추천 시스템

- 컨텐츠 기반
  - 아이템 컨텐츠 — 텍스트 분석
    - (컨텐츠 기반의) 유사도 (코사인 유사도 제일 많이 사용)
- 협업 필터링 — 사용자 - 아이템 평점 매트릭스 활용
  - 최근접 이웃 방식 (메모리 협업 필터링)
    - 사용자 기반 (최근접 이웃 방식)
      - 특정 사용자와 유사한 다른 사용자를 TOP-N으로 선정해 이 TOP-N 사용자가 좋아하는 아이템을 추천하는 방식.
    - 아이템 기반
      - 아이템이 가지는 속성과는 상관없이 사용자들이 그 아이템을 좋아하는지 / 싫어하는지의 평가 척도가 유사한 아이템을 추천하는 기초가 되는 알고리즘.
  - 잠재 요인 방식 — 사용자 - 아이템 평점 매트릭스 속에 숨어 있는 잠재 요인을 추출해 추천 예측을 할 수 있게 하는 기법이다. 대부분 다차원 행렬을 SVD와 같은 차원 축소 기법으로 분해하는 과정에서 잠재 요인을 추출하는데, 이러한 기법을 행렬 분해라고 한다.
    - 원본 행렬 → 사용자 - 잠재 요인 행렬 / 잠재 요인 - 아이템 행렬 → 내적곱 → RMSE (R, R')
    - SVD(NULL x) (SGD, ALS(NULL o))
    - L1: 라쏘 (예측 영향도 적은 피처의 회귀 계수 → 0)
    - L2: 릿지 (회귀 계수 값 작게)
    - 회귀
      - 과적합 방지 → 회귀 계수 크기 제어 (∵ 다중 공선성 완화, 노이즈 제거, 모델의 정확도↑, 속도↑)
      - RSS 비용함수 (기준) → 학습 데이터 잔차 오류 최소화 (하는데 사용하는 기법이 GD)
        - 사용은 비용함수 (목표) — 비용함수 식에서 W찾기 (페널티)
        - $= \min(RSS(w) + alpha * \|W\|_2^2$
        - 회귀 계수 값의 크기를 감소시켜 과적합 개선
        - → 규제

* 잠재 요인은 최소한의 오차를 RMSE 갖고 추출하고자 SVD, SGD를 활용한다.

- 사용자, 아이템 평점
  - 코사인 유사도
  - 개인화 추천 서비스
- 아이템 A와 가장 유사도가 높은 TOP-N개 아이템의 유사도 벡터 + 사용자 u의 A에 대한 실제 평점 벡터
  - 개인화된 예측 평점
  - RMSE