PURDUE UNIVERSITY | Department of Computer Science

CS47300: Web Information Search and Management

Collaborative Filtering

Prof. Chris Clifton

12 October 2020

Material adapted from course created by Dr. Luo Si, now leading Alibaba research group

PURDUE UNIVERSITY
Department of Computer Science

Collaborative Filtering

Outline

- Introduction to collaborative filtering
- Main framework
- Memory-based collaborative filtering approach
- Model-based collaborative filtering approach
  - Aspect model & Two-way clustering model
  - Flexible mixture model
  - Decouple model
- Unified filtering by combining content and collaborative filtering

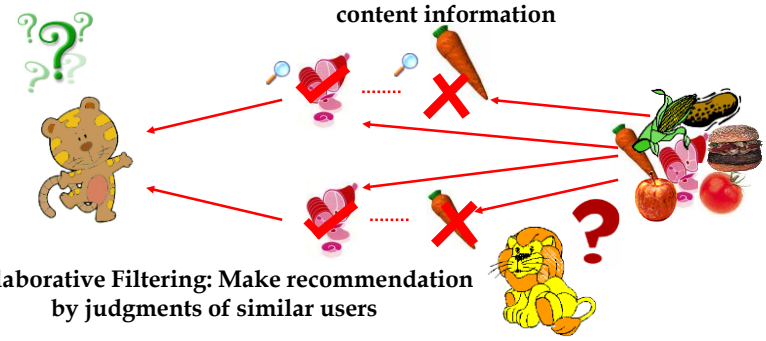
PURDUE UNIVERSITY
Department of Computer Science

What is Collaborative Filtering?

Collaborative Filtering (CF):

Making recommendation decisions for a specific user based on the judgments of users with similar tastes

Content-Based Filtering: Recommend by analyzing the content information

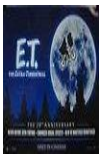


Collaborative Filtering: Make recommendation by judgments of similar users

PURDUE UNIVERSITY
Department of Computer Science

What is Collaborative Filtering?

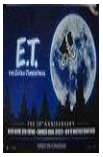
Collaborative Filtering (CF): Making recommendation decisions for a specific user based on the judgments of users with similar tastes

					
Train_User 1	1	5	3	3	4
Train_User 2	4	1	5	3	2
Test User	1	?	3		4

PURDUE UNIVERSITY
Department of Computer Science

What is Collaborative Filtering?

Collaborative Filtering (CF): **Making recommendation decisions for a specific user based on the judgments of users with similar tastes**

					
Train_User 1	1	5	3	3	4
Train_User 2	4	1	5	3	2
Test User	1	5	3		4

PURDUE UNIVERSITY
Department of Computer Science

Why Collaborative Filtering?

- Advantages of Collaborative Filtering
  - Collaborative Filtering does not need content information as required by CBF
  - The contents of items belong to the third-party (not accessible or available)
  - The contents of items are difficult to index or analyze (e.g., multimedia information)
- Problems of Collaborative Filtering
  - Privacy issues, how to share one's interest without disclosing too much detailed information?

PURDUE UNIVERSITY
Department of Computer Science

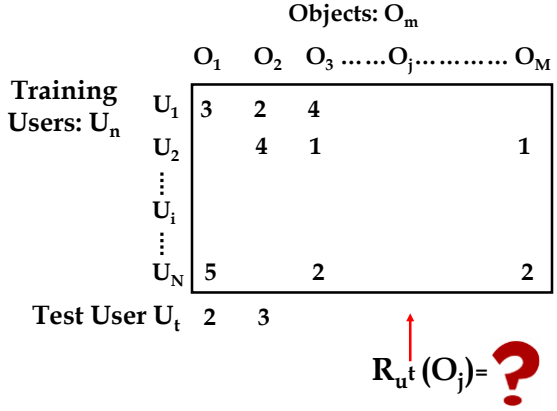
Why Collaborative Filtering?

- Applications Collaborative Filtering
  - E-Commerce

  - Email ranking: borrow email ranking from your office mates (be careful…)
  - Web search? (e.g., local search)

PURDUE UNIVERSITY
Department of Computer Science

Formal Framework for Collaborative Filtering



What we have:

- **Assume there are some ratings by training users**
- **Test user provides some amount of additional training data**

What we do:

- **Predict test user's rating based training information**

PURDUE UNIVERSITY
Department of Computer Science

Memory-Based Approaches

- Memory-Based Approaches
  - Given a specific user *u*, find a set of similar users
  - Predict *u*'s rating based on ratings of similar users
- Issues
  - How to determine the similarity between users?
  - How to combine the ratings from similar users to make the predictions (how to weight different users)?

PURDUE UNIVERSITY
Department of Computer Science

Memory-Based Approaches

- How to determine the similarity between users?
  - Measure the similarity in rating patterns between different users

Pearson Correlation Coefficient Similarity

$$w_{u,u'} = \frac{\sum (R_{u'}(o) - \bar{R}_{u'})(R_u(o) - \bar{R}_u)}{\sqrt{\sum (R_{u'}(o) - \bar{R}_{u'})^2} \sqrt{\sum (R_u(o) - \bar{R}_u)^2}}$$

Vector Space Similarity

$$w_{u,u'} = \frac{\sum R_{u'}(o) R_u(o)}{\sqrt{\sum R_{u'}(o)^2} \sqrt{\sum R_u(o)^2}}$$

Average Ratings

Prediction:

$$\hat{R}_{u'}(o) = \bar{R}_{u'} + \frac{\sum_u w_{u,u'} (R_u(o) - \bar{R}_u)}{\sum_u |w_{u,u'}|}$$

PURDUE UNIVERSITY
Department of Computer Science

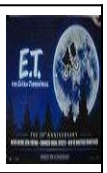
Memory-Based Approaches

- How to combine the ratings from similar users for predicting?
  - Weight similar users by their similarity with a specific user; use these weights to combine their ratings.

Prediction: $\hat{R}_{u'}(o) = \bar{R}_{u'} + \frac{\sum_u w_{u,u'}(R_u(o) - \bar{R}_u)}{\sum_u |w_{u,u'}|}$

PURDUE UNIVERSITY
Department of Computer Science

Memory-Based Approaches

					
Train_User 1	1	5	3	3	4
Train_User 2	4	1	5	3	2
Test User	1	?	3		4

Remove User-specific Rating Bias

PURDUE UNIVERSITY
Department of Computer Science

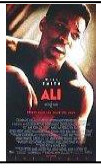
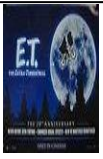
Memory-Based Approaches

					
Train_User 1	1	5	3	3	4
Sub Mean (Train1)	-2.2	1.8	-0.2	-0.2	0.8
Train_User 2	4	1	5	3	2
Sub Mean (Train2)	1	-2	2	0	-1
Test User	1	?	3		4
Sub Mean (Test)	-1.667		0.333		1.33

Normalize Rating

PURDUE UNIVERSITY
Department of Computer Science

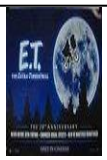
Memory-Based Approaches

					
Train_User 1	1	5	3	3	4
Sub Mean (Train1)	-2.2	1.8	-0.2	-0.2	0.8
Train_User 2	4	1	5	3	2
Sub Mean (Train2)	1	-2	2	0	-1
Test User	1	?	3		4
Sub Mean (Test)	-1.667		0.333		1.33

Calculate Similarity: **Wtrn1_test=0.92; Wtrn2_test=-0.44;**

PURDUE UNIVERSITY
Department of Computer Science

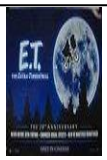
Memory-Based Approaches

					
Train_User 1	1	5	3	3	4
Sub Mean (Train1)	-2.2	1.8	-0.2	-0.2	0.8
Train_User 2	4	1	5	3	2
Sub Mean (Train2)	1	-2	2	0	-1
Test User	1	?	3		4
Sub Mean (Test)	-1.667		0.333		1.33

Make Prediction: 2.67+(1.8*0.92+(-2)*(-0.44))/(0.92+0.44)=4.54

PURDUE UNIVERSITY
Department of Computer Science

Memory-Based Approaches

					
Train_User 1	1	5	3	3	4
Sub Mean (Train1)	-2.2	1.8	-0.2	-0.2	0.8
Train_User 2	4	1	5	3	2
Sub Mean (Train2)	1	-2	2	0	-1
Test User	1	5	3		4
Sub Mean (Test)	-1.667		0.333		1.33

Make Prediction: 2.67+(1.8*0.92+(-2)*(-0.44))/(0.92+0.44)=4.54

PURDUE UNIVERSITY
Department of Computer Science

Memory-Based Approaches

- Problems with memory-based approaches
  - Associated a large amount of computation online costs (have to go over all users, any fast indexing approach?)
  - Heuristic method to calculate user similarity and make user rating prediction
- Possible Solution
  - Cluster users/items in offline manner, save for online computation cost
  - Proposal more solid probabilistic modeling method