Automatic Piano Music Transcription

Jianyu Fan Jianyu.Fan.Gr@dartmouth.edu
Qiuhan Wang Qiuhan.Wang.Gr@dartmouth.edu
Xin Li Xi.Li.Gr@dartmouth.edu

Introduction

Our aim is to do automatic transcription converting piano music recording to MIDI files. In other words, it is converting the audio information to visual data. In piano music recordings, audio signal contains information about the note's pitch and duration. The information is represented by symbolic data in corresponding MIDI files. This can be done by very experienced musicians. However, it is a time consuming and painful task. The computer can extract the information of a specific note from the audio signal and learn how to convert them automatically, and then finally, produce the music score.

The wav music recording files contain all the information about audio signals that can be used to extract audio features directly. "The Standard MIDI (Music Instrument Digital Interface) File is a file format that provides a standardized way for sequences to be saved, transported, and opened in other systems" [1]. MIDI file contains symbolic data which has simple representation of music information such as pitch, duration, rhythm, and dynamics. By statistically-based probability learning, the computer can find a best match between the audio note and symbolic expression.

Approach Method

We intend to solve this problem through a trained likelihood model HMM (Hidden Markov Model) to generate a hypothesis for each frame we segment from the audio signal and then construct the search graph out of these hypotheses [2]. In our project, the symbolic notation is the hidden state while the audio signal is the observed state in HMM. We denote the two state sequences as S=$S_1, S_2, \dots, S_k$ and $O = O_1, O_2, \dots, O_k$. S is the hidden sequence with every state being a portion of a note or chord. O is the observation sequence of the signal feature[3].

The first training process is to establish HMM based on the two given state sequences to compute two probabilities: $P(S_{n+1}|S_n)$ and $P(O_n|S_n)$ for all 1<=n+1<=k. With the assumption that S is a Markov chain, the probability of a certain state in the sequence only depends on the preceding state:

$$P(S_{n+1} = s_{n+1}|S_n = s_n, S_{n-1} = s_{n-1}, \dots, S_1 = s_1) = P(S_{n+1} = s_{n+1}|S_n = s_n)$$

HMM assumes that each S_n produces a O_k, and we assume:

$$P(O_n = o_n|S_n = s_n, S_{n-1} = s_{n-1}, \dots, S_1 = s_1) = P(O_n = o_n|S_n = s_n)$$

The second recognition step is searching the hidden state sequence which can most probably produce the given observed sequence based on the criterion of the maximum likelihood. In this part, we are going to apply Viterbi algorithm to look for the maximum probability in HMM model.

$$S^{*(1..k)} = \text{argmax}\, P(S^{(1..k)}|O^{(1..k)})$$

To extract the signal feature, we have decided to segment the audio signal by onset detection[4][5]. Then we would like to spate each segment into three frames and compute the spectrum energy of each frame as F1, and the Chroma feature as F2. Later we will add windows on each spectrum of frame to get harmonic features.

Dataset:

- Training Data— Saarland Music Data(SMD)
  SMD contains audio recordings along with the synchronized MIDI files. The recordings are MP3 files which will be converted to WAV files.
- Test Data—Classical Piano Midi Page Website (http://www.piano-midi.de/)
  This website provides classical piano MIDI files and corresponding MP3 files. It has more than 200 pieces of music, giving a vast database to test the system's performance.

Milestone:

We expect to finish HMM model by the milestone. The two major steps that we will follow in order to establish this model are probability training and the implementation of the searching algorithm, Viberti.

Time Schedule

Date	Task
24/01-03/02	Music Feature Extracted
04/02-09/02	HMM Probability Training
09/02-15/02	Searching Algorithm Implemented
16/02-19/02	Music Transcription Implemented Using HMM

References:

[1]. MIDI. In Wikipedia. Retrived Jan 24,2013, from http://en.wikipedia.org/wiki/MIDI

[2]. C. Raphael, "Automatic transcription of piano music," in Proc. Int. Conf. Music Information Retrieval (ISMIR), Paris, France, 2002.

[3]. Automatic transcription of piano music based on HMM Tracking of Jointly-Estimated Pitches. *16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP*

[4]. Bello, J.P.; Daudet, L.; Abdallah, S.; Duxbury, C.; Davies, M.; Sandler, M.B.; , "A Tutorial on Onset Detection in Music Signals," *Speech and Audio Processing, IEEE Transactions on* , vol.13, no.5, pp. 1035- 1047, Sept. 2005`

[5]. Complex Domain Onset Detection for Musical signals. *Proc. of the 6th Int. Conference on Digital Audio Effects (DAFx-03), London, UK, September 8-11, 2003*