
neue fische
School and Pool for Digital Talent

neue fische

Recommender Systems, an Introduction

By Alexandra Zimmermann-Rösner and Aljoscha Wilhelm

Code of Conduct

We expect all participants to our events and community to abide to this code of conduct:
LadyNerds Code of Conduct (http://bit.ly/LadyNerds-CoC).

We follow the **LadyNerds Code of Conduct** because we are dedicated to providing a safe, inclusive, welcoming, and harassment-free space and experience for all members and guests, regardless of gender identity and expression, sexual orientation, disability, physical appearance, socioeconomic status, body size, ethnicity, nationality, level of experience, age, or religion (or lack thereof).

The Code of Conduct exists because of that dedication. We do not tolerate harassment in any form and we prioritise marginalised people's safety over privileged people's comfort.

Team

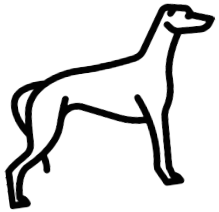
Alexandra Zimmermann-Rösner

- Biologist
- Coach Data Science Bootcamp
- the sea, crochet and my family

Team

Aljoscha Wilhelm

- Physicist
- Coach Data Science Bootcamp
- Chess, Hiking and his dog Ryu



Agenda

About

At the beginning...

Who are we and what is neuefische?

content based model

...shortly afterwards...

Intro and how to build a content based model

collaborative filtering

...roughly second half...

what is collaborative filtering and how do we predict user ratings?

Q&A

...when everyone is enlightened.

Any more questions? Ask us anything.

» neue fische

School and Pool for Digital Talent

- founded 2018 by Dalia Das
- large alumni and partner network

12 Week Bootcamps:

Web Development

Data Science

Data Analytics

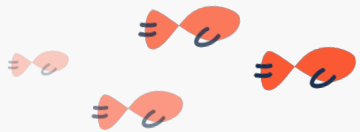
Java Development

Cloud Development

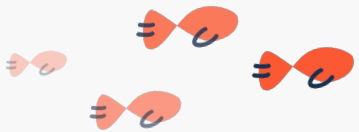
AWS

neuefische - upcoming Data Science Bootcamps

- Data Science Bootcamps:
  - 17.10., 24.10. (part-time)
- for more details of our Bootcamps, checkout our website:
  - neuefische.de
- or get in touch with us:
  - studienberatung@neuefische.de

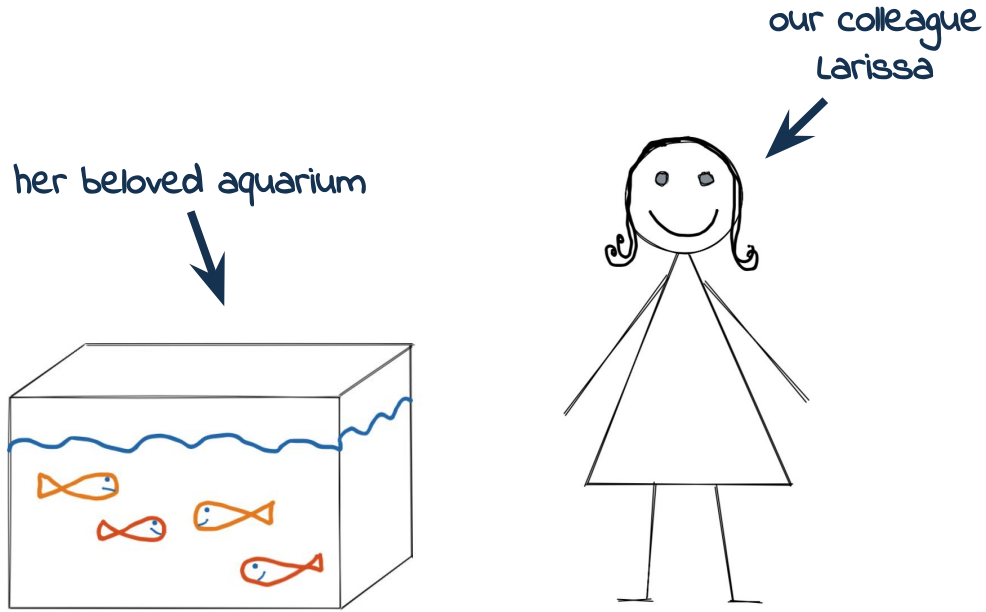


Previously on …



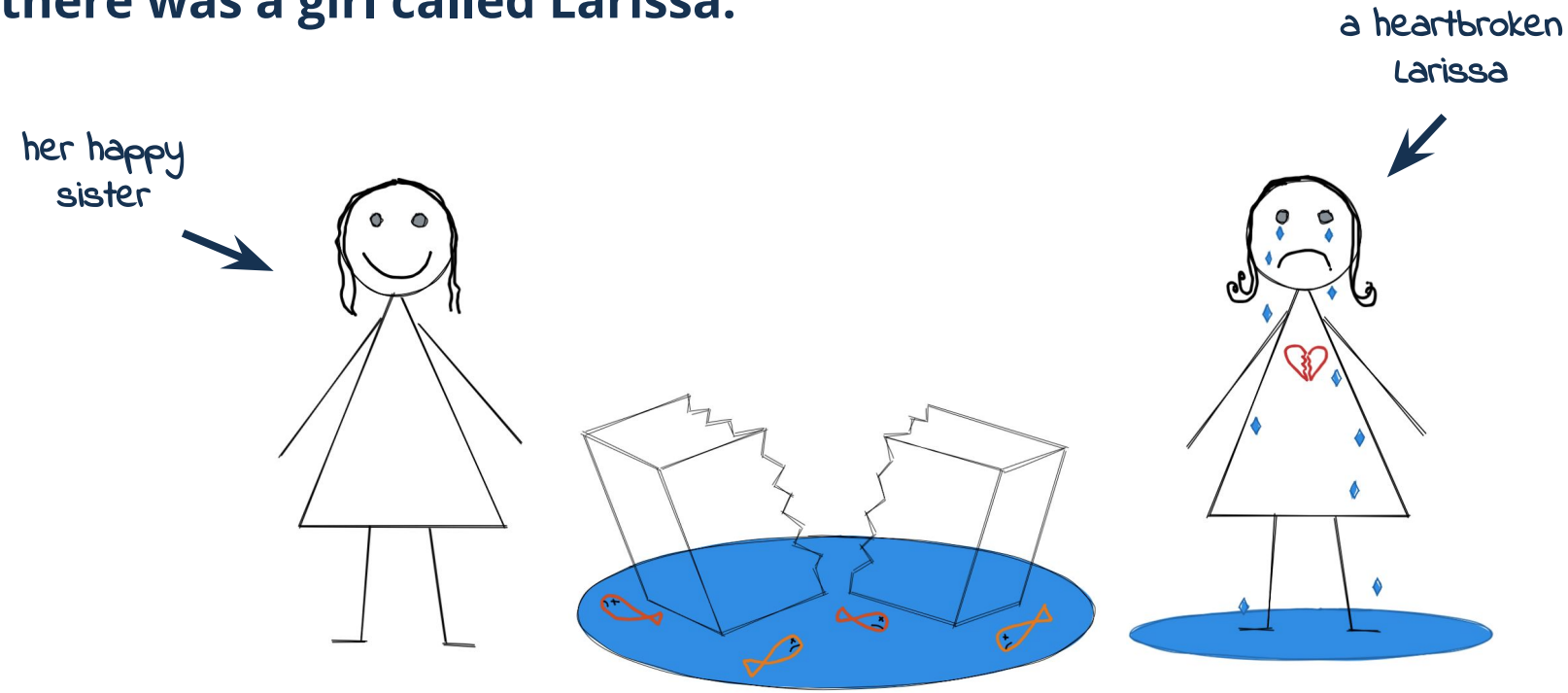
What happened before?

… there was a girl called Larissa.



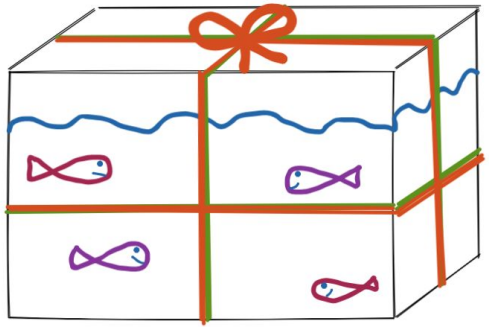
What happened before?

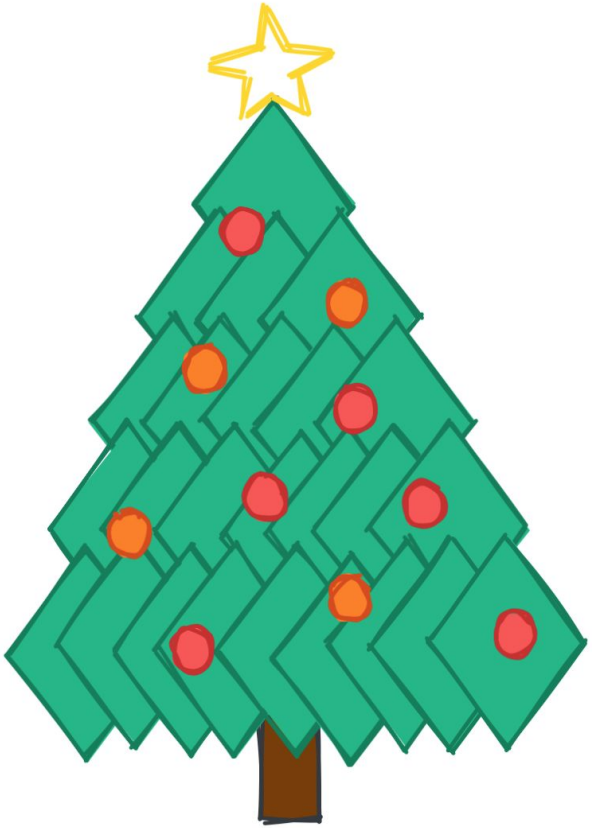
… there was a girl called Larissa.



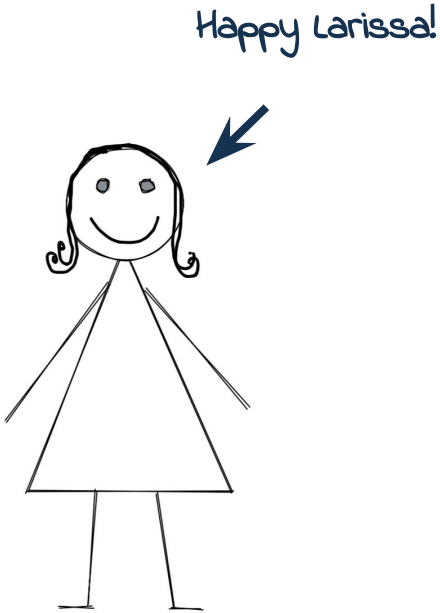
What happened before?

Larissa's
new aquarium



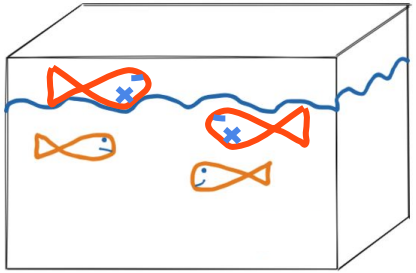


And Larissa was happy again

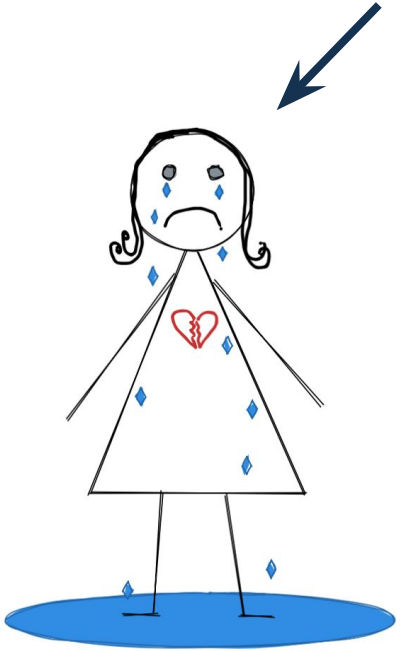


As time passes by ...

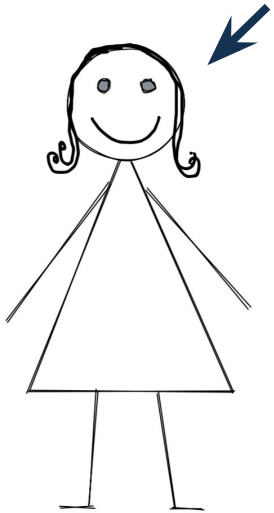
...



our colleague
Larissa

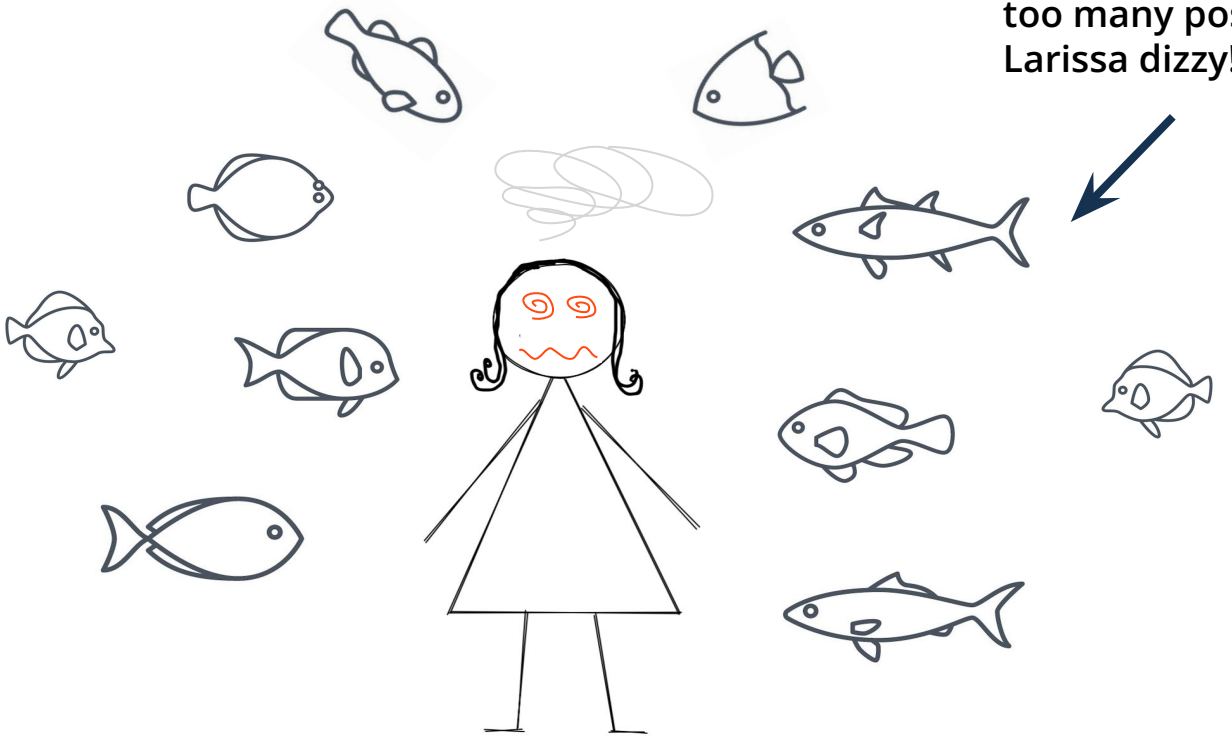


I'll buy new fish!



Similar to my old ones but also a little different

too many possible fishes make Larissa dizzy!



new fish for Larissa

Let’s help her to digest all the information by building a recommender!

What are recommenders?

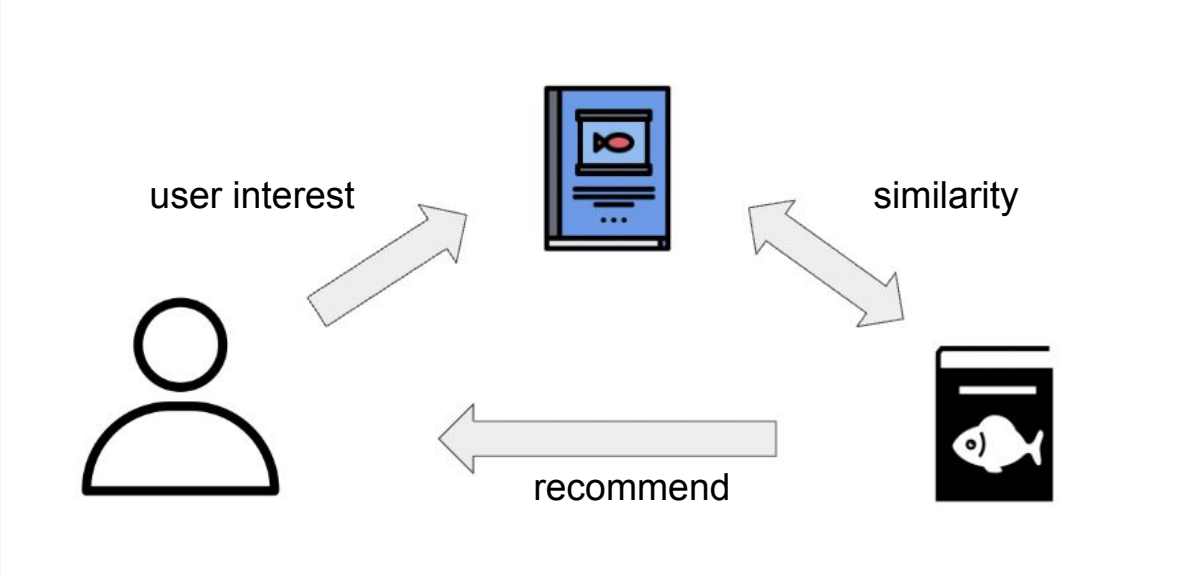
- Algorithms to find similar items and recommend them to user
- Examples: Youtube, Amazon, Netflix and Spotify

Types

- Content based
- Collaborative filtering
- Hybrid models

Content based model

item features only



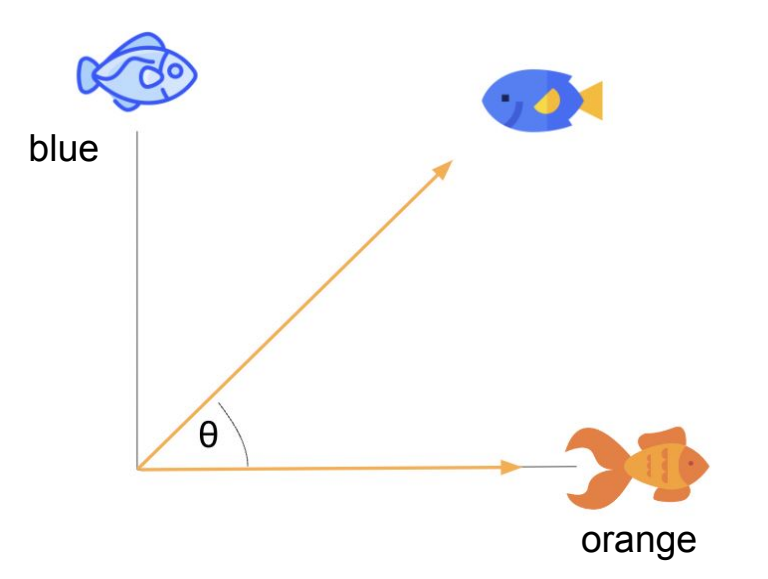
How to find similar items?

Similarity measures:

- euclidean distance
- pearson correlation
- cosine similarity ⇨ most used

Cosine similarity

can take values between 1 and 0



Cosine similarity

Dot product of the vectors divided by the product of their lengths

$$\cos(\theta) = \frac{\vec{x} \cdot \vec{y}}{\|x\| \cdot \|y\|} = \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2}\sqrt{\sum_i y_i^2}}$$

Example

	pink	blue	yellow	green
	1	1	0	1
	1	0	0	0
	0	0	1	0
	1	1	0	0

Calculation example?

vector has the values x = (1,1,0,1)

vector has the values y = (1,0,0,0)

$$\cos(\theta) = \frac{\vec{x} \cdot \vec{y}}{\|x\| \cdot \|y\|} = \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2}\sqrt{\sum_i y_i^2}}$$

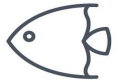
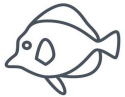
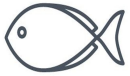
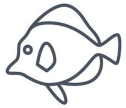
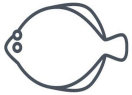
x * y = 1*1 + 1*0 + 0*0 + 1*0 = 1

||x|| = √ 1^2 + 1^2 + 0^2 + 1^2 = 1.732

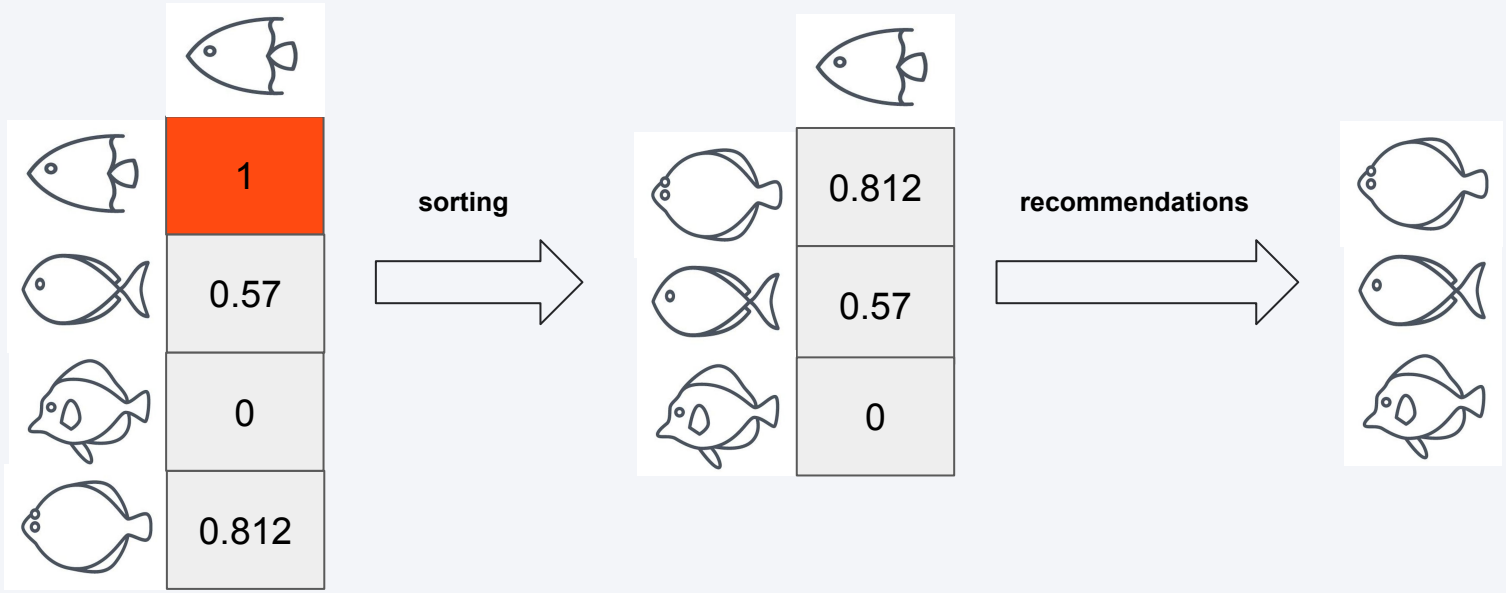
||y|| = √ 1^2 + 0^2 + 0^2 + 0^2 = 1

cos (x,y) = 1/(1.732 * 1) = 0.577

Similarity

				
	1	0.57	0	0.82
	0.57	1	0	0.71
	0	0	1	0
	0.812	0.71	0	1

Outcome



Notebook time

We shared the link with you. Please look into the chat :)

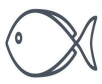
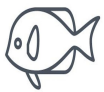
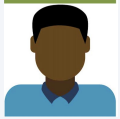
Collaborative filtering

based on user behavior

Types:

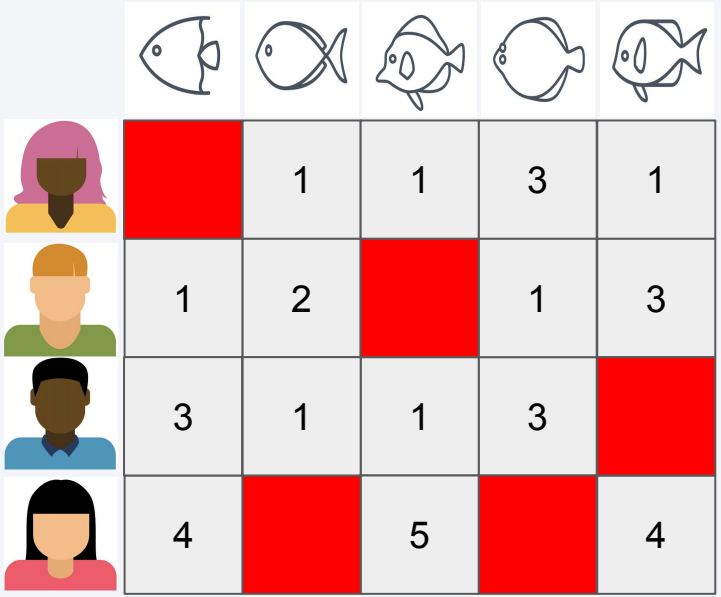
- item based
  - similarities between items based on user ratings
- user based
  - similarities between users based on their ratings

User-Item-Rating Matrix

					
	3	1	1	3	1
	1	2	4	1	3
	3	1	1	3	1
	4	3	5	4	4

User-Item-Rating Matrix

We usually deal with sparse matrices



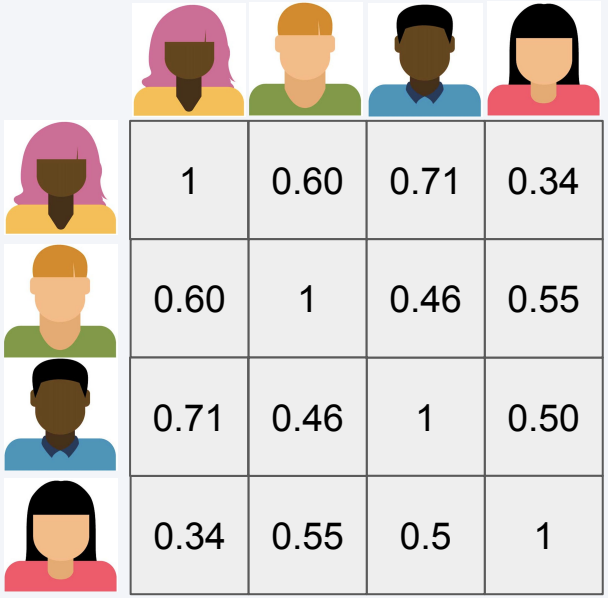
A simple user based Example

We want to use most similar user’s rating to predict rating of other user

- calculate similarity matrix (e.g. Cosine, Pearson, Euclidean)
- use most similar user to predict the rating

User-User-Similarity Matrix

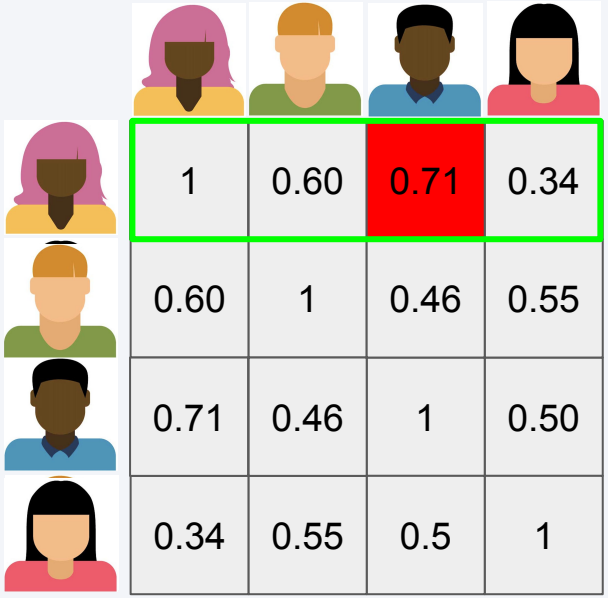
First we calculate the similarity matrix



1	0.60	0.71	0.34
0.60	1	0.46	0.55
0.71	0.46	1	0.50
0.34	0.55	0.5	1

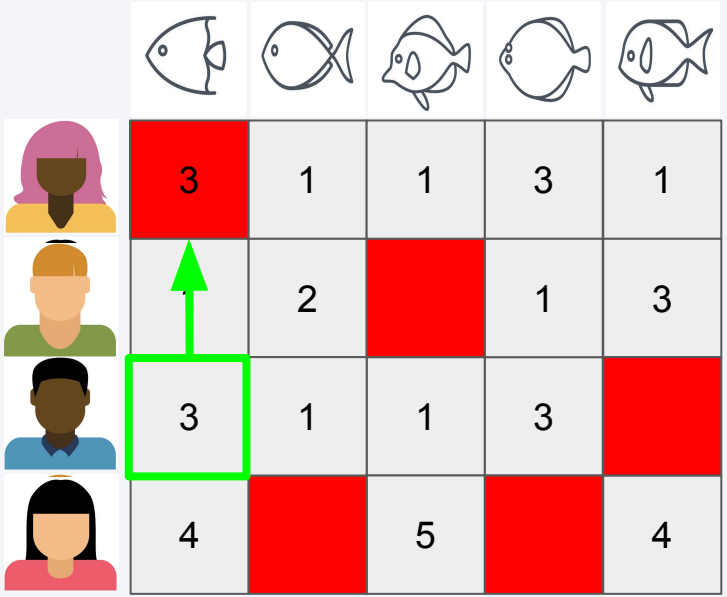
User-User-Similarity Matrix

Then we look for the most similar user



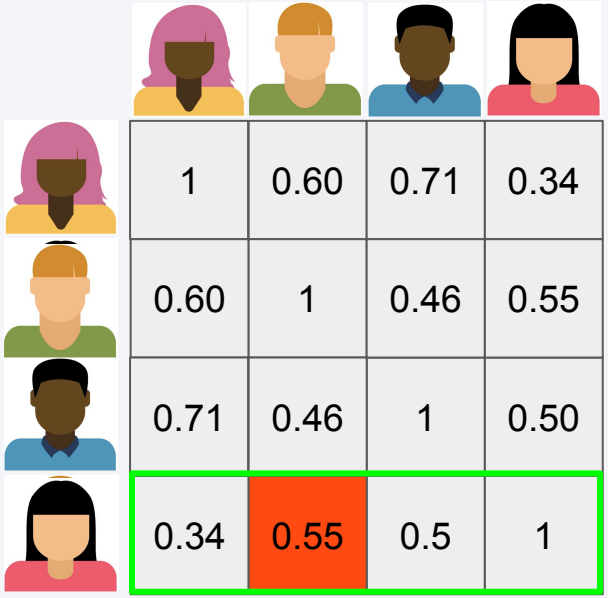
Predict Rating

Then we use the rating of the most similar user to predict the rating



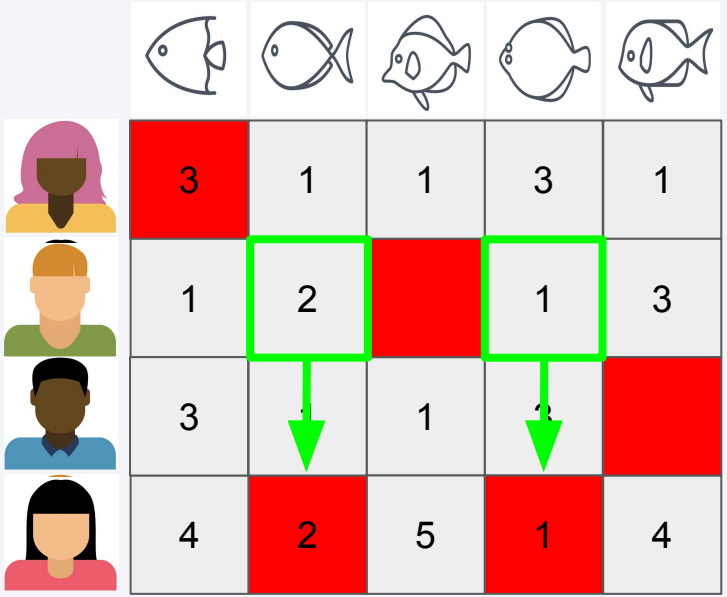
User-User-Similarity Matrix

Again we look for the most similar user



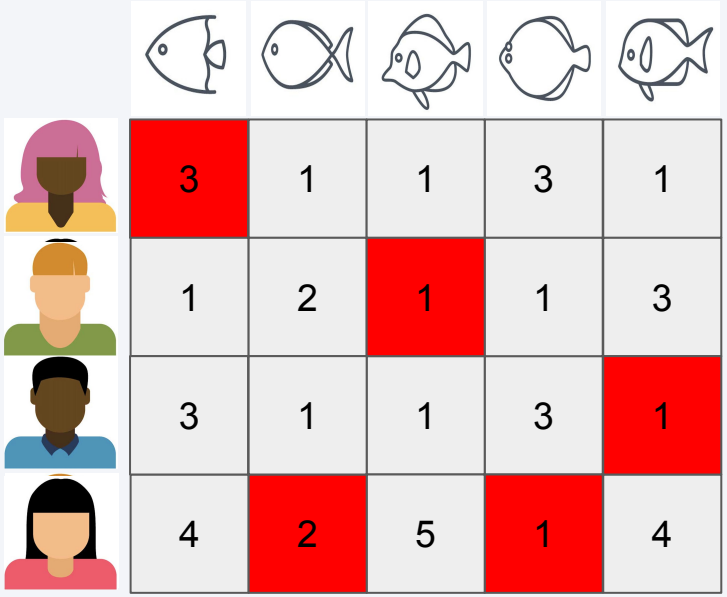
Predict Rating

Then we use the rating of the most similar user to predict the rating



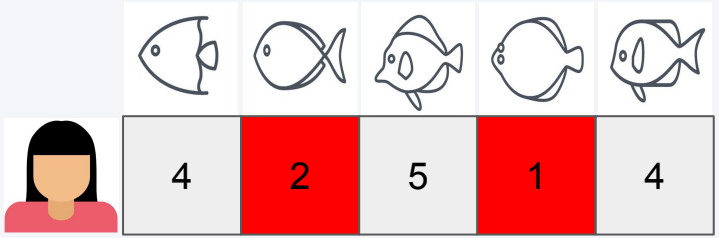
Predict Rating

Do this for all users

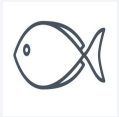


Recommend Item

recommend highest rated item which was not rated before



Dear Larissa,
this fish could be interesting for you!



A not so simple Example

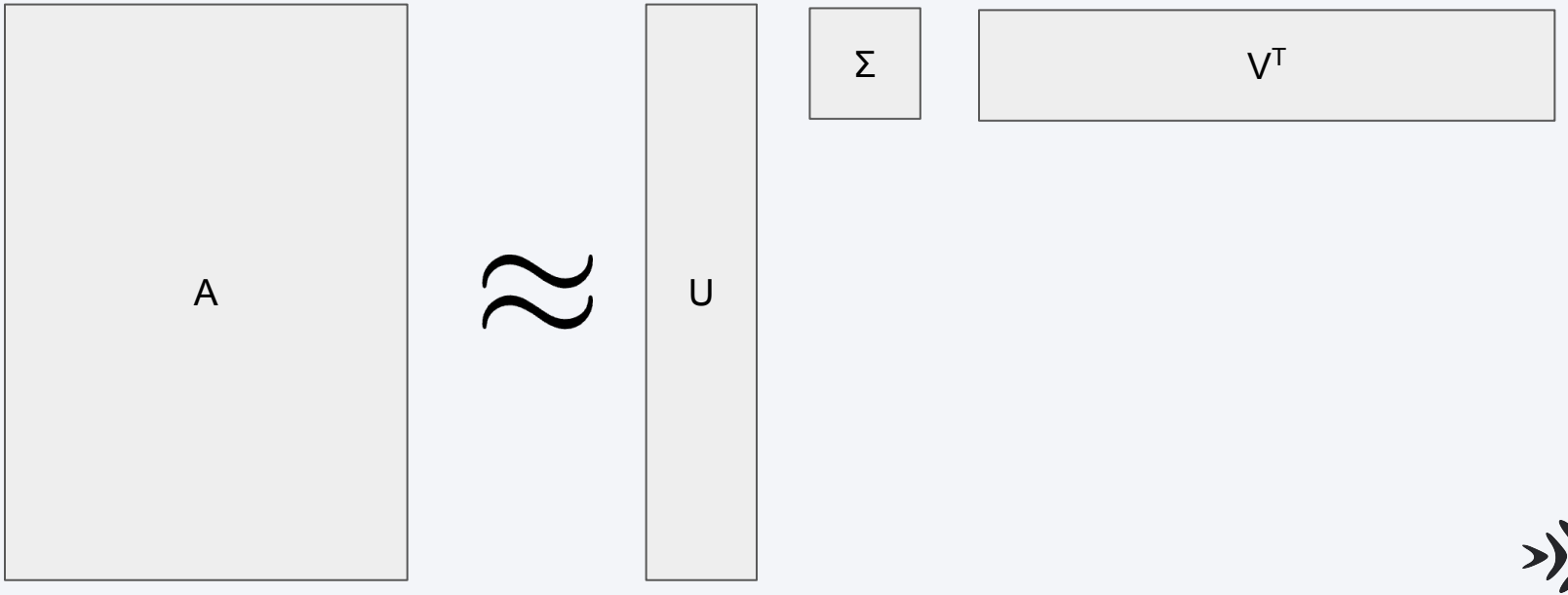
SVD - Singular Value Decomposition

- Approximate Rating Matrix by product of three matrices
- Latent features can often be interpreted (genre etc.)
- Won the Netflix Prize

SVD

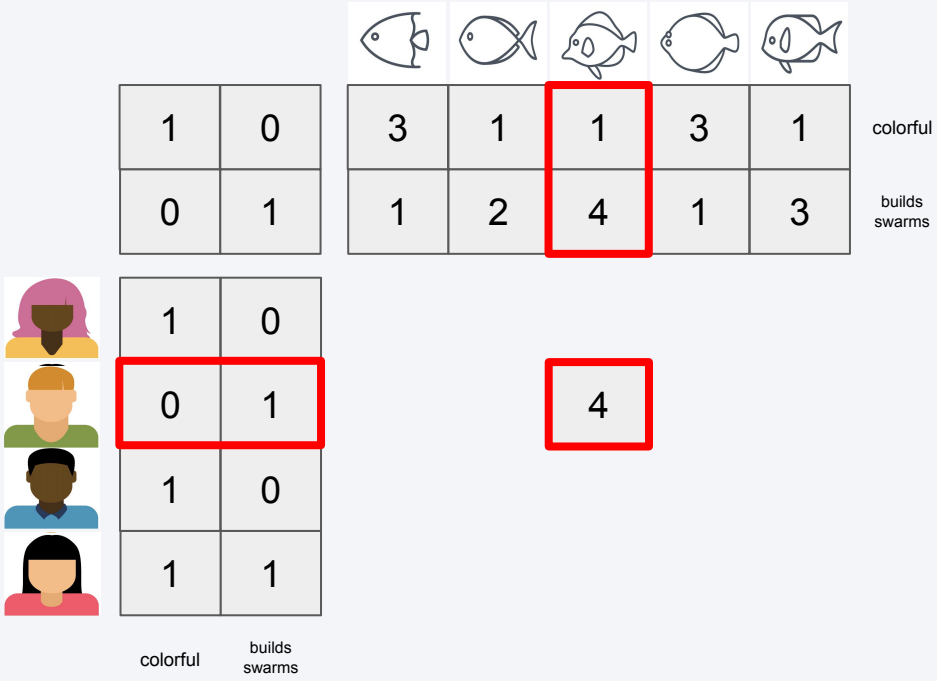
Singular Value Decomposition

$$A \approx U \Sigma V^T$$



SVD

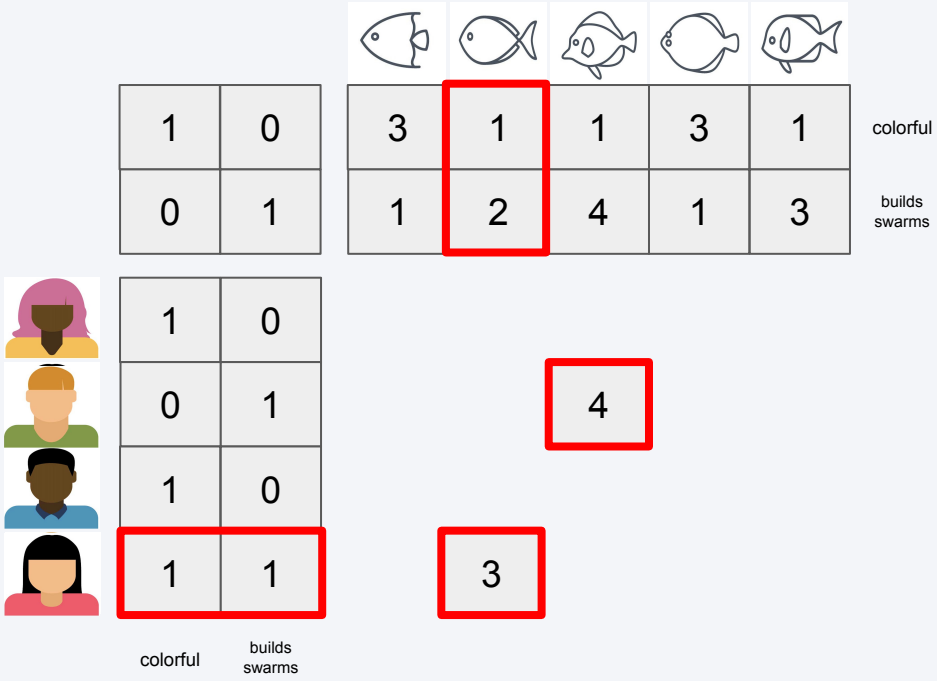
How to predict the missing ratings



$$0x1 + 1x4 = 4$$

SVD

How to predict the missing ratings



$$1x1 + 1x2 = 3$$

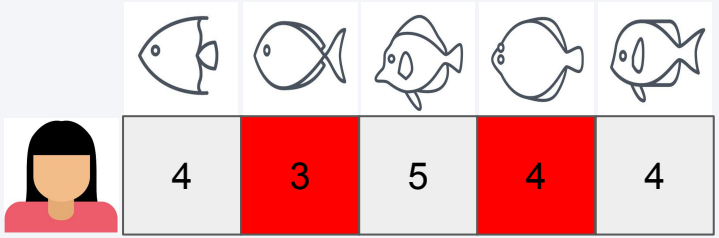
SVD

optimal decomposition can be found using gradient descent

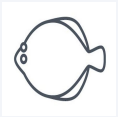
How to predict the missing ratings

Recommend Item

recommend highest rated item which was not rated before



Dear Larissa,
this fish could be interesting for you!



Evaluation Metrics

How to evaluate your model

- Offline metrics
  - regression metrics like MAE, RMSE
  - hit rate
  - diversity
  - novelty
  - churn

- online metrics
  - A/B test
  - feedback

Icons

www.vecteezy.com
www.flaticon.com

Notebook time

We shared the link with you. Please look into the chat :)

Kontakt

Noch Fragen? So könnt ihr uns erreichen:

Online:

studienberatung@neuefische.de
www.neuefische.de

Offline:

neue fische GmbH
Gasstraße 6a
22761 Hamburg

+49 40 2285 9616