Author(s): Rahul Sami, 2009

License: Unless otherwise noted, this material is made available under the terms of the Creative Commons Attribution Noncommercial Share Alike 3.0 License: http://creativecommons.org/licenses/by-nc-sa/3.0/

We have reviewed this material in accordance with U.S. Copyright Law and have tried to maximize your ability to use, share, and adapt it. The citation key on the following slide provides information about how you may share and adapt this material.

Copyright holders of content included in this material should contact open.michigan@umich.edu with any questions, corrections, or clarification regarding the use of content.

For more information about how to cite these materials visit http://open.umich.edu/education/about/terms-of-use.
Use + Share + Adapt

{ Content the copyright holder, author, or law permits you to use, share and adapt. }

- **Public Domain – Government**: Works that are produced by the U.S. Government. (USC 17 § 105)
- **Public Domain – Expired**: Works that are no longer protected due to an expired copyright term.
- **Public Domain – Self Dedicated**: Works that a copyright holder has dedicated to the public domain.
- **Creative Commons – Zero Waiver**
- **Creative Commons – Attribution License**
- **Creative Commons – Attribution Share Alike License**
- **Creative Commons – Attribution Noncommercial License**
- **Creative Commons – Attribution Noncommercial Share Alike License**
- **GNU – Free Documentation License**

Make Your Own Assessment

{ Content Open.Michigan believes can be used, shared, and adapted because it is ineligible for copyright. }

- **Public Domain – Ineligible**: Works that are ineligible for copyright protection in the U.S. (USC 17 § 102(b)) *laws in your jurisdiction may differ

{ Content Open.Michigan has used under a Fair Use determination. }

- **Fair Use**: Use of works that is determined to be Fair consistent with the U.S. Copyright Act. (USC 17 § 107) *laws in your jurisdiction may differ

  Our determination **DOES NOT** mean that all uses of this 3rd-party content are Fair Uses and we **DO NOT** guarantee that your use of the content is Fair.

  To use this content you should **do your own independent analysis** to determine whether or not your use will be Fair.
Lecture 5: User-User Recommender

SI583: Recommender Systems
Generating recommendations

- Core problem: predict how much a person “Joe” (is likely to) like an item “X”
- Then, can decide to recommend most likely successes, filter out items below a threshold, etc.
<table>
<thead>
<tr>
<th>user</th>
<th>item</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>.</th>
<th>.</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joe</td>
<td></td>
<td>7</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>?</td>
</tr>
<tr>
<td>Sue</td>
<td></td>
<td>7</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>John</td>
<td></td>
<td>2</td>
<td>3</td>
<td>7</td>
<td></td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>
User-User recommenders: Intuition

- Assumption: If Joe and another user agreed on other items, they are more likely to agree on X

- Collaborative filtering approach:
  - For each user, find how similar that user is to Joe on other ratings
  - Find the pool of users “closest” to Joe in taste
  - Use the ratings of those users to come up with a prediction
User-user algorithm: Details to be formalized

- How is similarity measured?
  - how are ratings normalized?
- How is the pool of neighbors selected?
- How are different users’ ratings weighted in the prediction for Joe?
CF Algorithms in the Literature

- Sometimes classified as *memory-based vs. model-based*

- **Model based:** statistically predict an unknown rating
  - Fit a statistical model, then estimate
  - E.g., SVD

- **Memory-based:** ad-hoc use of previous ratings
  - No explicit class of models, although sometimes retrofit
  - E.g., user-user, item-item
# Measures of similarity

<table>
<thead>
<tr>
<th>user</th>
<th>item</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joe</td>
<td></td>
<td>7</td>
<td>3</td>
<td>7</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Sue</td>
<td></td>
<td>6</td>
<td>4</td>
<td>6</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>John</td>
<td></td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Amy</td>
<td></td>
<td>9</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Bob</td>
<td></td>
<td>7</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Some possible similarity metrics

For all our metrics: focus on the ratings on items that both i and j have rated

- $\text{Similarity}(i,j) =$ number of items on which $i$ and $j$ have exactly the same rating
Some possible similarity metrics

- \( \text{Similarity}(i,j) = \) number of items on which \( i \) and \( j \) have the same rating
  - intuitive objection: we would have \( \text{similarity}(\text{Joe}, \text{John}) > \text{similarity}(\text{Joe}, \text{Sue}) \)
Some possible similarity metrics

- $Similarity(i,j) = \text{number of items on which } i \text{ and } j \text{ have the same rating}$
  - intuitive objection: we would have $similarity(Joe, John) > similarity(Joe, Sue)$

- $Similarity(i,j) = (i\text{'s rating vector}) \cdot (j\text{'s rating vector})^T$
Some possible similarity metrics

- $\text{Simila}rity(i,j) = \text{number of items on which } i \text{ and } j \text{ have the same rating}$
  - intuitive objection: we would have similarity(Joe, John) > similarity(Joe, Sue)

- $\text{Simila}rity(i,j) = (i\text{'s rating vector}) (j\text{'s rating vector})^T$
  - intuitive objection: we would have similarity(Joe, John) > similarity(Joe, Sue)
Some possibilities..

- Normalize for mean rating:
  - Let $\mu_i = i$’s average rating
  - Let $i$’s normalized rating vector
    \[ x_i = (\text{rating on A} - \mu_i, \text{rating on B} - \mu_i, \ldots) \]
  - Define similarity$(i,j) = x_i^T x_j$
# Mean-normalized ratings

<table>
<thead>
<tr>
<th>user</th>
<th>item</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joe</td>
<td></td>
<td>2</td>
<td>-2</td>
<td>2</td>
<td>-2</td>
<td></td>
</tr>
<tr>
<td>Sue</td>
<td></td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td></td>
</tr>
<tr>
<td>John</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Amy</td>
<td></td>
<td>5</td>
<td>-2</td>
<td>-1</td>
<td>-2</td>
<td></td>
</tr>
<tr>
<td>Bob</td>
<td></td>
<td>2</td>
<td>-2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Some possibilities..

- Normalize for mean rating:
  - Let $\mu_i = i$’s average rating
  - Let $i$’s normalized rating vector $x_i = (\text{rating on A} - \mu_i, \text{rating on B} - \mu_i, \ldots)$
  - Define $\text{similarity}(i,j) = x_i \cdot x_j^T$

- Objection:
  $\text{similarity}(\text{Joe, Amy}) > \text{similarity}(\text{Joe, John})$
Normalizing for mean and standard deviation

- Normalize for mean rating:
  - Let $\mu_i = i$’s average rating
  - Let $i$’s normalized rating vector
    $$x_i = (\text{rating on A} - \mu_i, \text{rating on B} - \mu_i, \ldots)$$

- Then, normalize for standard deviation
  - $z_i = (1/\sigma)x_i$
  - where $\sigma = ||x_i|| = \sqrt{x_i(A)^2 + x_i(B)^2 + \ldots/(#\text{items rated by } i)}$

- Define
  $$\text{similarity}(i,j) = z_i \cdot z_j^T$$
Mean-std.dev normalized ratings (z-scores)

<table>
<thead>
<tr>
<th>user</th>
<th>item</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joe</td>
<td></td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td></td>
</tr>
<tr>
<td>Sue</td>
<td></td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td></td>
</tr>
<tr>
<td>John</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Amy</td>
<td></td>
<td>1.7</td>
<td>-0.7</td>
<td>-0.3</td>
<td>-0.7</td>
<td></td>
</tr>
<tr>
<td>Bob</td>
<td></td>
<td>1</td>
<td>-1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Normalizing for mean and standard deviation

- Normalize for mean rating:
  - Let $\mu_i = i$’s average rating
  - Let $i$’s normalized rating vector $x_i = (\text{rating on A} - \mu_i, \text{rating on B} - \mu_i, \ldots)$

- Then, normalize for standard deviation
  - $z_i = (1/\sigma)x_i$
  - where $\sigma = ||x_i|| = \sqrt{x_i(A)^2 + x_i(B)^2 + \ldots/(\#\text{items})}$

- (Modified for different numbers of ratings):
  similarity($i,j$) = $z_i . z_j^T/(\#\text{items})$ [Pearson correlation coefficient]
Pearson correlation coefficient

- Intuitively: similarity measure that
  - adjusts for different average rating for different users
  - adjusts for different swing magnitudes for different users
  - adjusts for different numbers of common ratings

- Also has a good statistical justification
  - arises naturally in a statistical model.
Correlation: Statistical justification

Statistical model:
- Item $w$ drawn randomly from some space
- Each user’s rating is a random variable:
  - $i$’s rating can be represented by $r_i(w)$
- **Goal:** Estimate $r_{Joe}(item)$ from observing $r_{Sue}(item)$, $r_{John}(item)$, etc..
- If $r_j$ is independent of $r_i$, $r_j$ is useless for estimating $r_i$
- The more correlated $r_j$ is with $r_i$, the more useful it is (independence => correlation = 0)
- Correlation can be estimated from common ratings
Linear Algebra Representation

- **R**: \([n \times m]\) matrix representing \(n\) users’ ratings on \(m\) items
- **X**: \([n \times m]\) matrix representing ratings normalized by user means
- **Z**: \([n \times m]\) matrix representing z-scores (normalized ratings)
Mathematical representation

- **R**: \([n \times m]\) matrix representing \(n\) users’ ratings on \(m\) items
- **X**: \([n \times m]\) matrix representing ratings normalized by user means
- **Z**: \([n \times m]\) matrix representing \(z\)-scores (normalized ratings)

If matrices are complete:

- **C=XX^T** is an \([n \times n]\) matrix of covariances
  - \(C_{ij}/(\text{#items i&j rated})\) estimates covariance of \(r_i, r_j\)
- **P=ZZ^T** is an \([n \times n]\) matrix of correlations
  - \(P_{ij}/(\text{#items i&j rated})\) estimates correlation of \(r_i, r_j\)
Other similarity measures

- Any distance measure between vectors can be used to define a similarity

- e.g., “cosine similarity”
  - treat rating vectors as lines in space, similarity based on how small the angle between $i$ and $j$ is

- How do you decide which one is best?
Other similarity measures

- Any distance measure between vectors can be used to define a similarity

- e.g., “cosine similarity”
  - treat rating vectors as lines in space, similarity based on how small the angle between $i$ and $j$ is

- How do you decide which one is best?
  - intuitively judge what normalizations are important
  - try them out empirically on your data!
User-user algorithm: Details to be formalized

- How is similarity measured?
  - how are ratings normalized?
- How is the pool of neighbors selected?
- How are different users’ ratings weighted in the prediction for Joe?
Choosing a pool of neighbors

- **Common approach: k-nearest neighbors**
  - Pick up to k users who have rated X, in order of decreasing similarity to X
  - *parameter k is typically about 20-50*

- **Alternative: Thresholding**
  - Pick all users with correlation coefficients greater than t who have rated X
  - *threshold t >0 is recommended*
Weighting users

- Users’ ratings on X are weighted according to computed similarities
- Prediction for Joe is weighted average
  - $w_{ij} = \text{Pearson correlation similarity}(i,j)$
  - predicted $z_{\text{Joe}}(x) = \sum_{i \text{ in pool}} w_{i,\text{Joe}} z_i(x)$
- Denormalize to compute predicted rating
  - predicted $r_{\text{Joe}(x)} = \mu_{\text{Joe}} + z_{\text{Joe}}(x) \sigma_{\text{Joe}}$
**Example: Predict Joe’s rating for X**

<table>
<thead>
<tr>
<th>user</th>
<th>item</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>…</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joe</td>
<td></td>
<td>7</td>
<td>3</td>
<td>7</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sue</td>
<td></td>
<td>6</td>
<td>4</td>
<td>6</td>
<td>4</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>John</td>
<td></td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amy</td>
<td></td>
<td>9</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>Bob</td>
<td></td>
<td>7</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>6</td>
</tr>
</tbody>
</table>
Example: z-scores

<table>
<thead>
<tr>
<th>user</th>
<th>item</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>...</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joe</td>
<td></td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sue</td>
<td></td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td></td>
<td>-1</td>
</tr>
<tr>
<td>John</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amy</td>
<td></td>
<td>1.7</td>
<td>-0.7</td>
<td>-0.3</td>
<td>-0.7</td>
<td></td>
<td>0.8</td>
</tr>
<tr>
<td>Bob</td>
<td></td>
<td>1</td>
<td>-1</td>
<td></td>
<td></td>
<td></td>
<td>0.6</td>
</tr>
</tbody>
</table>
Example: weights and predictions

- similarity (Amy, Joe) = 0.95
- similarity (Sue, Joe) = 1
- similarity (Bob, Joe) = 1

- predicted $z_{Joe}(x) = -0.36$
- predicted rating = $5 - 2 \times 0.36 = 4.22$
### Recommendations [Herlocker et al, Information and Retrieval, 2002]

**Table 8.** A tabulation of recommendations based on the results presented in this chapter.

<table>
<thead>
<tr>
<th>Recommendation</th>
<th>Recommended</th>
<th>Not recommended</th>
</tr>
</thead>
<tbody>
<tr>
<td>Similarity weighting</td>
<td>Pearson correlation</td>
<td>Spearman, entropy, vector similarity, mean-squared difference</td>
</tr>
<tr>
<td>(Section 5.1)</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Significance weighting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Section 5.2)</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Selecting neighbors</td>
<td>Set max number of neighbors</td>
<td>Weight thresholding</td>
</tr>
<tr>
<td>(Section 6)</td>
<td>(potentially in the range of 20–60 nbors)</td>
<td></td>
</tr>
<tr>
<td>Rating normalization</td>
<td>Deviation-from-mean or z-score</td>
<td>No normalization</td>
</tr>
<tr>
<td>(Section 7.1)</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Weighting neighbor contributions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Section 7.2)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>