Recommendation systems

"what else to know about?"

••

USC Viterbi
School of Engineering

Recommendation Systems

Collaborative Filtering &
Content-Based Recommendations

••

Scarcity versus Abundance

- **Since online systems maintain large quantities of goods, systems that provide recommendations serve an important purpose**
  - In some cases items sold from the long tail, (i.e. those not particularly popular) can cumulatively outweigh the initial portion of the graph, an in effect produce the majority of sales
- **Recommendation systems are expected to increase diversity because they help us discover new products.**
  - However, some algorithms may unintentionally do the opposite.
  - Because they recommend products based on past sales or ratings, they cannot usually recommend products with limited historical data. This can create a rich-get-richer effect for popular products
  - This bias toward popularity can prevent what are otherwise better consumer-product matches.
    - Several collaborative filtering algorithms have been developed to promote diversity and the long tail by recommending novel, unexpected, and serendipitous items
  - See https://en.wikipedia.org/wiki/Collaborative_filtering

USC Viterbi School of Engineering

Two Types of Recommendation Systems

- **A *recommendation system* is any system which provides a recommendation/prediction/opinion to a user on items**

1. **A classic *content-based filtering system* uses item similarity/clustering to recommend items like ones you like**
2. **A classic *collaborative filtering system* uses user similarity, i.e. the links between users and the item they chose as the basis of recommendations**

- **Commonly many companies use *hybrid systems* in which both kinds of techniques are employed**

••

USC Viterbi School of Engineering

Difference between Collaborative and Content-based Filtering-An Example

- *Here are two early systems that recommended music*
- *Last.fm* creates a "station" of recommended songs by observing what bands and individual tracks ***the user*** has listened to on a regular basis (*user similarity*) and comparing those against the listening behavior of other users.
  - Last.fm will play tracks that do not appear in the user's library, but are often played by other users with similar interests.
  - As this approach leverages the behavior of users, it is an example of a ***collaborative filtering technique***.
- *Pandora* uses the ***properties of a song*** or artist (a subset of the 400 attributes provided by the Music Genome Project) to seed a "station" that plays music with similar properties (*item similarity*).
  - User feedback is used to refine the station's results, deemphasizing certain attributes when a user "dislikes" a particular song and emphasizing other attributes when a user "likes" a song.
  - This is an example of a ***content-based technique***.

 8

••

USC Viterbi School of Engineering

Input

Alice	Il Fornaio	Yes
Bob	Ming's	No
Cindy	Straits Café	No
Dave	Ming's	Yes
Alice	Straits Café	No
Estie	Zao	Yes
Cindy	Zao	No
Dave	Brahma Bull	No
Dave	Zao	Yes
Estie	Ming's	Yes
Fred	Brahma Bull	No
Alice	Mango Café	No
Fred	Ramona's	No
Dave	Homma's	Yes
Bob	Higashi West	Yes
Estie	Straits Café	Yes

••

USC Viterbi
School of Engineering

Algorithm 0

- Strategy: Recommend to you the most popular restaurants
  - say # positive votes minus # negative votes
- But this ignores
  - your culinary preferences
  - judgments of those with similar preferences
- How can we exploit the wisdom of "like-minded" people?
- Basic assumption
  - Preferences are not random
  - Assumption: if I like Il Fornaio, it's more likely I will also like Cenzo (another Italian restaurant)

••

USC Viterbi
School of Engineering

Cast the Input as a Matrix

	Brahma Bull	Higashi West	Mango	Il Fornaio	Zao	Ming's	Ramona's	Straits	Homma's
Alice		Yes	No	Yes				No	
Bob		Yes				No		No	
Cindy				Yes	No			No	
Dave	No			No	Yes	Yes			Yes
Estie				No	Yes	Yes		Yes	
Fred	No						No		

Called a *utility matrix*

Each row represents an individual and each column a restaurant
In this example, entries are either yes/no;
In the more general case they can be any value

USC Viterbi School of Engineering

The Utility Matrix

- In a recommendation-system application there are two classes of entities, which we shall refer to as *users* and *items*
  - Users have preferences for certain items, and these preferences must be teased out of the data.
- The data itself is represented as a ***utility matrix***, giving for each user-item pair, a value that represents what is known about the degree of preference of that user for that item.
- *Values might come from an ordered set*, e.g., integers 1–5 representing the number of stars that the user gave as a rating for that item.
- We assume that the matrix is *sparse*, meaning that most entries are “unknown.”
- An unknown rating implies that we have no explicit information about the user’s preference for the item.

 13

••

USC Viterbi
School of Engineering

Now That We Have a Matrix

	Brahma Bull	Higashi West	Mango	Il Fornaio	Zao	Ming's	Ramona's	Straits	Homma's
Alice		1	-1	1				-1	
Bob		1				-1		-1	
Cindy				1	-1			-1	
Dave	-1			-1	1	1			1
Estie				-1	1	1		1	
Fred	-1						-1		

View all other entries as zeros for now.

- To compute the similarity between individual's preference vectors we can use inner products as a good place to start, e.g.
    - Dave has similarity 3 with Estie,
        - e.g. (-1,0,0,-1,1,1,0,0,1) and (0,0,0,-1,1,1,0,1,0)
        - (i.e. there are three matching values of either 1 or -1)
    - but -2 with Cindy (0,0,0,1,-1,0,0,-1,0) (a zero value doesn't count).
- Perhaps recommend Straits Cafe to Dave and Il Fornaio to Bob, etc.

••

USC Viterbi
School of Engineering

Another Utility Matrix Example - Movies

	Avatar	LOTR	MATRIX	PIRATES
ALICE	1		0.2	
BOB		0.5		0.3
CAROL	0.2		1	
DAVID				0.4

- The blank spaces indicate either the user has not seen the movie or has decided not to rate it
- Main issue: how to fill in the values in the missing fields
- In general there are two basic techniques for populating the utility matrix
  1. Ask users to rate items
     E.g. movies, online stores from purchasers
  2. Make inferences from user behaviors
     Assumption: Users who watch a movie must "like" it, so above we can assign it a 1;
     We are mostly interested in fields for which the ratings are high

••

Recommending documents can be Viewed as a Form of Recommendation System

- **If the items we are considering are documents, then the profile will be the set of "important" words in the document**
- **How do we pick important words**
  - **We use the TF-IDF formulation seen earlier**

Profile of a document d_j is the vector of weights $w_{i,j}$ $Content(d_j) = (w_{1j}, \ldots w_{kj})$.

$$w_{i,j} = TF_{i,j} \times IDF_i \qquad TF_{i,j} = \frac{f_{i,j}}{\max_z f_{z,j}}, \qquad IDF_i = \log \frac{N}{n_i}.$$

- ***TF*** **: Term Frequency, IDF : Inverse Document Frequency**
- ***N*** **: Number of the documents**
- **n_i : How many times an element is seen in all of the documents**
- **$f_{i,j}$: Number of times an element is seen in the document d_j**

 16

••

Example 1
Boolean Utility Matrix

- **Items are movies, only feature is Actor**
  - **Item profile: vector with *0* or *1* for each actor**
- **Suppose user *X* has watched *5* movies**
  - ***2* movies featuring actor A (movies 1 and 3)**
  - ***3* movies featuring actor B (movies 2, 4, and 5)**
- **User profile = mean of item profiles**
  - **Feature A's weight = *2/5* = *0.4***
  - **Feature B's weight = *3/5* = *0.6***

	ActA	ActB	ActC	ActD	ActE
Movie1	1,	0,	0,	0,	0
Movie2	0,	1,	0,	0,	0,
Movie3	1,	0,	0,	0,	0
Movie4	0,	1,	0,	0,	0
Movie5	0,	1,	0,	0,	0

ActA's weight = Sum(ActA)/5

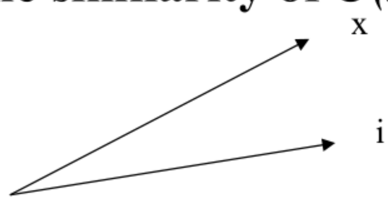
 17

••

USC Viterbi School of Engineering

Making Predictions

- **Given user profile *x* (movies he/she watched) and item profile *i (movies with actor profiles)***
- **Estimate the similarity of $U(x,i) = \cos(\theta) = (x.i)/(|x|\ |i|)$**



- **Technically the cosine distance is actually the angle θ and the cosine similarity is the angle $180 - \theta$**

- **For convenience we use $\cos(\theta)$ as our similarity measure and call it the "cosine similarity" in this context**

••

USC Viterbi School of Engineering

Pros: Content-based Approach

- No need for data on other users
- Able to recommend to users with unique tastes
- Able to recommend new and unpopular items
  - No first rater problem
- Explanations for recommended items
  - Content features that caused an item to be recommended

 20

USC Viterbi School of Engineering

Cons: Content-Based Approach

- Finding the appropriate features is not always obvious
  - E.g. movie features may be obvious but what about images and music
- Overspecialization
  - Never recommends items outside user's content profile
  - People might have multiple interests
  - Unable to exploit quality judgements of other users
- Cold-start problem for new users
  - How to build a user profile

USC Viterbi School of Engineering

Let's Switch Focus to Collaborative Filtering

- Collaborative filtering is a method of making automatic predictions (filtering) about the interests of a user by collecting preferences or taste information from many other users (collaborating).
  - The underlying assumption of the collaborative filtering approach is that if a person *A* has the same opinion as a person *B* on an issue, A is more likely to have B's opinion on a different issue than that of a randomly chosen person

••

Similar Users and Jaccard Similarity

movies

4 users

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

HP:Harry Potter
SW:Star Wars
TW:Twilight

- **Consider users *x* and *y* with rating vectors *rx* and *ry***
  - *The rating vector of user B is (5,5,4,0,0,0,0)*
- **We need a similarity metric *sim(x,y) between rating vectors***
- **The metric should capture the intuition that *sim(A,B) > sim(A,C)***
  - *A and B both liked HP1, but A and C had very different opinions about TW and SW1*
- **Recall *sim(A,B)* = |r_a *intersect* r_b| / |r_a *union* r_b |** ***try Jaccard Similarity***
- ***Sim (A,B) = 1/5; sim(A,C) = 2/4. since A&B rated only one movie in common***
  - *But using Jaccard Similarity we get a result we don't want, namely*
  - *Sim(A,B) < Sim (A,C)*
- **Problem: ignores rating values**

••

USC Viterbi School of Engineering

Cosine Similarity

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

- ***Instead of using Jaccard similarity, which ignored the rating values, let's use cosine similarity, the angle between the vectors; now we treat the unknown values as zero***
  - ***sim (A, B) = cos(r_a, r_b)***
- ***sim(A,B) = 0.38, sim(A,C) =0.32***
  - ***Now sim(A,B) > sim(A,C), which is what we wanted,*** **but not by much**
- **Problem: by treating missing ratings as zero, we sort of got the result we wanted, but actually the similarity of A and C should be farther apart than what we computed; using zero was not a great idea**

 25

••

USC Viterbi
School of Engineering

Centered Cosine Captures User Preferences

Solution: Normalize ratings by subtracting row mean

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

	HP1	HP2	HP3	TW	SW1	SW2	SW3
A	2/3			5/3	−7/3		
B	1/3	1/3	−2/3				
C				−5/3	1/3	4/3	
D		0					0

	HP1	HP2	HP3	TW	SW1	SW2	SW3
A	2/3			5/3	−7/3		
B	1/3	1/3	−2/3				
C				−5/3	1/3	4/3	
D		0					0

- *The average rating for A is 10/3;*
- *The average rating for B is 14/3;*
- *The solution is to subtract the average rating for each user's score; e.g. user A and HP1 we get 4 – (10/3)= 2/3*
- *The resulting matrix*

- *sim(A,B) = cos(r_a,r_b) = 0.09; sim(A,C)=-0.56*

Result: A and C are very DISsimilar

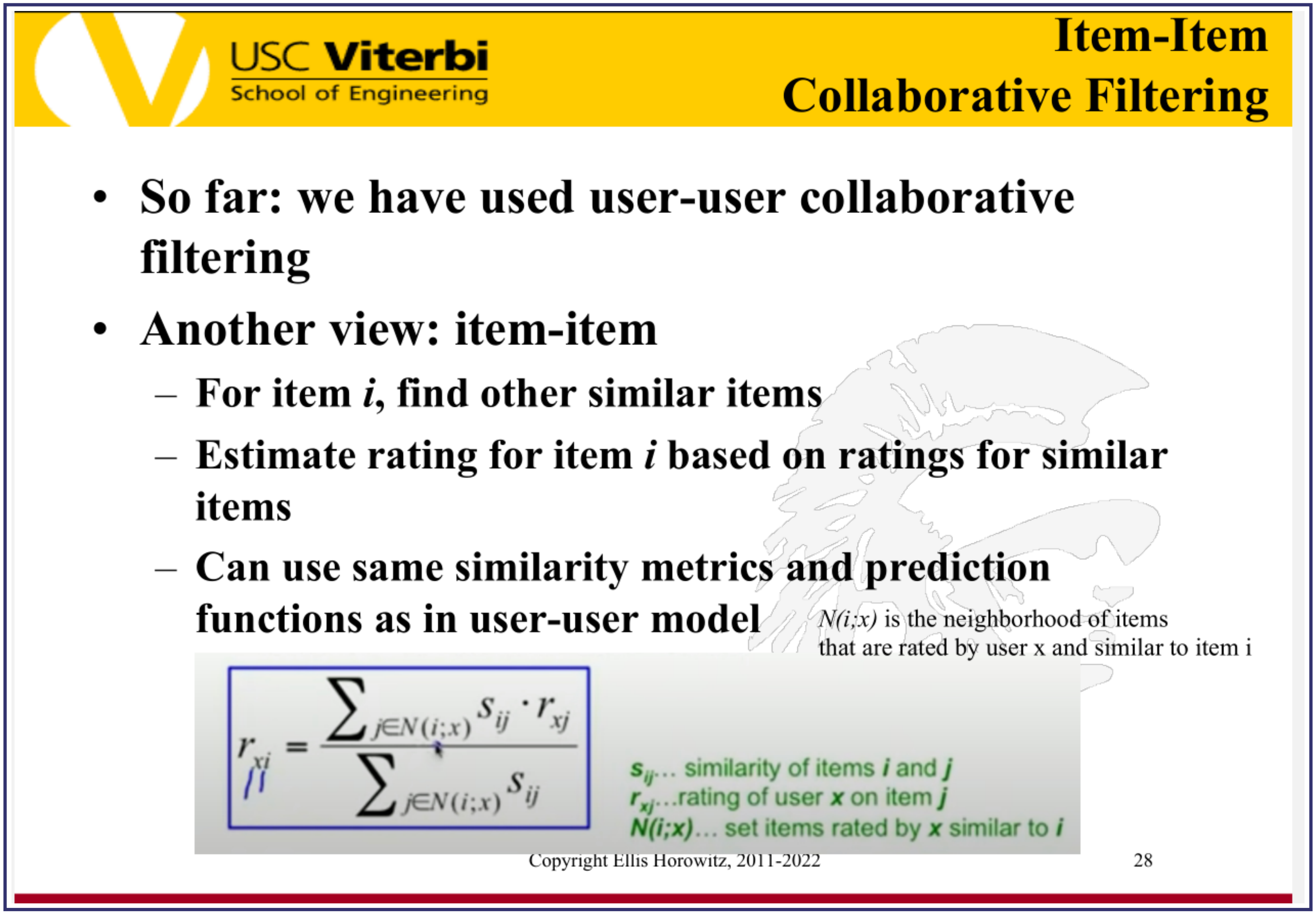
- *sim(A,B) > sim(A,C)*

- Captures intution better
  - Missing ratings treated as average
  - Handles "tough raters" and "easy raters"

Note: Summing the rows for any user gives zero, so positive ratings means they liked the movie
Another name for centered cosine is **Pearson Correlation**

 26

••


USC Viterbi
School of Engineering
Item-Item
Collaborative Filtering
• So far: we have used user-user collaborative filtering
• Another view: item-item
– For item *i*, find other similar items
– Estimate rating for item *i* based on ratings for similar items
– Can use same similarity metrics and prediction functions as in user-user model
N(i;x) is the neighborhood of items that are rated by user x and similar to item i
$r_{xi} = \frac{\sum_{j \in N(i;x)} s_{ij} \cdot r_{xj}}{\sum_{j \in N(i;x)} s_{ij}}$
s_{ij}… similarity of items *i* and *j*
r_{xj}…rating of user *x* on item *j*
N(i;x)… set items rated by *x* similar to *i*
Copyright Ellis Horowitz, 2011-2022
28

••

USC Viterbi School of Engineering

Let’s Do An Example Item – Item CF (|N| =2)

users

movies	1	2	3	4	5	6	7	8	9	10	11	12
1	1		3			5			5		4	
2			5	4			4			2	1	3
3	2	4		1	2		3		4	3	5	
4		2	4		5			4			2	
5			4	3	4	2					2	5
6	1		3		3			2			4	

☐ - unknown rating ■ - rating between 1 to 5

List of similarities of movie 1

users

movies	1	2	3	4	5	6	7	8	9	10	11	12	sim(1,m)
1	1		3		?	5			5		4		1.00
2			5	4			4			2	1	3	-0.18
3	2	4		1	2		3		4	3	5		0.41
4		2	4		5			4			2		-0.10
5			4	3	4	2					2	5	-0.31
6	1		3		3			2			4		0.59

Here we use Pearson correlation as similarity:
1) Subtract mean rating m_i from each movie i
$m_1 = (1+3+5+5+4)/5 = 3.6$
row 1: [-2.6, 0, -0.6, 0, 0, 1.4, 0, 0, 1.4, 0, 0.4, 0]
2) Compute cosine similarities between rows

Goal: Estimate rating of movie 1 by user 5

Movie 3 and 6 are the two recommendations that are most similar to movie 1
Compute similarity weights
$S_{13} = 0.41$ $S_{16} = 0.59$
Use the weighted average to compute
$R_{15} = (0.41*2+0.59*3)/(0.41+0.59)=2.6$

N is 2, looking at the two nearest neighbors

Conclusion: user 5 will like movie 1: 2.6

USC Viterbi School of Engineering

Item-Item vs. User-User

- In theory user-user and item-item are dual approaches
- In practice, item-item outperforms user-user in many use cases
- Items are “simpler” than users
  – Items belong to a small set of “genres”, users have varied tastes
  – *Item* similarity is more meaningful than *user* similarity

 30