Google Cloud

Content-Based Recommendation Engines

Advanced ML with TensorFlow on GCP

End-to-End Lab on Structured Data ML

Production ML Systems

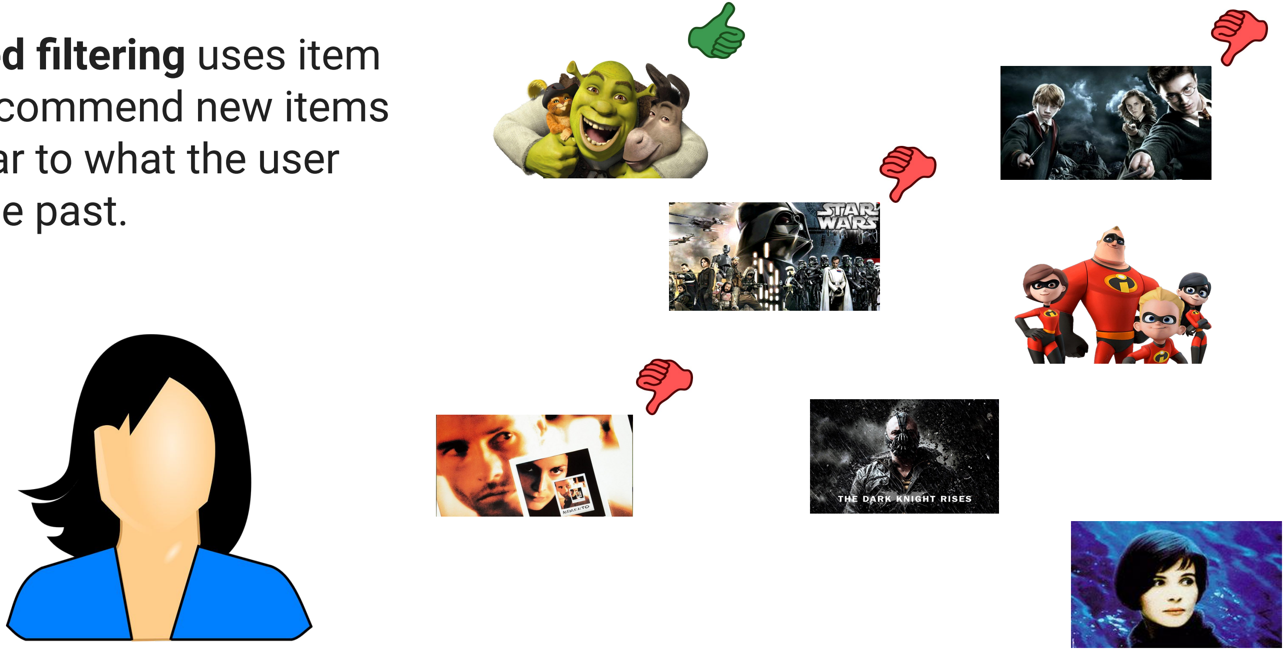
Image Classification Models

Sequence Models

Recommendation Systems

Content-based filtering doesn’t rely on information about other users or other user-item interactions

Content-based filtering uses item features to recommend new items that are similar to what the user has liked in the past.

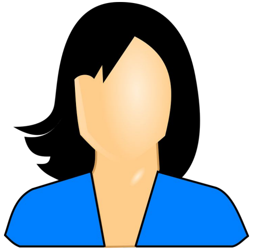


Quiz: Based on the ratings we currently have, which of the remaining, unrated movies should we recommend to this user?

A. The Dark Knight Rises

B. The Incredibles

C. Bleu



Quiz: Based on the ratings we currently have, which of the remaining, unrated movies should we recommend to this user?

A. The Dark Knight Rises

B. The Incredibles

C. Bleu

Learn how to...

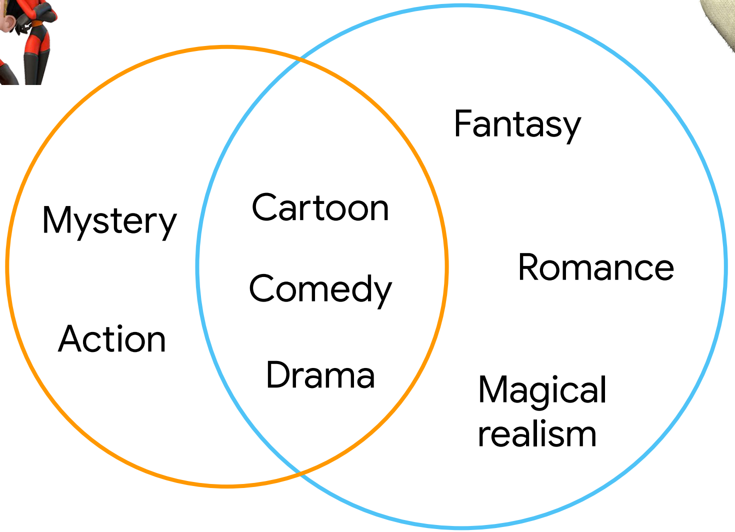
Measure the similarity of elements in an embedding space.

Discuss the mechanics of content-based recommendation systems.

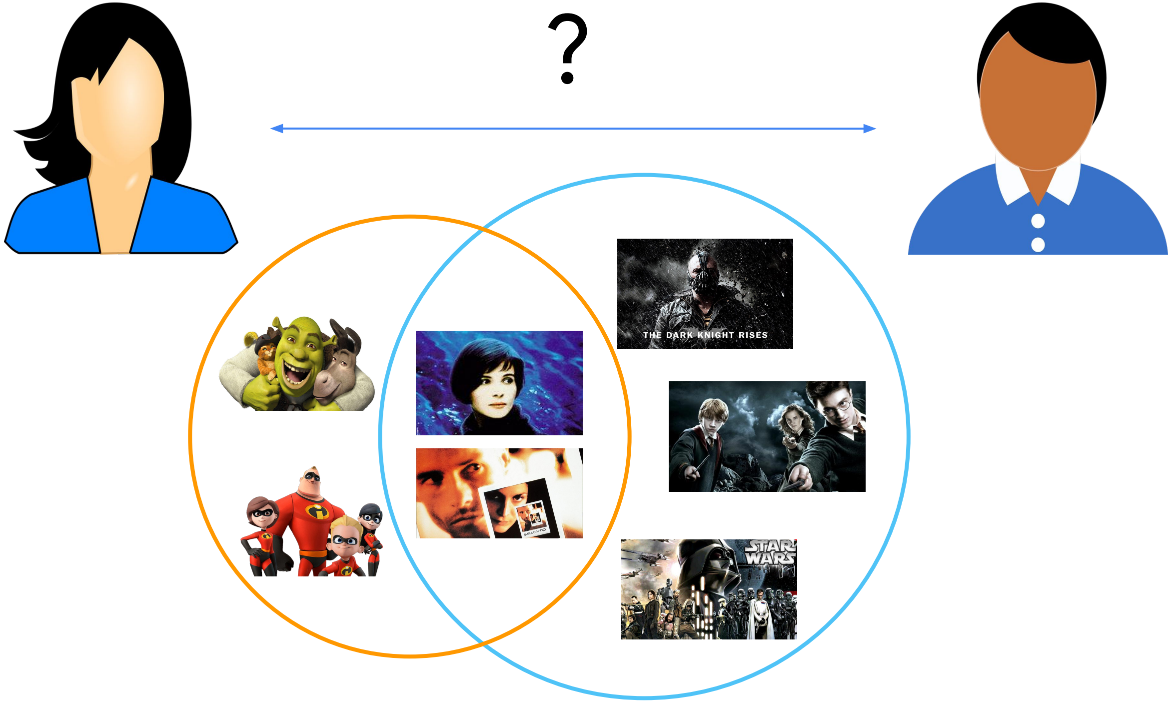
Build a content-based recommendation system.

What it means for two movies to be similar

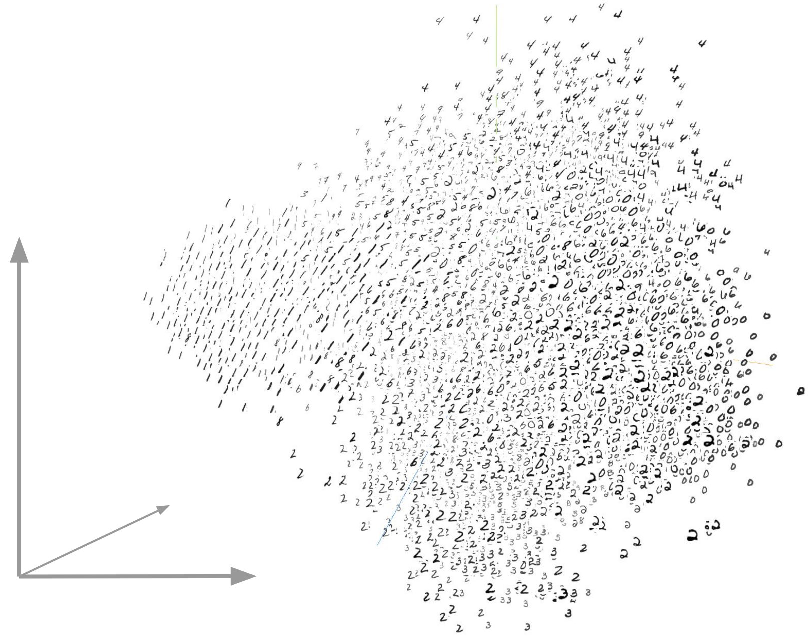
?



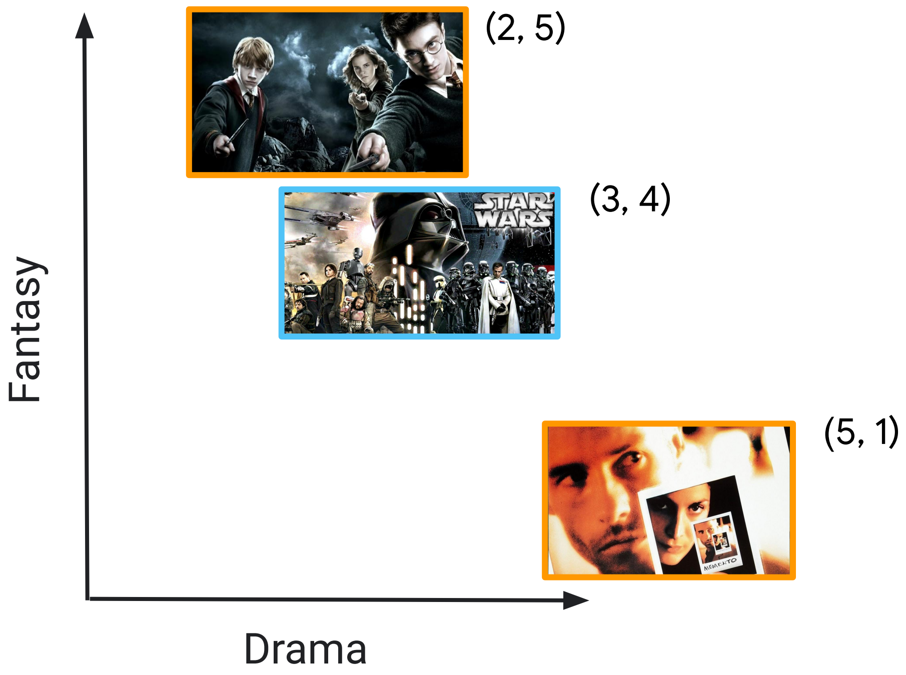
What it means for two users to be similar



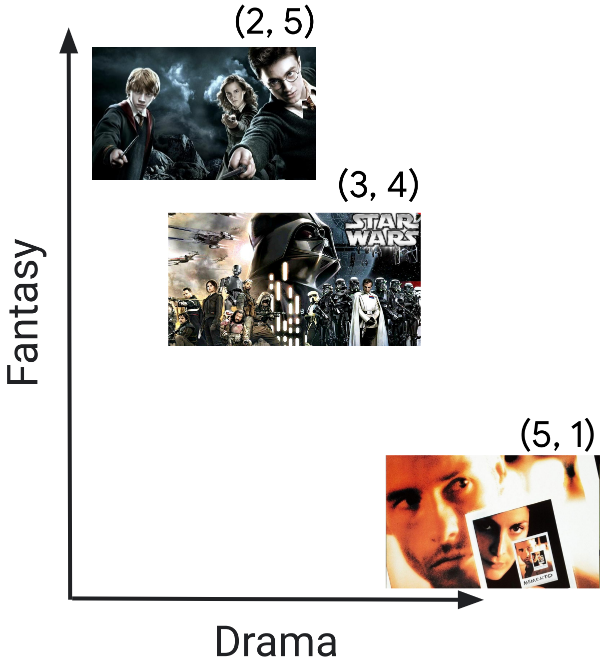
An embedding is a map from our collection of items to some finite dimensional vector space



An embedding is a map from our collection of items to some finite dimensional vector space



A similarity measure is a metric for items in an embedding space

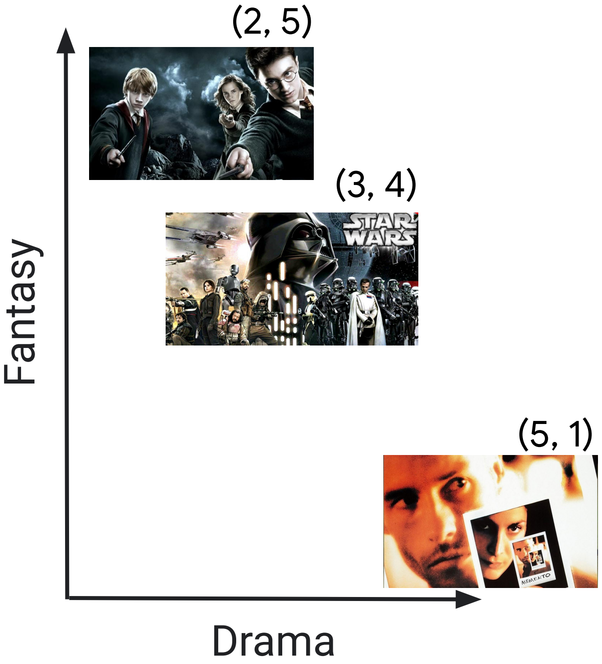


dot product: $s\left(\vec{a}, \vec{b}\right) = \sum_i a_i b_i$

$s\Big($  ,  $\Big) = 2 \cdot 3 + 5 \cdot 4 = 26$

$s\Big($  ,  $\Big) = 3 \cdot 5 + 4 \cdot 1 = 19$

A similarity measure is a metric for items in an embedding space



Cosine similarity: $s\left(\vec{a}, \vec{b}\right) = \frac{\sum_i a_i b_i}{|\vec{a}||\vec{b}|}$

$s\Big($  ,  $\Big) = \frac{2 \cdot 2 + 5 \cdot 4}{\sqrt{29}\sqrt{25}} = 0.97$

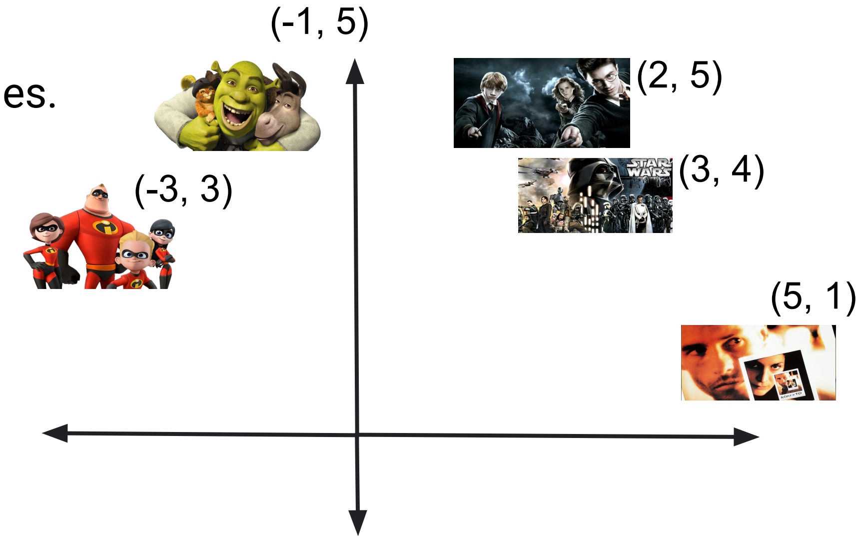
$s\Big($  ,  $\Big) = \frac{3 \cdot 5 + 4 \cdot 1}{\sqrt{25}\sqrt{26}} = 0.75$

Quiz: Compute the cosine similarity between Star Wars and Shrek and between Harry Potter and The Incredibles. Which pair is more similar?

A. Star Wars and Shrek.

B. Harry Potter and The Incredibles.

C. The pairs are equally similar.



Quiz: Compute the cosine similarity between Star Wars and Shrek and between Harry Potter and The Incredibles. Which pair is more similar?

A. Star Wars and Shrek.

B. Harry Potter and The Incredibles.

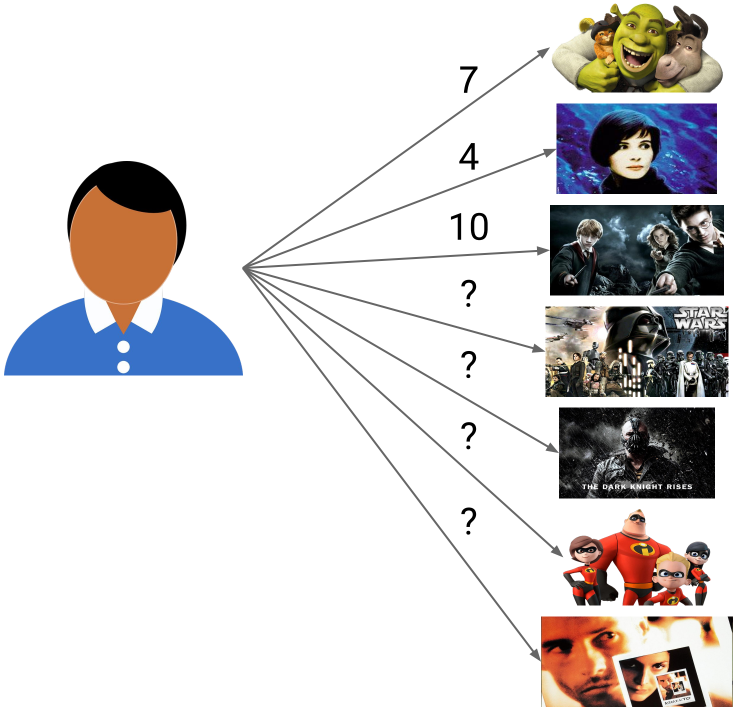
C. The pairs are equally similar.

$s\Bigg($ 

,  $\Bigg) = \dfrac{(3)(-1)+(4)(5)}{\sqrt{10}\sqrt{41}} = 1.78$

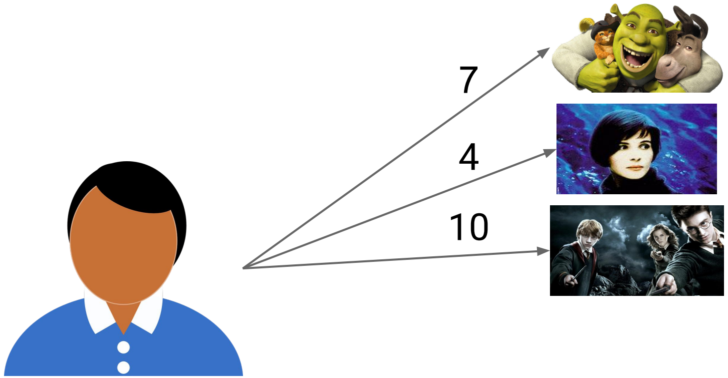
$s\Bigg($  ,  $\Bigg) = \dfrac{(2)(-3)+(5)(3)}{\sqrt{13}\sqrt{34}} = 0.95$

Create a recommendation using content-based filtering



Fantasy	Action	Cartoon	Drama	Comedy
		1		1
			1	
1				1
1	1			
1	1		1	
	1	1		1
			1	

Create a recommendation using content-based filtering



Fantasy	Action	Cartoon	Drama	Comedy
		1		1
			1	
1				1

$$\begin{bmatrix}7\\4\\10\end{bmatrix} \times \begin{bmatrix}0 & 0 & 1 & 0 & 1\\0 & 0 & 0 & 1 & 0\\1 & 0 & 0 & 0 & 1\end{bmatrix} \longrightarrow$$

user feature vector

Create a recommendation using content-based filtering

7
4
10

✖

Fantasy	Action	Cartoon	Drama	Comedy
0	0	1	0	1
0	0	0	1	0
1	0	0	0	1

=

Fantasy	Action	Cartoon	Drama	Comedy
0	0	7	0	7
0	0	0	4	0
10	0	0	0	10

$\downarrow \sum$

Fantasy	Action	Cartoon	Drama	Comedy
10	0	7	4	17

$\downarrow$

Fantasy	Action	Cartoon	Drama	Comedy
0.26	**0**	**0.18**	**0.11**	**0.45**

user feature vector

Create a recommendation using content-based filtering

7
4
10

×

Fantasy	Action	Cartoon	Drama	Comedy
0	0	1	0	1
0	0	0	1	0
1	0	0	0	1

=

Fantasy	Action	Cartoon	Drama	Comedy
0	0	7	0	7
0	0	0	4	0
10	0	0	0	10

$\downarrow \sum$

Fantasy	Action	Cartoon	Drama	Comedy
10	0	7	4	17

$\downarrow$

Fantasy	Action	Cartoon	Drama	Comedy
0.26	**0**	**0.18**	**0.11**	**0.45**

user feature vector

Quiz: Compute the user feature vector for this user based on the ratings below. Which category has the strongest influence for this user?

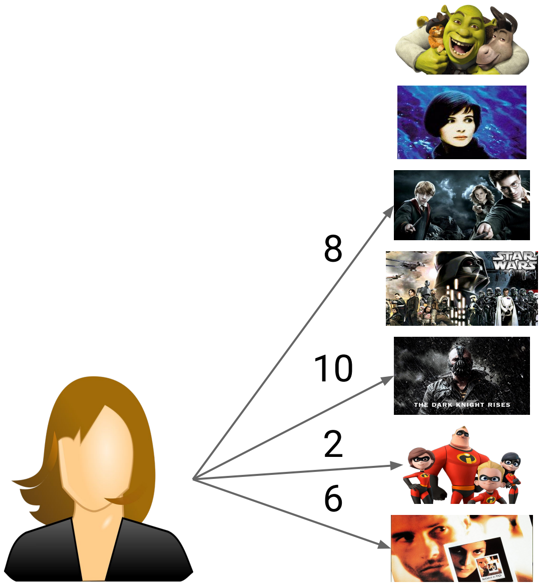
A. Fantasy

B. Action

C. Cartoon

D. Drama

E. Comedy



Fantasy	Action	Cartoon	Drama	Comedy
		1		1
			1	
1				1
1	1			
1	1		1	
	1	1		1
			1	

Quiz: Compute the user feature vector for this user based on the ratings below. Which category has the strongest influence for this user?

A. Fantasy
B. Action
C. Cartoon
D. Drama
E. Comedy

Fantasy	Action	Cartoon	Drama	Comedy
8	0	0	0	10
10	10	0	10	0
0	2	2	0	2
0	0	0	6	0

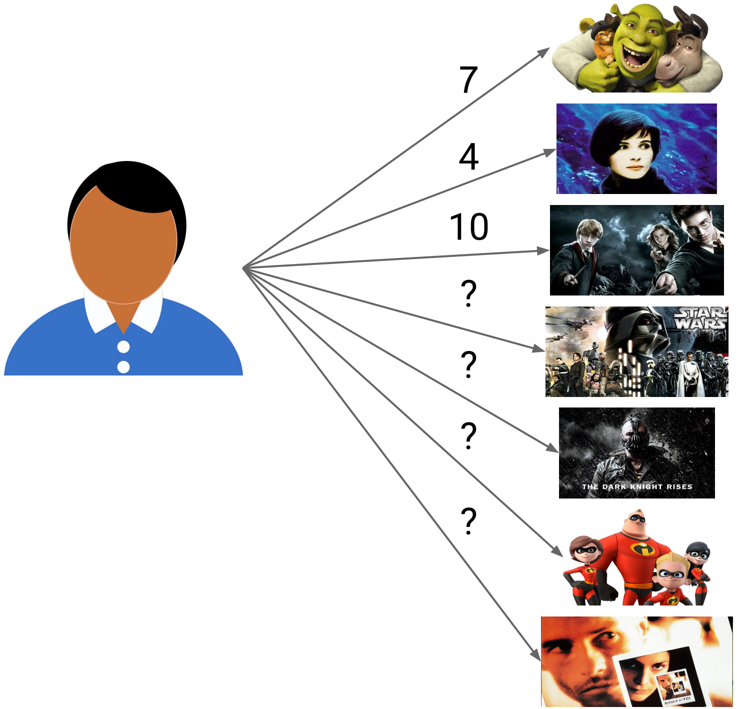
$\downarrow \sum$

Fantasy	Action	Cartoon	Drama	Comedy
18	12	2	16	12

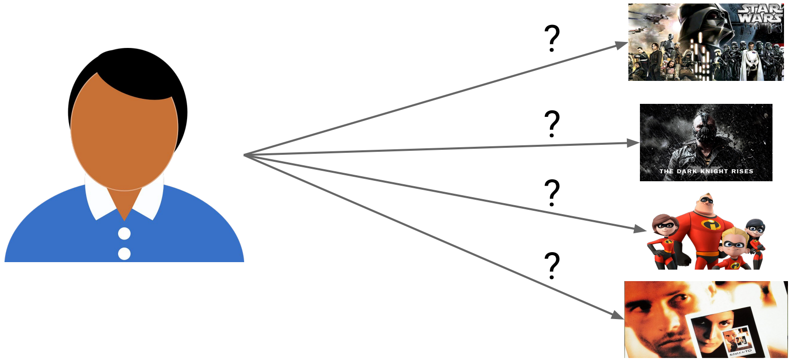
$\downarrow$

Fantasy	Action	Cartoon	Drama	Comedy
0.31	0.21	0.03	0.28	0.17

Make the best recommendation for a user based on their user-feature vector



Make the best recommendation for a user based on their user-feature vector



Fantasy	Action	Cartoon	Drama	Comedy
1	1			
1	1		1	
	1	1		1
			1	

0.26	0	0.18	0.11	0.45

✖

1	1	0	0	0
1	1	0	1	0
0	1	1	0	1
0	0	0	1	0

→ user movie ratings

Make the best recommendation for a user based on their user-feature vector

0.26	0	0.18	0.11	0.45

✖

1	1	0	0	1
1	1	0	1	0
0	1	1	0	1
0	0	0	1	0

=

0.26	0	0	0	0.45
0.26	0	0	0.11	0
0	0	0.18	0	0.45
0	0	0	0.11	0

$\xrightarrow{\sum}$

user movie ratings

0.71
0.37
0.63
0.11

Recommendations

Quiz: In the last quiz, we computed the user-feature vector to be <0.31, 0.21, 0.03, 0.28, 0.17>. Using this, which movie would be our top recommendation for this user?

A. Shrek

B. Bleu

C. Star Wars

	Fantasy	Action	Cartoon	Drama	Comedy
			1		1
				1	
	1	1			

Quiz: In the last quiz, we computed the user-feature vector to be <0.31, 0.21, 0.03, 0.28, 0.17>. Using this, which movie would be our top recommendation for this user?

A. Shrek

B. Bleu

C. Star Wars

	Fantasy	Action	Cartoon	Drama	Comedy
			1		1
				1	
	1	1			

Quiz: Do you know what movie to recommend 2nd?

A. Shrek

B. Bleu

C. Star Wars

	Fantasy	Action	Cartoon	Drama	Comedy
			1		1
				1	
	1	1			

Quiz: Do you know what movie to recommend 2nd?

0.31	0.21	0.03	0.28	0.17

×

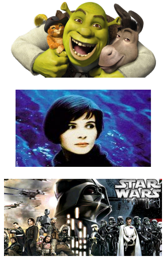
0	0	1	0	1
0	0	0	1	0
1	1	0	0	0

=

0	0	0.03	0	0.17
0	0	0	0.28	0
0.31	0.21	0	0	0

$\xrightarrow{\sum}$

0.20
0.28
0.52



Content-based filtering can be used to generate movie recommendations for multiple users at a time

					
	4	6	8		
			10		8
		6			3
	10	9		5	

user-item rating matrix

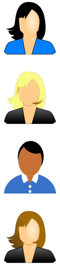
	Action	Sci-Fi	Comedy	Cartoon	Drama
	1	1			1
	1	1			
			1	1	
	1		1	1	
					1

item-feature matrix

Content-based filtering can be used to generate movie recommendations for multiple users at a time

```
# each row represents a user ['Vijay', 'Danielle', 'Ryan', 'Chris']
# each column represents a movie ['Star Wars', 'Dark Knight', 'Shrek',
                                  'The Incredibles', 'Bleu', 'Harry Potter']
users_movies = tf.constant([[4,  6,  8,  0, 0],
                            [0,  0, 10,  0, 8],
                            [0,  6,  0,  0, 3],
                            [10, 9,  0,  5, 0]])

# the columns represent ['Action', 'Sci-Fi', 'Comedy', 'Cartoon', 'Drama']
movies_feats = tf.constant([[1, 1, 0, 0, 1],
                            [1, 1, 0, 0, 0],
                            [0, 0, 1, 1, 0],
                            [1, 0, 1, 1, 0],
                            [0, 0, 0, 0, 1]])
```

Content-based filtering can be used to generate movie recommendations for multiple users at a time



4	6	8		
		10		8
	6			3
10	9		5	

Action	Sci-Fi	Comedy	Cartoon	Drama
1	1			1
1	1			
		1	1	
1		1	1	
				1

10	10	0	0	10
9	9	0	0	0
0	0	0	0	0
5	0	5	5	0
0	0	0	0	0

0	0	4
0	0	0
0	0	0
0	0	0
3	8	0

weighted feature matrix for each user

```
tf.stack(
    values,
    axis=0,
    name='stack'
)
```

Content-based filtering can be used to generate movie recommendations for multiple users at a time

```
...

wgtd_feature_matrices = [tf.expand_dims(tf.transpose(users_movies)[:,i],
                                        axis = 1) *
                         movies_feats for i in range(num_users)]

users_movies_feats = tf.stack(wgtd_feature_matrices, axis = 0)

...
```

Quiz: What is the shape of the tensor resulting from stacking together all of the weighted feature matrices?

A. (# movies, # features)

B. (# users, # movies, # features)

C. (# features, #movies, #users)

D. (# movies, # features, # users)

Quiz: What is the shape of the tensor resulting from stacking together all of the weighted feature matrices?

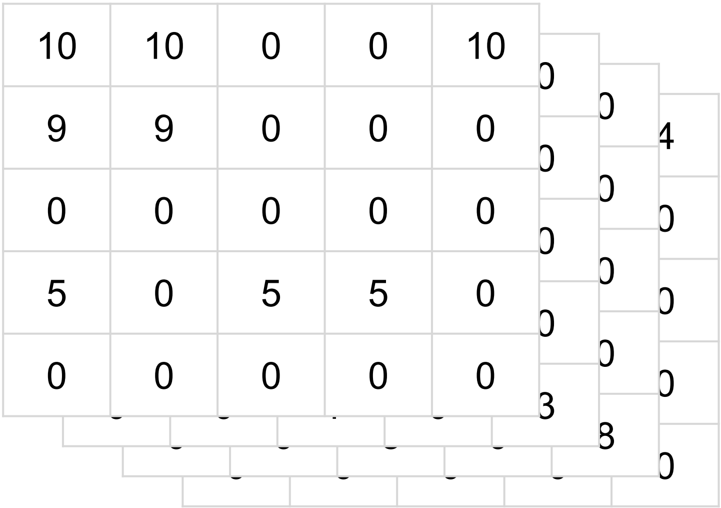
```
tf.stack(wgtd_feature_matrices, axis = 0)
```

A. (# movies, # features)

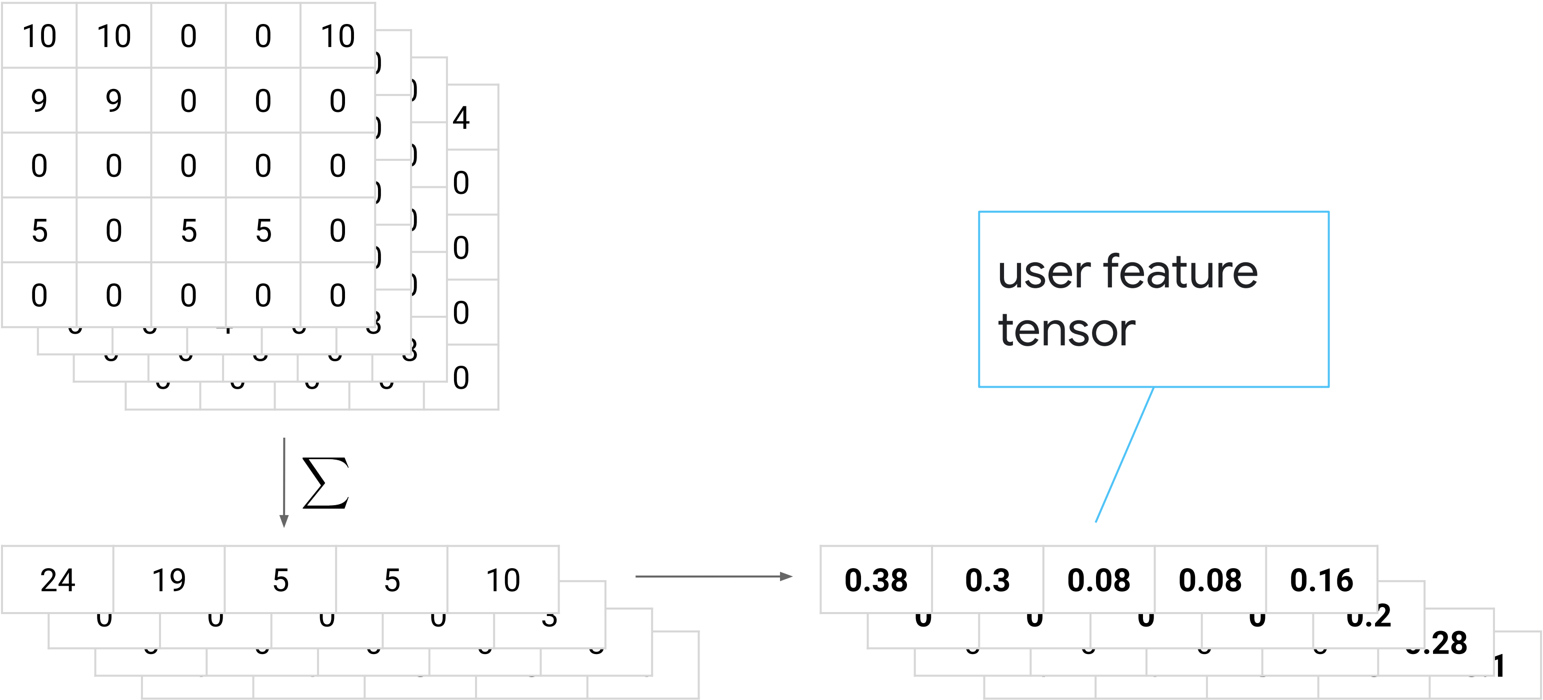
B. (# users, # movies, # features)

C. (# features, #movies, #users)

D. (# movies, # features, # users)



Finding the user feature tensor

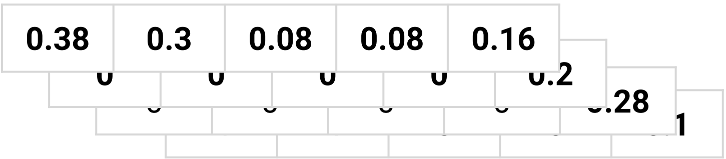


Finding the user feature tensor

```
...

users_movies_feats_sums = tf.reduce_sum(users_movies_feats, axis = 1)
users_movies_feats_totals = tf.reduce_sum(users_movies_feats_sums, axis = 1)

users_feats = tf.stack([users_movies_feats_sums[i,:]/users_movies_feats_totals[i]
                        for i in range(num_users)], axis = 0)


...
```

Finding the user feature tensor

user feature tensor



	Action	Sci-Fi	Comedy	Cartoon	Drama
	0.25	0.25	0.2	0.2	0.1
	0	0	0.36	0.36	0.29
	0.4	0.4	0	0	0.2
	0.38	0.3	0.08	0.08	0.16

Finding the inferred movie rankings for our users

	Action	Sci-Fi	Comedy	Cartoon	Drama
	0.25	0.25	0.2	0.2	0.1
	0	0	0.36	0.36	0.29
	0.4	0.4	0	0	0.2
	0.38	0.3	0.08	0.08	0.16

	Action	Sci-Fi	Comedy	Cartoon	Drama
	1	1	0	0	1
	1	1	0	0	0
	0	0	1	1	0
	1	0	1	1	0
	0	0	0	0	1

					
	0.6	0.5	0.4	0.65	0.1

Finding the inferred movie rankings for our users

	Action	Sci-Fi	Comedy	Cartoon	Drama
	0.25	0.25	0.2	0.2	0.1
	0	0	0.36	0.36	0.29
	0.4	0.4	0	0	0.2
	0.38	0.3	0.08	0.08	0.16

	Action	Sci-Fi	Comedy	Cartoon	Drama
	1	1	0	0	1
	1	1	0	0	0
	0	0	1	1	0
	1	0	1	1	0
	0	0	0	0	1

					
	0.3	0	0.71	0.71	0.29
	1	0.8	0	0.4	0.2
	0.84	0.68	0.16	0.54	0.16

					
	0.6	0.5	0.4	0.65	0.1

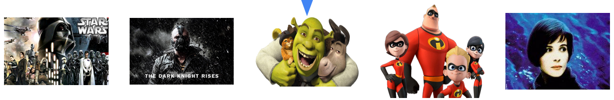
Use the map function to achieve this in TensorFlow

```
...


users_ratings = [tf.map_fn(lambda x: tf.tensordot(users_feats[i], x, axes = 1),
                           tf.cast(movies_feats, tf.float32))
                 for i in range(num_users)]

all_users_ratings = tf.stack(users_ratings)

...
```

Compare the user-movie ranking matrix with the original to see which movies to recommend to which user

0.84	0.68	0.16	0.54	0.16

`tf.stack(...)`



0.6	0.5	0.4	0.65	0.1
0.3	0	0.71	0.71	0.29
1	0.8	0	0.4	0.2
0.84	0.68	0.16	0.54	0.16

4	6	8		
		10		8
	6			3
10	9		5	

Quiz: Which TensorFlow operation could we use?

Which TensorFlow operation could we use to mask the previously rated movies in our user-movie ranking matrix, so we only focus on previously unrated movies when providing recommendations?

A. tf.mask
B. tf.strided_slice
C. tf.tile
D. tf.sparse_slice
E. tf.where

Quiz: Which TensorFlow operation could we use?

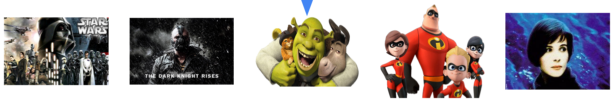
Which TensorFlow operation could we use to mask the previously rated movies in our user-movie ranking matrix, so we only focus on previously unrated movies when providing recommendations?

A. tf.mask
B. tf.strided_slice
C. tf.tile
D. tf.sparse_slice
E. $\boxed{\text{tf.where}}$

What condition would we use to mask out movies that do not have a rating?

0.84	0.68	0.16	0.54	0.16

tf.stack(...)



0.6	0.5	0.4	0.65	0.1
0.3	0	0.71	0.71	0.29
1	0.8	0	0.4	0.2
0.84	0.68	0.16	0.54	0.16

4	6	8		
		10		8
	6			3
10	9		5	

```
tf.where(
    condition,
    x=None,
    y=None,
    name=None
)
```

What condition would we use to mask out movies that do not have a rating?

0.84	0.68	0.16	0.54	0.16

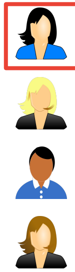
`tf.stack(...)`

-	-	-	0.65	0.1
0.3	0	-	0.71	-
1	-	0	0.4	-
-	-	0.16	-	0.16

4	6	8		
		10		8
	6			3
10	9		5	

```
tf.where(
    condition,
    x=None,
    y=None,
    name=None
)
```

Use the similarity rankings computed to suggest new movies for each user



			0.65	0.1
0.3	0		0.71	
1		0	0.4	
		0.16		0.16

Top Recommendations

Use the similarity rankings computed to suggest new movies for each user

					
				0.65	0.1
	0.3	0		0.71	
	1		0	0.4	
			0.16		0.16

					
	4	6	8		
			10		8
		6			3
	10	9		5	

	Action	Sci-Fi	Comedy	Cartoon	Drama
	1	1			1
	1	1			
			1	1	
	1		1	1	
					1

Lab

Create a content-based recommendation system in TensorFlow

Developing a content-based approach using tools from our supervised machine learning toolbox

$$f\left(\begin{array}{c}\text{user}\\\text{features,}\end{array}\quad\begin{array}{c}\text{movie}\\\text{features}\end{array}\right)\overset{?}{\to}\text{star rating}$$

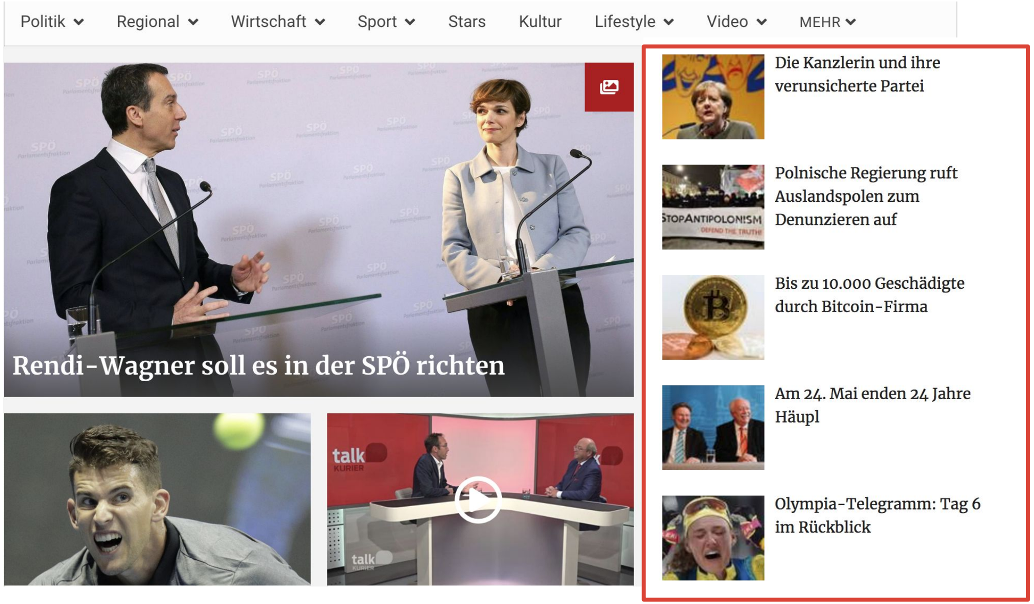
$$f\left(\begin{array}{c}\text{user}\\\text{features,}\end{array}\quad\begin{array}{c}\text{movie}\\\text{features}\end{array}\right)\overset{?}{\to}\text{movie id}$$

age
location
language
gender
demographics
...

genre
duration
director
MPAA rating
awards
...

Building a content-based filtering model to recommend articles to readers

kurier.at



Building a content-based filtering model to recommend articles to readers

kurier.at

label

visitor_id	content_id	category	title	author	months_since_epoch	next_content_id
1000196974485173657	299925700	Lifestyle	Nach Tod von Vater: Tochter bekommt jedes Jahr...	Marlene Patsalidis	574	299972194
1000196974485173657	299972194	News	LIVE: Spielstand bei Sturm - Admira	Mathias Kainz	574	299816215
1000196974485173657	299972194	News	Dominante Grazer nehmen Admira auseinander	Mathias Kainz	574	299410466
1007505561418545529	299407839	Stars & Kultur	Trump: 165 Millionen Dollar für 5 Tage Urlaub	Elisabeth Spitzer	574	299816215
1017855659516706306	298846345	Stars & Kultur	Meghan Markle: Lottogewinn veränderte ihr Leben	Elisabeth Spitzer	574	299814775

features

Building a content-based filtering model to recommend articles to readers

kurier.at

visitor_id	content_id	category	title	author	months_since_epoch	next_content_id
1000196974485173657	299925700	Lifestyle	Nach Tod von Vater: Tochter bekommt jedes Jahr...	Marlene Patsalidis	574	299972194
1000196974485173657	299972194	News	LIVE: Spielstand bei Sturm - Admira	Mathias Kainz	574	299816215
1000196974485173657	299972194	News	Dominante Grazer nehmen Admira auseinander	Mathias Kainz	574	299410466
1007505561418545529	299407839	Stars & Kultur	Trump: 165 Millionen Dollar für 5 Tage Urlaub	Elisabeth Spitzer	574	299816215
1017855659516706306	298846345	Stars & Kultur	Meghan Markle: Lottogewinn veränderte ihr Leben	Elisabeth Spitzer	574	299814775

```
content_id_column = tf.feature_column.categorical_column_with_hash_bucket(
                            key="content_id",
                            hash_bucket_size= len(content_ids_list))
embedded_content_column = tf.feature_column.embedding_column(
                           categorical_column=content_id_column,
                            dimension=10)
```

Building a content-based filtering model to recommend articles to readers

kurier.at

visitor_id	content_id	category	title	author	months_since_epoch	next_content_id
1000196974485173657	299925700	Lifestyle	Nach Tod von Vater: Tochter bekommt jedes Jahr...	Marlene Patsalidis	574	299972194
1000196974485173657	299972194	News	LIVE: Spielstand bei Sturm - Admira	Mathias Kainz	574	299816215
1000196974485173657	299972194	News	Dominante Grazer nehmen Admira auseinander	Mathias Kainz	574	299410466
1007505561418545529	299407839	Stars & Kultur	Trump: 165 Millionen Dollar für 5 Tage Urlaub	Elisabeth Spitzer	574	299816215
1017855659516706306	298846345	Stars & Kultur	Meghan Markle: Lottogewinn veränderte ihr Leben	Elisabeth Spitzer	574	299814775

```
category_column_categorical =
           tf.feature_column.categorical_column_with_vocabulary_list(
                key="category",
                vocabulary_list=categories_list,
                num_oov_buckets=1)
category_column = tf.feature_column.indicator_column(category_column_categorical)
```

Building a content-based filtering model to recommend articles to readers

kurier.at

visitor_id	content_id	category	title	author	months_since_epoch	next_content_id
1000196974485173657	299925700	Lifestyle	Nach Tod von Vater: Tochter bekommt jedes Jahr...	Marlene Patsalidis	574	299972194
1000196974485173657	299972194	News	LIVE: Spielstand bei Sturm - Admira	Mathias Kainz	574	299816215
1000196974485173657	299972194	News	Dominante Grazer nehmen Admira auseinander	Mathias Kainz	574	299410466
1007505561418545529	299407839	Stars & Kultur	Trump: 165 Millionen Dollar für 5 Tage Urlaub	Elisabeth Spitzer	574	299816215
1017855659516706306	298846345	Stars & Kultur	Meghan Markle: Lottogewinn veränderte ihr Leben	Elisabeth Spitzer	574	299814775

```
embedded_title_column = hub.text_embedding_column(
                              key="title",
                              module_spec="https://tfhub.dev/google/nnlm-de-dim50/1",
                              trainable=False)
```

Building a content-based filtering model to recommend articles to readers

kurier.at

visitor_id	content_id	category	title	author	months_since_epoch	next_content_id
1000196974485173657	299925700	Lifestyle	Nach Tod von Vater: Tochter bekommt jedes Jahr...	Marlene Patsalidis	574	299972194
1000196974485173657	299972194	News	LIVE: Spielstand bei Sturm - Admira	Mathias Kainz	574	299816215
1000196974485173657	299972194	News	Dominante Grazer nehmen Admira auseinander	Mathias Kainz	574	299410466
1007505561418545529	299407839	Stars & Kultur	Trump: 165 Millionen Dollar für 5 Tage Urlaub	Elisabeth Spitzer	574	299816215
1017855659516706306	298846345	Stars & Kultur	Meghan Markle: Lottogewinn veränderte ihr Leben	Elisabeth Spitzer	574	299814775

```
author_column = tf.feature_column.categorical_column_with_hash_bucket(
                      key="author",
                      hash_bucket_size=len(authors_list) + 1)
embedded_author_column = tf.feature_column.embedding_column(
                                 categorical_column=author_column,
                                 dimension=3)
```

Building a content-based filtering model to recommend articles to readers

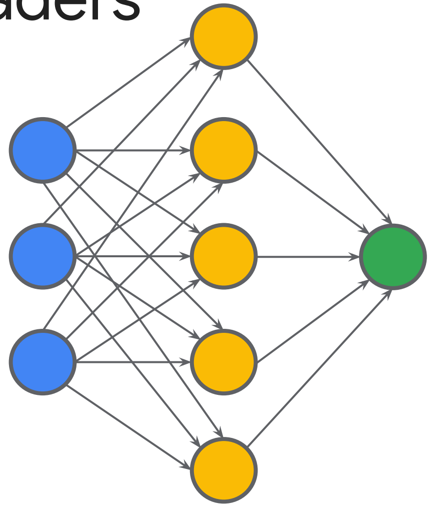
visitor_id	content_id	category	title	author	months_since_epoch	next_content_id
1000196974485173657	299925700	Lifestyle	Nach Tod von Vater: Tochter bekommt jedes Jahr...	Marlene Patsalidis	574	299972194
1000196974485173657	299972194	News	LIVE: Spielstand bei Sturm - Admira	Mathias Kainz	574	299816215
1000196974485173657	299972194	News	Dominante Grazer nehmen Admira auseinander	Mathias Kainz	574	299410466
1007505561418545529	299407839	Stars & Kultur	Trump: 165 Millionen Dollar für 5 Tage Urlaub	Elisabeth Spitzer	574	299816215
1017855659516706306	298846345	Stars & Kultur	Meghan Markle: Lottogewinn veränderte ihr Leben	Elisabeth Spitzer	574	299814775

```
months_since_epoch_column = tf.feature_column.numeric_column(
                                key="months_since_epoch")
months_since_epoch_bucketized = tf.feature_column.bucketized_column(
                                    source_column = months_since_epoch_column,
                                    boundaries = months_since_epoch_boundaries)
```

Building a content-based filtering model to recommend articles to readers

kurier.at

current article id
current article title
current article author
current article category
current article age
......



next article read

```
net = tf.feature_column.input_layer(features, params['feature_columns'])
  for units in params['hidden_units']:
        net = tf.layers.dense(net, units=units, activation=tf.nn.relu)
  # Compute logits (1 per class).
  logits = tf.layers.dense(net, params['n_classes'], activation=None)
```

Lab

Create a content-based recommendation system in TensorFlow using the Kurier.at dataset

In this lab, we'll build a content-based recommender using a neural network to recommend the next article to read for visitors of the Kurier website.

Summary: The pros and cons of content-based filters

Pros

Doesn't need information about other users.

Can recommend niche items.

Cons

Requires domain knowledge to hand-engineer features.

Difficult to expand interests of user.

Quiz: Which of the following are true of content-based recommendation engines?

(choose all that apply)

A. They rely on all user-item interactions.
B. They use the item features to recommend items for a user, based on items that user has liked in the past.
C. They are the best filtering method for providing recommendations.
D. They require hand-engineered features.
E. They don't do a good job expanding the interests of the user.

Quiz: Which of the following are true of content-based recommendation engines?

(choose all that apply)

A. They rely on all user-item interactions.

B. They use the item features to recommend items for a user, based on items that user has liked in the past.

C. They are the best filtering method for providing recommendations.

D. They require hand-engineered features.

E. They don't do a good job expanding the interests of the user.

Summary

Mechanics of content-based recommendation engine.

How to build your own content-based recommendation engine.

cloud.google.com