

Lecture 3: Recommender systems



Applied Data Science/Feb 2017/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.
- ... 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](#)

[12" Footstool](#)

[Noose](#)

★★★★★ (1) \$94.99 ★★★★★ (10) \$36.99 ★★★★★ (2) \$19.99

Motivation

YouTube
Broadcast Yourself™

Search

Home Videos Channels

Add / Remove Modules

Recommended for You

Edit

Video Title	Length	Published	Views	Reason Suggested
Three Days Grace - Just Like You	3:15	5 months ago	11,067,416 views	Because you watched Three Days Gay
Radiohead - 15 Step (played by C...)	3:56	1 year ago	31,645 views	Because you watched Sarah doing her g...
Hotel Mario Remixed	1:47	1 year ago	284 views	Because you watched Youtube Poop: Spo...
Miley Cyrus - Party In The U.S.A...	3:21	5 months ago	95,168,141 views	Because you watched Man Almost Dies.
Cannibal Holocaust part 2 of 10.avi	10:01	1 week ago	707 views	Because you watched Cannibal Holocaus...
Mortal Kombat Theme	3:21	4 years ago	16,597,849 views	Because you watched Fat Kid Kung Fu!
Call of Juarez: Bound in Blood ~..	10:01	8 months ago	64,335 views	Because you watched Call of Juarez: B...
Wanted Drug Dealer Getting Tasered	1:25	1 year ago	296,890 views	Because you watched University of Flo...

WTF?

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...

New York Restaurants

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

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

JG MELON Uptown, Upper East Side, Manhattan Burgers, American • \$ (212) 650-1310 Reserve

JOHN'S OF 12TH STREET Downtown, East Village, Manhattan Italian • \$ (212) 475-9531 Order Online

NICK'S PIZZA Queens Pizza • \$\$ (718) 263-1126

KESTE PIZZA & VINO

BOCCO'S Midtown, Upper West Side Italian • \$ (212) 580-2100

Map

Motivation

HOME PAGE | TODAY'S PAPER | VIDEO | MOST POPULAR | TIMES TOPICS

The New York Times

Recommendations

Articles Recommended for You

1 TECHNOLOGY
Apps Alter Reading on the Web
By JENNA WORTHAM | Feb 1, 2011
One service that removes ads from articles for easier reading is introducing a new way to compensate publishers.
Books and Literature; Newspapers; Magazines; Mobile Applications; Online Advertising; Wireless Communications;



2 TECHNOLOGY
Apple Moves to Tighten Control of App Store
By CLAIRE CAIN MILLER and MIGUEL HELFT | Feb 1, 2011
Sony said Apple had rejected its e-book app because it did not route book sales through Apple's system.
Electronic Books and Readers; Mobile Applications; iPad; iPhone;

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

Motivation



Top Ten Recommended Beers



Fullers London Porter



Fullers ESB



Anchor Liberty Ale



Samuel Smiths Nut Brown Ale



Fullers London Pride



Youngs Double Chocolate Stout



Samuel Smiths Winter Welcome Ale



Anchor Steam Beer

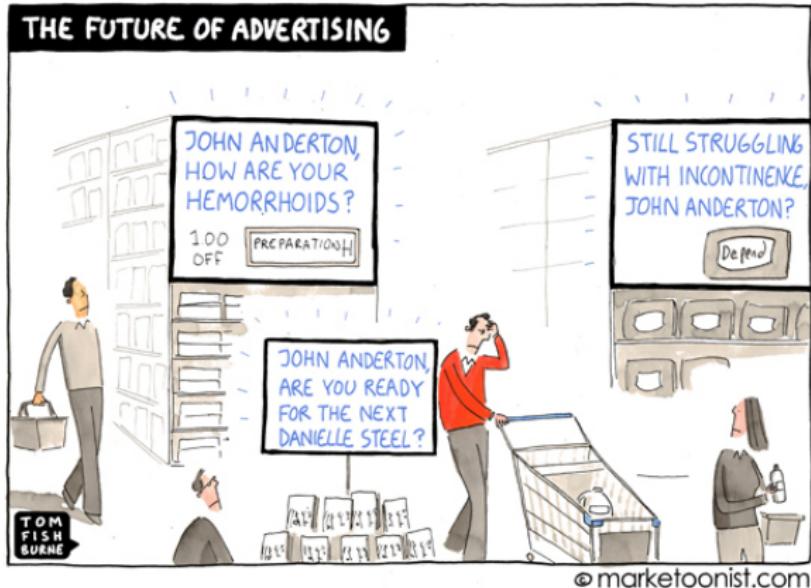


Samuel Smiths Imperial Stout



Samuel Smiths, The Famous Taddy Porter

Motivation: Personalization



Motivation: Music



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.,

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

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

Collaborative filtering

[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	0	0	0

Includes both:

- User-based methods
- Item-based methods

Collaborative filtering

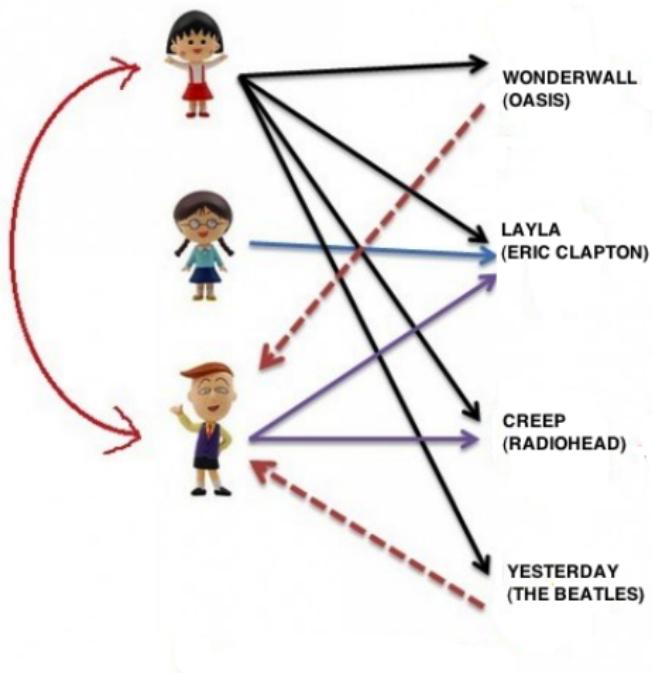
[Idea] predict the utility of songs for a user based on the items previously rated by other like-minded users!

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

$\mathbf{u} = (u_1, u_2, \dots, u_l)$ → profile of the *target* user

$\mathbf{v} = (v_1, v_2, \dots, v_l)$ → profile of a *neighbour* user

Similarity

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

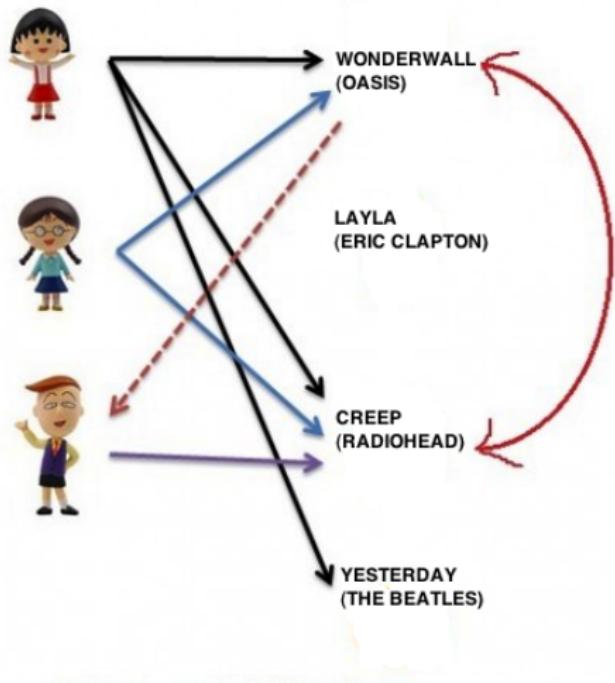
$$Sim(\mathbf{u}, \mathbf{v}) = \frac{\sum_i (u_i - \bar{u})(v_i - \bar{v})}{\sqrt{\sum_i (u_i - \bar{u})^2} \sqrt{\sum_i (v_i - \bar{v})^2}}$$

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 \mathbf{u}

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

Collaborative filtering: item-based



Collaborative filtering: item-based

$\mathbf{u} = (u_1, u_2, \dots, u_l) \rightarrow$ 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!

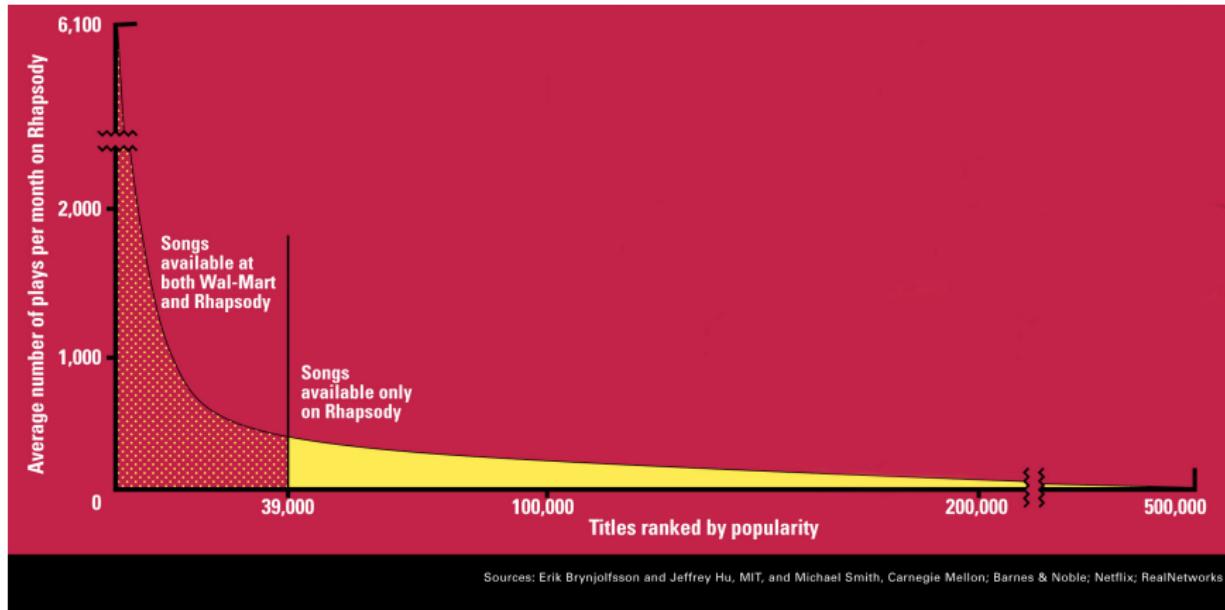
[Sparsity] many users, many songs

[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



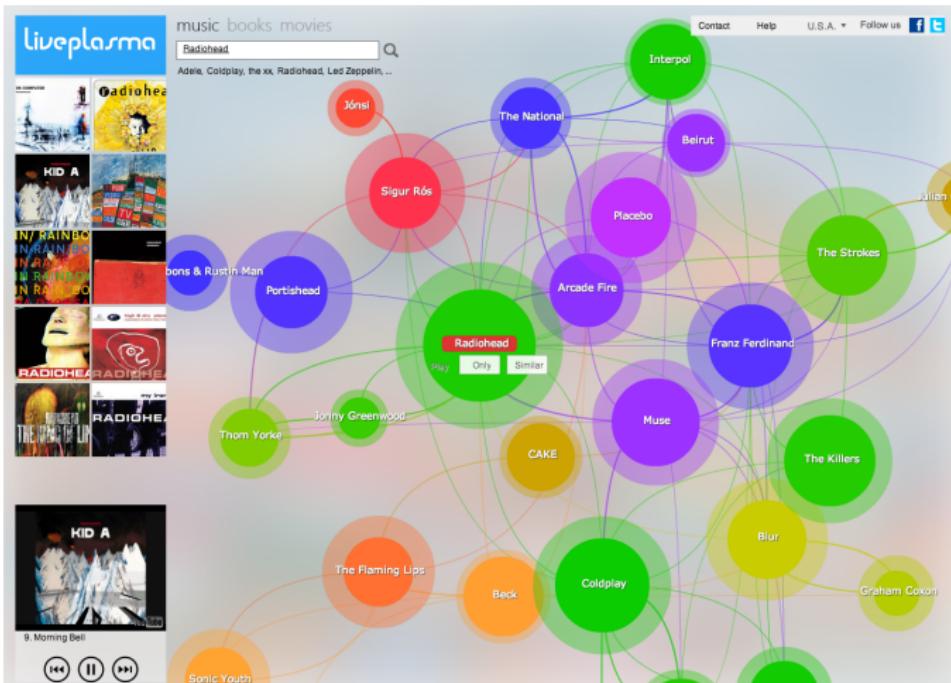
[Idea] song recommendations by predicting the utility of songs for a particular user based on how *similar* the songs are to those that he/she liked in the past.

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/>

Content-based: similarity

- 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?



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%



25% YOY Growth



36 Million subscribers
60-70% view results from recommendation

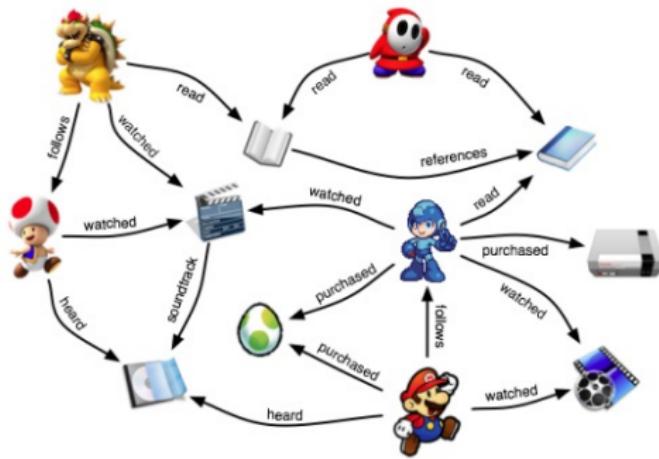


Trip Advisor collaborates with EBAY, ORBITZ and others.



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





Next lecture

We will have a closer look at different data storage solutions!

Have a look at:

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