

Personalized Cross-Domain Recommendation of Books Based on Video Consumption Data

Yea Bin Lim[†] · Hyon Hee Kim^{††}

ABSTRACT

Recently, the amount of adult reading has been continuously decreasing, but the consumption of video content is increasing. Accordingly, there is no information on preferences and behavior patterns for new users, and user evaluation or purchase of new books are insufficient, causing cold start problems and data scarcity problems. In this paper, a hybrid book recommendation system based on video content was proposed. The proposed recommendation system can not only solve the cold start problem and data scarcity problem by utilizing the contents of the video, but also has improved performance compared to the traditional book recommendation system, and even high-quality recommendation results that reflect genre, plot, and rating information-based user taste information were confirmed.

Keywords : Cross Domain, Hybrid Recommendation System, Word2Vec, Personalization, Deep Learning

영상 소비 데이터를 기반으로 한 교차 도메인에서 개인 맞춤형 도서 추천

임 예 빈[†] · 김 현 희^{††}

요 약

최근 성인 독서량은 지속적으로 감소하는데 비해 영상 콘텐츠 소비가 증가하고 있다. 이에 따라 새로운 사용자에게 대한 선호도 및 행동 패턴에 대한 정보가 없고 새로운 도서에 대한 사용자 평가나 구매 정보가 부족해 콜드 스타트 문제와 데이터 희소성 문제가 발생하고 있다. 본 논문에서는 영상물 콘텐츠 기반 도서 하이브리드 추천 시스템을 제안하였다. 제안하는 추천 시스템은 영상물의 콘텐츠를 활용하여 콜드 스타트 문제와 데이터 희소성 문제를 해결할 수 있을 뿐만 아니라, 전통적인 도서 추천 시스템에 비해 성능이 향상됨을 보여주었다. 또한 장르, 줄거리, 평점 정보 등 사용자 취향 정보까지 모두 반영한 개인 맞춤형 추천 결과를 제시하였다.

키워드 : 교차 도메인, 하이브리드 추천 시스템, Word2Vec, 개인화, 딥러닝

1. 서 론

2021년 통계청에서 발표한 연령대별 독서량에 따르면, 최근 교과서나 참고서를 제외한 일반 도서를 1년에 한 권 이상 읽은 성인의 비율은 46.9%에 불과하며, 특히 20대의 평균 독서량은 2011년엔 18.8권이었으나 2021년엔 8.8권으로 10년 사이 절반이 넘게 감소한 것으로 밝혀졌다[1]. 2022년 문화체육관광부에서 공개한 국민독서실태조사에 따르면 일과 공부

로 인한 시간 부족을 제외하고 성인과 학생 모두 다른 매체 또는 콘텐츠 이용을 독서 장애 요인으로 가장 많이 답변했다[2].

성인 독서량은 지속적으로 감소하는 데 비해 영상 콘텐츠 소비는 증가함에 따라서 도서 구매와 관련된 고객 데이터가 점점 부족해지고 있다. 따라서 새로운 사용자에게 대한 도서 구매 이력이 없거나 부족하여 추천이 어려운 콜드 스타트(Cold Start)문제가 발생하고, 데이터 부족으로 인해 효과적인 추천이 이루어지지 않는 데이터 희박성(Data Sparsity) 문제가 도서 추천에서 점차적으로 심화되고 있다[3].

이러한 문제를 해결하기 위해 본 논문에서는 교차 도메인 추천과 하이브리드 추천 방법을 사용했다. 성인 독서량 감소와 영상 콘텐츠 소비 증가라는 현대 사회의 실태를 고려해 영상물과 도서는 서로 다른 도메인 간 정보를 활용해 영상물을 활용한 도서 추천 시스템을 제안하여 데이터 부족으로 인

※ 이 논문은 2024년도 한국정보처리학회 춘계학술대회에서 '영상 콘텐츠 시청 데이터를 활용한 개인 맞춤형 도서 추천 시스템'의 제목으로 발표된 논문을 확장한 것이다.

† 비 회 원 : 동덕여자대학교 정보통계학과 학사과정

†† 중 심 회 원 : 동덕여자대학교 정보통계학과 부교수

Manuscript Received : July 2, 2024

Accepted : July 22, 2024

* Corresponding Author : Hyon Hee Kim(heekim@dongduk.ac.kr)

한 문제를 해결하고자 한다.

먼저 영상 및 도서의 콘텐츠만을 활용하여 추천하는 콘텐츠 기반 필터링 추천 시스템을 구축하였으며, 다음으로 비슷한 영화 취향을 가진 사용자들의 도서 정보들을 모아 관심사를 예측하는 협업 필터링 추천 시스템을 구현하였다. 마지막으로 두 개의 추천 시스템에서 각각 5권을 반환하도록 한 후 가중치를 조절하여 콘텐츠별 장르, 줄거리, 평점 정보 모두를 적절하게 반영한 최종 하이브리드 추천 시스템을 제안하였다.

본 논문의 공헌은 다음과 같다. 첫째, 본 연구에서 제안하는 교차 도메인에서의 하이브리드 도서 추천 시스템은 단순히 도서 정보를 활용하는 기존 도서 추천 시스템보다 성능이 뛰어나다. 둘째, 영화 시청 기록과 콘텐츠를 활용함으로써, 데이터 희박성과 콜드 스타트 문제를 완화하면서 개인 맞춤형 도서 추천이 가능하게 하였다. 셋째, 가중치를 조절하여 두 개의 추천 시스템을 반영하였기 때문에 장르, 줄거리, 사용자의 취향 모두를 적절히 반영된 결과가 추천된다.

본 논문은 다음과 같이 구성된다. 2장은 관련 연구에 대해 설명하고 3장은 데이터 수집과 전처리에 대해 설명한다. 4장에서는 추천 시스템 모델을 제안하고 성능 평가 결과를 보여준다. 5장에서는 분석 결과를 설명하고 6장에서는 결론 및 향후 연구를 제시한다.

2. 관련 연구

2.1 행렬 분해

딥러닝 기반의 추천 시스템에서 Matrix Factorization(MF)은 널리 사용되는 방법 중 하나이다. MF는 Netflix Prize[4] 통해 알려지게 되었다. 이 방식은 사용자와 아이템 간의 상호작용 데이터를 행렬로 표현하고, 이 행렬을 저차원의 잠재 요인 공간으로 분해하여 추천을 수행한다. 이러한 접근 방식은 사용자와 아이템 간의 잠재적인 관계를 효과적으로 모델링할 수 있어 추천 시스템의 성능을 크게 향상시킨다.

행렬 분해(Matrix Factorization)의 주요 아이디어는 사용자와 아이템 상호작용 행렬 R 을 두 개의 저차원 행렬 P 와 Q 로 분해하는 것이다. 여기서 P 는 사용자 잠재 요인 행렬, Q 는 아이템 잠재 요인 행렬이다. 이 두 행렬의 곱을 통해 원래의 상호작용 행렬을 근사할 수 있다. 이를 수식으로 나타내면 다음과 같다:

$$R \approx P \times Q^T$$

특히 대규모 데이터셋에서의 성능이 뛰어나며, 사용자와 아이템 간의 잠재적인 선호도를 효과적으로 학습할 수 있다.

2.2 협업 필터링 추천 시스템

협업 필터링 추천 기법(Collaborative Filtering)[5]은 사용자의 구매 이력이나 평점 기록 등을 사용하여 추천하는 방식이다. 추천 아이템에 대한 선호도가 유사하면 다른 아이템에서

도 유사한 취향을 갖고 선호도를 나타낼 것이라고 가정하여, 사용자 또는 아이템 간의 유사도를 기반으로 선호도를 예측하는 추천 방법이다. 평점 기록을 바탕으로 사용자 간의 유사도를 측정하는 방법을 사용자 기반의 협업 필터링(User-based Collaborative Filtering)이라고 한다. 음악추천이나 Facebook의 친구추천 서비스가 대표적인 예시이며, 사용자가 부여한 평점 정보를 이용하여 사용자와 유사한 성향을 갖는 사용자를 선별한 뒤, 선별된 사용자가 공통적으로 선호하는 아이템을 최종적으로 추천해 주는 방식이다.

아이템 기반 협업 필터링(Item-Based Collaborative Filtering)[6]은 Youtube와 Netflix의 영상 추천, 아마존과 같은 쇼핑 사이트의 제품 추천 등에 주로 사용된다. 사용자 기반 협업 필터링과는 달리 아이템의 유사도를 먼저 측정하여, 과거 아이템 선호도 데이터를 바탕으로 유사도가 높은 다른 아이템을 추천하는 방식이다[7].

앞서 언급한 문제점은 도서 추천 시스템에도 유사하게 적용된다. 미디어의 활발한 보급으로 인해 대부분의 사람들은 책 이외의 다른 매체나 콘텐츠(Netflix, YouTube 등)를 이용하며 독서를 하지 않는 경우가 많다. 그렇기에 도서 추천을 원하는 사용자들의 대부분은 책을 읽지 않은 경우가 대다수며 그 평가 또한 적기에 책에 비해 상대적으로 데이터가 풍부한 영화 평점 정보를 활용하여 추천을 수행하는 연구 또한 진행되었던 바가 있다[8].

2.3 하이브리드 추천 시스템

콘텐츠 기반 필터링은 추천을 위한 상품에 대한 특성을 추출해야 하는 반면, 사실상 효과적으로 이뤄지기 어렵기 때문에 분석의 깊이가 얕아지는(Shallow Analysis) 현상이 일어날 수밖에 없으며, 추천 결과가 이전에 좋게 평가한 상품과 비슷한 상품으로 추천이 이루어지기에 너무 과도하게 특정 부분에 치우치게 되는(Overspecialization) 문제 등의 한계점이 있다. 협업 필터링 기반 추천 시스템은 초기 사용자 문제와 데이터가 너무 과도하게 많을 경우, 연산 시간이 과도하게 소요되는 문제(Scalability Problem)의 단점이 있기도 하다.

이러한 한계점을 극복하기 위하여 새로운 정보의 분석 결과를 추천 알고리즘에 결합하는 하이브리드 접근법에 대한 연구들이 최근 다양하게 시도되고 있는 추세이다[9]. 평점 예측에서 우수한 성능을 보인 협업 필터링 추천 시스템과 높은 적중률을 보인 콘텐츠 기반 필터링을 결합하여 각 모형의 단점을 보완한 하이브리드 추천 시스템 연구가 진행되었던 바가 있다[10].

본 연구에서는 기존에 진행되었던 연구들과 달리, OTT 플랫폼의 영화 콘텐츠를 기반으로 하여 사용자들의 취향 파악이 유리했다는 점과 가중치를 조정해 콘텐츠별 장르 및 줄거리 정보와 유사한 취향을 가진 사용자들의 평점 정보 모두를 적절하게 반영하여 최종적으로 도서를 추천해 준다는 점에서 차별화를 두고 있다.

3. 데이터 수집 및 전처리

본 연구에서는 사용자의 영상 콘텐츠에 대한 취향을 기반으로 도서를 추천하기 위해 교보문고와 영화관 입장권 통합 전산망 사이트를 통해 각각 베스트셀러 도서 98권과 역대 박스오피스 순위 상위 100위의 영화를 활용했다. 왓차피디아에서 각 영화와 도서에 대한 평점을 남긴 사용자 데이터를 통해, 해당 사용자가 평점을 남긴 영화와 도서를 추가로 크롤링 하여 총 624개의 영상 콘텐츠와 716개의 도서 데이터, 2,251개의 사용자 평점 데이터를 수집했다.

콘텐츠 기반 필터링 알고리즘에서는 도서와 영화 각각에 대한 장르, 줄거리를 변수로 활용했다. 장르의 경우, 콘텐츠 별로 분류 기준이 상이하고 세부적으로 분류되어 있어 인문, 자기 계발, 시/에세이, 소설, 과학, 역사/문화, 예술/대중문화 총 7개의 장르로 재분류하였다. 줄거리 변수는 Mecab 형태소 분석기를 사용해 명사만 추출한 뒤 토큰화를 진행했다. 명사화된 단어 중 이미 다른 변수로 활용된 장르와 관련된 단어 등을 불용어 리스트로 설정해 변수의 특성을 높였다.

협업 필터링 알고리즘에서는 사용자의 평점을 변수로 하였으며, 사용자의 평점이 없는 콘텐츠의 경우, 딥러닝을 활용해 평점을 예측하여 결측치를 채웠다.

4. 추천 알고리즘

본 연구에서는 콘텐츠 기반 필터링 추천 알고리즘과 협업 필터링 추천 알고리즘을 구현한 후, 두 시스템의 가중치를 조절하여 하이브리드 추천 알고리즘을 구현하여 각 추천 알고리즘의 성능을 비교하였다.

4.1 콘텐츠 기반 추천

본 기법에서는 영화와 도서의 키워드, 장르 데이터를 사용하여 영화 콘텐츠를 기반으로 도서를 추천한다. Algorithm 1은 콘텐츠 기반 추천 시스템의 코드를 유사 코드로 표현한 것이다. 먼저, 모든 도서와 영화의 키워드들을 하나의 리스트로 합친다. 그 후 사용자에게 선호하는 영화 목록들을 입력받은 후 평점을 3.5이상 남긴 사용자들을 긍정적으로 평가했다고 정의하여 긍정적으로 평가한 사용자 목록을 추출했다.(3-5) 그리고 Word2Vec 알고리즘을 이용해 각 도서와 영화의 키워드 리스트를 임베딩한 뒤 코사인 유사도를 계산한다.(6-9) 그 후 추출된 사용자들이 읽은 도서 리스트 중에서 입력된 영화와 코사인 유사도가 높은 순으로 5권의 도서를 반환하도록 했다.(11-12)

4.2 협업 필터링 추천

협업 필터링 추천 시스템에서는 평점 데이터를 사용해 비슷한 영화 취향을 가진 사용자들의 도서 정보들을 모아 관심사를 예측하는 방식으로 추천하고자 했다. Algorithm 2는 협업

```

Algorithm 1: Content Based Recommendation

1  function content_based(items_df, cosine_df, movie_id, k)
      users <- list of all "user_id" from items_df
      where the rating is 3.5 or higher for the given
2  movie_id
      item_id <- list of item_id rated by users
3  cat_list <- list of category rated by users
4  book_list <- empty list
5  for i. item in item_list do
6      if cat_list[i] == 'book' :
7          sim <- similarity score from cosine_df for
8          movie_id and item
9          append (item, sim) to book_list
10 end for
11 sort book_list in descending order by similarity score
12 return top k elements from book_list
13 end function
    
```

Algorithm 1. Content Based Recommendation

```

Algorithm 2: Collaborative Filtering Recommendation

1  function Collaborative(movie_id, book_ratings_df,
      movie_ratings_df, k)
2      user_list <- list of all "user_id" from movie_ratings_df
      where "item_id" is equal to movie_id and rating >= 3.5
3      book_df <- select all rows from book_ratings_df where
      "user_id" is in user_list
4      book_ratings <- group book_df by "item_id" and calcul
      ate the average "rating" for each group
5      book_ratings <- sort book_ratings in descending order
      by "rating"
6      book_ids <- indices of the top k items in book_ratings
7      return user_list, name[name's "item_id" is in
      book_ids], top k items from book_ratings
8  end function
    
```

Algorithm 2. Collaborative Filtering Recommendation

필터링 추천 시스템의 코드를 유사 코드로 표현한 것이다. 먼저 사용자에게 선호하는 영화 목록들을 입력받는다. 평점을 3.5점 이상 남긴 사용자들을 긍정적으로 평가했다고 정의하여 입력된 영화에 대해 평점을 3.5점 이상 남긴 사용자들을 필터링한다.(2-3) 필터링된 사용자들이 평가한 각 책들의 평균 평점을 예측 평점으로 구한 뒤(4), 높은 평점 순으로 정렬하여 최종 도서 5권을 반환하도록 했다.(5-7)

4.3 하이브리드 추천

앞선 콘텐츠 기반 필터링과 협업 필터링 추천 결과를 동시에 반영한 하이브리드 추천 시스템을 구축하였다. 제안하는 하이브리드 모델로 추천 가능한 책 목록에 대해 예측 평점과 줄거리 유사도의 반영 비율을 달리하여 상위 5개의 도서를 반환하였다. 즉 콘텐츠 기반 필터링 추천과 협업 필터링 추천의 가중

치를 조절해 콘텐츠별 장르, 줄거리 정보와 유사한 취향을 가진 사용자들의 평점 정보 모두를 적절하게 반영하도록 했다.

이때, 하이브리드 추천 시스템의 예측 성능을 최대화하기 위해 콘텐츠 기반 필터링과 협업 필터링의 적절한 가중치를 찾는 과정을 수행하였다. 사용한 수식은 다음과 같다 :

$$(\lambda_1 * similarity) + (\lambda_2 * rating) \quad (1)$$

먼저, 협업 필터링에 가중치를 0.8을 부여한 결과, 콘텐츠에 대한 정보는 반영되지 못하고 사용자별 평점 데이터 위주로만 추천된 경향을 파악할 수 있었다. 따라서, 협업 필터링의 가중치를 낮춰 0.4를 부여한 결과, 사용자별 평점 정보는 반영되지 못하고, 콘텐츠에 대한 정보 위주로만 추천되었다. 실험 결과, 6 : 4로 반영했을 때 가장 성능도 높고 콘텐츠에 대한 정보와 사용자별 평점 정보가 적절하게 반영되어 추천되는 것을 확인할 수 있었다. 따라서, 이를 최종 추천 알고리즘의 가중치로 사용했다.

Algorithm 3은 하이브리드 추천 시스템의 코드를 유사 코드로 표현한 것이다. 먼저, 앞서 실험을 통해 설정한 가중치를 weight 리스트에 설정했다.(2) 그 다음, 4.1절에서 구현한 콘텐츠 기반 추천 시스템을 통해 추천된 도서 리스트를 추출한다.(3-9) 추출된 도서 리스트와 해당 점수를 이용해 'r1' 데이

터 프레임 생성하고, score에 가중치를 곱하여 'score2'라는 변수를 추가한다.(10-11) 이후 4.2절에서 구현한 협업 필터링에서 사용자 리스트, 추천된 도서, 도서 평점의 결과를 추출한다.(12) 추출된 도서 평점의 컬럼 이름을 score로 변경한 다음, 'r2' 데이터 프레임을 생성하여 마찬가지로 가중치를 곱하여 'score2'라는 컬럼을 추가한다.(13-15) 두 가지 추천 시스템의 데이터 프레임인 'r1'과 'r2'를 결합하여 하나의 데이터 프레임 'df'로 만든 후, 'item_id'로 그룹화하여 'score2'를 합산한다.(16-17) 마지막으로, 'score2'를 기준으로 내림차순 정렬하여 상위 5개의 항목을 반환하도록 했다.(18-19)

5. 성능 평가 및 결과 해석

5.1 성능 평가

본 연구에서는 추천 시스템의 성능을 추천하기 위해 Precision, MAP(Mean Average Precision), NDCG[11]를 사용하였다. Precision은 추천된 아이템 중 실제 사용자가 선호하는 아이템의 비율을 측정하는 지표로, 추천된 아이템 중에서 얼마나 많은 아이템이 실제 사용자에게 맞는지 측정해 정확성을 나타낸다. MAP는 Precision의 평균을 취한 값으로, 여러 추천 리스트에 대한 정밀도를 고려한다. 따라서 정확한 순위를 평가하는 데 유용하다. 마지막으로 NDCG(Normalized Discounted Cumulative Gain)는 순위에 가중치를 두고 추천 리스트의 품질을 측정하는 지표로, 해당 아이템의 실제 선호도와 함께 고려한다. 따라서 사용자에게 높은 순위의 아이템이 얼마나 중요한지를 반영하는 지표이다.

먼저, 본 논문에서 제안하는 영화 기반 도서 추천 시스템과 도서 데이터만을 가지고 추천한 전통 추천 시스템과의 성능을 비교했다. 아래 Fig. 1은 성능 평가 결과이다. 성능 평가 결과 전통 추천 시스템은 Precision 0.88, MAP, NDCG는 각각 0.89를 기록했다. 반면, 영화 기반 도서 하이브리드 추천 시스템은 Precision 0.96, MAP 0.98, NDCG 0.97로 전통 추천 시스템에 비해 모든 지표에서 성능 향상을 보였다.

```

Algorithm 3: Hybrid Recommendation
1  function HybridRec(items, cosine, movie_id, k,
   book_pred_ratings_df, movie_pred_ratings_df)
2      Set weight: [0.6, 0.4]
3      books <- Find similar books using content_filtering(ite
   ms, cosine, movie_id, k)
4      s <- 4.5
5      Initialize empty lists ids and rs
6      for book (i, v) in books do
7          append i to ids
8          append s to rs
9          s -= 0.1
10     end for
11     r1 = pd.DataFrame({'item_id': ids, 'score': rs})
12     r1['score2'] = r1['score'] * weight[0]
13     user_list, recommend_books, book_ratings <- collabor
   ative_rec
14     book_ratings.columns = ["score"]
15     r2 <- Reset index of book_ratings
16     r2['score2'] <- r2['score'] * weight[1]
17     df <- pd.concat([r1, r2], axis=1)
18     df <- df.groupby('item_id').sum().reset_index()
19     final <- Sort df by "score2" in descending order and sel
   ect top 5
20     return user_list, "item_id" values from final
end function
    
```

Algorithm 3. Hybrid Recommendation

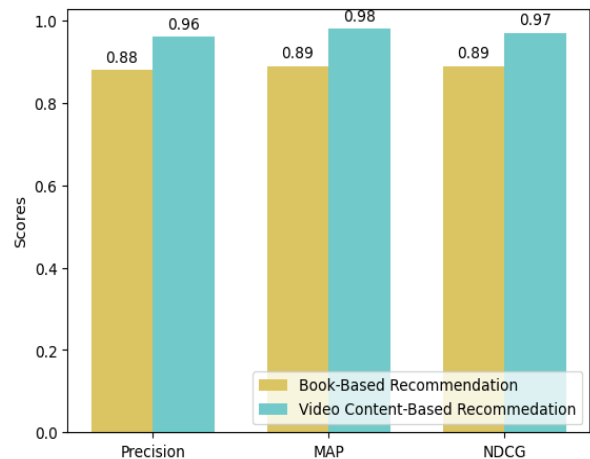


Fig. 1. Performance Evaluation

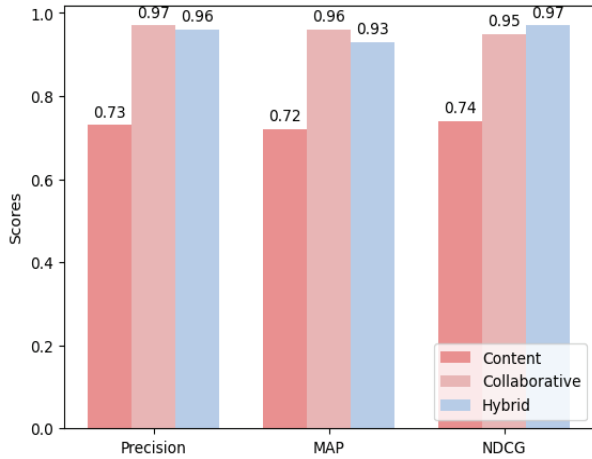


Fig. 2. Performance Evaluation of Video Content-Based Recommendation System Models

다음으로, Fig. 2는 영화 기반 도서 추천 시스템 3가지의 성능을 비교한 결과이다. 먼저, 콘텐츠 기반 필터링 추천을 진행한 결과 Precision 0.73, MAP 0.72, NDCG 0.74를 기록했다. 다음으로, 협업 필터링 추천을 진행한 결과 Precision은 0.97, MAP 0.93, NDCG 0.95로 콘텐츠 기반 추천을 진행한 경우에 비해 성능 향상을 보였다. 마지막으로 하이브리드 추천 시스템은 Precision 0.96, MAP 0.98, NDCG 0.97로 대부분의 지표에서 가장 우수한 성능을 보였다.

5.2 결과 해석

세 가지 추천 시스템에 여러 가지 영화를 입력하였을 때 추천받은 결과를 Table 1과 Table 2로 나타내었다. 먼저, Table 1은 범죄 영화인 ‘다만 악에서 구하소서’를 입력했을 때의 결과를 나타낸다. 최종 결과인 하이브리드 기반 추천 결과에서 1,2,4위로 추천된 도서들은 모두 범죄와 관련한 도서로 줄거리 기반으로 잘 추천이 된 것을 확인할 수 있었다. 특히, 사용자 평점 면에서 2위였던 ‘앵무새 죽이기’가 범죄와 관련한 줄거리는 아니지만 하이브리드 추천을 통해 최종 3위로 추천된 것을 확인할 수 있었다. 이를 통해 장르, 줄거리의 특색이 잘 보이며 줄거리 및 도서 평점으로 추천이 잘된 것을 확인할 수 있었다.

Table 2는 역사에 관한 영화인 ‘광해’와 ‘봉오동 전투’를 입력했을 때의 결과를 나타낸다. 먼저, 콘텐츠 기반 필터링에선 단순히 내용이 비슷한 책 위주로 추천이 이뤄졌다. ‘광해’와 ‘봉오동 전투’를 입력했을 때, 순서대로 ‘역사의 쓸모’, ‘슬기로운 세계사’, ‘최소한의 한국사’ 등이 추천되었다. 그러나 줄거리와 사용자 평점까지 포함한 하이브리드 추천에서는 ‘역사의 쓸모’, ‘슬기로운 세계사’, ‘나는 소망한다 내게 금지된 것들’, ‘최소한의 한국사’ 등이 추천되었다.

즉, 단순 키워드 매칭을 보면 ‘최소한의 한국사’가 ‘나는 소망한다 내게 금지된 것들’ 보다 더욱 일치하지만, 하이브리드 추천 시스템을 통해 ‘광해’와 ‘봉오동 전투’를 선호한 사용자

Table 1. Experimental Results

순위	협업 필터링 기반 추천	콘텐츠 기반 추천	하이브리드 기반 추천
1	여름 언덕에서 배운 것	암살자닷컴	암살자닷컴
2	앵무새 죽이기	사악한 소년	백조와 박쥐
3	심판	백조와 박쥐	앵무새 죽이기
4	블랙 쇼맨과 이름 없는 마을의 살인	끝없는 밤	푸른 불꽃

Table 2. Experimental Results

순위	협업 필터링 기반 추천	콘텐츠 기반 추천	하이브리드 기반 추천
1	죄와 벌	역사의 쓸모(3.9)	역사의 쓸모(3.9)
2	헤어질 결심	슬기로운 세계사(3.4)	슬기로운 세계사(3.4)
3	나는 소망한다 내게 금지된 것들(4.1)	최소한의 한국사(3.8)	나는 소망한다 내게 금지된 것들(4.1)
4	동물농장	30개 도시로 읽는 한국사(3.3)	최소한의 한국사(3.8)

들이 더욱 좋게 평가한 책인 ‘나는 소망한다 내게 금지된 것들’이 더 높은 순위가 된 것을 확인할 수 있다.

즉, 한 가지 영화를 입력했을 때의 결과가 좋았지만 두 가지 이상의 영화를 입력했을 때도 또한 추천의 질이 좋은 것을 결과를 통해 알 수 있었다. 다양한 영화 입력을 통해 각 영화의 장르와 줄거리에 맞는 도서가 추천되는 것을 확인할 수 있었고 하이브리드 추천 시스템이 추천의 질이 가장 좋음을 알 수 있었다.

하지만 선호하는 영화들의 장르가 서로 다른 경우, 장르의 유사성이 떨어져 사용자의 선호도가 크게 달라지고 줄거리 유사도가 많이 차이 나게 되어 줄거리보다는 도서 평점에 기반한 추천이 많았다.

6. 결 론

본 연구에서는 기존의 전통적인 도서를 통한 도서 추천 시스템의 한계를 탈피하기 위하여 사용자의 영화 취향을 기반으로 도서를 추천하는 모델을 제안하였다. 하이브리드 추천 시스템을 구현하기 위하여 콘텐츠 기반 필터링과 협업 필터링 추천을 각각 구현한 후 가중치를 조절하며 실험을 진행하였다.

기존의 전통적인 도서 추천 시스템의 경우, 도서를 추천받기 위해서는 사용자가 선호하는 도서를 입력해야 하기 때문에 독서량 부족이 문제시되는 콜드 스타트 문제가 발생하기 쉽다는 단점이 존재한다. 반면, 영상 콘텐츠를 기반으로 도서를 추천하는 하이브리드 추천 시스템을 활용하였을 경우, 새로운

도메인이 추가되었음에도 불구하고 기존 도서 추천 시스템과 비교해 보았을 때, 예측 성능이 크게 저하되지 않았다. 또한, 접근성이 뛰어난 영상 콘텐츠를 기반으로 도서를 추천하기 때문에 독서량 부족의 가장 큰 원인인 ‘책 이외의 다른 매체, 콘텐츠 이용’, ‘스마트폰, TV 등을 이용’, ‘읽을만한 책이 없음’ 등의 문제를 해결할 수 있을 것으로 기대된다. 나아가, 비교적 쉽게 수집할 수 있는 변수들을 예측 모델에 활용하기 때문에 기술 응용 및 적용 가능성도 높아 향후 도서 산업 부흥에 도움을 줄 수 있을 것으로 보인다.

References

[1] G. Mun, "The book-covered south korea, how about you?... 53% of adults haven't read a single book in a year," Mael Business Newspaper, 2022. [Internet], Available: <https://www.mk.co.kr/news/society/10482768>

[2] Ministry of Culture, Sports and Tourism. Cultural policy. pp.12-14, 2022.

[3] T-H. Kim and S. K. Kim, "SVD-based cross-domain recommendation using k-means clustering," *Journal of KIISE*, Vol.49, No.5, pp.360-368, 2022.

[4] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems." *Journal of Computer*, Vol.42, No.8, pp.30-37, 2009.

[5] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry, "Using collaborative filtering to weave an information tapestry", *Communications of the ACM*, Vol.35, No.12, pp.61-70, 1992.

[6] B. G. Sarwar, J. Karypis, J. Konstan, and J. Reidl, "Item-based collaborative filtering recommendation algorithms", in *Proceedings of the 10th International Conference on World Wide Web, Hong Kong*, pp.285-295, 2001.

[7] J. Hyun, S. Ryu and S-Y. T. Lee, "How to improve the accuracy of recommendation systems: Combining ratings and review texts sentiment score" *Journal of Intelligence and Information Systems*, Vol.25, No.1, pp.219-239, 2019.

[8] S. Kim, S. Han, H-E. Mok, and H. Choi, "Cross media-platform book recommender system: Based on book and movie ratings," *Journal of the Convergence on Culture Technology (JCCT)*, Vol.7, No.1, pp.582-587, 2021.

[9] M. J. Ku and H. Ahn, "A hybrid recommender system based on collaborative filtering with selective use of overall and multicriteria rating," *Journal of Intelligence and Information Systems*, Vol.24, No.2, pp.85-109, 2018.

[10] J. Yoon and J. Kim, "A hybrid recommendation method in cross domain based on user preferences and content information," *Journal of Korean Institute of Intelligent Systems*, Vol.32, No.6, pp.486-493, 2022.

[11] K. Järvelin and K. Jaana, "IR evaluation methods for retrieving highly relevant documents.," in *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '00)*. Association for Computing Machinery, New York, NY, USA, pp.41-48, 2000.



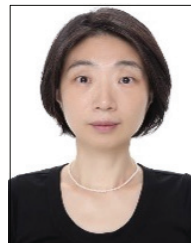
임 예 빈

<https://orcid.org/0009-0008-9393-9944>

e-mail : yell0315@naver.com

2024년 동덕여자대학교 정보통계학과
학사과정

관심분야 : Big Data Analysis, Deep Learning, Machine Learning, Data Engineering



김 현 희

<https://orcid.org/0000-0002-7507-8342>

e-mail : heekim@dongduk.ac.kr

1996년 이화여자대학교 컴퓨터학과(학사)

1998년 이화여자대학교 컴퓨터학과(석사)

2005년 이화여자대학교 컴퓨터공학과
(박사)

2005년 ~ 2006년 LG전자 디지털미디어 연구소 선임연구원

2006년 ~ 현 재 동덕여자대학교 정보통계학과 부교수

관심분야 : Machine Learning, DeepLearning, Big Data Analysis