

Week 13

Agenda

1. Recsys / Personalization Discussion

Discussion Questions

- What is a recommender system?
- Is Google a recommender system?
- What are some recommender systems you use? Which ones do you feel are best?
- What other systems do you interact with that are 'personalized'?

Is this a recommender system?



berkeley mids academic calendar



All

Shopping

News

Maps

Images

More

Settings

Tools

About 9,440 results (0.67 seconds)

I School Feeds | UC Berkeley School of Information

<https://www.ischool.berkeley.edu/feeds> ▾

Apr 13, 2018 - Academic calendar for students in on-campus degree programs. (From the UC ... Add MIDS & MICS Academic Calendar to my calendar.

Academic Calendar < University of California, Berkeley

guide.berkeley.edu/academic-calendar/ ▾

For PDFs of current and future **Berkeley Academic Calendars**, visit the Calendar page on the Office of the Registrar website. See the instructions at the bottom of ...

What about this?



Search



Pop Music Recommended videos for you



Kings Of Leon - Hands To
Myself (Selena Gomez cover)

BBCRadio1VEVO

5.2M views • 1 year ago



Fiona Apple - I Walk a little
faster (underwater)

Fiona Apple Rocks

105K views • 1 month ago



Harry Styles - The Chain
(Fleetwood Mac cover) in th...

BBCRadio1VEVO

10M views • 10 months ago



First Aid Kit - Running Up
That Hill (Kate Bush Cover)

Sheen Gekoo

44K views • 2 weeks ago

Personal Data Marketplaces

Basic Info

- Graduate Degree
Demographic > Education > Graduate Degree
- Gen X
Demographic > Age > Lifestages > Gen X

Location & Neighborhood

Professional Interests

Hobbies & Interests

What Others Know About You



DMPs: Personal data marketplaces

- Bluekai.com/registry
- Discuss.
- How do you feel about them?
- Which datasets about yourself do you assume are public?

Taste Domains



Taste Domains

Taste Domains

- Early research focused on 'taste' domains, particularly movies, music, and books.
- In these domains, 'finding' often involved suggestions from friends or tastemakers
- Researchers created collaborative filtering and other approaches as means of emulating this process
- Usually cast as ratings prediction problem in part because its relatively easy to collect ratings data

1996 MovieLens (Minnesota)

- <http://en.wikipedia.org/wiki/MovieLens>
- Research project collected ratings on movies, etc
- Very early Amazon and Netflix strongly influenced by this
- User based CF: Find k-nearest users and use their ratings
- 2001 Item based CF: Find k nearest items to those items a user prefers. http://files.grouplens.org/papers/www10_sarwar.pdf
- Still popular recommendation algorithm

What about recommending open houses?

movielens (Late 90's)

movielens
helping you find the *right* movies

Welcome abc
You're the

So far you have rated **15** movies.
MovieLens needs at least **15** ratings from you to generate predictions.
Please rate as many movies as you can from the list.

	Your Rating	Movie Information
???	Not seen	Beneath the Planet of the Apes (1970) Action, Sci-Fi
???	Not seen	Gift, The (2000) Thriller
???	Not seen	Great Muppet Caper, The (1981) Children, Comedy
???	Not seen	Heaven Can Wait (1978) Comedy
★★★☆☆	4.0 stars	Hitch (2005) Comedy, Romance
???	Not seen	Kate & Leopold (2001) Comedy, Romance
???	Not seen	Muppets Take Manhattan, The (1984) Children, Comedy, Musical
???	Not seen	Police Academy 4: Citizens on Patrol (1987) Comedy
???	Not seen	Saturday Night Fever (1977) Comedy, Drama, Romance
???	Not seen	Teenage Mutant Ninja Turtles II: The Secret of the Ooze (1991) Action, Children, Fantasy

To get a new set of movies click the [next>](#)

movielens - Microsoft Internet Explorer

Archivo Edición Ver Favoritos Herramientas Ayuda

Dirección <http://movielens.umn.edu/search/searchPhrase=&action=newSearch&hiddenParam=1&genre=All&date=All&domain=All&genreSearch=Search+Genre%2FDate%21>

Welcome dus@infovis.net
You've rated **48** movies.
You're the 24th visitor in the past hour.

★★★★★ = Must See
★★★★ = Will Enjoy
★★★★☆ = It's OK
★★☆☆☆ = Fairly Bad
★★☆☆☆ = Awful

[Home](#) | [Manage Buddies](#) | [Your Account](#) | [Help](#) | [Logout](#)

You've searched for **all titles**.
Found 7233 movies, sorted by Prediction.
Genres: All | Exclude Genres: None
Dates: All | Domain: All | Format: All | Language: All
[Show Printer-Friendly Page](#) | [Download Results](#) | [Suggest a Title](#)

Page 1 of 483 | Go to page: [1](#) ... [96](#) ... [192](#) ... [288](#) ... [384](#) ... [480](#) ... [last](#) | [page 2>](#)

Predictions for you	Your Ratings	Movie Information	Wish List
★★★★★	Not seen	Tainted (1998) info imdb Comedy, Thriller	<input type="checkbox"/>
★★★★★	Not seen	Friday Night Lights (2004) info imdb Action, Drama	<input type="checkbox"/>
★★★★★	Not seen	Harry Potter and the Prisoner of Azkaban (2004) info imdb Adventure, Children, Fantasy	<input type="checkbox"/>
★★★★★	Not seen	Spider-Man 2 (a.k.a. Spiderman 2) (2004) info imdb Action, Fantasy, Sci-Fi, Thriller	<input type="checkbox"/>
★★★★★	Not seen	Finding Nemo (2003) DVD , VHS , info imdb Adventure, Animation, Children, Comedy	<input type="checkbox"/>
★★★★★	Not seen	X-Men 2 (a.k.a. X2: X-Men United) (2003) DVD , VHS , info imdb Action, Adventure, Sci-Fi	<input type="checkbox"/>
★★★★★	Not seen	Oliver Twist (1948) info imdb Adventure, Crime, Drama	<input type="checkbox"/>
★★★★★	Not seen	Raiders of the Lost Ark (1981) DVD , info imdb Action, Adventure	<input type="checkbox"/>
★★★★★	Not seen	Indiana Jones and the Last Crusade (1989) DVD , info imdb Action, Adventure	<input type="checkbox"/>

movielens (Today)

MovieLens 10M Dataset

Stable benchmark dataset. 10 million ratings and 100,000 tag applications applied to 10,000 movies by 72,000 users.
Released 1/2009.

- [README.html](#)
- [ml-10m.zip](#) (size: 63 MB, [checksum](#))

Permalink: <http://grouplens.org/datasets/movielens/10m/>

The screenshot shows two sections of the MovieLens website. The top section, titled "top picks", displays a grid of movie cards for "Band of Brothers", "Casablanca", "One Flew Over the Cuckoo's Nest", "The Lives of Others", "Sunset Boulevard", "The Third Man", and "The Third Man". Below this is a "recent releases" section showing cards for "Cantinflas", "Felony", "What If", "Frank", "Sin City: A Dame to Kill For", "If I Stay", and "Are We There Yet?". Each card includes the movie title, year, rating, and duration.

Recommending Houses

Get suggested homes just for you BETA

Finding your suggestions
Don't see anything you like? Start a new search and like homes in the results.

Thumbnail	Price	Details
	\$35,000,000	7 bd / 13 ba / 14,395 sqft 1025 N Starwood Dr Aspen, CO 81611
	\$22,995,000	8 bd / 8 ba / 10,979 sqft 1055 Stage Rd Aspen, CO 81611
	\$37,500,000	6 bd / 9 ba / 11,722 sqft 1509 Owl Creek Ranch Rd Aspen, CO 81611
	\$21,500,000	5 bd / 9 ba / 9,000 sqft 314 E Hyman Ave Aspen, CO 81611
	\$19,950,000	5 bd / 8 ba / 13,927 sqft 360 S Starwood Dr Aspen, CO 81611
	\$18,850,000	6 bd / 10 ba / 12,002 sqft 360 Eagle Pines Dr
	\$17,900,000	7 bd / 9 ba / 13,167 sqft 412 Pioneer Springs Ranch Rd
	\$22,500,000	7 bd / 9 ba / 16,000 sqft 1518 W Buttermilk Rd
	\$47,500,000	12 bd / 15 ba / 23,649 sqft 201 565 Midnight Mine Rd
	\$21,995,000	7 bd / 8 ba / 9,737 sqft 979 Red Mountain Rd

Search vs. Discovery

- Real estate search used to involve looking at classifieds in newspaper, Sites like Trulia are the online equivalent.
- Real estate recommendation used to (and still does) involve agents passing along listings. These recommendations get better as the agent learns more about the homebuyers.
- There are no real estate sites that emulate agent recommendations particularly well

Listing Discovery

- Important decision in one's life
- Only partly a taste domain--utility also plays a role (e.g. space, commute distance)
- Decision making happens over a period of time
- Tastes sometimes evolve during search process
- Social component to both (family, agent, sometimes friends)

Recommending Houses

Get suggested homes just for you BETA

Step 1: Select the place you like best in Malibu, CA

[Pick a new city ▾](#)

Malibu, CA

Pick City



\$7,995,000 4 bd / 4 ba / 2,642 sqft
24604 Malibu Rd
Malibu, CA 90265

 Select



\$2,595,000 5 bd / 5,481 sqft
30010 Andromeda Ln
Malibu, CA 90265

 Select



\$1,999,000 6 bd / 8 ba / 5,012 sqft
4400 Encinal Canyon Rd
Malibu, CA 90265

 Select



\$23,800,000 7 bd / 8 ba
26848 Pacific Coast Hwy
Malibu, CA 90265

 Select

Recommending Houses

Your Homes

Recommendations 23

Followed 10

Liked 11

Hidden 200

From My Agent 121

From My Wife 1,123

Recently Viewed

These are homes you and your agent, Alice Agent, are sharing.



You and your wife may like this home because it's totally unique, unlike all of these homes on this mock.
-Alice Agent



\$899,000 / 2 bd / 2 ba / 1,215 sqft

235 Berry St. #611
San Francisco, CA 94105
South of Market (SoMA)

Hide Like

Show to...



\$899,000 / 2 bd / 2 ba / 1,215 sqft

235 Berry St. #611
San Francisco, CA 94105
South of Market (SoMA)

Hide Like

Show to...



\$899,000 / 2 bd / 2 ba / 1,215 sqft

235 Berry St. #611
San Francisco, CA 94105
South of Market (SoMA)

Hide Like

Show to...

Netflix DVD Recommendations

Movies You've Rated

Based on your 745 movie ratings, this is the list of movies you've seen. As you discover movies on the website that you've seen, rate them and they will show up on this list. On this page, you may change the rating for any movie you've seen, and you may remove a movie from this list by clicking the 'Clear Rating' button.

TITLE	MPAA	GENRE	
Add 12 Angry Men (1957)	UR	Classics	
Add The 39 Steps (1935)	UR	Classics	Clear Rating
Add An American in Paris (1951)	UR	Classics	Clear Rating
Add The Andromeda Strain (1971)	G	Sci-Fi & Fantasy	Clear Rating
Add Apollo 13 (1995)	PG	Drama	Clear Rating
Add The Battle of Algiers (1965) La Battaglia di Algeri	UR	Foreign	Clear Rating
Add Being There (1979)	PG	Drama	Clear Rating
Add Big Deal on Madonna Street (1958) I soliti ignoti	UR	Foreign	Clear Rating
Add The Birds (1963)	PG-13	Thrillers	Clear Rating
Add Blade Runner (1982)	R	Sci-Fi & Fantasy	Clear Rating

Rating Activity

of Ratings: 745 [Rate More](#)

Favorite Genres: 0 [Edit Favorites](#)

Recommendations: 428 [View All](#)

of Reviews Written: 5 [View](#)

[View All](#)

movielens - Microsoft Internet Explorer

Archivo Edición Ver Favoritos Herramientas Ayuda

Búsqueda Favoritos Multimedia

Dirección: <http://movielens.unn.edu/search/searchPhrase=baction=newSearch&hiddenParam=1&genre>AllGenre>AllDate=&alldomain=AllGenreSearch=Search+Genre%2FDate%21>

Welcome [dustaninfo.net](#)

You've rated 48 movies.

You're the 24th visitor in the past hour.

★★★★★ = Will See
★★★★ = Will Enjoy
★★★★ = OK
★★★★ = Fairly Bad
★★★★ = Awful

movielens

helping you find the right movies

[Home](#) | [Manage Buddies](#) | [Your Account](#) | [Help](#) | [Logout](#)

Shortcuts Search

Search Titles [Go!](#)

Use selected buddies!

Search by Genre/Date

All Genres All Dates

Domain: All movies

Use selected buddies!

[Advanced Search](#)

Select Buddies Test Buddy

[What are buddies?](#)

You've searched for all titles.

Found 7233 movies, sorted by Prediction

Genres: All | Exclude Genres: None

Dates: All | Domain: All | Format: All | Language: All

[Show Printer-Friendly Page](#) | [Download Results](#) | [Suggest a Title](#)

Page 1 of 483 | Go to page: 1...96...192...288...384...480...last page 2>

Predictions	Your Ratings	Movie Information	Wish List
★★★★★	<input type="checkbox"/> Not seen	Tainted (1998) info imdb Comedy, Thriller	<input type="checkbox"/>
★★★★★	<input type="checkbox"/> Not seen	Friday Night Lights (2004) info imdb Action, Drama	<input type="checkbox"/>
★★★★★	<input type="checkbox"/> Not seen	Harry Potter and the Prisoner of Azkaban (2004) info imdb Adventure, Children, Fantasy	<input type="checkbox"/>
★★★★★	<input type="checkbox"/> Not seen	Spider-Man 2 (a.k.a. Spiderman 2) (2004) info imdb Action, Fantasy, Sci-Fi, Thriller	<input type="checkbox"/>
★★★★★	<input type="checkbox"/> Not seen	Finding Nemo (2003) DVD, VHS, info imdb Animation, Children, Comedy	<input type="checkbox"/>
★★★★★	<input type="checkbox"/> Not seen	X-Men 2 (a.k.a. X2: X-Men United) (2003) DVD, VHS, info imdb Action, Adventure, Sci-Fi	<input type="checkbox"/>
★★★★★	<input type="checkbox"/> Not seen	Oliver Twist (1948) info imdb Adventure, Crime, Drama	<input type="checkbox"/>
★★★★★	<input type="checkbox"/> Not seen	Raiders of the Lost Ark (1981) DVD, info imdb Action, Adventure	<input type="checkbox"/>
★★★★★	<input type="checkbox"/> Not seen	Indiana Jones and the Last Crusade (1989) DVD, info imdb Action, Adventure	<input type="checkbox"/>

Netflix DVD Recommendations

Movies You've Rated

Based on your 745 movie ratings, this is the list of movies you've seen. As you discover movies on the website that you've seen, rate them and they will show up on this list. On this page, you may change the rating for any movie you've seen, and you may remove a movie from this list by clicking the 'Clear Rating' button.

TITLE	MPAA	GENRE	
Add  12 Angry Men (1957)	UR	Classics	
Add  The 39 Steps (1935)	UR	Classics	 Clear Rating
Add  An American in Paris (1951)	UR	Classics	 Clear Rating
Add  The Andromeda Strain (1971)	G	Sci-Fi & Fantasy	 Clear Rating
Add  Apollo 13 (1995)	PG	Drama	 Clear Rating
Add  The Battle of Algiers (1965) La Battaglia di Algeri	UR	Foreign	 Clear Rating
Add  Being There (1979)	PG	Drama	 Clear Rating
Add  Big Deal on Madonna Street (1958) I soliti ignoti	UR	Foreign	 Clear Rating
Add  The Birds (1963)	PG-13	Thrillers	 Clear Rating
Add Blade Runner (1982)	R	Sci-Fi & Fantasy	Clear Rating

 **Rating Activity** [View All](#)

of Ratings: 745 [Rate More](#)
Favorite Genres: 0 [Edit Favorites](#)
Recommendations: 428 [View All](#)
of Reviews Written: 5 [View](#)

 Clear Rating
 Clear Rating
 Clear Rating
 Clear Rating
 Clear Rating
 Clear Rating
 Clear Rating
 Clear Rating

2002 Netflix DVDs

- Mailed 1-3 DVDs to users
- Brand based in part on ratings and recommendations
- Users could rate any movie, not just those recently watched
- Users were told recommendations get better with more ratings

Search and Discovery

- Amazon historically incorporated recommendations into search results, including to extend results
- Recommendations used at different stages of transaction (E.g. Product page, cart)

Amazon Recommendations (2003)

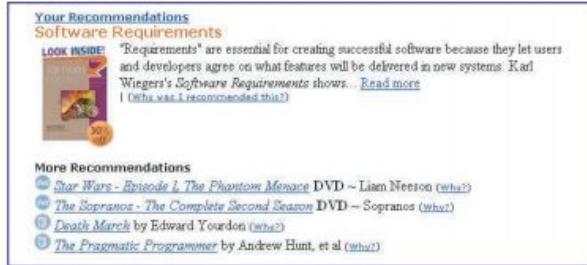


Figure 1. The "Your Recommendations" feature on the Amazon.com homepage. Using this feature, customers can sort recommendations and add their own product ratings.



Figure 2. Amazon.com shopping cart recommendations. The recommendations are based on the items in the customer's cart: The Pragmatic Programmer and Physics for Game Developers.

2003 Amazon

- Many people at time identified recommendations with Amazon
- Item-Item CF (<http://www.cs.umd.edu/~samir/498/Amazon-Recommendations.pdf>)
- Compares to User-Item based CF (as in GroupLens)
- Pointed to scalability: compute similarities between items offline, usually nightly. This is preferred as item similarities change less frequently than user similarities
- Use item similarities to recommend similar items to those a user has shown preference towards
- Item preference a weighted combination of views, ratings, cart adds, and purchases
- Added blended exponential decay to account for time
- Item similarities can also be used in non-personalized fashion ("users who considered this, also considered that...")
- Not fully real-time and suffers from cold-start problem
- Flawed problem setup: not directly optimizing sales

Amazon Recommendations

Item-based Collaborative Filtering Recommendation Algorithms

Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl

GroupLens Research Group/Army HPC Research Center
Department of Computer Science and Engineering
University of Minnesota, Minneapolis, MN 55455
{sarwar, karypis, konstan, riedl}@cs.umn.edu

Appears in WWW10, May 1-5, 2001, Hong Kong.

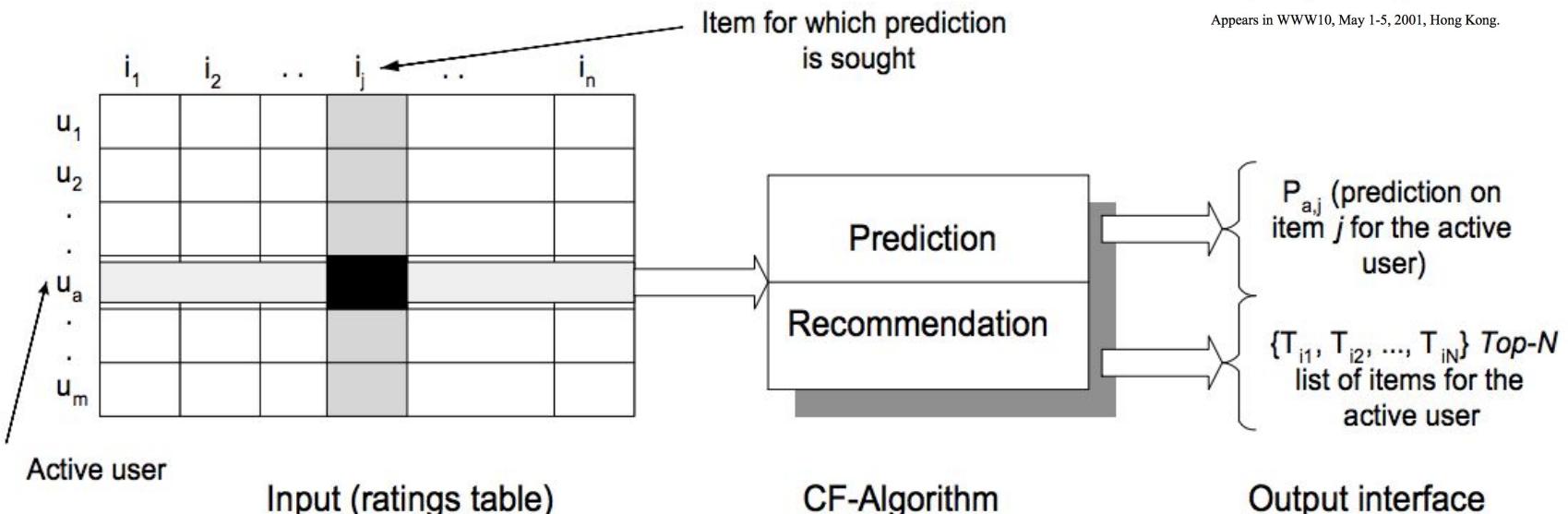


Figure 1: The Collaborative Filtering Process.

Amazon Recommendations

Item-based Collaborative Filtering Recommendation Algorithms

Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl

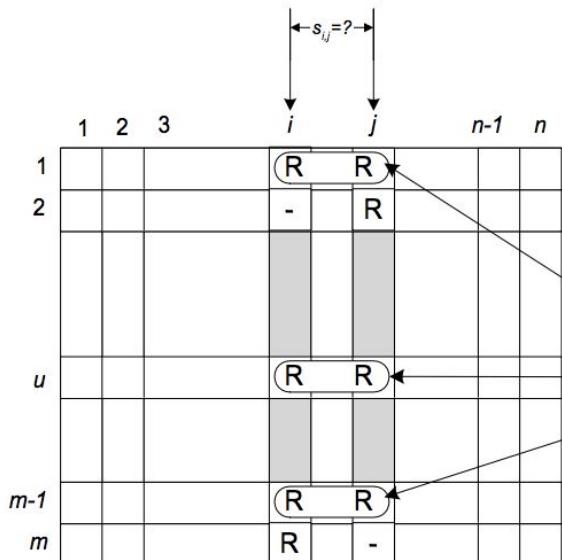
GroupLens Research Group/Army HPC Research Center

Department of Computer Science and Engineering

University of Minnesota, Minneapolis, MN 55455

{sarwar, karypis, konstan, riedl}@cs.umn.edu

pearls in WWW10, May 1-5, 2001, Hong Kong.



$$sim(i, j) = \frac{\sum_{u \in U} (R_{u,i} - \bar{R}_i)(R_{u,j} - \bar{R}_j)}{\sqrt{\sum_{u \in U} (R_{u,i} - \bar{R}_i)^2} \sqrt{\sum_{u \in U} (R_{u,j} - \bar{R}_j)^2}}.$$

Item-item similarity is computed by looking into co-rated items only. In case of items i and j the similarity s_{ij} is computed by looking into them. Note: each of these co-rated pairs are obtained from different users, in this example they come from users 1, u and $m-1$.

Figure 2: Isolation of the co-rated items and similarity computation

Amazon Recommendations

Item-based Collaborative Filtering Recommendation Algorithms

Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl

GroupLens Research Group/Army HPC Research Center
 Department of Computer Science and Engineering
 University of Minnesota, Minneapolis, MN 55455
 {sarwar, karypis, konstan, riedl}@cs.umn.edu

Appears in WWW10, May 1-5, 2001, Hong Kong.

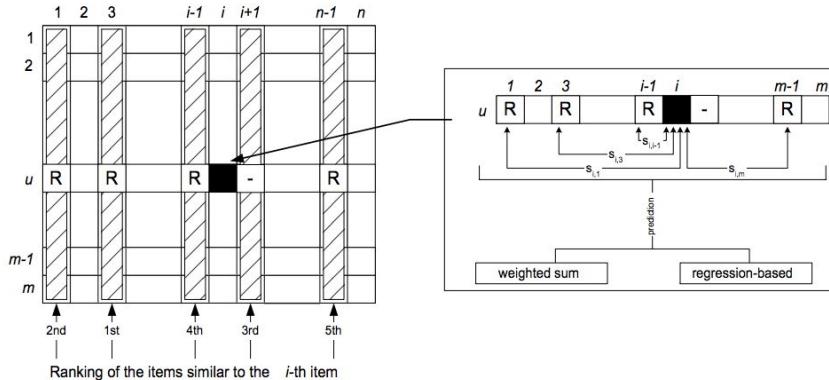


Figure 3: Item-based collaborative filtering algorithm. The prediction generation process is illustrated for 5 neighbors

the corresponding user average from each co-rated pair. Formally, the similarity between items i and j using this scheme is given by

$$sim(i, j) = \frac{\sum_{u \in U} (R_{u,i} - \bar{R}_u)(R_{u,j} - \bar{R}_u)}{\sqrt{\sum_{u \in U} (R_{u,i} - \bar{R}_u)^2} \sqrt{\sum_{u \in U} (R_{u,j} - \bar{R}_u)^2}}.$$

Here \bar{R}_u is the average of the u -th user's ratings.

Cold Start Problem

How do we handle new items?

How do we handle new users?

Pandora Recommendations



The Music Genome Project

Track Length (Min:Sec) 0:19

Head (Principal Melody)	
Fixed-to-Improvized [0-5]	1.0
Number of Secondary Themes [0-4]	0.0
Span Narrow-to-Wide [0-4]	4.5
Lyrical-to-Angular [0-4]	3.5
Melodic Rhythm Intensity Lo-to-Hi [0-4]	3.0
Contour Mono-to-Melismatic [0-5]	3.5
Phrase Repetitive-to-Thru [0-4]	2.5
Ornamentation [0-5]	0.0
Presentation Single-to-Ensemble [0-4]	1.0
Presentation Unison to Chordal [0-4]	1.0
Antiphony [0-5]	0.0
Counterpoint [0-5]	0.0

Harmony	
Modal [0-5]	4.0
Minor-to-Major [0-4]	1.5
Diatomic-to-Chromatic [0-5]	2.0
Overall Resonance Lo-to-Hi [0-4]	2.0
Chordal Patternning [0-5]	2.0
Chordal Rhythm Slow-to-Fast [0-4]	2.5
Pedal Point [0-5]	3.0
Fermi and Arrangement	
Multi-Sectioned [0-5]	0.0
Head-Solo-Head-to-Thru Composed [0-4]	1.0
12-bar Blues [0-5]	0.0
"Song" Form (ABA) [0-5]	2.5
Breaks [0-5]	0.0
Intro Incidental-to-Dominant [0-1-4]	1.0
Intro Faded-to-In Tempo [0-1-5]	5.0

Rhythmic Tempo, Meter, Feel, Groove	
Primary Tempo BPM [0-200]	172.0
Cut time [0-6]	0.0
Triple (3/4,3/8,9/8) [0-5]	0.0
Compound Duplet (6/8+2/8) [0-4]	0.0
Odd (5/4,7/4) P-S	5.0
Discernibility Lo-to-Hi [0-5]	5.0
Number of Shifts [0-4]	0.0
Swing or Shuffle [0-5]	5.0
Swing to Shuffle [0-1-5]	1.0
Swung Sixteenths [0-5]	0.0
Subdivisions Lo-to-Hi [0-5]	3.0
Latin p-g	0.0
Double-Time [0-3]	0.0
Back-beat Prominence [0-5]	0.0
Rhythmic Estimate-Based [0-5]	4.0
Motion-Inducing Lo-to-Hi [0-5]	3.5
Rhyth. Temp. Simple-to-Complex [0-5]	3.0
Syncopation Level Low-to-High [0-5]	3.0
Rhythmic Loops (Pre-recorded) [0-5]	0.0

Heads, Harmonies, Form, Rhythms 1 2 3 4 5 6 7
Low Vocal Expressive, Improv A, Improv B 1 2 3 4 5 6 7
Sax, Clarinet, Tuba, Harmonica, Trumpet 1 2 3 4 5 6 7
Trombone, Valve, Vibes, Flute, Organ, Synthesizer, Accordion, Ocarina 1 2 3 4 5 6 7
Bass, Double Bass, Electric Bass, Piano, Harpsichord, String Inst., Accordion, Ukelele 1 2 3 4 5 6 7
Lyric, Sound, Concept, Dynamics 1 2 3 4 5 6 7
Guitar, Overall 1 2 3 4 5 6 7

This analysis is NOT approved and should be excluded from the database.

The Netflix Prize

- Training data
 - 100,480,507 ratings
 - 480,189 users
 - 17,770 movies
 - <user, movie, date of grade, grade>
 - <integer, integer, date, 1–5>
- Test data
 - 2,817,131 ratings
 - <integer, integer, date>
- Objective: minimize RMSE:

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

Most Active Users

User ID	# Ratings	Mean Rating
305344	17,651	1.90
387418	17,432	1.81
2439493	16,560	1.22
1664010	15,811	4.26
2118461	14,829	4.08
1461435	9,820	1.37
1639792	9,764	1.33
1314869	9,739	2.95

Copyright AT&T

Data about the Movies

Most Loved Movies	Avg rating	Count
The Shawshank Redemption	4.593	137812
Lord of the Rings : The Return of the King	4.545	133597
The Green Mile	4.306	180883
Lord of the Rings : The Two Towers	4.460	150678
Finding Nemo	4.415	139050
Raiders of the Lost Ark	4.504	117456

Most Rated Movies

Miss Congeniality
Independence Day
The Patriot
The Day After Tomorrow
Pretty Woman
Pirates of the Caribbean

Highest Variance

The Royal Tenenbaums
Lost In Translation
Pearl Harbor
Miss Congeniality
Napoleon Dynamite
Fahrenheit 9/11

Copyright AT&T

The Netflix Prize



The Netflix Prize

SVD for Rating Prediction

- User factor vectors $p_u \in \Re^f$ and item-factors vector $q_v \in \Re^f$
- Baseline $b_{uv} = \mu + b_u + b_v$ (user & item deviation from average)
- Predict rating as $\hat{r}_{uv} = b_{uv} + p_u^T q_v$
- **SVD++** (Koren et. Al) asymmetric variation w. implicit feedback

$$\hat{r}_{uv} = b_{uv} + q_v^T \left(|R(u)|^{-\frac{1}{2}} \sum_{j \in R(u)} (r_{uj} - b_{uj}) x_j + |N(u)|^{-\frac{1}{2}} \sum_{j \in N(u)} y_j \right)$$

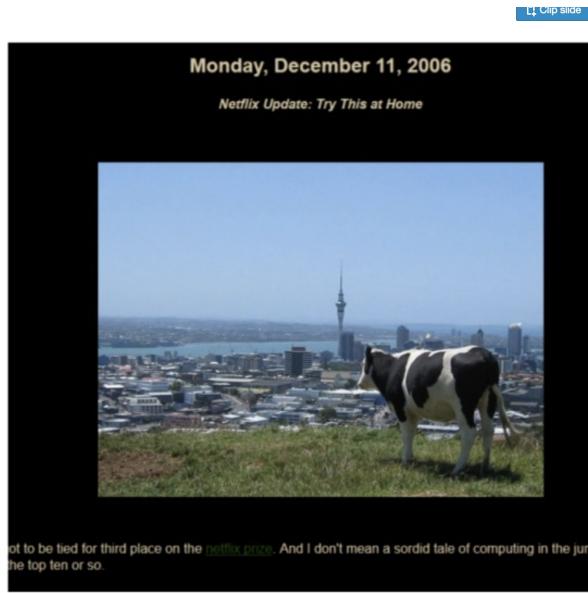
- Where
 - $q_v, x_v, y_v \in \Re^f$ are three item factor vectors
 - Users are not parametrized, but rather represented by:
 - $R(u)$: items rated by user u
 - $N(u)$: items for which the user has given implicit preference (e.g. rated vs. not rated)

The Netflix Prize

Simon Funk's SVD

- One of the most interesting findings during the Netflix Prize came out of a blog post
- Incremental, iterative, and approximate way to compute the SVD using gradient descent

NETFLIX



Clustering

- Item-item CF didn't work well
- <http://gigapan.com/gigapans/65469>

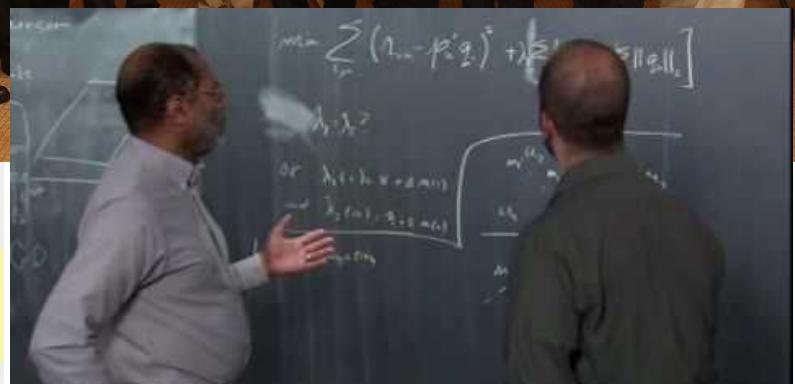
Matrix Factorization Approaches

- 'Simon Funk' released his matrix factorization solution, which supported data sparsity. Used in Netflix (for predicted rating) today
- Dim reduction on sparse data using SVD
- <http://sifter.org/~simon/journal/20061211.html>
- Later approaches added better formalization and incremental improvements

The Netflix Prize

What Happened

- October 2, 2006: contest launched—\$1 million first prize
 - *Cinematch* RMSE = 0.9514
 - Naïve "average rating" RMSE = 1.0540
- October 8, 2006: "WXYZ" team beats *Cinematch*
- June 2007: 20,000 teams in competition (150 countries)
- September 2007: "BellKor" RMSE = 0.8728 (\$50k)
- September 2008: "BellKor" RMSE = 0.8616 (\$50k)
- July 2009: two teams hit 10% margin—no more entries
- September 2009: "BellKor's Pragmatic Chaos" wins \$1 million with RMSE of 0.8567. Same RMSE matched by "The Ensemble," but submission made 20 minutes later. Tiebreaker went to BellKor's Pragmatic Chaos



The Netflix Prize

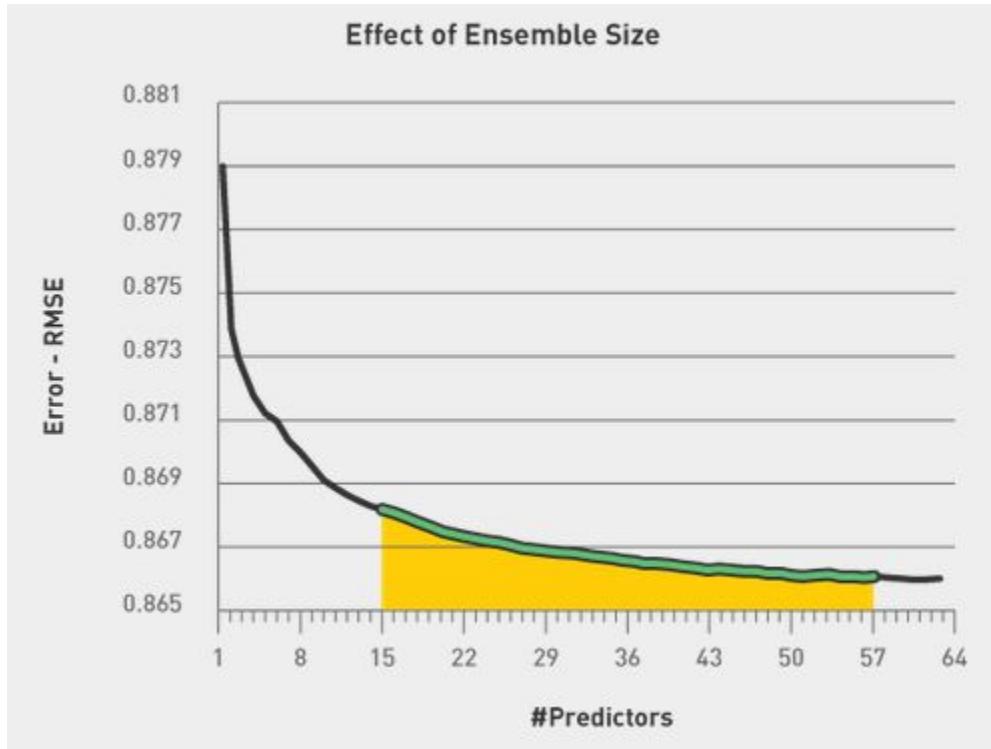
Netflix: The Winner (contd)

- Final product used ensemble methods, which combined results from several algorithms.
 - Neighborhood methods
 - Matrix decomposition methods
 - Regression
 - Boltzmann machines
 - *Et cetera*
- Combining results is topic of another lecture!
- Gradient-boosted decision trees combine 500+ models.

What Other Factors Matter?

- The number of other people who have rated the movie
- The (square root of the) number of days since first rating
- The number of days since the movie's first rating
- Plus a bunch of others!

The Netflix Prize



Netflix Today

Goodbye



& Cinematch

Hello



&

% Match

Why?

+200% ratings volume

Clear link to personalization

Netflix Today

Everything you
see is a
recommendation

House of Cards

★★★★★ 2013 TV-MA 1 Season HD 5.1

Sharks gliding ominously beneath the surface of the water? They're a lot less menacing than this Congressman.

 This winner of three Emmys, including Outstanding Directing for David Fincher, stars Kevin Spacey and Robin Wright.

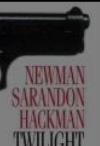


NETFLIX

Because you watched Orange Is the New Black



Because you watched Red Lights



Netflix Today



Netflix Today

Ranking

- Ranking = **Scoring + Sorting + Filtering** bags of movies for presentation to a user
- **Goal:** Find the best possible ordering of a set of *videos* for a *user* within a specific *context* in real-time
- **Objective:** maximize consumption
- **Aspirations:** Played & “enjoyed” titles have best score
- Akin to CTR forecast for ads/search results
- **Factors**
 - Accuracy
 - Novelty
 - Diversity
 - Freshness
 - Scalability
 - ...

Netflix Today

Ranking

- Popularity is the obvious baseline
- Ratings prediction is a clear secondary data input that allows for personalization
- We have added many other features (and tried many more that have not proved useful)
- What about the weights?
 - Based on A/B testing
 - Machine-learned

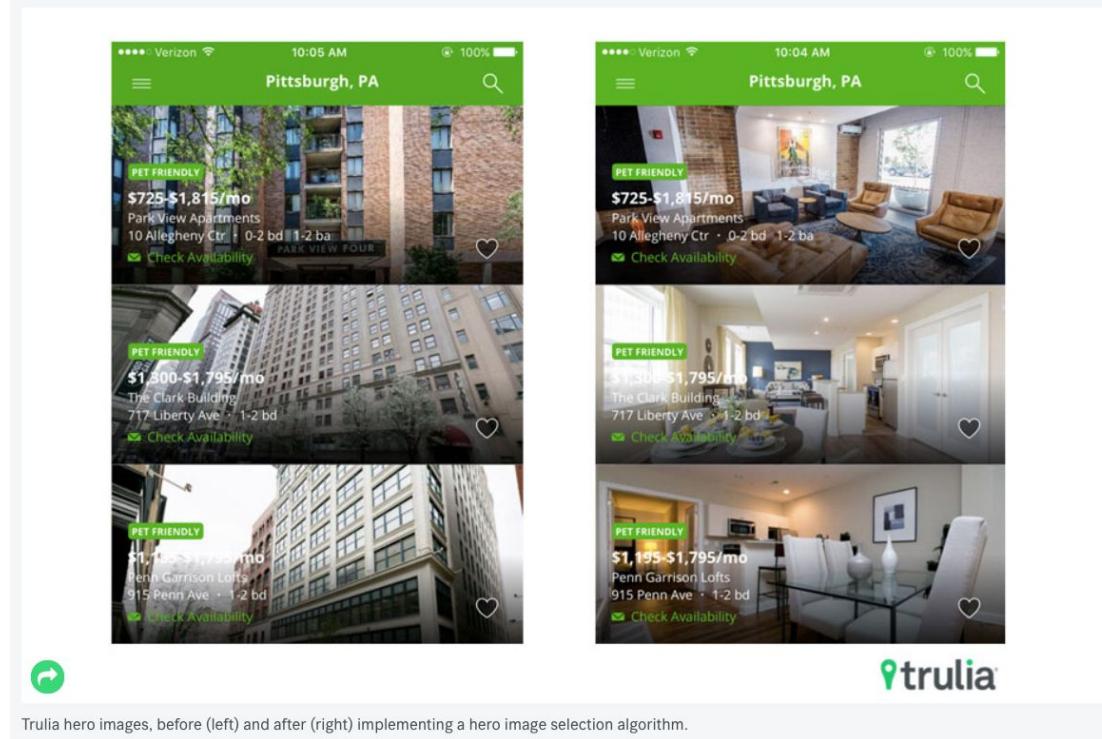
Netflix Today

Learning to Rank Approaches

1. Pointwise
 - Ranking function minimizes loss function defined on individual relevance judgment
 - Ranking score based on regression or classification
 - Ordinal regression, Logistic regression, SVM, GBDT, ...
2. Pairwise
 - Loss function is defined on pair-wise preferences
 - Goal: minimize number of inversions in ranking
 - Ranking problem is then transformed into the binary classification problem
 - RankSVM, RankBoost, RankNet, FRank...

Personalized Presentation

What Makes a Photo Click: Selecting Hero Images with Deep Learning



Personalized Presentation

Using Deep Learning to automatically rank millions of hotel images

At idealo.de we trained two Deep Neural Networks to assess the aesthetic and technical quality of images 😊😊😊



6.52 (1.44)



5.58 (1.37)



5.53 (1.42)



5.04 (1.35)



4.92 (1.48)



4.29 (1.45)

Aesthetic MobileNet predictions



8.04 (2.11)



4.61 (2.75)



1.92 (1.53)



5.73 (2.85)



4.31 (2.7)



4.22 (2.78)

Netflix Today

Here are some screenshots from the tool that we use to track relative artwork performance.
Dragons: Race to the Edge: the two marked images below significantly outperformed all others.

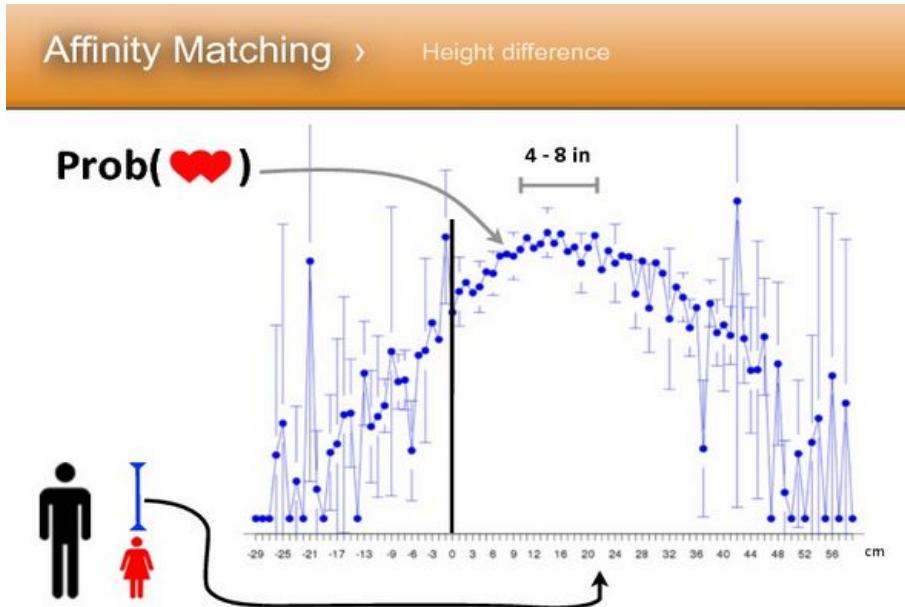


Unbreakable Kimmy Schmidt



Dating Recommendations - How would you do it?

EHarmony

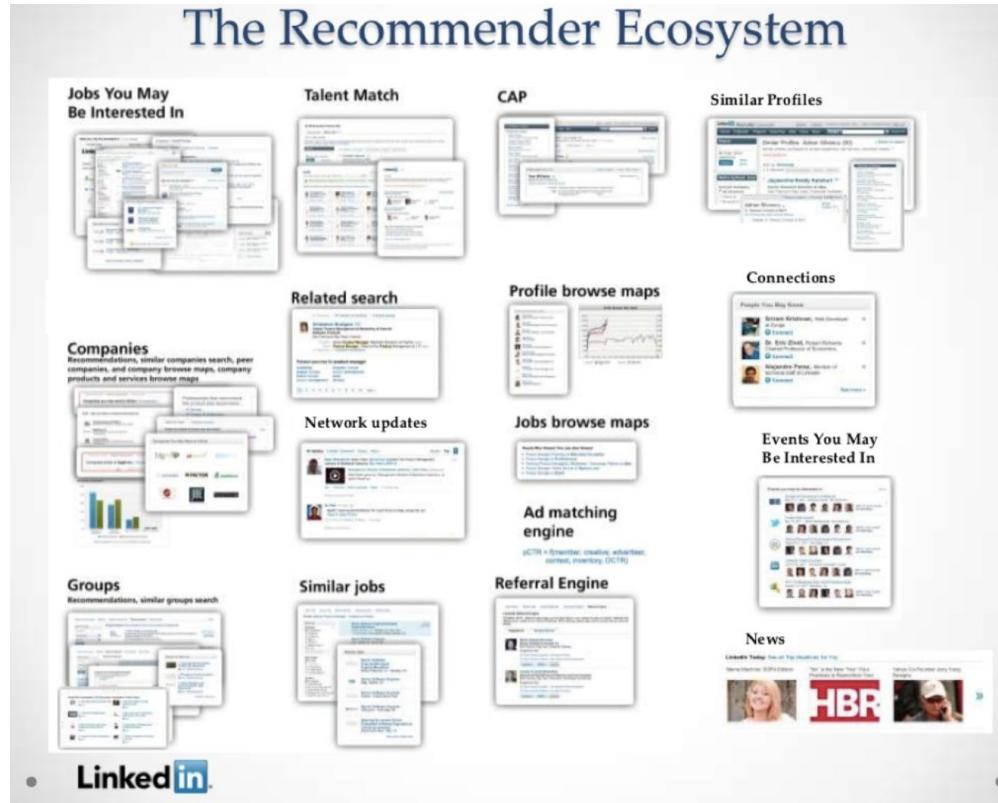


EHarmony

- Straightforward supervised learning approach to building a recommender system.
- Focus is on feature engineering
- 320 attributes (e.g. romantic, height, photo zoom, food preferences)
- Predict likelihood of 'successful match' using logistic regression (Vowpal Wabbit, GBM)
- For more info
<https://ieondemand.com/divisions/big-data/presentations/data-science-of-love-1>

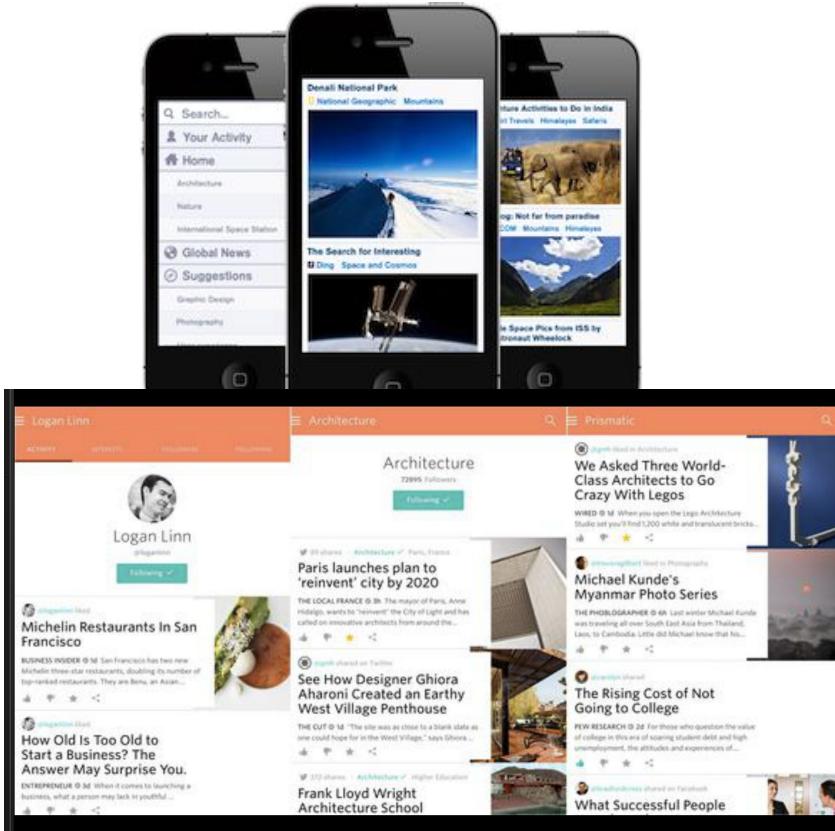
Job Recommendation - How would you do it?

LinkedIn



News Recommendation - How would you do it?

News Recommendation



Prismatic (Acquired by MS in 2015)

- Apple news is almost identical
- Problem: Provide personal news feed
- Popular on mobile devices as way to kill time (e.g. on bus)
- Mobile usage impacts implicit data

ML Approach

- Doesn't use Netflix Prize style collaborative filtering
- Step 1: Featurization using Topic Modeling (25K+ topics)
- Step 2: Directly optimize likelihood to read using logistic regression

Frontier: Deep Latent Factors

Variational Autoencoders

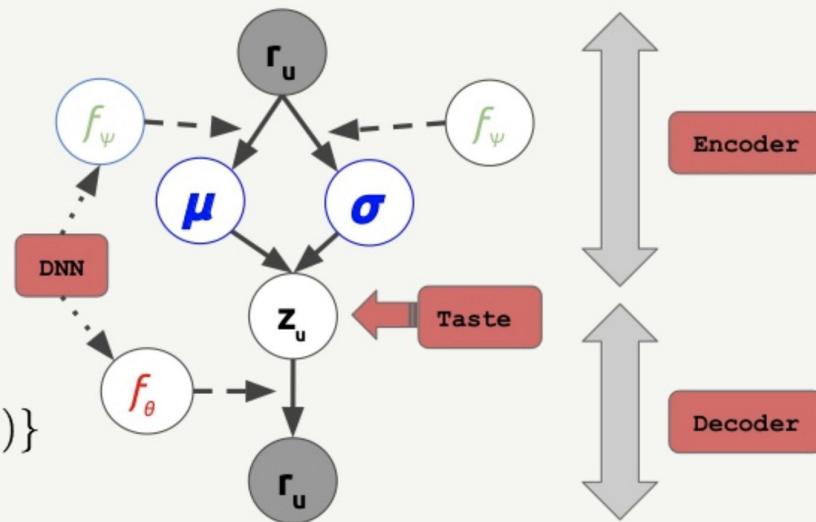
Inference model:

$$q_\psi(\mathbf{z}_u | \mathbf{r}_u) = \mathcal{N}(\mu_\psi(\mathbf{r}_u), \text{diag}\{\sigma_\psi^2(\mathbf{r}_u)\})$$

Generative model:

$$\mathbf{z}_u \sim \mathcal{N}(0, \mathbf{I}_K), \pi(\mathbf{z}_u) \propto \exp\{f_\theta(\mathbf{z}_u)\}$$

$$\mathbf{r}_u \sim \text{Multi}(N_u, \pi(\mathbf{z}_u))$$



Frontier: Whole Page Optimization

Beyond Ranking: Optimizing Whole-Page Presentation





LATEST NEWS

July 26, 2018: Tentative lists of accepted [short](#) and [long](#) papers for RecSys 2018 are online!

July 12, 2018: This year's RecSys conference will feature a record number of six [tutorials](#), taking place on Oct 2, prior to the main conference.

July 10, 2018: Acceptance notifications have been sent out: It was very competitive this year, with 18% acceptance rate for long and 25% for short papers.

July 4, 2018: Registration for RecSys 2018 is open!

June 12, 2018: The call for [late-breaking results \(posters\)](#) is out now, and the call for [demos](#) has been updated!

June 05, 2018: RecSys 2018 will feature three exciting [keynotes](#) by Elizabeth F. Churchill (Google), Lise Getoor (UCSC), and Christopher Berry (Canadian Broadcasting Corporation)!

May 30, 2018: The tutorial submission deadline has been extended to next Monday, June 4, 2018!

SHORTCUTS TO CONFERENCES

- [RecSys 2018 \(Vancouver\)](#)
- [RecSys 2017 \(Como\)](#)
- [RecSys 2016 \(Boston\)](#)
- [RecSys 2015 \(Vienna\)](#)
- [RecSys 2014 \(Silicon Valley\)](#)
- [RecSys 2013 \(Hong Kong\)](#)
- [RecSys 2012 \(Dublin\)](#)
- [RecSys 2011 \(Chicago\)](#)
- [RecSys 2010 \(Barcelona\)](#)
- [RecSys 2009 \(New York\)](#)
- [RecSys 2008 \(Lausanne\)](#)
- [RecSys 2007 \(Minnesota\)](#)

Final Thoughts?