

적응형 콘텐츠 추천 시스템 개발

김건희, 하성도, 최진우, 김태수, 박면웅
한국과학기술연구원 시스템연구부
e-mail : s.ha@kist.re.kr

Development of Adaptive Contents Recommender System

Gunhee Kim, Sungdo Ha, Jin-Woo Choi, Myon-Woong Park, Tae-Soo Kim
Systems Technology Division, Korea Institute of Science and Technology

요 약

웹을 통한 정보량의 폭발적인 증가로 인하여, 사용자에게 적합한 정보만을 제공할 수 있는 개인화 기술에 관심이 증가하고 있다. 정보를 선별하고 추천하는 대표적인 개인화 기술로서 Content-based Filtering(CBF) 기법과 Collaborative Filtering(CF) 기법이 널리 사용되고 있다.

본 논문에서는 위에서 언급한 CBF 기법과 CF 기법을 혼합하여, 사용자 선호도를 보다 정확하게 반영할 수 있는 새로운 모델을 제시한다. 또한, Demographic Filtering 기법과 전문가의 추천을 고려한 Fusion Model 을 제시한다. 그리고 사용자 선호 모델을 실시간으로 반영하기 위한 업데이트 방법을 Exponential Smoothing 기법을 사용하여 구성하였다.

1. 서론

폭발적인 정보량의 증가는 한정된 정보처리 능력을 갖고 있는 우리에게 새로운 부담이 되고 있다. 해마다 수많은 책, 음악, 영화, 웹사이트들이 생성되며, 이 중에 필요한 콘텐츠를 얻는 것은 쉽지 않다. 따라서, 사용자가 보다 쉽게 원하는 콘텐츠를 얻을 수 있도록 사용자에게 유용한 콘텐츠를 선별하는 기법이 필요하다. 그 중 한가지 기법이 추천 기법이다.

추천 시스템은 필터링 기법을 이용하여 사용자의 취향에 맞는 아이템을 추천하여 사용자의 의사결정을 도와주는 시스템이다. 현재, Amazon.com[1]을 비롯한 많은 온라인 사이트에서 추천 시스템을 도입하여 고객이 선호하는 상품을 추천하고 있다. 이러한 추천 시스템은 사용자의 구매 히스토리 데이터와 사용자가 직접 입력한 아이템에 대한 평가 데이터를 이용하여 사용자 모델을 구축하고 추천한다.

본 논문에서는 추천 기법 중 가장 널리 사용되고 있는 Content-based Filtering 기법과 Collaborative Filtering 기법을 중심으로 새로운 추천 방법을 제시한다. 또한, 사용자의 평가를 사용자 선호도 모델에 실시간으로 반영하여 최근의 사용자 취향을 보다 정확히 반영할 수 있도록 한다.

2. 관련연구

Content-based Filtering (CBF) 기법은 콘텐츠 속성에 대한 사용자의 선호도를 바탕으로 콘텐츠를 추천하는 방법이다[2][3]. 사용자로부터 선호도 정보를 직접적으로 입력 받거나, 사용자의 과거 데이터를 기반으로 선호 모델을 생성한다. 그리고 콘텐츠의 속성을 사용자 선호 모델과 비교하여 그 콘텐츠에 대한 점수를 결정한다. 이 접근법은 콘텐츠의 속성을 명확히 정의할 수 있는 대상에 대하여 효과적으로 사용할 수 있다는 장점이 있으나, 콘텐츠의 속성을 정의하고 사용자 성향을 파악할 수 있는 정보를 얻는 것이 쉽지 않은 단점이 있다.

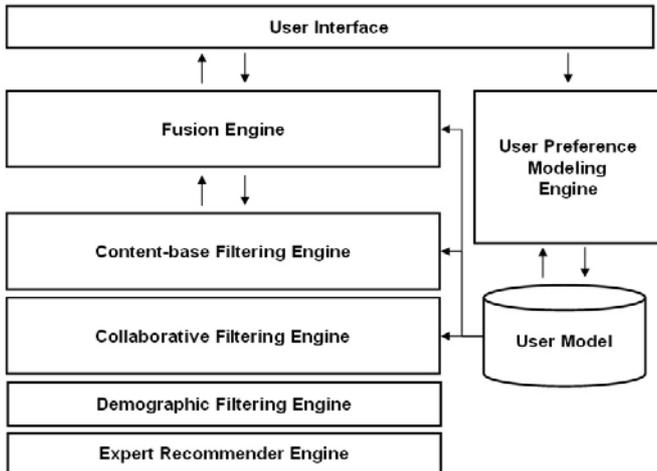
Collaborative Filtering (CF) 기법은 사용자와 비슷한 성향을 나타내는 사용자 집단을 선별하여, 그 집단에서 높은 점수를 받은 콘텐츠를 사용자에게 추천하는 방법이다[4][5]. 일반적으로 CF 기법에서는 아이템에 대한 분석은 필요하지 않으며, 사용자와 아이템 간의 평가정보(Rating)를 바탕으로 비슷한 사용자 집단을 구성한다. CF 기법은 아이템 속성에 대한 사용자의 선호도 모델을 구축할 필요성이 없으며, 아이템에 대한 분석과 속성 정의가 필요치 않아 적용할 수 있는 영

역이 넓다는 장점을 갖고 있다. 그러나 CF 기법은 사용자와 아이템 간에 충분한 평가정보를 갖고 있어야 적절한 추천이 가능하며, 평가정보가 없는 새로운 아이템에 대해서는 추천이 불가능하다는 단점이 있다.

CBF 기법과 CF 기법의 단점을 보완하고 보다 정확한 추천을 하기 위하여 두 가지 기법을 모두 사용하는 연구가 진행되었다. 혼합방법에 따라 한 기법의 결과를 다른 기법에 사용하는 Hybrid System[4][6]과 두 기법의 결과를 조합하여 추천하는 Combination System[7]이 있다.

3. 시스템 설계

본 논문에서는 사용자의 인포테인먼트 서비스를 위한 추천 시스템으로서 적응형 영화 콘텐츠 추천 시스템을 설계하였다. 적응형 영화 콘텐츠 추천 시스템은 TV 프로그램, 음악, 교육강좌와 같은 다양한 콘텐츠 추천에 적용할 수 있는 일반적인 구조를 갖는다. 아래 [그림 1]은 영화 콘텐츠 추천을 위한 적응형 시스템의 구조를 보여준다.



[그림 1] 적응형 콘텐츠 추천 시스템 구조

적응형 영화 콘텐츠 추천 시스템에서는 4 개의 추천 엔진을 Fusion 하여 사용자에게 적절한 영화를 추천한다. 또한 사용자의 영화선택 정보와 영화에 대한 반응을 학습하여 그에 따라 사용자 선호도 모델을 실시간 업데이트 한다. 이 사용자 모델은 다시 추천 엔진에서 영화를 추천하는데 필요한 선호도를 제시한다.

3.1 CBF Engine

CBF 엔진에서는 장르, 감독, 배우와 같이 객관적으로 설명 가능한 영화의 외적 요소 이외에 영화의 스토리, 연기력, 음악, 영상과 같은 내적 요소에 대한 사용자의 선호도를 평가하고 사용자의 각 요소에 대한 민감도 모델을 확립하여 영화를 추천한다. 또한 이러한 내적 요소와 각본, 배우, 음악, 영상감독과 같이 내적 요소를 결정하는 결정 인자와의 관계를 온톨로지를 통해 구축하여 보다 정확하고 다양하게 평가할 수 있도록 한다.

CBF Engine 은 먼저 영화정보와 사용자 모델로부터

각 영화에 대한 각 요소들의 점수를 계산한다. 영화들은 여러 장르의 속성을 갖고 있으며, 그 비율과 여러 장르들에 대한 선호도를 통해 그 영화의 장르 점수를 결정한다.

$$S_i = \sum_j R_{ij} P_{ij}$$

i : 영화의 요소 (장르, 감독, 배우)

j : i 의 하위분류 (i 가 장르이면, j 는 코미디, 액션 등)

S_i : i 에 대한 점수

R_{ij} : 요소 i 의 하위 분류 j 의 비율

(요소 i 가 장르이면, j 는 코미디, 액션 등)

P_{ij} : 요소 i 의 하위 분류 j 에 대한 사용자의 선호도

영화의 각 요소들에 대한 점수를 계산하면, 각 요소들에 대한 사용자의 민감도를 계산한다. 민감도는 사용자가 장르, 감독, 배우 등에 영향받는 정도를 의미하며, CBF 점수를 계산하는데 가중치로 사용된다. 민감도는 상위 50%의 하위분류 값들의 평균과 하위 50%의 하위분류 값들의 평균의 차이를 통해 계산된다. 즉, 사용자의 평가점수의 편차가 큰 요소에 대해서는 사용자 선호도의 차이가 크다고 가정하였다.

$$w_i = \frac{(\sum_{상위50\%j} P_{ij} - \sum_{하위50\%j} P_{ij})}{N_i}$$

w_i : i 에 대한 사용자의 민감도

N_i : i 의 하위 분류 j 의 개수

민감도의 값들을 합이 1 이 되도록 보정한 후, 앞에서 구한 요소들의 점수에 가중치로 사용하여 CBF score 를 계산한다.

$$sw_i = \frac{w_i}{\sum_i w_i}$$

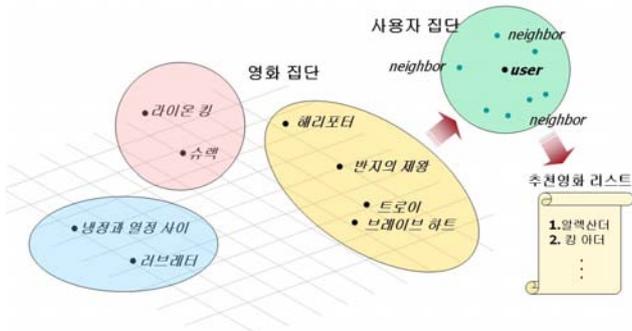
$$CBF_{SCORE} = \sum_i sw_i S_i$$

w_i : 요소 i 에 대한 사용자의 민감도

3.2 CF Engine

본 논문에서는 사용자와 비슷한 성향을 갖는 사용자 집단을 선별하여 그 집단에서 높은 점수를 받은 콘텐츠를 추천하는 방법(CF-user)과 사용자가 재미있게 본 영화와 유사한 영화 집단을 선별하여 사용자가 선호할 콘텐츠를 추천하는 방법(CF-item)을 혼합한 2-way Collaborative Filtering 기법[6][8]을 새로운 접근법으로 제시한다.

CF 엔진에서는 먼저 영화들에 대하여 아이템 CF 기법을 활용하여 추천 영화들을 선별하고 이에 대한 사용자 CF 기법을 적용하여 영화를 추천한다[그림 2].



[그림 2] CF 기법을 이용한 추천 방법

① 유사한 영화 찾기

사용자가 좋은 평가를 내린 영화목록을 Seed Movie 로 사용하여 그 영화들과 비슷한 영화들을 선별한다. 피어슨 상관계수를 계산하여 기준 상관계수 k 보다 높은 유사도를 갖는 영화들을 하나의 영화 집단으로 만든다.

$$Corr(A, B) = \frac{\sum_i (A_i - M_i)(B_i - M_i)}{\sqrt{\sum_i (A_i - M_i)^2} \sqrt{\sum_i (B_i - M_i)^2}}$$

$$-1 \leq Corr(A, B) \leq 1$$

- A: 기준 영화 (사용자가 좋은 평가를 내린 영화)
- B: 기준 영화에 대한 유사도를 측정할 영화
- A_i: 영화 A에 대한 사용자 i의 점수
- B_i: 영화 B에 대한 사용자 i의 점수
- M_i: 사용자 i의 모든 영화에 대한 평균 점수

영화 집단 간에 같은 영화가 있는 경우엔 두 집단을 합하여 하나의 집단으로 합한다.

② 비슷한 성향을 갖는 사용자 집단 구성

앞 단계에서 구한 각 영화 집단을 대상으로 유사한 사용자 집단을 구한다. 아이템 전체에 대하여 유사한 성향을 갖는 사용자를 찾아 추천하는 것보다 해당 하는 아이템에 대하여 비슷한 성향을 갖는 사용자를 대상으로 하는 것이 더 효과적이기 때문이다[5]. 유사 정도를 판별하기 위한 방법으로는 영화 찾기 단계와 마찬가지로 피어슨 상관계수를 이용한다.

③ 영화 추천

각 사용자 집단에서 결정된 점수들과 사용자 집단의 가중치를 고려하여 하나의 추천 영화 리스트를 작성한다. 여기서 사용자 집단의 가중치는 그 사용자 집단에 속한 seed movie 수로 결정한다.

3.3 Demographic Filtering Engine

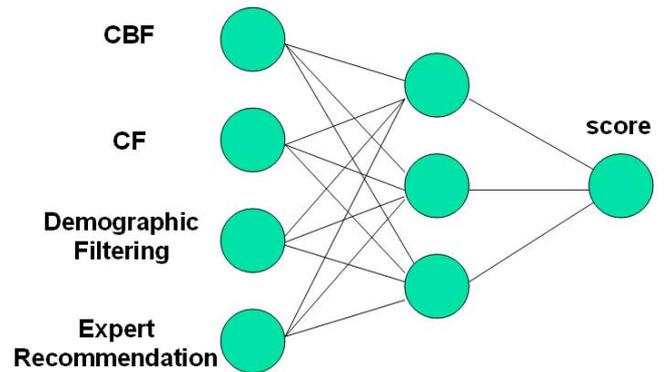
Demographic Filtering 기법은 사용자의 기본 정보인 나이, 성별, 직업을 바탕으로 비슷한 사용자들이 선호하는 영화를 추천한다.

3.4 Expert Recommender Engine

전문가 추천 엔진은 전문가들의 영화에 대한 평가를 바탕으로 영화를 추천한다.

3.5 Fusion Engine

Fusion Engine 은 각 엔진에서 추천된 결과를 하나의 추천 목록으로 Fusion 하는 역할을 한다. 사용자가 어떤 엔진으로부터 추천된 결과를 선호하는지에 대한 선호모델을 사용하여 각 엔진에서 나온 영화 목록을 재구성한다. 과거 사용자의 선택기록을 인공신경망(ANNs)의 EBP(Error Back-propagation) 알고리즘을 사용하여 현재 조건하에서의 최종 점수를 계산한다[그림 3].



[그림 3] 인공신경망을 이용한 Fusion Scoring

3.6 User Preference Modeling Engine

사용자의 영화 선택, 영화에 대한 반응을 관찰하여 사용자 로그를 축적한다. 이 사용자 로그를 바탕으로 사용자 선호도 모델을 업데이트 한다. 사용자의 평가를 반영하기 위해 Exponential Smoothing 기법을 사용한다.

$$W_t = \frac{w_1 P_t + w_2 P_{t-1} + \dots + w_n P_{t-n+1}}{w_1 + w_2 + \dots + w_n}$$

$$w_1 = 1, w_2 = a, w_3 = a^2, \dots, w_n = a^{n-1}$$

$$0 \leq a \leq 1$$

Exponential Smoothing 기법은 시간에 따른 가중치를 부여하여 성향의 변화를 반영할 수 있으며, 또한 업데이트 식이 간단하다는 장점이 있다.

4. 시스템 구현

적응형 영화 추천 시스템을 구현하기 위하여, 100 개의 영화를 대상으로 과거에 본 영화에 대한 점수를 입력하도록 설문조사를 실시하였다. 설문조사의 결과를 바탕으로 개인 사용자 선호도 모델을 구축하였으며, 이를 바탕으로 영화를 추천한다. 또한, 영화에 대한 사용자의 평가는 즉시 선호도 모델에 반영된다. [그림 4]은 영화 추천 엔진을 통해 나온 결과를 Fusion 하여 상위 6 개의 영화를 추천하는 화면이다.



[그림 4] 영화 추천 결과 화면

5. 결론

본 논문에서는 인포테인먼트 서비스 중의 하나인 영화 서비스에 대하여 적응형 콘텐츠 추천 시스템을 구현하였다. CBF 엔진에서는 콘텐츠 요소에 대한 가중치를 부여하여 사용자 선호도를 보다 정확히 반영하도록 하였으며, CF 엔진에서는 아이템 기반의 CF와 사용자 기반의 CF를 혼합한 새로운 방법론을 제시하였다. 사용자 선호도 모델 엔진에서는 Exponential Smoothing 방법을 사용하여 시간에 따른 사용자의 선호도 변화를 실시간으로 반영할 수 있게 하였다.

향후에는 영화 콘텐츠 추천 서비스와 함께 다른 콘텐츠 추천 서비스에 공통적으로 활용할 수 있는 사용자 선호도 모델을 구축하여 인포테인먼트 서비스를 위한 적응형 콘텐츠 추천 시스템을 개발하고자 한다.

Acknowledgement

이 논문은 산업자원부 지원으로 수행하는 21세기 프론티어 연구개발사업(인간기능 생활지원 지능로봇 기술개발사업)의 일환으로 수행되었습니다.

참고문헌

- [1] Amazon.com, available at <http://www.amazon.com>
- [2] C. Basu, H. Hirsh, & W. Cohen, "Recommendation as Classification: Using Social and Content-based Information in Recommendation," In Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI-98), pp. 11-15, 1998.
- [3] M. J. Pazzani, "A Framework for Collaborative, Content-based and Demographic Filtering," Artificial Intelligence Review, 13(5/6), pp. 393-408, 1999.
- [4] M. Balabanovic, Y. Shoham, "Fab: Contents-Based, Collaborative Recommendation," Communications of the ACM, vol. 40, no. 3, pp. 66-72, 1997.
- [5] J. Herlocker, J. Konstan, A. Borchers, and J. Riedle, "An Algorithmic Framework for Performing Collaborative Filtering," In Proceedings of the 22nd ACM-SIGIR International Conference on Research and Development

in Information Retrieval, 1999.

- [6] Y. Li, L. Lu and L. Xuefeng, "A Hybrid Collaborative Filtering Method for Multiple-interests and Multiple-contents Recommendation in E-commerce," Expert System with Applications, vol. 28, pp. 67-77, 2005.
- [7] 이수정, 이형동, 김형주, "사용자 경향에 기반한 동적 추천 기법 : 영화 추천 시스템을 중심으로," 한국정보과학회, 31(2), 2004.
- [8] 박지선 외 3명, "2-way 협동적 필터링 방법을 이용한 예측 알고리즘," 정보과학회지 논문지, 29권, 10호, pp. 669-675, 2002.