# Lecture 4: Recommender systems

University of BRISTOL

EMAT31530/Feb 2018/Raul Santos-Rodriguez

# Have a look at ...

... D. Janach, M. Zanker, A. Felfering and G. Friedrich: Recommender Systems: An Introduction. Cambridge University Press, Cambridge, 2011.

... F. Ricci, L. Rokach, B. Shapira and P. B. Kantor (eds.): Recommender Systems Handbook. Springer Verlag, New York, 2011.

... O. Celma: Music Recommendation and Discovery, Springer, 2010.

... `http://www.coursera.org/learn/recommender-systems-introduction`

... and some **code**!

`https://github.com/ocelma/python-recsys`

# Outline

This talk discusses how automatic recommender systems learn about your tastes and exploit all the information available from other users to present you with interesting suggestions.

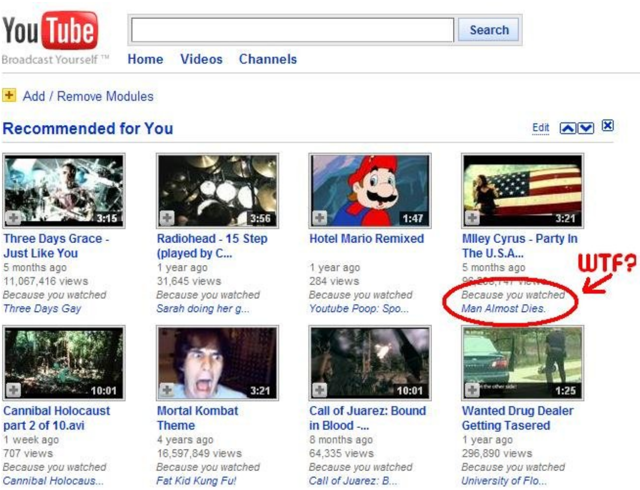
We will describe the main approaches to the design of music recommender systems.

- Collaborative filtering
- Content-based methods

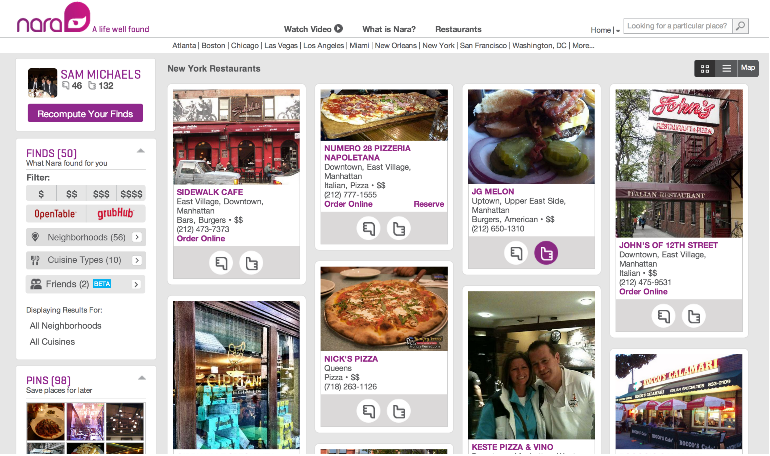
# Motivation

amazon.com
Hello. Sign in to get personalized recommendations. New customer? Start here.
Your Amazon.com | Today's Deals | Gifts & Wish Lists | Gift Cards
Shop All Departments
Search
Movies & TV
Movies & TV
Advanced Search
Browse Genres
New Releases
Bestsellers
DVD & Blu-Ray Deals
Jon and Kate Plus Ei8ht: The Complete Season 4 (6 DVD Set)
Starring: Jonathan Gosselin, Kate Gosselin Director: Jennifer Stocks Format: DVD
List Price: $49.98
Price: $29.99 & this item ships for FREE with Super Saver Shipping. Details
You Save: $19.99 (40%)
In Stock.
Ships from and sold by Amazon.com. Gift-wrap available.
Customers Who Bought This Item Also Bought
Hanging Ceiling Fan
(1) $94.99
12" Footstool
(10) $36.99
Noose
(2) $19.99

# Motivation

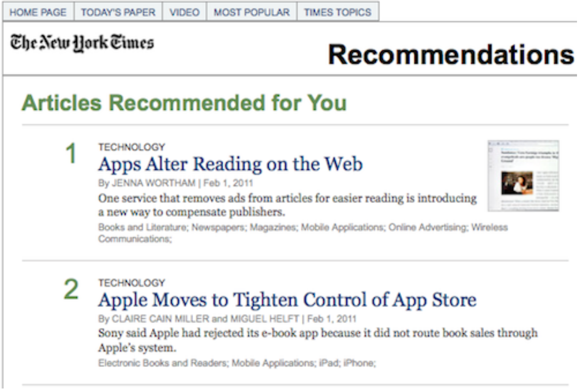

You Tube
Broadcast Yourself ™
Search
Home
Videos
Channels
Add / Remove Modules
Recommended for You
Edit
3:15
Three Days Grace - Just Like You
5 months ago
11,067,416 views
Because you watched
Three Days Gay
3:56
Radiohead - 15 Step (played by C...
1 year ago
31,645 views
Because you watched
Sarah doing her g...
1:47
Hotel Mario Remixed
1 year ago
284 views
Because you watched
Youtube Poop: Spo...
3:21
Miley Cyrus - Party In The U.S.A...
5 months ago
Because you watched
Man Almost Dies.
WTF?
10:01
Cannibal Holocaust part 2 of 10.avi
1 week ago
707 views
Because you watched
Cannibal Holocaus...
3:21
Mortal Kombat Theme
4 years ago
16,597,849 views
Because you watched
Fat Kid Kung Fu!
10:01
Call of Juarez: Bound in Blood -...
8 months ago
64,335 views
Because you watched
Call of Juarez: B...
1:25
Wanted Drug Dealer Getting Tasered
1 year ago
296,890 views
Because you watched
University of Flo...

# Motivation



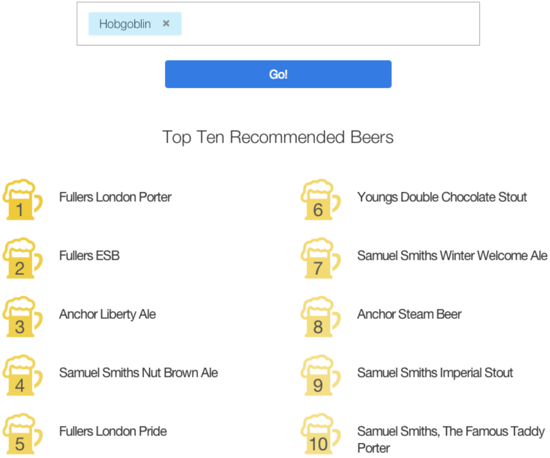
nara A life well found
Watch Video
What is Nara?
Restaurants
Home |
Looking for a particular place?
Atlanta | Boston | Chicago | Las Vegas | Los Angeles | Miami | New Orleans | New York | San Francisco | Washington, DC | More...
SAM MICHAELS
46 132
Recompute Your Finds
FINDS [50]
What Nara found for you
Filter:
$ $$ $$$ $$$$
OpenTable grubHub
Neighborhoods (56)
Cuisine Types (10)
Friends (2) BETA
Displaying Results For:
All Neighborhoods
All Cuisines
PINS [98]
Save places for later
New York Restaurants
Map
SIDEWALK CAFE
East Village, Downtown, Manhattan
Bars, Burgers • $$
(212) 473-7373
Order Online
NUMERO 28 PIZZERIA NAPOLETANA
Downtown, East Village, Manhattan
Italian, Pizza • $$
(212) 777-1555
Order Online
Reserve
NICK'S PIZZA
Queens
Pizza • $$
(718) 263-1126
JG MELON
Uptown, Upper East Side, Manhattan
Burgers, American • $$
(212) 650-1310
KESTE PIZZA & VINO
JOHN'S OF 12TH STREET
Downtown, East Village, Manhattan
Italian • $$
(212) 475-9531
Order Online

# Motivation



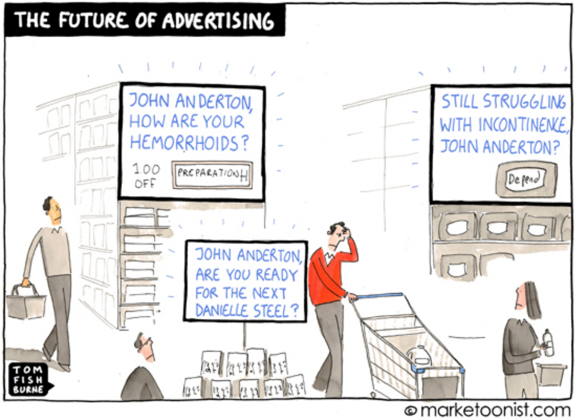

http://open.blogs.nytimes.com/2015/08/11/building-the-next-new-york-times-recommendation-engine/

# Motivation




http://blog.yhathq.com/posts/recommender-system-in-r.html

# Motivation: Personalization



# Motivation: Music

Musicroamer

Spotify

PANDORA

# Problem setting

## Standard scenario

1. A set of users has initially rated some subset of songs that they have already listened to.
2. The recommender system uses these known ratings to predict the ratings that each user would give to those not rated songs by him/her.
3. Recommendations of songs are made to each user based on the predicted ratings.

# Problem setting

## Notation

$\mathcal{U}$: set of users

$\mathcal{S}$: set of songs

$p(u, s)$: usefulness of song $s \in \mathcal{S}$ to user $u \in \mathcal{U}$, i.e.,

$$p : \mathcal{U} \times \mathcal{S} \to \mathcal{R}$$

where $\mathcal{R}$ is a totally ordered set (e.g., non-negative integers or real numbers in a range)

## Objective

- **Learn** $p$ based on the available data
- Use $p$ to **predict** the utility value of each song $s$ to each user $u$

# Two main approaches

### Collaborative filtering

The user will be recommended songs that people with similar tastes and preferences liked in the past.

### Content-based recommendations

The user will be recommended songs that are similar to the songs the user preferred in the past.

Hybrids: combine collaborative and content-based methods

[Idea] predict the utility of songs for a user based on the items previously rated by other like-minded users!

| Song / User | **Wonderwall** | **Layla** | **Creep** | **Yesterday** |
|---|---|---|---|---|
| **Mary** | 1 | 0 | 1 | 1 |
| **Jane** | 1 | 0 | 1 | 0 |
| **Jack** | 1 | 0 | 0 | 0 |

Includes both:

- User-based methods
- Item-based methods

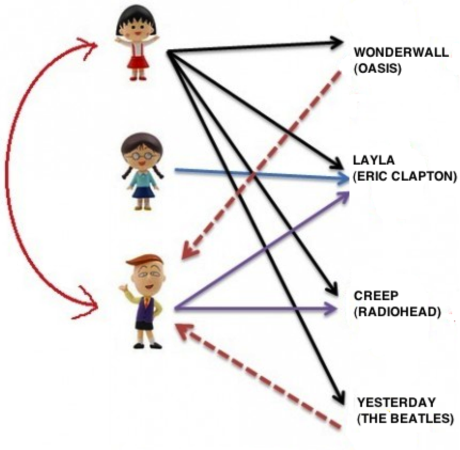
[Idea] predict the utility of songs for a user based on the items previously rated by other like-minded users!

| User \ Song | **Wonderwall** | **Layla** | **Creep** | **Yesterday** |
|---|---|---|---|---|
| **Mary** | 1 | 0 | 1 | 1 |
| **Jane** | 1 | 0 | 1 | 0 |
| **Jack** | 1 | ? | ? | ? |

Includes both:

- User-based methods
- Item-based methods

# Collaborative filtering: user-based



# Collaborative filtering: user-based

$\mathbf{u} = (u_1, u_2, \ldots, u_I) \rightarrow$ profile of the *target* user
$\mathbf{v} = (v_1, v_2, \ldots, v_I) \rightarrow$ profile of a *neighbour* user

## Similarity

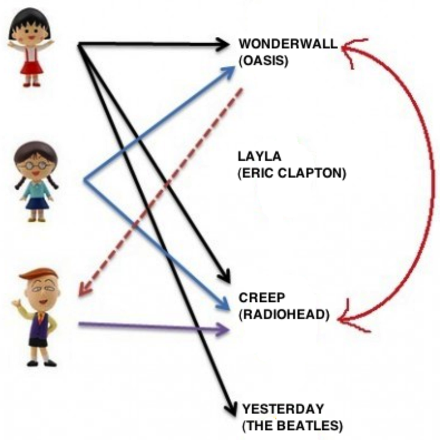
The similarity between **u** and **v** can be calculated using the **Pearson's correlation coefficient**:

$$Sim(\mathbf{u}, \mathbf{v}) = \frac{\sum_i (u_i - \bar{\mathbf{u}})(v_i - \bar{\mathbf{v}})}{\sqrt{\sum_i (u_i - \bar{\mathbf{u}})^2} \sqrt{\sum_i (v_i - \bar{\mathbf{v}})^2}}$$

Select the $K$ most similar users to the target user and generate a predicted value of user $u$'s rating

Rating prediction of song $i$ for target user **u**

$$\hat{u}_i = \bar{\mathbf{u}} + \frac{\sum_{k=1}^{K} Sim(\mathbf{u}, \mathbf{v}^k)(v_i^k - \bar{\mathbf{v}}^k)}{\sum_{k=1}^{K} |Sim(\mathbf{u}, \mathbf{v}^k)|}$$


WONDERWALL
(OASIS)
LAYLA
(ERIC CLAPTON)
CREEP
(RADIOHEAD)
YESTERDAY
(THE BEATLES)

## Collaborative filtering: item-based

$\mathbf{u} = (u_1, u_2, \ldots, u_I) \rightarrow$ profile of an user

### Similarity

Compare songs based on their pattern of ratings across $n$ users:

$$Sim(i,j) = \frac{\sum_n (u_i^n - \bar{\mathbf{u}}^n)(u_j^n - \bar{\mathbf{u}}^n)}{\sqrt{\sum_n (u_i^n - \bar{\mathbf{u}}^n)^2}\sqrt{\sum_n (u_j^n - \bar{\mathbf{u}}^n)^2}}$$

Select a set of $J$ most similar songs to the target song and generate a predicted value of user $u$'s rating

Rating prediction of song $i$ for target user $\mathbf{u}$

$$\hat{u}_i = \frac{\sum_{j=1}^{J} Sim(i,j) u_j}{\sum_{j=1}^{J} Sim(i,j)}$$

- **Rule-based**

  Find item association rules, e.g., $buy(flour), buy(sugar) \rightarrow buy(eggs)$

- **Matrix factorization**

  Find latent structures, e.g., SVD

- **Clustering**

  Find representatives, e.g., $K$-means

# Collaborative filtering: drawbacks

[Scalability] comparison of the target user to all user records!

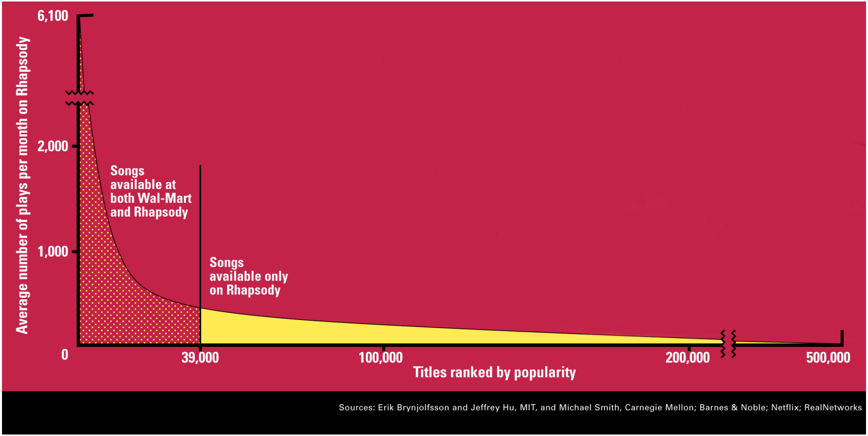
[Sparsity] many users, many songs

[Privacy] once a system has built your profile, who else can have access to it?

[Cold start] what if we don't know anything about you?

[Long tail] what is the effect of popularity?

# Collaborative filtering: long tail



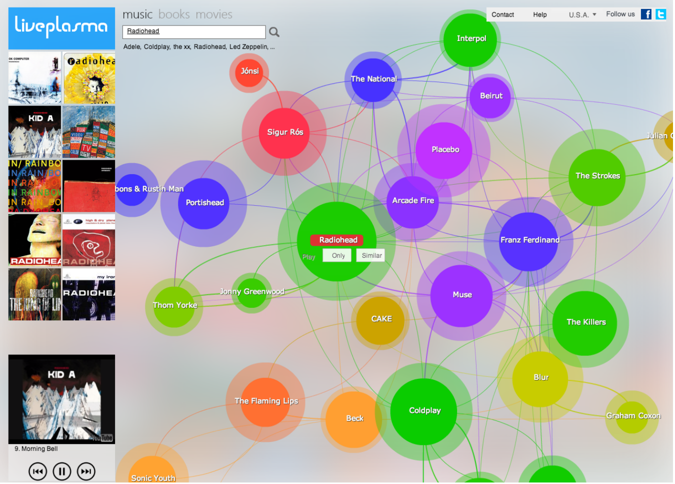
[Idea] song recommendations by predicting the utility of songs for a particular user based on how *similar* the songs are to those that he/she liked in the past.

**Example: Movie recommendation**

A movie may be represented by such **features** as specific actors, director, genre, subject matter, etc. The users interest or preference is also represented by the same set of features.

What features are relevant for songs?

# Content-based: similarity




http://liveplasma.com/

1. Analyse the content of each song: feature extraction
2. Compare the user profile with each songs
3. Recommend the top $K$ most similar songs

DEMO

# Content-based: drawbacks

[Effort] requires some effort from programmers (domain knowledge)

[Effort] requires some effort from users

[Changes] cannot cope with changes in users interests

[Privacy] once a system has built your profile, who else can have access to it?

Do we really want similar songs?

# Social media

# Evaluation

Sales Traffic (clicks, time, ...) Satisfaction Loyalty

- 5 – 20% increase in sales
- 60% use "recommendations" to determine suitable product
- In 2011 15% of customers admitted to buying recommended products, 2013 nearly 30%

yelp: 25% YOY Growth

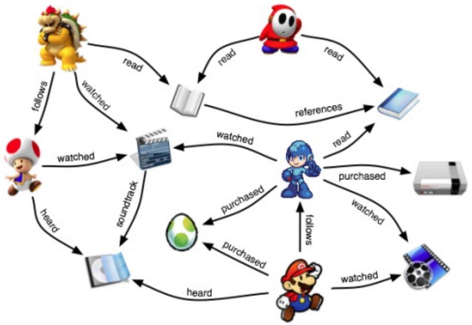
NETFLIX: 36 Million subscribers
60-70% view results from recommendation

tripadvisor: Trip Advisor collaborates with EBAY, ORBITZ and others.

PANDORA: Tens of Billions "Thumbs up"
60 Million active users
3.8 billion hours of music (last Qtr)
47% up-tic in active users
67% increase in music served

Hortonworks

# Future




http://blog.soton.ac.uk/hive/files/2012/05/recsys-data2.jpg

# Next lecture

We will have a closer look at different unsupervised learning approaches!

Have a look at:

http://dataconomy.com/2017/06/deep-learning-personalizing-internet/

https:
//www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html?_r=0