

영화 포스터의 장르 예측을 위한 멀티 레이블과 NLP 학습 기반의 네트워크 아키텍처

김수미^o, 김종현^{*}

^o강남대학교 소프트웨어응용학부,

^{*}강남대학교 소프트웨어응용학부

e-mail: jonghyunkim@kangnam.ac.kr

Network Architecture Based on Multi-label and NLP Learning for Genre Prediction of Movie Posters

Sumi Kim^o, Jong-Hyun Kim^{*}

^oSchool of Software Application, Kangnam University,

^{*}School of Software Application, Kangnam University

● 요약 ●

본 논문에서는 멀티 레이블을 이용한 CNN 구조 활용과 NLP 학습을 이용하여 한국 영화의 장르를 예측하는 방법을 제안한다. 포스터는 영화의 전반적인 내용을 한눈에 알아볼 수 있게 하는 매체이기 때문에 다양한 요소들로 구성되어 있다. 합성곱 신경망(Convolutional neural network)을 활용해, 한국 영화 포스터가 가지는 특징들을 추출하여 영화 장르 분류를 진행하였다. 하지만, 영화의 경우 감독이 생각하는 장르와 관객이 영화를 봤을 때, 느끼는 장르가 다를 수 있다. 그렇기 때문에 장르 예측에 있어서 문제가 발생할 수 있다. 이러한 문제를 완화하기 위해 본 논문에서는 합성곱 신경망 활용뿐만 아니라, 자연어 처리(Natural Language Processing)를 같이 활용한 방법을 제안한다.

키워드: 합성곱 신경망(Convolutional Neural Network, CNN), 자연어 처리(Natural language processing), 멀티 레이블(Multi-label), 영화 포스터(Movie poster)

I. Introduction

장르는 관객이 관람할 영화를 선택할 때 중요한 선택 기준 중 하나이기 때문에 중요하다. 하지만 한 영화의 장르는 여러 수 있고, 그 분류의 기준과 범위는 정확하게 구분 짓기 어렵다. 그래서 장르 분류 작업을 자동화하려는 시도들이 있었다. 포스터는 영화의 전반적인 내용을 한눈에 알아볼 수 있게 하는 매체이기 때문에 다양한 요소들로 구성되어 있다. 제목의 서체, 이미지의 색감에 따라 영화의 장르가 다르게 예상될 수 있다. 또한, 감독이 말하고자 하는 영화의 내용을 함축적인 글귀로 나타낼 수 있다. 이와 관련된 연구로 합성곱 신경망(Convolutional Neural Network)을 활용해, 한국 영화 포스터가 가지는 특징들을 추출하여 영화 장르 예측을 진행하였다.[1] 본 논문에서는 합성곱 신경망(Convolutional neural network)을 이용한 멀티 레이블 학습뿐만 아니라 영화의 리뷰와 OCR 알고리즘을 추출한 영화 포스터의 글귀를 이용한 자연어 처리(Natural Language Processing) 학습을 추가하여 한국 영화 장르 예측을 진행하였다.

II. The Proposed Scheme

1. Poster genre prediction network via CNN

본 논문에 사용한 합성곱 신경망 네트워크 구조는 Fig. 1과 같다. 이 네트워크 구조는 기존에 연구된 내용을 참고하였다. INPUT IMAGE를 넣게 되면 Convolution layer와 max pooling의 5개 층과 global average pooling 이후 fully connected layer를 거쳐 예측값을 얻게 된다. Convolution layer는 모두 동일한 크기의 필터를 사용했으며 층이 깊어짐에 따라 채널 개수의 변화를 해주었다. Max pooling 역시 모두 동일한 크기의 필터를 사용하였습니다. 그리고 오버피팅 문제를 줄이고자 global average pooling을 사용하였다[1].

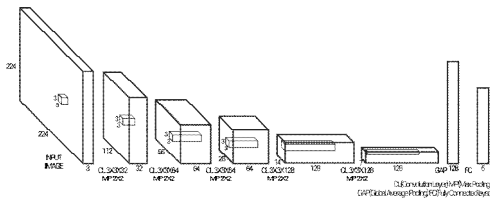


Fig. 1. Network structure.

장르 예측을 위해 네이버 영화 사이트에서 크롤링하여 데이터를 수집하였다. 한국 영화를 대상으로 300명 이상 평가한 영화만 데이터를 가지고 오게 하였으며, 총 1,311개의 영화 포스터를 포함한 데이터를 수집하였다. 영화는 여러 개의 장르를 가질 수 있다. 비슷한 장르끼리 합쳐서 총 6개의 장르인 ‘로맨스, 공포, 드라마, 스릴러, 액션, 코미디’로 분류하였고, MultiLabelBinarizer를 통해 멀티레이블-멀티클래스 분류하였다.

1,311개의 데이터셋에서 1,045쌍을 training data, 266쌍을 test data로 사용하였다. 장르 분류는 멀티 레이블뿐만 아니라 멀티 클래스 분류이다. 이에 맞게 손실함수는 binary cross-entropy를 사용하였다. optimizer는 Adam을 사용하였다.



Fig. 2. Test results for poster genre prediction.

이 네트워크의 정확도 평가는 binary accuracy를 사용하였고, 모델의 성능은 약 78%의 정확도를 보여주었다. Fig. 2 는 6개의 장르 클래스에 대하여 예측한 확률값 중에서 가장 예측값이 높은 3가지 장르를 나타낸 것이다. 또한 출력될 때, 네트워크 구조를 거치면서 포스터의 공간적인 특징 중 어떤 부분에 집중하여 활성화되는지 알아보기 위하여 class activation map을 만들어 시각화하였다. 배우들의 얼굴에 집중하기도 하고 포스터의 어두운 색감에 집중하는 것을 확인할 수 있다.

2. Review genre prediction network via NLP

본 논문에 사용한 영화 리뷰 장르 예측 네트워크 구조는 Fig. 3과 같다. 영화 리뷰 데이터를 INPUT 데이터로 넣게 되면 Word2vec의 학습 방식인 Skip-Gram를 통해 만들어진 모델을 가지고 썬트렌드를 통해 만들어진 장르별 단어들과 유사도를 비교하여 가장 일치하는 장르 2가지가 나오게 된다. skip-gram은 중간에 있는 단어들을 입력으로 주변 단어들을 예측하는 방법으로 모델 학습 시 데이터는 기존에 나와 있는 리뷰 데이터를 사용하였다.

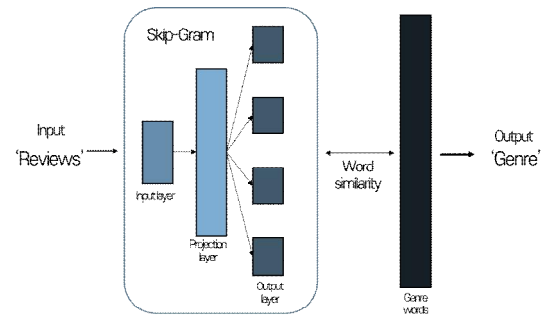


Fig. 3. Network structure.

이전에 크롤링했던 1,311개의 영화에 대한 리뷰를 네이버 영화에서 각 영화당 300개씩 수집했다. 장르별 단어 리스트를 만들기 위해서 Sometrend에서 각 장르를 검색했을 때 나오는 단어들을 20개씩 수집했다. 정규 표현식을 통해서 한글 외 문자를 제거하고, 불용어를 제거하였으며 형태소 분석기 OKT를 사용하여 토큰화 작업을 하였다. Word2Vec 모델을 만드는 데 사용된 입력 데이터는 ratings[2]에 포함된 영화 리뷰 데이터이며, 임베딩 된 벡터의 차원은 100, 윈도우 크기는 5, 단어 최소 빈도수 제한은 5, 학습을 위한 프로세스 수는 4로 설정하였다.



Fig. 4. Test results for review genre prediction.

이 네트워크를 통해서 실제 영화의 장르와 예측 영화의 장르의 일치성을 비교하였고, 모델의 성능은 약 62%의 정확도를 보여주었다. Fig. 4는 6개의 장르 클래스에 대하여 예측한 확률값 중에서 가장 예측값이 높은 2가지 장르를 나타낸 것이다. 장르별 단어들과 유사도를 비교할 때, 장르별 단어의 수가 적어서 각각의 유사도 값들의 차이가 적은 것을 알 수 있다.

3. Poster writing genre prediction network through NLP

본 논문에 사용한 영화 포스터 글귀 장르 예측 네트워크 구조는 Fig. 5와 같다. OCR로 추출된 영화 포스터 글귀를 INPUT 데이터로 넣게 되면 임베딩을 거친 다음 다대일 구조의 LSTM 네트워크를 통해 학습된 모델로 장르가 예측되도록 한다. 임베딩 벡터의 차원은 100, 은닉 상태의 크기는 128이다.

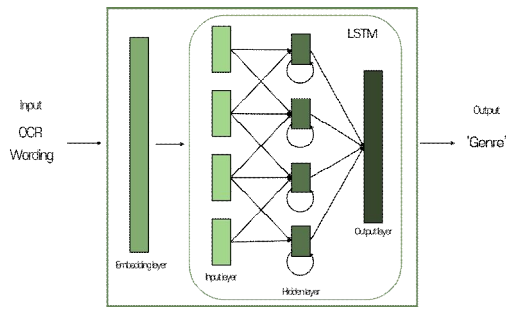


Fig. 5. Network structure.

1,311장의 포스터를 네이버 클로바 OCR 프로그램 이용하여 글귀를 추출하고 JSON 파일로 저장하였다. 포스터에서 인식되는 모든 글귀가 저장되기 때문에 영화 내용과 관련 없는 글귀는 제외하였다. 형태소 분석기 OKT를 사용하여 토큰화 작업을 하였다. 그리고 정수 인코딩을 하였다. 단어 집합의 크기는 4000으로 하였고, 최대 단어의 길이를 16으로 설정하였다. 손실함수는 categorical cross-entropy를 사용하였다. optimizer는 Adam을 사용하였다.



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

ModelCheckpoint를 사용하여 검증 데이터의 정확도가 이전보다 좋아질 경우에만 모델을 저장하였다. 모델의 정확도는 약 66%이다. Fig. 6은 6개의 장르 클래스에 대하여 예측한 확률값 중에서 가장 예측값이 높은 2가지 장르를 나타낸 것이다.

III. Result

본 논문에서 사용한 네트워크들을 종합하여 영화 포스터 장르 예측을 시도하였다. 각각의 네트워크들은 장르별로 예측값들을 가지고 있다. 그 예측값들의 평균을 내어서 가장 예측값이 높은 장르 2가지가 출력되도록 하였다. 장르별로 테스트 해 본 결과는 Table 1과 같다. 정확도는 약 75%이다. 포스터로만 예측했을 때보다 NLP를 더해서 장르를 예측하는 경우, Fig.7과 같이 포스터로만 예측했을 때 장르 예측에서 오류가 났던 것이 비로잡아진다. 또한, 영화 리뷰들은 관람객들의 주관적인 평가가 들어가기 때문에 감독이 생각했던 주 장르와 관객이 생각한 것이 달라질 수 있어서 영화에서 말하고자 하는 주 장르가 바뀌는 상황이 생길 수 있다.

Table 1. Test accuracy by genre.

로맨스 78%	공포 68%	드라마 70%
스릴러 83%	액션 78%	코미디 82%

Table 2. Test results of a model combining CNN and NLP networks.

	CNN을 통한 포스터 장르 분류	CNN과 NLP를 통한 포스터 장르 분류	실제 장르
	액션(0.282) 스릴러(0.247)	액션(0.478) 코미디(0.368)	액션 코미디
	드라마 (0.305) 코미디 (0.237)	코미디 (0.410) 액션 (0.391)	로맨스 코미디

IV. Conclusions

본 논문에서는 멀티 레이블을 이용한 CNN 구조 활용과 NLP 학습을 이용하여 한국 영화의 장르를 예측하는 것에 대한 연구를 진행하였다. 추후에 범죄, 판타지, 다큐멘터리 등 장르의 종류를 더 다양하게 하고자 하며, OCR 글귀에 경우 포스터에 직접적으로 ‘코미디’, ‘로맨스’ 등으로 쓰여 있는 경우 가중치를 두어 장르의 정확도를 높여 결과를 개선하고자 한다. 또한, 현재는 각각의 예측값의 평균을 내는 방법으로 장르를 예측하지만 각 모델의 값들을 이용한 방법들에 대해 생각해보고자 한다.

REFERENCES

[1] Soojin Jang, Jeonga Wi, Youngbin Kim. "Korean Movie Genre Classification based on Poster Images with Multi-label Training using CNN" The HCI Society of Korea (2019): 746-749.
 [2] <https://raw.githubusercontent.com/e9t/nsmc/master/ratings.txt>