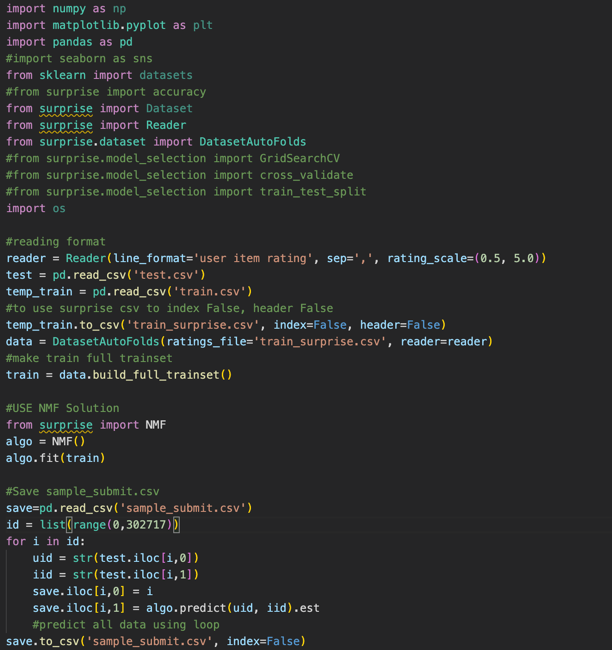
HW7\_20210815\_정재현



먼저 RatingPrediction을 할 때, NMF 기법을 사용하기 위해서 surprise 를 import 해주었다. 우리에게 주어진 파일을 분석해보면 test.csv, train.csv가 존재한다. 이때 csv 파일을 그대로 쓰지 않고 surprise 패키지를 사용하기 위해 index와 header를 false로 만든 csv 파일을 만든 다음 읽어줘야한다. 그래서 먼저 temp\_train으로 읽어주고 to\_csv를 통해 train\_surprise.csv 파일을 구성한다. 그 다음 reader 방식에 따라 line\_format을 user, item rating으로 읽을 수 있게 하고 rating\_scale은 0.5 – 5.0으로 잡았다. 그 뒤 reader 방식대로 dataset을 불러오고 이 데이터를 그대로 train하기 위해 build\_full\_trainset을 호출한다.

NMF를 사용하기 위해 surprise에서 NMF를 불러오고 fit을 통해 train 데이터를 학습 시킨다. 그 다음 predict를 위해 sample\_submit.csv를 save로 불러오고 데이터의 개수가 0~302717만큼 있으므로 id를 구성하도록 한다. 그 후 for를 이용하여 id만큼 반복시키고 그 때마다 uid,iid 를 업데이트하여 그에 맞는 id값과 est값을 save의 연속하는 row에 넣어준다. 그후 save를 to\_csv로 저장시켜서 마친다. 이때 샘플 값으로 주어진 파일에서는 index가 없었으므로 index=False라는 서식을 부여하여 정상적으로 rating prediction방법 중 NMF 사용을 완료하였다. 여기서 predict 대신에 test를 사용하여 결과를 만들 수 있지만 그 방법보다는 predict로 하나하나 uid와 iid를 대입하여 구하는 것이 간편하여 해당 방법대로 구현하였다.