

인공지능 기반 광고비 예측 알고리즘 개발

전경민¹⁺, 강제하², 배희재², 윤은수², 김종원^{3*}, 정대식^{4*}, 송영기^{4*}, 김경수^{4*}

¹⁺상명대학교 컴퓨터과학전공 학사과정

²상명대학교 컴퓨터과학과 석사과정

^{3*}상명대학교 지능 IOT 융합전공 교수

^{4*}상명대학교 융합공과대학 SW융합학부교수

^{4*}주식회사 11시 11분 CEO

^{4*}주식회사 11시 11분 CTO

wjsrudals69@naver.com jhkang3457@gmail.com baehj100@gmail.com hkhk0331@gmail.com

jungsoft97@smu.ac.kr jwkim@smu.ac.kr yksong@11h11m.com kelvin@11h11m.com

(+: 1 저자, *: 교신저자)

Development of AI-based advertising cost prediction algorithms

Kyung-Min Jeon¹⁺, Jae-Ha Kang², Hui-Jae Bae², Eun-Su Yun², Jong-weon Kim^{3*}, Dae-Sik Jeong^{4*}

¹⁺Dept. of Computer Science, Sang-myung University.

²Dept. of Computer Science Sang-myung University.

^{3*}Dept. of Electronics Engineering Sang-myung University.

^{4*}Dept. of Faculty of SW Convergence, College of Convergence Engineering, Sang-myung University.

^{4*}Eleven past eleven

^{4*}Eleven past eleven

요 약

시장 경쟁력을 확보하고 기업을 성장시키기 위해서는 광고 행위가 필수적이므로 현재까지 효율적으로 광고하기 위한 여러 가지 방안들이 활용되었다. 이 중에는 타 업체와의 경쟁전략을 위해서 경쟁업체의 광고비를 파악하려는 과정도 포함 되어있다. 이에 디지털 광고 측면에서는 상대적으로 광고의 노출, 클릭, 시간 대 등의 관련 정보를 획득하기 용이하므로 본 연구에서는 대량의 데이터를 이용하고 XGBoost(Extreme Gradient Boosting) 알고리즘을 활용하여 크롤링된 데이터 그룹을 분석하고, 클릭 수를 예측하는 모델을 구현하였다. 실험 결과 모델의 RMSE(Root Mean Squared Error) Average 가 1.13 정도 나온 것을 확인하였고 이에 따른 과적합을 피하기 위한 방안을 검토하였다.

1. 연구 배경 및 목적

최근 광고 업계에서는 AI(Artificial Intelligence) 기술이 핵심적 역할을 하고 있다. AI 는 타겟팅 및 개인화된 광고, 콘텐츠 생성 및 관리, 시장 예측 및 분석 등[1]의 다양한 전략에서 변화를 주도하고 있으며, 이러한 변화는 광고 업계의 방식에도 영향을 미치고, 관련 업자들에게 새로운 연구 주제를 제공하고 있다[2]. 광고와 AI 의 융합은 광고 시장의 확장에 있어서 중요한 기회를 제공하고 있지만, 경제적 측면, 특히 광고비 예측의 정확성 향상은 여전히 큰 도전과제로 남아 있다. 이에 AI 를 통한 광고비 예측 모델을 개발하는 것은 인적자원이 부족해지는 사회적 현상에 수많은 광고주의 광고비용을 관리해야 하는 광고대행사의 현실적 문제 해결에 이바지하며, 최소한의 비용으로 최대한의 매출을 창출해야 하는 광고주에게 있어 효과적인 의사결정에 도움을 줄 것이다.

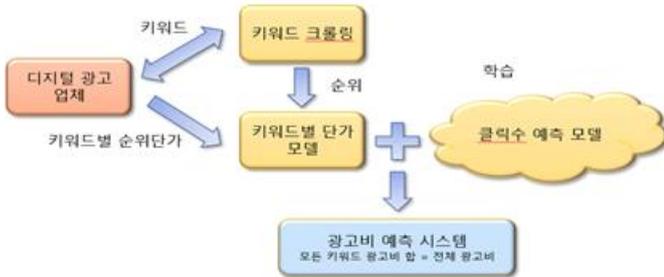
2. 연구 과정

1) 예측 기법

디지털 광고는 노출 순서를 높이기 위해서 입찰 형태로 광고비가 결정되는데, 이때 입찰이 된 후 클릭 할 때마다 입찰가를 지불하게 된다. 따라서 광고비를 도출하는 과정은 해당 키워드의 순위 별 입찰가와 클릭수의 곱이 된다. 크롤링하여 얻은 키워드에 관한 정보는 노출 수, 클릭 수, 시간 등이 있고, 각각의 키워드에 대한 클릭 수를 입찰가 데이터와 조합하여 보다 정확한 광고비 예측이 가능하다. 본 연구에서는 XGBoost(Extreme Gradient Boosting) 알고리즘을 활용하여 클릭 수를 예측하고자 한다. XGBoost 는 Gradient Boosting 기법을 확장한 것으로, 비선형 문제에 강력하며 과적합 방지가 가능하고 대규모 데이터셋에 적합하다는 장점이 있다[3]. 또한 XGBoost 는 결정 트리를 기반으로 한 앙상블 기법을 사용하여, 여러 개의 약한 예측 모델을 결합함으로써 최종적인 강력한 예측 모델을 생성한다.

2) 데이터 분석

본 연구에서 사용한 데이터는 네이버에서 하루에 약 100 만개의 키워드를 크롤링한 실제 데이터이며, 본 연구에서 크게 3 개의 그룹으로 나뉘었다. 그룹 1 은 키워드로 통합검색 영역을 각각 PC/MOBILE 로 크롤링한 결과, 그룹 2 는 키워드 광고의 실제 결과, 그룹 3 은 가상의 견적 데이터이다. 그룹 2 의 Column 들은 날짜, 키워드 명, 도메인, 디바이스 타입, 시간, 노출 수, 클릭 수, 노출 순위 평균, 광고비로 이루어져 있고, 그룹 3 은 날짜, 키워드 명, 디바이스 타입, 순위, 예상 입찰가, 예상 노출 수, 예상 클릭 수, 예상 광고비로 이루어져 있다. 광고비 예측 모델은 클릭 수를 정확히 예측하는 데 중점을 두고 있으며, 이 클릭 수는 광고주의 클릭당 입찰가와 직접 연결되어 있어 광고 비용을 계획하고 최적화하는 데 매우 중요하다. 디지털 광고 업체로부터 키워드를 크롤링한 결과, 키워드별 단가가 있는 데이터와 XGBoost 로 학습한 결과를 토대로 만든 클릭 수 예측 모델을 결합하여 광고비 예측 시스템의 흐름은 (그림 1)과 같다.



(그림 1) 광고비 예측 시스템

3) 연구 방법

먼저, 데이터 그룹 1 과 그룹 2 를 날짜, 키워드명, 디바이스 타입, 순위로 merge 시킨다. 이때 데이터 그룹 1 은 동일한 날짜에 각각의 시간대별로 구분되어 있으므로 merge 하기 위해 시간대를 삭제하고 같은 날짜끼리 노출 수, 클릭 수는 합한 뒤 순위를 평균 내었다. 그 다음 XGBoost 모델을 적용하기 위해 Column(날짜, 키워드)를 삭제하고, 데이터 그룹 2 의 클릭 수를 타겟 값으로 설정하였다. 따라서 최종 데이터의 X 는 디바이스 타입, 순위, 입찰금, 예상 노출 수, 예상 클릭 수, 노출 수가 되고, Y 는 클릭 수가 된다. XGBoost 모델 훈련 시 train 데이터셋과 test 데이터셋으로 분할하는 방법이 다르거나 난수가 다르면 RMSE(Root Mean Squared Error)도 달라지기 때문에 본 연구에서는 k 를 10 으로 설정한 k-Fold Cross Validation 방법을 사용하였다.

3. 연구 결과

XGBoost 모델을 적용하고 교차 검증한 클릭수 예측의 결과는 (표 1)과 같다. 결론적으로 본 연구에서는 클릭 수를 예측하는 데 있어 과적합이 발생하였다.

Algorithm	RMSE Average
XGBoost	1.13

(표 1) XGBoost 교차 검증 결과

과적합이 발생하는 경우는 일반적으로 모델이 학습 데이터셋에 지나치게 맞춰져 복잡해지기 때문에 새로운 데이터에 대해서는 정확하게 예측하지 못하는 문제가 발생한다. 과적합이 발생한 원인을 분석한 결과, 대부분의 타겟 값이 0 과 1 에 집중되어 있다는 것을 (표 2)를 통해 확인할 수 있다.

클릭 수	비율
0	0.932610
1	0.047618
2	0.009387
3	0.003565
4	0.001942
...	...
110	0.000001
137	0.000001
154	0.000001

(표 2) 훈련 데이터 클릭 수 비율

또한 학습 데이터셋의 양은 충분하였지만 X(변수)가 충분하지 못한 것을 확인할 수 있었다. 클릭 수를 예측하기 위해서는 보다 많은 변수들이 존재하는데, 훈련 데이터의 경우에는 6 가지밖에 되지 않아 과적합이 발생하게 되었다.

4. 결론 및 기대효과

본 연구의 한계점을 극복하기 위해서 두 가지 접근이 필요하다. 첫째, 클릭 수가 적은 키워드를 분류하고 별도 모델을 적용해야 한다. 대다수 클릭 수(97%)가 0 과 1 이므로, 이런 키워드는 다른 데이터셋으로 분리하고, 다른 모델을 사용할 필요가 있다. 둘째, 추가 변수를 도입해야 한다. 기존의 노출 수와 클릭 수 외에 키워드별 카테고리 분류를 통해 다양한 모델을 생성할 예정이다. 향후에는 이러한 접근으로 과적합 문제를 해결하고, 정확도를 향상시킬 것이다. 향상시킨 모델을 효과적으로 적용할 경우 특정 사이트에 국한되지 않고 다양한 광고 서비스에 적용될 수 있으며, 이를 통해 사업 영역을 확장할 수 있는 기대감을 제공할 것이다. 결과적으로 이 연구를 통해 제시된 방법론과 기술들은 광고 산업 내에서의 혁신적 발전을 이끌고, 광고 캠페인의 성공률을 높이는 데 기여할 것이다.

참고문헌

[1] 차영란, “Artificial Intelligence Strategy for Advertising and Media Industries: Focused on In-depth Interview”, 한국콘텐츠학회논문지, 제18권, 제9호, 105-107, 2018.
 [2] 윤성욱, “마케팅연구에서의 머신러닝/AI활용 동향”, KISDI AI TREND WATCH, 제6호, 2-7, 2021.
 [3] 한영진, “Prediction of Ad Clicks Using Early Stop Based on XGBoost”, 한국통신학회논문지, 제46권, 제6호, 994-996, 2021.