Recsys Data Analysis

February 17, 2017

1 Results

Values Row Labels	Average of RMSE.ByUser	Average of RMSE.ByRating	Average of Predict.nDCG	Average of MRR
Lastfm	20	20	20	20
Custom	1225.316632	3052.545397	0.811768122	0.000650303
ItemItem	1455.425755	4469.689135	0.734481196	0.001738894
PersMean	1225.316632	3052.545397	0.811768122	0.000650303
UserUser	1226.077364	3251.001417	0.743536709	0.001728617
Movielens	20	20	20	20
Custom	0.920828036	0.93823614	0.949940793	0.002643017
ItemItem	0.890329159	0.896982711	0.955260097	0.095015162
PersMean	0.920828036	0.93823614	0.949940793	0.002643017
UserUser	0.915197144	0.924923196	0.953311329	0.003773996
(blank)				
(blank)				
jester	4	4	4	4
Custom	0.826521	0.87686	0.944675	0.61766
ItemItem	0.790645	0.834773	0.951061	0.611854
PersMean	0.826521	0.87686	0.944675	0.61771
UserUser	0.796074	0.836992	0.951285	0.712654

2 Datasets

2.1 LastFM

#Users: 1892 users
#Items: 17632 artists
This dataset contains 92.834 user-artist relations. Each relation represents the number of times a user listened to an artist. In order to run the recommender system algorithms, we decided to consider this counter as a kind of rating.

2.2 Movielens

#Users: 943 users
#Items: 1682 movies
This dataset contains 100.000 user-movie relations. Each relation represents the evaluation of a movie by a user (in a scale from 1 to 5).

2.3 Jester

#Users: 73.495 users
#Items: 100 jokes
This dataset contains 4.1 million of user-joke relations. Each relation represents the evaluation of a movie by a user (in a scale from -10.0 to 10.0). As the scale is different to the used in Movielens, we decided to normalized the evaluations to a scale of [1,5]. Thus, the results of running the algorithm with Jester and Movielens can be comparable.

2.4 Bookcrossing

#Users: 278.858 users
#Items: 271.379 books
This dataset contains 1.149.780 of user-book relations. Each relation represents the evaluation of a book by a user. Also with this dataset, we normalized the ratings to a scale of [1,5].

3 Metrics

3.1 RMSE

RMSE reflects the difference between the real values and those predicted by the algorithm. It is calculated in two ways:

- Grouped by user: RMSE.byUser

$$\frac{\sum_{\forall\ user\ j} \sqrt{\frac{\sum_{\forall\ rating\ i} err_{ij}^2}{Total\ ratings\ user\ i}}}{Total\ users} \tag{1}$$

- Globally: RMSE.byRating

$$\sqrt{\frac{\sum_{\forall\ user\ j} \sum_{\forall\ rating\ i} err_{ij}^2}{Total\ ratings}} \tag{2}$$

In general, both ways gives similar results.

In the case of LastFM, while grouping by users, the results are approximately a third of those obtained using the Global expression.

3.2 MRR

This metric gives particular good results for Jester. That seems to be caused by the reduced number of items in the dataset (only 100 jokes). We could say that it is easier to "guess" the correct ranking in Jester than in the other datasets because of it has less items to order. Also it's curious that MRR result for Movielens using CF item-item are approximately 40 times better than the outcome for the same metric, the same dataset, but with other algorithms.

4 Algorithms

4.1 CF Item-item

For the rating evaluations, is the algorithm with the best performance in Jester and Movielens. But with LastFm it is worse than CF User-user.