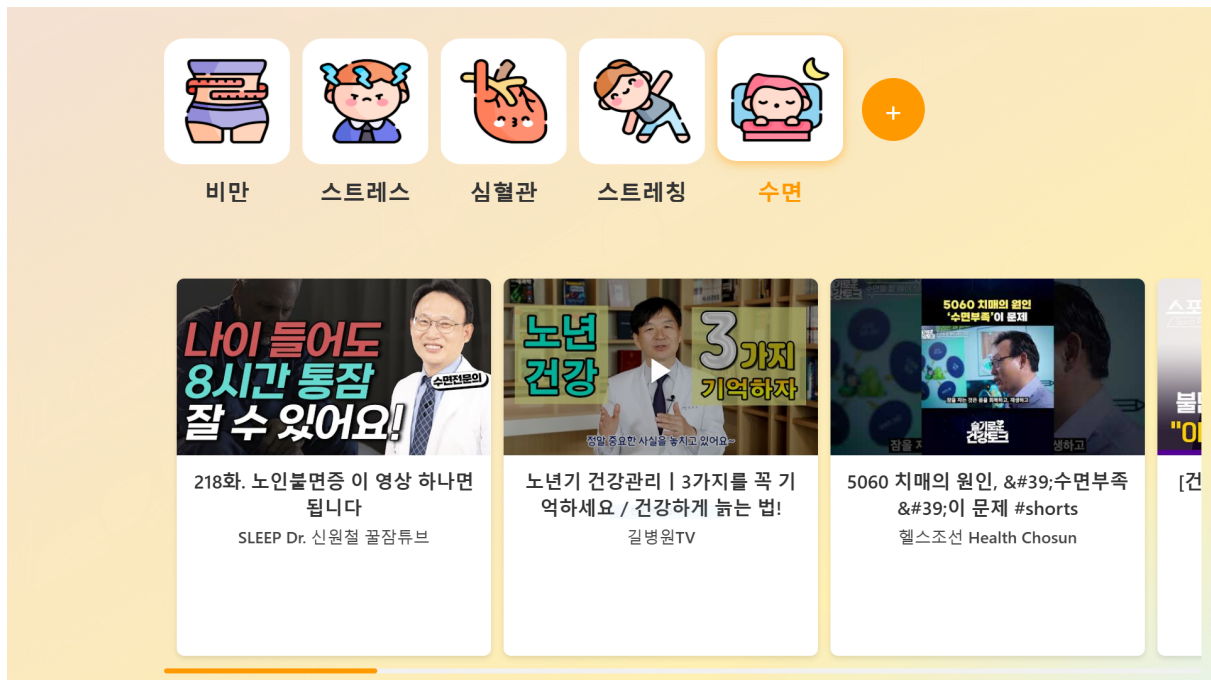


LG U+ 프로젝트 영상 추천 알고리즘

Step1. 개요

1. 현재 화면 배치는 아래와 같다.



- 카테고리 별로 영상은 사용자 시청 이력 데이터가 없을 때 배치는 Youtube API로 불러와 조회수와 좋아요 및 검색 Keyword를 조합하여 왼쪽부터 배치한다.
- 사용자의 시청 이력이 DB에 쌓여 있을 때 영상 배치를 개인에 맞게 왼쪽부터 배치하도록 추천 알고리즘을 적용한다.

2. 영상 추천 방식

- 영화 평점 기반 추천 알고리즘 방식 사용한다. 사용자가 보지 않은 영상의 Rating을 얼마나 매길지 예측한 뒤 가장 높은 Rating을 매길 영상을 왼쪽부터 배치하도록 한다.

Step2. 모델 개요

1. KNN 이웃 알고리즘

- 핵심 아이디어 : 데이터의 유사성(거리)을 측정하여 판단.
- 기본 과정
 - 모든 데이터 간의 거리를 계산.
 - 가장 가까운 K개의 이웃을 선택.
 - 이웃의 결과를 바탕으로 새로운 데이터 포인트를 예측.
- 선정 이유
 - KNN은 새로운 데이터가 추가될 때 학습이 필요하지 않고, 바로 계산에 활용 가능하다는 장점이 있다.
 - "유사한 사용자들이 좋아했던 콘텐츠를 추천 한다." 라는 직관적인 방식의 알고리즘
 - 데이터가 적을 때 유리한 알고리즘이다.

2. SVD 알고리즘

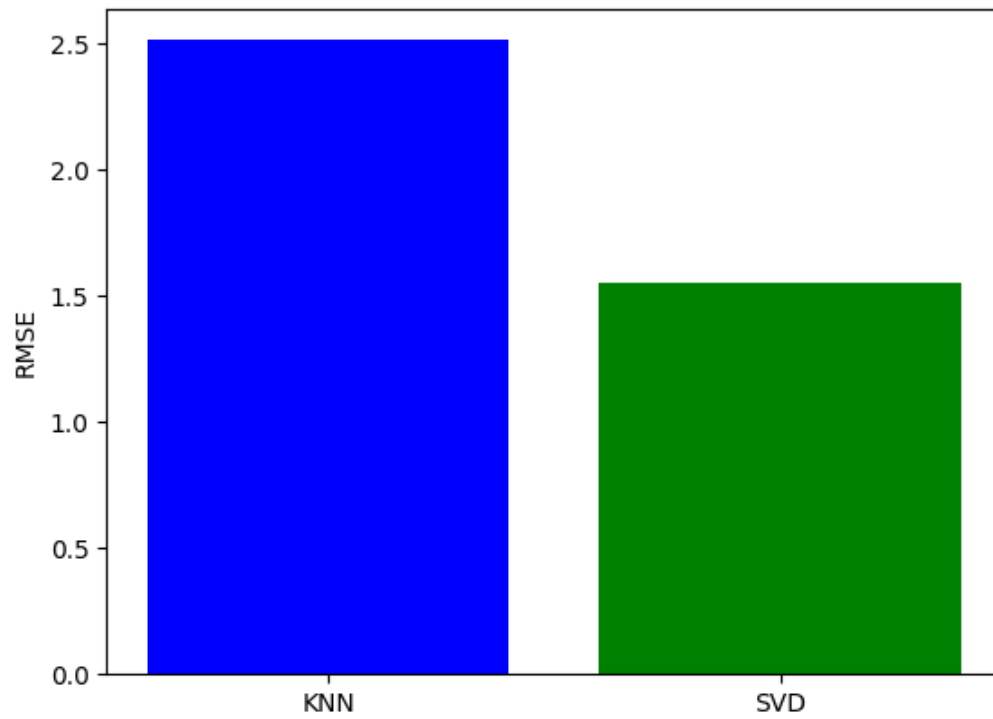
- 핵심 아이디어 : SVD는 사용자와 아이템(영상) 간의 잠재적 관계를 모델링 함으로써, 기존에 관측되지 않은 사용자 선호도를 예측할 수 있습니다.
- 선정 이유
 - 희소 데이터를 효과적으로 처리 : 한 사용자가 모든 영상을 평가하지는 않기 때문에 데이터 행렬의 대부분이 비어있을 수 있습니다.
 - 사용자와 영상의 수가 많아질수록 데이터의 차원이 커지는데, SVD는 이 데이터를 축소하여 계산 효율성을 높입니다.

Step3. 모델 선정 과정

1. 데이터 구성

- 가상의 데이터를 생성
 - User_id, Video_id, Rating 칼럼으로 구성
 - 사용자 10명, 영상 20개, Rating이 1~5점으로 구성된다.
 - 사용자가 모든 영상에 Rating을 매기지 않는다는 상황을 반영
 - Rating은 전체 영상 길이의 영상 시청 시간을 비율로 환산하여 Rating을 매김

2. SVD 알고리즘, KNN 알고리즘 적용 후 성능 비교



- RMSE 값이 SVD가 더 작으므로 성능이 더 우수한 것을 확인했다. 따라서 추천 알고리즘으로 SVD알고리즘을 선정했고 추후 시청 이력 데이터가 쌓이면 SVD추천 알고리즘을 적용한다.