Recommender Systems

Week 11 - Day 02

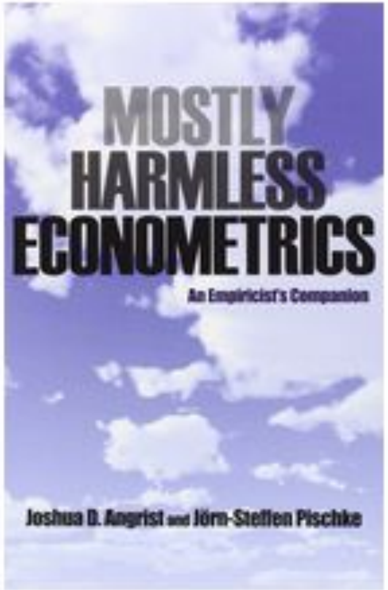
What is a recommender system?

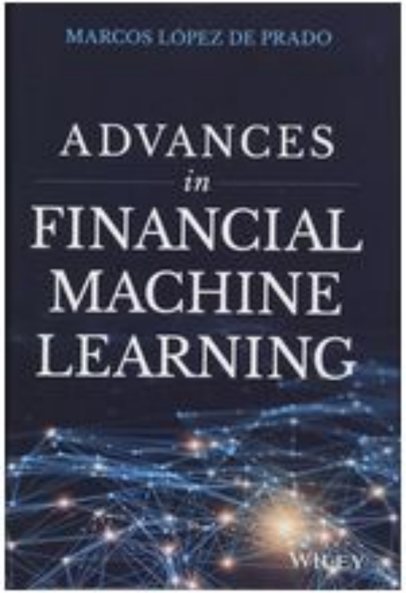
A model/system to recommend

(usually) objects to users

You use them very often

Recommendations for You, edoardo






Recommended For You
Learn more
LOVE, SIMON
BEAU TRAVAIL
ORSON WELLES
CITIZEN KANE
ROBERT DE NIRO
"RAGING BULL"
Back to the Future Part III
Back to the Future Part III
(1990)
PG Adventure | Comedy | Sci-Fi
7.4/10
Enjoying a peaceable existence in 1885, Doctor Emmet Brown is about to be killed by Buford "Mad Dog" Tannen. Marty McFly travels back in time to save his friend.
Add to Watchlist
Next »
◀ Prev 6 Next 6 ▶
Director: Robert Zemeckis
Stars: Michael J. Fox, Christophe...


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
Breaking Bad
HOUSE of CARDS
the L word
ALWAYS SUNNY PHILADELPHIA
Because you watched Red Lights
6 SOULS
ELECTRIC MIST

Suggested Post

Top Kickstarter Inventions

Sponsored ·

Like Page

You bring the books, and this elegant tool will make them more comfortable to read.

Deals recommended for you

See all deals

$13.99 - $15.99

Ends in 23:25:39

$7.69 - $371.25

Ends in 23:30:38

$79.99 - $499.99

Ends in 23:25:39

Daniel Kibblesmith ⛄

@kibblesmith

Follow

Amazon is a $250 billion dollar company that reacts to you buying a vacuum by going THIS GUY LOVES BUYING VACUUMS HERE ARE SOME MORE VACUUMS

Online Magazines, Youtube,

Carousell, Tripadvisor, Spotify, etc.

Can you think about a baseline?

The most popular object(s)!

Content-Based Filtering

Recommend similar objects

Lord of the ring 1

Lord of the ring 2

"Element of statistical learning"

↓

"Statistical learning: introduction"

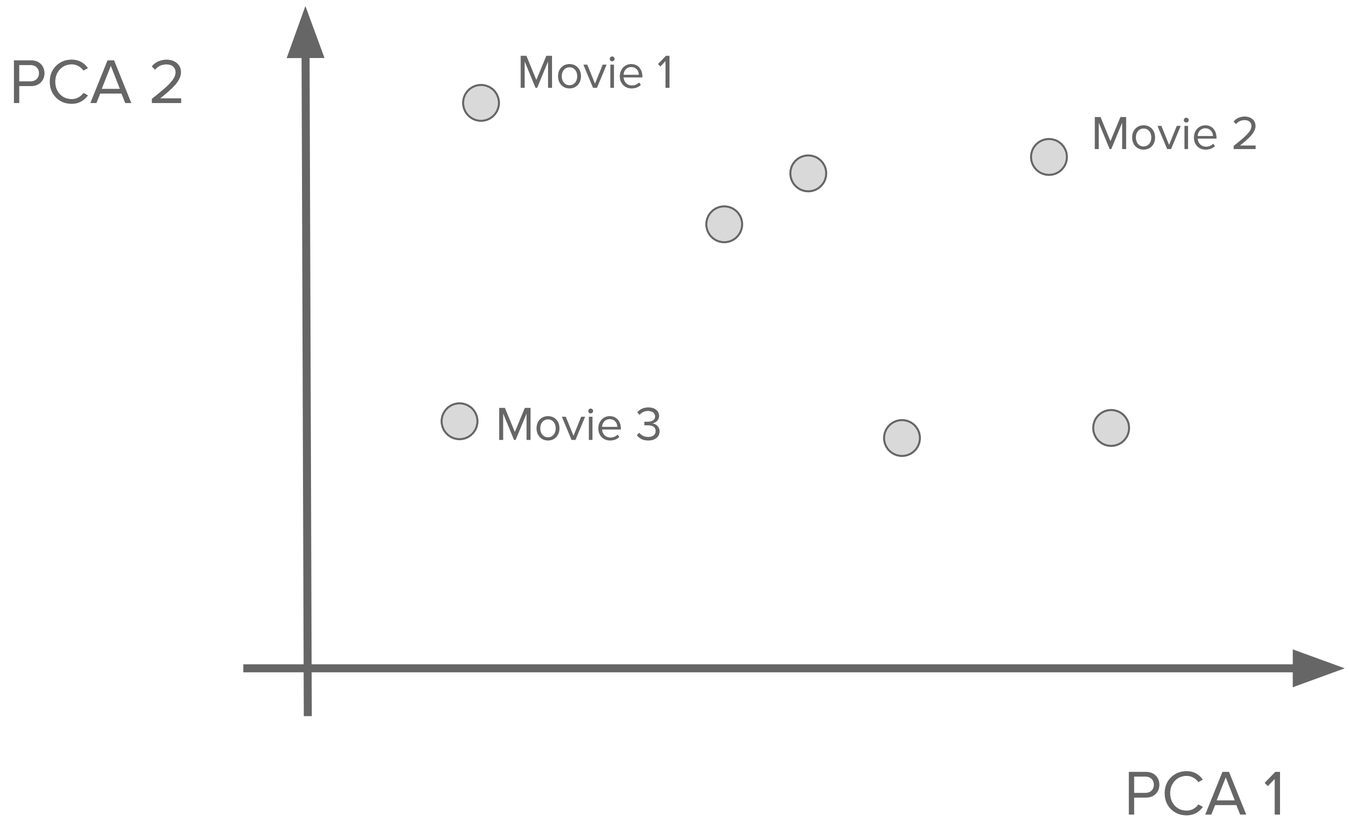
Romantic Movie

Good for kids

Movie 1 = [0, 8, 7, 1, 0, 2, 3]

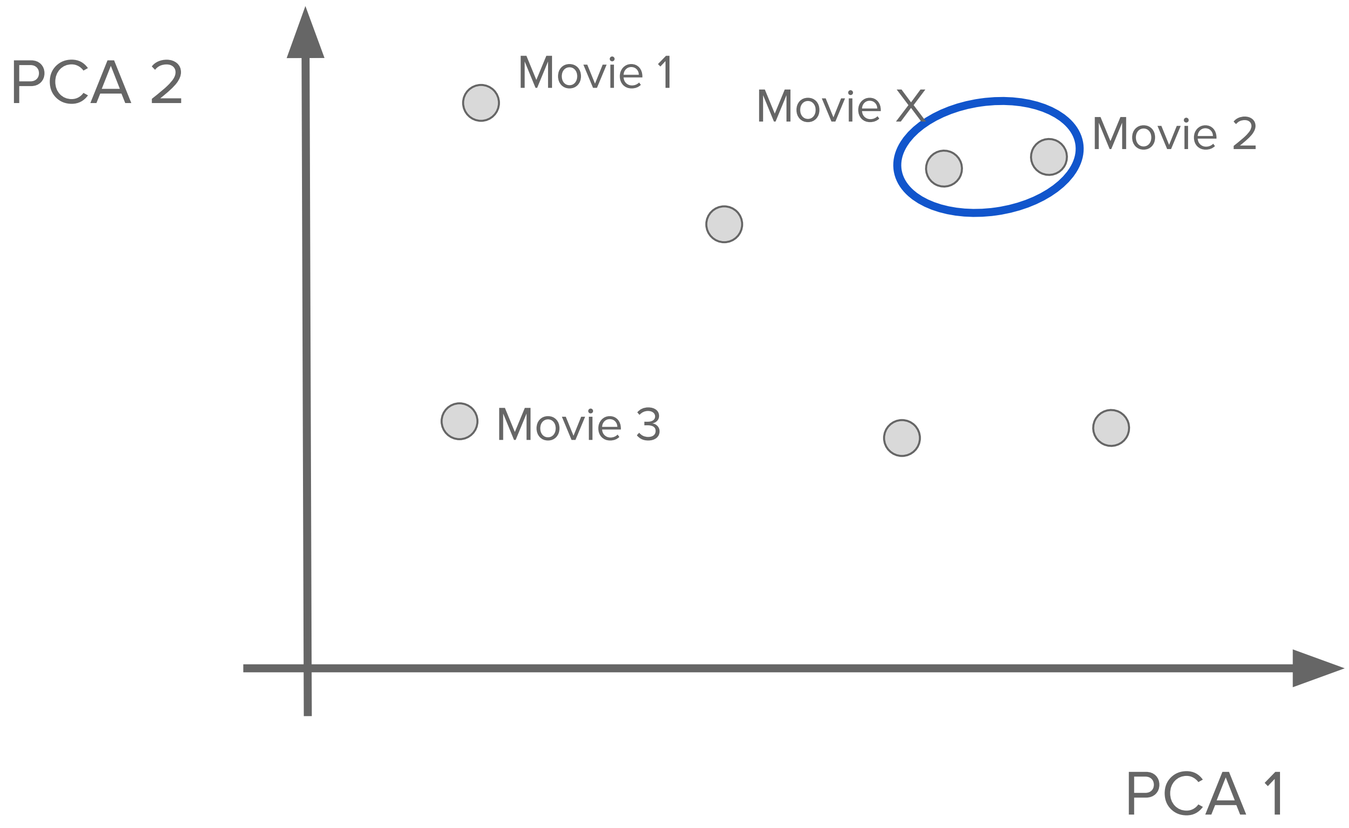
Movie 2 = [2, 1, 0, 0, 3, 3, 1]

Movie 3 = [8, 1, 2, 9, 0, 0, 2]


PCA 2
Movie 1
Movie 2
Movie 3
PCA 1

Use case 1

Distance (movie, movie)


PCA 2
Movie 1
Movie X
Movie 2
Movie 3
PCA 1

Use case 2

Distance (user, movie)

Movie 1 = [1, 0, 3]

Movie 2 = [3, 0, 0]

Movie 3 = [0, 0, 2]

User = [1, 1, 2]

User x Movie 1 = 1*1 + 0*1 + 3*2 = $\boxed{7}$

User x Movie 2 = 3*1 + 0*1 + 0*2 = 3

User x Movie 3 = 0*1 + 0*1 + 2*2 = 4

Problem

The object description

is not always straightforward

Collaborative Filtering

Recommend what similar users like

It doesn’t care about

describing the objects!

Rating “Lord of the ring”

Rating “La dolce vita”

User 1 = [0, 8, 7, 1, 0, 2, 3]

User 2 = [2, 1, 0, 0, 3, 3, 1]

User 3 = [8, 1, 2, 9, 0, 0, 2]

18,000 movies

480,000 users

x	1	1	x	...	x
x	x	x	5	...	x
x	x	3	x	...	x
x	4	3	x	...	2
...	x	x	x	...	x
x	5	x	1	...	x
x	x	3	3	...	x
x	1	x	x	...	2

Which movie would you recommend to User X?

	Movie 1	Movie 2	Movie 3	Movie 4	Movie 5
User 1	7	8	2	5	9
User 2	3	2	8	8	3
User X	8	7	3		

1. What are the most similar users?
2. What do they like?

Similarity scores

(jaccard, euclidean, etc.)

Content
Vs.
Collaborative

Content = describe objects and find similar objects

Collaborative = find most similar users and check what they like

Feedback

Collaborative Filtering Recom. Sys.

are based on preferences/feedback

Explicit feedback

Vs.

Implicit feedback

Explicit = actively provided by users

Implicit = provided by users “without intention”

Amazon review: explicit or implicit?

Tinder swap: explicit or implicit?

Click on a FB ad: explicit or implicit?

Implicit feedback are everywhere

- Email impressions
- Email click-throughs
- Conversions
- Demographic
- Session lengths
- Login attempts
- Track plays
- Money spent
- Ad impressions
- Ad clicks
- Ad click-purchase
- Web “click depth”
- # of swipes
- Profile views
- Message initiations
- Poll Votes
- Friend / unfriend
- Follow / unfollow
- *Like
- Post text
- Image EXIF
- Friends in common
- Message text
- Food purchases
- Geospatial data
- Store cameras
- Wifi logins / MAC
- Time series
- Objects in photos
- Driving record
- Credit history
- Topics most read

Binary feedback (0/1)

Vs.

Numeric feedback (1,2,3,4,5)

Cold Start

Collaborative:

User = [8,2,3,0,0,1]

What if a user has no ratings/reviews?

Cold start problem!

Jaccard Similarity

$$similarity(user1, user2) = 7.43$$

“Shared” Objects

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

All objects


IoU: 0.4034
IoU: 0.7330
IoU: 0.9264
Poor
Good
Excellent

What’s the similarity?

User 1 = {**Lotr1, Lotf2**, Bttf 1, Bttf 2}

User 2 = {**Lotr1, Lotf2**, Aven 1, Aven 2}

Shared objects = Lotr1, Lotr2

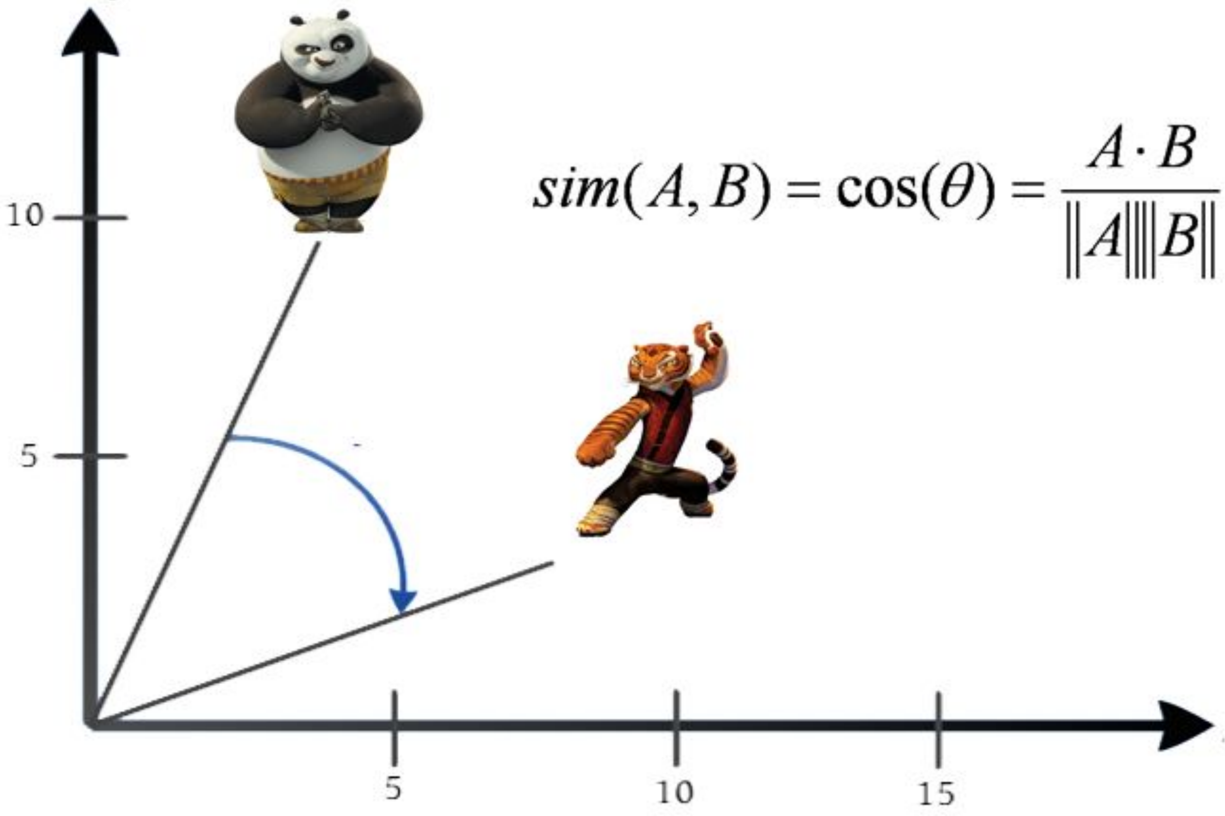
All objects = Lotr1, Lotr2, Bttf1, Bttf2, Av1, Av2

Jaccard = 2/6

Cosine Similarity

Angle between two vectors

Cosine Similarity



$$sim(A,B) = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|}$$

```
a = np.array([2, 0, 1, 1, 0, 2, 1, 1])
b = np.array([2, 1, 1, 0, 1, 1, 1, 1])

np.dot(a,b) / (np.sqrt(sum(np.square(a))) * np.sqrt(sum(np.square(b))))
```

Testing

1. Offline

Train + Test

Lord of the Ring

La Dolce Vita

User 1 = [4, 2, 3, 6, 9]

Let's try to predict this

Other common approach: Recall

2. Online

Put different models in production

(test phase)

A/B testing + CLTV

Nice article about evaluation techniques

Other aspects

$ Revenue $

Serendipity

Diversity

Privacy

Summary

- Rec.Sys. are everywhere
- Content vs. Collaborative
- Similarity Score
- Offline vs. Online testing
- Revenue, Serendipity, etc.