

협업 필터링 방안 분석 및 적용 분야 연구

이승희*, 박영호*

*숙명여자대학교 멀티미디어학과

e-mail : ariseshine@sookmyung.ac.kr, yhpark@sookmyung.ac.kr

A Study on Collaborative Filtering Analysis and Application

Seung-Hee Lee*, Young-Ho Park*

*Dept. of Multimedia Science, Sookmyung Women's University

요 약

최근 사용자의 취향에 맞는 콘텐츠를 필터링하여 자동으로 추천하는 연구가 활발히 진행되고 있다. 참여형, 개방형, 공유형 서비스들의 증가와 함께 웹 3.0 시대에는 더욱 지능화되고 개인화된 서비스가 중요시되고, 이를 위한 맞춤형 정보 제공 연구가 필수적이다. 본 논문에서는 사용자 맞춤형 추천 방법의 대표적인 기술인 협업 필터링(Collaborative Filtering) 방안 분석에 대해 설명하고, 협업 필터링 방법의 적용 연구를 설명한다.

1. 서론

콘텐츠의 증가로 인하여 사용자가 원하는 콘텐츠를 선택하기가 어려워짐에 따라, 사용자의 취향에 맞는 콘텐츠를 필터링하여 자동으로 제공하는 것이 요구되고, 이를 위한 사용자 맞춤형 추천방법에 관한 연구들이 활발히 진행되고 있다. 사용자의 특성을 파악하여 개인화된 정보를 제공하는 맞춤형 정보 제공은 사용자가 원하지 않는 검색 결과를 제외함으로써 검색의 정확성과 사용자의 만족도 향상을 가능하게 한다.

웹 3.0 시대에 지능화, 개인화된 서비스를 위한 사용자 중심의 검색 및 추천은 매우 중요한 문제이다. 이러한 사용자 맞춤형 추천 방법의 대표적인 기술이 전자상거래 분야에서 가장 보편적으로 사용되고 있는 협업 필터링(Collaborative Filtering) [1], [2] 기술이다.

협업 필터링 방법은 도서, 상품, 영화와 같은 인터넷 쇼핑몰에서의 상품 추천이나 웹 사이트, 웹 페이지 추천, 음악 추천, 디지털 TV의 프로그램 추천 등에서 유용하게 사용되고 있다.

본 논문에서는 협업 필터링의 방법을 설명하고, 그 방안을 분석한다. 또한 협업 필터링을 적용한 연구에 대하여 살펴본다.

본 논문의 구성은 다음과 같다. 2 장에서는 협업 필터링 방안의 분석 연구에 대하여 설명하고, 3 장에서는 협업 필터링을 적용한 분야 연구에 대하여 설명한다. 마지막으로 4 장에서는 본 논문의 결론을 맺는다.

2. 관련 연구

본 장에서는 사용자 맞춤 추천의 일반적인 분류와 그 방법을 설명한다.

사용자 맞춤 추천 방법은 주로 내용기반 추천(Content-based Recommendation) 과 인구통계학적 추천

(Demographic Recommendation), 그리고 협업적 추천(Collaborative Recommendation or Social Filtering)으로 분류한다 [3]. 내용 기반 추천은 사용자가 이전에 이용한 콘텐츠와 내용 정보가 유사한 다른 콘텐츠를 추천하는 방법이다. 그러나 콘텐츠들의 내용 정보들을 모두 비교하기 어려우므로 분석 시간이 많이 사용되며, 텍스트 이외의 내용에 대해서는 추천이 어렵다. 인구통계학적 추천 방법은 사용자에 대한 일반적인 정보를 기준으로 정보를 제공한다. 협업적 추천, 즉 협업 필터링 방법은 사용자들 간의 유사한 성향의 사용자들이 공통적으로 선호하는 콘텐츠를 추천하는 방법이다.

3. 본론

본 장에서는 협업 필터링 방안을 분석하고, 협업 필터링의 적용 분야를 설명한다.

3.1 협업 필터링 방법

협업 필터링(Collaborative Filtering)은 많은 사용자들의 기호(Preference) 정보를 추출하여, 사용자들의 관심사들을 자동적으로 예측하게 하는 방법이며, 특정 사용자의 정보만이 아니라 많은 사용자들로부터 수집한 정보를 사용한다는 것이 특징이다[4].

협업 필터링의 방법으로는 주로 사용자 기반 협업 필터링 기법(User-based Collaborative Filtering)과 아이템 기반 협업 필터링 기법(Item-based Collaborative Filtering)이 가장 많이 쓰이는 필터링 기법이라 할 수 있다 [1].

사용자 기반의 협업 필터링 방법은 모든 사용자들이 평가한 랭킹 데이터를 이용하여, 사용자 간의 유사도를 계산하는 방법이다. 그러나 제품의 수가 무수히 많은 경우에는 사용자간의 중첩이 발생할 경우가

드물기 때문에 효과적인 추천 결과가 나타나기 어렵다.

데이터가 많이 발생하는 추천 시스템의 경우에는 사용자 기반 방법보다 아이템 기반의 방법이 더 알맞게 사용된다.

사용자 기반의 협업 필터링은 기존의 어느 정도 예측이 가능한 고객들과 유사한 패턴을 가진 사용자들을 찾고, 다음 단계로 기존 고객들의 행동을 예측하기 위하여 유사하다고 생각된 사용자들의 행동을 수치화하여 사용하는 방법을 운영한다.

아이템 기반의 협업 필터링은 아이템 간의 상관관계를 결정하는 아이템 매트릭스를 만들고, 이 매트릭스를 사용하여 최신 사용자의 데이터를 기반으로 그 사용자의 기호를 유추한다.

3.2 유사도 평가 계산 방법

협업 필터링 방법에서 가장 중요한 유사도 평가를 위한 계산 방법은 다음과 같다. 이 유사도 평가 계산 방법은, 벡터의 내적 공식을 통해 $\cos\theta$ 값을 유도하여, 이 사이각인 θ 값의 크기를 측정한다. 사이각의 크기와 벡터의 방향으로 두 대상 간의 유사성을 분석하는 방법이다.

벡터 내적 공식을 이용하여 유사성을 비교하는 방법으로 Cosine-based Similarity 방법이 사용된다. 그림 1은 식을 유도하는 과정과 유사성을 비교하고자 하는 아이템 t 와 r 에 대한 식을 나타낸다.

$$\cos\theta = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| |\vec{b}|} \rightarrow -1 \leq \cos\theta = \frac{\sum_{k=1}^n a_k b_k}{\sqrt{\sum_{k=1}^n a_k^2} \sqrt{\sum_{k=1}^n b_k^2}} \leq 1$$

↓

Cosine-based Similarity

$$sim(t, r) = \frac{\sum_{i=1}^m R_{it} R_{ir}}{\sqrt{\sum_{i=1}^m R_{it}^2} \sqrt{\sum_{i=1}^m R_{ir}^2}}$$

(그림 1) Cosine-based Similarity 방법

이 방법보다 보다 객관적이고, 다양하고 많은 수의 사용자들의 평가를 고려한 방법으로 표준편차를 적용하여 응용한 Correlation-based Similarity 방법이 있다. 두 아이템 t 와 r 에 대한 유사성 비교를 그림 2와 같은 식을 통하여 계산한다.

$$-1 \leq \rho = \frac{\frac{1}{n} \sum_{k=1}^n (a_k - \bar{a})(b_k - \bar{b})}{\sqrt{\frac{1}{n} \sum_{k=1}^n (a_k - \bar{a})^2} \sqrt{\frac{1}{n} \sum_{k=1}^n (b_k - \bar{b})^2}} \leq 1$$

↓

Correlation-based Similarity

$$sim(t, r) = \frac{\sum_{i=1}^m (R_{it} - A_t)(R_{ir} - A_r)}{\sqrt{\sum_{i=1}^m (R_{it} - A_t)^2} \sqrt{\sum_{i=1}^m (R_{ir} - A_r)^2}}$$

(그림 2) Correlation-based Similarity 방법

두 방법에서 유사성 측정에 사용되는 값의 범위는 -1에서 1사이의 범위인데, 이는 $\cos\theta$ 의 범위 때문이며, 두 아이템을 비교한 계산 값이 1에 가까울수록 유사하며 -1에 가까울수록 유사하지 않다. -1인 경우는 벡터가 서로 반대 방향을 나타내기 때문이다.

3.3 협업 필터링 적용 연구

더욱 우수한 협업 필터링 기법을 개발하기 위한 많은 연구들이 진행되고 있으며, 기존 필터링 기법들을 결합한 하이브리드 기법 [5]도 연구되고 있다.

협업 필터링은 사용자들의 과거의 경향이 미래에서도 그대로 유지될 것이라는 전제로, 단순히 투표수로 평가하는 방법과 차별되어, 선호도와 관심 표현을 바탕으로 유사한 성향을 가진 사용자들을 구분한다. 유사 성향의 사용자를 그룹화하여, 그룹 안의 다른 사용자의 미래 행동을 예측할 수 있는 연구로 적용된다. 이는 서로 구매하지 않은 상품에 대한 추천은 물론, 사용자의 생활 패턴이나 기호 아이템, 이미지 추천 등 다양한 연구에 매우 알맞게 적용될 수 있다.

4. 결론

본 논문에서는 사용자 맞춤형 추천 방법인 협업 필터링의 방안 분석과 적용 연구를 설명하였다.

협업 필터링은 사용자 기반, 아이템 기반의 방법으로 사용되고 단순한 텍스트로 이루어진 속성 정보가 아닌 사용자의 선호도와 행동 양식을 바탕으로 추천하는 방법이다. 본 논문에서 설명한 유사도 평가 계산 방법을 통하여 유사 성향을 갖는 대상을 분별할 수 있으며, 협업 필터링이 갖는 단점을 보완하기 위하여 하이브리드 기법과 같은 연구들이 진행되고 있다. 협업 필터링 방법은 단순히 상품 아이템만이 아니라 무형의 아이템에 대해서도 적용할 수 있다는 장점을 가지므로 사용자들에게 더욱 만족스런 정보를 제공할 수 있을 것이다.

참고문헌

- [1] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based Collaborative Filtering Recommendation Algorithms," In Proc. the 10th International Conference on WWW, pp. 285-295, 2001.
- [2] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating Collaborative Filtering Recommender Systems," *ACM Transactions on Information Systems*, Vol.22, No.1, pp.5-53, 2004.
- [3] Pazzani, M., "A Framework for Collaborative, Content-Based and Demographic Filtering," *Artificial Intelligent Review*, Vol. 13, pp. 393-408, 1999.
- [4] N. Good, J. B. Schafer, J. A. Konstan, A. Borchers, B. Sarwar, J. Herlocker, and J. Riedl, "Combining Collaborative Filtering with Personal Agents for Better Recommendations," In Proc. the Conference on Artificial Intelligence, pp. 439-446, 1999.
- [5] Yu Li, Liu Lu and Li Xuefeng, "A hybrid collaborative filtering method for multiple-interests and multiple-content recommendation in E-Commerce," *Expert Systems with Applications*, Vol. 28, pp. 67-77, 2005.