

사용자의 항목별 선호도에 따른 장소 추천 모델의 설계 및 구현

박소은

과천외국어고등학교

e-mail : aliciasocunpark@gmail.com

Design and Implement of Location Recommendation Model based on Users' Preference of Items.

Soeun Park

요 약

사용자 위치 기반 방문지 장소 추천에 대한 연구가 활발히 진행되었다. 기존 장소 추천 모델은 모든 사용자가 방문지 선택을 결정하는 항목에 동일한 가중치를 부여한다고 가정하여 주어진 정보 안에서 정확도를 최대화 시키는데 초점이 맞춰져 있다. 이러한 모델은 사용자의 항목별 선호도가 모델의 가정과 일치하지 않을 때 추천 정확도가 급감한다는 문제점을 가진다. 본 연구에서는 방문지 선택을 결정하는 항목에 가중치를 사용자가 직접 설정하고, 이를 토대로 장소를 추천하는 모델 및 알고리즘을 제안 및 구현한다. 제안하는 기법을 통해, 상이한 요구사항을 지닌 방문 후보자가 자신의 경향성을 반영한 장소를 추천 받는 서비스를 제공할 수 있다.

1. 서론

모바일 기기와 정보처리 시스템의 발전으로 사용자 위치 기반 서비스가 다수 등장하였으며 사용자가 장소 방문을 결정할 때 다양한 항목을 고려한 추천 모델에 연구가 활발하게 진행되었다. 장소 추천을 위한 모델로 시간을 고려한 주제확률모델[1], 은닉 마르코프 기반 모델[2], 행동패턴모델[3] 등이 존재한다. 하지만 이들 모델은 장소를 선정하는 파라미터들의 중요도를 사용자가 조정할 수 없다는 문제점을 지닌다. 방문지를 선택할 때 가격, 거리, 동선, 편리성 등의 항목의 중요도는 사용자마다 상이하기 때문에[4], 임의의 사용자가 이들 장소 추천 모델을 사용할 때의 효용은 현저히 감소할 수밖에 없다.

본 연구에서는 기존 장소 추천 모델의 문제점을 해결하기 위해 사용자가 직접 설정한 선호도에 따라 후보 장소들의 점수를 산정하는 장소 추천 모델과 그 알고리즘을 제안한다. 또한 안드로이드 앱으로 이를 구현하여 사용자가 제안하는 모델을 별도의 프로그래밍 없이 사용할 수 있도록 한다.

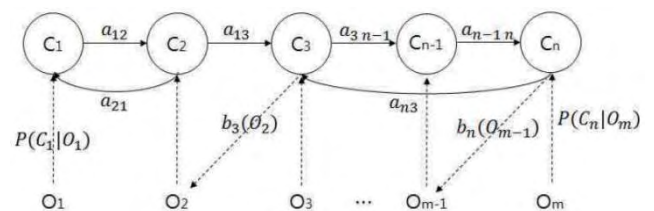
2. 관련연구

2.1 시간을 고려한 확률 기반 모델

사용자에게 방문할 장소를 추천하는 기법으로 시간을 고려한 확률 기반 모델[1]이 존재한다. 해당 모델은 베르누이 확률 변수의 값으로 사용자 고유의 주제 분포를, Gaussian mixture 을 통해 방문 시간을 결정하

여 방문할 가능성이 높은 위치를 top-k 방식으로 추천한다. 해당 모델은 사용자가 선호하는 고유 주제에 대한 정보가 많을 때 높은 정확도를 보인다는 장점이 있지만, 사용자가 빈번하게 접근하지 않은 고유 주제에 대해서는 추천이 불가능하고 사용자가 직접 방문지 선택에 고려요소를 결정할 수 없다는 문제점을 지닌다.

2.2 은닉 마르코프 기반 모델



<그림 0> 은닉 마르코프 기반 추천 모델

사용자에게 콘텐츠를 추천하는 기법으로 은닉 마르코프 기반 모델[2]이 존재한다. 해당 모델은 은닉 마르코프 모델에서 상태 정보 C_n 을 콘텐츠로 맵핑하고 상태 i 에서 상태 j 로 천이 되는 확률변수 a_{ij} 와 출현 확률 $b_i(O_m)$ 를 순차적으로 계산해 추천 콘텐츠 리스트를 만든다. 시간 정보와 사용자의 콘텐츠 선택 정보가 높은 정확도를 지닐 때 좋은 결과를 도출해낼 수 있다는 장점을 가지지만, 사전 정보가 부족하거나 선호도가 변화할 때 예측 정확도가 급격히 하락한다는 문제점이 있다.

2.3 행동패턴을 고려한 모델

사용자에게 사용자가 좋아할 만한 다른 장소를 추천하는 기법으로 사용자의 행동패턴을 고려한 확률기반 모델[3]이 존재한다. 해당 모델은 방문 기록이 유사한 사용자는 유사한 주제 선호도를 가질 것으로 보고 주제 z 를 선택한 사용자 u 가 장소 x 를 방문할 확률을 정규화하여, <식 1>로 나타내었다. 해당 모델은 정규화된 수식에 파라미터만을 입력하여 결과를 도출할 수 있어 빠른 질의 처리가 가능하다는 장점이 있지만, 사용자가 고려하는 여러 항목들에 대한 선호도를 반영하지 못한다는 문제도 있다.

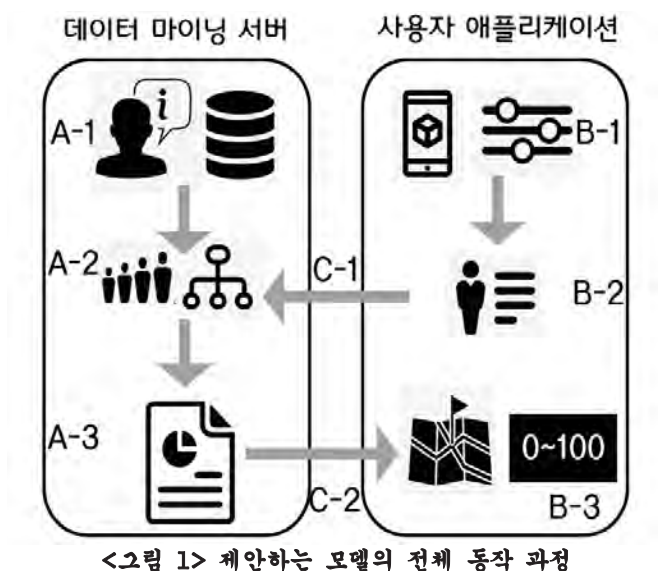
$$P(x|u) = \sum_{z=1}^Z P(z|u)P(x|z, x_u) \quad \text{<식 1>}$$

2.4 Travel Advisor

Travel Advisor[5]는 장소 추천 모델을 기반으로 하는 대표적인 응용 앱이다. Tavel Advisor 를 이용하여 사용자는 장소 추천 모델을 별도의 프로그래밍 과정 없이 사용할 수 있다. 그러나 해당 앱은 미리 입력된 정보만을 보여주거나 사람들의 주관적인 생각만을 여행지에 대해서 리뷰 형태로 보여줄 뿐, 사용자 개인이 가진 여행 장소 선택 시의 고려사항들의 중요도와 선호도를 반영하진 못한다.

3. 본문

3.1 제안하는 모델의 전체 동작 과정



제안하는 모델은 데이터 마이닝 서버와 사용자 애플리케이션으로 구성되어 있으며 그 동작 과정은 <그림 1>과 같다. B-1 에서 사용자가 어플리케이션을 실행하여 사용자가 항목별 중요도를 선택한다. B-2 에서 사용자가 입력한 정보를 취합하여 데이터 마이닝 서버에 전송(C-1)한다. A-1 에서 DB 와 IR engine 에서

후보 방문지를 검색하고, A-2 에서 사용자가 입력한 정보를 모델에 입력하여 100 점을 기준으로 한 여행지 점수의 결과 보고서를 작성한다(A-3). 해당 보고서가 사용자 어플리케이션으로 반환(C-2)되면 B-3 에서 사용자가 최종 결과를 확인하면 모든 동작은 완료된다.

3.2 제안하는 장소 추천 모델

사용자 U 의 경향성 표시를 위한 선호도 입력치 $\{e_1, e_2, \dots, e_n\}$ 와 방문하려는 날짜의 항목별 데이터 $I=\{d_1, d_2, \dots, d_n\}$ 를 처리하여 후보지 a 에 대한 최종 산정 점수 X_a 를 계산하는 수식 모델은 <식 2>와 같다. 점수 산정 함수 $F(d)$ 에 대한 자세한 내용은 본 논문 3.3 장에서 다룬다. $G(k, e)$ 는 입력변수 e 와 $k=F(d)$ 를 곱하여 계산하며, 차후 더 높은 정확도를 지닌 고차원 함수로 변경 가능하다.

$$X_a = \sum_{i=0}^n G(F(d_i), e_i) \quad \text{<식 2>}$$

3.3 선호도 경향성 입력 요인 및 그 점수 산정

Abesamis RA 의 연구[6]에 따르면 사용자가 여행시 가장 많이 고려하는 요소는 {방문자 수, 온도, 거리, 가격, 강수량, 평균 머무는 시간}이다. 본 장에서는 원활한 이해를 위해 유명 관광지 중 하나인 플로리다를 여행 후보지로 가정하여 서술한다.

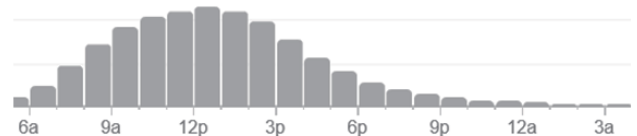
3.3.1 온도

Abesamis RA 의 연구 [6]에 따르면 플로리다의 평균이 28 도이고 평균 온도 보다 높을 때 사람들의 불쾌지수가 급격히 올라가고 낮을 때는 불쾌지수가 낮아진다. 이를 이용하여 <표 1 >과 같이 온도와 반비례하는 점수 배치를 만들 수 있다.

온도(°C)	~27	27~29	29~
산정 점수	10	8	2

<표 1> 온도와 산정 점수 관계

3.3.2 방문자수



<그림 2> 시간에 따른 방문자 수

플로리다의 방문자수는 <그림 2>와 같다[7]. 약 12 시~1 시는 사람이 가장 많은 시간이므로 2 점을 배치하고, 만약 9 시~10 시 이거나 2 시~3 시 일 때는 시간이 적절하고 사람도 거의 없는 시간이므로 10 점을 배치한다. 만약 10 시~11 시, 1 시~2 시이면 사람이 조금 있는 시간이므로 6 점을 배치하고, 만약 그 외 시간은 방문자수는 거의 없으나 너무 새벽이거나 너무 밤이

므로 치안상 낮은 점수인 4 점을 배치한다. 이때 헤밍웨이는 영업시간이 오전 9시부터 오후 5시까지이므로 영업시간이 아닐 때는 전체점수를 0 점으로 처리한다.

3.3.3 거리

JL Nicolau 의 연구[8]에 따르면 사람들이 여행장소를 선택 시 이동에 소요되는 시간이 보통 30 분 이내면 여행지까지의 이동이 매우 편리하고 가기 쉽다고 생각하고 2 시간 이상이라면 가기 어렵다고 생각한다. 이를 이용하여 <표 2 >과 같이 온도와 반비례하는 점수 배치를 만들 수 있다.

거리	30 분 이내	1 시간	1 시 30 분	2 시간 이내	2 시간 이상
산정 점수	10 점	8 점	6 점	4 점	2 점

<표 2> 거리와 산정 점수 관계

3.3.4 가격

JL Nicolau 의 연구[8]에 따라 가격과 방문지 선택을 위한 점수를 산정한다. 여행객은 방문지 입장이 무료라면 매우 실용적이라고 생각하고, 40 달러 이상이라면 비싸다고 생각한다. 이를 고려하여 점수 산정표를 작성하고 플로리다를 기준으로 대표적인 예시를 뽑으면 <표 3>과 같다.

가격(\$)	산정 점수	예시
40~	0	-
20~40	4	에버글레이즈
0~20	6	헤밍웨이 생가
0	10	리틀쿠바

<표 3> 가격과 산정 점수의 관계

3.3.5 강수량

플로리다의 평균 강수량이 59 인치이고 사람들은 보통 강수량이 높아지면 활동하기 어렵다고 생각하고 강수량이 낮으면 활동하기 쉽다고 생각하는 경향이 있다[9]. 플로리다 평균 강수량이 59 인치이점을 고려해서 만약 40-60 인치 라면 평균강수량이랑 비슷할 때로 8 점을 배치한다. 만약 40 인치 미만일 때는 강수량이 적어 습하지도 않고 활동하기 가장 좋을 때이므로 가장 높은 점수인 10 점을 배치한다. 만약 60 인치 이상 일 때는 비가 너무 많이 온다면 활동하기 힘들므로 가장 낮은 점수인 2 점을 배치한다.

3.3.6 평균적으로 머무는 시간

Xin Lu 의 연구[10]에 따르면 만족스러운 여행을 위해서는 임의의 장소에 머무를 수 있는 시간이 지나치게 짧을 경우 여행자는 만족감을 느끼지 못한다. 그리고 전동환 의 연구[11]에 따르면 오래 머무를 수 있는 관광지는 관광에 대한 전반적인 만족을 높이는 경향이 있다. 3 시간 이상으로 머무를 수 있다면 여행지로서 충분한 가치가 있고, 1 시간도 못 머무는 곳은 여행지로서의 가치가 없다. 따라서 이 논문에서는 평균 머무는 시간이 만약 30 분 이상 1 시간 미만이라면 0 점을 배치한다. 그리고 만약 1 시간 이상 2 시간 미만이라면 4 점을 배치하는데 헤밍웨이의 집 과 플로리다 리틀 쿠바가 여기에 해당한다. 만약 2 시간 이상 3 시간 미만이라면 6 점을 배치하는데 에버글레이즈 국립공원이 여기에 해당한다. 마지막으로 만약 3 시간 이상이라면 8 점을 배치하는데 키웨스트가 여기에 속한다.

머무는 시간	30 분이상~ 1 시간 미만	1 시간이상~ 2 시간 이상	2 시간이상~ 3 시간 미만	3 시간 이상
산정 점수	0 점	4 점	6 점	8 점
예시	-	헤밍웨이의 생가, 플로리다 리틀쿠바	에버글레이즈국립공원	키웨스트

3.4 제안하는 알고리즘

Algorithm 1.
Input : Date, UserPrior, Destination
Output : Score
1. /*Init*/
2. Integer prior [] = getParam ()
3. Data dt = getDate()
4.
5. /*Request Destination Information */
6. IF (getInfo(dt) == NULL):
7. return 0
8. ELSE :
9.
10. /*Make Score */
11. Info i_list [] = getInfo(dt)
12. Double score []
13.
14. FOR EACH i = 0 → LENGTH(prior) :
15. score[i] = prior[i] × to_score(i_list[i])
16.
17. Double score_avg = 0
18. FOR EACH i = 0 → LENGTH(prior):
19. score_avg+ = score[i]
20. return score_avg / LENGTH(prior)

<그림 3> 제안하는 알고리즘

제안하는 알고리즘은 [그림 3]과 같이 3 단계로 구성된다. 먼저 초기화 단계에서 사용자가 입력한 중요도와 날짜 데이터를 클라이언트에서 가져온다. (2~3 lines). 목적지 정보 요청 단계에서 해당하는 날짜에 후보 도착지에서 온도, 방문자수, 가격 등의 데이터를 검색한다. 이때 데이터를 찾을 수 없으면 사용자에게 0 점을 반환하고 알고리즘은 종료된다(6~7lines). 데이터를 찾을 수 있을 경우, 점수 산정 단계에서 선택한 날짜의 도착지 정보를 제안하는 알고리즘을 토대로 환산한다. 사용자가 입력한 중요도 곱하여 score 값을 최종적으로 산정할 수 있다(11~15lines). 점수 산정이 완료되면 이들 점수의 평균을 이용하여 클라이언트에게 최종 반환하고 알고리즘은 종료된다 (17~20lines).

4. 애플리케이션 구현



<그림 4> 어플리케이션

어플리케이션의 실행과정은 사용자가 여행을 갈 날짜(A)와 시간(B)을 선택하고 후 여행지를 갈 때 고려하는 요소들(온도, 방문자 수, 거리, 가격, 강수량, 평균적으로 머무는 시간)의 중요도만큼 Seek Bar 을 오른쪽으로 당겨주면(C) 제안하는 알고리즘에 따라 계산된 결과 값을 보여준다(D). 이 애플리케이션은 git 에 공개하도록 한다

5. 결론 및 향후 연구

본 연구에서는 사용자 선호도에 따른 여행지 추천 시스템을 개발함으로써 사용자경향성에 따라 사용자

가 지정한 여행지에 대한 점수를 보여준다. 제안하는 알고리즘은 사용자의 경향성을 반영함으로써 효율적인 여행지 선택을 도와준다.

향후 연구는 데이터 마이닝 기법을 적용하여 사용자별 방문지 추천 알고리즘의 정확도를 향상시키는 것이다.

참고문헌

[1]정우환, 심규석. (2015). 시간정보를 고려한 장소추천 주제확률모델. 한국정보과학회 학술발표논문집, 167-169.
 [2] 양영욱, 임희석. (2015). 사용자의 선호도 변화를 고려한 추천 모델. 한국정보과학회 학술발표논문집, 819-821.
 [3]송광호, 김영훈, 심규석. (2013). 확률모델링을 사용한 장소 추천 시스템. 한국정보과학회 학술발표논문집, 280-282.
 [4] 신운숙. (2005). 여행목적지 이미지형성과 선택에 영향을 주는 새로운 결정요인에 대한 탐색적 연구. 관광학연구, 28(4), 213-232.
 [5] Jannach D., Zanker M., Jessenitschnig M., Seidler O. (2007) Developing a Conversational Travel Advisor with ADVISOR SUITE. In: Sigala M., Mich L., Murphy J. (eds) Information and Communication Technologies in Tourism 2007. Springer, Vienna
 [6] Abesamis RA, Jadloc CRL, Russ GR (2015) Varying annual weather patterns of Florida and influence. Marine Biology 162:1993
 [7]Map.google.com – popular time
 [8]. JL Nicolau, Francisco J. Más (2006) The influence of distance and prices on the choice of tourist destinations: The moderating role of motivations-Juan L
 [9] Denissen, J. J. A., Butalid, L., Penke, L., & van Aken, M. A. G. (2008). The effects of weather on daily mood: A multilevel approach. Emotion, 8(5), 662-667.
 [10] Generating travel routes from geo-tagged photos for trip planning - Xin Lu, Jiang-Ming Yang , Lei Zhang.
 [11] 전동환, 심상화. (2009). 관광 중요도의 만족과 추천의도에 미치는 영향. 관광연구저널, 23(1), 129-144.