



Behavior Sequence Transformer for E-commerce Recommendation in Alibaba

Abstract

본 논문에서는 제안한다. 강력한 트랜스포머 모델을 사용하는 것을 연속적인 신호를 사로잡기 위해서 유저의 행동 시퀀스에 따른 알리바바의 추천을 위한

Introduction

알리바바의 추천 시스템은 두 스테이지의 파이프 라인이다. 매치와 랭크

매치에서는 유사한 아이템들의 셋이 아이템과 유저의 상호작용에 따라서 선택된다.

파인-튜닝된 예측 모델은 주어진 후보 아이템 셋의 유저 클릭 확률을 예측하기 위해 학습된다.

본 논문은 알리바바의 C2C 플랫폼인 타오바오에서의 랭크 스테이지에 초점에 맞춰져 있다.

와이드&딥 모델의 프레임워크이 성공했음에도 만족하지 못한다. 그 이유는 모델이 실전에서 중요한 한가지 시그널을 무시하기 때문이다. 시그널의 흐름을 무시하고 있기 때문이다. 그 것의 예로 핸드폰을 사고 핸드폰 케이스를 클릭하는 것, 바지를 사고 신발을 본다는 것

와이드&딥 에서는 모든 피쳐들을 간단하게 합치기만 한다. 유저의 행동 시퀀스에 대한 주 문 정보 없이

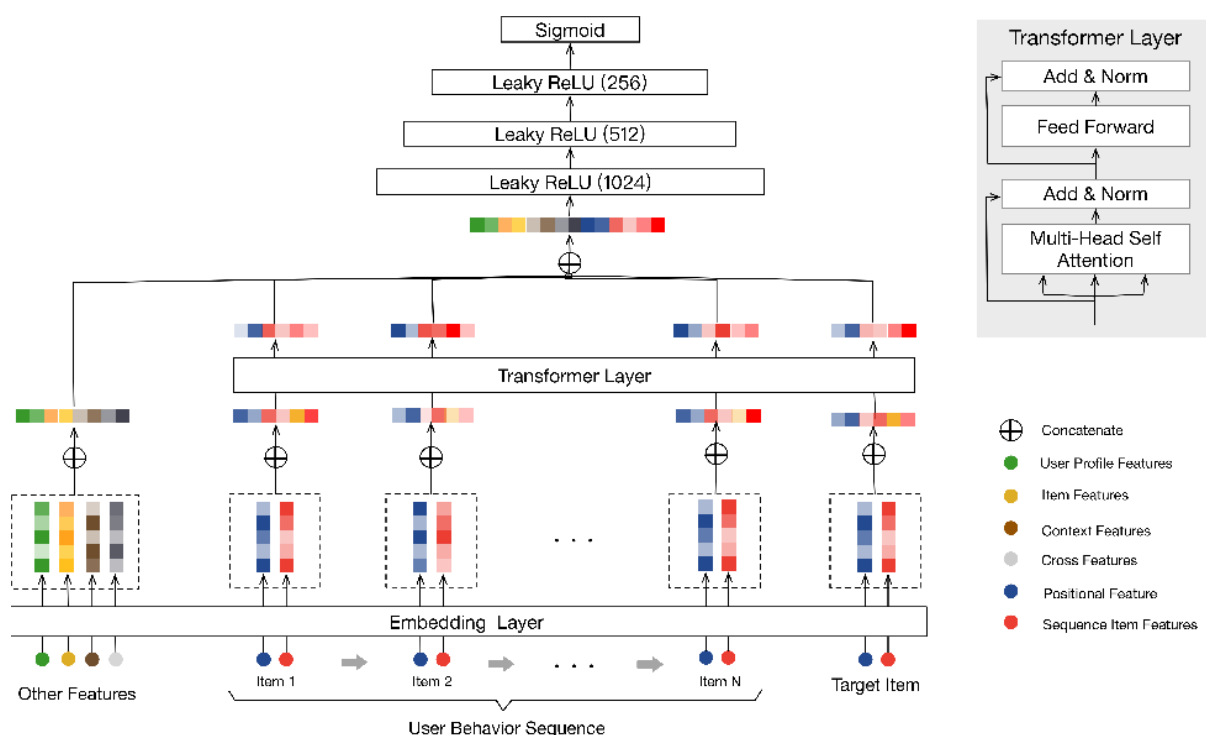
딥 인터레스트 모델 (Deep Interest Model)은 어텐션 메커니즘을 사용하는 것을 제안한 다. 후보 아이템과 과거에 유저가 클릭한 아이템 사이의 유사도를 사로잡기를 위해서 그러 나 유저의 행동 시퀀스에 기반한 연속적인 특성을 고려하지는 않았다.

셀프-어텐션 메커니즘을 적용한다. 각각의 아이템에 대한 유저의 연속적인 행동 정보를 임 베딩 표현식으로 학습하기 위해서 , 그리고 임베딩을 MLP에 피딩한다. 후보 아이템들에 대한 유저의 반응을 예측하기 위해서

Architecture

CTR을 예측 문제로 추천 과제를 모델링을 한다.

아이템 클릭 이력 데이터를 가지고 목표는 함수 f 를 학습. f 는 타겟 아이템 v 에 대한 유저 u 의 클릭 확률을 예측한다.



우선 입력 피쳐들을 낮은-차원 벡터들로 임베딩을 한다. 행동 시퀀스 내에서 아이템들 간 의 관계를 잘 잡아내기 위해, 시퀀스 내의 아이템 각각의 더 깊은 표현을 학습하는 트랜스 포머 레이어가 사용된다. 그리고 다른 피쳐들과 트랜스포머 레이어 출력을 합치고, 세 레 이어인 MLP들이 히든 피쳐 간의 상호작용 학습에 사용된다. 그리고 시그모이드 함수들이 마지막 출력을 생성에 사용된다. "위치적 특징 들"이 "시퀀스 아이템 피쳐들"로 합쳐진다.

와이드&딥 모델과 BST의 차이는 트랜스포머를 더한 점이다. 유저들이 클릭한 아이템들에 대한 표현을 잘 학습하기 위해서

Embedding Layer

아이템 들 각각에 대해 시퀀스 아이템 피처와 위치 피처 들을 합친다.

포지셔널 피처==포지셔널 임베딩

낮은 차원의 벡터로 투영되기 전에 "포지션"을 아이템 각각의 입력 피처로 더한다. 특정 아이템의 "포지션"은 특정 아이템의 추천 시간과 유저가 그 아이템을 클릭한 시간대를 뺀다.

Transformer Layer

아이템 각각에 대한 더 깊은 표현을 학습한다. 행동 시퀀스 내에서의 다른 아이템들 간의 관계를 사로 잡아서

셀프-어텐션은 아이템들의 임베딩을 입력으로 취하고, linear projection으로 세 개의 메트릭으로 변환한다. 변환한 것을 어텐션 레이어로 피딩한다. 세 개의 메트릭은 멀티 헤드 어텐션에 붙는다.

Point-wise Feed-Forward Networks을 모델을 비선형성을 위해 추가한다.

오버피팅을 피하고 계층적으로 의미있는 피처들을 학습을 위해, Drop-out과 LeakyReLU를 셀프-어텐션과 FFN에 사용한다.

셀프-어텐션 블록을 스택킹 한다. 첫번째 셀프-어텐션 블록 다음 이전의 모든 아이템 임베딩을 집계하여 아이템 시퀀스를 기반으로 complex relations를 모델링 한다.

MLP Layers and Loss function

다른 피처들의 임베딩과 트랜스포머 레이어 출력을 결합해 타깃 아이템에 적용함으로써, 세 Fully-Connected 레이어 들이 dense 피처 간의 상호작용을 학습하는 것은 추천 시스템의 표준이다. 특정 아이템의 클릭을 예측하기 위해서 이진 분류 문제로 모델링한다. 그래서 출력 유닛으로 시그모이드를 사용한다. 모델을 학습하기 위해 크로스-엔트로피 로스를 사용한다.

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/9e30fb-d1-a721-47a7-8382-e5dd63f0d5f8/transformers_recommendation.pdf


[https://s3-us-west-2.amazonaws.com/secure.notion-static.com/ff0f65f4-9ee7-42a4-9d86-c7498573de7c/Transformer_\(1\).pdf](https://s3-us-west-2.amazonaws.com/secure.notion-static.com/ff0f65f4-9ee7-42a4-9d86-c7498573de7c/Transformer_(1).pdf)



Seq2Seq + Attention

How Does the Recommendation System Work on Tmall?

Kaiwai September 9, 2019 14,216 By Chen Qiwei. As soon as you open the Tmall app, it starts to display content that is specifically designed for your online shopping experience based on the

 https://www.alibabacloud.com/blog/how-does-the-recommendation-system-work-on-tmall_595335




Attention? Attention!

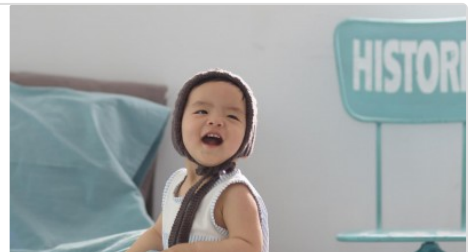
이 글은 lilianweng의 Attention? Attention! 포스트를 번역한 글입니다. Attention은 최근 딥러닝 커뮤니티에서 자주 언급되는 유용한 토픽입니다. 이 포스트에서는 어떻게 어텐션 개념과 다양한 어텐션 메커니즘을 설명하고 transformer와 SNAIL과 같은 모델들에 대해서 알아보려고 합니다. Attention은 우리가 이미지에서 어떤 영역을 주

<https://yjjucho1.github.io/attention/attention/>

dsindex/blog

personal blog posts. Contribute to dsindex/blog development by creating an account on GitHub.

 <https://github.com/dsindex/blog/wiki/%5Battention%5D-NLP-with-attention>



자동등록방지를 위해 보안절차를 거치고 있습니다.

<http://solarisailab.com/archives/2162>