

협업 필터링 알고리즘에 관한 비교연구

이가배* · 이효맹* · 이현창* · 신성윤**

*원광대학교

**군산대학교

A Comparative Study on Collaborative Filtering Algorithm

Jiapei Li^{1*} · Xiaomeng Li¹ · HyunChang Lee¹ · SeongYoon Shin²

¹Wonkwang University

²Kunsan National University

E-mail : *jiapei@gmail.com *hclglory@gmail.com

요 약

추천시스템 중 가장 대표적인 협업 필터링은 여러 아이템에 대한 사용자 평가 데이터를 이용하여 공통적 패턴을 찾고 특정 사용자에게 대한 선호 아이템을 예상하여 추천하는 기법이다. 본 논문에서는 모두 5가지 알고리즘을 사용하였다. Recall-Precision, FPR-TPR, RMSE, MSE, MAE 등 지표를 측정하였다. 실험 결과를 보면 MovieLens 데이터를 이용해 사용자에게 기반 협업 필터링 알고리즘을 적용해 영화를 추천하는 것이 좋은 효과를 얻고 있다.

ABSTRACT

In recommendation system, collaborative filtering is the most important algorithm. Collaborative filtering is a method of making automatic predictions about the interests of a user by collecting preferences or taste information from many users. In this paper five algorithms were used. Metrics such as Recall-Precision, FPR-TPR, RMSE, MSE, MAE were calculated. From the result of the experiment, the user-based collaborative filtering was the best approach to recommend movies to users.

키워드

협업 필터링; user-based CF; item-based CF; SVD

1. 서 론

추천시스템의 목적은 고객의 취향을 분석하여 높은 선호도가 예상되는 아이템을 추천 하는 것이다. 일반적으로 추천시스템은 두 가지 테크닉으로 나뉜다. 하나는 내용 기반의 필터링(content-based filtering)이고 다른 하나는 협업 필터링(collaborative filtering)이다. 내용 기반의 필터링은 내용을 분석하여 특정된 사용자가 아이템에 대한 선호도를 예측하는 것이다. 협업 필터링은 내용을 사용하지 않고 오직 사용자의 아이

템에 대한 과거 평가 데이터(rating)만을 분석하여 선호도를 판단하는 것이다. 협업 필터링은 내용 기반의 필터링 가지고 있는 여러 가지 문제점을 보완하며 좋은 성능을 보인다. 협업 필터링 기반의 추천 시스템은 음악, 영화, 도서 등의 분야에서 활용되고 있으며 Amazon.com, Last.fm 등 상업적으로 성공한 여러 전자상거래사이트에서 적용되고 있다.

협업 필터링은 사용자들의 기초정보와 고객들의 선호도를 바탕으로 선호도에서 비슷한 패턴을

보이는 고객들을 분류하는 기법이다. 즉 이 기법은 비슷한 취향을 가진 고객들에게 아직 구매하지 않은 상품들을 교차 추천하거나 분류된 고객의 프로파일(profile)에 따라 상품을 추천하는 형태이다. 협업 필터링은 정확한 개인별 추천과 예측하지 못한 상품을 추천할 수 있다는 장점을 가지고 있다. 또한 이 기법은 추천시스템에 적용할 때 사용자의 의견을 반영할 수 있다는 장점으로 인해 최근에 많은 연구가 진행되고 있다.

II. 본 론

협업 필터링은 일반적으로 이웃 기반의 협업 필터링(neighborhood-based collaborative filtering)과 모델 기반의 협업 필터링(model-based collaborative filtering)과 하이브리드 모델 기반의 협업 필터링(hybrid model-based collaborative filtering)으로 나눌 수 있다. Fig.1.

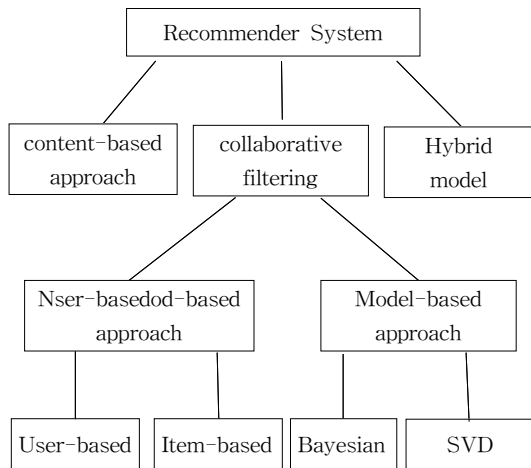


그림 1. 추천시스템 분류

이웃 기반의 협업 필터링은 일반적으로 사용자 기반의 협업 필터링(user-based collaborative filtering)과 아이템 기반의 협업 필터링(item-based collaborative filtering)으로 나눌 수 있다. 사용자 기반의 협업 필터링은 특정 사용자가 특정 아이템에 대한 선호도를 예측할 때 다른 아이템들에 대하여 대하여 비슷한 사용자가 특정 된 아이템에 대한 선호도에 의지하여 예측하는 기법이다. 즉 사용자 u 의 아이템 I 에 대한 평가점수 $r_{u,i}$ 를 예측할 때, 다른 아이템에 대한 평가점수의 유사도가 가장 높은 k 명의 사용자 v 를 추출

하고 I 에 대한 이웃 사용자 평가점수 $r_{v,i}$ 를 이용하여 평가점수 $r_{u,i}$ 를 예측한다.

$$r_{u,i} = \bar{r}_u + \frac{\sum_{v \in N(u;i)} s(u,v)(r_{v,i} - \bar{r}_v)}{\sum_{v \in N(u;i)} s(u,v)}$$

위 식에서 $N(u;i)$ 은 특정 사용자 u 와 i 비슷한 사용자의 집합이며, $s_{u,v}$ 는 사용자 u 와 사용자 v 의 유사도를 나타낸다. 본문은 코사인 유사도 기법을 이용하여 사용자 u 와 사용자 v 의 유사도를 계산한다.

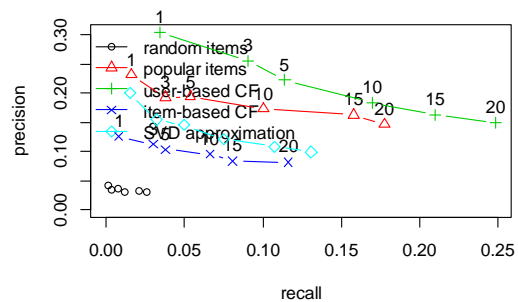
$$s(u,v) = \cos(\vec{u}, \vec{v}) = \frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\| \cdot \|\vec{v}\|}$$

모델 기반의 협업 필터링에서 사용자 선호도를 예측하기 위해 주로 Bayesian, clustering, singular value decomposition 등의 알고리즘이 사용된다. 본문에서는 특이값 분해 이용한다. $n \times m$ ($n \geq m$) 행렬 R 의 rank가 r 일 때, R 의 특이값 분해는 다음과 같이 정의한다.

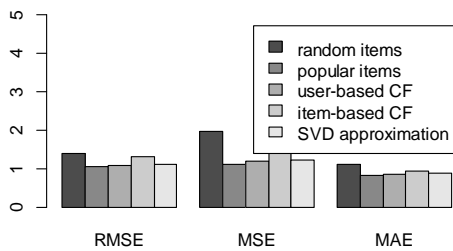
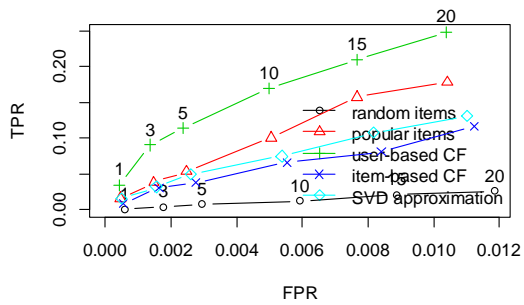
$$R = U \Sigma V'$$

III. 결론

실험에서 사용한 데이터는 최근에 많 연구 되고 있는 영화 평가 MovieLens 100K 데이터이다. MovieLens 데이터는 943개 사용자가 1682개 영화에 대해 약 100,000개 평가점수를 포함하고 있으며 평가점수의 범위는 1,2,3,4,5, 다섯 개 등급으로 나누어진다.



실험에서 모두 5가지 방법을 사용하였다. Recall-Precision, FPR-TPR, RMSE, MSE, MAE 등 지표를 측정하였다. 그림은 평가데이터에 대한 실



험 결과를 보여준다. 실험 결과를 보면 MovieLens 데이터를 이용해 사용자에게 기반 협업 필터링 알고리즘을 적용해 영화를 추천하는 것이 좋은 효과를 얻고 있다.

참고문헌

- [1] Candillier L, Meyer F, Boulle M. Comparing State-of-the-Art Collaborative Filtering Systems[J]. Machine Learning and Data Mining in Pattern Recognition, 2007, 11 (4) : 548-562.
- [2] Vucetic S, Obradovic Z. Collaborative Filtering Using a Regression-Based Approach[J]. Knowledge and Information Systems, 2005, 7 (1) :1-22.