
위치 정보를 이용한 오프라인 도서 추천 시스템 스마트폰을 이용한 사용자 행동추적 방법과 추천시스템의 설계

Offline Book Recommendation System using User Location Information

김선준, Sunjun Kim*, 오혜연, Alice Haeyun Oh**

요약 책은 빠른 기술의 발전에도 불구하고, 정보전달의 유효한 수단으로써 아직도 큰 영향을 미치고 있다. 추천 시스템은 수많은 서적 중 사용자가 책을 선택하는데 큰 도움을 준다. Collaborative Filtering 은 훌륭한 추천 알고리즘이지만, 오프라인 서점과 온라인 서점에서의 사용자 행동 차이로 인해 오프라인 서점에 바로 적용하기에는 무리가 있다. 본 논문에서는 위치 정보를 활용하는 도서 추천 알고리즘을 제안한다. 이 알고리즘은 교보문고의 3 년간의 판매 데이터에 기초하여 만들어졌다. 오프라인 서점에서 사용자의 위치를 추적하기 위하여 아이폰 앱을 개발하고, 사용자 스터디를 통하여 새로운 도서 추천 알고리즘의 성능을 검증하였다.

Abstract A book remains a powerful medium of transferring knowledge and experiences, even in this era of fast increases in e-books and other electronic media. Many recommendation systems assist users in choosing books they make like. A collaborative filtering technique is the most common type of recommendation system in e-commerce environment, but such a technique cannot be applied in a physical bookstore because of inherent differences in users' behaviors in online and offline environments. This paper suggests a novel recommendation system for the offline bookstore using a location information. This system is based on an analysis of 3-year sales records of Kyobo Bookstore, the most popular offline bookstore in Korea. The contributions of this paper include a new recommendation system, implemented for the iPhone, a systematic analysis of the Kyobo dataset for use in offline recommendation, and a user study of the performance of recommendation algorithm.

↓

핵심어: Recommendation, Collaborative Filtering, Human Computer Interaction, Mobile

*주저자 : 카이스트 전산학과 박사과정; e-mail: inornate@kaist.ac.kr

**교신저자 : 카이스트 전산학과 조교수; e-mail: alice.oh@kaist.edu

1. 서론

종이에 인쇄된 문자라는 형태는 예로부터 정보전달의 수단으로써 큰 가치를 지니고 있다. 최신기술의 발전에 힘입어, 종이를 대체하려는 시도가 여러 분야에서 이루어지는 현대에도 종이의 우위는 변하지 않는다. 2006년부터 2008년까지의 모든 교보문고 판매기록에서 보았을 때, 오프라인 서점에서의 구매 횟수는 전체 구매(72,814,910 회) 중 63.05% (45,907,472 회)에 달한다. 높은 할인율과 편리함을 제공하는 인터넷 쇼핑몰의 장점에도 불구하고, 아직도 많은 사람들은 서점에서 책을 직접 보고 사는 것을 선호한다는 사실을 알 수 있다.

한편, 현대로 접어들며 사용자들의 정보 소모량이 많아짐에 따라 도서의 발매량 역시 꾸준히 증가[1]하고 있다. 수많은 도서의 범람 속에 사용자들이 자신이 원하는 책을 찾기 위해 도서추천 시스템 역시 발전하고 있다. 대표적으로 Amazon.com 에서 사용하는 Collaborative Filtering(이하 CF) 을 들 수 있다 [2]. CF 는 사용자들의 구매 기록 등을 바탕으로, 통계적으로 사용자가 구매할만한 제품들을 추천해주는 방법으로써, 높은 선호도와 비교적 간단한 구현 난이도에 힘입어 여러곳에서 활용하고 있다.

한편, 도서추천 시스템의 측면에서 보았을 때, 오프라인 서점에서는 온라인 서점의 방법을 쉽게 적용하기 힘들다. 첫째로, 오프라인 서점에서는 온라인 서점과는 달리 사용자의 행동 추적이 힘들다. 온라인 서점에서는 로그인을 통해 사용자를 특정하고, 페이지이동 등을 통해 사용자의 관심 서적을 쉽게 추적할 수 있지만, 오프라인 서점에서 이와 같은 추적 시스템을 구축하기에는 어려움이 따른다. 둘째로, 시스템이 사용자에게 능동적으로 맞춤형 정보를 제공할 수 있는 수단이 없다. 기존의 오프라인 서점에서 사용자가 책을 추천 받는 방법은 고정된 키오스크를 이용하거나, 가판대 등에 놓여있는 책을 보는 방법이 고작이다. 최근 들어 스마트폰을 활용한 도서 추천 시스템 역시 늘어나고 있으나, 공간적으로 이동에 비용이 필요한 오프라인 서점의 경우, 사용자 물리자 제약을 간과한 추천으로 동선의 비효율성을 가져오는 경우가 있다.

본 논문에서는 오프라인 서점의 특성을 파악하고, 이를 활용한 도서추천 시스템을 제안한다.

2. 오프라인 서점에서의 사용자 특징

우선 온라인 서점과 오프라인 서점에서 사용자의 행동특성이 어떻게 달라지는 지에 대한 조사를 먼저 수행하였다. 3 년간의 교보문고 판매 기록 (표 1)을 사용하고, 분석의 용이성을 위해 표본 데이터를 추출하였다. 우선 모든 책들 중 인터넷 상점, 광화문점, 강남점의 구매 횟수가 각각 각 지점별 평균 구매 횟수보다 많은 책을 선택하고, 총 판매량으로 정렬 후 최상위 10000 권의 목록을 뽑아냈다. 또한, 온라인 점포에서의 구매 횟수와 오프라인 점포에서의 구매 횟수가 각각 20 회 이상인 사용자를 추려냈고, 이는 총 756,477 명의 사용자 중 57,993 명이 이에 해당되었다. 최종적으로 해당

사용자들의 구매기록 중 판매량 최상위 10000 권의 책들의 구매 기록만을 따로 추려내어 각 지점별로 표본 Dataset 을 만들었다. 결과, 인터넷 구매 기록은 전체 26,907,438 건 중 1,618,864 건, 광화문점은 전체 15,499,359 건 중 525,630 건, 강남점은 전체 7,136,808 건 중 248,807 건의 기록을 추출하였다.

기간	2006년 1월 1일 ~ 2008년 12월 31일
구매 횟수	72,814,910 회
데이터 종류	일시, 장소, 구매자 정보, 도서 정보, 수량, 가격
구매자 수	5,033,389 명
서적 수	1,118,395 권
서점	인터넷 상점, 광화문(KHM)점, 강남(GN)점

표 1. 교보문고 데이터의 특성

우선, 사용자의 섹션별 집중도를 우선 살펴보았다. 위 표본 집단에서 한 사용자가 하루 안에 구매한 책들의 모든 각 지점별로 Pair Combination 을 추출하였다. 예를 들어 사용자 A 가 한 지점에서 하루에 책 1, 2, 3 을 구매하였고, 또 다른 날에 4, 5, 6 를 구매하였다면 (1,2), (1,3), (2,3) / (4,5), (5,6), (4,6) Pair 들, 즉 같은 날 구매한 모든 도서간의 Combination 을 해당 지점의 기록에 추가하여, 각 지점(인터넷, 광화문점, 강남점)에서 사용자가 한 번에 구매한 책들의 Pair List 와 그 빈도수를 추출해 낼 수 있었다. 각 도서에 있는 서가 위치 정보를 바탕으로, 얼마나 많은 Pair 들이 같은 섹션에 위치하여 있는가를 살펴보았다. 결과, 인터넷 서점, 광화문점, 강남점에서 각각 28.15%, 37.34%, 35.98% 의 Pair 가 같은 섹션 내부에서의 구매였다. 오프라인 서점에서 같은 섹션 내부에서 구매 횟수가 많음을 알아볼 수 있다.

위에서 나타난 차이를 좀 더 명확히 검증하기 위하여, 한 명의 사용자가 하루에 몇 개의 섹션에서 책을 구매하였는지 알아보았다. 결과를 보면, 인터넷 상점에서는 평균 4.92 개(표준편차 2.11), 광화문점 3.67 개(표준편차 2.07), 강남점 3.27 개(표준편차 2.11)의 섹션을 이용하였다. 각각 인터넷과 광화문점, 인터넷과 강남점에 대한 t-test 결과 모두 통계적으로 유의한 차이를 보였다. (p<0.0001) 결과적으로, 오프라인 서점에서는 온라인 서점보다 좀 더 같은 섹션 내부에 구매가 집중됨을 보였다.

3. 오프라인 사용자 정보 수집 방법

서론에서 간략히 서술한 바와 같이, 오프라인 서점에서는 온라인 서점에 비해 사용자 정보를 수집하는 방법에 큰 제약이 있다. 위와 같은 제약은 기술의 발전에 힘입어, 스마트폰 등의 사용으로 극복할 수 있다. 스마트폰은 개인화된 기계로써 사용자에게 대한 정보를 담고 있으며, 항상 네트워크에 접속이 가능하기 때문에 정보를 쉽게 배포할 수 있는 특징을 지닌다. 다만, 일반적으로 스마트폰에서 위치 정보를 수집하는데 사용하는 GPS 장치는 실내에서의 정확도가 매우 떨어진다. 오프라인 서점

환경은 일반적으로 실내에 위치하여 있으므로, 이에 대한 극복 방안이 필요하다.

본 논문에서는 각 서점에서 제공하는 서가 정보를 토대로, 사용자의 위치를 추적하는 시스템을 제안한다. 이를 위해 아이폰용 앱(그림 1)을 개발하였으며, 도서의 바코드를 스캔하여 정보를 조회하는 기능을 제공하였다.



그림 1. 도서 검색 앱 화면.

좌측: 도서 정보 검색 화면 / 우측: 메인화면

우선 GPS 를 활용하여 서점에 있는지 여부를 먼저 확인하고, 책의 바코드를 스캔하여 조회하였을 때 해당 서가에 있는것으로 간주하는 방법을 사용하여 사용자의 위치를 추정하였다. 사용자 한 명에게 교보문고 광화문점에서 자유롭게 앱을 사용하게 하고, 검색 결과로 위치를 추정해 본 결과, GPS 를 사용하지 않고도 어느 정도 사용자의 동선 추적이 가능하였음을 확인할 수 있었다.



그림 2. 사용자 위치 추정 결과

4. 도서 추천 알고리즘

기존 Collaborative Filtering 알고리즘은 그림 3 과 같은 Item-User 행렬을 사용한다.

	U_1	U_2	...	U_{m-2}	U_{m-1}	U_m
I_1						
I_2						
...						
I_{n-1}						
I_n						

그림 3. $N \times M$ Item-User 행렬.

U_i = i 번째 유저, I_j = j 번째 아이템, $n = |I|$, $m = |U|$.

이 중, Item-based CF 는 각 item 간의 cosine similarity 를 사용하여 한 item 에 대해 similarity 가 높은 item 을 추천하는 방식으로 동작한다. 본 논문에서는, 이를 수정하여 Customer Vector based CF(이하 CV-CF) 추천 알고리즘을 제안한다.

CV-CF 에서는 Customer Vector(이하 CV) 개념을 도입하여 각 사용자의 섹션 선호도에 기반하여 추천 item 을 선정한다. 우선, CV는 다음과 같이 계산한다.

$$\overline{CV}_u = (S_{u,1}, S_{u,1}, \dots, S_{u,n-1}, S_{u,n}) \quad (1)$$

\overline{CV}_u : 사용자 u 에 대한 사용자 벡터 (Normalized)
 $S_{u,i}$: 사용자 u 가 섹션 i 에서 구매한 빈도

본 논문에서는 교보문고 Dataset 을 토대로 모든 사용자의 CV 에 대하여 100 개의 cluster 에 대하여 k-means 알고리즘을 적용하였고, 중복과 수가 적은 클러스터를 제외하여 최종적으로 총 31 개의 cluster 를 추출하였다. 실제 사용자에게 추천 도서 목록을 제공할 때, 3 장의 방법으로 수집한 사용자의 선호 섹션 정보를 바탕으로 가장 similarity 가 높은 cluster 를 선택한 뒤, 해당 cluster 에 소속된 사용자들만을 대상으로 item-based CF 를 다시 수행하였다.

CV-CF 를 평가하기 위하여 총 16 명의 사용자를 대상으로 설문조사를 수행하였다. Item-based CF 만을 사용한 추천 목록과, CV-CF 를 사용하여 각각 20 권씩의 책을 추천하였다. 사용자는 Item-based CF 와 CV-CF 로 추천된 각 20 권의 서적에 대하여 5-level likert scale 로 평가하였다.

	Item-based CF	CV-CF
Mean	3.43	3.78
Std. Dev.	1.256	1.177
N	280	225

표 2. 사용자 평가 결과

위 결과를 바탕으로 t-test 를 수행한 결과, 통계적으로 유의한 차이가 발생하였다. (p-value=0.0012)

4. 결론

본 논문에서는 온라인 서점과 오프라인 서점에서의 사용자 행동 양식 차이를 검증하였다. 오프라인 서점에서 사용자들은 좀 더 적은 섹션에서 구매가 이루어지는 경향을 보였으므로, 이 상황에서의 추천 알고리즘은 위치 정보를 활용해야 할 필요성이 있다. 따라서, 사용자의 위치 정보를 수집하기 위하여 오프라인 서점에서 유효하게 활용할 수 있는 사용자 행동 추적 시스템을 제안하였으며, 사용자의 위치 정보를 활용한 도서 추천 알고리즘(CV-CF)을 제안하였으며 사용자 평가를 이용하여 알고리즘의 성능을 검증하였다. 제안한 알고리즘은 위치정보를 활용하지 않은 알고리즘에 비하여 사용자 선호도가 더 높게 나타났고, 이 결과는 사용자의 위치 정보가 도서추천에 유용하게 활용될 수 있음을 시사한다.

참고문헌

[1] 국립중앙도서관. (2006-2009). 국립중앙도서관연보.
 [2] Linden, G. and Smith, B. and York, J. (2003). "Amazon.com Recommendations: Item-to-item collaborative filtering." IEEE Internet computing, Volume (7), pp. 76-80.