

드롭아웃-넷과 하버사인 공식을 활용한 위케이션 관광지, 업무공간 추천 서비스 개발에 대한 연구

김도연, 박지영, 신재용, 양슬빈, 이도희, 정성훈
명지대학교 ICT 융합대학 융합소프트웨어학부 데이터테크놀로지전공

초록

본 연구에서는 위케이션에 특화된 관광지와 업무공간 위케이션 관광지, 업무공간 추천 웹 서비스를 개발하고자 한다. 초기 유저의 정보가 없는 상태에서도 적절한 추천 기능을 제공하기 위해 드롭아웃-넷(Dropout-Net)을 사용하고 관광지와 가까운 업무공간 추천을 위해 하버사인 공식(Haversine Formula)을 활용한다. 실험 결과 해당 추천 서비스의 드롭아웃-넷 성능은 Recall@100 기준의 목표인 0.7을 능가한 0.9를 기록했다. 이를 통해 해당 연구가 위케이션 니즈가 있는 대상자들이 만족할 수 있는 위케이션 장소와 업무공간 추천 기능을 제공한다. 이는 본 연구의 개발이 성공적으로 진행되었음을 나타낸다.

1. 서론

1.1 배경

2019년 12월에 시작된 COVID-19로 인해 언택트(Untact) 사회가 빠르게 발전했고 팬데믹이 종식되어가면서 관광 시장이 회복되기 시작했다. 이러한 영향은 원격근무에 대한 기술의 발전과 복지 니즈의 증가로 이어졌고, 확산된 원격근무 트렌드는 일시적 현상이 아닌 뉴노멀(new normal)화 되었다. 또한 관광수요가 급증하면서 관광지 중심의 체류형 관광이 트렌드로 부상하게 되었고 장기간 여행의 니즈와 수요가 지속적으로 상승했다[1].

이러한 변화로부터 위케이션(workation)이 등장하게 되었다. 위케이션이란 ‘일(work)’과 ‘휴가(vacation)’의 합성어로, 근로자가 휴가지에서 일상적인 업무를 수행하면서 관광, 휴양을 동시에 실시하는 것을 의미한다. 위케이션이 등장함에 따라 지자체와 기업은 각각 인구감소로 침체된 지방 경제 활성화와 근로자의 니즈 반영,

인재 확보라는 목적을 위해 이를 도입하기 시작했다.

위케이션을 위해서는 근로자가 관광과 휴양을 즐길 수 있는 관광지와 업무에 집중할 수 있는 업무공간이 필요하다. 하지만 기존의 ‘마이로[2]’와 ‘여행콧콧[3]’ 같은 국내 여행지 추천 서비스는 위케이션을 위한 서비스로서는 문제점을 지닌다. 마이로는 직접 여행 계획을 작성하여 등록된 여행 장소가 정렬되어 선택할 수 있게만 제공될 뿐 여행 장소의 맞춤형 추천은 되지 않고 있다. 또한 마이로와 여행콧콧 모두 관광지에 대한 장소만 제시할 뿐 업무공간 제시는 불가능했다. 그리고 ‘디어먼데이[4]’, ‘데스커 위케이션[5]’, ‘달렘 웰니스 위케이션[6]’ 등의 위케이션 플랫폼은 제공된 업무 공간에서 위케이션 일정을 계획할 수 있는 서비스를 제공한다. 하지만 한정된 업무 공간과 기간 제한이 있었고 관광지 추천 서비스는 존재하지 않았다. 이는 근로자의 위케이션 특성 니즈의 미반영과 변동 가능성이 있는 일정에 대한 위케이션 계획 수립의 어려움이란 문제를 초래한다.

따라서 본 연구에서는 이러한 문제를 해결하기 위해 위케이션을 위한 관광지와 업무공간을 동시에 추천해주는 서비스를 개발한다. 또한 일하면서 휴가를 즐기는 사람들이 증가하고 있으므로 본 연구의 목표를 사용자 개인의 여행 스타일과 자유로운 기간 설정을 고려한 관광지, 업무공간 추천 서비스 제공으로 설정한다. 이를 통해 전국 장소에 대한 관광지와 업무공간을 사용자 맞춤형으로 추천하며 기간적 한계점의 대안을 제시한다.

목표 달성을 위해 세부 목표를 사용자 맞춤형 관광지 추천 기능 개발과 관광지와 가까운 업무공간 추천 기능 개발로 세우고자 한다. 각 세부 목표를 이루기 위해 드롭아웃-넷을 기반으로 관광지 장소를 추천하고, 하버사인 공식으로 업무공간을 탐색해 제안한다. 또한 해당 서비스는 웹 애플리케이션 UI 프레임워크 Vue.js, 서버사이드 프레임워크 Express.js, 데이터베이스 관리 시스템

템 MySQL 을 기반으로 구현되어 사용자에게 제공된다.

2. 관련 연구

2.1 추천 시스템

추천 시스템은 사용자에게 맞춤형으로 상품, 서비스, 또는 콘텐츠를 추천하는 시스템으로, 사용자의 선호 및 행동 패턴을 분석하여 최적의 추천을 제공한다. 추천 시스템에 필요한 데이터는 사용자와 아이템 간의 상호작용 정보로 구성된다. 사용자의 프로필, 이력, 평가, 구매 이력 등과 아이템의 속성, 이미지, 평가, 리뷰 등이 포함된다. 이 데이터를 종합하여 사용자에게 맞춤형 추천을 제공한다. 그리고 추천 방식의 기준과 특성에 따라 콘텐츠 기반 추천(Content-Based Recommendations)과 협력 필터링 방법(Collaborative Filtering) 이 있다. 주로 사용되는 추천 알고리즘에는 콘텐츠 기반 필터링(Content-Based Filtering), 협업 필터링(Collaborative Filtering), 그리고 이 둘을 결합한 하이브리드 방식이 있다.

콘텐츠 기반 필터링은 아이템 자체의 속성을 기반으로 사용자의 선호를 파악하고 유사한 아이템을 추천한다. 예를 들어, 영화 추천에서 장르, 감독, 배우 등의 특성을 고려하여 추천할 수 있다. 협업 필터링은 사용자 간의 유사성을 분석하여 아이템을 추천한다. 사용자 기반 협업 필터링은 유사한 취향을 가진 사용자가 선호한 아이템을 추천하고, 아이템 기반 협업 필터링은 유사한 아이템을 기반으로 추천을 수행한다. 하이브리드 방식은 콘텐츠 기반 필터링과 협업 필터링을 결합하여 두 알고리즘의 장점을 활용하는 방식이다. 이를 통해 추천의 다양성과 정확도를 높일 수 있다.

2.2 잠재 인수 모델(Latent Factor Model)

잠재 인수 모델은 의미처럼 정보 또는 벡터를 잠재적인 인수들로 표현하는 방법이다. 잠재 요인 모델(Latent Factor Model)은 추천 시스템에서 사용되는 통계적 모델 중 하나로, 사용자와 아이템 간의 상호작용을 잠재 요인을 통해 모델링 한다. 이 모델은 데이터의 차원을 축소하여 사용자 및 아이템을 잠재적인(latent) 요인으로 표현하여 예측을 수행한다.

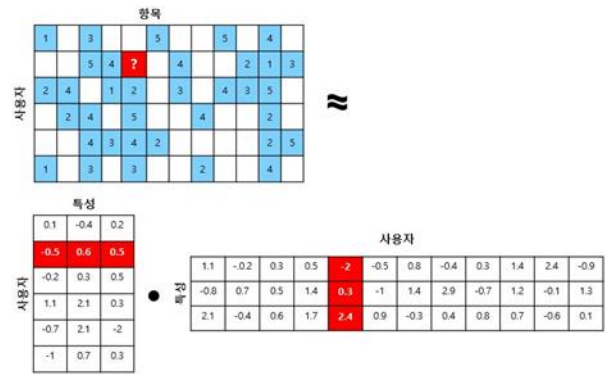


그림 1 : 잠재 인수 모델

[그림 1]은 잠재 인수 모델 예시 그림이다. 추천 시스템에서는 사용자와 항목을 특성화 함으로써 벡터로 표현한다. 따라서 사용자들이 상품에 부여한 평점이 있는 행렬이 있을 때 해당 행렬을 각각 사용자와 항목에 대한 행렬로 분해한다. 기본적으로 잠재 요인 모델은 사용자와 아이템을 특정한 차원의 벡터로 표현하고, 이 벡터들의 내적을 통해 평점 또는 상호작용을 예측한다.

예를 들어, 사용자 u 의 벡터를 P_u , 아이템 i 의 벡터를 Q_i 로 표현하면 해당 사용자가 해당 아이템에 대해 예측하는 평점은 $P_u \cdot Q_i$ 가 된다. 해당 방법은 선형대수학의 특이 값 분해(SVD, Singular Value Decomposition)와 유사한 방식으로 사용자-항목 평가 행렬의 특성에 맞게 조정된다. 이를 통해 최종적으로 사용자의 잠재적인 특성들이 반영된 사용자 행렬과 항목에 대한 항목 행렬을 얻게 된다.

2.3 Cold-Start

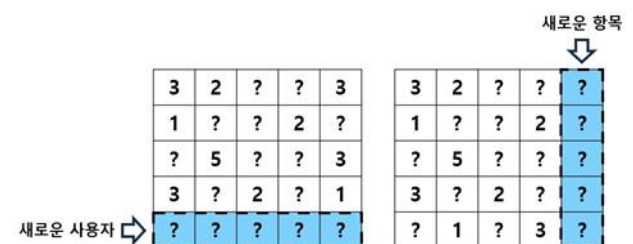


그림 2 : Cold-Start 예시

Cold-Start 문제는 추천시스템을 진행할 때 사용자와 아이템에 대한 정보의 부족으로 신뢰할 수 있는 추천을 할 수 없는 문제를 말한다. [그림-2]와 같이 Cold Start 문제는 추천 시스템에서 새로운 사용자나 아이템이 도입될 때 발생하는 문제를 나타낸다. 이러한 새로운 사용자

자나 아이템은 시스템이 이들에 대한 충분한 이력 정보를 갖고 있지 않아 정확한 추천을 수행하기 어렵다. 이러한 문제들을 해결하기 위해 다양한 방법을 사용해 추천 시스템을 구축하는 것이 중요하다. 이에 대한 해결책으로는 콘텐츠 기반 필터링 방식과 협력 필터링 방식을 결합한 하이브리드 방식이나 평가 점수 외 추가 정보를 활용하는 방법들이 있다. 이러한 접근법을 통해 Cold-Start 문제를 완화하고 사용자 경험을 향상시키는 데 기여할 수 있다.

3. 데이터

3.1 데이터셋

본 연구의 핵심 기능인 위케이션 관광지, 업무공간 추천 기능을 구현하기 위해 두 가지 데이터셋을 사용한다. 첫 번째 데이터셋은 관광지 추천을 위한 ‘국내 여행 로그 데이터[7]’이다. 해당 데이터셋은 국내 AI 인프라 플랫폼 AI Hub 에서 제공되며 여행객, 여행 기록, 여행 장소 등과 같은 정보로 이루어져 있다. 그리고 여행 데이터를 수도권, 동부권, 서부권, 제주도 및 도서지역으로 나눠 각 권역별로 3,600 세트씩 총 14,400 세트의 여행 로그 데이터로 구성된다. 해당 데이터셋에서 여행객 ID, 성별, 연령대, 여행 스타일 등 여행객 데이터와, 방문지 ID, 방문지명, 주소, 만족도, 방문지 유형, 위도/경도 등 방문지 정보의 데이터를 수집한다.

두 번째 데이터셋은 업무공간 추천을 위한 ‘업무공간 데이터’이다. 해당 데이터셋은 kakao developers [8]에서 제공하는 지도/로컬에 대한 ‘키워드로 장소 검색하기’ API 를 사용하여 “스터디카페”와 “공유오피스”이라는 키워드를 기반으로 크롤링 한 데이터이다. 해당 키워드를 바탕으로 장소명, 장소 ID, 도로명 주소, 지번 주소, 연락처, 위도, 경도, URL 의 정보를 크롤링 한다. 이후 하버사인 공식을 활용하기 위한 추가 전처리를 진행한다.

3.2 데이터 전처리

데이터 전처리는 사용자의 여행 취향과 선호도를 정확하게 파악하고, 개인화된 서비스의 품질을 보장하고자

추천 알고리즘의 학습 및 테스트 데이터를 준비하는 핵심 단계이다. 따라서 각 알고리즘에 적합한 형태로 데이터 전처리를 진행한다.

먼저 국내 여행 로그 데이터는 관광지 추천을 위해 드롭아웃-넷이 요구하는 데이터 형태에 맞게 전처리한다. 기존 드롭아웃-넷 모델에서 사용하는 CITEULike 데이터셋의 학습, 검증 데이터의 구조와 동일하게 데이터를 가공한다. 기본적으로 적용하는 데이터 전처리는 다음과 같다: 첫 번째는 관광지 데이터 필터링이다. 이 과정을 통해 AI Hub 에서 제공하는 여행로그 데이터 중 관광지가 아닌 데이터는 제거한다. 추가로 이 과정에서 관광지로 적합하지 않다고 판단되는 장소 또한 함께 제거한다. 두 번째는 방문지에 대해 만족도가 없는 값을 삭제한다. 세 번째로 공백과 같이 의미 없는 문자열 패턴을 제거한다. 이렇게 세 가지 데이터 전처리를 진행한 후 드롭아웃-넷 학습을 위한 데이터 구조를 생성한다.

드롭아웃-넷 학습을 위해 생성된 새로운 데이터들은 크게 세 가지다. 첫 번째는 여행객과 관광지에 대한 평점 행렬 데이터이다. 평점 행렬은 각 여행객이 방문 후 평가한 모든 관광지에 대한 평가 점수가 할당된 행렬이다. 따라서 여행객 ID, 관광지 ID, 평가점수 열로 구성된다. 해당 데이터셋은 8:2 비율로 훈련 데이터셋과 검증 데이터셋으로 나눈다. 두 번째는 여행객과 관광지에 대한 잠재 벡터(Latent Vector) 행렬 데이터다. 해당 데이터는 여행객-관광지 만족도 행렬 데이터에 교대 최소 제곱법(Alternating Least Squares, ALS)를 기반으로 Matrix Factorization 을 적용해 구해진 데이터다. 이는 여행객과 관광지 정보를 잠재적으로 표현한 데이터이다. 마지막 세 번째는 여행객 추가 정보에 대한 libsvm 포맷의 데이터이다. Libsvm 포맷 데이터는 범주형 변수를 표현하기에 적합한 포맷이다. 해당 포맷을 사용해 여행객의 ID, 성별, 나이, 여행 스타일에 대한 정보를 변환한다. 이러한 전처리 과정을 거친 데이터들은 드롭아웃-넷의 모델 학습과 전체 성능 평가를 위한 입력값으로 사용한다.

업무공간 데이터는 크롤링 된 데이터를 활용하여 하

버사인 알고리즘을 적용하기 위해 전처리를 진행한다. 각 장소의 주소를 도, 시/군으로 그룹화하여 각각 새로운 열에 추가한다.

추가적으로 웹페이지에서 사용자에게 보일 관광지의 사진은 구글에서 크롤링 되었으며, 결과에 부합하지 않는 사진은 적합한 이미지로 대체되었다. 이를 통해 사용자에게 품질 높은 시각적 정보를 제공하고, 추천 결과에 대한 사용자 만족도를 높일 수 있도록 한다.

4. 제안 방법

4.1 시스템 아키텍처

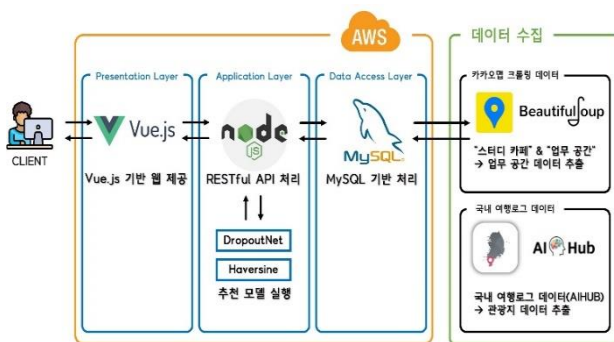


그림 3: 시스템 아키텍처

본 연구에서 진행하는 위케이션 관광지, 업무공간 추천 서비스의 전체 시스템 아키텍처는 [그림 3]처럼 표현된다. 사용자는 Presentation Layer와 상호작용한다. 그리고 해당 Layer는 웹 애플리케이션의 UI를 담당하며, Vue.js를 기반으로 구현한다. 메인 페이지를 통해 첫 로그인 시 추천 시스템에 사용될 성별을 입력할 수 있는 페이지를 구현하였다. 그리고 추천 페이지 진입 전 위케이션 기간과 관광지에 대한 취향을 입력 받는다. 이를 토대로 추천된 관광지 중 한 가지를 선택하면 가까운 업무공간 목록을 제공하며 관광지들과 업무공간들을 선택하여 저장할 수 있다. 저장된 공간들은 마이페이지의 내가 선택한 장소에서 목록을 조회할 수 있고, 지도에 선택한 장소들이 마커로 표시된다.

Presentation Layer는 Application Layer와 상호작용한다. Application Layer는 사용자의 서비스 요청과 관련된 부분을 처리하며 해당 요청에 대한 응답을 제공하기 위해 Data Access Layer의 DB와 드롭아웃-넷, 하버사인 공식

등 여러 기능들을 수행한다. 사용자에게 입력받은 위케이션 기간과 관광지 취향을 입력받아 드롭아웃-넷 모델을 통해 관광지 추천 목록을 제공한다. 그리고 선택한 관광지를 기준으로 하버사인 공식을 적용하여 근처의 업무공간을 제안한다. 해당 Layer는 Node.js 기반으로 개발한다. 여러 사용자는 웹페이지를 통해 개인의 선호 여행 취향을 입력하고 관광지 및 업무 공간 추천 서비스를 제공받을 수 있다.

서버는 Amazon 서버를 활용하였으며, DB Schema를 구성해 정보들을 구분할 테이블을 설계하였으며, 각 테이블을 연결시키는 ER Diagram을 구성하였다. Data Access Layer는 응답을 제공하기 위해 필요한 데이터 관련 CRUD 작업을 수행한다. 그리고 MySQL을 데이터베이스를 활용하며 지속적으로 발전하는 여행 및 업무공간 데이터를 안정적으로 저장하고 조회할 수 있다. 해당 DB에서 관리되는 데이터들은 관광지 데이터와 업무공간 데이터로서 전처리 이후 데이터베이스에 덤프되어 관리된다.

4.2 관광지 추천

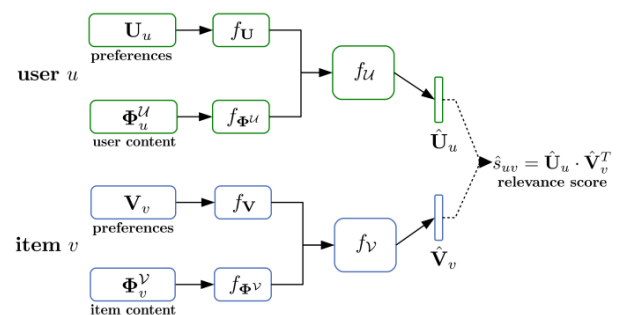


그림 4: 드롭아웃-넷 구조

관광지 추천 시스템을 구현하기 위해 드롭아웃-넷 모델을 사용한다. 드롭아웃-넷[9]은 Cold-Start 문제를 해결하기 위해 도입된 추천 시스템으로, 초기 단계에서 부족한 정보에 대한 대응을 위해 Dropout 기법을 활용한다. 이를 활용한 학습 환경 조성은 정확하고 효과적인 추천을 제공하는 데 기여한다.

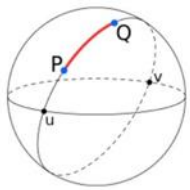
추천 시스템에서 나타나는 Cold Start 문제는 초기 입력 데이터의 부족으로 새로운 아이템이나 사용자에게 대한 정확한 추천을 어렵게 만드는 문제를 나타낸다.

DropoutNet 은 [그림 4]와 같이 Dropout 기법을 도입하여 Cold Start 학습 환경을 조성한 Latent Model 과 Denoising AutoEncoder 를 결합한 형태로 구성되어 있다.

DropoutNet 은 SVD(Singular Value Decomposition)를 활용한 Latent Model 을 도입한다. 기존에 사용되는 사용자-항목 평가 점수 외 사용자, 항목에 대한 추가 정보를 활용하는 방법이다. 이를 통해 데이터를 간결하게 표현하여 정확도가 높아 빠른 검색에 최적화되어 있다.

DropoutNet 의 또 다른 주요 구성 요소로는 Denoising AutoEncoder 가 활용된다. 이 모델은 데이터를 특정 벡터로 인코딩한 후, 노이즈가 있는 데이터를 정제하고 다시 복원한 값이 원래 데이터와 일치하도록 학습한다. Denoising AutoEncoder 는 사용자와 아이템 간의 상호작용에서 유의미한 특성을 추출하여 모델의 성능을 향상시키는 데 기여한다.

4.3 업무공간 추천



$$d = 2 \cdot R \cdot \arcsin \left(\sqrt{\text{hav}(\theta)} \right)$$

$$\text{hav}(\theta) = \sin^2 \left(\frac{\theta}{2} \right)$$

$$\text{hav}(\theta) = \sin^2 \left(\frac{\Delta\phi}{2} \right) + \cos(\phi_1) \cdot \cos(\phi_2) \cdot \sin^2 \left(\frac{\Delta\lambda}{2} \right)$$

그림 5 : 하버사인 공식

업무공간 추천 시스템을 구현하기 위해 하버사인 공식(Haversine Formula)을 사용한다. 하버사인 공식은 위도와 경도를 사용하여 지구상의 두 지점 사이의 거리를 계산하는 공식이며 거리 d는 두 지점 (P,Q) 간의 대원 거리로 [그림 5]의 공식을 활용하여 구할 수 있다. 하버사인 공식은 주로 지리적 위치 기반 서비스 및 지리 정보 시스템에 사용한다. 하버사인을 사용하여 추천된 관광지마다 관광지와 전체 업무공간들의 거리를 각각 측정하여 거리(d)가 가장 가까운 업무공간을 추천한다.

5. 실험

5.1 드롭아웃-넷(Dropout-Net)

Hyperparameter	Value
num_epoch	1000
learning_rate	0.005
user_batch_size	1000
data_batch_size	100
eval_batch_size	1000
n_scores	1000
max_data_per_step	2500000

그림 6 : 드롭아웃-넷의 하이퍼 파라미터

전처리 된 데이터를 사용하여 하이퍼 파라미터(Hyperparameter)는 [그림 6]과 같이 드롭아웃-넷 논문에서 제시한 기본 설정을 기반으로 user_batch_size 는 1000 으로, data_batch_size 는 100 으로, eval_batch_size 를 1000 으로 설정하여 학습을 진행하였다. 전처리된 데이터의 user 수를 기반으로 n_score_user 는 1000 으로 두어 학습을 진행하였다. 이후 새로운 유저의 정보 벡터를 입력받아 libsvm format 파일 형식에 맞게 추가한 뒤 학습시켜 기존 유저 중 새로운 유저와 가장 유사한 top k 개의 유저를 추출한다. 또한 가장 유사한 유저의 rating 을 top k 개 추출하여 최종 방문지에 대한 정보를 출력한다. 드롭아웃-넷 논문과 동일한 Recall@100 으로 검증한 결과 0.9 의 성능이 도출되었으며, 이는 기존의 목표인 0.7 을 뛰어넘는 수치이다.

5.3 QA 테스트

QA (Quality Assurance) 테스트는 소프트웨어 제품이나 서비스가 고객의 요구 사항을 충족하고, 안정적으로 동작하며, 품질 수준을 유지하는지 확인하기 위한 프로세스이다. QA 테스트는 버그를 식별하고 수정하여 제품의 품질을 향상시키는 데 기여한다. 기능테스트와 성능 테스트를 위해서 QA 테스트 항목을 작성하여 메인페이지, 로그인, 내 정보 페이지, 내가 선택한 장소 페이지, 추천 페이지의 UI 와 기능을 확인했다.

6. 결론(Conclusion)

본 논문에서는 위케이션을 위한 관광지 및 업무공간 추천 서비스를 제안한 뒤 구현하였다. 데이터를 전처리하여 Dropout-Net 으로 관광지를 추천하고, Haversine 라이브러리를 활용해 추천된 관광지과 가장 가까운 업무공간을 제공하는 웹 애플리케이션을 구축하였다. 본 논문의 연구는 관광지 추천에서 한정되지 않고 위케이션 사용자의 니즈를 만족할 수 있는 관광지 및 업무공간 추천 시스템을 웹 서비스로 제공했다는 점과 Recall@100 기준의 목표인 0.7 을 능가하여 0.9 의 성능까지 보였다는 점에서 적절한 연구동향임을 확인하였다. 추가로 본 연구를 기반으로 하여 사용자의 숙박 성향에 대한 정보를 입력 받아 추천된 관광지과 가까운 거리에 위치하면서도 사용자의 니즈에 알맞은 숙박업소를 추천해주는 시스템을 결합할 수 있을 것으로 기대한다. 또한 웹사이트 의 형식으로만 서비스를 제공했다는 한계점을 넘어 애플리케이션 서비스를 제공하는 등 다양한 발전 양상을 보일 수 있을 것이라 기대한다.

References

- [1] [한국관광공사. 2021]. “(지자체용) 위케이션 가이드.” 한국관광공사.
- [2] MYRO. 마이로 기업 소개. <https://help.myro.co.kr/>
- [3] 한국관광공사. 대한민국 구석 구석. 여행콕콕 . <https://korean.visitkorea.or.kr/main/main.do>
- [4] Deer Monday.디어먼데이-위케이션/ <https://dearmonday.io/>
- [5] 일룸. 데스크 양양 위케이션. <https://www.deskerworkation.com/>
- [6] 웰니스 위케이션. 달렘 서비스별 소개서 <https://workation.dallem.com/>
- [7] AI Hub 국내 여행 로그 데이터 (동부권/서부권/수도권/제주도 및 도서지역)
- [8] Kakao Developers. 로컬(Local) API 이해하기. <https://developers.kakao.com/docs/latest/ko/local/common>
- [9] Maksims Volkovs, Guangwei Yu, Tomi Poutanen, DropoutNet: Addressing Cold Start in Recommender Systems