Multimedia Search and Retrieval (MMSR), Winter Term 2023/24

Task 3: Video-based retrieval, fusion, evaluation and user interface.
Introduction: 13.12.23
Q&A: 10.01.24
Deadline for submission: 17.01.24, 12:00

The third task complements the project with

- the implementation and evaluation of three additional music retrieval systems,
- the implementation of a user interface for retrieving music similar to a given query track.

Finally, you will present your project.

Retrieval systems: The input (query) to the system is the title and artist of a song (track). The output of the system should be a list of songs (title and artist) of length N that are similar to the query song. In this exercise, we will investigate a video-based retrieval system and two fusion techniques.

- **Video-based(<similarity>, <feature>):** similar to Text-based(<similarity>, <feature>) and Audio-based(<similarity>, <feature>). Choose as feature one of the video features (i.e., the representations of the videoclips obtained with one of the neural network architectures, VGG19, ResNet, or inception, and provided at the link below).

$$sim(query,\ target\_track)\ =\ <similarity>(<video\,feature\ >(query),\ <video\,feature>(target\_track))$$

- **Early fusion:** from all the features provided so far (textual, audio, and video) select two and use an early fusion aggregation technique to combine them into a single feature. Then use the feature resulting from the aggregation in a retrieval similar to Video-based(<similarity>, <feature>). Motivate your methodological choices, such as the choice of the features to combine from different modalities, the pre-processing of the features, if any is applied (standardization, normalization, PCA, …).

$$sim(query,\ target\_track)\ =\ <similarity>(<aggregated\,feature>(query),\ <aggregated\,feature>(target\_track))$$

- **Late fusion:** this retrieval system combines the **results** of two of the retrieval algorithms developed so far. Select two of the algorithms you already developed, motivate the choice of the selected algorithms (features and similarities), and of the late fusion techniques (e.g., rank or score aggregation), as well as any additional methodological choices (i.e., scaling or weighting of the ranks, or of the scores). **Hint:** use precomputed scores / retrieval results .

Evaluation: evaluate and present results as a table including all systems from tasks 1, 2, and 3 (11 systems in total) and all implemented metrics:

- ○ Accuracy: **Precision@10**, **Recal@10**, **nDCG@10**
- ○ Beyond accuracy: **Genre_coverage@10**, **Genre_diversity@10**

Please include a **Precision-Recall plot** with curves for the three new systems (varying **k** in the interval [1, 100]).

Hint: you can reuse your implementation and evaluation results from Task 2.

User Interface: Choose at least one of your implemented retrieval systems and create a publicly available web interface allowing to pass different queries and retrieve relevant tracks. The interface should allow users to interact with the query and retrieved tracks e.g. through YouTube. Minimum functionalities:

- select the query song from a search box or list
- display the top 10 most similar songs in a list
- the elements of the list [give access to / embed] the youtube video of the song, so that the user can play them, the query track should be accessible in the same way
- The system should be deployed and publicly available (e.g. hosted via GitHub pages)

Feel free to implement additional useful features (e.g. allowing to switch between different systems, compare their results side by side, display additional track information, such as genre, popularity, …).

Hint: You can save precomputed retrieval results, accessing them on demand as the queries are passed to the system.

Presentation: On January 17th, 2024, you will present the project developed within the course. The presentation should last no more than 15 minutes and should cover the methodological and scientific aspects of the project, as well as include a demonstration of your user interface.

Hints:

- For a good scientific presentation, it is good practice to include
  - the methodological details (e.g., choice of feature and similarity function)
  - the easily graspable presentation and concise analysis of the results (e.g., comparison of performances of different retrieval systems)
- Ideally, the methodological choices are also motivated. We understand that 15 minutes is a short time for that amount of detail. Therefore, you might choose to omit the *motivation* of your methodological choices from the presentation, to favor clarity. However, be prepared for corresponding questions.
- The implementation details (e.g., code snippets, names of libraries, …) are *not* required
- The grade is given to the group, it will only be affected by the quality of the presentation, and not be given individually to the speaker(s). We encourage you to pick the best speaker(s) among you,for a maximum of two.

Hint on task 3:

The data required to complete task 3 are available at the following link:
https://drive.google.com/file/d/1sWf4zmP8gsYWUp01AMvb37N564aK5VIw/view?usp=sharing

Each video feature is stored as a .tsv file with a column containing the identifier of the track, and the remaining columns containing the components of the feature vector. Each file is named as id_<*name_of_feature*>_mmsr.tsv.
Genres are stored in the same file of Task 2, i.e., as a .tsv file with a column containing the identifier of the track, and another column containing the list of genres corresponding to the track.
YouTube Links are stored as a .tsv file with a column containing the identifier of the track, and another column containing the link.

As in Task 2, for precision-recall plots, you will have to vary the length **k** of the list of retrieved tracks. To speed things up you can save the longest list (e.g. **k**=100) and then only select the top **k** for each **k** ≤ 100.

For the remaining metrics (nDCG, coverage, and diversity), only consider a list of **k**=10 retrieved tracks.
To compute the IDCG of a given query track, you will have to compute the relevance of all possible retrieved tracks.
In your report, include a table consisting of 11 rows, one for each retrieval system, and 5 columns, one for each evaluation metric computed for a list of **k**=10 retrieved tracks (precision@10, recall@10, nDCG@10, coverage@10, and diversity@10).

To optimize the functionality of the user interface to the retrieval system, you are encouraged to precompute and store the similarities.

Lab report: Prepare a consolidated report describing the aim of the project, your approach, the experimental setup, and the results/findings. Make sure that your report also includes the retrieval systems from Tasks 1 and 2 in the analysis.

If your final report consists of an extension and revision of the previous lab reports, distinguish the new text from the old (e.g., writing the old text in gray, or another sensible font color, and the new one in black).

Files:

1. Your report as a .pdf
2. Link to the user interface
3. Source code (e.g., link to Github)
4. (Optional): link to Overleaf version of the report

Submission: Deadline is January 17 at 12:00, via mail to *markus.schedl@jku.at*, *oleg.lesota@jku.at* **and** *marta.moscati@jku.at*. Project presentations will take place on the same day, January 17, in the lecture timeslot, i.e., from 13:45 to 16:15.