

CPSC 481

Artificial Intelligence

Dr. Mira Kim
Mira.kim@fullerton.edu

What we will cover this week

- Recommender systems

Introduction

- Recommender System
 - To infer customer interests by utilizing the various sources of data
 - User Feedback such as Rating
 - Act of a user buying or browsing an item
 - To analyze the previous interaction between users and items
- Key Principles of Recommendation Algorithms
 - Significant dependencies/correlations exist between user- and item-centric activity.

Example) Movie Ratings

- Users rate movies using 0-5 stars

Movie	Jenny (1)	Bob (2)	Grace (3)	Joseph (4)	
Serendipity	5	5	0	0	
The notebook	5	?	?	0	
Love actually	?	4	0	?	
Gladiator	0	0	5	4	
300	0	0	5	?	

Example) Movie Ratings

- Users rate movies using 0-5 stars

Movie	Jenny (1)	Bob (2)	Grace (3)	Joseph (4)	
Serendipity	5	5	0	0	
The notebook	5	?	?	0	
Love actually	?	4	0	?	
Gladiator	0	0	5	4	
300	0	0	5	?	

Example) Movie Ratings

- Users rate movies using 0-5 stars

Movie	Jenny (1)	Bob (2)	Grace (3)	Joseph (4)	
Serendipity	5	5	0	0	
The notebook	5	? 4?	? 0?	0	
Love actually	?	4	0	? 0?	
Gladiator	0	0	5	4	
300	0	0	5	? 4?	

Primary Model of Recommender

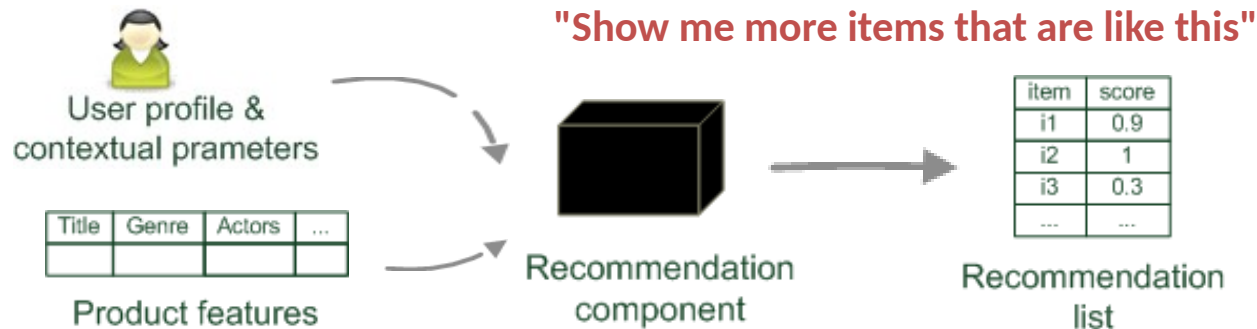
- Prediction Version of Problem
 - Matrix Completion Problem
 - To predict the rating value for a user-item combination
 - For m users and n items $\hookrightarrow (m \times n)$ Matrix
 - Specified (or observed) values are used for training.
 - Missing (or unobserved) values predicted using this training model.
- Ranking Version of Problem
 - Top-k Recommendation Problem
 - To recommend top-k items for a particular user, or determine top-k users to target for a particular item
 - Determination of top-k items more common

Goal of Recommender Systems

- Relevance
 - To recommend items that are relevant to the user
- Novelty
 - To recommend items that the user has not seen in the past
- Serendipity
 - To recommend items that are somewhat unexpected and are surprising to the user
- Increasing Recommendation Diversity

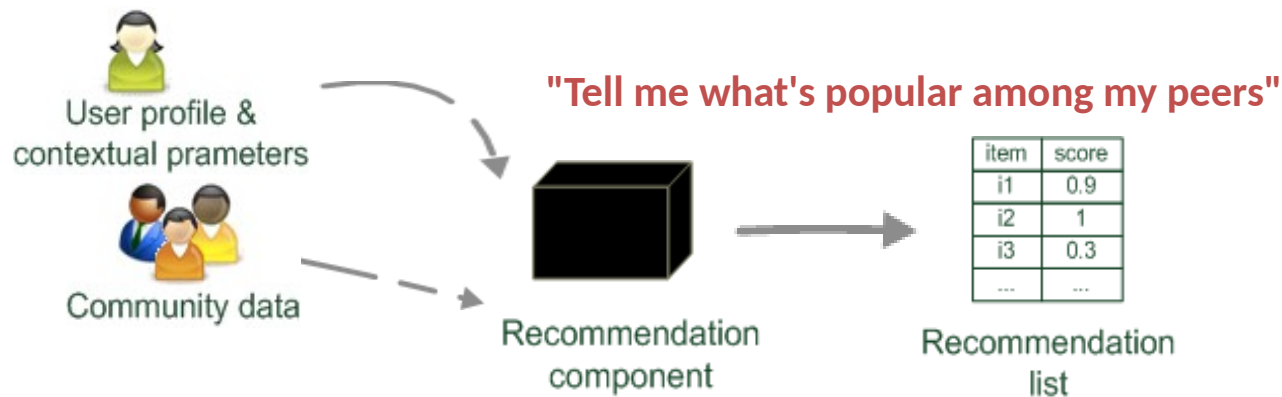
Types of Recommender Systems

- Simple Recommenders
 - To offer generalized recommendations to every user, based on popularity and/or genre
 - No personalization
- Content-based Recommenders
 - To suggest similar items based on a particular item
 - To use item metadata
 - Ex) genre, director, description, actors, etc. for movies
 - If a person liked a particular item, he or she will also like an item that is similar to it.



Types of Recommender Systems

- Collaborative Filtering Recommenders
 - To predict the rating or preference that a user would give an item-based on **past ratings and preferences of other users**



- Hybrid Recommender Systems
 - Combinations of various inputs and/or composition of different types of recommender systems

Simple Recommender

- Basic systems that recommends the top items based on a certain metric or score
 - Ex) based on popularity
- Steps
 - 1. Decide what metric or score to use for rating items.
 - 2. Calculate the score for every item.
 - 3. Sort the items based on the score and output the top results.

Content-based Recommenders

- To use only information about the description and attributes of the items users has previously consumed to model user's preferences
 - Based on a description of the item and a profile of the user's preferences
- Various candidate items are compared with items previously rated by the user and the best-matching items are recommended.
- Pros
 - Not depend on other users
 - Possible to recommend new items that may not be well-known
 - Easy to explain recommendations
- Cons
 - Not possible to recommend to new users
 - Over-specialization
 - Limited content analysis

Content-based Recommenders

- Steps
 - 1a. Analyze contents to give a classification of items
 - 1b. Learn a user profile that represents each user's preference
 - 2. Generate a list of recommendations for each user by taking item classification & user preference
- Applicable ML Techniques
 - Classifications like nearest neighbor
 - Find the nearest neighbors of a not-yet-seen item in a set of already rated by the user
 - Linear Classification
 - User's past ratings as labels and item features as input

Content-based Recommenders

- Content Representation & Item Similarities
 - Structured Representation

Title	Genre	Author	Type	Price	Keywords
The Night of the Gun	Memoir	David Carr	Paperback	29.90	Press and journalism, drug addiction, personal memoirs, New York
The Lace Reader	Fiction, Mystery	Brunonia Barry	Hardcover	49.90	American contemporary fiction, detective, historical
Into the Fire	Romance, Suspense	Suzanne Brockmann	Hardcover	45.90	American fiction, murder, neo-Nazism

- Unstructured Representation (like keyword)
 - By using 'Bag of Words' – counting the frequency of each word
 - 'Term Frequency/Inverse Document Frequency (TF/IDF)
 - how important is the word in the document (local importance), with how important is the word in the corpus (global importance)
- Measure Similarity
 - Euclidian distance
 - Others

Collaborative Filtering Recommenders

- To make automatic predictions (filtering) about the interests of a user by collecting preferences or taste information from many users (collaborating)
 - If a person A has the same opinion as a person B on a set of items, A is more likely to have B's opinion for a given item than that of a randomly chosen person.
- Pros
 - Works with any kind of item
 - Not depend on attributes of items
- Cons
 - Cannot recommend items that are not already rated (i.e. new-item problem)
 - Usually recommend more popular items
 - Need a minimum number of users to match similar users (cold start problem for new users)

Collaborative Filtering Recommenders

- Input
 - A matrix of given user-item ratings
- Output types
 - A (numerical) prediction indicating to what degree the current user will like or dislike a certain item
 - A top-N list of recommended items

Collaborative Filtering Recommenders

- 2 Types of Collaborative Filtering
 - User-based filtering
 - Finds similar users based on their behavior and preferences
 - “You may like it because your friends liked it”
 - Item-based filtering
 - Identifies similarities between items based on how users interact with them
 - Many users who like item X also like item Y, system recommends item Y to users who liked item X

Collaborative Filtering Recommenders

- User-based Collaborative Filtering
 - With ratings of users, determine whether Alice will like or dislike Item 5.

	Item1	Item2	Item3	Item4	Item5
Alice	5	3	4	4	?
User1	3	1	2	3	3
User2	4	3	4	3	5
User3	3	3	1	5	4
User4	1	5	5	2	1

- **Pros:**
 - Can provide highly personalized recommendations since it relies on the preference of similar users
 - Recommendations can include a diverse set of items across different categories
- **Cons:**
 - Can suffer from the cold start problem for new users
 - Scalability as the computation grows with the number of users
 - Sparsity of the user-item matrix -> less reliable similarity measure

Collaborative Filtering Recommenders

- Item-based Collaborative Filtering
 - Use the similarity between items (and not users) to make predictions
 - Look for items that are similar to Item5
 - Take Alice's ratings for these items to predict the rating for Item5

	Item1	Item2	Item3	Item4	Item5
Alice	5	3	4	4	?
User1	3	1	2	3	3
User2	4	3	4	3	5
User3	3	3	1	5	4
User4	1	5	5	2	1

- **Pros:**
 - More stable over time as items don't change their own characteristics like users
 - Scalable to large user bases
 - Can handle the new user problem better since it relies on item similarities.
- **Cons:**
 - May not capture the user's current interests
 - Can struggle with new items that have few ratings (cold start problem for items).

Example) Amazon

Frequently Bought Together



+



Price for both: **\$158.15**

Add both to Cart

Add both to Wish List

One of these items ships sooner than the other. [Show details](#)

☒ **This item:** Introduction to Data Mining by Pang-Ning Tan Hardcover **\$120.16**

☒ Data Science for Business: What you need to know about data mining and data-analytic thinking by Foster Provost Paperback **\$37.99**

Customers Who Bought This Item Also Bought

Page 1 of 15



Data Science for Business:
What you need...

> Foster Provost

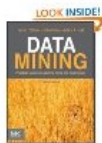
★★★★☆ 102

#1 Best Seller in Data

Mining

Paperback

\$37.99 ✓Prime



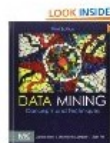
Data Mining: Practical
Machine Learning Tools...

> Ian H. Witten

★★★★☆ 52

Paperback

\$40.65 ✓Prime



Data Mining: Concepts and
Techniques, Third...

> Jiawei Han

★★★★☆ 28

Hardcover

\$60.22 ✓Prime



Regression Analysis by
Example

> Samprit Chatterjee

★★★★☆ 9

Hardcover

\$92.39 ✓Prime

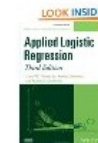


SAS Statistics by Example
Ron Cody

★★★★☆ 10

Perfect Paperback

\$44.37 ✓Prime



Applied Logistic
Regression

David W. Hosmer Jr.

★★★★☆ 9

Hardcover

\$62.33 ✓Prime



An Introduction to
Statistical Learning:...

> Gareth James

★★★★☆ 56

#1 Best Seller in

Mathematical & Statistical...

Hardcover

\$72.79 ✓Prime



Example) Amazon



Introduction to Data Mining

\$120.16 FREE Shipping. | Temporarily out of stock. Order now and we'll deliver when available. We'll e-mail you w

What Other Items Do Customers Buy After Viewing This Item?



Data Science for Business: What you need to know about data mining and data-analytic thinking Paperback

> Foster Provost

★★★★☆ 102

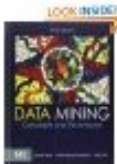
\$37.99 ✓Prime



Introduction to Data Mining Paperback

Pang-ning Tan

★★★★☆ 4



Data Mining: Concepts and Techniques, Third Edition (The Morgan Kaufmann Series in Data Management Sy

> Jiawei Han

★★★★☆ 28

\$60.22 ✓Prime



Data Mining: Practical Machine Learning Tools and Techniques, Third Edition (The Morgan Kaufmann Series i

> Ian H. Witten

★★★★☆ 52

\$40.65 ✓Prime



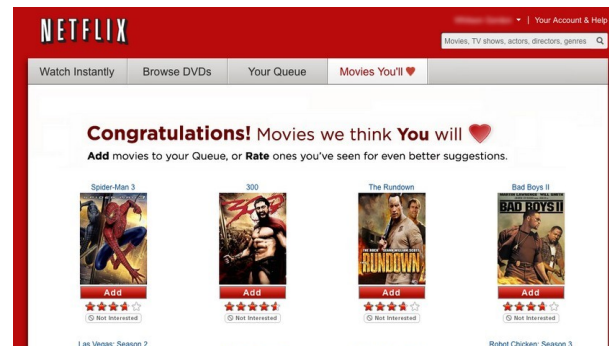
Example) Pandora Music Recommender

- Goal
 - To recommend music
 - Recommendation based on data from Music Genome Project
 - Assigns 400 attributes (melody, rhythm, etc) for each song, done by musicians



Example) Netflix Movie Recommender

- Goal
 - To recommend movies by comparing the watching and the searching habits of similar users as well as by offering movies that share characteristics with films that a user has rated highly
 - Ensemble Method of 107 different Algorithmic Approaches
 - Collaborative, content-based, knowledge-based, and demographic technique as the Basis



References

- Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems 3rd Edition