

Item-based Collaborative Filtering Recommendation Algorithms

https://www.researchgate.net/publication/2369002_Item-based_Collaborative_Filtering_Recommendation_Algorithms

0. Introduction

- 추천 시스템은 정보 과부하 문제를 해결하기 위한 핵심 기술로 자리 잡음
- Collaborative Filtering(CF)은 사용자들의 집단적 행동(평점, 구매 기록)을 활용하여 개인화 추천을 제공하는 대표적인 방법
- 기존의 User-based CF는 특정 사용자와 유사한 다른 사용자들을 찾아서 그들의 선호를 기반으로 추천하는 방식
- User-based CF는 사용자가 많아질수록 계산량이 폭발적으로 증가하고 희소 행렬(sparse matrix) 문제로 인해 정확도가 떨어지는 단점 존재
- 이러한 한계를 극복하기 위해 Item-based CF 알고리즘 제안
- 핵심 아이디어 : 사용자 대신 아이템 간의 유사도를 미리 계산하여 추천 시점에서는 단순한 매칭으로 빠른 추천 가능하게 하는 것

1. Overview

- Item-based CF는 두 가지 기본 가정을 전제로 함 :
 1. 사용자의 과거 행동은 미래 행동을 예측하는 데 유용하다.
 2. 사용자가 과거에 선호한 아이템과 유사한 아이템을 미래에도 선호할 가능성이 높다.
- 프로세스 :
 - 사전 처리 단계에서 아이템 간 유사도(similarity) 행렬 계산

- 추천 단계에서 사용자가 평가한 아이템을 기반으로, 유사도가 높은 다른 아이템을 추천
- 장점 :
 - 아이템 수는 사용자 수보다 변동성이 적어 계산 안정적
 - 온라인 추천 시 속도가 빠름(실시간 응답 가능)
 - 대규모 데이터셋에 잘 확장됨

2. Challenges

- 데이터 희소성 문제 : 대부분의 사용자-아이템 행렬은 비어 있는 값이 많음 → 유사도 계산 신뢰성 낮아짐
- 유사도 측정 방법 : 어떤 유사도 척도(Cosine, Pearson, Adjusted Cosine)를 쓰느냐에 따라 결과 크게 달라짐
- 정확성과 효율성 균형 : 사전 계산 과정은 시간이 오래 걸릴 수 있으며, 추천 품질을 보장하려면 trade-off 필요
- 실시간성 문제 : 새로운 아이템이나 사용자가 들어왔을 때 어떻게 빠르게 업데이트할 것인지 고려해야 함

3. Method

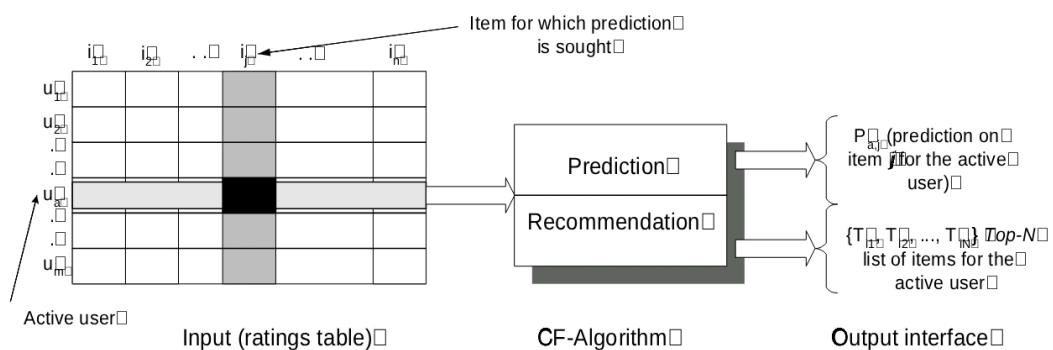
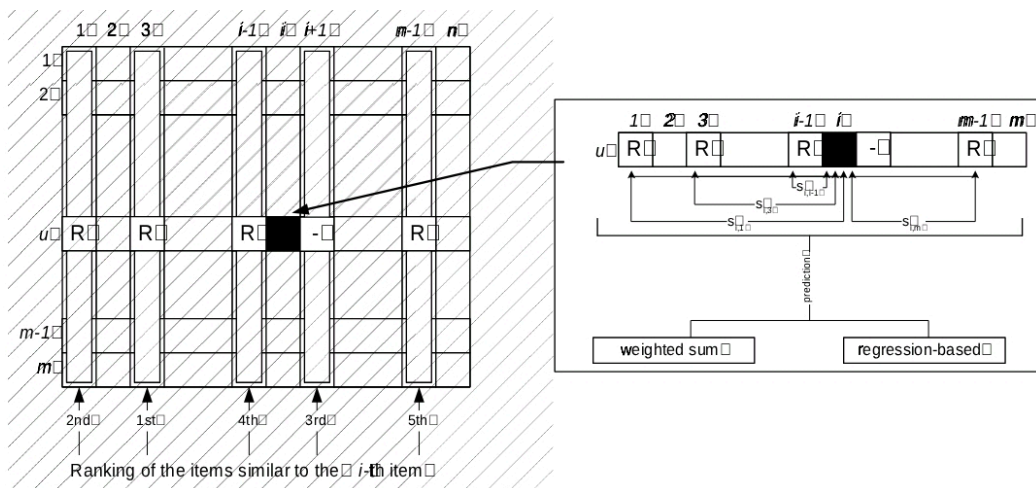
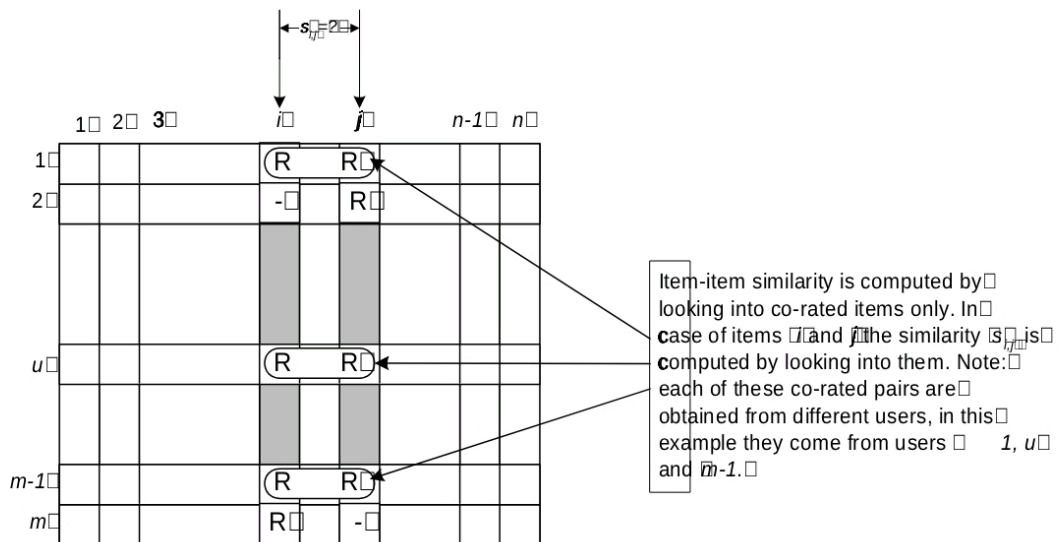


Figure 1: The Collaborative Filtering Process.



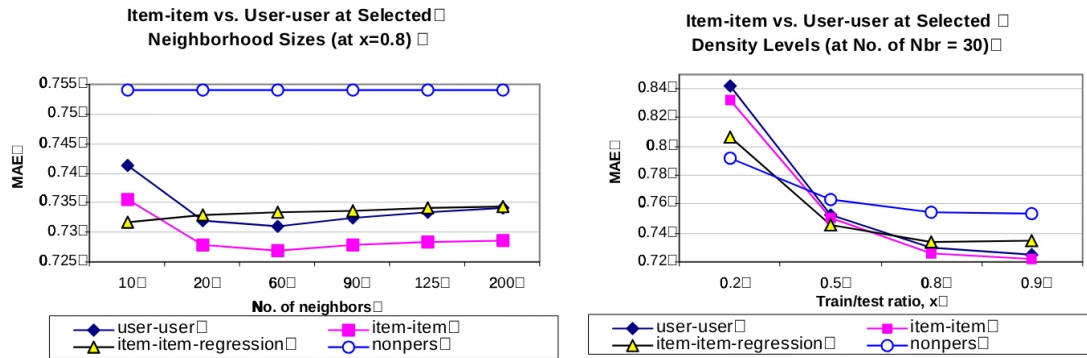
- 사용자가 아이템에 남긴 평점을 활용해 아이템 간 유사도를 계산하고, 이를 기반으로 추천 생성
- User-based CF와 달리 유저-유저 관계 대신 아이템-아이템 관계 활용 → 확장성 높음
- 아이템 유사도 계산
 - 사용자-아이템 평점 행렬 에서 아이템 에 대한 열 벡터 추출
- 평점 예측
 - 사용자가 아직 평가하지 않은 아이템 의 평점 예측은 그 사용자가 평가한 아이템 집합 과의 유사도를 활용
- 추천 리스트 생성
 - 사용자가 아직 소비하지 않은 아이템들 중, 예측 평점이 높은 순서대로 정렬
 - 상위 개 아이템을 최종 추천

- 계산 효율성
 - 아이템 수는 사용자 수보다 변동이 적으므로, 아이템-아이템 유사도 행렬을 오프라인에서 미리 계산 가능
 - 추천 시점에는 단순히 유사도 조회 및 가중합 계산만 하면 됨 → 실시간 응답 가능

4. Experiments

- 실험 데이터 :
 - MovieLens 데이터셋 사용
 - 약 943명의 사용자, 1,682개의 영화, 총 100,000개 이상의 평점 기록
 - 평점 범위는 1~5점
- 비교 대상 :
 - User-based CF
 - 무작위(Random) 추천
- 평가 지표 :
 - Mean Absolute Error (MAE)
- 실험 목표 :
 - Item-based CF가 User-based CF보다 정확성과 효율성 측면에서 우수한지 검증
 - 유사도 측정 방법(Cosine, Pearson, Adjusted Cosine)에 따른 성능 비교

5. Results



- Item-based CF가 User-based CF보다 전반적으로 낮은 MAE 기록 → 더 정확한 추천 제공
- 특히 대규모 데이터에서 확장성이 뛰어나고 계산 효율성이 높음
- User-based CF는 사용자 수가 증가할수록 성능 저하와 계산 부담이 커지는 반면 Item-based CF는 아이템 수 고정적이므로 안정적
- 유사도 척도 중 Adjusted Cosine이 가장 좋은 결과 도출
- 실험 결과는 실제 전자상거래 환경(예: 아마존 상품 추천)에서 Item-based CF가 효과적임을 뒷받침

6. Insight

- Item-based CF는 추천 시스템의 확장성 문제를 크게 개선한 기념비적 연구
- 사전 계산된 아이템-아이템 유사도를 활용해 실시간 추천 가능
- 오늘날 대형 추천 시스템(아마존, 넷플릭스)의 기반 기술 중 하나
- 본 연구는 CF의 새로운 패러다임 제시
- 사용자-사용자 관계보다 아이템-아이템 관계가 더 안정적
- cold-start 문제는 여전히 해결되지 않음