Melodify: Music Similarity Mapping

Group Name: Melodify
Group Leader: Gautham Satyanarayana

Group Info

- Group Members: Saumya Chaudhury, Pradeep Raj, Prabhu Raj, Yukta Salvi, Parikha Goyanka
- Group Leader: Gautham Satyanarayana

1 Introduction

Identifying similar songs is a complex problem as songs can vary along many dimensions that are not always predefined or easily quantifiable. Traditional music recommender systems typically rely on user listening history or explicit feedback using algorithms like collaborative filtering to suggest songs. However, such methods often overlook the intrinsic characteristics of the music itself, instead focusing primarily on user behavior patterns. This limitation is particularly evident in cold start scenarios where a lack of user history makes recommendations unreliable. While some systems incorporate basic audio features such as tempo, key, and rhythm, these approaches are often too simplistic to capture a song's full semantic and artistic value. Music is inherently multifaceted, involving a combination of lyrical meaning, vocal delivery, instrumentation, and other auditory elements that contribute to its emotional and aesthetic impact on listeners. As a result, a more holistic method is required to effectively understand and compute song similarity in a meaningful way.

2 Problem and Goal

In modern music recommendation systems, finding songs similar to a user's preference remains challenging, particularly in scenarios lacking user history where collaborative filtering techniques face the "cold start" problem. While content-based approaches such as those used by Spotify rely on feature extraction (e.g. audio profiles, instruments, and tempo), they often fail to capture the underlying semantics and emotional "feel" of a song. Since a song's impact on listeners is a subjective and multidimensional experience, there is a need for more nuanced representations. This paper proposes an algorithmic framework that generates embeddings or feature vectors based on a song's inherent content, including its audio characteristics, lyrical content, instrumentation, and vocal qualities. By leveraging these features, we aim to embed songs into a vector space that encapsulates their semantic and emotional essence. Subsequently, clustering techniques can be applied to these vectors to identify songs with similar "vibes" or emotional resonance. This approach seeks to enhance the accuracy of music recommendations by moving beyond traditional feature-based methods, offering a more comprehensive representation of a song's affective and experiential qualities.

3 Formalization into an ML task

Data: The data would be in the form of audio obtained from the Million Song dataset. The project also utilizes Last.fm for song tags and EchoNest Taste Profile Subset for data on songs and users. The datasets will generate embeddings that will later be embedded into vectors and clustered into groups.

Function: The function aims to group the songs according to similar features. Initially, filtering based on the most tagged songs will take place. The transformer will be fed with the filtered set as the input. The transformer output, comprising embeddings, will be embedded into vectors and finally the data will be clustered based on the feature vectors.

4 Data Plan

We take both user-item interactions and music audio content end-to-end to train the model. The dataset used will be the EchoNest Taste Profile Subset and Last.fm. To check the accuracy of the recommendation experiments, we validate them with two tasks: music recommendation and music auto-tagging. For music recommendation, we will use a subset of the Echo Nest Task Profile Subset, which contains about 20,000 active users and 10,000 songs. In the case of auto-tagging, we filter the song list to those containing at least one of the 50 most used tags. While training in user-item interactions, we will use a lookup table and for music content, each interaction will be paired up with one listened-to song and several unlisted songs. Together, this will be fed to the transformer network.

5 Project Schedule

The tentative project schedule is as follows:

Week	Task Description
Week 1	Exploratory Data Analysis (EDA) and Data Preprocessing. Perform detailed analysis of the dataset, including genre distribution, song counts, audio quality assessment, and the availability of lyrics. Preprocess both the audio files and lyrics for further analysis.
Week 2	Transformer Architecture and Embedding Model Training. Select appropriate feature sets for analysis. Train a transformer-based architecture to extract features from the cleaned dataset, including both audio and lyrical content. Embed audio and lyrics into a unified feature space.
Week 3	Model Evaluation and Fine-Tuning. Analyze preliminary results and fine-tune the model for enhanced feature extraction, particularly focusing on sentiment analysis from the audio. Expand the feature set and assess model performance.
Week 4	Clustering and Validation. Perform clustering of songs based on the generated feature vectors. Validate the clusters by comparing them with existing audio embedding models for consistency and accuracy.
Week 5	Re-evaluation and Optimization. Integrate results from clustering and embedding models. Optimize the model by further fine-tuning feature extraction mechanisms to improve overall performance.
Week 6	Comparison with State-of-the-Art (SOTA) Methods. Compare the developed model against state-of-the-art algorithms for song similarity measurement, conducting comprehensive tests to evaluate relative performance.
Week 7	Conclusion and Documentation. Visualize the results, provide an in-depth analysis of the outcomes, and document the entire process, including insights, challenges, and future work.

6 References

1. Hongwei Wang, Miao Zhao, Xing Xie, Wenjie Li, Minyi Guo. Knowledge Graph Convolutional Networks for Recommender Systems. arXiv:1904.12575v1 [cs.IR] 18 Mar 2019.

2. Jongpil Lee, Kyungyun Lee, Jiyoung Park, Jangyeon Park, Juhan Nam. Deep Content-User Embedding Model for Music Recommendation. arXiv:1807.06786v1 [cs.IR] 18 Jul 2018.

3. Seokgi Kim, Jihye Park, Kihong Seong, Namwoo Cho, Junho Min, Hwajung Hong. Can Music Be Represented With Numbers? arXiv:2102.13350 26 Feb 2021.