

Recommender Systems

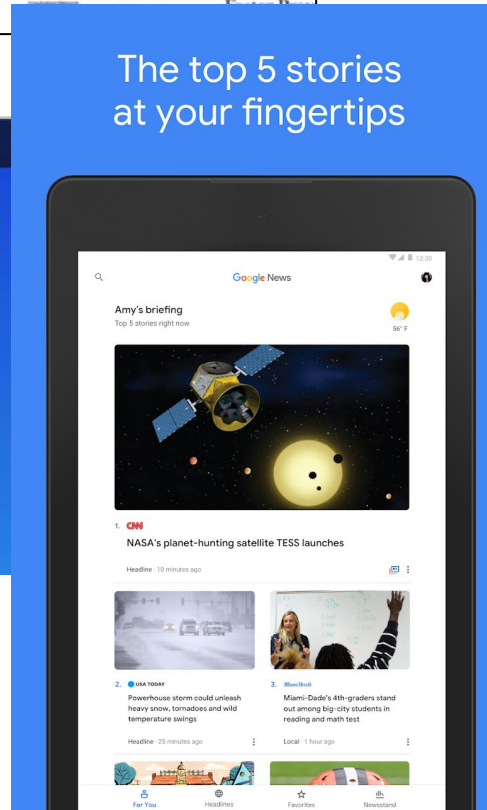
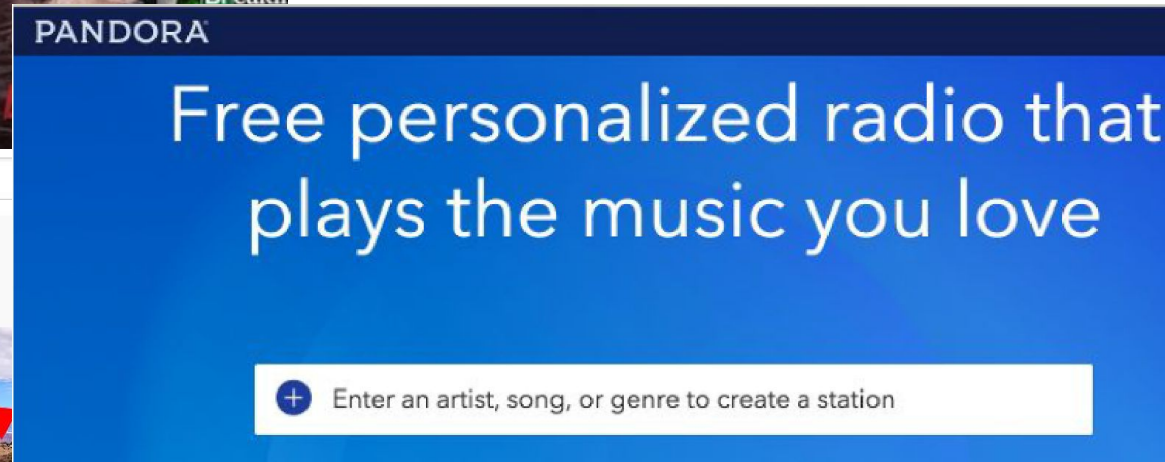
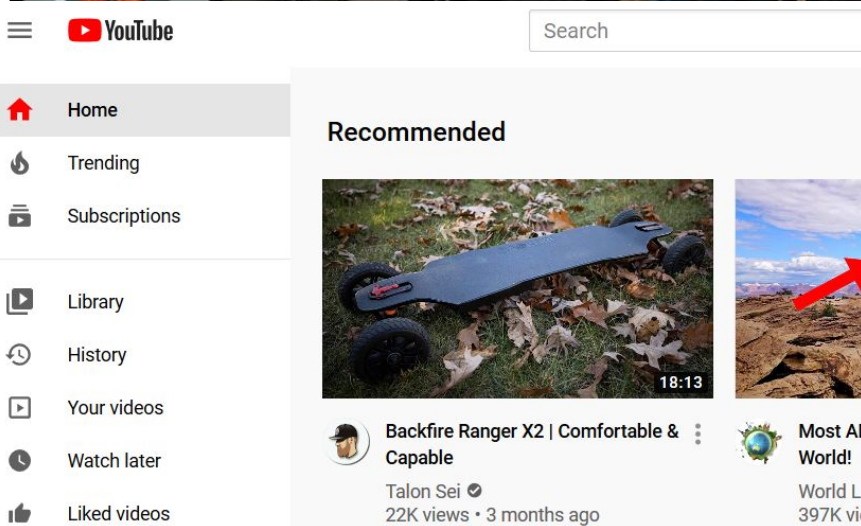
Recommender systems

- What is a Recommender System
- Two popular approaches

Content-based, Collaborative filtering

- Similarity measures
- Things to consider

What is a recommender system?



Recommendation approaches

Popularity

- Recommend the most popular or trending item(s) to everyone.

Content-based

- Items are similar if their attributes are similar
- Often hand-engineered (domain-specific) attributes

Collaborative filtering

- Recommends items chosen by similar users
- domain-free

Content-Based Filtering

- Creates profile of each user and items
- Need to collect user demographics or questionnaire
- Need domain-specific info about the items
- Features are hand-engineered by the domain experts

Content-Based Filtering Example

pandora® | Music Genome Project

About

Contact

Press

Management

Board

About The Music Genome Project®

The Music Genome Project powers Pandora. It's the most comprehensive analysis of music ever undertaken.

For over a decade, we've been gathering musical knowledge to bring you the best, most personalized listening experience out there.

We believe each individual has a unique relationship with music – no one has tastes that are exactly the same. So delivering a great experience to every listener requires a broad and deep understanding of music.

Our team of trained musicologists has been listening to music across all genres and decades, including emerging artists and new releases, studying and collecting musical details on every track– 450 musical attributes altogether.

The result of all our work is a personalized listening experience filled with both old favorites and new discoveries.

Item features examples

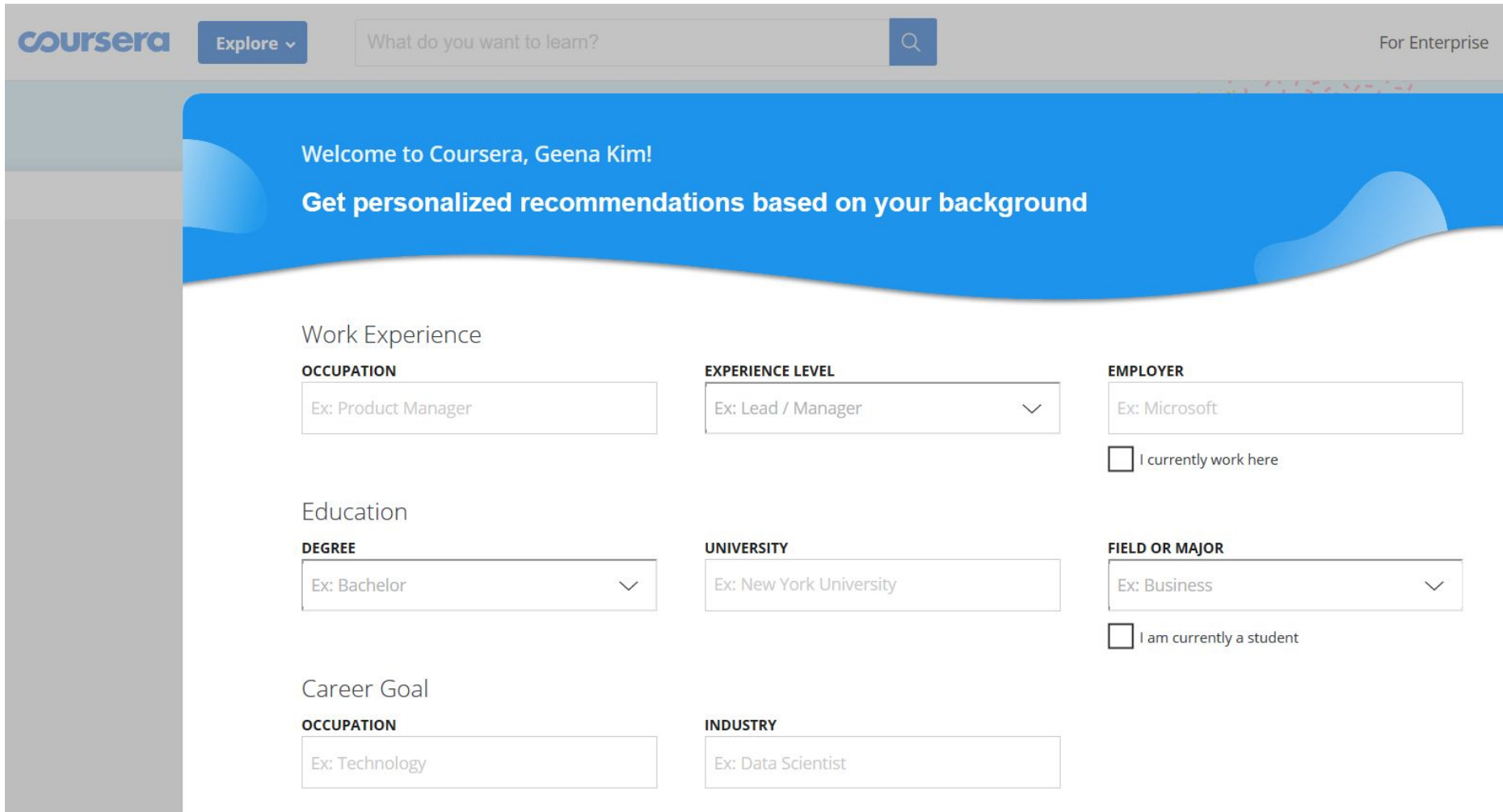
Laptop features: Hard disk size, CPU speed, RAM size, Weight, monitor size, GPU specs

Shirt features: Size, color, fabric, style, collar, finish, sleeve

Movie features: actors, director, year, genre, series, average rating

User profiling

Oftentimes Recommender systems collect user data



The screenshot shows the Coursera user interface for user Geena Kim. The header includes the Coursera logo, an 'Explore' dropdown, a search bar with the placeholder 'What do you want to learn?', and a 'For Enterprise' link. Below the header is a blue banner with the text 'Welcome to Coursera, Geena Kim!' and 'Get personalized recommendations based on your background'. The main content area contains several form fields for user profiling:

- Work Experience**
 - OCCUPATION**: Text input field with placeholder 'Ex: Product Manager'.
 - EXPERIENCE LEVEL**: Dropdown menu with placeholder 'Ex: Lead / Manager'.
 - EMPLOYER**: Text input field with placeholder 'Ex: Microsoft'.
 - ☐ I currently work here
- Education**
 - DEGREE**: Dropdown menu with placeholder 'Ex: Bachelor'.
 - UNIVERSITY**: Text input field with placeholder 'Ex: New York University'.
 - FIELD OR MAJOR**: Dropdown menu with placeholder 'Ex: Business'.
 - ☐ I am currently a student
- Career Goal**
 - OCCUPATION**: Text input field with placeholder 'Ex: Technology'.
 - INDUSTRY**: Text input field with placeholder 'Ex: Data Scientist'.

Collaborative Filtering

- No need of hand-engineered features
- Domain-free
- Learns from also other users' interaction with items
- May suffer from **cold-start** problem

Collaborative Filtering Approaches

Memory-based

- Customers who bought this item also bought

Using Similarity

- item-item similarity
- user-user similarity

Using Latent factor modeling

- Matrix Factorization

Other

- Supervised approaches, graphs

Utility Matrix

What does the data look like?

	Star Wars I	Star Wars II	Squid Game	Lord of ...	Harry Potter I
Amy			2		5
Bob		3	5		
Cathy	1			4	
Dave	4	5		5	

Utility Matrix

User explicitly rate
products

Explicit
ratings

What does the data look like?

	Item 1	Item 2	Item 3	Item 4	Item 5
A			1		1
B		1	1		
C	1			1	
D	1	1			1

User buy or not buy the product
implicit ratings

	Item 1	Item 2	Item 3	Item 4	Item 5
A			1		1
B		1	1		1
C	1			1	
D		1	1		1

User-user
similarity
Item-item
similarity

Similarity measures

Cosine similarity

$$\cos(a, b) = \frac{a \cdot b}{||a|| ||b||}$$

$$\text{sim}(a, b) = 0.5 + 0.5 \times \cos(a, b)$$

Similarity measures

Jaccard similarity

set of users who rated item a



$$\text{Jaccard}(a, b) = \frac{|S_a \cap S_b|}{|S_a \cup S_b|}$$

Similarity measures

Distance-based

- Manhattan distance
- Euclidean distance
- Minkowski distance

$$\text{sim}(a, b) = \frac{1}{1 + \text{dist}(a, b)}$$

Similarity measures

Pearson Correlation

$$\text{Pearson}(a, b) = \frac{\text{cov}(a, b)}{\text{std}(a)\text{std}(b)} = \frac{\sum_i (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_i (a_i - \bar{a})^2} \sqrt{\sum_i (b_i - \bar{b})^2}}$$

$$\text{sim}(a, b) = 0.5 + 0.5 \times \text{Pearson}(a, b)$$

Examples- Cosine

	M1	M2	M3	M4
A	5.0	NaN	1.0	4.0
B	2.0	3.0	5.0	NaN
C	4.0	4.0	NaN	4.0

Examples- Cosine

	M1	M2	M3	M4	M5
A	5.0	NaN	1.0	4.0	3.0
B	2.0	3.0	5.0	NaN	3.0
C	5.0	4.0	5.0	4.0	NaN
D	3.0	2.0	NaN	2.0	1.0

A	3.25
B	3.25
C	4.50
D	2.00

Normalize by 3

AB	-0.89
AC	0.105
AD	-0.13
BC	0.282
BD	0.0
CD	-0.25

Normalize by user avg

AB	-0.94
AC	-0.21
AD	0.478
BC	0.172
BD	-0.32
CD	0.353

Examples- Jaccard

	M1	M2	M3	M4	M5	M6
A	5.0	NaN	1.0	4.0	NaN	NaN
B	2.0	3.0	5.0	NaN	1.0	NaN
C	4.0	4.0	NaN	4.0	NaN	3.0

	M1	M2	M3	M4	M5	M6
A	1	0	0	1	0	0
B	0	1	1	0	0	0
C	1	1	0	1	0	1

Things to consider in Recommender Systems

- Time complexity of operations
- The data is often large scale
- The data is sparse

Recommender system in large scale

Where to place components?

- Example: Matrix Factorization
- Offline:
 - Collect sample of play data
 - Run batch learning algorithm to produce factorization
 - Publish item factors
- Nearline:
 - Solve user factors
 - Compute user-item products
 - Combine
- Online:
 - Presentation-context filtering
 - Serve recommendations

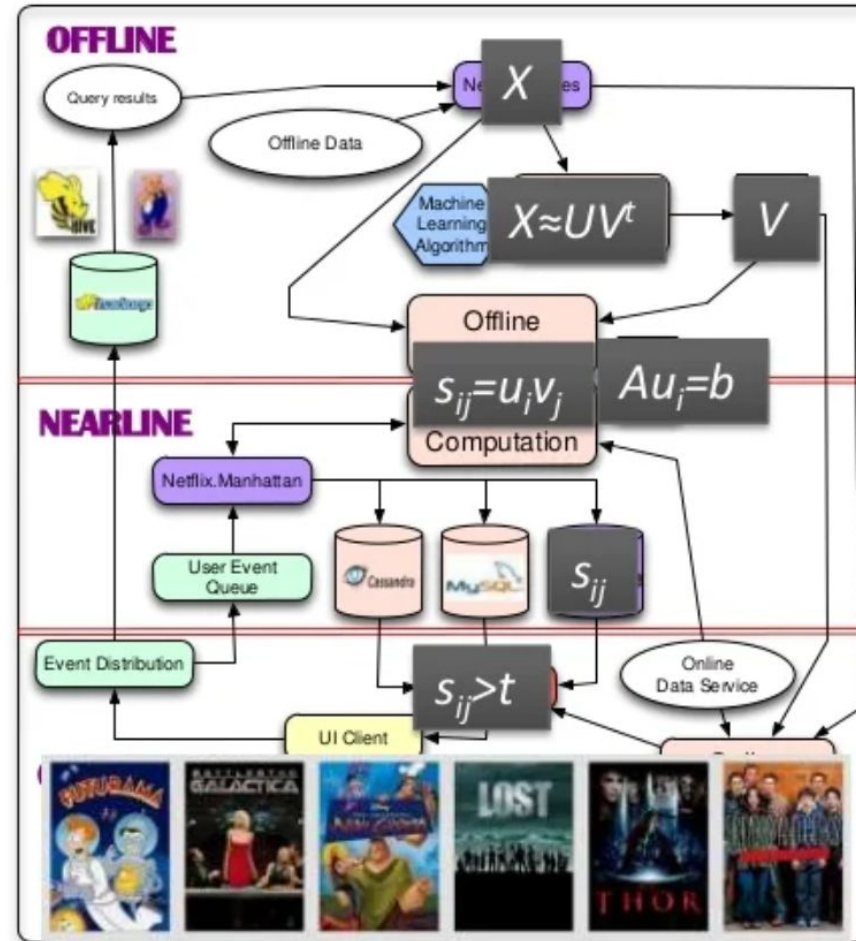


Image credit:
Justin Basilico
RecSys 2013