Understanding Algorithms for Recommendation Systems

UNDERSTANDING TASKS PERFORMED BY RECOMMENDATION SYSTEMS

Swetha Kolalapudi
CO-FOUNDER, LOONYCORN

www.loonycorn.com

Overview

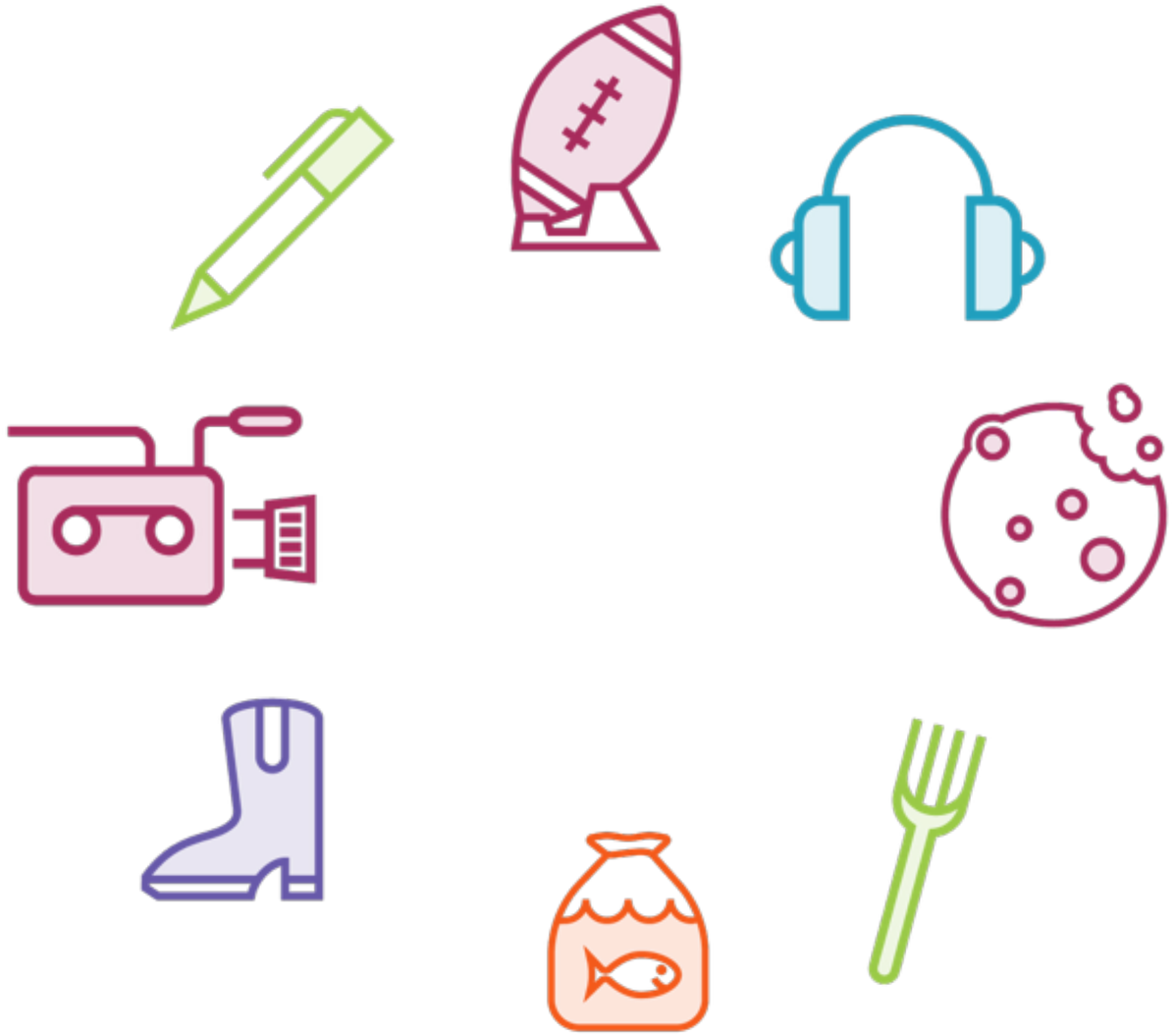
Finding relationships among users and products

Monetizing the insight from these relationships

Understanding the different types of recommendation algorithms

- Content based Filtering
- Collaborative filtering
- Association rules learning

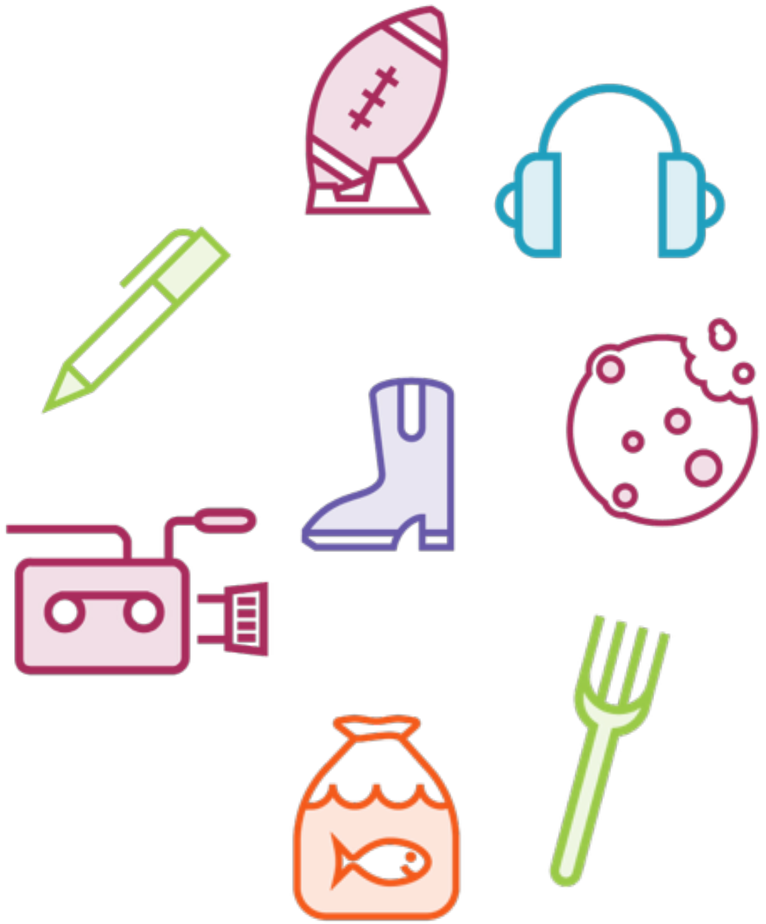
Products Come in a Variety of Forms



People Come in a Variety of Forms

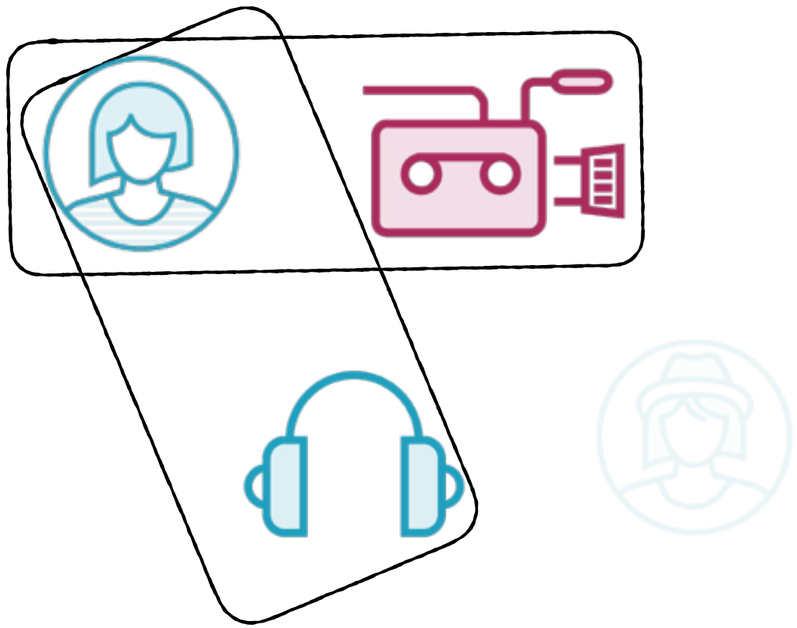
People and Products

Different people have different preferences and requirements



People and Products

Some people have a preference for specific products



Some people have a preference for specific products

This is measured using

User-Product Relationships

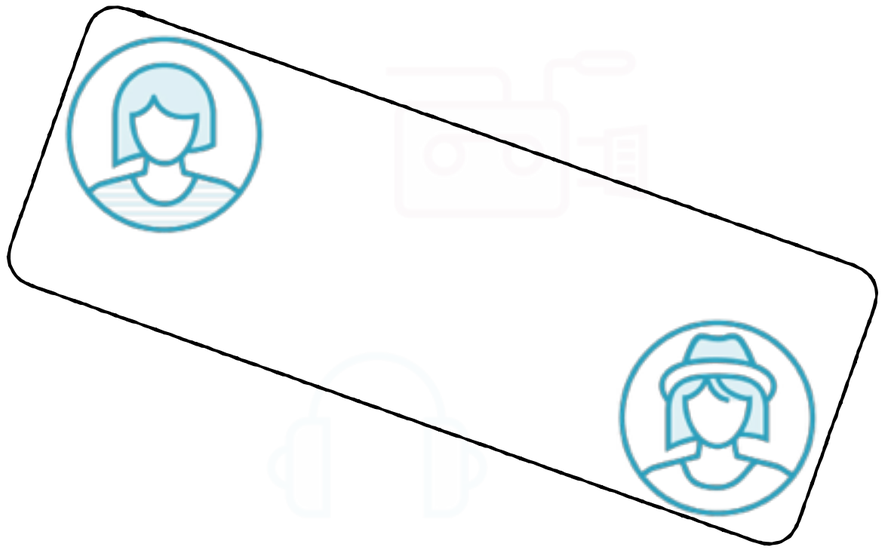
Some products are similar in nature

- Books of the same genre
- Dishes from the same cuisine
- News articles about an event

Some products are similar in nature

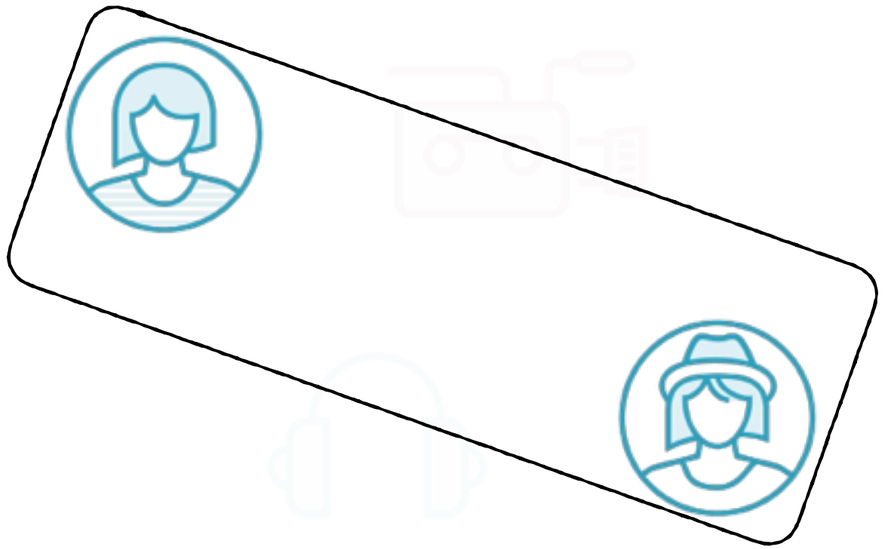
This is measured using

Product-Product Relationships



Some people are similar in nature

- They like the same books
- They have common friends
- They have similar backgrounds



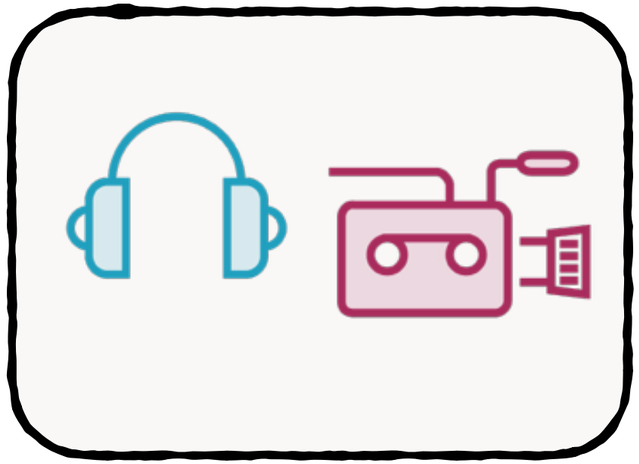
Some people are similar in nature

This is measured using

User-User Relationships

Relationships Among Users and Products

User-Product Relationships



Product-Product Relationships

User-User Relationships

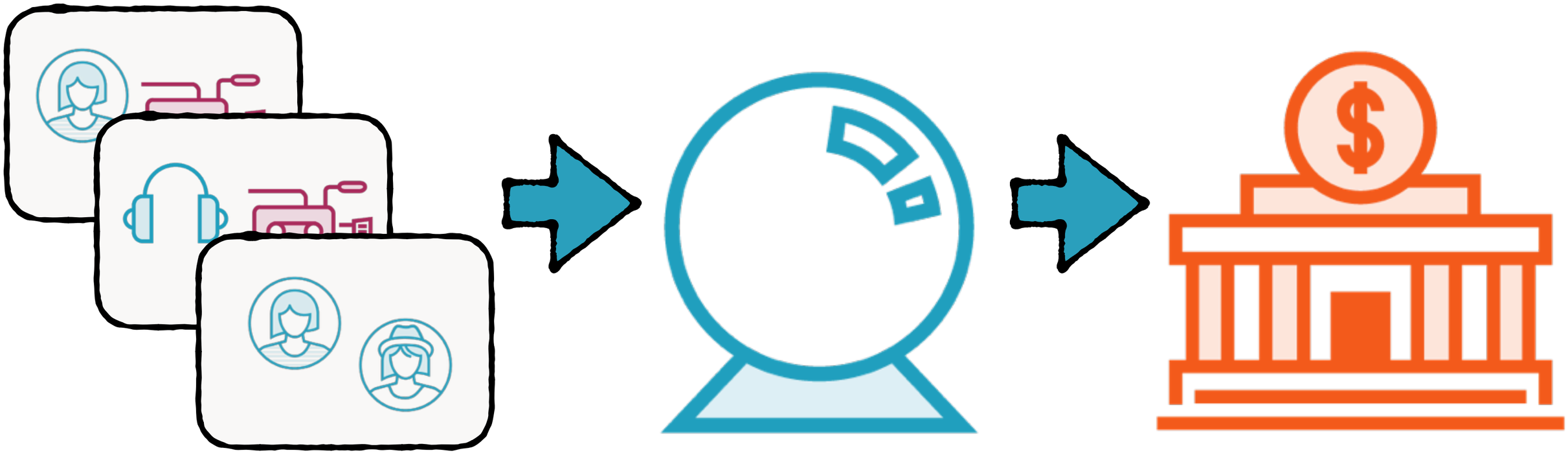
Relationships Among Users and Products

Relationships

These relationships can provide tremendous insight

- What books will a person like?
- If a person buys a phone, what else will they buy?
- If A knows B, and B knows C, does A know C?

Insight is Monetizable



Relationships

Insight

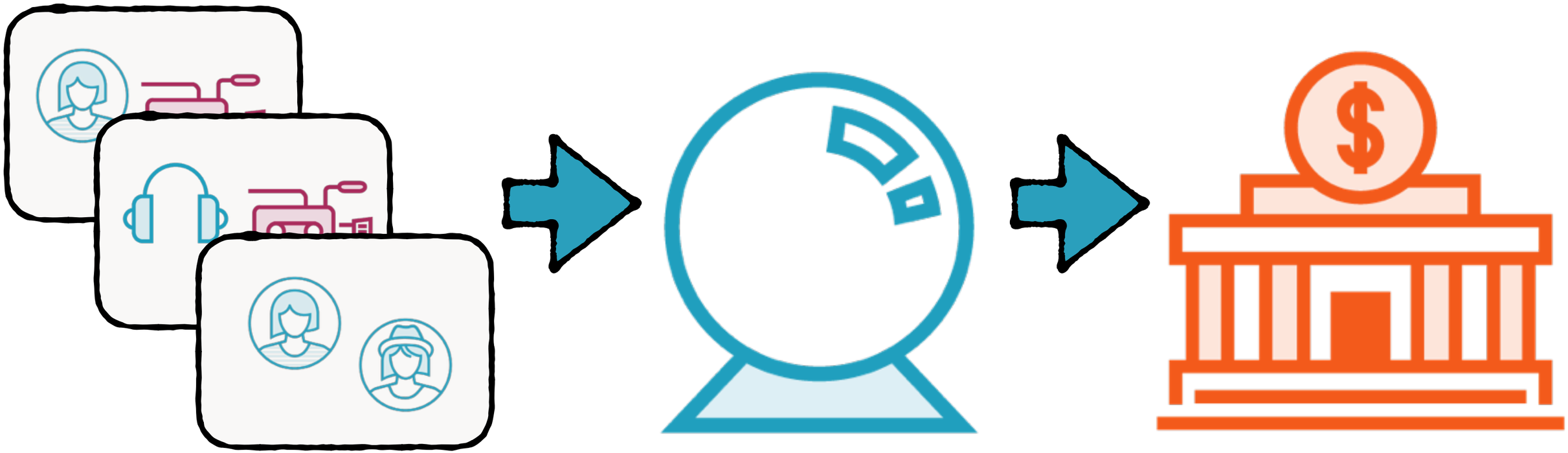
Monetization

Insight is Monetizable

Monetization

- **Personalized** promotional emails
- **Personalized** homepage
- **Personalized** notifications

Insight is Monetizable

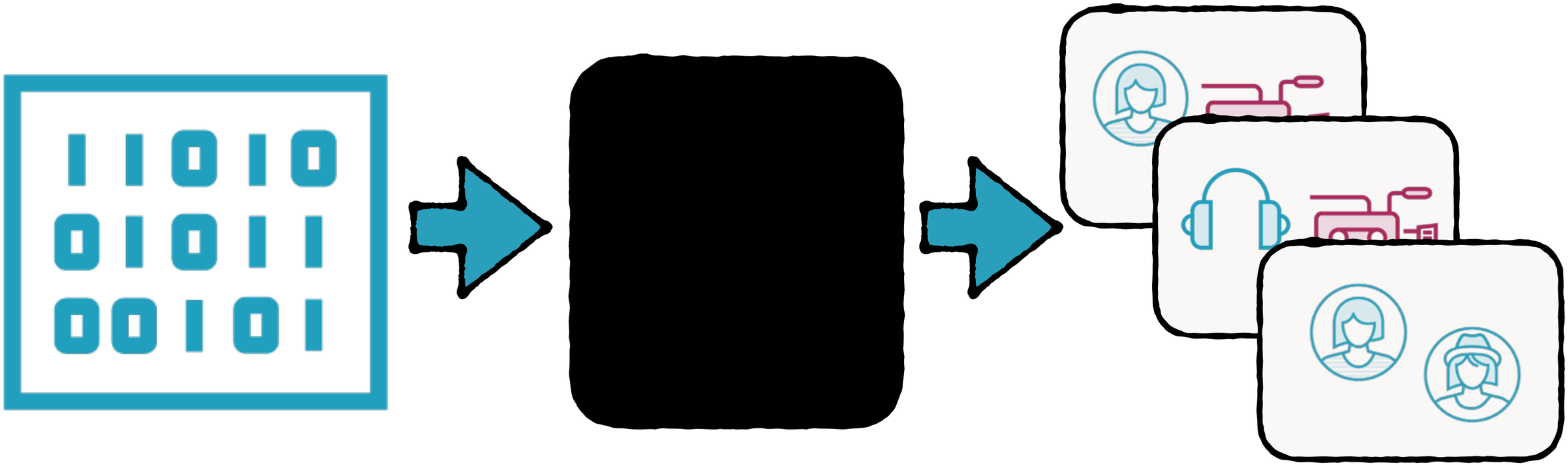


Relationships

Insight

Monetization

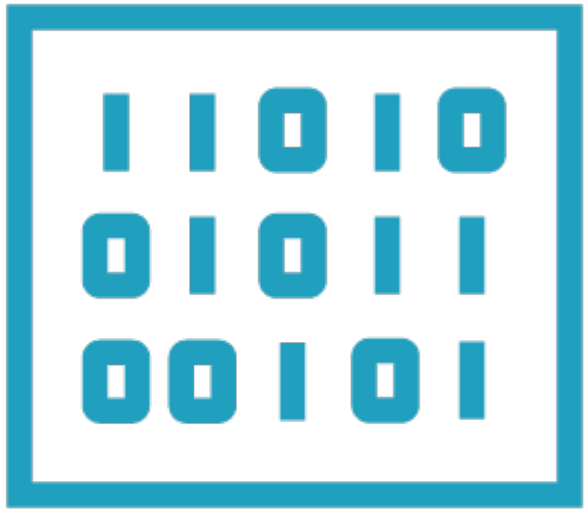
Identifying Relationships



Data

Recommendation Algorithms

Relationships



Data

User Behavior Data

- Ratings, Clicks, Purchases

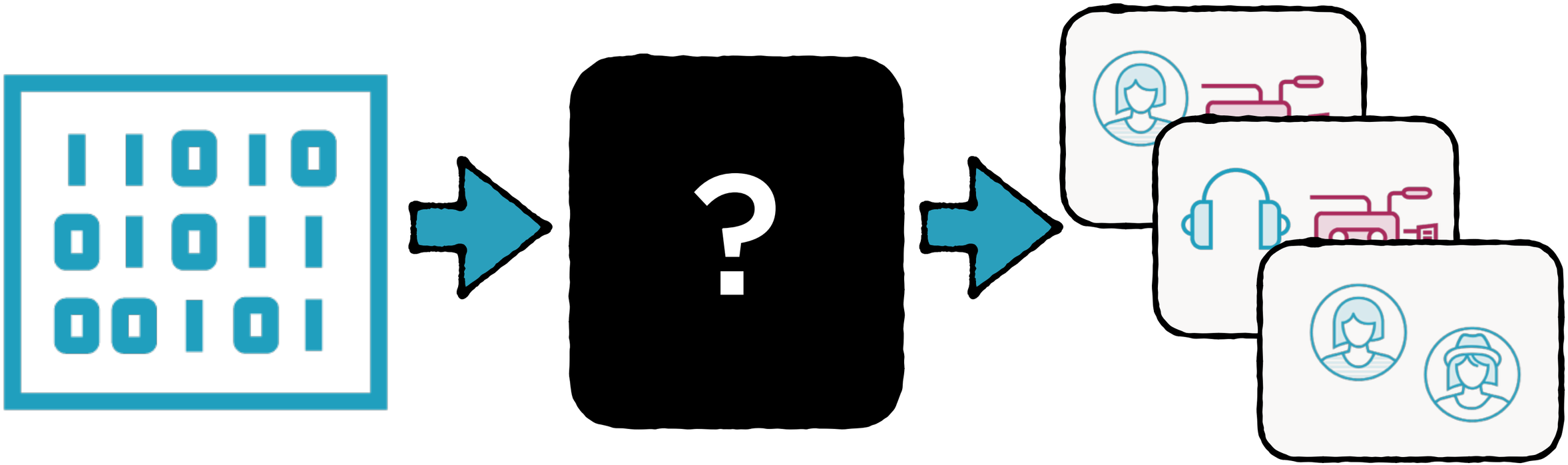
User Demographic Data

- Age, Education, Income, Location

Product Attribute Data

- Genre (Books), Cast (Movies), Cuisine (Food)

Identifying Relationships



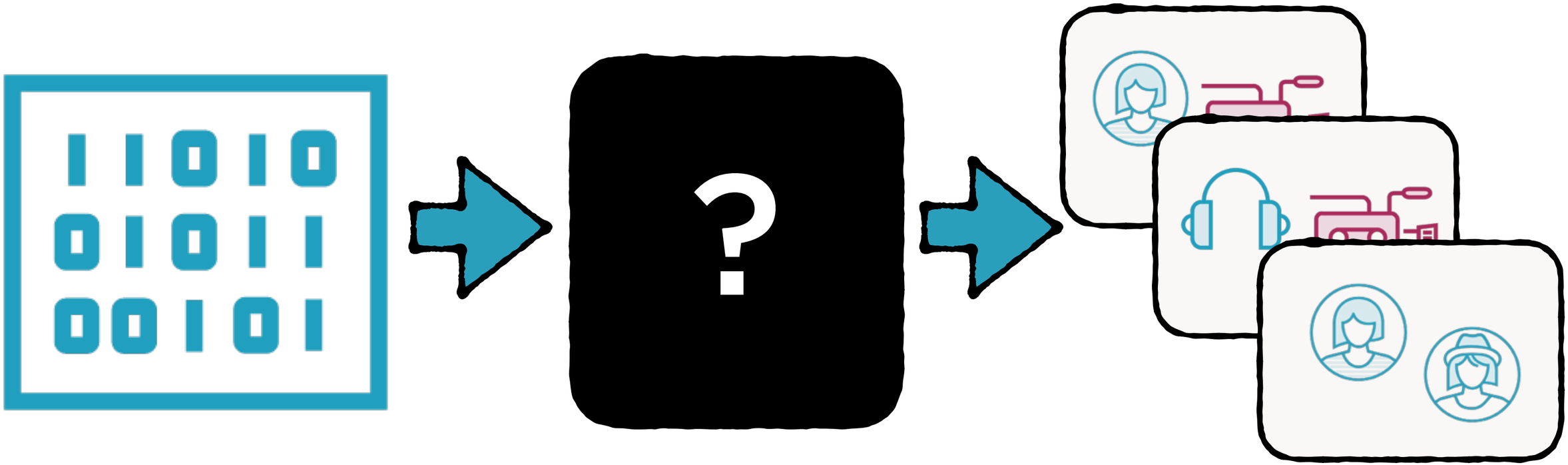
Data

Recommendation Algorithms

Relationships

Types of Recommendation Algorithms

Identifying Relationships

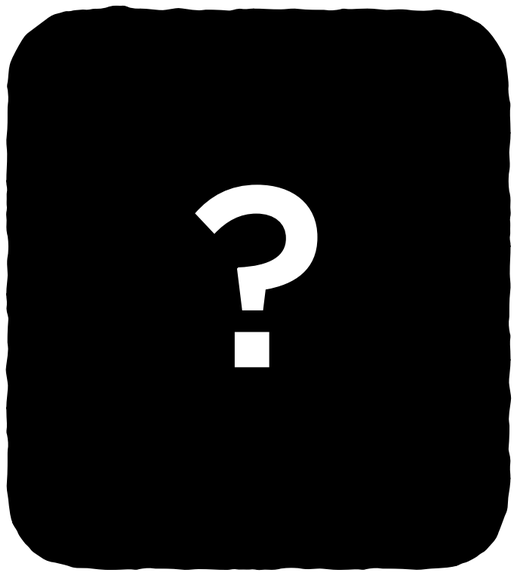


Data

Recommendation Algorithms

Relationships

Recommendation Algorithms



Recommendation Algorithms

You know what products a user already likes

- Ratings, Purchases, Clicks

What other products should you recommend to that user?

Recommendation Algorithms

Option 1

Find products with “similar” attributes

Option 2

Find products liked by “similar” users

Option 3

Find “complementary” products

Recommendation Algorithms

Content based filtering

Option 1

Find products with “similar” attributes

Collaborative filtering

Option 2

Find products liked by “similar” users

Association rules learning

Option 3

Find “complementary” products

Recommendation Algorithms

Content based filtering

Find products with “similar” attributes

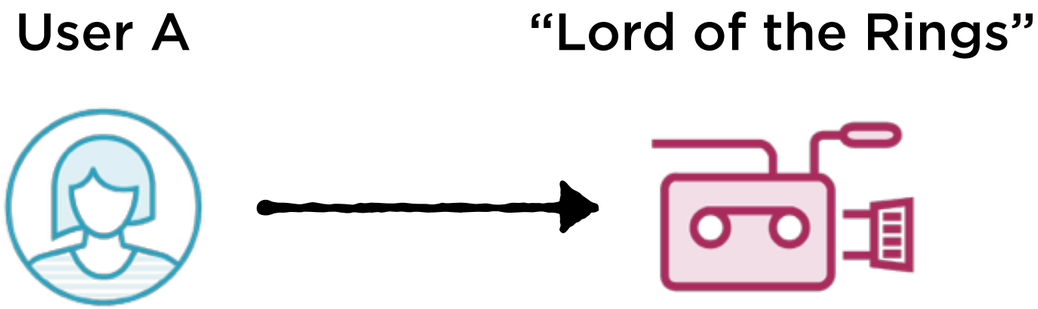
Collaborative filtering

Find products liked by “similar” users

Association rules learning

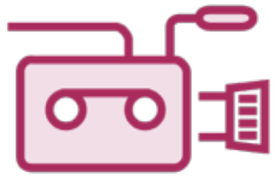
Find “complementary” products

User A likes the movie
“Lord of the Rings”



A database has ratings against different attributes for “Lord of the Rings”

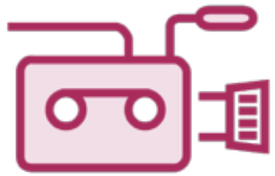
“Lord of the Rings”



	Rating
Direction	10
Cast	8
Cinematography	10
Story	9

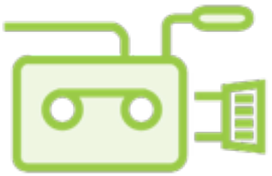
“The Hobbit” has very similar ratings against these same attributes

“Lord of the Rings”



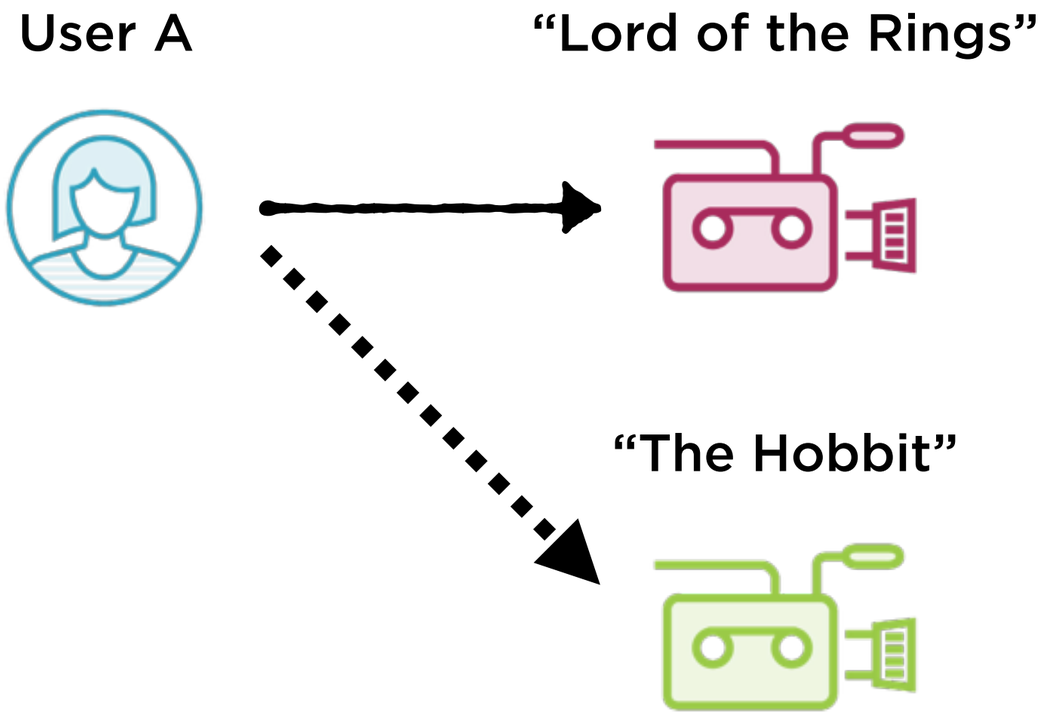
	Rating
Direction	10
Cast	8
Cinematography	10
Story	9

“The Hobbit”



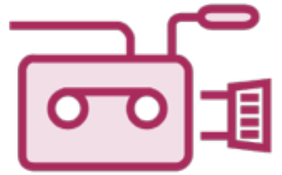
Rating
9.5
8
9
10

Recommend “The Hobbit” to User A



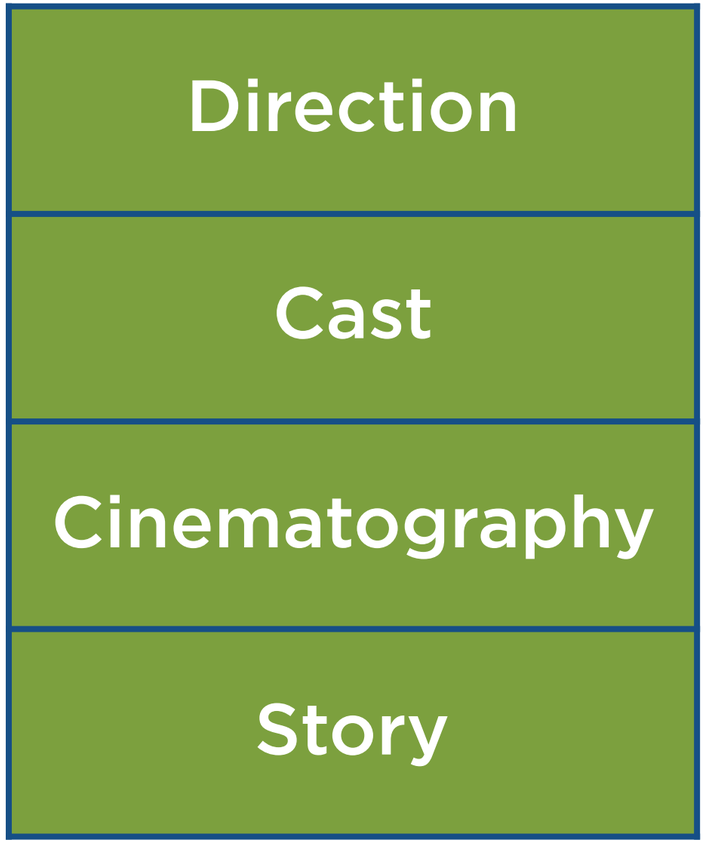
Recommendation is based on the “similarity” of products

“Lord of the Rings”



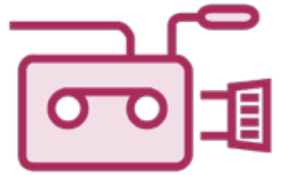
“The Hobbit”

Here, “similarity” is measured against product attributes



“Similarity” can be measured in other indirect ways

“Lord of the Rings”



“The Hobbit”

Recommendation Algorithms

Content based filtering

Find products with “similar” attributes

Collaborative filtering

Find products liked by “similar” users

Association rules learning

Find “complementary” products

Recommendation Algorithms

Content based filtering

Find products with “similar” attributes

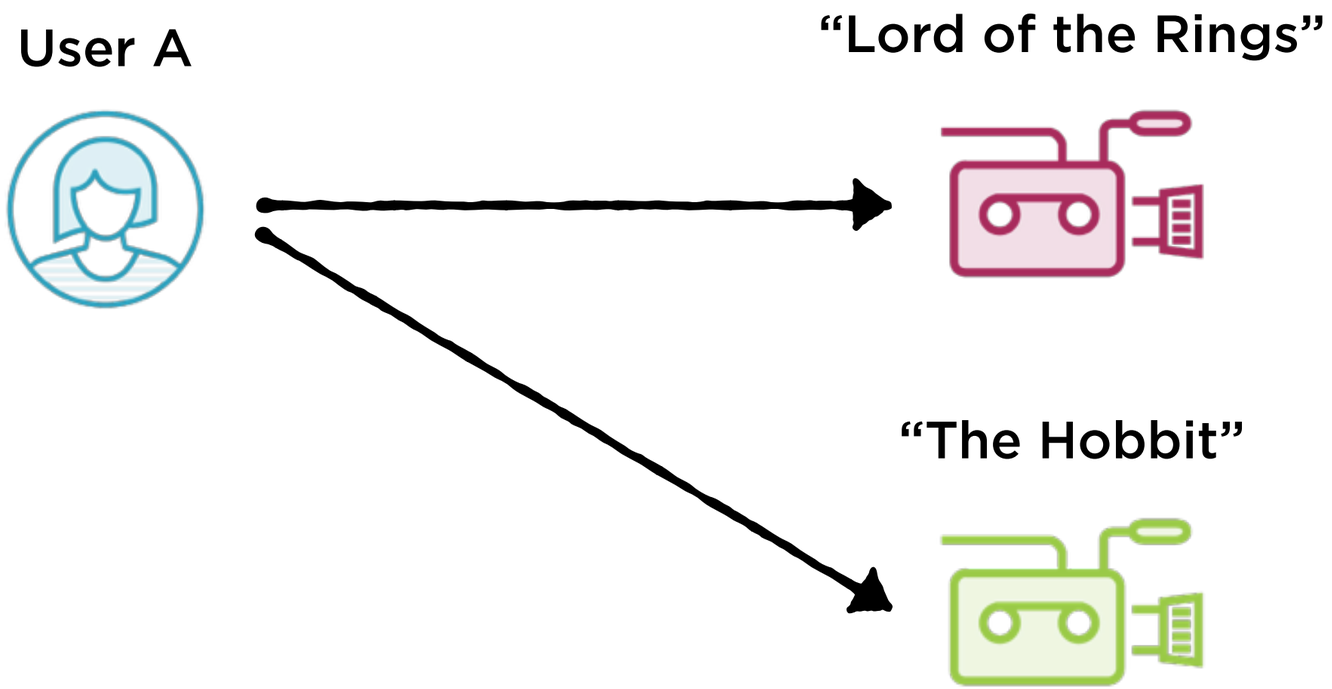
Collaborative filtering

Find products liked by “similar” users

Association rules learning

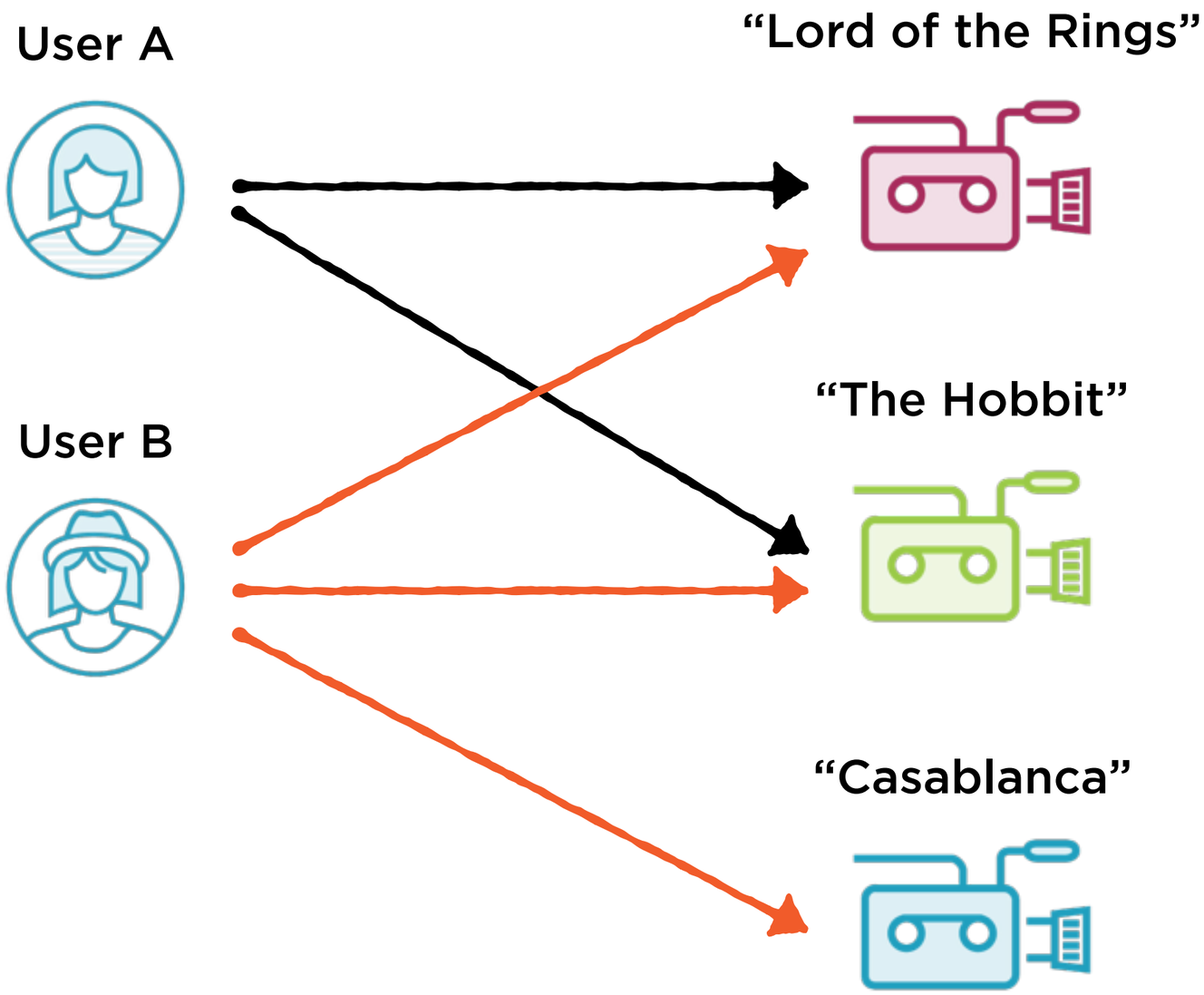
Find “complementary” products

User A likes the movies “Lord of the Rings” and “The Hobbit”

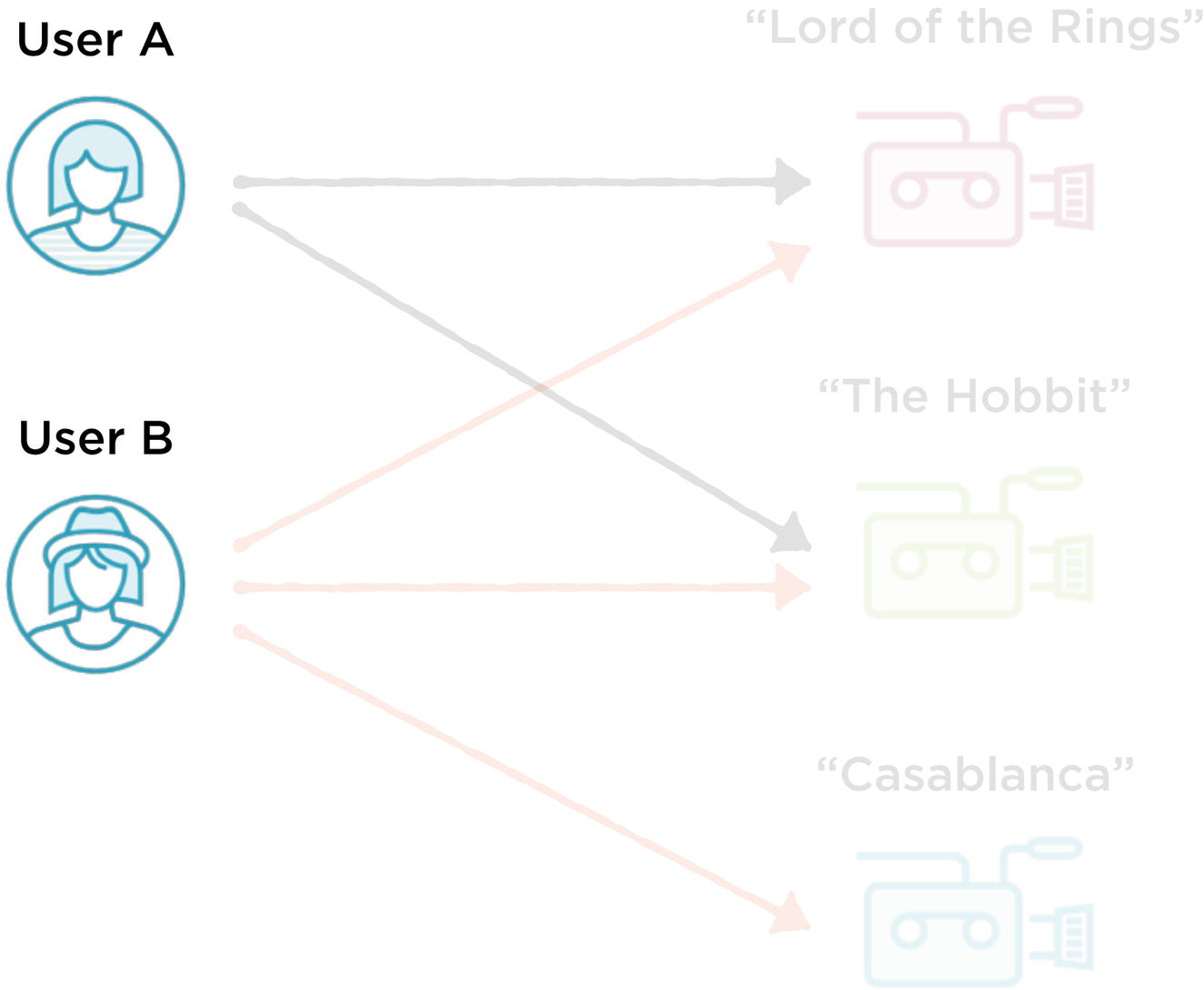


User B likes some of the same movies as User A

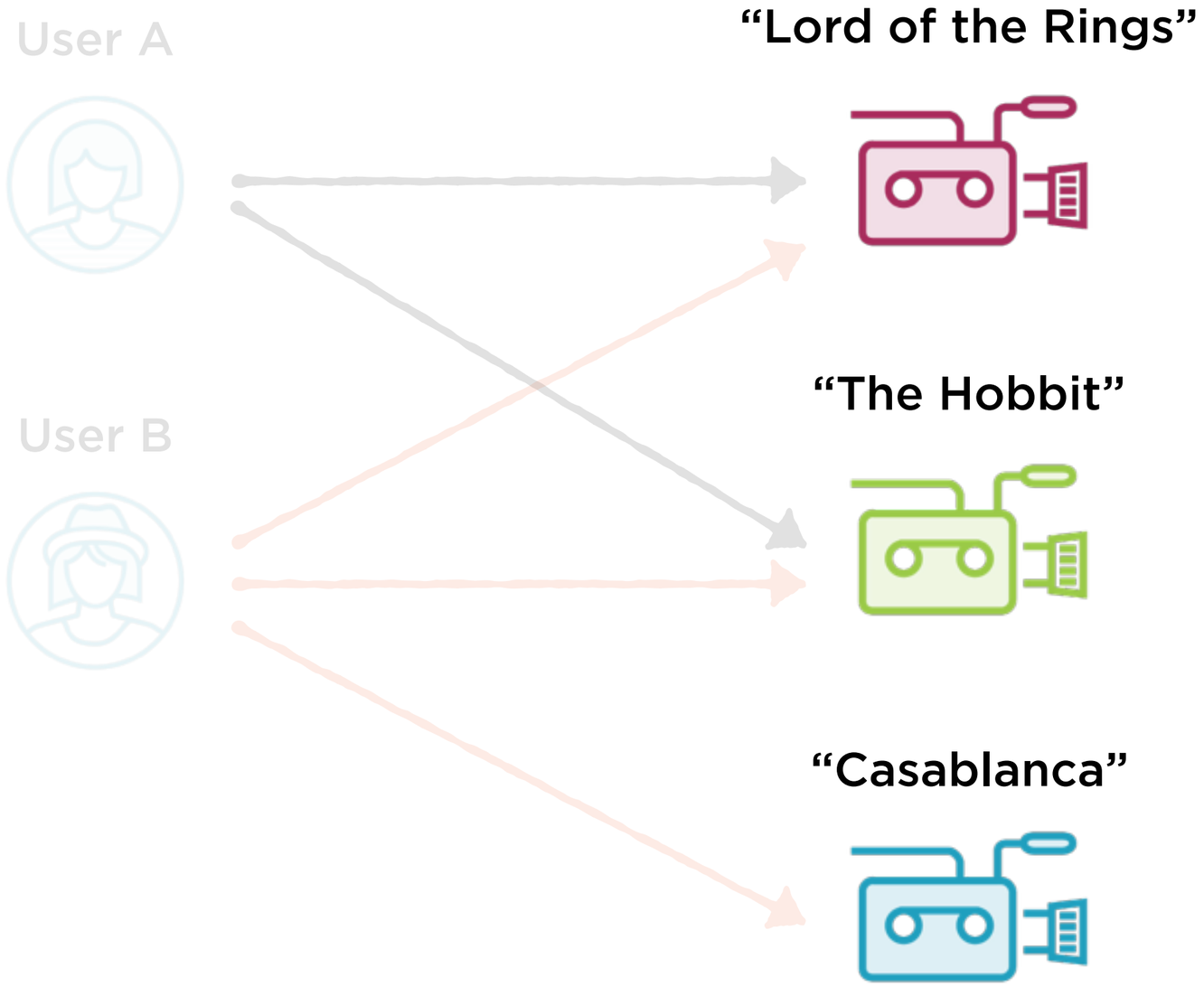
.. and some other movies



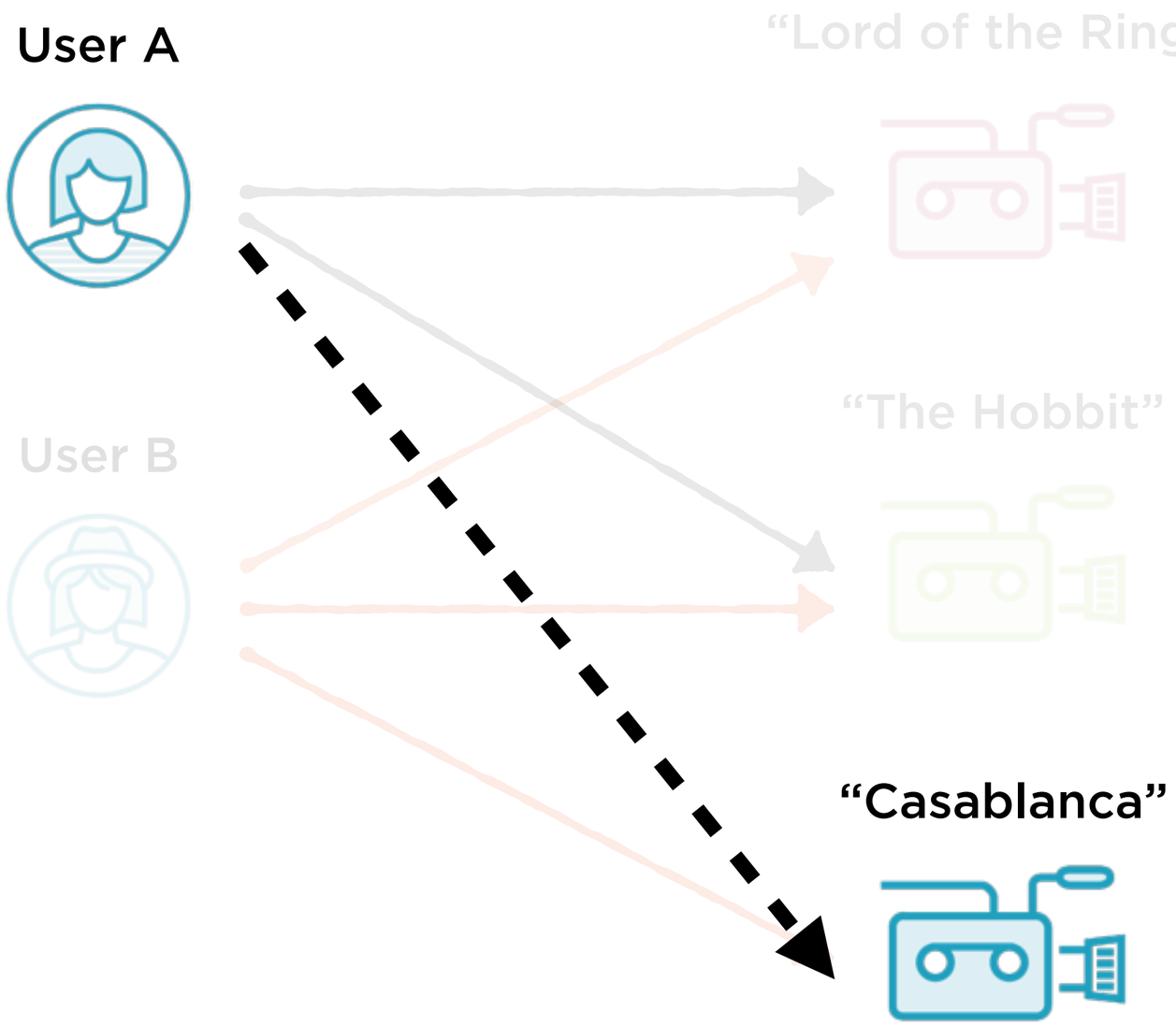
User B is “similar” to User A



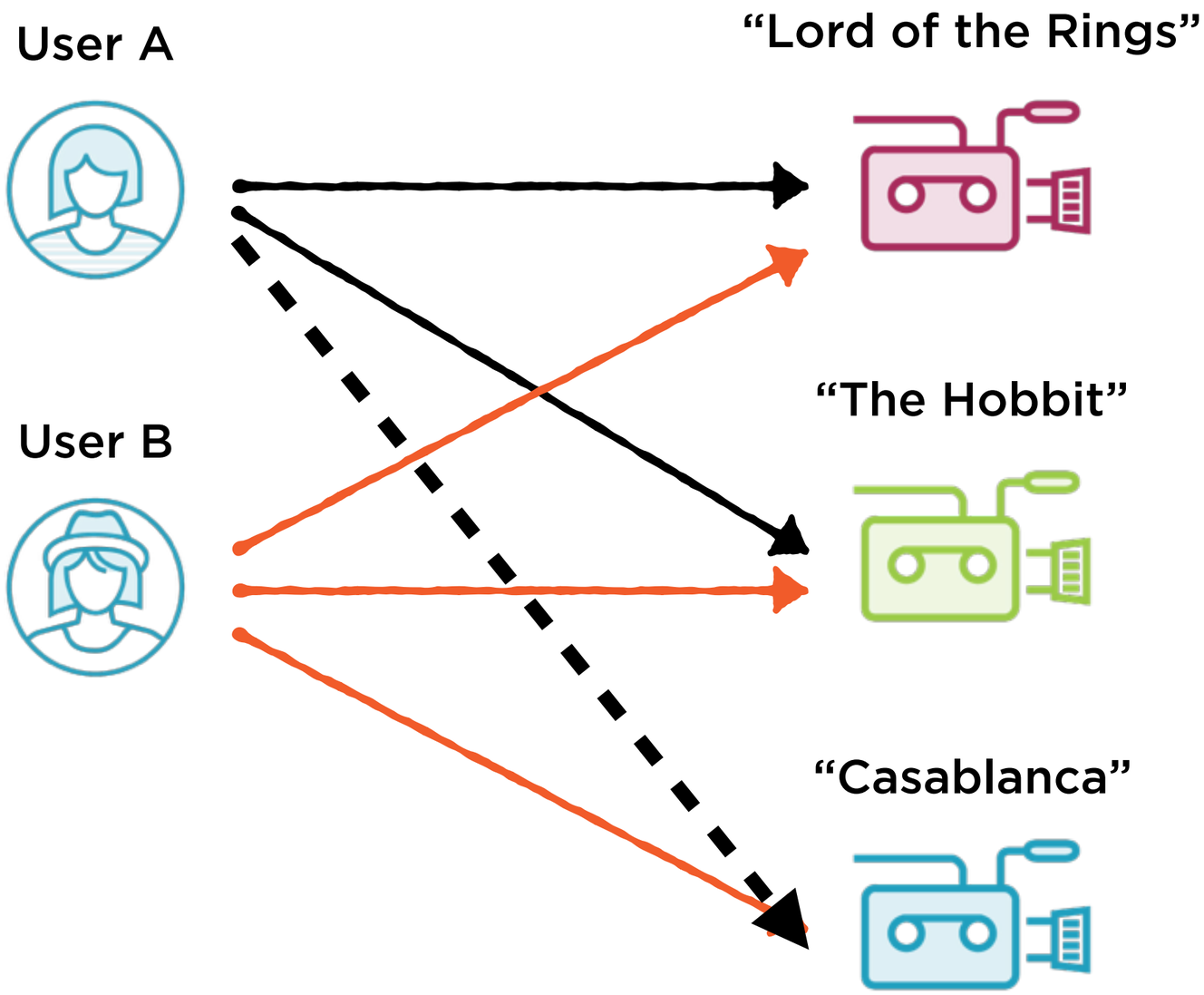
Movies that User B likes are “similar” to movies that User A likes



Recommend movies that User B likes to User A



"Similarity" of products is indirectly measured using "similarity" of users



Recommendation Algorithms

Content based filtering

Find products with “similar” attributes

Collaborative filtering

Find products liked by “similar” users

Association rules learning

Find “complementary” products

Recommendation Algorithms

Content based filtering

Find products with “similar” attributes

Collaborative filtering

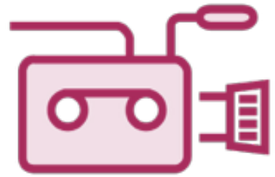
Find products liked by “similar” users

Association rules learning

Find “complementary” products

In movie recommendations, the products are substitutable for each other

“Lord of the Rings”



“The Hobbit”

“Casablanca”

You can also recommend complementary products

Users who buy smartphones also like to buy headphones

These are “associated” products

Recommendation Algorithms

Content based filtering

Find products with “similar” attributes

Collaborative filtering

Find products liked by “similar” users

Association rules learning

Find “complementary” products

Recommendation Algorithms

Content based filtering

Find products with “similar” attributes

Collaborative filtering

Find products liked by “similar” users

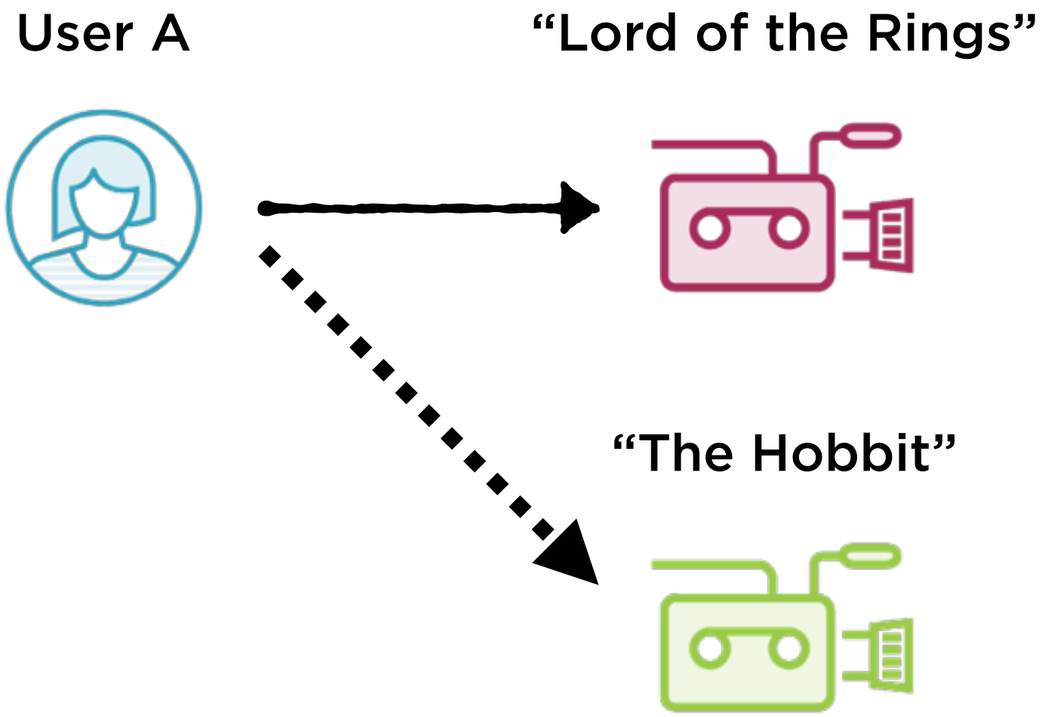
Association rules learning

Find “complementary” products

Digging Deeper into Content Based Filtering

Content Based Filtering

Recommend products which have “similar” attributes



“Similar” Attributes

“Similar” Attributes

What attributes should be used?

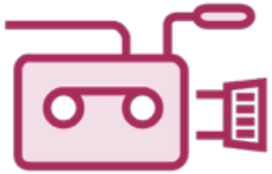
“Similar” Attributes

What attributes should be used?

How do we measure “similarity”?

What attributes should be used?

"Lord of the Rings"



Direction	✓
Cast	✓
Cinematography	✓
Story	✓
# Extras used	✗
Production Time	✗

What attributes should be used?

"The Hard Thing about Hard Things"

Author	✓
Story	✓
Genre	✓
ISBN number	✗

What attributes should be used?

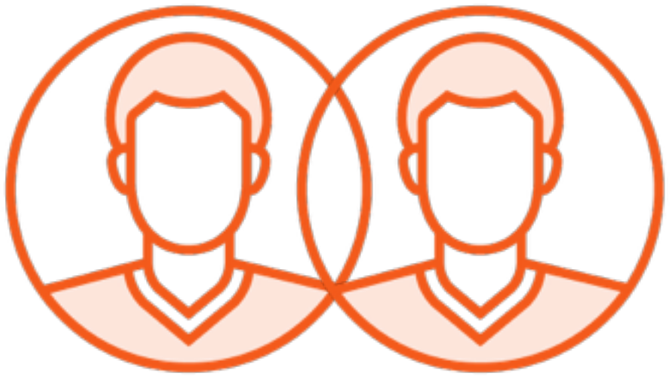
Identify attributes/ factors that influence user preferences

How do we measure “similarity”?

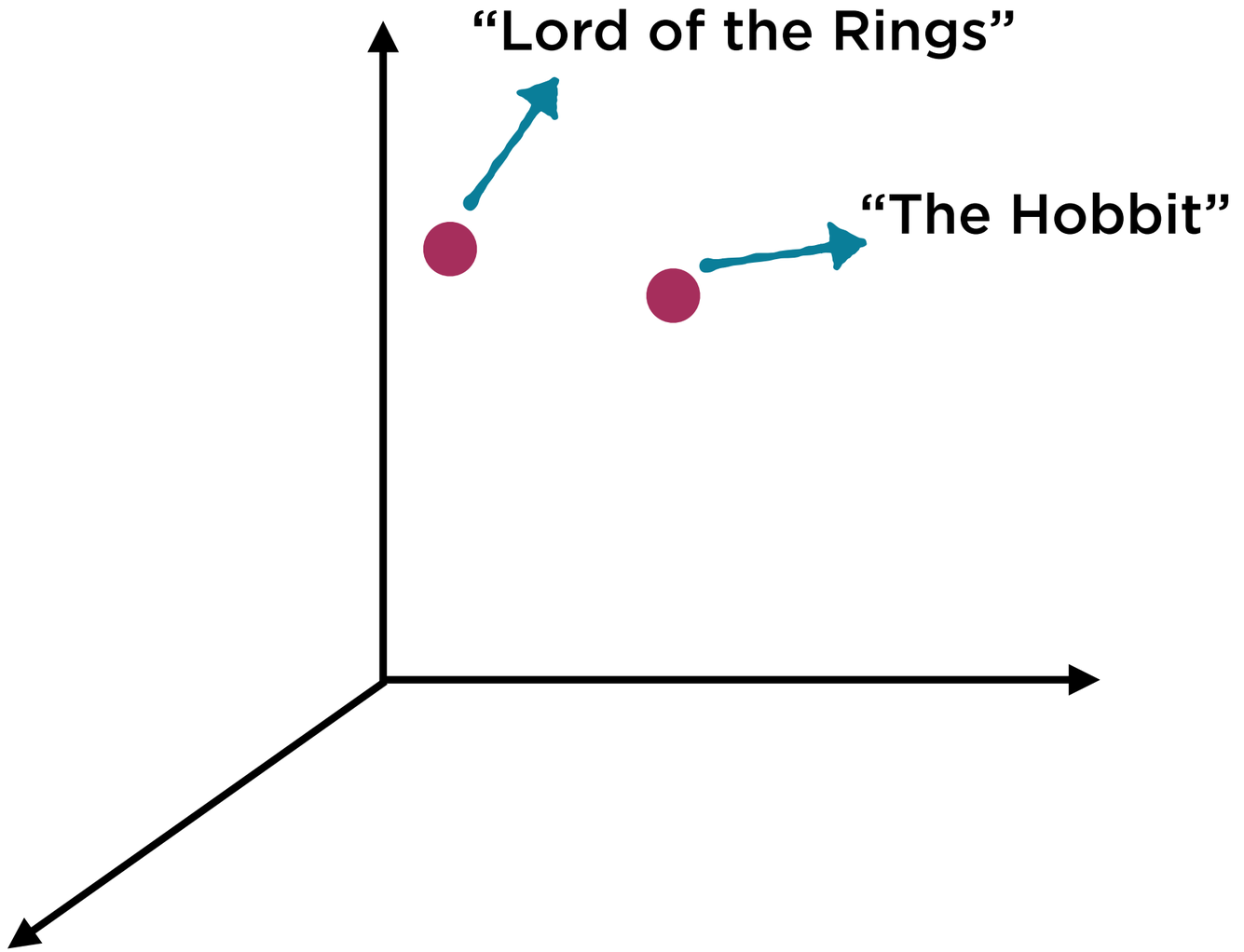
Rate the products against chosen attributes

	“Lord of the Rings”	“The Hobbit”
Direction	10	9.5
Cast	8	8
Cinematography	10	9
Story	9	10

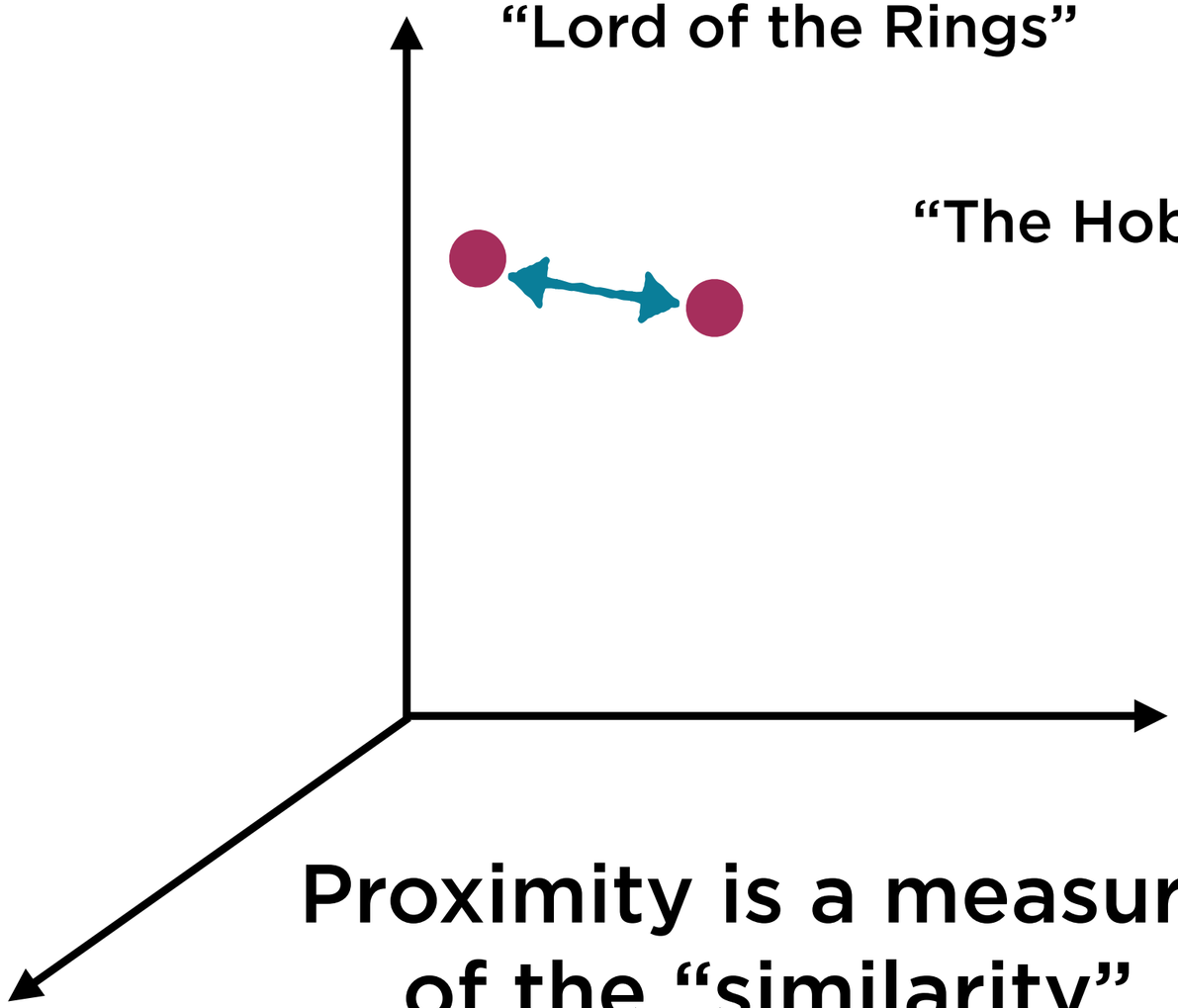
Points in N-Dimensional Space



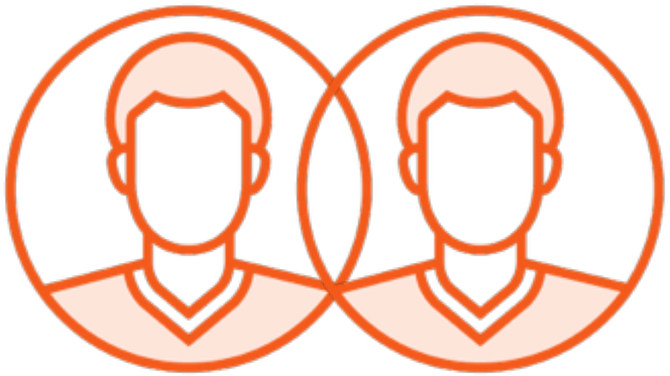
How do we measure “similarity”?



How do we measure “similarity”?



Proximity is a measure of the “similarity”



How do we measure “similarity”?

“Similarity” is measured using distance metrics

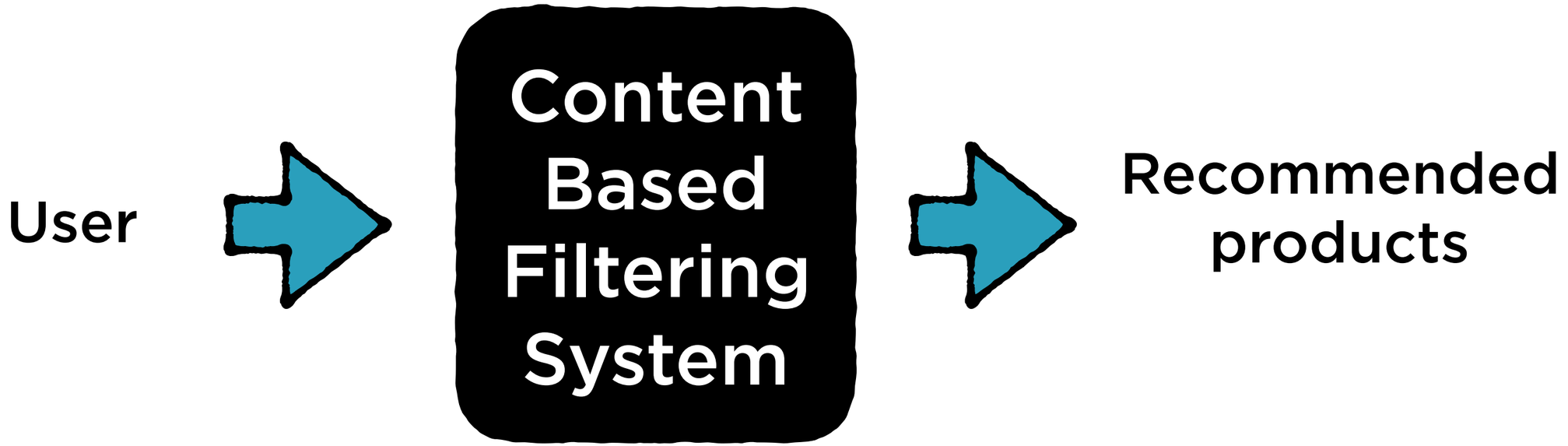
Examples:

Euclidean distance

Hamming distance

Correlation distance

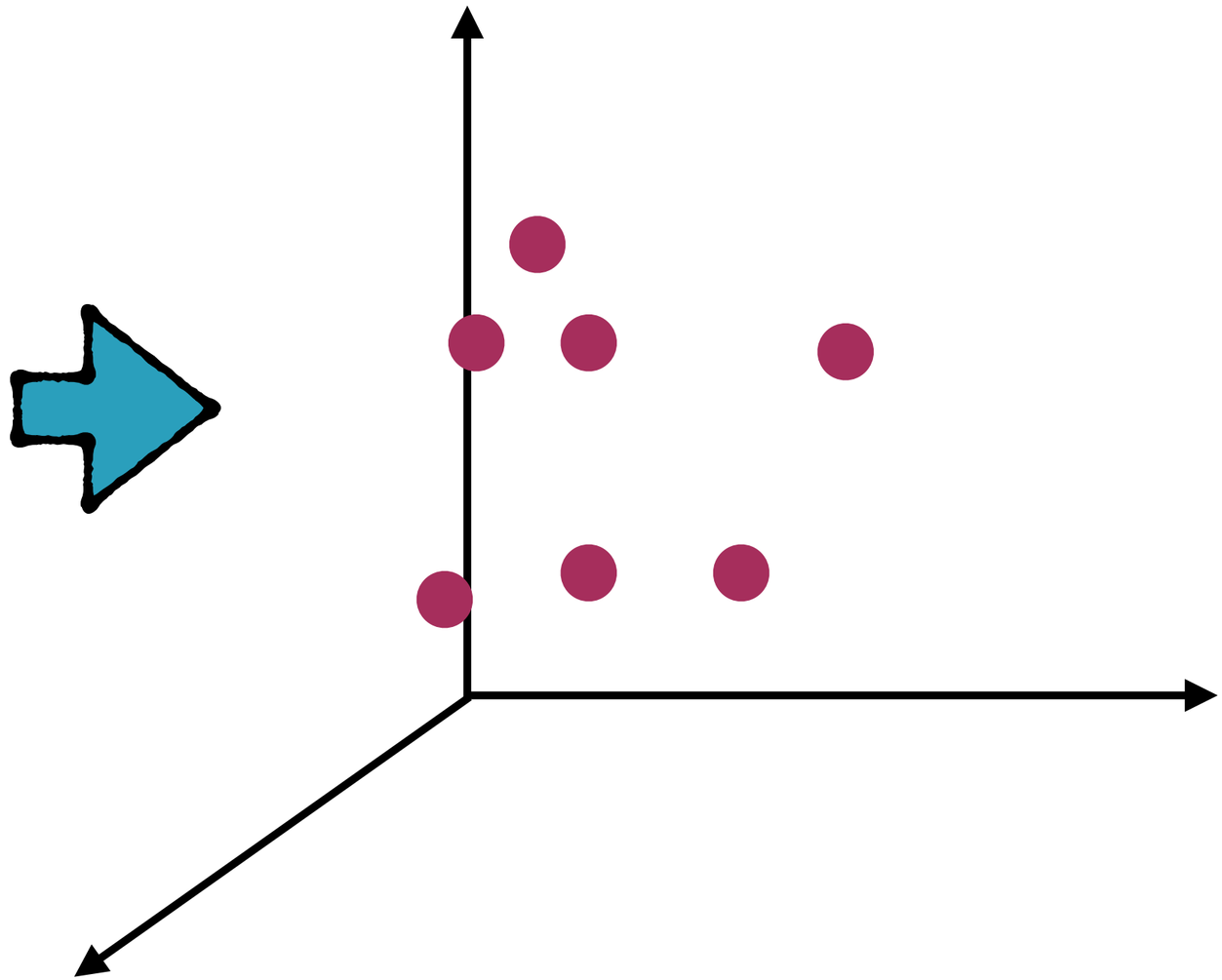
A Full-Fledged Content Based Filtering System



A Full-Fledged Content Based Filtering System

Rate every product against the relevant attributes

Product	F1	F2	F3	F4
A	0	3	2	5
B	5	2	3	4
C	4	5	2	1
D	3	4	5	2

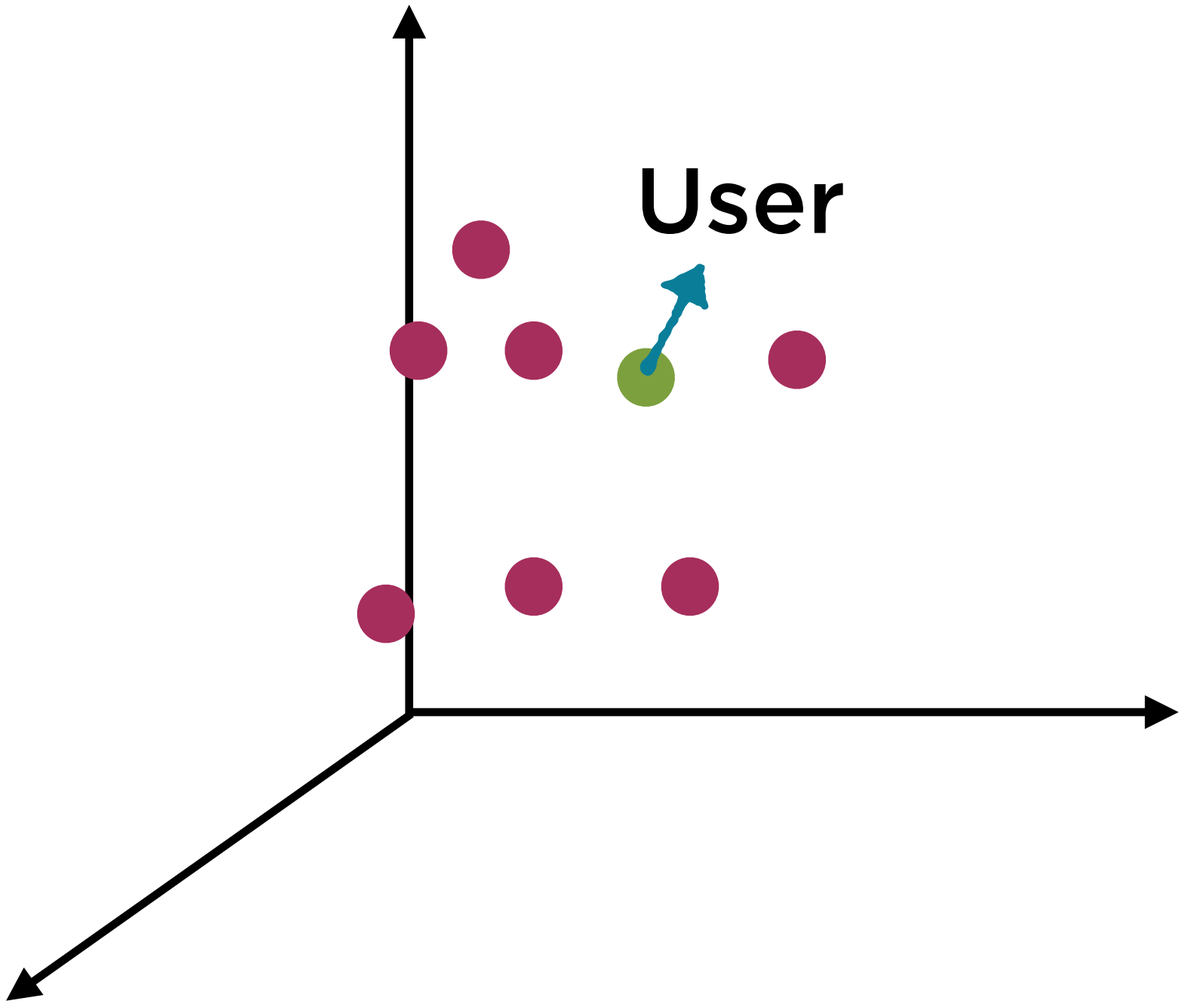


A Full-Fledged Content Based Filtering System

Rate the user on the importance he/she gives to these factors

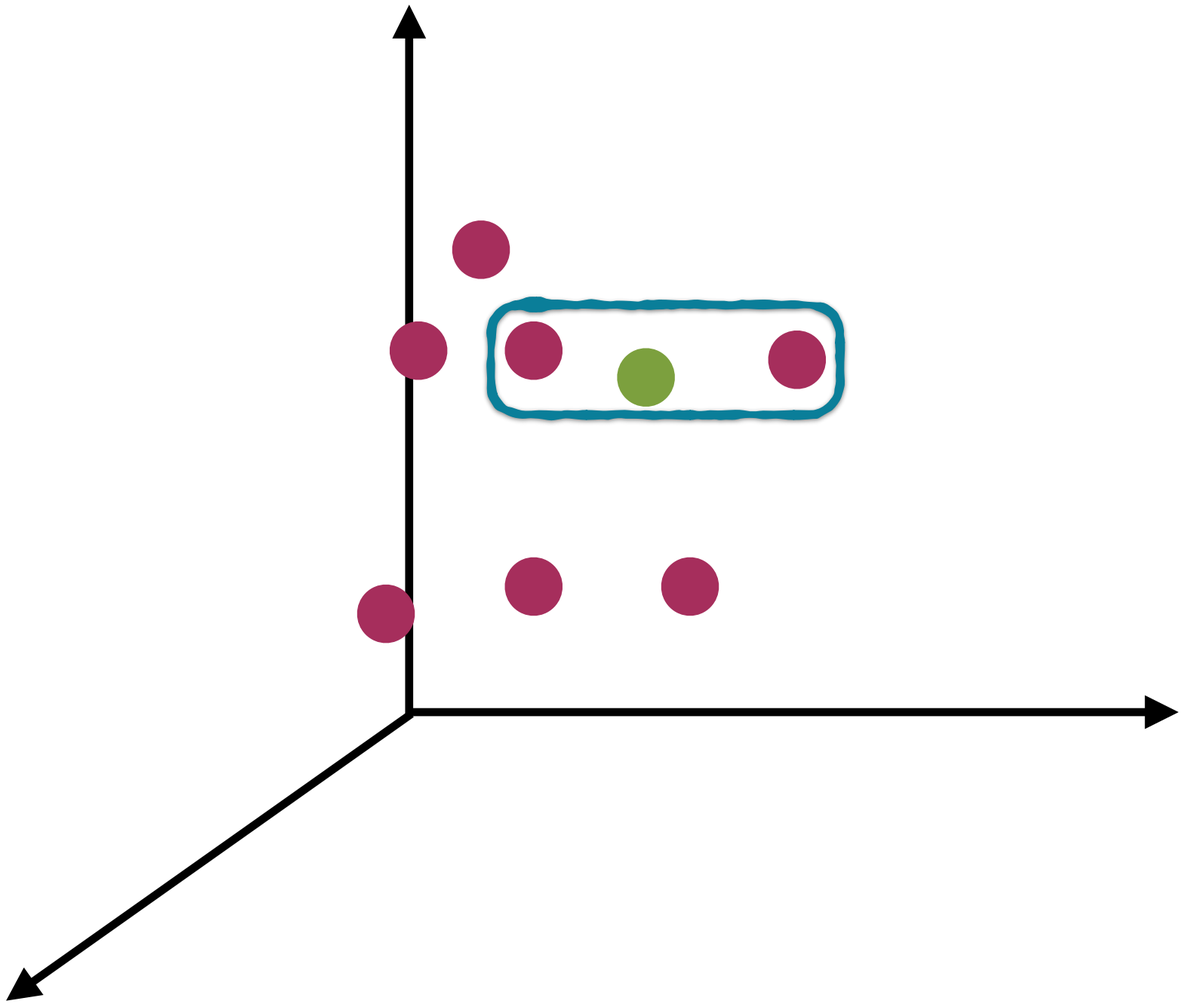
User	F1	F2	F3	F4
A	0	3	2	5

Ex : Average of ratings of products that the user already likes



A Full-Fledged Content Based Filtering System

Find the “nearest” neighbors of the user



Digging Deeper into Collaborative Filtering

Collaborative Filtering

Content based filtering requires a product attribute database

Collaborative filtering uses easily captured user behavior data

User's affinity for some products

Purchases

Pageviews

Clicks

Ratings

All of these are easily captured by the business owner

Purchases

Pageviews

Clicks

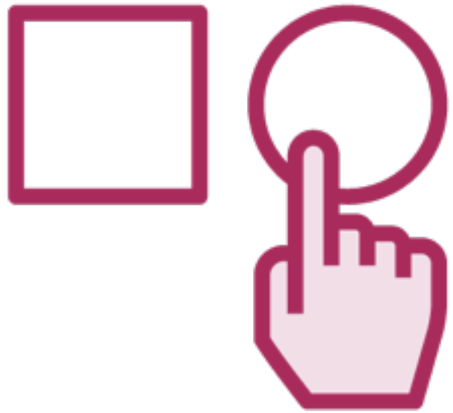
Ratings

Rating

Pick one or a combination of these metrics

Types of Ratings

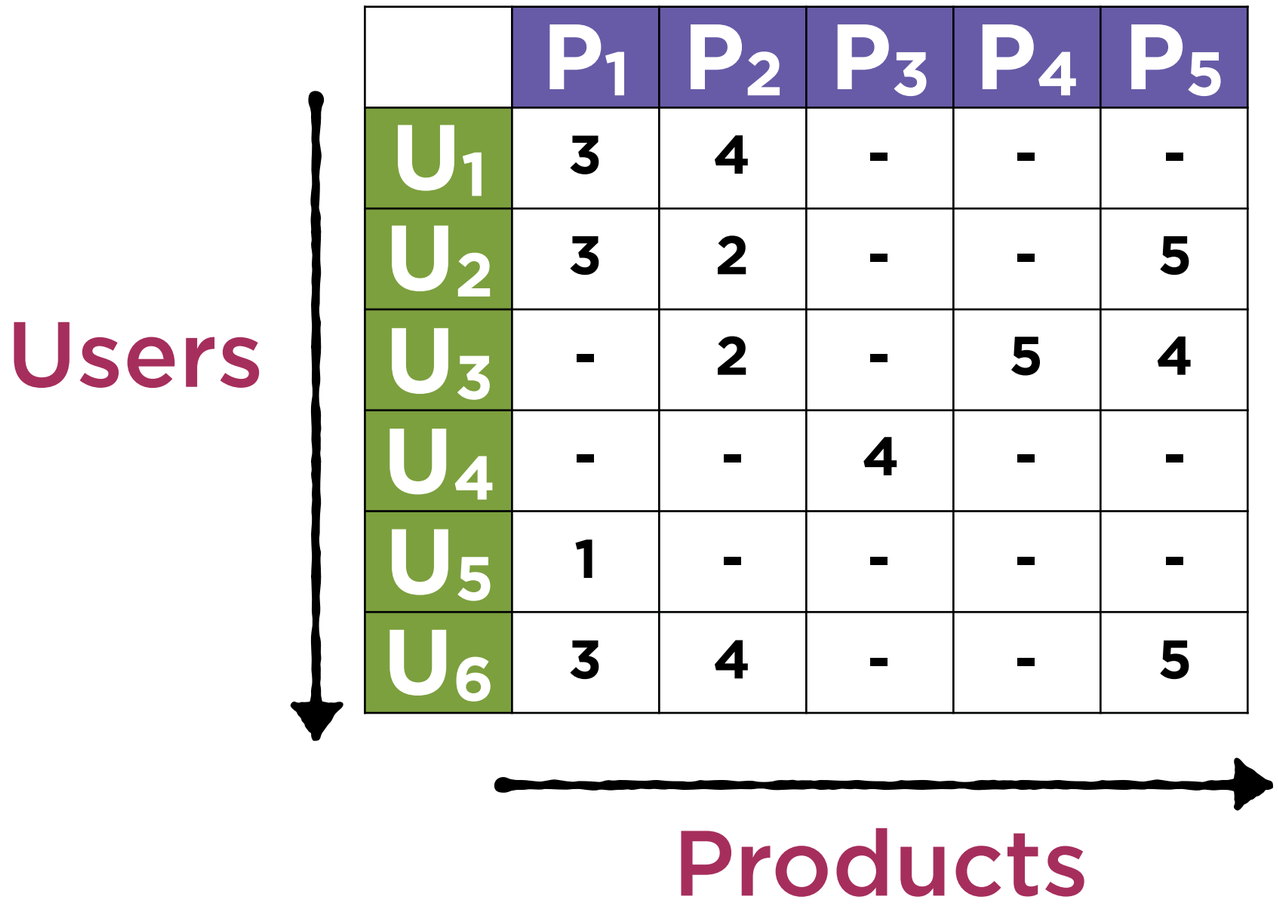
Explicit



Implicit

The rating data is represented using a matrix

Rating Matrix

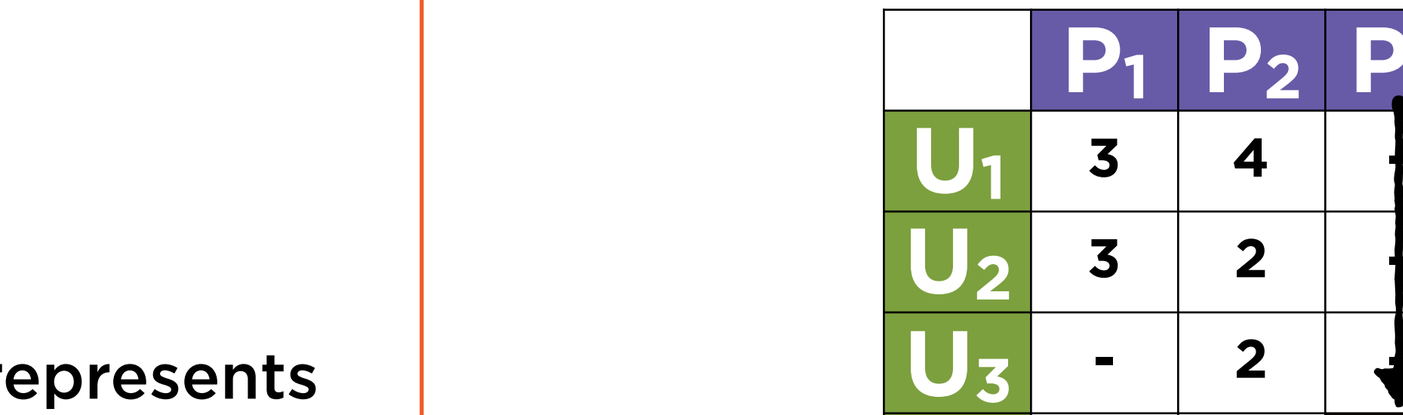


	P_1	P_2	P_3	P_4	P_5
U_1	3	4	-	-	-
U_2	3	2	-	-	5
U_3	-	2	-	5	4
U_4	-	-	4	-	-
U_5	1	-	-	-	-
U_6	3	4	-	-	5

Each cell represents one rating

Rating Matrix

	P_1	P_2	P_3	P_4	P_5
U_1	3	4	-	-	-
U_2	3	2	-	-	5
U_3	-	2	-	5	4
U_4	-	-	4	-	-
U_5	1	-	-	-	-
U_6	3	4	-	-	5



User 4’s affinity for product 3

Blank cells represent the ratings for unseen products

Rating Matrix

	P_1	P_2	P_3	P_4	P_5
U_1	3	4	-	-	-
U_2	3	2	-	-	5
U_3	-	2	-	5	4
U_4	-	-	4	-	-
U_5	1	-	-	-	-
U_6	3	4	-	-	5

Use the filled cells to predict the value of the blank cells

Rating Matrix

	P_1	P_2	P_3	P_4	P_5
U_1	3	4	-	-	-
U_2	3	2	-	-	5
U_3	-	2	-	5	4
U_4	-	-	4	-	-
U_5	1	-	-	-	-
U_6	3	4	-	-	5

Different techniques to help fill the blank cells

Nearest neighbor model

Latent factor analysis

Rating Matrix

	P_1	P_2	P_3	P_4	P_5
U_1	3	4	-	-	-
U_2	3	2	-	-	5
U_3	-	2	-	5	4
U_4	-	-	4	-	-
U_5	1	-	-	-	-
U_6	3	4	-	-	5

Contrasting Different Recommendation Algorithms

Recommendation Algorithms

Content based filtering

Find products with “similar” attributes

Collaborative filtering

Find products liked by “similar” users

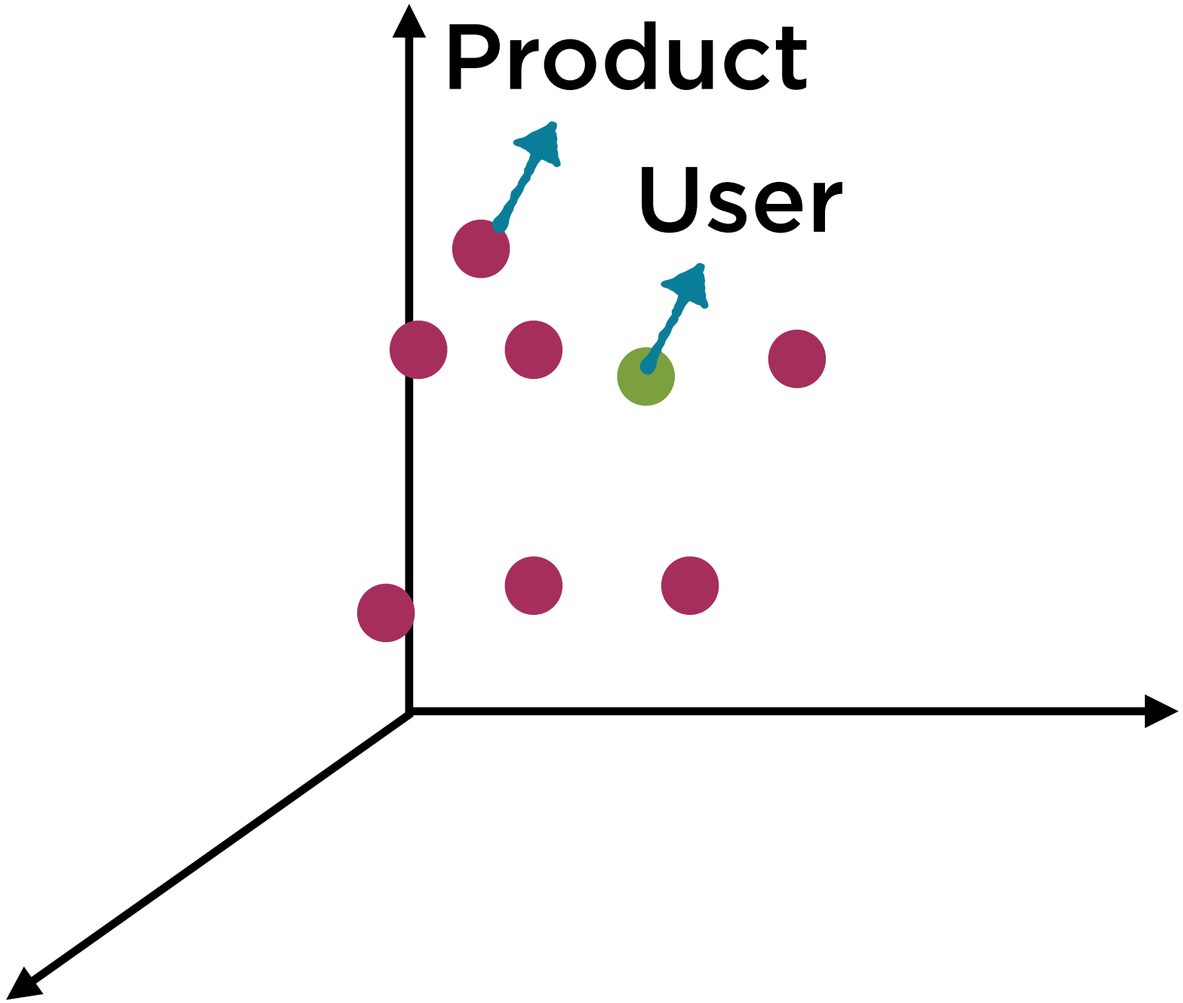
Association rules learning

Find “complementary” products

Data Representation

Content based filtering

Find products with “similar” attributes



Data Representation

Collaborative filtering

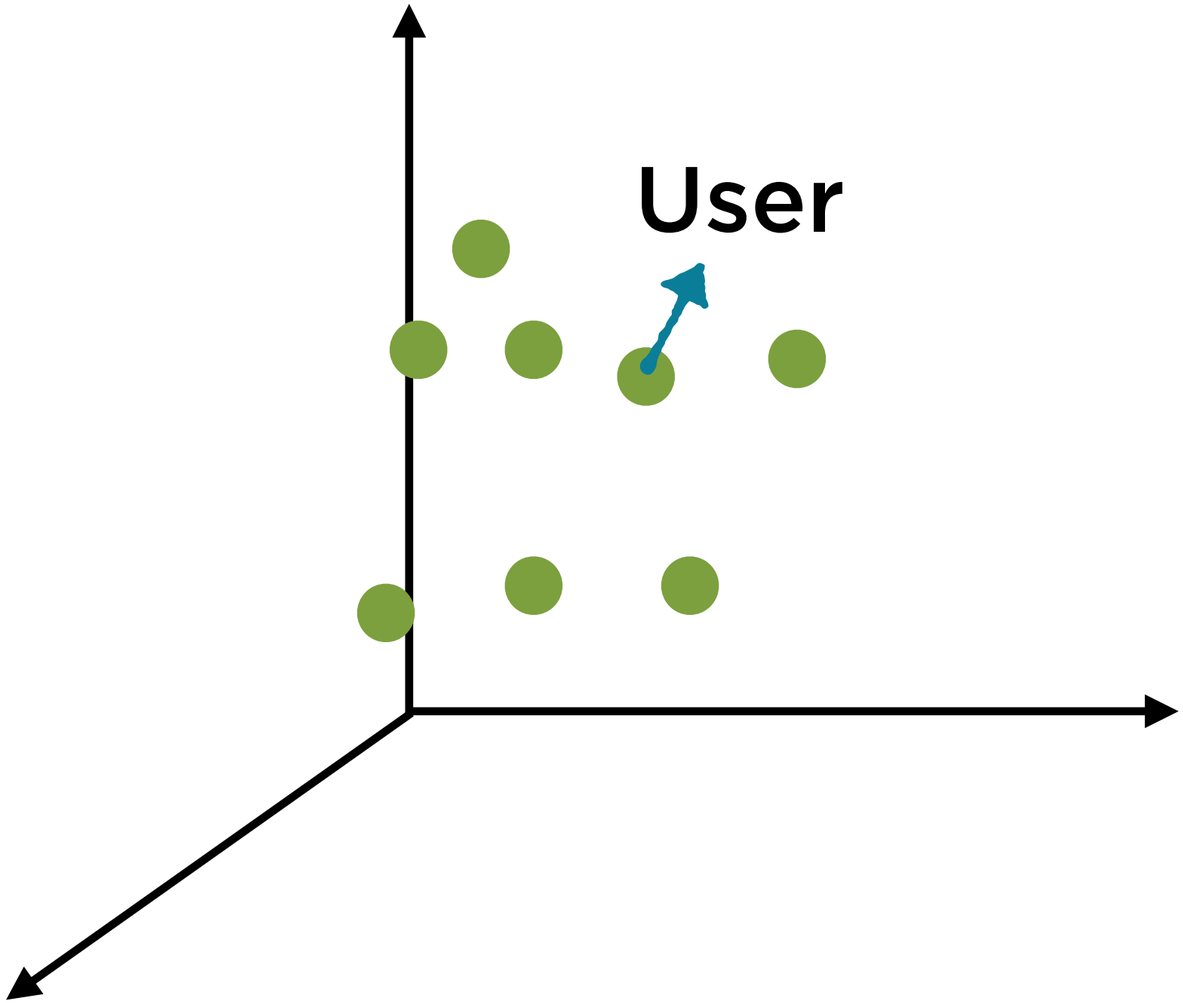
Find products liked by “similar” users

	P_1	P_2	P_3	P_4	P_5
U_1	3	4	-	-	-
U_2	3	2	-	-	5
U_3	-	2	-	5	4
U_4	-	-	4	-	-
U_5	1	-	-	-	-
U_6	3	4	-	-	5

Data Representation

Collaborative filtering

Find products liked by “similar” users



Data Representation

Association rules learning

Find “complementary” products

Conditional Probabilities

Recommendation Algorithms

Content based filtering

Find products with “similar” attributes

Collaborative filtering

Find products liked by “similar” users

Association rules learning

Find “complementary” products

Recommendation Algorithms

Find products associated with another product

Help create offers for buyers of a certain product

Association rules learning

Find “complementary” products

Recommendation Algorithms

Content based filtering

Find products with “similar” attributes

Collaborative filtering

Find products liked by “similar” users

Find products relevant to a user

Help create personalized experiences for a user

Content Based vs Collaborative Filtering

Find products with “similar” attributes

“Similarity” measured directly against products

Find products liked by “similar” users

“Similarity” measured indirectly through other users

Content Based vs Collaborative Filtering

Seems more direct and intuitive

Yet, collaborative filtering is more commonly used

Pre-Requisite for Content Based Filtering

Product	F1	F2	F3	F4
A	0	3	2	5
B	5	2	3	4
C	4	5	2	1
D	3	4	5	2

A database with products rated against relevant attributes

Creating an Attribute Rating Database

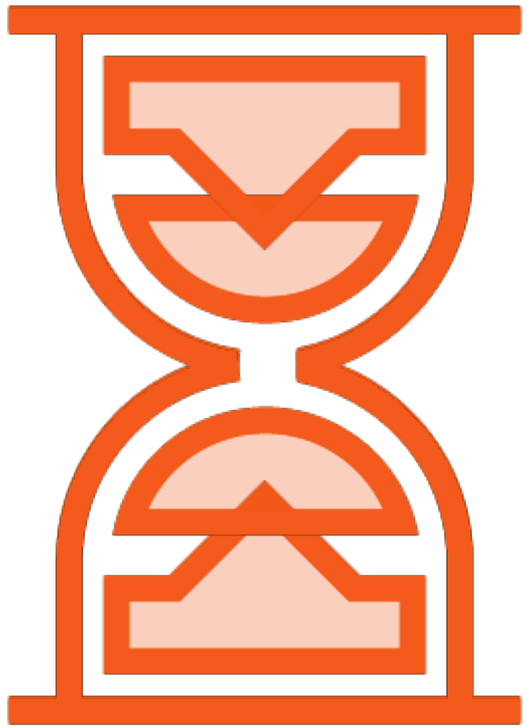
All products

Analyzed by human experts

Product	F1	F2	F3	F4
A	0	3	2	5
B	5	2	3	4
C	4	5	2	1
D	3	4	5	2

Creating an Attribute Rating Database

Analyzed by human experts



The Music Genome Project

Collaborative Filtering

Purely based on user behavior

Agnostic to product attributes

No human intervention required

Content Based vs Collaborative Filtering

Requires a 2 step process

1. Manual process with human intervention

2. An algorithm to extract information

1 step process

Extract information directly from user ratings

Summary

Recommendation algorithms find relationships among users and products

Content based filtering

- Rating database built by experts

Collaborative filtering

- User provided ratings

Association rules learning

- Conditional probabilities