

[추천 시스템]

콘텐츠 기반 필터링 : 사용자가 특정한 아이템을 매우 선호하는 경우, 그 아이템과 비슷한 콘텐츠를 가진 아이템 추천.

협업 필터링

- 최근접 이웃 협업 필터링**
- 잠재 요인 협업 필터링**

} 사용자 - 아이템 평점 행렬 데이터에 의지해 추천 수행.

사용자가 아이템에 매긴 평점 점수나 상품 구매 이력과 같은 사용자 행동 양식만을 기반으로 추천을 수행.

[주요 목표] : 사용자 - 아이템 평점 매트릭스와 같은 축적된 사용자 행동 데이터를 기반으로 사용자가 아직 평가하지 않은 아이템을 예측 평가.

사용자 - 아이템 행렬 데이터.

↳ 행 : 개별 사용자, 열 : 개별 아이템

판다스의 pivot-table()

↳ 일반적으로 다차원 행렬이며, 최소 행렬 특성을 가진다.

< 최근접 이웃 협업 필터링 >

사용자 기반 최근접 이웃 방식 : 특정 사용자와 유사한 다른 사용자를 Top-N으로 선정해 이 Top-N 사용자가 좋아하는 아이템을 추천하는 방식.

↑ 정확도

아이템 기반 최근접 이웃 방식 : 아이템이 가지는 속성과는 상관없이 사용자들이 그 아이템을 좋아하는지 / 싫어하는지의 평가 척도가 유사한 아이템을 추천하는 기준이 되는 알고리즘.

* 유사도 측정을 위해 주로 코사인 유사도 이용.

< 잠재 요인 협업 필터링 >

: 사용자 - 아이템 평점 매트릭스 속에 숨어있는 잠재 요인을 추출해 추천 예측을 할 수 있게하는 기법.

대규모 다차원 행렬을 SVD와 같은 차원 감소 기법으로 분해하는 과정에서 잠재 요인을 추출하는데, 이러한 기법을 행렬 분해라고 함.

⇒ 잠재 요인을 기반으로 다차원 최소 행렬인 사용자 - 아이템 행렬 데이터를 저차원 밀집 행렬의 사용자 - 잠재 요인 행렬과 아이템 - 잠재 요인 행렬의 전치 행렬로 분해할 수 있으며, 이렇게 분해된 두 행렬의 내적을 통해 새로운 예측 사용자 - 아이템 평점 행렬 데이터를 만들어서 사용자가 아직 평점을 부여하지 않는 아이템에 대한 예측 평점을 생성하는 것이 잠재 요인 협업 필터링의 골자.

[행렬 분해]

: 사용자들의 행렬을 저차원 매트릭스로 분해하는 기법으로서 대표적으로 SVD, NMF 등이 있음.
NaN 값이 없는 행렬에만 적용할 수 있음.

if. 분할 문제
=> 확률적 경사 하강법이나 AIS 방식을 이용해 SVD 수행.

[확률적 경사 하강법]

: P와 Q 행렬로 계산된 예측 R행렬 값이 실제 R 행렬 값과 가장 최소의 오차를 가질 수 있도록 반복적인 비용 함수 최적화를 통해 P와 Q를 유추해내는 것.

(L2 규제)

$$\min \sum (r_{(u,i)} - p_u q_i^t)^2 + \lambda (\|q_i\|^2 + \|p_u\|^2)$$

(P', q')

$$p'_u = p_u + \eta (e_{(u,i)} * q_i - \lambda * p_u)$$

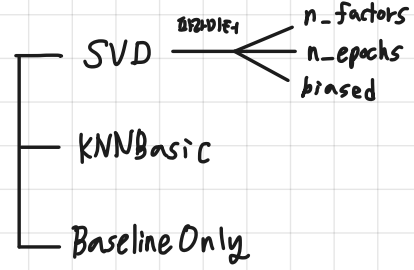
$$q'_i = q_i + \eta (e_{(u,i)} * p_u - \lambda * q_i)$$

< Surprise 모듈 >

: user-id, item-id, rating 데이터가 모두 레벨로 된 데이터 세트에만 적용할 수 있음.

* 데이터 컬럼 순서가 반드시 사용자 아이디, 아이템 아이디, 평점 순으로 되어있어야 함.

Surprise에서 추천 예측을 위해 자주 사용되는 추천 알고리즘 클래스.



* 베이스라인 평점

=> 개인의 성향을 반영해 아이템 평가에 편향성 요소를 반영하여 평점을 부과하는 것.

ex. 전체 평균 평점 + 사용자 편향 점수 + 아이템 편향 점수

