UFMG UNIVERSIDADE FEDERAL DE MINAS GERAIS

Recommender Systems

Non-Personalized Collaborative Filtering

Rodrygo Santos
rodrygo@dcc.ufmg.br

Recommender systems

UFMG COMPUTER SCIENCE

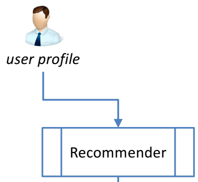
- *"Recommender systems are software applications that aim to support users in their* ***decision-making*** *while interacting with* ***large information spaces****. They recommend* ***items*** *of interest to* ***users*** *based on* ***preferences*** *they have expressed, either explicitly or implicitly."*
  – ACM Conference on Recommender Systems

What to recommend?

UFMG COMPUTER SCIENCE

How to recommend?

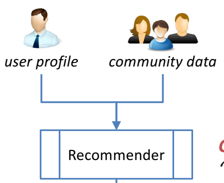
UFMG COMPUTER SCIENCE



item	score
1	0.7
2	0.3
…	…

How to recommend?

UFMG COMPUTER SCIENCE

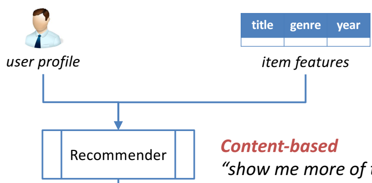


Collaborative filtering
"tell me what's popular among my peers"

item	score
1	0.7
2	0.3
…	…

How to recommend?

UFMG COMPUTER SCIENCE

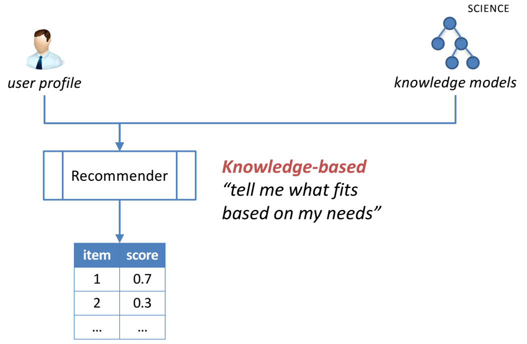


Content-based
"show me more of the same what I've liked"

item	score
1	0.7
2	0.3
…	…

How to recommend?

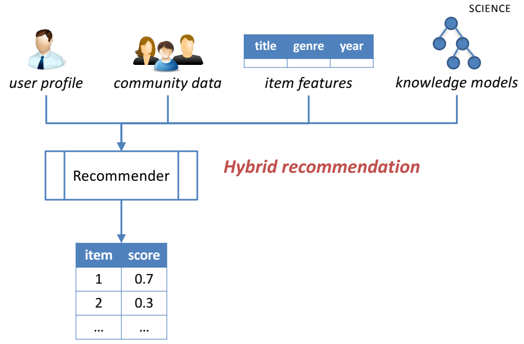
UFMG COMPUTER SCIENCE



Knowledge-based
"tell me what fits based on my needs"

How to recommend?

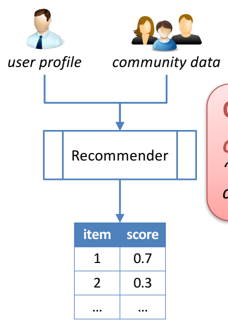
UFMG COMPUTER SCIENCE



Hybrid recommendation

How to recommend?

UFMG COMPUTER SCIENCE



OUR FOCUS
Collaborative filtering
"tell me what's popular among my peers"

UFMG COMPUTER SCIENCE

Recommender Systems

MODELING PREFERENCES

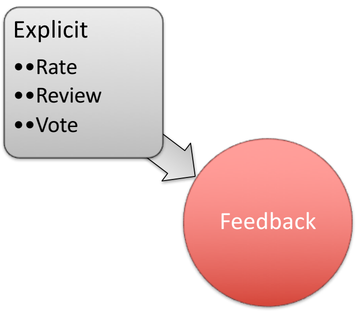
Acquiring feedback

UFMG COMPUTER SCIENCE

- We want to know
  - What users consider relevant
- We can observe
  - What users tell us (ratings)
  - What users do (actions)
- These are *noisy measurements*
  - Relevance is a user's prerogative

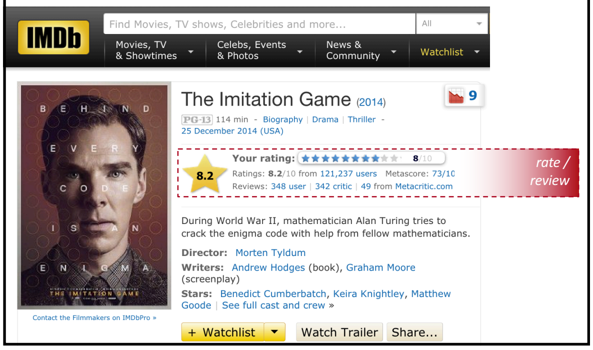
Feedback model

UFMG COMPUTER SCIENCE



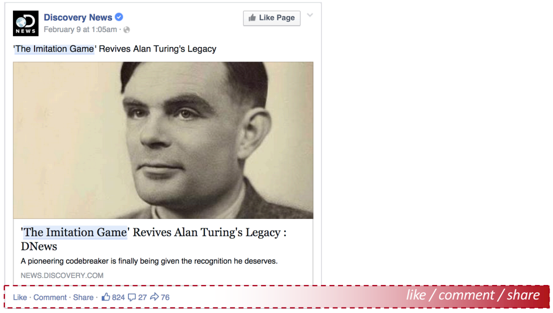
Explicit feedback

UFMG COMPUTER SCIENCE



Explicit feedback

UFMG COMPUTER SCIENCE



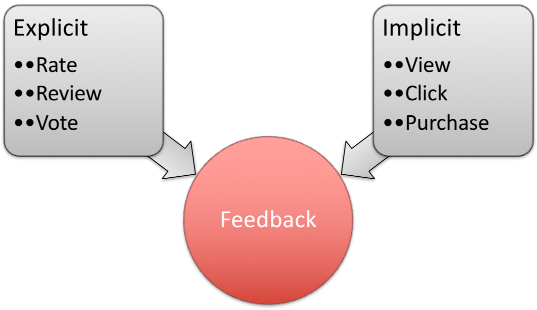
Explicit feedback

UFMG COMPUTER SCIENCE

- Are ratings reliable and accurate?
  - *Are my 8/10 stars equivalent to yours?*
- Do user preferences change?
  - *Will I still like the item after 10 years?*
- What does a rating mean?
  - *Will I ultimately consume the rated item?*

Feedback model

UFMG COMPUTER SCIENCE



Implicit feedback

UFMG COMPUTER SCIENCE

- Abundant data from user actions
  - Views, clicks, reads, buys, etc.
- Actions typically for some other purpose
  - Not direct expressions of preference
- But actions say a lot!
  - Binary signals (click, skip, play, purchase)
  - Attention signals (reading / listening / watching)
  - Cognitive signals (eye-tracking, brain imaging)

Implicit feedback

UFMG COMPUTER SCIENCE

- What does the action mean?
  - Purchase: they might still hate it
  - Don't click: bad, or didn't see?
- How to factor in cognitive biases?
  - Position, presentation, popularity, etc.

How to leverage user feedback to produce recommendations?

UFMG COMPUTER SCIENCE

Recommender Systems

COLLABORATIVE FILTERING

Collaborative filtering (CF)

UFMG COMPUTER SCIENCE

- Most prominent recommendation approach
  - Used by large commercial e-commerce sites
  - Applicable in many domains (books, movies, etc.)
- Key idea
  - Leverage the “wisdom of the crowds”
- Basic assumption
  - Users’ future preferences can be predicted by their past preferences (acquired via feedback)

Stable preferences

UFMG COMPUTER SCIENCE

- Some examples
  - News: I prefer technology, travel
  - Music: I prefer rock, grunge, folk
  - Clothing: I prefer cotton, casual
  - Movies: I prefer sci-fi, thrillers

CF formulation

UFMG COMPUTER SCIENCE

- Input
  - A matrix of user-item ratings
    - Ratings could be either implicit or explicit
    - Items could be pretty much anything
- Output (for a target user)
  - A number indicating the user’s predicted preference for an item ***(prediction)***
  - A list of items in decreasing order of predicted preference for the user ***(top-n recommendation)***

CF formulation

UFMG COMPUTER SCIENCE

- *Predict how much user **u** will like item **i***

			i					
			1	3		2		
	1	2	5		4		1	
	4			3	5		4	
u		2	?		5	4		4
		3	4		5		3	

buys
clicks
views
rates
reviews
…

CF formulation

UFMG COMPUTER SCIENCE

- *Recommend items that user **u** will like*

			1	3		2		
	1	2	5		4		1	
	4			3	5		4	
u	?	2	?	?	5	4	?	4
		3	4		5		3	

buys
clicks
views
rates
reviews
…

CF formulation

- *What if we know nothing about user u?*

			i					
			1	3		2		
	1	2	5		4		1	
	4			3	5		4	
u			?					
		3	4		5		3	

Non-personalized CF

- *What if we know nothing about user u?*

			i					
			1	3		2		
	1	2	5		4		1	
	4			3	5		4	
u			3.3					
		3	4		5		3	

We can use the average rating!

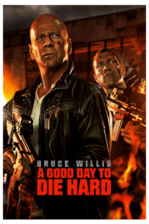
$$\hat{r}_{ui} = \frac{1}{|U_i|} \sum_{x \in U_i} r_{xi}$$

$$\hat{r}_{ui} = \frac{1}{3}(1+5+4)$$

$$\hat{r}_{ui} = 3.33$$

Non-personalized CF

- Problem?
  - Predicted utility of *i* will be the same for all users



Non-personalized CF

- Problem?
  - Predicted utility of *i* will be the same for all users
- Solution
  - We could compute a segmented average
    - e.g., by age, gender, income, location
  - Better, but still not fully personalized
    - Prediction will be the same for a given segment

Non-personalized CF

- Problem?
  - Predicted utility of *i* ignores context

Non-personalized CF

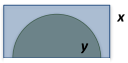
- Problem?
  - Predicted utility of *i* ignores context
- Solution
  - We could compute non-personalized associations
    - e.g., what sauce goes along with ice cream?
  - Great, but what associations to leverage?
    - ***Historical profiles*** may introduce spurious associations
    - ***Transaction data*** may limit follow-up sales
    - ***Time-constrained profiles*** offer a compromise

Associative recommendations

UFMG COMPUTER SCIENCE

- Start simple
  - Percentage of ***x***-buyers who also bought ***y***

$$\hat{r}_{uy} \propto \frac{f(x \wedge y)}{f(x)}$$



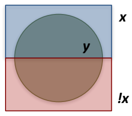
- Problem:
  - What if ***y*** is extremely popular?

Associative recommendations

UFMG COMPUTER SCIENCE

- Start simple
  - Percentage of ***x***-buyers who also bought ***y***

$$\hat{r}_{uy} \propto \frac{f(x \wedge y)}{f(x)}$$



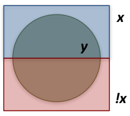
- Problem:
  - What if ***y*** is extremely popular?

Associative recommendations

UFMG COMPUTER SCIENCE

- Take two
  - Does ***x*** make ***y*** more likely?

$$\hat{r}_{uy} \propto \left[\frac{f(x \wedge y)}{f(x)} \Big/ \frac{f(!x \wedge y)}{f(!x)} \right]$$



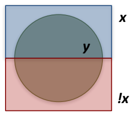
- Intuitively:
  - Is ***y*** more likely with ***x*** than without it?

Associative recommendations

UFMG COMPUTER SCIENCE

- More generally
  - Does ***x*** make ***y*** more likely?

$$\hat{r}_{uy} \propto \frac{p(x \cap y)}{p(x)p(y)}$$



- Intuitively:
  - Are ***x*** and ***y*** more likely to occur together than separately? (aka "lift" in association rule mining)
  - Lift = 1: ***x*** and ***y*** are independent

Summary

UFMG COMPUTER SCIENCE

- Recommenders mine what users *say* and what they *do* to learn preferences
  - Ratings provide explicit expressions of preference
  - Implicit data benefits from greater volume
- Non-personalized recommenders are a good first approximation of users' preferences
  - May be the only possibility in some cases
- To go beyond, we need more data
  - Personalized recommenders