Chord Estimation from Audio using Hidden Markov Models

Roy Ben Haroosh
Department of Computer Science
Your Institution
roy.benharoosh@post.runi.ac.il
Gal Davidi
Department of Computer Science
Your Institution
gal.davidi@post.runi.ac.il

March 17, 2025

Abstract

Automatic chord recognition is a significant task within Music Information Retrieval (MIR). Hidden Markov Models (HMMs) using beat-synchronous chroma features have proven effective for chord estimation. This report explores the combination of Neural Networks (NN) and Hidden Markov Models (HMM) and investigates pre-trained audio embeddings, such as VGGish and L3net, to enhance acoustic modeling for chord recognition.

1 Introduction

Automatic chord recognition aims to identify and segment chords within audio recordings automatically. Since 2008, it has been a task in the Music Information Retrieval Evaluation eXchange (MIREX), achieving notable success. Hidden Markov Models (HMM) coupled with chroma vectors have established themselves as effective methods. This work evaluates extensions to classical HMM methods by integrating neural network-based acoustic models and leveraging pre-trained embeddings.

2 Related Works

Previous studies have shown HMMs combined with Expectation-Maximization (EM) as robust techniques for chord recognition. Notably, studies like *Chord Segmentation and Recognition using EM-Trained HMM* have highlighted the

efficacy of such models. Recent advancements leverage deep neural networks and deep embeddings, such as Deep Belief Networks (DBNs), OpenL3 embeddings, and VGGish features, significantly improving acoustic modeling quality.

3 Architecture

Our proposed architecture leverages an NN/HMM hybrid approach:

1. **Feature Extraction:** Extract beat-synchronous chroma vectors, potentially augmented by pre-trained embeddings (VGGish/OpenL3).
2. **Acoustic Model:** Neural network-based models trained to map audio embeddings to chord probabilities.
3. **Temporal Model:** Hidden Markov Model integrates temporal dependencies between chord progressions.

The probability of an observed sequence O given a state sequence Q in an HMM is defined by:

$$P(O|Q,\lambda) = \prod_{t=1}^{T} P(o_t|q_t,\lambda) \tag{1}$$

The forward algorithm computes the probability of the partial observation sequence up to time t as:

$$\alpha_t(i) = \left[\sum_{j=1}^{N} \alpha_{t-1}(j)a_{ji}\right] b_i(o_t) \tag{2}$$

where a_{ji} represents transition probabilities and $b_i(o_t)$ emission probabilities.

4 Experiments

We evaluated various configurations of acoustic models:

- Baseline: Classical HMM with chroma features.
- NN-based acoustic models combined with HMM.
- Pre-trained embeddings from VGGish and L3net as input features.

Experiments utilized well-known datasets (e.g., Beatles dataset).

5 Results

Preliminary results indicate:

- Classical HMM with chroma achieved accuracy around X%.
- Neural network acoustic models significantly improved accuracy, reaching Y%.
- Incorporation of pre-trained embeddings (OpenL3, VGGish) further increased accuracy to approximately Z%.

Detailed results are available in the provided notebook and will be continuously updated.

6 Training Data

The quality and composition of training data play a crucial role in the performance of our machine learning model. This section describes the dataset used for training, its characteristics, and the preprocessing steps applied.

6.1 Dataset Overview

Our model was trained on a comprehensive dataset consisting of [describe data source, e.g., "customer transactions collected over a 12-month period" or "publicly available dataset from XYZ repository"]. The dataset contains [number] samples with [number] features, covering [describe what the data represents].

6.2 Data Characteristics

The training data exhibits the following key characteristics:

- **Size:** [Describe the size of the dataset, e.g., number of samples, storage size]
- **Features:** [List and briefly describe the most important features]
- **Class distribution:** [Describe the distribution of target classes if applicable]
- **Temporal coverage:** [Describe the time period covered by the data if relevant]

6.3 Preprocessing Steps

Before training, we applied several preprocessing techniques to enhance the quality of the data:

- **Cleaning:** [Describe how missing values, outliers, and noise were handled]

- **Normalization:** [Describe any normalization or standardization applied]
- **Feature engineering:** [Describe any feature creation, selection, or transformation]
- **Data augmentation:** [Describe any techniques used to expand the dataset if applicable]

6.4 Data Splitting

The dataset was divided into training, validation, and test sets using a [describe splitting strategy, e.g., "70-15-15 random split" or "chronological split"]. This approach ensures that our model evaluation reflects its performance on unseen data while maintaining the statistical properties of the original dataset.

7 Conclusion

Our findings demonstrate the effectiveness of integrating neural network acoustic modeling and pre-trained embeddings with traditional HMMs for chord recognition tasks. Further experimentation and hyperparameter tuning may yield additional improvements.

Code and Media

The code for this project is available at: `https://github.com/caiomiyashiro/music_and_science/tree/master/Chord%20Recognition`

Media and additional results can be accessed here: `https://drive.google.com/drive/folders/1YmfEPtX_QLlpo0sR0kQwFV4-Lz6Sd01W`