는 다중 표현 학습 기반 장르 예측 방법을 제안한다는 점에서 최근 연구 흐름과 맥락을 같이한다.

III. The Proposed Scheme

1. Poster Genre Prediction Network via CNN

본 연구에서 사용한 CNN 기반 포스터 장르 예측 네트워크 구조는 Fig. 1과 같다. 제안한 네트워크는 선행 연구에서 검증된 구조를 참고하여 설계하였으며, 입력 포스터 이미지를 받아 5개의 Convolution-Max Pooling 계층을 순차적으로 통과시킨 뒤, Global Average Pooling과 Fully Connected Layer를 거쳐 최종 장르 예측값을 출력한다. Convolution 계층은 모두 동일한 크기의 필터를 사용하되, 네트워크의 깊이가 증가함에 따라 채널 수를 점진적으로 증가시켜 고수준 특징을 학습하도록 구성하였다. Max Pooling 역시 동일한 필터 크기를 적용하여 공간 해상도를 점차 줄이면서 핵심적인 시각 패턴만을 남기도록 하였다. 마지막으로, 과적합을 완화하고 파라미터 수를 줄이기 위해 Global Average Pooling을 적용한 뒤, 이를 Fully Connected Layer에 입력하여 장르별 확률을 산출하였다[1].

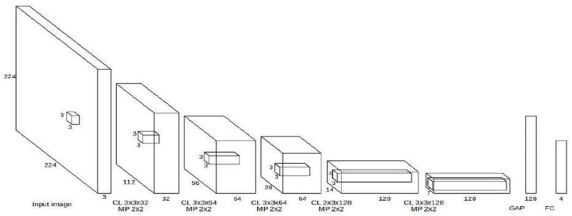

Fig. 1. CNN structure.

데이터셋은 네이버 영화 사이트에서 크롤링을 통해 구축하였다. 한국 영화를 대상으로 관람객 평가 인원이 300명 이상인 작품만을 선별하여 총 1,311개의 영화 포스터 데이터를 수집하였다. 각 영화는 복수의 장르를 가질 수 있으므로, 유사한 장르를 통합하여 '로맨스, 공포, 드라마, 스릴러, 액션, 코미디'의 6개 장르로 재분류하였으며, MultiLabelBinarizer를 사용해 멀티레이블-멀티클래스 형태의 장르 벡터를 생성하였다. 전체 1,311개 샘플 중 1,045쌍을 학습용 데이터로, 266쌍을 시험용 데이터로 사용하였다. 장르 분류는 멀티레이블 특성을 가지므로 손실함수로는 Binary Cross-Entropy를 사용하였고, 최적화 알고리즘은 Adam을 채택하였다.

본 연구에서는 리뷰 기반 분석의 안정성과 최소 표본 신뢰도를 확보하기 위해 관람객 평가 인원이 300명 이상인 영화만을 대상으로 선정하였다. 평가 인원이 적은 경우 리뷰 수가 부족하거나 소수 의견에 의해 텍스트 표현이 편향될 가능성이 크므로, 일정 규모 이상의 관람객 반응이 축적된 작품으로 데이터셋을 구성하였다. 다만 이 기준은 상대적으로 인지도가 높은 작품이 포함될 가능성을 높여 장르 분포 편향이 발생할 수 있으며, 향후에는 임계값 변화에 따른 민감도 분석 및 장르별 표본 균형화 등을 통해 이를 완화할 계획이다.

Fig. 2. Test results for poster genre prediction.

모델의 성능 평가는 Binary Accuracy를 기준으로 수행하였으며, 약 78%의 정확도를 얻었다. Fig. 2는 6개 장르 클래스에 대해 예측된 확률값 중 상위 3개의 장르를 예시로 나타낸 결과이다. 또한, 네트워크가 포스터의 어떤

영역에 주로 주목하는지를 분석하기 위해 Class Activation Map(CAM)을 생성하여 시각화하였다. 그 결과, 배우의 얼굴이나 인물 클로즈업 영역, 포스터의 어두운 색감이 두드러지는 부분 등 장르를 암시하는 시각적 요소에 집중하는 경향을 확인할 수 있었다.

본 연구에서 사용한 정답 장르 레이블은 네이버 영화에서 제공하는 장르 정보를 기준으로 구축한 것이다. 해당 장르 정보는 제작사 및 배급사의 공식 분류를 기반으로 하지만, 실제 관람객 리뷰와 평가가 함께 축적되는 공개 플랫폼에서 널리 사용되는 기준이라는 점에서 실험을 위한 공통적인 참조 레이블로 활용하였다. 본 연구에서 제안하는 소비자 관점 장르 예측은 이러한 공식 장르 체계를 대체하기 위한 것이 아니라, 관람객 리뷰와 포스터 문구 등 소비자 반응 기반 정보를 활용하여 공식 장르 기준에서 발생할 수 있는 인지적 차이나 오분류 사례를 보완하는 것을 목적으로 한다. 따라서 본 연구에서의 장르 레이블은 평가를 위한 기준으로 사용되었으며, 제안 방법의 의의는 소비자 관점 정보를 추가적으로 반영함으로써 장르 해석의 안정성과 활용 가능성을 높이는 데 있다.

2. Review Genre Prediction Network via NLP

본 연구에서 사용한 영화 리뷰 기반 장르 예측 네트워크 구조는 Fig. 3과 같다.

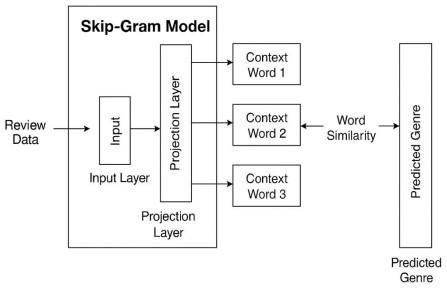

Fig. 3. Architecture of the Skip-Gram-based review genre prediction model.

영화별 리뷰 텍스트를 입력으로 받아 Word2Vec의 Skip-gram 학습 방식으로 사전에 학습된 임베딩 모델을 구성하고, 이를 바탕으로 각 장르별 대표 단어와의 유사도를 계산하여 장르를 예측한다. Skip-gram은 중심 단어를 입력으로 주변 단어를 예측하는 방식으로 단어 임베딩을 학습하는 기법이며, 본 연구에서는 기존에 공개된 영화 리뷰 코퍼스를 활용하여 Word2Vec 임베딩 모델을 학습하였다. 이후 각 영화에 대한 리뷰 벡터와 장르별 단어 리스트의 평균 벡터 간 코사인 유사도를 계산하고, 유사도가 가장 높은 상위 2개의 장르를 해당 영화의 예측 장르로 출력한다.

리뷰 데이터는 포스터 데이터와 동일한 1,311편의 한국 영화에 대해, 네이버 영화 사이트에서 각 영화당 300개씩 크롤링하여 수집하였다. 장르별 대표 단어 리스트를 구축하기 위해 Sometrend에서 각 장르 키워드(로맨스, 공포, 드라마, 스릴러, 액션, 코미디)를 검색하여 상위 연관 단어 20개씩을 수집하였다. 수집된 텍스트 데이터에 대해서는 정규표현식을 이용해 한글 이외의 문자를 제거하고, 불용어(stopwords)를 제거한 후, 형태소 분석기 OKT를 사용하여 토큰화 작업을 수행하였다. Word2Vec 모델 학습에는 공개된 영화 리뷰 데이터셋인 ratings[2]를 사용하였으며, 임베딩 차원은 100, 윈도우 크기는 5, 단어 최소 빈도수는 5, 학습에 사용된 프로세스 수는 4로 설정하였다.

Fig. 4. Test results for review genre prediction.

학습된 리뷰 기반 장르 예측 네트워크를 이용하여 실제 영화 장르와 예측 결과 간의 일치도를 평가한 결과, 약 62%의 정확도를 얻었다. Fig. 4는 6개 장르 클래스에 대해 계산된 장르별 유사도 중 상위 2개의 장르를 예시로 나타낸 것이다. 장르별 대표 단어 리스트의 크기가 상대적으로 작기 때문에, 각 장르 간 유사도 값의 차이가 크지 않다는 한계도 관찰되었으며, 이는 향후 장르별 단어 풀 확장 및 가중치 조정을 통해 개선할 수 있을 것으로 판단된다.

3. Poster Writing Genre Prediction Network through NLP

본 연구에서 사용한 영화 포스터 글귀 기반 장르 예측 네트워크 구조는 Fig. 5와 같다. 영화 포스터에서 OCR을 통해 추출한 문구를 입력으로 받아 임베딩 층을 거친 뒤, 다대일 구조의 LSTM 네트워크를 통해 장르를 예측하도록 구성했다. 임베딩 벡터의 차원 수는 100으로 설정했으며, LSTM 은닉 상태의 크기는 128로 설정하여 포스터 문구에 내재된 시퀀스 정보를 효과적으로 학습하도록 하였다.

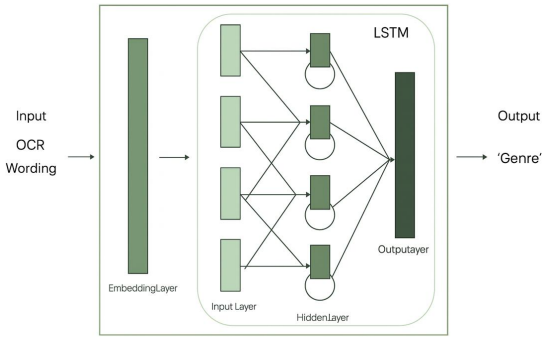

Fig. 5. Architecture of the LSTM-based genre prediction network using OCR-extracted poster wording.

포스터 문구 데이터는 총 1,311장의 영화 포스터에 대해 네이버 클로바 OCR 프로그램을 사용하여 추출하였으며, 인식된 결과는 JSON 형식으로 저장하였다. OCR 과정에서 포스터 내 모든 텍스트가 인식되므로, 영화 내용 및 장르와 직접적인 관련이 없는 문구(제작사 로고, 배급사 정보 등)는 후처리 과정에서 제거하였다. 이후 형태소 분석기 OKT를 이용해 토큰화를 수행하고, 각 단어를 정수로 인코딩하였다. 단어 집합 크기는 4,000으로 제한하였으며, 문장 길이는 최대 16 단어로 고정하여 길이가 짧은 문장은 패딩 처리하였다. 학습 시 손실 함수는 다중 클래스 분류에 적합한 categorical cross-entropy를 사용하였고, 최적화 알고리즘으로는 Adam을 적용하였다.

또한, ModelCheckpoint 기법을 사용하여 검증 데이터의 정확도가 이전보다 향상될 때에만 최적의 가중치를 저장하도록 하여 과적합을 완화하고 학습 안정성을 높였다. 최종적으로 포스터 글귀 기반 장르 예측 모델은 약 68%의 정확도를 보였으며, Fig. 6은 6개 장르 클래스에 대해 예측된 확률값 중 상위 2개의 장르를 예시로 나타낸 것이다. 이 결과는 짧은 포스터 문구만을 이용하더라도 장르적인 분위기와 키워드를 어느 정도 포착할 수 있음을 보여준다.

(a) Drama (0.793), Romance (0.156) (b) Thriller (0.320), Action (0.269)

Fig. 6. Test results for poster writing genre prediction.

IV. Experiment and Results

본 절에서는 포스터 이미지 기반 CNN 모델, 영화 리뷰 기반 Skip-gram NLP 모델, 포스터 문구 기반 LSTM 모델, 그리고 세 예측 결과를 통합한 최종 다중 표현 학습 모델의 성능을 비교·분석하였다. 또한 모델이 특정 장르를 예측할 때 어떤 시각·언어적 단서를 활용하는지를 정성적으로 해석하고, 오분류 사례 분석을 통해 각 접근 방식의 장단점을 종합적으로 평가하였다.

본 연구는 영화가 복수의 장르를 동시에 가질 수 있는 멀티 레이블 분류 문제를 다루므로, 성능 평가는 각 장르에 대한 이진 분류 결과를 기준으로 정확도를 계산하는 방식을 사용하였다. 각 영화 i에 대해 실제 장르 벡터를 $y_i \in \{0,1\}^C$, 예측 장르 벡터를 $\hat{y_i} \in \{0,1\}^C$라 할 때, 전체 정확도는 다음과 같이 정의한다.

$$Accuracy = \frac{1}{N \times C} \sum_{i=1}^{N} \sum_{j=1}^{C} 1\left(y_{i,j} = \hat{y}_{i,j}\right) \qquad (1)$$

여기서 N은 전체 영화 수, C는 장르의 개수이며, $1(\bullet)$는 두 값이 일치할 때 1, 그렇지 않을 때 0을 번환하는 지시 함수이다. 이 정의에 따라 하나의 영화에서 다수의 정답 장르 중 일부만 정확히 예측된 경우에는 해당 장르 항목에 대해서만 정답으로 반영하고, 나머지 항모은 오답으로 계산된다. 이러한 방식은 멀티 레이블 문제에서 각 레이블에 대한 예측 성능을 균등하게 반영하기 위한 것이다.

본 연구에서 사용한 데이터셋의 장르별 분포를 분석한 결과, 6개 장르 간 샘플 수에 일부 차이가 존재하였으나 특정 장르에 과도하게 편중되는 수준은 아니었다. 각 장르는 전체 데이터에서 일정 비율 이상 포함되어 있어 모델 학습과 성능 비교에 큰 영향을 줄 정도의 심각한 클래스 불균형은 관찰되지 않았다. 따라서 본 연구에서는 별도의 oversampling이나 class weighting을 적용하지 않고 동일한 조건에서 모델 간 성능을 비교하였다.

1. Performance of the Poster-Based CNN Model

포스터 이미지 단독으로 학습된 CNN 모델은 테스트 데이터셋 266개에 대해 약 78%의 binary accuracy를 기록하였다. 이는 포스터의 색감, 인물 배치, 촬영 구도 등 시각적 특징이 장르 판별에 상당한 영향을 미침을 보여준다.

장르별 예측 경향은 다음과 같은 특징을 보였다:

- 드라마·로맨스 계열: 인물 클로즈업, 밝은 톤의 색감, 감성 중심 구성에서 높은 예측률을 보임.

- 호러·스릴러 계열 : 어두운 색감, 대비가 강한 명암, 공포 분위기 요소가 있는 경우 정확도가 높음.
- 코미디·액션 계열 : 포스터 디자인의 다양성으로 인해 예측 난이도가 상대적으로 높음.

예측 결과에 대한 Class Activation Map(CAM) 분석(예: Fig. 2)은 모델이 다음 영역에 주목함을 보여주었다:

- 등장인물의 표정이나 얼굴 영역
- 포스터의 전체적인 색조(따뜻함/차가움)
- 장르적 분위기를 암시하는 배경 요소(도시, 실내, 폐허, 조명 등)

CAM 시각화는 모델이 실제 사람의 판단과 유사한 시각 단서를 활용하고 있음을 시사한다.

2. Performance of the Review-Based NLP Model

영화 리뷰 기반 Skip-gram Word2Vec 모델은 리뷰 임베딩과 장르별 대표 단어 리스트 간 유사도를 비교하여 장르를 산출하도록 구성되었으며, 약 62%의 정확도를 기록하였다. 이는 리뷰가 영화의 분위기·서사·감정 등을 직접적으로 반영한다는 장점이 있으나, 다음과 같은 제약이 존재하기 때문이다:

- 리뷰는 관객의 주관성이 강하게 반영됨 → 감독 의도 장르와 불일치 가능
- 특정 장르(예: 드라마)는 단어의 감성적 중첩이 커 유사도 기반 구분이 어려움
- Sometrend 기반 장르 키워드 수가 제한적이라 유사도 차이가 충분히 벌어지지 않음

그럼에도 불구하고, 액션·공포·스릴러와 같은 강한 키워드가 포함된 리뷰에서는 유사도 기반 예측이 비교적 안정적으로 동작하였다.

3. Performance of the Poster Text (OCR) LSTM Model

포스터 문구를 OCR로 추출하여 LSTM 기반 분류 모델로 학습한 경우, 68%의 정확도를 달성하였다. 이는 짧은 포스터 문구(tagline)나 핵심 메시지가 장르적 분위기를 강하게 전달하기 때문이며, 특히:

- 스릴러/공포 계열에서는 긴장감·위험 요소 관련 단어가 자주 등장해 예측 정확도가 높음
- 로맨스/드라마에서는 감성적 단어·관계 중심 문구가 장르를 잘 반영
- 코미디에서는 문구의 다양성 때문에 예측이 불안정

OCR 기반 접근은 문장 길이가 매우 짧고 특정 장르 단어가 포함될 경우 성능이 급상승하는 경향이 있다.

4. Comparative Evaluation of the Three Models

세 모델의 개별 성능을 비교하면 다음과 같다:

Table 1. Performance comparison of the three models.

Model	*Accuracy*	*Strength*	*Limitation*
Poster CNN	**78%**	Strong ability to capture visual patterns	Misclassification may occur due to visual similarity across genres
Review NLP (Skip-gram)	**62%**	Reflects narrative and emotional context	Subjective reviews; limited genre-specific keywords
OCR-LSTM	**68%**	Short text often contains strong genre cues	Insufficient information when the extracted text is too short

각 모델은 서로 다른 정보를 담고 있으므로 단일 모델의 한계를 보완하기 위해 통합 전략이 필요함을 확인하였다.

5. Integrated Multi-Representation Model

세 예측 모델의 장르별 확률값을 평균하여 통합한 최종 모델은 약 75% 정확도를 달성하였다. 단일 CNN(78%)보다 약간 낮아 보이지만, 개별 장르별 성능에서는 다음과 같은 개선이 관찰되었다:

- CNN 단독 모델이 오분류한 사례를 리뷰·OCR 텍스트가 보정
- 스릴러/공포/액션과 같은 강한 분위기 장르는 텍스트 기반 보정 효과가 큼
- 로맨스/드라마의 경우 포스터 이미지 중심 예측을 텍스트가 안정화
- 다중 장르(복합 장르) 영화의 경우 세 모델의 상호 보완 효과가 특히 크게 나타남

특히 포스터 이미지만 보고 잘못 분류된 영화가 리뷰·문구 정보 결합을 통해 장르가 다시 바로잡히는 현상을 여러 사례에서 확인했다. 이는 통합 모델이 단일 모달 모델보다 보다 일관되고 안정적인 장르 예측이 가능함을 의미한다.

통합 모델의 전체 정확도는 포스터 기반 CNN 단일 모델보다 다소 낮게 나타났으나, 이는 서로 다른 표현 정보를 단순 평균으로 결합하는 과정에서 발생하는 확률 희석 효과에 기인한다. 그러나 통합 모델은 단일 모달리티에서 발생하는 체계적 오분류를 완화하고, 복합 장르 영화나 시각적 단서가 약한 사례에서 예측을 보완하는 경향을 보였다. 이러한 결과는 멀티 표현 기반 접근이 단순한 정확도

향상뿐 아니라 예측의 안정성과 해석 가능성을 높이는 데 의미가 있음을 보여준다.

6. Error Case Analysis

오분류 사례 분석 결과, 다음 패턴이 두드러졌다:

1. 포스터 디자인이 장르 정체성과 무관한 경우
   • 미니멀리즘 포스터, 배우 중심 포스터 등
2. 리뷰의 감정적 표현이 실제 장르와 불일치하는 경우
   • 드라마에 대해 "소름 돋는다", "쫄깃하다" 등 스릴러적 표현 등장
3. OCR 텍스트가 영화 내용과 직접적 관련이 없는 경우
   • 제작사 정보, 슬로건, 홍보 문구 포함
4. 복합 장르 영화
   • 로맨스+코미디, 드라마+스릴러 등 경계가 모호한 영화에서 혼동 증가

이러한 분석은 향후 모델 개선 방향—예: 장르 가중치 조정, 텍스트 필터링 강화, 모달 융합 방식 개선—에 대한 중요한 시사점을 제공한다.

멀티레이블 장르 분류 문제에서는 전체 정확도만으로 각 장르에서의 예측 성능 변화를 충분히 설명하기 어렵기 때문에, 본 연구에서는 6개 장르 클래스(로맨스, 공포, 드라마, 스릴러, 액션, 코미디)에 대해 Precision, Recall, F1-score를 추가로 산출하여 비교하였다. 분석 결과, 포스터 기반 CNN 모델이 시각적 특징이 뚜렷한 장르에서 높은 성능을 보인 반면, 통합 모델은 복합 장르 영화나 시각적 단서가 약한 사례에서 Recall이 개선되는 경향을 확인할 수 있었다. 특히 스릴러 및 공포 장르와 같이 텍스트 기반 단서가 강하게 나타나는 경우, 리뷰 및 포스터 문구 정보가 예측을 보완하는 효과가 관찰되었다. 이러한 결과는 통합 모델의 기여가 전체 평균 정확도의 단순 비교뿐 아니라 장르별 성능 변화와 오분류 보완 효과를 통해 해석될 필요가 있음을 보여준다.

V. Conclusion

본 연구에서는 영화 포스터 이미지, 관람객 리뷰, 포스터 내 OCR 문구라는 세 가지 이질적 표현 정보를 통합하여 한국 영화 장르를 예측하는 다중 표현 학습(Multi-representation Learning) 기반 모델을 제안하였다. 포스터 이미지는 CNN으로, 리뷰 텍스트는 Skip-gram 기반 NLP 모델로, 포스터 문구는 LSTM 기반 모델로 각각 학습하고, 이들의 장르별 확률값을 통합하여 최종 예측을 수행하였다. 실험 결과, 포스터 기반 CNN이 가장 높은 단일 성능(78%)을 보였고, 리뷰 NLP(62%)와 OCR-LSTM(68%)은 특정 장르에서 오분류를 보정하는 역할을 하여 통합 모델은 약 75% 수준의 안정적인 성능을 나타냈다. CAM 분석을 통해 모델이 주목하는 시각적 단서를 확인했으며, 오분류 사례를 통해 포스터 디자인, 리뷰의 주관성, OCR 텍스트의 정보 한계 등 각 표현 방식의 구조적 제약도 파악하였다.

한편, 본 연구에서 사용한 통합 방식은 세 모델의 장르별 확률값을 단순 평균하는 방법으로 구현되었으며, 이 경우 각 모델의 예측 신뢰도나 장르별 특성을 충분히 반영하지 못하는 한계가 존재한다. 특히 특정 장르에서 상대적으로 성능이 낮은 모델의 예측 확률이 결합 과정에서 영향을 미쳐 전체 정확도가 감소하는 현상이 관찰되었다. 이러한 결과는 단순 평균 기반 융합 방식이 항상 최적의 결합 전략이 아님을 시사한다. 향후 연구에서는 장르별 가중치를 학습하는 방식이나 confidence 기반 gating, attention 기반 융합과 같은 고도화된 결합 전략을 적용하여 각 모달리티의 기여도를 적응적으로 조절함으로써 통합 모델의 성능을 보다 향상시킬 수 있을 것으로 기대된다.

본 연구에는 몇 가지 한계가 존재한다. 먼저, 데이터셋이 특정 플랫폼에서 수집된 한국 영화에 한정되어 있어 장르 분포와 표현 방식이 제한적일 가능성이 있다. 또한 관람객 리뷰와 OCR 기반 문구는 텍스트 품질과 전처리 과정에 영향을 받기 때문에, 노이즈가 포함될 경우 예측 성능이 저하될 수 있다. 마지막으로 본 연구에서는 세 모달리티의 결합을 위해 단순 평균 기반의 융합 방식을 사용하였으며, 이는 각 모델의 신뢰도를 충분히 반영하지 못할 수 있다. 향후 연구에서는 데이터 범위의 확장, 장르별 단어 집합의 정교화, 그리고 가중치 학습이나 attention 기반 융합 전략을 적용하여 이러한 한계를 보완할 필요가 있다.

본 연구에서 제안한 다중 표현 기반 장르 예측 방법은 실제 영화 산업 환경에서도 활용 가능성이 있다. 예를 들어 스트리밍 서비스에서는 신규 콘텐츠 등록 시 포스터 이미지와 홍보 문구, 초기 관람객 반응 데이터를 기반으로 장르 태그를 자동으로 생성하거나 기존 장르 태그를 보완하는 데 활용할 수 있다. 또한 복합 장르 콘텐츠에 대해 보다 세분화된 장르 정보를 제공함으로써 추천 시스템의 정확도 향상이나 콘텐츠 탐색의 효율성을 높이는 데 기여할 수 있다. 이러한 점에서 본 연구는 영화 장르 분석의 자동화뿐 아니라 실제 서비스 환경에서 활용 가능한 보조 도구로서의 가능성을 제시한다.

한국 영화 실데이터를 대상으로 포스터-리뷰-문구를

결합한 다중 표현 구조와 그 기여도를 계량적으로 분석했다는 점에서 본 연구의 의의가 있다. 향후에는 장르 세분화, ViT·CLIP·BERT 등 최신 표현 모델 도입, 그리고 단순 평균을 넘어선 attention 기반·gating 기반 융합 전략을 적용하여 멀티모달 영화 장르 예측의 성능과 설명력을 더욱 향상시키고자 한다.

REFERENCES

[1] W.-T. Chu and H.-J. Guo, "Movie genre classification based on poster images with deep neural networks," Proc. Workshop on Multimodal Understanding of Social, Affective and Subjective Attributes, pp. 39–45, 2017. DOI: 10.1145/3132515.3132516

[2] S. Park and H. Shim, "Movie poster classification into genres via convolutional neural network," Proc. Korean Institute of Information Scientists and Engineers, pp. 890–892, 2017.

[3] S. Sung and R. Chokshi, "Classification of movie posters to movie genres," Proc. Workshop on Multimodal Understanding of Social, Affective and Subjective Attributes, 2017.

[4] M. A. Korai, A. H. Bouk, and A. H. Sindhi, "Movie genre classification from RGB movie poster image using deep feed-forward network," Yanbu Journal of Engineering and Science, vol. 18, no. 1, pp. 73–80, 2021.

[5] Y. Cho, K. Kang, and O. Kwon, "Deep learning-based box office prediction using the image characteristics of advertising posters in performing arts," Journal of Society for e-Business Studies, vol. 26, no. 2, 2022.

[6] J. A. Wi, S. Jang, and Y. Kim, "Poster-based multiple movie genre classification using inter-channel features," IEEE Access, vol. 8, pp. 66615–66624, 2020.

[7] V. Battu, V. Batchu, R. R. R. Gangula, M. M. K. R. Dakannagari, and R. Mamidi, "Predicting the genre and rating of a movie based on its synopsis," Proc. Pacific Asia Conference on Language, Information and Computation, 2018.

[8] C. van der Lee, "Text-based video genre classification using multiple feature categories and categorization methods," Master's thesis, Tilburg University, 2017.

[9] M. R. Cuomo, "Movie Genre Classification Using Script Texts," Master's thesis, Florida Atlantic University, 2025.

[10] F. Shaukat, N. Ejaz, Z. Ashraf, M. M. Alnfiai, N. N. Alotaibi, and S. M. M. Alnefaie, "An interpretable multi-transformer ensemble for text-based movie genre classification," PeerJ Computer Science, vol. 11, e2945, 2025.

[11] J. Wang, "Using machine learning to identify movie genres through online movie synopses," Proc. 2nd International Conference on Information Technology and Computer Application (ITCA), pp. 1–6, 2020.

[12] J. Li, G. Qi, C. Zhang, Y. Chen, Y. Tan, C. Xia, and Y. Tian, "Incorporating domain knowledge graph into multimodal movie genre classification with self-supervised attention and contrastive learning," Proc. ACM International Conference on Multimedia, pp. 3337–3345, 2023. DOI: 10.1145/3581783.3612417

[13] D. K. Vishwakarma, M. Jindal, A. Mittal, and A. Sharma, "Multilevel profiling of situation and dialogue-based deep networks for movie genre classification using movie trailers," arXiv preprint arXiv:2109.06488, 2021.

[14] S. Sulun, P. Viana, and M. E. P. Davies, "Movie trailer genre classification using multimodal pretrained features," Expert Systems with Applications, vol. 258, 125209, 2024. DOI: 10.1016/j.eswa.2024.125209

[15] A. Nair, "Multi-modal movie genre classification," Kaggle Notebook, 2023. Available: https://www.kaggle.com/code/avinashnair76/multi-modal-movie-genre-classification

Authors

Jong-Hyun Kim received the B.A. degree in the Department of Digital Contents at Sejong University in 2008. He received M.S. and Ph.D. degrees in the Department of Computer Science and Engineering at Korea University, in 2010 and 2016. Prof. Kim is an Associate Professor in the College of Software and Convergence (Dept. of Design Technology) in Inha University. His current research interests include fluid animation and virtual reality.