퓨처스리그 영화관객수 예측

KHU낙지 팀
팀원 : 서훈, 추봉군, 최우빈

-목차

1. 2004 ~ 2016 영화 데이터 수집
2. 영화 관객수 데이터 수집(8일치, 14일치)
-8일치 : 넛잡 2 , 남한산성(10월 3일 ~ 10월 10일)
-14일치 : 킹스맨 2 (9월 27일 ~ 10월 10일)
3. 네이버 영화 평점 데이터 수집
4. Deep Learning을 위해 최종 CSV 파일 전처리
5. 넛잡 2, 킹스맨 2 , 남한산성 관객수 예측

그림 1

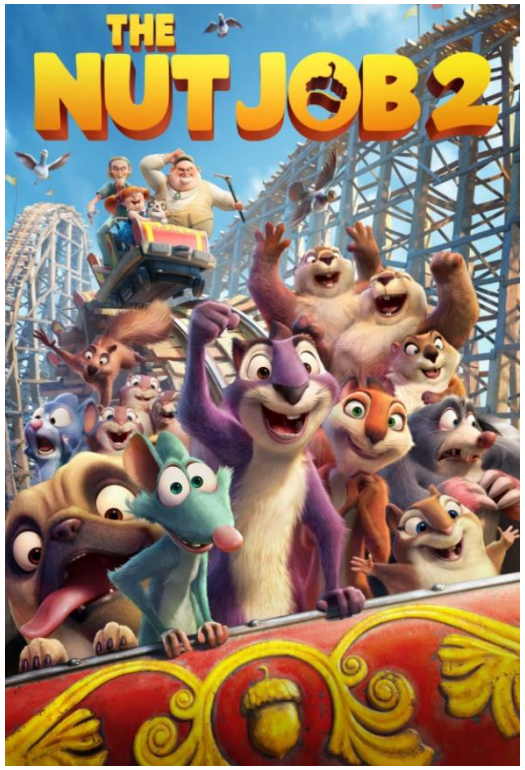


그림 2

그림 3

그림1 :킹스맨2 포스터 (http://www.spotvnews.co.kr/?mod=news&act=articleView&idxno=158080)

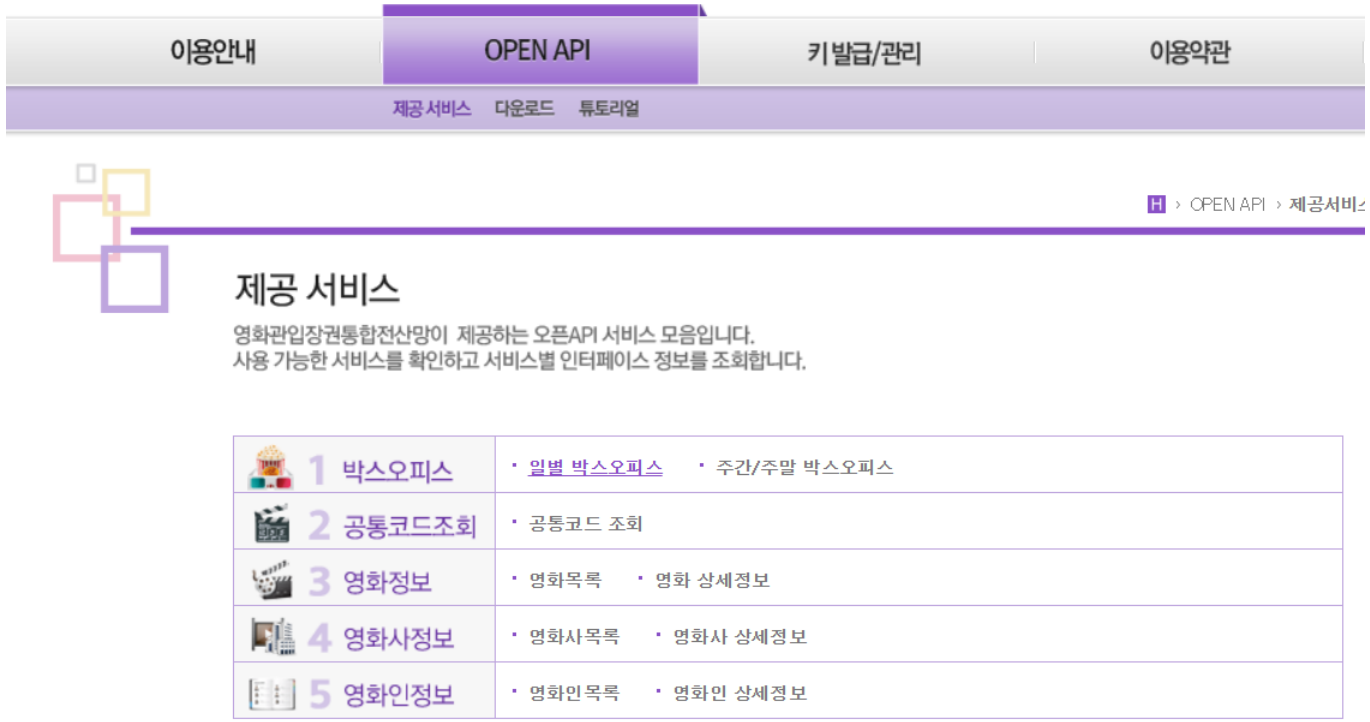
그림2 : 넛잡2 포스터 (http://stock.hankyung.com/news/app/newsview.php?aid=2016051254696)

그림3 : 남한산성 포스터 (http://bbs.ruliweb.com/av/board/300013/read/2324983)

1. 2004 ~ 2016 영화 데이터 수집

- 영화진흥위원회 API에서 KEY값을 사용하여 데이터 수집(학습용)

KOFIC 영화관입장권통합전산망 오픈API

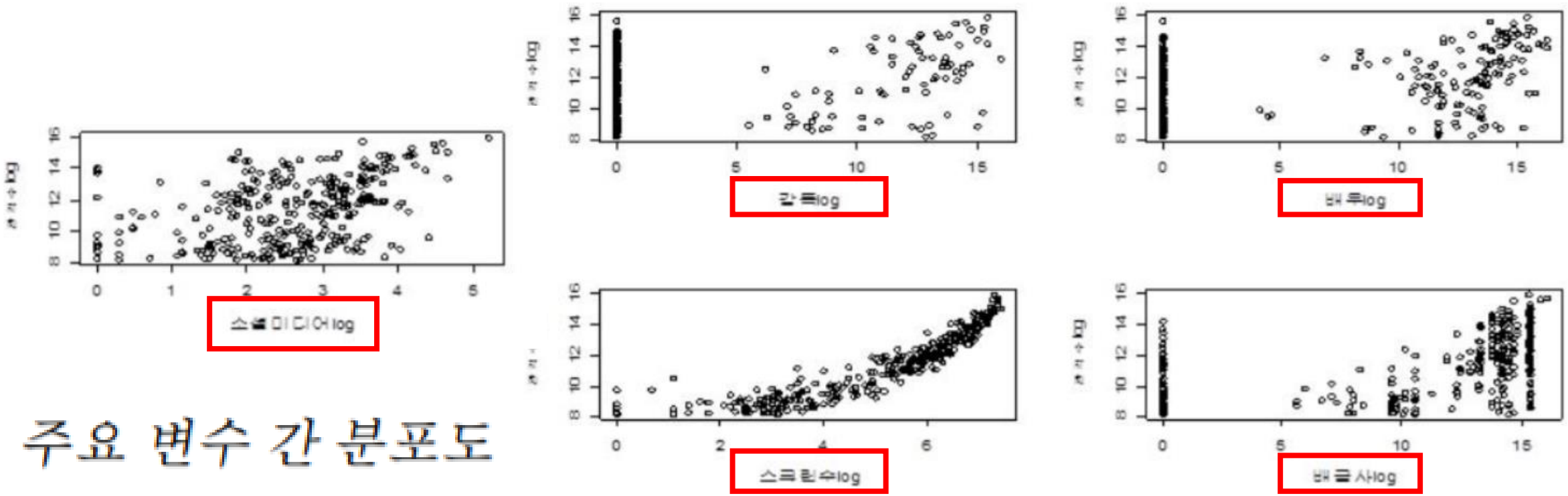


boxoffice.csv 파일로 저장

–영화진흥위원회 오픈API (http://www.kobis.or.kr/kobisopenapi/homepg/main/main.do)

사용할 데이터 변수 선택:

영화 감독, 영화 코드, 영화 이름, 개봉 날짜, 제작 년도, 제작 국가, 상영 시간, 배급사, 상영 등급



주요 변수 간 분포도

-영화 흥행 결정 요인과 흥행 성과 예측 연구 :김연형, 홍정한 /전주대학교 통계학과 /한국 통계학회 논문집 2011, 18권, 6호, 859–869

Boxoffice.csv에 영화 감독, 영화 코드, 영화 이름, 개봉 날짜, 제작 년도, 제작 국가, 상영 시간, 배급사, 상영 등급을 저장.

앞의 도표에서 알 수 있듯이 스크린 수, 배우, 감독,

소셜미디어(네이버 평점), 배급사와 최종 관객수 간에 양의 상관 관계가 있음

2. 영화 관객수 수집(8일치, 14일치)

- total_movie_d8.csv : 영화코드를 이용하여 영화 이름, 감독, 배우, 관람 등급, 8일간 일일 관객수,
스크린 점유율, 상영 점유율, 좌석 점유율, 8일간 누적 관객수
수집

- total_movie_d14.csv :영화코드를 이용하여 영화 이름, 감독, 배우, 관람 등급, 14일간 일일 관객수,
스크린 점유율, 상영 점유율, 좌석 점유율, 14일간 누적 관객수
수집

3. 네이버 영화 평점 데이터 수집

total_movie_d8.csv와 total_movie_d14.csv에서
영화 이름과 영화 코드 url 이용하여, 영화 별 네이버 평점 추출

각각, star_score_d8.csv, star_score_d14.csv 로 저장

4. Deep Learning을 위해 최종 CSV 파일 전처리

Tensorflow 기반 DNN(Deep Neural Network) 이용
➡ 영화관객수 Regression

<최종 CSV>
-result_df_d8.csv
-result_df_d14.csv

그림4 : tensorflow 사진 (https://www.tensorflow.org/)

그림 4

- 최종 CSV파일 변수 분류

-categorical features : 영화이름, 감독, 제작 국가, 장르 이름, 상영 등급, 배우들, 배급사

-numerical features : 개봉 날짜, 제작 연도, 상영 시간, 개봉 전 시사회 관객 수, 네이버 영화 평점

-최종 CSV파일 예측 값 지정

-result_df_d8.csv : 넛 잡 2 , 남한산성

관객수(예측 값) : 시사회 관객수 +1일차 관객 수 + ... + 8일차 관객 수

-result_df_d14.csv : 킹스맨 2

관객수(예측 값) : 시사회 관객수 + 1일차 관객 수 + ... + 14일차 관객 수

5. 넛잡 2, 킹스맨 2 , 남한산성 관객수 예측

<최종 CSV>
-result_df_d8.csv(넛잡 2, 남한산성) - 8일치 관객수
-result_df_d14.csv(킹스맨 2) - 14일치 관객수

<최종관객수>
넛잡 2 :　　172,375 명
남한산성 :　2,697,307 명
킹스맨 2 :　3,783,177 명

[넛잡2, 남한산성] 8일간 누적 관객 수

```
train_df.shape =  (2000, 17)
evaluate_df.shape =  (400, 17)
test_df.shape =  (2, 17)
2017-09-29 23:33:23.416864: W C:₩tf_jenkins₩home₩workspace₩rel-win₩M₩windows₩PY₩35₩tensorflow₩core₩platform₩cpu_feature_guard.cc:45] The TensorFlow library wasn't compiled to use AVX instructions, but these are availa
2017-09-29 23:33:23.417299: W C:₩tf_jenkins₩home₩workspace₩rel-win₩M₩windows₩PY₩35₩tensorflow₩core₩platform₩cpu_feature_guard.cc:45] The TensorFlow library wasn't compiled to use AVX2 instructions, but these are avail
Evaluating ...
global_step: 14500
loss: 4.82159e+13
[172375.4, 2697306.8]

Process finished with exit code 0
```

[킹스맨2] 14일간 누적 관객 수

```
train_df.shape =  (2000, 17)
evaluate_df.shape =  (400, 17)
test_df.shape =  (1, 17)
2017-09-29 23:38:16.072384: W C:₩tf_jenkins₩home₩workspace₩rel-win₩M₩windows₩PY₩35₩tensorflow₩core₩platform₩cpu_feature_guard.cc:45] The TensorFlow library wasn't compiled to use AVX instructions, but these are av
2017-09-29 23:38:16.072846: W C:₩tf_jenkins₩home₩workspace₩rel-win₩M₩windows₩PY₩35₩tensorflow₩core₩platform₩cpu_feature_guard.cc:45] The TensorFlow library wasn't compiled to use AVX2 instructions, but these are a
Evaluating ...
global_step: 15000
loss: 5.04635e+13
[3783177.4]

Process finished with exit code 0
```

-첨부 파일(CSV) 목록

- boxoffice.csv
- total_movie_d8.csv
- total_movie_d14.csv
- star_score_d8.csv
- star_score_d14.csv
- result_df_d8.csv
- result_df_d14.csv

 csv 파일은 utf-8 encoding 형식으로 저장되어 있습니다.

-참고 문헌

-영화 흥행 결정 요인과 흥행 성과 예측 연구 :김연형, 홍정한 /전주대학교 통계학과 /한국통계학회 논문집 2011, 18권, 6호, 859–869

-그림1 :킹스맨2 포스터 (http://www.spotvnews.co.kr/?mod=news&act=articleView&idxno=158080)

-그림2 : 넛잡2 포스터 (http://stock.hankyung.com/news/app/newsview.php?aid=2016051254696)

-그림3 : 남한산성 포스터 (http://bbs.ruliweb.com/av/board/300013/read/2324983)

그림4 : tensorflow 사진 (https://www.tensorflow.org/)

–영화진흥위원회 오픈API (http://www.kobis.or.kr/kobisopenapi/homepg/main/main.do)