Video Clip Retrieval Based on Similarity

Sideras Andreas, Boura Tatiana

Instructors: Giannakopoulos T. & Maglogiannis I.

July 3, 2023

MSc in Artificial Intelligence, NCSR Demokritos & University of Piraeus
Multimodal Machine Learning

Table of contents

Project Definition

Project Definition

- Our objective in this assignment is to create a model capable of suggesting similar video clips based on a query.
- This query is a single YouTube video clip, and by using a similarity metric, we retrieve the top-k most relevant video clips.
- Each video clip is composed of 3 modalities (*audio - video - text/lyrics*).
- We train a model that **learns representations** for each video clip by **combining the features extracted from the 3 modalities**.
- Our data collection consists of 821 video clips from *9 different music genres.*

Feature extraction

Audio features

- We computed **257** mid-term audio features using Librosa.
- The aggregations we made were **mean**, **median**, **std**, and **75th-percentile**.

Features
BPM
Zero crossing rate
Mel spectrogram and dB-scaled spectrogram
Harmonics
Perceptual shock wave
Spectral centroids and their 1st and 2nd derivatives
Spectral flux and Spectral rolloff
Spectral bandwidth 2, 3, and 4
12 Chroma features
13 MFCCs and their 1st and 2nd derivatives

- We used a pre-trained model that performs 3D convolutions. We end up with a **400-d feature vector**.
- The model is a 3D RESNET that captures the spatial semantics of an input video by using 3D convolutions.
- In **3D convolution**, the kernel becomes a 3D tensor, and the convolution operation is applied to a 3D input volume, which can be thought of as a stack of 2D images or as a video sequence.

- We obtained the lyrics of our videos using the YouTube API.
- For each video clip, we obtained a single vector representation of all its lyrics (**768 dimensions**) using the **BERT** model.
- BERT is a pre-trained language model that utilizes a Transformer architecture.
- It outputs a vector representation for each token, including an extra one called the **[CLS]** token.
- During the training process, the [CLS] token is used as an input to the NSP task.
- The [CLS] token captures the overall information and contextual understanding of the entire input sequence, allowing it to serve as a summary representation of the sentence pair.

Representation Learning

Representation Learning

- We concatenated all the feature vectors of the 3 modalities, resulting in a single vector of **1425** dimensions.
- We need a way to uncover the intrinsic dependencies between the 3 modalities.
- We use an **AutoEncoder** architecture that maps the video clip features to a lower-dimensional latent space.

AutoEncoder

```
Autoencoder(
  (encoder0): Linear(in_features=1425, out_features=1024, bias=True)
  (activation): Tanh()
  (encoder1): Linear(in_features=1024, out_features=1024, bias=True)
  (activation): Tanh()
  (encoder2): Linear(in_features=1024, out_features=768, bias=True)
  (activation): Tanh()
  (encoder22): Linear(in_features=768, out_features=512, bias=True)
  (activation): Tanh()
  (encoder3): Linear(in_features=512, out_features=256, bias=True)
  (activation): Tanh()
  (decoder0): Linear(in_features=256, out_features=512, bias=True)
  (activation): Tanh()
  (decoder11): Linear(in_features=512, out_features=768, bias=True)
  (activation): Tanh()
  (decoder1): Linear(in_features=768, out_features=1024, bias=True)
  (activation): Tanh()
  (decoder2): Linear(in_features=1024, out_features=1024, bias=True)
  (activation): Tanh()
  (decoder3): Linear(in_features=1024, out_features=1425, bias=True)
)
```

- After training the AutoEncoder, we compute the **embedding** for each video clip. This embedding is just the *output of the encoder*, a **256-d** vector.
- We compute the cosine similarity between the embeddings of two video clips.

Experiments

- We **compared** the AutoEncoder method against just taking the *cosine similarity over the initial feature vectors* and *after applying PCA*.
- We also performed **clustering** on the representations of the latent space.
- It is quite challenging to provide a **quantitative metric** for this task.
- We attempted to gain insights into the performance of our model by manually assessing some video clips of our choice.
- The assessment is primarily based on the similarity of music genres and resemblance of videos.

AutoEncoder

```
205    Avenged Sevenfold - Shepherd Of Fire [Official...
466    Nim Vind - 'Where I'm From' House of Vind - A ...
640              WATERPARKS - RITUAL (VISUAL)
394           K.Flay - Raw Raw (Official Music Video)
279    Disturbed - Another Way To Die [Official Music...
```

Raw Features

```
527                    Parklife (2012 Remaster)
543    Sonic Martini Music - User User User (Official...
551    The Calling - Wherever You Will Go (First Vide...
709        Sum 41 - In Too Deep (Official Music Video)
624           Green Day - Hitchin' A Ride (Video)
```

After PCA

```
737    Yellowcard - Southern Air (+ Lyrics & HQ_HD Au...
543    Sonic Martini Music - User User User (Official...
496    Matchbox Twenty - Don't Get Me Wrong [Official...
184      Destiny's Child - Soldier ft. T.I., Lil' Wayne
201    Ice Cube, Mack 10, Ms. Toi - You Can Do It (Of...
```

Query : Falling In Reverse - ZOMBIFIED

Similar 1 : Avenged Sevenfold - Shepherd Of Fire

Similar 2 : Nim Vind - "Where I'm From" House of Vind

Similar 3 : WATERPARKS - RITUAL

Similar 4 : K.Flay - Raw Raw

Similar 5 : Another Way To Die

Video Clip : STORMZY - LONGEVITY FLOW

AutoEncoder

250	Central Cee - Retail Therapy [Music Video]
733	Mike Jones - Back Then Part 2 (Video)
617	Female Special - Plugged In w_ Fumez The Engin...
732	Cristale x Teezandos - Plugged In w_ Fumez The...
441	Ying Yang Twins - Salt Shaker (feat. Lil Jon &...

Raw Features

215	Lloyd Banks - On Fire (Explicit)
118	Old Dominion - One Man Band
531	Ricky Montgomery - Line Without a Hook (Offici...
160	Baby Bash - Suga Suga ft. Frankie J
146	Zac Brown Band - Chicken Fried (Official Music...

After PCA

531	Ricky Montgomery - Line Without a Hook (Offici...
160	Baby Bash - Suga Suga ft. Frankie J
215	Lloyd Banks - On Fire (Explicit)
47	Dan + Shay - Speechless (Wedding Video)
173	Chris Brown - Yo (Excuse Me Miss) (Official HD...

Clustering

cluster		
1	695	Luke Bryan - Roller Coaster (Official Music Vi...
	57	Choking Victim - War Story
	111	Craig Cardiff - Emm & May (Official Video)
	393	Coldplay - Adventure Of A Lifetime (Official V...
	626	Murray McLauchlan _ Down by the Henry Moore
2	667	Billy Ray Cyrus - Achy Breaky Heart (Official ...
	436	One Direction - You & I
	253	Chris Young - The Man I Want To Be (Official V...
	820	Sam Smith - Stay With Me ft. Mary J. Blige (Live)
	402	Maddie & Tae - Die From A Broken Heart (Offici...
3	351	(OOAK) Russ Millions x Buni x YV - Reggae & Ca...
	132	Florence + The Machine - Dog Days Are Over (20...
	407	Pennywise - 'Same Old Story'
	541	Coldplay - Hymn For The Weekend (Official Video)
	760	Sly Slick and Wicked - Somebody Please
4	440	Hate It Too - 'Twelve's the New Eight' Officia...
	794	Counting Crows - Mr. Jones (Official Music Video)
	23	DragonForce - Three Hammers (Live)
	379	OutKast -- Hey Ya lyrics
	121	Chris Lane - I Don't Know About You (Official ...
5	750	Grey Kingdom 'Sun Like Moon Light' - www.strea...
	590	Madison Violet - Small Of My Heart
	67	Johnny Flynn - Raising the Dead
	502	Mystery Train _ Playing For Change _ Live Outside
	237	Bukka White - Miss Leola
6	20	Huey - Pop, Lock & Drop It (Video Edit)
	41	50 Cent - Outta Control ft. Mobb Deep
	722	Lil Wayne - Fireman (Official Music Video)
	360	Eminem - Stan (Long Version) ft. Dido
	391	Ludacris - Rollout (My Business) (Official Mus...
7	19	Blur - Girls And Boys (Official Music Video)
	250	Central Cee - Retail Therapy [Music Video]
	725	The Struts - Kiss This (2014)
	362	Matchbox Twenty - Wild Dogs (Running in a Slow...
	129	Steel Panther - Party Like Tomorrow Is The End...
8	758	Morgan Evans - Over For You (Official Music Vi...
	526	Kelsie Kimberlin - Cosmopolitan Girl _ #stand...
	245	Queens Of The Stone Age - No One Knows (Offici...
	373	You Are The First, My Last, My Everything (Bar...
	336	Maroon 5 - Won't Go Home Without You (Official...
9	329	Motörhead – God Save The Queen (Official Video)
	18	Foo Fighters - Best Of You (Official Music Video)
	423	Jet - Are You Gonna Be My Girl
	170	Sex Pistols - God Save The Queen
	782	Social Distortion - Ball and Chain

Thank you!

Any questions?