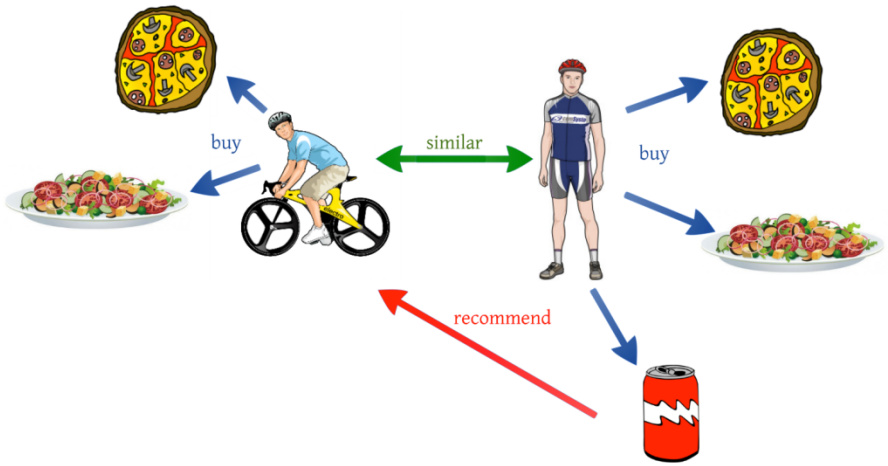
Practical Data Science – Tute 10 / Week 11

PDS Teaching Team

Activity 1 – Collaborative Filtering

- Give a high-level explanation of what collaborative filtering is, and how it works.

Predict the rating on an unknown item based on the ratings on that item by **like minded users**. It is the most successful recommendation technique to date.



http://dataconomy.com/wp-content/uploads/2015/03/Beginners-Guide-Recommender-Systems-Collaborative-Filtering.png

Activity 1 – Collaborative Filtering

- What is the input? Output? Example?

Input:

- Set of users
- Set of items
- Set of opinions, ratings, reviews, or purchases

Output:

- Automatic predictions about the interests of a user

Activity 1 – Collaborative Filtering

- What is the underlying assumption of Collaborative Filtering?

✓ if person ***A*** has the same opinion as person ***B*** on an issue

✓ ***A*** is more likely to also have ***B***'s opinion on a different issue than that of a randomly chosen person.

Activity 1 – Collaborative Filtering

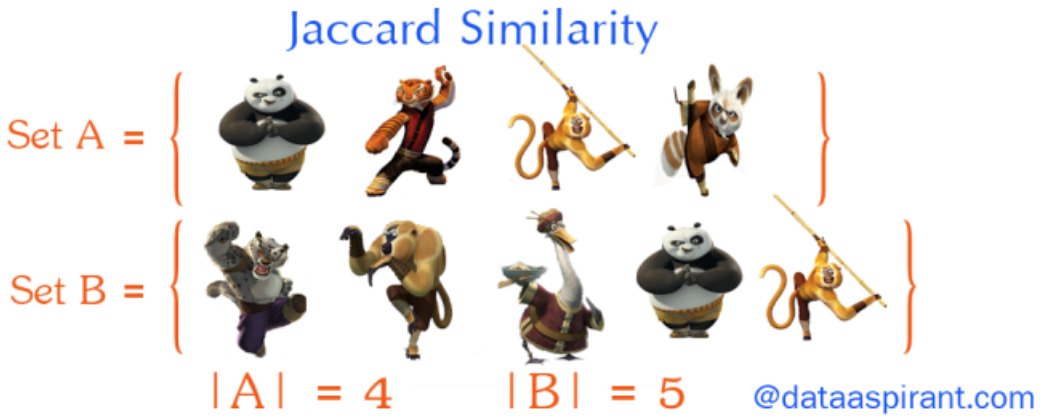
- What are the main steps of collaborative filtering?

1. Measure similarities of the active user to other users, i.e. find like minded users with similar ratings to common items.
2. Make predictions about the active user's rating to an unrated item based on the similarities measured in step 1.

Activity 1 – Similarity Measures

- What is the Jaccard similarity, and why its usefulness is limited in this application?

$$J(A,B) = \frac{INSTERCTION}{UNION}, \quad 0 \leq J(A,B) \leq 1$$



| Intersection (A,B)| = 2

| Union (A,B) | = 7

$$J(A,B) = \frac{2}{7} = 0.286$$

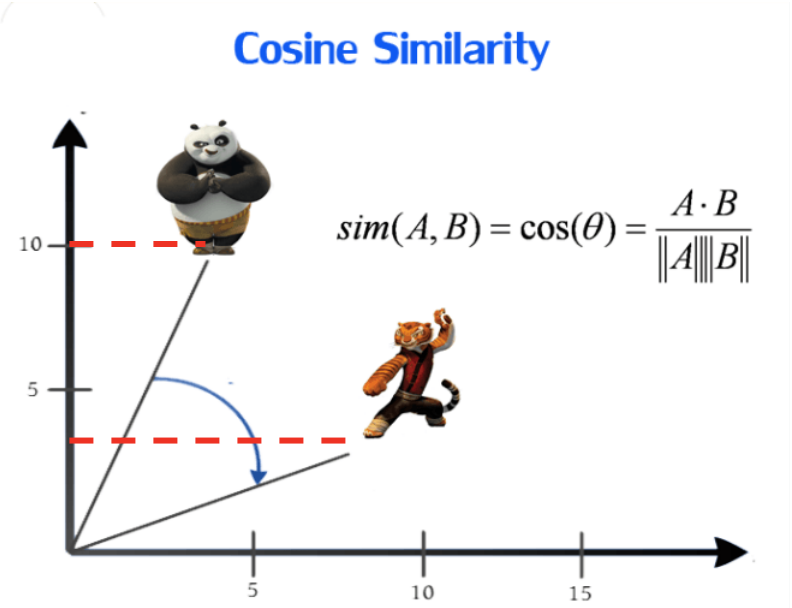
http://dataaspirant.com/2015/04/11/five-most-popular-similarity-measures-implementation-in-python/

Do you see any **ratings** for the characters?!

LIMITATION → Ignores rating values

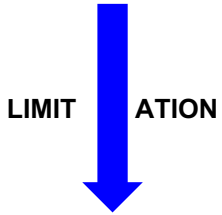
Activity 1 – Similarity Measures

- How does the cosine similarity treat the missing ratings? As positive? As negative?



http://dataaspirant.com/2015/04/11/five-most-popular-similarity-measures-implementation-in-python/

- Assume scoring scheme is 0 (least-liked) – 10 (favourite)
- Cosine similarity treats missing ratings as zeros
- What does zero mean in the context of the scoring scheme?



Treats missing ratings as negative

Activity 1 – Similarity Measures

- What is the Centred Cosine Similarity?

- Normalize ratings by subtracting row mean, i.e. centres ratings around zero
- Missing ratings are still treated as zeros, but what does zero mean in the scoring context?
- Missing ratings are treated as average
- Handles “tough raters” and “easy raters”

Activity 1 – Similarity Measures

- After finding the similar users, there are two options to make the final rating prediction. What are they?

Input:
- Number of users in the neighbourhood ($k = 2$)
- r_1, r_2 → ratings of user 1 and user 2 in the neighbourhood to item i
- s_1, s_2 → similarity values of user 1 and user 2 in the neighbourhood

Output:
- $r_{predicted}$ → predicted rating of the active user for item i

❑ Option 1

Average rating in the neighbourhood

$$r_{predicted} = \frac{r_1 + r_2}{2}$$

❑ Option 2

Weighted average rating in the neighbourhood

$$r_{predicted} = \frac{s_1 * r_1 + s_2 * r_2}{s_1 + s_2}$$

Activity 2 – User-User vs. Item-Item

- What are the key differences between user-user collaborative filtering and item-item collaborative filtering?

User-User

- for user x, find other similar users
- estimate rating of user x to item i based on ratings of similar users to item i.

Item-Item

- for item i, find other similar items
- estimate rating for item i based on ratings for similar items by user x
- can use same similarity metrics and prediction functions as in user-user model

Activity 2 – User-User vs. Item-Item

- In practice, does the item-item method perform similarly to user-user method? If not, why?

- NO, item-item outperforms user-user in many user cases
- Because 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

Questions?