

- 프로젝트 요약보고서 -

데이터분석을 활용한 여행지 추천 웹사이트

팀명: [또바기]

과 목 명 :	졸업프로젝트
지도교수 :	박양재 교수님
제 출 일 :	2020. 11. 10. (화)
소 속 :	컴퓨터공학과
성 명 :	김재이, 홍건호, 김예원, 강영준, 윤건우

목차

1.서론3

2.배경지식5

3.시스템구축8
4. 구현결과 및 실행13
5.시스템 설치 및 사용18
6.결론 및 개선사항19

1. 서론

1.1 배경

이 프로젝트의 목적은 데이터 분석을 활용하여 개인에게 맞춤형 국내 여행지와 여행 스타일에 따른 여행 타입 결과를 제공해 여행의 질을 향상시키고 커뮤니티를 활용해 유저 간의 소통을 활성화하여 새로운 여행지에 관심을 가지게 하는 것에 있다. 이때 사용한 기술은 유저 기반 협업 필터링으로 사용자들로부터 얻은 데이터를 이용해 사용자의 관심사를 예측하는 방법이다. 예를 들어 사용자 A가 강릉, 부산을 다녀왔고 사용자 B가 서울, 강릉, 여수를 다녀왔을 때 서울, 강릉을 다녀온 사용자 C에게 “여수”라는 여행지를 예측해서 추천해 주는 것이다. 추천 시스템의 중점은 데이터에 있다. 많은 정확도가 높은 데이터를 가지고 있느냐에 따라 예측의 정확도가 달려있다. 이를 보조하기 위해 KnnBasic 알고리즘과 Pearson Similarity를 사용해 정확도를 더욱 높이도록 하였다. 최근 국내외 기업에서 사용자에게 정보를 제공할 때 추천 시스템을 사용하는 경우가 늘어나고 있다. 쇼핑, OTT 서비스 등에서 추천 시스템을 사용해 사용자에게 맞춤형 정보를 제공해 만족도를 높이고 있다. 이 기술에서 가장 중요한 것은 사용자에게 보다

정확한 정보를 제공해 주는 것으로 좀 더 정확한 예측 정보를 사용자에게 제공할 수 있어야 한다.

1.2 관련기술 동향

정보 필터링은 1990년대 인터넷의 등장과 더불어 당시 대량의 정보로부터 관심 주제를 선별하기 위한 목적으로 활용되었다. 초기 정보 필터링은 문헌에 포함된 단어 등에 기반을 두어 유사 문헌을 탐색하는 콘텐츠 기반 필터링이 이루어졌으나 이후 특정 주제 분야에 관심이 있는 이용자들의 평가에 기반을 두어 적합 문헌을 추천해 주는 협업 필터링 시스템이 등장하였다. 협업 필터링은 아이템 기반과 유저 기반 협업 필터링으로 나눌 수 있다. 유저 기반 협업 필터링의 경우 아이템의 내용이나 특징이 아닌 특정 아이템을 사용한 경험이 있는 사용자들의 선호도를 이용해서 개인에게 맞춤 추천을 해주는 시스템이다. 1990년대 인터넷의 등장으로 멀티미디어 자료가 급증하면서 단순 콘텐츠 기반 검색의 전통적 방식으로는 사용자들의 요구를 충족시키기 어려워지게 되면서 협업 필터링을 이용한 추천 시스템이 새로운 형식의 정보를 검색하기 위한 적합한 방법으로 각광받고 있다. 또한 더 나은 협업 필터링을 제공하기 위해 클러스터링 기법 데이터 마이닝 기법을 응용하는 등의 추가 연구도 활발하게 이루어지고 있다.

1.3 국내·외 동향

협업 필터링에 기반을 둔 추천 시스템은 실생활 각 분야에 많이 활용되고 있다. 대표적인 사례가 넷플릭스의 영화 추천 시스템인 시네 매치인데 이용자가 선택한 영화 별점 등의 데이터를 수집하여 개인에게 맞춤 영화를 추천해 준다. 또한 넷플릭스는 추천 시스템의 성능 향상을 위해 머신러닝 및 데이터마이닝 기법을 적극 활용하고 있다. 아마존은 주력 상품인 도서에 그치지 않고 전자제품, 의류 등의 다양한 상품 및 서비스를 제공하는 온라인 쇼핑몰로 발전함에 따라 다양한 제품에 협업 필터링을 적극적으로 이용하는데 사용자의 제품에 대한 별점 구매 이력 등 사용자가 페이지에 접속한 후에 한 활동을 관찰하여 파악한 데이터를 활용하고 있다. 국내의 경우 왓차라는 영화 추천 서비스 역시 넷플릭스와 같은 방식으로 이용자 선호도 및 취향을 분석하여 새로운 영화에 대해 사용자의 예상 별점을 예측하는 서비스를 제공하고 있다.

1.4 차별성

여행 목적 및 여행 인원, 교통수단까지 고려한 개인 맞춤형 추천 시스템 구현 및 여행 스타일 별 여행 타입 추천 구현을 통해 사용자의 여행의 질이 향상시키겠다는 목적을 가지고 있다. 또한 커뮤니티 기능을 통해 유저 간의 소통을 활성화해 새로운 국내 여행지에 대한 관심을 이끌어 내도록 해 사용자가 보다 많은 여행 경험을 할 수 있도록 한다.

1.5 주요 기능 및 범위

구분	최종 결과물	세부 내용
추천시스템	유저기반 협업필터링	<ul style="list-style-type: none">• KnnBasic알고리즘 사용• Pearson Similarity 유사도 사용• 직접 얻은 여행지 데이터가 기반

1.6 프로젝트 일정

Activity	Task	추진일정 (주차)								
		1	2	3	4	5	6	7	8	9
프로젝트 설계	팀장 선정 &역할분담									
	비즈니스 시나리오									
자료 조사	설문조사									
프론트 엔드	웹 디자인									
	메인 구현 (백 엔드의 5가지 기능 구현 계획과 함께 진행)									
데이터 분석	협업 필터링 테스트									
	데이터 분석 모듈 구현									
백 엔드	회원관리									
	커뮤니티									
	여행지 추천									
	성격 유형별 추천									
	공지사항									
테스트	서버 배포 테스트									

몇 번 놀리든 정해진 소리만을 낼 수 있다. 이것이 전통적인 컴퓨터 프로그램의 예이다. 전통적인 컴퓨터 프로그램은 정해진 규칙만을 따르기 때문에 경험을 많이 한다고 성능이 더 좋아지거나 나빠지지 않는다. 반대로 ‘페이스북’은 쓰면 쓸수록 점점 더 자신의 취향에 맞는 글들을 상단에 보여준다. 이는 ‘페이스북’이 머신 러닝 기법들을 적용하고 있기 때문이다. 이처럼 데이터를 가지고 스스로 학습하는 것을 ‘기계 학습’이라고 부른다.

2) AWS(아마존 웹 서비스)

image

아마존 웹 서비스(Amazon Web Service)는 아마존의 자회사로 퍼블릭 클라우드 컴퓨팅 서비스를 제공하고 있다. 클라우드 컴퓨팅은 직접 서버 장비를 구매하거나 임대 계약을 하지 않고도, 요청하는 즉시 컴퓨팅 자원을 제공해 주는 서비스로 원하는 시간 동안 원하는 만큼 컴퓨팅 자원을 이용할 수 있는 것을 말한다. 가상 서버를 제공해 주는 'EC2'가 AWS의 대표적인 서비스이다. 아마존 웹 서비스는 누구나 사용 가능한 퍼블릭 클라우드 서비스이다. 이를 사용하기 위해서는 계정 생성이 필수이다. 본래 유료 서비스이지만 체험 서비스를 이용하면 일정 기간 무료로 이용할 수 있다.

3) Pandas(판다스)

image

컴퓨터 프로그래밍에서 pandas는 데이터 조작 및 분석을 위해 Python 프로그래밍 언어로 작성된 소프트웨어 라이브러리이다. 행과 열로 이루어진 데이터 객체를 다룰 수 있고 더욱 안정적으로 대용량의 데이터들을 처리하는데 매우 편리한 도구이다. 인덱스(index)에 따라 데이터를 나열하므로 사전(Dictionary) 자료형에 가깝다. 쉽게 말해, Python(파이썬)의 엑셀 같다고 볼 수 있는데 엑셀은 대용량 처리를 할 때 속도가 아주 느리지만, pandas를 이용하면 좀 더 빠르게 할 수 있다. 그리고 똑같은 작업을 반복해야 할 때 pandas 코딩을 해 놓으면 생산성이 훨씬 높다.

4) 장고(Django)

image

장고는 파이썬으로 작성된 오픈소스 웹 프레임워크로 MVT 패턴(Model-View-Template)을 따르고 있다. 고도의 데이터베이스 기반 웹사이트를 작성하는 데 있어서 사용자에게 편리함을 주는 것이 장고의 장점이다. 보안이 우수하고 유지보수가 편리한 웹사이트를 신속하게 만들 수 있어 새로 웹 개발을 시작할 필요 없

이 그저 프레임워크를 활용하여 앱 개발에만 집중할 수 있다. 쉽게 말해, 장고는 회원가입, 로그인, 로그아웃과 같이 사용자 인증을 다루는 방법이나 웹사이트의 관리자 패널, 폼, 파일 업로드와 같은 것들을 스스로 해준다.

image

[참고 그림 - MVT패턴]

2.2 적용 기술

1) 협업 필터링(collaborative filtering)

협업 필터링은 많은 사용자로부터 얻은 기호 정보에 따라 사용자들의 관심사들을 자동으로 예측하게 해주는 방법이다. 협업 필터링에는 ‘사용자 기반(User-Based) 추천’과 ‘아이템 기반(Item-Based) 추천’으로 크게 두 가지 종류가 있다. 첫 번째로 ‘사용자 기반’은 나와 비슷한 성향을 지닌 사용자를 기반으로, 그 사람이 구매한 상품을 추천하는 방식을 의미한다. 두 번째로 ‘아이템 기반’은 내가 이전에 구매했던 아이템을 기반으로, 그 상품과 유사한 다른 상품을 추천하는 방식을 말한다. 위 두 가지 종류 중 첫 번째인 ‘사용자 기반 추천’을 사용하였다.

image

위 그림은 ‘사용자 기반(User-based)’ 추천 시스템을 도식화한 것이다. 예를 들어, 유저 1은 온라인 물에서 피자와 샐러드, 그리고 콜라를 함께 구매하고, 유저 2는 피자와 샐러드를 구매했다고 가정해보면, 알고리즘은 구매 목록이 겹치는 이 둘을 유사하다고 인식하고, 두 번째 사용자에게 콜라를 추천해 주는 방식이다. 협업 필터링 알고리즘을 사용하기 위해서는 ‘기존 데이터’가 필수 요소이다. 이 때문에 협업 필터링은 ‘콜드 스타트(Cold Start)’라고 부른다. 시스템이 아직 충분한 정보를 모으지 못했다면 새로 시작할 때의 곤란함으로 인해 사용자에게 대한 추

론을 끌어낼 수 없다.

2) KNN 알고리즘

KNN은 K-Nearest Neighbor의 약자로 새로운 데이터를 입력받았을 때 K에 가장 가까운 것을 중심으로 새로운 데이터의 종류를 정해주는 알고리즘이다. 한마디로 유유상종이라고 할 수 있다. 하지만 단순히 K에 가까운 것을 고른다면 오차가 발생할 수 있다. 이 때문에 주변에 있는 몇 개의 것들을 같이 봐서 가장 많은 것들을 골라내는 방식을 사용하는 데 그것을 KNN이라고 한다. 여기서 K 값은 주변의 개수를 의미한다.

image

위 그림에서 보이는 것처럼 K 값에 맞춰 새로운 데이터의 종류를 결정해 줄 수 있다. 예를 들어, K 값이 1일 때는 새로운 데이터 ‘물음표’의 종류가 파란 동그라미로 결정되지만, K 값을 4로 두고 측정한다면 ‘물음표’는 빨간 세모가 된다. 이처럼 기존 데이터의 유사도를 구해 K 개의 데이터를 비교하여 새로운 데이터를 정확도 높은 종류로 지정하는 KNN 알고리즘으로 협업 필터링 추천 시스템을 구현할 수 있다.

3. 시스템 구축

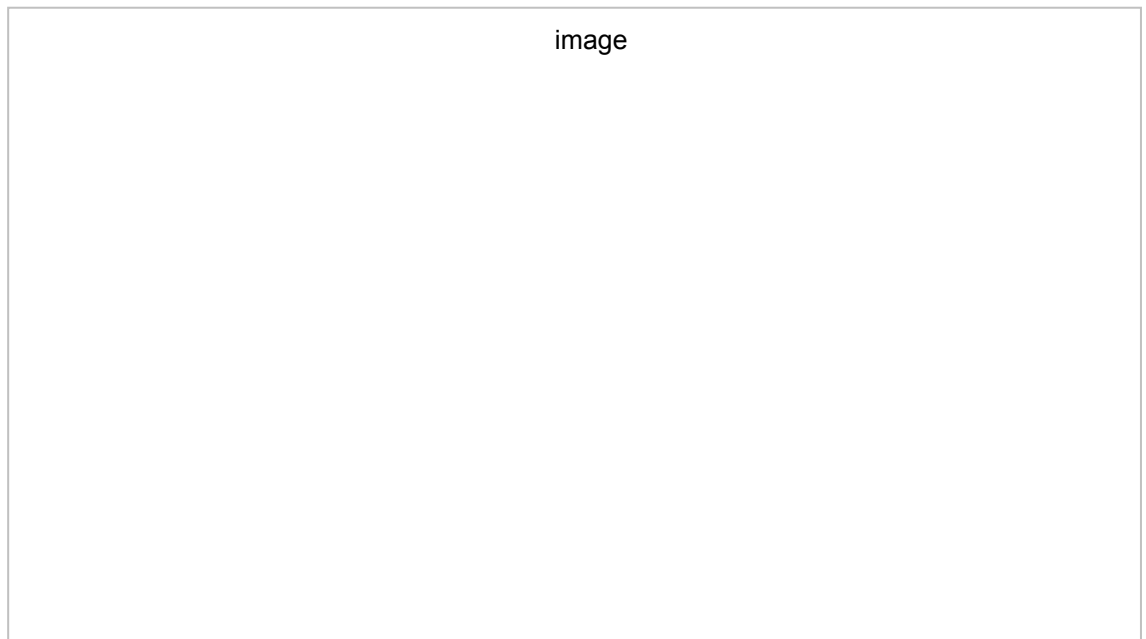
3.1 개발 환경

구분	개발환경	규격
개발툴	편집툴	Visual Studio Code
	개발IDE	PyCharm, Jupyter Notebook
	프레임워크	Django
	개발 언어	Python, JavaScript
	OS	Windows 10

서버		
	DB	SQLite3
서버 하드웨어	CPU	Intel(R) Core(TM) i5-7300HQ CPU @ 2.50GHz
	RAM	8.00GB
	DISK	
	Network	

3.2 시스템 구성

1)서비스 흐름도



- ① 사용자는 웹에서 제공하는 콘텐츠에 따라 회원가입을 하거나 여행 타입 추천 페이지 등에서 설문을 진행한다.
- ② 서버로 전달된 데이터를 Django 프레임워크와 연결된 SQLite3 DB로부터 불러온다.
- ③ 추천 콘텐츠에 필요한 데이터는 초기 설문 폼 결과를 데이터를 기반으로 유저 기반 협업 필터링으로 구현한다.
- ④ Surprise 패키지의 KNN Basic 알고리즘 중 피어슨 상관계수를 사용하여 사용자 간 유사도를 측정한다.
- ⑤ 실행된 알고리즘에 따라 추천 결과값을 웹으로 리턴한다.
- ⑥ 커뮤니티의 경우 Disqus Blog를 사용하여 유저의 댓글 기능을 실행한다.

2) 시퀀스 다이어그램

a) 회원가입

image



- A. 사용자가 서버(AWS EC2)에 Username, Password 등의 회원 정보를 기입한다.
- B. Django Administration에 회원정보를 저장한다.
- C. 저장된 회원 정보로부터 로그인 여부를 사용자에게 출력한다.

b) 여행타입 테스트

image

- A. 사용자가 서버(AWS EC2)에 여행 타입 설문 결과를 요청하면 사용자의 설문 정보가 boolean 값으로 넘어간다.
- B. 사용자의 설문 결과를 SQLite3에 저장한다.
- C. 저장된 사용자의 설문 결과에 맞는 결과를 사용자에게 제공한다.

c) 추천 테스트

image

- A. 사용자가 서버(AWS EC2)에 맞춤 여행지 정보를 요청하면 사용자의 여행 경험이 boolean과 text 값으로 넘어간다.
- B. 사용자의 설문 결과를 SQLite3에 저장한다.
- C. 사용자의 설문 결과를 function에 넣어 사용자 여행 경험에 맞는 맞춤 여행지 정보를 return 한다.
- D. function의 return 값을 사용자에게 제공한다.

d) 커뮤니티 게시글 작성

image

- A. 사용자가 커뮤니티 페이지의 글쓰기 폼에 게시글을 작성한다.
- B. 게시글의 Title, 내용 등을 Administration에 저장한다.
- C. 정해진 형식에 맞춰 저장된 게시물을 등록한다.

e) 공지사항

image

- A. Admin이 Administration 페이지에 공지사항 글을 작성한다.
- B. 형식에 맞춰 공지사항 페이지에 공지사항을 등록한다.
- C. 사용자에게 등록된 공지사항을 출력하고 검색 기능도 제공한다.

3) SQLite3 Database 설계

1) Admin (관리자 테이블)

Name	Data Type	비고
1 id	TEXT	관리자아이디
2 password	TEXT	관리자비밀번호

) : 관리자 계정을 생성한다.

--	--	--

Name	Data Type	비고
1 username	TEXT	사용자 이름
2 email	TEXT	사용자 이메일
3 firstname	TEXT	이름
4 lastname	TEXT	성
5 password	TEXT	사용자 비밀번호

2) User (사용자 테이블) : 사용자 계정을 생성한다.

3) C_type (성격유형 테이블)

Name	Data Type	비고
1 q1	BOOLEAN	성격유형테스트질문1
2 q2	BOOLEAN	성격유형테스트질문2
3 q3	BOOLEAN	성격유형테스트질문3
4 q4	BOOLEAN	성격유형테스트질문4
5 q5	BOOLEAN	성격유형테스트질문5
6 q6	BOOLEAN	성격유형테스트질문6
7 q7	BOOLEAN	성격유형테스트질문7
8 q8	BOOLEAN	성격유형테스트질문8

) : 성격유형 테스트에서 입력값을 받는다.

4) Test (메인테스트 테이블)

Name	Data Type	비고
1 name	TEXT	사용자 이름
2 local	TEXT	지역명
3 rating	TEXT	점수
4 q1	BOOLEAN	메인테스트 질문1
5 q2	BOOLEAN	메인테스트 질문2
6 q3	BOOLEAN	메인테스트 질문3

) : 메인 테스트에서 입력값을 받는다.

5) Photo (커뮤니티 게시글 테이블)

Name	Data Type	비고
1 author	TEXT	게시자
2 photo	BLOB	사진
3 text	TEXT	내용
4 created	NUMERIC	생성날짜
5 updated	NUMERIC	수정날짜

) : 커뮤니티 게시글 작성에서 입력값을 받는다.

6) Notice (공지사항 테이블)

Name	Data Type	비고
1 title	TEXT	공지사항제목
2 pub_date	NUMERIC	생성날짜
3 body	TEXT	내용

) : 공지사항 작성에서 입력값을 받는다.

4. 구현결과 및 실행

4.1 메인

image

4.2 회원가입 및 로그인

image

image

4.3 맞춤형 여행지 추천

image

image

4.4 여행스타일 별 여행타입

image

image

4.5 커뮤니티 및 공지

image

image

image

5. 시스템 설치 및 사용

5.1 시스템 설치

1) 또바기

별도의 설치 필요 없이 아래 링크로 접속하여 이용한다.

<http://ec2-13-125-230-42.ap-northeast-2.compute.amazonaws.com/>

5.2 시스템 사용법

1) 회원가입

여행지 추천 서비스와 성격유형별 여행 타입 테스트는 따로 회원가입이 필요하지 않다. 하지만 커뮤니티 이용을 위해서는 회원가입이 필요하다. 회원가입은 사용자 이름과 비밀번호만으로 가입할 수 있다. 다른 부가정보들은 입력해도 되고 하지 않아도 된다. 비밀번호는 사용자 이름과 같거나 반복되는 숫자들은 사용할 수 없다.

2) 맞춤 여행지 추천

여행지 추천 서비스를 이용하기 위해서는 교통수단, 동행자, 선호하는 여행 테마, 이름, 만족했던 여행지, 그 여행지의 평점을 입력해야 한다.

3) 성격유형별 여행타입 테스트

또바기에서 제시하는 8개의 YES or NO 질문들에 답하여 여행 타입을 알아볼 수 있다.

4) 커뮤니티

회원가입 및 로그인이 필요하다. 글 목록과 글 작성을 선택할 수 있다. 글 목록은 게시된 글, 사진 그리고 작성자를 볼 수 있다. 사진을 크게 보고 싶거나 댓글을 달고 싶으면 댓글 달기 버튼을 클릭하면 된다. 자신이 게시한 게시물을 수정하거나 삭제할 수 있다. 새로운 게시글을 작성하고 싶으면 글 작성 버튼을 클릭하여 작성한다.

5) 공지사항

관리자가 작성한 공지사항을 볼 수 있다. 검색하고 싶은 공지사항이 있으면 검색 기능을 사용한다.

6. 결론 및 개선방안

6.1 문제점 및 개선방안

1) 메인 추천 시스템 테스트 문항의 한계

단기간 프로젝트라는 시간적 제약으로 인해 사용자의 여행 스타일을 파악하기 위한 질문이 단순하다는 한계가 있다. 단 3가지 질문만으로는 정확도 높은 여행지를 추천할 수 있다는 취지와 달리 신뢰도가 높지 않다. 하지만 이미 우리 프로

젝트 목적을 위한 핵심 로직을 중심으로 구현을 완료하였기 때문에 응용할 수 있는 범위가 매우 넓다. 좀 더 정확한 정보를 수집하여 학습시키기 위해 테스트 문항을 세분화할 수 있고 주관식 문항을 객관식으로 변경하여 오류를 최소화할 수 있다.

2) 추천 여행지 범위의 한계

지금 우리가 개발한 시스템은 '지역'만을 추천해 줄 수 있는 시스템이다. 하지만 이 역시 여행지의 범위를 구체화시켜 추천해 줄 수 있도록 응용할 수가 있는 가능성이 있다. 예를 들어, 현재는 단순히 '제주도'라고 추천해 주지만, 시스템 발전을 통해 '제주도 올레길'과 같이 구체화시킬 수 있다.

6.2 결론

코로나로 인해 해외여행이 사실상 불가능해진 이 시점에서 사람들의 관심은 국내 여행으로 쏠리고 있다. 이 때문에 우리 조는 현 상황에 초점을 맞춰 사람들의 흥미를 유발할 수 있는 것이 무엇이 있을까 고민했다. 해당 여행지에 대해 수많은 정보를 제공해 주는 사이트와 정해진 여행지를 가는 데 필요한 수단에 대한 정보를 제공하는 웹 사이트는 많지만 '어디'로 떠나고 싶은데 '어디'로 떠나야 할지 알려주는 웹 사이트는 없다는 생각이 들었다. 코로나에 지친 사람들의 흥미를 유발할 수 있고 현재까지는 완벽하게 존재하지 않는 시스템인 개인 맞춤형 추천 시스템 구현을 목표로 프로젝트를 진행했다.

우리 웹 사이트에서는 사용자에게 맞춤형 여행지를 추천해 줄 수 있는 여행지 추천 테스트와 번외 코너인 성격 유형별 여행지 추천 테스트 그리고 여행 정보를 공유할 수 있는 커뮤니티가 핵심 기능이다. 이러한 기능들을 활용하여 '또바기'에 방문하는 사용자들은 우리가 개발한 테스트를 통해 본인에게 어울리는 여행지가 어디인지 추천받고 커뮤니티를 통해 맛집, 여행 후기 등과 같은 부가적인 정보를 얻을 수 있음을 목표로 하고 있다.

image