

# 개인성향과 협업필터링을 이용한 숙박업소 추천 시스템

김민기\*, Xayvilakone\*, 박두순\*  
\*순천향대학교 컴퓨터소프트웨어공학과  
e-mail : kmk6715k@naver.com

## Accommodations Recommendation System Based on the Personal Propensity and Collaborative Filtering

\*Min-ki Kim, \*Xayvilakone, \*Doo-soon Park  
\*Dept. of Computer Software Engineering, SoonChunHyang University

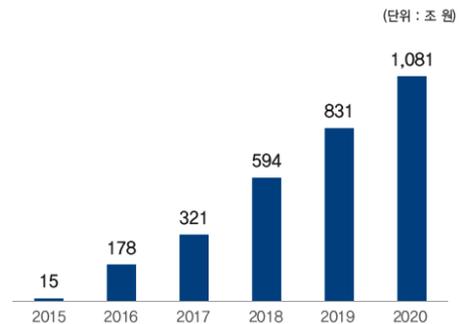
### 요 약

최근 현대인들은 불륜 등 부정적이고 은밀한 공간으로 생각하던 중소형 호텔에 대한 부정적인 인식이 낮아지며 누구나 즐길 수 있고 친구들끼리 추억을 만들 수 있는 공간인 파티 룸에 대한 긍정적인 개념이 더 많이 생겼다. 이에 따라 최근 숙박 어플리케이션들이 중소형 호텔 시장을 진두지휘 하면서 관련 시장이 나날이 성장하고 있다. 본 논문은 기존에 있던 가격, 거리, 평점 중심의 시스템과 달리 개인화 요소인 나이, 직업, 성별, 소득분위, 소비성향을 반영하여 사용자의 주변에 있는 숙박업소 중 사용자들에게 가장 적합한 숙박업소를 추천해주는 시스템을 제안한다.

### 1. 서론

최근 인터넷과 스마트폰의 성장세를 통해 사회 전반적으로 많은 변화가 이루어 졌다. 스마트폰의 성장세는 SNS에서 멈추지 않고, O2O(Online to Offline) 서비스로 확장되어 졌다. 이러한 O2O 서비스 중에 대표적으로 숙박 예약 서비스가 대중들에게 쉽게 접해지게 되었다. 따라서 불륜 등 부정적이고 은밀한 공간으로 생각하던 중소형 호텔에 대한 부정적인 인식이 낮아지며 누구나 즐길 수 있고 친구들끼리 추억을 만들 수 있는 공간인 파티 룸에 대한 긍정적인 개념이 더 많이 생겼다. (그림 1) 은 국내 O2O 시장규모 전망이다.

국내 O2O 시장규모 전망



(그림 1) 국내 O2O 시장규모 전망 ('15~'20).[1].

이에 따라 최근 숙박 어플리케이션들이 중소형 호텔 시장을 진두지휘 하면서 관련 시장이 나날이 성장하고 있다. (그림 2) 는 주요 숙박 O2O 순 이용자 수 추이 이다.

※ 본 연구는 미래창조과학부 및 정보통신기술진흥 센터의 대학ICT연구센터육성 지원사업의 연구결과로 수행되었음 (IITP-2016-H8601-16-1009)



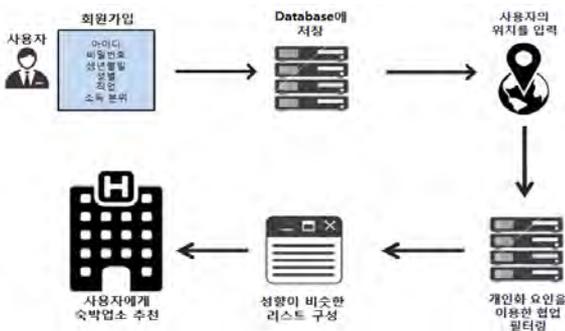
(그림 2) 주요 숙박 O2O 순 이용자 수 추이.[2].

본 논문은 기존의 사용자 기반 협업필터링의 문제 중 Cold Start 문제에 대해 더 나은 방안을 제시한다. 협업 필터링에 대한 데이터가 적을 시 발생하는 Cold Start에서는 사용자의 유사도를 생년월일(나이), 성별, 직업 그리고 소득분위와 같은 개인 성향을 Jaccard Similarity로 계산하고, 여러 가중치를 두어 결과를 비교하여 사용자들의 선호도를 구하여 기존의 시스템보다 정확도 높게 숙박업소를 추천해주는 시스템을 제안한다.

## 2. 숙박업소 추천 시스템의 구성

숙박업소 추천 시스템에서 사용자에게 숙박업소를 추천해주는 방법은 개인성향과 협업필터링을 이용한 방법이다. 협업 필터링 추천 방법이란, 사용자와 성향이 비슷한 사용자들이 공통적으로 좋아하는 콘텐츠를 추천하는 방법으로 과거에 이용한 콘텐츠가 비슷하면 유사한 성향이 있다고 판단하는 근거를 토대로 이루어진다[1]. 추천시스템을 이용하려면 필수적으로 사용자의 회원가입이 필요하다. 개인화 요인으로 사용되어질 생년월일(나이), 성별, 직업 그리고 소득분위를 입력하면 회원가입이 완료된다. 그 받아온 정보를 토대로 사용자에게 적합한 숙박업소를 추천해 준다.

(그림 3)은 숙박업소 추천 시스템이 진행되는 절차이다. 데이터베이스에 저장된 데이터의 양에 따라 두 가지 방법으로 나누어서 진행한다.



(그림 3) 숙박업소 추천 시스템

사용자에게 숙박업소를 추천하기 위해서는 사용자에게

받은 사용자 개인화요인을 통해 사용자와 유사한 근접이웃을 구성하거나, 숙박업소간의 유사도를 측정하여 유사한 숙박업소를 추천하는 방식이 있을 수 있는데, 본 논문에서는 숙박업소와 숙박업소 사이의 유사도를 측정하는 것이 아닌 사용자간의 유사도를 개인화 요인을 통해 계산하여 근접이웃을 구성하는 방식을 사용한다.

첫 번째로 근접이웃을 구성하기 위한 사용자의 개인화요소를 회원가입을 통해 받게 되는데 이때 개인화 요소는 나이, 직업, 성별, 소득분위를 입력받는다. 입력받은 사용자의 DB는 서버 데이터베이스에 저장한다.

두 번째로 회원가입 시 입력받은 개인화 요소에 협업 필터링을 이용하여 최 근접 이웃을 구성한다.

세 번째로 최 근접 이웃의 숙박업소 데이터를 추출한 뒤 그들이 이용한 숙박업소 중 가장 긍정적인 반응을 보인 숙박업소 3개를 추출하여 사용자에게 추천한다.

본 논문에서는 사용되는 개인화 요소를 나이, 직업, 성별, 소득분위로 분류하였으며 나이는 10살 단위로 연령을 분류하였고, 직업은 통계청의 자료를 토대로 대분류 항목을 기준으로 구성하였고 소득분위는 국민건강보험공단의 자료를 토대로 대분류 항목을 기준으로 구성하였다.

(표 1)은 개인화 요소의 분류 결과이다.

(표 1) 개인화 요인 분류 결과

번호	소득분위	직업	나이
1	1분위	학생	20세 미만
2	2분위	관리자	20세 ~ 30세
3	3분위	전문가	30세 ~ 40세
4	4분위	사무직	40세 ~ 50세
5	5분위	서비스업	60세 이상
6	6분위	판매직	
7	7분위	농림어업 종사자	
8	8분위	기능직	
9	9분위	기계 조작 및 조립 종사자	
10	10분위	노무직	
11		군인	

## 3. 숙박업소 추천 시스템의 구현

사용자가 개인 맞춤형 추천 시스템을 사용하려면 회원가입이 필요하다. 회원가입은 추천 시스템에서 사용될 개인화 요인인 생년월일(나이), 성별, 직업, 소득분위를 필수적으로 입력받는다. 숙박업소 추천 시스템의 회원 가입 양식은 (그림 4)와 같다.

**회원가입**

아이디  중복체크

비밀번호  : 4~12자 이상 /  
: 특수문자 사용불가

비밀번호 확인

생년월일 2015 ▼ 1 ▼ 15 ▼

성별  남성  여성

소득분위  선택 ▼

직업  선택 ▼

(그림 4) 숙박업소 추천 시스템 회원가입 양식

회원가입 시 입력한 정보는 회원관리 데이터베이스에 저장되고, 회원관리 데이터베이스는 (그림 5)에서 확인 할 수 있다.

user_id	user_password	user_age	user_gender	user_career	user_income
102464876	1241	23	2	11	1
103325123	52142	15	2	3	4
103241712	1247	35	1	4	5
103215451	1485	27	1	5	7
102251235	13115	22	1	8	2
102685123	21452	42	2	3	3
102541452	1541236	33	2	1	1
102531524	671503	18	2	1	5
102514231	369570	31	1	8	6
101542153	8541211	29	2	6	10
104125412	22145	26	1	6	9

(그림 5) 회원관리 데이터베이스

본 논문에서는 사용자가 회원가입 할 때 입력한 개인정보 중에서 개인화 요인으로 쓰이는 정보들을 나이, 직업, 소득분위, 성별 다섯 가지로 Jaccard Similarity를 이용해 근접 이웃을 구성한다. 사용한 Jaccard Similarity 공식은 다음 식과 같다.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

상기 식에서 J = (A, B)는 사용자 A와 사용자 B의 유사도이다. Jaccard Similarity 방법은 주로 아이템 기반 추천 시스템에서 사용하는 기법인데, 본 논문에서는 이 식을 사용자의 지역이라는 개인화 요소에 적용하였고, 그에 따라 구성된 유사도와 근접이웃을 구상한 데이터베이스의 일부는 (표 2)와 같다.

(표 2) User\_1에 대한 유사도와 근접이웃 구성

ID	Union	Intersection	Similarity	Recommand
user_1	x	x	x	
user_20	1.10	0.90	0.8182	메리어트
user_14	1.20	0.80	0.6667	피카소
user_18	1.20	0.80	0.6667	양카블
user_28	1.20	0.80	0.6667	상그리라
user_29	1.40	0.60	0.4286	씨티
user_2	1.40	0.60	0.4286	펠리스
user_4	1.40	0.60	0.4286	티롤리아
user_5	1.50	0.50	0.3333	미라보
user_7	1.50	0.50	0.3333	강남
user_8	1.50	0.50	0.3333	프라임
user_22	1.50	0.50	0.3333	소풍

또한 개인화 요소를 각각 나이(20%), 직업(20%), 소득분위(20%), 성별(40%)로 가중치를 주었을 때 근접 이웃의 추천 숙박업소는 (표 3)이고, 나이(20%), 직업(20%), 소득분위(25%), 성별(35%)로 가중치를 주었을 때 근접 이웃의 추천 숙박업소는 (표 4)와 같다.

(표 3) User\_1에 대한 유사도와 근접이웃 구성

ID	Union	Intersection	Similarity	Recommand
user_1	x	x	x	
user_28	1.20	0.80	0.6667	양카블
user_29	1.30	0.70	0.5385	메리어트
user_14	1.30	0.70	0.5385	피카소
user_18	1.30	0.70	0.5385	안단테
user_10	1.40	0.60	0.4286	상그리라
user_5	1.40	0.60	0.4286	터미널
user_17	1.40	0.60	0.4286	벨리노
user_9	1.40	0.60	0.4286	태조
user_7	1.50	0.50	0.3333	J
user_13	1.50	0.50	0.3333	티롤리아
user_22	1.50	0.50	0.3333	미라보

(표 4) User\_1에 대한 유사도와 근접이웃 구성

ID	Union	Intersection	Similarity	Recommand
user_1	x	x	x	
user_14	1.20	0.80	0.6667	양카블
user_20	1.30	0.70	0.5385	메리어트
user_18	1.30	0.70	0.5385	피카소
user_21	1.40	0.60	0.4286	안단테
user_5	1.40	0.60	0.4286	상그리라
user_13	1.40	0.60	0.4286	터미널
user_9	1.40	0.60	0.4286	벨리노
user_7	1.50	0.50	0.3333	태조
user_5	1.50	0.50	0.3333	J
user_10	1.50	0.50	0.3333	티롤리아
user_24	1.50	0.50	0.3333	미라보

(표 3)은 가중치 없이 4개의 개인화 요소를 Jaccard Similarity로 유사도를 구하였을 때 나타난 결과로, 가장 높은 점수가 나타났다. 이는 소득분위에 사용자들이 민감한 것을 알 수 있었다.

본 논문에서는 여러 가지 가중치를 실험한 후 요인을 나이, 직업, 소득분위, 성별 4가지의 개인화 요소를 사용하였으며, Jaccard Similarity 유사도 공식을 이용하여 근접 사용자를 추출한다. (그림 6)은 계산된 유사도 점수가 높고 가장 유사한 사용자 3명을 추출하여 숙박업소를 각각 하나씩 사용자에게 추천하는 Top-3 리스트 이다[2].

user_id	Accommodation
102541423	리버 호텔
105214785	소풍
105878712	엠파이어

(그림 6) 사용자와 성향이 비슷한 3명의 유저 추출

사용자와 성향이 비슷한 3명이 좋아하는 숙박업소 정보가 들어있는 데이터베이스는 (그림 7)와 같다.

숙박업소	주소	입실시간	퇴실시간	가격
리버 호텔	충청남도 천안시 서북구 두정동 116-2	13:00	11:00	standard: 80000원
소풍	충청남도 천안시 서북구 두정동 12	14:00	12:00	standard: 65000원
엠파이어	충청남도 천안시 서북구 두정동 192-1	13:00	11:00	standard: 60000원

(그림 7) 숙박업소 정보 데이터베이스

(그림 8)은 숙박업소 추천 결과를 띄운 화면이고 추천받은 숙박업소를 선택하면 (그림 9)와 같이 숙박업소의 상세 정보를 확인할 수 있다.



(그림 8) 숙박업소 추천 결과



(그림 9) 숙박업소의 상세정보

## 참고문헌

- [1] 국내 O2O 시장규모 전망 , <http://www.econovill.com/news/articleView.html?idxno=307340>
- [2] 주요 숙박 O2O 순 이용자 수 추이 , <http://www.etoday.co.kr/news/section/newsview.php?idxno=1260200>
- [3] 김은규, “카테고리 협업 필터링 기반의 모바일 광고 추천 시스템 설계” 한양대학교 공학대학원 2012.2
- [4] 김민기, 박두순, “개인화 요인과 협업필터링을 이용한 드라마 추천 시스템” 순천향대학교 컴퓨터소프트웨어공학 과 2015.11

## 4. 결론

본 논문은 기존에 있던 가격, 거리, 평점 중심의 시스템과 달리 개인화 요소인 나이, 직업, 성별, 소득분위, 소비 성향을 반영하여 사용자의 주변에 있는 숙박업소 중 사용자들에게 가장 적합한 숙박업소를 추천하는 프로그램을 구현하였다.

향후 연구 과제로는 본 논문에서 사용한 개인화요인 이외에 다른 요인들을 찾아 더 적합하게 숙박업소를 추천할 수 있는 것이 필요할 것으로 보인다. 현재는 사용자의 성별과 소득분위에 높은 가중치를 두어 숙박업소를 추천하였지만, 향후에 숙박업소를 추천할 더 많은 개인화 요소를 추가하여 사용자가 다른 숙박업소를 추천 받을 수 있도록 해야 할 것이다.