

CF Assignment 1

User – Based and Item – Based Recommender System with Variance & Significance Weighting

The complete code is in present in CF_Assignment1.ipynb. The associated data_set is in the folder ml-100k. For 5-fold evaluation, used the various (80:20 split) data files named u1.base & u1.test ...u5.

Calculated Cosine Similarity without any module, and also adjusted the missing values in the dataset accordingly. The code checks for the suggested reading by the system and then checks it against the actual rating given by the user.

Observations:

It was found that in the user-based recommender system, the NMAE was less than Item-Based Recommender system, suggesting a better & accurate performance. While using significance weights & variance weights systems, the accuracy remained nearly the same because of the proper preprocessing of data, ensuring to reduce the sparsity. Therefore the penalties are not implemented strictly, and the results remain similar. But we do see some decrease in NMAE in the later case.

Note: Bonus has been implemented for both cases and both types

Note: Similarity matrices were saved at the end to avoid any future high complicated data calculation. Using pickle module

Results:

The NMAE (Normalized Mean Absolute Error) for different k (k nearest neighbors), across 5 folds is given below. The average across multiple folds has also been mentioned for collective analysis.

User-based

	k = 10	k = 20	k = 30	k = 40	k = 50
Fold 1	0.693418	0.662175	0.652435	0.655838	0.660308
Fold 2	0.705651	0.680454	0.664431	0.667499	0.661360
Fold 3	0.744569	0.691484	0.674257	0.671427	0.672585
Fold 4	0.734123	0.689136	0.667382	0.666401	0.664177
Fold 5	0.736034	0.693678	0.671269	0.665892	0.667274
Average	0.722759	0.683385	0.665955	0.665412	0.665141

Item-based

	k = 10	k = 20	k = 30	k = 40	k = 50
Fold 1	0.581941	0.764364	0.827649	0.879297	0.901199
Fold 2	0.572676	0.747578	0.838136	0.875136	0.904646
Fold 3	0.561732	0.729817	0.824063	0.881374	0.927508
Fold 4	0.475176	0.654886	0.762559	0.837193	0.873354
Fold 5	0.459266	0.642181	0.744372	0.792142	0.830131
Average	0.530158	0.707765	0.799356	0.853028	0.887368

Variance Weighting

User-based

	k = 10	k = 20	k = 30	k = 40	k = 50
Fold 1	0.697750	0.662793	0.652578	0.655713	0.660304
Fold 2	0.711063	0.681948	0.665133	0.668097	0.661857
Fold 3	0.753394	0.695552	0.676268	0.673272	0.674681
Fold 4	0.742759	0.692744	0.670121	0.669279	0.667015
Fold 5	0.745316	0.697468	0.673963	0.668439	0.669686
Average	0.730056	0.686101	0.667613	0.666960	0.666709

Item Based

	k = 10	k = 20	k = 30	k = 40	k = 50
Fold 1	0.589934	0.776178	0.839057	0.890759	0.911341
Fold 2	0.590419	0.762300	0.851549	0.889109	0.917253
Fold 3	0.571901	0.739963	0.834269	0.890820	0.936596
Fold 4	0.485871	0.667017	0.773763	0.848065	0.883699
Fold 5	0.472261	0.655274	0.757143	0.805678	0.843018
Average	0.542077	0.720147	0.811156	0.864886	0.898381

Significance Weighting

User-based

	k = 10	k = 20	k = 30	k = 40	k = 50
Fold 1	0.697750	0.662793	0.652578	0.655713	0.660304
Fold 2	0.711063	0.681948	0.665133	0.668097	0.661857
Fold 3	0.753394	0.695552	0.676268	0.673272	0.674681
Fold 4	0.742759	0.692744	0.670121	0.669279	0.667015
Fold 5	0.745316	0.697468	0.673963	0.668439	0.669686
Average	0.730056	0.686101	0.667613	0.666960	0.666709

Item Based

	k = 10	k = 20	k = 30	k = 40	k = 50
Fold 1	0.599772	0.778604	0.839884	0.890504	0.911626
Fold 2	0.591470	0.762694	0.851203	0.887105	0.915618
Fold 3	0.581075	0.744009	0.836042	0.891120	0.936258
Fold 4	0.494712	0.670192	0.775604	0.849063	0.884399
Fold 5	0.479154	0.657809	0.757501	0.804112	0.841346
Average	0.549237	0.722661	0.812047	0.864381	0.897849