

Spark Version: 2.3.1
Scala Version: 2.11.0

Program can be run same as shown in description document for scala.
Output of program will be in same path program run (Relative Path).

Task1: ModelBasedCF

>=0 and <1: 29455
>=1 and <2: 13014
>=2 and <3: 2381
>=3 and <4: 337
>=4: 49
Mean Squared Error = 1.0795904674566204
Time: 42.557427053

Task2: ItemBasedCF

>=0 and <1: 29821
>=1 and <2: 12462
>=2 and <3: 2498
>=3 and <4: 451
>=4: 4
Mean Squared Error = 1.0765185752857487
Time: 113.579725711

Improvements on Base Models

For Task1(ModelBasedCF) computed a weighted mean for cold-start items/users. For example, for users with only 1 other item and for predicting an item never seen before (cold-start item), I took the average between user's rating and total average rating of the training data.

For Task2(ItemBasedCF) I computed a weighted mean with the total average rating for items where there were fewer than 5 items used to calculate the rating..