

딥러닝 모델을 활용한 관광지 활동 정보 공유 애플리케이션

신지호, 권은혜, 류병욱, 이병정
서울시립대학교 컴퓨터과학부

sjm010529@gmail.com, enehyejjang@gmail.com, rk97989@naver.com, bjlee@uos.ac.kr

An Application for Sharing Travel Activities Information by Using Deep Learning Models

Jiho Shin, Eunhye Gwon, Byungook Ryu, Byungjeong Lee
Dept. of Computer Science and Engineering, University of Seoul

요 약

일반적인 여행 커뮤니티는 사진과 텍스트 기반의 사용자 리뷰를 바탕으로 정보 공유를 한다. 본 연구에서는 관광지에서 수행한 활동을 한 문장의 형태로 공유하는 애플리케이션을 제안한다. ChatGPT를 활용하여 활동을 산책, 사진, 음식 등 9가지 태그로 분류하여 관광지가 가지는 특징을 용이하게 파악한다. 또한, 사용자가 작성한 활동을 임베딩하고 관광지 소개 글 벡터와 유사도를 비교하여 관광지를 추천한다. 본 애플리케이션을 통해 사용자가 긴 설명이나 사진 없이 관광지가 가지는 정보를 쉽게 공유하고 관광지 추천을 하는 새로운 여행 커뮤니티를 제공할 수 있을 것으로 기대한다.

1. 서론

국내 여행 분야에서 현재까지 사용되는 관광지 추천 서비스는 주로 방문자 수와 관련된 후기를 기반으로 한다[1]. 따라서 이미 많은 관광객이 다녀간 장소들이 주로 추천되는 반면, 알려지지 않은 숨은 관광지들은 방문자 수와 후기 부족으로 인해 무시되는 경향이 존재한다. 이런 상황에서 본 연구는 한국관광공사의 Tour API를 활용하여 사용자들이 새로운 관광지와 다양한 여행 경험을 찾을 수 있도록 하는 서비스를 제안한다. 본 연구의 목표는 사용자의 관심과 취향에 따라 여행지를 추천하고, 선택한 여행지에 대한 활동 목록을 제공함으로써 사용자들이 더 풍부하고 맞춤형 여행 경험을 즐길 수 있도록 하는 것이다. 이를 통해 관광객들은 알려지지 않은 관광지를 쉽게 찾아갈 수 있고, 새로운 경험을 통해 더 풍부한 여행을 즐길 수 있도록 하고자 한다.

2. 관련 연구

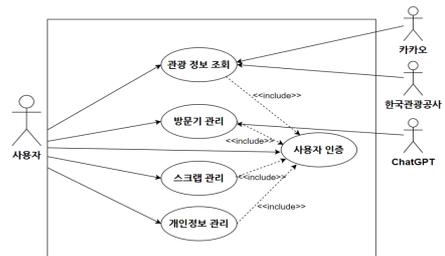
ChatGPT(Generative Pre-trained Transformer)는 거대 자연어 처리 모델인 GPT3, GPT4를 기반으로 하는 대화형 인공지능 서비스이다[2]. GPT3는 1,750억 개의 파라미터로 in-context-learning 방식으로 훈련하였기 때문에 추가 학습 없이 범용 자연어를 처리할 수 있다. 다만 한국어 학습 비중이 작고 원하는 형식의 출력을 얻고 싶다면 파인 튜닝 과정이 필요하다. OpenAI는 GPT3.5 모델의 파인 튜닝을 API로 지원한다. 본 연구에서는 관광지 활동 정보를 9가지의 태그로 분류하는 GPT3.5 모델을 파인 튜닝하고, API로 접근하여 사용한다.

3. 딥러닝 기반 관광지 활동 정보 공유

3.1 시스템 개요

그림 1은 본 연구에서 개발하는 앱 서비스의 유스케이스도이다. 본 서비스는 사용자가 자신의 취향에 맞는 관광지 정보를 카카오톡과 한국관광공사 API를 활용하여 조회한다. 조회한 방문지를 사용자가 방문한 후 사용자는 해당 관광지에 대한 방문 후기를 작성할 수 있다. 사용자의 방문 후기 작성을 보조하기 위해

생성 AI 모델인 ChatGPT를 활용한다. 또한 사용자는 각 관광지 및 방문 후기를 스크랩하여 관광지 정보를 쉽게 찾아볼 수 있다.



(그림 1) 유스케이스도

3.2 추천 시스템

사용자는 애플리케이션 메인화면에서 관광지를 추천받는다. 관광지를 추천하기 위해 사전에, 한국관광공사에서 제공하는 쇼핑, 관광지, 문화시설, 레포츠 타입의 관광지 개요 28,052개를 활용한다. 데이터셋은 한국관광공사의 SPARQL Endpoint로 질의를 작성하여 관광지 이름, ID, 개요로 분류된 CSV 파일을 얻는다. pandas 패키지로 빈 데이터를 제외하거나, 데이터를 가공한다. 가공된 데이터는 ChatGPT가 제공하는 API로 text-embedding-ada-002 모델을 사용하여 워드 임베딩한다[3]. 관광지 추천 요청이 오면 사용자의 관광지에서 수행한 활동을 불러와 동일하게 워드 임베딩하여 관광지 개요와 코사인 유사도를 비교한다. Top-5로 1등부터 5등까지 추천 관광지를 선정한다. 결과적으로 사용자가 작성한 관광지에서 수행한 활동 내용과 관광지 개요가 유사한 장소가 추천된다. 그림 2는 "바다에서 스노클링하기"를 입력값으로 얻은 추천 관광지이다. 클립씨피싱을 제외한 4곳은 모두 스노클링을 하는 관광지이다.

```

name resource
5429 "판포포구" <http://data.visitkorea.or.kr/resource/2785869>
18441 "드림카약투어" <http://data.visitkorea.or.kr/resource/2782810>
4883 "등산항" <http://data.visitkorea.or.kr/resource/2752126>
17759 "클립씨피싱" <http://data.visitkorea.or.kr/resource/2769653>
3406 "신홍해수욕장" <http://data.visitkorea.or.kr/resource/129542>
INFO: 127.0.0.1:60977 - "GET /recommand HTTP/1.1" 200 OK
    
```

(그림 2) 추천 시스템 결과 예시

3.3 관광지 활동 태그 분류

```

1 {"role": "system", "content": "너는 다음과 같은 분류 기준에 따라 1~9까지 숫자 중 하나로 대답해야 해.
2 (1) [사진] 어디에서/무엇과(들) 사진찍기, 포착하기
3 (2) [체험] 무엇을 만들기, 어떤 프로그램 참여하기, 어떤 것 타기
4 (3) [산책] 어디를 걷기, 산책하기
5 (4) [운동] 어디를 뛰기, 완주하기, 배드민턴 치기
6 (5) [관람] 어디를 보기
7 (6) [음식] 어떤 것 먹기
8 (7) [힐링] 어디에서 쉬기, 얘기하기
9 (8) [지식] 무엇에 대해 알기
10 (9) [쇼핑] 어떤 기념품 사기"},
11 {"role": "user", "content": "todo: 광안대교의 야경 뷰와 함께 음료 하나 들고 산책하기 \n ->"},
12 {"role": "assistant", "content": "3"}
    
```

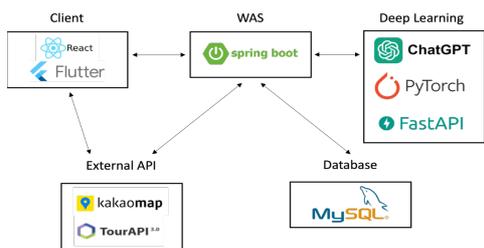
(그림 3) 태그 분류 학습 데이터

사용자가 관광지에서 수행한 활동을 작성하면 내용에 따라 (1)사진, (2)체험, (3)산책, (4)운동, (5)관람, (6)음식, (7)힐링, (8)지식, (9)쇼핑으로 태그 분류를 한다. 추천 수를 많이 받은 활동 태그는 그 관광지를 대표하는 태그가 된다. 태그 분류를 위해 gpt-3.5-turbo 모델을 파인 튜닝한다. 그림 3은 ChatGPT를 파인 튜닝하기 위한 Jsonl 타입의 학습 데이터의 한 예시이다. 각 학습 데이터는 system, user, assistant로 구성된다. user는 ChatGPT에 질의하는 내용으로 사용자가 입력한 관광지에서 한 활동이다. "\n ->"처럼 구분자를 사용하면 파인 튜닝 후 ChatGPT에 질의할 때 더욱 정확한 값을 얻을 수 있다. assistant는 기대하는 ChatGPT의 답변이다. 관광지에서 수행한 활동을 직접 분류하여 1~9 사이 정수로 채운다. system은 assistant의 값을 정할 때 참고하는 데이터로서 대답 형식과 분류 기준을 정의한다. 약 100개의 데이터를 ChatGPT에서 제공하는 API로 파인 튜닝한다. 결과적으로 같은 내용의 프롬프트를 작성하여 ChatGPT에 답을 얻는 것보다 파인 튜닝으로 두 가지 이점을 갖는다. 첫 번째로 파인 튜닝 한 모델에 "\n ->" 구분자를 넣고 질의하면 " 7"과 같이 "(1~9 사이 정수)" 고정된 형식의 답변을 얻는다. ChatGPT의 답변을 별다른 데이터 처리 없이 DB에 저장할 수 있다. 두 번째로 그림 3에서 system이 분류 기준을 설명하고 있지만 "OO 느끼기"처럼 분류 기준에 없는 데이터는 분류하기 어려울 수 있다. 파인 튜닝을 통해 본 시스템에 더 적합한 ChatGPT를 만들 수 있다.

4. 시스템 구현

4.1 시스템 구성

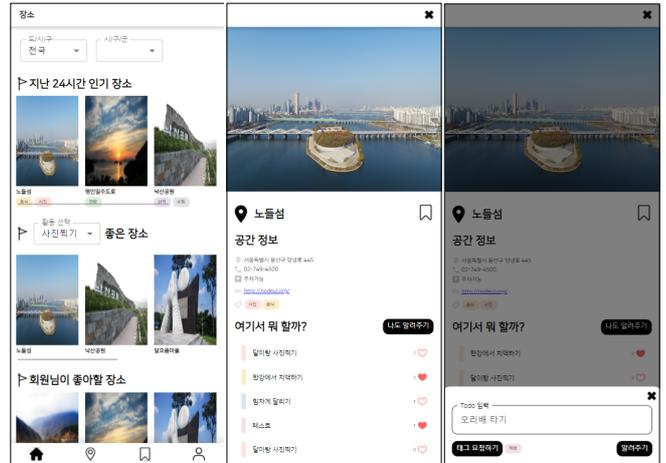
그림 4는 본 연구의 시스템 구성도이다. 사용자 인터페이스는 React로 개발한 웹 뷰 형태이며 Flutter를 이용해 모바일 애플리케이션으로 구동되도록 한다. WAS(Web Application Server)는 DB와 연결되어 사용자, 게시글 등 데이터를 관리한다. 관광지와 관련된 모든 정보는 DB에 따로 저장하지 않고 한국관광공사에서 제공하는 Tour API를 사용하여 제공한다. 또한, 카카오 맵 API는 관광지의 위치정보를 클라이언트에게 제공한다. 딥러닝 시스템은 ChatGPT API와 Python 기반의 Fast API를 사용하여 관광지 추천과 관광지 활동 태그 분류 서비스를 제공한다. 관광지 추천은 Python 코드로 작성 후 Fast API로 래핑하여 Python 환경의 서버를 구축하여 API로 통신한다. 관광지에서 수행한 활동의 태그 분류는 ChatGPT API로 결과를 받아 클라이언트에게 제공한다.



(그림 4) 시스템 구성도

본 시스템의 소스 코드는 깃허브에서 확인할 수 있다.¹⁾

4.2 UI 구현



(그림 5) 홈 (그림 6) 관광지 상세 (그림 7) 태그 분류

그림 5-7은 본 애플리케이션의 UI 일부이다. 그림 5와 같이 홈 화면에서 사용자에게 맞게 추천되는 관광지 목록을 확인할 수 있다. 원하는 관광지를 선택하면, 그림 6과 같이 해당 관광지 상세 화면으로 이동한다. 해당 관광지에서 할 수 있는 활동 목록들을 통해 어떤 관광지인지 쉽게 파악할 수 있다. '나도 알려주기' 버튼을 누르면, 그림 7과 같이 직접 입력하여 관광지에서 할 수 있는 활동을 공유할 수 있다. 입력받은 활동 내용에 맞춰서 자동으로 태그로 분류된다.

5. 결론

본 연구의 애플리케이션을 통해 사용자는 방문한 관광지에서 할 수 있는 활동을 다른 사용자들과 공유할 수 있다. 기존 유명 플랫폼에서 한국관광공사의 관광지 목록들의 후기 게시판에는 다양한 내용이 없었다. 하지만 본 연구에서 태그 분류를 통해 구체적인 관광지 특징을 제공하여 사용자는 관광지와 관광지에서 수행할 수 있는 활동을 더 잘 이해할 수 있을 것으로 기대한다.

참고문헌

[1] 이종민, 권영준, 김울, 김경석, 장재준, 강현규, "딥러닝 예측 알고리즘 기반의 맞춤형 추천 모바일 앱 '눈치코칭_여행딥러닝' 개발," 한국정보과학회 연어공학연구회 학술대회논문집, pp. 498-503, 2018.
 [2] Tom Brown et. al. "Language models are few-shot learners," In Proc. of 34th Conference on Neural Information Processing Systems, 2020.
 [3] Recommender System with Python, "https://github.com/ljsjsj92/recommender_system_with_Python."

1) https://github.com/tripcube