

Negative Sampling과 Deep Interest Network를 활용한 오프라인 옥외광고 추천시스템 개발

서현우¹, 김수혁¹, 유상기², 조승규², 조수필², 손종수², 임치현¹

¹UNIST 산업공학과, ²CJ 올리브네트웍스 AI Core 연구소

{ta57xr, sooo, chlim}@unist.ac.kr,

{sanggi.ryu1, seungkyu.jo, suphil.cho, jongsoo.sohn}@cj.net

초록

옥외광고 시장은 영업사원의 노하우에 의해 폐쇄적으로 이루어져온 분야로, 과학적인 의사결정이 어렵고 광고주에게 여러 옵션을 체계적으로 제공하기 어렵다. 본 연구는 과거 옥외광고 이력 데이터를 분석해 옥외광고 추천시스템을 개발한 사례를 소개한다. 옥외광고 할당 문제는 학습 데이터가 암시적 피드백인 동시에, 오프라인 광고판 하나에 하나의 광고만 게시 가능하다는 특수성을 가지고 있다. 이에 본 연구는 negative sampling과 Deep Interest Network를 활용하여 옥외광고 이력 데이터에 적합한 추천시스템을 제안한다. 제안한 추천시스템은 벤치마크 대상들보다 우월한 성능을 보였으며, 본 연구는 유사한 추천 문제의 해결을 위한 시사점을 제시한다.

1. 서론

2019 광고산업조사에 따르면 작년 대비 디지털 옥외광고 시장 규모는 62.3%가 성장하여 전체 옥외광고 시장의 규모가 확대되었다 [1]. 하지만 광고주들은 여전히 영업사원의 노하우에 의해서 옥외광고 매체를 제안받고 있다. 각 영업사원 별 상이한 영업 방식으로 인해 옥외광고 시장은 과학적이고 체계적인 옵션들을 제시하지 못해왔다. 광고주는 할당된 매체에 계약 기간 동안 광고를 수행해야 하기 때문에 광고주에 적합하지 않은 매체는 많은 손실을 야기할 수 있다. 즉, 폐쇄적인 옥외광고 산업은 광고주에게 최적의 광고 옵션을 고려할 기회를 충분히 제시하지 못했다. 이에 객관적이고 과학적인 광고 집행을 위해 데이터 기반의 옥외광고 추천시스템을 구축하는 것이 필요하다 (그림 1).

한편, 옥외광고 이력 데이터는 영화의 평점 등 명시적인 피드백과 달리 사용자의 구체적인 선호를 반영하기 어렵다 [2]. 그렇기 때문에 옥외광고 추천시스템은 과거 광고주의

광고 이력을 통해 선호를 추론할 필요가 있다. 하지만 옥외광고 할당 문제는 한 광고판에 한 광고만 게시 가능하다는 ‘배타적 점유’의 특수성을 가지고 있기 때문에 광고주들은 점유된 광고판에는 광고하기 어렵다. 즉 광고주가 사용하지 않은 매체에는 선호하나 ‘배타적 점유’에 의해 옵션으로 제공받지 못한 매체와 광고주가 비선호하는 매체가 혼재된다. 이에 광고주가 광고한 기록이 없는 매체 중에 광고주가 비선호할 매체를 판별하는 것이 필요하다. 또한, 옥외광고 매체는 광고주 산업군 범주 정보, 매체 지역 범주 정보 등 여러 메타 정보들을 고려하여 제안되어야 한다.

본 연구에서는 과거 옥외광고 이력 데이터를 분석해 옥외광고 산업에 적합한 추천시스템을 개발하였다. 본 연구는 딥러닝 기반 Click-through rate (CTR) 예측 모델 중 Deep Interest Network (DIN) [3]를 활용하였다. 또한, 옥외광고의 ‘배타적 점유’ 특수성과 ‘같은 산업군의 광고주들은 비슷한 매체를 선호한다’는 협업 필터링 아이디어를 반영하여 옥외광고 추천에 적합한 negative sampling 방법을 개발하여 모델을 학습하였다. DIN은 광고주의 메타 정보, 옥외광고 사용 이력과 후보 옥외광고 매체의 연관성을 고려하여 해당 옥외광고 매체의 선호도를 예측한다. 그래서 과거 옥외광고 이력을 반영하고 광고주와 매체의 특성을 고려하여 최적의 옥외광고 매체를 추천할 수 있다. 또한, 추천 옥외광고 매체와 과거 옥외광고 이력들의 연관성을 수치화하여 광고주에게 제공함으로써 신뢰성 있는 추천을 할 수 있다. negative sampling은 광고 기록이 없는 매체 중에 ‘배타적 점유’에 의해 제공되지 않은 매체를 효과적으로 제외하여 광고주가 비선호하는 매체를 판별한다. negative sampling으로 학습된 DIN은 옥외광고 집행 특성을 반영하게 된다.

본 연구가 제안하는 negative sampling과 DIN 기반의 오프라인 광고 추천시스템은 실제 옥외광고 이력 데이터를 기반으로 검증되었다. 벤치마크 모델들과 비교 실험을 수행한 결과, 제안된 시스템은 기존 모델보다 우월한

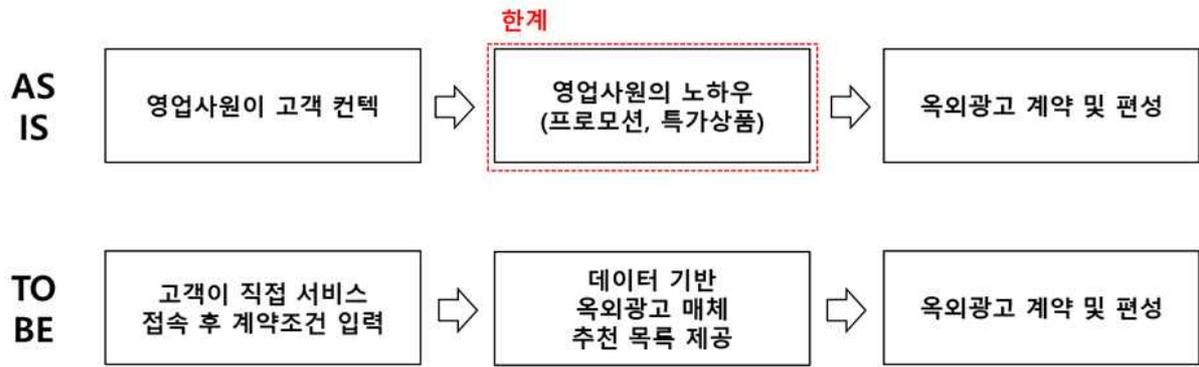


그림 1. 옥외광고 산업 프로세스

성능을 보였고 DIN의 지역적 활성화 네트워크로 과거 옥외광고 기록에 대한 연관성 정보를 제시하여 추천시스템의 설명력과 신뢰성을 높일 수 있음을 검증하였다. 또한, 광고한 기록이 없는 신규 광고주에 대한 추론에도 높은 성능을 보여 콜드 스타트 (Cold start) 문제에도 강건함을 보였다. 옥외광고 추천 문제에 맞추어 개발된 negative sampling 방법은 기존 negative sampling 방법보다 DIN 및 벤치마크 모델 성능의 향상을 이끌어 옥외광고 추천 문제에 적합함을 실험적으로 검증하였다. 본 연구는 오프라인 옥외광고 할당을 위한 추천시스템을 개발한 첫 연구로서, 본 연구는 유사한 오프라인 추천 문제 해결을 위한 시사점을 학계에 제공할 것이다.

본 논문은 다음과 같은 구조로 서술된다. 섹션 2에서 오프라인 추천과 관련된 연구와 추천 모델 학습을 위한 negative sampling과 관련된 연구를 분석한다. 섹션 3에서 옥외광고 매체 추천을 위한 추천시스템을 설명하고, 섹션 4에서 실제 데이터 기반 비교 실험 결과를 제시한다. 섹션 5에서 논의 및 결론을 제시한다.

2. 관련 연구

2.1. 오프라인 상품 추천시스템

오프라인 상품 추천시스템은 사용자의 구매 이력과 함께 오프라인 환경의 물리적 특성을 추가로 고려한다. 김남기 외 (2017) [4]은 오프라인 쇼핑몰에서 고객의 브랜드 구매 이력 데이터를 통해 고객에게 오프라인 브랜드를 추천하는 방법을 제안했다. 구매 이력을 통해 실내 위치, 이동 동선, 체류 시간 등을 추론했고 협업 필터링을 응용하여 고객에게 브랜드를 추천했다. 김남기 외 (2016) [5]은 오프라인 쇼핑몰에서 고객의 동적 컨텍스트 정보를 활용하여 고객에게 적합한 상품을 추천한

다. 최영환 외 (2006) [6]은 개인화된 오프라인 상품을 추천하기 위해 사용자의 이동 패턴을 분석하였다. 개인 사용자의 구매 이력을 바탕으로 아이템을 추천하는 오프라인 상품 추천시스템 연구는 많은 반면, 옥외광고 매체 추천을 위한 연구는 부족하다. 옥외광고 시장은 기업 사용자에게 추천되기 때문에, 기업 사용자에 대한 다양한 메타 정보들을 고려해야 한다. 또한, 옥외광고 매체는 지역 정보, 유동인구 등 고려해야 할 오프라인 환경의 물리적 특성이 다양하다.

2.2. 구매 데이터 기반 추천시스템 학습을 위한 negative sampling

오프라인 구매 데이터 기반 추천시스템은 구매 데이터에 부정적 피드백이 없다는 한계 때문에 학습이 어렵다 [7]. 구매 데이터 등 암시적 피드백에서 비선호 아이템을 판별하기 위해 negative sampling 방법들이 개발되어 왔다. 일반적인 negative sampling 방법은 구매한 이력이 없는 아이템으로부터 균등 분포를 따라 사용자가 비선호하는 아이템을 추출하는 것이다 [8, 9, 10]. He 외 (2016) [11]은 전체 구매 빈도 기반의 확률 분포를 따라 비선호 아이템을 추출하였다. Zhang 외 (2013) [12]은 협업 필터링의 추천 목록으로부터 추정된 비선호 아이템을 다시 학습에 활용하여 협업 필터링 모델을 학습하였다. 또한, Ding 외 (2019) [13]와 Wang 외 (2020) [14]은 사용자 아이템 노출 데이터, 아이템 특성 정보 그래프 등을 활용하여 추천시스템 학습을 위한 negative sampler 모델을 같이 학습하였다. 하지만 오프라인 옥외광고 이력 데이터는 긴 광고 주기로 인해 데이터가 적어 현실적으로 추천시스템과 sampler 모델을 같이 학습하기 어렵다. 또한, 옥외광고 시장의 특성인 ‘배타적 점유’와 광고주 간의 산업군 관계를 반영하여

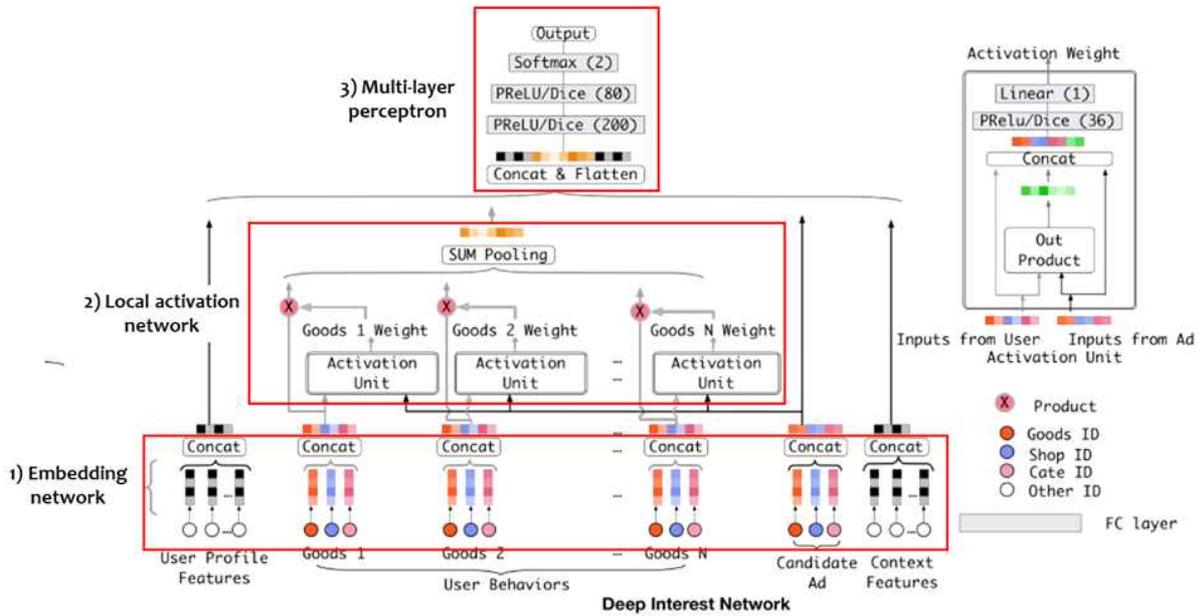


그림 2. Deep Interest Network [3]

비선호 매체를 추출해야 한다.

3. 오프라인 옥외광고 추천시스템

3.1. Deep Interest Network

옥외광고 이력 데이터는 고객 광고주와 광고주가 수행한 옥외광고 매체 정보, 그리고 광고주와 매체를 표현하는 다양한 연속형, 범주형 메타 정보들을 포함한다. 단지 광고주와 옥외광고 매체의 수행 이력만을 활용하는 협업 필터링은 다양한 메타 정보들로 유사성을 갖는 광고주와 옥외광고 매체들의 관계를 명시적으로 반영하기 힘들다. 또한, 광고주의 과거 옥외광고 이력의 매체들과 후보 옥외광고 매체의 연관성을 고려하여 후보 매체에 대한 선호도를 추론하여야 한다.

이에, 본 연구에서는 오프라인 옥외광고 추천을 위해 딥러닝 기반의 CTR 예측 모델인 DIN을 활용하였다. CTR은 온라인 추천에서 아이템을 클릭할 확률이고 일반적으로 아이템을 선호할 확률로 해석된다. DIN은 광고주와 광고주 메타 정보, 옥외광고 이력과 후보 옥외광고 매체를 입력받아 후보 옥외광고 매체의 선호도를 계산한다. 이때 DIN은 옥외광고 이력을 후보 옥외광고 매체와 연관성을 계산해 선택적으로 (Adaptively) 반영한 표현 벡터로 결합한다. DIN은 광고주 정보, 옥외광고 이력, 후보 옥외광고 매체를 각 표현 벡터로 결합하여 신경망을 통해 다양한 정보 간 복잡한 관

계를 반영하여 광고주의 옥외광고 선호도를 계산할 수 있다. DIN의 구조는 3가지 서브 네트워크로 이루어진다 (그림 2).

1) 임베딩 네트워크 (Embedding network)

DIN은 광고주와 옥외광고 매체, 그리고 범주형 메타 정보들을 One-hot encoding으로 입력받는다. 입력값이 고차원의 이진 벡터이기 때문에, 임베딩 네트워크는 각 입력 변수를 저차원의 밀집 벡터로 변환한다. 예를 들어, K_i 개의 범주를 가진 i 제 입력 변수 t_i 는 $W^i = [w_1^i, w_2^i, \dots, w_{K_i}^i] \in \mathbb{R}^{D \times K_i}$ 의 임베딩 네트워크를 가진다. 임베딩 네트워크를 통해 각 범주형 입력 변수는 D 차원의 밀집 벡터로 변환된다.

2) 지역적 활성화 네트워크 (Local activation network)

지역적 활성화 네트워크는 광고주의 옥외광고 이력에 적용되어 각 이력과 후보 옥외광고 매체의 연관성을 계산한다. 그리고 각 이력은 연관성에 따른 가중합으로 한 표현 벡터로 결합된다. 즉, 후보 매체 A 에 대한 광고 이력 벡터 v_U 는 아래와 같이 계산된다.

$$v_U(A) = f(v_A, e_1, e_2, \dots, e_H) = \sum_{j=1}^H a(e_j, v_A) e_j = \sum_{j=1}^H w_j e_j$$

$\{e_1, e_2, \dots, e_H\}$ 은 광고주의 옥외광고 이력 임베딩 벡터의 목록이고 v_A 는 후보 옥외광고 매체

의 임베딩 벡터이다. 지역적 활성화 네트워크 $a(\cdot)$ 는 활성화 가중치 w 를 출력하는 피드 포워드 네트워크로 옥외광고 이력은 후보 매체에 따라 다른 표현 벡터로 변환된다. $a(\cdot)$ 는 두 입력 벡터로부터 외적을 계산하여 다음 층에 입력하여 연관성 모델링에 대한 명시적인 정보를 반영한다.

3) 다층 퍼셉트론 (Multi-layer perceptron)

임베딩 네트워크로 변환된 광고주 벡터, 지역적 활성화 네트워크로 계산된 광고이력 벡터와 후보 옥외광고 매체 벡터는 하나의 벡터로 결합되어 다층 퍼셉트론으로 입력된다. 다층 퍼셉트론은 결합된 밀집 벡터로 광고주의 후보 옥외광고 선호도를 계산한다. 여기서 목적함수로 음의 로그우도 L 를 사용한다.

$$L = -\frac{1}{N} \sum_{(x,y) \in S} (y \log p(x) + (1-y) \log(1-p(x)))$$

S 는 크기 N 의 훈련 집합이며 x 는 모델의 입력값, $y \in \{0, 1\}$, $p(x)$ 는 후보 옥외광고 매체 선호도 예측값이다. L 은 딥러닝 기반의 이진 분류 문제에서 일반적으로 사용되는 손실함수이다.

3.2. 동 산업군 선호 및 점유 매체 제외 negative sampling

옥외광고 할당 문제는 한 광고판에 한 광고만 게시 가능하다는 '배타적 점유'의 특수성을 갖는다. 광고를 계약할 때 목표 광고 기간에 점유된 광고판은 고객 광고주에게 제공되지 않을 것이다. 또한, 본 연구는 같은 산업군의 광고주는 비슷한 연령 및 직업을 갖는 고객을 타겟한다는 논리 하에 '같은 산업군의 광고주는 비슷한 매체를 선호한다'고 가정한다. 이는 '비슷한 특성을 가진 사용자는 비슷한 선호를 가진다'는 협업 필터링의 가정과 유사하다. 본 연구는 '배타적 점유'의 특수성과 협업 필터링의 가정을 반영하여 아래와 같이 동 산업군 선호 및 점유 매체 제외 negative sampling을 제안한다.

동 산업군 선호 및 점유 매체 제외 negative sampling은 광고주의 산업군 메타정보를 활용해 동 산업군이 옥외광고를 수행한 매체를 후보 negative sample에서 제외한다. 옥외광고 구매 이력 데이터에 포함된 광고 계약 기간을 활용해 광고주의 각 옥외광고 구매 이력에 대해 목표 광고 기간에 이미 점유된 매체는 후보 negative sample에서 제외한다. 이 과정은 광고주는 '배타적 점유'에 의해

알고리즘 1) 동 산업군 선호 및 점유 매체 제외 negative sampling

Data: out-of-home advertising history

Input: num_negative t

Output: out-of-home advertising history with negative samples

For each user i do

cand_neg_media = non-advertised media of user i

cand_neg_media = cand_neg_media - advertised media of users in the same industry of user i

For each media history j of user i do

non_occupied_media = neg_media - occupied media at advertising period of a media history j

neg_sample = select t samples uniformly from non_occupied_media

append neg_sample to out-of-home advertising history

return out-of-home advertising history

해당 옥외광고를 수행할 당시, 기점유된 매체를 옵션으로 제공받지 못해 고려할 수 없는 상황을 반영한다. 즉, 제안한 negative sampling 방법은 옥외광고 시장의 영업 특성을 반영한 negative sampling 방법으로 고객 광고주가 선호할 수 있으나 옵션으로 제공받지 못한 매체를 효과적으로 제외하여 비선호 매체를 판별한다. DIN은 제안한 negative sampling으로 생성된 학습 데이터를 입력받아 광고주의 선호 및 비선호 패턴을 학습한다.

4. 실험

본 섹션에서는 제안 추천시스템을 평가하기 위해 진행한 실험에 대해 자세히 설명한다. 본 실험은 실제 옥외광고 이력 데이터를 기반으로 진행하였고 제안 추천시스템은 기존 모델보다 우월한 성능을 보였다.

4.1. 옥외광고 이력 데이터

옥외광고 이력 데이터는 2016년 1월부터 2021년 3월까지 계약된 옥외광고 이력이며, 광고주의 산업군, 광고주 ID, 계약 옥외광고 매체 ID, 옥외광고 매체 지역, 계약 기간을 포함한다. 세부적으로 본 데이터는 542명의 광고주, 61개의 매체, 94개의 광고주 산업군, 12개의 매체 지역, 그리고 총 2869개의 옥외광고 이력을 포함한다. 그 중 133명의 광고주가 하나의 광고 이력을 가지고 있다. (표 1)

학습 데이터와 테스트 데이터는 계약 순

광고주 ID	산업군	start_date	end_date	매체 ID	지역
A	병원	18.03.01	18.08.31	item_1	하남시
A	병원	19.02.01	19.03.31	item_3	고양시
B	문화	19.03.01	20.02.28	item_2	강남구
...

표 1. 옥외광고 이력 샘플 데이터

서를 기준으로 나누었다. 광고주의 광고 이력을 (b_1, b_2, \dots, b_n) 이라고 할 때, b_k , $k = 1, 2, \dots, n-1$ 의 광고 이력을 학습 데이터로 정의하였고 광고주의 마지막 광고 이력 b_n 을 테스트 데이터로 정의하였다. 또한, 광고주의 각 광고 이력 b_k 마다 negative sampling을 하여 광고 시점에서의 비선호 아이템을 학습 데이터에 추가하였다. 본 실험에서는 negative sample을 4개 선택하였다. DIN은 과거 기록을 입력받기 때문에, 각 광고주의 광고 이력에서 $k-1$ 번째까지의 광고 이력이 주어질 때 b_k 의 선호도를 맞추도록 학습하였다.

4.2. 비교 모델

- CF-IF [1]: CF-IF (Collaborative filtering for implicit feedback)은 암시적 피드백에 협업 필터링을 적용한 대표적인 모델이다. 협업 필터링은 머신러닝 기반의 전통적인 추천시스템 방법이다. 행렬 분해(MF, Matrix factorization)를 기반으로 사용자-아이템 상호작용 행렬을 분해하여 사용자 잠재 벡터와 아이템 잠재 벡터를 학습한다.
- NCF [15]: NCF (Neural collaborative filtering)은 암시적 피드백 기반 추천시스템의 대표적인 모델인 협업 필터링을 딥러닝 기반으로 확장한 모델이다. 사용자 잠재 벡터와 아이템 잠재 벡터를 MF와 MLP에 입력하여 사용자 정보와 아이템 정보의 비선형 관계를 반영하여 아이템에 대한 선호도를 계산한다. NCF는 MF에서의 사용자와 아이템 정보의 얇은 상호작용뿐만 아니라 MLP의 깊은 상호작용을 활용하여 선호도를 계산한다.
- DeepFM [16]: DeepFM (Deep factorization machines)는 FM (Factorization machines)의 얇은 변수간 상호작용과 DNN (Deep neural network)의 깊은 상호작용을 통합하여 아이템의 선호도를 예측한다. NCF가 사용자 벡터와 아이템 벡터만을 입력받는 것과 달리, DeepFM은 사용자와 아이템에 대한 메타 정보를 자유롭게 입력받아 아이템의 선호도를 계산할 수 있다.

4.3. 평가 방법

제안한 추천시스템과 비교 모델들의 비교를 위해 recall, precision at K ($P@K$), mean average precision at K ($MAP@K$)를 사용하였다. recall은 옥외광고 추천 문제를 이진 분류의 문제로 간주하여 분류 모델로서의 성능을 검증한다. 옥외광고 산업에서 보통 광고주는 영업사원이 적합하다고 판단한 몇 개의 옥외광고 목록을 제공받기 때문에, 본 실험에서는 $P@K$ 와 $MAP@K$ 를 통해 상위 K개의 추천 목록의 적합성을 판단한다.

옥외광고 추천 문제에 제안한 negative sampling의 적합성을 판단하기 위해 NCF, DeepFM, DIN에 random sampling과 제안한 negative sampling을 적용하여 학습하였다. Random sampling을 적용하여 학습한 모델은 '모델명-RS'으로 지칭하였고, 제안한 negative sampling으로 학습한 모델은 '모델명-NS'로 지칭하였다. CF-IF는 상호작용이 없는 아이템은 모두 비선호 아이템으로 가정하기 때문에 negative sampling을 적용하지 않았다.

4.4. 평가 지표 비교

	Recall	P@3	MAP@3
CF-IF	0.860	0.219	0.240
NCF-RS	0.805	0.315	0.313
DeepFM-RS	0.447	0.218	0.234
DIN-RS	0.775	0.277	0.283
NCF-NS	0.916	0.313	0.315
DeepFM-NS	0.558	0.248	0.257
DIN-NS	0.929	0.338	0.363

표 2. 비교 모델 평가 결과

표 2는 비교 모델들의 Recall, $P@3$, $MAP@3$ 를 나타낸다. 제안한 추천시스템 DIN-NS의 recall이 0.929로 가장 높은 것을 알 수 있다. CF-IF가 0.860의 비교적 높은 recall을 보였고, 이는 다수의 광고주가 과거에 광고한 광고판에 광고하는 특징이 있기 때문이다. 하지만, 3개의 추천 목록을 제공하는 것에 있어서 낮은 성능을 보여 옥외광고 추천시스템에는 적합하지 않다. 딥러닝 기반 비교 모델 중 DeepFM, NCF, DIN 순으로 recall이 높다. 이를 통해 광고주에 대한 산업군, 아이템에 대한 지역 범주 메타 정보를 모델에 입력하고, 과거 이력을 반영하는 것이 옥외광고 선호 추론의 정확도 향상에 기여함을 알 수 있다. 또한, DIN-NS의 $P@3$, $MAP@3$ 이 가장 높은 것을 알 수 있다. 이는 몇 개의 옥외광고 목록을 광고주에게 제공하는 옥외광고 산업에서 DIN-NS가 적절한 옥외광고 목록을 제공

할 수 있음을 의미한다.

표 2에서 NCF-RS, DeepFM-RS, DIN-RS보다 NCF-NS, DeepFM-NS, DIN-NS가 대체적으로 성능이 높음을 알 수 있다. 옥외광고 이력이 없는 광고판 중 무작위로 비선호 광고판을 추출하는 것보다 옥외광고 시장의 특성을 고려하여 광고주의 비선호 광고판을 선택하여 모델을 학습한 것이 모델의 성능 향상에 기여함을 알 수 있다.

4.5. 옥외광고 이력과 연관성

DIN의 지역적 활성화 네트워크는 옥외광고 이력과 후보 옥외광고의 연관성을 피드포워드 네트워크로 계산하여 후보 옥외광고에 대한 과거 이력의 정보를 선택적으로 결합한다. 제안한 추천시스템은 지역적 활성화 네트워크의 활성화 가중치를 통해 고객 광고주에게 추천의 근거를 제시할 수 있다.

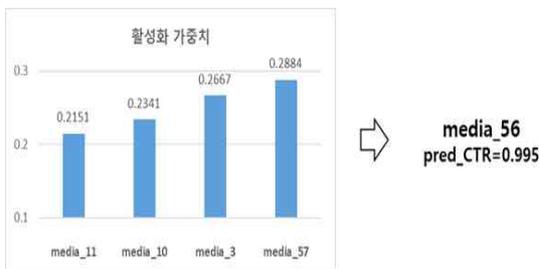


그림 3. 활성화 가중치와 후보 옥외광고 예측

그림 3은 테스트 데이터에서 한 광고주의 후보 옥외광고의 선호도를 계산할 때 계산된 활성화 가중치이다. 예를 들어, media_56은 광고주의 각 이력과 0.2151, 0.2341, 0.2667, 0.2884의 연관성을 통해 0.995의 선호도를 가진다고 할 수 있다. DIN의 활성화 가중치를 통한 과거 이력과의 연관성으로 광고판의 지역 정보 등 정성적인 근거뿐만 아니라 모델을 통한 정량적인 근거를 제시하여 옥외광고 추천의 설명력과 신뢰성을 높일 수 있다.

4.6. 콜드 스타트

콜드 스타트에 강건한 추천시스템은 신규 광고주에게 광고주의 산업군 범주 정보만으로 적합한 추천을 할 수 있어야 한다. 본 실험에서는 제안한 추천시스템이 신규 광고주에게 적합한 옥외광고를 추천하는가를 평가하였다. 옥외광고 이력 데이터에서 광고 이력이 1개인 133명의 광고주를 제외하고 모델을 학습하였고 제외한 광고주들의 이력을 광고 이력이 없는 신규 광고주로 가정하였다.

	New-user test set	Test set
Recall	0.887	0.953

표 3. 신규 사용자 테스트 결과

표 3은 광고 이력이 없는 신규 광고주의 첫 선호 매체를 대상으로 하는 new-user test set과 광고 이력이 있는 기존 광고주의 최종 선호 매체를 대상으로 하는 test set에 제안 추천시스템의 recall을 제시한다. New-user test set의 recall이 0.887로 광고주의 산업군 범주 정보만으로도 신규 광고주에게 적절한 옥외광고를 제안할 수 있음을 알 수 있다. 즉, 제안한 추천시스템은 콜드 스타트에 강건하다고 할 수 있다. 또한, 광고 이력이 1개인 광고주들의 데이터를 학습에서 제외했음에도 불구하고 기존 광고주들에 대한 추천시스템의 recall이 0.953으로 높다.

5. 결론 및 토의

옥외광고 산업은 중요성에 비해 체계적이고 과학적인 옥외광고 계약 및 편성이 이루어지지 않았다. 이에, 본 연구는 과거 옥외광고 이력을 기반으로 오프라인 옥외광고 추천시스템을 개발하였다. 본 연구는 광고주와 옥외광고의 메타 정보를 반영하고 옥외광고 이력과 후보 옥외광고의 연관성을 통해 선호도를 계산하는 DIN을 활용하고 사용 옥외광고 산업의 특성을 반영한 negative sampling을 통해 모델을 학습하였다. 실제 옥외광고 이력 데이터에서 제안한 추천시스템은 비교 모델들보다 우월한 성능을 보였고, DIN의 지역적 활성화 네트워크를 통해 추천의 설명력과 신뢰성을 높였으며, 콜드 스타트에도 강건함을 보였다. 제안한 추천시스템은 유사한 오프라인 추천 문제에 시사점을 제시할 것이라 예상된다.

후속 연구에서는 옥외광고의 지역 범주 정보뿐만 아니라 추가적인 메타 정보를 포함하여 모델을 학습할 것이다. 예를 들어, 본 모델에 인구 정보를 추가하여 광고판 간의 연관성을 다각도로 반영해 광고판의 선호도를 계산할 것이다. 또한, 제안한 negative sampling을 확률적인 sampler 모델로 발전시켜 추천시스템의 성능을 향상할 것이다. 마지막으로 본 연구를 환경의 물리적 메타 정보를 반영할 필요가 있는 오프라인 매장 상품 추천 등 다른 분야에 적용하여 제안한 추천시스템을 평가할 것이다.

6. 참조 문헌

- [1] 문화체육관광부 (2020) 2019 광고산업조사. https://www.mcst.go.kr/kor/s_policy/dept/depView.jsp?pSeq=1300&pDataCD=0417000000&pType=08에서 검색
- [2] Hu, Y., Koren, Y., & Volinsky, C. (2008, December). Collaborative filtering for implicit feedback datasets. In 2008 Eighth IEEE International Conference on Data Mining (pp. 263-272). Ieee.
- [3] Zhou, G., Zhu, X., Song, C., Fan, Y., Zhu, H., Ma, X., ... & Gai, K. (2018, July). Deep interest network for click-through rate prediction. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 1059-1068).
- [4] 김남기, 정석봉. (2017). 협업 필터링을 이용한 오프라인 쇼핑몰에서 개인 맞춤형 브랜드 추천시스템에 관한 연구. 한국경영학회 통합학술발표논문집, (), 600-619.
- [5] 김남기, 정석봉(2016).오프라인 쇼핑몰에서 고객 행위에 기반을 둔 맞춤형 브랜드 추천에 관한 연구. Journal of Information Technology Applications & Management,23(4),55-70,.
- [6] Choi, Y. H., & Lee, S. Y. (2006). Users' Moving Patterns Analysis for Personalized Product Recommendation in Offline Shopping Malls. Journal of Korean Institute of Intelligent Systems, 16(2), 185-190.
- [7] Pan, R., Zhou, Y., Cao, B., Liu, N. N., Lukose, R., Scholz, M., & Yang, Q. (2008, December). One-class collaborative filtering. In 2008 Eighth IEEE International Conference on Data Mining (pp. 502-511). IEEE.
- [8] Xu, Z., Chen, C., Lukasiewicz, T., Miao, Y., & Meng, X. (2016, October). Tag-aware personalized recommendation using a deep-semantic similarity model with negative sampling. In Proceedings of the 25th ACM International on Conference on Information and Knowledge Management (pp. 1921-1924).
- [9] Liu, Y., Wang, S., Khan, M. S., & He, J. (2018). A novel deep hybrid recommender system based on auto-encoder with neural collaborative filtering. Big Data Mining and Analytics, 1(3), 211-221.
- [10] He, X., Liao, L., Zhang, H., Nie, L., Hu, X., & Chua, T. S. (2017, April). Neural collaborative filtering. In Proceedings of the 26th international conference on world wide web (pp. 173-182).
- [11] He, X., Zhang, H., Kan, M. Y., & Chua, T. S. (2016, July). Fast matrix factorization for online recommendation with implicit feedback. In Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval (pp. 549-558).
- [12] Zhang, W., Chen, T., Wang, J., & Yu, Y. (2013, July). Optimizing top-n collaborative filtering via dynamic negative item sampling. In Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval (pp. 785-788).
- [13] Ding, J., Quan, Y., He, X., Li, Y., & Jin, D. (2019, August). Reinforced Negative Sampling for Recommendation with Exposure Data. In IJCAI (pp. 2230-2236).
- [14] Wang, X., Xu, Y., He, X., Cao, Y., Wang, M., & Chua, T. S. (2020, April). Reinforced negative sampling over knowledge graph for recommendation. In Proceedings of The Web Conference 2020 (pp. 99-109).
- [15] He, X., Liao, L., Zhang, H., Nie, L., Hu, X., & Chua, T. S. (2017, April). Neural collaborative filtering. In Proceedings of the 26th international conference on world wide web (pp. 173-182).
- [16] Guo, H., Tang, R., Ye, Y., Li, Z., & He, X. (2017). DeepFM: a factorization-machine based neural network for CTR prediction. arXiv preprint arXiv:1703.04247.

