카카오톡에서의 텍스트 데이터 마이닝 기반의 사용자별 적합 광고 키워드 도출

전예림 [1], 소다영 [2], 이지민 [1], 조은진 [3], 문지훈 [4]
[1] 순천향대학교 AI·빅데이터학과 학부생
[2] 순천향대학교 ICT 융합학과 석사과정
[3] 카카오 스토리개발팀 창작자앱개발파트
[4] 순천향대학교 AI·빅데이터학과 교수

{9261190, sodayeong, dlwlals7359, jmoon22}@sch.ac.kr, jinny.jo@kakaocorp.com

Extracting User-Specific Advertising Keywords Based on Textual Data Mining from KakaoTalk

Yerim Jeon[1], Dayeong So[2], Jimin Lee[1], Eunjin (Jinny) Jo[3], and Jihoon Moon[1,2]
[1] Department of AI and Big Data, Soonchunhyang University
[2] Department of ICT Convergence, Soonchunhyang University
[3] Story Development Team, Kakao Corp.

요 약

대화 데이터 기반 광고 추천은 광고 마케팅에서 고객 맞춤형 광고 제공, 마케팅 효과 극대화 등을 위한 중요한 기술로 주목받고 있다. 본 논문에서는 모바일 인스턴스 메신저인 카카오톡 대화창에서 발생한 텍스트 데이터를 기반으로 대화 내용을 분석하여 대화 주제별 적절한 광고 키워드를 제안한다. 이를 위해 주제별 대화 내용을 미용, 식음료, 상거래로 세분하고 KoNLPy 의 Okt 를 이용하여 텍스트 전처리를 수행하고 키워드별로 빈도수를 뽑아 워드 클라우드를 제시한다. 또한, 잠재 디리클레 할당(Latent Dirichlet Allocation, LDA)을 기반으로 대화 주제를 세분화한 뒤 라벨링을 통해 주제별 대화 키워드를 분석한다. 실험 결과, 대화 주제를 온라인 쇼핑, 헤어, 뷰티 관리, 음식으로 나눌 수 있었으며, 토픽별 상위 키워드를 Word2Vec 을 통해 특정 단어와 유사한 키워드를 도출하여 적절한 광고 키워드를 제시할 수 있었다.

1. 서 론

최근에 주목받는 광고매체는 이용자의 스마트 디바이스를 통해 언제 어디서든 광고 메시지를 쉽게 전달할 수 있는 모바일 광고이다[1]. 카카오톡(카톡)은 국민 대다수가 일상적으로 사용하고 있는 대표적인 모바일 메신저로, 기업들은 최적의 광고 성과를 도출하고자 카톡 채팅 상단에 있는 배너에 광고를 넣는다. 이뿐만 아니라, 최근에는 소셜 네트워크 서비스(Social Network Service, SNS) 자체적으로 쇼핑 플랫폼으로 발전하는 양상을 보인다[2]. 구체적으로, 카카오톡 선물하기 카테고리에서는 식료품과 전자기기, 뷰티와 명품 등을 쉽게 구매하고 선물할 수 있다.

본 논문은 기업의 마케팅 효과와 카카오톡 사용자의 맞춤형 광고를 제공하기 위해 대화 주제에 맞는 배너 광고를 제시하는 기법을 제안한다. 이를 위해, 카카오톡에서 발생한 텍스트 데이터를 분석하여 대화 주제별로 빈도가 높은 키워드를 제시한다.

2. 본 론

2.1 데이터 소개

본 논문의 실험을 위해 AI Hub 에서 제공하는 카카오톡에서 발생한 텍스트 데이터를 사용하였다. 이 중 대화 주제가 20 여 개인 데이터에서 미용, 식음료, 상거래만 선별하여 134,262 건을 분석에 활용하였다.

2.2 데이터 전처리

카카오톡에서 발생하는 텍스트 데이터는 의성어와 이모티콘, 줄임말 등이 빈번히 등장하는 특징이 있다. 이러한 데이터에서 유의미한 텍스트를 추출하기 위해, Okt 사전에서 텍스트를 토큰으로 분할하고 불용어를 제거하였다. 다음으로 수치형 데이터를 제거하고, 각 단어의 빈도를 분석하였다.

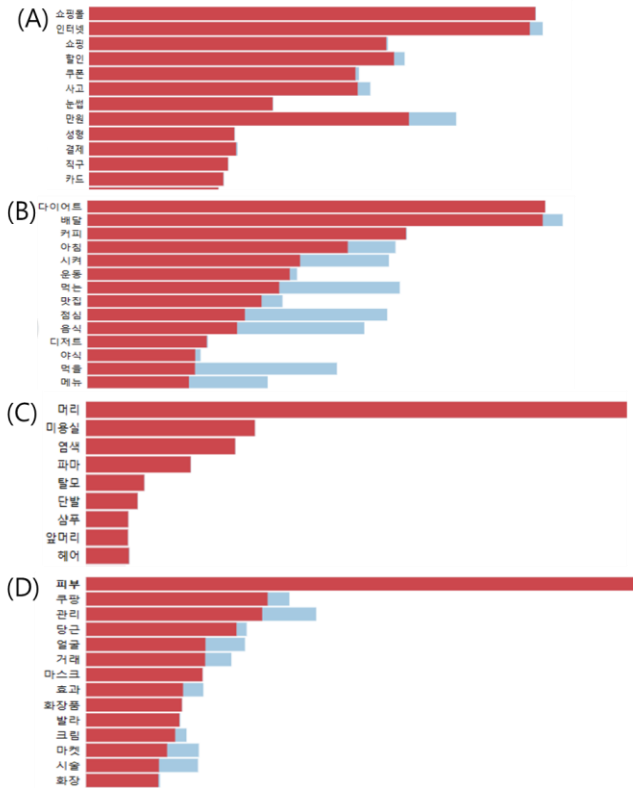
(a) 미용 (b) 식음료 (c) 상거래

(그림 1) 주제별 워드 클라우드 시각화

그림 1 은 미용, 식음료, 상거래 주제에서 발생한 텍스트 데이터를 워드 클라우드로 시각화한 것이다. 미용은 주로 다이어트, 헬스장, 요가와 같은 운동에 관한 키워드가 상대적으로 높은 빈도로 언급되었으며, 식음료 주제는 대중적인 인스턴트 음식인 햄버거나 치킨과 같은 키워드가 주로 등장하였다. 마지막으로, 상거래 주제에서는 워치와 함께 세일, 할부, 옷 등의 키워드가 높은 빈도로 언급되었다.

2.3 분석 및 결과

워드 클라우드를 시각화한 결과, 대주제 안에 소분류의 대화 주제가 군집되어 있었으며, 이를 바탕으로 LDA (Latent Dirichlet Allocation) 기반 토픽 모델링을 수행하였다. 먼저, 2 회 미만으로 언급된 단어와 10% 이상의 빈도로 등장하는 단어는 제외하였으며, Topic Coherence 를 활용하여 토픽의 개수를 인터넷 쇼핑, 헤어, 뷰티 관리, 음식 등 총 4 개로 결정하였다.



(그림 2) 주제별 토픽 모델링 시각화

그림 2(A)에서는 쇼핑몰, 인터넷, 쇼핑, 할인 등의 키워드가 추출되었으며, 이는 인터넷 쇼핑과 관련된 대화 주제로 유추할 수 있다. (B)에서는 다이어트, 배달, 커피, 아침, 운동, 음식, 야식 등의 키워드가 추출되었으며, 이는 음식에 관한 대화 주제로 유추할 수 있다. 또한, (C)에서는 머리, 미용실, 염색, 파마, 탈모 등의 키워드가 추출되었으며, 이는 헤어 관련 대화 주제로 유추할 수 있다. 마지막으로 (D)에서는 피부, 쿠팡, 관리, 당근, 얼굴, 화장품, 크림, 시술, 여드름 등의 키워드가 추출되어, 이는 뷰티 관리에 관한 대화 주제로 유추할 수 있다.

또한, 대화 주제에 적합한 광고 키워드를 추천하기 위해 Word2Vec 을 활용하였다. 이를 위해, LDA 결과에서 도출한 4 가지 주제의 상위 키워드를 기반으로 Word2Vec 를 통해 추천 키워드를 분석하였다. 먼저, 인터넷 쇼핑 주제는 온라인, 인스타그램, 구매 등과 같은 단어가 추출되었으며, 음식 주제에서는 운동, 조절, 탄수화물 등의 단어가 높은 유사성을 보였다. 헤어 주제에서는 스타일, 펌, 트렌디, 패셔너블 등이 추출되었으며, 뷰티 관리 주제에서는 여드름, 트러블, 주름, 모공 등과 같은 단어가 유사하게 추출되었다. 이러한 추천 키워드를 기반으로 광고주는 제품이나 서비스를 효과적으로 홍보할 수 있다.

3. 결 론

본 논문은 카카오톡에서 발생한 텍스트 데이터를 기반으로 대화 주제별로 도출된 키워드를 활용하여 워드 클라우드를 시각화하였다. 또한, LDA 모델을 활용하여 주제별 상위 키워드를 추출 및 Word2Vec 를 기반으로 적합한 광고 키워드를 도출하여 사용자별 맞춤형 광고 전략을 제시하였다. 이를 통해 기업은 효과적인 타겟팅을 수행하여 마케팅 효율성을 높일 수 있다. 향후 기업들의 성공적인 마케팅 전략 수립을 위해 자연어 처리를 활용하여 마케팅 분야에서 적합한 광고 키워드를 제공할 예정이다.

사 사 문 구

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW 중심대학사업의 연구 결과로 수행되었음(2021-0-01399).

참 고 문 헌

[1] J.-S. Bae, “Structural Relationship between Mobile Advertisement Characteristics, Personal Characteristics, Purchasing Attitude and Purchasing Behavior -Focus on Youth Consumers-,” *Journal of the Korea Contents Association*, Vol. 20, No. 5, pp. 303–317, 2020.

[2] J. H. Na and W. C. Lee, “What Makes Top Influencers Different?: Content Analysis and Text Mining of 1-person Market Instagram,” *Journal of Practical Research in Advertising and Public Relations*, Vol. 16, No. 1, pp. 64–96, 2023.