

고객 맞춤형 웹툰 추천 시스템 제안

CUAI 4 기 문화 콘텐츠 C 팀 권송아(소프트웨어학부), 신수현(수학과), 이상화(소프트웨어학부), 이하은(소프트웨어학부)

본 논문에서는 일상 속의 많은 문화 콘텐츠 중 웹 툰에 초점을 맞추고 사용자가 좋아할 만한 웹툰을 추천한 결과를 간단히 보여주는 서비스를 개발하였다. 사용한 추천 방식은 크게 콘텐츠 기반 추천 시스템과 협업 필터링을 이용한 추천 시스템으로 나눌수 있으며, 콘텐츠 기반 추천에는 스토리, 그림체 기반 알고리즘, 협업 필터링에는 아이템, 잠재 요인 기반 알고리즘이 있다.

1. 서 론

20세기는 더이상 인간만이 컴퓨터의 언어를 공부하는 시대가 아니다. 컴퓨터가 인간의 지식과 경험을 학습하고 이해하여 상호 작용하는 인공지능이라는 개념이 탄생하였다. 인공지능은 컴퓨터가 인간의 반복적인 일을 대신해주는 역할을 함과 동시에 다양한 기술에 접목되어 우리 삶을 한층 더 편리하게 바꾸어 주고 있다.

추천 시스템은 개인의 취향과 선호도에 따른 상품 정보를 제공하는 것으로, 주로 쇼핑몰과 같은 온라 인상에서 사용된다. 이와 마찬가지로 문화 콘텐츠를 소비하는 과정에서도 추천 시스템이 접목된 사례가 많이 나타나고 있다. 도서나 영화 같은 경우에는 추 천 시스템이 많이 상용화되어 있고, 이와 관련된 개 발 프로젝트를 빈번하게 볼 수 있다. 그러나 웹툰 시장을 살펴보면, 웹툰 플랫폼들의 접근성이 매우 높음에도 불구하고 플랫폼들이 갖추고 있는 추천 시 스템은 보유한 기능이 매우 간단하며 각 플랫폼 안 에서만 국한되어 있다는 점에서 고객의 니즈를 만족 시키기엔 부족하다.

이에 따라 본 논문에서는 여러 문화 콘텐츠들 중에서 웹툰을 선택하여 고객 맞춤형 추천 시스템을 제안하고자 하였다. 본 논문의 추천 시스템에서 사용한 추천 방식은 크게 콘텐츠 기반(Content- based) 추천 시스템과 협업 필터링(Collaborative Filtering)을 이용한 추천 시스템으로 나눌 수 있다.

콘텐츠 기반 추천 시스템은 해당 콘텐츠가 가진 속성에 기반하여 사용자가 좋아할 상품을 추천해 주는 방법을 사용한다. 웹툰이 가진 속성으로는 웹툰을 그린 작가, 장르, 스토리, 그림체 등이 있는데, 작가와 장르는 사용자가 쉽게 파악할 수도 있고 추천알고리즘이 매우 간단할 것으로 예상되어 제외하였다. 이에 스토리와 그림체를 기반으로 한 추천 알고리즘을 제작하여 그 결과를 나타냈다.

협업 필터링은 각 사용자가 특정 상품을 선호하는 정도를 알고 있을 때, 그 사용자가 아직 접하지 않은 다른 상품에 대한 선호를 예측하여 추천하는 방법이다. 이를 이용한 추천 시스템에는 아이템 기반 (Item-based) 협업 필터링 알고리즘과 잠재 요인 (Latent factor) 기반 협업 필터링 알고리즘을 사용하였다. 아이템 기반 협업 필터링은 각 아이템들(본 논문에서는 웹툰)에 대한 유사도를 계산하는 방식이고, 잠재 요인 기반 협업 필터링은 명확하지 않은 잠재요인에 따라 행렬 분해를 진행하여 유사도가 계산되는 방식이다. 만들어진 알고리즘을 테스트하기 위해서 실제 고객들의 웹툰 선호도 데이터가 사용되었다.

이러한 추천 알고리즘들의 정확도를 높이기 위해여러 유사도 계산 방식과 모델을 이용한 실험을 통하여 본론과 같은 결과를 얻었다. 다양한 소스와 알고리즘을 사용해 결과를 도출하고 사용자가 중요하게 여기는 요인에 따라 결과를 세분화하여 확인할수 있기 때문에 실제 고객에게 사용되었을 때 높은 정확도를 보일 것이라 기대한다. 추후 다양한 웹툰플랫폼에서 고객 맞춤형 추천 시스템을 도입하여 고객이 편리하게 선택할 수 있도록 도울 것이다.

2. 본 론

가. 콘텐츠 기반 추천 시스템

1) 스토리 기반 추천 시스템

보통 웹툰에 있어서 자신의 취향을 결정짓는 요소가 있다면 그것은 해당 웹툰의 스토리와 그림체

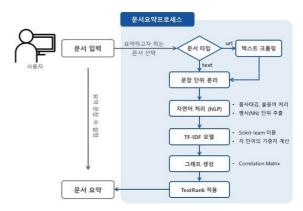
이다. 그 중에서도 스토리의 경우 웹툰 전체의 흐름과 맥락을 결정짓는 요소로써 독자들 각각의 취향과 가장 직결되는 부분이기도 하다. 따라서 각웹툰의 스토리를 키워드별로 세분화하게 되면, 특정 웹툰들의 공통점을 파악할 수 있고, 해당 키워드에 기반하여 독자의 취향을 찾을 수 있게 되는 것이다. 또한 독자들은 자신의 취향인 웹툰과 비슷한 키워드를 추천받을 수 있게 되니 더 취향에 적합한 웹툰을 추천받을 확률이 올라가게 된다.

스토리를 기반으로 한 추천 시스템을 구현하기 위해서 프로그램을 크게 전처리 부분과 코사인 유 사도를 활용하여 한 웹툰과 유사한 웹툰을 추출하 는 부분으로 나눌 필요가 있다. 우리는 전처리 부 분에서 각 웹툰 별 키워드를 추출할 것이고, 이 키워드를 활용하여 코사인 유사도를 계산한 후 유 사 웹툰을 찾을 것이다. 이 프로그램에는 두 가지 알고리즘이 사용되었다. TextRank와 TF-IDF 알 고리즘이다. TextRank는 구글의 PageRank를 활 용한 알고리즘으로 텍스트에 관한 graph-based ranking model이다. PageRank는 웹 문서에 상대 적 중요도에 따라 가중치를 부여하는 알고리즘인 데, 이때 사이트간 인용과 참조로 연결된 임의의 묶음을 적용할 수 있다. 즉 PageRank가 높은 웹 페이지는 다른 웹 사이트로부터 링크를 많이 받게 되고. 이는 다른 사이트가 참조를 많이한 것으로 해석할 수 있다. 이러한 PageRank의 특성에 착안 하여 TextRank는 긴 글을 문장 단위로 분리하고 각 문장 별 명사를 추출한 후 각 단어의 Ranking 을 계산하는 알고리즘이다. TextRank 식은 다음과 같다. [1]

$$TR(V_i) = (1-d) + d * \sum_{V_j \in In(V_j)} \frac{w_{ji}}{\sum_{V_k \in out(V_j)} W_{jk}} TR(V_j)$$

이 식에서 $TR(V_i)$ 는 단어 (V_i) 에 대한 TextRank 값이며, w_{ij} 는 단어 i와 j 사이의 가중치를 의미한다. 또한 d는 damping factor로 PageRank알고리즘에서 웹서핑을 하는 사람이 해당 페이지를 만족하지 못하고 다른 페이지로 이동하는 확률을 말하는데 이 값을 TextRank 알고리즘에도 그대로 적용한다. 이러한 수식을 통해 각각의 단어의 가중치를 계산한 후 높은 순으로 정렬하면 주어진 문장에서 중요 단어를 확인할 수 있게 되는 것이다. 따라서 다음 그림과 같은 프로세스를 거쳐 문서를

요약하게 되는 것이다.



[그림 1] TextRank 알고리즘 프로세스

다음으로 TF-IDF는 특정 문서 내에서 단어 빈도(Text Frequency)와 전체 문서에서의 단어 빈도인 역문서 빈도(Document Frequency)를 통해 해당 문서에서 자주 사용되는 단어를 찾아 문서 내단어의 가중치를 계산하여 문서의 핵심어를 추출하는 알고리즘이다. [2] 이 두 알고리즘을 사용하여 각 웹툰 별 주어진 스토리의 키워드를 뽑아내는 것이 스토리 기반 추천 알고리즘의 핵심이다.

키워드를 추출하기 위해 각 웹툰 플랫폼에서 전 체 웹툰 7300여개의 제목, 작가, 스토리, 썸네일 등을 크롤링 한 후, 각 웹툰의 스토리를 TextRank 알고리즘을 이용하여 키워드를 추출한 다. 이 알고리즘을 적용할 때 중요한 점은 stopword이다. stopword는 불용어로써 문장 내에 서 내용을 나타내는데 의미를 갖지 않는 단어들의 집합을 의미한다. 특히 웹툰의 스토리 소개의 경 우 여러 미사여구, 대명사, 연결어구 등이 포함된 경우가 많기 때문에 이러한 단어를 제거해주어야 더욱 유의미한 결과를 도출해낼 수 있다. 스탑워 드를 제거한 후 추출된 단어를 대상으로 TF-IDF 계산 및 가중치 그래프를 만들고, 그 그래프를 이 용하여 TextRank 알고리즘을 적용하는 방식이다. 그리고 알고리즘을 통해 Ranking 값이 높은 순으 로 정렬하여 3개의 단어를 추출해준다. 이것이 각 웹툰 별 키워드가 되는 것이다.

title : 다시피는 꽃 keywords : ['이야기', '위안부', '할머니']
title : 트리니티 원더 keywords : ['고수', '마법사', '무협']
title : 피그말리온 컴플렉스 keywords : ['고나', '변신', '소녀']
title : 오늘의 초능력 keywords : ['고녀', '봉하', '예측']
title : 골든왕츠 keywords : ['기적', '사랑', '아미스댄싱']
[그림 2] 웹툰 별 키워드 추출

이렇게 각 웹툰의 키워드를 추출한 후 이를 기



준으로 전체 웹툰들의 코사인 유사도를 비교하고, 유사도가 높은 순으로 상위 10개의 인덱스를 추 출해준다.

술론이 말라하옵니다.
('시작업 선부', 0.6751755550000089), ('회하의 무용 위 [전제]', 0.48441496755761628), ('100일의 결국지박', 0.406702588
[그림 3] 유사웹툰 상위 목록

2) 그림체 기반 추천 시스템

추천 알고리즘을 만들기에 앞서 웹툰의 썸네일이 해당 웹툰의 그림 스타일을 함축적으로 담고 있다고 가정하였다. 따라서 네이버, 다음, 카카오페이지, 레진코믹스 플랫폼에서 연재 중이거나 완결된 웹툰들의 썸네일 이미지 링크를 크롤링하여정리한 데이터를 사용하였다. 이 데이터에서 각행의 인덱스, 웹툰의 제목과 웹툰이 속한 플랫폼, 썸네일 이미지의 링크를 불러와 각 웹툰의 썸네일이미지를 로컬 폴더에 다운받아 사용하였다.

추천에 사용될 모델은 Pytorch에서 제공하는 세개의 모델들인 MobileNetV3- Small, ResNet18, ResNet50 중에서 한 개를 선택하기로 하였다. 이모델들은 모두 ImageNet dataset을 사용해 미리학습되어 있는 모델이다. 본 연구에서 사용하는 썸네일 이미지 dataset을 사용하기 전에 전체 이미지 dataset을 대상으로 직접 RGB 평균과 표준편차값을 계산하여 이를 이용해 dataset을 정규화하였다. 또한, 이미지 전처리 과정에서는 플랫폼별로 썸네일 이미지의 크기가 다르다는 이유로 같은 플랫폼의 웹툰이 우선적으로 추천될 가능성이 있다고 생각하여 이미지의 크기를 일정하게 설정하였다. 본 논문에서는 크기를 정사각형 모양의 125*125로 결정하였으며, 고유한 비율을 유지하면서 resize하였고 여백은 흰색 바탕으로 패딩하였다

이후 학습된 모델에서 각 웹툰의 feature를 추출하였다. 이 값을 바탕으로 특정 웹툰과 코사인 유사도가 높은 다른 웹툰을 출력하는 것을 결과로 나타냈다. 이 과정에서 얻은 정확도와 소요 시간 등을 종합적으로 고려하여 세 개의 모델 중 최종 적으로 사용할 모델을 결정하였는데, 이때 플랫폼 별로 3~4개의 웹툰을 무작위로 선정하여 결과를 비교하는 데 사용하였다.

결과는 코사인 유사도를 계산한 값을 바탕으로 도출되는데 이미지의 수많은 특성들 중에서 각 모 델별로 초점을 두는 특성이 다르기 때문에 썸네일 이미지 간의 유사함에 대한 정확도를 판단하는 평 가 지표를 만들기 어려웠다. 따라서 본 연구에서 웹툰의 그림체가 유사하다고 판단하기에 필요한 요소들을 바탕으로 주관적인 평가를 진행하였다.

[그림 4], [그림 5], [그림 6]은 네이버 플랫폼에서 인기 있는 완결 웹툰인 "대학일기"의 썸네일이미지에 대해 각 모델이 계산한 유사도와 소요시간을 나타낸 출력이다. 가장 크게 나타난 이미지는 "대학일기"의 썸네일 이미지이다. 결과로 나온 10개의 이미지 중에서 윗줄은 유사도가 가장높게 계산된 5개의 이미지이고, 아랫줄은 유사도가가장 낮게 계산된 5개의 이미지이다.

MobileN etV 3- Small의 결과에서 윗줄을 보면, 앞 3개의 이미지들에 나오는 캐릭터들의 눈이 "대학일기"의 캐릭터와 다르게 동그란 모양을 갖는다. 반면에 ResN et18과 ResN et50의 결과에서 윗줄을 보면, 캐릭터들의 생김새가 "대학일기"의 캐릭터처럼 단순하고 이목구비가 작은 모습을 확인할 수 있다. 더불어 같은 작가의 작품들도 높은 유사도를 가졌다. 다른 웹툰에 대한 결과들에서도 ResN et18과 ResN et50이 MobileN etV 3- Small보다 높은 정확도를 보였다.

또한, 출력된 그림 위에 출력된 소요 시간들을 보면, 같은 웹툰에 대한 결과에서 MobileN etV 3-Small은 171066.53125ms가 소요되었고 ResN et18은 179147.5ms, ResN et50은 406031.15625ms가 소요되었다. 이처럼 ResN et50 을 사용했을 때는 결과가 도출되는 데 다른 모델 에 비해 매우 오래 걸리는 것을 볼 수 있다. 따라 서 본 연구에서는 그림체 기반 추천 시스템을 만 드는 데에 ResN et18을 사용하기로 결정하였다.





[그림 4] 네이버 웹툰 "대학일기"의 썸네일 이미지에 대한 MobileN etV 3- Small을 사용한 유사도와 소요 시간 측정 결과

179147.5



[그림 5] 네이버 웹툰 "대학일기"의 썸네일 이미지에 대한 ResN et18을 사용한 유사도와 소요 시간 측정 결과



[그림 6] 네이버 웹툰 "대학일기"의 썸네일 이미지에 대한 ResN et50을 사용한 유사도와 소요 시간 측정 결과

이러한 실험 결과를 바탕으로 ResNet18을 사용하기로 결정하고 실제 사용자에게 보여줄 결과를 출력하였다. 이미지 간의 유사도가 0.8 이상이어야 신뢰할 수 있는 결과라고 판단하여 유사도가 높은 순서대로 0.8 이상인 것만 20개를 결정하고 사용자가 선호하는 장르의 웹툰부터 유사도가 높은 순서대로 보여주었다. 이때 각 웹툰의 정보 중제목과 썸네일 이미지를 출력하게 하였다.













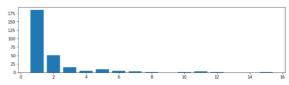
[그림 7] 그림체 기반 추천 결과

나. 협업 필터링을 이용한 추천 시스템

1) 아이템 기반 협업 필터링

[3] 아이템 기반 협업 필터링은 "이 아이템을 좋아하는 사람들이 좋아하는 다른 아이템"으로 설명할 수 있다. 각 아이템에 대한 사용자들의 행동양식(평점, 좋아요/싫어요, 구매이력 등)을 기반으로 아이템 간의 유사도를 계산하는 방식이다.

실습에 앞서, 104명의 사용자를 대상으로 선호 장르, 취향에 맞는 웹툰, 취향에 맞지 않는 웹툰을 조사했다. 설문조사 결과, 취향에 맞는 웹툰 277 개, 취향에 맞지 않는 웹툰 51개, 총 303개의 웹툰을 응답으로 받았다. 사용자 별 평균 응답 웹툰수는 6.44개였으며, 웹툰 별 사용자의 응답 수가 1개인 경우가 184개로 전체 응답 웹툰 수의 절반이상을 차지했다.



[그림 8] 웹툰 별 사용자의 응답 수

응답 받은 데이터를 가공해 '취향에 맞는 웹툰'을 1, '취향에 맞지 않은 웹툰'을 -1, '보지 않은 웹툰'을 0으로 표현한 사용자-아이템 선호도 행렬을 구성했고, 이 값으로 각 웹툰들 사이의 코사인 유사도(Cosine similarity)를 계산했다. 코사인 유사도 이외에 유클리디안 유사도(Euclidean similarity), 자카드 유사도(Jaccard similarity)를 고려했지만 코사인 유사도가 -1과 1로 표현된 선호도 데이터 세트에 가장 적합하다고 판단했다. 최

종 예측값은 사용자의 선호도 데이터와 코사인 유 사도를 곱한 후, 각 아이템에 대한 코사인 유사도

벡터의 합을 나누는 정규화 과정을 거쳐 구했다. 데이터 수가 부족했기 때문에 예측값이 큰 값이 많지 않았다. 따라서 예측값이 0.2 이상인 웹툰 전 기리기소요로 취임 2개기기 출전했다. 다르 사

중 내림차순으로 최대 3개까지 추천했다. 다른 사람들이 보지 않은 웹툰을 많이 작성한 사용자([그림 9]의 user53)는 추천 결과가 보이지 않았다.

설문조사에 응한 사용자들에게 피드백을 받을 수 없는 상황이었기 때문에 추천결과에 대한 평가가 제대로 이루어지지 못했다. 그렇지만 추천된 웹툰과 사용자가 좋아하는 웹툰을 대략적으로 비교해 보았을 때, 좋아하는 웹툰에 로맨스물이 포함되어 있는 사용자들([그림 9]의 user0, user37)에게는 로맨스물이 추천된 것을 볼 수 있었다. 또한로맨스물을 좋아하지 않는 사용자([그림 9]의 user1, user7)에게는 로맨스물이 아닌 웹툰이 추천된 것을 볼 수 있었다.

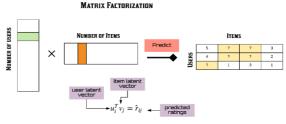
```
user10(가) 좋아하는 웹툰
하르모니아 1.0
                                                                           하르모니아
하이브
당신의 과녁
유미의 세모들
추기
쌍갑포차
파이게임
머니게임
노네임드(NoNameD)
Name: 1, dtype: f
       er0이(가) 좋아하는 웹툰
항설 1.0
의 일지 1.0
user(00)(가)
가담항설
모죠의 일지
내일 1
남과 여
호려와 로맨스
하나의 하루
싸우자귀신이
                   1.0
                                                                            pred_score
냄새를 보는 소녀 0.345124
신도림 0.334273
친애하는 X 0.327199
   me: 0, dtype: float64
                                                                            신도림
친애하는 X
pred_score
악마와 계약연애 0.593068
                                                                         user3701(가) 좋아하는 웹뿐
                                                                     모표의 일지
                                                                     모호의 일시
유미의 세포들
죽음에 관하여
이번 생도 잘 부탁해
                                                                                                      1.0
                                                                                                1.0
                                                                    연해혁명 1.0
화장 지워주는 남자
살아남은 로맨스
피라미드 게임 1
바른연애 길잡이
세기말 풋사과 보습학원
아홉수 우리들 1
                                                                                                      1.0
                                                                                                    1.0
                                                                                                    1.0
                                                                                                        1.0
                                                                                                   1.0
                                                                     성스러운 아이를
합격시켜주세용
                                                                     다하당자
나한당자
간 떨어지는 동거
연우의 순정
Name: 37, dtype:
                                                                                               1.0
                                                                                                    1.0
    user7이(가) 좋아하는 웹툰
대학일기
신과함께
목욕의 신
Name: 7, dtype: float64
                                                                    새벽 두 시의 신데렐라 0.50
언터처블 0.467763
섯업앤댄스 0.255287
                                                                                                      0.506066
                                                                    세탁 후 시-
언터처음
섯업앤댄스
pred_score
놓지마 정신줄 0.203682
user53이(가) 좋아하는 웹툰
전지적 독자 시점
                                                  1.0
전시역 목사 시점
신의 탑
하루만 네가 되고 싶어
비밀 사이
왕년엔 용사님
하렘의 남자들
                                            1.0
                                                     1.0
                                             1.0
                                                 1.0
                                                 1.0
하임의 임사들 1.0
화산귀환 1.0
정령왕 엘퀴네스 1.0
내 동생 건들면 너희는 다 죽은 목숨이다
황제와 여기사 1.0
            : 여주인공의 오빠를 지키는 방법
                                                               1.0
악녀를 죽여 줘
                                            1.0
 시____
백작가의 망나니가 되었다
                                                        1.0
                                           1.0
미성년
Name: 53, dtype: float64
Columns: [pred_score]
Index: []
```

[그림 9] 아이템 기반 협업 필터링을 이용한 추천 결과

2) 잠재 요인 기반 협업 필터링

잠재 요인 모델이란 관찰된 값을 설명하는 잠재된 요인들을 밝혀 이를 이용하여 문제를 해결하기위한 모델이다. 추천 시스템에서 잠재 요인 협업필터링은 사용자-아이템 평점 행렬 속 숨어있는 잠재 요인을 추출하여 추천 예측을 할 수 있게 하는 기법이다. '잠재 요인'이 어떤 것인지 특정 지을 수는 없으나, 이러한 '잠재 요인'을 기반으로다차원 행렬 데이터를 두개의 저차원 행렬로 분해할 수 있으며, 분해된 두 행렬 내적을 통해 새로운 예측 행렬 데이터를 만들어 사용자가 아직 부여하지 않은 예측 값을 생성하는 것이 잠재 요인 협업 필터링 알고리즘의 핵심이다. [3], [4]

행렬 분해 기법으로는 대표적으로 SVD가 있다. [그림 10]과 같이 사용자- K차원 잠재요인 행렬(P)과 K차원 잠재요인- 아이템(Q.T)의 내적은 $M \times N$ 차원으로 구성된 사용자- 아이템 행렬(R)이 된다(본 논문에서 아이템은 웹툰이고, 사용자는 웹툰을 시청하는 이용자들이다). 즉, R 행렬의 1행 사용자와 2열 웹툰에 해당하는 $r_{(1,2)}$ 은 아직 사용자가 평점을 매기지 않은 미정 데이터이지만, $r_{(1,2)} = p_1 \times q_2^t$ 으로 유추 가능하다. [3]



[그림 10] SVD 행렬 분해 (ICHI.PRO, 추천 시스템 시리즈 4 부 : 협업 필터링을 위한 MF의 7 가지 변형)

아이템 기반 협업 필터링에서 사용한 가공이 완료된 데이터를 잠재 요인 협업 필터링에 적용하였다. 실제 R 행렬과 예측 행렬(P• Q.T)의 RMSE 계산 시 오류가 가장 적게 나온 파라미터를 바탕으로 모델을 수행하였다. 각 파라미터 값은 잠재요인의 차원수(K)는 8, SGD의 반복 횟수(steps)는 200, 학습률(learning_rate)은 0.006, 규제 계수(r_lambda)는 0.01로 설정하였다.

현저히 적은 데이터 수로 인해 사용자수 중 웹 툰이 추천되지 않은 경우가 존재한다. 아래의 [표 1]과 같이 104명의 사용자들 중에서 추천이 되지 않은 user_id가 있음을 알 수 있다.

[표 1] 잠재 요인 협업 필터링을 이용한 예측 결과

	pred_score	user_id
바른연애 길잡이	1.165272	2
피라미드 게임	1.071151	2
재혼 황후	1.060038	2
소녀의 세계	1.044844	3
여주실격!	0.886652	3
전지적 독자 시점	0.836209	3
세기말 풋사과 보습학원	0.997757	8
마음의소리	0.875702	8

아이템 기반 협업 필터링과 같이 즉각적인 사용자들의 피드백이 없어 추천 시스템 결과에 대한 완벽한 평가와 대처가 이루어지지 않았지만, 사용자들의 선호하는 웹툰 장르와 비슷하게 웹툰이 추천되었음을 알 수 있다. [표 1]에서와 같이 'user_id'가 3이 선호하는 장르인 판타지를 기반으로 한 "전지적 독자 시점" 웹툰을 추천하는 것을 볼 수 있다.

3. 결 론

우리는 보다 효율적인 웹툰 추천 서비스를 개발하기 위해 총 4가지의 시스템을 설계하였다. 콘텐츠기반 추천으로는 TextRank를 활용해 웹툰의 스토리에서 뽑아낸 키워드를 기반으로 추천을 진행하는 스토리기반 추천, Resnet18로 썸네일의 특징을 추출해그림체가 유사한 웹툰을 추천하는 그림체 기반 추천을 제시하였다. 협업 필터링 기반 추천으로는 특정아이템을 좋아하는 사람들의 취향을 반영하는 아이템 기반 협업 필터링, SVD를 활용해 개개인의 잠재적 취향을 고려하는 잠재요인 기반 협업 필터링을 제시하였다. 각각의 추천 시스템은 어떤 요소가 사용자의 취향을 반영하는지를 고려하여 이를 중심으로 모델링 하는 것에 중점을 둔다.

이렇게 개발된 네 개의 추천 시스템은 서로 다른 요소에 집중하고 있어 사용자에게 기존의 웹툰 추천 시스템보다 폭 넓은 선택지가 주어질 것이라 기대된 다. 또한, 콘텐츠 기반 추천 시스템들은 서비스 초기 에 사용자에 대한 데이터가 부족할 때 활용할 수 있 다는 강점이 있고, 협업 필터링은 개인의 취향에 맞 춘 결과를 얻을 수 있다는 강점이 있다. 추후에 이 시스템들을 통합적으로 이용할 수 있는 웹을 구현하면, 사용자의 행동 양식에 관한 데이터와 추천 결과에 대한 피드백을 대량으로 모을 수 있어 더 효과적인 추천을 할 수 있을 것이라 예상한다.

참고 문헌

[1] P. Wongchaisuwat, "Automatic Keyword Extraction Using TextRank," 2019 IEEE 6th International Conference on Industrial Engineering and Applications (ICIEA), 2019, pp. 377-381
[2] 이지훈, 엄태현, 이혁준. "TF- IDF 와 CNN을 사용한 법률문서 분류 시스템에 관한 연구." 한국통신학회 학술대회논문집. (2019): 982-983.
[3] 권철민, 파이썬 머신러닝 완벽 가이드, 2020, pp. 567-579, 591-606
[4] 김형도. "잠재 요인 모델의 원리를 이용한 협업

[4] 김형도. "잠재 요인 모델의 원리를 이용한 협업 태그 기반 추천 방법." 한국전자거래학회지 14.4 (2009): 47-57.

[5] 이미지 탐지기 쉽게 구현하기, 당근마켓 팀블로그,

https://medium.com/daangn/%EC%9D %B4%EB% AF%B8%EC%A7%80- %ED %83%90%EC%A7%80 %EA%B8%B0- %EC%89%BD %EA%B2%8C- %EA %B5%AC%ED %98%84%ED %95%98%EA%B8%B0- abd967638c8e