Explanations of Recommendations

- recommendations: selection (ranked list) of items
- explanations: (some) reasons for the choice
Goals of Providing Explanations

Why explanations?
Goals of Providing Explanations

Why explanations?

- transparency, trustworthiness, validity, satisfaction (users are more likely to use the system)
- persuasiveness (users are more likely to follow recommendations)
- effectiveness, efficiency (users can make better/faster decisions)
- education (users understand better the behaviour of the system, may use it in better ways)
Examples of Explanations

- knowledge-based recommenders
  - “Because you, as a customer, told us that simple handling of car is important to you, we included a special sensor system in our offer that will help you park your car easily.”
- algorithms based on CSP representation
Examples of Explanations

- knowledge-based recommenders
  - “Because you, as a customer, told us that simple handling of car is important to you, we included a special sensor system in our offer that will help you park your car easily.”
- algorithms based on CSP representation
- recommendations based on item-similarity
  - “Because you watched X we recommend Y”
Explanations – Collaborative Filtering

Your Neighbors’ Ratings for this Movie

<table>
<thead>
<tr>
<th>Rating</th>
<th>Number of Neighbors</th>
</tr>
</thead>
<tbody>
<tr>
<td>★</td>
<td>1</td>
</tr>
<tr>
<td>★★</td>
<td>2</td>
</tr>
<tr>
<td>★★★</td>
<td>7</td>
</tr>
<tr>
<td>★★★★</td>
<td>14</td>
</tr>
<tr>
<td>★★★★★</td>
<td>9</td>
</tr>
</tbody>
</table>

Your Neighbors' Ratings for this Movie

![Bar Chart]

- 1's and 2's: 3 neighbors
- 3's: 7 neighbors
- 4's and 5's: 23 neighbors

Explaining Collaborative Filtering Recommendations, Herlocker, Konstan, Riedl
Explanations – Collaborative Filtering

Figure 4. A screen explaining the recommendation for the movie “The Sixth Sense.” Each bar represents a rating of a neighbor. Upwardly trending bars are positive ratings, while downward trending ones are negative. The x-axis represents similarity to the user.
## Explanations – Comparison

<table>
<thead>
<tr>
<th>#</th>
<th>Explanation</th>
<th>N</th>
<th>Mean Response</th>
<th>Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Histogram with grouping</td>
<td>76</td>
<td>5.25</td>
<td>1.29</td>
</tr>
<tr>
<td>2</td>
<td>Past performance</td>
<td>77</td>
<td>5.19</td>
<td>1.16</td>
</tr>
<tr>
<td>3</td>
<td>Neighbor ratings histogram</td>
<td>78</td>
<td>5.09</td>
<td>1.22</td>
</tr>
<tr>
<td>4</td>
<td>Table of neighbors ratings</td>
<td>78</td>
<td>4.97</td>
<td>1.29</td>
</tr>
<tr>
<td>5</td>
<td>Similarity to other movies rated</td>
<td>77</td>
<td>4.97</td>
<td>1.50</td>
</tr>
<tr>
<td>6</td>
<td>Favorite actor or actress</td>
<td>76</td>
<td>4.92</td>
<td>1.73</td>
</tr>
<tr>
<td>7</td>
<td>MovieLens percent confidence in prediction</td>
<td>77</td>
<td>4.71</td>
<td>1.02</td>
</tr>
<tr>
<td>8</td>
<td>Won awards</td>
<td>76</td>
<td>4.67</td>
<td>1.49</td>
</tr>
<tr>
<td>9</td>
<td>Detailed process description</td>
<td>77</td>
<td>4.64</td>
<td>1.40</td>
</tr>
<tr>
<td>10</td>
<td># neighbors</td>
<td>75</td>
<td>4.60</td>
<td>1.29</td>
</tr>
<tr>
<td>11</td>
<td>No extra data – focus on system</td>
<td>75</td>
<td>4.53</td>
<td>1.20</td>
</tr>
<tr>
<td>12</td>
<td>No extra data – focus on users</td>
<td>78</td>
<td>4.51</td>
<td>1.35</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>#</th>
<th>Description</th>
<th>N</th>
<th>Mean Response</th>
<th>Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>MovieLens confidence in prediction</td>
<td>77</td>
<td>4.51</td>
<td>1.20</td>
</tr>
<tr>
<td>14</td>
<td>Good profile</td>
<td>77</td>
<td>4.45</td>
<td>1.53</td>
</tr>
<tr>
<td>15</td>
<td>Overall percent rated 4+</td>
<td>75</td>
<td>4.37</td>
<td>1.26</td>
</tr>
<tr>
<td>16</td>
<td>Complex graph: count, ratings, similarity</td>
<td>74</td>
<td>4.36</td>
<td>1.47</td>
</tr>
<tr>
<td>17</td>
<td>Recommended by movie critics</td>
<td>76</td>
<td>4.21</td>
<td>1.47</td>
</tr>
<tr>
<td>18</td>
<td>Rating and agreement of closest neighbor</td>
<td>77</td>
<td>4.21</td>
<td>1.20</td>
</tr>
<tr>
<td>19</td>
<td># neighbors with std. deviation</td>
<td>78</td>
<td>4.19</td>
<td>1.45</td>
</tr>
<tr>
<td>20</td>
<td># neighbors with avg correlation</td>
<td>76</td>
<td>4.08</td>
<td>1.46</td>
</tr>
<tr>
<td>21</td>
<td>Overall average rating</td>
<td>77</td>
<td>3.94</td>
<td>1.22</td>
</tr>
</tbody>
</table>

**Table 1.** Mean response of users to each explanation interface, based on a scale of one to seven. Explanations 11 and 12 represent the base case of no additional information. Shaded rows indicate explanations with a mean response significantly different from the base cases (two-tailed $\alpha = 0.05$).
Attacks on Recommender System

- Why?
- What type of recommender systems?
- How?
- Countermeasures?
Attacks

susceptible to attacks: collaborative filtering

reasons for attack:

- make the system worse (unusable)
- influence rating (recommendations) of a particular item
  - push attacks – improve rating of “my” items
  - nuke attacks – decrease rating of “opponent’s” items
Fig. 2  Simplified system database showing authentic user profiles and a number of attack profiles inserted. In this example, user $h$ is seeking a prediction for item 7, which is the subject of a product nuke attack.
Types of Attacks

more knowledge about system → more efficient attack

random attack generate profiles with random values
  (preferably with some typical ratings)
average attack effective attack on memory-based systems
  (average ratings → many neighbors)
bandwagon attack high rating for “blockbusters”, random
  values for others
segment attack insert ratings only for items from specific
  segment
special nuke attacks love/hate attack, reverse bandwagon
Fig. 3 Prediction shift (left) and hit ratio (right) for product push attacks mounted against the user-based collaborative recommendation algorithm. Hit ratio results relate to a 10% attack size.
Countermeasures

- more robust techniques: model based techniques, additional information
- increasing injection costs: Captcha, limited number of accounts for single IP address
- automated attack detection
Attacks and Educational Systems

- cheating $\sim$ false rating
  example: Problem Solving Tutor, Binary crossword
- gaming the system – using hints as solutions

can have similar consequences as attacks
Context Aware Recommendations

describes context:

- **physical** – location, time
- **environmental** – weather, light, sound
- **personal** – health, mood, schedule, activity
- **social** – who is in room, group activity
- **system** – network traffic, status of printers
Contextualization

- pre- post- filtering
- model based
  - multidimensionality: user × item × time ×...
  - tensor factorization
Context – Applications

- tourism, visitor guides
- museum guides
- home computing and entertainment