

Personalization and Recommendation

Here are some movie ratings.

	Movie 1	Movie 2	Movie 3	Movie 4
Alice	4	4		1
Bob		2	2	3
Carol	1	5	3	
Dennis	3		4	1
Emma	5	2	1	4
Flora	3	1		5

Predict Alice's rating for movie 3. What's your reason?

	Movie 1	Movie 2	Movie 3	Movie 4
Alice	4	4	???	1
Bob		2	2	3
Carol	1	5	3	
Dennis	3		4	1
Emma	5	2	1	4
Flora	3	1		5

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.



Netflix: How did it know I was bi before I did?

After BBC reporter Ellie House came out as bisexual, she realised that Netflix already seemed to know. How did that happen?

How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did

Kashmir Hill Former Staff

Welcome to The Not-So Private Parts where technology & privacy collide

What types of data do firms collect?

Data that firms collect

Search history: Almost every search engine collects data on your search histories. Whenever you search on Google, it knows what you are looking for, the items you viewed on Google, the webpages you clicked...

Data that firms collect

Geolocation and device information: A GPS and Wi-Fi chip are installed in every smartphone. The dozens of apps on our phones, most of them free, aren't just serving up information and entertainment. They are collecting and selling your data to digital marketers who will then offer you personalized ads.

Data that firms collect

Purchase histories: Online sellers frequently collect data on your purchase histories. They know what you have purchased and what you have not purchased. As we will figure out later, the purchase history data is the most important data for online sellers.

Data that firms collect

IP addresses: You have to get an IP address to search the internet. You can find your IP address [here](#). Then, even if you are not using a mobile device, firms can still know your geolocation (e.g., your country and even neighborhood).

Google also collects...

Number of email exchanges you've had in Gmail; number of files in Drive; number of photos Google stores for you.

Your location or searches or browsing history: Google Maps keeps track of everywhere you go and when, alongside the photos taken that day and travel times down to the minute.

Your Google Account: your photo and birthdate.

How about traditional firms?

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...)

How about traditional firms?

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.

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

But most importantly, firms use consumer data for two main purposes: pricing and recommendation.

Personalized Pricing

With personalized pricing, a seller offers each consumer an individualized price, and two persons can receive two different prices at the same time.

To implement personalized pricing, we need to satisfy three conditions. Which three?

Personalized Pricing

To implement personalized pricing, we need to satisfy the following three conditions:

- **Different valuations:** Some consumers are willing to pay more and some are willing to pay less.
- **Identifiability:** The firm must be able to identify consumers' willingness to pay.
- **No Resales:** there are no-resales among consumers.

Which of the conditions is most challenging to the firm?

Personalized Pricing

Identifiability: The firm must be able to identify consumers' willingness to pay.

The challenge: No one will tell you "I am willing to pay more." Firms must find a way to figure it out themselves.

Personalized Pricing

The airline industry is famous for personalized pricing. Imagine that you are running an airline, how can you learn about consumers' valuations to charge them personalized prices?

Personalized Pricing

Consider two persons buying air tickets from Hong Kong to Singapore on Dec 15:

- Alice books the ticket on 10 September.
- Bob books the ticket on 13 December.

Who is going to be charged a higher price? Why?

Personalized Pricing

Consider two persons buying air tickets from Hong Kong to Singapore on Dec 15:

- Alice books the ticket on 10 September.
- Bob books the ticket on 13 December.

Who is going to be charged a higher price? Why?

Early birds enjoy lower prices.

Personalized Pricing

Consider two persons buying air tickets from San Francisco to Hong Kong at the same date:

- Alice books from the airline's official website / APP.
- Bob books from an online travel agency (OTA).

Who is going to be charged a higher price? Why?

When booking from an OTA

OTA



Air ticket

\$450



**AIRLINE
OFFICIAL SITE**



Air ticket

\$500

When you book from the official website, you cannot easily compare the prices with other airlines. When booking from a third-part OTA, you can easily make price comparisons.

My Conference Trip

I attended an academic conference in Rome, Italy in 2019. The conference collaborated with a nearby hotel, and we can directly book the hotel from the conference website at a special discount. The out-of-pocket price was around 300 euros.

My Conference Trip

I attended an academic conference in Rome, Italy in 2019. The conference collaborated with a nearby hotel, and we can directly book the hotel from the conference website at a special discount. The out-of-pocket price was around 300 euros.

It turned out that when booking from another OTA, the price was only 150 euros! Why?

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

Are you using a Mac or PC? (2012)

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

Websites Vary Prices, Deals Based on Users' Information

By Jennifer Valentino-DeVries, Jeremy Singer-Vine and Ashkan Soltani

Updated December 24, 2012



It was the same Swingline stapler, on the same Staples.com website. But for Kim Wamble, the price was \$15.79, while the price on Trude Frizzell's screen, just a few miles away, was \$14.29.

A key difference: where Staples seemed to think they were located.

A Wall Street Journal investigation found that the Staples Inc. website displays different prices to people after estimating their locations. More than that, Staples appeared to consider the person's distance from a rival brick-and-mortar store, either OfficeMax Inc. or Office Depot Inc. If rival stores were within 20 miles or so, Staples.com usually showed a discounted price.

Staples charges you lower prices when you are close to its rival's offline store (2012)

< Apple iPhone 17 Pro

概述 價格 新聞 評價 規格



Apple iPhone 17 Pro 搭載最新 A19 Pro 3nm 晶片，配備 6.3 吋 LTPO Super Retina XDR OLED 螢幕，支援 120Hz 更新率，峰值亮度高達 3000 尼特。三鏡頭系統包含 48MP 主鏡頭、48MP 潛望式望遠鏡頭和 48MP 超廣角鏡頭，支援 4K 120fps 錄影。配備 12GB RAM，儲存容量最高 1TB，支援 IP68 防水。搭載 iOS 26，支援 Wi-Fi 7 和 5G 網路，提供 25W MagSafe 無線充電。

[詳細規格 >](#)

價格

全部 256GB 512GB 1TB

	博客來	\$38,190	4 件商品 >
	Yahoo購物中心 80~1%	\$38,400	12 件商品 >
	myfone網路門市	\$38,688	15 件商品 >
	誠品線上	\$38,700	6 件商品 >
	全聯	\$38,703	14 件商品 >
	全國電子	\$38,800	3 件商品 >
	momo購物網 天天下單抽【LUDEYA 微電流磁石美容儀】	\$38,900	34 件商品 >
	PChome 24h購物 註冊新會員輸碼領\$200折價券【PCNEW306】	\$38,900	11 件商品 >

A 2013 study shows that if you visit a company's website from a discount aggregator or a price-comparison website, you will get a lower price.

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. (2019)

There are rumors that some websites customize prices based on your mouse movement: Users who scrolled down the page more slowly are "deliberate buyers." They receive higher prices compared to fast scrollers.

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.

Behavior Based Pricing

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? Why?

Amazon's old customers 'pay more'



Some Amazon customers are refusing to accept some DVD prices

In 2000, behavior-based pricing first appeared to the public.
You can find the link to this phenomenon [here](#).

TRAVEL

Airfare Expert: Do cookies really raise airfares?

Rick Seaney, special for USA TODAY

Published 5:00 a.m. ET Apr. 30, 2013

This is also evidence that airlines offer higher prices to frequent travelers.

In China, this is a very vivid description of this kind of behavior, i.e., “杀熟” --- “killing existing consumers.”

同样的订单，同一家外卖平台、同一家商户、同一处送餐地址、同一个时间段，会员却比非会员支出更多——近日，有网民几次测试发现，在注册成为美团会员后，相比非会员，外卖满减优惠力度不仅有所降低，配送费也不减反增。此事再次引发舆论对互联网平台“杀熟”现象的强烈关注。

After becoming a member of Meituan, an online food delivery platform in China, you will have to suffer from higher prices and receive lower price discounts.

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

Behavior-Based Pricing

The rationale is as follows. Consumers who really like the product will make the purchase early. So, compared to new consumers, existing consumers are likely to be fans of the product and are willing to pay a higher price for it.

Following the logic, the firm can take advantage of this and charges existing consumers a higher price, i.e., punishing existing consumers (杀熟).

Nonetheless, in some other cases, firms offer high prices to new consumers and lower prices to old consumers. What makes the difference?

A Customer Management Dilemma: When Is It Profitable to Reward One's Own Customers?

Jiwoong Shin, K. Sudhir

Yale School of Management, Yale University, New Haven, Connecticut 06520
{jiwoong.shin@yale.edu, k.sudhir@yale.edu}

This study attempts to answer a basic customer management dilemma facing firms: when should the firm use behavior-based pricing (BBP) to discriminate between its own and competitors' customers in a competitive market? If BBP is profitable, when should the firm offer a lower price to its own customers rather than to the competitor's customers? This analysis considers two features of customer behavior up to now ignored in BBP literature: heterogeneity in customer value and changing preference (i.e., customer preferences are correlated but not fixed over time). In a model where both consumers and competing firms are forward-looking, we identify conditions when it is optimal to reward the firm's own or competitor's customers and when BBP increases or decreases profits. To the best of our knowledge, we are the first to identify conditions in which (1) it is optimal to reward one's own customers under symmetric competition and (2) BBP can increase profits with fully strategic and forward-looking consumers.

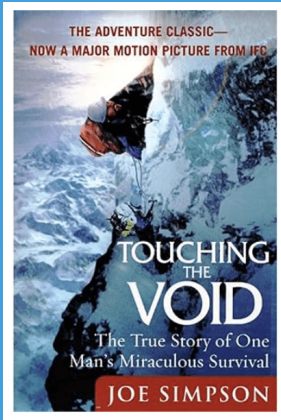
Nonetheless, in some other cases, firms offer high prices to new consumers and lower prices to old consumers. What makes the difference?

In some markets consumers are variety-seeking, i.e., their preferences change over time. For instance, when you eat sushi today probably you don't want to try sushi again tomorrow. Studies found out that for variety-seeking products, we should offer lower prices to old consumers.

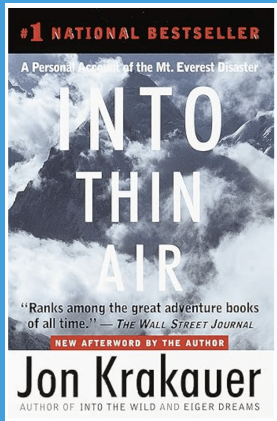
As consumers, do you like behavior-based pricing? Why?

Actually, studies show that behavior-based pricing typically hurts firms but benefits consumers. Why?

Consider the case of competition, with firm A and firm B. Firm A wants to poach firm B's consumers, and charges these consumers a low price to induce them to switch. Firm B does not want firm A to poach its own consumers, and lowers price to retain its own consumers as well. Overall, behavior-based pricing intensifies competition, benefits consumers but hurts firm profit.



In 1988, British climber Joe Simpson wrote a book called *Touching the Void* about his near-death experience in the Peruvian Andes. It received good reviews but didn't sell very well and was soon forgotten.



About ten years later, Jon Krakauer's book *Into Thin Air*, another story about a mountain-climbing disaster, became very popular. This renewed interest caused *Touching the Void* to start selling again.

Touching the Void then became so popular: It was sold out for a long time and was later adapted into a film. Here is its **trailer**!

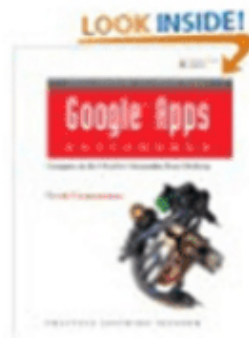
But how did *Touching the Void* become so popular?
Because Amazon recommended to that readers who liked *Into Thin Air* also *Touching the Void*.

Recommendation is everywhere!



Recommended for You

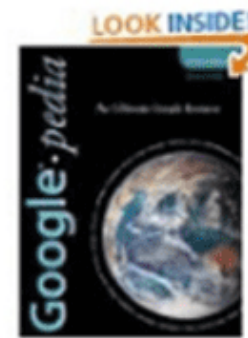
Amazon.com has new recommendations for you based on [items](#) you purchased or told us you own.



[Google Apps
Deciphered: Compute in
the Cloud to Streamline
Your Desktop](#)



[Google Apps
Administrator Guide: A
Private-Label Web
Workspace](#)



[Googlepedia: The
Ultimate Google
Resource \(3rd Edition\)](#)



Arizona Border Ranchers Torn in Support for Trump's Wall

172,275 views

683

249

SHARE

SAVE

...

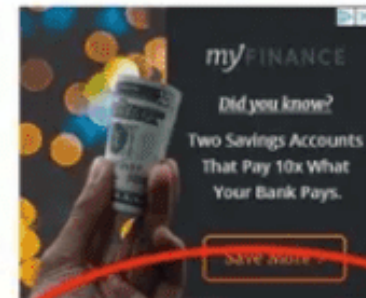


Wall Street Journal
Published on Mar 16, 2017

SUBSCRIBE 1.2M

Despite enthusiastic backing for President Donald Trump and pleas for a stronger border, Arizona ranchers are conflicted in their support for Trump's promise to build a wall along the border with Mexico. Photo/Video: Jake Nicol/The Wall Street Journal

SHOW MORE



Up next

AUTOPLAY



(Part II) A Day in the Life of Arizona Rancher: Fences, II
Center for Immigration Studies
43K views



CNN
You promised Mexico would...
2.7M views
New



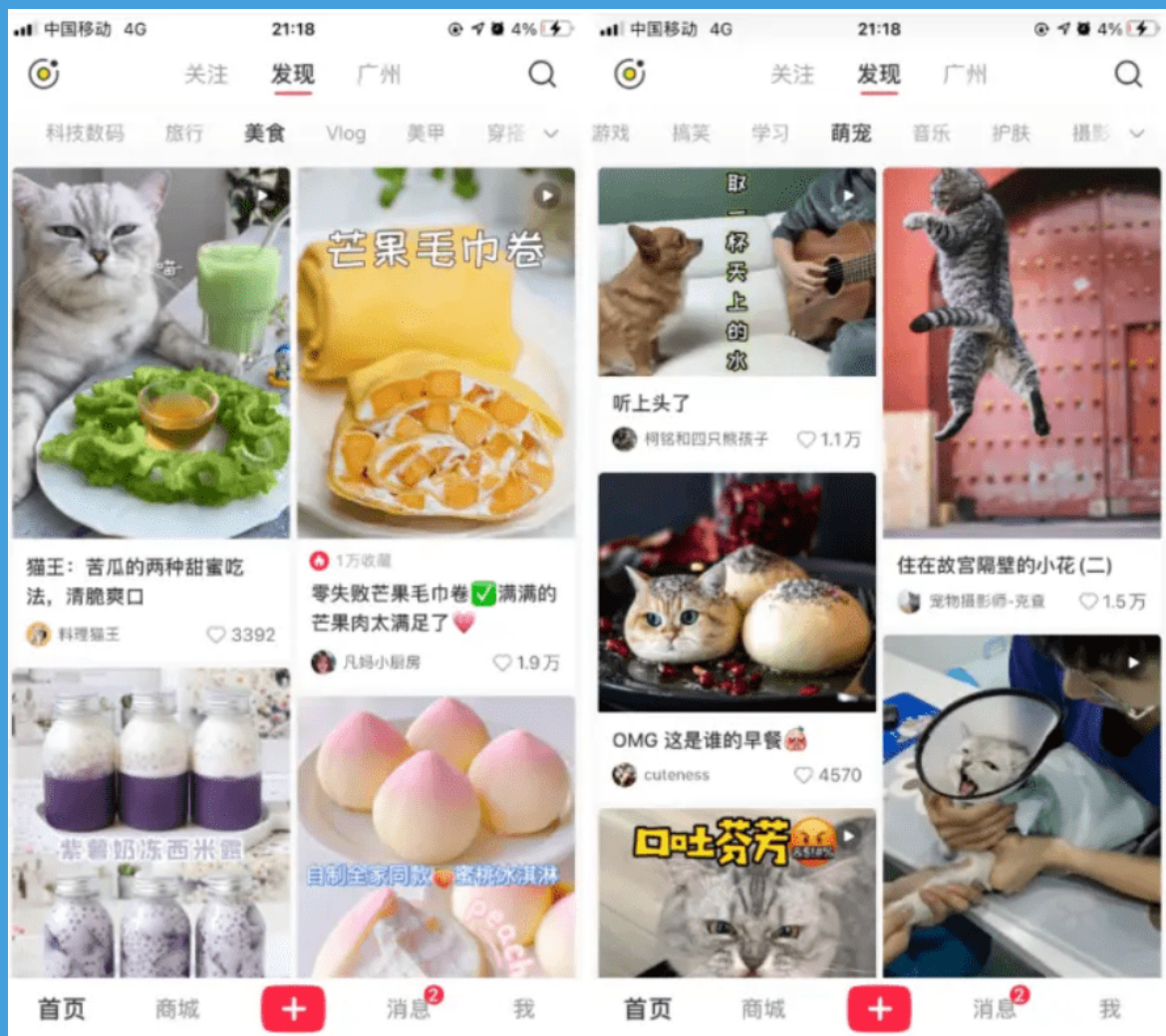
NBC News
People Are Fleeing President Trump's America To This T...
308K views



Sky News
Scrambling onto trucks for better life
2.4M views



BBC Planet Earth | BBC Studios
Polar Bear vs Walrus colony
Recommended for you



NETFLIX

Home TV Shows Movies Latest My List



KIDS



Continue Watching for SmartTV



Trending Now



Korean TV Shows



Top job picks for you

Data Scientist



kaam.work · Bangalore, IN



Skills match: Data Science, TensorFlow, Data Mining, and 5 more

Decision Point - Data Analyst (0-1 yrs) Gurgaon/Gurugram



Decision Point · Gurgaon, IN



Skills match: Analytics, Data Visualization, SQL, and 5 more



Associate Analyst II - Analyst - Analytics

Colgate-Palmolive · Mumbai, IN

KGS: MC : Data Science



KPMG India · Bangalore, IN



1 connection works here



Junior Data Scientist (Mumbai)

MightyHive · Mumbai, IN

Senior Data Analyst - Ad Tech



JioSaavn · Gurgaon, IN



Skills match: Data Visualization, Leadership, Data Mining, and 2 more



Easy Apply

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.

How to recommend?

A recommendation system must have three inputs:

- **Items** to be recommended: songs, movies, products, restaurants etc. (often many thousands)
- **Users** of the items: watchers, listeners, purchasers, shoppers etc. (often many millions)
- **Feedback** of users on items: 5-star ratings, upvotes/downvotes, clicking “next” or “skipping the ad”, purchases or clicks.

Collaboratives Filtering

Collaborative filtering is not something new. We have done it in many places in the past. Here are a few examples:

- Bestseller list for books
- Top 50 music list
- The “recent returns” shelf at libraries

The intuition behind: People's tastes are correlated.

Collaboratives Filtering

However, in the above examples, recommendations are not personalized, i.e., everybody receives the same recommendation. How to make recommendations personalized?

The intuition: If Alice and Bob both like X and Alice also likes Y , then Bob is more likely to like Y , especially when Alice and Bob know each other.

Suppose that you want to recommend a movie to Emma,
which movie will you recommend?

	1	2	3	4	5	6
Alice	2			4	5	
Bob	5		4			1
Carol			5		2	
Dennis		1		5		4
Emma			4			2
Flora	4	5		1		

User-Based Collaborative Filtering: The Neighbourhood Method

Step 1: Find all the movies rated by Emma before, we get
movies 3 and 6

	1	2	3	4	5	6
Alice	2			4	5	
Bob	5		4			1
Carol			5		2	
Dennis		1		5		4
Emma			4			2
Flora	4	5		1		

Step 2: Identify other users that have rated the same movie,
we get Bob, Carol, and Dennis

	1	2	3	4	5	6
Alice	2			4	5	
Bob	5		4			1
Carol			5		2	
Dennis		1		5		4
Emma			4			2
Flora	4	5		1		

Step 3: Compare the similarity between Emma and her “neighbors” to see who are close to Emma.

	1	2	3	4	5	6
Alice	2			4	5	
Bob	5		4			1
Carol			5		2	
Dennis		1		5		4
Emma			4			2
Flora	4	5		1		

Step 4: Select the top k most similar neighbors and use their average ratings to predict Emma's rating.

	1	2	3	4	5	6
Alice	2			4	5	
Bob	5		4			1
Carol			5		2	
Dennis		1		5		4
Emma			4			2
Flora	4	5		1		

Item-Based Collaborative Filtering

Item-based collaborative filtering

Suppose that we are predicting the who will like movie 5.

Step 1: Who have rated movie 5 before? We get Alice and Carol.

	1	2	3	4	5	6
Alice	2			4	5	
Bob	5		4			1
Carol			5		2	
Dennis		1		5		4
Emma			4			2
Flora	4	5		1		

Step 2: Identify other movies that have rated the same users, we get movies 1 and 3.

	1	2	3	4	5	6
Alice	2			4	5	
Bob	5		4			1
Carol			5		2	
Dennis		1		5		4
Emma			4			2
Flora	4	5		1		

Step 3: Compare the similarity between movie 5 and its “neighbors” to see which movie is close to movie 5.

	1	2	3	4	5	6
Alice	2			4	5	
Bob	5		4			1
Carol			5		2	
Dennis		1		5		4
Emma			4			2
Flora	4	5		1		

Step 4: Select the top k most similar neighbors and use their average ratings to predict movie 5's rating.

	1	2	3	4	5	6
Alice	2			4	5	
Bob	5		4			1
Carol			5		2	
Dennis		1		5		4
Emma			4			2
Flora	4	5		1		

Model-based Collaborative Filtering

What did Netflix do to make recommendations?

In general, how much do you like watching movies from the following genres?						
	Really dislike	Dislike	Neither like nor dislike	Like	Really like	Not sure of genre definition
Action	☐	☒	☐	☐	☐	☐
Adventure	☐	☐	☒	☐	☐	☐
Animation	☐	☐	☐	☐	☒	☐
Comedy	☐	☐	☐	☒	☐	☐
Crime/Gangster	☐	☒	☐	☐	☐	☐
Documentary	☐	☐	☐	☐	☐	☐
Drama	☐	☐	☐	☐	☐	☐
Fantasy	☐	☐	☐	☐	☐	☐
Film-Noir	☐	☐	☐	☐	☐	☐
Foreign	☐	☐	☐	☐	☐	☐
Horror	☐	☐	☐	☐	☐	☐

Issues with this approach:

- It takes users much effort to complete the questionnaire, and users do not always bother to answer the questions carefully and seriously.
- Sometimes users themselves cannot clearly state their preferences.
- Maybe there are some hidden factors that are not captured by the questionnaire, e.g., the duration of the movie.

The Matrix Factorization Method

Matrix Factorization

Here, we assume that each movie has a number of “latent” or “hidden” factors that affect user preferences. Examples of the factors include the length of the movie, the amount of actions in the movie, the seriousness of the movie, the orientation of the movie for children etc.

The factors are “latent” or “hidden,” implying that we do not know which these factors are, and we do not need to know either.

Matrix Factorization

Each user also has his or her own preference for each factor. For instance, some users prefer long movies over short movies, and some users prefer to have more actions in their movie. If we know the preferences of a user and a movie's attribute values, we can match the movie with the user to see whether the user will like the movie.

Matrix Factorization

For instance, suppose that a movie is long and contains a lot of actions. We also know that

- Alice likes short movies and hates action movies,
- Bob prefers long movies and enjoys action movies.

Then, we can predict that Alice will hate the movie and Bob will like the movie.

Mathematically, our model is as follows:

$$\begin{aligned} \text{Your rating} = & \text{Your preference for length} \times \text{Movie's length} \\ & + \text{Your preference for action} \times \text{Movie's amount of action} \end{aligned}$$

Suppose that the movie's length is 1 and amount of action is 2. Alice's preference for length is 0.5, for action is 0; Bob's preference for length is 1, for action is 1.5. We can predict:

- Alice's rating: $0.5 \times 1 + 0 \times 2 = 0.5$.
- Bob's rating: $1 \times 1 + 1.5 \times 2 = 4$.

Let's generalize the above discussion. Suppose that Alice, Bob, Carol's preferences for length and action are as follows:

	Length	Action
Alice	0.5	0
Bob	1	1.5
Carol	1.5	0.5

There are two movies, whose length and action values are

	Length	Action
1	1	2
2	0	3

We can multiply the two matrix to get user-movie ratings:

$$\begin{bmatrix} 0.5 & 0 \\ 1 & 1.5 \\ 1.5 & 0.5 \end{bmatrix} \times \begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 0.5 & 0 \\ 4 & 4.5 \\ 2.5 & 1.5 \end{bmatrix}$$

In other words, our prediction is as follows.

	Movie 1	Movie 2
Alice	0.5	0.0
Bob	4.0	4.5
Carol	2.5	1.5

Overall, if we know the user matrix and the movie matrix, we can multiply the two to get the user-movie rating matrix. The issue is: **We do not yet know the user matrix and the movie matrix.**

But how to get the user matrix and the movie matrix?

The short answer is, we can guess. We guess different user matrices and movie matrices, and see if the predicted rating matrix is close to the actual rating given by users. When the two are close enough, we can use the two matrices to construct the new rating matrix. This is known as matrix factorization.

We can do better than guessing. There are some advanced statistical methods for estimating the user matrix and movie matrix, but this is beyond the scope of our class. If you are interested, you can search for “stochastic gradient descent.”

The matrix factorization algorithm

```
1 matrix_factorization <- function(R, P, Q, K, steps=5000, alpha=0.0002,  
  beta=0.02) {  
2   Q <- t(Q)  
3   for (step in 1:steps) {  
4     for (i in 1:nrow(R)) {  
5       for (j in 1:ncol(R)) {  
6         if (R[i, j] > 0) {  
7           eij <- R[i, j] - sum(P[i,] * Q[,j])  
8           for (k in 1:K) {  
9             P[i, k] <- P[i, k]+alpha*(2*eij*Q[k, j] - beta*P[i, k])  
10            Q[k, j] <- Q[k, j]+alpha*(2*eij*P[i, k] - beta*Q[k, j])  
11          }  
12        eR <- P %*% Q  
13        e <- 0  
14        for (i in 1:nrow(R)) {  
15          for (j in 1:ncol(R)) {  
16            if (R[i, j] > 0) {  
17              e <- e + (R[i, j] - sum(P[i,] * Q[,j]))^2  
18              for (k in 1:K) {  
19                e <- e + (beta/2) * (P[i, k]^2 + Q[k, j]^2)  
20              }  
21            }  
22          if (e < 0.001){break}  
23        return(list(P = P, Q = t(Q)))  
24      }
```

```

1  set.seed(123)
2  R <- matrix(c(5, 3, 0, 1,
3               4, 0, 0, 1,
4               1, 1, 0, 5,
5               1, 0, 0, 4,
6               0, 1, 5, 4,
7               2, 1, 3, 0), nrow = 6, ncol = 4, byrow = TRUE)
8  # N: num of User
9  N <- nrow(R)
10 # M: num of Movie
11 M <- ncol(R)
12 # Num of Features
13 K <- 2
14 P <- matrix(runif(N * K), nrow = N, ncol = K)
15 Q <- matrix(runif(M * K), nrow = M, ncol = K)
16 result <- matrix_factorization(R, P, Q, K)
17 nP <- result$P
18 nQ <- result$Q
19 nR <- nP %*% t(nQ)
20 print(nP)
21 print(nQ)
22 print(nR)

```

	Movie 1	Movie 2	Movie 3	Movie 4
Alice	4	4		1
Bob		2	2	3
Carol	1	5	3	
Dennis	3		4	1
Emma	5	2	1	4
Flora	3	1		5

$$\begin{bmatrix} 4 & 4 & ? & 1 \\ ? & 2 & 2 & 3 \\ 1 & 5 & 3 & ? \\ 3 & ? & 4 & 1 \\ 5 & 2 & 1 & 4 \\ 3 & 1 & ? & 5 \end{bmatrix} \approx \begin{bmatrix} 1.71 & 0.74 \\ 0.84 & 1.35 \\ 1.92 & -0.69 \\ 2.05 & 0.46 \\ 0.70 & 1.95 \\ 0.13 & 1.93 \end{bmatrix} \times \begin{bmatrix} 1.24 & 2.44 & 1.77 & -0.20 \\ 1.83 & 0.14 & 0.10 & 2.32 \end{bmatrix}$$

$$= \begin{bmatrix} 3.47 & 4.28 & 3.09 & 1.37 \\ 3.52 & 2.26 & 1.63 & 2.97 \\ 1.14 & 4.63 & 3.34 & -1.99 \\ 3.41 & 5.09 & 3.67 & 0.66 \\ 4.48 & 1.99 & 1.43 & 4.40 \\ 3.69 & 0.60 & 0.43 & 4.46 \end{bmatrix}$$

<https://www.youtube.com/embed/n3RKsY2H-NE?enablejsapi=1>

The Cold Start Problem

The cold start problem

The collaborative filtering algorithm works very well in general, yet it suffers from the issue of cold start. Recall that in the recommendation algorithm, we need to know the users' past interaction with the items to make recommendations. **However, if a user or an item is completely new without any historical interactions, what would you do to make recommendations?**

One solution is to use surveys. In the case of movie recommendations, we can ask new users about their preferences for different movie characteristics. Then, we can match the movies with user preferences.

In general, how much do you like watching movies from the following genres?						
	Really dislike	Dislike	Neither like nor dislike	Like	Really like	Not sure of genre definition
Action	☐	☒	☐	☐	☐	☐
Adventure	☐	☐	☒	☐	☐	☐
Animation	☐	☐	☐	☐	☒	☐
Comedy	☐	☐	☐	☒	☐	☐
Crime/Gangster	☐	☒	☐	☐	☐	☐
Documentary	☐	☐	☐	☐	☐	☐
Drama	☐	☐	☐	☐	☐	☐
Fantasy	☐	☐	☐	☐	☐	☐
Film-Noir	☐	☐	☐	☐	☐	☐
Foreign	☐	☐	☐	☐	☐	☐
Horror	☐	☐	☐	☐	☐	☐

This strategy does not work well. Users just randomly complete the survey if you force them to do so. Their choices do not reveal much information to you.

In general, how much do you like watching movies from the following genres?						
	Really dislike	Dislike	Neither like nor dislike	Like	Really like	Not sure of genre definition
Action	☐	☒	☐	☐	☐	☐
Adventure	☐	☐	☒	☐	☐	☐
Animation	☐	☐	☐	☐	☒	☐
Comedy	☐	☐	☐	☒	☐	☐
Crime/Gangster	☐	☒	☐	☐	☐	☐
Documentary	☐	☐	☐	☐	☐	☐
Drama	☐	☐	☐	☐	☐	☐
Fantasy	☐	☐	☐	☐	☐	☐
Film-Noir	☐	☐	☐	☐	☐	☐
Foreign	☐	☐	☐	☐	☐	☐
Horror	☐	☐	☐	☐	☐	☐

Another strategy is to cluster users or movies based on observed characteristics. For users, we can collect information such as country, age, gender, nationality... For movies, we can also construct a profile for them (e.g., language, genre, duration, director). Then we apply a clustering algorithm to divide the users or movies into a few clusters.

Even though we may not know a specific user, we know other users in the same cluster. We can make recommendations based on these users' preferences.