UFMG UNIVERSIDADE FEDERAL DE MINAS GERAIS

Recommender Systems

Diversity and Novelty

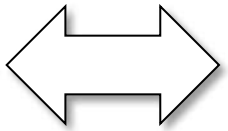
Rodrygo L. T. Santos
rodrygo@dcc.ufmg.br

You bought
(or browsed)

Revolver

Abbey Road

Users
also bought



So you are recommended…


Rubber Soul


With The


Beatles


Let it be

Help!

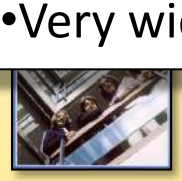
A Hard Day's
Night

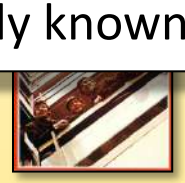
The recommended items are…

- Very similar to each other
- Very similar to what the user has already seen
- Very widely known

The White
Album

Please
Please me


1967-1970
(Blue)


1962-1966
(Red)

Past Masters

Past Masters
Vol 2


Dark Side
of the Moon


Some Girls


Bob Dylan

…

More Beatles' albums

Diversity in search

> *Avoid **redundancy** of possible user intents to cope with the **ambiguity** in the query*

glass

Sign in

All Images Maps Shopping News More Settings Tools

Google

glass

Sign In

All Images Maps Shopping News More Settings Tools

About 2,320,000,000 results (0.72 seconds)

Glass - Wikipedia
https://en.wikipedia.org/wiki/Glass
Glass is a non-crystalline amorphous solid that is often transparent and has widespread practical, technological, and decorative usage in, for example, window ...
Glass ionomer cement: A glass ... · Glass (disambiguation) · History of glass · Sand

material

Glass - X – The Moonshot Factory
https://www.x.company/glass/
Glass Enterprise Edition is a hands-free device, for hands-on workers that removes distractions and helps you focus on what's most important.

smart eyeglasses

Glassdoor Job Search | Find the job that fits your life
https://www.glassdoor.com/index.htm
Search millions of jobs and get the inside scoop on companies with employee reviews, personalized salary tools, and more. Hiring? Post a job for free.
Sign In · Glassdoor Jobs · Companies & Reviews · Know Your Worth

recruiting website

Glass (2019) - IMDb
www.imdb.com/title/tt6823368/
Thriller · The imprisoned Elijah Price holds secrets critical to both David Dunn and Kevin Crumb.

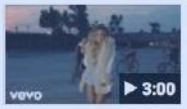
thriller film

Philip Glass
philipglass.com/
Glass holds the Richard and Barbara Debs Composer's Chair at Carnegie Hall for the 2017-2018 season. Highlights will include performances by the Pacific ...

classical composer

Glass | Definition of Glass by Merriam-Webster
https://www.merriam-webster.com/dictionary/glass
Define **glass**: any of various amorphous materials formed from a melt by cooling to rigidity without crystallization: such as — **glass** in a sentence.

word definition

Rachel Platten - Broken Glass - YouTube


https://www.youtube.com/watch?v=b2390GAm4d0
Aug 18, 2017 - Uploaded by RachelPlattenVEVO
Rachel Platten - "Broken **Glass**" (Official Video) Get "Broken **Glass**" when you pre-order her upcoming album ...

pop song

Google Glass 2.0 Is a Startling Second Act | WIRED
https://www.wired.com/story/google-glass-2-is-here/

Glass

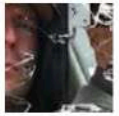
Glass is a non-crystalline amorphous solid that is often transparent and has widespread practical, technological, and decorative usage in, for example, window panes, tableware, and optoelectronics. Wikipedia

View 10+ more

Design Furniture

Feedback

Philip Glass (American composer)
Born: January 31, 1937 (age 80), Baltimore, MD
Compositions: Einstein on the Beach, Satyagraha, Akhnat...



Director: M. Night Shyamalan

Rachel Platten - 'Broken Glass' (Official Video) Get 'Broken Glass' when you pre-order her upcoming album ...

Initial release: January 18, 2019
Director: M. Night Shyamalan

Google Glass 2.0 Is a Startling Second Act | WIRED
https://www.wired.com/story/google-glass-2-is-here/ ▾
Jul 18, 2017 - Google **Glass** flopped. Then Alphabet realized that the future of wearables was in factories and warehouses. Welcome to Google **Glass** 2.0.

Glass Enterprise Edition | Glass Explorer Edition | Google Developers
https://developers.google.com/glass/distribute/glass-enterprise ▾
Jul 18, 2017 - **Glass** Partners are authorized to develop and deliver enterprise solutions for **Glass** customers. Learn more here. Except as otherwise noted, the ...

People also ask

Who invented the glass?

What are some of the properties of glass?

How the glass is manufactured?

What are the Google Glasses?

Feedback

Why Google Glass Broke - The New York Times
https://www.nytimes.com/2015/02/05/style/why-google-glass-broke.html
This is the story of Google **Glass**. Before we begin, this is the part in the tale where I should probably explain what Google **Glass** is. Except ...

Boston wants to fight climate change. So why is every new building ...
https://www.bostonglobe.com/ideas/2017/07/14/boston-wants...glass/.../story.html
Yet **glass** buildings also take a lot of energy to heat and cool. When New York started tracking energy use by skyscrapers, the gleaming 7 World ...

Three-dimensional printing of transparent fused silica glass : Nature ...
www.nature.com/articles/nature22061
Glass is one of the most important high-performance materials used for scientific research, in industry and in society, mainly owing to its ...

news + Q&A

Searches related to glass

glass **movie**	glass **chemical formula**
what is glass **made of elements**	**types of** glass
glass **definition chemistry**	glass **imdb**
how is glass **made from sand**	**properties of** glass

related searches

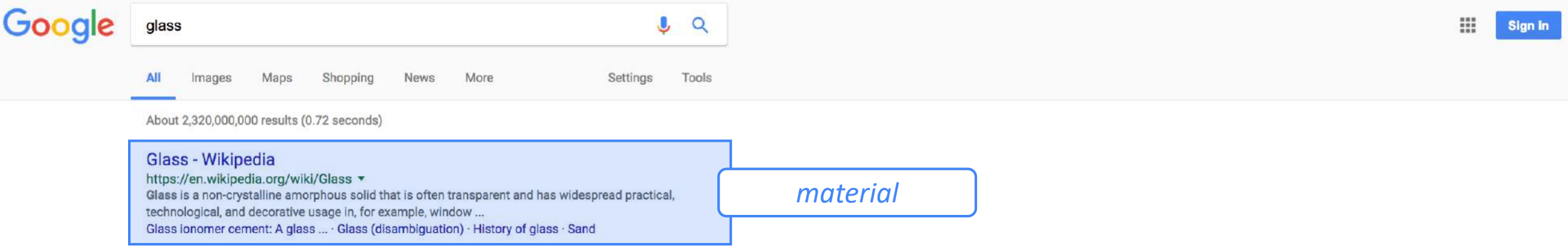
Query ambiguity

Wikipedia lists over 30 meanings for ‘glass’…

- … but we can only display ‘10 blue links’

Ambiguity is inherent to user queries...

- ... but not all queries are equally ambiguous



Any need for more results about the material?

- Users are unlikely to inspect the results any further once they find something relevant (Craswell et al., 2008)

Diversity and novelty
(Clarke et al., SIGIR 2008)

Diversity

- “the need to resolve ambiguity”
  [in the retrieval request]

Novelty

- “the need to avoid redundancy”
  [in the retrieval response]

Diversity in recommendation

> *Avoid **redundancy** of possible user intents to cope with the **ambiguity** ~~in the query~~ in the observed user interests*

Why diversify recommendations
(Vargas, Castells & Vallet SIGIR 2011, 2012)

For better system effectiveness (“a safer bet”)

Uncertainty in user preferences

- Ambiguity, underspecification
- Preferences are multiple, dynamic, contextual…
- Much broader needs than in search

Increase chances of at least some relevant item

Why diversify recommendations
(McAlister 1982, and many more…)

For the sake of it: direct user satisfaction

- Natural variety-seeking drive in human behavior, within a recommendation and over time

Ideal level of stimulation

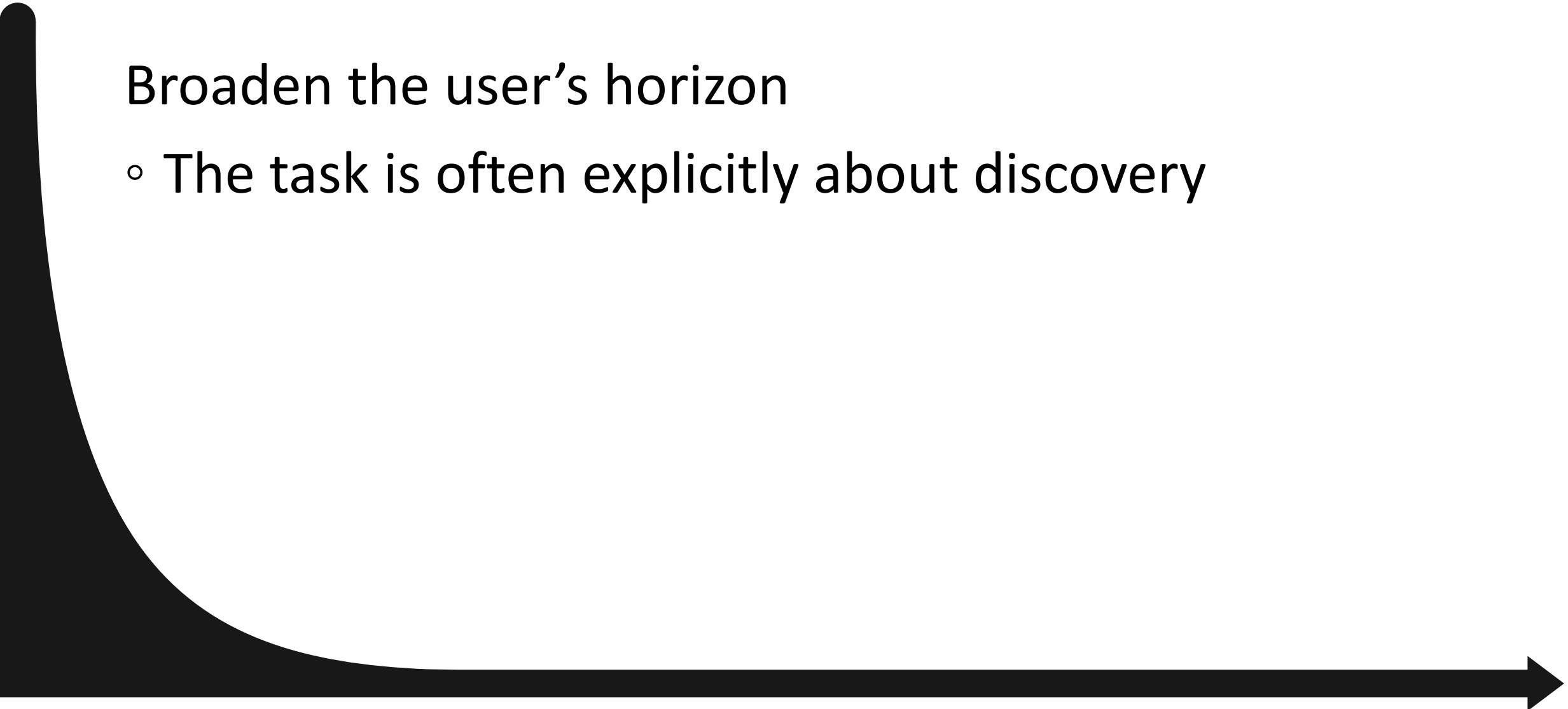
- Desire for the unfamiliar, variety among the familiar

Multiple simultaneous needs and tastes

Why diversify recommendations
(McAlister 1982, and many more…)

Broaden the user's horizon

- The task is often explicitly about discovery



Why diversify recommendations
(Fleder Mgt. Sci. 2009, McNee CHI 2006)

For enhanced business performance

Sales diversity: mitigate risk, expand the business

Long tail: draw revenues from market niches

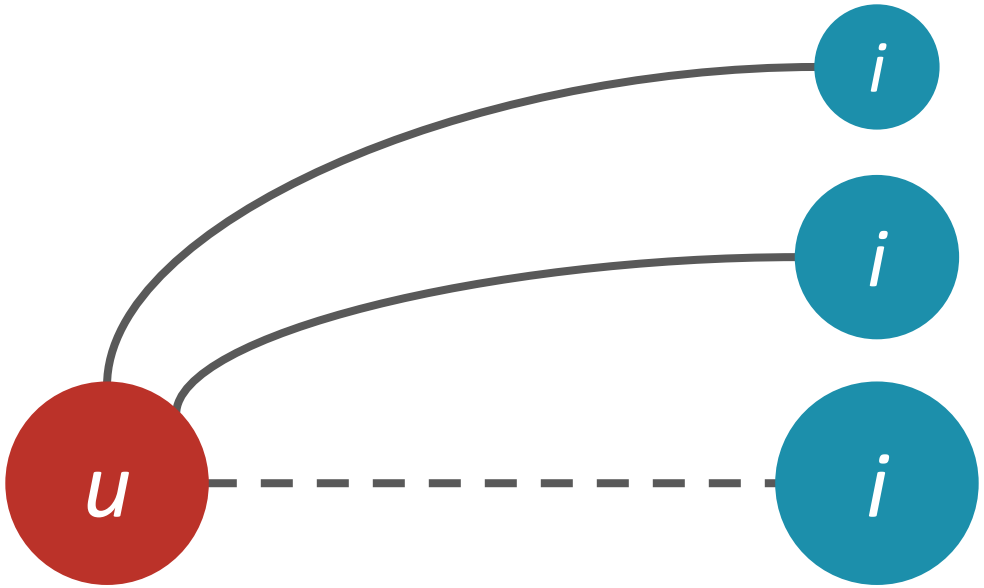
- “Sell less of more”
- Higher profit margin on cheaper long-tail products

The recommendation problem

$$f(u, i)$$

The recommendation problem

The diversitication problem



$$f(u, i, D)$$

The diversification problem

DIVERSIFY

- Given an initial ranking R and an integer τ
- Return a permutation $D \in 2^R$ such that

$$D = \text{argmax}_{D' \in 2^R} \left| \bigcup_{i \in D'} A_u \cap A_i \right|, \text{s. t. } |D'| \leq \tau$$

  - A_u are aspects underlying u
  - A_i are aspects covered by i

Greedy approximation

DIVERSIFY is NP-hard

- Reduction from MAXIMUM COVERAGE

Constant-factor ($1 - 1/e \approx 0.632$) approximation

- Iteratively select an item that covers the most aspects yet uncovered by the previous items

Greedy approximation

$D \leftarrow \emptyset$

while $|D| < \tau$ **do**

$i^* \leftarrow \text{argmax}_{i \in R \setminus D} f(u, i, D)$

$R \leftarrow R \setminus \{i^*\}$

$D \leftarrow D \cup \{i^*\}$

end while

return D

Greedy approximation

Approximation effective in practice (Carterette, 2009)

- Minor deviations from the optimal solution

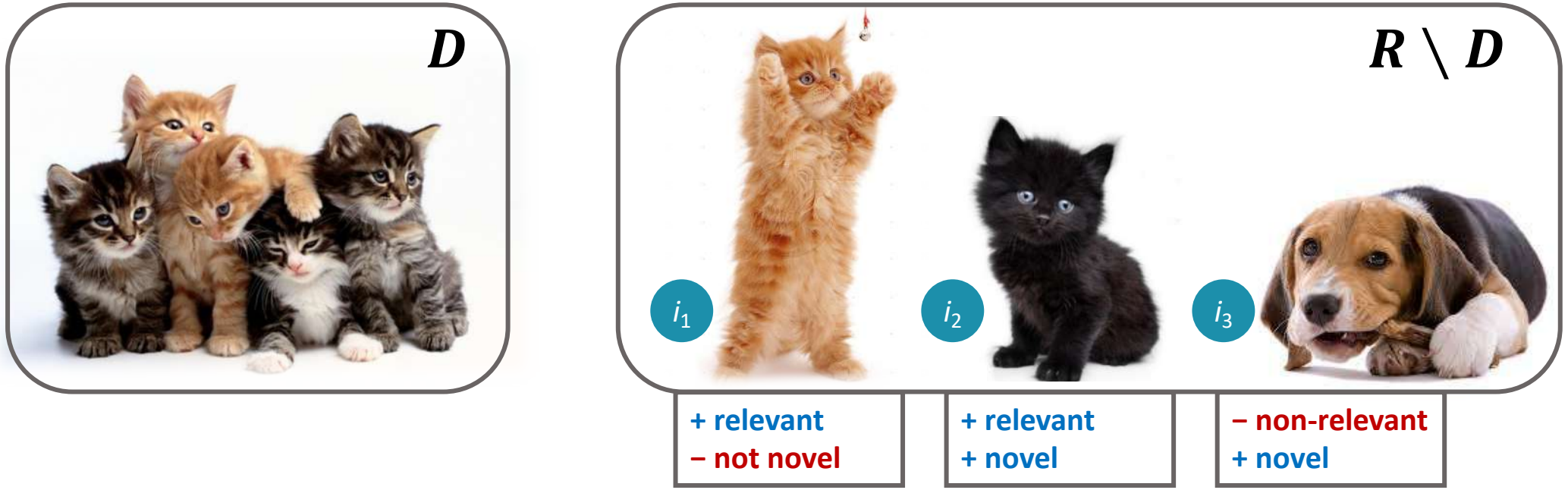
Most approaches focus on producing effective diversification objectives $f(u, i, D)$

How to diversify?

Maximal Marginal Relevance
(Carbonell and Goldstein, 1998)

Promote relevant but dissimilar items

- $f(u, i, D) = \lambda\, f(u, i) - (1 - \lambda) \max_{j \in D} sim(i, j)$



search term SEARCH

Web Directory of High-quality Resources

Arts
Movies, Television, Music...

Business
Jobs, Real Estate, Investing...

Computers
Internet, Software, Hardware...

Games
Video Games, RPGs, Gambling...

Health
Fitness, Medicine Alternative...

Home
Family, Consumers, Cooking...

Kids and Teens
Arts, School Time, Teen Life...

News
Media, Newspapers, Weather...

Recreation
Travel, Food, Outdoors, Humor...

Reference
Maps, Education, Libraries...

Regional
US, Canada, UK, Europe...

Science
Biology, Psychology, Physics...

Shopping
Clothing, Food, Gifts...

Society
People, Religion, Issues...

Sports
Baseball, Basketball, Soccer...

World
Deutsch, Français, 日本語, Italiano, Español, Русский, Nederlands, Polski, Türkçe, Dansk, 简体中文...

IA-Select

(Agrawal et al., WSDM 2009)

Categories as aspects

- $f(u, i, D) = \sum_{c \in T} f(c|u, D) f(i|u, c)$

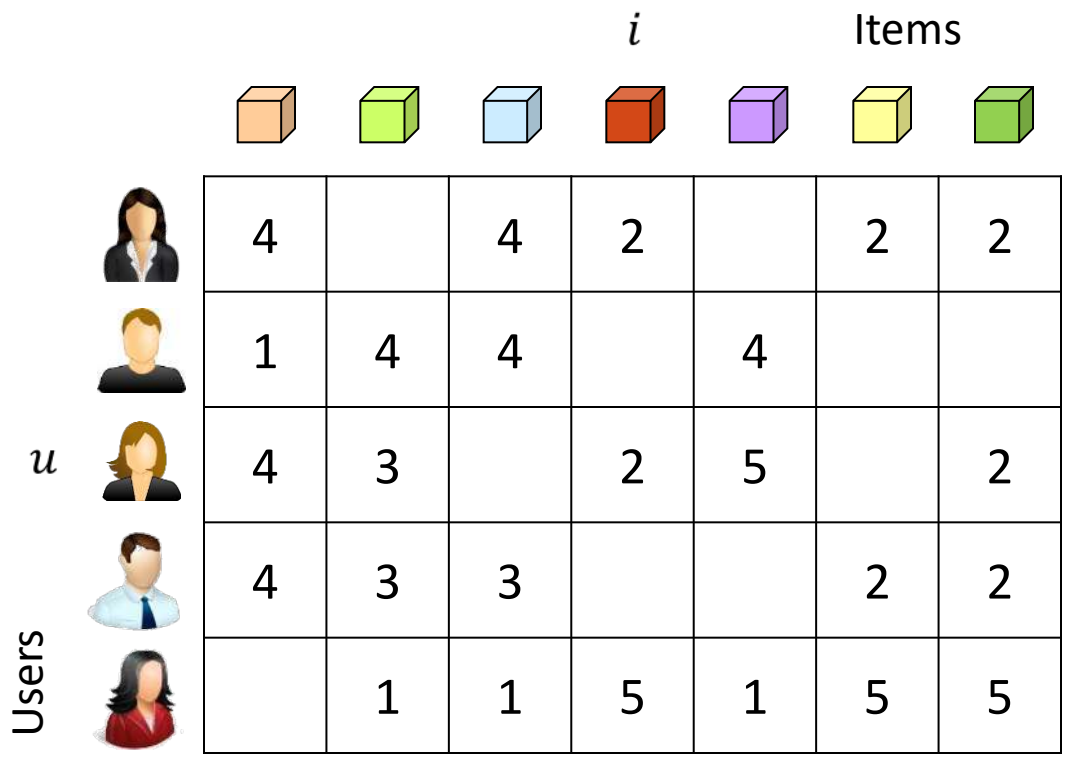
user categories

category utility

category coverage

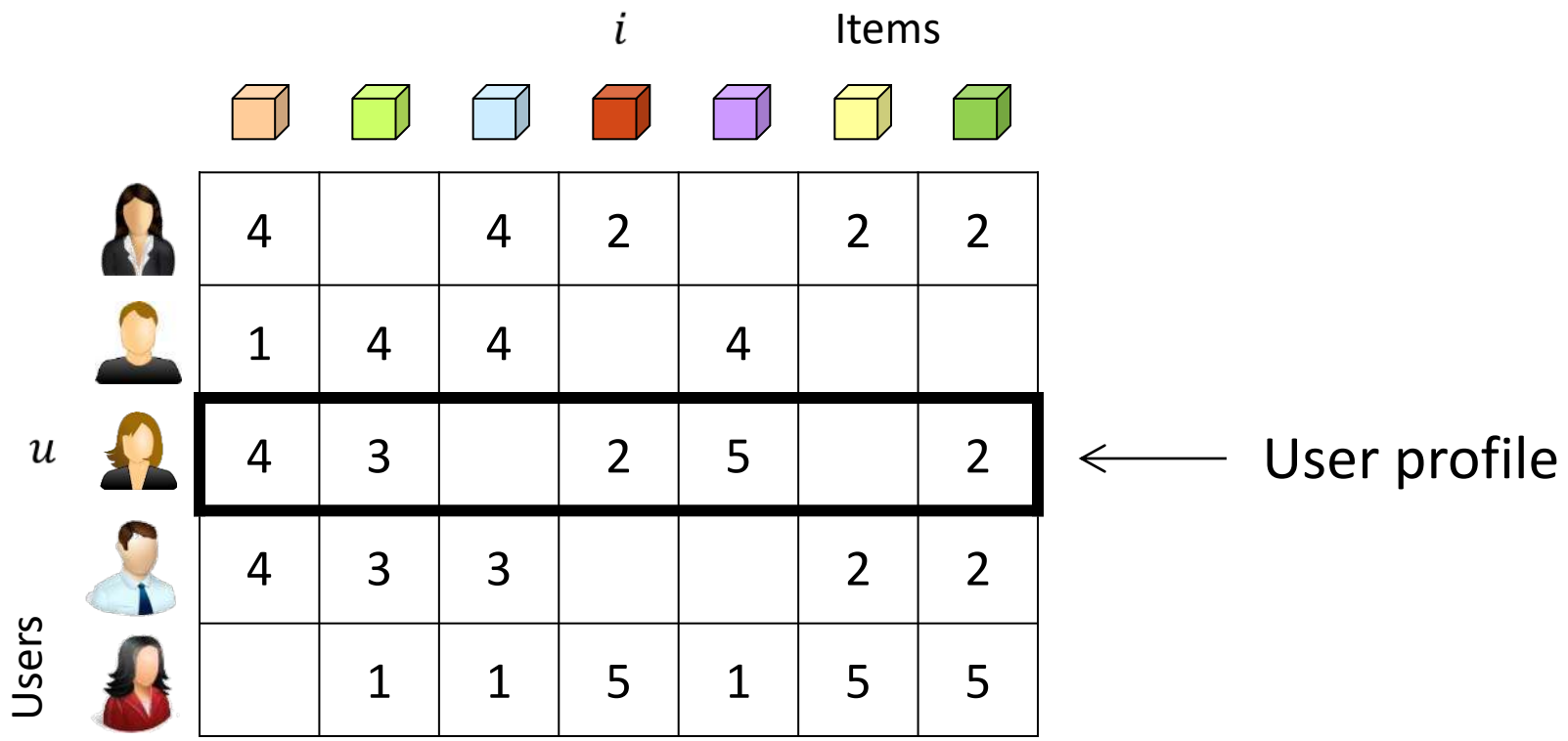
Aspect diversity applied to recommendation

(Shi et al SIGIR 2012; Vargas et al SIGIR 2011, 2012; …)



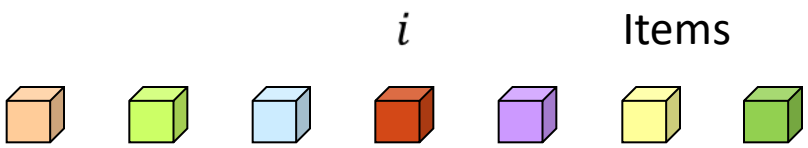
Aspect diversity applied to recommendation

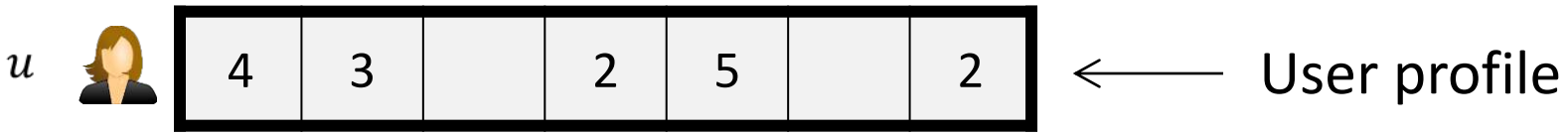
(Shi et al SIGIR 2012; Vargas et al SIGIR 2011, 2012; …)



Aspect diversity applied to recommendation

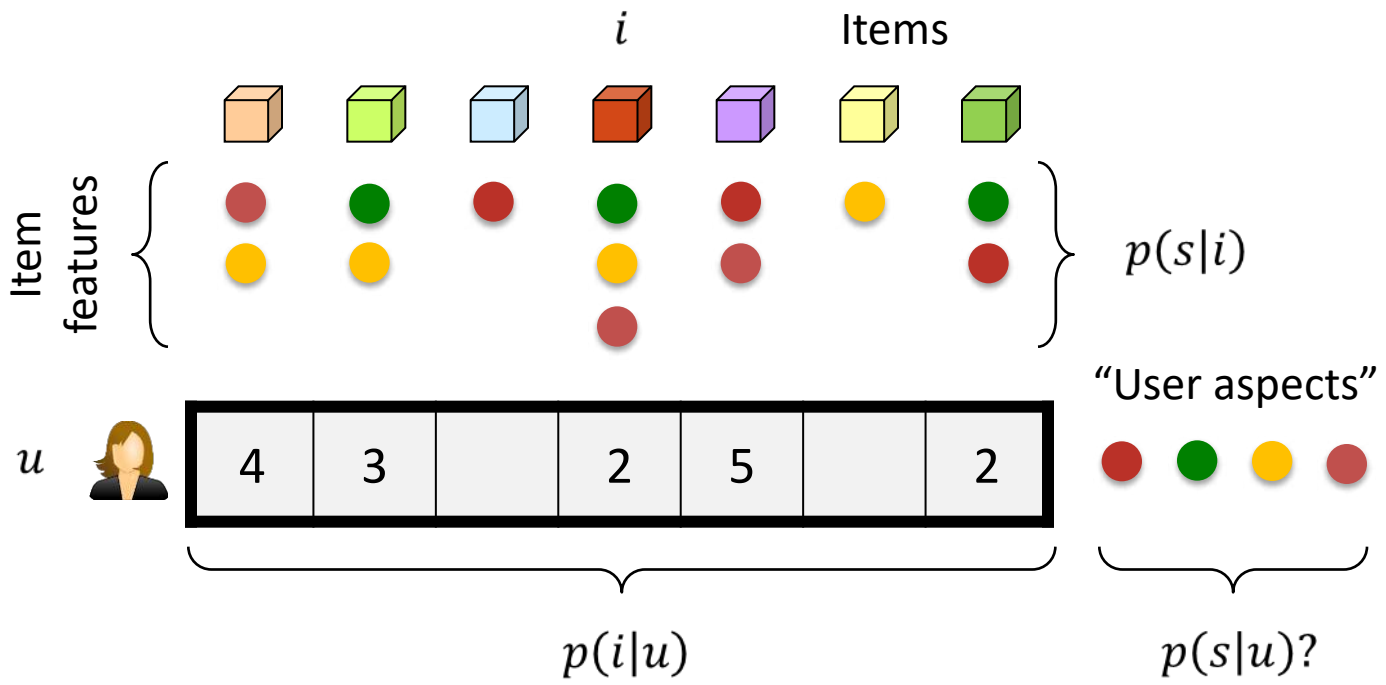
(Shi et al SIGIR 2012; Vargas et al SIGIR 2011, 2012; …)





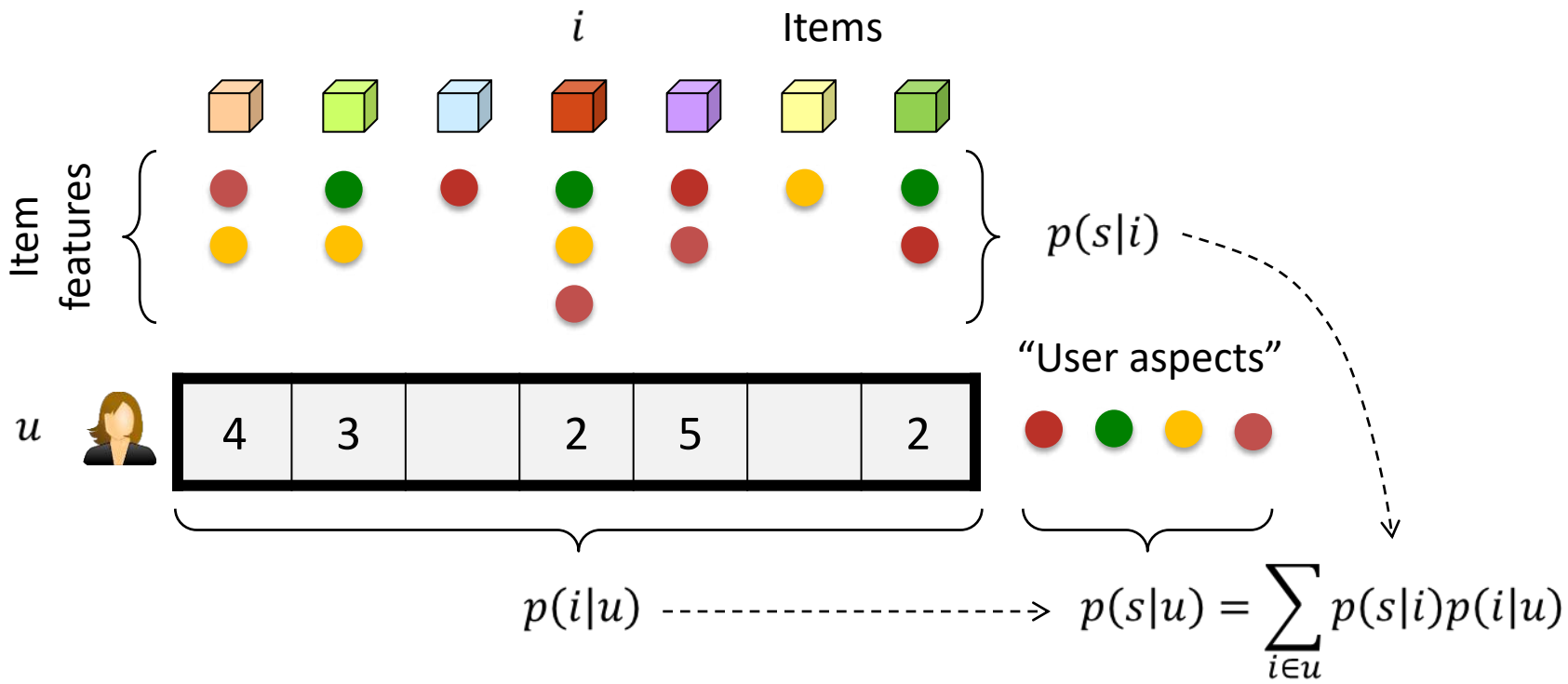
Aspect diversity applied to recommendation

(Shi et al SIGIR 2012; Vargas et al SIGIR 2011, 2012; …)



Aspect diversity applied to recommendation

(Shi et al SIGIR 2012; Vargas et al SIGIR 2011, 2012; …)



How to evaluate diversity and novelty?

Diversity and novelty metrics

Diversification evaluation in search

- Assumes relevance at the query-aspect level

Hard to use in recommendation directly

- User aspects are somewhat latent

Alternative metrics have been proposed

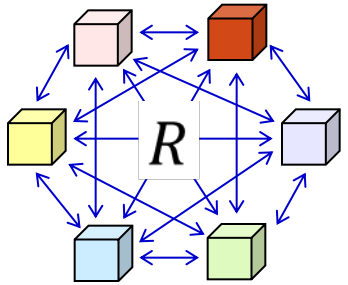
- Relevance and diversity assessed independently

Intra-list diversity

(Smyth & McClave ICCBR 2001, Ziegler et al WWW 2005, etc.)

Average pairwise distance among returned items

$$\text{ILD} = \frac{2}{|R|(|R|-1)} \sum_{\substack{i,j \in R \\ i \neq j}} d(i,j)$$



Internal diversity

Some distance measure, e.g. $d(i,j) = 1 - sim(i,j)$

with sim = cosine, Jaccard, etc. on item features

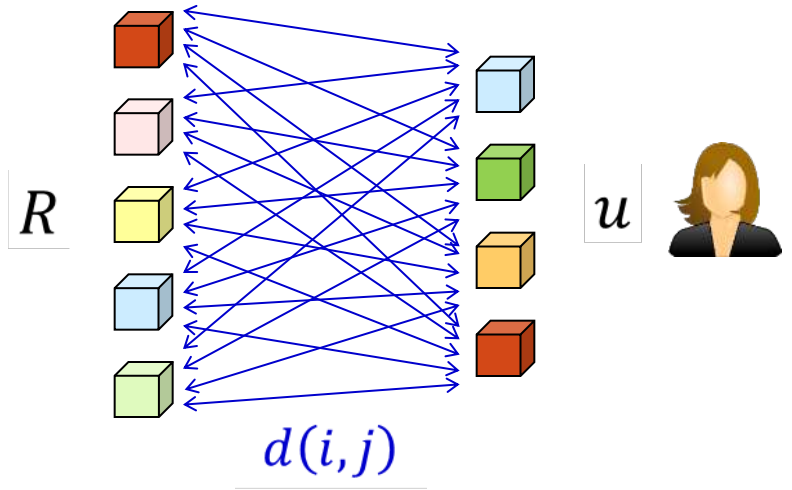
Unexpectedness

(Hurley & Zhang TOIT 2011, Zhang et al WSDM 2012, etc.)

Average distance to items in user profile

$$\text{Unexp} = \frac{1}{|R||u|} \sum_{\substack{i \in R \\ j \in u}} d(i,j)$$

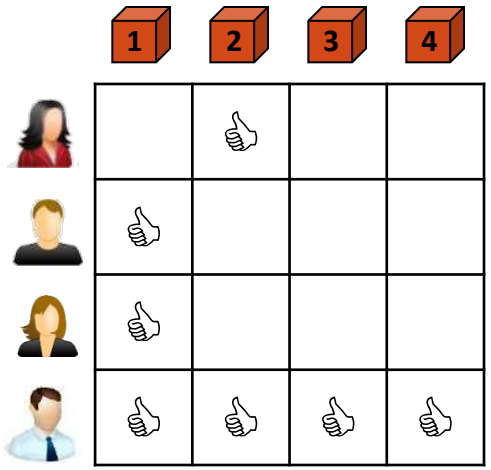
User-specific novelty



Toy example

System's input

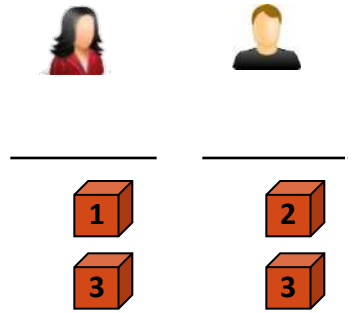
User-item interaction
(e.g. items = music)



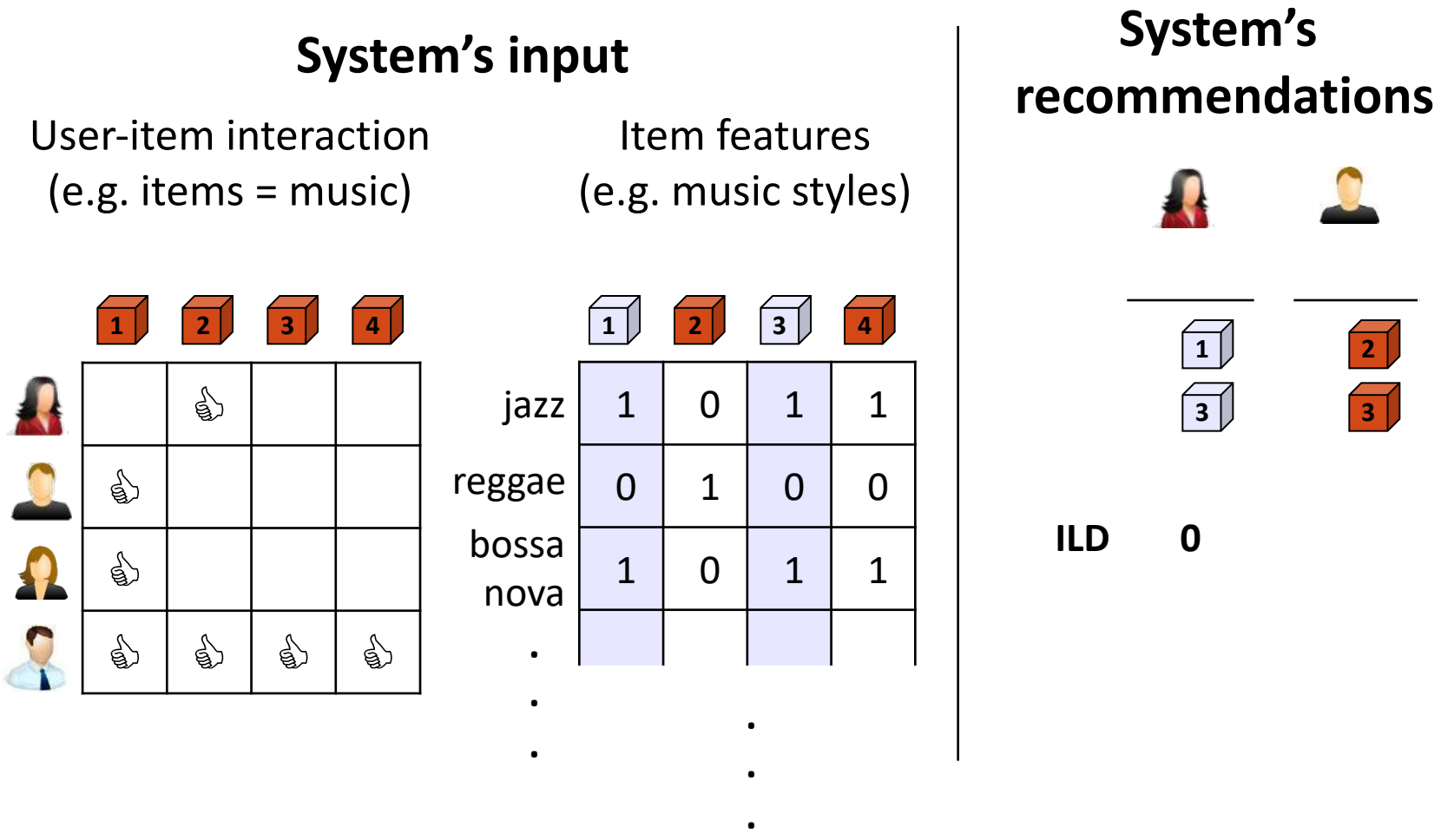
Item features
(e.g. music styles)

	1	2	3	4
jazz	1	0	1	1
reggae	0	1	0	0
bossa nova	1	0	1	1
.				
.				
.				

System's recommendations



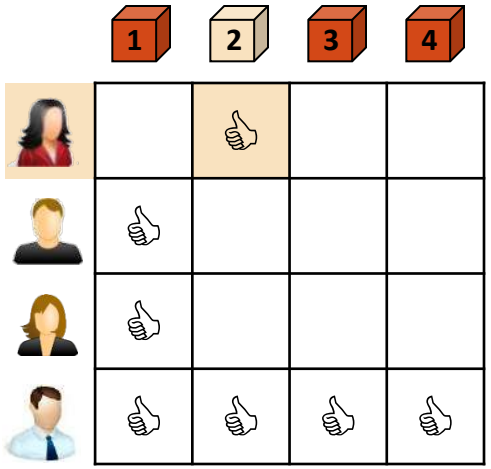
Toy example


System's input
User-item interaction
(e.g. items = music)
Item features
(e.g. music styles)
1
2
3
4
jazz
reggae
bossa
nova
1 0 1 1
0 1 0 0
1 0 1 1
System's
recommendations
1
3
2
3
ILD
0

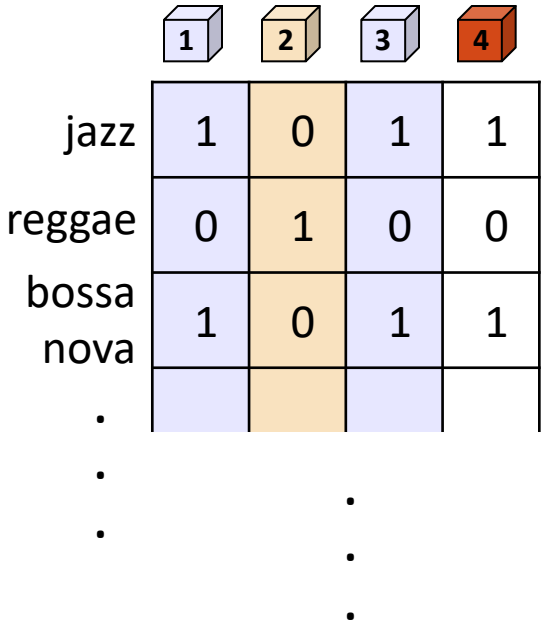
Toy example

System's input

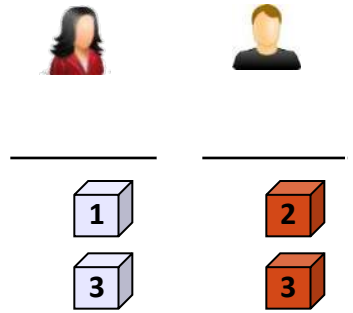
User-item interaction
(e.g. items = music)



Item features
(e.g. music styles)



System's recommendations



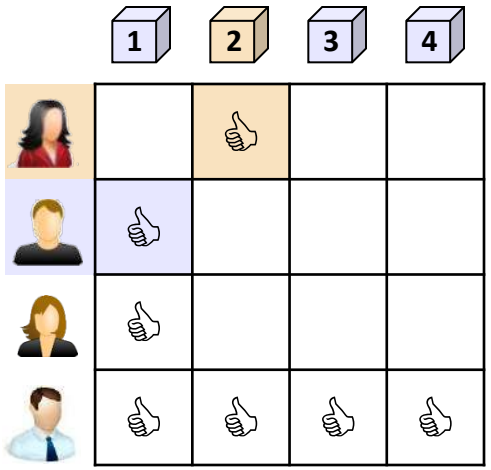
ILD 0

Unexp 1

Toy example

System's input

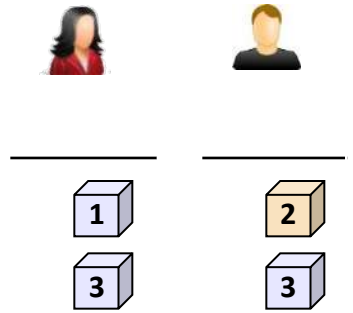
User-item interaction
(e.g. items = music)



Item features
(e.g. music styles)

	1	2	3	4
jazz	1	0	1	1
reggae	0	1	0	0
bossa nova	1	0	1	1
.				
.				
.				

System's recommendations



ILD	0	**1**
Unexp	1	**0.5**

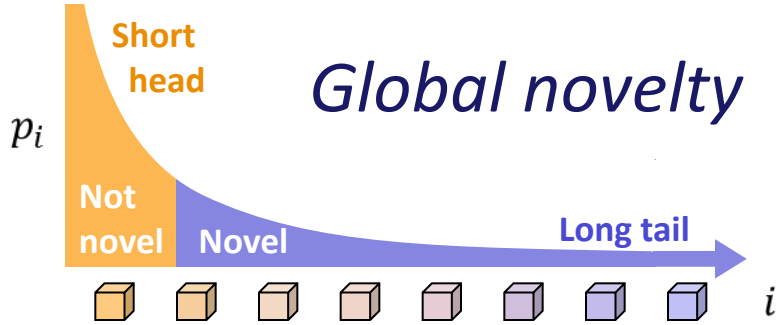
Inverse popularity

(Zhou et al PNAS 2010, Vargas & Castells RecSys 2011,, etc.)

Mean self-information ("average unpopularity")

$$\mathrm{MSI} = -\frac{1}{|R|}\sum_{i \in R} \log_2 p_i$$

Popularity of i



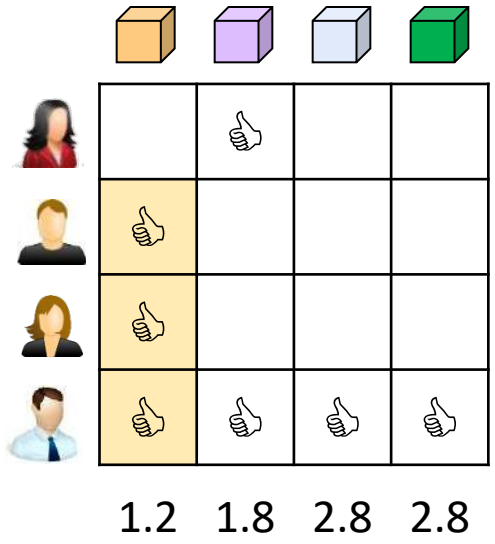
$$p_i = p(known|i) \sim |\{u \in \mathcal{U} | r(u,i) \in \mathcal{R}\}| / |\mathcal{U}|$$
$$p_i = p(i|known) \sim |\{u \in \mathcal{U} | r(u,i) \in \mathcal{R}\}| / |\mathcal{R}|$$

Set of all observed ratings

Toy example

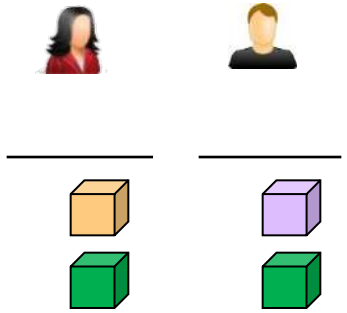
System's input

User-item interaction



e.g. $p(a|known) \sim 3/|\mathcal{R}| = 3/7$

System's recommendations

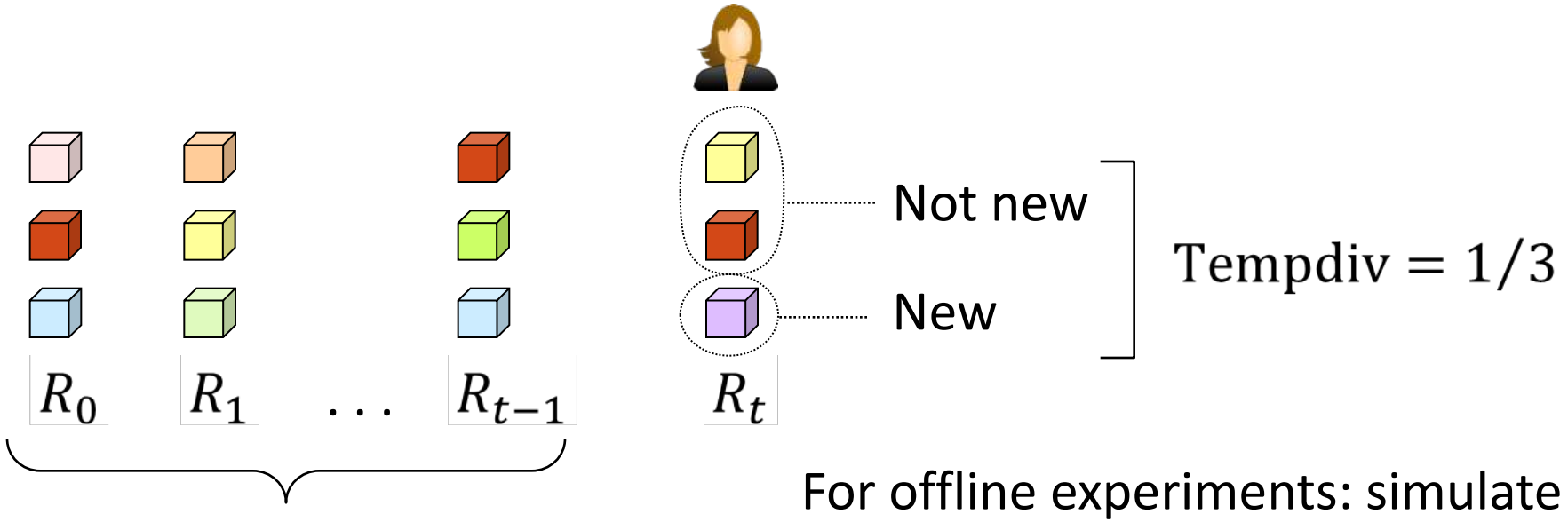


ILD	0	1
Unexp	1	0.5
MSI	**2.01**	**2.30**

Other metrics: temporal diversity
(Lathia et al SIGIR 2010)

Ratio of different recommended items over time

$$\text{Tempdiv} = \left|R_t - \cup_{t'<t} R_{t'}\right| / |R_t|$$



Unified perspective

Diversity

- Generally applies to a set of items
- How different the items are from one another?

Novelty

- Applies to an item, aggregated to sets
- Variants: unexpected, surprising, …

Summary

Diversity and novelty are important quality aspects

- As important in recommendation as in search

Search diversification principles can be applied

- Plus further particular motivation and techniques

Wide variety of metrics and methods

- Most can be unified in a common scheme

References

Diversity and novelty on the Web: search, recommendation, and data streaming aspects

Santos et al., SIGIR 2013, WSDM 2014, WWW 2015

Recommender Systems Handbook (Ch. 26)