Modeling Music as Markov Chains: Composer Identification

Music 254 Final Report
Yi-Wen Liu (jacobliu@ccrma.stanford.edu)

Professor Eleanor Selfridge-Field
TA: Craig Sapp

11 June 2002 (Edt. 20 October 2006)

Abstract

In this research, it is proposed that music can be thought of as random processes, and music style recognition can be thought of as a system identification problem. Then, a general framework for modeling music using Markov chains is described, and based on this framework, a two-way composer identification scheme is demonstrated. The scheme utilizes the Kullback-Leibler distance as the metric between distribution functions, and it is shown that under the condition when the marginals are identical, the scheme gives maximum likelihood identification. Experiments of composer identification are conducted on all the string quartets written by Mozart and Haydn, and the results are documented and discussed.

1   Motivation

The goal of this research is to develop a model that enables a computer to symbolically examine a piece of classical music which is never exposed to it, and tell who is the composer. As a classical music fan myself, I observed that when we listen to unknown music, we almost always like to classify the style of the music. Actually, we even make guesses on who is the composer, and a person with higher accuracy is often considered as a better listener. While in this research, it is avoided to address on the controversy of quantifying how well a person listens to music, I am still curious if music appreciation, and composer identification in particular, can be statistically learned by a computer.

I claim that music can be dealt with as random processes. The claim is based on the fact that music consists of time-sequences of musical events, such as notes, chords, dynamics, rhythmic patterns, etc. It is already known that other types of time-sequences, such as stock prices, are quite successfully modeled by random processes. Therefore, I claim that the process of musical composition can be modeled as a realization of an underlying random process, and

the underlying random process *is* what we fuzzily call "musical styles". If we further hypothesize that each composer has his/her unique style, then the composer identification problem becomes what is known by engineers as a "system identification" problem.

Among the many system identification techniques that already exist, this research explores the Markov chain model due to its simplicity. Due to the simple assumption that the future is conditionally independent to the past given the present, a Markov chain can be characterized by its state-transition probability matrix. However, we have to point out that the assumption is ignorant of macroscopic musical structures, and it is therefore valid to question the validity of Markov modeling of music in general.

Although, in this research, I am awared of the general validity issue, I remain optimistic that Markov models may be sufficiently sophisticated to teach a computer to appreciate music, as far as composer identification is concerned. The rest of the paper is organized as follows. Section 2 describes the general methods. Section 3 documents the experiments that have been done so far. Results are discussed in Section 4, and possible future directions are pointed out in Section 5.

2   Methods

To describe how music is modeled by Markov chains, let's first define the terminologies and notations –

A first order, discrete time Markov chain $\mathcal{C}$ is a random walk $X_t, t = 1, 2, 3, ...$, in a state space $\mathcal{S} = \{s_1, s_2, .., s_N\}$ according to a $N \times N$ state-transition matrix $P_{i,j} = p\left(X_{t+1} = s_j | X_t = s_i\right)$, where $P_{i,j}$ denotes the element on the ith row and j^{th} column, and $p(\cdot|\cdot)$ is the usual notation of the conditional probability distribution function.

Mathematically, it suffices to say that a Markov chain $\mathcal{C}$ is characterized by its state-transition matrix P, up to one-to-one mappings between homeomorphic state spaces. One can even sloppily write C $= P$. However, how the transition matrix actually means depends upon how the state-space is defined. For example, if the (pentatonic) state space is defined as $\mathcal{S} = \{\text{C4}, \text{D4}, \text{E4}, \text{A4}, \text{G4}\}$, then $P_{1,5} = 0.3$ reads "the next note is G4 thirty percents of the times when the current note is C4."

Since higher-order Markov chains are beyond the scope of this research, in the rest of this paper, "Markov chains" means first-order ones unless otherwise mentioned.

2.1   Computational models of music styles

The following steps describe how to build up a Markov chain model for a music style.

- Step 1: Define the repertoire of that style.

- Step 2: Encode all the works of that repertoire.

- Step 3: Define the state space $\mathcal{S}$ that consists of musical events.

- Step 4: From all the encoded works available in that repertoire, calculate the matrix P of conditional event-transition probabilities.

- Step 5: Let $\mathcal{C}$ be the Markov chain corresponding to $\{P, S\}$. Call $\mathcal{C}$ a Markov chain model for that particular style of music.

Step 1 can be a pain, since musical styles are often of a vague notion to human beings. Is Beethoven's Moonlight Sonata classical or romantic? Here, it is supposed that the human researcher somehow clearly define the repertoire, anyway. Step 2 is not quite straightforward, either. It all depends on what musical features are relevant. Data entry is another practical difficulty in this step, because it takes forever if the repertoire is huge. Step 3 involves defining an interesting prominent feature associated with that style, and quantifying that feature in a clever way. The state space consists of all the possible values to which the feature is quantified. In step 4, the transition matrix is obtained by normalizing the histogram of state-transition instances in all the works of that repertoire.

2.2 Composer identification by comparing Kullback-Leibler distances

Assuming that composer A and composer B each has a unique style that can be recognized in a certain repertoire, one can select a feature that looks interesting, define the state space, and build up the Markov chain models $\mathcal{C}_{A} = P_{A}$, $\mathcal{C}_{B} = P_{B}$, for composer A and B, respectively. Based on these two Markov models, for each given unknown work U, two-way composer identification can be achieved by the following hypothesis test,

$$\texttt{My\_Author} = \begin{cases} A, & D(P^{U}||P^{A}) < D(P^{U}||P^{B}); \\ B, & \text{otherwise}, \end{cases}$$

where P_{U} is the transition matrix of the repertoire that contains only the unknown work U, `My_Author` is the composer that is more likely than the other one to have written the work, and D is a distance metric on the space of $N \times N$ probability matrice. In this research, we use the Kullback-Leibler distance metric,

$$D(P^{(1)}||P^{(2)}) = \sum_{i=1}^{N} \sum_{j=1}^{N} P_{i,j}^{(1)} \log_{2} \frac{P_{i,j}^{(1)}}{P_{i,j}^{(2)}}$$

Although the Kullback-Leibler distance metric is not symmetric and does not obey the trangular inequality, it is useful to interpret of the metric as the distance between distributions [1]. Accepting this interpretation, the above

hypothesis test reads "if the transition matrix of the unknown work is closer to that of composer A's repertoire, then it is more likely to have been written by A." This hand-waving statement can be justified by the following theorem.

THEOREM (likelihood ratio interpretation) For the two-way composer identification test, if the two marginal distributions are identical,

$$P_i^{\mathrm{A}} = \sum_j P_{i,j}^A = \sum_j P_{i,j}^B = P_i^{\mathrm{B}}, \quad \forall i,$$

then, the difference of the two distances has a likelihood ratio interpretation,

$$2^{D(P^{\mathrm{U}}||P^{\mathrm{A}}) - D(P^{\mathrm{U}}||P^{\mathrm{B}})} = \left(\frac{p(\mathrm{U}|\mathrm{B})}{p(\mathrm{U}|\mathrm{A})} \right)^{1/L}$$

where $p(\mathrm{U}|\mathrm{A})$ is the *a priori* probability that U$= \{U_1, U_2, ...\}$ is generated according to P_{A}, and L is the length of U.

Proof:

$$
\begin{aligned}
D(P^{\mathrm{U}}||P^{\mathrm{A}}) - D(P^{\mathrm{U}}||P^{\mathrm{B}}) &= \sum_i \sum_j P_{i,j}^{\mathrm{U}} \log_2 \frac{P_{i,j}^{\mathrm{U}}}{P_{i,j}^{\mathrm{A}}} - \sum_i \sum_j P_{i,j}^{\mathrm{U}} \log_2 \frac{P_{i,j}^{\mathrm{U}}}{P_{i,j}^{\mathrm{B}}} \\
&= \sum_i \sum_j P_{i,j}^{\mathrm{U}} \log_2 \frac{P_{i,j}^{\mathrm{B}}}{P_{i,j}^{\mathrm{A}}} \\
&= \sum_i \sum_j \frac{n_{i,j}}{L} \log_2 \frac{P_{i,j}^{\mathrm{B}}}{P_{i,j}^{\mathrm{A}}} \\
&= \frac{1}{L} \sum_i \sum_j \log_2 \left(\frac{P_{i,j}^{\mathrm{B}}}{P_{i,j}^{\mathrm{A}}} \right)^{n_{i,j}}
\end{aligned}
$$

where $n_{i,j}$ denotes how many times the transition between i^{th} and j^{th} states occurs in U. Therefore,

$$2^{D(P^{\mathrm{U}}||P^{\mathrm{A}}) - D(P^{\mathrm{U}}||P^{\mathrm{B}})} = \left(\prod_{t=1}^{L-1} \frac{P_{U_t, U_{t+1}}^{\mathrm{B}}}{P_{U_t, U_{t+1}}^{\mathrm{A}}} \right)^{1/L} \tag{1}$$

$$= \left(\frac{P_{U_1}^{\mathrm{B}}}{P_{U_1}^{\mathrm{A}}} \prod_{t=1}^{L-1} \frac{P_{U_t, U_{t+1}}^{\mathrm{B}} / P_{U_t}^{\mathrm{B}}}{P_{U_t, U_{t+1}}^{\mathrm{A}} / P_{U_t}^{\mathrm{A}}} \right)^{1/L} \quad \text{(identical marginals)} \tag{2}$$

$$= \left(\frac{p(\mathrm{U}|\mathrm{B})}{p(\mathrm{U}|\mathrm{A})} \right)^{1/L} \tag{3}$$

The above theorem states that, for the purposes of composer identification, if the marginals of are identical and hence don't help, then

- it is necessary to apply Markov modeling;

- the above mentioned hypothesis test is a maximum likelihood test.

For example, if two composers use as frequently each of the notes as one another but concatenate them differently, then Kullback-Leibler distance is the right metric to choose for the purposes of composer identification.

3   Experiments and discussion

3.1   String quartets: Mozart vs. Haydn

Markov chain models are built up for Mozart's and Haydn's scale degree class transitions. The repertoire consists of all of their string quartets, 100 movements from Mozart and 212 movements from Haydn. The scale degree is defined relative to the tonics, i.e., all works are transposed into the same key. Also, for the sake of reducing the dimensionality of transition matrice, the system does not distinguish between the same degree class at different octaves. The system does not distinguish between major keys and minor keys, either. Then, the two-way identification tests based on Kullback-Leibler distances were conducted for each of the 4 voices. The tests were conducted in a bootstrap manner. In other words, it is always assumed that the computer has been exposed to the all the works in the repertoire except the one to be identified.

Part	Mozart	Haydn
ViolinI	**68.0%**	**64.2%**
ViolinII	58.0%	**64.2%**
Viola	**61.0%**	53.8%
Cello	57.0%	52.8%

Table 1: Mozart vs. Haydn identification tests.

Table 1 shows the result of two-way identification tests. The boldfont indicates data that are statistically significant. To qualify for statistical significance, a recognition rate has to be higher than $\mu + 2\sigma$ of random flips of fair coins, where μ is the mean, 50%; and σ is the standard deviation, which is inversely proportional to the square root of the number of flips, according to the central limit theorem.

As a controlled experiment, repertoires are defined as two randomly assigned, and mutually exclusive lists of quartets. Then, the two-way style identification tests are conducted, and the result is shown in Table 3.1. Each of the data shows an average of ten runs.

Human listening tests are conducted via web survey. A user is played, the MIDI piano version of, a quartet from either composer by equal chance, and is asked to identify the composer. As of June 11, 2002, web users average 59.0% of accuracy in 1865 attempts. How experienced the users are with the repertoire

Part	Random A	Random B
ViolinI	42.6%	56.1%
ViolinII	44.5%	53.4%
Viola	47.4%	51.9%
Cello	43.1%	53.4%

Table 2: Random repertoires identification tests.

ranges from novices of classical music to string players that have played some of the works.

4 Future directions

- Tests on other features, such as intervallic values, rhythmic patterns, harmonic rhythms, chord progression, etc.

- Combination of opinions from multiple agents.

- Tests on other repertoires that are similar to human beings, such as Schumann's vs. Schubert's lieders, J.S. Bach's vs. Handel's choral music, etc.

5 Conclusion

We proposed a scheme for modeling music using Markov chains and applied it to two-way composer identification. Experiments were conducted by examining the scale degree transition of Haydn's and Mozart's string quartets, and the results showed that, as far as the recognition rate is concerned, the computer's performance can be statistically significant.

References

[1] T. Cover and J. Thomas, *"Elements of information theory"*, Wiley Interscience, New York (1991)