

Machine Learning 0xJ01 Recommendation

Given data about a user, his environment, and some items of interest (training data), determine items to recommend.

We don't have to find the max k .

It's enough to find k within some max n .

Examples:

- Amazon
- Google News (or Le Monde, etc.)
- Facebook
- Twitter
- Medical testing
- App store / play store
- Youtube
- Advertising
- Netflix, last.fm, spotify, pandora, ...
- Browser (URL recommendations)
- Search

Client value proposition:

- Find opportunities
- Reduce choice
- Explore options
- Discover long tails
- Recreation

Provider value proposition:

- Offer unique or additional service (beyond competitors)
- Customer trust, loyalty
- Increase sales, CTR, conversions
- Better understand customers

Roughly:

Content-based filtering (<i>filtrage basée sur le contenu</i>)	More things similar to what I like
Collaborative filtering (<i>filtrage collaboratif</i>)	More of what other people who like what I like like
Knowledge-based filtering (<i>filtrage basée sur connaissance</i>)	More of what I need.

Content-based filtering

More things similar to what I like

Plus de ce qui ressemble à ce que j'aime

Advantages

yes! No need for community

yes! Possible to compare items

Disadvantages

no Understand content

yes Cold start problem

no Serendipity

Collaborative filtering

More of what other people who like what I like like

Plus de ce que d'autres qui aiment ce que j'aime aiment

Advantages

yes! No need to understand content

yes! Serendipity

yes! Learn market

Disadvantages

no User feedback

yes Cold start problem (users)

yes Cold start problem (items)

Knowledge-based filtering

More of what I need

Plus de ce qu'il faut

Advantages

yes! Deterministic

yes! Certainty

no! Cold start problem

yes! Market knowledge

Disadvantages:

yes Studies to bootstrap

yes Static model, doesn't learn from trends

Utility Matrix

- Users (utilisateurs)
- Items (objets)

The goal is to fill in the blanks.

Example: books sales at Amazon.

	I_1	I_2	I_3	I_4	I_5
U_1	1				
U_2			1	1	1
U_3		1		1	1

Example: film advice at Netflix.

	I_1	I_2	I_3	I_4	I_5
U_1	3				
U_2			5	1	4
U_3		2		5	1

But thousands or millions of columns and rows.

How do we make the matrix?

- Ask users
- Observe users

That's usually expensive...

Examples: you are the users, what are the objects

- Films
- Books
- News
- Images

For tomorrow:

- Create a git repo, ideally identifiable as you. Send a PR to add it to GIT.md.
- Add your MNIST code to your repo, in a directory called something recognisable.
- If GDPR concerns, tell me. But your career...