CS 572: Information Retrieval

Recommender Systems 2: Implementation and Applications

Acknowledgements
Many slides in this lecture are adapted from Xavier Amatriain (Netflix), Yehuda Koren (Yahoo), and Dietmar Jannach (TU Dortmund), Jimmy Lin, Foster Provost
## Recap of Recommendation Approaches

<table>
<thead>
<tr>
<th>Method</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collaborative</td>
<td>No knowledge-engineering effort, serendipity of results, learns market segments</td>
<td>Requires some form of rating feedback, cold start for new users and new items</td>
</tr>
<tr>
<td>Content-based</td>
<td>No community required, comparison between items possible</td>
<td>Content descriptions necessary, cold start for new users, no surprises</td>
</tr>
<tr>
<td>Knowledge-based</td>
<td>Deterministic recommendations, assured quality, no cold-start, can resemble sales dialogue</td>
<td>Knowledge engineering effort to bootstrap, basically static, does not react to short-term trends</td>
</tr>
</tbody>
</table>
Recap: What works (in “real world”)

- Depends on the **domain** and particular **problem**
- However, in the general case it has been demonstrated that the best isolated approach is CF.
  - Other approaches can be hybridized to improve results in specific cases (cold-start problem...)
- **What matters:**
  - **Data preprocessing:** outlier removal, denoising, removal of global effects (e.g. individual user's average)
  - “Smart” dimensionality reduction using MF/SVD
  - Combining methods
Today: Applications

• Products: Netflix challenge
• Information: News recommendation
• Social: Who to follow
• Human issues (how recommendations are used)
Welcome!

The Netflix Prize seeks to substantially improve the accuracy of predictions about how much someone is going to love a movie based on their movie preferences. Improve it enough and you win one (or more) Prizes. Winning the Netflix Prize improves our ability to connect people to the movies they love.

Read the Rules to see what is required to win the Prizes. If you are interested in joining the quest, you should register a team.

You should also read the frequently-asked questions about the Prize. And check out how various teams are doing on the Leaderboard.

Good luck and thanks for helping!
# Movie rating data

## Training data

<table>
<thead>
<tr>
<th>user</th>
<th>movie</th>
<th>score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>21</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>213</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>345</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>123</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>768</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>76</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>45</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>568</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>342</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>234</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>76</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>56</td>
<td>4</td>
</tr>
</tbody>
</table>

## Test data

<table>
<thead>
<tr>
<th>user</th>
<th>movie</th>
<th>score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>62</td>
<td>?</td>
</tr>
<tr>
<td>1</td>
<td>96</td>
<td>?</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>?</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>?</td>
</tr>
<tr>
<td>3</td>
<td>47</td>
<td>?</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>?</td>
</tr>
<tr>
<td>4</td>
<td>41</td>
<td>?</td>
</tr>
<tr>
<td>4</td>
<td>28</td>
<td>?</td>
</tr>
<tr>
<td>5</td>
<td>93</td>
<td>?</td>
</tr>
<tr>
<td>5</td>
<td>74</td>
<td>?</td>
</tr>
<tr>
<td>6</td>
<td>69</td>
<td>?</td>
</tr>
<tr>
<td>6</td>
<td>83</td>
<td>?</td>
</tr>
</tbody>
</table>
Netflix Prize

• Training data
  – 100 million ratings
  – 480,000 users
  – 17,770 movies
  – 6 years of data: 2000-2005

• Test data
  – Last few ratings of each user (2.8 million)
  – Evaluation criterion: root mean squared error (RMSE)
  – Netflix Cinematch RMSE: 0.9514

• Competition
  – 2700+ teams
  – $1 million grand prize for 10% improvement on Cinematch result
  – $50,000 2007 progress prize for 8.43% improvement
Overall rating distribution

- Third of ratings are 4s
- Average rating is 3.68

From TimelyDevelopment.com
#ratings per movie

- Avg #ratings/movie: 5627
#ratings per user

- Avg #ratings/user: 208
Average movie rating by movie count

More ratings to better movies

From TimelyDevelopment.com
Most loved movies

<table>
<thead>
<tr>
<th>Title</th>
<th>Count</th>
<th>Avg rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Shawshank Redemption</td>
<td>137812</td>
<td>4.593</td>
</tr>
<tr>
<td>Lord of the Rings: The Return of the King</td>
<td>133597</td>
<td>4.545</td>
</tr>
<tr>
<td>The Green Mile</td>
<td>180883</td>
<td>4.306</td>
</tr>
<tr>
<td>Lord of the Rings: The Two Towers</td>
<td>150676</td>
<td>4.460</td>
</tr>
<tr>
<td>Finding Nemo</td>
<td>139050</td>
<td>4.415</td>
</tr>
<tr>
<td>Raiders of the Lost Ark</td>
<td>117456</td>
<td>4.504</td>
</tr>
<tr>
<td>Forrest Gump</td>
<td>180736</td>
<td>4.299</td>
</tr>
<tr>
<td>Lord of the Rings: The Fellowship of the ring</td>
<td>147932</td>
<td>4.433</td>
</tr>
<tr>
<td>The Sixth Sense</td>
<td>149199</td>
<td>4.325</td>
</tr>
<tr>
<td>Indiana Jones and the Last Crusade</td>
<td>144027</td>
<td>4.333</td>
</tr>
</tbody>
</table>
Important RMSEs

- **Grand Prize:** 0.8563; 10% improvement
- **BellKor:** 0.8693; 8.63% improvement
- **Cinematch:** 0.9514; baseline
- **Movie average:** 1.0533
- **User average:** 1.0651
- **Global average:** 1.1296
- **Inherent noise:** ????

Personalization

->

erroneous

accurate

4/6/2016
Challenges

• Size of data
  – Scalability
  – Keeping data in memory

• Missing data
  – 99 percent missing
  – Very imbalanced

• Avoiding overfitting

• Test and training data differ significantly
The Rules

• Improve RMSE metric by 10%
  – RMSE = square root of the averaged squared difference between each prediction and the actual rating
  – amplifies the contributions of egregious errors, both false positives ("trust busters") and false negatives ("missed opportunities").

• 100 million “training” ratings provided
  – User, Rating, Movie name, date
Leaderboard

<table>
<thead>
<tr>
<th>Rank</th>
<th>Team Name</th>
<th>Best Score</th>
<th>% Improvement</th>
<th>Last Submit Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BellKor in BigChaos</td>
<td>0.8816</td>
<td>9.44</td>
<td>2008-09-30 21:17:40</td>
</tr>
<tr>
<td>2</td>
<td>BigChaos</td>
<td>0.8629</td>
<td>9.30</td>
<td>2008-11-13 01:06:01</td>
</tr>
<tr>
<td>3</td>
<td>PragmaticTheory</td>
<td>0.8642</td>
<td>9.17</td>
<td>2008-10-31 10:48:25</td>
</tr>
<tr>
<td>4</td>
<td>BellKor</td>
<td>0.8643</td>
<td>9.15</td>
<td>2008-08-28 15:12:06</td>
</tr>
<tr>
<td>5</td>
<td>Gravity</td>
<td>0.8585</td>
<td>9.00</td>
<td>2008-10-05 15:08:10</td>
</tr>
<tr>
<td>6</td>
<td>My Brain and His Chain</td>
<td>0.8668</td>
<td>8.89</td>
<td>2008-09-30 02:19:47</td>
</tr>
<tr>
<td>7</td>
<td>When Gravity and Dinosaurs Unite</td>
<td>0.8575</td>
<td>8.82</td>
<td>2008-10-05 14:16:53</td>
</tr>
<tr>
<td>8</td>
<td>acmehill</td>
<td>0.8577</td>
<td>8.80</td>
<td>2008-11-02 10:47:35</td>
</tr>
<tr>
<td>9</td>
<td>Just a guy in a garage</td>
<td>0.8577</td>
<td>8.80</td>
<td>2008-11-07 17:24:20</td>
</tr>
<tr>
<td>10</td>
<td>Opera Solutions</td>
<td>0.8578</td>
<td>8.79</td>
<td>2008-11-13 00:31:57</td>
</tr>
<tr>
<td>11</td>
<td>scientist</td>
<td>0.8684</td>
<td>8.72</td>
<td>2008-10-24 20:38:27</td>
</tr>
<tr>
<td>12</td>
<td>Ces</td>
<td>0.8711</td>
<td>8.44</td>
<td>2008-08-25 05:00:23</td>
</tr>
<tr>
<td>13</td>
<td>KorBell</td>
<td>0.8712</td>
<td>8.43</td>
<td>2007-10-01 23:25:23</td>
</tr>
<tr>
<td>14</td>
<td>basho</td>
<td>0.8714</td>
<td>8.41</td>
<td>2008-05-21 22:06:00</td>
</tr>
<tr>
<td>15</td>
<td>pengpengzhou</td>
<td>0.8714</td>
<td>8.41</td>
<td>2008-10-05 11:13:13</td>
</tr>
<tr>
<td>16</td>
<td>blednotik</td>
<td>0.8718</td>
<td>8.37</td>
<td>2008-11-09 10:57:10</td>
</tr>
<tr>
<td>17</td>
<td>Dan Tillberg</td>
<td>0.8723</td>
<td>8.31</td>
<td>2008-03-28 20:58:05</td>
</tr>
</tbody>
</table>

Progress Prize 2007 - RMSE = 0.8712 - Winning Team: KorBell
Lessons from the Netflix Prize

Yehuda Koren
The BellKor team
(with Robert Bell & Chris Volinsky)

movie #15868

at&t
The BellKor recommender system

- Use an ensemble of complementing predictors
- Two, half tuned models worth more than a single, fully tuned model
Effect of ensemble size

<table>
<thead>
<tr>
<th>#Predictors</th>
<th>Error - RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.89</td>
</tr>
<tr>
<td>8</td>
<td>0.8875</td>
</tr>
<tr>
<td>15</td>
<td>0.8825</td>
</tr>
<tr>
<td>22</td>
<td>0.8775</td>
</tr>
<tr>
<td>29</td>
<td>0.8725</td>
</tr>
<tr>
<td>36</td>
<td>0.87</td>
</tr>
<tr>
<td>43</td>
<td>0.885</td>
</tr>
<tr>
<td>50</td>
<td>0.8925</td>
</tr>
</tbody>
</table>
The BellKor recommender system

- Use an ensemble of **complementing** predictors
- **Two, half tuned** models worth more than a **single, fully tuned** model
- But:
  Many seemingly different models expose similar characteristics of the data, and won’t mix well
- **Concentrate efforts along three axes...**
The three dimensions of the BellKor system

The first axis:

- Multi-scale modeling of the data
- Combine top level, regional modeling of the data, with a refined, local view:
  - k-NN: Extracting local patterns
  - Factorization: Addressing regional effects
Multi-scale modeling – 1st tier

Global effects:

- Mean rating: 3.7 stars
- *The Sixth Sense* is 0.5 stars above avg
- *Joe* rates 0.2 stars below avg

→ Baseline estimation: 
  *Joe* will rate *The Sixth Sense* 4 stars
Multi-scale modeling – 2\textsuperscript{nd} tier

Factors model:

- Both *The Sixth Sense* and *Joe* are placed high on the “Supernatural Thrillers” scale

\[ \Rightarrow \text{Adjusted estimate: } \]

*Joe* will rate *The Sixth Sense* 4.5 stars
Multi-scale modeling – 3rd tier

Neighborhood model:
- *Joe* didn’t like related movie *Signs*
- ➔ Final estimate: *Joe* will rate *The Sixth Sense* 4.2 stars
The second axis:

- Quality of modeling
- Make the best out of a model
- Strive for:
  - Fundamental derivation
  - Simplicity
  - Avoid overfitting
  - Robustness against #iterations, parameter setting, etc.
- Optimizing is good, but don’t overdo it!
The three dimensions of the BellKor system

- The third dimension will be discussed later...
- Next: Moving the multi-scale view along the quality axis
Local modeling through k-NN

• Earliest and most popular collaborative filtering method
• Derive unknown ratings from those of “similar” items (movie-movie variant)
• A parallel user-user flavor: rely on ratings of like-minded users (not in this talk)
### k-NN

#### Movie Ratings

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td></td>
<td>2</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Users**: 1 to 6
- **Movies**: 1 to 5

- Unknown rating
- Rating between 1 to 5
- estimate rating of movie 1 by user 5
**k-NN**

Neighbor selection:
Identify movies similar to 1, rated by user 5
### k-NN

#### Compute similarity weights:
\[ s_{13} = 0.2, \ s_{16} = 0.3 \]
### k-NN

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>users</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td></td>
<td>3</td>
<td>2.6</td>
<td>5</td>
<td></td>
<td>5</td>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>5</td>
<td>4</td>
<td></td>
<td>4</td>
<td></td>
<td>2</td>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td></td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>2</td>
<td>4</td>
<td>5</td>
<td></td>
<td>4</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td></td>
<td>2</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Predict by taking weighted average:**

\[
\frac{(0.2 \times 2 + 0.3 \times 3)}{(0.2 + 0.3)} = 2.6
\]
Properties of k-NN

- Intuitive
- No substantial preprocessing is required
- Easy to explain reasoning behind a recommendation
- Accurate?
k-NN on the RMSE scale

Global average: 1.1296
User average: 1.0651
Movie average: 1.0533
Cinematch: 0.9514
BellKor: 0.8693
Grand Prize: 0.8563
Inherent noise: ????

erroneous

accurate
k-NN - Common practice

1. Define a **similarity measure** between items: $s_{ij}$

2. Select **neighbors** -- $N(i;u)$:
   items most similar to $i$, that were rated by $u$

3. Estimate unknown rating, $r_{ui}$, as the **weighted average**:

   $$
   r_{ui} = b_{ui} + \frac{\sum_{j \in N(i;u)} s_{ij} (r_{uj} - b_{uj})}{\sum_{j \in N(i;u)} s_{ij}}
   $$

   **baseline estimate for $r_{ui}$**
k-NN - Common practice

1. Define a similarity measure between items: \( s_{ij} \)
2. Select neighbors -- \( N(i;u) \): items similar to \( i \), rated by \( u \)
3. Estimate unknown rating, \( r_{ui} \), as the weighted average:

\[
 r_{ui} = b_{ui} + \frac{\sum_{j \in N(i;u)} s_{ij} \left( r_{uj} - b_{uj} \right)}{\sum_{j \in N(i;u)} s_{ij}}
\]

Problems:

1. Similarity measures are arbitrary; no fundamental justification
2. Pairwise similarities isolate each neighbor; neglect interdependencies among neighbors
3. Taking a weighted average is restricting; e.g., when neighborhood information is limited
Interpolation weights

- Use a **weighted sum** rather than a **weighted average**:

\[
    r_{ui} = b_{ui} + \sum_{j \in N(i; u)} w_{ij} (r_{uj} - b_{uj})
\]

(We allow \( \sum_{j \in N(i; u)} w_{ij} \neq 1 \))

- Model relationships between item i and its neighbors
- Can be learnt through a **least squares problem** from all other users that rated i:

\[
    \text{Min}_w \sum_{v \neq u} \left( (r_{vi} - b_{vi}) - \sum_{j \in N(i; u)} w_{ij} (r_{vj} - b_{vj}) \right)^2
\]
Interpolation weights

\[
\min_w \sum_{v \neq u} \left( (r_{vi} - b_{vi}) - \sum_{j \in N(i;u)} w_{ij} (r_{vj} - b_{vj}) \right)^2
\]

- Interpolation weights derived based on their role; no use of an arbitrary similarity measure
- Explicitly account for interrelationships among the neighbors

Challenges:
- Deal with missing values
- Avoid overfitting
- Efficient implementation

Estimate inner-products among movie ratings
Latent factor models

- The Color Purple
- Amadeus
- Braveheart
- Ocean’s 11
- Lethal Weapon
- The Lion King
- Independence Day
- The Princess Diaries
- Sense and Sensibility
- Dumb and Dumber

Geared towards females: The Princess Diaries, The Color Purple, Sense and Sensibility

Geared towards males: Amadeus, Independence Day, Dumb and Dumber

serious vs. escapist
Latent factor models

A rank-3 SVD approximation

<table>
<thead>
<tr>
<th>items</th>
<th>.1</th>
<th>-.4</th>
<th>.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>-.5</td>
<td>.6</td>
<td>.5</td>
<td></td>
</tr>
<tr>
<td>-.2</td>
<td>.3</td>
<td>.5</td>
<td></td>
</tr>
<tr>
<td>1.1</td>
<td>2.1</td>
<td>.3</td>
<td></td>
</tr>
<tr>
<td>-.7</td>
<td>2.1</td>
<td>-.2</td>
<td></td>
</tr>
<tr>
<td>-.1</td>
<td>.7</td>
<td>.3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>users</th>
<th>1.1</th>
<th>-.2</th>
<th>.3</th>
<th>.5</th>
<th>-2</th>
<th>-.5</th>
<th>.8</th>
<th>-.4</th>
<th>.3</th>
<th>1.4</th>
<th>2.4</th>
<th>-.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>-.8</td>
<td>.7</td>
<td>.5</td>
<td>1.4</td>
<td>.3</td>
<td>-1</td>
<td>1.4</td>
<td>2.9</td>
<td>-.7</td>
<td>1.2</td>
<td>-.1</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>2.1</td>
<td>-.4</td>
<td>.6</td>
<td>1.7</td>
<td>2.4</td>
<td>.9</td>
<td>-.3</td>
<td>.4</td>
<td>.8</td>
<td>.7</td>
<td>-.6</td>
<td>.1</td>
<td></td>
</tr>
</tbody>
</table>
Estimate unknown ratings as inner-products of factors:

A rank-3 SVD approximation
Estimate unknown ratings as inner-products of factors:

A rank-3 SVD approximation
Estimate unknown ratings as inner-products of factors:

A rank-3 SVD approximation
Latent factor models

Properties:

• SVD isn’t defined when entries are unknown ➔ use specialized methods

• Very powerful model ➔ can easily overfit, sensitive to regularization

• Probably most popular model among contestants
  – 12/29/2006: Free implementation at timelydevelopment.com
Factorization on the RMSE scale

- Grand Prize: 0.8563
- BellKor: 0.8693
- Cinematch: 0.9514
- Movie average: 1.0533
- User average: 1.0651
- Global average: 1.1296
- Inherent noise: ????
BellKore Approach

• **User factors:**
  Model a user $u$ as a vector $p_u \sim N_k(\mu, \Sigma)$

• **Movie factors:**
  Model a movie $i$ as a vector $q_i \sim N_k(\gamma, \Lambda)$

• **Ratings:**
  Measure “agreement” between $u$ and $i$: $r_{ui} \sim N(p_u^Tq_i, \sigma^2)$

• **Maximize model’s likelihood:**
  – Alternate between recomputing user-factors, movie-factors and model parameters
  – Special cases:
    • Alternating Ridge regression
    • Nonnegative matrix factorization
Combining multi-scale views

Residual fitting

- global effects
- regional effects
- local effects

Weighted average

- factorization
- k-NN

A unified model

- factorization
- k-NN
Localized factorization model

• Standard factorization:
  User $u$ is a linear function parameterized by $p_u$
  $$r_{ui} = p_u^T q_i$$

• Allow user factors – $p_u$ – to depend on the item being predicted
  $$r_{ui} = p_u(i)^T q_i$$

• Vector $p_u(i)$ models behavior of $u$ on items like $i$
Results on Netflix Probe set

RMSE vs. #Factors

More accurate

Factorization vs. Localized factorization

CS 572: Information Retrieval. Spring 2016
The third dimension of the BellKor system

- A powerful source of information: Characterize **users** by *which* movies they rated, rather than *how* they rated

- ➔ A binary representation of the data

<table>
<thead>
<tr>
<th>users</th>
<th>movies</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>users</th>
<th>movies</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

4/6/2016
The third dimension of the BellKor system

Great news to recommender systems:

• Works even better on real life datasets (?)
• Improve accuracy by exploiting implicit feedback
• Implicit behavior is abundant and easy to collect:
  – Rental history
  – Search patterns
  – Browsing history
  – ...
• Implicit feedback allows predicting personalized ratings for users that never rated!

movie #17270
Google News Personalization: Scalable Online Collaborative Filtering
Abhinandan Das, Mayur Datar, Ashutosh Garg
WWW 2007, May 8-12, 2007
Recommended

TODAY'S NEWS
Washington Post - 11 hours ago
In 2002 he became the first person to fly around the world in a balloon by himself. In 2004 he set the record for sailing around the world faster than anyone else (36 days 9 hours).

Scientists Uncover Earliest Known Ancestor to Tyrannosaurus Rex - Chosen Ilbo
Scientists discover possible ancestor of T-rex - People's Daily Online

Melbourne Herald Sun - Independent - Boston Globe - Earthtimes.org - all 263 related »

Apple unveils cut-price iPod nano
BBC News - Feb 8, 2006
Apple has introduced a new version of its iPod nano with a smaller capacity of one gigabyte, enough for 240 songs. The move is latest tweak to the company's evolving iPod line-up, and means the ultra-slim nano now comes in 1GB, 2GB and 4GB sizes.

Apple unveils cut-price iPods - Guardian Unlimited
Gadgets, Apple offers £109.99 iPod nano - Digit Magazine
DigitTimes - BusinessWeek - MarketWatch - Fast Company - all 378 related »

Blu-Ray movies get priced
Ve3.com - Feb 8, 2006
February 08, 2006 - For those of you planning on grabbing a Blu-Ray player later this year, you may want to check out ArsTechnica’s article discussing Sony Pictures Blu-Ray pricing info released yesterday.

Sony Pictures set price for Blu-ray discs - Pocket-lint.co.uk
Sony prices Blu-ray movies - PC Pro
Sony prices Blu-ray movies - PC Pro - all 89 related »

Google Updates Desktop Search App
PC World - 3 hours ago
Google plans to release on Thursday a new beta version of its free, downloadable PC and Web search application. The new version will expand the functionality of the product's Sidebar feature, a panel that...

Google betas version of Desktop search tool - PC Pro
Google connects desktop search - VNUnet.com

Rod Herring - Reuters - ABC News - Search Engine Watch - all 200 related »

Cert asks for telco curbs
Inquirer - 11 hours ago
FATHER OF the Internet, Vint Cerf, has asked the US Congress to forbid telcos from discriminating against rival web applications, computer devices and websites.

Free happy broadband cable shows its true “net neutrality” colors - ZDNet
Google Telecoms Clash Over Net Neutrality - Mediapost

Regulator - Reuters - CNET News.com - MarketWatch - all 242 related »

US assessing Windows Vista antitrust concerns
Seattle Post Intelligencer - 6 hours ago
Antitrust concerns have been raised over aspects of Windows Vista, the next version of the operating system, according to a report Wednesday by the US Justice Department and the states that participated in Microsoft’s US antitrust settlement, as noted in ...

Microsoft under pressure as Vista ship date looms - Wall Street Journal
Challenges

• Scale
  – Number of unique visitors in last 2 months: several millions
  – Number of stories within last 2 months: several millions

• Item Churn
  – News stories change every 10-15 mins
  – Recency of stories is important
  – Cannot build models ever so often

• Noisy ratings
  – Clicks treated as noisy positive vote
Approach

• Content-based vs. Collaborative filtering
• Collaborative filtering based on **co-visitation**
  – Content agnostic: Can be applied to other application domains, languages, countries
  – Google’s key strength: Huge user base and their data.
• Could have used content based filtering
• Focus on algorithms that are scalable
Algorithm Overview

• Obtain a list of candidate stories

• For each story:
  – **Obtain 3 scores** \((y_1, y_2, y_3)\)
  – **Final score** = \(\sum w_i * y_i\)

• User clustering algorithms (Minhash \((y_1)\), PLSI \((y_2)\))
  – Score \(\sim\) number of times this story was clicked by other users in your cluster

• Story-story co-visititation \((y_3)\)
  – Score \(\sim\) number of times this story was co-clicked with other stories in your recent click history
Algorithm Overview ...

• User clustering done offline as batch process
  – Can be run every day or 2-3 times a week

• Cluster-story counts maintained in real time

• New users
  – May not be clustered
  – Rely on co-visititation to generate recommendations
Architecture

- Personalization Servers
- Cache/Buffer
- UserTable (user clusters, click hist)
- StoryTable (cluster + covisit counts)
- Bigtables
- User clustering (Offline) (Mapreduce)
- Update profile
- News Frontend Webserver
- Rank Request
- Rank Reply
- Click Notify
- Read Stats
- Update Stats
- Read user profile
- Update profile
User clustering - Minhash

• Input: User and his clicked stories
  \[ S_u = \{ s_1^u, s_2^u, ..., s_m^u \} \]

• User similarity = \[ \frac{| S_{u_1} \cap S_{u_2} |}{| S_{u_1} \cup S_{u_2} |} \]

• Output: User clusters.
  – Similar users belong to same cluster
Minhash ...

- Randomly permute the universe of clicked stories
  \[ \{ s_1, s_2, \ldots, s_m \} = \{ s'_1, s'_2, \ldots, s'_m \} \]
- \( MH(u) = \min( s^u_j ) \)  
  min defined by permutation
- \( P\{ MH(u_1) = MH(u_2) \} = \frac{| S_{u_1} \cap S_{u_2} |}{| S_{u_1} \cup S_{u_2} |} \)
- Pseudo-random permutation
  - Compute hash for each story and treat hash-value as permutation index
- Treat MinHash value as ClusterId
- Probabilistic clustering
MinHash as Mapreduce

• Map phase:
  – Input: key = userid, value = storyid
  – Compute hash for each story (parallelizable across all data)
  – Output: key = userid value = storyhash

• Reduce phase:
  – Input: key = userid value = <list of storyhash values>
  – Compute min over the story hash-values (parallelizable across users)
PLSI Framework

$$P(s|u) = \sum_z P(s|z) \cdot P(z|u)$$
Clustering - PLSI Algorithm

- Learning (done offline)
  - ML estimation
    - Learn model parameters that maximize the likelihood of the sample data
    - Output = $P[z_j|u]$’s $P[s|z_j]$’s
  - $P[z_j|u]$’s lead to a soft clustering of users
- Runtime: we only use $P[z_j|u]$’s and ignore $P[s|z_j]$’s
PLSI (EM estimation) as Mapreduce

• E step: (Map phase)
  \[ q^*(z; u,s,\emptyset) = \frac{p(s|z)p(z|u)}{\sum_z p(s|z)p(z|u)} \]

• M step: (Reduce phase)
  \[ p(s|z) = \frac{\sum_u q^*(z; u,s,\emptyset)}{\sum_z \sum_u q^*(z; u,s,\emptyset)} \]
  \[ p(z|u) = \frac{\sum_s q^*(z; u,s,\emptyset)}{\sum_z \sum_s q^*(z; u,s,\emptyset)} \]

• Cannot load the entire model from prev iteration in a single machine during map phase

• Trick: Partition users and stories. Each partition loads the stats pertinent to it
PLSI as Mapreduce
Covisitation

• For each story $s_i$ store the covisitation counts with other stories $c(s_i, s_j)$
• Candidate story: $s_k$
• User history: $s_1, ..., s_n$
• $score(s_i, s_j) = c(s_i, s_j)/\sum_m c(s_i, s_m)$
• $total\_score(s_k) = \sum_n score(s_k, s_n)$
Experimental results
Live traffic clickthrough evaluation

![Graph showing clickthrough rates over days for different categories: Popular, CSBiased, CVBiased. The x-axis represents days ranging from 0 to 160, and the y-axis represents the number of clicks as a fraction of Popular clicks, ranging from 0.5 to 3.0. The graph displays fluctuating trends over time.](image-url)
Summary

• Most effective approaches are often **hybrid** between collaborative and content-based filtering
  – c.f. Netflix, Google News recommendations

• Lots of data available, but algorithm scalability critical
  – (use Map/Reduce, Hadoop)

• Suggested rules-of-thumb (take w/ grain of salt):
  – If have moderate amounts of rating data for each user, use collaborative filtering with kNN (reasonable baseline)
  – When little data available, use content (model-) based recommendations
Who to follow · refresh · view all

- **freshbooks** FreshBooks · Follow
  - Promoted · Followed by @zappos and others.
- **alanwarms** Alan Warms · Follow
  - Followed by @fredwilson and others.
- **Mozzie21** Moises Henriquez · Follow
  - can eat

Similar to @ryanhall3 · view all

- **RunnerSpace_com** RunnerSpace.com · Follow
  - RunnerSpace.com has the latest in news and media...
- **chrislieto** chris lieto · Follow
  - Chris Lieto is a top ranked World Class Triathlete, ...
- **RUNNINGTIMES** runningtimes · Follow

Launched summer 2010
MG Siegler @parislemon · Jul 27
@kevinweil @elizabeth OMG just seeing you secured @thirdweil for baby. Most amazing handle ever. Well played. (via @amy)

Elizabeth Weil
@elizabeth

@parislemon Got it in January. Then Twitter recommended the account to my uncle. Family guessed we were pregnant. Oops. Denied it all. ;)

3:39 PM - 27 Jul 2014
“Circle of Trust”

Ordered set of important neighbors for a user

Result of egocentric random walk
Computed online based on various input parameters

One of the features used in search
SALSA for Recommendations

“hubs”

“authorities”

hubs scores: similarity scores to $u$

authority scores: recommendation scores for $u$

CoT of $u$

users LHS follow
Goel, Lin, Sharma, Wang, and Zadeh. WTF: The Who to Follow Service at Twitter. WWW 2013
What about new users?
Cold start problem: they need recommendations the most!
EXPLAINING RECOMMENDATIONS
Why was Mariko shown this Pottery Barn ad?

How can I explain the reason for a decision made by a complex predictive model with 100K or 1 million or 10 million variables?
Explanations in recommender systems

Additional information to explain the system’s output following some objectives
Objectives of explanations

- Transparency
- Validity
- Trustworthiness
- Persuasiveness
- Effectiveness

- Efficiency
- Satisfaction
- Relevance
- Comprehensibility
- Education
Explanations in general

• **How? and Why?** explanations in expert systems

• Form of abductive reasoning
  – Given: $KB \models_{RS} i$ (item i is recommended by method RS)
  – Find $KB' \subseteq KB$ s.t. $KB' \models_{RS} i$

• Principle of succinctness
  – Find smallest subset of $KB' \subseteq KB$ s.t. $KB' \models_{RS} i$
    i.e. for all $KB'' \subset KB'$ holds $KB'' \not\models_{RS} i$

• But additional filtering
  – Some parts relevant for deduction, might be obvious for humans

[Friedrich & Zanker, AI Magazine, 2011]
Taxonomy for generating explanations in RS

Major design dimensions of current explanation components:

- Category of reasoning model for generating explanations
  - White box
  - Black box

- RS paradigm for generating explanations
  - Determines the exploitable semantic relations

- Information categories
RS paradigms and their ontologies

- Classes of objects
  - Users
  - Items
  - Properties

- N-ary relations between them

- Collaborative filtering
  - Neighborhood based CF (a)
  - Matrix factorization (b)
    - Introduces additional factors as proxies for determining similarities
RS paradigms and their ontologies

• Content-based
  – Properties characterizing items
  – TF*IDF model

• Knowledge based
  – Properties of items
  – Properties of user model
  – Additional mediating domain concepts
• Similarity between items

• Similarity between users

• Tags
  – Tag relevance (for item)
  – Tag preference (of user)
Results from testing the explanation feature

- Knowledgeable explanations significantly increase the users’ perceived utility
- Perceived utility strongly correlates with usage intention etc.

** sign. < 1%, * sign. < 5%
The Evidence Counterfactual

“The explanation trick”

see (Martens & FP MISQ 2014; Chen, Moakler, Fraiberger, FP, SSRN 2015)

- Models can be viewed as evidence-combining systems
- Especially models based on massive fine-grained behavior data
- Thus, for any specific decision* from any model we can ask:

What is a minimal set of evidence such that if it were not present, the decision* would not have been made?

*The “decision” can be a threshold crossing for a prob. estimation, scoring or regression model
Example

Why was Mariko shown this Pottery Barn ad?

Because she visited:

- www.diningroomtableshowroom.com
- www.mazeltovfurniture.com
- www.realtor.com
- www.recipezaar.com
- www.americannidol.com

And now we can start also to understand why using these fine-grained data works so well.
Consumers Increasingly Concerned About Inferences Made About Them


Private traits and attributes are predictable from digital records of human behavior

Michal Kosinski*, David Stillwell*, and Thore Graepel

*Free School Lane, The Psychometrics Centre, University of Cambridge, Cambridge CB2 3RQ United Kingdom; and "Microsoft Research, Cambridge CB1 2FB, United Kingdom

We show that easily accessible digital records of behavior, such as Facebook Likes, can be used to automatically and accurately predict a range of highly sensitive personal attributes including: sexual orientation, ethnicity, religious and political views, personality traits, intelligence, happiness, use of addictive substances, parental separation, age, and gender. The analysis presented here is based on a dataset of over 58,000 volunteers who provided their Facebook Likes, detailed demographic profiles, and the results of several psychometric tests. The proposed model uses dimensionality reduction for pre-processing the Likes data, which are then entered into logistic/linear regression to predict individual psychodemographic profiles from Likes. The model correctly discriminates between homosexual and heterosexual men in 88% of cases, African Americans and Caucasian Americans in 95% of cases, and between Democrat and Republican in 85% of cases. For the personality trait "Openness," prediction accuracy is close to the test--test accuracy of a standard personality test. We give examples of associations between attributes and Likes and discuss implications for online personalization and privacy.

A growing proportion of human activities, such as social interactions, entertainment, shopping, and gathering information, are now mediated by digital services and devices. Such digitally mediated behaviors can easily be recorded and analyzed, fueling the emergence of computational social science (1) and new applications.

2.4% females), whereas those interested in users of the opposite gender were labeled as heterosexual.

**Results**

**Prediction of Dichotomous Variables.** Fig. 2 shows the prediction accuracy of dichotomous variables expressed in terms of the area under the receiver-operating characteristic curve (AUC), which is equivalent to the probability of correctly classifying two randomly selected users one from each class (e.g., male and female). The highest accuracy was achieved for ethnic origin and gender. African Americans and Caucasian Americans were correctly classified in 95% of cases, and males and females were correctly classified in 93% of cases, suggesting that patterns of online behavior as expressed by Likes significantly differ between those groups allowing for nearly perfect classification.

Christians and Muslims were correctly classified in 82% of cases, and similar results were achieved for Democrats and Republicans (85%). Sexual orientation was easier to distinguish among males (88%) than females (75%), which may suggest a wider behavioral divide (as observed from online behavior) between hetero- and homosexual males.

Good prediction accuracy was achieved for relationship status and substance use (between 65% and 73%). The relatively lower accuracy for relationship status may be explained by its temporal variability compared with other dichotomous variables (e.g., gender or sexual orientation).

The model’s accuracy was lowest (60%) when inferring whether users’ parents stayed together or separated before users were 21 years old. Although it is known that parental divorce does have long-

![Area Under Curve](chart.png)
Critical Evidence ("Counter-factual")

- Models can be viewed as evidence-combining systems.
- For a specific decision, can ask:
  - What is a minimal set of evidence such that if it were not present, the decision would not have been made?

<table>
<thead>
<tr>
<th>Task</th>
<th>Num. Users</th>
<th>% Positive</th>
<th>Avg. Likes</th>
</tr>
</thead>
<tbody>
<tr>
<td>age ≥ 37</td>
<td>145,400</td>
<td>0.127</td>
<td>216</td>
</tr>
<tr>
<td>agreeableness ≥ 5</td>
<td>136,974</td>
<td>0.014</td>
<td>218</td>
</tr>
<tr>
<td>conscientiousness ≥ 5</td>
<td>136,974</td>
<td>0.018</td>
<td>218</td>
</tr>
<tr>
<td>extraversion ≥ 5</td>
<td>136,974</td>
<td>0.033</td>
<td>218</td>
</tr>
<tr>
<td>iq ≥ 130</td>
<td>4,540</td>
<td>0.130</td>
<td>186</td>
</tr>
<tr>
<td>iq &lt; 90</td>
<td>4,540</td>
<td>0.073</td>
<td>186</td>
</tr>
<tr>
<td>is democrat</td>
<td>7,301</td>
<td>0.596</td>
<td>262</td>
</tr>
<tr>
<td>is drinking</td>
<td>3,351</td>
<td>0.485</td>
<td>262</td>
</tr>
<tr>
<td>is female</td>
<td>164,285</td>
<td>0.616</td>
<td>209</td>
</tr>
<tr>
<td>is gay</td>
<td>22,383</td>
<td>0.046</td>
<td>192</td>
</tr>
<tr>
<td>is homosexual</td>
<td>51,703</td>
<td>0.035</td>
<td>257</td>
</tr>
<tr>
<td>is lesbian</td>
<td>29,320</td>
<td>0.027</td>
<td>307</td>
</tr>
<tr>
<td>is muslim</td>
<td>11,600</td>
<td>0.050</td>
<td>238</td>
</tr>
<tr>
<td>is single</td>
<td>124,863</td>
<td>0.535</td>
<td>226</td>
</tr>
<tr>
<td>is smoking</td>
<td>3,376</td>
<td>0.237</td>
<td>261</td>
</tr>
<tr>
<td>life satisfaction ≥ 6</td>
<td>5,958</td>
<td>0.125</td>
<td>252</td>
</tr>
<tr>
<td>network density ≥ 65</td>
<td>32,704</td>
<td>0.012</td>
<td>214</td>
</tr>
<tr>
<td>neuroticism ≥ 5</td>
<td>136,974</td>
<td>0.004</td>
<td>218</td>
</tr>
<tr>
<td>num friends ≥ 585</td>
<td>32,704</td>
<td>0.140</td>
<td>214</td>
</tr>
<tr>
<td>openness ≥ 5</td>
<td>136,974</td>
<td>0.043</td>
<td>218</td>
</tr>
<tr>
<td>ss belief = 1</td>
<td>13,900</td>
<td>0.178</td>
<td>229</td>
</tr>
<tr>
<td>ss belief = 5</td>
<td>13,900</td>
<td>0.079</td>
<td>229</td>
</tr>
<tr>
<td>uses drugs</td>
<td>2,490</td>
<td>0.172</td>
<td>264</td>
</tr>
</tbody>
</table>
Removing Evidence: Cloacking

Effect of removing selected Likes from consideration by the predictive model

Opportunity for offering users control via a “cloaking device”?

My IQ intelligence test
- The Big Bang Theory
- I don’t sleep enough because I stay up late for no reason
- Coldplay
- Brain Buddies
- Magic: The Gathering
- Fight Club
- IQ Test
- Megan Fox
- Lord Of The Rings

Probability (IQ ≤ 130)

Number of top-wise likes removed
How Many Likes Have to Remove?

<table>
<thead>
<tr>
<th>Task</th>
<th>All</th>
<th>$\eta_{0.9}$</th>
<th>FP</th>
</tr>
</thead>
<tbody>
<tr>
<td>is lesbian</td>
<td>3.075</td>
<td>5.437</td>
<td>2.829</td>
</tr>
<tr>
<td>is homosexual</td>
<td>3.493</td>
<td>6.572</td>
<td>2.888</td>
</tr>
<tr>
<td>extraversion $\geq 5$</td>
<td>4.428</td>
<td>5.944</td>
<td>4.300</td>
</tr>
<tr>
<td>conscientiousness $\geq 5$</td>
<td>4.746</td>
<td>6.746</td>
<td>4.670</td>
</tr>
<tr>
<td>agreeableness $\geq 5$</td>
<td>4.985</td>
<td>6.508</td>
<td>4.957</td>
</tr>
<tr>
<td>num friends $\geq 585$</td>
<td>5.043</td>
<td>6.556</td>
<td>4.197</td>
</tr>
<tr>
<td>life satisfaction $\geq 6$</td>
<td>5.128</td>
<td>7.214</td>
<td>4.642</td>
</tr>
<tr>
<td>is gay</td>
<td>5.653</td>
<td>10.944</td>
<td>3.161</td>
</tr>
<tr>
<td>ss belief = 1</td>
<td>5.738</td>
<td>6.880</td>
<td>4.946</td>
</tr>
<tr>
<td>iq $\geq 130$</td>
<td>6.566</td>
<td>3.429</td>
<td>7.283</td>
</tr>
<tr>
<td>openness $\geq 5$</td>
<td>6.674</td>
<td>7.677</td>
<td>6.571</td>
</tr>
<tr>
<td>is drinking</td>
<td>6.771</td>
<td>7.463</td>
<td>3.875</td>
</tr>
<tr>
<td>iq $&lt; 90$</td>
<td>6.867</td>
<td>16.318</td>
<td>4.582</td>
</tr>
<tr>
<td>ss belief = 5</td>
<td>8.251</td>
<td>11.098</td>
<td>7.760</td>
</tr>
<tr>
<td>is smoking</td>
<td>8.357</td>
<td>9.800</td>
<td>5.621</td>
</tr>
<tr>
<td>is democrat</td>
<td>8.462</td>
<td>8.533</td>
<td>2.000</td>
</tr>
<tr>
<td>neuroticism $\geq 5$</td>
<td>9.140</td>
<td>5.667</td>
<td>9.173</td>
</tr>
<tr>
<td>is female</td>
<td>9.971</td>
<td>10.015</td>
<td>5.475</td>
</tr>
<tr>
<td>age $\geq 37$</td>
<td>10.259</td>
<td>13.011</td>
<td>5.847</td>
</tr>
<tr>
<td>network density $\geq 65$</td>
<td>10.545</td>
<td>15.308</td>
<td>10.388</td>
</tr>
<tr>
<td>is muslim</td>
<td>11.706</td>
<td>27.804</td>
<td>2.930</td>
</tr>
<tr>
<td>uses drugs</td>
<td>12.161</td>
<td>12.143</td>
<td>12.176</td>
</tr>
<tr>
<td>is single</td>
<td>13.665</td>
<td>15.514</td>
<td>7.888</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td><strong>7.465</strong></td>
<td><strong>9.851</strong></td>
<td><strong>5.572</strong></td>
</tr>
</tbody>
</table>
Explanations in recommender systems: Summary

• There are many types of explanations and various goals that an explanation can achieve.

• Which type of explanation can be generated depends greatly on the recommender approach applied.

• Explanations may be used to shape the wishes and desires of customers but are a double-edged sword.
  – On the one hand, explanations can help the customer to make wise buying decisions;
  – On the other hand, explanations can be abused to push a customer in a direction which is advantageous solely for the seller.

• As a result, a deep understanding of explanations and their effects on customers is of great interest.
Resources

• Recommender Systems Survey:
  http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1423975

• Koren et al. on winning the NetFlix challenge:
  – http://videolectures.net/kdd09_koren_cftd/

• http://www.recommenderbook.net/