

Homework #4
406.424 인터넷응용
2016년도 봄학기

제출기한: 2016년 6월 7일 23:59까지

1. (50점) **MovieLens** 데이터를 활용하여, 아래의 작업을 수행하시오.
(1-1과 1-2의 경우 결과 출력에 사용한 *BasicRecommender.java* 파일과, 문제에서 요구하는 *txt* 파일을 함께 제출하시오. 1-3에서는 추가로 제시한 유사도 함수에 대하여 1-1과 1-2에서 요구한 *txt* 파일을 다시 제출할 필요는 없으며, 결과 비교 및 의견 서술 내용만을 MS-Word 문서 파일 형태로 작성하여 제출하시오.)
 - 1-1. (20점) User based recommender를 완성하고, 이를 사용하여 1~10번 user에 대한 상위 10개 item의 추천 결과를 **user-based-result.txt** 파일로 출력하시오. 이 때, 아래의 유사도 함수를 사용하시오.
 - **Pearson correlation similarity**
 - 1-2. (20점) Item based recommender를 완성하고, 이를 사용하여 1~10번 item에 대한 상위 10개 item의 추천 결과를 **item-based-result.txt** 파일로 출력하시오. 이 때, 아래의 유사도 함수를 사용하시오.
 - **Pearson correlation similarity**
 - 1-3. (10점) 아래의 유사도 함수를 사용하여 1-1, 1-2 과정을 반복한 뒤, 그 결과를 서로 비교하시오. 각 유사도 함수를 사용했을 때의 차이점에 대해 본인의 의견을 서술하시오.(u.user와 u.item 파일의 정보를 함께 활용 가능).
 - **Cosine similarity**
 - **Log likelihood similarity**
2. (50점) **MovieLens** 데이터를 활용하여, 아래의 작업을 수행하시오.
(2-1의 경우 결과 출력에 사용한 *MFRecommender.java* 파일과, 문제에서 요구하는 *txt* 파일을 함께 제출하시오. 2-2와 2-3의 경우 결과 분석 및 의견 서술 내용만을 MS-Word 문서 파일 형태로 작성하여 제출하시오.)
 - 2-1. (20점) Matrix factorization(MF) based recommender를 제작하고, 이를 사용하여 1~10번 user에 대한 상위 10개 item의 추천 결과를 **mf-based-result.txt** 파일로 출력하시오.
 - 2-2. (20점) 2-1의 추천 결과 성능을 아래의 지표를 종합적으로 활용하여 평가하시오. MF에 사용되는 SVD++ factorizer의 feature 수와 iteration 수를 변화시키면서 추천 결과 성능의 변화 양상을 분석하시오.
 - **AAD(average absolute difference), RMS(root mean squared difference)**
 - 2-3. (10점) 2-2에서 찾은 최적의 MF based recommender를 사용한 추천 결과의 성과, 일반적인 User based recommender를 사용한 추천 결과의 성능을 아래의 지표를 종합적으로 활용하여 비교하시오. 성능에 차이를 보인 경우, 이에 대한 본인의 의견을 서술하시오.
 - **AAD(average absolute difference), RMS(root mean squared difference)**

ETL 사이트에 업로드된 예시 코드를 응용하시오. 과제는 MS-Word나 PDF의 파일 형태로, *txt* 파일 및 Java 코드 결과물과 함께 압축하여 ETL 사이트에 제출하시오.