

## 단어 임베딩 모델 기반 캡티브 포털 메뉴 추천 시스템

여동훈<sup>○</sup>, 황병일\*, 김동주\*

<sup>○</sup>포항공과대학교 인공지능연구원,

\*포항공과대학교 인공지능연구원

e-mail: tgc01098@postech.ac.kr<sup>○</sup>, {bihwang, kkb0320}@postech.ac.kr\*

## Captive Portal Recommendation System Based on Word Embedding Model

Dong-Hun Yeo<sup>○</sup>, Byung-Il Hwang\*, Dong-Ju Kim\*

<sup>○</sup>Postech Institute of Artificial Intelligence, Pohang University of Science and Technology,

\*Postech Institute of Artificial Intelligence, Pohang University of Science and Technology

### ● 요약 ●

본 논문에서는 상점 내 캡티브 포털을 활용하여 수집된 주문 정보 데이터를 바탕으로 사용자가 선호하는 메뉴를 추천하는 시스템을 제안한다. 이 시스템은 식품 관련 공공 데이터셋으로 학습된 단어 임베딩 모델(Word Embedding Model)로 메뉴명을 벡터화하여 그와 유사한 벡터를 가지는 메뉴를 추천한다. 이 기법은 캡티브 포털에서 수집되는 데이터 특성상 사용자의 개인정보가 비식별화 되고 선택 항목에 대한 정보도 제한되므로 기존의 단어 임베딩 모델을 추천 시스템에 적용하는 경우에 비해 유리하다. 본 논문에서는 실제 동일한 시스템을 사용하는 상점들의 구매 기록 데이터를 활용한 검증 데이터를 확보하여 제안된 추천 시스템이 Precision@k(k=3) 구매 예측에 유의미함을 보인다.

**키워드:** 캡티브 포털(captive portal), 단어 임베딩(word embedding), 추천 시스템(recommendation system)

## I. Introduction

캡티브 포털은 공용 네트워크에서 사용자를 식별하고 접근 권한을 부여하기 위해 사용되는 장치이며 외부에서 접속이 불가능하여 실제 신호 범위 내에 접근하는 사람들을 위한 서비스 구축에 용이하다. 현재 KTX 열차와 스타벅스 등 다양한 분야에서 해당 서비스를 활용하고 있다. 만약 상권 내에서 캡티브 포털을 활용한다면 매장을 직접 방문한 사용자만을 위한 주문 예약 시스템을 제공할 수 있다[1].

본 논문에서는 상권 내 캡티브 포털을 통해 비식별화된 사용자의 주문 기록을 수집 및 활용하여 사용자와 메뉴에 대한 사전 정보가 제한될 때 메뉴명을 기반으로 메뉴를 추천하는 시스템을 제안한다. 그리고 제안된 추천 시스템의 성능 확인을 위해 실제 상점들의 구매 기록 데이터와 구매 예측 결과를 비교하고자 한다.

존재한다[2]. 협업 필터링은 사용자나 아이템에 대한 선택과 선호도 데이터를 활용하여 항목을 추천한다. 이 기법은 사용자나 아이템에 대한 명시적인 정보 없이도 추천이 이루어질 수 있으나 충분한 사전 데이터가 존재하지 않는다면 Cold-Start 문제가 발생할 수 있다. 그리고 콘텐츠 기반 필터링은 선택과 선호도 정보가 아닌 아이템의 특성을 분석하여 각 사용자에 대한 프로파일을 만들고 시스템과의 상호작용 정보를 통하여 추천을 수행한다. 이 경우 다른 사용자의 데이터에 대한 영향은 적지만 아이템에 대한 정보가 부족한 경우 적합하지 않은 추천이 이루어질 수 있다. 캡티브 포털은 통해 수집된 비식별화 사용자 주문 기록은 충분한 사전 데이터와 아이템의 특성을 나타내는 정보가 존재하지 않는다. 이를 해결하기 위해 단어 임베딩 모델을 활용한 메뉴 추천 시스템을 제안한다.

기존의 단어 임베딩 모델은 문장 내 토큰을 임의의 벡터 공간으로 투영하며 Transformer 구조 기반의 BERT로부터 파생된 다양한 모델이 존재한다[3]. 그 중 KoBERT 모델과 본 연구에서 자체적으로 학습시킨 단어 임베딩 모델을 비교한다.

## II. Preliminaries

일반적으로 추천 시스템에는 협업 필터링(Collaborative Filtering)과 콘텐츠 기반 필터링(Content-based Filtering) 추천 시스템이

### III. The Proposed Scheme

본 논문에서 활용한 단어 임베딩 모델은 공공 데이터셋인 전국통합 식품영양성분정보 표준데이터를 기반으로 학습되었다. 먼저 데이터셋에 Keras 토큰라이저를 활용하여 식품명을 의미 단위로 분리 및 불필요한 정보를 제거하는 전처리를 수행하였다. 그리고 단어 임베딩 모델을 설계하여 식품명을 식품 카테고리라 성분 정보로 벡터화할 수 있도록 학습시켰다. 해당 모델을 통해 입력된 메뉴명을 벡터화하고 다른 메뉴와 코사인 유사도를 비교하여 높은 유사도 순으로 추천한다. Fig. 1. 은 제안된 추천 시스템의 구성도이다.

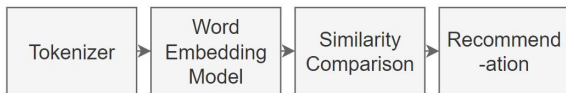


Fig. 1. Proposed Recommendation System

모델이 형성하는 임의의 벡터 공간으로 학습이 이루어지는 기존 모델에 비하여 본 연구에서 학습된 모델은 명시적인 벡터 공간을 형성하여 항목에 대하여 설명가능한 벡터로 사용자가 이해하는 메뉴명과 유사한 Feature를 갖도록 하였다.

추천 시스템의 검증에 위하여 향후 수집될 캡티브 포털의 구매 기록과 유사한 전국 휴게소에서의 직원 구매 기록을 확보하였다. 이 데이터는 5백 개의 메뉴에 대하여 1천5백 명의 사용자의 구매 기록 3천 건으로 구성되어 있으며 이 중 메뉴 추천을 위해 2개 이상의 휴게소에서 구매 기록이 있는 사용자 데이터만 활용하였다.

Table 1. Results by Embedding Model

Embedding Model	Precision@k(k=3)
KoBERT	0.12
Our Model	0.45

KoBERT 모델과 본 연구에서의 개발한 모델에 기존 구매 기록을 입력하여 도출되는 추천 메뉴와 실제 구매 기록 간의 Precision@k(k=3)를 측정하였다. Table 1. 과 같이 KoBERT 모델보다 본 연구에서 학습된 모델이 0.33 높은 수치를 보이며 이는 메뉴명 기반의 카테고리화 하는 모델이 단순 단어 임베딩 모델보다 메뉴 추천에 적합함을 의미한다. 제안된 추천 시스템은 일반적으로 활용되는 추천 시스템의 성능에 비해 낮은 수치를 보이지만 메뉴명이 사용자의 선택과 상관관계가 있음을 시사한다.

### IV. Conclusions

캡티브 포털 내 비식별화된 데이터에서 얻을 수 있는 정보는 상당 부분 제한되어 있다. 해당 데이터를 활용하기 위해 본 연구는 식품명을 식품 카테고리라 성분 정보로 벡터화하는 임베딩 모델의 개발을 통해 추천 시스템을 구성하였다. 제안된 추천 시스템은 입력되는 정보가 한정적인 경우 단독으로 활용할 수도 있으며 향후 메뉴명의 다른 정보와 함께 콘텐츠 기반 필터링을 고도화하고 사용자의

선택과 선호도 정보를 결합한 하이브리드 시스템을 개발한다면 더욱 개인화된 메뉴 추천 시스템을 구현할 수 있을 것이라 기대된다.

## ACKNOWLEDGEMENT

이 논문은 2023년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업(No.2022R1A6A1A03052954)이며, 2023년도 정부(과학기술정보통신부)의 재원으로 정보통신산업진흥원(S0442-23-1002, 캡티브 포털을 활용한 지능형 상권 서비스 개발 및 사업화)의 지원을 받아 수행된 연구임

## REFERENCES

- [1] Ali et al. "On privacy risks of public wifi captive portals," Data Privacy Management, Cryptocurrencies and Blockchain Technology: ESORICS 2019 International Workshops, DPM 2019 and CBT 2019, September 2019.
- [2] Isinkaye, F. Olubusola, Y. O. Folajimi, and B. A. Ojokoh, "Recommendation systems: Principles, methods and evaluation," Egyptian informatics journal, Vol. 16, No. 3, pp. 261-273, 2015.
- [3] Devlin, Jacob, et al, "Bert: Pre-training of deep bidirectional transformers for language understanding," arXiv preprint arXiv:1810.04805, 2018.