Report

Collaborative filtering is a rapidly advancing research area. Every year new techniques are proposed. Classic methods include neighborhood methods such as memory based or user based collaborative filtering. More recent methods often revolve around matrix factorization including singular value decomposition and non- negative matrix factorization.

Collaborative filtering systems are usually categorized into two subgroups: memory-based and model-based methods.
Memory-based methods simply memorize the rating matrix and issue recommendations based on the relationship between the queried user and item and the rest of the rating matrix. Model-based methods fit a parameterized model to the given rating matrix and then issue recommendations based on the fitted model.

Observations: 1)Matrix factorization methods in general show better performance when the number of users gets sufficiently large (>3,000). 2)Overall, the best performing algorithm is regularized SVD. 3)When the number of users is sufficiently small there is very little difference between matrix factorization methods and the simpler neighborhood- based methods. 4)Item average, item-based, regularized SVD, PMF, BPMF, and NLPFM tend to be the most sensitive to variation in user count. 5)Constant baseline, user average, user-based, NMF, NPCA, and rank-based are relatively insensitive to the number of users. **Difficulty:** One difficulty that is undoubtedly central to the lack of clarity is that the performance of different methods differ substantially based on the problem parameters. **Factors:** Factors such as number of users, number of items, and sparsity level (ratio of observed to total ratings) affect different collaborative filtering methods in different ways. Some methods perform better in sparse settings while others perform better in dense settings, and so on.

Conclusion: The conclusions stated that we have identified seven groups of CF methods, where CF methods in the same group share certain experimental properties:

- Baselines: Constant, User Average, Item Average
- Memory-based methods: User-based, Item-based (with and without default values)
- Matrix-Factorization I: Regularized SVD, PMF, BPMF, NLPFM
- Matrix-Factorization II: NMF
- Others (Individually): Slope-one, NPCA, Rank-based CF.

Matrix-Factorization-based methods generally have the highest accuracy. Specifically, regularized SVD, PMF and its variations perform best as far as MAE and RMSE, except in very sparse situations, where NMF performs the best. All algorithms vary in their accuracy, based on the user count, item count, and density.