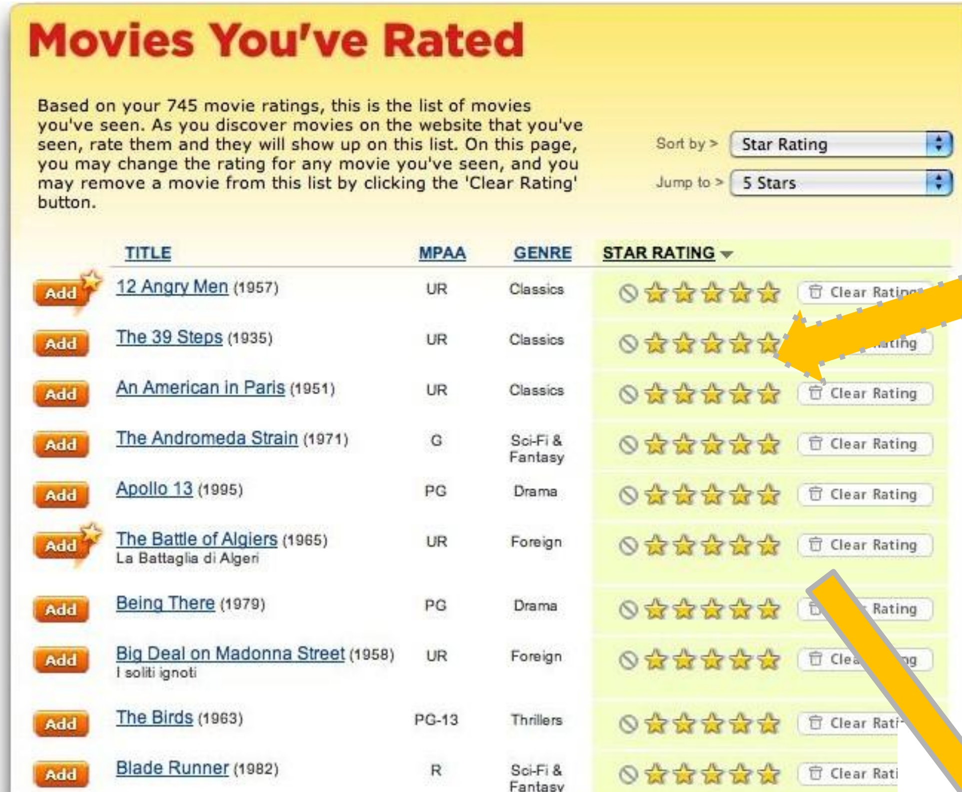
Collaborative Filtering

- Collaborative filtering is the process of selecting information or patterns using techniques involving collaboration among multiple agents, viewpoints, data sources, etc.
- The main advantage of this method is that the recommender system *does not need to have additional information about the users or content of the items*
- Users' rating or purchase history is the only information that is needed to work

Rating Matrix: An Example



Movies You've Rated

Based on your 745 movie ratings, this is the list of movies you've seen. As you discover movies on the website that you've seen, rate them and they will show up on this list. On this page, you may change the rating for any movie you've seen, and you may remove a movie from this list by clicking the 'Clear Rating' button.

Sort by > Star Rating

Jump to > 5 Stars

	TITLE	MPAA	GENRE	STAR RATING
Add	12 Angry Men (1957)	UR	Classics	Clear Rating
Add	The 39 Steps (1935)	UR	Classics	Clear Rating
Add	An American in Paris (1951)	UR	Classics	Clear Rating
Add	The Andromeda Strain (1971)	G	Sci-Fi & Fantasy	Clear Rating
Add	Apollo 13 (1995)	PG	Drama	Clear Rating
Add	The Battle of Algiers (1965) La Battaglia di Algeri	UR	Foreign	Clear Rating
Add	Being There (1979)	PG	Drama	Clear Rating
Add	Big Deal on Madonna Street (1958) I soliti ignoti	UR	Foreign	Clear Rating
Add	The Birds (1963)	PG-13	Thrillers	Clear Rating
Add	Blade Runner (1982)	R	Sci-Fi & Fantasy	Clear Rating

Value	Graphic representation	Textual representation
5	☆ ☆ ☆ ☆ ☆	Excellent
4	☆ ☆ ☆ ☆	Very good
3	☆ ☆ ☆	Good
2	☆ ☆	Fair
1	☆	Poor

Table 9.1: User-Item Matrix

	Lion King	Aladdin	Mulan	Anastasia
John	3	0	3	3
Joe	5	4	0	2
Jill	1	2	4	2
Jane	3	?	1	0
Jorge	2	2	0	1

Rating Matrix

Users rate (rank) the items (purchased, watched)

- **Explicit ratings:** entered by a user directly
  - (i.e., “Please rate this on a scale of 1-5”)

- **Implicit ratings:** inferred from other user behavior
  - Play lists or music listened to, for a music Rec system
  - The amount of time users spent on a webpage

Collaborative Filtering

Types of Collaborative Filtering Algorithms:

- **Memory-based**: Recommendation is directly based on previous ratings in the stored matrix that describes user-item relations

- **Model-based**: Alternatively, one can assume that an underlying model (hypothesis) governs the way users rate items.
  This model can be approximated and learned. The model is then used to recommend ratings.
  A model, for example, is that users rate low budget movies poorly

Memory-Based Collaborative Filtering

The most important assumptions of collaborative filtering are:

- **User-based CF**

  Users with similar **previous** ratings for items are likely to rate future items similarly

	I1	I2	I3	I4
U1	1	2	4	4
U2	1	2	4	?
U3	2	5	2	2
U4	5	2	3	3

- **Item-based CF**

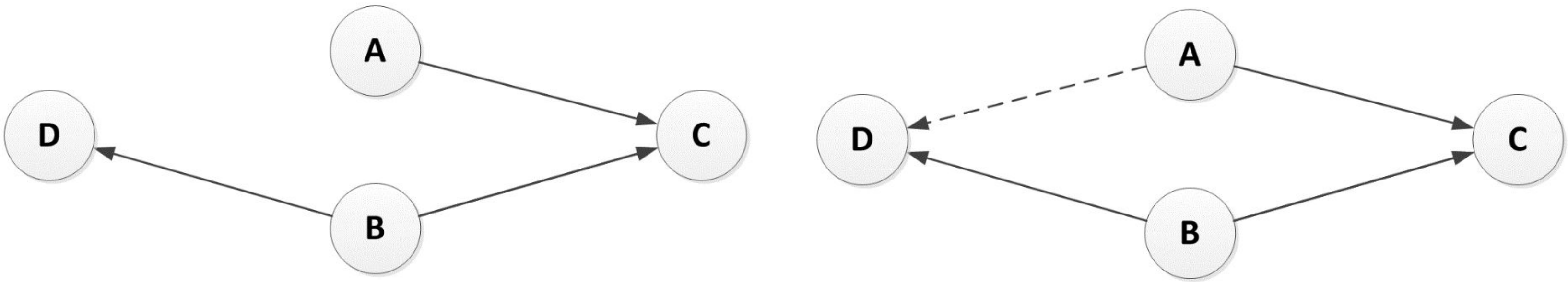
  Items that have received similar ratings **previously** from users are likely to receive similar ratings from future users (*item-based CF)*

	I1	I2	I3	I4
U1	1	2	4	4
U2	1	2	4	?
U3	2	5	2	2
U4	5	2	3	3

Collaborative Filtering: Algorithm

1. Weigh all users/items with respect to their similarity with the current user/item
2. Select a subset of the users/items (neighbors) as recommenders
3. Predict the rating of the user for specific items using neighbors’ ratings for the same (or similar) items
4. Recommend items with the highest predicted rank

Collaborative Filtering: Illustration



- A and B are following C, and B is following D
  - A might be interested in following D

Measure Similarity between Users (or Items)

- Cosine Similarity

$$sim(U_i, U_j) = cos(U_i, U_j) = \frac{U_i \cdot U_j}{||U_i||\ ||U_j||} = \frac{\sum_k r_{i,k} r_{j,k}}{\sqrt{\sum_k {r_{i,k}}^2} \sqrt{\sum_k {r_{j,k}}^2}}.$$

- Pearson Correlation Coefficient

$$sim(U_i, U_j) = \frac{\sum_k (r_{i,k} - \bar{r}_i)(r_{j,k} - \bar{r}_j)}{\sqrt{\sum_k (r_{i,k} - \bar{r}_i)^2} \sqrt{\sum (r_{j,k} - \bar{r}_j)^2}}.$$

User-based Collaborative Filtering

- In user-based collaborative filtering, the system finds the most **similar user (users) to the current user** and uses their preferences for recommendation
- The user-based approach is *not as popular* as the item-based approach
  - For a system that handles a large user base, even the smallest change in the user data is likely to reset the entire group of similar users

User-based CF

Updating the ratings:

User v's mean rating

User u's mean rating

$$r_{u,i} = \bar{r}_u + \frac{\sum_{v \in N(u)} sim(u,v)(r_{v,i} - \bar{r}_v)}{\sum_{v \in N(u)} sim(u,v)},$$

Predicted rating of user u for item i

Observed rating of user v for item i

User-based CF, Example

$$r_{u,i} = \bar{r}_u + \frac{\sum_{v \in N(u)} sim(u,v)(r_{v,i} - \bar{r}_v)}{\sum_{v \in N(u)} sim(u,v)}$$

	Lion King	Aladdin	Mulan	Anastasia
John	3	0	3	3
Joe	5	4	0	2
Jill	1	2	4	2
Jane	3	?	1	0
Jorge	2	2	0	1

Predict Jane's rating for Aladdin

3- Calculate Jane's rating for Aladdin, Assume that neighborhood size = 2

1- Calculate average ratings

$$\bar{r}_{John} = \frac{3+3+0+3}{4} = 2.25$$

$$\bar{r}_{Joe} = \frac{5+4+0+2}{4} = 2.75$$

$$\bar{r}_{Jill} = \frac{1+2+4+2}{4} = 2.25$$

$$\bar{r}_{Jane} = \frac{3+1+0}{3} = 1.33$$

$$\bar{r}_{Jorge} = \frac{2+2+0+1}{4} = 1.25$$

2- Calculate user-user similarity

$$sim(Jane, John) = \frac{3\times3+1\times3+0\times3}{\sqrt{10}\sqrt{27}} = 0.73$$

$$sim(Jane, Joe) = \frac{3\times5+1\times0+0\times2}{\sqrt{10}\sqrt{29}} = 0.88$$

$$sim(Jane, Jill) = \frac{3\times1+1\times4+0\times2}{\sqrt{10}\sqrt{21}} = 0.48$$

$$sim(Jane, Jorge) = \frac{3\times2+1\times0+0\times1}{\sqrt{10}\sqrt{5}} = 0.84$$

User-based CF, Example- continued

3- Calculate Jane's rating for Aladdin,
Assume that neighborhood size = 2

$$
\begin{aligned}
r_{Jane,Aladdin} &= \bar{r}_{Jane} + \frac{sim(Jane, Joe)(r_{Joe,Aladdin} - \bar{r}_{Joe})}{sim(Jane, Joe) + sim(Jane, Jorge)} \\
&\quad + \frac{sim(Jane, Jorge)(r_{Jorge,Aladdin} - \bar{r}_{Jorge})}{sim(Jane, Joe) + sim(Jane, Jorge)} \\
&= 1.33 + \frac{0.88(4 - 2.75) + 0.84(2 - 1.25)}{0.88 + 0.84} = 2.33
\end{aligned}
$$

Item-based CF

Calculate the similarity between items and then predict new items based on the past ratings for similar items

$$r_{u,i} = \bar{r}_i + \frac{\sum_{j \in N(i)} sim(i,j)(r_{u,j} - \bar{r}_j)}{\sum_{j \in N(i)} sim(i,j)},$$

Item i's mean rating

i and j are two items

Item-based CF, Example

1- Calculate average ratings

$$\bar{r}_{Lion\ King} = \frac{3+5+1+3+2}{5} = 2.8$$

$$\bar{r}_{Aladdin} = \frac{0+4+2+2}{4} = 2.$$

$$\bar{r}_{Mulan} = \frac{3+0+4+1+0}{5} = 1.6$$

$$\bar{r}_{Anastasia} = \frac{3+2+2+0+1}{5} = 1.6$$

2- Calculate item-item similarity

$$sim(Aladdin, Lion\ King) = \frac{0\times 3+4\times 5+2\times 1+2\times 2}{\sqrt{24}\sqrt{39}} = 0.84$$

$$sim(Aladdin, Mulan) = \frac{0\times 3+4\times 0+2\times 4+2\times 0}{\sqrt{24}\sqrt{25}} = 0.32$$

$$sim(Aladdin, Anastasia) = \frac{0\times 3+4\times 2+2\times 2+2\times 1}{\sqrt{24}\sqrt{18}} = 0.67$$

3- Calculate Jane's rating for Aladdin, Assume that neighborhood size = 2

$$\begin{aligned} r_{Jane,Aladdin} &= \bar{r}_{Aladdin} + \frac{sim(Aladdin, Lion\ King)(r_{Jane,Lion\ King} - \bar{r}_{Lion\ King})}{sim(Aladdin, Lion\ King) + sim(Aladdin, Anastasia)} \\ &\quad + \frac{sim(Aladdin, Anastasia)(r_{Jane,Anastasia} - \bar{r}_{Anastasia})}{sim(Aladdin, Lion\ King) + sim(Aladdin, Anastasia)} \\ &= 2 + \frac{0.84(3-2.8) + 0.67(0-1.6)}{0.84+0.67} = 1.40 \end{aligned}$$

Model-Based Collaborative Filtering

- **In memory-based methods** (either item-based or user-based), we aim to predict the missing ratings based on similarities between users or items.
- **In model-based collaborative filtering**, we assume that an underlying model governs the way users rate.
- We aim to learn the model and then use that model to predict the missing ratings.
  - Among a variety of model-based techniques, we focus on a well-established model-based technique that is based on singular value decomposition (SVD).

Models in Collaborative Filtering

- What is a model?
  - "Readers rate a low budget movie poorly."
  - A latent model that governs the way users rate
    - Singular Value Decomposition is a method to find it
- SVD is a linear algebra technique of lossless matrix factorization to
  - Decompose X into 3 matrices ($U\Sigma V^T$)
- Obtain a low-rank (k) approximation C of X
  - Minimizes Frobenius norm of (X-C)
  - Maps X into a k-d space X_k (=C)
- Recommend using X_k

Singular Value Decomposition (SVD)

- SVD is a linear algebra technique that, given a real matrix $X \in R^{m \times n}$, $m \geq n$, and factorizes it into three matrices,

$$X = U \Sigma V^T$$

where $U \in R^{m \times m}$ and $V \in R^{n \times n}$ are orthogonal matrices and $\Sigma \in R^{m \times n}$ is a diagonal matrix.

- The product of these matrices is equivalent to the original matrix; therefore, no information is lost.

Low-rank Matrix Approximation

- A Low-rank matrix approximation of matrix X is another matrix $C \in R^{m\times n}$.
- C approximates X, and C's rank (the maximum number of linearly independent columns) is a fixed number k << min(m, n)

$$\mathrm{Rank}(C) = k$$

- The best low-rank matrix approximation is a matrix C that minimizes $||X-C||_F$

$$||X||_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij}^2}$$

  - Low-rank approximations of matrices remove noise by assuming that the matrix is not generated at random and has an underlying structure.
  - SVD can help remove **noise** by computing a low-rank approximation of a matrix.

Low-rank Matrix Approximation with SVD

1. Create Σ_k from Σ by keeping only the first k elements on the diagonal. This way, $\Sigma_k \in \mathbb{R}^{k \times k}$.

2. Keep only the first k columns of U and denote it as $U_k \in \mathbb{R}^{m \times k}$, and keep only the first k rows of V^T and denote it as ${V_k}^T \in \mathbb{R}^{k \times n}$.

3. Let $X_k = U_k \Sigma_k {V_k}^T$, $X_k \in \mathbb{R}^{m \times n}$.

X_k is the best low-rank approximation of a matrix X

Theorem 9.1 (Eckart-Young-Mirsky Low-Rank Matrix Approximation). *Let X be a matrix and C be the best low-rank approximation of X; if $||X - C||_F$ is minimized, and rank(C) = k, then $C = X_k$.*

Model-based CF, Example

Table 9.2: An User-Item Matrix

	Lion King	Aladdin	Mulan
John	3	0	3
Joe	5	4	0
Jill	1	2	4
Jorge	2	2	0

$$U = \begin{bmatrix} -0.4151 & -0.4754 & -0.7679 & 0.1093 \\ -0.7437 & 0.5278 & 0.0169 & -0.4099 \\ -0.4110 & -0.6626 & 0.6207 & -0.0820 \\ -0.3251 & 0.2373 & 0.1572 & 0.9018 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 8.0265 & 0 & 0 \\ 0 & 4.3886 & 0 \\ 0 & 0 & 2.0777 \\ 0 & 0 & 0 \end{bmatrix}$$

$$V^T = \begin{bmatrix} -0.7506 & -0.5540 & -0.3600 \\ 0.2335 & 0.2872 & -0.9290 \\ -0.6181 & 0.7814 & 0.0863 \end{bmatrix}$$

Considering a rank 2 approximation (i.e., k = 2), we truncate all three matrices:

$$U_k = \begin{bmatrix} -0.4151 & -0.4754 \\ -0.7437 & 0.5278 \\ -0.4110 & -0.6626 \\ -0.3251 & 0.2373 \end{bmatrix}$$

$$\Sigma_k = \begin{bmatrix} 8.0265 & 0 \\ 0 & 4.3886 \end{bmatrix}$$

$$V_k^T = \begin{bmatrix} -0.7506 & -0.5540 & -0.3600 \\ 0.2335 & 0.2872 & -0.9290 \end{bmatrix}$$

What is Jill's rating for Lion King?

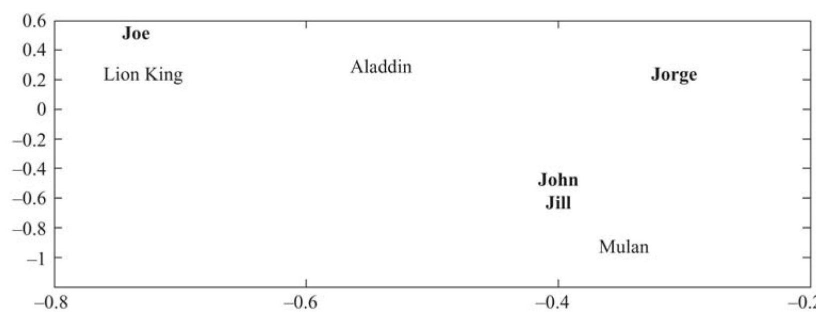


Figure 9.1: Users and Items in the 2-D Space.

Recommendation to a Group

Recommendation to Groups

- Find content of interest to all members of a group of socially acquainted individuals

- Examples:
  - A movie for friends to watch together
  - A travel destination for a family to spend a holiday break
  - A good restaurant for colleagues to have a working lunch
  - A music to be played in a public area

Tasks of a Group Recommender System

- Acquiring preferences
- Generating recommendations
- Explaining recommendations
- Helping group members to achieve consensus

Aggregation Strategies

- **Maximizing Average Satisfaction**
  – Average everyone's ratings and choose the max

$$R_i = \frac{1}{n} \sum_{u \in G} r_{u,i}$$

- **Least Misery**
  – This approach tries to minimize the dissatisfaction among group's members (Max of the mins of all)

$$R_i = \min_{u \in G} r_{u,i}$$

- **Most Pleasure**
  – The maximum of individuals' maximum ratings is taken as group's rating

$$R_i = \max_{u \in G} r_{u,i}$$

Recommendation to Group, an Example

Table 9.3: User-Item Matrix

	Soda	Water	Tea	Coffee
John	1	3	1	1
Joe	4	3	1	2
Jill	2	2	4	2
Jorge	1	1	3	5
Juan	3	3	4	5

Group: John Jill, Juan

Average Satisfaction

$$R_{Soda} = \frac{1+2+3}{3} = 2.$$
$$R_{Water} = \frac{3+2+3}{3} = 2.66$$
$$R_{Tea} = \frac{1+4+4}{3} = 3.$$
$$R_{Coffee} = \frac{1+2+5}{3} = 2.66$$

Least Misery

$$R_{Soda} = \min\{1,2,3\} = 1$$
$$R_{Water} = \min\{3,2,3\} = 2$$
$$R_{Tea} = \min\{1,4,4\} = 1$$
$$R_{Coffee} = \min\{1,2,5\} = 1$$

Most Pleasure

$$R_{Soda} = \max\{1,2,3\} = 3$$
$$R_{Water} = \max\{3,2,3\} = 3$$
$$R_{Tea} = \max\{1,4,4\} = 4$$
$$R_{Coffee} = \max\{1,2,5\} = 5$$

What to recommend? The maximum in each measure