Data Mining II / Adv. Topics in Data Science

Web Mining: Recommender Systems

Álvaro Figueira
Rita Ribeiro
2023/2024

U.PORTO FACULDADE DE CIÊNCIAS UNIVERSIDADE DO PORTO
[dcc] DEPARTAMENTO DE CIÊNCIA DE COMPUTADORES FACULDADE DE CIÊNCIAS DA UNIVERSIDADE DO PORTO

1

Summary

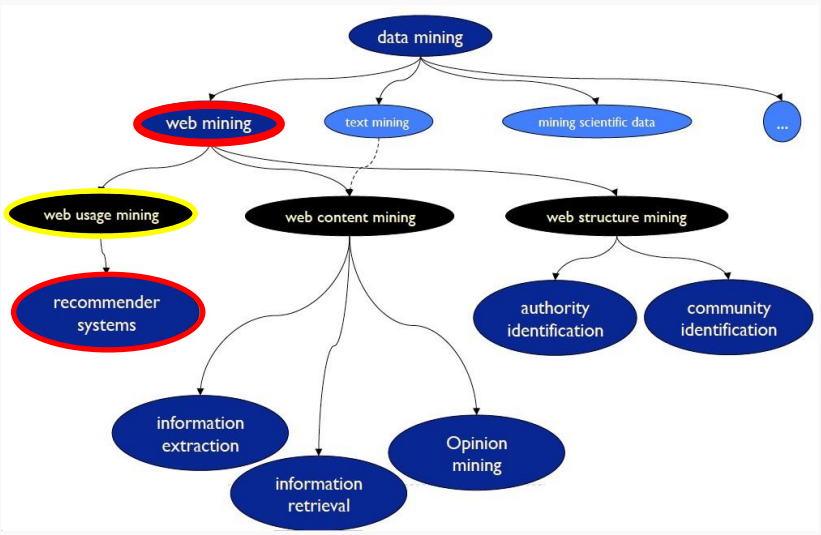
1. Web Mining Introduction
2. Web Usage Mining - Recommender Systems
   - Basic Concepts
   - Association Rules
   - Collaborative Filtering
   - Evaluation
3. Web Structure Mining - Link Analysis
4. Web Content Mining - Information Retrieval

2

Basic Concepts

3

Data Mining - a structured view



4

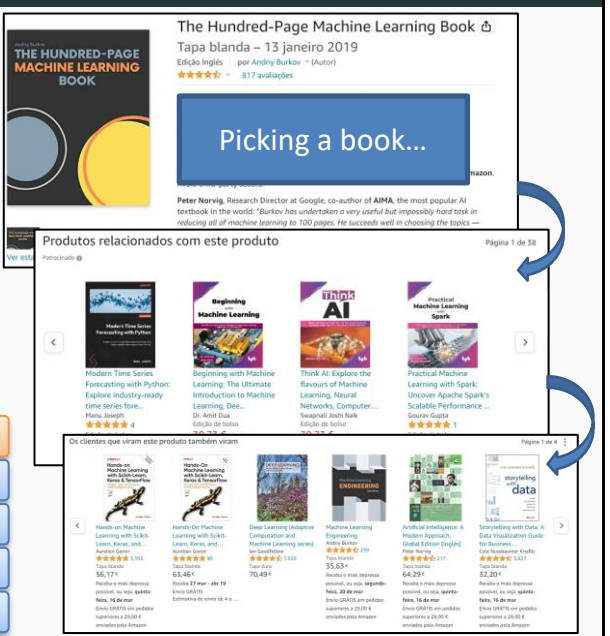
Recommender Systems

- Widely used on the Web for recommending products and services to users.

5

Recommender Systems

- It is widely used on the Web for recommending products and services to users.



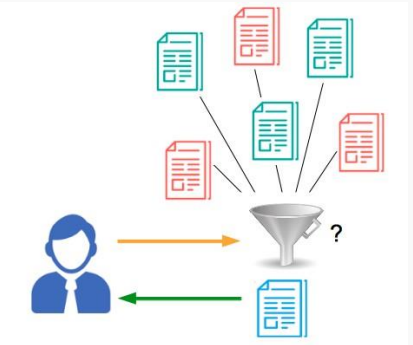
Amazon has

- Very large catalogue
- Has ratings
- A base of buying customers
- Uses "wisdom" of the crowd

6

Recommender Systems (cont.)

- Serve two important functions:
  1. help businesses make more profits;
  2. help users deal with information overload by giving them personalized recommendations.



7

Recommender Systems (cont.)

- Recommender Systems are about:
  - Information filtering
  - Web intelligence
  - Data mining
  - Big Data
  - Sales

8

Recommender Systems (cont.)

- The Recommendation Problem:

- Set of users *U*
- Set of items *S* to be recommended to the users
- Each user in *U* is defined with a **user profile** that includes user characteristics, tastes, preferences, etc.
- The task is to learn the **utility function** that measures the usefulness of item *s* to user *u* and predicts the utility value of each item to each user.

Recommender Systems (cont.)

- Two associated prediction tasks can be performed:

  a) Item Prediction: **predict a ranked list of items** that a user is likely to buy or use.

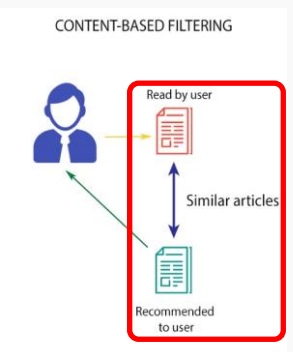
  b) Rating prediction: assuming the user is rating items, predict the **rating score that a user is likely to give** to an item **not seen before**.

Recommender Systems Approaches

- Content-based filtering
- Collaborative filtering
- Hybrid approaches

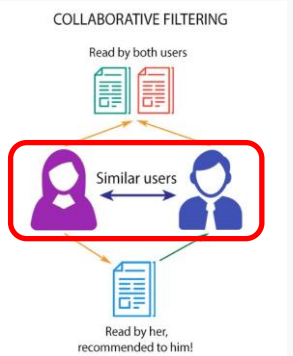
Recommender Systems Approaches (cont.)

- **Content-based filtering**:
  the user will be recommended items similar to the ones he preferred in the past.



Recommender Systems Approaches (cont.)

- **Collaborative filtering:**
  the user will be recommended items that people with similar tastes and preferences liked in the past; we do not need to know the content.



Recommender Systems Approaches – Working example

- Collaborative filtering with binary (or unary) data

USER	ITEM
1	A
1	B
1	G
2	A
2	C
3	B
3	G
3	F
3	I
4	B
4	C
5	G
5	F
5	I
5	J
6	A
6	C

- Collect data is cheaper than hiring people
- Quality might be a problem (*did the user liked it?*)
- There are implicit ratings (*did the user watch the whole movie?)*
- The data is sparse
- Example: **user1** interactions with
  - A
  - B
  - G

A	B	C	D	E	F	G
1	1	0	0	0	0	1

Recommender Systems Approaches – Working example

- Collaborative filtering with ratings-based data (using a five-star system)

USER	ITEM	RATING
1	A	1
1	B	3
1	G	4
2	A	4
2	C	2
3	B	4
3	G	5
3	F	3
3	I	4
4	B	5
4	C	4
5	G	3
5	F	4
5	I	5
5	J	3
6	A	5
6	C	3

- **Not all users rate all items**
- example: **User1** ratings
  - A: ⋆
  - B: ⋆ ⋆ ⋆
  - C: not seen or not rated
  - G: ⋆ ⋆ ⋆ ⋆

User 1:

A	B	C	D	E	F	G
1	3	0	0	0	0	4

User 2:

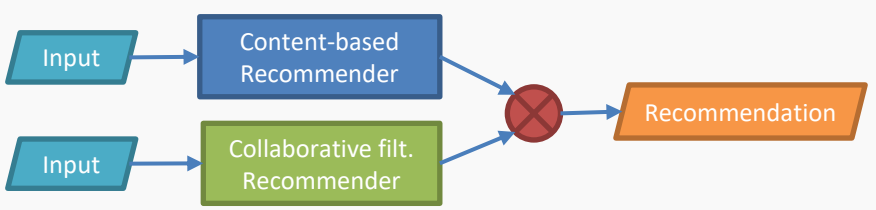
A	B	C	D	E	F	G
4	0	2	0	0	0	0

15

Recommender Systems Approaches (cont.)

- **Hybrid approaches:**

  For the recommendation, content-based and collaborative filtering methods are combined, incorporating characteristics of both: they use background information, social links, context information, etc. together with collaborative filtering.



16

Web Usage Mining: Recommendation's vocabulary

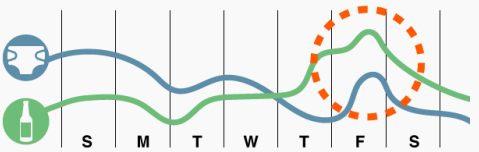
- Stating the problem
  - **given the current activity of a user** (the active user) recommend relevant items
- Recommender algorithm
  - **produces a recommender model** by learning from past actions of the users
- Session
  - **set/sequence of items "transitioned" by a user** in a period of time.
- Recommender model
  - **relates the active session with potentially interesting items**
- Simple recommenders
  - **Top-ranked**, **clustering**

17

Nerd culture - the example of Market Basket Analysis

The origins of the diapers and beers link can be traced back to a story from the 1980s. As the story goes, a Walmart in the United States noticed that **sales of beer and diapers were consistently high on Fridays and Saturdays**.

After analyzing their sales data, they discovered that many fathers, who were responsible for buying diapers, were also picking up a pack of beer while they were at the store.



18