

Study Comparing the Performance of Linear and Non-linear Models in Recommendation Systems

Da-Hun Seong[†] · Yujin Lim^{††}

ABSTRACT

Since recommendation systems play a key role in increasing the revenue of companies, various approaches and models have been studied in the past. However, this diversity also leads to a complexity in the types of recommendation systems, which makes it difficult to select a recommendation model. Therefore, this study aims to solve the difficulty of selecting an appropriate recommendation model for recommendation systems by providing a unified criterion for categorizing various recommendation models and comparing their performance in a unified environment. The experiments utilized MovieLens and Coursera datasets, and the performance of linear models(ADMM-SLIM, EASER, LightGCN) and non-linear models(Caser, BERT4Rec) were evaluated using HR@10 and NDCG@10 metrics. This study will provide researchers and practitioners with useful information for selecting the best model based on dataset characteristics and recommendation context.

Keywords : Recommendation System, Collaborative Filtering, Linear Model, Non-linear Model, Performance Evaluation

추천 시스템에서의 선형 모델과 비선형 모델의 성능 비교 연구

성 다 훈[†] · 임 유 진^{††}

요 약

추천 시스템은 기업의 매출 증가로 이어질 만큼 핵심적인 역할을 하기에 추천 시스템에 대한 연구는 과거부터 다양한 접근법과 모델들이 연구되어왔다. 그러나 이러한 다양성으로 인해 추천 시스템의 종류 또한 복잡하게 구성되고 있어 추천 모델을 선택하는 데 어려움이 따른다. 따라서 본 연구는 추천 시스템에서 적절한 추천 모델 선택의 어려움을 해결하고자, 다양한 추천 모델을 구분하는 통합적인 기준을 제공하고, 통일된 환경에서 이들의 성능을 비교 평가하였다. 실험은 MovieLens와 Coursera 데이터셋을 활용하였으며, 선형 모델(ADMM-SLIM, EASER, LightGCN)과 비선형 모델(Caser, BERT4Rec)을 HR@10과 NDCG@10 지표를 통해 성능을 평가하였다. 본 연구는 연구진과 실무자들에게 데이터셋 특성과 추천 상황에 맞는 최적의 모델을 선택하는 데 유용한 정보를 제공할 것이다.

키워드 : 추천 시스템, 협업 필터링, 선형 모델, 비선형 모델, 성능 평가

1. 서 론

추천 시스템은 사용자의 평점과 같은 명시적 데이터 또는 클릭, 구매 등의 암시적 데이터인 행동 정보를 기반으로 사용자 선호도를 파악하여 상품이나 콘텐츠를 추천해준다[1]. 기업과

다양한 서비스들은 이러한 추천 시스템을 적극 활용하고 있으며, 예를 들어, Amazon과 같은 전자 상거래 사이트는 고객의 구매 이력과 탐색 패턴을 분석하여 맞춤형 제품 추천을 제공하고, Netflix는 사용자의 시청 기록을 기반으로 영화를 추천한다. 이러한 시스템은 사용자의 데이터 분석을 통해 더 정교하고 정확한 추천을 가능하게 하여 개인 맞춤형 경험을 제공함으로써, 사용자의 만족도를 높이고 사용자와 서비스 간의 상호작용을 증대시킬뿐더러, 성공적인 마케팅 전략 수립에도 큰 도움이 되어 기업의 매출 증가까지 이어진다. 이에 따라 추천 시스템은 과거부터 여러 접근법을 가진 다양한 모델들이 제안되고 연구되어 왔다. 그러나 이러한 다양성으로 인해 추천 시스템의 종류 또한 복잡하게 구성되고 있으며 제시된 모델이 어떤 추천 시스템에 해당되는지 명시되어 있지 않은 경우가 많다.

※ 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 ICT혁신인재 4.0 사업의 연구결과로 수행되었음(IITP-2024-RS-2022-00156299).

※ 이 논문은 2024년 ACK 2024의 일반논문으로 "추천 시스템에서의 선형 모델과 딥러닝 모델의 데이터 크기에 따른 성능 비교 연구"의 제목으로 발표된 논문을 확장한 것임.

† 준 회원 : 숙명여자대학교 IT공학과 석·박통합과정

†† 종신회원 : 숙명여자대학교 인공지능공학부 교수

Manuscript Received : July 11, 2024

Accepted : July 22, 2024

*Corresponding Author : Yujin Lim(yujin91@sookmyung.ac.kr)

추천 시스템은 크게 추천 방식, 추천 분야, 기반 모델 등으로 구분할 수 있는데 각 구분에 따라라도 세부적으로 구분하는 기준이 다양하기 때문에 추천 상황에서 적합한 모델 선택에 어려움이 생긴다. 또한, 이렇게 복잡하게 구분된 추천 모델들이 독자적으로 연구되고 있어 통합된 성능 결과를 명확하게 파악하기 어려운 실정이다. 이는 연구자들이나 실무자들이 특정 상황에 적합한 추천 모델을 선택하는 데 혼란을 초래할 수 있다.

따라서 본 연구에서는 첫째, 추천 모델을 구분하는 통합적인 기준을 제공하여 추천 모델의 구분을 명확하게 하고자 한다. 이를 통해 연구자들과 실무자들이 각자의 필요에 맞는 모델을 더 쉽게 식별하고 선택할 수 있도록 돕고자 한다. 둘째, 같은 데이터셋과 성능 지표로 추천 모델을 통일된 환경에서 비교하여 추천 모델의 통합된 성능 평가를 진행하고자 한다. 이때 데이터셋의 특성 분석도 진행하여 이에 기반한 성능 결과를 분석한다. 이를 통해 다양한 추천 모델들의 성능을 객관적으로 평가하고 각 모델의 강점과 약점을 명확히 파악할 수 있을 것이다.

이 논문은 다음과 같이 구성된다. 2장에서는 추천 시스템 분야에서의 기존 연구 동향과 한계를 분석한다. 3장에서는 본 연구에서 사용된 데이터셋과 5가지의 비교 모델에 대한 실험 환경을 자세히 설명한다. 4장에서는 실험 결과를 제시하고, 각 추천 모델의 성능을 비교 분석한다. 마지막으로 5장에서는 본 연구의 결론을 요약하고, 향후 연구 방향을 제안한다.

2. 관련 연구

2.1 추천 시스템

1) 추천 방식별 구분

a) 개인화, 비개인화 추천 여부

추천 시스템에 대한 연구는 오랜 역사를 가지고 있으며, 다양한 방법론과 접근법이 개발되었다. 먼저 추천 방식별 구분 기준 중 첫 번째는 개인화 추천과 비개인화 추천이다(Fig 1). 초기에는 베스트 셀러를 추천해주는 비개인화 추천에 대해 연구되었으나, 최근에는 개인의 특성을 고려한 개인화 추천에 대한 연구가 활발히 진행되고 있다. 개인화 추천 시스템은 콘텐츠 기반 필터링(Content-based Filtering)과 협업 필터링 방식(Collaborative Filtering)을 포함하는, 정보 필터링 방식

으로 구분하는 것이 가장 일반적이다. 초기의 추천 시스템은 주로 콘텐츠 기반 필터링에 의존했는데, 이는 아이템의 자체적인 특성 및 메타데이터를 이용하여 추천을 제공하며, 새로운 아이템이 추가되더라도 해당 아이템의 특성 정보를 이용하여 즉시 추천이 가능하다는 장점이 있다. 그러나 이는 사용자가 이미 선호하는 특성과 유사한 아이템을 추천하기 때문에 추천의 다양성이 떨어질 수 있다. 이후 협업 필터링 방법론이 등장하면서 추천의 정확성과 다양성이 크게 향상되었다. 이 방식은 추천 분야와 관계없이 사용자와 아이템 간의 상호작용 데이터를 기반으로 추천을 생성하기 때문에, 특정 도메인 지식이 없어도 단순히 상호작용 데이터를 분석함으로써 추천을 제공할 수 있고, 동일한 알고리즘을 다른 도메인에서도 쉽게 적용할 수 있다. 그러나 이는 사용자나 아이템 수가 많아질수록 계산 비용이나 메모리 사용량이 급격히 증가하는 문제, 새로운 사용자나 아이템에 대한 데이터 부족할 때 발생하는 콜드 스타트 문제, 사용자들이 한정된 아이템에만 평점을 남기기 때문에 데이터의 분포가 고르지 않은 문제가 발생한다. 따라서 하이브리드 추천 시스템은 앞의 두 가지 방법론의 장점을 결합하여 보다 향상된 성능을 제공하고 있다.

본 연구에서 비교할 모델은 모두 협업 필터링 기반 모델로, 이는 다시 사용자 기반 협업 필터링(User-based CF)과 아이템 기반 협업 필터링(Item-based CF)으로 나뉜다. 사용자 기반 협업 필터링은 유사한 취향을 가진 사용자를 찾고, 그들이 좋아하는 아이템을 추천하는 방식이다. 반면에 아이템 기반 협업 필터링은 유사한 아이템을 찾고, 사용자가 좋아할 가능성이 높은 아이템을 추천하는 방식이다. 그런데 사용자 기반 방식은 많은 사용자 데이터를 처리해야 하므로, 사용자 수가 많아질수록 계산량이 급격히 증가하는 문제가 있어, 사용자 수보다 적은 아이템 기반 방식이 사용되는 경우가 일반적이다. 따라서 본 연구에서도 아이템 기반 협업 필터링 모델을 비교하고자 한다.

b) 사용자의 아이템 상호작용 순서 고려 여부

또 다른 추천 시스템 구분 기준은, 사용자가 아이템과의 과거 상호작용을 순차적으로 분석했는지 여부에 따라 구분할 수 있다. 과거에는 사용자와 아이템 상호작용의 순서를 무시하고 단순히 해당 아이템에 대한 상호작용 존재 여부만을 고려하였지만, 이러한 상호작용은 순서에 따라 인과관계가 성립할 수

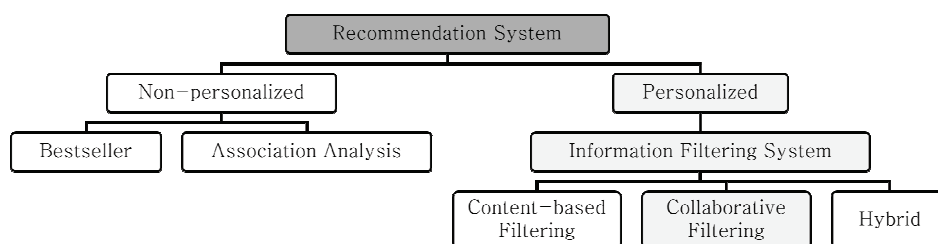


Fig. 1. The Categorization of Recommendation Systems

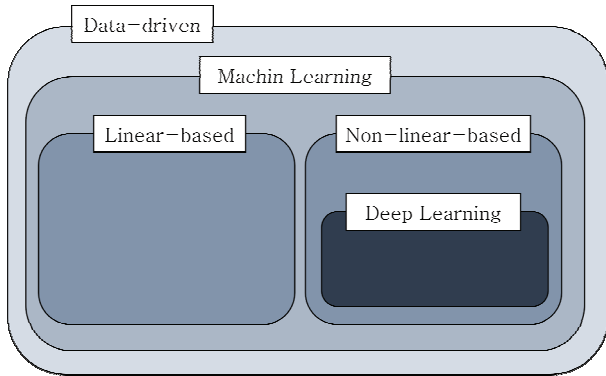


Fig. 2. The Categorization of Recommendation Systems Model

있으므로 순서를 고려하는 추천이 중요해졌다. 이렇게 순서를 고려한 추천 시스템을 순차적 추천 시스템(SRS; Sequential Recommendation System)이라고 하며, 이는 사용자와 아이템 간의 상호작용 순서와 사용자의 최근 행동을 고려하기 때문에, 사용자의 선호가 시간에 따라 변화하는 것을 인식하고 반영하여 추천의 정확도가 향상될 수 있다. 이러한 특징은 특히 뉴스나 엔터테인먼트와 같은 동적인 콘텐츠가 많은 영역에 적합하다는 장점이 있다. 그러나 단점으로는 복잡성과 콜드 스타트 문제와, 사용자의 최근 행동에만 의존할 경우 일반화 능력이 떨어질 수 있는 과적합 문제가 있다. 따라서 본 연구에서도 순차적 추천 모델을 포함하여 성능을 분석해보고자 한다.

2) 기반 모델별 구분

추천 시스템을 구현하는 방식에 있어, 사람이 일일이 규칙을 지정해주는 규칙 기반 방식(Rule-based)이 아닌, 데이터 기반(Data-driven) 방식인 기계학습(Machine Learning) 방식이 사용되며, 이러한 모델들은 크게 선형(Linear) 모델과 비선형(Non-linear) 모델로 나눌 수 있다(Fig. 2).

선형 모델은 회귀 계수를 선형 결합으로 표현할 모델이며, 비선형 모델은 회귀 계수를 선형 결합으로 표현할 수 없는 것으로 대표적으로 딥러닝(Deep Learning) 모델이 해당된다. 딥러닝 모델은 일반적으로 선형 모델에 비하여 높은 성능을 보이나, 충분한 데이터가 없는 경우 높은 차원의 임베딩을 사용하면 성능이 저하되는 문제가 발생한다. 그렇기에 딥러닝 모델과 비교하여 실험적으로 성능 차이가 있음에도, 실제 환경에서 딥러닝 모델의 대안이 될 수 있는 선형 모델 추천 시스템도 여전히 활발히 연구되고 있다. 따라서 본 연구에서는 두 모델을 모두 포함하여 실험하였다.

2.2 성능 지표

추천 시스템의 평가 방식은 평점 예측 기반 평가(Rating Prediction Evaluation)와 순위 기반 평가(Ranking-based Evaluation)로 구분할 수 있다. 평점 예측 기반 평가는 추천 시스템이 사용자가 개별 아이템에 부여할 평점을 얼마나 정확하게 예측하는지를 측정하는 평가 방법이며, 순위 기반 평가

는 사용자에게 제공한 추천의 유효성 여부를 측정한다. 본 연구에서는 추천할 아이템의 평점이 아닌, 순위 기반 추천으로 추천의 유효성을 평가하고자 한다. 순위 기반 평가에서는 추천된 아이템의 추천 순서 고려 여부에 따라 세부적으로 구분되는데, 추천된 아이템의 순서를 고려하지 않고 평가하는 성능 지표의 대표적인 방식인 HR(Hit Rate)과 순서를 고려한 NDCG(Normalized Discounted Cumulative Gain)를 기반으로 추천 모델을 평가하고자 한다. HR@K, Equation (1)은 사용자가 실제로 선택한 아이템이 K개의 아이템으로 구성된 추천 목록에 포함된 비율을 알려주며, NDCG@K, Equation (2)는 사용자가 실제로 선택한 아이템을 K개의 아이템으로 구성된 추천 목록에 얼마나 높은 순위에 두었는지를 측정하는 지표이다. 이는 Equation (3)인 IDCG(Ideal Discounted Cumulative Gain) 대비 Equation (4)인 DCG(Discounted Cumulative Gain)로, K개의 아이템을 추천할 때, 추천 순서를 포함한 가장 이상적인 추천 조합(IDCG)에 대비하여 현재 모델의 결과인 추천 목록의 순서 적합도(DCG)를 평가한다. Equation (3)과 Equation (4)에서의 rel_i 은 각 아이템과 사용자의 관련도(Relevance)를 나타내는 값으로, 평점을 관련도로 사용하였다. 두 성능 지표는 모두 0에서 1 사이의 값을 가진다.

$$Hit\ Rate@K = \frac{Number\ of\ Hit\ User}{Number\ of\ User} \tag{1}$$

$$NDCG@K = \frac{DCG}{IDCG} \tag{2}$$

$$IDCG@K = \sum_{i=1}^K \frac{rel_i^{opt}}{\log_2(i+1)} \tag{3}$$

$$DCG@K = \sum_{i=1}^K \frac{rel_i}{\log_2(i+1)} \tag{4}$$

3. 실험 환경

3.1 데이터셋

본 연구에서는 모델의 성능을 비교하기 위해, 공개되어 누구나 사용 가능한 MovieLens[2]와 Coursera[3] 데이터셋을 사용하였다. MovieLens는 추천 시스템에서 많이 활용되는 데이터셋으로 사용자의 영화 평점 데이터를, Coursera는 강의 평점 데이터를 포함하고 있다. 비교된 모델 중 순차적 추천 모델이 포함되어있으므로 사용자가 아이템을 선택한 시간 정보가 포함된 데이터셋으로 선정하였다. 사용자와 상호작용한 아이템의 개수가 10개 미만인 사용자의 데이터는 제거하고, 평점은 0.5단위의 1~5점 사이로 전처리하였다. 전처리 후 데이터셋의 통계 정보는 Table 1과 같다.

데이터셋의 희소성(Sparsity)은 사용자와 아이템 행렬에서 실제로 채워진 항목의 비율을 나타내며, 전체 아이템 수에 비

Table 1. Statistics of Datasets After Preprocessing

Datasets	# Users	# Items	# Ratings	Sparsity
① MovieLens(0.1M)	610	2,269	81,116	94.14%
② MovieLens(1M)	6,743	6,775	977,419	97.86%
③ Coursera(0.01M)	824	440	12,679	96.50%
④ Coursera(0.5M)	22,047	595	517,263	96.06%

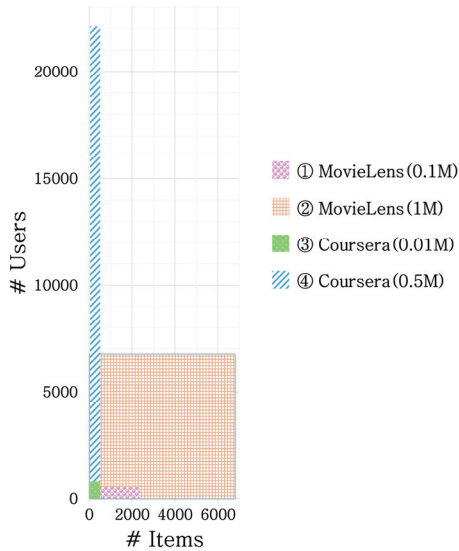


Fig. 3. Size of the User-Item in the Datasets

- 사용자 수 : ① → ③ → ② → ④
- 아이템 수 : ③ → ④ → ① → ②
- 평점 수 : ③ → ① → ④ → ②
- 희소성 : ① → ④ → ③ → ②

하여 사용자가 상호작용한 아이템 수는 매우 적기 때문에 대부분의 추천 시스템 데이터셋은 매우 희소하다는 특징을 가지고 있다. 각 데이터셋의 사용자와 아이템의 크기 형태는 Fig. 3과 같으며, 데이터셋을 각각 사용자 수(# Users), 아이템 수(# Items), 평점 수(# Ratings), 희소성(Sparsity)을 기준으로 오름차순으로 나열하면 아래와 같다. 본 연구에서는 이러한 데이터의 특성에 따라 결과가 어떻게 달라지는지 분석하고자 한다.

3.2 비교 모델

실험에 사용된 추천 모델은 아래의 5가지 모델이다. 이는 크게 선형 모델과 비선형 모델로 구분되며, 순차적 추천 모델로는 비선형 모델의 두 가지 모델이 해당된다. 5가지 모델은 모두 사용자 행동 데이터를 활용하여 높은 추천 정확도를 제공하는 협업 필터링 모델이다. 이 모델들을 통일된 데이터와 성능 지표에서 비교한 실험이 없으므로 본 연구에서는 통일된 환경에서 모델들의 성능을 측정하여 비교해보고자 한다.

1) 선형 모델

- **ADMM-SLIM[4]:** 기존 SLIM(Sparse Linear Method)[5] 모델의 최적화 함수로 ADMM(Alternating Directions Method of Multipliers)[6]을 적용하였으며, 학습 시간이 사용자 수에 독립적이고, 제약 조건과 정규화 항목이 조정 가능하다.
- **EASER[7]:** EASE(Embarrassingly Shallow Autoencoders for Sparse Data)[8]를 확장한 모델로, L1-정규화와 비음수 제약을 제거하여 구현된 얇은 오토인코더 기반 모델이다. 이는 사용자가 상호작용한 아이템 간의 관계도 고려하여 고차원의 사용자 행동 양식과 선호도를 포착한다.
- **LightGCN[9]:** GCN(Graph Convolution Network)[10] 모델에서 네트워크를 제거한 간소화 버전으로, 사용자와 아이템 상호작용을 그래프로 임베딩하며, 컴퓨팅 성능을 높이면서도 추천 성능도 높은 모델이다.

2) 비선형 모델

- **Caser[11]:** CNN 기반 순차적 추천 모델로, 수평 및 수직 컨볼루션 필터(Horizontal and Vertical Convolution Filters)를 사용하여 사용자 행동의 단기적, 동적인 순차적 패턴을 포착한다.
- **BERT4Rec[12]:** 트랜스포머(Transformer) 기반의 BERT(Bidirectional Self-attention Mechanism)[13] 모델을 추천에 적용한 순차적 추천 모델로, 사용자의 행동 시퀀스를 양방향으로 모델링하여 시퀀스 내의 아이템 간의 모든 관계를 포착한다.

4. 실험 결과

Table 1의 ①-④의 네 가지 데이터셋을 이용하여 5개의 모델을 학습시킨 후 각각 10개의 아이템을 추천하였고, 추천 결과를 성능지표 HR@10과 NDCG@10으로 평가하였다. Table 2와 Table 3에서 각 데이터셋별로 가장 높은 성능을 보인 결과는 진하게 표시하고 밑줄을 그었고, 두 번째로 높은 성능을 보인 결과는 진하게만 표시하였다. 또한, 모든 결과 중에서 가장 좋은 결과를 보인 것은 *표시를 하였다. 먼저 HR@10의 결과는 Table 2와 Fig. 4와 같다.

모델별 성능 결과를 데이터셋별로 보자면, ①-③까지 BERT4Rec 모델이 가장 높은 성능을 보였고, ④에서는 Caser 모델이 가장 높은 성능을 보였으며, 이 경우 HR 지표의 모든 결과에서도 최고 성능을 보였다. 그런데 Caser는 ③, ④에서는 높은 성능을 보였으나 이에 반해 ①, ②에서는 성능이 저조한 것을 볼 수 있다. 두 데이터 그룹의 차이는 Fig. 3에서 볼 수 있듯이, 크게 사용자의 수와 아이템 수 중 어느 것이 더 크지에 따라 구분할 수 있다. 우선 ①, ②는 사용자의 수가 아이템의 수보다 적으며, ③, ④는 사용자의 수가 아이템의 수보다 더 많다. 따라서 Caser는 사용자의 수가 아이템 수보

Table 2. Experimental Results of HR@10

Datasets	ADMM-SLIM	EASER	LightGCN	Caser	BERT4Rec
①	0.1777	0.1724	0.0864	0.1852	0.5782
②	0.2114	0.2203	0.0730	0.0474	0.8293
③	0.0149	0.0409	0.0080	0.8477	0.9102
④	0.0750	0.0779	0.0096	*0.9489	0.9104

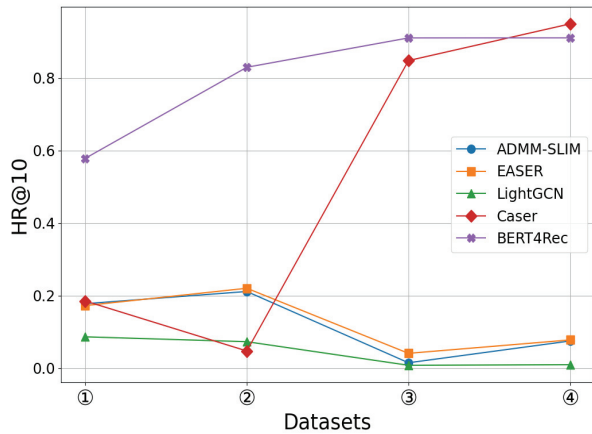


Fig. 4. Experimental results of HR@10

다 많을수록 높은 결과를 보였는데, 이는 이 모델이 다양한 사용자 데이터 처리에서 더 높은 효율성을 보인다는 것을 의미한다. 이는 수평 및 수직 컨볼루션 필터가 사용자-아이템 상호작용을 다양한 각도에서 학습할 수 있도록 해주어, 사용자가 많은 상황에서도 효과적으로 패턴을 학습하고 일반화할 수 있기 때문인 것으로 보이며, 그중에서도 Caser가 ④에서 가장 높은 성능을 보인 이유는 이 경우 아이템의 수가 사용자의 약 4배라는 특성 때문으로, Caser 모델이 이러한 필터를 통해 다양한 사용자들의 아이템에 대한 선호도와 연관성을 동시에 고려하여 더 정확한 추천을 할 수 있게 되기 때문인 것으로 보인다.

다만 ④에서 BERT4Rec과 Caser의 성능 결과 차이가 크지 않고, Caser는 특히 ①과 ②에서 BERT4Rec에 비해 성능이 확연히 낮으므로, 전반적인 성능은 HR에서 BERT4Rec이 가장 높은 결과를 보였다. Table 3과 Fig. 5는 NDCG@10으로 평가한 결과이며, 여기서도 BERT4Rec이 가장 높은 결과를 보였는데, 이렇게 BERT4Rec이 두 성능 지표에서 전반적으로 높은 성능을 보인 이유는 이 모델이 양방향 자기 주의 (self-attention) 메커니즘을 사용하여 사용자 행동의 양방향 문맥 정보를 효과적으로 학습하고, 클로즈(Cloze) 기법을 사용하여 학습 데이터의 일부를 마스킹하고 나머지 문맥을 기반으로 해당 아이템을 예측하는 방식이 모델의 성능 향상에 기여한 것으로 생각된다.

각 데이터셋에서의 HR과 NDCG의 최고 결과는 ④에서 가장 높았으며, 순위를 비교하자면, ④→③→②→① 순서이다. 이러한 결과에 <Table 1>에서 데이터셋의 4가지 특성 중 어떤

Table 3. Experimental Results of NDCG@10

Datasets	ADMM-SLIM	EASER	LightGCN	Caser	BERT4Rec
①	0.2708	0.2600	0.2923	0.0426	0.3293
②	0.3174	0.3293	0.2561	0.1952	0.5792
③	0.0316	0.0670	0.0053	0.2365	0.8594
④	0.1149	0.1185	0.0043	0.3001	*0.8730

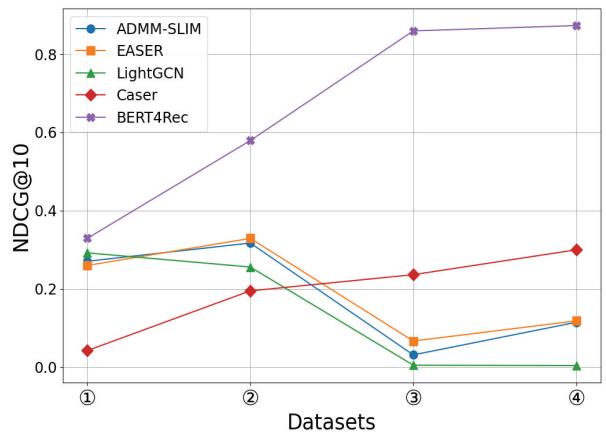


Fig. 5. Experimental Results of NDCG@10

Table 4. Correlation Coefficient Between Dataset Characteristics and Model Performance Ranking

Datasets	①	②	③	④	Correlation Coefficient
# Users	4	2	3	1	0.80
# Items	2	1	4	3	-0.60
# Ratings	3	1	4	2	0
Sparsity	4	1	2	3	0.20
Model Performance	4	3	2	1	

특성이 가장 영향을 많이 끼쳤는지 살펴보기 위해 4가지 특성의 순위를 기준으로, 데이터의 각 특성과 모델 성능 간의 상관관계를 피어슨 상관계수(Pearson Correlation Coefficient)로 분석하였으며 결과는 Table 4와 같다.

상관계수가 높은 순서대로 정리하면 사용자 수, 아이템 수, 희소성, 평점 수 순서이나, 희소성과 평점 수는 모델 성능 순위와의 상관관계가 거의 없는 것으로 나타났다. 사용자의 수는 1과 가까운 0.8의 값으로 성능과 강한 상관이 있고, 아이템은 -0.6의 값으로 모델 성능과 상관이 있는 것으로 나타나며 음의 상관관계를 갖는다. 즉, 아이템에 비하여 사용자의 데이터가 많을수록 모델 성능이 올라간다는 것을 유추할 수 있다. 이는 사용자가 아이템보다 많을수록, 추천을 위한 협업 필터링 기반 연산에서, 한 사용자당 아이템의 연관성 계산 비용이 줄어들기 때문인 것으로 생각된다.

또한, 선형 모델과 비선형 모델 각각 HR과 NDCG의 결과를

비교하면, 비선형 모델인 Caser와 BERT4Rec은 NDCG가 HR보다 전반적으로 낮게 나온 반면, 선형 모델인 ADMM-SLIM, EASER, LightGCN에서는 반대로 NDCG가 HR보다 전반적으로 더 높게 나온 것을 확인할 수 있다. 일반적으로 NDCG가 HR보다 평가에서 고려하는 요소가 많으므로 NDCG의 성능이 더 낮는데, 선형 모델의 경우 오히려 예측 결과에 순서와 정확성을 모두 고려했을 때 성능이 높게 나왔다. 이는 선형 모델이 사용자의 다양한 상호작용 패턴을 보다 일관되고 안정적으로 반영하기 때문인 것으로 보인다. 특히, 선형 모델들은 특정 상호작용을 과대평가하지 않고, 전체적인 패턴을 학습하여 순서와 관련된 정보를 효과적으로 처리할 수 있어 높은 NDCG 성능을 나타낸다.

비선형 모델만 순차 정보를 고려하고 있으므로 선형 모델과 비선형 모델을 동일 선상에서 비교하는 것은 어려우나, 결과적으로 순차 정보가 반영된 비선형 모델이 선형 모델보다 높은 성능을 보이는 것을 확인할 수 있다. 이러한 이유는 비선형 모델이 심층 네트워크를 통해 순서 정보를 포함한 복잡한 사용자-아이템 상호작용 패턴을 효과적으로 학습하고, 높은 표현력을 갖기 때문인 것으로 보인다. 그러나 선형 모델이 비교적 높은 결과를 보이는 NDCG에서는, ①, ②에서 최고 성능을 보이는 순서 정보 포함 비선형 모델과, 그 다음을 잇는 선형 모델의 성능 차이가 비교적 적은 것으로 나타나며, 특히 ①에서는 그 차이가 미미한 모습으로, 선형 모델도 순서 정보를 포함한 비선형 모델 못지않은 성능을 보여준다. 또한, 선형 모델과 순서 정보를 포함한 비선형 모델은 HR과 NDCG 결과에서 모두, ①, ②에서의 결과와 ③, ④에서의 결과가 반대로 나온 것을 확인할 수 있다. 선형 모델은 ①, ②에서의 결과가 ③, ④에서의 결과보다 높게 나왔고, 순서 정보 포함 비선형 모델은 그 반대이다. 앞서 언급했듯이 ①, ②와 ③, ④의 차이는 사용자의 수와 아이템의 수의 대소비교에 따른 차이이다. 다시 말해, ①, ②는 ③, ④와 반대로 사용자 수가 아이템 수보다 적은데, 이처럼 선형 모델이 ①, ②에서 ③, ④에서 보다 높은 성능을 보이는 이유는, 이 모델이 아이템 수가 많은 상황에서 단순하고 명확한 상호작용 패턴을 일관되도록 학습하여, 사용자 수가 많지 않아 복잡하지 않은 상호작용에서 높은 예측 정확도를 유지할 수 있기 때문이다.

선형 모델 내에서도 HR과 NDCG에서 EASER의 결과가 전반적으로 가장 높는데 이는 이 모델의 고차 상호작용(High-order Interaction)과 관련된 특성 때문인 것으로 보인다. EASER는 단순한 쌍(Pairwise) 관계뿐만 아니라 삼중(Triplet) 관계까지 학습하여 더욱 정밀한 추천을 제공할 수 있다. 이는 특히 사용자 활동이 낮은 그룹에서 큰 성능 향상을 가져오며, 높은 활동을 보이는 사용자 그룹에서도 일관된 성능을 유지하게 한다. 또한, EASER는 상호작용 데이터를 효율적으로 최적화하기 위해 ADMM을 사용하여 대규모 데이터셋에서도 빠르고 효과적인 학습을 가능하게 한다.

또한, 비선형 모델이자 순차 추천 모델인 Caser와 BERT4Rec을 각 모델의 순차적인 특성을 중점적으로 비교하자면, Caser

는 CNN을 사용하여 시퀀스 내의 연속적 패턴을 학습하는 반면, BERT4Rec은 트랜스포머 기반의 모델로, 양방향 자기주의 메커니즘을 통해 시퀀스의 양쪽 문맥 정보를 모두 활용하여 더 정교한 예측이 가능하다. 따라서 BERT4Rec은 긴 시퀀스에서도 강력한 성능을 유지하며, 특히 사용자의 복잡한 행동 패턴을 학습할 수 있기 때문에 BERT4Rec이 Caser에 비하여 더 높은 성능이 나타난 것으로 보인다.

5. 결 론

본 연구는 추천 상황에서 적절한 추천 모델 선정의 어려움을 해결하고자, 복잡한 추천 시스템의 모델을 구분하는 기준을 통합적으로 제공하였으며, 통일된 환경에서 추천 모델을 비교하여 다양한 특징을 갖는 추천 모델의 통합된 성능 평가를 진행하였다. 결과적으로 BERT4Rec이 HR@10과 NDCG@10 모두에서 다양한 데이터 세트에서 일관되게 좋은 성능을 보였으며, Caser는 사용자의 수가 아이템의 수보다 많은 데이터셋에서는 강력한 성능을 보이지만 반대의 데이터셋에서는 낮은 성능을 보여주었다. ADMM-SLIM과 EASER는 중간 정도의 성능을 보여주며, LightGCN은 일반적으로 5개 모델 중 가장 낮은 성능을 보여주었다.

또한, 본 연구는 데이터의 특성을 크기 및 상관관계 분석을 통해 면밀히 분석하였으며, 협업 필터링 추천 모델 성능에 사용자의 수가 양의 강한 상관관계를 갖고, 아이템의 수가 음의 상관관계를 갖는다는 것을 밝혀내었다. 이를 통해 연구자들과 실무자들이 다양한 추천 모델들의 성능을 객관적으로 평가하고 각 모델의 강점과 약점을 명확히 파악할 수 있을 것이며, 각자의 데이터셋과 추천 상황에서 필요에 맞는 모델을 더 쉽게 식별하고 선택할 수 있을 것이다. 추가적으로, 본 연구를 통해 선형 모델은 간단한 상호작용 패턴을 처리하는 데 유리하며, 순서 정보를 포함한 비선형 모델은 순서 정보를 포함한 복잡한 정보를 잘 학습할 수 있음을 확인하였다.

추후 연구에서는 선형 모델에서도 순서 정보를 고려한 추천을 고려할 것이며, 본 연구에서는 지도학습 기반 추천 모델만 다루었으나, 앞으로는 선형 모델과 비선형 모델의 비지도 학습 추천과 강화학습 추천 모델을 포함하여 비교 실험을 진행할 예정이다. 이러한 연구는 추천 시스템의 성능을 더욱 향상시키고, 다양한 학습 방법을 활용하여 보다 정확하고 효율적인 추천을 가능하게 할 것이다.

References

- [1] D.-H. Seong, and Y. Lim, "A study comparing the performance of Linear and deep learning models in recommender systems as a function of data size," *Proceedings of the Annual Symposium of Korea Information Processing Society Conference (KIPS) 2024*, Vol.31, pp.715-718, 2024.

[2] Dataset of MovieLens, Grouplens [Internet], <https://grouplens.org/datasets/movielens/>.

[3] Dataset of Coursera, Kaggle [Internet], <https://www.kaggle.com/datasets/imuhammad/course-reviews-on-coursera>.

[4] H. Steck, M. Dimakopoulou, N. Riabov, and T. Jebara, "ADMM SLIM: Sparse recommendations for many users," in *Proceedings of the 13th ACM International Conference on Web Search and Data Mining (WSDM '20)*, Feb. 3-7, 2020.

[5] X. Ning and G. Karypis, "SLIM: Sparse linear methods for top-N recommender systems," in *Proceedings of the IEEE International Conference on Data Mining(ICDM)*, pp.497-506, 2011.

[6] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, Vol.3, No.1, pp.1-122, 2011.

[7] H. Steck and D. Liang, "Negative interactions for improved collaborative filtering: Don't go deeper, go higher," in *Proceedings of the 15th ACM Conference on Recommender Systems (RecSys)*, pp.34-43, Sep. 27-Oct. 1, Amsterdam, Netherlands, 2021.

[8] H. Steck, "Embarrassingly shallow autoencoders for sparse data," in *Proceedings of the World Wide Web Conference (WWW)*, pp.3251-3257, May. 13-17, CA, USA, 2019.

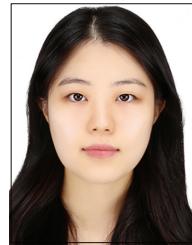
[9] X. He, K. Deng, X. Wang, Y. Li, Y. Zhang, and M. Wang, "LightGCN: Simplifying and powering graph convolution network for recommendation," in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval(SIGIR '20)*, pp. 25-30, 2020.

[10] N. T. Kipf and M. Welling, "Semi-supervised Classification with graph convolutional networks." in *Proceedings of the ArXiv Preprint*, 2016, arXiv:1609.02907.

[11] J. Tang and K. Wang, "Personalized Top-N sequential recommendation via convolutional sequence embedding," in *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining (WSDM '18)*, pp.565-573, Feb. 5-9, New York, 2018.

[12] F. Sun, J. Liu, J. Wu, C. Pei, X. Lin, W. Ou, and P. Jiang, "BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management (CIKM)*, pp. 1441-1450, Nov. 3-7, Beijing, China, 2019.

[13] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the ArXiv Preprint*, 2018, arXiv:1810.04805.



성 다 훈

<https://orcid.org/0009-0003-4733-6708>
 e-mail : ekgs324@sookmyung.ac.kr
 2023년 숙명여자대학교
 교육학부/IT공학과(학사)
 2024년 ~ 현 재 숙명여자대학교
 IT공학과 석·박통합과정

관심분야 : Artificial Intelligence, Recommendation System,
 Traffic Signal Control, EduTech



임 유 진

<https://orcid.org/0000-0002-3076-8040>
 e-mail : yujin91@sookmyung.ac.kr
 2000년 숙명여자대학교 전산학과(박사)
 2013년 일본 Tohoku University,
 Department of Information
 Sciences(박사)

2022년 ~ 2015년 수원대학교 정보미디어학과 부교수
 2016년 ~ 현 재 숙명여자대학교 인공지능공학과 교수
 관심분야 : Reinforcement Learning, Internet of Things,
 Edge Computing