**추천 시스템의 유형**

1. **콘텐츠 기반 필터링(Content based filtering)**

사용자가 특정한 아이템 선호하는 경우, 그 아이템과 비슷한 콘텐츠 가진 다른 아이템 추천

콘텐츠의 예로는 장르, 감독, 영화배우, 영화 키워드 등

이들 요소를 결합하여 하나의 콘텐츠 특징으로 피처 벡터화한 뒤에 이들 피처 벡터와 가장 유사한 다른 피처 벡터 가진 것 추천

1. **협업 필터링(Collaborative)**

사용자가 아이템에 매긴 평점 정보나 상품 구매 이력과 같은 사용자 행동 양식만을 기반으로 추천 수행

사용자-아이템 평점 매트릭스와 같은 축적된 사용자 행동 데이터를 기반으로 사용자가 아직 평가하지 않은 아이템을 예측 평가 🡪 사용자-아이템 평점 행렬에서 행은 개별 사용자, 열은 개별 아이템(희소 행렬)

* **최근접 이웃(Nearset Neighbor) 협업 필터링**
* 사용자 기반 최근접 이웃 방식: 사용자 간의 유사도 측정한 뒤 가장 유사도가 높은 TOP-N 사용자 추출해 그들이 선호하는 아이템 추천
* **아이템 기반 최근접 이웃 방식**: 아이템 속성과는 상관 X, 사용자들이 그 아이템을 좋아하는지/싫어하는 지의 평가 척도가 유사한 아이템 추천
* **잠재 요인(Latent Factor) 협업 필터링**

사용자-아이템 평점 매트릭스 속에 숨어있는 잠재 요인 추출해 추천 예측할 수 있도록

대규모 다차원 행렬을 SVD같은 차원 축소로 분해 – 행렬 분해

사용자-아이템 행렬 🡪 (사용자-잠재요인 행렬) 과 (잠재요인-아이템 행렬)로 분해

SVD는 널이 없는 행렬에만 사용할 수 있어서 확률적 경사 하강법이나 ALS 방식 이용

**파이썬 추천 시스템 패키지 – Surprise**

장점 – 다양한 추천 알고리즘, 사이킷런의 핵심 API와 유사한 API명으로 작성됨

Surprise의 주요 모듈

**Dataset**

Surprise는 user\_id, item\_id, rating 데이터가 로우 레벨로 된 데이터 세트만 적용할 수 있음

Dataset.load\_builtin: 무비렌즈 아카이브 FTP 서버에서 무비렌즈 데이터 내려 받음

Dataset.load\_from\_file: OS 파일에서 데이터 로딩할 때 사용

Dataset.load\_from\_df: 판다스 DataFrame에서 데이터 로딩

* 무비 렌즈 데이터 형식이 아닌 다른 OS 파일은 Reader 클래스 설정해야 함
* Line\_format: 칼럼을 순서대로 나열, sep: 칼럼 분리하는 분리자, rating\_scale: 평점 값의 최소~최대 평점 설정

Surprise의 추천 알고리즘 클래스

SVD: 행렬 분해 통한 잠재 요인 협업 필터링

* N\_factors: 잠재 요인 K의 개수, 디폴트 100
* N\_epochs: SGD 수행 시 반복 횟수, 디폴트 20
* Biased: 베이스라인 사용자 편향 적용 여부, 디폴트 True

KNNBasic: 최근접 이웃 협업 필터링

BaselineOnly: 사용자 Bias와 아이템 Bias를 감안한 SGD 베이스라인 알고리즘

SVD++, NMF 등등

Baseline Rating: 각 개인이 평점을 부여하는 성향을 반영해 평점 계산하는 방식, 개인의 성향을 반영해 아이템 평가에 편향성 요소를 반영하여 평점 부과하는 것

전체 평균 평점 + 사용자 편향 점수(사용자별 아이템 평점 평균 값 – 전체 평균 평점) + 아이템 편향 점수(아이템별 평점 평균 값 – 전체 평균 평점)