UNIVERSITÀ DELLA CALABRIA
DIPARTIMENTO DI
INGEGNERIA INFORMATICA,
MODELLISTICA, ELETTRONICA
E SISTEMISTICA (DIMES)

Spotify

Dipartimento di Eccellenza 2018-2022

Master degree in

Computer Engineering for the Internet of Things

Course Name

Big Data Analytics

Project

Recommendation System for Spotify Users relying on Content-based Filtering approach

Professors
Andrea Tagarelli
Antonio Caliò

GitHub Link for the code
click here

Speaker
Fabio Capparelli

Recommendation System

- Its aim is to **suggest/recomend** or **predict** possible user preferences or rating to a specific item.

- They are used in the most known social media and content streaming platform such as *Facebook*, *Instagram*, *YouTube*, *Spotify*, *Netflix*, *Amazon*.

- It uses many *algorithms* and *approaches*:
  - **Collaborative** filtering.
  - **Content-based** filtering.
  - **Session-based** recommender systems.
  - **Hybrid** recommender systems.

Content-based filtering

- It suggests items only relying on their content and does not leverage on user behavior, interactions, rating on the platform/social media, two principal concepts:

  - The **items content** could be identified with its *attributes*, *features or description*.
  - The **user profile** is built following only user *preferences and items* in its platform database.

Content-based filtering

- The use of this kind of approach has ensured to us the following PROS:
  - Major **privacy**, because we are not interested in the user actions, behavior.
  - No user rating is required.
  - The system is scalable, if the user preferences change (he/she adds new songs or new playlist) the system changes in a few time.
- Otherwise, some “CONS” are:
  - It requires very accurate item description;
  - It is strictly depended on the item description/attributes and not in the real user like/rating, and sometimes these are not sufficient to perform an accurate recommendation.

Content-based filtering

- This approach relies on **three main steps**:

1. Acquiring **items/objects** knowledge.
2. Acquiring **user profile** knowledge.
3. Performing **dot product** to select similar items.

Acquiring items/objects knowledge.

- It relies on selecting proper **attributes**, **features** of the specific item and assigning to those a value that can be *binary*, *numerical* that will represent the "**Content**".

Essentially a generic item is described as a **vector**, of **n** components:

$$\boldsymbol{X} = \{\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}, \ldots, \boldsymbol{x_n}\}$$

where components are the value (numerical, binary) assigned to the attribute of that item.

Acquiring user profile knowledge.

- Similarly, this step defines a user profile record/vector having the same **attributes**, **features** of the items that we want to recommend and assigning to those a *binary*, *numerical* value that will represent the "**User Content**" depending on user preferences.

It is a generic **vector**, of **n** components:

$$Y = \{ y_1, y_2, y_3, \dots, y_n \}$$

where components are the value (numerical, binary).

Performing the dot product

- On this step is performed the "**Dot Product**" among each **item vector** and **user profile vector** for gaining a Recommended measure for each item.

  A Recommended measure, **R** of item **k** will be defined

$$R_k = \langle X_k \cdot Y \rangle = x_{k,1} \cdot y_1 + \cdots + x_{k,n} \cdot y_n = \sum_{i=1}^{n} x_{k,i} \cdot y_i$$

- Now it is necessary to define a **support S** value for giving a proper threshold to define if an item is affine and should be recommended.

$$\begin{cases} R_k < S \rightarrow Not\ Recommended \\ R_k \geq S \rightarrow Recommended \end{cases} \quad \text{V} \quad \begin{cases} R_k \leq S \rightarrow Not\ Recommended \\ R_k > S \rightarrow Recommended \end{cases}$$

Our System

Our solution is based on the following **three steps**, according to what has been said before:

1. Building the *user profile* getting the "audio_features" of the song inside the user playlists.
2. Getting the *items* "audio_features" of about **500 tracks** of the newest albums uploaded on spotify.
3. Building the final recommender system performing the dot product among the user profile to all newest tracks and selecting the tracks that have the result above a **support** provided as input via console.

The Content

We rely on the tracks' audio_features, that are parameter/attributes which characterize each song inside spotify, we considered the following:

- **Danceability**
- **Energy**
- **Liveness**
- **Speechiness**
- **Acousticness**
- **Instrumentalness**
- **Tempo**
- **Valence**

The Content Preprocessing

Most of them stand **in a range [0, 1]**, and they are represented by a float number, to convert them to integer value the following operation is performed:

$$Integer_{Attr} = round\,(\,Float_{attr} \cdot 5\,)$$

Exception is made for «TEMPO», it is just and float number greater than 1. We have to perform another step before.

We put as "*upper bound*" for bpms the **190**. This because after some research on the spotify and songs in general the higher bmp stands around 180 185, finally to put "Tempo" inside a range [0,1] we performed the following operation

$$Integer_{Attr} = round\left(\frac{Float_{attr}}{190} \cdot 5\right) = round\left(\frac{Float_{attr}}{38}\right)$$

STEP 1

BUINLDING THE USER_RECORD

On this step we get the user's songs inside his/her playlists, the steps followed have been:

1. Collect tracks inside an array deleting duplicates.
2. Retrieve the attributes we are interested in, in combination with some header attributes (id, name, artists, reference link).
3. Perform the preprocessing stage described before and store the inside proper dataset.
4. Perform the mean for each attribute (danceability, energy, … , tempo)
5. Obtain the user_record.

STEP 1

BUINLDING THE USER_RECORD

Attributes of a track inside a user playlist:

ID	Name	Artist	Link	Danceability	Energy	Liveness	Speechiness	Acousticness	Instrumentalness	Tempo	Valence
f3e2039043474108	Zitti e buoni	Maneskin	click	0.625	0.939	0.424	0.0669	0.00138	0	102.999	0.644

After preprocessing:

ID	Name	Artist	Link	Danceability	Energy	Liveness	Speechiness	Acousticness	Instrumentalness	Tempo	Valence
f3e2039043474108	Zitti e buoni	Maneskin	click	3	5	2	0	0	0	3	3

STEP 1

BUINLDING THE USER_RECORD

User record example for ‘Fabio Capparelli’

ID	Name	Link	Danceability	Energy	Liveness	Speechiness	Acousticness	Instrumentalness	Tempo	Valence
11102339711	Fabio Capparelli	click	3	3	1	1	1	0	3	2

User record example for ‘Francesco Raco’:

ID	Name	Link	Danceability	Energy	Liveness	Speechiness	Acousticness	Instrumentalness	Tempo	Valence
prp468n1n5qp2sdr1ps5hk8t0	Francesco Raco	click	3	3	1	0	1	0	3	3

User record example for ‘Luigi Rachiele’:

ID	Name	Link	Danceability	Energy	Liveness	Speechiness	Acousticness	Instrumentalness	Tempo	Valence
gixs	Luigi Rachiele	click	3	3	1	0	1	1	3	2

STEP 2

RETRIEVE THE NEWEST SONGS

On this step we get the spotify new releases:

1. Get newest published albums
2. Up to 500 hundred songs
3. Perform preprocessing for each song
4. Store the into a similar dataset used for user's playlists songs

STEP 3

SUGGEST NEW SONGS

On this step we recommend new tracks for the spotify user:

1. Perform the “Dot Product” among user record and each new song
2. Set a proper support S value in range [0,1].
3. Select only the songs that has as result a recommended value above the support S
4. Print suggested song.

STEP 3

OUR RECCOMENDATION FOR FABIO CAPPARELLI

```
> Please insert a user id:
11102339711

> Building a recommendation system for:
Fabio Capparelli

> USER_PROFILE:
danceability 3 energy 3 liveness 1 speechiness 1 acousticness 1 instrumentalness 0 tempo 3 valence 2

> Looking for new releases . . .

> Please insert support in range [0, 1]:
0.25

> Recommented songs:

Song: Theme For The People
Artist:  Logic
Recommendation Support: 0.255
Reference Link: https://open.spotify.com/track/0UVbDDuCr1weCpZCAxTFa5

Song: BATHROOM (skit)
Artist:  The Kid LAROI
Recommendation Support: 0.265
Reference Link: https://open.spotify.com/track/4BhDYmhCmMZCUmPxjzWtDX

Song: Lay Wit Ya (feat. Duke Deuce)
Artist:  Isaiah Rashad
Recommendation Support: 0.27
Reference Link: https://open.spotify.com/track/5KW5AYiCyi5auXXZR2cvxM

Song: El Primo
Artist:  Diferente Nivel
Recommendation Support: 0.26
Reference Link: https://open.spotify.com/track/6xTSszx1dP0LZ92gxKNdFi
```

STEP 3

OUR RECCOMENDATION FOR FRANCESCO RACO

```
> Please insert a user id:
prp468n1n5qp2sdr1ps5hk8t0

> Building a recommendation system for:
prp468n1n5qp2sdr1ps5hk8t0

> USER_PROFILE:
danceability 3 energy 3 liveness 1 speechiness 0 acousticness 1 instrumentalness 0 tempo 3 valence 3

> Looking for new releases . . .

> Please insert support in range [0, 1]:
0.25

> Recommented songs:

Song: Theme For The People
Artist:  Logic
Recommendation Support: 0.275
Reference Link: https://open.spotify.com/track/0UVbDDuCr1weCpZCAxTFa5

Song: BATHROOM (skit)
Artist:  The Kid LAROI
Recommendation Support: 0.255
Reference Link: https://open.spotify.com/track/4BhDYmhCmMZCUmPxjzWtDX

Song: From These Heights
Artist:  Jelani Aryeh
Recommendation Support: 0.255
Reference Link: https://open.spotify.com/track/74IoFNuC2nDDFkrt4MAgpS

Song: Contigo Mami
Artist:  Ramon Vega
Recommendation Support: 0.27
Reference Link: https://open.spotify.com/track/119A9Ea2mqoGTUUUILsQZF
```

```
Song: Lay Wit Ya (feat. Duke Deuce)
Artist:  Isaiah Rashad
Recommendation Support: 0.29
Reference Link: https://open.spotify.com/track/5KW5AYiCyi5auXXZR2cvxM

Song: Perla
Artist:  Zion & Lennox
Recommendation Support: 0.26
Reference Link: https://open.spotify.com/track/74C728ToDJFoZ4Huws0YlN

Song: Sistema
Artist:  Zion & Lennox
Recommendation Support: 0.255
Reference Link: https://open.spotify.com/track/1BRCOxoDsForww8nCBqfVH

Song: No Más
Artist:  Zion & Lennox
Recommendation Support: 0.26
Reference Link: https://open.spotify.com/track/2W1OyE9v5kibecNNBnjFCe

Song: Riata Dada
Artist:  EST Gee
Recommendation Support: 0.26
Reference Link: https://open.spotify.com/track/6e7VjMcOvEiF6no2bTxcGb

Song: Daily Routine
Artist:  GRiZ
Recommendation Support: 0.265
Reference Link: https://open.spotify.com/track/67tdgyCcsTd4qRK4BmscLO

Song: El Primo
Artist:  Diferente Nivel
Recommendation Support: 0.28
Reference Link: https://open.spotify.com/track/6xTSszx1dP0LZ92gxKNdFi
```

STEP 3

OUR RECCOMENDATION FOR LUIGI RACHIELE

```
> Please insert a user id:
gixs

> Building a recommendation system for:
Luigi Pss Rachiele

> USER_PROFILE:
danceability 3 energy 3 liveness 1 speechiness 0 acousticness 1 instrumentalness 1 tempo 3 valence 2

> Looking for new releases . . .

> Please insert support in range [0, 1]:
0.25

> Recommented songs:

Song: Theme For The People
Artist:  Logic
Recommendation Support: 0.255
Reference Link: https://open.spotify.com/track/0UVbDDuCr1weCpZCAxTFa5

Song: Lay Wit Ya (feat. Duke Deuce)
Artist:  Isaiah Rashad
Recommendation Support: 0.265
Reference Link: https://open.spotify.com/track/5KW5AYiCyi5auXXZR2cvxM

Song: El Primo
Artist:  Diferente Nivel
Recommendation Support: 0.255
Reference Link: https://open.spotify.com/track/6xTSszx1dP0LZ92gxKNdFi
```

THANKS FOR THE ATTENTION