

1. 프로젝트 개요

지난 10여 년 동안 매출 7천억원 규모에 묶여 있었던 한국만화 시장이 머지않아 만화 시장 1 조원 시대에 접어들게 될 것으로 예상되며, 이를 통해 웹툰의 성장 가능성이 무궁무진함을 확인해 볼 수 있다. 현재 국내에서 서비스되고 있는 웹툰은 네이버, 카카오 등 포털을 중심으로 매우 빠른 성장을 보이고 있으며, 잠재적 시장성이 매우 높은 산업으로 촉망받고 있다. 웹툰의 특성 중 가장 두드러지는 특징은 다양한 요구를 가진 고객층을 대상으로 서비스를 제공할 수 있는 것이다. 이 때문에 국내뿐만 아니라 해외에서도 웹툰 산업을 긍정적으로 판단하고 관련 산업 시장을 확충하고 있다. 뿐만 아니라 국내 웹툰 플랫폼 중 하나인 네이버웹툰의 월간 전 세계 웹툰 이용자 수가 1억 8000만 명에 달한다고 한다. 웹툰의 인기가 많아지는 만큼 작가들 간의 경쟁도 심해지고, 독자는 원하는 웹툰을 찾기 어려워지고 있다. 웹툰 이용자를 대상으로 평소 웹툰 이용행태를 살펴보면 웹툰을 선택하는 가장 중요한 기준은 장르(52.5%)이며 그림체는(44.0%)로 높은 비중을 차지했다. 이에 따라 우리는 장르와 그림체를 기준으로 사용자가 좋아할 만한 웹툰을 추천해주는 시스템을 만들어 작가들에게는 홍보를, 독자들에게는 편의를 제공하려 한다.

2. 프로젝트 진행 과정

2.1. 데이터 수집

국내 웹툰 플랫폼 1,2위에 해당하는 네이버와 카카오 웹툰의 기본정보 데이터 크롤링을 통해 정보를 수집하였다.

| 네이버 | 카카오 |
|-------------------------|-----------------------|
| Title[제목] | Id[작품 Id] |
| Artist[작가] | Title[제목] |
| Genre[장르] | Writer[스토리 작가] |
| Score[최근 3회차 평점 평균] | Illustrator[그림 작가] |
| Story[줄거리] | Synopsis[줄거리] |
| Image[대표 썸네일 이미지] | Keywords_1[키워드1] |
| Link[웹툰 주소] | Keywords_2[키워드2] |
| Thumbs[회차 별 썸네일 이미지 모음] | Keywords_3[키워드3] |
| Comments[베스트 댓글 모음] | Keywords_4[키워드4] |
| | Adult[연령제한] |
| | Thumbnail[대표 썸네일 이미지] |
| | Timg[회차 별 썸네일 이미지] |
| | Like[좋아요 수] |
| | View[조회수] |
| | Genre[장르] |

[표 2.1. 변수명]

2.2. 데이터 전처리

2.2.1. 카카오 웹툰 평점 계산

카카오 웹툰의 조회수를 이용하여 네이버 평점과 호환되도록 평점을 계산하였다. 평점 계산 방법으로는,
a. 네이버 웹툰과 카카오 웹툰 모두 랭킹이 높을 경우 평점이 높은 경향이 있다는 것을 확인
b. 3 ~ 4일치의 누적 조회수를 수집하여 카카오 웹툰 전체를 랭킹 순으로 나열
c. 네이버 웹툰 평점 별 분포 확인
d. 카카오 랭킹순으로 높은 평점 부여 이때 카카오 웹툰에 부여한 평점과 네이버 웹툰의

평점 분포는 동일
순으로 진행하였다.

2.2.2. 카카오 웹툰과 네이버 웹툰의 장르 통일

웹툰 플랫폼마다 장르명이 다르기 때문에 통일시켜주었다.

- a. 네이버
 - 무협/사극 -> 무협
 - 옴니버스, 에피소드, 스토리 -> 삭제
 - 개그 -> 코믹
 - 무협/사극, 액션 -> 액션/무협
 - 스포츠 -> 학원
 - 감성 -> 드라마
- b. 카카오
 - 액션/무협 -> 무협
 - 코믹/일상 -> 코믹, 일상
 - 판타지 드라마 -> 판타지, 드라마
 - 공포/스릴러 -> 스릴러
 - 로맨스 판타지 -> 로맨스, 판타지

2.2.3. 변수 설정

필요한 변수 외에 사용하지 않는 변수들은 제외해주었다.

2.3. 추천시스템 모델링

2.3.1. 장르 기반 추천 시스템 모델링

웹툰별 장르의 유사성을 구하기 위해 코사인 유사도를 사용하였다. 이때 코사인 유사도란 내적공간의 두 벡터간 각도의 코사인값을 이용하여 측정된 벡터간의 유사한 정도를 의미한다.

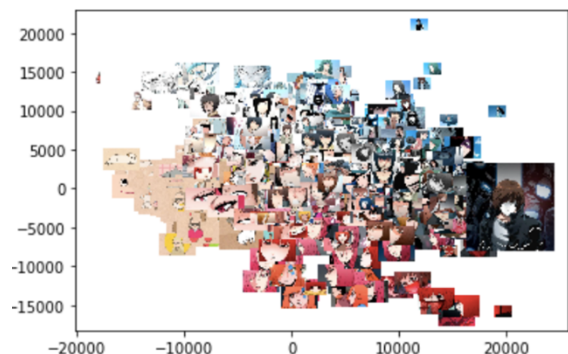
같은 유사도를 가지는 웹툰들에 대해서는 더 높은 평점의 결과를 반환하도록 하고, 총 10개의 웹툰을 추천하도록 모델을 구성하였다.

단, 같은 유사도와 평점을 가진 웹툰 중 sorting된 순서에 의해 상위에 있는 웹툰만 보이는 문제점이 발생하였는데 이를 해결하기 위해 추천되는 웹툰이 랜덤으로 보이는 작업을 진행해주었다.

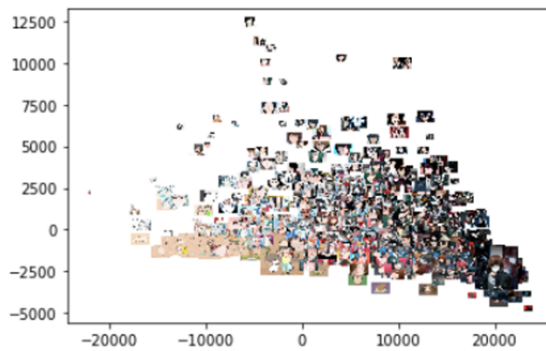
2.3.2. 그림체 기반 추천 시스템 모델링

뉴럴 스타일(Neural-Style) 알고리즘을 이용하여 style loss 값을 추출해 그림체 분류에 적용하였다. 이때, 뉴럴 스타일 알고리즘은 입력 이미지, Content 이미지, Style 이미지 3개의 이미지를 가져와서 Content 이미지의 Content와 Style 이미지의 예술적 Style을 닮도록 입력을 변환하는 알고리즘이다. 우리는 이 알고리즘에서 style loss값을 이용해 그림체 스타일을 분류하였다.

각 웹툰의 회차별 썸네일을 크롤링하여 PCA로 차원 축소한 뒤 썸네일 별 style loss의 평균을 구해 각 웹툰별 style loss로 사용하였다. 또한, style loss의 유클리디안 거리를 구해 그림체 유사도를 판단하였다. 유클리디안 거리는 거리가 가까울수록 그림 스타일이 같음을 의미한다. 2차원 공간에서 비슷한 이미지끼리 모이는 방식을 이용하였기 때문에 이와 같은 방법을 사용하였다. 이때 모델이 색깔에 따라 분류하는 편향을 없애기 위해 GrayScale 적용을 추가해주었다. 여러 개의 input이 들어올 경우 거리 데이터들을 종합하여 그 중 가장 거리가 짧은 웹툰 10개를 반환하도록 만들어주었다.



[그림 2.1. GrayScale 적용 전]



[그림 2.2. GrayScale 적용 후]

웹툰을 중심으로. *한국빅데이터학회지*, 3(1), 21-32.

[3] 최건호, 파이토치 첫걸음, 한빛미디어 2019

3. 프로젝트 결과

streamlit을 이용하여 장르 기반 추천 시스템과 그림체 기반 추천 시스템을 웹 상에 배포하였다. 독자가 평소 선호하는 웹툰을 선택하면 유사한 장르와 그림체의 웹툰을 각각 추천해 주는 시스템이다. 이를 통해 독자는 자기만의 만화를 추천 받을 수 있다.

4. 참고문헌 및 출처

출처

[1] 김영학 (2018) [마크로밀 트렌드 조사:웹툰] 만화를 양지로 끌어올린 웹툰, 생태계에 대한 관심 필요

<https://www.startuptoday.kr/news/articleView.html?idxno=10395>

[2] 김해욱 (2022) '네이버웹툰, 분사 5년만에 월 이용자수·매출 4배 넘게 성장' UPI뉴스기사

<https://www.upinews.kr/newsView/upi202205020041>

참고문헌

[1] 양지훈, 이지영, & 이상우. (2016). 웹툰 (Webtoon) 의 흥행 결정요인

연구. *한국콘텐츠학회논문지*, 16(5), 194-204.

[2] 도상범, & 강주영. (2018). 한, 미, 일 웹툰 분석을 통한 구독자 선호 요인 탐색: 네이버