

09-8

* Surprise?

: 추천시스템 구축 전용 pkg

skit-learn과 비슷

+

추천 algo. 적용 - 행렬필터링 적용 가능

sklearn에서 사용하던 API 사용

실습) 데이터 분할 → SVD 행렬분해 → 잠재행렬필터링

• Surprise의 dataset - load_builtin('ml-100k')

※ 3개의 카테고리 분류

• 추천 예측 메소드

- test() : 임의된 dataset 전체에 대하여 추천

- predict() : 개별 데이터에 대하여 추천

• details 속성: was_impossible = True 면

여러값 만들 수 없는 data

• RMSE 평가지표 사용

* Surprise 주요 모듈

• Dataset : 행은 user_id, item_id, rating → row

순인 데이터만 적용 가능

- load_from_file 으로

OS 파일에서 로딩

- load_from_df 으로

DataFrame에서 데이터 로딩

* Surprise의 추천 algo.

- SVD : 행렬분해 · 잠재행렬필터링

- KNNBasic : 최근접 이웃 행렬 필터링

- BaselineOnly : Bias 경향한 SVD 베이스라인 algo.

* 베이스라인 평정 : 개인 상황(필터링) 고려하여

가중치 부여

* 교차검증 - cross_validate 사용

→ 폴딩된 샘플 평가치 + 검증치 제공

(RMSE로 확인)

* data 불러서 train-test-split으로

TrainSet 개체로 변환 필요