

협업 필터링과 개인 성향을 이용한 개인화 자취방 추천 시스템

*장태훈, *Khamphaphone, *박두순
 *순천향대학교 컴퓨터소프트웨어공학과
 e-mail : wkdxogns23@naver.com

A Rented room Recommendation System using Collaborative Filtering and Personal Propensity

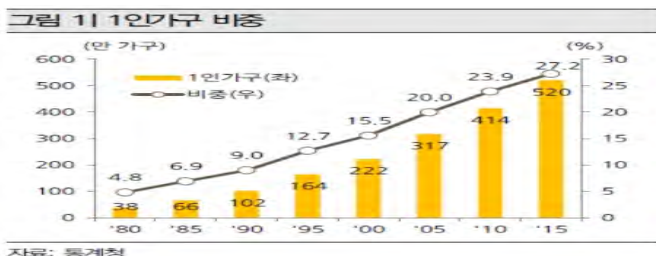
Tae-Hoon Jang, Khamphaphone, Doo-Soon Park
 Dept. of Computer Software Engineering, Soonchunghyang University

요 약

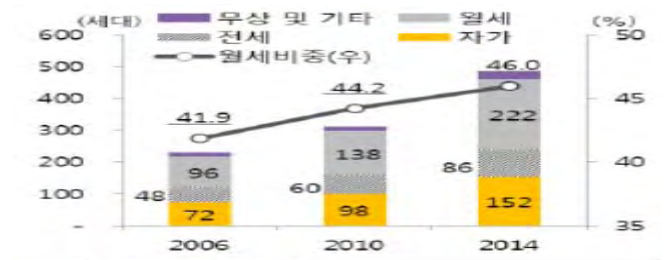
최근 1인 가구의 급격한 증가로 인해 자취방을 찾는 인구가 늘어나고 있다. 그에 따라 수많은 자취방이 등장했으며 매우 다양한 가격과 조건으로 자취방들이 형성되고 있다. 이에 따라 사용자에게 만족되는 조건과 가격을 만족하는 자취방을 추천하기란 어려운 일이다. 따라서 본 논문에서는 개인화 요인과 협업 필터링 방법을 이용하여 사용자에게 보다 적합한 자취방을 추천하는 시스템을 제안한다.

1. 서론

최근 1인 가구가 급격하게 증가하고 있다. 통계청 조사에 따르면 현재 전체 약 1,900만 가구 중 520만 가구가 1인가구로 27.2%에 달할 정도로 많은 비중을 차지한다. 이에 따라 원룸, 투룸, 복층, 오피스텔 등 다양한 종류의 자취방들이 등장했고 자취방 인근의 교통 및 편의시설의 유무와 특정 지역과의 접근성 등에 따라서 다양한 가격으로 자취방이 형성되고 있다. 그렇기 때문에 자취방을 구할 때 따져봐야 할 것이 매우 많고 자신이 생각하는 가격에 적합한 자취방을 찾기란 쉬운 일이 아니다. 1980년 ~ 2015년까지의 1인 가구 증가율은 아래의 (그림 1)과 같고 2006년 ~ 2014년까지의 자취방 증가율은 다음의 (그림 2)와 같다.[1]



(그림 1) 1980년 ~ 2015년 1인 가구 증가율



(그림 2) 2006년 ~ 2014년 자취방 증가율

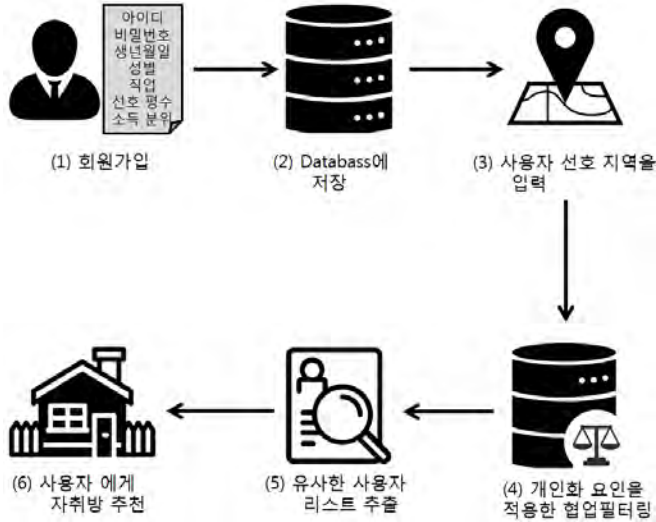
현재 자취방을 알아보는데 도움을 주는 직방(www.zigbang.com)다방(www.dabangapp.com)과 같은 부동산 앱이 존재하지만 단지 특정 지역과의 거리와 해당 집의 사진을 보여주는 것이 전부이기 때문에 사용자들의 기호에 부합하기에는 한계가 있다. 또한 기존의 부동산 앱은 자취방을 추천해주는 것이 아닌 단지 자취방을 포함한 건축물의 정보를 보여주는 것이 전부다.

본 논문에서는 기존의 사용자 기반 협업필터링의 문제 중 희소성 문제에 대해서 보다 나은 방안을 제시한다. 협업 필터링에 대한 데이터가 적을 시 발생하는 희소성 문제를 개인화 요인과 사용자가 원하는 특정 지역과의 거리에 가중치를 부여하고 이러한 데이터를 바탕으로 협업 필터링을 사용하여 사용자들 사이에 유사성을 판별한다. 이렇게 형성된 데이터를 기반으로 사용자에게 가장 적합한 자취방을 추천하는 자취방 추천 시스템을 제안한다.

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

2. 자취방 추천 시스템의 구성

본 논문에서 구현하게 될 자취방 추천 시스템의 시나리오는 (그림 3)과 같다.[2]



(그림 3) 추천 시스템 시나리오

(그림 3)의 추천 시스템 시나리오의 설명은 다음과 같다. 사용자에게 자취방을 추천하기에 앞서 사용자에게 입력 받은 개인화 요인을 통해 사용자와 유사한 사용자를 추출하는 방법과 자취방간의 유사도를 측정하여 유사한 자취방을 추천하는 방법이 있다. 그 중 본 논문에서는 사용자간의 유사도를 개인화 요인을 통해 계산하고 유사한 사용자 리스트를 추출하는 방법을 선택했다.

(1) 먼저 사용자는 추천 시스템을 사용하기에 앞서 회원가입을 통해 개인화 요인으로 사용하게 될 생년월일, 성별, 직업, 자취방의 선호 평수, 소득 분위를 입력한다. (2) 이렇게 입력 받은 개인화 요인 데이터를 데이터베이스에 저장한다. (3) 그 후 사용자는 자신의 자취방이 위치하게 될 선호 지역 혹은 특정 구조물을 입력한다. (4) 그 후 협업 필터링의 문제 중 하나인 희소성 문제를 해결하기 위해 저장된 데이터와 사용자가 입력한 선호 지역에 가중치를 부여하고 기존의 데이터를 이용해 사용자와 유사도를 측정하며 협업필터링을 통해 사용자와 유사한 사용자 리스트를 추출한다.

협업 필터링이란, 사용자들의 선호도와 관심 표현을 바탕으로 선호도, 관심도가 비슷한 사용자들을 식별해 내는 방법으로 과거에 이용한 콘텐츠가 비슷하다면 사용자 간에 유사한 성향을 가지고 있다고 판단하고 그 근거를 토대로 이루어진다.[3]

(5) 이렇게 판별된 데이터를 바탕으로 추천 받을 사용자와 가장 유사한 사용자 2명을 추출하며 (6) 추출된 유사 사용자가 추천하는 자취방을 추천하게 된다.

개인화 요인으로 사용하게 될 요소들은 아래의 (표 1)과 같다

번호	연령대	성별	직업	선호 평수	소득 분위	선호 지역과의 거리
1	10대	남성	관리자	5평 이내	1분위	100M 이내
2	20대	여성	전문가	10평 이내	2분위	200M 이내
3	30대		사무직	15평 이내	3분위	500M 이내
4	40대		서비스업	20평 이내	4분위	1KM 이내
5	50대		판매직	25평 이내	5분위	2KM 이내
6	60대		농림어업 종사자	30평 이내	6분위	5KM 이내
7	70대		기능직	35평 이내	7분위	10KM 이내
8	80대		장치·기계 조작 및 조립 종사자	40평 이내	8분위	10KM 외
9			군인	40평 이상	9분위	
10			학생		10분위	

(표 1) 개인화 요인 사용 요소

위의 그림과 같이 분류한 요소는 선호 지역과의 거리를 제외하고 회원가입을 할 때 사용자에게 입력 받는다. 그 후 추천 직전에 입력 받은 선호 지역과의 거리 정보와 개인화 요소를 기반으로 협업필터링을 통해 자취방을 추천해 주는 절차를 갖는다.

3. 소설 추천 시스템의 구현

본 논문에서 구현한 자취방 추천 시스템은 협업 필터링의 희소성 문제를 해결하기 위해 회원가입이 필수로 되어야 이용이 가능하다. 회원가입을 통해 사용자는 개인화 요인을 입력하게 되는데 요소는 연령대, 성별, 직업, 선호 평수, 소득 분위로 이루어져있다. 회원가입의 양식은 (그림 4)와 같다.

(그림 4) 회원 가입 양식

다음과 같은 양식에 따라서 사용자가 입력한 정보는 회원 관리 테이블에 저장되고, 회원관리 테이블은 소득 분위, 선호 평수, 직업, 성별, 생년월일, ID 순으로 저장된다. 회원관리 데이터베이스 (그림 5)와 같다.

	BIRTH	SEX	JOB	AREA	INCOME	ID
	2	1	4	3	2	20134039
	2	2	6	2	3	20147714
	4	2	3	2	5	20115484
	3	1	2	5	2	20091252
	5	1	7	2	3	20014230
	3	2	4	3	8	20074595
	4	2	3	4	7	20134036
	3	1	10	2	2	20113594
	5	2	3	4	10	20032265
	6	1	4	3	8	20122654
	4	1	1	1	3	20112656
	2	1	2	3	6	20078052

(그림 5) 회원정보 데이터베이스

본 논문에서는 사용자가 회원가입 할 때 입력한 개인화 요인을 각각 선호 지역과의 거리(30%), 소득 분위(20%), 연령대(15%), 선호 평수(15%), 성별(10%), 직업(10%)로 가중치를 부여하였다. 이러한 데이터를 토대로 사용자와 다른 사용자간의 유사도를 피어슨 상관계수를 통해 구한다. 피어슨 상관계수는 -1부터 1의 값을 가지며, 관계가 크면 1에 가까워지고 서로의 유사함이 적으면 -1에 가까워진다. 피어슨 상관계수를 이용한 수식은 아래 (그림 6)과 같다.

$$r = r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

(그림 6) 피어슨 상관계수 공식

이러한 피어슨 상관계수를 이용하여 사용자의 유사도 측정을 한 결과를 토대로 사용자에게 Top-N 리스트를 나타내면 (그림 7)과 같다.

ID	Compare User	Similarity
My ID: 20134036	ID:20154023	0.597
	ID:20143265	0.512
	ID:20123243	0.678
	ID:20172345	0.223
	ID:20091235	0.354
	ID:20144248	0.724
	ID:20129545	0.452
	ID:20132541	0.465
ID:20142214	0.554	

(그림 7) 사용자에게 대한 Top-N List

Top-N List를 정렬하여 추천받을 사용자와 가장 유사한 사용자 2명을 추출하면 아래의 (그림 8)과 같다.


자취방 추천	ID:20144248	추천 자취방: 레드빌
	ID:20123243	추천 자취방: 시몬의 집

(그림 8) 사용자와 가장 유사한 2명의 소셜 추천 리스트


이렇게 추천된 자취방은 사용자에게 자취방의 설명과 사진을 통해 제공되며 자취방을 추천한 사용자들의 수와 실제 사용했던 사용자가 부여한 평점을 볼 수 있도록 구현되었다. 사용자에게 추천된 자취방은 아래의 (그림 9)과 같다.

추천 시스템 실행 자취방 정보 조회 회원 정보 관리

장태훈님에게 추천된 자취방



상권장르	레드빌
면적	010-532-4565
평세	당세 450만원, 계약금 30만원 (1인 주거비용 50만원)
관리비	도시가스/전기료, 인터넷 무료, 관리비 없음
층	42명 추천
평점	★★★★☆ 8.72



상권장르	시몬의 집
면적	041-542-4025
평세	당 200만원 (2명 주거 : 2인/월 300만원 /보증금 X)
관리비	가스, 인터넷, 가스비, 렌트비, 인터넷, 무료
층	32명 추천
평점	★★★★★ 9.48

(그림 9) 추천 결과 화면

4. 결론

본 논문에서는 수많은 자취방 중에서 사용자에게 가장 적합한 자취방을 보다 편리하고 효과적으로 추천하기 위해 개인화 요인과 협업 필터링을 기반으로 추천해주는 프로그램을 구현했다. 또한 기존의 사용자 기반 협업필터링의 문제 중 희소성 문제에 대해서 보다 적합한 추천을 위해 사용자 간에 유사도 측정에 있어 개인화 요소에 가중치를 부여하고 피어슨 상관계수 공식을 이용하였다.

향후 연구 과제는 보다 정확한 추천을 위해 본 시스템을 이용한 사용자에게 평점을 보다 편리하게 부여 할 수 있도록 하는 것과 협업 필터링의 또 다른 문제점인 Cold Start문제를 보완 할 수 있다면 본 논문에서 제안한 것 보다 한층 더 발전된 방법으로 사용자의 성향에 부합하는 자취방을 추천할 수 있도록 할 것이다.

참고문헌

- [1] “한국표준직업분류와_국제분류_연계표”, 통계청 - https://kssc.kostat.go.kr:8443/ksscNew_web/index.jsp 2014.05.21.
- [2] 장태훈, 김한이, 박두순, “협업 필터링을 이용한 개인화 소셜 추천 시스템”, 순천향대학교 컴퓨터소프트웨어공학과 2016. 11
- [3] 김영아, 박두순, “협업 필터링 기반 드라마 추천 시스템”, 순천향대학교 컴퓨터소프트웨어공학과 2013.11