Million Playlist - Recommender Systems

Nurlan, Zahara, Irina

Dataset Creation date : 2019-03-20

Million Playlist : Describe the chosen dataset

- Created specifically for recommender systems
- Two tables:
  - user-item interactions (91,986,831 x 2),
    - 1 column for user id
    - 1 column for item ⇒ *index*
  - track names and music bands (12,431,127 x 3).
    - 1 column for song_id ⇒ *index*
    - 1 column for song_name
    - 1 column for band name

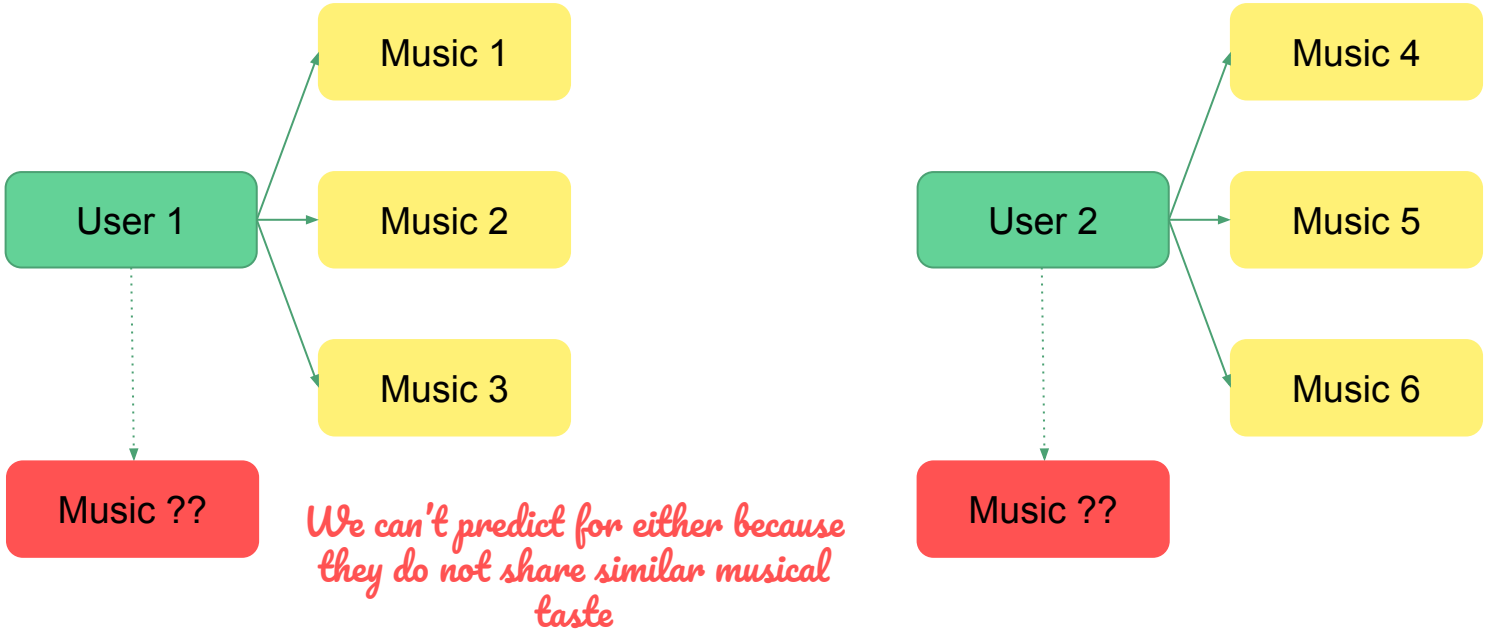
Merge

No null values

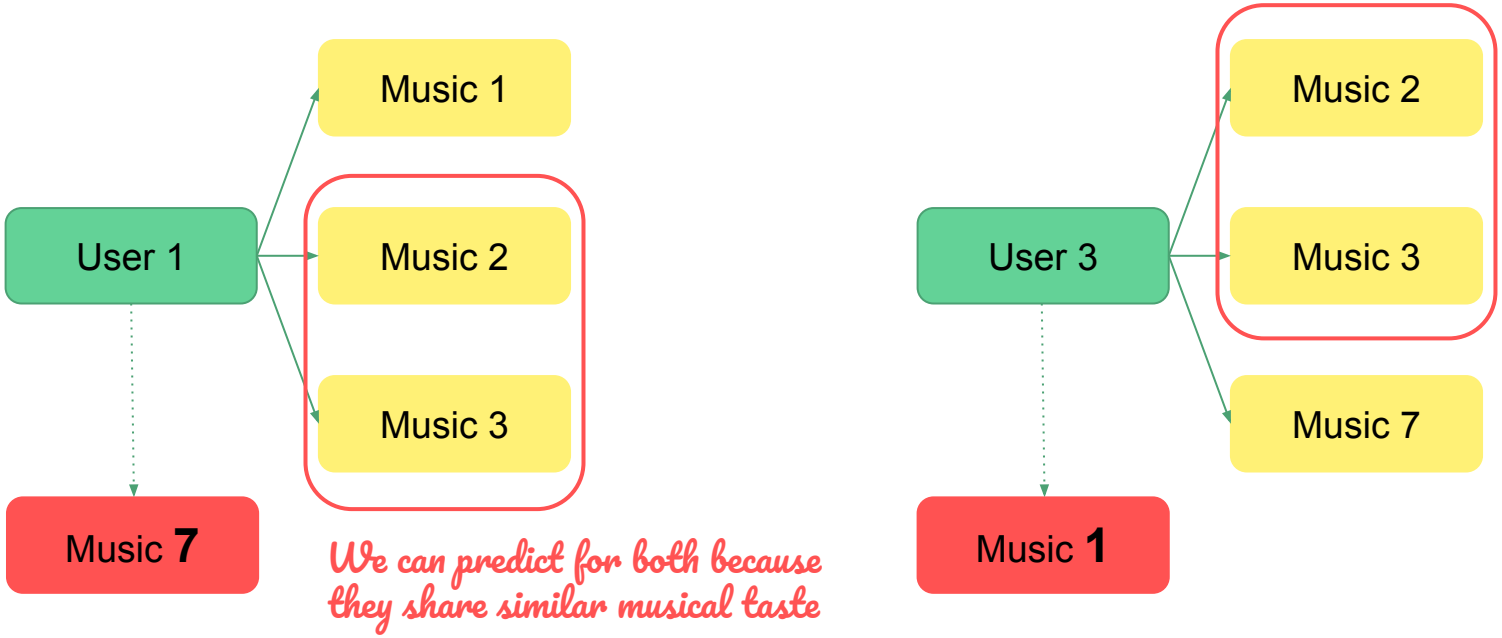
Null values :

song_id	0
song_name	38
band	3043

Million Playlist : Describe the chosen dataset



Million Playlist : Describe the chosen dataset



Recommender system: *Systems to do recommendation to user*

Used a lot in marketing to predict user preferences *(e.g. what user will buy based on what product they already bought)*

3 types of Recommender system:

- Demographic filtering :

  Generalized recommendations based on **item popularity**. User grouping is only based on generalized knowledge (**not user personal taste**) like demography.

- Content based filtering:

  Recommendation based on items similarities using **item metadata** (**not user personal taste**). Item grouping.

- Collaborative filtering

  Groups persons with similar interests, uses **user metadata** (**user personal taste**).

Matrix Factorization based recommendations

- class of collaborative filtering algorithms used in recommender systems.
- work by decomposing the user-item interaction matrix into the product of two lower dimensionality rectangular matrices

Recommendation for user id 4:
-911 by Wyclef Jean . Mary J. Blige
-If I Ain't Got YOu by Alicia Keys
-Is Sample Track 2 by Simon Thompson
-Kennedy Rag by Suzy Thompson
-Clocks by Coldplay
-Yellow by Coldplay
-Hey_ Soul Sister by Train
-The Scientist by Coldplay
-Eco by Jeorge Down

Collaborative filtering with user based Recommendation system

Identify other musics that a person might like based on his taste and comparing them with other users

Training data:

- Merge of user metadata and music dataset
- 80% of individual users preferences

Evaluating data:

- Merge of user metadata and music dataset
- 20% of individual users preferences

Testing data:

- Prediction of new songs an individual users will like

Collaborative filtering with user based Recommendation system

Identify other musics that a person might like based on his taste and comparing them with other users

We worked on 50% of the data (to much data to handle)

- Predict Band names (13 min 48s)
- Predict Song names (9 min 50s)

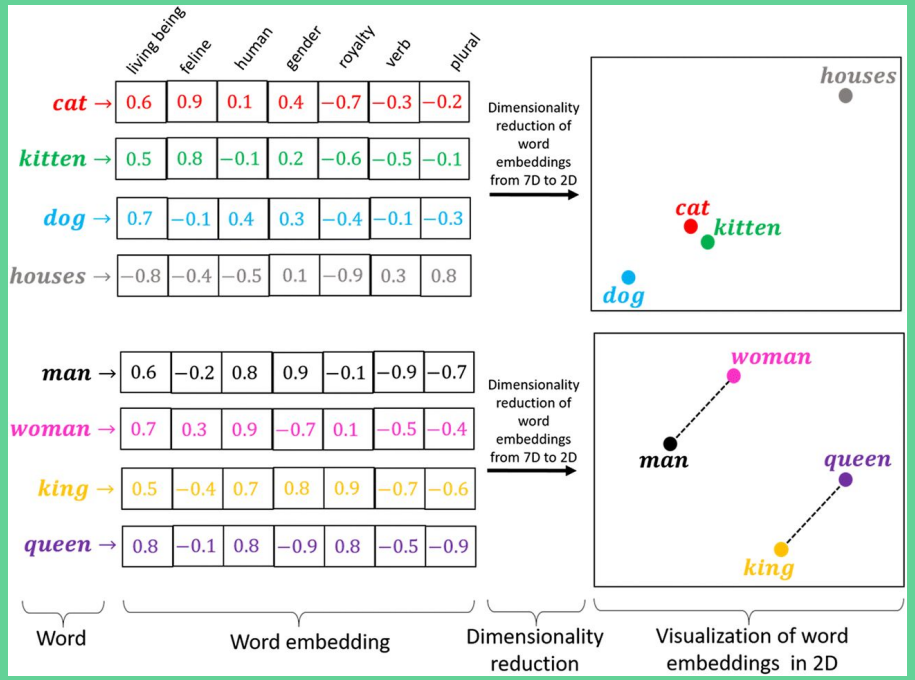
We used **gensim Word2Vec models = word embedding because we didn't have ratings**

We could have created (0,1) ratings but rather ...

Word embedding

Language processing technic

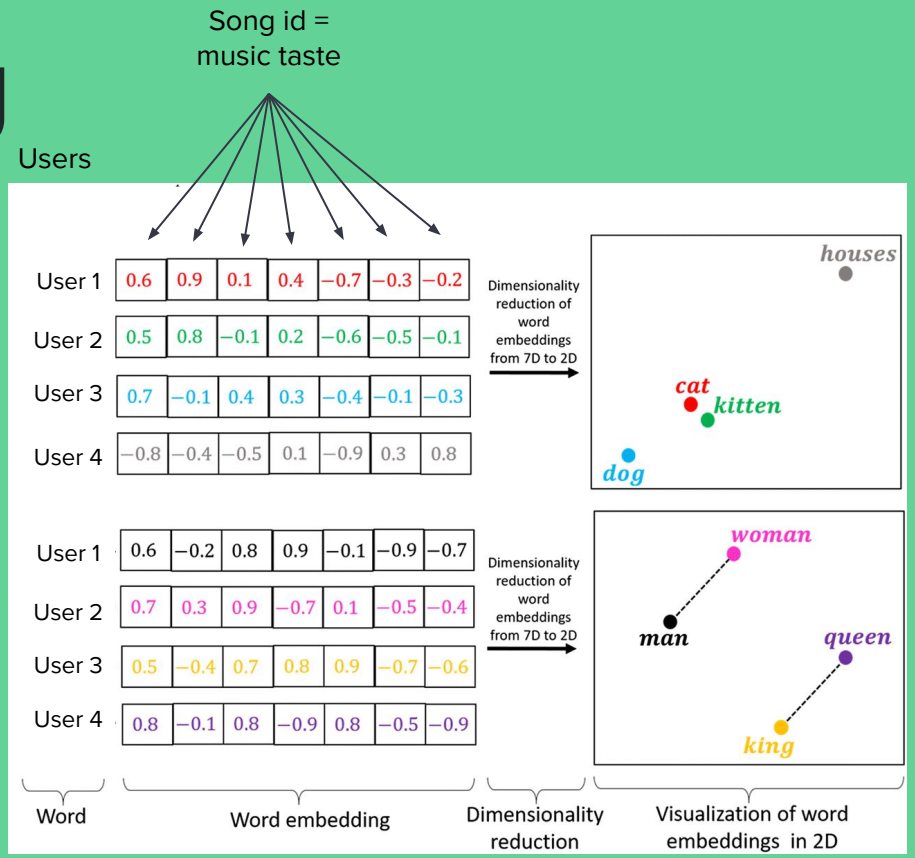
Representation of words for text analysis, by using vector that encodes the meaning of the word. The words that are closer in the vector space are expected to be similar in meaning



Word embedding

Language processing technic

Representation of **users** for **recommender system**, by using vector that encodes the **music taste of each user**. The **users** that are closer in the vector space are expected to **have similar musical taste.**



User 40638 : Band recommendation

User 40638 likes 65 bands:

['Frenchi Fer',
 'ТОП-10',
[.....]
 'Александр Константиновский']

The prediction propose the following bands:

['жека',
 'олег винник',
 'скрябин',
 'стас михайлов',
 'шансон',
 'виктор калина',
 'скрябін']

User 40638 : Music recommendation

User 40638 likes 76 musics:

```
['Komsi_Komsa',
 '2000 год',
  [...]
 ' крейсер на иголки',
 'ВОЙСКА ДЯДИ ВАСИ(кавер)']
```

The prediction propose the following musics:

```
['не плачь',
 'мама',
 'исповедь',
 'ах',
 'голуби',
 'люди',
 'журавли',
 'помнишь',
 'война',
 'ой']
```

User X (creation of a user by picking bands)

Input bands :

['ABBA',
'Blue Wednesday',
'Sum41',
'Skillet',
'Green Day',
'twenty one pilots']

Result :

['linkin park',
'nickelback',
'sum 41',
'papa roach',
'my chemical romance',
'a day to remember',
'starset']

User X (creation of a user by picking musics)

Input bands :

['Chiquitita',
'Mr. Blue Sky',
'In the End',
'Hail to the King',
'High Hopes']

Result :

['hell to pay',
'judgement
day',
'checkmate',
'vodka',
'the
awakening',
'so alive',
'superhero',
'falling apart',
"don't cry"]

Other recommender system?

Our merged dataset columns:

- User id
- Song id
- Song name
- Band name

No metadata on music (type of music, instruments, tempo...)

No metadata on users (genre they like, instrument they like...)

So we can’t do content based filtering or demographic filtering