

추천 시스템 연구 동향 분석 및 신규 응용 분야 적용을 위한 고찰

최성운*, 서범준**, 윤성로**
*고려대학교 모바일솔루션학과
**고려대학교 전기전자전파공학부

e-mail: sryoon@korea.ac.kr (corresponding author)

A Survey on Recommender Systems and Perspectives for Emerging Applications

Sungwoon Choi*, Bumjoon Seo**, Sungroh Yoon**

*Dept. of Advanced Mobile Solutions, Korea University, Seoul 136-713. Korea

**School of Electrical Engineering, Korea University, Seoul 136-713. Korea

요 약

인터넷의 발달과 유통되는 정보의 양이 증가함에 따라, 그 속에서 사용자가 원하는 정보를 효과적으로 얻기 위한 노력이 계속되어 왔다. 그 중에서 사용자에게 적합한 상품을 추천하고 정보를 제공하기 위한 추천 시스템이 매우 중요하게 인식되고 있으며, 이에 대한 연구가 지속되고 있다. 본 논문에서는 추천 시스템의 핵심적인 방법들과 발전 방향을 다양한 측면에서 체계적으로 분석하였다. 최근에 자주 사용되어지는 효과적인 추천 시스템의 장단점을 비교하여, 추천 시스템이 적용될 수 있는 다양한 환경에서 최적화된 방법을 제시하였다. 더불어 기존의 추천 시스템이 가지고 있는 한계와 그 극복 방안에 대하여 모색해 보았다.

1. 서론

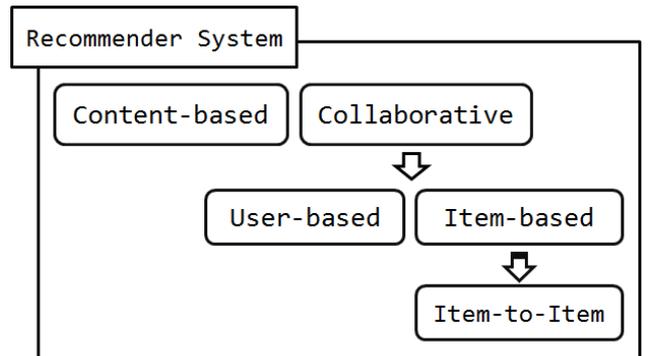
인터넷의 발달로 인하여 사람들은 원하는 정보를 쉽고 빠르게 얻게 되었다. 하지만 시간이 흐름에 따라 인터넷 상의 정보는 급격하게 방대해 졌고, 수많은 정보 속에서 개인이 필요로 하는 정보를 얻기란 더욱 힘들어 졌다. 이에 따라 사용자에게 특화된 추천 시스템이 필요하게 되었다. 추천 시스템은 사용자의 정보 검색 노력을 줄여 줄 뿐만 아니라, 사용자의 충성도 제고에도 도움을 주기 때문에 중요하게 인식되고 있으며, 개인에게 좀 더 적합한 정보와 상품을 추천하기 위한 수많은 연구들이 계속 진행되고 있다.

추천 시스템은 <그림 1>과 같이 크게 Content-based 필터링 방법 [1]과 협업(Collaborative) 필터링 방법 [2][3]으로 나눌 수 있다. Content-based 필터링 방법은 아이템의 기본 성능 또는 특성에 기반하여, 가장 비슷한 속성을 가지는 아이템을 추천하는 방법이고, 협업 필터링 방법은 사용자들의 평점 또는 구매 자료를 바탕으로 유사한 평가를 받은 아이템을 찾아준다. Content-based 필터링 방법은 같은 종류의 아이템들 사이에서만 비교가 가능하지만, 협업 필터링 방법에서는 사용자 또는 아이템의 경향성을 살펴 추천하기 때문에 종류가 다른 아이템들도 비교가 가능하다.

이러한 장점을 가지는 협업 필터링 방법은 크게 두 가지 방법으로 발전이 되었다. 두 방법으로는 사용자를 기반으로 추천하는 사용자 기반(User-based) 협업 필터링 방법

[4]과 아이템을 기반으로 추천하는 아이템 기반(Item-based) 협업 필터링 방법 [5]이 있다. 사용자 기반 방법은 사용자들끼리의 상관관계를 계산하여 특정 사용자에게 자신과 가장 높은 상관관계를 가진 사용자들이 구매 하였던 아이템을 추천하게 된다. 반면에, 아이템 기반 방법은 각 아이템들에 매겨진 사용자의 구매 또는 평점 기록을 바탕으로, 상관도가 높은 아이템들끼리 묶어 추천하는 방식이다. 최근에는 데이터의 크기가 커지고 비교할 아이템 수가 증가함에 따라 성능을 유지하면서 계산량을 줄인 Item-to-Item 방식 [6]도 제안되었다. 이 두 가지 방법은 현재 상용화된 전자상거래 시스템에서 가장 많이 쓰이고 있다.

다양한 곳에서 추천 시스템을 사용하게 됨에 따라 추천 기법들에 대한 비교와 분석이 필요하게 되었다. 기존에 사용자 기반 협업 필터링과 아이템 기반 협업 필터링을 비



<그림 1> 추천 시스템의 분류

교한 논문 [5]이 있기는 하지만, 평가 방법으로 추천 아이템의 정확도만을 우선시하는 문제가 있었으며, 추천 시스템이 필요한 여러 환경에 대하여 효과적인 방법에 대한 정보를 제공해 주지 못하였다.

이 논문에서는 다양한 측면에서 추천 시스템 기법을 비교 분석하고 각 전자상거래 시스템의 상황에 맞는 알고리즘을 제안하였다. 더 나아가 새로운 환경에서 추천 시스템의 적용 가능성에 대하여 예측해 보았다.

2. 추천시스템의 방법

2.1 Content-based 필터링

추천시스템은 크게 Content-based 방법과 협업 필터링 방법으로 나눌 수 있다. Content-based 방법은 기본적인 아이템의 특징만을 고려하여 추천하는 방식이다. 예를 들면, 장난감은 아이들이 대부분 선호한다는 특징을 가지고 있다. 따라서 특정 구매자가 장난감을 사면, 비슷한 특징을 가지는 또 다른 인형 등을 추천해준다. 이 방법은 이해하기 쉽고 구현이 간단하다는 장점을 가지지만, 비슷한 종류의 아이템만을 추천해주는 단점이 있다.

2.2 협업 필터링

협업 필터링은 Content-based와는 다르게 아이템의 기본적인 특징을 고려하지 않는다. 사용자들의 취향을 평점 또는 구매내역을 바탕으로 수집하여 통계를 내고 그 통계를 이용하여 사용자에게 가장 적합한 아이템을 찾아준다. 협업 필터링은 사용자의 유사도를 이용한 사용자 기반 협업 필터링과 아이템의 유사도를 이용한 아이템 기반 협업 필터링으로 나눌 수 있다.

2.2.1 사용자 기반 협업 필터링

사용자 기반 협업 필터링은 모든 사용자들 사이의 유사도를 구하여 특정한 사용자와 취향이 비슷한 사용자들을 분류한다. 그리고 그 사용자에게 분류되어진 취향이 비슷한 사용자들이 선호하는 아이템들을 추천하여 준다. 이때, 선호하는 아이템들은 사용자들의 평점 또는 구매내역에 기반하여 정해진다.

자세한 과정은 다음과 같다. 사용자 기반 협업 필터링의 첫 번째 단계로 <그림 2>와 같이 각 사용자의 평점 데이터를 이용하여 사용자와 아이템 사이의 $n \times m$ 행렬을 만든다. n 명의 사용자는 각각 m 차원의 아이템 평점 정보를 가지고 있으며, 각 차원의 값은 아이템의 평점이 된다. 여기서 m 은 총 아이템의 개수를 의미한다. 두 번째 단계는 각 사용자의 아이템 평점 내역을 바탕으로 사용자들 사이의 유사성을 판단한 후 비슷한 성향을 가진 사용자들 (User i 와 User n)을 이웃으로 묶는 것이다. 이 때 유사성을 판단하는 방법으로는 주로 피어슨 상관계수를 이용한다 [7]. 마지막 단계에서는 추천 대상이 된 사용자와 성향이 비슷한 이웃의 평점 정보를 바탕으로 추천 대상의

	Item 1	...	Item j	...	Item m
User 1	0	0	0	0	9
::	0	0	0	3	0
User i	9	0	7	0	3
::	0	0	3	0	0
User n	8	0	8	0	2
User x	9	0	0	0	3

<그림 2> 사용자 기반 협업 필터링

평점을 예측한 후 몇 개의 상위 아이템을 추천해 준다. 예상 평점계산 방식은 다음의 (식 1)과 같다 [7]. 식에서 $r_{c,s}$ 는 아이템 s 에 대한 사용자 c 의 예상 평점이고, c' 은 c 를 제외한 유사사용자의 집합인 \hat{C} 의 사용자이며 $r_{c',s}$ 는 아이템 s 에 대한 사용자 c' 의 평점이다. 또한 N 은 c 를 제외한 \hat{C} 의 총 사용자의 수이다. (식 1)을 이용하면 <그림 2>의 User x 의 Item j 의 예상 평점은 7.5가 되므로 User x 에게 Item j 를 추천하여 준다.

$$r_{c,s} = \frac{1}{N} \sum_{c' \in \hat{C}} r_{c',s} \quad (\text{식 1})$$

하지만 비슷한 성향을 가진 이웃이라도 개개인의 성향이 조금씩 다르고, 사용자의 평점을 주는 기준도 다르기 때문에 추가적으로 사용자 사이의 상관계수와 개인의 평점평균까지 고려하여 정확도를 높이는 방법이 제안되었다 [7].

2.2.2 아이템 기반 협업 필터링

아이템 기반 협업 필터링은 사용자 기반 협업 필터링과는 다르게 모든 아이템들 사이의 유사도를 바탕으로 성향이 비슷한 아이템들을 묶어낸다. 각각의 아이템은 유저들의 평점 정보를 포함하며 이를 이용하여 유사도를 판단한다. 마지막으로 특정 사용자가 과거에 높은 평점을 준 아이템 또는 구매한 아이템에 대하여 가장 유사도가 높은 아이템들을 추천하여 준다.

아이템 기반 협업 필터링의 첫 번째 과정은 <그림 3>에서 보듯이 아이템과 사용자 사이의 $m \times n$ 행렬을 만드는

	Item 1	...	Item j	...	Item m
User 1	0	0	0	0	9
::	0	0	0	3	0
User i	9	0	7	0	3
::	0	0	3	0	0
User n	8	0	8	0	2
User x	9	0	0	0	3

<그림 3> 아이템 기반 협업 필터링

것이다. 이 때 각각의 아이터은 n차원의 사용자의 평점 정보를 가진다. 두 번째 단계로는 각 아이터들 사이의 평점 정보를 바탕으로 유사도를 계산하여 상관관계가 높은 아이터들(Item 1 과 Item j)을 찾아낸다. 마지막 단계에서는 추천 대상(User x)의 평점 정보(Item 1 의 9점) 또는 구매내역을 바탕으로 그 아이터과 가장 유사도가 높은 아이터(Item j)을 추천하여 준다.

최근의 연구에서는 초 대용량 데이터 셋을 위하여 좀더 빠르게 아이터들 사이의 유사도를 측정하는 방법인 Item-to-Item 협업 필터링이 제안되었다 [6]. 이 방법은 기존의 아이터 기반 협업 필터링과는 다르게 모든 아이터들 사이의 유사도를 계산하지 않고, 차원 값을 가지는 아이터들 사이에만 유사도를 구한다. 일반적으로 모든 사용자들의 구매내역은 총 아이터의 개수보다 훨씬 작으므로, 이 방법을 이용하면 상당수의 차원 값을 가지지 않는 아이터에 대하여 계산을 하지 않아도 되기 때문에 알고리즘의 수행시간을 줄일 수 있다.

3. 새로운 환경에서의 적용

3.1 추천시스템의 장단점 비교

과거와는 달리 최근에는 추천 시스템을 적용하는 분야가 다양해지고 있다. 따라서 상황에 맞는 적절한 추천 시스템을 적용하는 것이 중요해지고 있으며, 이를 효과적으로 사용하기 위하여 각각의 추천시스템이 가지는 장단점을 분석해 볼 필요가 있다.

우리는 최근에 가장 많이 사용되는 추천시스템 방법인 사용자 기반 협업 필터링과 아이터 기반 협업 필터링에 대하여 비교하고 더불어 Item-to-Item 협업 필터링에 대해서도 장단점을 분석해 보았다. <표 1>과 같이 사용자 기반의 협업 필터링은 항상 추천을 받을 사용자와 다른 모든 사용자 사이의 유사도를 계산해야 하기 때문에 실시간 추천으로 인한 최신 경향을 반영 가능하다. 하지만 Amazon.com과 같은 대규모 아이터 셋을 보유하고 있는

회사에서는 급격하게 늘어나는 계산량으로 인하여 실시간 추천이 불가능 하고, 대부분의 사용자들은 총 아이터에 비하여 미미한 구매 또는 평점 내역을 가지기 때문에 아이터의 Sparsity가 높아 사용자들을 분류하기 힘든 문제점을 가지고 있다 [7]. 아이터 기반의 협업 필터링은 아이터들의 유사도를 Off-Line 계산으로 미리 계산하여 메모리에 가지고 있으므로, 사용자의 평점 내역에 따라 실시간 추천을 할 수 있다는 장점을 가지고 있다. 따라서 사용자의 구매 내역이 적더라도 사용자 기반 협업 필터링보다 질 좋은 추천을 해줄 수 있는 반면, 최신 아이터은 추천을 하기 힘들다는 단점이 있다.

3.2 새로운 환경에서의 추천시스템

오늘날에는 다양한 환경에서 추천시스템을 필요로 하고 있다. 글로벌화에 따라 전 세계 사람들이 동일한 구매시스템에 접근함으로 인하여 초 대규모 데이터 셋을 다루는 추천시스템이 필요로 해지고 있다. 초 대규모 셋을 다루는 시스템에서는 많은 데이터를 빠르게 처리해야 하기 때문에, 실시간 처리와 빠른 계산을 할 수 있는 Item-to-Item 협업 필터링 방법이 적합할 것이다.

또한, 스마트폰의 급격한 성장에 따라 많은 사용자들이 스마트폰 내에서의 전자상거래 활동에 참여하고 있다. 따라서 기존의 일반 PC에서의 추천시스템과는 다른 모바일 전용의 추천 시스템을 필요로 한다. 스마트폰 환경의 경우에는 많은 사람들이 저렴한 가격의 어플리케이션을 대량 구매하거나 또는 체험 버전의 어플리케이션을 빈번하게 구매하고 있다. 따라서 일반 웹의 전자상거래 시스템보다 구매한 아이터의 수가 많아 기존 Sparsity를 활용하여 계산량을 줄이는 방식의 알고리즘의 경우 이에 대한 개선이 필요할 것으로 예측된다.

계속해서 대규모의 아이터 셋을 이용한 환경이 등장하고 있고 기존의 방법들은 여전히 한계를 가지고 있다. 따라서 대규모 아이터 셋을 실시간으로 처리할 수 있는 분산 시

<표 1> 추천시스템의 장단점 비교 [5][6][7]

	장점	단점
사용자 기반	- 실시간 추천으로 인한 최신 트렌드 반영 가능	- 대규모 아이터 셋의 실시간 추천 불가능 - Sparsity가 높아 사용자 분류 힘들 - 신규유저의 추천 질의 낮음
아이터 기반	- 실시간 추천 가능 - 대규모 데이터 셋에 적합 - Sparsity가 높아도 추천의 질이 좋음	- 같은 종류의 아이터를 추천 - 아이터 추천을 위한 캐쉬 테이블 필요 - 최신 트렌드 반영 불가
Item-to-Item	- 아이터 기반 협업 필터링의 장점 포함 - 빠른 계산 속도 - 시간복잡도 : O(nm) - 샘플링에 의한 연산량 감소	- 기존 대비 정확도의 감소

시스템에서 사용 가능한 알고리즘의 개발이 요구된다. 구매 또는 평점 데이터를 활용하여 소량의 데이터로도 높은 추천의 질을 얻을 수 있는 알고리즘의 개발 역시 필요하다.

4. 결론 및 향후 연구

사용자에게 적절한 상품과 정보를 제공하기 위하여, 추천 시스템의 활용이 중요해 지고 있다. 지금까지 여러 서로 다른 방법으로 추천 시스템의 성능을 향상시키기 위한 연구가 계속 되었으며, 각각은 장점과 그 한계를 가지고 있었다.

본 논문에서는 여러 추천 시스템 방법들의 특징과 장단점을 알아보았다. 더불어 추천 시스템을 요구하는 환경에 대하여 분석하고, 그 환경에 맞는 적절한 추천 시스템 방법을 이용하여 효과적으로 적용 할 수 있었다.

지금까지 제안되어진 추천 시스템들은 저마다의 한계를 가지고 있었으며, 앞으로 추천 시스템이 필요한 많은 환경에서 원활히 추천이 가능하도록 보다 체계적인 정리와 현재의 단점을 보완하기 위한 연구가 요구되어 진다.

사사

본 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임. (No. 2010-0000407 및 No. 2010-0000631)

참고문헌

- [1] M. Balabanovic, and Y. Shoham, "Content-Based Collaborative Recommendation," Communications of the ACM, Vol.40, Issue.1, 1997.
- [2] J. Konstan, B. Miller, D. Maltz, J. Herlocker, L. Gordon, and J. Riedl, "GroupLens: Applying Collaborative Filtering to Usenet News," Communications of the ACM, Vol.40, Issue.3, pp.77-87, 1997.
- [3] P. Resnjck, N. Iacovou, M. Suchak, P. Bergstrom and J. Riedl, "GroupLens: An Open Architecture for Collaborative Filtering of Netnews," Proceedings of ACM CSCW94, pp.175-186, 1994.
- [4] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Analysis of Recommendation Algorithms for E-Commerce," The ACM E-Commerce 2000 Conference, 2000.
- [5] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-Based Collaborative Filtering Recommendation Algorithms," Prpceedings of the 10th international conference on World Wide Web, pp.285-295, 2001.
- [6] G. Linden, B. Smith and J. York, "Amazon.com Recommendations," IEEE Internet Computing, Vol.7, No.1, pp.76-80, 2003.

- [7] G. Adomavicius, A. Tuzhilin, "Toward the next Generation of recommender systems: A survey of the state-of-the-art and possible extensions," IEEE Transactions on Knowledge and Data Engineering, Vol.17, Issue 6, pp.734-749, 2005.