

## 정보 필터링을 사용한 개인화된 추천시스템

곽미라\*, 조동섭

이화여자대학교 과학기술대학원 컴퓨터학과

## Personalized Recommender System Using Information Filtering

Mira Kwak\*, Dong-sub Cho

Dept. of Computer Science and Engineering EIST Ewha Womans University

**Abstract** - 본 논문에서는 웹기반 쇼핑몰에서 사용자들에게 새로운 상품을 추천하는 시스템을 제안한다. 추천시스템이란 사용자의 필요와 취향을 고려하여 그에게 적합한 새로운 상품이나 대신할만한 상품 등을 추천하는 시스템이다. 지금까지 제안된 대부분의 추천시스템들은 협력적인 필터링 기법을 쓰고 있는데, 이러한 시스템의 경우 사용자들의 선호도 점수 정보가 부족하면 정확한 추천결과를 기대할 수 없다. 본 논문에서는 내용기반 필터링 기법을 협력적 필터링 기법과 함께 사용하여 이와 같은 문제를 해결하고자 한다.

## 1. 서 론

전자상거래의 성장은 사용자를 위한 추천 알고리즘들이 제안되게 하였고, 추천 알고리즘으로 정보 필터링 알고리즘들이 널리 쓰이게 되었다. 정보 필터링에는 두개의 주된 패러다임이 있는데, 내용기반 필터링과 협력적 필터링이 그것이다. 내용기반 필터링 기법을 사용하는 추천시스템은 사용자가 과거에 선호했던 것들과 유사한 아이템들을 추천한다. 따라서 이러한 시스템에는 아이템의 내용 자체에 대한 분석과 사용자의 취향 파악을 자동으로 신뢰할만한 수준에서 처리할 수 있는 방법이 요구된다. 협력적 필터링 기법을 사용하는 추천시스템은 현 사용자에게 그와 취향이 비슷한 다른 사용자들이 선호한 아이템들을 추천한다. 협력적 필터링의 성능은 고려대상이 되는 아이템에 대한 선호정보들의 양에 의존적이다.

내용기반 필터링 기법과 협력적 필터링 기법을 혼합하여 필터링 시스템의 성능을 향상시킬 수 있다. 최근의 연구에서 두 기법들을 혼합하여 추천시스템에 적용한 예들을 찾아볼 수 있다[1,2,3]. 본 논문에서, 현재의 사용자에게 그의 취향과 필요에 부합하는 상품을 추천하기 위해 내용기반 필터링과 협력적 필터링을 결합한 기법을 사용하는 추천시스템을 제안한다.

먼저 기존의 연구들을 2절에서 살펴보고, 우리가 제안하는 시스템을 3절에서 기술한다. 4절에서 결론을 기술함으로써 본 논문을 맺는다.

## 2. 정보 필터링

2.1과 2.2절에서 내용기반 필터링과 협력적 필터링에 관한 기존의 연구들을 살펴보고, 2.3절에서는 이들을 혼합하여 더 나은 성능을 내도록 하는 접근방법을 고찰한다.

## 2.1 내용기반 필터링

내용기반 필터링은 고려의 대상이 되는 아이템에 대한 내용분석에 기반한다. 예를 들어, 문서들을 대상으로 하는 필터링 시스템의 경우 용어의 등장빈도와 그것이 사

용자의 해당문서에 대한 선호도와 어떤 관계에 있는 지 등이 파악되어야 할 내용이다.

내용기반 필터링이 적용된 추천시스템은 단지 해당 사용자가 과거에 선호도를 표시한 아이템들의 내용 분석만으로 그 내용이 이루어진 프로파일에 기초하여 추천한다. InfoFinder와 NewsWeeder가 이러한 시스템의 예다[4,5].

내용기반 필터링 기법만을 사용하는 추천시스템은 다음과 같은 한계를 가진다. 첫 번째 문제는 아이템의 내용에 대한 단편적인 분석에 의존하여, 다른 복잡한 측면에서 볼 때 있을 수 있는, 해당 아이템이 사용자가 선호하는 정도에 미치는 영향들을 고려하지 못하게 된다는 것이다. 예를 들어, 아이템이 그림인 경우 색상, 명도 등 뿐 아니라 예술적인 수준 웹브라우저에 로드되는데 걸리는 시간 등이 무시될 수 있다. 지나치게 개별 사용자에게 특수화된다는 것이 또다른 문제점이다. 사용자는 자신이 과거에 선택하고 높은 선호도를 보였던 것과 아이템 자체의 특징이 비슷한 아이템들만 추천 받게 된다.

## 2.2 협력적 필터링

협력적인 필터링은 사용자에게 그와 취향이 비슷한 사용자들이 선호한 아이템들을 추천하며, 대개의 경우 이때 아이템 내용 자체는 무시된다.

협력적인 필터링을 사용하는 추천시스템은 아이템 내용 분석은 전혀 하지 않고, 사용자들 사이의 취향 유사성에만 의존한다. GroupLens와 Ringo가 이러한 시스템의 예다[6,7].

협력적인 필터링 기법만을 사용하는 추천시스템은 내용기반 필터링만을 사용하는 경우에 발생하는 문제들을 해결한다. 다른 사용자들의 선호도 정보를 바탕으로 다른 종류의 아이템들을 추천 대상으로 고려할 수 있고, 내용자체의 특징이 과거에 해당사용자에 의해 선호되었던 아이템의 것과 다른 경우에도 추천의 대상이 될 수 있다.

그러나 역시 이 시스템에도 그 자체의 문제점들이 있다. 선호도 정보가 부족한 경우 추천 성능이 매우 낮아질 것이다. 새로운 아이템이 추가된 경우나 아이템들의 개수에 비해 사용자 수가 너무 적은 경우 이와 같은 문제가 발생할 수 있다. 또한 취향이 독특한 사용자에게는 추천기능이 적용되기 힘들다는 문제점이 있다.

## 2.3 내용기반 필터링과 협력적 필터링의 혼합

2절에서 설명한 바와 같이, 내용기반 필터링만을 사용하는 추천시스템이나 협력적 필터링만을 사용하는 추천시스템은 모두 그와 전형적인 문제들을 가지고 있다. 이를 해결하기 위해 두 기법을 혼합하여 추천시스템에 적용하는 혼합기법의 접근방법들이 제안되고 있다. 많은 연구에서 협력적 필터링 기법을 기본으로 사용하고 내용기반 필터링의 장점들을 더하여 취하는 방법을 보인다.

이 논문은 2001년도 두뇌한국21사업에 의하여 지원되었음.

이는 협력적 필터링 기법이 내용기반 필터링 기법보다 성능 면에서 더 우수한 것으로 나타났기 때문이다[8].

Fab, GroupLens 그리고 Active WebMuseum 시스템은 이러한 혼합 정보 필터링 기법을 사용하고 있다. Fab 시스템은 웹을 위한 에이전트 기반 문서 필터링 시스템이다[1]. 이 시스템에서 사용자들 사이의 취향 유사성은 선호되는 문서들의 키워드들로 결정된다. 이 때 문서의 키워드는 문서 추출 기법을 통해 발견된다. GroupLens 시스템은 유즈넷을 위한 필터링 시스템으로서, 에이전트들은 감지된 특징들에 기반하여 인위적으로 아이템들의 점수를 매긴다[3]. Active WebMuseum은 화가들의 그림에 대한 추천시스템이다[9]. Active WebMuseum은 협력적 필터링을 사용하여 사용자들이 자신에게 적합하게 설계된 이동경로를 따라 그림들을 탐색할 수 있게 하는 서비스를 제공한다. 이 시스템은 멀티미디어 데이터를 컨텐츠로 하는 웹 서비스에서 어떻게 내용기반과 협력적 필터링 기법이 결합되어 사용자 적응적인 서비스를 제공할 수 있는지를 보인다.

3절에서는 본 연구에서 취한 접근방법을 상세히 설명한다.

### 3. 시스템 설계

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

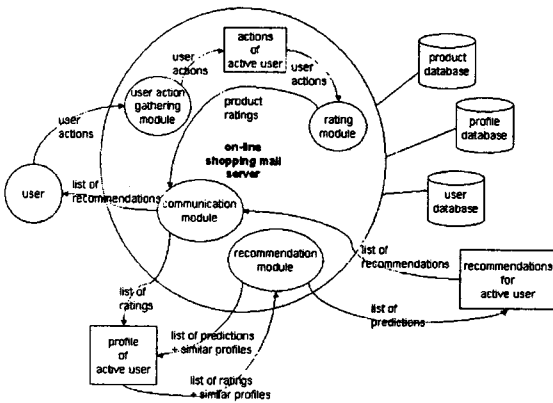


그림 1. 추천 시스템의 구조

- (사용자 행위 수집 모듈) 상품의 상세한 내용을 보거나 보고 있는 상품을 구입하는 등, 사용자의 상품 구입에 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.

하위 절들에서, 적용 기법들을 상세히 설명한다.

#### 3.1 내용기반 필터링

내용 면에서 아이템들 사이의 비슷하고 다른 정도와 그에 따른 사용자의 아이템들에 대한 선호도를 알기 위해서는 그 상관관계를 결정할 수 있어야 한다. 이러한 상관관계는, 대개 내용기반 분석 결과 비슷한 아이템들은 한 사용자에 의해 비슷한 선호도 점수를 받는다는 생각을 바탕으로 한다. 이러한 상관관계를 통해 어떤 사용자의 아이템들에 대한 예상 선호도를 구하고 그에게 적합한 아이템들을 선택할 수 있다.

웹 쇼핑물의 내용은 상품 아이템으로, 컴퓨터 시스템이 분석할 수 있는 상품의 속성으로 색상, 재료, 가격대 등이 있다. 이 논문에서는 가격대만을 다룬다. 두 아이템 사이의 가격대 거리는 다음과 같이 구할 수 있다.

$price_i(j)$ : 아이템  $j$ 의 가격대

$$distance^{price}(p, p') = |price_p(j) - price_{p'}(j)|$$

거리간격은 다음과 같다.

$$j = 1..n_j$$

$$interval_1 = [0, 1), interval_2 = [1, 2)$$

거리 클래스는 다음과 같다.

$$C_j(i) = \{i' \in I_u : distance^{price}(i, i') \in interval_j\}$$

$I_u$ : 사용자  $u$ 에 의해 점수 매겨진 아이템

사용자  $u$ 의 아이템  $i$ 에 대한 선호도 예상 점수는 다음과 같다.

$$p^{price}(u, i) = \sum_{j=1..n_j} \lambda_j \cdot \frac{\sum_{i' \in C_j(i)} r_{u,i'}}{|C_j(i)|}$$

$r_{u,i}$ : 사용자  $u$ 의 아이템  $i$ 에 대한 예상 선호도

#### 3.2 협력적 필터링

지금까지 협력적 필터링 시스템에 관한 연구들이 많이 있어왔다. 이러한 시스템들의 대부분은 사용자들의 의견을 아이템에 대한 수치형태의 점수로서 받아들여 행렬 ratings(user,item) 형태로 표현한다. 이 행렬을 통해 예측이 이루어진다. 이러한 행렬을 사용하여 선호도 점수 예측하는 알고리즘들이 몇 가지 제안되었다[10,7].

본 논문에서는 GroupLens 프로젝트에서 제안되어 널리 사용되고 있는 Pearson 상관계수를 사용한다. 다음은 사용자  $u$ 의 아이템  $i$ 에 대한 선호도를 예측하는 식이다.

$I_u$ : 사용자  $u$ 가 점수매긴 아이템들의 집합

$r_{u,i}$ :  $u$ 의 아이템  $i$ 에 대한 점수

$\bar{r}_u$ : 사용자  $u$ 의 평균 점수

$$\bar{r}_u = \frac{1}{|I_u|} \sum_{i \in I_u} r_{u,i}$$

두 사용자들 사이의 유사도는 사용자  $u$ 와  $u'$ 의 점수 벡터들의 상관관계에 의해 결정된다.

$$\rho(u, u')$$

$$= \frac{\sum_{i \in I_u \cap I_{u'}} (r_{u,i} - \bar{r}_u)(r_{u',i} - \bar{r}_{u'})}{\sqrt{(\sum_{i \in I_u \cap I_{u'}} (r_{u,i} - \bar{r}_u)^2) (\sum_{i \in I_u \cap I_{u'}} (r_{u',i} - \bar{r}_{u'})^2)}}$$

$\rho$ 는 -1과 +1 사이의 값을 가진다. 1에 가까운 값은

높은 유사도를 0에 가까운 값은 낮은 유사도를 나타낸다. -1에 가까운 값은 대개 대립되는 의견을 보임을 의미한다.

예상되는 값은 다른 사용자들이 매긴 점수들의 가중치 주어진 평균이다. 가중치는 사용자  $u$ 와 다른 이들의 취향이 유사한 정도를 의미한다.

$U_i$ : 아이템  $i$ 에 점수를 매긴 사용자들

$$p^{collab}(u, i) = \overline{r_u} + k \sum_{u' \in U_i} \rho(u, u') (r_{u, u'} - \overline{r_{u'}})$$

$$\text{with } k = \frac{1}{\sum_{u' \in U_i} \rho(u, u')}$$

### 3.3 내용기반 필터링과 협력적 필터링 기법의 결합

내용기반 예측값  $\rho^{price}$ 와 협력적 예측값  $\rho^{collab}$ 를 다음과 같이 결합한다.

$$\rho^{comb}(u, i) = \mu^{price} \cdot \rho^{price}(u, i) + \mu^{collab} \cdot \rho^{collab}(u, i)$$

$$\text{with } \sum \mu = 1$$

## 4. 결 론

본 논문에서 정보 필터링 기법을 사용하는, 웹기반 쇼핑물을 위한 추천시스템을 설계하였다. 우리는 필터링을 위하여 협력적 필터링 기법을 주된 기법으로 하고 내용기반 필터링 기법을 추가 적용하였다. 그리하여 저장되어 있는 사용자의 선호도 점수 정보가 불충분한 경우에도 현 사용자의 아이템에 대한 선호도 예측을 수행할 수 있도록 하였다.

### (참 고 문 헌)

- [1] M. Balabanovic and Y. Shoham, "Fab: Content-based, collaborative recommendation," *Communications of the ACM*, 40(3):66-72, March 1997.
- [2] J. Delgado, N. Ishii, and T. 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.
- [3] B. M. Sarwar, J. A. Konstan, A. Bochers, J. Herlocker, B. Miller and J. 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.
- [4] B. Krulwich and C. Burkey, "Learning user information interests through extraction of semantically significant phrases," In *Proceedings of the AAAI Spring Symposium on Machine Learning in Information Access*, Stanford, Calif., March 1996.
- [5] K. Lang, "Newsweeder: Learning to filter netnews," In *Proceedings of the 12th International Conference on Machine Learning*, Tahoe City, Calif., 1995.
- [6] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom and J. Riedl, "GroupLens: An open architecture for collaborative filtering of netnews," In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work*, Chapel Hill,

NC, 1994.

- [7] U. Shardanand and P. Maes, "Social information filtering: Algorithms for automating "word of mouth"," In *Proceedings of Human Factors in Computing Systems*, CHI '95.
- [8] J. Alspector, A. Kolcz and N. Karunanithi, "Comparing feature-based and clique-based user models for movie selection," In *Proceedings of the third ACM Conference on Digital libraries*, pages 11-18, 1998.
- [9] A. Kohrs and B. Merialdo, "Improving collaborative filtering with multimedia indexing techniques to create user-adapting Web sites," In *Proceedings of the seventh ACM international conference on Multimedia*, Pages 27 - 36, 1999.
- [10] A. Kohrs and B. Merialdo, "Clustering for collaborative filtering applications," In *Computational Intelligence for Modeling, Control & Automation*. IOS Press, 1999.