Final group assignment: Objective and subjective metrics in Spotify recommendations

The group assignment will allow you to work with music recommender systems data, use content-based or collaborative filtering, calculate metrics, rerank or generate new recommendation lists and evaluate the system based on user feedback using subjective (and objective) metrics. We follow a three step process: first collect preference data from a large group of participants, second calculate recommendations on this data, and third go back to the users to ask their feedback on the recommendations.

You will be split into groups of 3 to 4 students. Each group chooses their own focus in the project. Groups will be supervised during short feedback sessions.

We will collect rating data from a representative set of spotify songs, as well as musical expertise (MSI) from a pool of participants you will collect yourself, using an online survey we have prepared. Each group should try to get data from about 20-30 people. Participants indicate which group invited them. In this way you can go back to the participants in your group to deliver them personalized recommendations and collect their feedback. However, as we collect rating data from a large group of participants across all groups, we can also build are recommendations on all data from all participants. Based on the rating data, and the spotify feature data we already have for each song, and MSI scores of each user you can build new recommendation lists for each participant . For example you can use collaborate filtering on the ratings, use content-based filtering on the audio features or a mix of both.

The idea is to generate new lists and get back to the participant and present this list to them, for example using a simple HTML page or microsoft form. You can change the list by changing the ranking, filter based on specific spotify features, diversifying the list or anything else that you can come up with based on what you learned in the lectures.

Let the participant review their new list and provide feedback. You can collect ratings, rankings, and/or user experience measures such as perceived quality, perceived diversity, etc.

This allows you to ask questions like:
- Will people rate songs from a new genre higher if they are more personalized?
- Will people rate popular songs (use spotify popularity feature) higher than less popular ones, even if they are personalized?
- Do people experience lists that have been filtered for a lower valence (audio feature) also as darker/low mood?
- Does a diversified list bring higher perceptions of diversity? How strong is the correlation between subjective and objective diversity? Are people more or less satisfied with a diverse list?
- Evaluate different explanations or visualizations of the playlist (perhaps even use an LLM to explain recommendations and compare it against a feature-based explanation).
-

Ideally you will construct different versions and check how they affect user experience differently (e.g. high or low diversity list) using an AB test in a within- or between subjects design.

Project outline:
Oct 22: introduction to the project, create groups, collect data
Oct 29: CF tutorial, assignment 2.
Nov 5: Lecture. Data collection finished, we will share the data. after the lecture: discuss the plans for the project with the supervisors to get feedback, start doing the data manipulation/recommendations
Nov 12: MF / diversification tutorial
Nov 19: feedback session on the result of your data manipulations
Nov 20-25: Send out a list of recommendations to the original participants to get their feedback
Nov 26: Lecture and feedback session. analyze the data and start writing the report
Dec 3: present your results in class, some opportunity for feedback
Dec 19: deadline for the report