P9120 - Statistical Learning and Data Mining

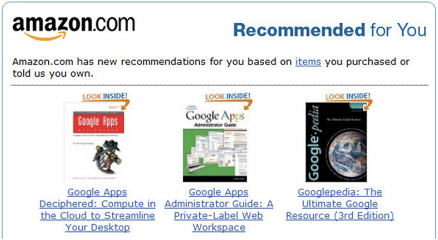
Lecture 13 - Recommender Systems

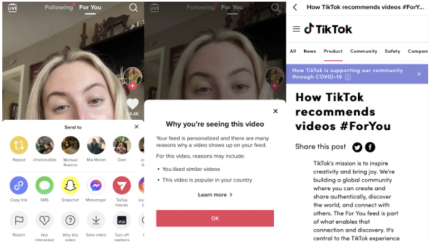
Min Qian

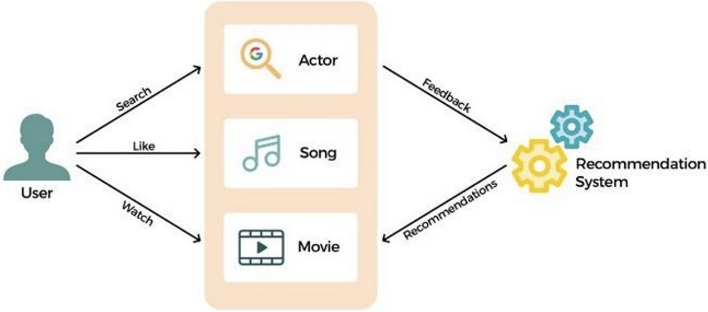
Department of Biostatistics, Columbia University

December 5th, 2024

Recommender Systems







Outline

1. Collaborative Filtering
   - User-based collaborative filtering
   - Item-based collaborative filtering
   - Model-based collaborative filtering

2. Content-based Filtering
   - Content-based filtering using only item features
   - Content-based filtering using both user and item features

Collaborative Filtering

Collaborative filtering uses similarities between users and items simultaneously to provide recommendations.

Question: Which movie should be recommended to user A?

	Harry Potter			Twilight	Star Wars		
Users	HP1	HP2	HP3	TW	SW1	SW2	SW3
A	4	?	3	4	1	?	?
B	5	4	4			3	
C		1		2	4	5	
D		3	3	3		3	2
E	5	5	5		2		3

Solution: Predict user A's movie ratings for HP2, SW2, and SW3, and recommend the one(s) with the highest ratings.

User-based Collaborative Filtering: Similarity Measure

	Harry Potter			Twilight	Star Wars			
Users	HP1	HP2	HP3	TW	SW1	SW2	SW3	avg
A	4	?	3	4	1	?	?	3
B	5	4	4			3		4
C		1		2	4	5		3
D		3	3	3		3	2	2.8
E	5	5	5		2		3	4

Mean-centered cosine similarity:

1. Center the ratings for each user.
2. Impute unknown ratings by 0.
3. Calculate cosine similarity.

Users	HP1	HP2	HP3	TW	SW1	SW2	SW3
A	1	0	0	1	-2	0	0
B	1	0	0	0	0	-1	0
C	0	-2	0	-1	1	2	0
D	0	0.2	0.2	0.2	0	0.2	-0.8
E	1	1	1	0	-2	0	-1

$sim(A, B) = 0.29$

$sim(A, C) = -0.39$

$sim(A, D) = 0.09$

$sim(A, E) = 0.72$

Other similarity measures, e.g. cosine without mean-centering and zero imputation.

User-based Collaborative Filtering: Rating Prediction

Users	HP1	HP2	HP3	TW	SW1	SW2	SW3	avg	
A	4	?	3	4	1			3	
B	5	4	4			3		4	$sim(A,B) = 0.29$
C		1		2	4	5		3	$sim(A,C) = -0.39$
D		3	3	3		3	2	2.8	$sim(A,D) = 0.09$
E	5	5	5		2		3	4	$sim(A,E) = 0.72$

$\mathcal{K}$: the set of K users most similar to user A who have rated item j

Y_{kj} : user-k's rating for item-j.

- Naive: $\widehat{Y}_{Aj} = \frac{1}{K}\sum_{k\in\mathcal{K}} Y_{kj} = \frac{4+5}{2} = 4.5.$
- Weighted average: $\widehat{Y}_{Aj} = \frac{\sum_{k\in\mathcal{K}} \text{sim}(A,k) Y_{kj}}{\sum_{k\in\mathcal{K}} \text{sim}(A,k)} = \frac{0.29*4+0.72*5}{0.29+0.72} = 4.7.$
- Mean-adjusted weighted average:
$\widehat{Y}_{Aj} = \bar{Y}_{A\cdot} + \frac{\sum_{k\in\mathcal{K}} \text{sim}(A,k)(Y_{kj}-\bar{Y}_{k\cdot})}{\sum_{k\in\mathcal{K}} \text{sim}(A,k)} = 3 + \frac{0.29*(4-4)+0.72*(5-4)}{0.29+0.72} = 3.7.$

Item-based Collaborative Filtering

Users	HP1	HP2	HP3	TW	SW1	SW2	SW3
A	4	?	3	4	1		
B	5	4	4			3	
C		1		2	4	5	
D		3	3	3		3	2
E	5	5	5		2		3
avg	4.67	3.25	3.75	3	2.33	3.67	2.5

$\mathcal{J}$: set of J items rated by user A and similar to HP2.

Users	HP1	HP2	HP3	TW	SW1	SW2	SW3
A	-0.67	0	-0.75	1	-1.33	0	0
B	0.33	0.75	0.25	0	0	-0.67	0
C	0	-2.25	0	-1	1.67	1.33	0
D	0	-0.25	-0.75	0	0	-0.67	-0.5
E	0.33	1.75	1.25	0	-0.33	0	0.5
avg	4.67	3.25	3.75	3	2.33	3.67	2.5

$$sim(HP2, HP1) = 0.35$$
$$sim(HP2, HP3) = 0.52$$
$$sim(HP2, TW) = 0.54$$
$$sim(HP2, SW1) = -0.68$$
$$sim(HP2, SW2) = -0.69$$
$$sim(HP2, SW3) = 0.48$$

Item-based Collaborative Filtering

Users	HP1	HP2	HP3	TW	SW1	SW2	SW3
A	4	?	3	4	1		
B	5	4	4			3	
C		1		2	4	5	
D		3	3	3		3	2
E	5	5	5		2		3
avg	4.67	3.25	3.75	3	2.33	3.67	2.5

Predicted rating:

$$\hat{Y}_{A,HP2} = \frac{\sum_{j\in\mathcal{J}} sim(HP2,j) Y_{A,j}}{\sum_{j\in\mathcal{J}} sim(HP2,j)} = \frac{0.52*3+0.54*4}{0.52+0.54} = 3.5$$

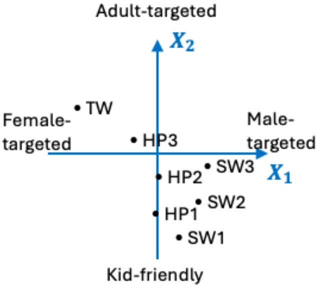
$\mathcal{J}$: set of J items rated by user A and similar to HP2.

Users	HP1	HP2	HP3	TW	SW1	SW2	SW3
A	-0.67	0	-0.75	1	-1.33	0	0
B	0.33	0.75	0.25	0	0	-0.67	0
C	0	-2.25	0	-1	1.67	1.33	0
D	0	-0.25	-0.75	0	0	-0.67	-0.5
E	0.33	1.75	1.25	0	-0.33	0	0.5
avg	4.67	3.25	3.75	3	2.33	3.67	2.5

$sim(HP2, HP1) = 0.35$
$sim(HP2, HP3) = 0.52$
$sim(HP2, TW) = 0.54$
$sim(HP2, SW1) = -0.68$
$sim(HP2, SW2) = -0.69$
$sim(HP2, SW3) = 0.48$

Model-based Collaborative Filtering

	A	B	C	D	E	X_1	X_2
HP1	4	5			5		
HP2		4	1	3	5		
HP3	3	4		3	5		
TW	4		2	3			
SW1	1		4		2		
SW2		3	5	3			
SW3				2	3		



Idea: predict user-i's rating for item-j by $\mathbf{x}_{j\cdot} w_i + b_i$, where

- $\mathbf{x}_{j\cdot} = (x_{j1}, \ldots, x_{jk})$ is a row vector of feature values of item-j
- $w_i \in \mathbb{R}^k, b_i \in \mathbb{R}$ are parameters for user-i

Latent Model

- Y_{ij} : user-i's rating of item-j.
- $1_{ij} = 1$ if user-i has rated item-j, and 0 o.w.
- n : # of users, m : # of items

Assume $\{\mathbf{x}_{j\cdot} : j = 1, \ldots, m\}$ are known.

- Estimate user-i's parameters (w_i, b_i) by minimizing

$$\sum_{j=1}^{m} 1_{ij}[Y_{ij} - (\mathbf{x}_{j\cdot} w_i + b_i)]^2 + \lambda \|w_i\|_2^2$$

- Sum over all users, estimate $\{(w_i, b_i) : i = 1, \ldots, n\}$ by minimizing

$$\sum_{i=1}^{n} \sum_{j=1}^{m} 1_{ij}[Y_{ij} - (\mathbf{x}_{j\cdot} w_i + b_i)]^2 + \lambda \sum_{i=1}^{n} \|w_i\|_2^2 \tag{1}$$

For binary outcome, use binomial deviance loss.

Latent Model

Assume $\{(w_i, b_i) : i = 1, \ldots, n\}$ are known.

- Estimate item-j's feature values $\mathbf{x}_{j\cdot}$ by minimizing

$$\sum_{i=1}^{n} 1_{ij}[Y_{ij} - (\mathbf{x}_{j\cdot} w_i + b_i)]^2 + \lambda \|\mathbf{x}_{j\cdot}\|_2^2$$

- Sum over all items, estimate $\{\mathbf{x}_{j\cdot} : j = 1, \ldots, m\}$ by minimizing

$$\sum_{j=1}^{m} \sum_{i=1}^{n} 1_{ij}[Y_{ij} - (\mathbf{x}_{j\cdot} w_i + b_i)]^2 + \lambda \sum_{j=1}^{m} \|\mathbf{x}_{j\cdot}\|_2^2 \tag{2}$$

Combining (1) and (2), estimate $\{(w_i, b_i) : i = 1, \ldots, n\}$ and $\{\mathbf{x}_{j\cdot} : j = 1, \ldots, m\}$ by minimizing

$$\sum_{i=1}^{n} \sum_{j=1}^{m} 1_{ij}[Y_{ij} - (\mathbf{x}_{j\cdot} w_i + b_i)]^2 + \lambda_1 \sum_{i=1}^{n} \|w_i\|_2^2 + \lambda_2 \sum_{j=1}^{m} \|\mathbf{x}_{j\cdot}\|_2^2$$

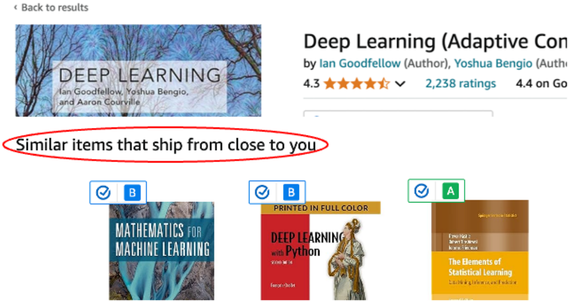
Mean Normalization

How to predict the rating for a new user / new item?

	A	B	C	D	E	F	X_1	X_2
HP1	4	5			5	?		
HP2		4	1	3	5	?		
HP3	3	4		3	5	?		
TW	4		2	3		?		
SW1	1		4		2	?		
SW2		3	5	3		?		
SW3				2	3	?		
Frozen	?	?	?	?	?	?		

- new users: normalize rating by item-specific mean
- new items: normalize rating by user-specific mean

Finding related items



The learned features $X_1, X_2, \dots$ can help find similar items.

- For item-j, find item-k such that $\mathbf{x}_{k\cdot}$ is similar to $\mathbf{x}_{j\cdot}$, i.e., $\arg\min_k \|\mathbf{x}_{k\cdot} - \mathbf{x}_{j\cdot}\|_2^2$.

Content-based Filtering

Use features of user and item to find good match.

- With only item features:
  - Construct user's feature profile using weighted item feature profile, where the weights depend on the user's previous rating of different items.
  - Recommend items with feature profile(s) that are similar to user's profile.
- With both item features and user features
  - Use both user's features and item's features to predict user's rating for the item.
  - Recommend items to user with high predicted rating.

Item Feature Profile

Profile is a set of features

- Genre
- Actor
- Theme
- Avg rating
- ...

Features	HP1	HP2	HP4	TW	SW1	SW2	SW3
Romance	0	0	0	1	0	0	0
Mystery	1	1	1	1	0	0	0
PG-13	0	0	0	1	0	0	1
Watson?	1	1	1	0	0	0	0
Avg rating	4.7	3.3	3.8	3	2.4	3.7	2.5

Use a scaling factor α to balance between binary and numeric features

Features	HP1	HP2	HP4	TW	SW1	SW2	SW3
Romance	0	0	0	1	0	0	0
Mystery	1	1	1	1	0	0	0
PG-13	0	0	1	1	0	0	1
Watson?	1	1	1	0	0	0	0
Avg rating	4.7α	3.3α	3.8α	3α	2.4α	3.7α	2.5α

$$cos(HP2, HP4) = \frac{2+12.5\alpha^2}{\sqrt{(2+10.9\alpha^2)(3+14.4\alpha^2)}} = 0.97\ \alpha = 1,\ 0.85 \text{ for } \alpha = 0.2$$

User Profile based on Item Profile

	User A's rating							
	4		3	4	1			
Features	**HP1**	**HP2**	**HP4**	**TW**	**SW1**	**SW2**	**SW3**	**User A profile**
Romance	0	0	0	1	0	0	0	
Mystery	1	1	1	1	0	0	0	
PG-13	0	0	1	1	0	0	1	
Watson?	1	1	1	0	0	0	0	
Avg rating	0.94	0.66	0.76	0.6	0.48	0.74	0.5	

Build user-A's feature profile using weighted average of item profiles that user A has rated.

- Simple: user A's ratings
- Variant: user-specific mean normalized ratings

Making Recommendations

- For each candidate item j, calculate cosine similarity between user-A's profile, $\mathbf{x}_u^{(A)}$, and item j's profile $\mathbf{x}_m^{(j)}$:

$$\cos(A, j) = \frac{\langle \mathbf{x}_u^{(A)}, \mathbf{x}_m^{(j)} \rangle}{\|\mathbf{x}_u^{(A)}\| \|\mathbf{x}_m^{(j)}\|}.$$

- Recommend items with high similarity to user-A's profile.

Content-based Filtering with User and Item Features

User-i's feature $\mathbf{x}_u^{(i)}$:

- Age
- Gender
- movies watched
- expressed preferences
- ...

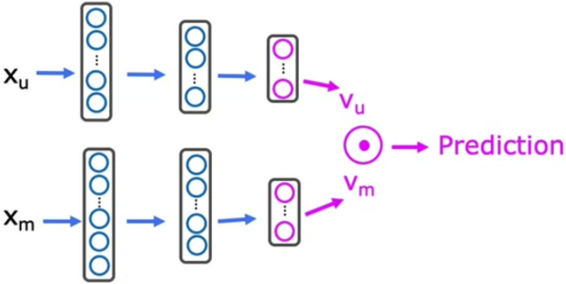
Item-j's features $\mathbf{x}_m^{(j)}$:

- Genre
- Actor
- Theme
- Average rating
- ...

Idea:

- Embed $\mathbf{x}_u^{(i)}$ and $\mathbf{x}_m^{(j)}$ into $\mathbf{v}_u^{(i)}$ and $\mathbf{v}_m^{(j)}$, respectively, where $\mathbf{v}_u^{(i)}$ and $\mathbf{v}_m^{(j)}$ are vectors of the same length.
- Predict user i's rating of item j by $\langle \mathbf{v}_u^{(i)}, \mathbf{v}_u^{(i)} \rangle$

Neural Network Architecture to Learn the Embeddings



Learn parameters in NN by minimizing

$$\sum_{i=1}^{n} \sum_{j=1}^{m} 1_{ij} (\mathbf{v}_u^{(i)} \cdot \mathbf{v}_m^{(j)} - Y_{ij})^2 + \text{ NN regularization terms}$$

Similar items can be found by minimizing $\|\mathbf{v}_m^{(k)} - \mathbf{v}_m^{(j)}\|$

Summary

Collaborative Filtering

- Need enough users in the system to find a good match.
- Not good recommendation for new items/users (cold start problem).
- Tends to recommend popular items.
- Does not use side information about items or users.

Content-Based Filtering

- No need for data from other users.
- Able to recommend new and unpopular items.
- Content features can help explain recommended items.
- Need careful design of features.
- Unable to exploit quality judgments of other users.

Reminder: Quiz for Lecture 13 is due at 9pm on Monday, Dec. 9th.
Hw #4 is due at 12pm on Monday, Dec. 9th.

Final Project

- In-depth exploration of one methodology.
- If the method you plan to explore is not covered in class, a detailed introduction to it must be included in the final report.
- Include theoretical results and/or simulation studies and/or data analysis.
- Conclude with your results and discussion.
- Do not simply reproduce the results of a paper.
- Judged based on clarity, thoroughness, and originality.
- Reports must be $\leq$5 pages (not including tables, figures, and references), single spaced, 12 point font.
- The Final Project Report is due at 9pm on December 22nd.