

Lecture 16: Recommender Systems

Announcements



Show of Hands

What percentage of watch time on YouTube do you think comes from recommendations?

- a. 30%
- b. 40%
- c. 60%
- d. 70%
- e. 90%

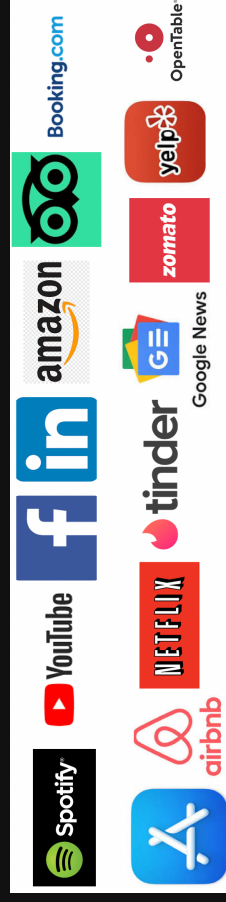
Based on YouTube Product Chief at CES 2024. The number may have changed since then!

What is a recommendation system?

A recommendation system suggests products, services, or content that a user is likely to consume or enjoy.

Example: Recommender systems

Why should we care?



- Recommendations shape almost everything we buy or watch.
- Central to the success of companies like Amazon, Netflix, YouTube, and Spotify.

Why recommendation systems?

- They help users navigate information overload.
- Without them, finding the right item would require sifting through thousands of options.
- Recommendations reduce effort and improve user experience.



The other side: Filter bubbles

- Recommenders often amplify what users already like or what similar users like.
- This can create filter bubbles, limiting exposure to diverse content.
- Probably harmless in shopping, but can have **serious consequences in domains like news, politics, or science, where diverse viewpoints matter.**

Data and problem setup

What data do we need?

To build a recommender, we typically use:

- User-item interactions (ratings, clicks, views, purchases)
- Item or user features (e.g., genre, price, age)
- Historical data (purchase or viewing history)

Problem formulation

- We have N users and M items.
- Observed data = interactions:
 - Movie ratings (Netflix)
 - Song plays (Spotify)
 - Product purchases (Amazon)

Goal: predict unobserved interactions.

The utility matrix

Sparsity

- The utility matrix is mostly empty.
- Each user interacts with only a small fraction of all items.
- Examples:
 - Netflix users rate only a few shows out of thousands.
 - Amazon shoppers review only a handful of products.

What do we predict?

We aim to fill in the missing entries. In other words, predict ratings or preferences the user hasn't expressed yet.

Rating prediction $\neq$ regression

Regression

Rating prediction

✓	✓	✓	✓	✓	✓	✓
✓	✓	✓	✓	✓	✓	✓
✓	✓	✓	✓	✓	✓	✓
✓	✓	✓	✓	✓	✓	?
✓	✓	✓	✓	✓	✓	✓
✓	✓	✓	✓	✓	✓	✓
✓	✓	✓	✓	✓	✓	✓

?	?	✓	?	✓	?	✓
✓	?	?	?	?	?	?
?	?	✓	✓	✓	?	✓
?	?	?	?	?	?	?
?	?	?	?	?	✓	?
?	✓	✓	✓	✓	✓	✓
?	?	✓	✓	✓	✓	✓

Main approaches

- Collaborative filtering
 - “Unsupervised” learning
 - We only have labels y_{ij} (rating of user i for item j).
 - We learn latent features.
- **Content-based recommenders** (today’s focus)
 - Supervised learning
 - Extract features x_i of users and/or items building a model to predict rating y_i given x_i .
 - Apply model to predict for new users/items.
- Hybrid
 - Combining collaborative filtering with content-based filtering

Evaluating Recommender Systems

How do we evaluate recommendations?

- Is there a **single correct answer** for what should be recommended or what rating should be predicted?
 - **Not really!**
- Still, we need ways to **compare different methods** and measure their usefulness.

Why evaluation matters?

- We'll experiment with different ways to **fill in missing entries** in the utility matrix.
- Even though recommendations are subjective, we still need **quantitative metrics** to judge:
 - How well do our predictions match real user behavior?
 - Which model performs better?

RMSE for rating prediction

- RMSE, which measures how close predicted ratings are to actual ratings, is one of the commonly used metric to evaluate recommendation systems
- In **2006**, Netflix launched the **Netflix Prize** competition.
- They released a dataset of **100 million movie ratings** and offered **\$1 million** to the first team that improved Netflix's existing algorithm by **at least 10%** in **RMSE** on a held-out test set.



COMPLETED

Netflix Prize

Home Rules Leaderboard Update Download

Leaderboard

Showing Test Score. [Click here to show guide score](#)

Display top 20 leaders.

Rank	Team Name	Best Test Score	% Improvement	Best Submit Time
Grand Prize - RMSE = 0.8567 - Winning Team: BellKor's Pragmatic Chaos				
1	BellKor's Pragmatic Chaos	0.8567	10.06	2008-07-26 18:18:28
2	The Ensemble	0.8567	10.06	2008-07-26 18:38:22
3	Grand Prize Team	0.8568	9.90	2008-07-10 21:24:40
4	Opera Solutions and Vandelay United	0.8588	9.84	2008-07-10 01:12:31
5	Vandelay Industries I	0.8591	9.81	2008-07-10 00:32:20
6	PragmaticTheory	0.8594	9.77	2008-06-24 12:06:56
7	BellKor in BioChaos	0.8601	9.70	2008-06-13 08:14:09
8	Dice	0.8612	9.59	2008-07-24 17:18:43

Source: Netflix Tech Blog

Clicker 16.1

Select all of the following statements which are **True**

- a. In the context of recommendation systems, the shapes of validation utility matrix and train utility matrix are the same.
- b. RMSE perfectly captures what we want to measure in the context of recommendation systems.
- c. It would be reasonable to impute missing values in the utility matrix by taking the average of the ratings given to an item by similar users.
- d. In KNN type imputation, if a user has not rated any items yet, a reasonable strategy would be recommending them the most popular item.

Clicker 16.2

Select all of the following statements which are **True**

- a. In content-based filtering we leverage available item features in addition to similarity between users.
- b. In content-based filtering you represent each user in terms of known features of items.
- c. In the set up of content-based filtering we discussed, if you have a new movie, you would have problems predicting ratings for that movie.
- d. In content-based filtering if a user has a number of ratings in the training utility matrix but does not have any ratings in the validation utility matrix then we won't be able to calculate RMSE for the validation utility matrix.

Baseline approaches

- Global average baseline
- Per-user average baseline
- Per-item average baseline
- Average of 2 and 3
 - Take an average of per-user and per-item averages.
- *k*-Nearest Neighbours imputation

Content-based filtering

							
Eva	?	?	?	5.0	?	5.0	1.0
Jim	?	?	4.0	5.0	?	5.0	?
Pat	5.0	5.0	?	?	4.0	?	?
Sam	5.0	5.0	?	?	5.0	1.0	4.0

Eva's profile

Animation	Drama	Documentary	Rating
0	0	1	5.0
0	0	1	5.0
1	1	0	1.0

Pat's profile

Animation	Drama	Documentary	Rating
1	1	0	5.0
1	1	0	5.0
0	0	1	1.0
...	...	...	...

iClicker

Select all of the following statements which are True

- a. In content-based filtering we leverage available item features in addition to similarity between users.
- b. In content-based filtering you represent each user in terms of **known** features of items.
- c. In the set up of content-based filtering we discussed, if you have a new movie, you would have problems predicting ratings for that movie.
- d. In content-based filtering if a user has a number of ratings in the training utility matrix but does not have any ratings in the validation utility matrix then we won't be able to calculate RMSE for the validation utility matrix.

Ethics of Recommender Systems

We should be very mindful of the drawbacks of relying on recommender systems.

What are some ethical considerations of recommender systems?

What comprises a recommender system?

What does the problem formulation look like?

What tools would/should we use to create a recommender system?

Group Work: Class Demo & Live Coding

For this demo, each student should [click this link](#) to follow along with the demo from today.

All credit to Dr. Varada Kolhatkar for putting this together!