

SBERT 임베딩을 활용한 행렬 분해 추천 방법론

강우승*, 이수원**, 최상민***

A Matrix Factorization-based Recommendation Approach with SBERT Embeddings

Woo-Seung Kang*, Suwon Lee**, and Sang-Min Choi***

본 논문은 교육부와 한국연구재단의 재원으로 지원 수행된 3단계 산학연협력 선도대학 육성사업(LINC 3.0), 정부(과학기술정보통신부) 재원의 한국연구재단 지원(RS-2022-00165785)과 2023년도 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 지자체-대학 협력기반 지역혁신 사업의 결과입니다(2021RIS-003).

요약

본 논문은 SBERT(Sentence-BERT) 임베딩을 MF(Matrix Factorization)모델에 통합한 영화 추천 방식을 제안한다. 전통적으로 MF 기반 추천 방식은 사용자 아이템 상호 작용 데이터에만 의존하기 때문에 영화의 메타데이터(제목, 장르, 개요 등)를 활용하지 않는다. 본 연구에서는 MF에서의 아이템 잠재 벡터의 표현을 SBERT 임베딩으로 대체하여 의미 정보를 효과적으로 활용할 수 있는 방식을 제안한다. 제안 방식에 대한 정확도 검증을 실제 영화 데이터에 대한 실험을 수행하고 SBERT를 활용한 메타데이터 벡터를 MF 방법에 주입한 모델과 전통적인 MF 모델과 비교한다. 그리고 실험 결과 분석을 통해 제안 방식의 우수성을 입증한다. 본 연구는 아이템의 메타데이터를 대상으로 SBERT 임베딩을 활용하여 협업 필터링 기술의 성능을 향상하는데 새로운 가능성을 제시한다.

Abstract

We propose a movie recommendation approach that integrates Sentence-BERT(SBERT) embeddings into Matrix Factorization(MF) models. Conventional MF-based recommender systems rely solely on rating data and do not utilize metadata such as titles, genres, or summaries. In this study, we suggest a method to effectively leverage semantic information by replacing item latent vectors in MF with SBERT embeddings. To validate the accuracy for the proposed approach, experiments are conducted on real movie dataset. We also compare the model incorporating metadata vectors generated by SBERT with conventional MF models. Through the analysis of experimental results, the superiority of the proposed approach is demonstrated. This research introduces new possibilities for enhancing the performance of collaborative filtering techniques using SBERT embeddings for item metadata.

keywords

recommender system, collaborative filtering, matrix factorization, sentence-bert

* 경상국립대학교 컴퓨터과학과 학사과정
- ORCID: <http://orcid.org/0000-0009-4568-5887>
** 경상국립대학교 컴퓨터과학과 부교수
- ORCID: <http://orcid.org/0000-0003-2603-1385>
*** 경상국립대학교 컴퓨터과학과 조교수(교신저자)
- ORCID: <http://orcid.org/0000-0001-5950-3081>

· Received: Oct. 11, 2023, Revised: Nov. 01, 2023, Accepted: Nov. 04, 2023
· Corresponding Author: Sang-Min Choi
Dept. of Computer Science, Gyeongsang National University, Jinju-si, 52828, Korea
Tel.: +82-55-772-1384, Email: jerassi@gnu.ac.kr

1. 서론

현대 사회의 디지털 기술 발달과 포스트 코로나 시대에 미디어 시장의 급격한 성장으로 웹 사용자들은 수많은 콘텐츠에 노출되었다[1]. 이에 따라 미디어 분야의 경우, 사용자들은 다양한 콘텐츠를 즐길 수 있게 되었지만, 넘쳐나는 콘텐츠 중에서 자신의 취향에 맞는 콘텐츠를 찾기 위해서 더 많은 시간이 소비되는 새로운 문제가 나타났다. 이러한 문제 완화를 위해 다양한 웹 서비스들에서 추천 시스템을 활용하고 있다. 이러한 문제 완화를 위해 다양한 웹 서비스들에서 추천 시스템을 활용하고 있다 [2]. 추천 방법론 중 대표적으로 행렬 분해(Matrix factorization) 방법론이 있다[3].

MF 방법론은 사용자와 아이템의 잠재 특징들 (Latent features)을 저차원 공간 벡터에 포함한다. 이를 통해 사용자의 선호도와 해당 아이템의 특징을 모두 통합된 방식으로 나타내어 사용자-아이템 상호 작용의 관련성 점수를 잠재 특징 공간에서 벡터의 내적 곱으로 간단히 측정한다. 전통적인 MF 방법론은 사용자의 선호도 정보를 활용하여 빠르게 예측 결과를 도출할 수 있는 장점이 있지만 실제 서비스에 적용 시 데이터 희소성으로 인하여 추천 결과의 신뢰성 및 콜드스타트와 같은 문제가 발생할 수 있다[4]-[6].

희소성 문제 이외에 MF 방법론이 갖는 또 다른 문제는 메타데이터와 같은 부가정보 활용에 폐쇄적이라는 점이다. 예를 들어, 영화 도메인은 영화 제목, 장르, 줄거리 요약과 같은 풍부한 메타데이터를 보유하고 있어 이러한 정보를 사용한다면 추천 성능을 더욱 향상할 수 있는 가능성이 존재한다. 그러나 MF 방법론은 사용자-아이템 상호작용 행렬만을 기반으로 학습하여 본질적으로 이러한 상황 정보를 통합하는 능력이 제한되어 있다[7].

본 논문에서는 이러한 문제점의 해결 방안으로 MF 방법론에서 메타데이터의 사용성 제고를 위한 방법론을 제안한다. 이를 위해 영화 도메인에서 MF 방법론에 메타데이터를 임베딩하여 주입하는 방법을 제안한다. 제안 방식을 통해 MF 방법론의 추천 성능의 향상을 기대할 수 있다.

본 논문은 다음과 같이 구성된다. 제2장에서는 MF 방법론의 선행 연구에 관해 설명한다. 제3장에서는 제안 방법론인 MF 방법론에 SBERT 기반의 메타데이터를 주입한 Metadata Injected Matrix Factorization을 설명한다. 제4장에서는 kaggle의 The Movies Dataset를 활용하여 본 논문에서 제안한 방법론과 전통적인 MF 방법론과의 성능 비교 실험 및 분석 결과를 제시한다. 마지막으로 제5장에서는 논문의 의의와 향후연구를 서술한다.

II. 관련연구

2.1 MF 기반 추천 시스템

MF 방법론의 주요 아이디어는 사용자-아이템 상호작용 행렬 R 을 일반적으로 사용자 잠재 특징 행렬 P 와 아이템 잠재 특징 행렬 Q 로 표시되는 두 개의 저차원 행렬로 분해하는 것이다. 이를 식 (1)과 그림 1과 같이 나타낼 수 있다[3].

$$R = P^T Q \tag{1}$$

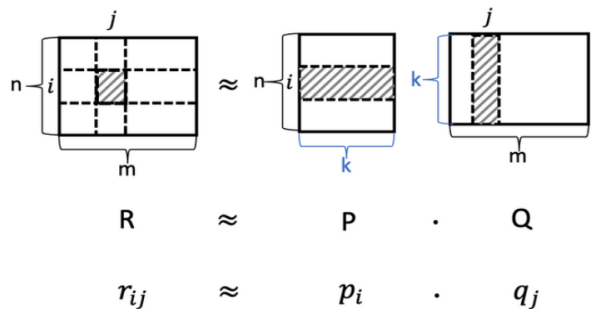


그림 1. MF 알고리즘 도식화
Fig. 1. Visualization of MF algorithm

그림 1에서 p_i 와 q_j 의 곱으로 표현되는 r_{ij} 는 실제 사용자의 아이템 평가로 영화 도메인에서는 평점으로 나타낼 수 있다. 그리고 k 는 잠재된 특징의 차원 크기를 나타낸다. 따라서 MF 방식은 행렬 분해를 통해 P 와 Q 로 분해하고, 각 행렬의 특정 행과 열을 기반으로 내적을 계산하여 R 의 누락된 값들을 추정할 수 있다. 즉, 사용자와 아이템의 잠재된 특징을 학습하여 사용자가 이전에 상호작용한 적이 없는 아이템을 얼마나 선호하는지 예측할 수 있다.

이는 손실함수를 정의한 후, 경사 하강법을 이용해 손실함수가 최소가 되는 사용자, 아이템의 잠재 특징 행렬을 찾는 과정이다. 행렬 분해의 손실함수는 아래 식 (2)과 같이 정의할 수 있다.

$$\min_{q,p,b} \sum_{(u,i) \in k} (r_{ui} - m - b_u - b_i - p_u^T q_i)^2 + \lambda (\|p_u\|^2 + \|q_i\|^2 + b_u^2 + b_i^2) \quad (2)$$

식 (2)에서 k 는 r_{ij} 의 값이 존재하는 사용자(u) 아이템(i)의 집합이다. m 은 전체 아이템 평점의 평균이다. 그리고 b_u , b_i 는 각각 사용자와 아이템의 편향을 나타내며 이를 통해 사용자나 아이템별로 평점에 편향이 있을 수 있는 점을 고려할 수 있다. $p_u^T q_i$ 는 사용자, 아이템 특징 벡터 곱을 나타낸다. 따라서 식 (2)를 통해 MF 방식은 사용자, 아이템의 잠재 특징 행렬(p, q)와 사용자 와 아이템의 편향(b_u , b_i)을 학습한다. 그리고 나머지 부분인 $\lambda (\|p_u\|^2 + \|q_i\|^2 + b_u^2 + b_i^2)$ 은 과적합을 방지하기 위한 정규화 부분이다.

2.2 추천 시스템 성능 개선 연구

MF 방법론을 이용한 추천시스템 연구에서는 기존 문제들을 해결하기 위하여 여러 연구가 진행되었다. 특히 콘텐츠 기반 필터링 방법과 MF 방법론을 결합한 하이브리드 추천 방법 등을 통해 평점 이외의 요인을 이용한 연구를 수행하고 있다[8]-[10]. 이와 더불어 아이템이 소비되는 상황 문맥을 고려하는 방법, 아이템 리뷰의 감성 점수를 고려하는 방법 등을 사용하여 연구를 진행하고 있다[11]-[13].

본 논문에서도 추천 성능을 개선하기 위하여 MF 방법론을 활용한다. 특히 기존의 MF 방법론에 메타데이터를 활용하지 못하는 점을 해소하기 위하여 SBERT를 통한 메타데이터 임베딩 벡터를 MF 방법론과 통합하여 추천 성능의 개선을 시도한다.

III. 제안 방식

본 연구의 제안 방식은 메타데이터 주입 행렬 분해(MIMF)라는 MF 추천 시스템에 대한 새로운 접근

방식이다. MIMF 모델은 영화 메타데이터, 특히 제목, 장르 및 개요를 아이템의 잠재 벡터로 통합함으로써 전통적인 MF 방법이 갖는 메타데이터 활용성 문제를 개선한다. 이를 위해 문장 임베딩에 최적화된 SBERT(Sentence-BERT)를 사용한다.

3.1 SBERT를 사용한 영화 메타데이터 임베딩

메타데이터는 다양한 콘텐츠 도메인에서 중요한 역할을 하며, 추천 시스템을 향상할 수 있는 가치 있는 정보를 제공한다[14]. 본 연구에서는 영화 메타데이터를 활용한다. 영화 메타데이터는 제목, 장르, 감독, 배우, 개요 등과 같은 다양한 정보를 포함하고 있다. 이 중에서도 제목, 장르, 줄거리와 같은 요소들은 텍스트로서 영화와 관련된 풍부한 정보를 담고 있다[15]-[17].

본 논문에서는 사용자의 선호 아이템이 존재한다면 아이템 메타데이터가 선호를 포함한다는 가설을 기반한다. 이러한 가설을 기반으로 메타데이터 선호 정보가 포함된 추천 모델을 구성하여 추천의 정확도를 제고한다. 텍스트 메타데이터 의미를 MF 방법론 추천 시스템에 활용하기 위해, SBERT 모델을 사용한다[18].

SBERT는 사전 훈련된 BERT(Bidirectional Encoder Representations from Transformers) 모델의 특수한 형태로, BERT 모델의 문제점을 해결한다. SBERT는 두 개의 BERT에 문장을 하나씩 입력하여 BERT 간에 서로 같은 weight를 공유하는 Siamese Network를 구성한다. 그리고 BERT의 출력값 벡터에 pooling 연산을 추가하여 문장의 임베딩 벡터의 평균값으로 풀링하여 벡터를 생성한다. 이를 통해 기존 BERT와 비교하여 연산량을 줄이고 더 나은 문장 임베딩 값을 얻는다.

그림 2는 Regression을 목적으로 학습된 SBERT 모델이다. 이 모델은 두 문장의 벡터(u, v)를 생성하여 cosine-similarity를 계산하고 MSE(Mean Squared Error) loss를 통해 weight를 갱신하며 학습된다.

본 논문에서는 영화 메타데이터의 텍스트 요소, 특히 영화 제목, 장르, 그리고 줄거리를 고차원 벡터로 변환하기 위해 이 SBERT를 사용한다. 이러한 임베딩 과정은 텍스트 정보와 수치적 특성 간의 연

결 역할을 하며, 이를 활용하여 메타데이터를 MF 기반 추천 시스템에 원활하게 통합시킬 수 있다.

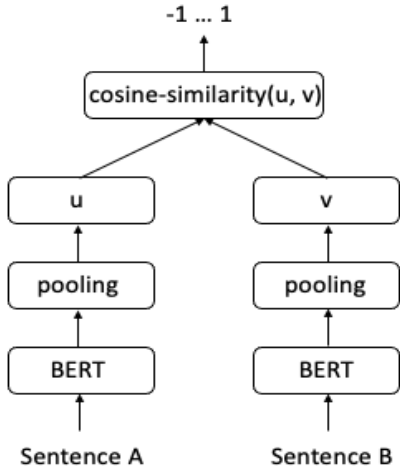


그림 2. SBERT 모델 구조
Fig. 2. Structure of SBERT model

3.2 행렬 분해와 메타데이터 임베딩 통합

본 논문에서는 SBERT를 활용하여 얻은 임베딩 벡터 ie 를 사용자-아이템 상호 작용 행렬의 행렬 분해 중에 아이템 특징 벡터로 고려한다. 전통적인 MF 방법론에서는 아이템과 사용자가 사용자-아이템 상호작용 행렬을 분해하여 학습된 잠재 특징으로 표현한다. 그러나 제안된 MIMF 모델에서는 이러한 아이템 잠재 특징을 처음부터 학습하는 대신 SBERT에서 미리 계산된 메타데이터 임베딩을 아이템 특징 벡터의 초깃값으로 설정한다. 그림 3은 제안 방법론의 예시를 나타낸다.

그림 3은 i 번째 사용자와 j 번째 아이템에 제안 방식을 적용한 예시이다. 그림 1의 전통적인 MF 방법론에서 j 번째 아이템 특징 벡터값을 영화 메타데이터 제목, 장르, 개요를 SBERT모델로 임베딩한 벡터 ie_j 로 대체하였다. 따라서 제안된 모델의 손실 함수는 식 (1)의 아이템 잠재 특징 벡터 p_i 를 아이템 메타데이터 임베딩 벡터 ie_i 로 대체하여 아래의 식 (3)으로 표현할 수 있다.

$$\min_{q, i, e, b} \sum_{(u, i) \in k} (r_{ui} - m - b_u - b_i - p_u^T i e_i)^2 + \lambda (\|p_u\|^2 + \|ie_i\|^2 + b_u^2 + b_i^2) \quad (3)$$

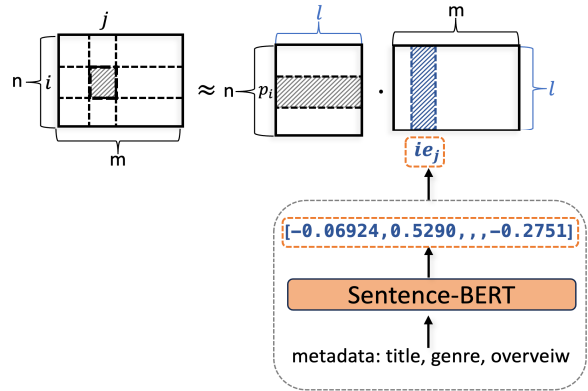


그림 3. 행렬 분해와 메타데이터 임베딩 통합 방법의 예시

Fig. 3. Example of integrating matrix factorization and metadata embedding method

식 (3)의 활용을 통해 본 연구에서는 아이템 잠재 특징의 초깃값을 영화의 메타데이터 임베딩 값으로 대체한다. 이는 사용자의 선호도에 아이템 메타데이터의 적용을 의미한다.

IV. 실험 및 결과

4.1 데이터 수집

본 논문에서는 kaggle의 The Movies Dataset 데이터셋[19]을 이용하여 실험을 진행하였다. 본 데이터셋은 700명의 사용자, 9,000편의 영화, 총 100,00개의 영화 평점으로 이루어져 있으며 사용자별로 20개 이상의 영화 리뷰 평점 데이터를 포함하고 있다. 또한 약 45,000편의 영화 메타데이터도 포함되어 있다. 메타데이터로 제목, 관람등급 성인 유무, 수위, 장르, 줄거리 등이 존재한다. 표 1은 The Movies Dataset의 메타데이터 정보를 나타낸다.

표 1. 영화 메타데이터
Table 1. Movie Metadata

ID	Meaning
id	Movie id
title	Movie title
genres	Movie genres
overview	Movie overview
language	Movie language
budget	Movie budget
adult	Viewer age restriction (Adult Content)

4.2 성능 평가

추천 성능 평가를 위해 본 연구에서는 추천 성능 지표로 활용되는 RMSE(Root Mean Squared Error)를 사용한다. RMSE를 사용하면 추천 시스템이 예측한 평점이 실제 사용자 평점에서 벗어나는 정도를 수량화할 수 있다. RMSE가 낮을수록 모델의 예측이 실제 사용자 선호도에 더 가깝다는 것을 의미한다. 식 (4)는 RMSE를 나타낸다. 식 (4)에서 r_{ui} 은 실제 사용자 평점을 의미하며 \hat{r}_{ui} 은 예측 평점을 의미한다.

$$SE = \sqrt{\frac{1}{n} \sum (r_{ui} - \hat{r}_{ui})^2} \tag{4}$$

4.3 데이터 분석

본 논문에서는 영화 메타데이터로 제목, 장르, 개요를 이용한다. 이를 SBERT로 임베딩했을 때 영화의 특징들이 잘 포착되었는지 확인하기 위하여 영화 데이터셋에서 28개의 영화를 추출하고 t-SNE(t-distributed Stochastic Neighbor Embedding) 알고리즘을 적용하여 벡터 공간에 사상된 데이터들을 시각화 한다. t-SNE는 비선형 차원 축소 기법으로, 고차원 데이터를 2차원 또는 3차원의 저차원으로 축소하여 시각화한다[20].

그림 4는 장르 데이터만을 임베딩 하여 시각화한 그림이다. 장르는 범주형 데이터로 같은 장르를 가진 영화 간의 거리가 가까이 형성되어 있다. 그리고 그림 5는 개요 데이터만을 임베딩 하여 시각화한 그림이다. 그림 5에서는 유사한 영화들 간의 거리가 가까이 형성 되어있다. 예를 들어 그림 5에서 ‘Batman’은 ‘Batman Forever’와 가까이 위치하지만, ‘Toy Story’와는 먼 지점에 있다. 마지막으로 그림 6은 제목, 장르, 개요 모두 같이 임베딩 하여 시각화한 그림이다.

그림 6은 그림 5와는 달리 장르와 제목의 데이터가 같이 임베딩 되어 유사한 영화 간의 거리가 더 근접하게 형성 되어있다. 예를 들어, 그림 5에서 ‘The mask’ 영화는 중앙에 위치하지만 그림 6에서는 ‘batman Forever’, ‘Robocob’과 가까운 거리에 위

치한다. 3가지 영화 모두 범주의 내용을 다루며 액션 영화임을 고려했을 때 유사한 영화임을 알 수 있다. 이를 통해서 SBERT를 이용하여 영화의 메타데이터를 임베딩 하였을 때 영화의 특징을 포착하고 있음을 알 수 있다.

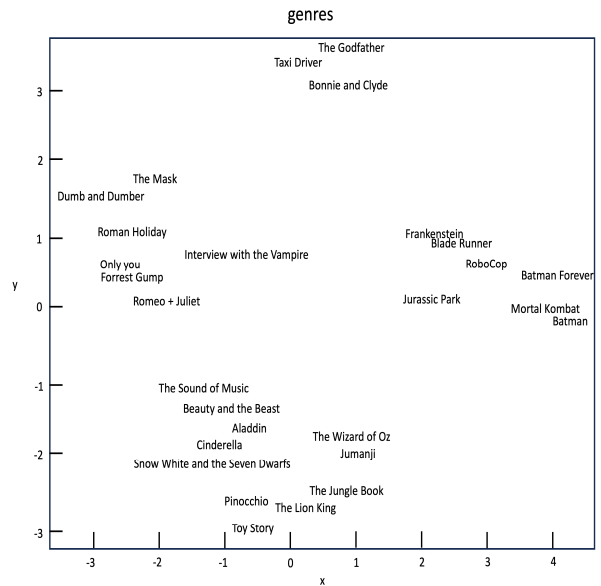


그림 4. t-SNE을 이용한 장르 임베딩 시각화
Fig. 4. Genre embedding visualization using t-SNE

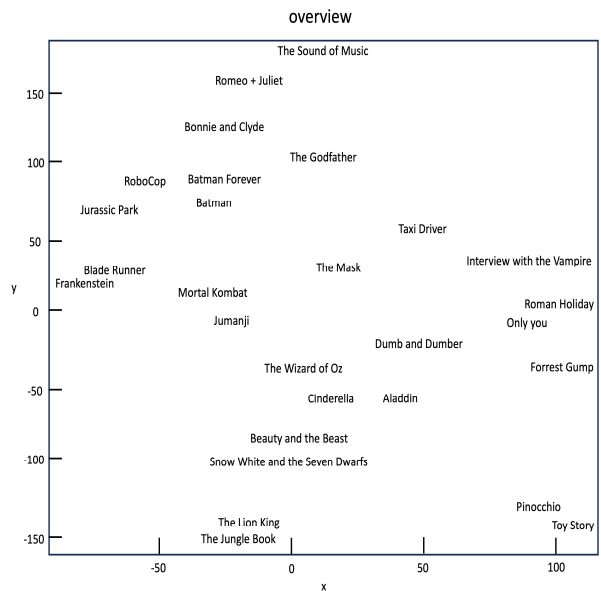


그림 5. t-SNE을 이용한 영화 줄거리 임베딩 시각화
Fig. 5. Movie overview embedding visualization using t-SNE

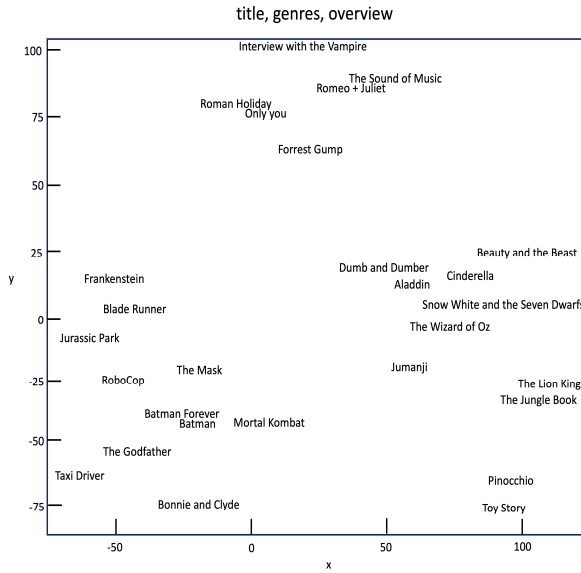


그림 6. t-SNE를 이용한 영화 제목, 장르, 줄거리 임베딩 시각화

Fig. 6. Movie title, genre and overview embedding visualization t-SNE

4.4 실험 결과 및 분석

제안 방법론인 MIMF의 성능 실험을 위해 전통적인 MF 방법론의 결과와 비교하였다. MIMF 방법론에 활용한 영화 메타데이터(제목, 장르, 개요)를 모든 가능한 조합을 표 2와 같이 구성한다.

표 2. 메타데이터 조합 방법론
Table 2. Method for combining metadata

Title	Genres	Overview	Title, Genres	Title, Overview	Genres, Overview	all factors
MIMF_T	MIMF_G	MIMF_O	MIMF_TG	MIMF_TO	MIMF_GO	MIMF

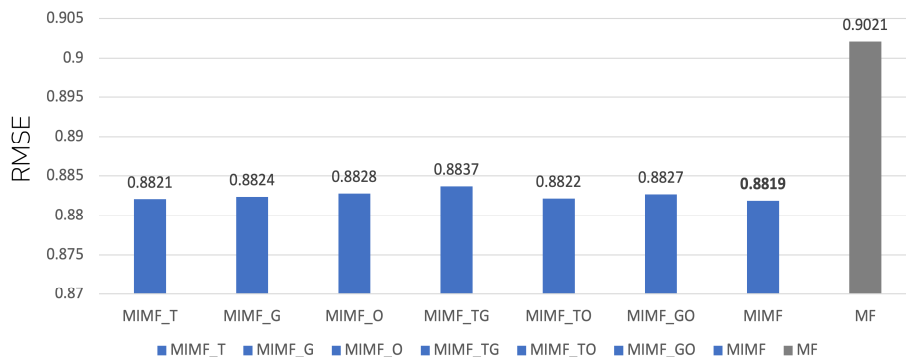


그림 7. 조합 MIMF 방법과 전통적인 MF 방법 RMSE 평가 그래프

Fig. 7. RMSE evaluation graph of combined MIMF method and traditional MF method

실험을 위한 기본 환경은 표 3과 같다. 메타데이터의 임베딩 벡터의 차원을 512로 두어 MF 방법론의 잠재 요인 512, 학습률 0.001, 과적합 방지를 위한 정규화 파라미터 0.005, 목적함수를 최소화하는 최적화 기법으로 SGD(Stochastic Gradient Descent)를 활용하였다. 또한 제안한 방법론(MIMF)의 성능평가를 위하여 10차 교차검증(10-fold cross validation)을 수행하였다.

표 3. 실험 기본 환경
Table 3. Experiment basic environment

Epoch	Latent factor	Regularized factor	Learning rate	optimazation
1000	512	0.005	0.001	SGD

실험 결과는 그림 7과 같다. 본 논문에서 제안하는 MIMF 방법론의 RMSE가 0.8819로 가장 낮으며, 메타데이터의 조합별 결과 모두 전통적인 MF 방법론의 0.9021보다 낮은 RMSE를 기록하였다. 이 관찰된 개선은 본 논문에서 제안한 MIMF 모델에 고유한 몇 가지 요소에 기인할 수 있다.

첫 번째는 SBERT를 사용하여 문서 메타데이터를 임베딩 하면 영화 제목과 장르 및 개요와 같은 텍스트 데이터 내에서 의미를 포착할 수 있다는 점이다. 이는 전통적인 MF 방식에서 포착하지 못하는 아이템의 추가적인 맥락을 제공할 수 있다. 두 번째는 영화 메타데이터를 아이템 잠재 특징 벡터에 포함함으로써 사용자가 선호하는 영화의 특성을 더욱 상세하게 나타낼 수 있다는 점이다.

결과에서 보이듯이 메타데이터가 하나 이상 포함되면 전통적인 MF 방법론보다 더 나은 추천 성능을 가질 수 있음을 보였다. 이는 기존의 랜덤한 아이템 잠재 특징 벡터보다 사용자 선호도를 더 정확하게 파악하는 데 도움이 될 수 있다는 것을 의미한다.

V. 결론 및 향후 과제

정보통신 기술의 발전과 더불어 미디어 콘텐츠 시장이 성장하면서 대량의 콘텐츠가 생산되고 있다. 이러한 환경에서 사용자에게 가장 적합한 콘텐츠를 제공하는 추천 시스템은 핵심적인 기술로 부상하였다[2]. 그러나 여전히 여러 문제점이 존재한다.

본 논문에서는 전통적인 MF 추천 시스템 방법론에서 메타데이터를 활용하지 못하는 단점을 보완할 수 있는 MIMF(Metadata Injected Matrix Factorization)를 제안한다. 제안 방식의 경우, SBERT를 이용하여 영화의 메타데이터(제목, 장르, 개요)를 임베딩하여 MF 추천 시스템에 주입한 방식이다. 그리고 실험을 통하여 본 논문에서 제안한 MIMF 방법론이 기존 MF 방법론과 비교하여 예측 정확도 관점에서 우수한 성능이 보임을 확인하였다.

본 연구의 의의는 다음과 같다. 첫 번째는 영화의 메타데이터를 SBERT 임베딩을 통해 고품질의 벡터 형태로 변환하여, 사용자와 아이템 간 상호 작용 데이터만을 활용하는 전통적인 방식에 비해 더욱 상세하고 정확한 아이템 특성 정보를 포착할 수 있다는 점이다. 두 번째는 영화 메타데이터의 종류별 활용검증을 통해 풍부한 영화 메타데이터를 활용하는 접근 방식이 전통적인 MF 방법론보다 더 나은 추천 성능을 보인다는 것을 입증한 점이다.

그러나 본 연구에서 제안한 방법론은 다음과 같은 향후 과제를 갖고 있다. 첫 번째로 메타데이터 삽입의 효과는 사용할 수 있는 메타데이터의 풍부성과 품질에 따라 다를 수 있다. 따라서 다양한 아이템에 대해 각각 비교분석을 통해 메타데이터를 검증할 필요가 있다.

두 번째로 본 연구의 방법론은 영화 도메인에만 적용한 것이므로, 다른 도메인의 추천을 위해 적용

하기는 어려움이 있을 수 있다. 다른 도메인의 추천 시스템에 적용하기 위해서 메타데이터의 변수 중 아이템 특성을 잘 드러내는 변수를 선정하여 군집 분석을 통해 적절한 메타데이터 변수를 선정하여 다른 도메인의 적용 가능성을 확인해야 할 것이다.

마지막으로 대규모 데이터 처리에 있어 확장성과 계산 복잡성이 실시간 추천 시스템에서 문제가 될 수 있다. MF기반의 추천 시스템에서는 실시간 추천에 대한 문제는 고려하지 않지만 향후 방법론의 확장을 위해 실시간 관점에 대한 고려가 필요하다. 이러한 한계점을 극복하기 위한 연구를 본 논문의 향후 과제로 한다.

References

- [1] M. J. Song, "Media Industry Ecosystem Changes in the Post-COVID-19 Era", *Broadcasting and Media*, Vol. 25, No. 4, pp. 9-17, 2020.
- [2] M. B. Dias, D. Locher, M. Li, W. El-Deredy, and P. J. G. Lisboa, "The value of personalised recommender systems to e-business: a case study", In *RecSys '08: Proceedings of the 2008 ACM conference on Recommender systems*, New York, NY, United States, pp. 291- 294, Oct. 2008. <https://doi.org/10.1145/1454008.1454054>.
- [3] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems", *Computer*, Vol. 42, No. 8, pp. 30-37, Aug. 2009. <https://doi.org/10.1109/mc.2009.263>.
- [4] T. Hastie, J. Friedman, and R. Tibshirani, "The elements of statistical learning: data mining, inference, and prediction", *Springer Series in Statistics*, New York: springer, Vol. 2, pp. 1-758, 2009. <https://doi.org/10.1007/978-0-387-21606-5>.
- [5] S. O. Kim and S. J. Lee "The Effect of Data Sparsity on Prediction Accuracy in Recommender System", *Journal of Internet Computing and Services*, Vol. 8, No. 6, pp. 95-102, Dec. 2007.
- [6] T. S. Kumari and K. Sagar, "A Semantic Approach to Solve Scalability, Data Sparsity and

- Cold-Start Problems in Movie Recommendation Systems", *International Journal of Intelligent Systems and Applications in Engineering*, Vol. 11, No. 6s, pp. 825-837, Mar. 2023.
- [7] P. Ingavélez-Guerra, S. Otón-Tortosa, J. Hilera-González, and M. Sánchez-Gordón, "The use of accessibility metadata in e-learning environments: a systematic literature review", *Universal Access in the Information Society*, Vol. 22, pp. 445-461, Dec. 2021. <https://doi.org/10.1007/s10209-021-00851-x>.
- [8] H. Parvin, P. Moradi, S. Esmacili, and N. N. Qader, "A scalable and robust trust-based nonnegative matrix factorization recommender using the alternating direction method", *Knowledge-Based Systems*, Vol. 166, pp. 92-107, Feb. 2019. <https://doi.org/10.1016/j.knosys.2018.12.016>.
- [9] S. Baek and D. Min, "Contents Preference Model Combined with Matrix Factorization for Movie Recommendation", *Journal of the Korean Institute of Industrial Engineers*, Vol. 47, No. 3, pp. 280-288, Jun. 2021. <https://doi.org/10.7232/JKIIIE.2021.47.3.280>.
- [10] H.-J. Xue, X. Dai, J. Zhang, S. Huang, and J. Chen, "Deep matrix factorization models for recommender systems", *IJCAI'17: Proceedings of the 26th International Joint Conference on Artificial Intelligence*, Melbourne, Australia, Vol. 17, pp. 3203-3209, Aug. 2017. <https://doi.org/10.24963/ijcai.2017/447>.
- [11] E. Frolov and I. Oseledets, "Tensor methods and recommender systems", *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Vol. 7, No. 3, May 2017. <https://doi.org/10.1002/widm.1201>.
- [12] A. Chauhan, D. Nagar, and P. Chaudhary, "Movie Recommender system using Sentiment Analysis", *2021 International Conference on Innovative Practices in Technology and Management (ICIPTM)*, Noida, India, pp. 190-193, Feb. 2021. <https://doi.org/10.1109/ICIPTM52218.2021.9388340>.
- [13] Y. Son, Y. Roh, S. Kim, and Y. Choi, "A Matrix Factorization Based Manuscript Editor Recommendation Method Considering Ratings and Sentiment Score of Clients", *Journal of KIISE*, pp. 835-837, 2019.
- [14] G. Behera and N. Nain, "Handling data sparsity via item metadata embedding into deep collaborative recommender system", *Journal of King Saud University-Computer and Information Sciences*, Vol. 34, No. 10, pp. 9953-9963, Nov. 2022. <https://doi.org/10.1016/j.jksuci.2021.12.021>.
- [15] S. Choi, S. Ko, and Y. Han, "A movie recommendation algorithm based on genre correlations", *Expert Systems with Applications*, Vol. 39, No. 9, pp. 8079-8085, Jul. 2012. <https://doi.org/10.1016/j.eswa.2012.01.132>.
- [16] G. Bae and H. Kim, "The impact of movie titles on box office success", *Journal of Business Research*, Vol. 103, pp. 100-109, Oct. 2019. <https://doi.org/10.1016/j.jbusres.2019.06.023>.
- [17] L. Chen and P. Pu, "Trust building in recommender agents", *Proc. of the Workshop on Web Personalization, Recommender Systems and Intelligent User Interfaces, Networks*, pp. 135-145, 2005. <https://doi.org/10.5220/0001422901350145>.
- [18] N. Reimers and I. Gurevych, "Sentence-bert: Sentence embeddings using siamese bert-networks", *Proc. of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, Hong Kong, China, pp. 3982-3992, Nov. 2019. <https://doi.org/10.18653/v1/d19-1410>.
- [19] "The Movie Dataset", <https://www.kaggle.com/datasets/rounakbanik/the-movies-dataset> [accessed: Sep. 14,, 2023]
- [20] L. V. D. Maaten and G. Hinton, "Visualizing data using t-SNE", *Journal of Machine Learning Research*, Vol. 9, Nov. 2008.

저자소개

강 우 승 (Woo-Seung Kang)



2019년 3월 ~ 현재 :
경상국립대학교 컴퓨터과학과
학사과정
관심분야 : 추천 시스템, 머신러닝,
빅데이터분석

이 수 원 (Suwon Lee)



2012년 2월 : 한국과학기술원
전산학과(공학석사)
2017년 2월 : 한국과학기술원
전산학과(공학박사)
2018년 3월 ~ 현재 :
경상국립대학교 컴퓨터과학과
부교수

관심분야 : 증강현실, 컴퓨터비전

최 상 민 (Sang-Min Choi)



2015년 2월 : 연세대학교
컴퓨터과학과(공학박사)
2018년 3월 ~ 8월 : 연세대학교
컴퓨터과학과 박사후연구원
2018년 9월 ~ 2019년 4월 :
Fantom Foundation, Senior
Principal Researcher

2021년 7월 ~ 2022년 2월 : Soldoc CTO

2022년 3월 ~ 현재 : 경상국립대학교 컴퓨터과학과
조교수

관심분야 : 추천 시스템, 알고리즘