

이중 스트림 융합 트랜스포머를 활용한 속성 정보 반영 순차 추천

(Sequential Recommendation with Attribute Information Using Dual-stream Fusion Transformer)

장승화*, 심춘보**, 정세훈***

(Seung-hwa Jang, Chun-bo Sim, Se-hoon Jung)

요약

오늘날 추천 시스템은 사용자 맞춤형 아이템을 제시하여 방대한 선택지 앞에서의 의사결정 문제를 완화한다. 특히 사용자 와 아이템 간 상호작용이 발생한 시간적 순서를 고려하는 순차 추천 분야에서는 트랜스포머 아키텍처를 통해 아이템 간 의존성을 학습하며, 최근에는 확산 모델을 접목한 CaDiRec이 높은 성능을 기록했다. 그러나 이러한 모델들은 주로 아 이템 ID 시퀀스를 통한 순차 정보에만 의존해 아이템의 속성 정보를 활용하지 못하는 한계를 지닌다. 본 연구는 CaDiRec 의 문맥 인식 확산 기반 구조를 기반으로, 순차 정보와 속성 정보를 함께 고려하여 추천 성능을 향상시키는 DAF-CaDiRec 을 제안한다. 제안하는 모델은 이중 스트림 융합 트랜스포머를 활용하여 아이템 ID 시퀀스와 속성 시퀀스를 독립된 경로로 처리하며, 양방향 크로스어텐션을 통해 두 스트림 간 정보 교환을 수행한다. 또한 다중의 속성값을 갖는 아이템을 처리하기 위해 합산 기반의 집계 방식을 적용한다. MovieLens-1M 데이터셋을 사용한 실험 결과, DAF-CaDiRec은 베이스라인 모델인 CaDiRec 대비 모든 평가지표에서 8~9% 향상된 추천 성능을 보였다. 이를 통해 이중 스트림 융합 트랜스포머를 활용한 순차 정보와 속성 정보의 융합이 추천 성능 향상에 유효함을 확인했다.

■ 중심어 : 순차 추천 ; 확산 모델 ; 속성 정보 ; 이중 스트림 융합 트랜스포머 ; 순차 및 속성 정보 융합

Abstract

Today, recommendation systems alleviate decision making difficulties in the face of vast options by providing personalized items. In particular, the field of sequential recommendation, which considers the temporal order of interactions between users and items, learns dependencies between items through Transformer architectures, and recently, CaDiRec, incorporating diffusion models, has recorded high performance. However, these models primarily rely on sequential information from item ID sequences, failing to fully utilize item attribute information. This study proposes DAF-CaDiRec, which improves recommendation performance by considering both sequential and attribute information based on the context-aware diffusion structure of CaDiRec. The proposed model utilizes a dual-stream fusion Transformer to process item ID and attribute sequences through independent paths and performs information exchange between the two streams via bidirectional cross-attention. Additionally, a sum-based aggregation method is applied to handle items with multiple attribute values. Experimental results using the MovieLens-1M dataset show that DAF-CaDiRec improved recommendation performance by 8-9% across all evaluation metrics compared to the baseline model, CaDiRec. These results confirmed that the fusion of sequential and attribute information using the dual-stream fusion Transformer is effective in improving recommendation performance.

■ keywords : Sequential Recommendation ; Diffusion Model ; Attribute Information ; Dual-stream Fusion Transformer ; Fusion of Sequential and Attribute Information

I. 서론

정보가 많은 오늘날에 사람들은 방대한 선택지 앞에서 의사결정의 어려움을 겪는다. 추천 시스

템은 사용자 맞춤형 아이템을 제시함으로써 이 문제를 완화한다. 일반적인 추천 방법들은 사용자 와 아이템 간 상호작용이 발생한 시간적 순서

* 준회원, 국립순천대학교 컴퓨터공학전공

** 정회원, 국립순천대학교 인공지능공학전공

*** 종신회원, 국립순천대학교 컴퓨터공학전공

이 논문은 2025학년도 국립순천대학교 글로벌대학 사업비의 지원을 받아 연구되었음.

접수일자 : 2026년 04월 13일

수정일자 : 2026년 05월 14일

게재확정일 : 2026년 05월 29일

교신저자 : 정세훈 e-mail : shjung@snu.ac.kr

를 고려하지 않는다. 그러나 사용자 선호는 시간에 따라 변화하며, 순차 추천은 상호작용의 시간적 흐름을 학습하여 사용자의 현재 관심사를 더 정확히 파악한다.

트랜스포머 아키텍처는 셀프어텐션 메커니즘을 통해 아이템 간 복잡한 의존성을 효과적으로 학습하며 순차 추천 분야에서 중추적인 구조로 자리 잡았다[1, 2]. 최근에는 이를 확장하여 확산 모델을 접목한 CaDiRec(Context-aware Diffusion-based Contrastive Learning for Sequential Recommendation)이 의미적으로 일관된 데이터 증강을 통해 높은 성능을 기록하였다[3]. 순차 추천은 사용자의 상호작용 이력, 즉 시간 순서에 따른 아이템 시퀀스를 입력으로 받는다. 이때 각 아이템은 고유 식별자인 아이템 ID와 영화의 장르, 상품의 브랜드와 같은 아이템 속성으로 구성된다. 하지만 대부분의 순차 추천 모델은 아이템 ID 시퀀스를 통해 상호작용의 흐름인 순차 정보만을 학습하며, 아이템 특징을 담은 속성 정보는 고려하지 않는다. 이러한 방식은 두 가지 측면에서 한계를 드러낸다. 상호작용 기록이 부족한 아이템은 충분한 표현을 학습하기 어렵고, 서로 다른 아이템 간의 내용적 유사성을 간과하는 것이다.

속성 정보 활용은 콘텐츠 기반 필터링에서 중요한 역할을 해왔으나, 순차 추천에서 순차 정보와 속성 정보를 효과적으로 융합하는 방법은 충분히 탐구되지 않았다[4]. 순차 정보만으로는 상호작용이 적은 아이템의 콜드 스타트 문제를 해결하기 어렵고, 속성 정보만으로는 사용자의 동적 선호 변화를 포착할 수 없다. 따라서 두 정보원이 상호 보완적으로 작용하는 융합 방법이 필요하며, 선행 연구에서도 이러한 필요성을 바탕으로 순차 정보와 속성 정보의 융합을 시도한 바 있다[5]. 본 연구는 선행 연구를 확장한 것으로, 선행 연구에서 단일 속성값만을 처리하던 한계를 보완하여 확산 기반 집계 방식을 통해 다중 속성값 처리가 가능하도록 개선하였다.

본 연구는 순차 정보와 속성 정보의 효과적 통

합을 위해 DAF-CaDiRec(Dual-stream Attribute Fusion CaDiRec)을 제시한다. DAF-CaDiRec은 CaDiRec의 확산 기반 증강과 대조학습 구조에 추가적으로 이중 스트림 융합 트랜스포머를 활용한다[6]. 이 구조는 아이템 ID 시퀀스와 속성 시퀀스를 별도 경로로 처리하면서, 셀프어텐션과 크로스어텐션을 적용하여 양방향 정보 흐름을 구현한다. 이때 영화처럼 하나의 아이템이 여러 장르를 가지는 경우, 이를 단일 속성 표현으로 집계하는 최적 방법을 함께 탐색했다. MovieLens-1M 데이터셋 실험 결과, DAF-CaDiRec은 CaDiRec 대비 전 지표에서 8~9% 향상된 추천 정확도를 보였다[7].

II. 관련 연구

순차 추천은 트랜스포머 구조의 도입으로 큰 발전을 이루었다. SASRec(Self-Attentive Sequential Recommendation)은 셀프어텐션을 활용하여 시퀀스 내 아이템 간 관계를 학습하며, BERT4Rec(Sequential Recommendation with Bidirectional Encoder Representations from Transformer)은 양방향 트랜스포머로 좌우 문맥을 모두 활용한다. 이들은 순환 신경망의 장기 의존성 문제를 해결하고 병렬 처리가 가능하다는 장점을 지닌다. 그러나 SASRec과 BERT4Rec은 모두 아이템을 ID로만 표현하여 아이템의 속성 정보를 활용하지 못하며, 이로 인해 상호작용이 적은 아이템에 대한 표현 학습이 제약되고 아이템 간 의미적 유사성을 충분히 포착하지 못한다.

트랜스포머 기반 순차 추천의 발전과 함께, 확산 모델을 활용한 접근도 제안되었다. CaDiRec은 문맥 인식 확산 모델로 데이터 증강을 수행한다. 기존의 무작위 마스킹이나 드롭아웃 기반 증강과 달리, CaDiRec은 시퀀스의 일부 위치를 마스킹하고 확산 모델로 의미적으로 유사한 아이템을 샘플링하여 여러 증강 시퀀스를 생성한다. 생성된 증강 시퀀스들은 대조학습에 활용되어

같은 원본에서 파생된 시퀀스는 임베딩 공간에서 가깝게, 다른 시퀀스는 멀리 위치하도록 학습한다. CaDiRec은 MovieLens-1M을 포함한 여러 데이터셋에서 높은 성능을 달성하였으나, 아이템을 ID로만 표현하여 속성 정보를 활용하지 못한다는 한계를 지닌다.

서로 다른 정보원을 융합하는 방법은 멀티모달 분야에서 연구되어 왔다. 이러한 융합 접근 중 CrossViT(Cross-Attention Multi-Scale Vision Transformer)는 이미지 분류에서 서로 다른 크기의 패치를 독립적인 스트림으로 처리하고 크로스어텐션으로 상호작용시킨다[8]. 각 스트림은 셀프어텐션으로 자신의 표현을 학습한 후, 크로스어텐션을 통해 다른 스트림의 정보를 참조한다. 이러한 이중 스트림 구조는 두 정보원의 독립성을 유지하면서도 양방향 정보 교환을 가능하게 한다. 본 연구는 이러한 양방향 융합 개념을 순차 추천에 적용하되, 이미지 패치가 아닌 아이템 ID 시퀀스와 속성 시퀀스를 두 스트림으로 처리하며, 셀프어텐션과 크로스어텐션을 교대로 수행하는 구조를 통해 순차 정보와 속성 정보 간의 복잡한 상호작용을 학습한다는 점에서 차별화된다.

순차 추천에서 속성 정보를 활용한 연구로 FDSA(Feature-level Deeper Self-Attention Network)가 있다[9]. FDSA는 아이템 ID 시퀀스와 속성 시퀀스에 각각 셀프어텐션을 적용한 뒤, 양쪽 출력을 연결하는 방식을 사용한다. 이는 속성 정보를 순차 추천에 활용한 연구로, 두 정보를 융합하는 방향성에서 본 연구와 유사한 동기를 공유한다. 다만 본 연구는 셀프어텐션과 크로스어텐션을 교대로 수행하는 이중 스트림 융합 트랜스포머를 통해 두 정보를 학습 과정에서 상호 참조하도록 한다는 점에서 구조적으로 차별화된다.

본 연구는 CaDiRec의 확산 기반 증강과 대조 학습 구조를 기반으로 하되, CrossViT의 양방향 융합 개념에 착안하여 순차 정보와 속성 정보를

통합하는 이중 스트림 융합 트랜스포머를 제안한다. 본 연구의 목표는 순차 정보와 속성 정보를 학습 과정에서 상호 참조하도록 함으로써, 속성 정보를 활용하지 않는 기존 순차 추천 모델의 한계와 두 정보를 독립적으로 학습하는 방식의 제약을 완화하는 데 있다.

III. 본 론

3.1. 제안하는 모델의 개요

본 연구에서 제안하는 DAF-CaDiRec은 아이템 ID 기반의 순차 정보만을 활용하던 기존의 CaDiRec을 확장하여, 순차 정보와 속성 정보를 융합 및 반영하도록 개선한 모델이다. 제안하는 모델은 이중 스트림 융합 트랜스포머를 사용하여 순차 정보와 속성 정보를 융합하는데, 이는 순차 인코더의 역할도 수행한다. 제안하는 모델은 융합 과정, 학습 과정, 예측 과정으로 구성된다.

그림 1은 DAF-CaDiRec의 융합 과정을 나타낸다. 입력 시퀀스의 아이템 ID 시퀀스와 속성 시퀀스는 임베딩 층을 거쳐 아이템 ID 시퀀스 임베딩과 속성 시퀀스 임베딩이 된다. 두 시퀀스 임베딩은 이중 스트림 융합 트랜스포머로 전달되어 융합 아이템 시퀀스 임베딩이 생성된다.

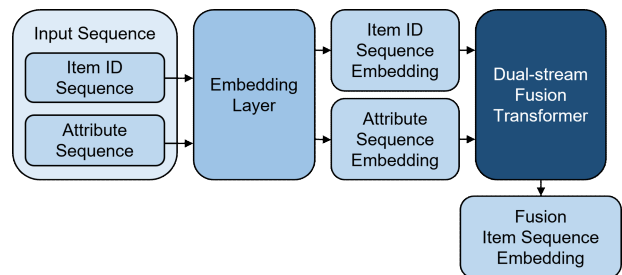


그림 1. DAF-CaDiRec의 융합 과정

그림 2는 DAF-CaDiRec의 학습 과정을 나타낸다. 학습 과정에서는 문맥 인식 확산 모델이 원본 ID 시퀀스로부터 증강 시퀀스를 생성하고, 원본과 증강 시퀀스가 각각 융합 과정을 거쳐 대조 학습에 활용된다.

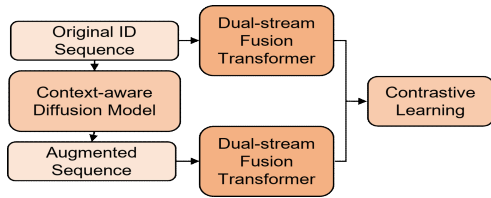


그림 2. DAF-CaDiRec의 학습 과정

그림 3은 DAF-CaDiRec의 예측 과정을 나타낸다. 입력 시퀀스가 융합 과정을 거쳐 융합 아이템 시퀀스 임베딩이 된다. 이 중 마지막 위치의 임베딩은 사용자의 최신 선호를 나타내며, 이는 다음 아이템을 예측하는 데 활용된다.



그림 3. DAF-CaDiRec의 예측 과정

3.2. 입력 시퀀스와 임베딩 층

제안하는 모델은 사용자의 상호작용 이력을 입력으로 받으며, 입력 시퀀스의 각 아이템들은 아이템 ID 시퀀스와 속성 시퀀스로 구성된다.

아이템 ID 시퀀스는 아이템 ID 임베딩 층을 거친 다음에, 시퀀스의 순서 정보를 반영하기 위해서 위치 임베딩이 더해져 아이템 ID 시퀀스 임베딩이 된다.

아이템 ID 시퀀스와 비슷한 과정을 거치며 속성 시퀀스도 속성 시퀀스 임베딩이 된다. 이때 하나의 아이템이 여러 속성값을 가지면 아이템 별로 속성 임베딩층을 거친 속성 임베딩들을 합산한다. 그 다음에 위치 임베딩을 더하여 속성 시퀀스 임베딩을 생성한다. 이를 통해 정규화 없이 모든 속성 정보를 보존하여 정보 손실을 최소화할 수 있다.

3.3. 이중 스트림 융합 트랜스포머

제안하는 모델은 이중 스트림 융합 트랜스포머를 활용하여 순차 정보와 속성 정보를 융합한다. 이는 아이템 ID 시퀀스 임베딩과 속성 시퀀스 임베딩을 입력으로 받아서, 각 정보를 ID 스트림과 속성 스트림을 별도 경로로 처리한다. 이중 스트

림 융합 트랜스포머는 2개 층으로 구성되며, 각 층에서 셀프어텐션과 크로스어텐션, 순방향 신경망이 순차적으로 수행된다. 그림 4는 이중 스트림 융합 트랜스포머를 나타낸다.

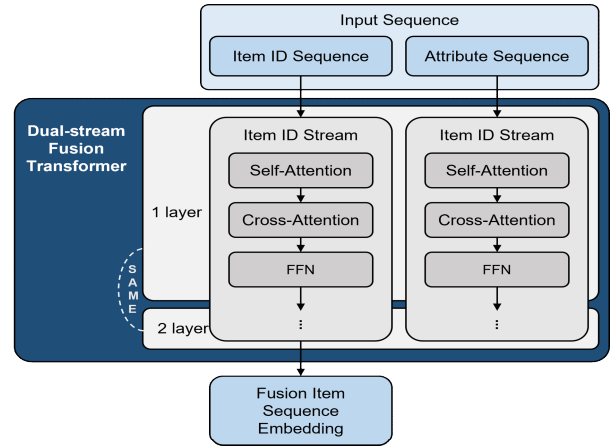


그림 4. 이중 스트림 융합 트랜스포머 (Dual-stream Fusion Transformer)

먼저 셀프어텐션 단계에서는 각 스트림이 독립적으로 동작한다. 아이템 ID 스트림은 아이템 간 순차 패턴을 학습하고, 속성 스트림은 속성 간 연관성을 학습한다. 각 스트림은 시퀀스 내 모든 위치 간 관계를 고려하여 자신의 표현을 학습한다. 이때 각 위치의 아이템은 시퀀스 내 다른 아이템들을 참조하여 자신의 표현을 생성한다.

크로스어텐션 단계에서는 두 스트림이 상대 스트림의 정보를 참조한다. 각 층의 셀프어텐션을 거친 후, ID 스트림은 자신의 출력을 Query로 설정하고 속성 스트림의 출력을 Key와 Value로 참조한다. 반대로 속성 스트림은 자신의 출력을 Query로, ID 스트림의 출력을 Key와 Value로 참조한다. 이러한 양방향 크로스어텐션을 통해 각 스트림은 다른 스트림의 특성을 통합하면서 학습한다.

이때 셀프어텐션은 각 스트림 내부의 패턴을 독립적으로 학습하고, 크로스어텐션은 두 스트림 간 정보를 교환한다. 이 두 연산을 교대로 수행함으로써 각 스트림이 자신의 표현을 충분히 학습한 후 상대 스트림의 정보를 참조할 수 있어, 단방향 크로스어텐션만 사용하는 방식보다 균형 잡힌 양방향 정보 융합이 가능하다.

순방향 신경망은 크로스어텐션의 출력에 비선형 변환을 적용하여 각 스트림의 표현력을 향상시킨다. 2개 층의 처리가 완료되면, 아이템 ID 스트림의 최종 출력을 융합 아이템 시퀀스 임베딩으로 사용한다. 이는 추천의 대상이 아이템 ID이므로, 속성 정보가 내재화된 아이템 ID 표현을 최종 입력으로 활용하는 것이 타당하기 때문이다.

3.4. 학습 과정 및 손실 함수

이중 스트림 융합 트랜스포머를 거쳐 생성된 융합 아이템 시퀀스 임베딩은 모델 학습과 예측의 입력으로 활용된다. 제안하는 모델은 문맥 인식 확산 모델을 통한 시퀀스 증강, 대조학습, 그리고 손실 함수의 최적화를 통해 학습된다.

문맥 인식 확산 모델은 아이템 ID 시퀀스 임베딩을 입력받아 증강 시퀀스를 생성한다. 시퀀스의 특정 위치들을 무작위로 선택하여 마스킹하고, 마스킹되지 않은 부분의 문맥을 활용하여 마스킹된 위치에 적합한 아이템을 생성한다. 이 과정을 반복하여 원본과 의미적으로 유사한 여러 증강 시퀀스를 만든다. 원본 시퀀스와 증강 시퀀스는 모두 동일한 이중 스트림 융합 트랜스포머를 통과하여 융합된 표현으로 변환한다.

대조학습은 동일한 원본에서 파생된 증강 시퀀스의 표현이 임베딩 공간에서 서로 가깝게 위치하도록 학습하며, 서로 다른 사용자의 시퀀스 표현은 멀어지도록 학습한다. 이를 통해 시퀀스의 본질적 특성을 유지하면서, 증강으로 인한 변화에도 일관된 표현을 학습한다.

제안하는 모델은 세 가지 손실을 결합한 손실 함수로 학습된다. 추천 손실 L_{rec} 은 BPR(Bayesian Personalized Ranking) 손실을 사용하여 양성 샘플과 음성 샘플 간 상대적 순위를 학습한다[10]. L_{rec} 은 수식 1과 같다.

$$L_{rec} = - \sum_{(u,i,j) \in D_s} \ln \sigma(\hat{x}_{uij}) \quad (1)$$

이때 D_s 는 사용자 u 가 아이템 i 를 아이템 j 보다

선호한다고 가정하는 학습 트리플 집합이며, \hat{x}_{uij} 는 아이템 i 와 j 에 대한 예측 점수의 차이이다.

대조학습 손실 L_{cl} 은 InfoNCE 손실을 기반으로 원본과 증강 시퀀스 간 일관성을 강화한다[11]. L_{cl} 은 수식 2와 같다.

$$L_{cl} = - \log \frac{A}{A+B} \quad (2)$$

$$A = \exp(\text{sim}(\tilde{h}_1^u, \tilde{h}_2^u)) \quad (2-1)$$

$$B = \sum_{\tilde{h}^- \in H^-} \exp(\text{sim}(\tilde{h}_1^u, \tilde{h}^-)) \quad (2-2)$$

이때 \tilde{h}_1^u 와 \tilde{h}_2^u 는 동일한 원본 시퀀스에서 파생된 두 증강 시퀀스의 표현이며, H^- 는 배치 내 다른 사용자의 증강 시퀀스 표현 집합이다. sim 은 내적 유사도를 나타낸다.

확산 손실 L_{diff} 은 마스킹된 위치의 복원 정확도를 NLL(Negative Log-Likelihood) 손실로 평가한다[3]. L_{diff} 은 수식 3과 같다.

$$L_{diff} = A + B - C \quad (3)$$

$$A = \sum_{t=2}^T \| \tilde{z}_0 - \tilde{f}_\theta(z_t, t) \|^2 \quad (3-1)$$

$$B = \| \tilde{e} - \tilde{f}_\theta(z_1, 1) \|^2 \quad (3-2)$$

$$C = \log p_\theta(s|z_0) \quad (3-3)$$

이때 \tilde{z}_0 와 \tilde{e} 는 마스킹된 위치에 해당하는 잠재 벡터와 임베딩이며, \tilde{f}_θ 는 역방향 과정의 신경망, $p_\theta(s|z_0)$ 는 복원된 시퀀스의 로그 우도이다.

본 연구는 CaDiRec 모델이 제안한 손실 함수 구조를 차용하며, 최종 손실 함수는 수식 4와 같다[3].

$$L = L_{rec} + \alpha \cdot L_{cl} + \beta \cdot L_{diff} \quad (4)$$

이때 α 와 β 는 각 손실의 상대적 중요도를 결정하는 하이퍼파라미터이다. 제안하는 모델은 추천 손실과 대조학습 손실 계산 시 융합 아이템 시퀀스 임베딩을 사용하며, 확산 손실은 연산 효율성을 위해 아이템 ID 시퀀스 임베딩을 사용한다.

예측 과정에서 이중 스트림 융합 트랜스포머는 순차 정보와 속성 정보를 융합하는 동시에 시퀀

스의 순차적인 패턴을 학습하는 순차 인코더 역할을 수행한다. 이때 사용자의 최신 선호를 나타내는 융합 아이템 시퀀스 임베딩의 마지막 위치 임베딩을 추출하여, 이를 전체 아이템 임베딩과 내적하고 계산된 유사도를 바탕으로 최적의 다음 아이터를 예측한다.

IV. 실험 및 결과

4.1. 실험 설정

본 연구는 제안하는 모델의 성능을 검증하기 위해 MovieLens-1M 데이터셋으로 실험을 수행했다. 이 데이터셋은 추천 연구의 표준 벤치마크로, 6,040명 사용자와 3,900개 영화에 대한 1,000,209건의 평점 기록을 포함한다. 데이터 품질 확보를 위해 상호작용 5회 미만인 사용자와 아이터를 반복 제거하는 전처리 과정을 거쳤다 [3]. 최종적으로 6,040명 사용자, 3,416개 아이터, 999,611건 상호작용이 남았다. 데이터 분할은 각 사용자의 마지막 아이터를 테스트용으로, 그 직전 아이터를 검증용으로, 나머지를 학습용으로 사용했다.

또한 본 연구는 아이터 속성 정보로 영화 장르를 활용했다. 장르 정보는 전체 아이터의 고유 장르 18개를 알파벳 순으로 정렬해 1부터 정수 ID로 인코딩했고, 아이터당 다중 장르를 최대 10개까지 처리할 수 있도록 패딩해 모델에 입력했다.

모델 성능 측정을 위해 HR@K(Hit Rate)와 NDCG@K(Normalized Discounted Cumulative Gain) 지표를 활용했다[12, 13]. HR@K는 상위 K개 추천 목록에 정답이 포함되는지를 측정하고, NDCG@K는 정답의 순위를 고려하여 추천의 상위 배치 시 더 높은 점수를 부여한다. K는 5와 10으로 설정했다. 평가는 전체 3,416개 아이터를 대상으로 순위를 매기는 전체 아이터 랭킹 방식을 사용했고, 랜덤 시드를 고정한 단일 실행에서 전체 아이터에 대해 1회 순위를 매긴 결과를 사용했다.

본 연구는 제안하는 모델의 성능과 추가된 구조의 유효성을 검증하기 위해 CaDiRec을 기반 모델이자 성능 비교를 위한 베이스라인 모델로 선정했다. CaDiRec은 SASRec, BERT4Rec 등 기존의 순차 추천 모델들을 상회하며 SOTA(State-of-the-Art)를 달성한 모델이다. 본 연구는 이처럼 검증된 최상위 성능 모델의 구조를 확장하여 순차 정보와 속성 정보 융합 모델을 활용했을 때 나타나는 추가적인 성능 향상을 분석하고자 했다. 따라서 CaDiRec 대비 성능 향상이 이루어질 경우, 이는 다른 모델들과의 비교 없이도 제안하는 구조의 우수성이 입증된 것으로 판단할 수 있는 근거가 된다.

제안하는 모델과 베이스라인 모델 간의 공정한 비교를 위해서 하이퍼파라미터는 CaDiRec의 설정을 준용했다. 이에 따라 임베딩 차원은 64, 최대 시퀀스 길이는 200, 학습률은 0.001, 배치 크기는 256, 드롭아웃은 0.3으로 설정했다. 또한 손실 함수 가중치인 α 와 β 는 CaDiRec이 MovieLens-1M에서 최적 성능을 달성한 값인 0.2로 각각 설정했다.

또한 학습 반복 횟수는 400 에포크로 설정했고, 10 에포크마다 검증 및 테스트 성능을 측정했다. 조기 종료는 적용하지 않았고, 설정된 에포크동안 학습을 완료했다.

4.2. 성능 비교 결과

표 1은 제안하는 모델 DAF-CaDiRec과 베이스라인 모델 CaDiRec의 성능 비교 결과다.

표 1. DAF-CaDiRec과 CaDiRec의 평가 지표별 성능 비교

모델명	HR@5	NDCG@5	HR@10	NDCG@10
CaDiRec	0.1357	0.0870	0.2100	0.1108
DAF-CaDiRec	0.1467	0.0953	0.2280	0.1215
변화율	8.1%	9.5%	8.6%	9.7%

DAF-CaDiRec은 모든 지표에서 CaDiRec을 상회했다. 특히 각 K 값에서 HR@K의 향상률은 8%대, NDCG@K의 향상률은 9%대였다.

NDCG@K는 순위를 반영하므로, 이는 속성 정보 융합이 정답을 추천 목록에 포함시킬 뿐 아니라 더 높은 순위에 배치하는 데 기여했음을 보여준다.

4.3. 융합 방식 비교

본 연구는 제안하는 모델을 만들기 위해서 순차 정보와 속성 정보를 결합하는 최적의 방식을 찾기 위한 비교 연구를 수행했다. 이때 모든 실험은 동일한 하이퍼파라미터에서 DAF-CaDiRec의 융합 모듈 부분만을 교체하며 수행했다.

비교 대상은 다음 여섯 가지 방식이다. 요소별 덧셈은 두 임베딩을 요소별로 더하는 방식이며, 게이트 융합은 학습된 가중치로 두 임베딩의 가중합을 수행하는 방식이다[14]. 연결은 두 임베딩을 이어붙인 후 선형 변환하는 방식이며, 투영 후 곱셈은 각 임베딩을 선형 변환한 후 요소별 곱셈을 수행하는 방식이다. 크로스어텐션은 한 쪽을 Query로, 다른 쪽을 Key와 Value로 사용하는 단방향 어텐션 방식이며, 이중 스트림 융합 트랜스포머는 3.3절에서 설명한 양방향 구조이다. 각 융합 방식은 동일한 하이퍼파라미터를 적용했고, 각 아이템 당 장르 1개만을 속성 정보로 활용했다. 표 2는 해당 비교 연구의 결과이다.

표 2. 융합 방식에 따른 추천 성능(HR@K, NDCG@K) 비교

융합 방식	HR@5	NDCG@5	HR@10	NDCG@10
요소별 덧셈	0.1303	0.0832	0.2070	0.1076
게이트 융합	0.1314	0.0860	0.2111	0.1114
연결	0.1191	0.0765	0.1896	0.0992
투영후곱셈	0.1238	0.0795	0.1996	0.1037
크로스어텐션	0.0740	0.0461	0.1268	0.0629
이중 스트림 융합 트랜스포머	0.1449	0.0936	0.2322	0.1218

실험 결과, 이중 스트림 융합 트랜스포머가 모든 평가지표에서 최고 성능을 기록했다. 해당 방식은 독립적인 정보 스트림을 유지하면서 셀프

어텐션과 크로스어텐션을 교대로 수행하며, 순차 정보와 속성 정보 간 양방향 정보 교환을 구현한다. 이러한 구조적 이점은 다른 융합 방식 대비 두 정보 사이의 복잡한 관계를 더욱 효과적으로 학습할 수 있게 한 요인으로 분석된다.

반면, 크로스어텐션은 가장 낮은 성능을 보였다. 이는 Query와 Key, Value가 고정된 단방향 구조로 인해 정보 흐름의 비대칭성이 발생하며, 두 정보의 상호보완적 결합이 제한되었기 때문으로 분석된다. 요소별 덧셈과 게이트 융합은 원본 정보를 비교적 잘 보존하여 중간 수준 성능을 보였으나, 연결과 투영 후 곱셈은 차원 축소나 비선형 변환 과정에서 발생하는 정보 손실로 인해 상대적으로 낮은 성능을 보였다.

V. 결 론

본 연구에서는 순차 정보와 속성 정보를 융합하여 추천 성능을 향상시키는 DAF-CaDiRec을 제안했다. 기존의 순차 추천 모델들이 아이템 ID 기반의 순차 정보에만 의존하여 아이템 속성을 반영하지 못하는 한계를 보완하고자, 이중 스트림 융합 트랜스포머 구조를 활용했다.

본 연구를 통해 도출한 주요 내용은 다음과 같다. 첫째, CaDiRec을 확장하여 순차 정보와 속성 정보를 동시에 고려하는 DAF-CaDiRec을 제안하고, 이를 통해 추천 성능을 향상시킬 수 있음을 확인했다. 둘째, 순차 정보와 속성 정보를 별도로 처리하며 상호 참조하는 이중 스트림 융합 구조가 단순 결합이나 단방향 참조 방식보다 우수함을 확인했다. 셋째, 하나의 아이템이 여러 장르를 갖는 경우와 같이 다중의 속성값을 처리할 때, 합산 기반의 집계 방식을 적용함으로써 정보 손실 없이 속성 정보를 임베딩 공간에 투영하는 방안을 제시했다.

본 연구는 CaDiRec의 확산 기반 구조를 기반으로, 이중 스트림 융합 트랜스포머를 통해서 아이템의 속성 정보를 고려하는 순차 추천 모델을 제안했다는 점에서 의의가 있다. 특히 순차 정보

와 속성 정보를 독립적으로 처리하면서도 양방향 크로스어텐션을 통해 상호 참조하도록 함으로써 추천 성능 향상을 입증했다.

다만 본 연구는 아이템 속성을 영화 장르라는 단일 범주로 한정하고 MovieLens-1M 데이터셋 하나에서만 실험을 진행했다는 한계가 있다. 이로 인해 속성 정보 융합이 특히 효과적인 상황과 제한적인 상황을 충분히 분석하지 못하였으며, 다른 도메인으로의 일반화 가능성 또한 검증되지 않았다. 또한 본 실험은 단일 실행 기반으로 수행되어 통계적 유의성이 충분히 검증되지 않았다. 따라서 향후 연구에서는 브랜드, 감독, 가격 등과 같이 서로 다른 성격을 가진 속성 정보를 동시에 수용할 수 있는 다중 스트림 구조로의 확장을 고려할 필요가 있다. 또한 다양한 초기화 조건에서의 반복 실험을 통한 통계적 유의성 검증을 병행한다면 보다 실용적인 연구로 발전할 수 있을 것으로 기대된다.

REFERENCES

- [1] W. Kang and J. McAuley, "Self-Attentive Sequential Recommendation," *Proceeding of 2018 IEEE International Conference on Data Mining*, pp. 197-206, Singapore, Dec. 2018.
- [2] F. Sun, J. Liu, J. Wu, C. Pei, X. Lin, W. Ou, et al., "BERT4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Transformer," *Proceeding of the 28th ACM International Conference on Information and Knowledge Management*, pp. 1441-1450, Beijing, China, Nov. 2019.
- [3] Z. Cui, H. Wu, B. He, J. Cheng, and C. Ma, "Context Matters: Enhancing Sequential Recommendation with Context-aware Diffusion-based Contrastive Learning," *Proceeding of the 33rd ACM International Conference on Information and Knowledge Management*, pp. 404-414, Boise, USA, Oct. 2024.
- [4] S. Raza, M. Rahman, S. Kamawal, A. Toroghi, A. Raval, F. Navah, and A. Kazemini, "A Comprehensive Review of Recommender Systems: Transitioning from Theory to Practice," arXiv preprint, arXiv:2407.13699v4, 2025.
- [5] S.H. Jang, C.B. Sim, and S.H. Jung, "Fusing Sequential Behavior and Item Attributes: Sequential Recommendation using Dual-Stream Fusion Transformer," *Journal of Korea Multimedia Society*, vol. 29, no. 1, pp. 140-150, Jan. 2026.
- [6] S.H. Jang, "Diffusion-based Sequential Recommendation Model with Dual-stream Fusion Transformer," *Sunchon National University Master's Thesis*, Feb. 2026.
- [7] F. M. Harper and J. A. Konstan, "The MovieLens Datasets: History and Context," *ACM Transactions on Interactive Intelligent Systems*, Vol. 5, No. 4, pp. 1-19, Dec. 2015.
- [8] C. F. R. Chen, Q. Fan, and R. Panda, "CrossViT: Cross-Attention Multi-Scale Vision Transformer for Image Classification," *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 357-366, Montreal, Canada, Oct. 2021.
- [9] T. Zhang, P. Zhao, Y. Liu, V. S. Sheng, J. Xu, et al., "Feature-level Deeper Self-Attention Network for Sequential Recommendation," *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, pp. 4320-4326, Macao, China, Aug. 2019.
- [10] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, "BPR: Bayesian Personalized Ranking from Implicit Feedback," *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, pp. 452-461, Montreal, Canada, Jun. 2009.
- [11] A. V. D. Oord, Y. Li, and O. Vinyals, "Representation Learning with Contrastive Predictive Coding," arXiv preprint, arXiv:1807.03748v2, 2019.
- [12] A. Gunawardana and G. Shani, "A Survey of Accuracy Evaluation Metrics of Recommendation Tasks," *Journal of Machine Learning Research*, Vol. 10, pp. 2935-2962, Dec. 2009.
- [13] K. Järvelin and J. Kekäläinen, "Cumulated Gain-Based Evaluation of IR Techniques," *ACM Transactions on Information Systems*, Vol. 20, No. 4, pp. 422-446, Oct. 2002.
- [14] J. Arevalo, T. Solorio, M. Montes-y-Gómez, F. A. González, "Gated Multimodal Units for Information Fusion," arXiv preprint, arXiv:1702.01992, 2017.

 저 자 소 개

**장승화(준회원)**

2017년 부산대학교 경제학부 졸업(학사).

2026년 국립순천대학교 컴퓨터공학과 졸업(공학석사).

<주관심분야 : 추천 시스템, 딥러닝, 머신러닝>

**심춘보(정회원)**

1996년 전북대학교 컴퓨터공학과 졸업(공학사).

1998년 전북대학교 컴퓨터공학과 졸업(공학석사).

2003년 전북대학교 컴퓨터공학과 졸업(공학박사).

2005년~현재 국립순천대학교 인공지능공학부 교수.

<주관심분야 : 빅데이터, 딥러닝, 생성모델, 자연어처리, 강화학습>

**정세훈(종신회원)**

2012년 국립순천대학교 멀티미디어공학과 졸업(공학석사).

2017년 국립순천대학교 멀티미디어공학과 졸업(공학박사).

2018년 영산대학교 빅데이터융합전공 조교수.

2020년 안동대학교 창의융합학부 조교수.

2022~현재 국립순천대학교 컴퓨터공학과 부교수.

<주관심분야 : 소프트웨어공학, 강화학습, 블록체인, 딥러닝, 데이터 마이닝, 빅데이터 분석 및 예측>