DAI-Labor
TU Berlin

Adapting Ratings in Memory-Based Collaborative Filtering using Linear Regression

DAI-Labor of
Technische Universität Berlin

Jérôme Kunegis
kunegis@dai-labor.de

A|O|T

Agententechnologien in betrieblichen Anwendungen und der Telekommunikation

Agenda

Collaborative Filtering

Baseline Algorithm

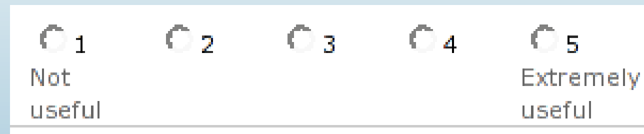
Linear Regression

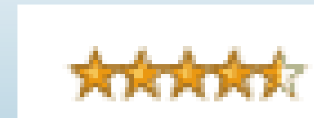
Evaluation

- ⇨ Collaborative Filtering: Rating Prediction
- ⇨ Baseline Algorithm: Pearson Correlation
- ⇨ Linear Regression: Adapting Ratings
- ⇨ Evaluation

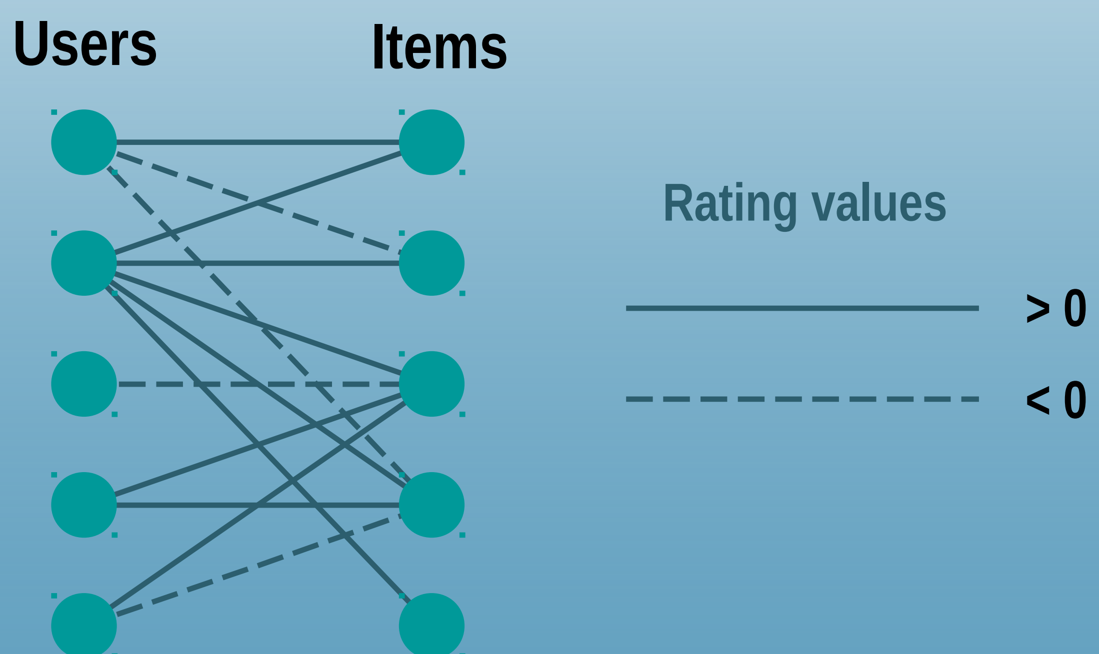
The Bipartite Rating Graph

- Database of user and items
- Users rate items





- Rating database as sparse bipartite graph



Collaborative Rating Prediction

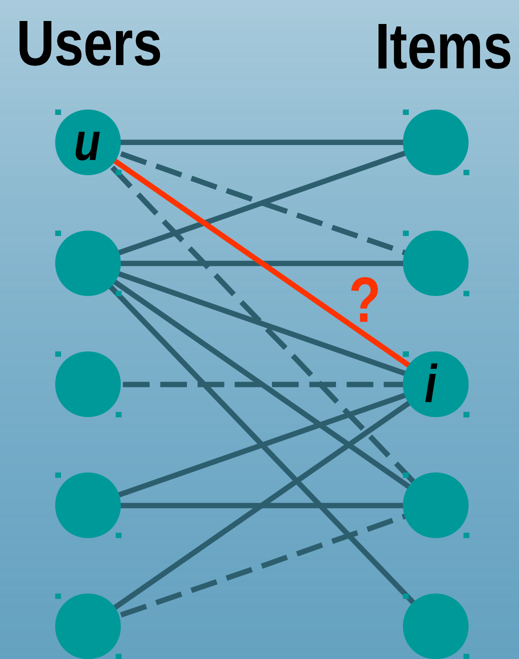
Collaborative Filtering

Baseline Algorithm

Linear Regression

Evaluation

- ⇨ Predict a missing rating
- ⇨ Use: Recommend new items (rank unrated items by predicted rating)

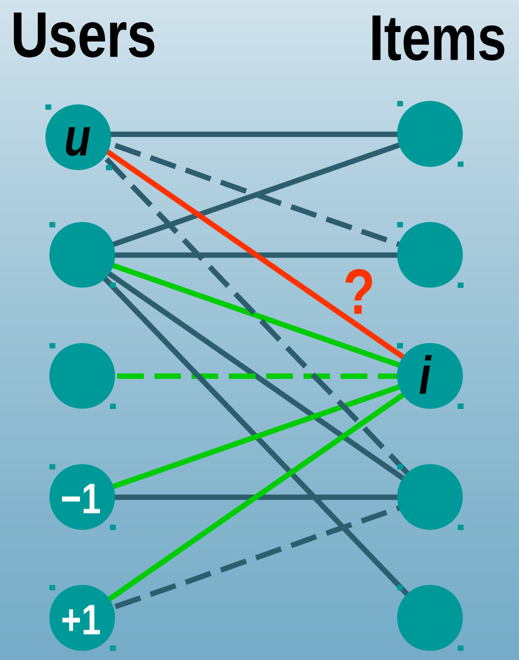


Will user *u* like item *i* ?

Weighted Mean of Ratings

- Collaborative Filtering
- Baseline Algorithm
- Linear Regression
- Evaluation

- ⇨ Item *i* has been rated by many users: Take the average of all these ratings
- ⇨ Give high weight to ratings by users similar to *u*
- ⇨ Similarity measure: use the Pearson correlation between two users' ratings (note: can be negative)
- ⇨ Works also as user-based algorithm



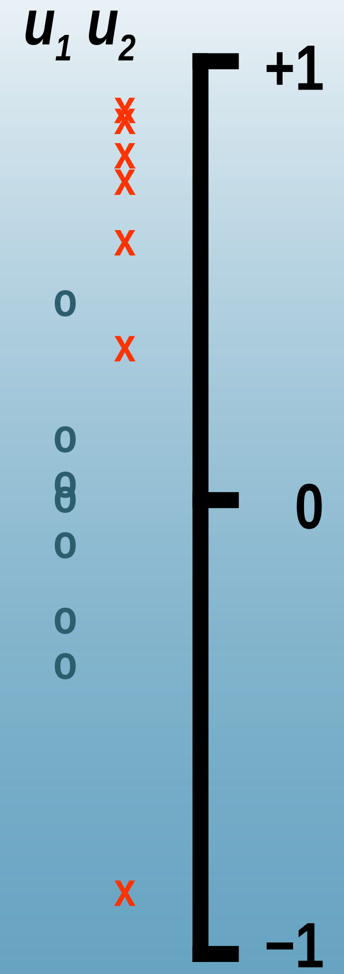
User Rating Habits and Taste

Users have different rating habits:

- ⇨ Different mean rating
- ⇨ Different variance

Weighted mean assumes the same rating scale for all users, although the Pearson correlation takes into account different scales

Solution: Weighted means of ratings scaled according to each user's rating mean and variance



Scaling Ratings

Collaborative Filtering

Baseline Algorithm

Linear Regression

Evaluation

Two Users *u* and *v* have common rating vectors *U* and *V* using linear regression

- ⇨ **Determine factors *a* and *b* minimizing sum of squared errors in $U - (V\ 1)\ (a\ b)^T$**
- ⇨ **When U and V are negatively correlated, *a* is negative**
- ⇨ **Therefore, use absolute value of correlation for weight**
- ⇨ **Correlation and (*a* *b*) can be calculated in one pass**

Evaluation

Collaborative Filtering

Baseline Algorithm

Linear Regression

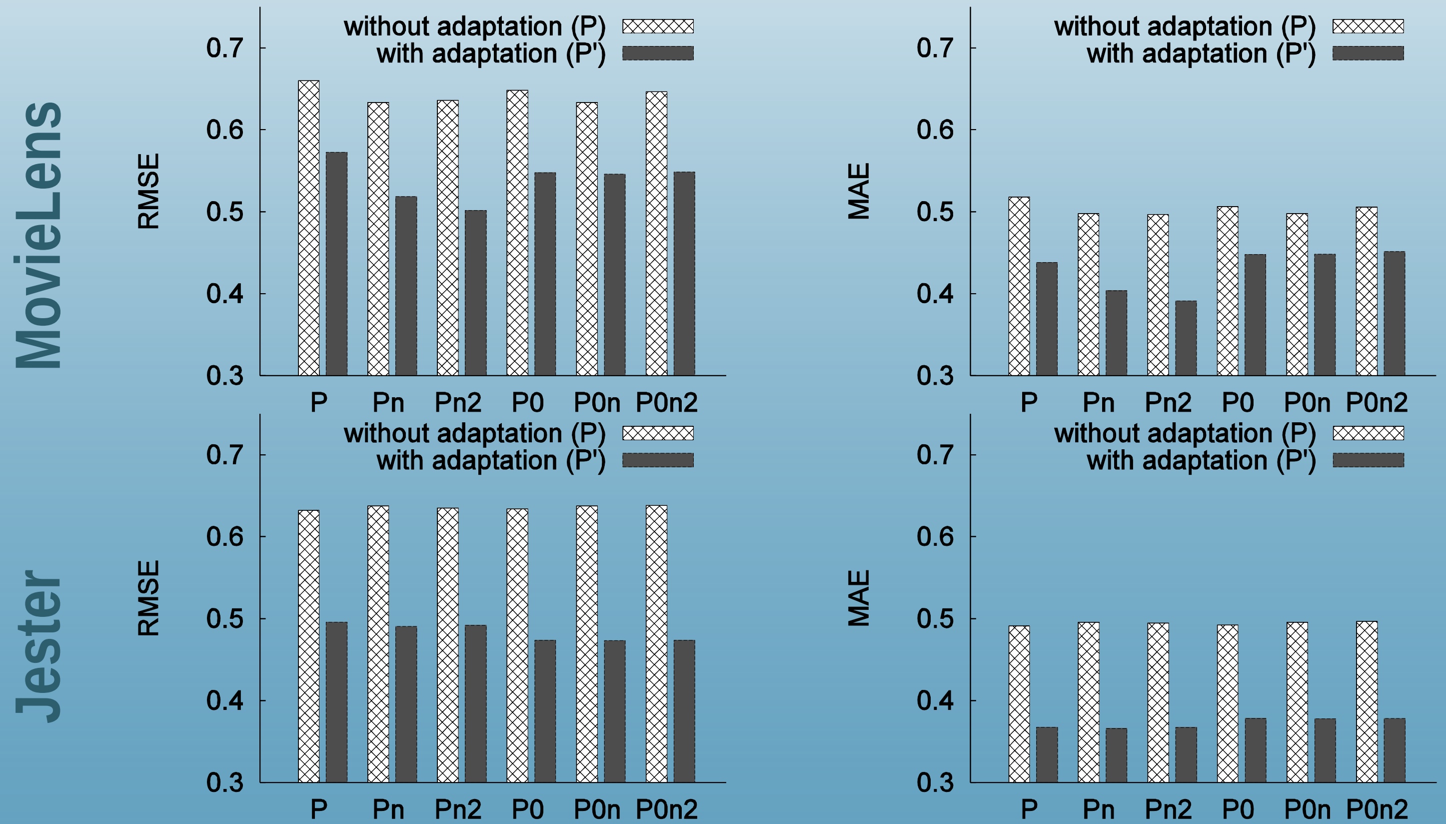
Evaluation

Linear Regression adaptation can be used in variants of weighted mean algorithm:

- ⇨ User-based, item-based
- ⇨ With/out normalized ratings
- ⇨ Fill missing ratings with default parameter
- ⇨ Weight users by number of common ratings (1, n, n^2)

Evaluation Results

Both error measures (mean average error, mean root squared error) reduced by ~0.1 points



The End

Thank you!

Questions?

Comments?

Backup

- A: Why doesn’t normalization make this obsolete?
- Q: Normalization corrects differences in rating habits (scales), whereas regression maps different tastes to each other