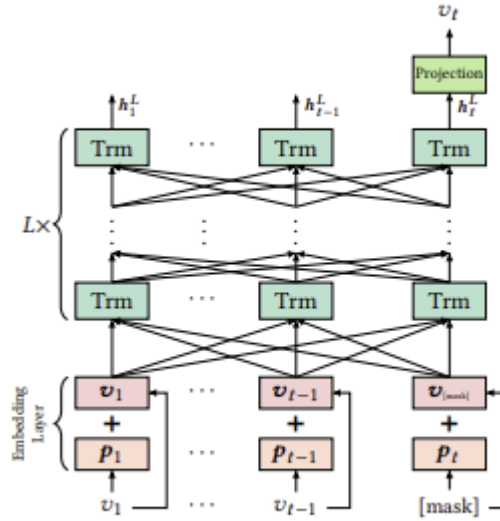


# BERT4Rec : Sequential Recommendation with Bidirectional Encoder Representations from Transformer

## Introduction

- 이전에 좋은 성능을 보였던 시퀀셜 추천시스템의 경우 유저의 이전 행동 패턴만들 고려한 단방향 추천모델들을 주로 사용했다
- 이 경우 유저가 과거에 구매했던 아이템의 정보만으로 학습하므로 성능에 제약이 존재한다
- 유저의 행동패턴을 단방향으로만 학습하지 않고 양방향으로 학습하는 모델인 BERT4Rec을 제안한다.

## BERT4Rec Model



(b) BERT4Rec model architecture.

BERT4Rec 모델은 embedding layer, transformer layer, output layer로 구성되어 있다. 모델을 양방향으로 학습하기 위해 BERT의 학습 방법처럼 유저의 행동시퀀스에 대해 [mask]토큰을 사용하여 앞 뒤 정보로부터 [mask]의 정보를 파악할 수 있도록 했다

- 모델이 양방향으로 학습하면서 target 아이템의 정보를 직접 보게되는 information leakage가 발생할 수 있기 때문이다

또한 기존의 단방향 모델은 n크기만큼의 시퀀스 길이가 있다고 가정할 때, 모델 학습 시 시퀀스마다 마지막 아이템을 맞추는 방식으로 학습샘플을 n개 구할 수 있다

반면 BERT4Rec 모델은 랜덤하게 마스킹을 처리하는 k개 역시 학습샘플로 설정하여 기존 보다 많은 샘플을 학습에 사용할 수 있다

모델의 세부 구조는 다음과 같다 :

- Embedding layer의 경우 아이템의 정보와 아이템의 위치정보를 더해 [mask]와 함께 transformer 모델의 입력으로 들어가게 된다.
- 이때 유저의 시퀀스 길이가 전체 시퀀스 길이인 하이퍼파라미터 N보다 크면 잘라내고, 작으면 0패딩을 한다
- Transformer layer의 경우 기존 transformer와 동일하게 multihead attention, point-wise feed forward를 사용하여 레이어의 수 만큼 반복연산을 수행한다.

$$H^l = \text{Trm}(H^{l-1}), \quad \forall i \in [1, \dots, L] \quad (4)$$

$$\text{Trm}(H^{l-1}) = \text{LN}\left(A^{l-1} + \text{Dropout}(\text{PFFN}(A^{l-1}))\right) \quad (5)$$

$$A^{l-1} = \text{LN}\left(H^{l-1} + \text{Dropout}(\text{MH}(H^{l-1}))\right) \quad (6)$$

- output layer는 transformer 모델로부터 받은 final output HL로부터 softmax를 통해 [mask]토큰의 확률값을 구하게된다. 이때 아이템에 대한 임베딩 매트릭스는 공유된 임베딩 매트릭스를 사용하여 모델의 사이즈를 줄일 수 있도록 했다.
- 정리하면, 모델의 입력에 유저의 시퀀스 중 p의 확률 만큼 [mask]를 수행하여 들어가게 되며, 출력으로는 [mask]된 아이템의 확률값이 나오게 된다.
- 이때 mask의 비율 p의 경우 데이터셋마다 다르지만 너무 큰 값으로 설정할 경우 오히려 성능은 악화된다.
- 모델의 loss의 경우 negative log-likelihood를 사용하여 [mask]가 반영된 유저의 행동 시퀀스가 주어졌을 때 [mask]아이템과 실제[mask]의 아이템을 비교하여 낮은 확률을 가질수록 weight를 더 많이 업데이트하는 방식으로 학습이 진행된다.

$$\mathcal{L} = \frac{1}{|S_u^m|} \sum_{v_m \in S_u^m} -\log P(v_m = v_m^* | S_u')$$

## Experiments

- 전체적인 모델의 성능 비교 결과를 보게 되면 저자가 제안한 모델이 모든 데이터셋에서 좋은 성능을 보인다.
- 성능표를 보면, 단방향으로 모델을 학습하는 것 보다 양방향으로 유저의 행동패턴을 학습하는 것이 더 좋은 추천성능을 나타내고 있음을 알 수 있다

## Conclusion

- 기존 단방향 추천모델의 한계를 극복하기 위해 유저의 행동 시퀀스에 [mask]를 반영한 양방향 학습모델인 BERT4Rec모델을 제안했다.
- NLP 분야에서 BERT는 문장들의 representation을 학습하기 위해 pre-training의 목적으로 주로 사용되며 next sentence loss, segment embeddings도 같이 사용되는 특징

이 있다

- 하지만 BERT4Rec은 유저의 행동패턴만을 바탕으로 시퀀셜 추천을 위한 end-to-end 방식의 추천모델로 차별점이 있다