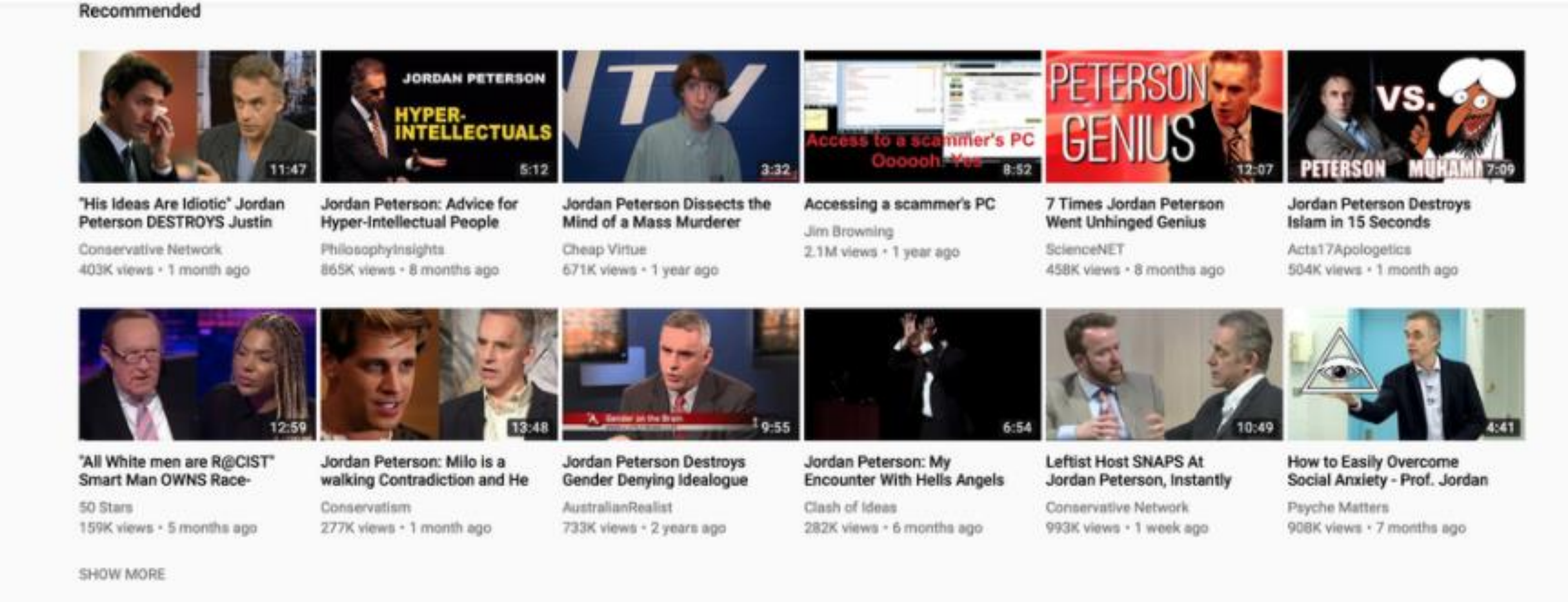
Personalization

A Collaborative Filtering Approach

The YouTube recommendations algorithm is way too reactive. I watched one Jordan Peterson video and this is my home page.



Nakilis 2 years ago

I couldn't agree more. After watching one Tyler Perry interview on Jimmy Fallon, all of my recommendations are now Tyler Perry and Jimmy Fallon related.. And not all of the other content that I watch in ungodly amounts. But sure, Youtube still knows what they're doing.

1 Share Report Save

Poenaconda 2 years ago

I once watched ONE video from a creator I enjoy. The next day my ENTIRE recommended was their videos. I understand that YouTube thinks I will watch them but that is insane.

1 Share Report Save

有人只是去参加了一次北京婚博会，晚上回到家打开微博和微信，发现信息流广告全部变成了婚纱照、婚庆公司、婚礼礼服等。令他感到恐怖的是在此之前从未在手机进行过结婚相关的任何搜索。这一切发生改变的原因仅仅是因为他去了一次婚博会这个地方而已。

有人在知乎看到除甲醛的相关问题，只是百度了一下，结果连一个美食app都开始推荐除甲醛公司。在百度上打开某理财网站，不到半小时推销电话就打过来了。

有时候你在网上搜了一本小说，然后突然就会有很多假网站在百度上显示他们网站有这本小说可以下载，然后让百度把他推在首页，你打开链接一看其实里面没有，但是有其他东西的广告。

TECHNOLOGY

Google Knows You Better Than You Know Yourself

Predictive analysis combs through calendars and search histories—and gets in the way of routine self-deception.

JAMES CARMICHAEL AUGUST 19, 2014

Facebook Knows You Better than You Know Yourself

Erman Misirlisoy, PhD Oct 18, 2018 · 7 min read ★

The Internet Knows You Better Than You Know Yourself

When Amazon or eBay recommend us something we like but were not looking for, they effectively know us better than we know ourselves.

QUESTION

What data do firms collect from us?

DATA THAT FITMS COLLECT

Firms also collect other types of data. Let's take Google as an example. Do you know what data that Google collects from you?

I'M
WATCHING YOU
facebook

DATA THAT FITMS COLLECT

Even the traditional brick-and-mortar (offline) shops are also collecting your data.

- Your payment method (Credit? Mobile pay? Cash?)
- Loyalty program information (Are you using Yuu?)
- Personal profile (If you ever registered there…)

DATA THAT FITMS COLLECT

With new technologies, brick-and-mortar stores can also get much more information than what they had before.

As described in the video, if you use the free Wi-Fi they provide you, they will be able to collect data from your smartphone!

Facial recognition and mobile payments help collect data from you ("刷脸支付").


BUYING FOOD
WITH FACIAL
RECOGNITION

QUESTION

Suppose that you are an Internet company, and you have access to all this consumer data, what would you do?

Price Discrimination

Broadly speaking, personalized pricing is a form of price discrimination. Let's review types of price discrimination (video here):

1st degree: The firm sells a product at the maximum price that every consumer is willing to pay.

2nd degree: price varies according to quantity demanded.

3rd degree: charging a different price to different consumer groups.

Examples

Are you using a Mac or a PC?

THE WALL STREET JOURNAL.

English Edition ▾ | Print Edition | Video | Podcasts | Latest Headlines

Home World U.S. Politics Economy Business Tech Markets Opinion Life & Arts Real Estate WSJ. M

On Orbitz, Mac Users Steered to Pricier Hotels

On Orbitz, Mac users spend as much as 30% more a night on hotels that PC users do.

Examples

Websites Vary Prices, Deals Based on Users' Information

Getting Different Deals Online

A Journal examination found online retailers adjusted prices by a shopper's location, among other factors

Staples.com
SnapSafe Titan safe
HIGHER PRICE $1,199.99
DISCOUNT PRICE $1,099.99
DIFFERENCE: 9.1%

Homedepot.com
A 250-foot spool of electrical wiring
Six pricing groups, including:
$70.80 in Ashtabula, Ohio
$72.45 in Erie, Pa.
$77.87 in Monticello, N.Y

Rosettastone.com
A 20% DISCOUNT
...for buying multiple levels of German lessons, when test-shopping from the U.S. or Canada. But not from the U.K. or Argentina.

Photos: l to r: SnapSafe; Home Depot; Rosetta Stone Source: WSJ testing The Wall Street Journal

MOST POPULAR NEWS

1. What You Can and Can't Do if You've Been Vaccinated: Travel, Risk Factors, What You Need to Know
2. Europe Confronts Covid Rebound as Vaccine Hopes Recede
3. Biden's $1,400 Stimulus Checks Hit Bank Accounts Starting Today
4. Schumer and Gillibrand Call for Cuomo to Resign

The US retailer *Office Depots* use customers' browsing history and location data to vary prices

Examples

These Brands Have Some of the Best Abandoned Cart Email Strategies

Aug 28, 2019 5:03:58 PM

When you abandon an item from your online shopping cart, e-tailers may issue you a discount to lure you to make a purchase.

Behavior-Based Pricing

The more common approach is pricing with consumers' purchase history, a practice known as "behavior-based pricing".

The idea is very simple: The price you receive depends on whether or not you have purchased the products before. In other words, we offer new and existing consumers different prices.

QUESTION

Suppose that a firm uses “behavior-based pricing”, how should the firm charge its prices? Should the firm offer new consumers a higher or lower price?

QUESTION

In most cases, firms offer high prices to existing consumers and lower prices to new consumers. But why?

Is It Legal?

While consumers often object to personalized pricing, it is legal in most countries.

In 1996, a consumer living in Manhattan sued Victoria's Secret for distributing different versions of catalogs with identical items but different prices. However, the New York Court dismissed the claim by noting that it was an accepted business practice to reward repeat consumers or to draw in new consumers with special savings.

Is It Legal?

Any form of price discrimination is legal in the United States, as long as the basis of discrimination is not race, religion, national origin, gender, and the like.

Recently, China banned behavior-based pricing in the traveling and hospitality industry. According to a 2020 regulation by the Ministry of Culture and Tourism, online traveling website is not allowed to offer consumers discriminated prices (see news here).

In the EU, there is a recent GDPR regulation on big data.

EU’s GDPR regulation

QUESTION

As an individual consumer, do you like personalized pricing? What should you do when you know firms are using personalized pricing?

The SAD fact

Firms are spending more and more money collecting, storing and analyzing consumer data.

Consumers are also spending money and effort to avoid being recognized by firms and to outsmart firms' big data algorithm.

Prediction: As more consumers become aware of personalized pricing and take measures to avoid it, both firms and consumers can be worse off with big data and personalized pricing.

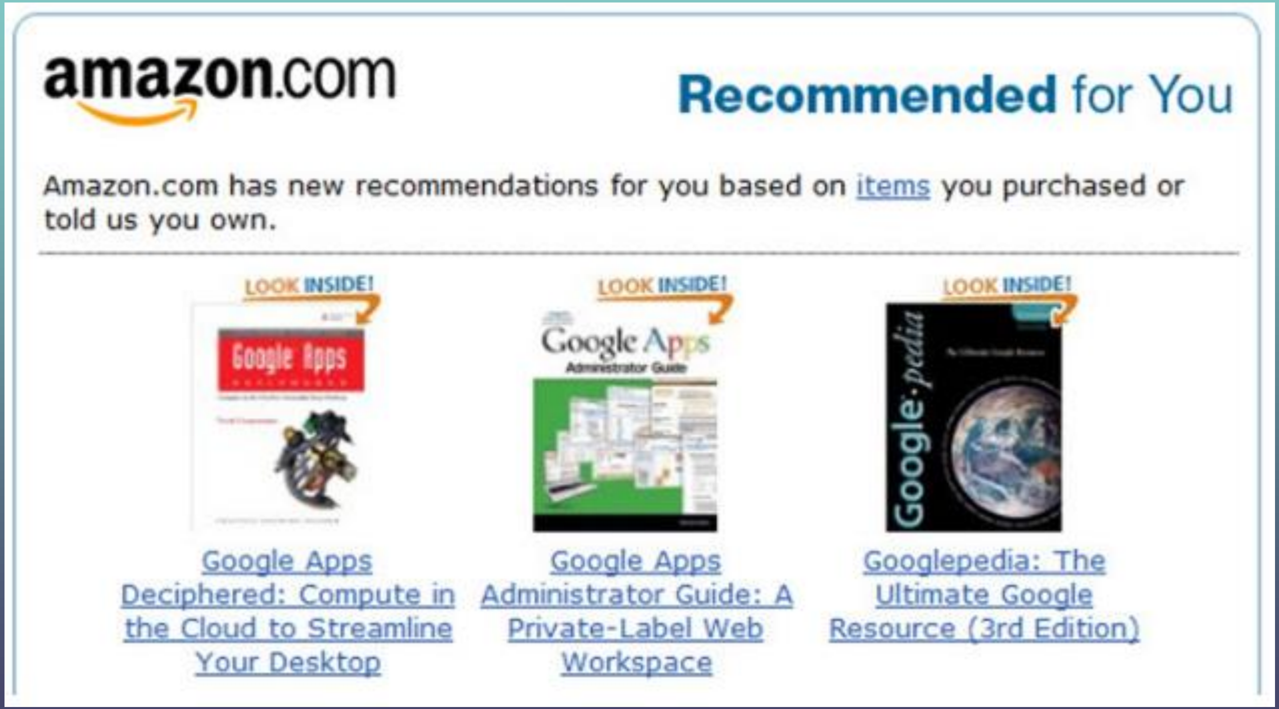
Solution: Regulation by the government.

How Firms Use Your Data

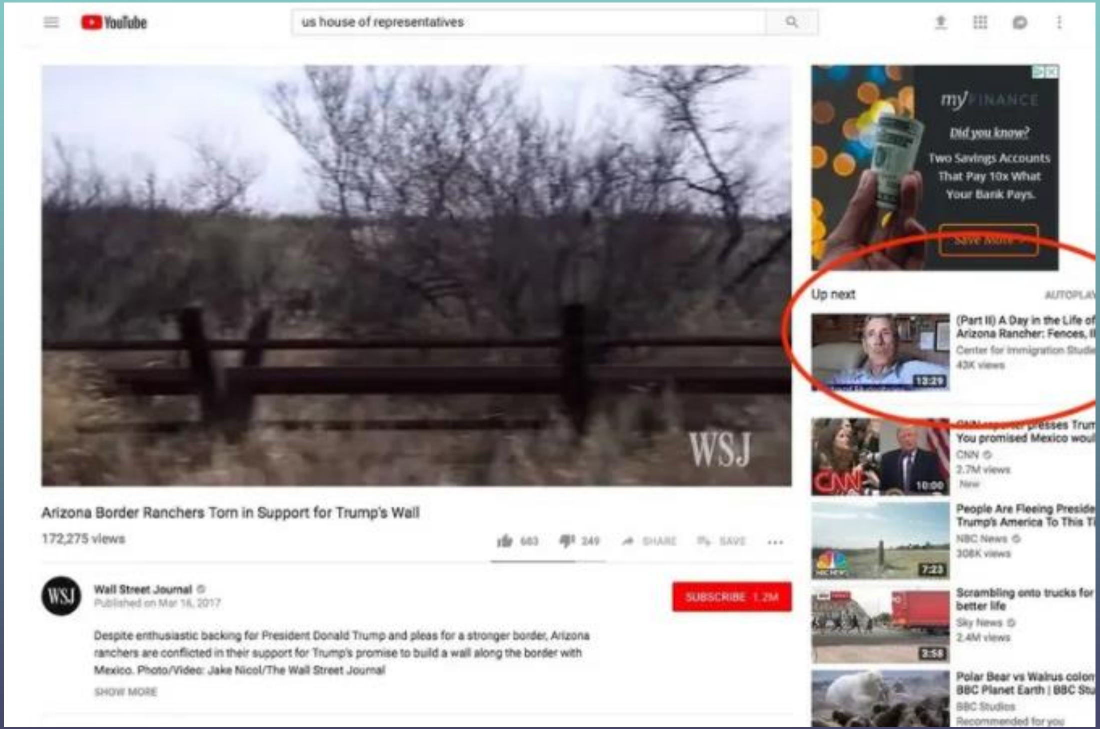
Personalized Pricng

Personalized Recommendation

Recommendation is everywhere



Recommendation is everywhere



Recommendation is everywhere

The Importance of Recommendation

Netflix: 2/3 of the movies watched are recommended.

Google News: recommendations generate 38% more click-throughs.

Amazon: 35% sales from recommendations.

ChoiceStream: 28% of the people would buy more music if they found what they liked.

QUESTION

Have you ever thought about how these platforms and APPs make recommendations to you? Any idea?

Items

movies, songs, products, etc. (often many thousands)

Users

watchers, listeners, purchasers, etc. (often many millions)

Feedback

5-star ratings, not-clicking ‘next’, purchases, etc.

Collaborative Filtering

Everyday examples of collaborative filtering:

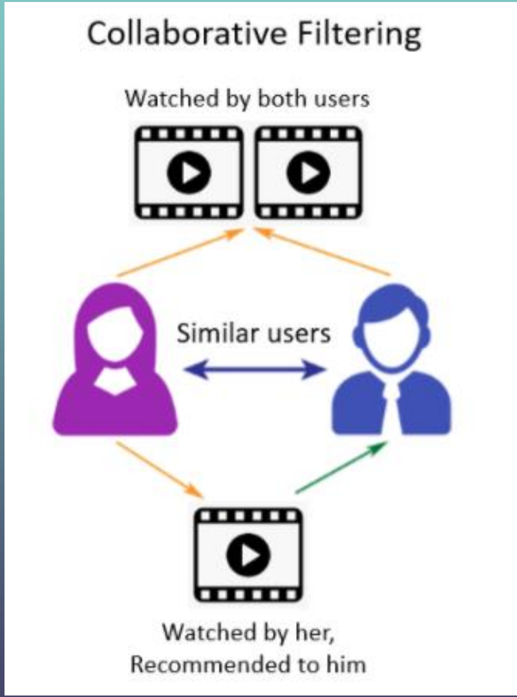
Bestseller lists, Top 40 music lists, The "recent returns" shelf at the library, "Read any good books lately?"

The intuition behind collaborative filtering: personal tastes are correlated

If Alice and Bob both like X and Alice likes Y, then Bob is more likely to like Y – especially (perhaps) if Bob knows Alice

Collaborative Filtering

In collaborative filtering, we make recommendation to one user based on the preference of similar users.

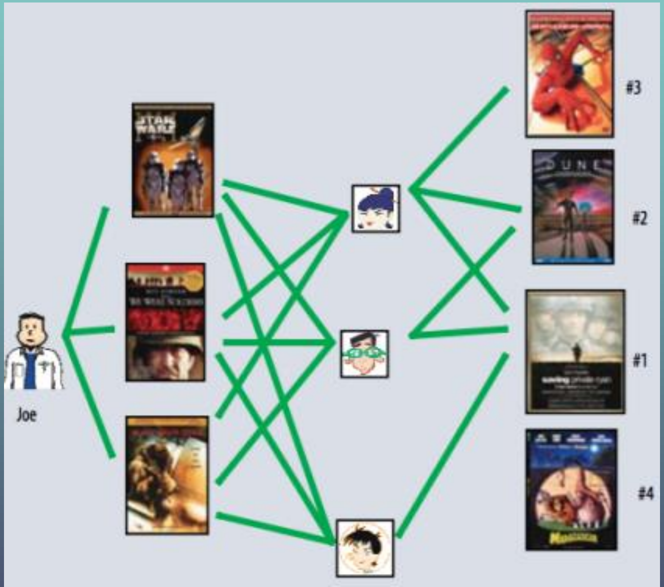


User-Based Collaborative Filtering

Item-Based Collaborative Filtering

Model-Based Collaborative Filtering

Neighborhood Method



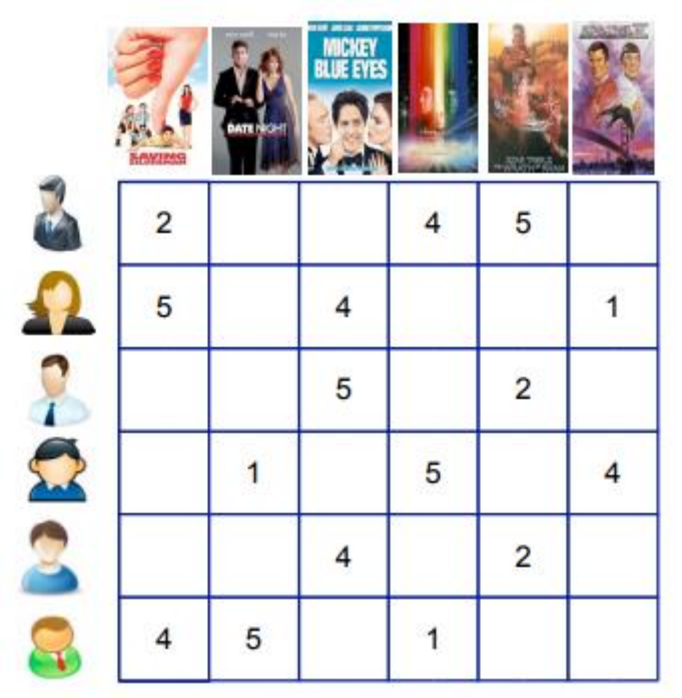
In the figure, assume that a green line indicates the movie was watched.

Algorithm:

1. Find neighbors based on similarity of movie preferences

2. Recommend movies that those neighbors watched

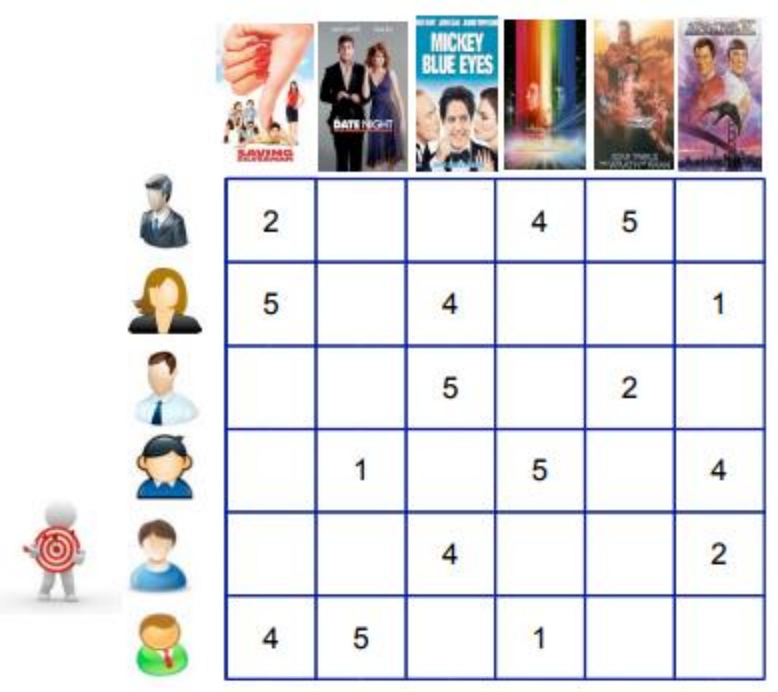
User-Based Collaborative Filtering



Each user has reviewed some items, but not every item.

We want to know their preferences for the unrated items.

User-Based Collaborative Filtering



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

Suppose that you want to under this specific user's preferences.

User-Based Collaborative Filtering

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

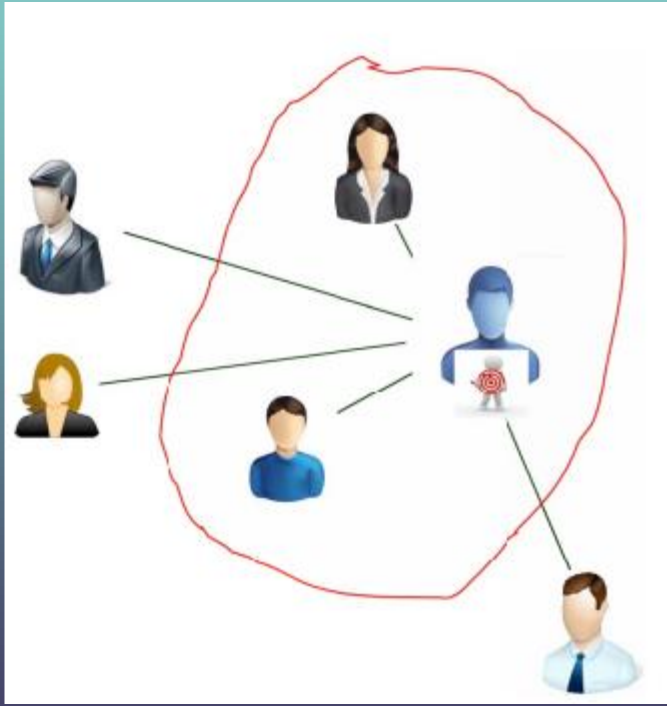
Identify items that have been rated by this user.

User-Based Collaborative Filtering

Identify items that have been rated by this user.

Identify other users that have rated the same items.

User-Based Collaborative Filtering



Compute how similar each neighbor is to the target user (similarity function). This is usually done by calculating the correlation between their ratings.

In case, select k most similar neighbors.

Make predictions based the similar neighbors' preferences.

User-Based Collaborative Filtering

Item-Based Collaborative Filtering

Model-Based Collaborative Filtering

Item-Based Collaborative Filtering

The idea is very similar to user-based collaborative filtering.

1. Identify set of users who rated the target item.
2. Identify which other items (neighbors) were rated by the users set.
3. Compute similarity between each neighbor & target item.
4. In case, select k most similar neighbors.
5. Predict ratings for the target item.

User-Based Collaborative Filtering

Item-Based Collaborative Filtering

Model-Based Collaborative Filtering

Matrix Factorization Method

Here, we assume that each individual and each movie has some “latent factors”. For movies, these factors can measure dimensions such as comedy versus drama, amount of action, or orientation to children; depth of character development or quirkiness, …

Each user has his or her preference for the factors and each movie has its value on each of these factors.

Matrix Factorization Method

Let us consider a very simple example. Suppose that there are two factors, amount of action (X) and seriousness (Y). A user also has her preference for action β_X and preference for seriousness β_Y.

When a movie has a large X, it means the movie has more actions, and when a movie has a large Y, it means the movie is more serious. Similarly, if β_X is large, it means the user prefers more actions in the movie.

Matrix Factorization Method

Then, if we know X, Y, β_X, β_Y, we can predict the user's preference for the movie, which is given by

$$\text{Preference score} = \beta_X X + \beta_Y Y$$

And we should recommend movies with the highest preference score.

Rating Matrix as a product of its factors

	1	1	F1
	0	1	F2

	F1	F2
	3	2

3	5

DOT PRODUCT →

(3x1) + (2x0)	(3x1) + (2x1)


2000 movies
Item Matrix
k features
User
Matrix
1000 users
Rating Matrix
1000 users
k features
2000 movies

Matrix Factorization Method

	Doctor Strange	Star Trek: Beyond	Zootopia
Alice	1		5
Bob	3	4	
Charlie	3	5	2

Usually, each user has only watched or rated a few movies.

So, the entire rating matrix has a lot of missing values.

We want to fill these missing values.

Matrix Factorization Method

Based on the data that we already have (i.e., existing ratings from users), we can decompose the rating matrix into the user matrix and the movie matrix.

> How to decompose? One approach is to minimize the sum of squares of errors like we do in linear regression.

Then, we can multiple these two matrix to predict a user's preference for an unwatched movie.

Matrix Factorization Method

In sum, based on what users have already watched, we can infer the user's preference for various movie attributes.

In addition, based on the ratings from the users who have watched the movie, we can infer the movie's attributes.

Finally, based on the user's preferences and the movie's attributes, we can predict a user's preference for this movie.

Matrix Factorization Method

We can compare it with linear regression:

In linear regression, we infer the value of α and β, and then we can use the regression formula $Y = \alpha + \beta X$ to make predictions.

In matrix factorization, we infer both β_X and β_Y for each user, and X and Y for each movie, and use the formula $\beta_X X + \beta_Y Y$ to predict the user's preference for the movie.

Summary Video