MUSIC COPYRIGHT RECOGNITION USING CHROMAPRINTS AND RANDOM FOREST CLASSIFIER

Lý Phúc Thành

Trường Đại học Công nghệ Thông tin - Đại học Quốc gia TP.HCM

Introduction

- Explosion of digital music content demands robust copyright protection, automated detection systems help safeguard creators' rights at scale.
- **Chromaprint** provides fast, reliable audio fingerprinting.
- **Random Forest** offers high-accuracy classification of copyrighted tracks.

Target

- Develop an automatic music copyright detection system using Chromaprints and Random Forest
- Optimize feature extraction for better accuracy and reduced errors in copyright detection.
- Evaluate and improve the Random Forest model for accurate classification of copyrighted music.

Overview

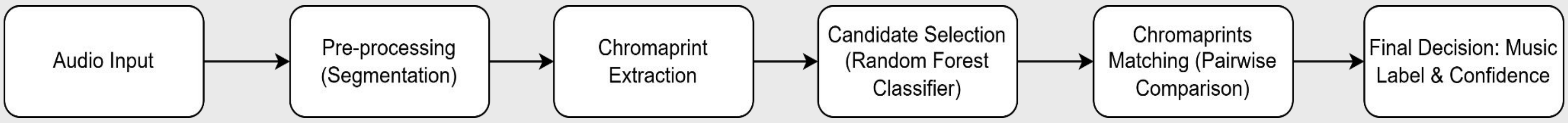

Fig 1: System Architecture

- **Input:** Song or links of song in local or Youtube video (Ex: youtube.com/watch?v=X-yIEMduRXk)
- **Output:** Similarity of the song to registered songs

- **Random Forest Classifier:** Training a RandomForestClassifier on the extracted features
- **Chromaprints:** Preprocessing audio with Essentia to normalize and denoise, generating audio fingerprints

Description

1. Extract Features

- We begin by loading raw audio files and applying preprocessing steps such as normalization, noise reduction, and silence trimming to ensure consistent quality. Using the Essentia library, we then compute spectral and temporal descriptors before generating robust Chromaprint fingerprints for each track. These fingerprints distill each song's unique audio signature into concise feature vectors, ready for classification.

2. Chromaprint Matching

- We begin by loading raw audio files and applying preprocessing steps such as normalization, noise reduction, and silence trimming to ensure consistent quality. Using the Essentia library, we then compute spectral and temporal descriptors before generating robust Chromaprint fingerprints for each track. These fingerprints distill each song's unique audio signature into concise feature vectors, ready for classification.

3. Predicted Youtube Video

- For real-world testing, audio streams from YouTube videos are downloaded via yt-dlp and passed through the same extraction pipeline. The trained Random Forest classifier evaluates each video's fingerprint, outputting a confidence score and predicted label. If the confidence exceeds a predefined threshold, the system flags the video as containing copyrighted content and logs the result for review.

4. RabbitMQ

- To handle large-scale fingerprint extraction efficiently, we integrated **RabbitMQ** as a message broker to orchestrate distributed workloads. Audio preprocessing tasks publish Chromaprint generation jobs to RabbitMQ queues, while multiple worker nodes consume and process these jobs in parallel. This architecture ensures high throughput and fault tolerance: if a worker fails, its unacknowledged messages are re-queued and retried automatically. As a result, the overall pipeline scales seamlessly to thousands of audio files without bottlenecks.

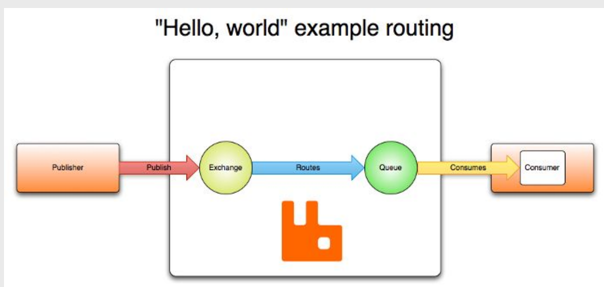

Fig 4: RabbitMQ Architexture

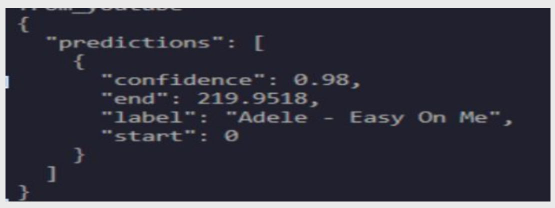

Fig 2: Recognized Output

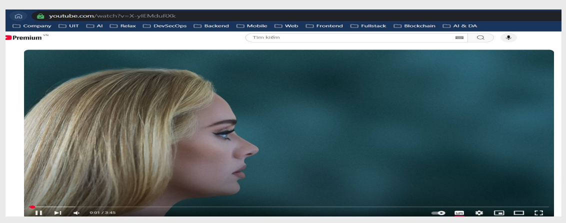

Fig 3: Match with Youtube's song

NII

Tên tác giả chính– Lý Phúc Thành

TEL : 0392000164 Email : thanhlp.19@grad.uit.edu.vn