UFMG UNIVERSIDADE FEDERAL DE MINAS GERAIS

Recommender Systems

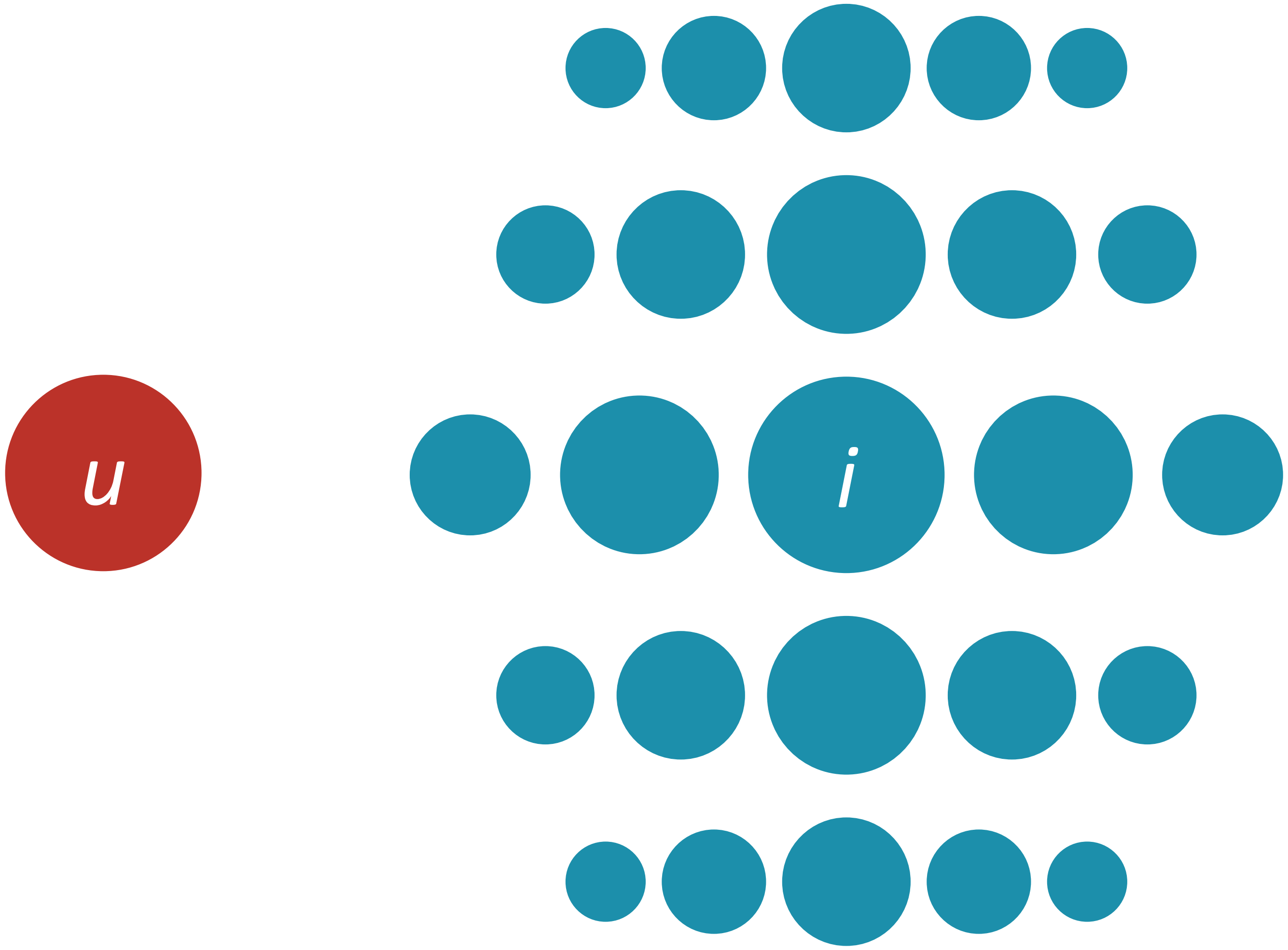
Non-Personalized Collaborative Filtering

Rodrygo L. T. Santos
rodrygo@dcc.ufmg.br

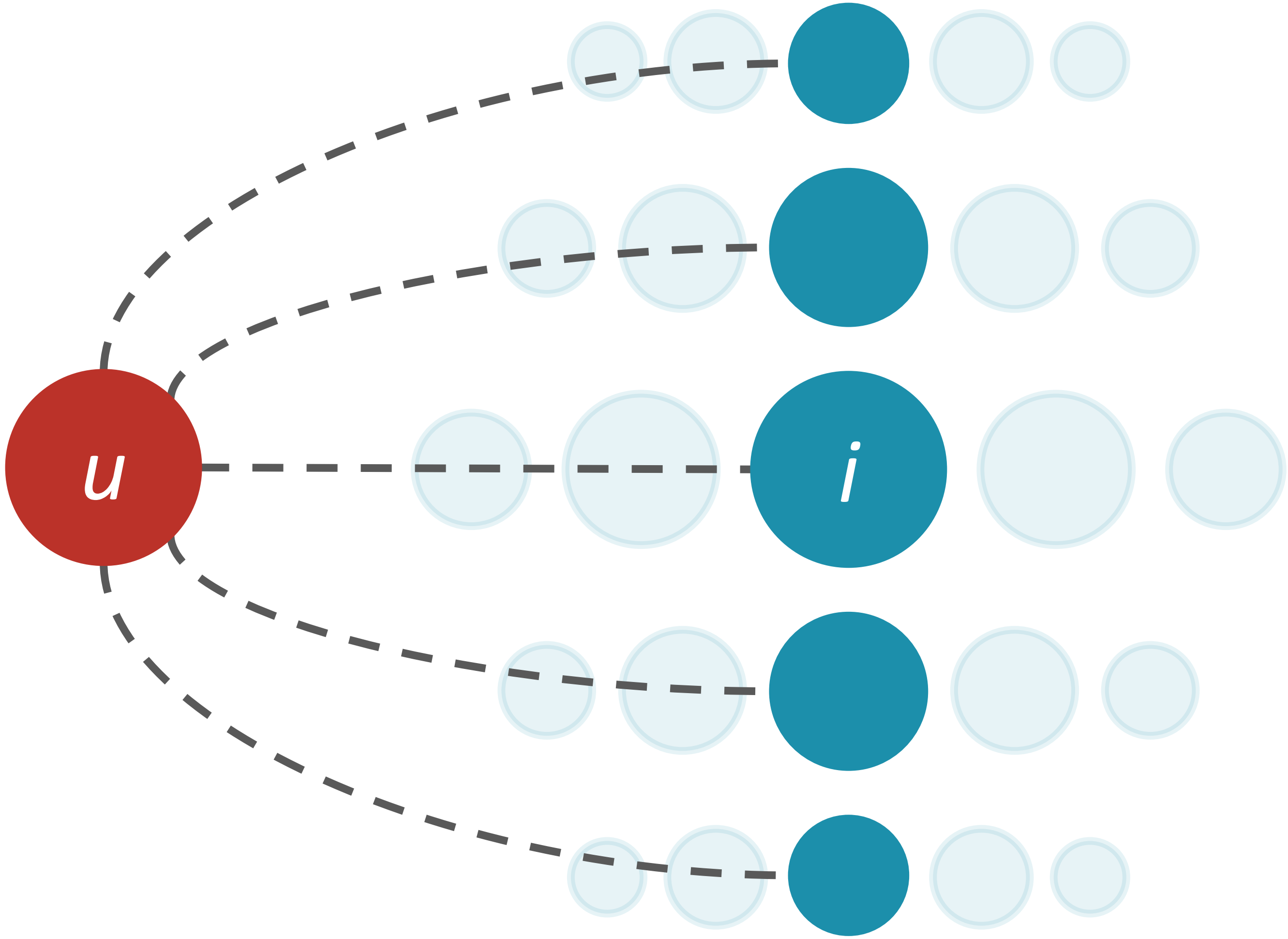
The recommendation problem

$$f(u,i)$$

The recommendation problem



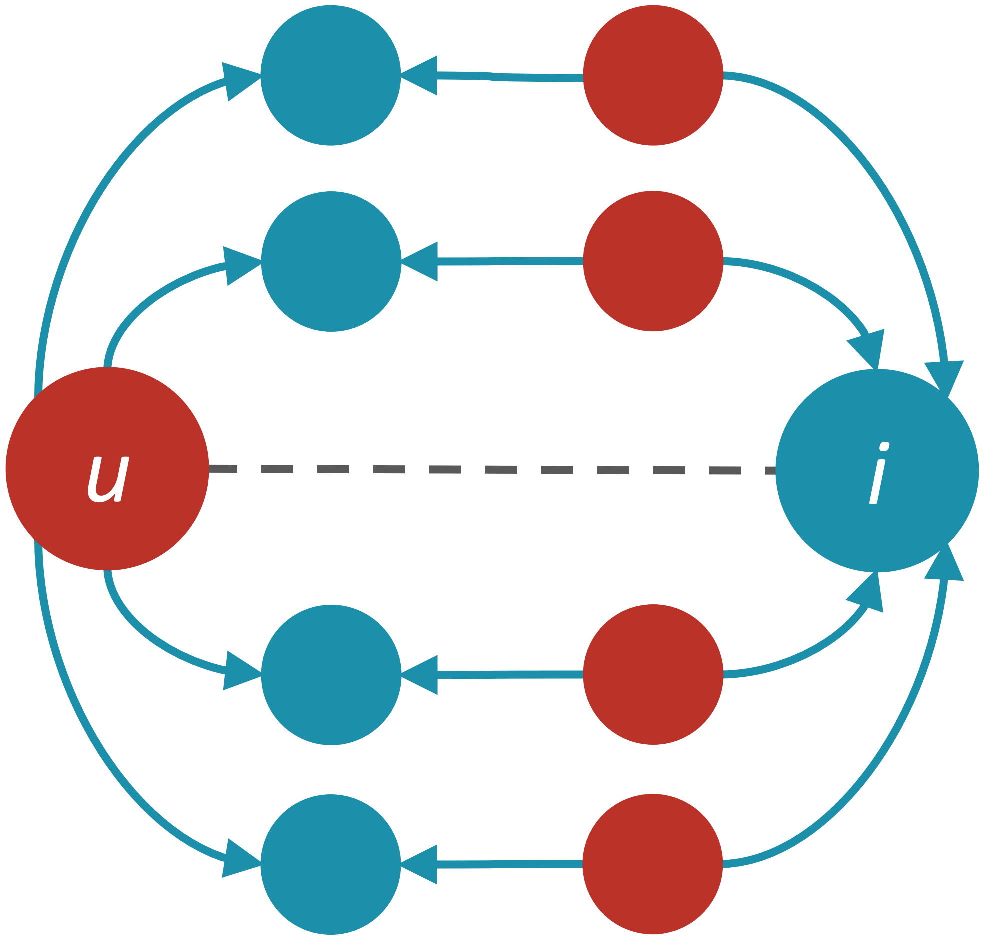
The recommendation problem



The recommendation problem

$$f(u,i)$$

Collaborative recommendation



Collaborative recommendation

Key idea

- Leverage the “wisdom of the crowds”

Most prominent recommendation approach

- Used by large commercial e-commerce sites
- Applicable in many domains (books, movies, etc.)

Stable preferences

Basic assumption

- Past preferences indicate future preferences

Some examples

- News: I prefer technology, travel
- Music: I prefer rock, grunge, folk
- Movies: I prefer sci-fi, thrillers

Modeling preferences

We want to know

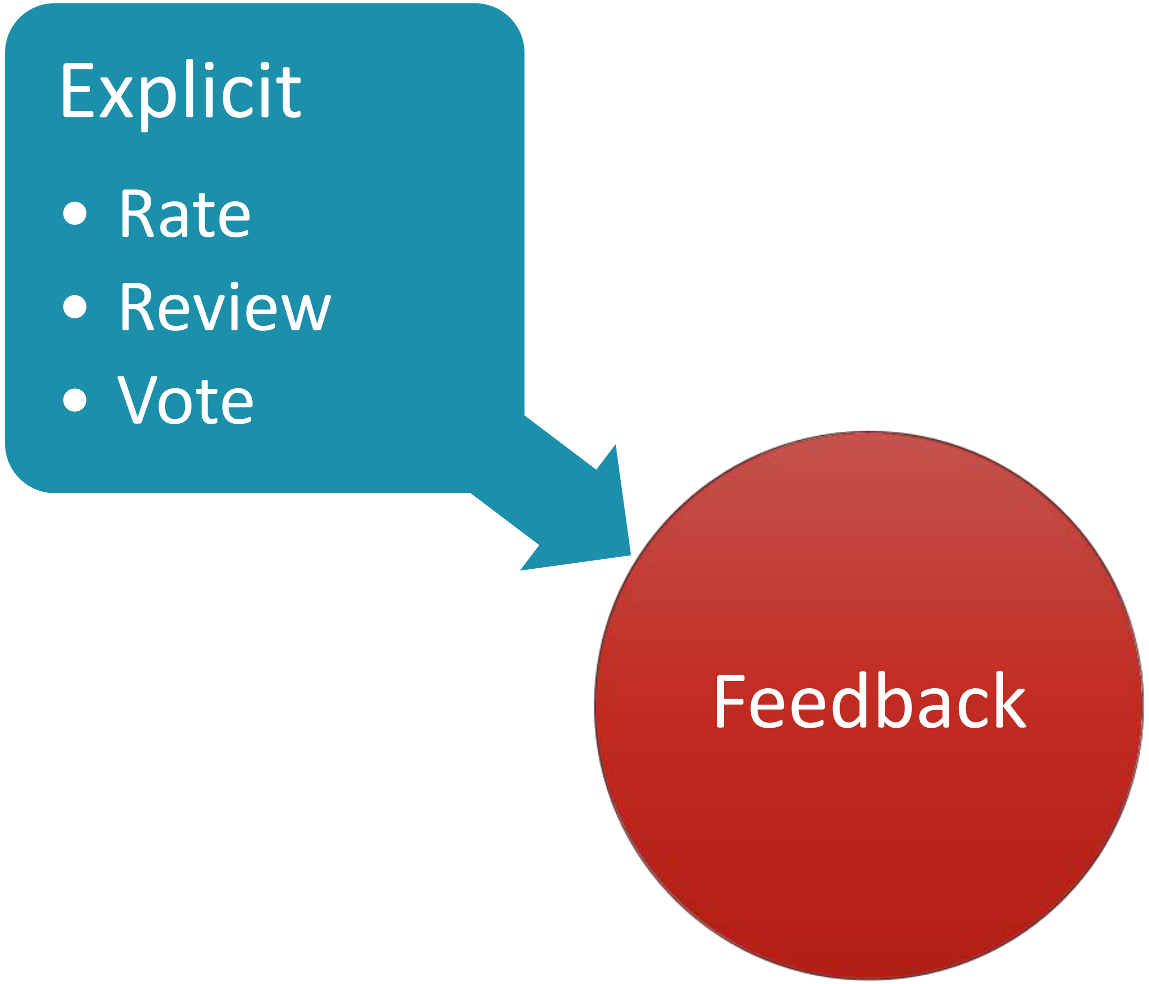
- What users consider relevant

We can observe

- What users tell us (ratings)
- What users do (actions)

These are ***noisy*** measurements

Feedback model




IMDb
Find Movies, TV shows, Celebrities and more...
All
Movies, TV & Showtimes
Celebs, Events & Photos
News & Community
Watchlist
B E H I N D
E V E R Y
C O D E
I S A N
E N I G M A
THE IMITATION GAME
Contact the Filmmakers on IMDbPro »
The Imitation Game (2014)
9
PG-13 114 min - Biography | Drama | Thriller -
25 December 2014 (USA)
8.2
Your rating: 8/10
Ratings: 8.2/10 from 121,237 users Metascore: 73/100
Reviews: 348 user | 342 critic | 49 from Metacritic.com
During World War II, mathematician Alan Turing tries to crack the enigma code with help from fellow mathematicians.
Director: Morten Tyldum
Writers: Andrew Hodges (book), Graham Moore (screenplay)
Stars: Benedict Cumberbatch, Keira Knightley, Matthew Goode | See full cast and crew »
+ Watchlist
Watch Trailer
Share...

rate / review

Explicit feedback

Are ratings reliable and accurate?

- *Are my 8/10 stars equivalent to yours?*

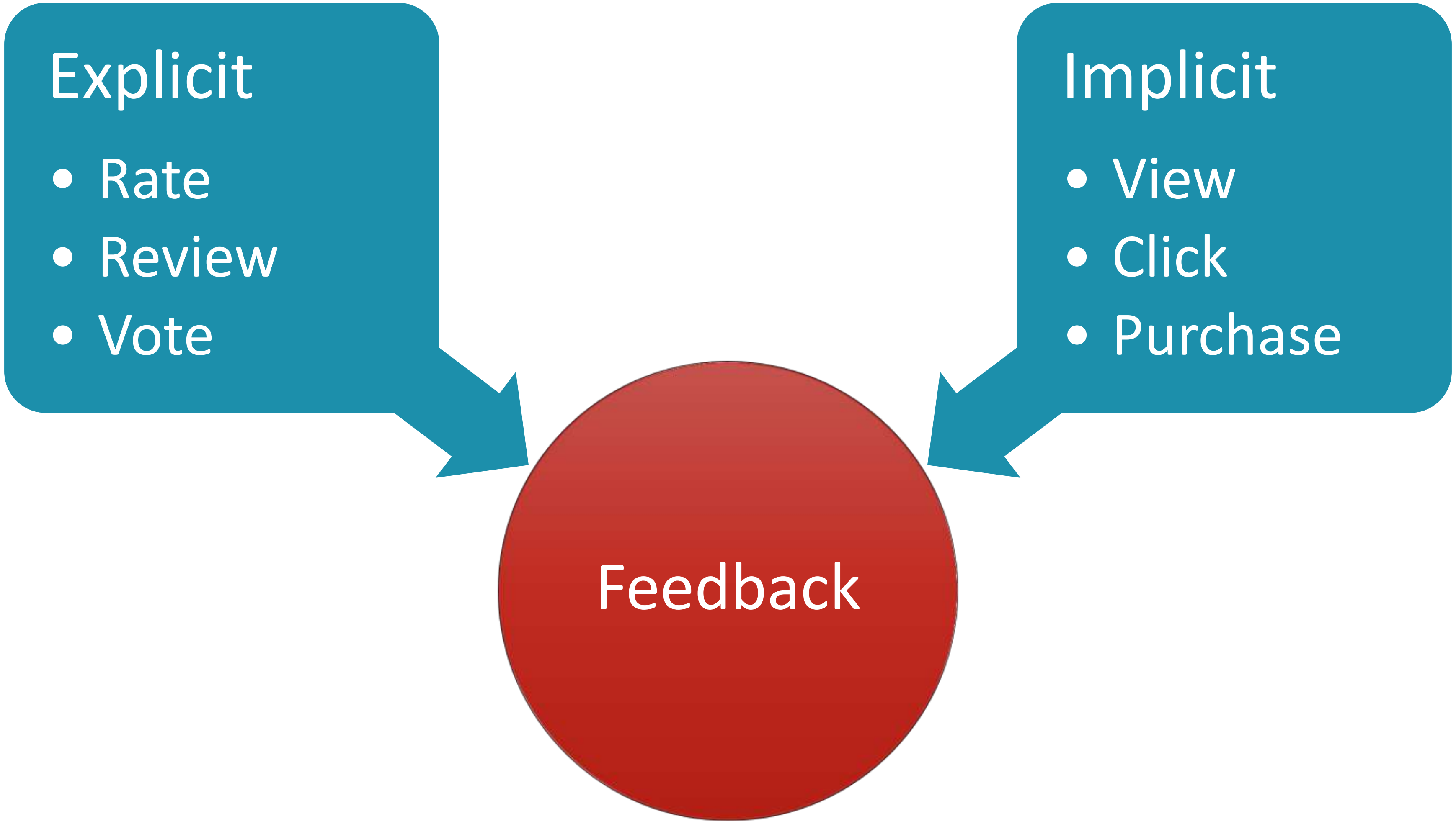
Do user preferences change?

- *Will I still like the item after 10 years?*

What does a rating mean?

- *Will I ultimately consume the rated item?*

Feedback model



Implicit feedback

Abundant data from user actions

- Not direct expressions of preference

But indirectly say a lot!

- Action signals (click, skip, play, purchase)
- Attention signals (reading, listening, viewing time)
- Cognitive signals (eye-tracking, brain imaging)

Implicit feedback

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?

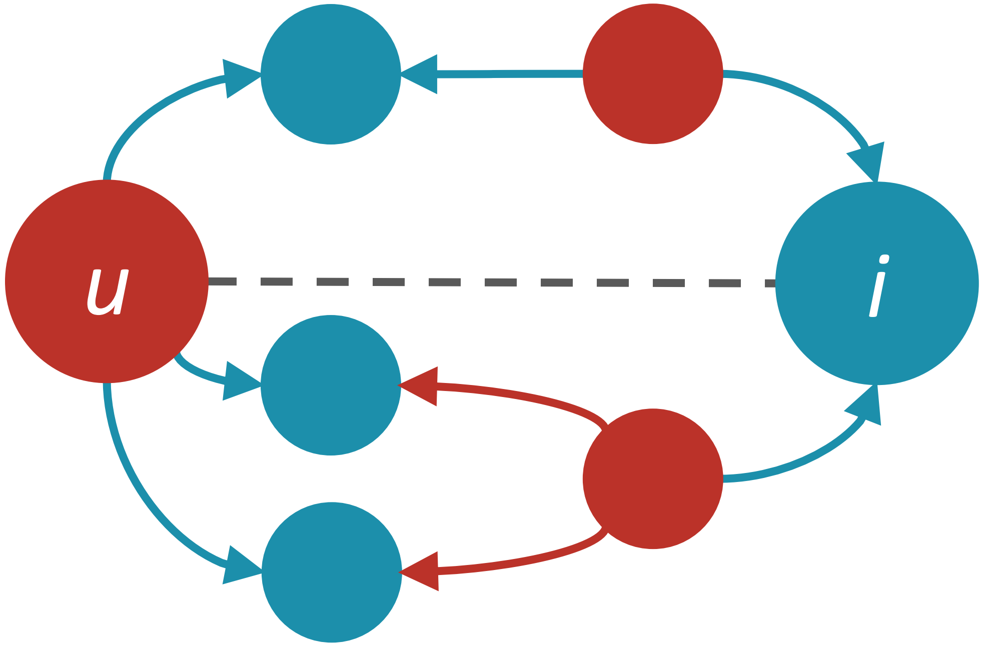
- Position, presentation, popularity, etc.

How to leverage user feedback?

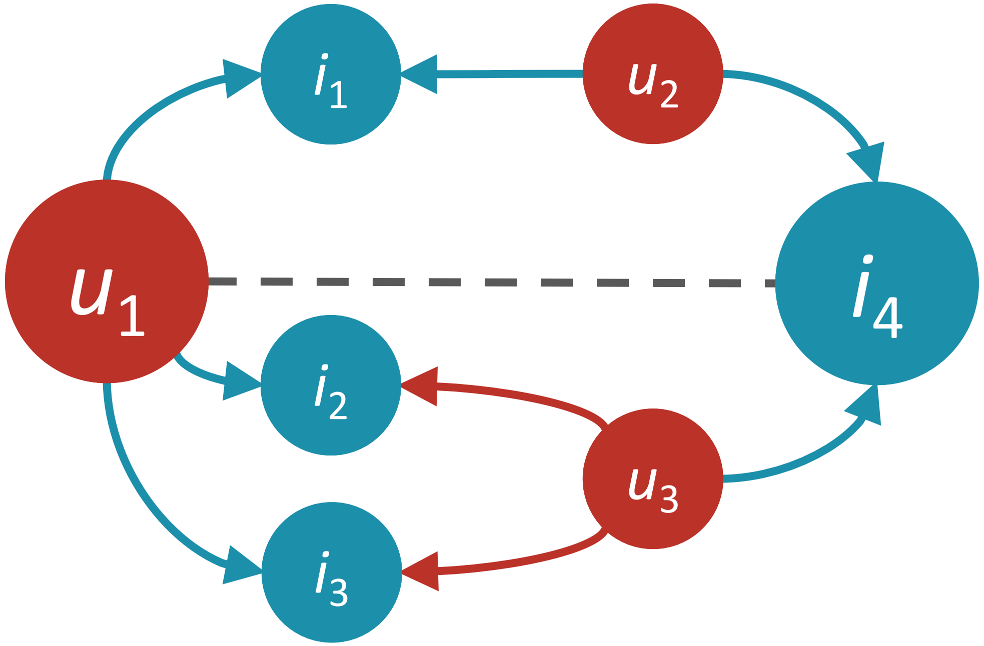
Leveraging user feedback

$$f(u,i)$$

Leveraging user feedback



Leveraging user feedback



Leveraging user feedback

	i_1	i_2	i_3	i_4
u_1	+1	+1	+1	
u_2	+1			+1
u_3				+1

unary feedback
(e.g., click)

Leveraging user feedback

	i_1	i_2	i_3	i_4
u_1	+1	+1	+1	
u_2	+1			+1
u_3		-1	-1	+1

binary feedback
(e.g., like / dislike)

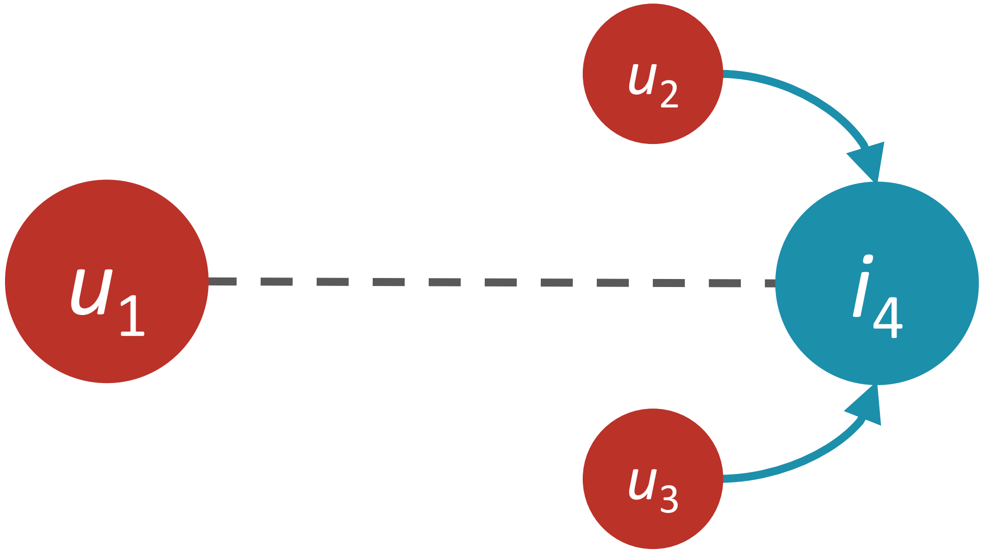
Leveraging user feedback

	i_1	i_2	i_3	i_4
u_1	5	3	3	?
u_2	5			3
u_3		1	2	1

graded feedback
(e.g., rate)

What if we don’t know the user?

Cold-start user



Cold-start user

	i_1	i_2	i_3	i_4
u_1				?
u_2	5			3
u_3		1	2	1

we can use some **global property** of the target item

Non-personalized collaborative filtering

	i_1	i_2	i_3	i_4
u_1				?
u_2				3
u_3				1

most recent
most popular
best rated

Best-rated recommendation

	i_1	i_2	i_3	i_4
u_1				?
u_2				3
u_3				1

$$\hat{r}_{u_1 i_4} = \frac{1}{|U_{i_4}|} \sum_{v \in U_{i_4}} r_{v i_4}$$

$$= \frac{1}{2}(3 + 1)$$

$$= 2$$

Most-popular recommendation

	i_1	i_2	i_3	i_4
u_1				?
u_2				3
u_3				1

$$\hat{r}_{u_1 i_4} = |U_{i_4}|$$
$$= 2$$

What could go wrong?

Problem #1: imbalanced feedback

Best rated denture cleaners
(ranked by average rating)

Problem #1: imbalanced feedback

Best rated definitions for "geek"
(ranked by proportion of positive ratings)

Geek

One who passionately engages in one or more things to extreme levels. A commonly ascribed term to people in the field of computer programming, but one does not have to be in a technical field to be a geek. The only criteria is an intense level of interest in something, often to a highly specialized degree.

by Mr. ZAP February 13, 2011

👍 52 | 👎 9

net ratings: 52 – 9 = 43 (85% positive)

Geek

Not to be confused with Nerd. A geek does not have to be smart, a Geek is someone who is generaly not athletic, and enjoys Video Games; Comic Books; being on the internet, and etc.

by Sknywhtboy88 November 05, 2004

👍 10086 | 👎 2990

net ratings: 10,086 – 2,990 = 7,096 (77% positive)

Problem #1: imbalanced feedback

Predicted utility of i ignores feedback imbalance

- We could add a confidence factor to balance the observed feedback with the incurred uncertainty

Typically (similar methods for proportions)

- $\hat{r}_{ui} \propto \bar{r}_i - z \frac{\sigma}{\sqrt{|U_i|}}$ (penalize proportionally to variance) (penalize inversely proportionally to sample size)

Problem #1: imbalanced feedback

Best rated definitions for "geek"
(ranked by proportion of positive ratings)

Geek

One who passionately engages in one or more things to extreme levels. A commonly ascribed term to people in the field of computer programming, but one does not have to be in a technical field to be a geek. The only criteria is an intense level of interest in something, often to a highly specialized degree.

by Mr. ZAP February 13, 2011

52 | 9

net ratings: 52 – 9 = 43 (85% positive)

lower bound for 95% confidence: 74%

Geek

Not to be confused with Nerd. A geek does not have to be smart, a Geek is someone who is generaly not athletic, and enjoys Video Games; Comic Books; being on the internet, and etc.

by Sknywhtboy88 November 05, 2004

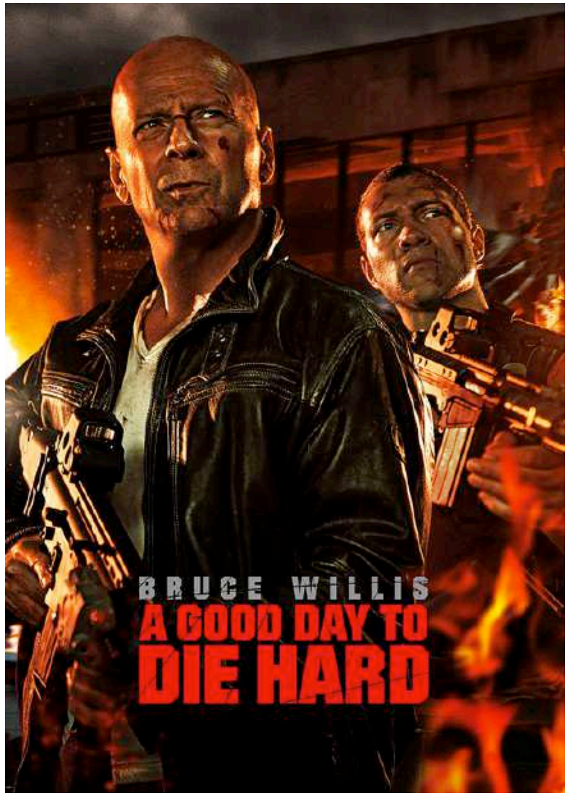
10086 | 2990

net ratings: 10,086 – 2,990 = 7,096 (77% positive)

lower bound for 95% confidence: 76%

Problem #2: user agnosticism

Most popular shows…



Problem #2: user agnosticism

Predicted utility of i will be the same for all users

- We could compute segmented statistics
  - e.g., by age, gender, income, location
- Better, but still not fully personalized
  - Prediction will be the same for a given segment

Problem #3: context agnosticism

Best rated sauce

… to go along with your ice cream?

Problem #3: context agnosticism

Predicted utility of i ignores context

- 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

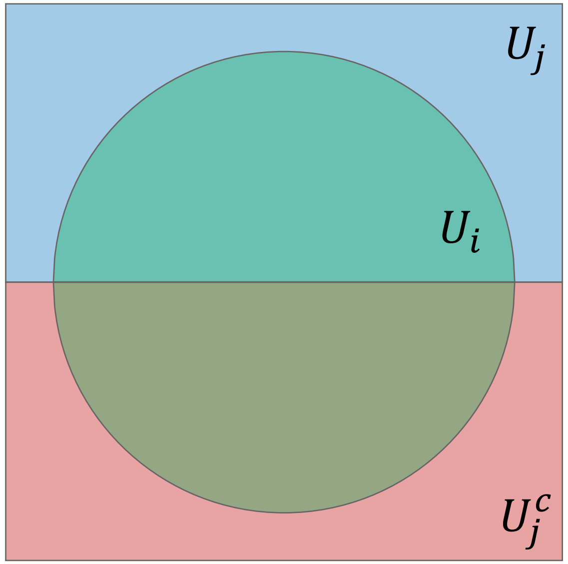
Associative recommendation

Percentage of j-buyers who also bought i

- $\hat{r}_{ui} \propto \frac{|U_i \cap U_j|}{|U_j|}$

Problem:

- What if i is extremely popular?



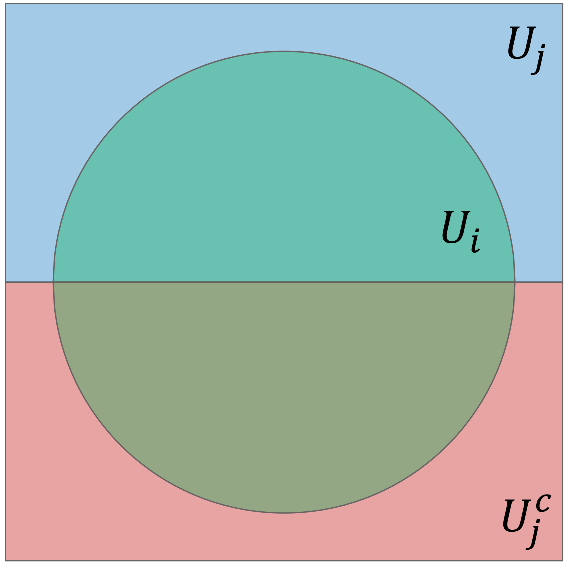
Associative recommendation

Odds of j-buyers to also buy i

- $\hat{r}_{ui} \propto \frac{|U_i \cap U_j|}{|U_j|} \bigg/ \frac{|U_i \cap U_j^c|}{|U_j^c|}$

Intuitively:

- Is i more likely with j than without it?



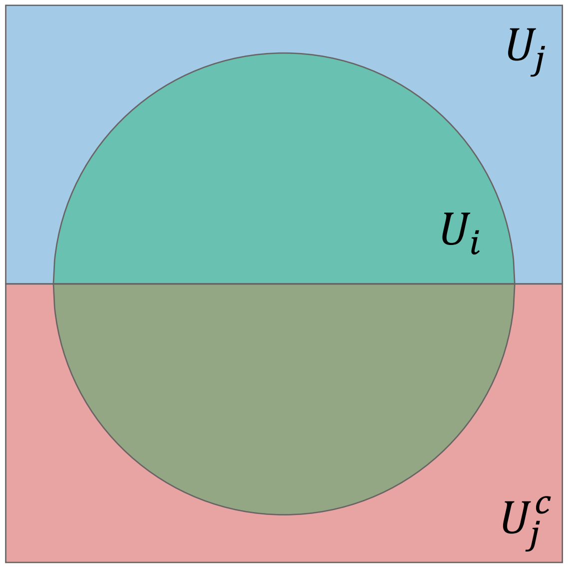
Associative recommendation

More generally:

- $\hat{r}_{ui} \propto \frac{p(i,j)}{p(i)p(j)}$ (aka “lift”)

Intuitively:

- Are i and j more likely to occur together than separately?
- Lift = 1: i and j are independent



Summary

Recommenders mine what users *say* and what they *do*

- Ratings provide explicit expressions of preference
- Implicit data benefits from greater volume

Non-personalized recommenders are a good first try

- May be the only possibility in some cases

Summary

To get personal, we need more data

- Personalized recommendation models
- Preference elicitation models