Posted by Contents&Education

2022.12.22.17:02

1 UP

[이경민팀 솔루션 공유]

사용한 모델 알고리즘

아이템 기반 최근접 이웃 협업 필터링 (데이터 전처리 방식 1번, 2번 적용)

잠재요인 협업 필터링 (데이터 전처리 방식 1번, 2번 적용)

newMF (데이터 전처리 방식 1번, 2번 적용)

newMF (데이터 전처리 방식 1번, 3번 적용)

데이터 전처리 방식

학습 데이터에 존재하는 album\_id 만으로 학습 및 추론 진행

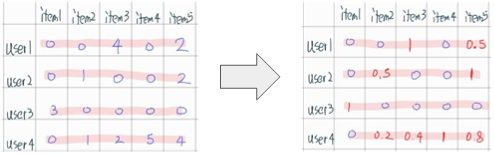
전체 albumid의 개수는 약 4만 개, user가 1번이라도 시청한 albumid의 개수는 약 2만 개

→ user가 1번이라도 시청한 album\_id 약 2만 개만 이용

모든 유저마다 각 album 별로 몇 번씩 봤는지 횟수를 측정하여 user-item count 행렬 생성

(평점 예측 문제로 변경 : 시청을 많이 한 경우 그 album에 대해 높은 평점을 주었다고 생각)

User마다 전체 시청 기록의 개수는 다르기 때문에 user 별로 count 횟수를 Max 1, Min 0으로 정규화 진행



전체 시청 기록 개수가 작은 User에 대해서는 약점 존재 : 위의 예시에서

user 2번을 전처리를 진행하게 될 경우 시청 기록 수가 총 3건밖에 없어 item2와 item5에 대해 평점 차이가 2배나 나게 되는 약점 존재

2번 데이터 전처리 방식의 약점을 보완하기 위하여 다른 방식의 전처리 적용

Count 방식이 아닌 단순히 시청 기록이 존재한다면 무조건 1로 처리



앙상블 방식

각 Model 들을 통해 나온 예측 행렬을 Max 1, Min 0으로 정규화를 진행

정규화된 각 예측 행렬들을 모두 더해주어 새로운 예측 행렬 생성(Soft Voting)