

MovieLens-1M, Avazu4, Criteo 데이터셋에 기반한 클릭 률 예측을 위한 어텐션 네트워크

안자건¹, 조인휘²

¹한양대학교 컴퓨터소프트웨어학과 석사과정

²한양대학교 컴퓨터소프트웨어학과 교수

azj535041051@hanyang.ac.kr, iwjoe@hanyang.ac.kr

Attention Network For Click-through Rate Prediction Based On MovieLens-1M, Avazu4, Criteo Datasets

Zijian An¹, Inwhee Joe²,

¹Dept. of Computer Science, Hanyang University

²Dept. of Computer Science, Hanyang University

요 약

CTR(Click Through Rate) 예측은 사용자가 광고나 아이템을 클릭할 확률을 예측하는 데 사용되는 용어로, 광고 분야에서 중요한 연구 분야로 자리 잡았다. 인터넷 데이터의 양이 증가함에 따라, 전통적인 피쳐 엔지니어링의 인건비는 계속해서 상승하고 있다. 특징 상호 작용에 대한 의존도를 줄이기 위해, 본 논문은 TMH(Two-Tower Multi-Headed Attention Neural Network) 접근법이라고 하는 명시적인 특징 상호 작용과 암시적인 특징 상호 작용을 결합한 융합 모델을 제안한다. CTR 예측에서 TMH의 효과를 평가하기 위해 3개의 실제 데이터 세트를 사용하여 많은 수의 실험을 수행하였다. 성능은 3개의 데이터 세트에서 0.12%, 0.41% 및 0.68%로 향상되었다.

1. 서론

수년 전의 협업 필터링[1]부터 오늘날의 머신러닝 및 딥러닝 기반 CTR 모델에 이르기까지 특징 상호 작용이 CTR 예측 연구에서 중심적인 역할을 한다. 이미지 및 오디오에서 볼 수 있는 연속적인 기능과는 달리, 실제 기능은 대부분 다중 도메인 카테고리 기능이다. 추천 시스템에서 좋은 품질의 특징을 얻기 위해서는 데이터 과학자들이 데이터의 기본 정보를 탐색하고 의미 있는 특징 상호작용을 추출하는 데 많은 시간과 비용[2]을 들여야 한다. 그리고 원시 특징의 수가 많아 모든 특징 상호작용을 수동으로 추출하는 것이 불가능하다. 따라서 상호작용 특징의 자동 학습은 중요한 과제이다.

본 논문에서는 투타워 멀티-헤드 어텐션 신경망을 기반으로 한 심층 융합 모델 TMH를 제안하여 고차 명시적 및 암시적 특징 상호 작용의 자동 학습을 융합하여 특징 상호 작용이 벡터 및 비트 수준 모두에서 발생하도록 한다. 우리가 제안한 모델은 복잡한 특징 상호 작용을 명시적이고 암시적이며 유연한 방식으로 모델링할 수 있으며 또한 좋은 기억력과 일반화를 제공한다.

2. 본론

A. 방법

TMH 방법은 사용자 타워와 아이템 타워 두 부분으로 진행된다. 먼저 사용자와 아이템의 수치적, 범주적 특성을 모델의 입력으로 사용한 후 사용자와 아이템 타워에 입력한다. 임베딩 레이어에서 범주형 변수는 범주형 및 수치형 특징 간의 상호 작용을 허용하고 이를 밀도 높은 벡터로 변환하기 위해 지정된 차원의 밀도 높은 수치 표현에 매핑된다. 특징 상호 작용 계층으로 멀티-헤드 어텐션 신경망이 구현된다. 각 상호작용 계층에서, 더 높은 차수의 특징들은 주의 메커니즘을 통해 결합될 수 있거나, 또는 멀티-헤드 어텐션 메커니즘을 통해 상이한 유형들의 결합된 특징들에 집중함으로써 특징들이 각 서브공간에 매핑될 수 있다. 그리고 여러 상호작용층의 중첩에 의해, 서로 다른 순서로 특징들의 조합을 모델링하는 것이 가능할 것이다. 다음으로, 특징 벡터는 상호작용 계층에서 사용자 타워와 아이템 타워에서 동일한 차원을 유지하고, 특징 정보는 고차 암묵적 상호작용을 위해 심층 신경망을 통해 학습된다. 마지막으로, 스칼라값에 의해 클릭 스루율이 추가로 추정된다.

우리는 MovieLens-1M, Avazu4, Criteo의 세 가지 데이터 세트를 사용하여 실험을 수행했습니다. 구체적인 매개변수는 표 I에 나와 있습니다.

표 I. 데이터셋 통계

Datasets	Sample	Fields	Features
MovieLens-1M	739012	7	3529
Avazu4	40428967	23	1544499
Criteo	45840617	39	998960

B. 모델훈련 및 평가

우리는 방법을 구현하기 위해 TensorFlow 를 사용한다. 최적의 하이퍼파라미터를 찾기 위해 그리드 검색 전략이 사용된다. 우리의 방법은 임베딩 벡터 차원을 12 벡터 차원으로, 배치 크기를 1024 로 설정하고, 어텐션 헤드의 수를 12 로 설정한다.

표 II 에는 다양한 방법의 성능을 요약하고 다음과 같은 결론을 도출합니다. NFM 은 FM 모델보다 우수하며, 이는 2 차 숨겨진 벡터 대신 FM 에서 신경망의 결합에 기인할 수 있으며, 이는 기능적 상호작용의 암묵적 학습 형태로서의 성능을 향상시키고 DNN 에 결합하는 효과를 입증한다. CIN 은 수동 기능 엔지니어링 없이 명시적 및 암묵적 고급 기능 교집합을 공동으로 훈련할 수 있어 모델 결과를 크게 향상시킬 수 있다. AutoInt 는 트랜스포머에서 멀티-헤드 어텐션 메커니즘을 갖는, 특징들 사이의 고차 명시적 상호작용을 달성하기 위해 트랜스포머를 도입한다.

표 II. 방법별 성능 비교

Model	Avazu	MovieLens-1M	Criteo
	Auc/Logloss	Auc/Logloss	Auc/Logloss
NFM	0.7708/0.3864	0.8357/0.3883	0.7957/0.4562
CIN	0.7758/0.3892	0.8286/0.4108	0.8009/0.4517
AutoInt	0.7752/0.3824	0.8415/0.3797	0.8061/0.4455
TMH	0.7764/0.3817	0.8456/0.3297	0.8069/0.4450

MovieLens-1M 데이터셋에 대한 결과는 그림 I 과 같다. 더 큰 차원 벡터가 더 많은 정보를 나타낼 수 있기 때문에, 우리는 차원이 증가함에 따라 성능이 증가함을 알 수 있다. 성능은 벡터 크기가 각각 16 와 24 에 이를 때 최적이다. 16 및 24 를 넘어서, 이 차원의 상태들은 이미 너무 많은 파라미터들이 생성될 경우 모델이 과적합되어 정확도가 떨어지고 로그 손실이 증가할 것이라는 충분한 정보를 포함하고 있기 때문에 성능이 감소하기 시작한다.

그림 I. MovieLens-1M 데이터를 사용하여 측정된 AUC 및 LogLoss.

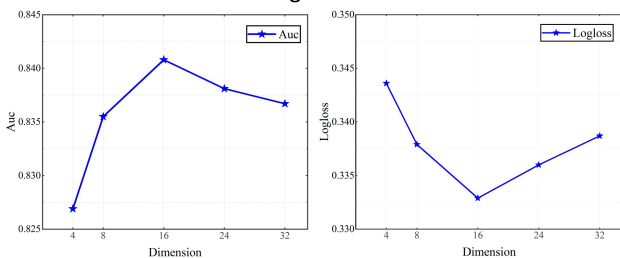


그림 II 에는 실험을 통해 고차 암묵적 상호 작용 구성요소를 제거하면 모델의 성능이 감소한다는 결론을 도출하였다. 더욱이 대형 Avazu4 및 Criteo 데이터 세

트에서 숨겨진 특징 상호 작용을 포함한 것은 상당한 성능 향상을 보였다. 이는 암묵적 특징 상호 작용이 제안된 모델의 예측력을 향상시키고 대규모 데이터 세트에서 우수한 성능을 발휘함을 시사한다.

그림 II. 고차 암묵적 상호 작용이 모델 성능에 영향을 미치는지 여부를 테스트

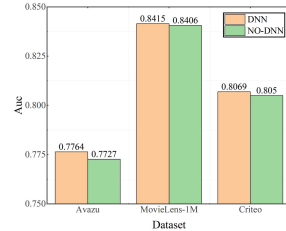
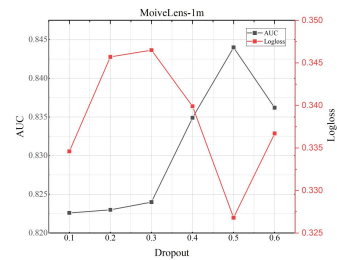


그림 III MovieLens-1m 데이터셋 탈락률



탈락률의 주된 기능은 훈련 데이터에 대한 과도한 의존을 방지하고 파라미터 일반화를 향상시키는 것이다. 탈락률이 작으면 과적합이 효과적으로 감소하지 않는 반면, 높은 값은 중요한 정보의 손실을 초래하여 권장사항에 영향을 미칠 수 있다. 그림 III. 에서 보는 바와 같이 MovieLens-1m 데이터 세트에서 모델의 AUC 와 손실 값이 증가하다가 감소하는 추세를 보인다. 우리는 MovieLens-1m 데이터 세트에서 0.5 의 탈락률을 적용하였다.

3. 결론

이 연구에서 모델은 고차 명시적 상호 작용과 고차 암묵적 상호 작용을 통해 기능을 자동으로 학습한다. 우리는 세 가지 실험 데이터 세트에서 실험을 수행했으며 결과는 제안된 방법이 효과적이고 Logloss 및 AUC 점수에서 좋은 결과를 나타냄을 분명히 보여준다. 후속 작업에서 우리의 방법은 다른 데이터에서 테스트될 것이다. 우리는 CTR 예측 수준을 높일 것이다.

참고문헌

[1] Su X, Khoshgoftaar T M. A survey of collaborative filtering techniques. Advances in artificial intelligence, 2009.

[2] He X, Zhang H, Kan M Y, et al. Fast matrix factorization for online recommendation with implicit feedback. Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval. 2016: 549-558.