

개인화된 추천 시스템의 선호도 계산을 위한 정보 필터링

곽미라⁰ 조동섭
이화여자대학교 컴퓨터학과
mirakwak@ieee.org, dscho@mm.ewha.ac.kr

Information Filtering for Preference Prediction of Personalized Recommender System

Mira Kwak⁰ Dong-sub Cho
Dept. of Computer Science and Engineering Ewha Womans University

요 약

웹 기반의 쇼핑몰 사이트의 수가 많아지고 그 이용량이 증가하면서, 차별화된 고객 서비스를 위해 다양한 데이터마이닝 기술들이 적용되고 있다. 특히 고객의 취향에 부합하며 그의 필요를 만족하는 상품을 고객에게 제안하는 추천 시스템을 위해 정보 필터링(information filtering) 알고리즘들이 사용되고 있다. 많은 추천시스템들은 고객들이 상품에 대해 부여한 선호도 정보를 기반으로, 현재 사용중인 고객에게 그와 취향이 비슷한 고객들이 선택했으며, 아직 그가 선택한 적이 없는 상품을 추천하는 협력적 필터링(collaborative filtering) 방법을 사용하고 있다. 본 연구에서는 보통의 협력적 필터링 방법에 내용기반 필터링(content-based filtering) 방법을 적용하고, 고객의 상품에 대한 선호도 점수를 자동으로 계산할 수 있도록 하는 방법을 제안하여 적용함으로써 협력적 필터링 방법을 개선하였다.

1. 서론

인터넷 기반의 상거래가 성장하면서 고객에게 적합한 상품을 추천하는 시스템의 필요성이 생겼고, 이를 위해 정보 필터링(information filtering) 기법들이 널리 사용되게 되었다.

정보 필터링은 두 가지로 나뉘는데, 내용기반 필터링(content-based filtering)과 협력적 필터링(collaborative filtering)이 그것이다. 내용기반 필터링은 고려할 대상의 내용 분석에 기반하는 방법이다. 문서 데이터를 대상으로 할 경우, 용어가 그 안에 얼마나 자주 나타나는가, 그리고 그러한 빈발도는 사용자의 취향과 어떤 관계가 있는가를 고려하여 필터링 하는 것이 이에 속한다. 내용기반 필터링에서 내용 분석과 사용자 취향 파악은 자동으로 이루어진다. 협력적 필터링에서, 특정 사용자를 위해 그 사용자와 취향이 비슷한 사용자들이 좋은 점수를 준 아이템들이 선택된다. 여기에서 아이템의 내용은 고려되지 않는다. 협력적 필터링의 성능은 고려 대상들에 대해 얼마나 많은 사람들이 평가를 했는지에 의존적이며, 따라서 그 양이 너무 적거나 없는 경우 좋은 결과를 낼 수 없다.

보다 좋은 성능을 위해서는 이러한 filtering 기법들

을 결합하고 보완할 필요가 있다. 내용기반 필터링과 협력적 필터링을 결합하여 더 좋은 예측결과를 얻고자 하는 연구가 최근에 이루어지고 있다[1,3,5]. 본 연구에서는 협력적 필터링 기법을 기반으로, 내용기반 필터링 방법과 상품에 대한 점수주기 과정을 사용자의 직접적인 입력 없이 자동으로 이루어지게 하는 방법이 추가 적용된 추천 시스템을 제안한다.

2장에서는 제안하는 시스템의 전체적인 설계물, 3장에서는 내용기반 필터링과 협력적 필터링의 이전까지 연구내용을 소개한다. 4장에서 본 논문에서 제안하는 정보 필터링 방법을 기술하고, 5장에서 향후 연구과제를 보이고 결론을 맺는다.

2. 추천 시스템

본 논문에서는 웹 기반의 쇼핑몰을 위한 추천 시스템을 제안한다. 이 시스템은 크게 네 개의 모듈들로 구성된다(그림 1). 쇼핑몰 사이트에 로그인한 각 사용자에게 대해 시스템은 다음의 작업을 수행한다.:

- (사용자 행위 수집 모듈) 상품의 상세한 내용을 보거나 보고 있는 상품을 구입하는 등, 사용자의 상품구입에 관계된 행위들을 수집한다

- (상품 평가 모듈) 사용자의 상품에 대한 호감도를 계산한다
- (통신 모듈) 상품 평가 모듈에 의해 사용자의 행위로부터 이끌어낸 그의 상품점수들을 수집한다
- (추천 모듈) 현 사용자와 취향이 같은 사용자들을 찾는다
- (추천 모듈) 같은 취향의 사용자들의 취향정보를 사용하여 현 사용자에게 상품을 추천한다

이러한 모든 작업은 카테고리 내 모든 상품에 대해 수행된다.

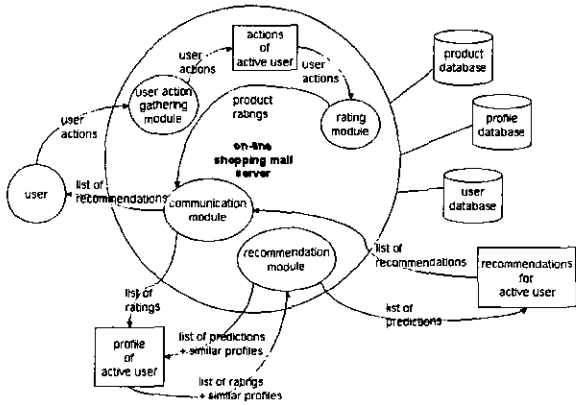


그림 1. 추천 시스템의 구조

3. 관련 연구

3.1 협력적 필터링

협력적 필터링은 취향을 기준으로 묶인 사용자 그룹의 취향정보에 기반하여 현 사용자의 취향을 예측하는 데 널리 사용되는 정보 필터링 기술이다. 예를 들어, 현 사용자의 몇몇 상품에 대한 점수를 알고 있을 때, 이 기법을 통해 그 사용자가 아직 선택한 적이 없는 상품에 어떤 점수를 줄 것인지 예상할 수 있다. 어떤 사용자는 그와 비슷한 취향인 사람들이 좋아하는 것을 좋아할 것이라고 가정하는 것이 이 기법의 핵심 생각이다. 이 기법의 적용은 Amazon.com 등의 유명 상거래 사이트에서 성공적인 결과를 보이고 있다.

대부분의 협력적 필터링 시스템은 일정 범위의 수치 점수로 표현된 사용자의 의견을 수집한다. 수집된 점수들은 성긴 행렬 형태로 저장되고, 협력적 필터링 시스템은 이 점수 행렬을 사용하여 예측한다. 지금까지 몇몇 연구에서 점수 예측을 위한 점수 행렬의 사용에 관한 알고리즘들을 제안하였다[4.6,2]. 본 연구에서 제안하는 시스템에서는 GroupLens 프로젝트에서 제안된, 일반적으로 사용되는 알고리즘을 사용하며, 특히 기능으로 자동 점수주기 방법을 고안하여 적용하였다.

대개, 협력적 필터링 시스템이 수행하는 작업은 사용

자 u 의 아이템 i 에 대한 점수를 예측하는 것이다. 시스템은 사용자 u 의 점수들을 이미 i 를 점수매긴 다른 사용자들의 점수들과 비교한다. 그런 다음 다른 사용자들의 점수에 대한 가중치를 고려한 평균을 사용하여 예측이 이루어진다. 가중치는 사용자 u 와 다른 사용자들 사이의 취향 유사성을 수치적인 양으로 표현한 것이며, 가중치 고려된, 여러 사용자들의 i 에 대한 점수 평균이 사용자 u 가 i 에 대해 줄 것으로 예상되는 점수이다.

3.2 내용기반 필터링

상품들은 색, 크기, 재질, 가격대 등의 속성을 가진다. 한 사용자에게 있어, 이전에 구매한 상품과 비슷한 속성을 가지는 상품은 그의 취향에 맞을 가능성이 높다. 상품들의 속성에 있어 비슷한 정도를 사용하여 상품 유사도를 계산하는 것이 내용기반 필터링이다.

4. 개선된 협력적 필터링

전형적인 협력적 필터링 시스템을 사용하는 경우 몇 가지 문제가 있을 수 있다. 그 중 하나는 저장된 점수 정보의 양이 너무 적은 경우 발생하는 문제이다. 이 경우, 현 사용자의 항목에 대한 점수 예측이 오히려 이루어지기 어렵다. 또 다른 문제는 사용자들이 정직하지 않게 점수를 매기는 경우에 발생한다. 이 또한 옳지 않은 예측 결과를 가져온다. 본 연구에서는 내용기반 필터링을 협력적 필터링에 접목함으로써 비슷한 속성의 상품을 상품을 선택하는 사용자들을 취향이 비슷한 것으로 간주하도록 하여, 점수 정보가 부족한 경우 그 내용을 보충할 수 있게 하였다. 또한 서버측 모듈 구현을 통한 상품보기, 상품구입 등의 사용자 행위정보의 수집을 바탕으로 해당 상품에 대한 호감도를 측정하여, 그 사용자의 해당 항목에 대한 점수로 삼았다. 이를 통해 사용자들이 점수를 거절로 매겨거나 점수를 주지 않는 경우 잘못된 예측이 이루어지는 문제의 발생 가능성을 낮추고자 하였다.

4.1 내용기반 필터링과 협력적 필터링의 결합

상품들에 대한 사용자 점수정보량의 불충분함으로 인해 협력적 필터링이 좋은 성능을 보이지 못하는 경우, 내용기반 필터링을 보조 방법으로 사용하여 나은 결과를 보일 수 있다.

사용자가 상품들에 대해 행한 개별적인 점수주기와 내용기반 필터링 방법으로 구한 상품들 사이의 유사한 정도를 함께 사용함으로써, 사용자가 점수를 매긴 두 상품 사이의 유사도와 그에 매겨지는 점수차 사이의 관계를 측정할 수 있다. 본 연구에서는 내용상, 즉 속성값들이 비슷한 상품들은 한 사용자로부터 비슷한 점수를 받는다고 가정한다.

사용자에 의해 이미 점수 매겨진 상품들의 점수 정보와 그 상품들과 새로운 상품 사이 유사도를 통해 새 상품에 대해 사용자가 매길 점수를 예측한다.

4.2 자동 점수주기를 통한 협력적 필터링

사용자가 상품에 대해 점수를 매길 것을 요구 받았을 때 이를 거부하거나 솔직하지 않게 응답한다면, 예측에 사용되는 점수데이터가 부정확하거나 불충분하게 되어 올바른 추천이 불가능하게 된다. 이를 피하기 위해 사용자가 직접 점수를 입력하게 하지 않고, 서버가 관찰하여 수집한 세션이 유지되는 동안의 사용자의 행위정보를 바탕으로 그의 상품에 대한 호감도를 파악하고 자동으로 점수를 매기는 방법을 제안한다.

자동 점수주기는 사용자의 상품 탐색에 관련된 클릭 스트림에서 추출한 데이터를 기반으로 이루어지는데, 이러한 데이터의 수집은 사용자 행위 수집 모듈에 의해 다음의 과정을 거쳐 이루어진다.:

- 사용자가 로그인 하면 쿠키를 생성한다
- 사용자가 클릭 이벤트를 발생시킬 때마다 그 행위가 뜻하는, '제품 상세히 보기', '주문하기'와 같은 행위 내용을 쿠키에 기록한다
- 사용자가 사이트를 떠날 때 쿠키에 기록된 내용으로부터 알 수 있는 제품 정보가 그 사용자에게 노출된 횟수, 노출시간, 주문여부를 바탕으로 해당 상품에 대한 사용자의 호감도를 계산하여 점수로 매긴다
- 쿠키를 삭제한다

5. 결론 및 향후과제

본 논문에서, 정보 필터링 기법을 사용하는 웹 기반의 쇼핑 물음 위한 추천 시스템을 설계하였다. 이를 위해 협력적 필터링을 기본으로 내용기반 필터링과 사용자 행위정보 수집 결과를 이용한 자동 점수주기 방법을 추가 적용하였다. 이렇게 하여 사용자의 점수정보의 부정확성과 불충분으로 인한 예측과 추천의 실패 가능성을 줄인다.

본 연구에서 보인, 모든 높은 점수의 새 상품을 제시하는 단순한 추천의 수준을 벗어나, 연관규칙탐사와 순차패턴탐사 기법을 적용을 통해 보다 지능적인 추천 서비스를 설계할 수 있다.

참고 문헌

[1] Marko Balabanovic and Yoav Shoham. Fab: Content-based, collaborative recommendation. *Communications of the ACM*, 40(3):66-72, March 1997.

[2] Jack Breese, David Heckerman, and Carl Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, Madison, WI, July 1998. Morgan Kaufmann Publisher.

[3] Joaquin Delgado, Naohiro Ishii, and Tomoki Ura.

Content-based collaborative information filtering. In *Cooperative Information Agents II, Lecture Notes in Artificial Intelligence 1435*, pages 206-215. Springer Verlag Berlin Heidelberg New York, 1998.

[4] Arnd Kohrs and Bernard Merialdo. Clustering for collaborative filtering applications. In *Computational Intelligence for Modeling, Control & Automation*. IOS Press, 1999.

[5] Badrul M. Sarwar, Joseph A. Konstan, Al Bochers, Jon Herlocker, Brad Miller, and John Riedl. Using filtering agents to improve prediction quality in the grouplens research collaborative filtering system. In *Proceedings of ACM CSCW'98 Conference on Computer-Supported Cooperative Work*, 1998.

[6] Uprendra Shardanand. Social information filtering: Algorithms for automating "word of mouth". In *Proceedings of Human Factors in Computing Systems, CHI '95*, 1995.