

Recommender Systems

Data Science Dojo

Overview

- What are Recommender Systems?
- How do they work?
 - Collaborative Recommendation
 - Content-Based Recommendation
- How do we evaluate them?
- Example using Azure ML

Overview

- **What are Recommender Systems?**
- How do they work?
 - Collaborative Recommendation
 - Content-Based Recommendation
- How do we evaluate them?
- Example using Azure ML

Recommender Systems

- What are Recommender Systems?
 - Automated systems to filter and recommend products based on users' interest and taste.
 - Designed to solve the information overload problem

Why recommendation systems?

For customer

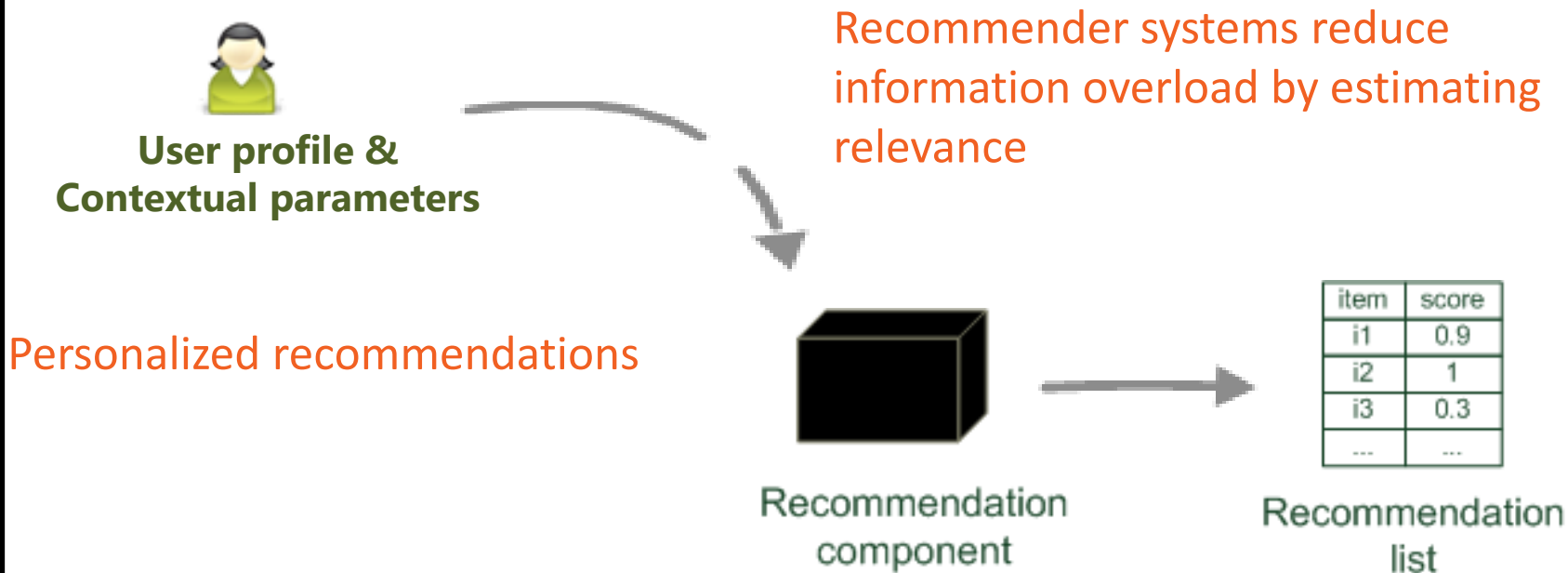
- Narrow down the set of choices
- Discover new, interesting things
- Save time

Why recommendation systems?

For businesses

- Increase the number of items sold
- Sell more diverse items
- Better understand what the user wants
- Increase user satisfaction

Recommender Systems

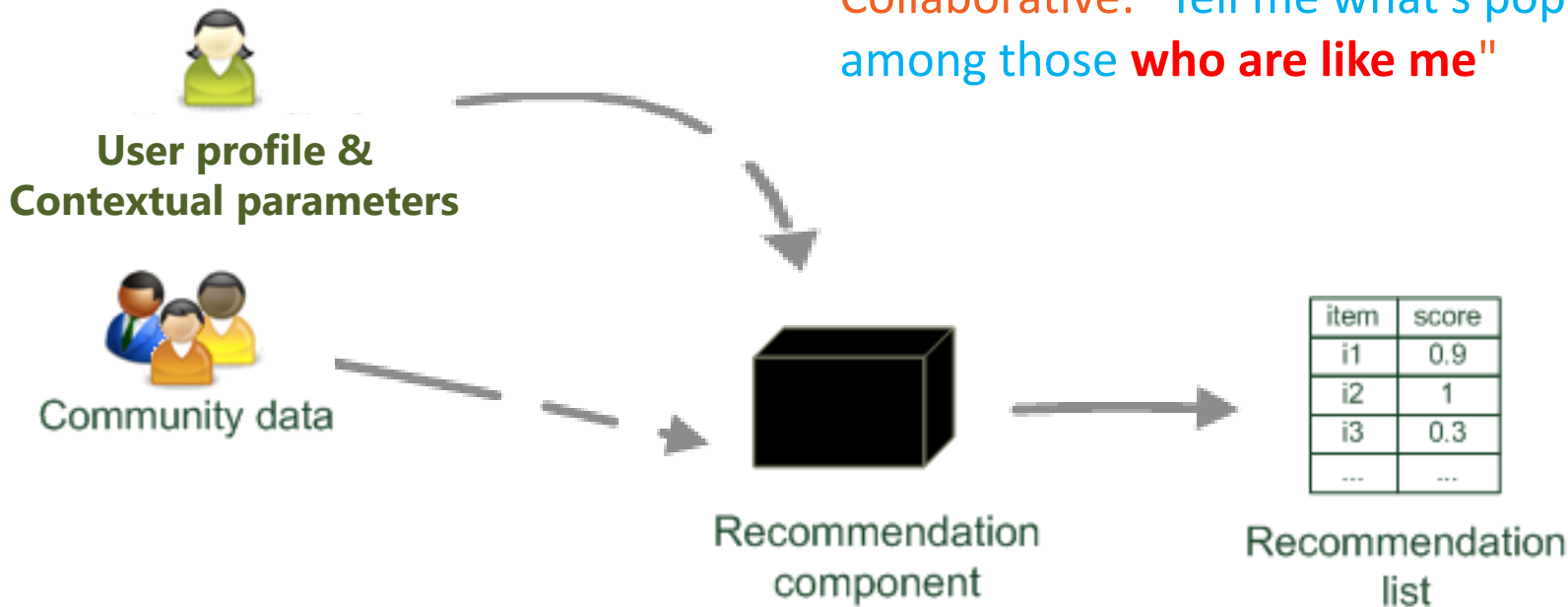


Overview

- What are Recommender Systems?
- How do they work?
 - **Collaborative Recommendation**
 - Content-Based Recommendation
- How do we evaluate them?
- Example using Azure ML

Collaborative Filtering (CF)

Collaborative: "Tell me what's popular
among those **who are like me**"



Example: Netflix

Top Picks for Cassandra



Frasier

★★★★★ 200 TV-PG 11 Seasons

Frasier Crane is a snooty but lovable Seattle psychiatrist who dispenses advice on his call-in radio show while ignoring it in his own relationships.

Starring: Kelsey Grammer, Jane Leeves, David Hyde Pierce

Genres: TV Shows, TV Comedies, Sitcoms

This show is: Witty

Winner of more than 37 Emmys, including three for Best Comedy and four Best Actor awards for Kelsey Grammer.

NETFLIX

Browse

Personalize

KIDS


DVDs

Top Picks for jodi




Example: Retail & Social Media

People You May Know




Cheryl Jamison
The Old School Of Hard Knocks
and 2 other mutual friends

Add Friend Remove



Susan D. Curtis
The new guy at Dailcon FSD
and 23 other mutual friends

Add Friend Remove



Dave DeWitt
Works at The Home Depot

Add Friend Remove

Ads You May Be Interested In



Big Data in 2015

Learn about 5 emerging big data trends in 2015 that help sustain high ROI.



Attn: Successful Women

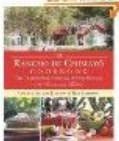
You're Invited to Join the National Association of Professional Women.

Pa

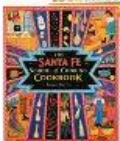
Editor

islatiional Science

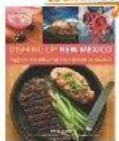
Customers Who Bought This Item Also Bought



Rancho de Chimayo Cookbook: The...
Cheryl Jamison
★★★★☆ 10
Paperback
\$19.05 Prime



The Santa Fe School of Cooking Cookbook
Susan D. Curtis
★★★★☆ 16
Paperback
\$21.14 Prime



Dishing Up® New Mexico: 145 Recipes from the...
Dave DeWitt
★★★★☆ 7
Paperback
\$15.45 Prime

Collaborative Filtering

- Most popular recommendation algorithm
 - Used by large, commercial e-commerce sites
 - Well-understood, variety of algorithms
 - Applicable to many domains (books, movies, songs,...)
- Approach: borrow the “wisdom of the crowd” to recommend items

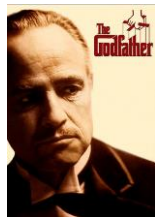
Collaborative Filtering

- Assumption:
 - Users give ratings to items
 - Users who have similar tastes in the past will have similar tastes in the future.
- User-based collaborative
- Item-based collaborative

Collaborative Filtering

- Assumption:
 - Users give ratings to items
 - Users who have similar tastes in the past will have similar tastes in the future.
- **User-based collaborative**
- Item-based collaborative

Movie Rating Example



Alice	5	3	4	4	?
Bob	3	1	2	3	3
Chris	4	3	4	3	5
Donna	3	3	1	5	4
Evi	1	5	5	2	1

User-Based Collaborative Filtering

Goal: Given Alice is an “active” user, we want to predict the rating of movie p Alice hasn’t seen.

- Find a set of users who liked the same items as Alice in the past and also had rated movie p
- Predict Alice’s rating on movie p
- Repeat for all movies Alice has not seen and recommend the best rated.

User-Based Collaborative Filtering

- How many neighbors should we include?
 - Choose a number – depends on size of data
- How do we define similarity?
- How to do we generate predictions from the neighbors' ratings?

Similarity Measurement

■ Pearson correlation

j, k : users

$r_{j,p}$: rating of user j for item p

\bar{r}_j and \bar{r}_k are the average ratings of user j and user k over all items

P : set of items, rated both by j and k

Possible similarity values between -1 and 1

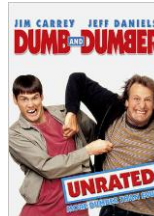
j : Alice


k : Bob

P : set of items, rated by Alice and Bob

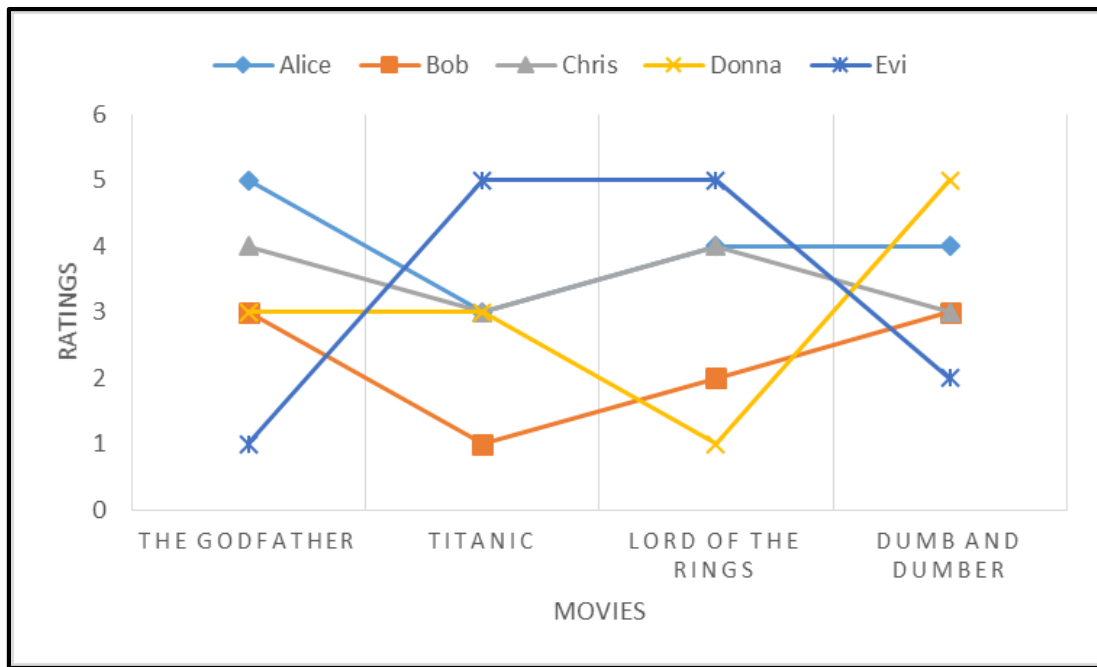
$$\text{sim}(j, k) = \frac{\sum_{p \in P} (r_{j,p} - \bar{r}_j)(r_{k,p} - \bar{r}_k)}{\sqrt{\sum_{p \in P} (r_{j,p} - \bar{r}_j)^2} \sqrt{\sum_{p \in P} (r_{k,p} - \bar{r}_k)^2}}$$

Pearson Correlation

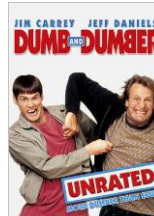


Alice	5	3	4	4	?	 sim=?
Bob	3	1	2	3	3	
Chris	4	3	4	3	5	
Donna	3	3	1	5	4	
Evi	1	5	5	2	1	

Pearson Correlation



Pearson Correlation



Alice	5	3	4	4	?	
Bob	3	1	2	3	3	sim=0.85
Chris	4	3	4	3	5	sim=0.90
Donna	3	3	1	5	4	sim=0.70
Evi	1	5	5	2	1	sim=0.79

Making Predictions

- Use "Aggregation Function"

- Choose N neighbors
- Simple

- $r_{j,p} = \frac{1}{N} \sum_{k \in U} r_{k,p}$

- Weighted & Centered

- $r_{j,p} = \bar{r}_j + \alpha \sum_{k \in U} \text{simil}(j, k) (r_{k,p} - \bar{r}_k)$

Making recommendations

- Prediction is typically not the ultimate goal
 - Rank items based on their predicted ratings
 - This might lead to the inclusion of (only) niche items
 - Optimize according to a given rank evaluation metric

Collaborative Filtering

- Assumption:
 - Users give ratings to items
 - Users who has similar tastes in the past, have similar tastes in the future.
- User-based collaborative
- **Item-based collaborative**

Item-based collaborative filtering

- Alternate idea:
 - Use the similarity between items (and not users) to make predictions
 - Look for movies that are similar to movie p
 - Take **Alice**'s ratings for these items to predict the rating for movie p

Similarity Measurement

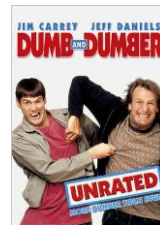
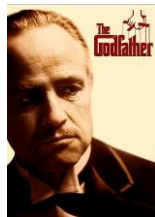
- Cosine similarity

$$\text{sim}(\vec{a}, \vec{b}) = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| * |\vec{b}|} = \frac{\sum_{u \in U} r_{u,a} * r_{u,b}}{\sqrt{\sum_{u \in U} r_{u,a}^2} \sqrt{\sum_{u \in U} r_{u,b}^2}}$$

- Adjusted cosine similarity

$$\text{sim}(\vec{a}, \vec{b}) = \frac{\sum_{u \in U} (r_{u,a} - \bar{r}_u)(r_{u,b} - \bar{r}_u)}{\sqrt{\sum_{u \in U} (r_{u,a} - \bar{r}_u)^2} \sqrt{\sum_{u \in U} (r_{u,b} - \bar{r}_u)^2}}$$

Movie Rating Example



Alice	5	3	4	4	?
Bob	3	1	2	3	3
Chris	4	3	4	3	5
Donna	3	3	1	5	4
Evi	1	5	5	2	1

sim=0.99

sim=0.74

sim=0.72

sim=0.93

Making Predictions

- Sum over items rather than users

- Simple

- $r_{j,p} = \frac{1}{N} \sum_{q \in P} r_{j,q}$

- Weighted & Centered

- $r_{j,p} = \overline{r_p} + \alpha \sum_{q \in P} \text{simil}(p, q)(r_{j,q} - \overline{r_q})$

Collaborative Filtering Pros

- **Wide applicability**
 - Usable in wildly different domains
- **Well-understood**
 - Most well studied type of recommender
- **Simple**
 - No knowledge engineering required
- **Serendipity**
 - Odd recommendations that are very good

Collaborative Filtering Cons

- **Data sparsity & Cold Start**

- New users need to indicate preferences for sufficient number of items before recommendations are good
- Need initial customer/rating database

- **Scalability**

- Millions of customers (M) and millions of items (N)

- **Grey Sheep and Black Sheep**

- Grey sheep are users with inconsistent recommendations.
- Black sheep are the users with idiosyncratic preferences.

Collaborative Filtering Cons

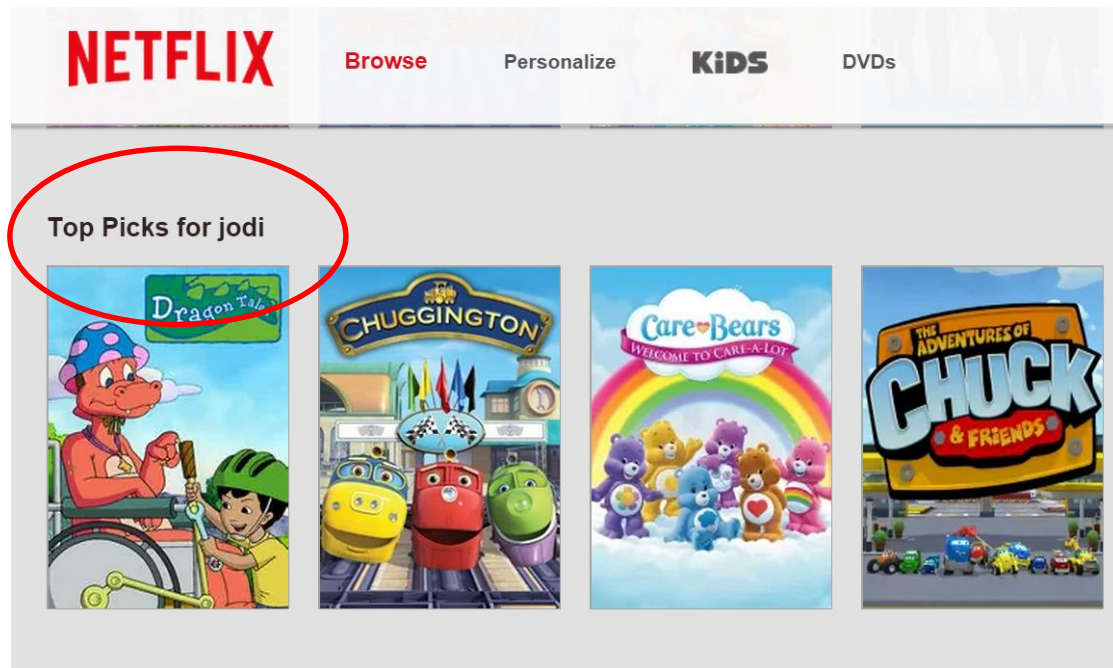
- **Shilling**

- Intentional manipulation of ratings of your own products and competitors products

- **Diversity and Long Tail**

- Rich tend to get richer

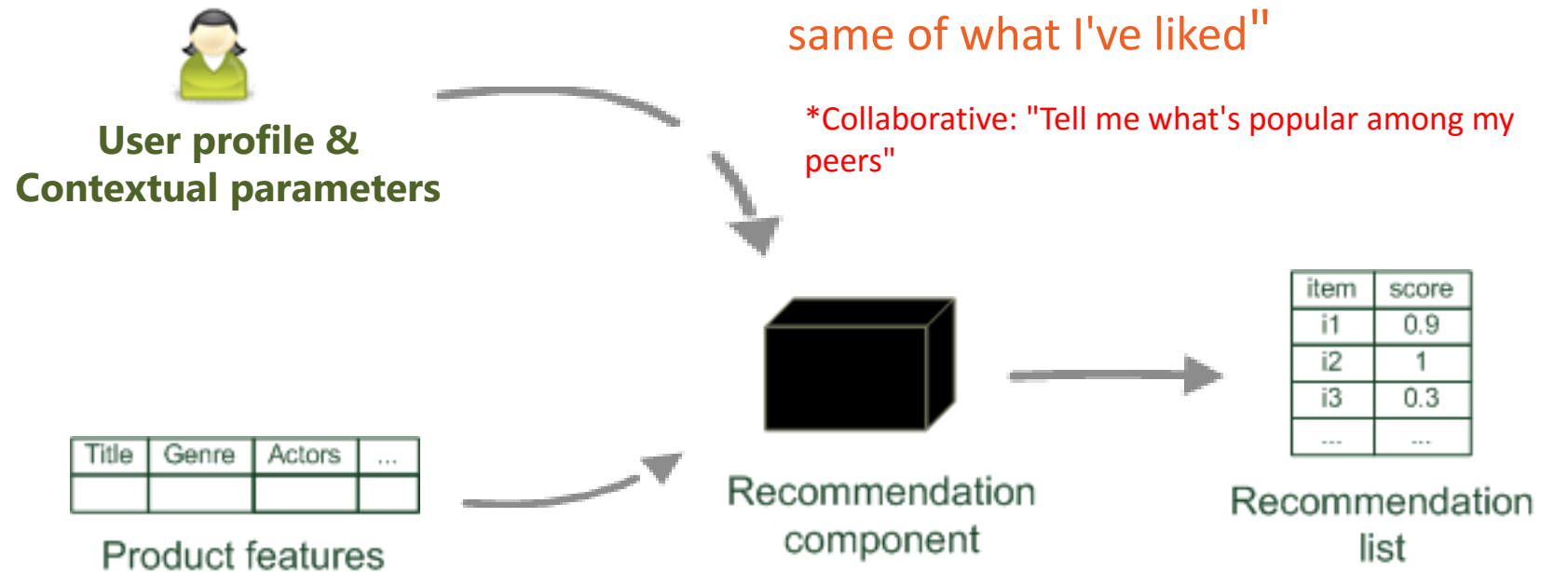
Back to Netflix



Overview

- What are Recommender Systems?
- How do they work?
 - Collaborative Recommendation
 - **Content-Based Recommendation**
- How do we evaluate them?
- Example using Azure ML

Content-based recommendation



Examples

Mind-bending Movies



Quirky Comedies



Cerebral TV Shows



NETFLIX Browse **KIDS**

Taste Preferences

How often do you watch Never Sometimes Often

Moods	Never	Sometimes	Often
Absurd	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Adrenaline Rush	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Bawdy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Campy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Cerebral	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Chilling	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

PANDORA

+ Type in artist, genre or composer

Enter an artist, genre or composer. We'll create a radio station featuring that music and more like it.

Based on your stations you might want to try:
Stone Sour, Lacuna Coil, Within Temptation, White Lion

Try one of these genre stations:
Viking Metal, 80s Pop

[Browse Genres](#)

Examples

Related to Items You've Viewed [See more](#)



Data Science



[Web](#) [News](#) [Images](#) [Books](#) [Videos](#) [More ▾](#) [Search tools](#)

About 288,000,000 results (0.42 seconds)

[Data science - Wikipedia, the free encyclopedia](#)

https://en.wikipedia.org/wiki/Data_science ▾ Wikipedia ▾

Data Science is an interdisciplinary field about processes and systems to extract knowledge or insights from large volumes of data in various forms, either ...

[Overview](#) - [History](#) - [Domain specific interests](#) - [Criticism](#)

[Data Science | Coursera](#)

<https://www.coursera.org/specializations/jhudatascience> ▾ Coursera ▾

Become an expert with Data Science Specialization offered by Johns Hopkins University. Take free online classes from 120+ top universities and educational ...

[Certificate in Data Science - UW Professional & Continuing ...](#)

www.pce.uw.edu/certificates/data-science.html ▾

University of Washington offers a certificate program in data science, with flexible evening and online classes to fit your schedule.

Jan 14, 2016 [Online](#)

Mar 28, 2016 [Bellevue](#)

Content-based recommendation

Recommend items that are “similar” to the user preferences

What do we need?

- Item Profiles: list of content-based keywords
- User profiles: preferences of the user.
 - User specified or based on past behavior

Item Profile Strategies

▪ Expert Labeling

- Assign keywords based on content
- Good for songs, movies, etc
- May be provided by creators/distributors
- Crowd sourcing?

Item Profile Strategies

▪ **Automated Indexing**

- Used for text documents (web pages, books, tweets)
- Based on word content of document set
- No expert knowledge involved
- Can be keyword or full dictionary based

Content-based recommendation

■ Prediction: Simple approach

- Compute the similarity of an item and user profile based on keyword overlap

- $$\text{sim}(b_i, b_j) = \frac{2 * |\text{keywords}(b_i) \cap \text{keywords}(b_j)|}{|\text{keywords}(b_i)| + |\text{keywords}(b_j)|}$$

Simple approach: drawbacks

- Not every word has similar importance
- Longer documents have a higher chance to have an overlap with the user profile
- **Solution:** TF-IDF

TF-IDF Search Exercise

- **D1** = "If it walks like a duck and quacks like a duck, it must be a duck."
- **D2** = "Beijing Duck is mostly prized for the thin, crispy duck skin with authentic versions of the dish serving mostly the skin."
- **D3** = "Bugs' ascension to stardom also prompted the Warner animators to recast Daffy Duck as the rabbit's rival, intensely jealous and determined to steal back the spotlight while Bugs remained indifferent to the duck's jealousy, or used it to his advantage. This turned out to be the recipe for the success of the duo."
- **D4** = "6:25 PM 1/7/2007 blog entry: I found this great recipe for Rabbit Braised in Wine on cookingforengineers.com."
- **D5** = "Last week Li has shown you how to make the Sechuan duck. Today we'll be making Chinese dumplings (Jiaozi), a popular dish that I had a chance to try last summer in Beijing. There are many recipies for Jiaozi."
- **Dictionary:** {beijing, dish, duck, rabbit, recipe}

Query: "Beijing duck recipe"

Document Matrix

Query: "Beijing duck recipe"

	Beijing	Dish	Duck	Rabbit	Recipe
D1	0	0	0.097	0	0
D2	0.199	0.199	0.097	0	0
D3	0	0	0.097	0.199	0.111
D4	0	0	0	0.398	0.222
D5	0.398	0.398	0.097	0	0.222
Query	1*.398	0	1*.097	0	1*.222

Word	IDF
Beijing	.398
Dish	.398
Duck	.097
Rabbit	.398
Recipe	.222

TF-IDF Search Exercise

- Cosine similarity of query and each doc
- $D1 = [0, 0, 0.097, 0, 0]$
- $Q = [0.398, 0, 0.097, 0, 0.222]$
- $$\cos(D1, Q) = \frac{0*0.398+0*0+0.097*0.097+0*0+0*0.222}{\sqrt{0.097^2}*\sqrt{0.398^2+0.097^2+0.222^2}}$$
- $$\cos(D1, Q) = \frac{0.00941}{0.0452} = 0.208$$

Cosine similarities

	Beijing	Dish	Duck	Rabbit	Recipe	Cos(D,Q)
D1	0	0	0.097	0	0	0.208
D2	0.199	0.199	0.097	0	0	0.639
D3	0	0	0.097	0.199	0.111	0.256
D4	0	0	0	0.398	0.222	0.232
D5	0.398	0.398	0.097	0	0.222	0.760
Query	.398	0	.097	0	.222	1

Final ordered list

- **D5** = "Last week Li has shown you how to make the Sechuan duck. Today we'll be making Chinese dumplings (Jiaozi), a popular dish that I had a chance to try last summer in Beijing. There are many recipes for Jiaozi."
- **D2** = "Beijing Duck is mostly prized for the thin, crispy duck skin with authentic versions of the dish serving mostly the skin."
- **D3** = "Bugs' ascension to stardom also prompted the Warner animators to recast Daffy Duck as the rabbit's rival, intensely jealous and determined to steal back the spotlight while Bugs remained indifferent to the duck's jealousy, or used it to his advantage. This turned out to be the recipe for the success of the duo."
- **D4** = "6:25 PM 1/7/2007 blog entry: I found this great recipe for Rabbit Braised in Wine on cookingforengineers.com."
- **D1** = "If it walks like a duck and quacks like a duck, it must be a duck."

Recommending items

- Advanced Methods
 - Classification algorithms
 - Predict either ratings or like/dislike
 - Information retrieval techniques
 - Well studied field, wide diversity of models

Content-based recommenders

Advantages

- **No community required**
 - Only need the items and a single user profile for recommendation.
- **Transparency**
 - CB models can tell you why they recommend an item, not subject to vagaries of human taste
- **Good cold start**
 - New items can be suggested before being rated by a substantial number of users.

Content-based recommenders

Disadvantages

- **Limited content analysis**
 - Requires well annotated content for good recommendations.
- **Over-specialization**
 - Users will tend to be recommended items very similar to what they have enjoyed in the past
 - Very limited discoverability
- **New users**
 - Limited user information results in bad recommendations.

Overview

- What are Recommender Systems?
- How do they work?
 - Collaborative Recommendation
 - Content-Based Recommendation
- **How do we evaluate them?**
- Example using Azure ML

Evaluating Recommendation

- Error Rate Metrics

- **Mean Absolute Error (MAE)** computes the deviation between predicted ratings and actual ratings

$$MAE = \frac{1}{n} \sum_{i=1}^n |p_i - r_i|$$

- **Root Mean Square Error (RMSE)** is similar to *MAE*, but places more emphasis on larger deviation

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (p_i - r_i)^2}$$

Metrics

- Order matters, not exact ranking value
- Graded Relevance
 - Have humans assign scores to possible results
 - Ideal results will be ordered by relevance, high to low
- Discounted cumulative gain (DCG)
 - Logarithmic reduction factor

$$DCG_{pos} = rel_1 + \sum_{i=2}^{pos} \frac{rel_i}{\log_2 i}$$

Where:

- *pos* is the length of the recommendation list
- *rel_i* returns the relevance of recommendation at position *i*

Metrics

- **Ideal discounted cumulative gain (IDCG)**

- DCG value when items are ordered perfectly

$$IDCG_{pos} = rel_1 + \sum_{i=2}^{|h|-1} \frac{rel_i}{\log_2 i}$$

- **Normalized discounted cumulative gain (nDCG)**

$$nDCG_{pos} = \frac{DCG_{pos}}{IDCG_{pos}}$$

- Normalized to the interval [0..1]

QUESTIONS

Overview

- What are Recommender Systems?
- How do they work?
 - Collaborative Recommendation
 - Content-Based Recommendation
- How do we evaluate them?
- **Example using Azure ML**