

A Neural Matrix Decomposition Recommender System Model based on the Multimodal Large Language Model

<https://arxiv.org/pdf/2407.08942.pdf>

0. Introduction

- 추천 시스템에서 사용자-아이템(interaction) 데이터를 효과적으로 학습하는 것이 핵심 문제임
- 기존 협업 필터링(Collaborative Filtering, CF)과 Factorization 기반 모델은 희소한 interaction matrix에서 성능 한계가 존재
- Neural Factorization 접근법을 통해 비선형 관계 학습과 고차원 특징 결합 가능성을 탐구

1. Overview

- 제안 모델은 Interaction Matrix를 Neural Network에 입력하여 latent feature를 학습
- 사용자-아이템 간 예측 점수와 실제 상호작용 간 오차 최소화
- 희소 데이터 환경에서도 추천 정확도 향상 및 feature 상호작용 포착
- 모델 구조 : embedding layer → MLP → 예측 layer

2. Challenges

- Interaction matrix 대부분이 0으로 희소하여 학습 어려움
- 기존 FM 모델은 2차 상호작용만 학습 가능, 비선형 관계 포착 제한적

- 대규모 사용자-아이템 데이터 처리 시 모델 복잡성과 학습 비용 증가
- Overfitting 위험 존재, 특히 희소 데이터

3. Method

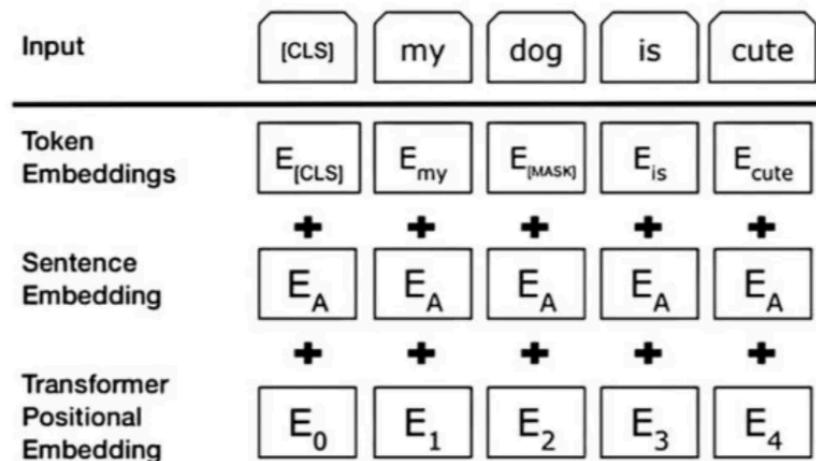


Figure 2. Structure of BERT embedding

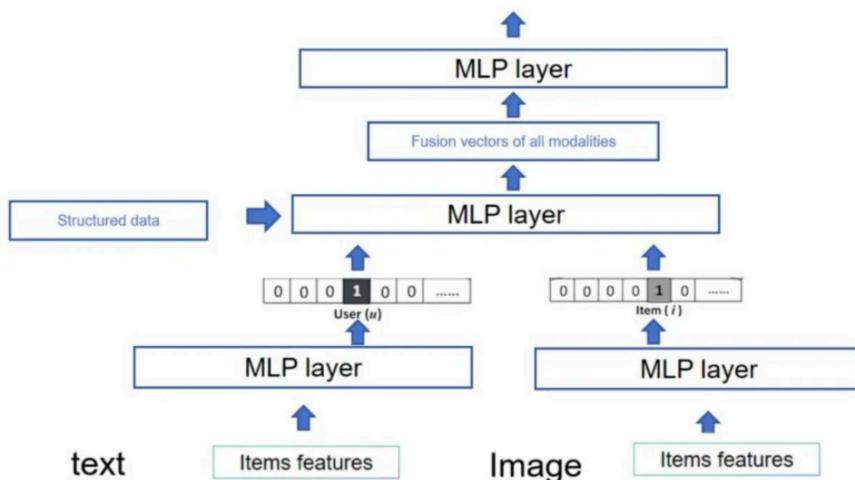


Figure 3. The structure of BoNMF model

- 모델 구조
 - 사용자/아이템 embedding → latent vectors

- Multi-Layer Perceptron(MLP)을 통해 비선형 interaction 학습
- 출력: interaction score 예측
- 학습 전략
 - 손실함수: Mean Squared Error (MSE) 또는 Binary Cross-Entropy
 - Optimizer: Adam 사용, learning rate 조정
 - Regularization: Dropout 및 L2 정규화로 overfitting 방지
- 특징 처리
 - 사용자 및 아이템의 side information optional
 - sparse matrix 효율적 처리 위해 embedding 활용

4. Experiments

- 데이터셋: MovieLens 100K, 1M 등, implicit/explicit feedback 포함
- 실험 설계
 - Train/Test split: 80/20
 - 비교 모델: FM, Wide & Deep, DeepFM
 - 평가 지표: RMSE, Precision@K, Recall@K
- 변수 설정
 - embedding dimension: 16~64
 - MLP layer: 23 layer, hidden units 64128
 - batch size: 256

5. Results

Table 1. Ablation experiment results

Model	MSE	Precision@K	NDCG
BoNMF	0.6917	0.8865	0.6911
NMD	0.8075	0.8355	0.6367
SVD	5.011	0.8465	0.6184
Tensor product-based	0.7212	0.8510	0.6559
Text Early Fusion	0.7212	0.8418	0.6559
CNN-based	0.9421	0.8210	0.6224
BoNMF -no Figure	0.7112	0.8677	0.6331

- Neural Factorization 모델, 기존 FM 대비 RMSE 5~10% 개선
- Precision@K, Recall@K 모두 기존 baseline 대비 향상
- 희소 환경에서도 latent feature 표현력 개선 확인
- Ablation study : MLP layer depth 증가 시 예측 성능 증가, embedding dimension 영향 분석
- 모델 확장성 : side information 추가 시 성능 향상 가능

6. Insight

- Neural Factorization 접근법은 기존 FM과 Neural Network의 장점을 결합한 효율적 추천 모델
- 비선형 사용자-아이템 상호작용 학습 가능, 희소 데이터 환경에서도 안정적 성능
- 실무 적용 시 embedding과 MLP 구조 조합으로 유연한 feature 설계 가능