Admin

Sign up for project demo

- 2mins: briefly talk about your baseline and step 2
- 6 mins: show step 3 code and results, main improvement
- Edit the shared file to sign up
  - Wednesday 9-10, 12-1
  - Friday 9-10, 10-11, 12-1, 1-2

Today

- Recommender systems

Recommender systems

- A **recommender system** is a subclass of information filtering system that seeks to predict the "rating" or "preference" a user would give to an item
- Two approaches
  - Collaborative Filtering
    - Based on feedback from other users who have rated a similar set of items in the past
  - Content based filtering
    - Based on how well the contend of the target item matches the user's preferred content pattern, which is learnt from the user's own past ratings and the content pattern of the rated items.
    - More like personalized search
- Hybrid

Example of movie ratings:

Name	Avengers	Star wars	Thor	Spider-man	Iron Man
Alex	4	2	?	5	4
Bob	5	3	4	?	3
Tom	3	?	4	4	3

Objectives of the recommender systems

- Recommend Top-N movies to a specific user, e.g. recommend 2 movies for Alex
  - Predict the ratings of movies that are not rated by Alex
  - Sort the movies based on the ratings of Alex
  - Select the top-2 rated movies

Name	Avengers	Star wars	Thor	Spider-man	Iron Man
Alex	4	2	?	5	4
Bob	5	3	4	?	3
Tom	3	?	4	4	3

User-Based Collaborative Filtering

Step 1: Calculate the similarity between Alex and all other users.

$$sim(x,y) = \cos(\vec{x},\vec{y}) = \frac{\vec{x}\cdot\vec{y}}{\|\vec{x}\|_2 \times \|\vec{y}\|_2} = \frac{\sum_{s\in S_{xy}} r_{x,s} r_{y,s}}{\sqrt{\sum_{s\in S_{xy}} r_{x,s}^2}\sqrt{\sum_{s\in S_{xy}} r_{y,s}^2}}, \tag{13}$$

User-Based Collaborative Filtering

Step 1: Calculate the similarity between Alex and all other users.

sim(Alex, Bob) = (4 * 5 + 2 * 3 + 4 * 3) / [sqrt(4^2 + 2^2 + 4^2) * sqrt(5^2 + 3^2 + 3^2)] = 0.97

Name	Avengers	Star wars	Thor	Spider-man	Iron Man
Alex	4	2	*?*	*5*	4
Bob	5	3	*4*	*?*	3

sim	Bob	Tom
Alex	0.97	1

sim(Alex, Tom)=1

User-Based Collaborative Filtering

Step 2: Predict the ratings of movies that are not rated by Alex

$$\text{(a) } r_{c,s} = \frac{1}{N}\sum_{c' \in \hat{C}} r_{c',s},$$

$$\text{(b) } r_{c,s} = k\sum_{c' \in \hat{C}} sim(c,c') \times r_{c',s}, \qquad (10)$$

$$\text{(c) } r_{c,s} = \bar{r}_c + k\sum_{c' \in \hat{C}} sim(c,c') \times (r_{c',s} - \bar{r}_{c'}),$$

where multiplier k serves as a normalizing factor and is usually selected as $k = 1/\sum_{c' \in \hat{C}} |sim(c,c')|$, and where the average rating of user c, $\bar{r}_c$, in (10c) is defined as[1]

User-Based Collaborative Filtering

Step 2: Predict the ratings of movies that are not rated by Alex

- $k = 1/ (0.97 + 1) = 0.51$
- $R_{(alex, thor)} = k * [sim(alex, bob) * R_{(bob, thor)} + sim(alex, tom) * R_{(tom, thor)}]$

$= 0.51 * (0.97 * 4 + 1 * 4) = 4.02$

sim	Bob	Tom
Alex	0.97	1

sim(Alex, Tom)=1

Name	Thor
Alex	?
Bob	4
Tom	4

User-Based Collaborative Filtering

Step 2: Predict the ratings of movies that are not rated by Alex

Name	Avengers	Star wars	Thor	Spider-man	Iron Man
Alex	4	2	*4.02*	5	4

User-Based Collaborative Filtering

Step 3: Select top-2 rated movies

- Spider-man and Thor will be selected as the recommendations for Alex

Name	Avengers	Star wars	Thor	Spider-man	Iron Man
Alex	4	2	*4.02*	5	4

Item-Based Collaborative Filtering

Step 1: transpose the user-item matrix to the item-user matrix

	Alex	Bob	Tom
Avengers	4	5	3
Star wars	2	3	?
Thor	?	4	4
Spider-man	5	?	4
Iron Man	4	3	3

Item-Based Collaborative Filtering

Step 2: Calculate the similarity between any two items to fill up the item-item similarity matrix

	Avengers	Star wars	Thor	Spider-man	Iron Man
Avengers					
Star wars					
Thor					
Spider-man					
Iron Man					

Item-Based Collaborative Filtering

Step 2: Calculate the similarity between any two items to fill up the item-item similarity matrix

Example - Avenger VS Start wars:

- Find the ratings of these two movies from the users who rated both of them

	Alex	Bob	Tom
Avengers	4	5	3
Star wars	2	3	?

Item-Based Collaborative Filtering

Step 2: Calculate the similarity between any two items to fill up the item-item similarity matrix

Example - Avenger VS Start wars:

- Calculate the cosine similarity between these two movies

$$sim(i,j) = \cos(\vec{i},\vec{j}) = \frac{\vec{i}\cdot\vec{j}}{\|\vec{i}\|_2 * \|\vec{j}\|_2}$$

Item-Based Collaborative Filtering

Step 2: Calculate the similarity between any two items to fill up the item-item similarity matrix

Example - Avenger VS Start wars:

- Calculate the cosine similarity between these two movies

$$sim(Avengers, Star\ wars) = (4 * 2 + 5 * 3) / [sqrt(4^2 + 5^2) * sqrt(2^2 + 3^2)] = 0.99624059$$

	Alex	Bob	Tom
Avengers	4	5	3
Star wars	2	3	?

Item-Based Collaborative Filtering

Step 2: Calculate the similarity between any two items to fill up the item-item similarity matrix

	Avengers	Star wars	Thor	Spider-man	Iron Man
Avengers					
Star wars	0.996240				
Thor	0.970142	0			
Spider-man	0.999512	0	0		
Iron Man	0.970142	0.942990	1	0.999512	

Item-Based Collaborative Filtering

Step 3: Predict the ratings of movies that are not rated by Alex

$$P_{u,i} = \frac{\sum_{\text{all similar items, N}} (s_{i,N} * R_{u,N})}{\sum_{\text{all similar items, N}} (|s_{i,N}|)}$$

Item-Based Collaborative Filtering

Step 3: Predict the ratings of movies that are not rated by Alex

- $R_{(alex,\ thor)} = (sim(t,a) * R_{(alex,a)} + sim(t,\ i) * R_{(alex,i)}) / (sim(t,a) + sim(t,\ i))$

$= 4$

Name	Avengers	Star wars	Thor	Spider-man	Iron Man
Alex	4	2	?	5	4

	Avengers	Star wars	Spider-man	Iron Man
Thor	0.9701425	0	0	1

Item-Based Collaborative Filtering

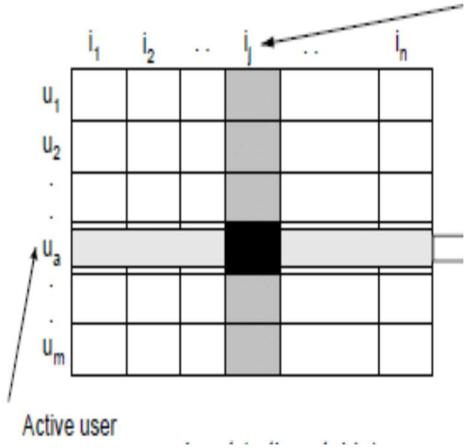
Step 4: Select top-2 rated movies

- Spider-man and Thor or Avengers or Iron Man will be selected as the recommendations for Alex

Name	Avengers	Star wars	Thor	Spider-man	Iron Man
Alex	4	2	*4*	5	4

Variants

- Correlation as the similarity measure
  - User-based Collaborative Filtering
    - Calculate user similarities
      - Pearson's correlation
- Nearest Neighbor Collaborative Filtering
  - Define the effective neighborhood
  - Compute the predicted ratings



The correlation of two users ken and lee, they both rated n items K(1..n) L (1..n)

$$r_{ken,lee} = \frac{\frac{\sum_{i=1}^{n}(K_i - \mu_{Ken})(L_i - \mu_{lee})}{n}}{\sigma_{Ken}\sigma_{Lee}}$$

Prediction on Ken's rating for m

$$R_{CF_{Ken,m}} = \mu_{Ken} + \frac{\sum_{J \in Users}(J_m - \mu_j) \cdot r_{ken,J}}{\sum_{J \epsilon Users} |r_{ken,j}|}$$

Other similarity measures

- Item-based Collaborative filtering
  - The same Item rating Matrix
  - Item vectors: the columns
- Item similarity
  - Pearson's Correlation

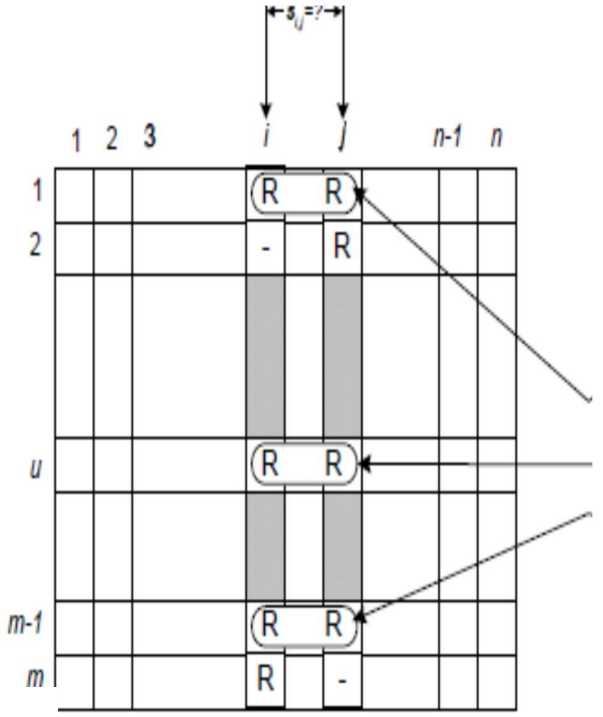
$$sim(i,j) = \frac{\sum_{u \in U}(R_{u,i} - \bar{R}_i)(R_{u,j} - \bar{R}_j)}{\sqrt{\sum_{u \in U}(R_{u,i} - \bar{R}_i)^2}\sqrt{\sum_{u \in U}(R_{u,j} - \bar{R}_j)^2}}$$

  - Cosine similarity

$$sim(i,j) = \cos(\vec{i},\vec{j}) = \frac{\vec{i} \cdot \vec{j}}{\|\vec{i}\|_2 * \|\vec{j}\|_2}$$

  - Adjusted Cosine similarity

$$sim(i,j) = \frac{\sum_{u \in U}(R_{u,i} - \bar{R}_u)(R_{u,j} - \bar{R}_u)}{\sqrt{\sum_{u \in U}(R_{u,i} - \bar{R}_u)^2}\sqrt{\sum_{u \in U}(R_{u,j} - \bar{R}_u)^2}}$$



Discussion

- User-based collaborative filtering
  - Advantages
  - Disadvantages/limitations
- Item-based collaborative filtering
  - Advantages
  - Disadvantages/limitations
- Which one is better? Which one is commonly used and why?
  - Item-based
  - User-based