Tackling the Problem of Cover Song Similarity Metric Design

Overview

The goal is to design a similarity metric that quantifies the relationship between original songs and their cover versions. The proposed metric should minimize distances between cover songs (same "work") and maximize distances otherwise. Since the **Da-TACOS dataset** provides pre-extracted features, our proposed approach leverages these features for accurate and efficient similarity computation.

The **8 transformation metrics** (timbre, tempo, timing, structure, key, harmonization, lyrics, and noise) outlined in the research paper are adapted based on the features available. Additionally, recommendations for future expansions (e.g., raw audio and lyrics) are provided.

Step-by-Step Methodology

1. Understanding the Available Features

The dataset contains two main components:

1. **Metadata**:
    - Performance title, performance artist, work title, work artist, release year, etc.
    - Useful for capturing high-level structural or contextual similarity.
2. **Pre-Extracted Features**:
    - **Chroma Features** (e.g., chroma_cens, HPCP): Capture harmonic and melodic content (key and harmonization).
    - **MFCC (Mel-Frequency Cepstral Coefficients)**: Represent timbre-related characteristics.
    - **Key Extractor**: Provides key, scale, and key strength.
    - **Madmom Features**: Include tempo, onset, and novelty curves for rhythmic analysis.
    - **Tags**: Provide genre, style, or instrumentation information.

2. Mapping Features to the 8 Metrics

We propose the mapping of the 8 transformation metrics to the features provided:

Metric	Relevant Features	Explanation
Timbre	**MFCC**	Models the spectral envelope, capturing timbral changes due to instrumentation or production.
Tempo	**Madmom tempo**	Captures variations in speed or tempo between original and cover songs.
Timing	**Madmom onsets and novelty curves**	Analyzes rhythmic patterns and timing differences.
Structure	**Metadata (tags, artists, titles, release year)**	Metadata provides context on stylistic and structural similarities.
Key	**Key extractor (key, scale, strength)**	Measures pitch and tonality similarity.
Harmonization	**Chroma Features (HPCP, chroma_cens)**	Captures harmonic relationships and melodic similarities.
Lyrics	**Not available**	Future expansion: add lyrics data to enhance similarity metrics.
Noise	**Tags (genre, instrumentation)**	Tags provide insights into noise-related variations due to production techniques or environments.

3. Feature Engineering

Implementing feature engineering will optimize the provided features to align with the similarity task. We propose the following processes:

1. **Dimensionality Reduction**:
    - The use of **Principal Component Analysis (PCA)** on high-dimensional features (e.g., chroma_cens, HPCP, MFCC) to retain only the most relevant components.
2. **Standardization**:
    - The standardization of numerical features (e.g., MFCC, tempos), to have zero mean and unit variance for consistent scaling.

3. **Key Transposition Invariance**:
   - The normalization of chroma features (HPCP, chroma_cens) to account for key shifts (e.g., transposing all tracks to the same reference key).
4. **Temporal Aggregation**:
   - The aggregation of temporal features (e.g., MFCC, onset curves) into summary statistics (mean, variance) to reduce sequence complexity.
5. **Tag Encoding**:
   - The tag conversion into a numerical format using one-hot encoding for similarity comparisons.

4. Developing the Similarity Metric

The proposed similarity metric combines multiple feature-based comparisons into a single distance measure.

1. **Pairwise Feature Similarity**:
   - **Timbre (MFCC)**:
     - Cosine similarity or dynamic time warping (DTW) on MFCC sequences.
   - **Harmonic Content (Chroma, Key)**:
     - Cosine similarity on chroma features and penalize large differences in key/scale values.
   - **Tempo and Timing (Madmom)**:
     - DTW application on tempo curves and onset features.
   - **Metadata (Tags)**:
     - **Jaccard similarity** use on genre/style tags.
   - **Overall Similarity**:

$$S = w1 \cdot S_{timbre} + w2 \cdot S_{harmonization} + w3 \cdot S_{tempo} + \cdots$$

   - Weights (w1,w2,...w_1, w_2, \ldots) are tuned based on validation data.
2. **Validation**:
   - We propose using the dataset's **WID** and **PID** labels for validation:
     - Cover songs (same WID) should have high similarity.
     - Non-related songs (different WIDs) should have low similarity.

3. **Clustering/Ranking**:
    - Implementing clustering (e.g., K-means) to group songs by similarity.
    - Using ranking metrics like Mean Reciprocal Rank (MRR) to evaluate performance.

5. Evaluation

We propose the evaluation the effectiveness of the similarity metric using:

- **Precision and Recall**: For classifying covers vs. non-covers.
- **MRR**: For ranking covers higher than unrelated songs.
- **Qualitative Analysis**: Review top matches for interpretability.

Recommendations for Future Expansions

1. Incorporation of Raw Audio Data

While the proposed approach is based on pre-extracted features, adding raw audio would enable:

- **End-to-End Learning**: Using deep learning models (e.g., CNNs) to learn representations directly from audio.
- **Fine-Grained Features**: Extracting more nuanced audio features beyond what's pre-extracted.

2. Adding Lyrics Data

Including lyrics for analyzing semantic and stylistic similarity. Tools like **NLTK** or **transformers** (e.g., BERT) can be used for text analysis.

3. Cross-Dataset Validation

Validation of the metric on other datasets to ensure generalizability.

4. Scalability

Optimize feature extraction and similarity computation by employing approximate nearest neighbor search for faster similarity lookups.