

Search and Recommendation

Hao Sheng
August 9th, 2023



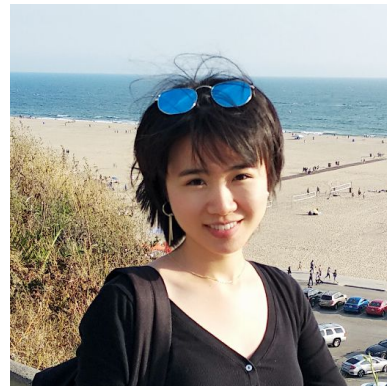
hello!

Hao Sheng

ICME Summer Workshop Instructor

Rui Yan

ICME Summer Workshop Assistant



Requirements

- **Laptop with networks**
- **Attention is all you need**

Agenda

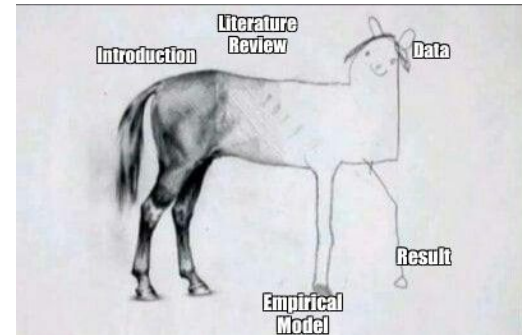
- **Introduction**
- **Lecturing:** The history of recommendation system
- **Break-out:** Recommendation system in daily life
- **Lecturing:** Recommendation as an ML problem
- **Lecturing:** Evaluate recommendation systems
- **Lab:** Recommendation system notebook I


60% Lecturing

25% Lab

10% Discussion

Introduction






[Events](#) [Tour](#) [About](#) [Help](#) [Join](#) [Log In](#)

[+ Provide Feedback](#)

Share housing. Save money. Meet cool people.
The fun, affordable alternative to hotels for an upcoming event.



Start by finding your event!

Event:

City:

When:

[Search →](#)

Host

Going to an event in your city?
Make money by turning your place into a DIY bed & breakfast. [» See how](#)

Attend

Traveling to another city for an event?
Save money by staying at another attendee's AirBed & Breakfast. [» See how](#)

Join!

Tour!

Airbeds!

Press Coverage

- Green Options
- Media Bistro
- Josh Spear

[» Read all](#)

"Think of it as Craigslist meets Hotels.com, but a lot less creepy."

JoshSpear.com

"Excellent idea!"

Swissmiss.com

[» Read all testimonials](#)

Featured Event:



Spread the Word:
Meet awesome people like you while making extra bucks! [Share this site now.](#)

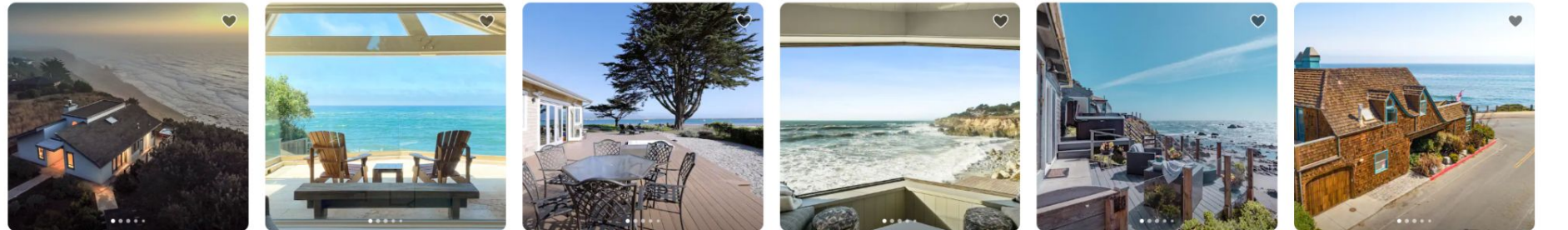
New Users:
Dsc05459 Photo_58 Andrew3 Jules_tap_144_Richard1 Deborah_chesky_web Elevator-icon Biopic2 Face Portrait2 Moi Bws

New Beds:
Rausch_004 Desk Ausapt4 Ausapt3 Ausapt2 Ausapt1 Ausapt4 Ausapt3 Ausapt2 Ausapt1 Ausapt4 Ausapt3

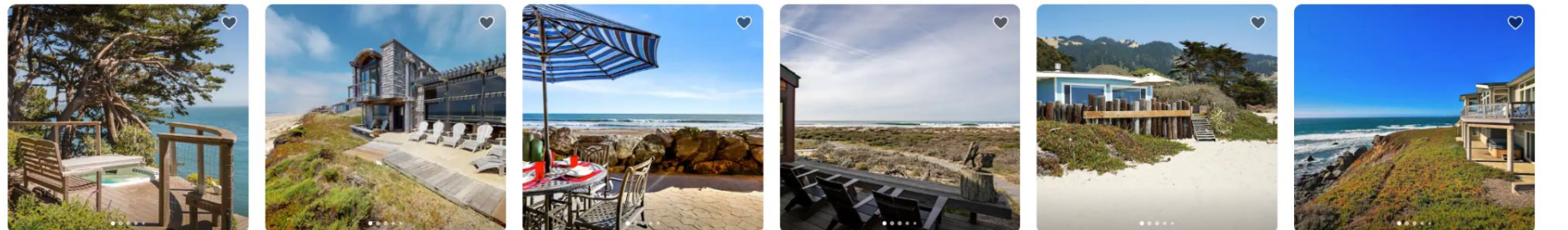
Introduction - Airbnb @ 2023

Anywhere Any week Add guests
Airbnb your home

Display total price | Includes all fees, before taxes



<p>Manchester, California ★ 4.96 Manchester State Park 7 nights · Jul 7 – 14 \$3,404 total before taxes</p>	<p>Moss Beach, California ★ 4.94 Moss Beach 5 nights · Aug 27 – Sep 1 \$7,286 total before taxes</p>	<p>Half Moon Bay, California ★ 4.89 Mavericks Beach 5 nights · Sep 4 – 9 \$4,855 total before taxes</p>	<p>Moss Beach, California ★ 4.85 Moss Beach 5 nights · Aug 1 – 6 \$23,965 total before taxes</p>	<p>Bodega Bay, California ★ 4.87 Scotty 5 nights · Oct 20 – 25 \$4,285 total before taxes</p>	<p>Santa Cruz, California ★ 4.89 Pleasure Point 5 nights · Aug 24 – 29 \$5,609 total before taxes</p>
---	--	---	--	---	---



<p>Muir Beach, California ★ 4.91</p>	<p>Watsonville, California ★ 4.84</p>	<p>Aptos, California ★ 4.91</p>	<p>Moss Landing, California ★ 4.96</p>	<p>Stinson Beach, California ★ 4.98</p>	<p>Dillon Beach, California ★ 5.0</p>
---	--	--	---	--	--

pizzahut.com



ORDER ONLINE

Online ordering is coming to you soon!

- Check to see if it's in your city...

ORDER ONLINE



click to register or login

STORE FINDER

Find your nearest Pizza Hut® store instantly by entering your zipcode below



enter zip and go

GET COUPONS

Enter your zip code below and receive the best deals from your neighborhood Pizza Hut®



enter zip and go

MORE PIZZA HUT®



click here for more on Pizza Hut®

click and go

Introduction - Pizza Hut @ 2023



[📍 Start here](#) Find your store to see local deals [FIND DEALS](#)

Featured [SEE MORE >](#)



75¢ Boneless Wings
Tossed in one of our nine signature sauces & rubs [>](#)



\$6.99 NEW Pizza Hut Melts
Crispy. Dippable. Loaded with toppings & cheese. [>](#)



\$9.99 Large 1-Topping Pizza
Our best delivery deal [>](#)

[Home](#) | [My Videos](#) | [My Favorites](#) | [My Messages](#) | [My Profile](#)

[Sign Up](#) | [Log In](#) | [Help](#)



Your Digital Video Repository

Search Videos

[Upload Your Videos](#)

[nansheng](#) : [azlan](#) : [wereldband](#) : [ny](#) : [superbike](#) : [japan](#) : [sinceretheory](#)
: [jozef](#) : [party](#) : [amazon](#) : [board](#) : [skate](#) : [buckley](#) : [shubs](#) : [falls](#) : [de](#) :
[stockshot](#) : [cubbyhole](#) : [burnout](#) : [satellite](#) : [poughkeepsie](#) : [cruise](#) : [heritage](#)
: [orgel](#) : [chin](#) : [themed](#) : [mill](#) : [music](#) : [new](#) : [live](#) : [to](#) : [farmer](#) : [mtv](#) :
[puenbrouck](#) : [sicily](#) : [fairfield](#) : [musical](#) : [coffeehouse](#) : [bud](#) :
[2005](#) : [trip](#) : [jfk](#) : [woordjes](#) : [death](#) : [xlanz](#) : [skill](#) : [olle](#) : [nature](#) : [ads](#) :
[dance](#) :

[See More Tags](#)

- ☰ YouTube
- 🏠 Home
- 🔥 Trending
- 📁 Subscriptions
- 📺 Library
- 🕒 History

Sign in to like videos, comment, and subscribe.

[SIGN IN](#)

BEST OF YOUTUBE

- 🎵 Music
- 🏆 Sports
- 🎮 Gaming
- 📰 News
- 📺 Live
- 👗 Fashion
- 🎓 Learning
- 📹 360° Video

[+ Browse channels](#)

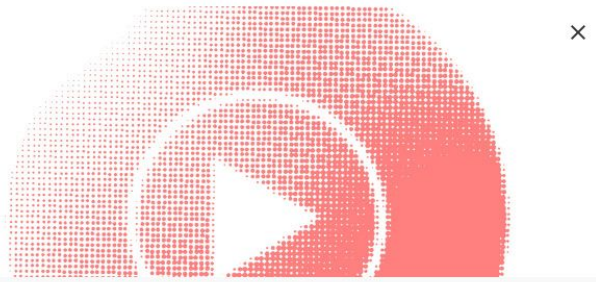
Search

[SIGN IN](#)

YouTube Music


Get family plan. YouTube Music ad-free and background play for up to 6 household members.


LET'S GO



Recommended

<p>SQUIRREL NINJA OBSTACLE COURSE</p> <p>21:40</p>	<p>UFC Free Fight: Khabib Nurmagomedov vs Conor...</p> <p>UFC - Ultimate Fighting... 4.3M views • 4 days ago</p>	<p>POLICE GOT MAD</p> <p>12:24</p>	<p>4K UHD GIANT ANTONOV AN-225 "Mriya" - Amazing...</p> <p>KNIGHT FLIGHT VIDEO 5.8M views • 3 months ago</p>
<p>Building the Perfect Squirrel Proof Bird Feeder</p> <p>Mark Rober 21M views • 1 week ago</p>	<p>I Built A Working Car Using Only LEGOS</p> <p>MrBeast 31M views • 1 year ago</p>		
<p>STREET MAGICIAN Does</p> <p>5:26</p>	<p>35 SIMPLE PAINTING</p> <p>15:09</p>	<p>Navy SEAL Jocko Willink</p> <p>17:34</p>	<p>Camilo - Favorito (Official)</p> <p>3:31</p>






STEAM

Steam delivers Valve's games to your desktop and connects you to a massive gaming community. Check out the full [feature list](#) now.

GET STEAM NOW!

NEWS
GET STEAM NOW
CYBER CAFÉS
SUPPORT
FORUMS
STATUS

GET THE GAMES



ORDER NOW >

Click [here](#) for more details.

HALF-LIFE 2 BRONZE

\$49.95

INCLUDES:

- Half-Life® 2*
- Counter-Strike™: Source™

With Steam installed
[CLICK HERE to PURCHASE](#)

Without Steam,
[CLICK HERE to sign up](#)

HALF-LIFE 2 SILVER

\$59.95

INCLUDES:

- Half-Life 2*
- Counter-Strike: Source
- Half-Life 1: Source*
- Day of Defeat™: Source*
- PLUS: Valve's back catalog available on Steam!

With Steam installed
[CLICK HERE to PURCHASE](#)

Without Steam,
[CLICK HERE to sign up](#)

HALF-LIFE 2 GOLD

\$89.95

INCLUDES:

- Half-Life 2*
- Counter-Strike:Source
- Half-Life 1: Source*
- Day of Defeat: Source*
- Valve's back catalog available on Steam.
- PLUS: HL2 posters, full strat guide, soundtrack, hat, collector's box, postcard & more

With Steam installed
[CLICK HERE to PURCHASE](#)

Without Steam,
[CLICK HERE to sign up](#)

Latest News

Half-Life 2 Pre-Loading Phase 6
The sixth phase of the Half-Life 2 preload has begun for Steam account holders. This will allow users to download the Half-Life 2 maps in encrypted form.

Half-Life 2 Steam Offers Ready
The Half-Life 2 Steam offers are now ready for purchase. Those who purchase via Steam, will receive the final version of Counter-Strike: Source immediately.

Technical Support

Questions, Answers, Etc...

Steam's support pages offer message boards, a list of frequently asked questions, and the Steam Troubleshooter to help identify and resolve any technical support issues

Cyber Café Licensing

Games Your Customers Want...

If you run a cyber café or gaming venue, Steam makes it easy for you to bring Valve's games to your customers. Over 1000 gaming venues have signed up for Valve's Cyber Café

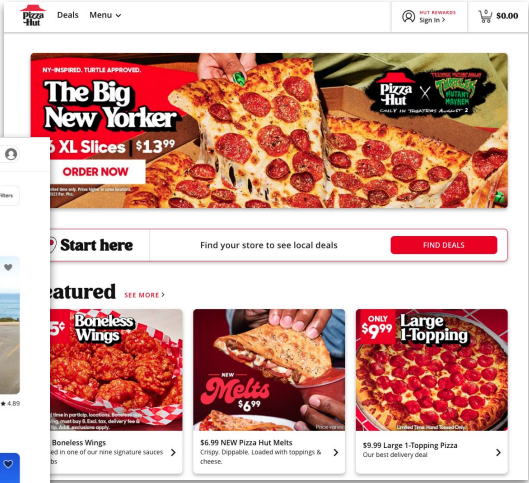
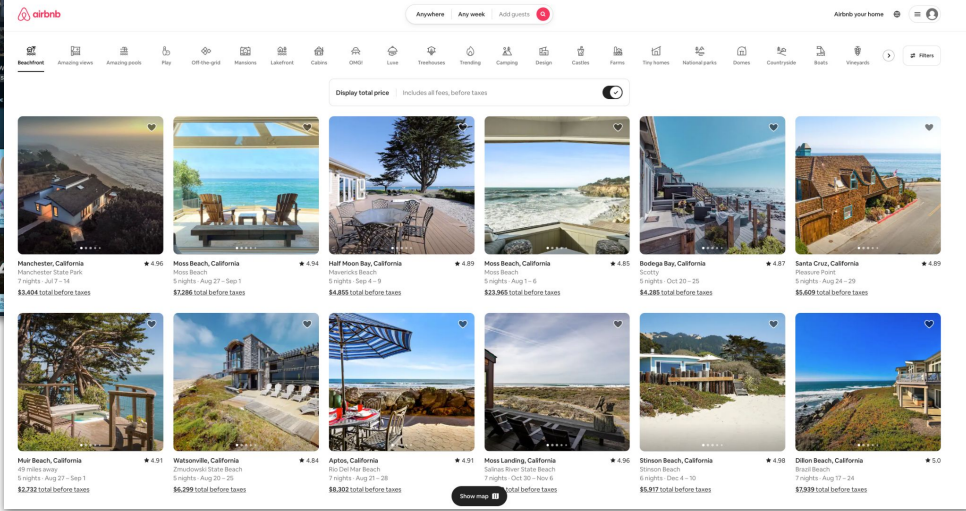
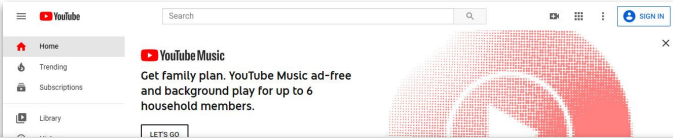
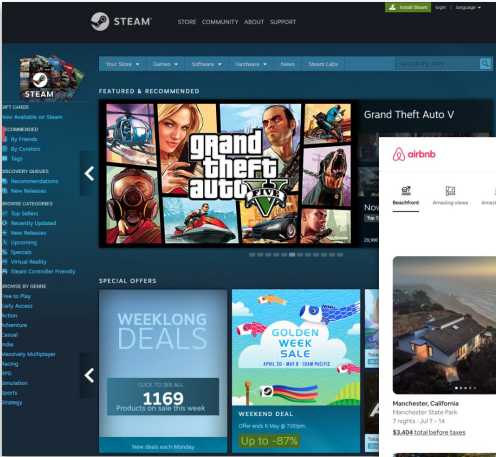
Get Steam Now

Sign Up and Play Games Today!

Start playing Valve's award-winning games within minutes. With Steam, you'll also get access to an instant messenger, automatic updates, and more. If you don't already have Steam.

The screenshot displays the Steam website's store page for Grand Theft Auto V. At the top, the Steam logo is on the left, and navigation links for 'STORE', 'COMMUNITY', 'ABOUT', and 'SUPPORT' are on the right. A utility bar includes 'Install Steam', 'login', and 'language' options. Below the navigation is a search bar and a category menu with options like 'Your Store', 'Games', 'Software', 'Hardware', 'News', and 'Steam Labs'. The main content area is titled 'FEATURED & RECOMMENDED' and prominently features Grand Theft Auto V. The game's cover art is shown in a collage, and a large text overlay reads 'Grand Theft Auto V'. To the right of the cover, the text 'Grand Theft Auto V' is displayed, followed by 'Now Available' and 'Top Seller' badges. The price is listed as 29,99€. Below this, there are smaller images of game scenes. At the bottom of the featured section, there are navigation arrows and a 'BROWSE MORE' button. The lower part of the page shows 'SPECIAL OFFERS' with a 'WEEKLONG DEALS' banner and a 'GOLDEN WEEK SALE' banner for 'METAVERS KEEPER' with a 37% discount. A left-hand sidebar contains various filters such as 'GIFT CARDS', 'RECOMMENDED', 'DISCOVERY QUEUES', 'BROWSE CATEGORIES', and 'BROWSE BY GENRE'.

Homepages @ 2023



Everyone gives user recommendations on the first impression!

- Every website gives user recommendations right on visiting.
- The recommendation is getting very “aggressive”?
 - 8 videos from 800 millions videos [1] on YouTube (used to recommend categories only)
 - 8-10 cities (and all in CA) from 100,000 cities [2] on AirBnb
 - GTA V from more than 50k games [3]
 - The Big New Yorker Pizza from 10 crusts x 21 toppings x 6 sources

[1] <https://www.globalmediainsight.com/blog/youtube-users-statistics>

[2] <https://www.searchlogistics.com/learn/statistics/airbnb-statistics>

[3] <https://backlinko.com/steam-users>

- Every websites gives user recommendations right on visiting.
- The recommendation is getting very “aggressive”?
 - 8 videos from 800 millions videos on YouTube (used to recommend categories only)
 - 8-10 cities (and all in CA) from 100,000 cities on AirBnb
 - GTA V from more than 50k games
 - The Big New Yorker Pizza from 10 crusts x 21 toppings x 6 sources
- What makes them so confident that it is a good suggestion?

The History of Recommendation System



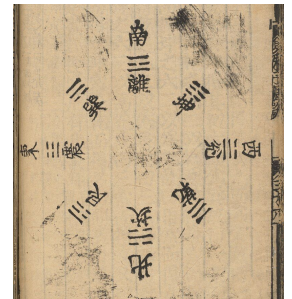
“I know of no people ... that does not consider that future things are indicated by signs and that it is possible ... to recognize those signs and predict what will happen.”

--- Marcus Tullius Cicero (44 BC)





Oracle bones (Chinese: 甲骨; pinyin: *jiǎgǔ*) are pieces of ox [scapula](#) and turtle [plastron](#), which were used for [pyromancy](#) – a form of [divination](#) – in ancient China, mainly during the late [Shang dynasty](#). [Scapulimancy](#) is the specific term if ox scapulae were used for the divination, [plastromancy](#) if turtle plastrons were used.



- Step 1: **Journey to Delphi**
- Step 2: **Preparation of the supplicant**
- Step 3: **Visit to the Oracle**
- Step 4: **Return home**



Divinetech: 450 BC - Today?



Aquarius
January 20 - February 18



Aries
March 21 - April 19



Cancer
June 22 - July 22



Capricorn
December 22 - January 19



Gemini
May 21 - June 21



Leo
July 23 - August 22



Libra
September 23 - October 22



Pisces
February 19 - March 20



Sagittarius
November 22 - December 21



Scorpio
October 23 - November 21



Taurus
April 20 - May 20

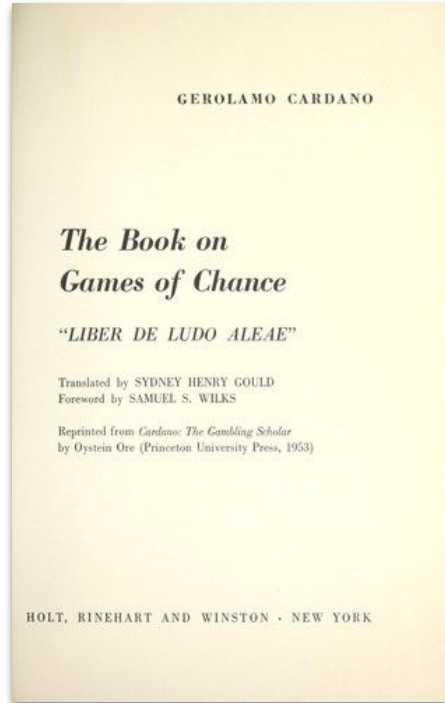


Virgo
August 23 - September 22

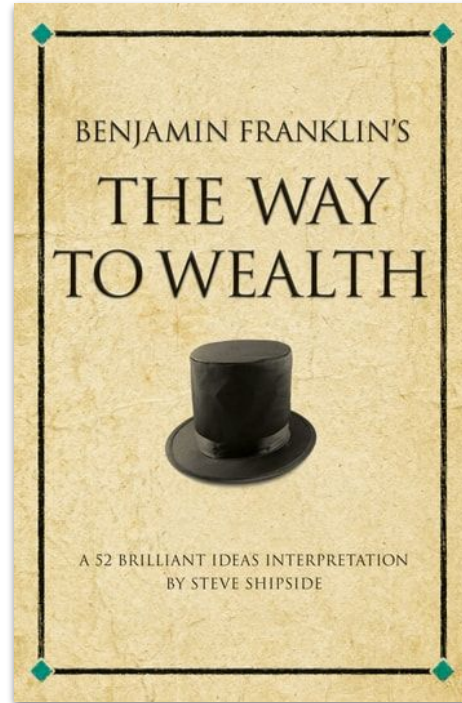
Aries March 21 - April 19

Today you should put good conversation at the top of your priority list! But in order to make it happen, you have to be ready to take things to a deeper level than you usually do with people you don't usually talk to. Small talk is for small minds right now, and since you certainly don't have one of those, why not prove it? Instead of asking about someone's weekend, ask them how they feel about international politics. You might get an odd look, but you'll also get great insight into another person.





Cardano, 1663



Benjamin Franklin, 1757



Samuel Smiles, 1859

- How to Win Friends and Influence People (1936)
- Think and Grow Rich (1937)
- ...
- Oracle of the Coffee House (1972)
- The 7 Habits of Highly Effective People (1989)
- Awaken the Giant (1992)
- Chicken Soup of the Soul (1993)
- ...
- The One Minute Manager (2001)
- A Guide to the Good Life (2009)
- Personal Development for Smart People (2009)

- People all over the world seek tools/techniques in their personal quest for actionable advice.
- Personalization and quantitative models are introduced.
- Self-help literature reflect the historical strand of recommendation: Curation.

“.. is a subclass of information filtering system that provide suggestions for items that are most pertinent to a particular user.” --- *Wikipedia*

“Recommender systems (RSs) are software tools and techniques that provide suggestions for items that are most likely of interest to a particular user. ” --- *Introduction to Recommender Systems Handbook*

“A recommender system can be described as a system which automatically selects personally relevant information for users based on their preferences.” --- *Intelligent and Relevant*

“.. is a subclass of information filtering system that provide suggestions for **items** that are most pertinent to a particular **user**.” --- *Wikipedia*

“Recommender systems (RSs) are software tools and techniques that provide suggestions for **items** that are most likely of interest to a particular **user**.” --- *Introduction to Recommender Systems Handbook*

“A recommender system can be described as a system which automatically selects personally relevant **information** for **users** based on their preferences.” --- *Intelligent and Relevant*

Recommendation System: Definition?

“.. is a subclass of **information filtering system** that provide suggestions for items that are most pertinent to a particular user.” --- *Wikipedia*

“Recommender systems (RSs) are **software tools and techniques** that provide suggestions for items that are most likely of interest to a particular user. ” --- *Introduction to Recommender Systems Handbook*

“A recommender system can be described as a system which **automatically** selects personally relevant information for users based on their preferences.” --- *Intelligent and Relevant*

First Recommendation System -- Grundy: 1979

COGNITIVE SCIENCE 3, 329-354 (1979)

User Modeling via Stereotypes *

ELAINE RICH

The University of Texas at Austin

This paper addresses the problems that must be considered if computers are going to treat their users as individuals with distinct personalities, goals, and so forth. It first outlines the issues, and then proposes stereotypes as a useful mechanism for building models of individual users on the basis of a small amount of information about them. In order to build user models quickly, a large amount of uncertain knowledge must be incorporated into the models. The issue of how to resolve the conflicts that will arise among such inferences is discussed. A system, Grundy, is described that builds models of its users, with the aid of stereotypes, and then exploits those models to guide it in its task, suggesting moves that people may find interesting. If stereotypes are to be useful to Grundy, they must accurately characterize the users of the system. Some techniques to modify stereotypes on the basis of experience are discussed. An analysis of Grundy's performance shows that its user models are effective in guiding its performance.

I. INTRODUCTION

Scene I

So someone walks into a large library, tells the librarian that he is interested in China, and asks for some books. What sort of books does the librarian recommend? That depends. Is the person a small child who just saw a TV show about China and wants to see more pictures of such an exotic place? Is the person a high school student doing a term paper? Or maybe a prospective tourist? Or a scholar interested in Eastern thought? Can the person read Chinese? The librarian needs to know these things before he can point the reader to the right books. Some of what he needs to know he'll know before he even thinks about it, such as the approximate age of the person. Some things he'll assume until he has evidence to the contrary, such as that the person does not read Chinese. To find out other things, he'll ask a few specific questions. Only after he has a rough model of the person he's talking to can he answer the question.

Scene II

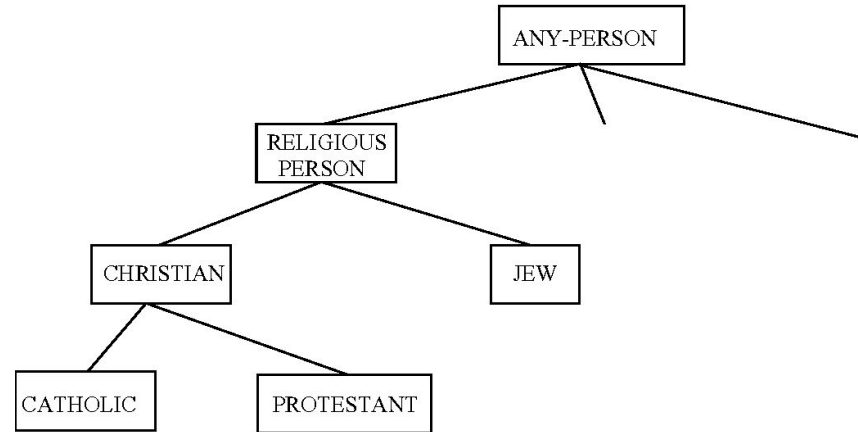
The phone rings in the information division of a large pharmaceutical firm. The caller wants information about a drug the company makes. What sort of information should be provided? That depends. Is the caller a doctor, a patient, or an FDA representative? To provide the right information, the person answering the phone needs to know some facts about the caller.

The scenes above illustrate some kinds of situations in which people need to form a *model* of the person with whom they are dealing before they can behave appropriately. They form their model by collecting a few specific pieces of information and then invoking the knowledge they have about the groups to which the current person belongs, such as scholar or medical patient.

As computers come to be used by a larger number of people to help perform a great variety of tasks, it is becoming more and more important for them to be easy for people to use. There are many factors that can contribute to the ease of use of a computer system, ranging from the good design of input devices such as terminals to the speed of the system's response, the appropriateness of its response, and the naturalness of its input and output languages. Appropriate models of the users of a system can be an important contribution because they can simultaneously affect several of those factors, such as speed and quality of response and habitability of the language interface.

Most systems that interact with human users contain, even if only implicitly, some sort of model of the creature they will be dealing with. For example, the central assumption behind the *mini-max* strategy used by game playing programs is that the opponent is trying to win and will therefore make his best possible move. Although it is almost always valid to assume that the opponent wants to win, it is much less often true that he will therefore make the best move. He may, and probably does, have idiosyncrasies of style or strategy that preclude that. Of course, human players know that and watch for evidence of such quirks in their opponents.

The term "user model" can be used to mean several different things. The three major dimensions along which user models can be classified are:



Rich, Elaine. "User modeling via stereotypes." *Cognitive science* 3.4 (1979): 329-354.

First Recommendation System Used: 1992, Palo Alto

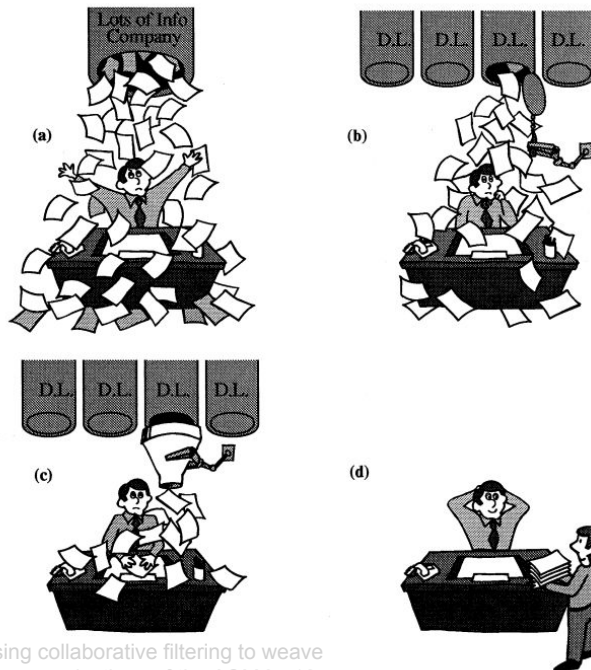
Palo Alto Research Center

Using Collaborative Filtering to Weave an Information Tapestry

David Goldberg, David Nichols, Brian Oki, and Douglas Terry

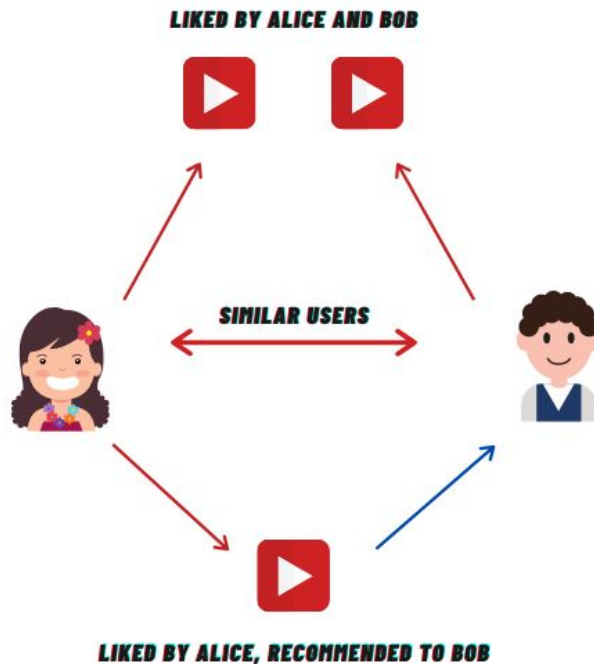
XEROX

Figure 1.
 (a) electronic mail overload
 (b) using distribution lists
 (c) conventional filtering
 (d) collaborative filtering



Goldberg, David, et al. "Using collaborative filtering to weave an information tapestry." *Communications of the ACM* 35.12 (1992): 61-70.

COLLABORATIVE FILTERING



Active Messages								
MoveTo	Display	Delete	AddTo	NewMail	Places	Levels	MsgOps	SortBy
	99>	18 Mar 93	Maria+Eblin...	Re: marked-up thesis proposal				
	99>	18 Mar 93	elib@eclipse...	Elib Response				
	99>	19 Mar 93	Dan Swineh...	Here's a topic for a future meeting				
	94>	18 Mar 93	To: elib%cs....	Test of Stanford SDI Service				
	85>	13 Mar 93	tapestry@pa...	ICDCS'13 program				
	70>	12 Mar 93	weiser:PAR...	here is a new service being offered by the stanford CS library				
?	99>	22 Mar 93	weiser:PAR...	<i>hiring meeting after dealer</i>				
?	99>	22 Mar 93	Bellotti@eur...	<i>Re: UbiCore meeting on Security</i>				
?	99>	22 Mar 93	chauser:PA...	<i>Request for more help in tapbrowser</i>				
?	90>	22 Mar 93	Nancy Frei...	<i>HIRING MTG. SCHEDULED...</i>				
?	90>	22 Mar 93	weiser:PAR...	<i>Re: IMPORTANT AGAIN: YES hiring meeting after Dealer...</i>				
?	85>	22 Mar 93	saraswat@pa...	<i>[arjun@cs.stanford.edu: Two talks in Concurrency Theory]</i>				
?	85>	21 Mar 93	tapestry@pa...	<i>Briefs California (Mar 17 11 am PST)</i>				
?	80>	22 Mar 93	Gary Eman...	<i>Equipment Shipping crates and boxes in warehouse</i>				
?	80>	22 Mar 93	Andrea Spi...	<i>Fry's Electronics Blanket Order (PARC Only)</i>				
?	75>	21 Mar 93	tapestry@pa...	<i>****OnLine Bookstore Offers "Electronic Lit" over Internet 03/19/93</i>				

Figure 1. My Active folder showing old messages followed by prioritized new mail.

MoveTo	Display	Delete	AddTo	MessageOps	gOps: SortBy
70>	12 Mar 93		weiser:PARC	Sender	g offered by
				UndoLastDelete	
				Size Of MsgSet	
? 90>	22 Mar 93		Nancy Fre...	Msg Info	...
? 90>	22 Mar 93		weiser:PARC	Append Msg	YES
				Interpress3.0PrintMsgSet	o talks in
? 85>	22 Mar 93		saraswat@p...	Interpress3.0PrintSelectedMsg	
				Interpress3.0PrintTOC	1 am PST)
? 80>	22 Mar 93		Gary Emar...	R.O.T. MsgSet	and boxes in
				Explain Msg's Priority	Order (PARC
? 80>	22 Mar 93		Andrea Spa...	Reappraise Selected Msg(s)	
? 75>	21 Mar 93		tapestry@p...	Reappraise MsgSet	rs "Electronic

Lit Over Internet 03/19/93

Figure 2. Requesting an explanation for a message's priority.

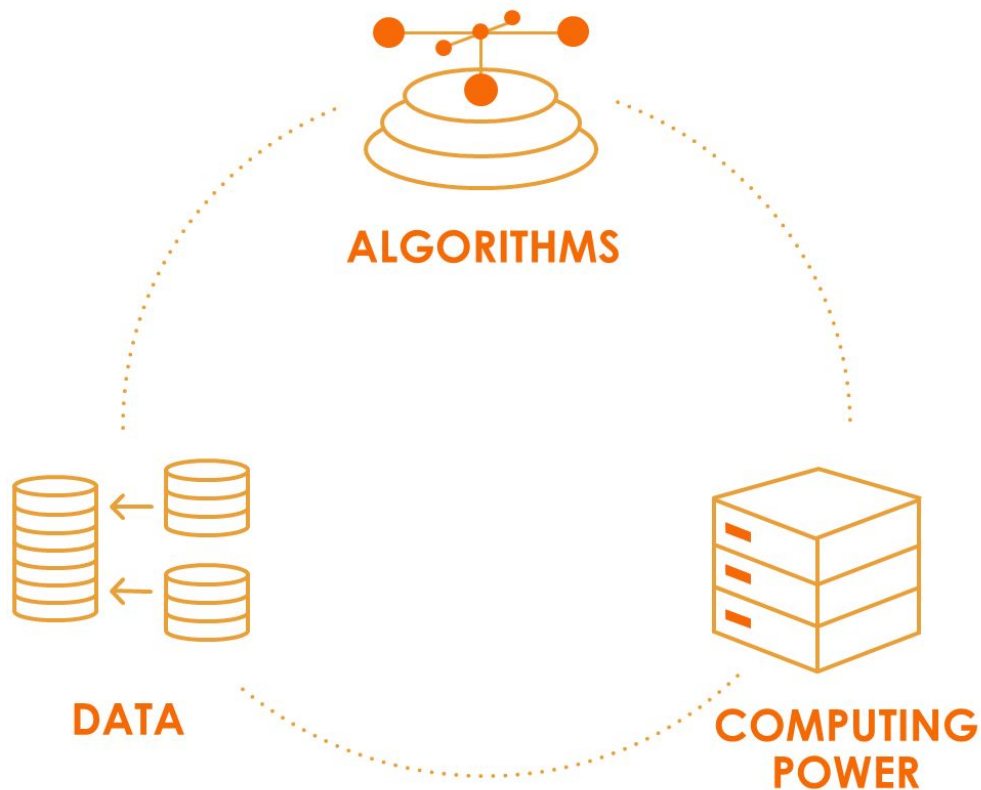
```

Annotations for message $ XNS-SMTP-Gateway:Parc:Xerox
appraiser terry$text:Bakersfield => priority 85
appraiser terry$$Subject:Briefs<California => priority 55
appraiser terry$sender:tapestry => priority 10
    
```

Figure 3. An explanation of priorities assigned to a message by various appraisers.

- Recommendation system: RS automatically selects personalized information based on users' preferences.
- Grundy:
 - Ask user questions and assign stereotype.
 - **Content-based filtering.**
- Tapestry:
 - Find similar users and recommend their choices.
 - **Collaborative filtering.**

Backbone of Recommendation System: Not Only the Algorithm



Take a 10 minute break



At Your Service: Coffee Beans Recommendation From a Robot Assistant

Jacopo de Berardinis^{*}, Gabriella Pizzuto¹, Francesco Lanza¹, Antonio Chella[‡], Jorge Meira[§], Angelo Cangelosi^{*}

^{*}School of Engineering, The University of Manchester

¹School of Informatics, The University of Edinburgh

[‡]Department of Engineering, University of Palermo

[§]Department of Computer Science and Information Technologies, University of A Corua

Recommendation System in Daily Life



Our next recommendation
is a 25 minute video on the
history of Parmesan
cheese.

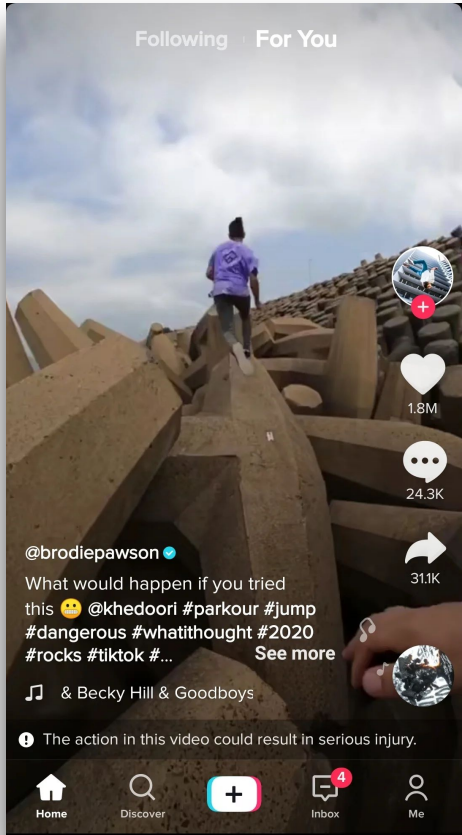


Yes.

- Q: What is the input/outputs of the system?
- Q: Why is this useful?
- Q: Can you guess what's behind the system?



Following · For You



1.8M



24.3K



31.1K

@brodiepawson

What would happen if you tried this 😊 @khedoori #parkour #jump #dangerous #whatithought #2020 #rocks #tiktok #...

See more

🎵 & Becky Hill & Goodboys

⚠️ The action in this video could result in serious injury.



Home



Discover



+

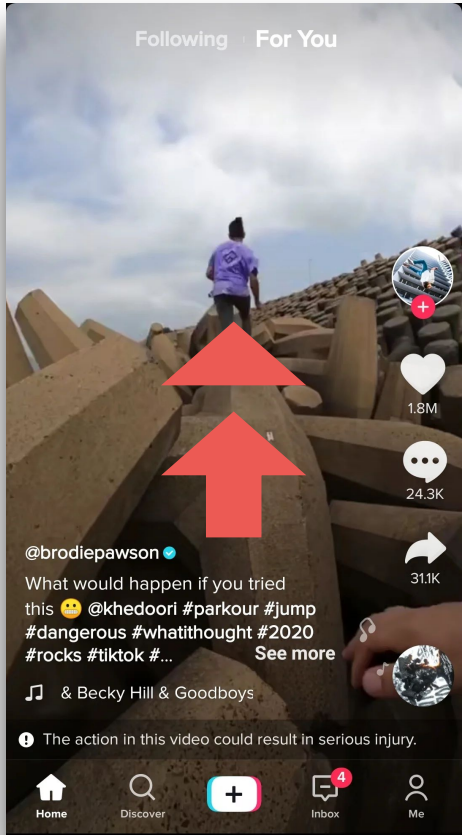


Inbox



Me

Following · For You



1.8M



24.3K



31.1K

@brodiepawson

What would happen if you tried this 😊 @khedoori #parkour #jump #dangerous #whatithought #2020 #rocks #tiktok #... See more

🎵 & Becky Hill & Goodboys

⚠️ The action in this video could result in serious injury.



Home



Discover

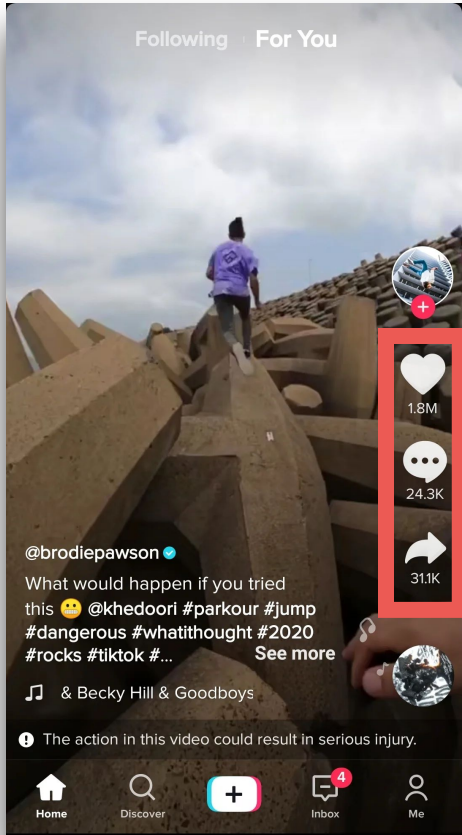





Inbox



Me

Following · For You



-  1.8M
-  24.3K
-  31.1K

@brodiepawson

What would happen if you tried this 😊 @khedoori #parkour #jump #dangerous #whatithought #2020 #rocks #tiktok #... See more

🎵 & Becky Hill & Goodboys

⚠️ The action in this video could result in serious injury.









- **Q: What is the input/outputs of the system?**
 - Input: Historical user behavior (e.g. like/comment/share or not; How long did I stay; Did I finish the video)
 - Outputs: Next short video I would like to watch. Or some Ads that I have high chance to spend my money on.
- **Q: Why is this useful?**
 - Well, it helps to.. kill my time more effectively(?). I don't need to search for videos I am interested in. Moreover, I explored my interests in a way I could never think about.
- **Q: Can you guess what's behind the system?**
 - Hmm. Maybe just like the Grundy example? They assigned me a “stereotype”?



Recommendation as an ML Problem

- How to model recommendation problem mathematically.
- Classic recommendation algorithms
 - Collaborative filtering
 - Memory-based filtering
 - Model-based filtering
 - *Content-based filtering
- How to evaluate recommendation modeling.

Recommendation System an ML Problem

						
	5		1	1		2
		2		4		4
	4	5		1	1	2
			3	5	2	
	2		1		4	4



prediction: $r_{u,i}$



recommend top-N items

Recommendation System as an ML Problem

	1	2	3	4	5	6
a	+	?	-	-	?	-
b		-		+		+
c	+	+		-	-	-
d			+	+	-	
e	-		-		+	+



prediction: $r_{u,i}$



recommend top-N items

Recommendation System as an ML Problem

	1	2	3	4	5	6
a	+	?	-	-	?	-
b		-		+		+
c	+	+		-	-	-
d			+	+	-	
e	-		-		+	+

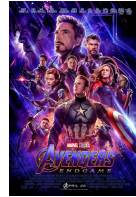


prediction: $r_{u,i}$



recommend top-N items

Recommendation System an ML Problem



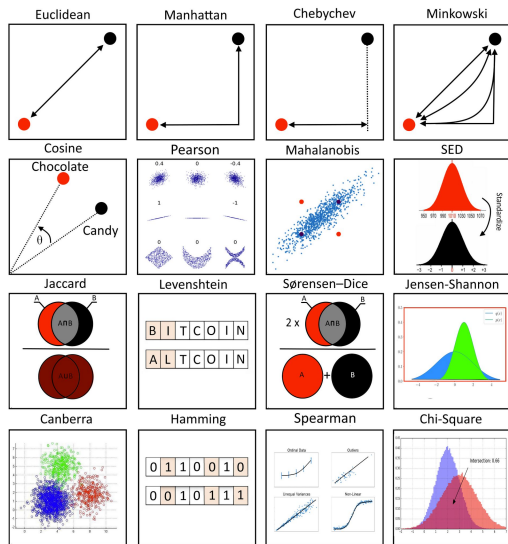
Sara	5	3		2	2	2
Jesper	4	3	4		3	3
Therese	5	2	5	2	1	1
Helle	3	5	3		1	1
Pietro	3	3	3	2	4	5
Ekaterina	2	3	2	3	5	5



- How to model recommendation problem mathematically.
- Classic recommendation algorithms
 - Collaborative filtering
 - Memory-based filtering
 - Model-based filtering
 - *Content-based filtering

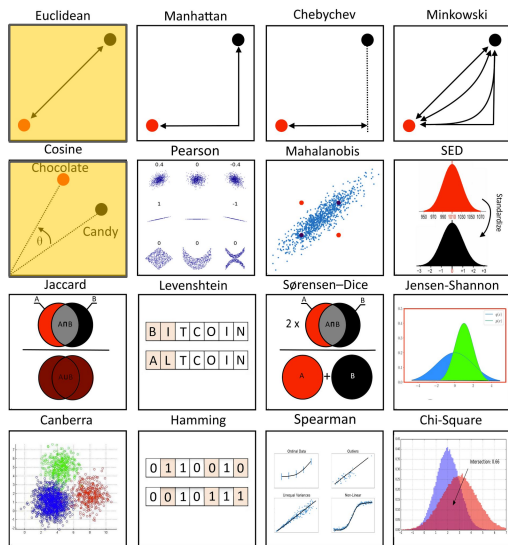
Collaborative Filtering: Similarity Function

You can calculate similarity in many ways, but the overall problem can be defined as follows: Given two items, i_1 and i_2 , the similarity between them is given by the function $\text{sim}(i_1, i_2)$.

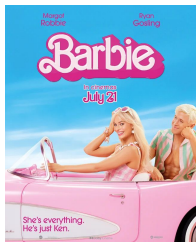


Collaborative Filtering: Similarity Function

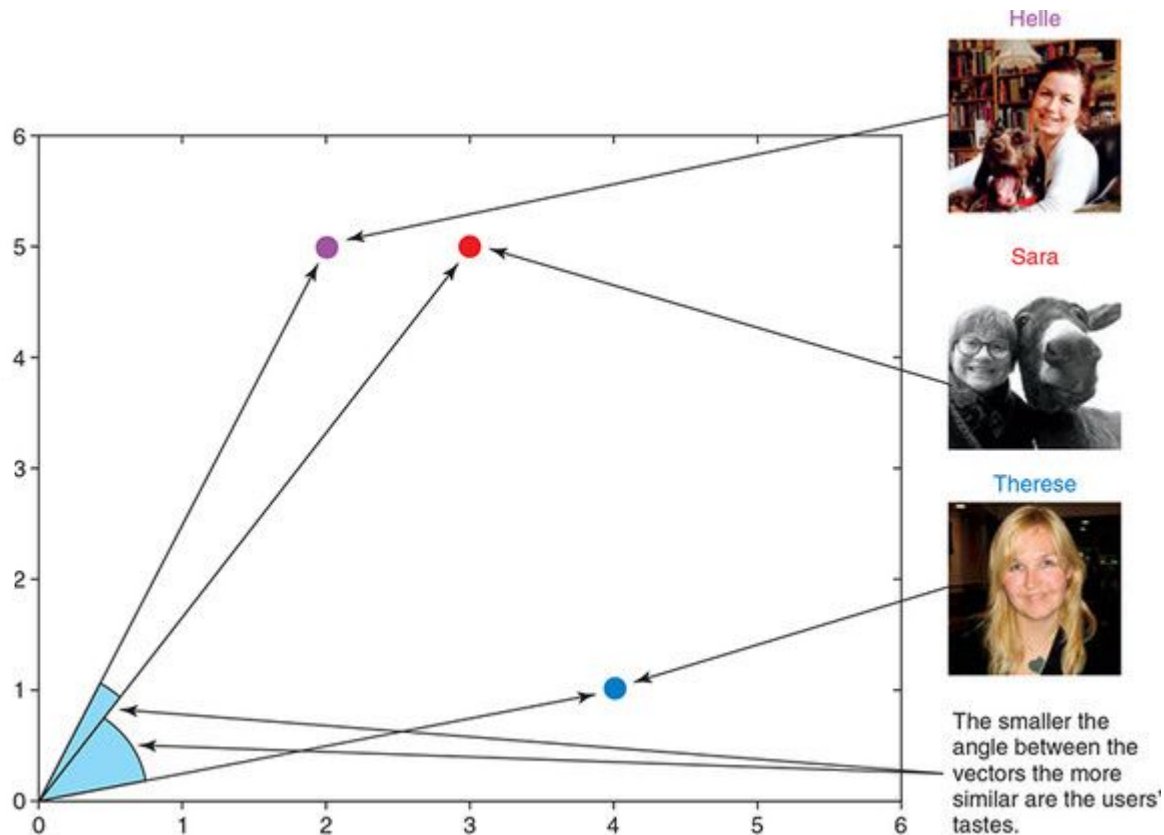
You can calculate similarity in many ways, but the overall problem can be defined as follows: Given two items, i_1 and i_2 , the similarity between them is given by the function $\text{sim}(i_1, i_2)$.



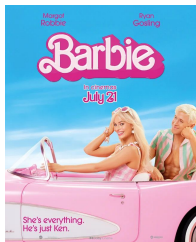
Collaborative Filtering: Cosine Similarity



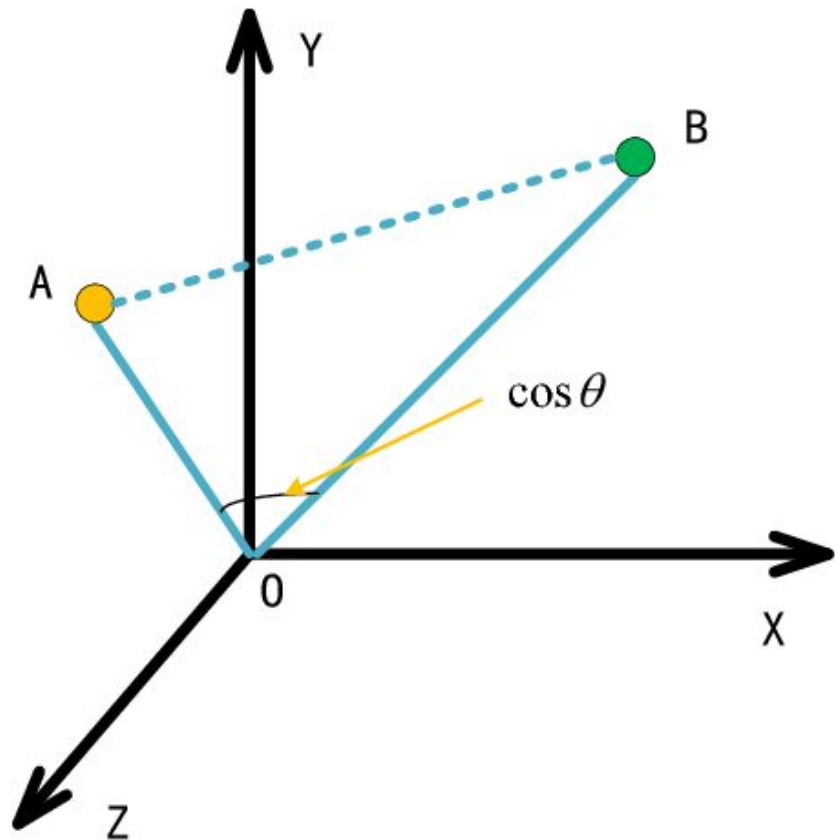
	Barbie	Oppenheimer
Helle	5	2
Sara	5	3
Therese	1	4

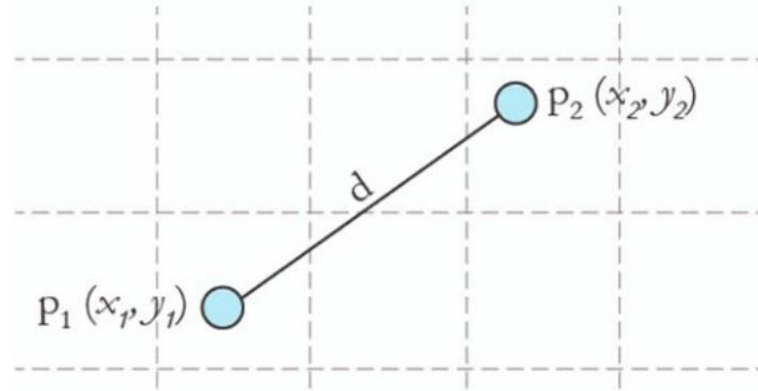


Collaborative Filtering: Cosine Similarity



	Barbie	Oppenheimer	Spider-Man: Across the Spider-Verse
Helle	5	2	3
Sara	5	3	4
Therese	1	4	3





$$\text{Euclidean distance (d)} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

What if we have 20 million active users?

Calculate similarities of current user \leftrightarrow 20 million users?

What if we have 20 million active users?

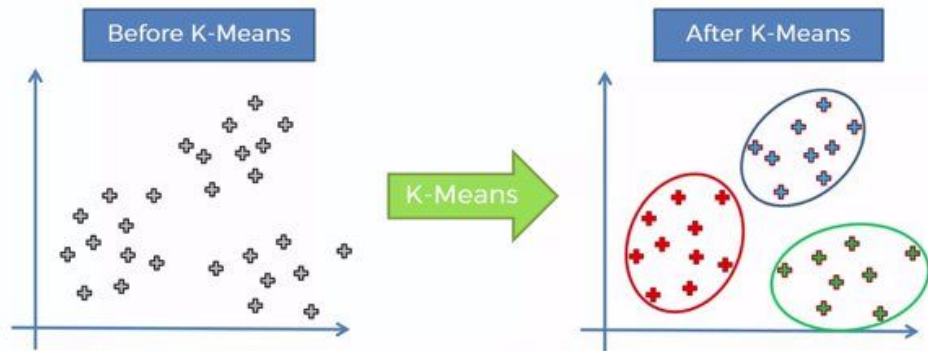
Calculate similarities of current user \leftrightarrow 20 million users?



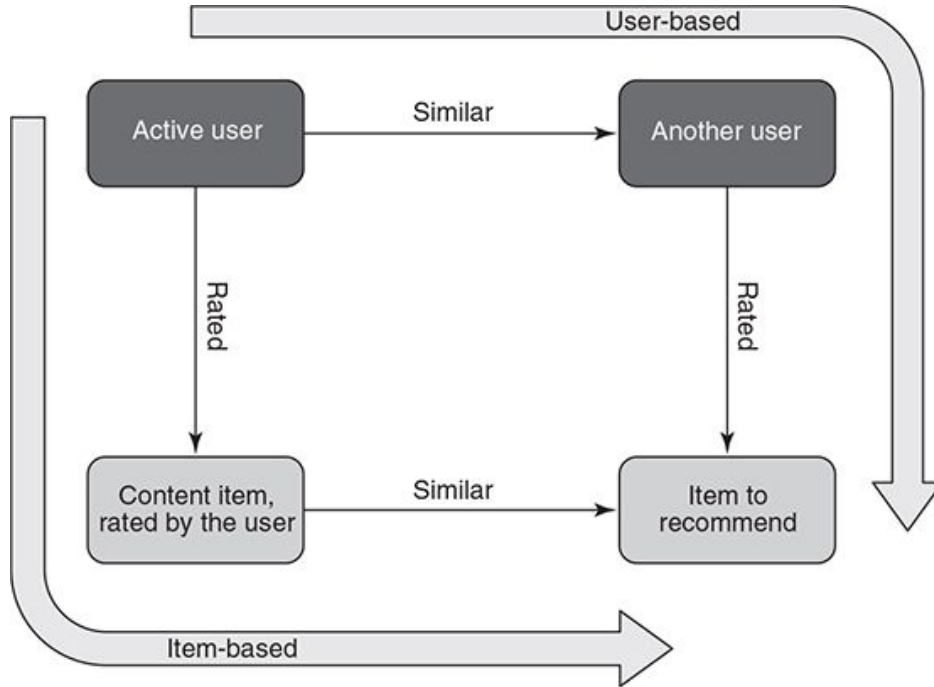
Collaborative Filtering: Scale-up Similarity Calculation

What if we have 20 million active users?

Calculate similarities of current user \leftrightarrow 20 million users?



Collaborative Filtering: Memory-based



- How to model recommendation problem mathematically.
- Classic recommendation algorithms
 - Collaborative filtering
 - Memory-based filtering
 - Model-based filtering
 - *Content-based filtering

Collaborative Filtering: Motivation of Model-based Filtering

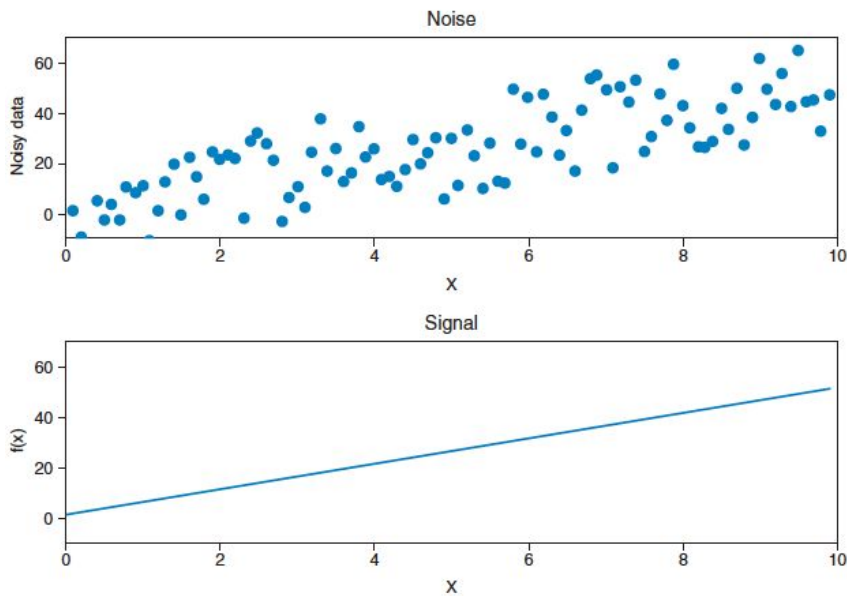


Figure 11.1 A scatter plot of noisy data (top) and the signals that uncover the information in the data (bottom)

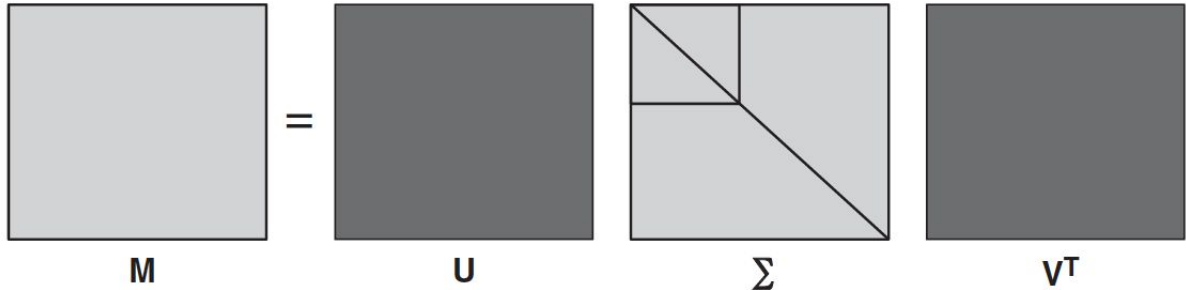
Collaborative Filtering: Factorization



Sara	5	3		2	2	2
Jesper	4	3	4		3	3
Therese	5	2	5	2	1	1
Helle	3	5	3		1	1
Pietro	3	3	3	2	4	5
Ekaterina	2	3	2	3	5	5

$$R = \begin{bmatrix} 5 & 3 & 0 & 2 & 2 & 2 \\ 4 & 3 & 4 & 0 & 3 & 3 \\ 5 & 2 & 5 & 2 & 1 & 1 \\ 3 & 5 & 3 & 0 & 1 & 1 \\ 3 & 3 & 3 & 2 & 4 & 5 \\ 2 & 3 & 2 & 3 & 5 & 5 \end{bmatrix}$$

Collaborative Filtering: SVD Factorization



$$\begin{bmatrix}
 5 & 3 & 0 & 2 & 2 & 2 \\
 4 & 3 & 4 & 0 & 3 & 3 \\
 5 & 2 & 5 & 2 & 1 & 1 \\
 3 & 5 & 3 & 0 & 1 & 1 \\
 3 & 3 & 3 & 2 & 4 & 5 \\
 2 & 3 & 2 & 3 & 5 & 5
 \end{bmatrix}$$

=

-0.34	0.05	0.91	0.11	0.19	-0.00	17.27	0	0	0	0	0	-0.50	-0.44	-0.41	-0.22	-0.40	-0.43
-0.43	0.16	-0.31	-0.12	0.74	0.35	0	5.84	0	0	0	0	0.46	0.17	0.42	-0.22	-0.49	-0.55
-0.39	0.56	-0.19	0.63	-0.32	0.02	0	0	3.56	0	0	0	0.50	0.22	-0.78	0.26	-0.08	-0.13
-0.33	0.42	0.02	-0.76	-0.37	-0.05	0	0	0	3.13	0	0	0.34	-0.77	0.17	0.51	-0.02	-0.01
-0.48	-0.34	-0.18	0.03	0.10	-0.78	0	0	0	0	1.67	0	0.41	-0.36	-0.16	-0.76	0.19	0.25
-0.46	-0.61	-0.06	0.02	-0.40	0.51	0	0	0	0	0	0.56	-0.01	-0.03	0.01	-0.02	0.75	-0.66
U						Σ						V^t					

Collaborative Filtering: SVD Factorization

```
def rank_k(k):
    U_reduced= np.mat(U[:, :k])
    Vt_reduced = np.mat(Vt[:k, :])
    Sigma_reduced = Sigma_reduced = np.eye(k)*Sigma[:k]
```

Returns the reduced matrices

```
    return U_reduced, Sigma_reduced, Vt_reduced,
U_reduced, Sigma_reduced, Vt_reduced = rank_k(4)
M_hat = U_reduced * Sigma_reduced * Vt_reduced
```

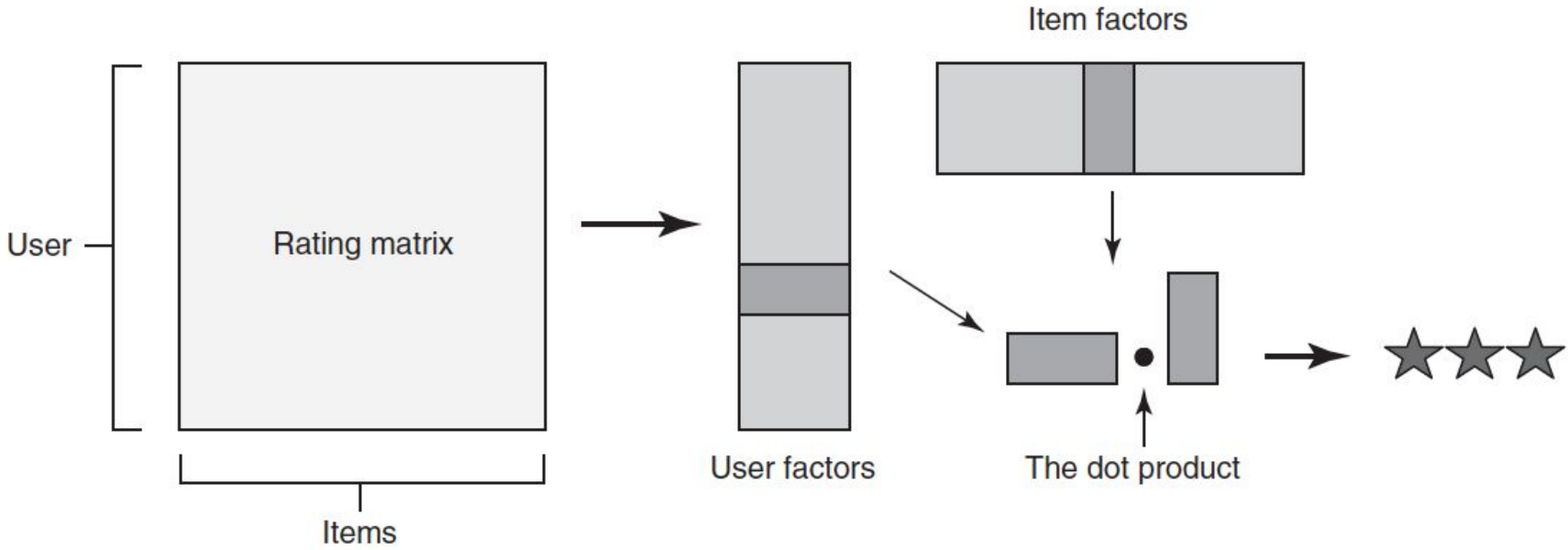
Uses rank_k to return the reduced matrices

Calculates the deduced matrix M_hat

$$R = \begin{bmatrix} 5 & 3 & 0 & 2 & 2 & 2 \\ 4 & 3 & 4 & 0 & 3 & 3 \\ 5 & 2 & 5 & 2 & 1 & 1 \\ 3 & 5 & 3 & 0 & 1 & 1 \\ 3 & 3 & 3 & 2 & 4 & 5 \\ 2 & 3 & 2 & 3 & 5 & 5 \end{bmatrix}$$

Sara	4.87	3.11	0.05	2.24	1.94	1.92
Jesper	3.49	3.46	4.19	0.95	2.62	2.82
Therese	5.22	1.80	4.92	1.59	1.10	1.14
Helle	3.25	4.77	2.90	-0.47	1.14	1.13
Pietro	2.93	3.05	3.03	2.11	4.30	4.67
Ekaterina	2.27	2.77	1.89	2.50	4.92	5.35

Collaborative Filtering: Model-based



$$\begin{bmatrix} 5 & 3 & 0 & 2 & 2 & 2 \\ 4 & 3 & 4 & 0 & 3 & 3 \\ 5 & 2 & 5 & 2 & 1 & 1 \\ 3 & 5 & 3 & 0 & 1 & 1 \\ 3 & 3 & 3 & 2 & 4 & 5 \\ 2 & 3 & 2 & 3 & 5 & 5 \end{bmatrix} = \begin{bmatrix} u_{1,1} & u_{1,2} \\ u_{2,1} & u_{2,2} \\ u_{3,1} & u_{3,2} \\ u_{4,1} & u_{4,2} \\ u_{5,1} & u_{5,2} \\ u_{6,1} & u_{6,2} \end{bmatrix} \begin{bmatrix} v_{1,1} & v_{1,2} & v_{1,3} & v_{1,4} & v_{1,5} & v_{1,6} \\ v_{2,1} & v_{2,2} & v_{2,3} & v_{2,4} & v_{2,5} & v_{2,6} \end{bmatrix}$$

$$RMSE = \sqrt{\frac{1}{|known|} \sum_{(u,i) \in known} (r_{ui} - u_u v_i)^2}$$

Collaborative Filtering: More about Funk



Netflix provided a *training* data set of 100,480,507 ratings that 480,189 users gave to 17,770 movies



[<< | Prev | Index | Next | >>]

Monday, December 11, 2006

Netflix Update: Try This at Home



At one point Simon Funk was #3 on the list. However, Simon is an independent software developer who works on Netflix prize in his spare time between his trips around New Zealand! He freely published his code and ideas – the first top leader to do so!

More to read:

<https://www.thrillist.com/entertainment/nation/the-netflix-prize>

<https://sifter.org/simon/journal/20061211.html>

https://www.kdd.org/exploration_files/simon-funk-explorations.pdf

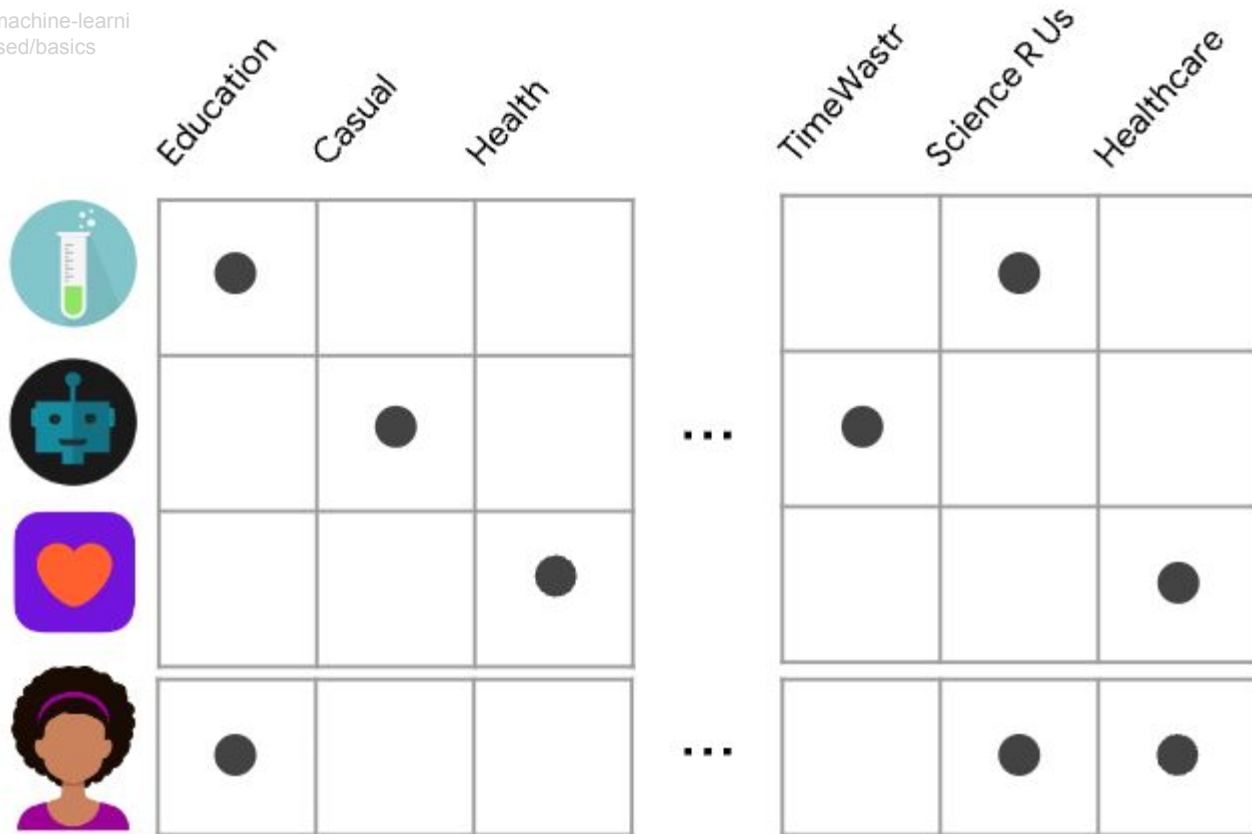
- No domain knowledge necessary
- Exploratory

- Sparsity
- Side features
- Cold start
- Gray sheep
- Not using popularity

- How to model recommendation problem mathematically.
- Classic recommendation algorithms
 - Collaborative filtering
 - Memory-based filtering
 - Model-based filtering
 - *Content-based filtering

Content-based Filtering

Example from:
<https://developers.google.com/machine-learning/recommendation/content-based/basics>

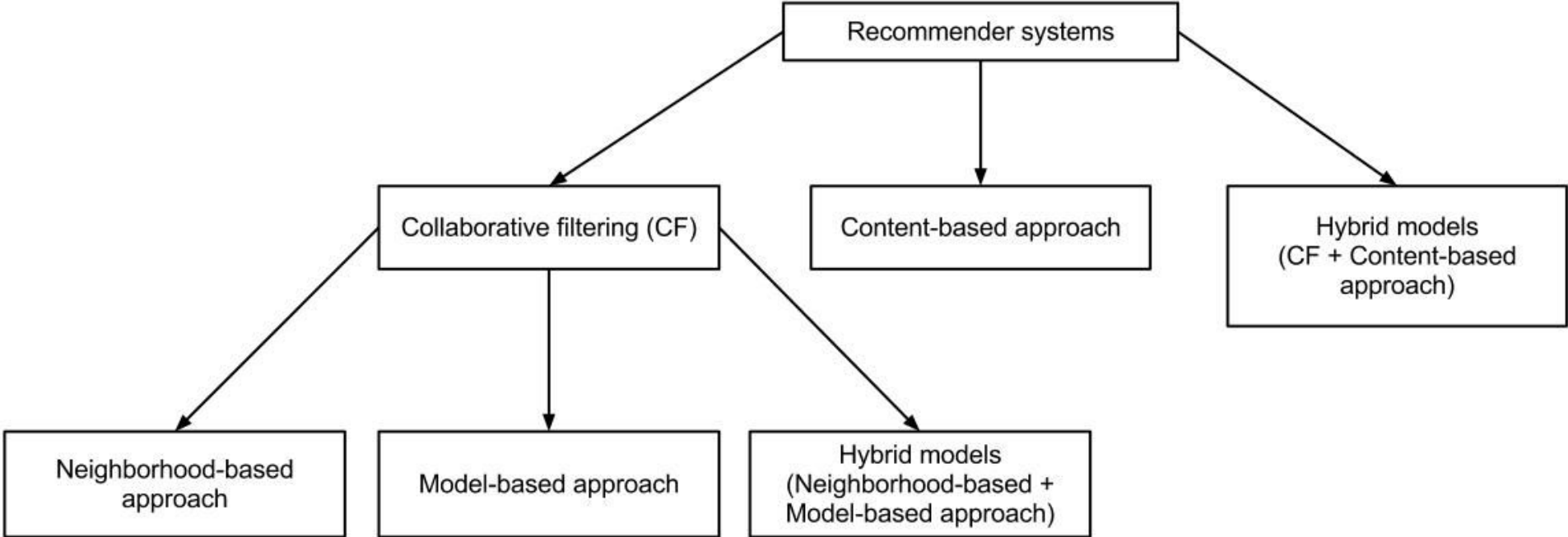


Content-based Filtering: Pros & Cons

- Easy to scale
- Explainable

- Hand-engineered features
- Not exploratory

Recommendation System an ML Problem



Evaluate Recommendation Systems

How likely are you to recommend Windows 10 to a friend or colleague?

1 2 3 4 5

Not at all likely

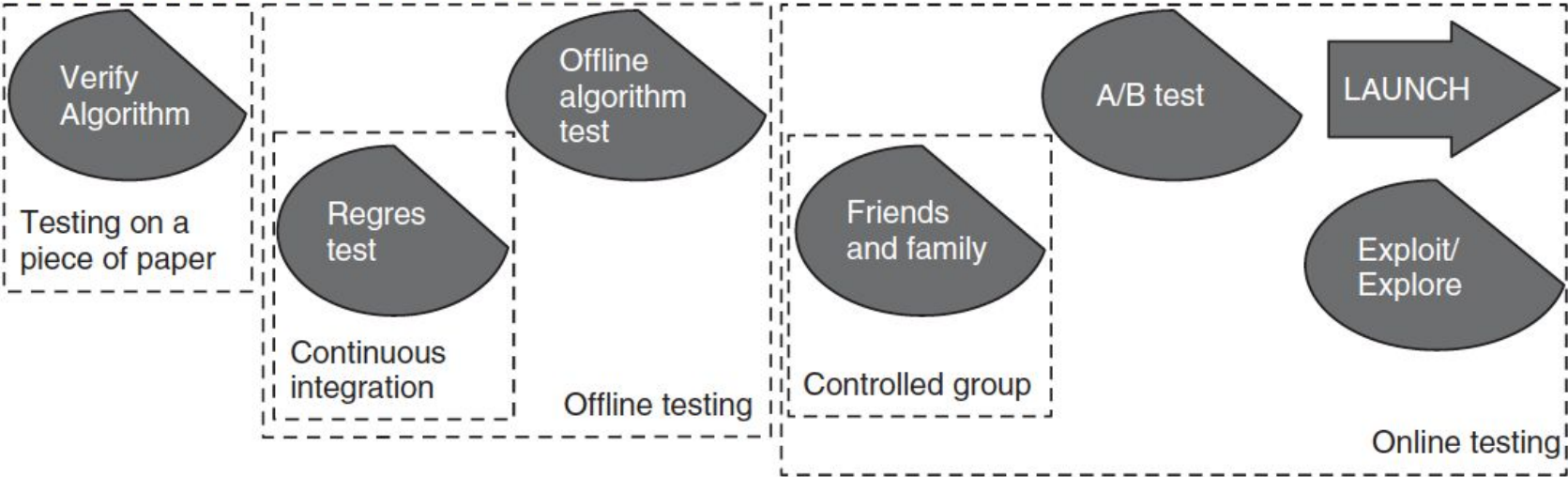
Extremely likely

Please explain why you gave this score.

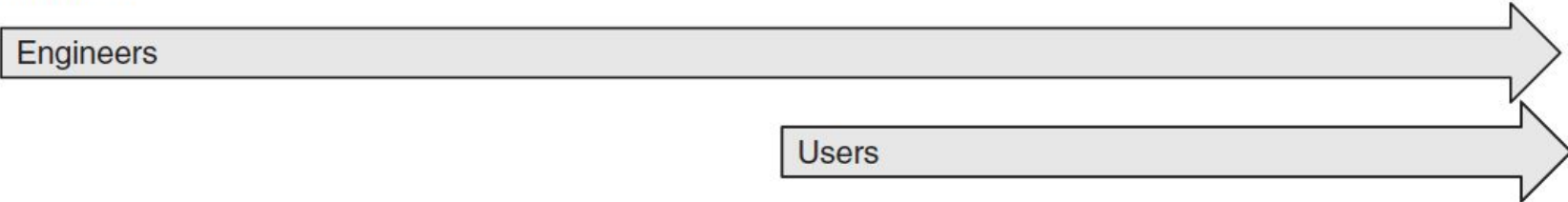
I need you to understand that people don't have conversations where they randomly recommend operating systems to one another

Evaluate Recommendation Systems

Recommender algorithm evaluation:



Involved:



- Goals and Metrics
- Offline Evaluation
- Online Evaluation

Evaluate Recommendation Systems: Goals

- Accuracy
- Diversity and coverage
- Serendipity
- Scalability

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

MSE = mean squared error

n = number of data points

Y_i = observed values

\hat{Y}_i = predicted values

$$\text{MAE} = \frac{\sum_{i=1}^n |y_i - x_i|}{n}$$

MAE = mean absolute error

y_i = prediction

x_i = true value

n = total number of data points

Evaluate Recommendation Systems: Goals

- Accuracy
- Diversity and coverage
- Serendipity
- Scalability

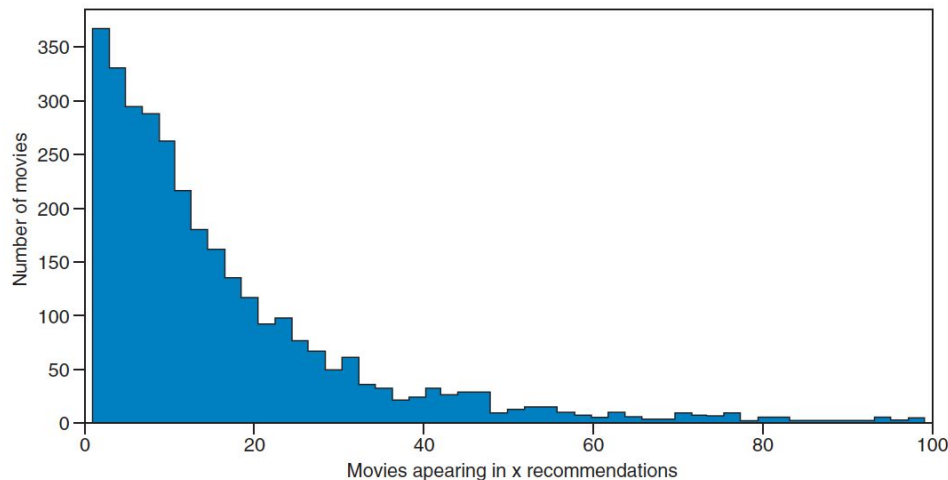
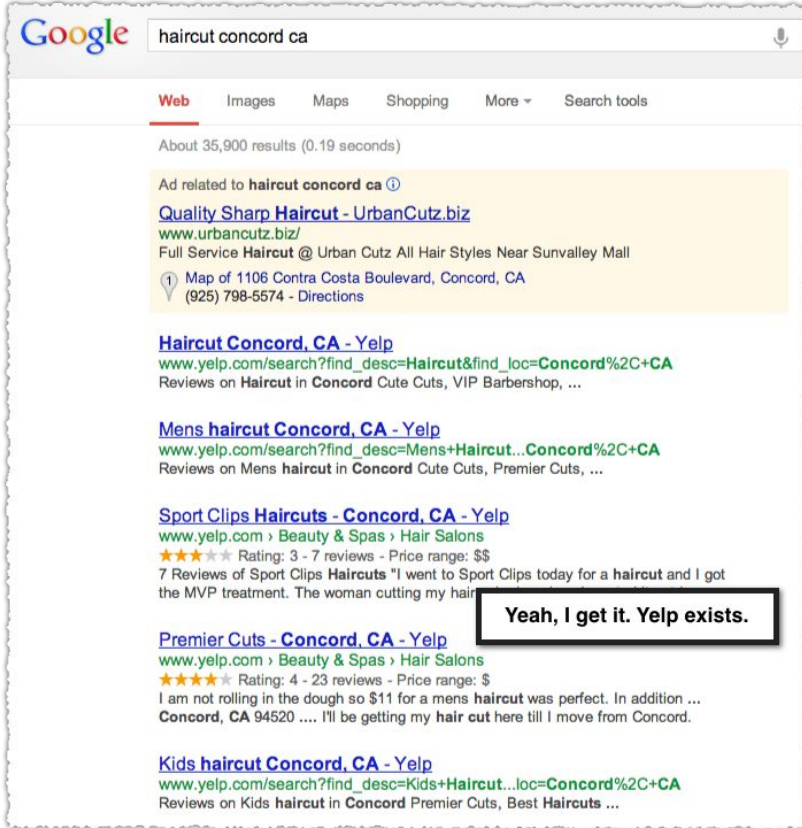


Figure 9.3 How many movies are shown X number of times. More than 350 movies are shown in only one recommendation. Counterintuitively, the movies that are most popular are the ones that are in the long tail.

Evaluate Recommendation Systems: Goals

- Accuracy
- Diversity and coverage
- Serendipity
- Scalability



Google haircut concord ca

Web Images Maps Shopping More Search tools

About 35,900 results (0.19 seconds)

Ad related to **haircut concord ca**

[Quality Sharp Haircut - UrbanCutz.biz](#)
www.urbancutz.biz/
Full Service **Haircut** @ Urban Cutz All Hair Styles Near Sunvalley Mall

Map of 1106 Contra Costa Boulevard, Concord, CA
(925) 798-5574 - Directions

[Haircut Concord, CA - Yelp](#)
www.yelp.com/search?find_desc=Haircut&find_loc=Concord%2C+CA
Reviews on **Haircut** in **Concord** Cute Cuts, VIP Barbershop, ...

[Mens haircut Concord, CA - Yelp](#)
www.yelp.com/search?find_desc=Mens+Haircut...Concord%2C+CA
Reviews on Mens **haircut** in **Concord** Cute Cuts, Premier Cuts, ...

[Sport Clips Haircuts - Concord, CA - Yelp](#)
www.yelp.com > [Beauty & Spas](#) > [Hair Salons](#)
★★★★★ Rating: 3 - 7 reviews - Price range: \$\$
7 Reviews of Sport Clips **Haircuts** "I went to Sport Clips today for a **haircut** and I got the MVP treatment. The woman cutting my hair

[Premier Cuts - Concord, CA - Yelp](#)
www.yelp.com > [Beauty & Spas](#) > [Hair Salons](#)
★★★★★ Rating: 4 - 23 reviews - Price range: \$
I am not rolling in the dough so \$11 for a mens **haircut** was perfect. In addition ...
Concord, CA 94520 I'll be getting my **hair cut** here till I move from Concord.

[Kids haircut Concord, CA - Yelp](#)
www.yelp.com/search?find_desc=Kids+Haircut...loc=Concord%2C+CA
Reviews on Kids **haircut** in **Concord** Premier Cuts, Best **Haircuts** ...

Yeah, I get it. Yelp exists.

Evaluate Recommendation Systems: Goals

- Accuracy
- Diversity and coverage
- Serendipity
- Scalability

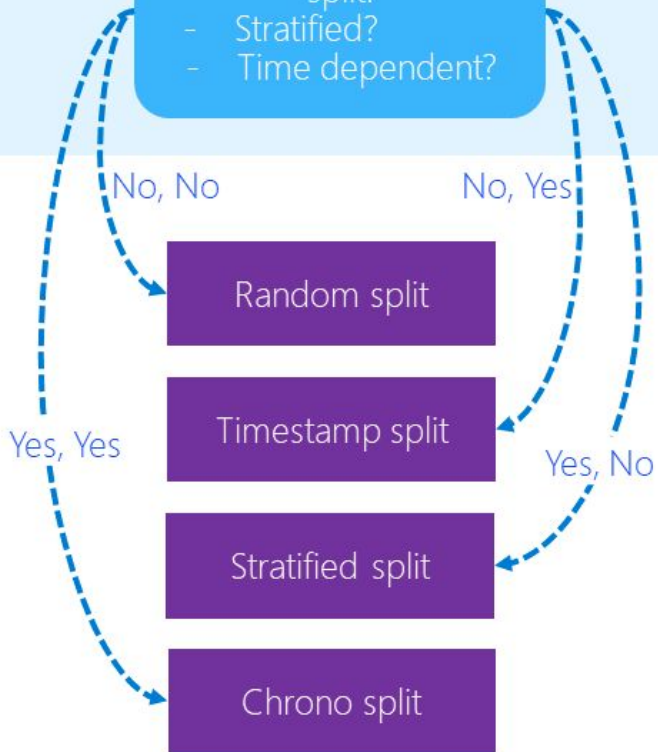


- Accuracy
- Diversity and coverage
- Serendipity
- Scalability -> More to come tomorrow!

Evaluate Recommendation Systems: Offline Evaluation - Split

How do you want to split:

- Stratified?
- Time dependent?



Left diagram:
<https://github.com/microsoft/recommenders/tree/main/examples>



Sara	5	3		2	2	2
Jesper	4	3	4		3	3
Therese	5	2	5	2	1	1
Helle	3	5	3		1	1
Pietro	3	3	3	2	4	5
Ekaterina	2	3	2	3	5	5

Evaluate Recommendation Systems: Offline Evaluation - Split

How do you want to split:

- Stratified?
- Time dependent?

No, No

No, Yes

Random split

Timestamp split

Stratified split

Chrono split

Yes, Yes

Yes, No



Sara	5	3		2	2	2
Jesper	4	3	4		3	3
Therese	5	2	5	2	1	1
Helle	3	5	3		1	1
Pietro	3	3	3	2	4	5
Ekaterina	2	3	2	3	5	5

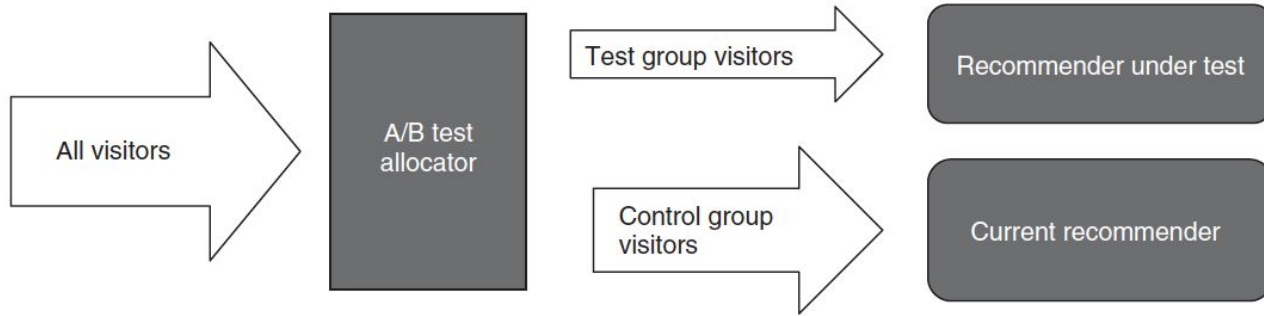


Figure 9.16 In an A/B test, visitors are split into two groups: the test group that sees the new feature and a control group that continues as usual.

Evaluate Recommendation Systems: Online Evaluation

✔ **Success! Results are ready, but the test is still running**
 This experiment has been running for 18 days and collected data for 1200 users. Remote config parameters specified for each variant are still being sent to players. Please stop the test to remove players from the experiment. COPY STOP TEST

Experiment details ▲

GA Test Game
1 variation(s) of led_place

Config key
button_color

Target users
100% of new users

Build(s)
0.9

Filter(s)
- No filters -

Detailed test results for Retention (Day 1)

	Users exposed	Retention (Day 1) ⌵	Probability to be the best ⌵	Improvement over Control ⌵	Distribution ⌵
Control group Default value	571	33.84%	2.22%	0% - 0%	--
Stop - Winner button_color = red	614	39.42%	97.78%	-7.5% - 40.8%	--

Retention (Day 1) All variants selected ⌵

Date	Control group (%)	Stop (%)
Thu 25 Jun 2020	0.0	0.0
Fri 26	25.3	0.0
Sat 27	38.0	45.0
Sun 28	25.3	50.6
Mon 29	25.3	45.0
Tue 30	25.3	45.0
Wed 1 Jul 2020	25.3	45.0
Thu 2	25.3	45.0
Fri 3	12.7	25.3
Sat 4	0.0	0.0
Sun 5	0.0	0.0
Mon 6	0.0	0.0
Tue 7	0.0	0.0
Wed 8	0.0	0.0
Thu 9	0.0	0.0
Fri 10	0.0	0.0
Sat 11	0.0	0.0
Sun 12	0.0	0.0
Mon 13	0.0	0.0

Evaluate Recommendation Systems: Online Evaluation



RETENTION REPORT

Report Settings

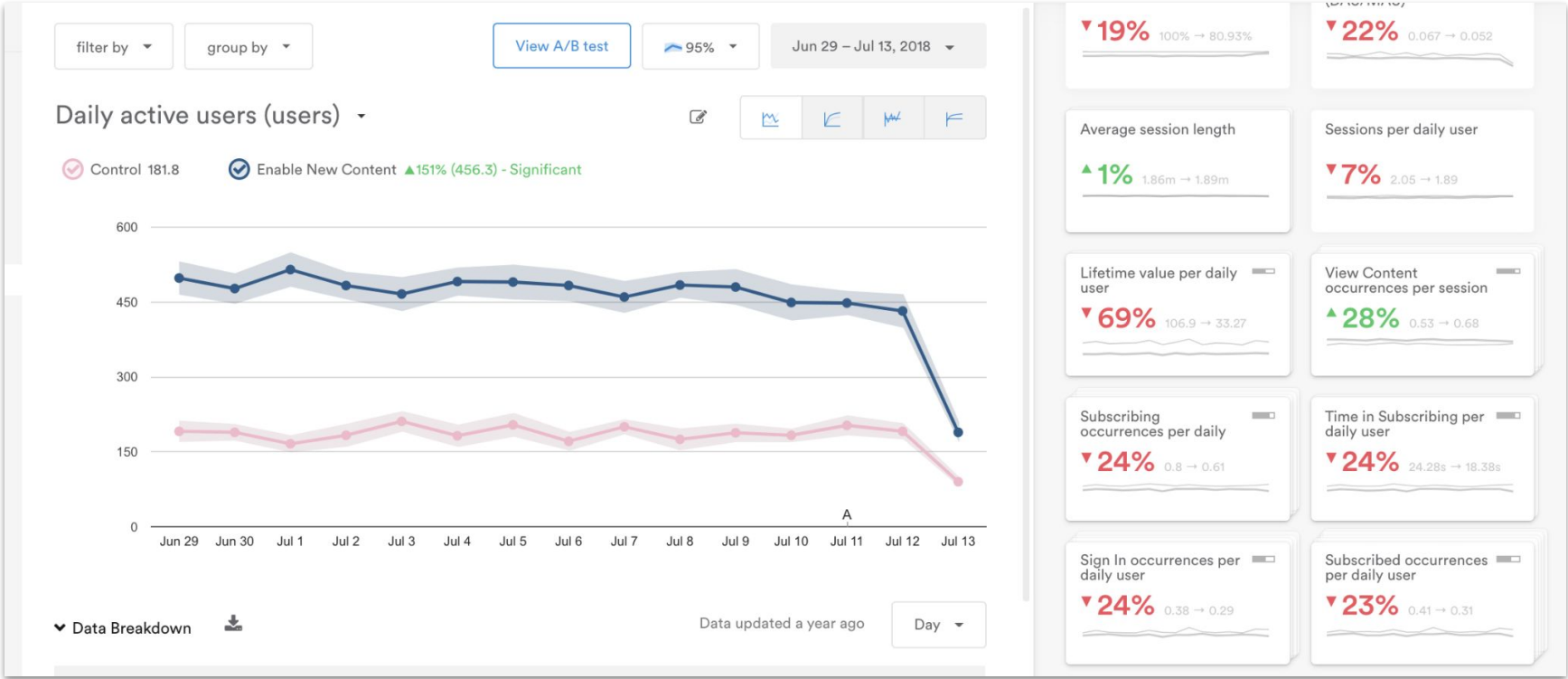
Media Source: All Geo: All

Group By: Media Source Campaign Adset Ad Min Cohort Size ✓

[Add More Filters](#)

Media Source	Campaign	Adset	Ad	Install Day ↓	Day 1	Day 2	Day 3	Day 4
organic	N/A	N/A	N/A	100% 5,212	1719% 896	10.57% 551	8.65% 451	7.77% 405
liftoff_int	iphone	N/A	320x480_... 1	100% 651	15.36% 100	8.45% 55	7.22% 47	6.3% 41
liftoff_int	iphone	N/A	320x480_... 2	100% 649	15.41% 100	9.24% 60	6.78% 44	4.31% 28
Facebook Ads	iOS	iOS Interest AEP	20190404_MWC_2kL... 3	100% 544	12.68% 69	6.99% 38	4.78% 26	3.86% 21

Evaluate Recommendation Systems: Online Evaluation



▼ **19%** 100% → 80.93%

▼ **22%** 0.067 → 0.052

Average session length

▲ **1%** 1.86m → 1.89m

Sessions per daily user

▼ **7%** 2.05 → 1.89

Lifetime value per daily user

▼ **69%** 106.9 → 33.27

View Content occurrences per session

▲ **28%** 0.53 → 0.68

Subscribing occurrences per daily

▼ **24%** 0.8 → 0.61

Time in Subscribing per daily user

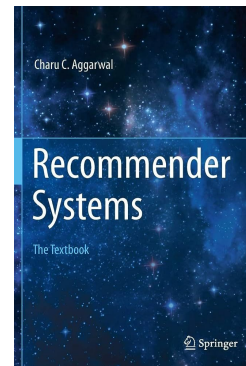
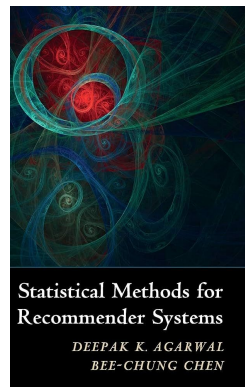
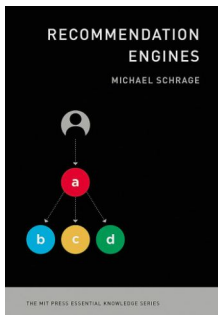
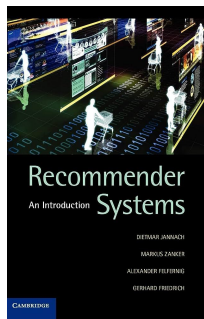
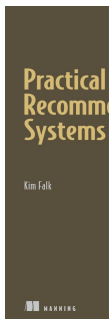
▼ **24%** 24.28s → 18.38s

Sign In occurrences per daily user

▼ **24%** 0.38 → 0.29

Subscribed occurrences per daily user

▼ **23%** 0.41 → 0.31



	PRS	IRS	RE	SERS	RST
Related Chap.	Chap. 7-11.	Chap. 2, 7	Chap. 1-5.	Chap. 1, 2, 4.	Chap 1-4.
Hao's Rating	5	3	4	2	4

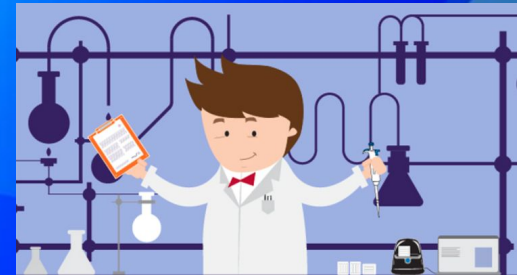
Take a 5 minute break



Local Coffee Outlet Recommender
Based on Yelp data

An illustration featuring a laptop with several small human figures standing on its screen, representing data analysis or user interaction. Below the laptop is a magnifying glass icon. To the left of the laptop is a small icon of a coffee cup. The text 'Local Coffee Outlet Recommender' is in blue, and 'Based on Yelp data' is in red.

Lab Time!



Instruction for running notebooks on Deepnote

Please click on the "Open project" button found in the email titled "ICME Summer 2023 - Search and Recommendation System - Invitation to Collaborate."

If you don't have a Deepnote account already, you might need to sign up for one.

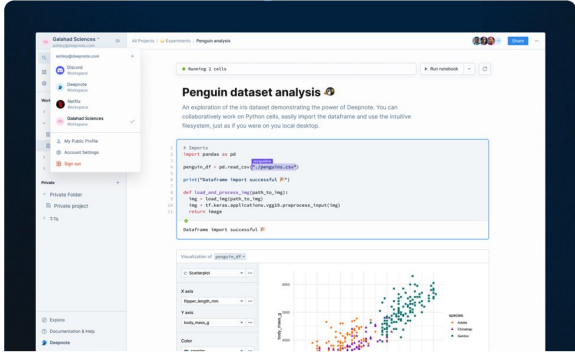
ICME Summer 2023 - Search and Recommendation System - Invitation to collaborate



Deepnote <info@mg.deepnote.com>

To: Rui Yan

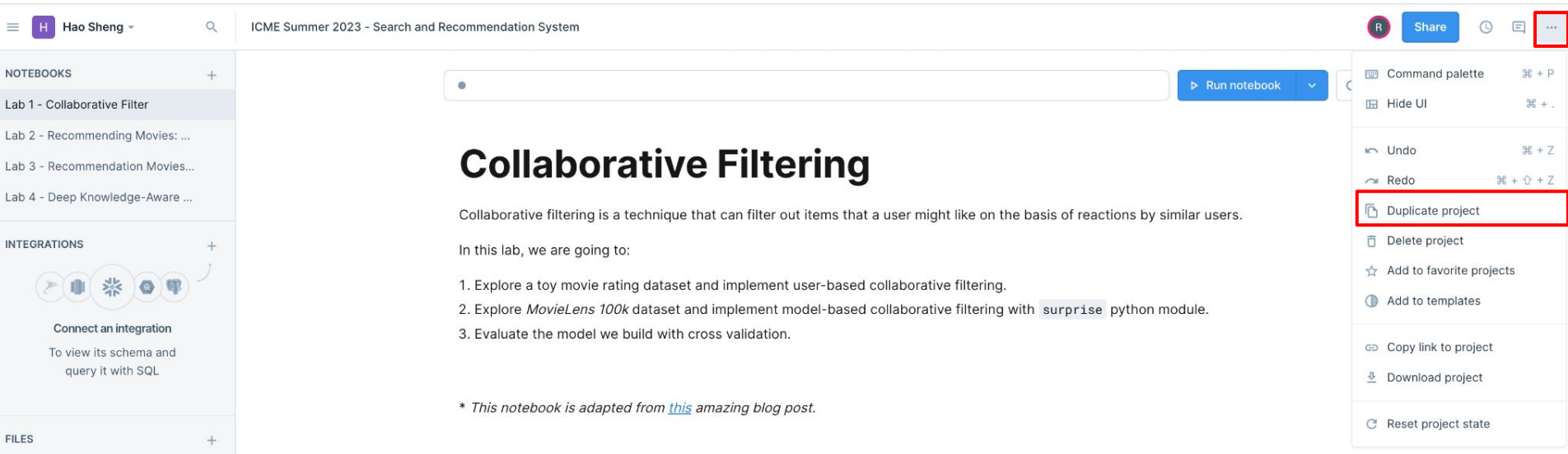
Hao Sheng has invited you collaborate on **ICME Summer 2023 - Search and Recommendation System**



Open project

Instruction for running notebooks on Deepnote

Once you have a Deepnote account, create a workspace with any name you prefer and duplicate the project to your own workspace and run the notebook.



The screenshot shows the Deepnote workspace interface. At the top, the user is identified as 'Hao Sheng' and the workspace is named 'ICME Summer 2023 - Search and Recommendation System'. The left sidebar contains a 'NOTEBOOKS' section with four labs: 'Lab 1 - Collaborative Filter', 'Lab 2 - Recommending Movies: ...', 'Lab 3 - Recommendation Movies...', and 'Lab 4 - Deep Knowledge-Aware ...'. Below this is an 'INTEGRATIONS' section with icons for various services and a 'Connect an integration' button. The main content area displays the notebook 'Collaborative Filtering' with a 'Run notebook' button. The notebook content includes a title, a description of collaborative filtering, a list of tasks, and a note that the notebook is adapted from a blog post. A context menu is open on the right side of the interface, with the 'Duplicate project' option highlighted by a red box.

NOTEBOOKS +

- Lab 1 - Collaborative Filter
- Lab 2 - Recommending Movies: ...
- Lab 3 - Recommendation Movies...
- Lab 4 - Deep Knowledge-Aware ...

INTEGRATIONS +

Connect an integration

To view its schema and query it with SQL

FILES +

ICME Summer 2023 - Search and Recommendation System

Share

Run notebook

Collaborative Filtering

Collaborative filtering is a technique that can filter out items that a user might like on the basis of reactions by similar users.

In this lab, we are going to:

1. Explore a toy movie rating dataset and implement user-based collaborative filtering.
2. Explore *MovieLens 100k* dataset and implement model-based collaborative filtering with `surprise` python module.
3. Evaluate the model we build with cross validation.

* This notebook is adapted from [this](#) amazing blog post.

- Command palette ⌘ + P
- Hide UI ⌘ + .
- Undo ⌘ + Z
- Redo ⌘ + ⌘ + Z
- Duplicate project
- Delete project
- Add to favorite projects
- Add to templates
- Copy link to project
- Download project
- Reset project state