Chord Recognition

Brad Girardeau Toki Migimatsu Pranav Rajpurkar

November 17, 2013

1 Introduction

Recognizing chords in popular music is a multifaceted challenge, involving segmenting signals of different instruments, determining the location of chord changes, and performing chord or pitch analysis. This project focuses on the last step of this process by recognizing and classifying chords. A learning system is trained on segmented computer generated chords representing a variety of instruments.

2 Data Generation

Chord data for the learning system is initially created programmatically. The data set consists of each inversion, octave, and major or minor type of every chord, giving 288 examples. A MIDI file containing the pitches for each chord type is generated in Python. Audio data is then created by sythesizing the MIDI files using a given instrument. This allows flexibility in the duration and instrument type of the resulting data set. Training and testing data is generated for several instruments, though for simplicity only one instrument dataset is currently being used.

After exploring results using computer sythesized MIDI chords, non-computer generated audio data will be used. Chords will be recorded played on instruments, and eventually databases of annotated popular music like the McGill billboard corpus will be incorporated. A combination of these datasets will allow creation of a robust, accurate chord recognition system.

3 Feature Extraction

The choice of features greatly influences the success of musical information retrieval algorithms. An audio sample is loaded and divided into frames of a given window size. A feature extraction engine can then compute a variety of metrics or characteristics of each frame. The results can be used independently or concatenated with the other frames from the sample to describe a feature

vector. In this system, the Yaafe python package was used to generate MFCC features. While MFCC are not optimal for chord recognition as MFCC discards much of the tonal pitch data and focuses on timbre, it has been effective in speech recognition as well as other music information retrieval tasks, such as artist or genre classification and pitch detection. As such, MFCC features are useful as part of a baseline system before further feature exploration.

The next step is extracting pitch based features. In particular, the chroma of a pitch refers to the root pitch class that the note belongs to, invarient to octave. For example, the C chroma class contains the C pitch in each octave. Chroma features are ideal for chord recognition because a chord type is also invarient to octave shifts, determined only by the chroma of the contained notes. The Chroma Toolbox is a Matlab toolkit developed by Meinard Müller and Sebastian Ewert that calculates advanced pitch and chroma based features. The Chroma Toolbox inputs WAV audio files and divides them into frames to compute the energy in each MIDI pitch for every frame. It can then calculate the energy in a chroma from the energy in the corresponding pitch bins. More complicated approaches are also supported and may be explored after learning with basic chroma features.

After features are computed for each frame, they can be combined to form a feature vector for a collection of frames or used independently. Training with individual frames does not allow learning about interactions between frames in a given chord, but gives a lower dimensional feature vector, so less data is needed. However, since the window of each frame is short, it is useful to concatenate chroma vectors for several frames into one higher dimensional feature vector. Concatenating the features is chosen as the first approach, but training on individual frames will also be attempted to see if it improves performance.

4 Model Training and Testing

To begin experimenting with classification systems, the chord recognition problem is simplified to the binary classification problem of separating major chords from minor chords. An SVM with RBF kernel trains using features extracted from one instrument's chord dataset. It has high training error, suggesting the data is not seperable using the given features.

References