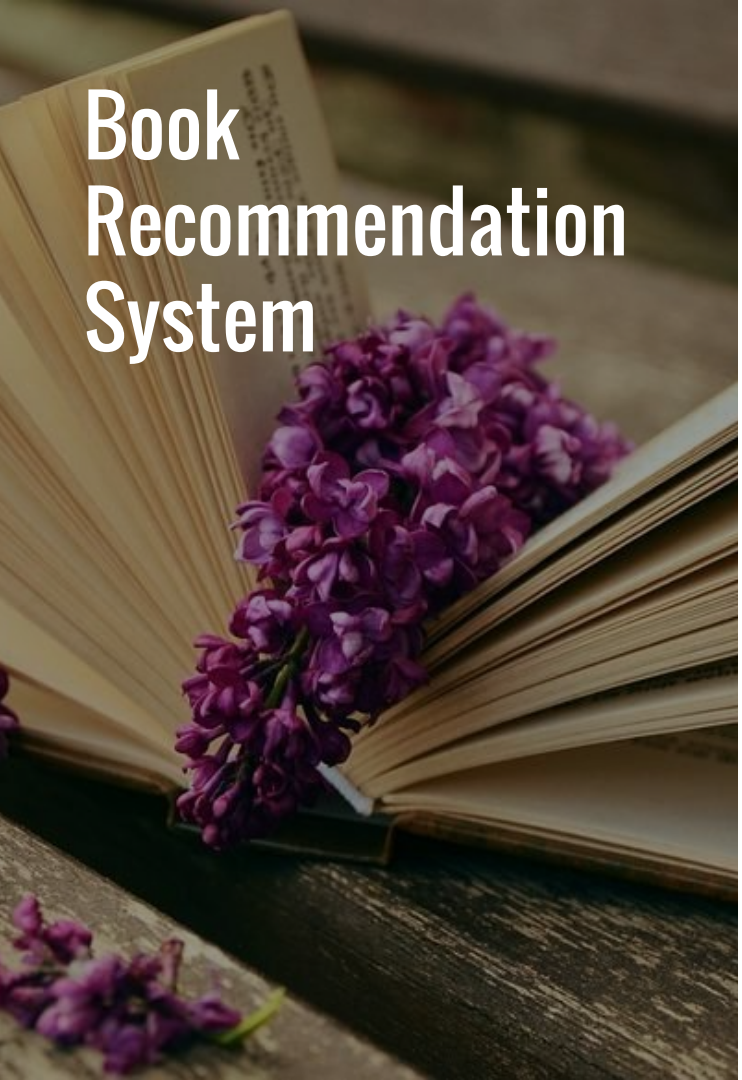
Frehiwot Bayou

Data Science Immersive (GA)



Book Recommendation System

Recommendation systems are for websites, businesses that have users looking at their product items.

Problem: Lots of users and lots of items. How to match?

: To be able to provide content while the user is still engaged

: Most recommender algorithms have cold start problem when a new user signed up or a new item is introduced.

Collaborative Algorithms

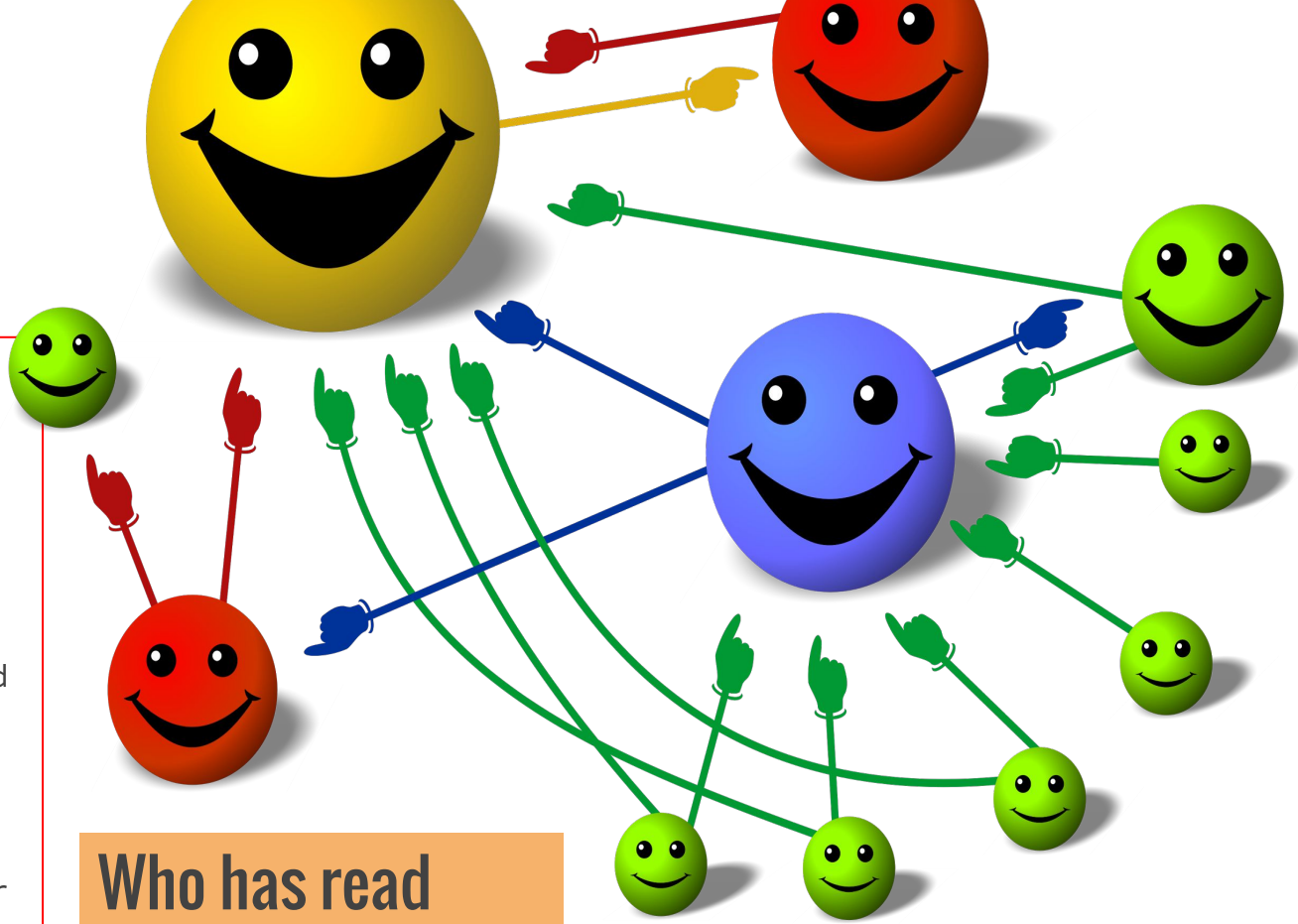
Users collaborate to review and rate products

Item to Item (similarity)

Find users that have read item A and find other books they have read and recommend accordingly.

User-Item

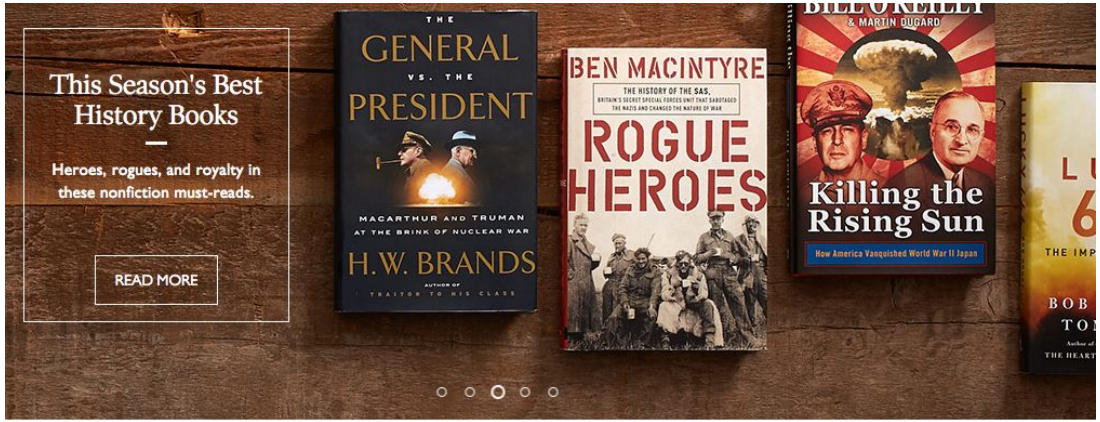
Find users that have similar ratings for items and recommend the items



Who has read what?

Content based Algorithms

Product information/content based. Author, category, year published …..



The Best Bookstore Online
Over 6 million books ready for shipment within 24 hours and 4.5 million eBooks to download now.

Recommended For You

Bestselling Popular

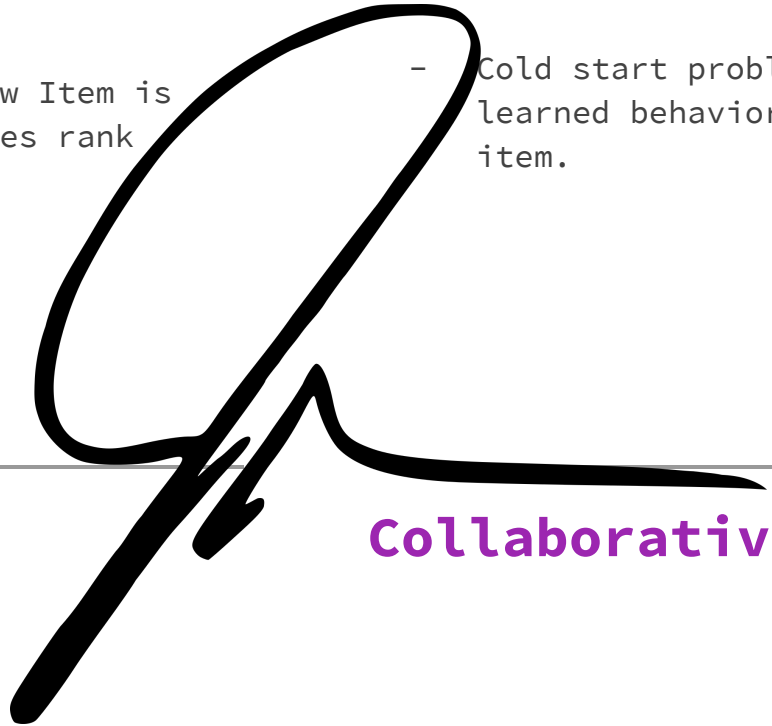
Problem

Content based

- dependent on User Ratings or some ranking as sales rank.
- not personalized
- Cold start problem: when new Item is introduced, there is no sales rank or rating.

Collaborative

- Highly dependent on User Ratings
- Cold start problem there is no learned behavior for the new user or item.



Exploring Co-occurrence Matrices and Recommenders

Work-in-Progress by Brad Lyon

This tool is intended to provide additional insight into how some Recommender systems work by showing the intermediate details.

New Random Values

Entries of matrices are blank where the value is zero. You can mouseover matrix entries to see connections with other entities.

Raw Data

User, Item Interacted With

user2, item1
user3, item5
user4, item2
user1, item6
user4, item4
user3, item7
user2, item3
user3, item2
user2, item2
user4, item1
user4, item7
user4, item6
user3, item4
user1, item7
user4, item8

This is the user-item matrix **A**, which summarizes the interactions between users and items. It is derived from the raw data, with one row for each user, and one column for each item.

A =

	item1	item2	item3	item4	item5	item6	item7	item8
user1						1	1	
user2	1	1	1					
user3		1		1	1		1	
user4	1	1						

This entry is nonzero because there is an interaction between user3 and item4 in the raw data.

The Item-Item Co-occurrence Matrix is calculated from $A^T A$

A^T A =

	user1	user2	user3	user4
item1		1		1
item2		1	1	1
item3		1		
item4			1	1
item5			1	
item6	1			1
item7	1		1	1
item8				1

	item1	item2	item3	item4	item5	item6	item7	item8
user1						1	1	
user2	1	1	1					
user3		1		1	1		1	
user4	1	1		1		1	1	1

=

	item1	item2	item3	item4	item5	item6	item7	item8
item1	2	2	1	1		1	1	1
item2	2	3	1	2	1	1	2	1
item3	1	1	1					
item4	1	2		2	1	1	2	1
item5		1		1	1		1	
item6	1	1		1		2	2	1
item7	1	2		2	1	2	3	1
item8	1	1		1		1	1	1

Recommendations for Users are Obtained by Multiplying the Co-occurrence Matrix $A^T A$ by the User's Interaction Vector

$(A^T A)$ u =

Co-occurrence Matrix $A^T A$

	item1	item2	item3	item4	item5	item6	item7	item8
item1	2	2	1	1		1	1	1
item2	2	3	1	2	1	1	2	1
item3	1	1	1					
item4	1	2		2	1	1	2	1
item5		1		1	1		1	
item6	1	1		1		2	2	1
item7	1	2		2	1	2	3	1
item8	1	1		1		1	1	1

User Interaction Vector **u**

Item	Has Interaction?
item1	☐
item2	☐
item3	☑
item4	☐
item5	☐
item6	☑
item7	☐
item8	☐

=

Recommendation Weights

Item	Value
item1	2
item2	2
item3	1
item4	1
item5	0
item6	2
item7	2
item8	1

Content Based: model on amazon book reviews

```
model = graphlab.recommender.create(reviews, user_id='reviewerID', item_id='title')
```

```
similar_items=model.get_similar_items()
similar_items.head(10)
```

title	similar	score	rank
Master Georgie	Breakfast on Pluto	0.0571428537369	1
Master Georgie	Arcadia	0.0454545617104	2
Master Georgie	BRIXTON BEACH	0.0454545617104	3
Master Georgie	N. C. Wyeth: A Biography	0.0434782505035	4
Master Georgie	Inventing Memory: A Novel of Mothers and Daughters ...	0.0434782505035	5
Master Georgie	Without Reservation: The Making of America's Most ...	0.0400000214577	6
Master Georgie	Catholics	0.0400000214577	7
Master Georgie	Fanny: A Fiction	0.0400000214577	8
Master Georgie	Andrew Wyeth: A Secret Life ...	0.0384615659714	9
Master Georgie	The Cloud Sketcher	0.0384615659714	10

```
Optimization Complete: Maximum number of passes through

Computing final objective value and training RMSE.

       Final objective value: 1.25533

       Final training RMSE: 1.12042

Training RMSE 1.12041630822
Test RMSE 1.0051544861
```

```
Optimization Complete: Maximum number of passes thr

Computing final objective value and training RMSE.

       Final objective value: 0.175321

       Final training RMSE: 0.418712

Training RMSE 0.432119024967
Test RMSE 1.0051544861
```

```
Training RMSE None
Test RMSE 4.3320616433
```

Similarity Model


turi
create intelligence™


amazon
web services

Contact

Your Name
no_reply@example.com
www.example.com