추천시스템은 사용자의 필요와 취향을 고려하여 그에
적합한 새로운 상품이나 대신할만한 상품 등을 추천하는
시스템이다. 지금까지 제안된 대부분의 추천시스템들은
학습적인 필터링 기법을 쓰고 있는데, 이러한 시스템의
경우 사용자의 선호도 점수 정보가 부족하면 정확한
추천결과를 도출할 수 없다. 본 논문에서는 내장기반 필
터링 기법을 학습적 필터링 기법과 함께 사용하여 이와
같은 문제를 해결하고자 한다.

1. 서 론

전자상거래의 성장은 사용자에 대한 추천 알고리즘들
이 제안되게 하였고, 추천 알고리즘으로 정보 필터링 알고리즘
이 널리 쓰이게 되었다. 정보 필터링은 독자
의 주관적이며, 내장기반 필터링과 학습적
필터링이 그것이다. 내장기반 필터링 기법을 사용하는
추천시스템은 사용자가 선호도에 따라 선호할 것들
와 유사한 아이템들을 추천한다. 따라서 이러한 시스템들이 아이템
의 사용 자세에 대한 분석과 사용자의 취향과 관리
으로 인해 발생할 수 있는 문제를 해결하기
가능하다. 학습적 필터링 기법을 사용하는 추천시스템은 개
두 사용자에 그 취향이 비슷한 다른 사용자들이 선호하는
아이템들을 추천한다. 학습적 필터링의 성능은 고려대상
이 되는 아이템에 대한 선호정보들의 양에 의존한다.
내장기반 필터링 기법과 학습적 필터링 기법을 합쳐
하여 필터링 시스템의 성능을 향상시킬 수 있다. 최근의
연구에서 두 기법들을 혼합하여 추천시스템에 적용한 예
를 제공해 보았다.(1,2,3,4) 본 논문에서는 사용자
에 대한 취향과 필요에 부합하는 상품을 추천하기
위해 내장기반 필터링과 학습적 필터링을 결합한 기법을
사용하는 추천시스템을 제안한다.

먼저 기존의 연구들은 2절에서 살펴보고, 우리가 제안하
는 시스템을 3절에서 기술한다. 4절에서 결론을 기술함
으로 본 논문을 마무리한다.

2. 정보 필터링

2.1과 2.2절에서 내장기반 필터링과 학습적 필터링에
관한 기존의 연구들을 살펴보고, 2.3절에서는 이들을 혼
합하여 더욱 나은 성능을 도출하는 접근법을 고찰한
다.

2.1 내장기반 필터링

내장기반 필터링은 고려의 대상이 되는 아이템에 대한
내용에 기반한다. 예를 들어, 문헌들을 대상으로 하
는 필터링 시스템의 경우 용어의 동정방도와 그것이 사
용자의 해당문서에 대한 선호도와 어떤 관계에 있는지
가 파악되어야 한 내용이다.

내장기반 필터링이 적용된 추천시스템은 단지 해당 사
용자가 괜히 선호도를 표시한 아이템들의 내용 분석만
으로 그 내용이 이루어진 프로세스에 기초하여 추천한
다. InfoFnder와 NewsWeeder가 이러한 시스템의 예
이다.(4,5)

내장기반 필터링 기법만을 사용하는 추천시스템은 다
음과 같은 한계를 겪는다. 첫 번째 문제는 아이템의 내
용에 대한 단면적인 분석에 의존하며, 다른 복잡한 측면
에 몰 때 있을 수 있는, 해당 아이템이 사용자가 선호
하는 정도로 미치는 영향을 고려하지 못하게 된다.
예를 들어, 아이템이 그림의 경우 사진, 영화
등과 같이 예술적인 수준 뽑아라우저에 보도되며
걸리는 시간 등이 무시될 수 있다. 나머지 개별 사용자에게 특수화된다는 것이 또다른 문제점이다. 사용자는
자신이 괜히 선호하는 아이템이 보였던 것과 아
이템 자세의 특성이 비슷한 아이템들만 추천 받게 된다.

2.2 학습적 필터링

학습적 필터링은 사용자에게 그와 취향이 비슷한 사
용자들이 선호한 아이템들을 추천하여, 대개의 경우 이
때 아이템 내용 자세는 무시한다.

학습적 필터링은 추천시스템의 아이템 내
용 분석은 전혀 하지 않고, 사용자들 사이의 취향을 합
성하여 추천한다. GroupLens와 Ringo는 이러한 시스
템의 예이다.(6,7)

학습적 필터링 기법만을 사용하는 추천시스템은 내
용기반 필터링만을 사용하는 경우에 발생하는 문제들을
해결한다. 다른 사용자들의 선호도 정보를 바탕으로 다
른 종류의 아이템들만 추천 대상으로 고려할 수 있고,
내용기반의 특성이 주어져 해당사용자에 의해 선호되었
던 아이템의 것과 다른 경우에도 추천내용의 대상이 될
수 있다.

그러나 이 시스템에도 그 자체의 문제점이 있다.
선호도 정보가 부족한 경우 추천 성능이 매우 낮아
질 것이다. 새로운 아이템이 추가된 경우나 아이템들의
개수에 비해 사용자의 수가 너무 적은 경우 이와 같은
문제가 발생할 수 있다. 또한 취향이 복잡한 사용자에게는
추천기반이 적용되기 어렵다는 문제점이 있다.

2.3 내용기반 필터링과 학습적 필터링의 혼합

교과서 설명의 바와 같이, 내용기반 필터링을 사용
하는 추천시스템이나 학습적 필터링을 사용하는 추천
시스템은 모두 그의 전형적인 문제점들을 가지고 있다. 이
를 해결하기 위해 두 기법들을 혼합하여 추천시스템에 적
용하는 혼합기법의 접근법들이 제안되고 있다. 많은
연구에서 학습적 필터링 기법을 기본으로 사용하고 내용
기반 필터링의 장점을 더욱이 취하는 방법을 보인다.
이는 협력적 필터링 기법이 사용가능한 필터링 기법보다 성능 면에서 더 우수한 것으로 나타났기 때문이다.

Fab, GroupLens 그리고 Active WebMuseum 시스템은 이러한 혼합 정보 필터링 기법을 사용하고 있다. 
Fab 시스템을 위해 헬브에 이예트 기반 문서 필터링 시스템이다[1]. 이 시스템에서 사용자와의 일치 유사성은 선호되는 문서들끼리의 시간적 유사성을 사용자들의 선호도들로 결정된다. 이 때 문서의 카드와 문득의 원래 기반을 통해 발전된다. 
GroupLens 시스템은 유니버스를 위한 필터링 시스템으로, 에이전트들이 감지된 초기가임에 기반하여 인위적으로 
로 아이템들의 정수를 예측한다[3]. 
The Active WebMuseum은 화가들의 그림에 대한 경우시스템이다[9]. 
Active WebMuseum은 협력적 필터링을 사용하고 
여러 사용자들이 자신에게 적합하게 설계된 이동경로를 따 
라 그림들을 탐색할 수 있게 하는 서비스를 제공한다. 
이 시스템은 벌써미더 데터를 전송하는 핸드폰에서 사용 
하는 사람들과 가정이나 인터넷을 사용하는 사람들 사이 
로 서비스를 제공할 수 있는지를 보 
인한다.
3. 시스템 설계

본 논문에서는 협력적 소핑몰을 위한 추천 시스템
을 제안한다. 이 시스템은 크게 세 개의 모듈들로 구성 
된다(그림 1). 소핑몰 사이트에 로그인한 각 사용자에 
대해 시스템은 다음의 작업을 수행한다. 

![그림 1. 추천 시스템의 구조](image)

- (사용자 행위 수집 모듈) 상품의 상세한 내용을 보 
근나 보고 있는 상품을 구입하는 등, 사용자의 상품 
구입에 관계된 행위들을 수집한다.
- (상품 평가 모듈) 사용자의 상품에 대한 호감도를 
계산한다.
- (통신 모듈) 상품 평가 모듈에서 사용자의 행동 
로부터 뽑아낸 그룹의 상품정보들을 수집한다.
- (추천 모듈) 실행의 취향이 같은 사용자들을 찾는다.
- (추천 모듈) 같은 취향의 사용자들의 취향정보들을 
사용하여 현 사용자에게 상품을 추천한다.

이러한 모든 작업은 캐테고리 내 모든 상품에 대해 수행 
된다. 하위 절에서, 적용 기법들을 상세히 설명한다.

3.1 내용기반 필터링

내용 면에서 아이템들 사이의 비슷함과 다른 정도와 
그에 따른 사용자의 아이템들에 대한 선호도를 알기 위 
해서는 그 상관관계를 결정할 수 있어야 한다. 이러한 
상관관계는, 대개 내용기반 분석 결과 비슷한 아이템들 
은 사용자에 의해 비슷한 선호도 점수를 받다는 성 
과가 바탕으로 한다. 이러한 상관관계를 통해 어떤 사용 
자들의 아이템들에 대한 선호 선호도를 구하고 그에 적 
합한 아이템들을 선택할 수 있다.

예 소핑몰의 내용은 상품 아이템으로, 컴퓨터 시스템 
이 분석할 수 있는 문제는 채택, 계약, 가격대 
등이 있다. 이 논문에서는 가격대만을 다룬다. 두 아이 
템 사이의 가격대가 다른 것과 같이 구할 수 있다.

\[ price_j : \text{아이템 } j \text{의 가격대} \]

\[ distance_{price}(p, p') = |price_j - price_p| \]

근거간격은 다음과 같다.

\[ j = 1..N, \quad interval_i = [0, 1], \quad interval_j = [1, 2] \]

거리 값은 다음과 같다.

\[ C_i(j) = \{ i' \in I_u : distance_{price}(i, i') \in interval_j \} \]

\[ I_u : \text{사용자 } u \text{에 의해 점수 예측된 아이템} \]

사용자 u의 아이템 j에 대한 선호도 예측 점수는 다음과 같다.

\[ p_{\text{price}}(u, j) = \sum_{i \in I_u} \alpha_i \cdot \frac{r_{u, i}}{|C_i(j)|} \]

\[ r_{u, i} : \text{사용자 } u \text{의 아이템 } j \text{에 대한 예측 선호도} \]

3.2 협력적 필터링

지금까지 협력적 필터링 시스템에 관한 연구들이 많이 
있었다. 이러한 시스템들의 대부분은 사용자들의 의견 
을 아이템에 대한 수치형태의 정수로서 받아들이는 형 
으로 표현된다. 이 형식은 사용자에 대한 선호도 정 
수 예측하는 알고리즘들이 많이 제안되었다(10,7).

본 논문에서는 GroupLens 프로젝트에서 제안되어 
있는 사용되고 있는 Pearson 상관계수를 사용한다. 다 
음은 사용자 u의 아이템 j에 대한 선호도를 예측하는 식 
이다.

\[ I_u : \text{사용자 } u \text{가 점수 매긴 아이템들의 집합} \]

\[ r_{u, i} : \text{사용자 } u \text{의 아이템 } j \text{에 대한 점수} \]

\[ r_{u} : \text{사용자 } u \text{의 평균 점수} \]

\[ r_{u} = \frac{1}{|I_u|} \sum_{i \in I_u} r_{u, i} \]

두 사용자들 사이의 유사도는 사용자 u와 u의 점수 
베터들을 상관관계에 의해 결정된다.

\[ \rho(u, u') = \frac{\sum_{i \in I_u} (r_{u, i} - r_{u}) (r_{u', i} - r_{u})}{\sqrt{\sum_{i \in I_u} (r_{u, i} - r_{u})^2 \sum_{i \in I_u} (r_{u', i} - r_{u'})^2}} \]

\[ \rho \text{는 } -1 \text{과 } +1 \text{ 사이의 값을 가집니다. } 1 \text{에 가까운 값은} \]
높은 유사도를 0에 가까운 값을 낮은 유사도를 나타낸다. -1에 가까운 값은 대개 대립되는 의견을 보임을 의미한다.

예상되는 값은 다른 사용자들이 매진 점수들의 가중치 추정 평균이다. 가중치는 사용자 $u$와 다른 이들의 평

\[ U_i; \text{ 아이템 } j \text{에 점수를 매긴 사용자들} \]
\[ p_{\text{collab}}(u, i) = r_u + k \sum_{u' \in U_i} \rho(u, u')(r_{u', u} - r_u) \]
\[ \text{with } k = \sum_{u' \in U_i} \rho(u, u') \]

3.3 내용기반 필터링과 협력적 필터링 기법의 결합

내용기반 예측값 $p_{\text{proc}}$와 협력적 예측값 두 값과 같이 결합한다.

\[ p_{\text{comb}}(u, i) = \mu \cdot p_{\text{proc}}(u, i) + \mu_{\text{collab}} \cdot p_{\text{collab}}(u, i) \]
\[ \text{with } \sum_{\mu} = 1 \]

4. 결 론

본 논문에서 정보 필터링 기법을 사용하는, 웹기반 소

평을 위한 추천시스템을 설계하였다. 우리는 필터링을

위하여 협력적 필터링 기법을 추천 기법으로 하고 내용

기반 필터링 기법을 추가 적용하였다. 그리하여 저장되

어 있는 사용자의 선호도 점수 정보가 부족한 경우에

도 한 사용자의 아이템에 대한 선호도 예측을 수행할 수

있도록 하였다.

(참고 문헌)


