

Matrix Factorization Techniques for Recommender Systems

<https://ieeexplore.ieee.org/document/5197422>

0. Introduction

- 추천 시스템은 사용자와 아이템 간 상호작용 데이터를 기반으로 개인화 제공
- 기존 Collaborative Filtering은 주로 메모리 기반(user-based, item-based) 방법 사용
- Netflix Prize를 계기로 Matrix Factorization(MF) 방법이 각광받음
- MF는 사용자-아이템 평점 행렬을 저차원 잠재 요인(latent factor)으로 분해하여 추천
- MF 기반 알고리즘(Latent Factor Model, SVD, Regularized SVD 등)을 체계적으로 소개하고 추천 시스템 성능 향상 가능성 제시

1. Overview

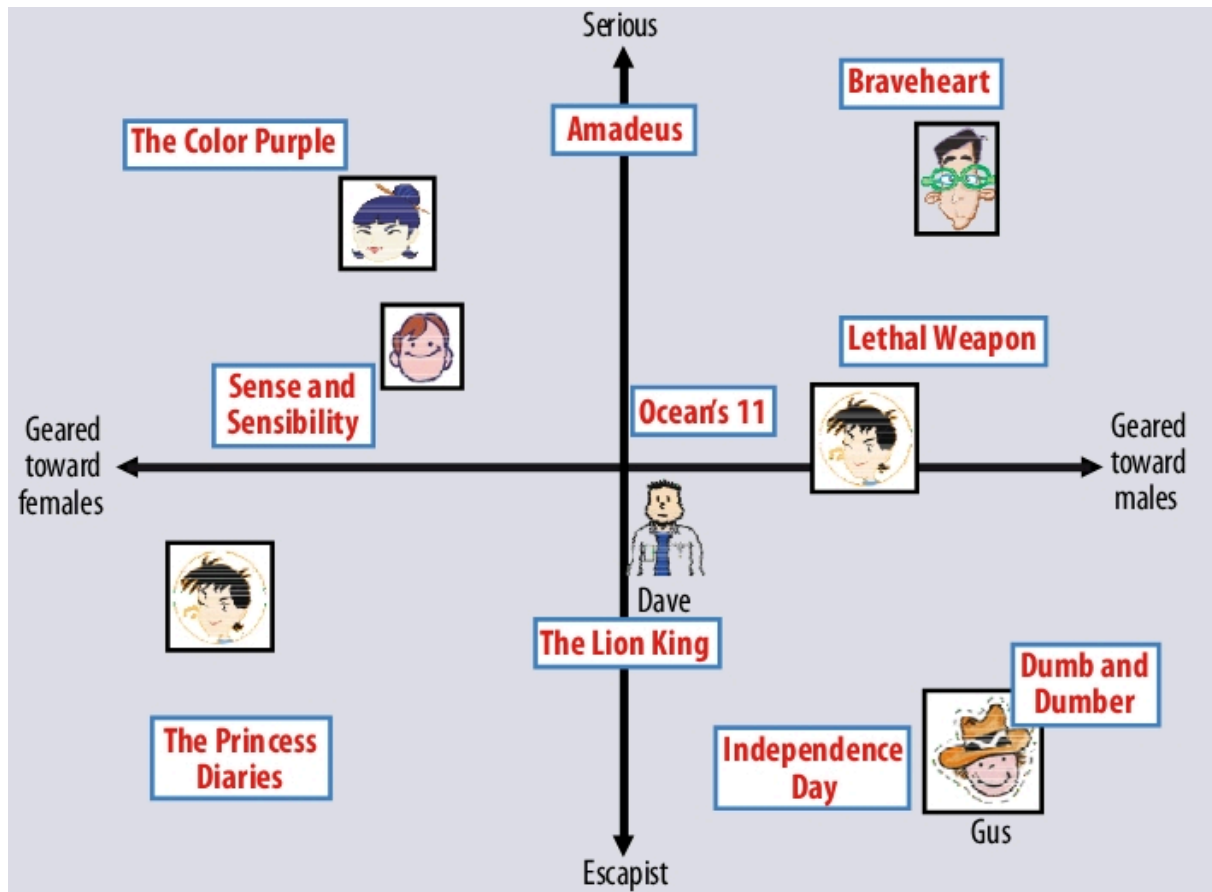
- 아이디어 : 사용자와 아이템을 잠재 공간(latent space)에 매핑
- 장점 : 희소 행렬에서 일반화 가능, 스케일링 용이, 사용자·아이템 특징 자동 학습
- 사용자 와 아이템 각각을 k 차원 잠재 벡터로 표현 → 내적(dot product)으로 예측 평점 계산

2. Challenges

- 평점 행렬은 대부분 희소 → 정확한 패턴 학습 어려움
- Cold-start 문제(새로운 사용자·아이템)
- 과적합 방지 위해 정규화 필요

- 대규모 데이터셋에서 빠른 학습 및 업데이트 방법 필요

3. Method

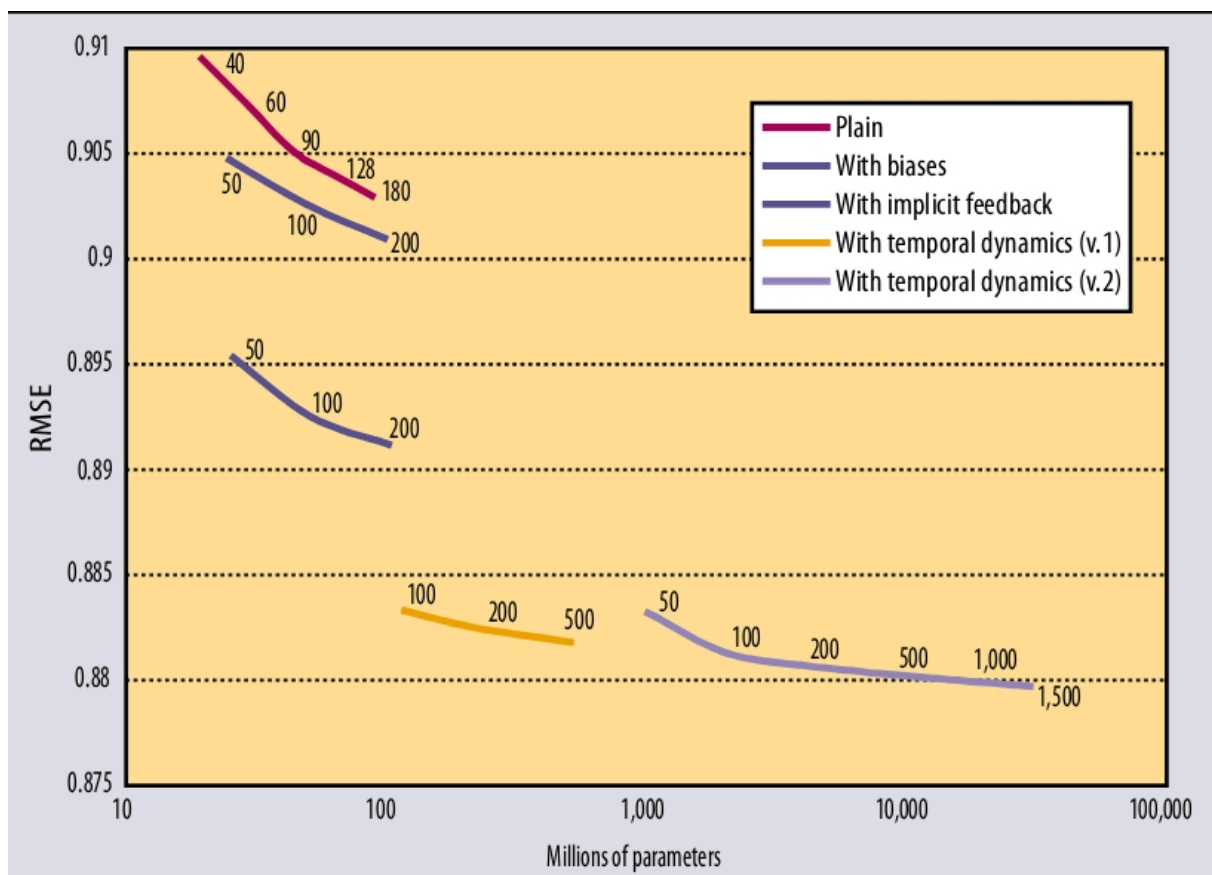


- 기본 모델 :
 - 사용자, 아이템 bias
 - 사용자, 아이템 latent vector
- 학습 :
 - SGD(Stochastic Gradient Descent) 또는 ALS(Alternating Least Squares) 사용
 - Regularization 추가하여 과적합 방지
- Netflix Prize 당시 확장 : Temporal dynamics 반영, implicit feedback 활용

4. Experiments

- 대표 데이터셋 :
 - Netflix Prize Dataset
 - 약 480,000명 사용자, 17,000편 영화, 1억 개 이상의 평점 기록
 - 평점 범위 : 1~5점
- 비교 :
 - 기존 User-based CF
 - Item-based CF
 - Latent Factor Models
- 평가 지표 :
 - RMSE 사용

5. Results



- MF 기반 모델, User-based/Item-based CF 대비 RMSE 크게 개선
- Latent factor 수(k) 증가 시 성능 향상, 단 과적합 가능 → 정규화로 제어
- Bias term(사용자, 아이템 편향) 포함 시 성능 개선
- Netflix Prize 대회에서 MF 계열 모델이 압도적 성능 → Ensemble 기법과 함께 우승
- 실험 결과:
 - MF는 대규모 추천 시스템에서 확장성과 정확도 모두 보장

6. Insight

- MF는 추천 시스템의 패러다임 전환을 이끈 기법
- 단순 CF를 넘어 사용자와 아이템을 잠재 공간에 투영하는 방식으로 일반화 능력 확보
- Netflix Prize 이후 거의 모든 대규모 추천 시스템에서 MF 또는 그 변형이 핵심
- Cold-start 문제 여전히 존재