

AD Finder: 머신러닝을 활용한 실시간 광고성 블로그 리뷰 탐지 프로그램

조민서[○], 박상근

경희대학교 소프트웨어융합대학 소프트웨어융합학과
choms0209@khu.ac.kr, sk.park@khu.ac.kr

AD Finder: Real-time Advertising Blog Review Detection Program using Machine Learning

Minseo Jo[○], Sangkeun Park

Department of Software Convergence, Kyung Hee University

요 약

온라인 쇼핑에서 제품 리뷰의 영향력이 커지면서 험찬 사실을 숨긴 과장된 광고 리뷰가 등장하고 있다. 이로 인한 소비자 문제를 해결하기 위해 광고 및 가짜 리뷰 탐지와 관련된 여러 연구가 진행되었으나, 광고 및 가짜 리뷰의 특성을 숨긴 리뷰 탐지에는 한계가 있다. 본 연구에서는 이런 한계점을 극복하기 위해 네이버 블로그를 대상으로 광고 리뷰를 탐지하는 ‘AD Finder’ 앱을 개발하였다. 이를 위해 광고와 비광고 블로그 간의 유의미한 차이가 있는 6 가지 특성을 기반으로 데이터 증강 기법을 활용한 광고 블로그 분류 모델을 개발했다. 이 모델을 활용하여 사용자가 네이버 블로그 리뷰를 통해 정보를 탐색할 때 광고성 블로그 글과 일반 글을 구분해 주는 크롬 익스텐션 앱을 개발하고 사용자 스터디를 통해 광고 리뷰 탐지 모델의 유용성을 검증하였다.

1. 서 론

다른 구매자의 제품 리뷰는 새로운 소비자가 해당 제품 구매 여부를 결정할 때 결정적인 참고 요인으로 작용한다. 2021 년 한국소비자연맹에서 진행된 ‘소비자 리뷰 영향력’ 조사에 따르면 응답자 중 97.2%가 상품 구매 전 리뷰를 확인한다고 답변했으며, 이 중 82.3%는 ‘구매 후 불만족을 줄이기 위해서’ 리뷰를 참고한다고 답하였다 [1].

리뷰의 영향력이 커지면서 일부 판매자들은 특정 리뷰어에게 금전적 대가나 혜택을 제공하여 의도적인 긍정적 리뷰를 양산하기 시작했다. 심지어 상업적인 리뷰임을 숨기고 소비자들의 신뢰를 얻기 위해 험찬 사실을 제대로 밝히지 않는 등, 소위 ‘뒷광고’ 리뷰로 인한 소비자 피해가 증가하고 있다 [2, 3].

이에 따라 광고 및 가짜 게시글을 탐지하는 다양한 연구가 진행되었다. 게시글의 텍스트 분석을 기반으로 하여 머신러닝과 딥러닝 기법을 활용해 광고 여부를 판단하는 연구[4]뿐만 아니라 게시글의 구조적 특징과 작성자의 행동 분석을 기반으로 한 메타데이터를 이용해 광고 게시글을 탐지하는 연구[5,6,8]가 진행되었다. 이와 더불어 광고 탐지 기능을 제공하는 모바일 앱과 크롬 익스텐션 앱 형태의 서비스[7,8]도 개발되었다.

본 연구에서는 블로그 리뷰가 광고성으로 작성된 리뷰인지 분류하는 데 효과적인 특징들을 찾고 이를 기반으로 광고성 블로그 리뷰를 분류할 수 있는 머신러닝 모델을 개발했다. 이를 활용해 사용자가 읽고자 하는 블로그 리뷰의 상업성

여부를 실시간으로 판단하고 표시하는 크롬 익스텐션 앱을 개발하였으며 사용자 스터디로 본 연구의 효과를 검증하였다.

2. 관련 연구

광고 및 가짜 게시글 탐지에 대한 다양한 연구가 수행되었다. Yoo et al.[4]은 딥러닝을 활용한 자연어 처리 방법을 사용해 불법 금융 광고 탐지 딥러닝 모델을 구축했다. Lee et al.[5]는 중고 거래 플랫폼에서의 사기 게시글과 정상 게시글을 비교 분석하여 텍스트뿐 아니라 ‘이미지 평균 크기’, ‘가격’, ‘핸드폰 번호 반복’ 등의 글의 구조적 특징을 활용하여 효과적으로 사기 게시글을 분류할 수 있음을 밝혔다. 임문영 & 박승범[6]은 ‘가입 후 글쓰기까지 걸린 개월 수’, ‘글을 게시한 시각’ 등이 가짜 뉴스 작성자를 찾는 데 효과적인 특징임을 확인하는 등, 텍스트 외에 메타 데이터도 광고 및 가짜 게시글 탐지에 유용하게 활용되고 있다.

사용자에게 광고 및 가짜 리뷰를 구분해서 제공하기 위한 다양한 연구가 수행되었다. 정다비치 & 노영주[7]는 사용자가 특정 콘텐츠 키워드를 검색하면 텍스트 분석을 통해 광고 리뷰를 탐지하고 이를 필터링 해주는 모바일 앱을 개발하였으며, Han et al.[8]은 현재 접속 중인 쿠팡 제품의 리뷰 정보를 크롤링하여 가짜 리뷰 계정의 특성을 추출하고 KNN 모델을 통해 가짜 리뷰어를 식별하여 표시해 주는 크롬 익스텐션 앱을 개발하였다.

본 연구는 광고와 비광고를 구분하는 다양한 특성을 발견하고 데이터 증강 기법과 머신러닝 및 딥러닝을 활용해 단일 특성에만 의존하지 않고 다양한 변수들을 학습한 모델을

개발하였다. 이는 부당 광고 리뷰 작성자가 게시글 작성 시 특정 특성이 드러나지 않게 의도적으로 숨기려고 해도 광고임을 식별해 낼 수 있다는 점에서 기존 연구의 한계를 보완했다. 또한 이를 실제 서비스로 제공하고자 네이버 블로그 리뷰의 광고 여부를 실시간으로 분석해 주는 크롬 익스텐션 앱을 개발하고 사용자 스터디를 통해 앱 서비스의 활용성을 확인했다.

3. 데이터 수집 및 전처리

3.1 학습 데이터 수집

파이썬 requests¹ 모듈을 활용해 다양한 주제의 10 가지 키워드(‘홍대 맛집’, ‘신발’, ‘화장품’, ‘스마트폰’, ‘가방’, ‘자동차’, ‘여행’, ‘알고리즘’, ‘에버랜드’, ‘치킨’)에 대한 네이버 블로그 리뷰를 크롤링했다. 총 494 개의 블로그 리뷰를 수집했으며, 각 블로그 리뷰를 직접 읽어보면서 비광고 블로그(283 개)와 상업적 광고 블로그(211 개)를 분류했다. 블로그 리뷰가 상업적 광고인지 아닌지를 판별할 때는 ‘공정거래 위원회 표시 광고법²’에 따라 ‘업체로부터 금전적 지원을 받았다.’, ‘후원을 받았다’, ‘원고료 지원을 받았다.’ 등의 언급이 있는 리뷰를 광고 리뷰로 판단했다.

3.2 광고/비광고 리뷰 분류를 위한 특징 추출 및 가공

수집한 블로그 리뷰 데이터를 기반으로 상업적 광고 여부를 판단하는데 주요 역할을 하는 ‘협찬 문구 포함 여부’, ‘내돈내산’ 키워드 포함 여부, 그리고 ‘최근 일주일 내 작성된 리뷰’ 변수를 생성했다. 이를 통해 1) 리뷰 텍스트 길이(int), 2) 리뷰에 달린 댓글 수(int), 3) 리뷰에 포함된 이미지 수(int), 4) 리뷰에 포함된 외부 링크 수(int), 5) ‘내돈내산’ 키워드 포함 여부(bool), 6) 협찬 문구 포함 여부(bool)의 총 6 가지 특징을 도출하였다. 광고 및 비광고 글을 분류하는 데 있어 추출한 변수들의 적합성을 확인하기 위해 변수별로 t-test 와 카이제곱 검정을 실시하였다. 가설검정 결과, 모든 변수의 p-value 가 유의 수준(0.05) 보다 낮게 나타났다.

상품 리뷰에 대한 블로그 게시글에 협찬 문구가 포함되어 있다면 해당 게시글은 광고임을 확신할 수 있다. 하지만 ‘협찬 문구 포함 여부’ 변수에 지나치게 의존하여 협찬 문구가 없는 광고 데이터에 대한 모델의 예측 정확도가 떨어지는 문제가 발생할 수 있다. 이러한 문제를 해결하기 위해 데이터 증강 기법을 활용하여 협찬 문구가 없는 비광고 데이터에 ‘광고 여부’ 레이블 값을 1 로 변경한 데이터를 추가함으로써 총 777 개의 데이터셋을 구축하였다. 이 방법을 통해 모델이 ‘협찬 문구 포함 여부’ 변수뿐만 아니라 다양한 변수 학습하여

협찬 문구가 없는 부당 광고 글도 탐지할 수 있는 성능을 가질 수 있게 된다.

4. 광고/비광고 분류 모델 구축 및 평가

전처리 과정과 가설 검정 과정을 거친 최종 6 개의 변수를 기반으로 광고/비광고 분류 모델을 구축하기 위해 딥러닝과 머신러닝 기법을 활용하였다. TabNet³(Ver.4.1.0), XGBoost⁴(Ver. 2.0.3), LightGBM⁵(Ver. 4.1.0), CatBoost⁶(Ver. 1.2.3)의 총 4 가지 알고리즘을 선정하였다. scikit-learn 에서 제공하는 GridSearchCV 를 사용하여 각 예측 모델의 최적 하이퍼 파라미터 탐색 및 교차 검증을 수행한 결과는 표 1 과 같다.

표 1 모델별 평가 결과

Model	Accuracy	Precision	Recall	F1	AUC
CatBoost	0.846	0.837	0.939	0.885	0.876
XGBoost	0.808	0.810	0.909	0.857	0.866
LightGBM	0.807	0.810	0.909	0.857	0.865
TabNet	0.737	0.779	0.818	0.798	0.759

5. 광고 리뷰 판별 서비스 개발

5.1 시스템 개발

JavaScript 기반의 크롬 익스텐션 앱을 개발하고 Python 웹 프레임워크인 Flask⁷를 이용하여 서버를 구축하였다. 서버에서 사용자의 페이지에 있는 블로그 포스트 데이터를 수집하고 가장 높은 성능을 보인 CatBoost 모델을 활용하여 수집된 블로그의 광고 여부를 판별한다. 예측 결과값은 클라이언트에게 반환되며 이에 따라 웹 페이지 UI 에 광고로 의심되는 블로그 포스팅에 표시 기능을 제공한다. 앱 시스템 구조는 그림 1 과 같다.

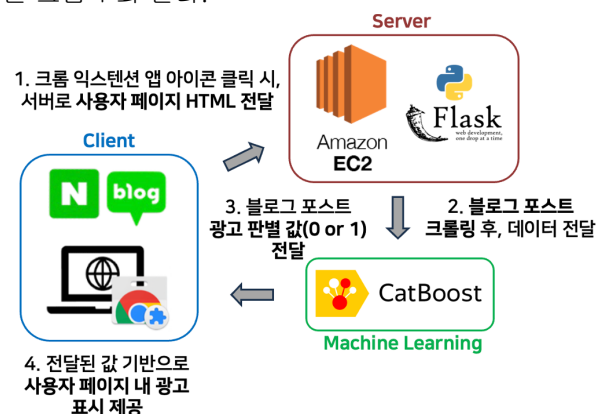


그림 1 ‘AD Finder’ 시스템 구조

¹ <https://requests.readthedocs.io/en/latest/>

² <https://tinyurl.com/y27qhk6m>

³ <https://tinyurl.com/275rjd89>

⁴ <https://xgboost.readthedocs.io/en/stable/>

⁵ <https://lightgbm.readthedocs.io/en/v4.1.0/>

⁶ <https://catboost.ai/en/docs/>

⁷ <https://flask.palletsprojects.com/en/3.0.x/>

5.2 ‘AD Finder’ 크롬 익스텐션 앱 개발

‘AD Finder’ 크롬 익스텐션 앱 서비스 화면은 그림 2 와 같다. 크롬 브라우저의 우측 상단에 존재하는 앱 아이콘을 클릭하면 앱이 실행된다. 광고로 의심되는 블로그 포스팅에는 HTML 과 CSS 요소를 삽입하여 ‘AD’ 텍스트 배지와 블로그 포스팅 리스트 박스의 색상을 주어 광고 표시를 제공한다. 이를 통해 사용자는 포스팅을 직접 확인하지 않아도 광고 여부를 쉽게 알 수 있으며 신뢰성 있는 후기 정보를 얻는 데 도움을 받을 수 있다.

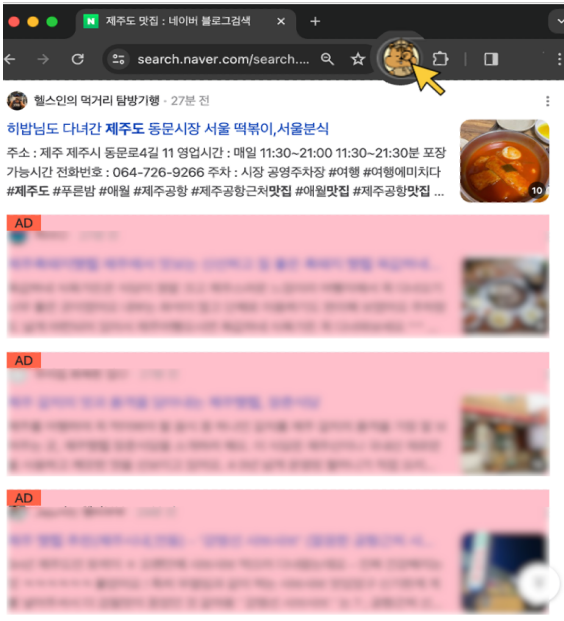


그림 2 ‘AD Finder’ 크롬 익스텐션 앱 서비스 화면

6. 사용자 스터디

‘AD Finder’의 사용성을 검증하기 위해 평소 네이버 블로그를 통해 상품 정보를 검색하는 20 대 여성 2 명, 20 대 남성 1 명, 50 대 여자 1 명을 대상으로 사용자 스터디를 진행했다. 4 명의 참여자에게 원하는 키워드를 검색하게 했으며 이들은 각각 ‘데이트 맛집’, ‘대전 맛집’, ‘여행’, ‘남자 운동화’ 키워드를 검색하였다. 그리고 ‘AD Finder’ 앱을 사용해 광고와 비광고 표시가 4:6 또는 6:4 의 비율로 나온 10 개의 블로그에서 정보를 탐색했다. 그 후, 앱을 사용하여 정보를 탐색한 경험에 대해 질문하였다.

광고 표기 효과성에 대해 검증하고자 광고 및 비광고 구분의 명확성에 대해서 5 점 척도 기반 조사를 진행하였다. 그 결과, 평균 5.0 점의 점수를 얻었다. 또한 “포스팅 마지막에 위치하는 협찬 문구 때문에 전체를 읽기 전까진 광고인지 모르는 경우가 많았는데, 포스팅을 클릭하기 전에도 광고 여부를 알 수 있어 편리했다.” 등 광고 표기에 긍정적인 반응을 확인할 수 있었다. 정보 탐색의 효율성 평가를 위해 앱을 통해 원하는 정보를 찾는 데 얼마나 효과적이었는지에 대해 5 점 척도 기반 조사를 진행하였다. 그 결과, 평균 4.25 점의 점수를 받았다. 4 명의 참여자 중 ‘데이트 맛집’, ‘남자 운동화’를 검색한 2 명은 광고 게시글임에도 불구하고 이를 의도적으로 확인했다. 그 이유는 “좋아하는 메뉴의 맛집이 있었기 때문”, “신발의 코디 활용도,

사이즈 정보를 얻기 위해서”라고 답변하였다. 이를 통해 일부 사용자들이 자신에게 필요한 정보를 얻기 위해 광고 게시글을 의도적으로 탐색하는 경향이 있음을 확인할 수 있었다.

7. 결론 및 제언

본 연구는 온라인 쇼핑에서 제품 리뷰의 중요성이 커지면서 발생하는 부당 광고 리뷰 문제를 해결하기 위해 네이버 블로그 리뷰 게시글을 대상으로 광고와 비광고를 구별할 수 있는 머신러닝 모델을 개발하였다. 그리고 이를 기반으로 ‘AD Finder’ 크롬 익스텐션 앱을 개발하여 사용자 스터디를 통해 앱 서비스의 효과성을 확인하였다. 하지만 사용자 스터디 결과, 협찬문구가 포함된 일부 글이 광고로 분류되지 않는 등의 한계점이 나타났다. 특히 협찬 문구가 없는 부당 광고 글을 식별하는 데는 여전히 어려움이 존재한다. 향후 연구에서는 모델의 F1-score 를 향상시켜 광고 판단의 정확성을 높이고자 한다. 또한 광고 글에서도 정보를 찾고자 하는 사용자를 위해 해당 글 내에서 유용한 정보만 필터링하여 제공하는 기능을 추가하고자 한다. 반면, 광고를 전혀 보고 싶지 않은 사용자를 위해서는 광고로 표시된 글을 사용자가 선택적으로 숨길 수 있는 필터링 기능을 추가할 계획이다.

8. 참고 문헌

- [1] 김세라, 온라인 쇼핑 이용후기 ‘실제 구매에 큰 영향’ 미쳐, 소비자 경제, <https://tinyurl.com/2356bmst>, 2023.
- [2] 정진호, 내돈내산? 점점 교묘해지는 SNS 뒷광고. 중앙일보, <https://tinyurl.com/2cktp6f>, 2023.
- [3] 김태준, ‘솔직 후기’라더니 뒷광고였다...피해 사례 5 년새 5 배. 조선일보, <https://tinyurl.com/2az99rwv>, 2023.
- [4] 유길상, 장진희, 김성주, 김광용. 딥러닝을 활용한 인터넷 불법금융광고 판별 모델에 관한 연구. 한국컴퓨터정보학회논문지 28, 8, 21-30, 2023.
- [5] 이동우, 황요셉, 민진영. C2C 중고거래 환경에서의 사기 게시물 탐지 모델에 관한 연구. 한국경영정보학회 정기 학술대회, 183-193, 2019
- [6] 임문영, 박승범. 데이터 마이닝을 활용한 가짜뉴스의 선제적 대응을 위한 연구: M 온라인 커뮤니티 게시물을 중심으로. 한국 IT 서비스학회지, 18, 1, 219-234, 2019
- [7] 정다비치, 노영주. 빅데이터 분석을 활용한 가짜 리뷰 필터링 시스템 ADDAVICHI. 한국인터넷방송통신학회 논문지, 19, 6, 1-8, 2019.
- [8] 한철현, 권세빈, 박상근. KNN 기반의 가짜 리뷰 계정 분류 모델 및 서비스 개발. 한국소프트웨어공학 학술대회 논문집, 26, 1, 146-151, 2024.