



BERT4Rec

BERT4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Transformer(2019)

BERT4Rec



개요

- 유저들에게 아이템을 추천하기 위해 유저들의 선호도를 모델링하는 것은 중요
- BERT4Rec 이전에는 **Left To Right** 방식으로 유저의 **interaction** 히스토리를 인코딩, 추천을 생성
- LTR 방식은 몇가지 한계가 있음

1. 유저의 행동 시퀀스에 숨겨진 표현 요소들의 영향을 제한
2. 실제로 맞지않는 순서를 지어버림

이를 해결하기위해 **BERT4Rec** 제안

BERT4Rec



개요

- 연속적인 **interaction**에 대한 아이템을 예측
예를 들어 방에 가구를 들이기 위해 침대를 구매했는데 또 침대를 추천해주는 것은 옳지 않음
- 기존 패러다임은 유저의 히스토리를 벡터로 인코딩, **LTR** 모델을 적용하여 추천
그러나 유저의 행동시퀀스를 표현하는 것이 충분하지 않음
- 시퀀스 내에 숨겨진 영향을 표현하는데 한계가 있으며, 엄격하게 순서를 따르는 시퀀스가 필요
그렇다면 **User Item Interaction** 데이터는 엄격한 순서가 정해져있는가? -> 그것은 아님
- 따라서 양방향으로 **context**를 살필 필요가 있음(**deep bidirectional model** 적용)

BERT4Rec



개요

- 양방향 추천 모델을 훈련 시키는 것은 그리 간단하지는 않음
- 전형적인 시퀀스 모델은 **LTR** 방식, 입력 시퀀스에서 각 위치에서 다음 아이템을 예측
셀프 어텐션 시 **target item**을 참고하기 때문에 낮은 성능을 보일 수 밖에 없음
- 이를 해결하기 위해 **Cloze task** 적용 -> 랜덤 마스크를 적용하여 **iid**를 예측
즉, **target item**을 마스킹하고 **information leakage**를 피하고, 더 많은 훈련 샘플을 얻을 수 있음
(데이터 증진)
- **Cloze task**는 꼭 최종 아이템 추천을 예측하는 것이 목적은 아님
테스트 시, 마지막에 **[mask]**를 추가하여 **final hidden vector**를 예측

BERT4Rec



3가지 제안

- Bidirectional Self-Attention in User behavior sequence
- Compare BERT4Rec with other SoTA models, demonstrate the effectiveness of bidirectional self-attention
- Ablation Study

BERT4Rec



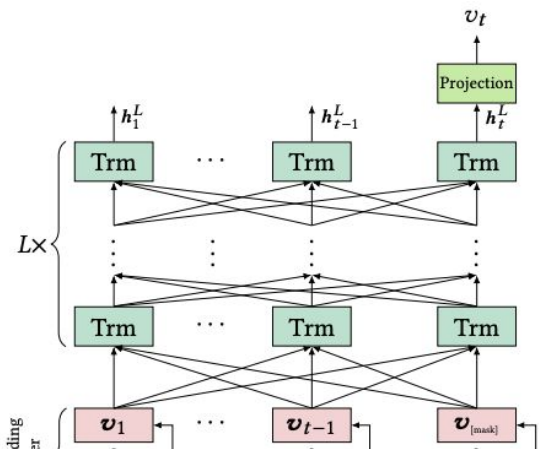
3가지 제안

- Bidirectional Self-Attention in User behavior sequence
- Compare BERT4Rec with other SoTA models, demonstrate the effectiveness of bidirectional self-attention
- Ablation Study

BERT4Rec

모델의 구조

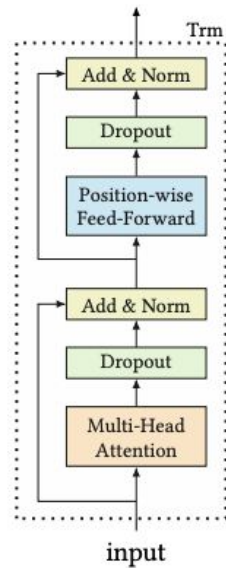
- BERT 모델에서 착안, 트랜스포머 인코더로 구성
- 이전 층으로부터 모든 포지션에 있는 표현들을 교차 연산
cf) RNN은 이전 출력이 현재의 입력으로, SAS4Rec은 이전 입력들과 현재 입력이 현재 출력을 생성
- Bidirectional 특징은 더욱 강력한 유저 행동들의 표현
- CNN과 다르게 전역적인 수용영역을 얻고, RNN과 다르게 Self-Attentive 병렬화가 가능하다는 의미: 한 번에 모든 시점 입력을 연산 가능



BERT4Rec

트랜스포머의 구조

- 입력 시퀀스 S 에 대해 그 길이가 t
- i 번째 포지션에서 l 번째 은닉층 통과, 출력 ...
- t 개의



(a) Transformer Layer.