Recommender Systems and the Netflix Prize

Andreas Töscher, Georg Preßler

18 February 2009, San Francisco

Founding Members

Georg Preßler, Michael Schrotter, Andreas Töscher, Michael Jahrer

Georg Preßler is on site and is prepared to answer your questions

commendo

- specialized in customized recommender systems
- top contender in the Netflix prize
- Netflix Progress Prize 2008

Goal of a recommender system

- building customized websites
- suggestions for interesting products
- improved sorting of search results
- display of similar products

Content vs Collaborative Filter

Content filtering

- based on item meta data
  for example:
  - category or genre
  - actor or director of a movie

Collaborative filtering (CF)

- take taste of many users to infer taste of an individual
- *Assumption:* Users who agreed in the past will agree in the future.

Good recommender systems are mostly a combination of both!

Sources of information for CF

Explicit information

- ratings
- rankings
- comments

Implicit information

- purchase information
- wish list
- clicks
- time spent on a page

Sources of information for CF

Explicit information

- ratings
- rankings
- comments

Implicit information

- purchase information
- wish list
- clicks
- time spent on a page

Netflix Prize

NETFLIX

- Netflix is an US online movie rental service
- hiring of movies via mail
- over 100.000 titles
- 55 million DVDs total
- productive start at 1997
- have their own recommendation system called “Cinematch”

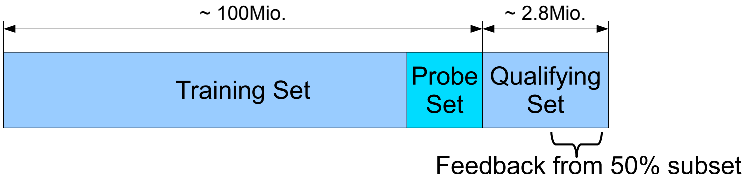
Netflix Prize

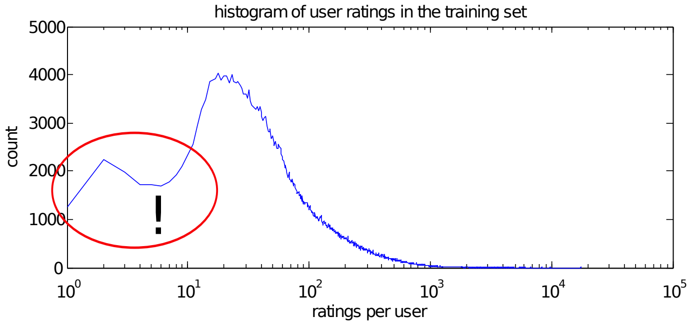
- Grand Prize, 1 Mio. US-Dollar for 10% improvement in prediction accuracy
- Progress Prize, 50.000 US-Dollar, October every year
- started Oct, 2 2006
- ends Oct, 2 2011
  - or when a 10% improvement in RMSE is reached by a team
- www.netflixprize.com

Netflix Prize Facts

- dataset consists of 100 Mio. entries
- quadruples of $< movie, user, rating, date >$
- integer ratings from 1 to 5
- error measure: RMSE (root mean square error)
  $RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(r_i - t_i)^2}$ $(r_i = prediction \;\; t_i = target)$
- over 45.000 contestants from 180 countries

Dataset Details





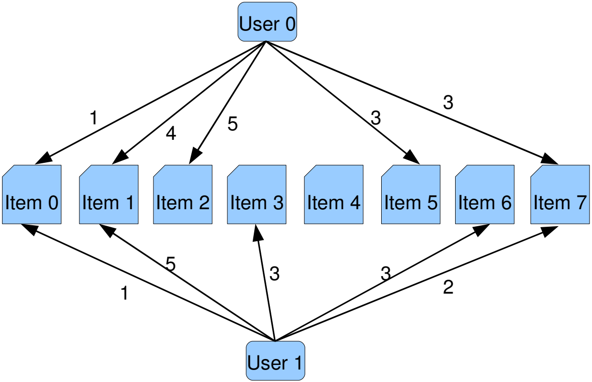
Overview of Models

- Similarity based
  - item based
  - user based
- Latent factor model (e.g. SVD)
- Asymetric factor models
- Neural Networks
- Restricted Boltzmann Machines
- Hybrid approaches

Similarity based methods

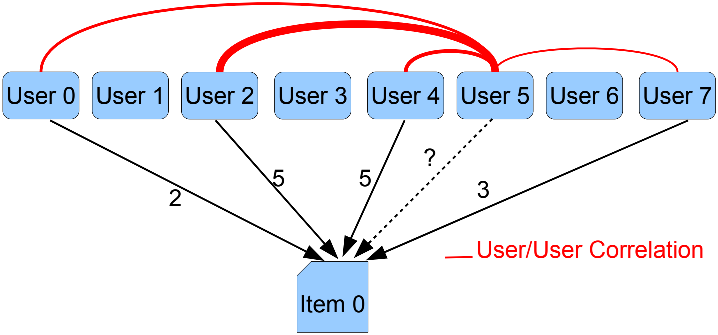
- rating prediction based on neighborhood information
- standard algorithm
- very simple to implement
- easy to understand
- easy to customize

Similarity between users



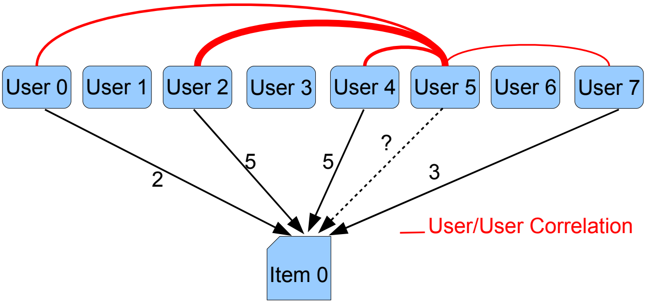
- $U_0 = [1, 4, 3]$
- $U_1 = [1, 5, 2]$

Prediction from user similarity



- Pearson correlation between users
- rating prediction for user 5 on item 0:
  - select k-best correlating users which voted item 0
  - weight the ratings of the k-best users with their correlation to user 5

Prediction from user similarity



- $N(u,i)$ ... k-best correlating users
- c_{uv} ... Pearson correlation between users u and v

$$r_{ui} = \frac{\sum_{v \in N(u,i)} c_{uv} r_{vi}}{\sum_{v \in N(u,i)} |c_{vi}|} \tag{1}$$

Remarks on user similarity models

- in general two users have very few common ratings
- Netflix dataset consists of 500,000 users
- precomputation of all user/user correlations is hard (takes over 1 TByte storage)
- due to sparsity predictions for some user/item combinations are not possible (coverage)

Item/Item Correlations

- in general better defined
- 18,000 movies in the Netflix dataset
- precomputable (for the Netflix dataset)
- predictions are more accurate

Remarks on user similarity models

- in general two users have very few common ratings
- Netflix dataset consists of 500,000 users
- precomputation of all user/user correlations is hard (takes over 1 TByte storage)
- due to sparsity predictions for some user/item combinations are not possible (coverage)

Item/Item Correlations

- in general better defined
- 18,000 movies in the Netflix dataset
- precomputable (for the Netflix dataset)
- predictions are more accurate

Ranking Quality

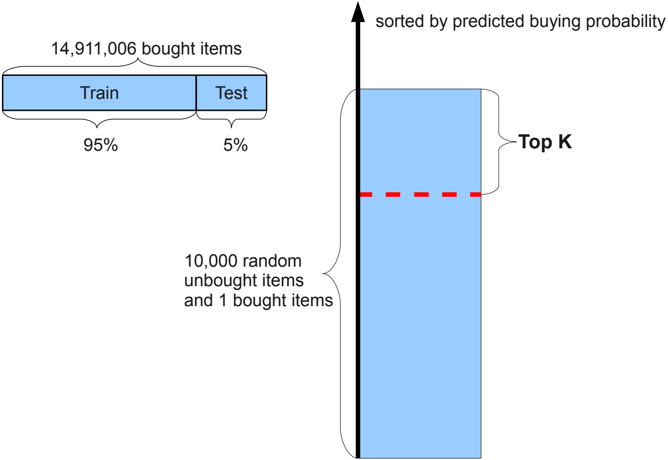
Ranking

- use collaborative filtering for personalized sorting
- small RMSE improvements lead to much better rankings

Dataset details

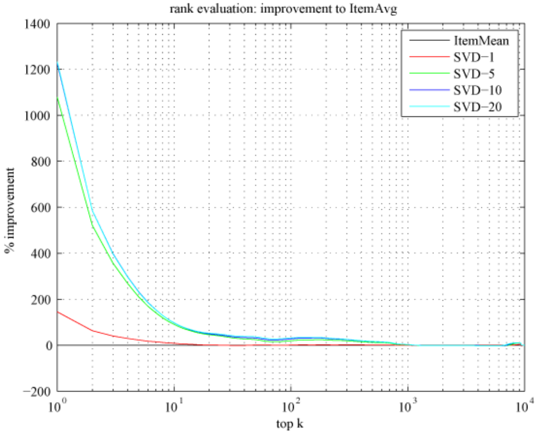
- 855,949 users
- 96,198 items
- 14,911,006 bought items

The Experiment



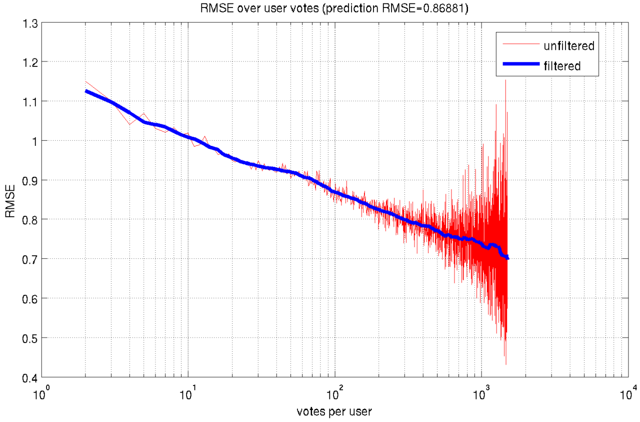
- the experiment is done on the last 5% of rating data

Improved Ranking



	item mean	SVD 1	SVD 5	SVD 10	SVD 20
Improvement over item mean [RMSE]	0.0%	2.14%	6.22%	8%	8.6%
Hit probability for top-1	0.66%	1.6%	7.8%	8.7%	8.8%

Average Error as Function of User Votes



- Netflix Grand Prize level at $\sim$ 100 votes/user

Ways to Increase accuracy

- more sophisticated models
- combining different models
- more information
  - ratings
  - clicks
  - wish list
  - purchase information

Small improvements in accuracy dramatically improves the sorting quality!

An improved RMSE of 8.6% leads to over **1,200%** of improvement on the hit rate for the top-1 item!

Thank you for your attention.

commendo
research & consulting
www.commendo.at