BayesianBand: Jam Session System based on Mutual Prediction by User and System

Tetsuro Kitahara[12], Naoyuki Totani[1],
Ryosuke Tokuami[1], and Haruhiro Katayose[12]

[1] School of Science and Technology, Kwansei Gakuin University
2-1 Gakuen, Sanda 669-1337, Japan
{t.kitahara,katayose}@kwansei.ac.jp
http://ist.ksc.kwansei.ac.jp/~kitahara/
[2] CrestMuse Project, CREST, JST, Japan

Abstract. One kind of pleasure that jam sessions bring is deciding a melody or an accompaniment while mutually predicting what the other participants are going to play. We propose a jam session system, called *BayesianBand*, which provides this kind of musical pleasure through sessions with computers. With this system, the chord progression in a session is not fixed in advance but rather is determined in real time by predicting the user's melody. The user, while improvising, is also expected to predict the chord progression generated by the system; accordingly, a cooperative jam session based on the mutual prediction will be achieved. To build this system, we constructed a model for melody prediction and chord inference based on a Bayesian network.

1 Introduction

The entertaining quality of music resides in the fact that music can be partly, but not fully, predicted. Predictability is indispensable for listeners' understanding of a piece of music, but if they can *completely* predict it, they cannot enjoy it. Musical pieces composed by professional musicians are therefore organized so as to achieve a satisfying tradeoff between predictability and unpredictability. Shimojo formed a hypothesis that predictability (he calls it *congruency*) in music psychologically rewards the listener for successful prediction based on his or her internal model of the music, while unpredictability (he calls it *novelty*) also brings psychological rewards, in this case as a result of the listener's detection of new information that enables his or her internal model of the music to be modified [1].

In jam sessions, these two kinds of enjoyment play an important role. During a jam session, each participant determines the melody or accompaniment to be played, while predicting what the other participants will play. When the musicians' predictions succeed and their performances sound harmonious, congruency-based pleasure (psychological reward) is obtained. When the prediction fails, but the performance nevertheless sounds harmonious, participants may attain the novelty-based psychological reward or enjoyment. The goal of

our study is to provide these two kinds of enjoyment through jam sessions with computers.

In this paper, we propose a jam session system, called *BayesianBand*, in which the user and system mutually predict each other's performance. The principal feature of this system is that the chord progression, rather than being decided in advance, is decided by the system in real time. The user determines and plays the main melody by predicting what chord the system will generate in the next measure, while the system determines the next chord by predicting what melody the user will play next. Through this mutual prediction, which often succeeds but sometimes does not, the user can obtain the two kinds of musical pleasure described by Shimojo [1].

2 Technical Requirements and Related Work

BayesianBand is a jam session system that determines a chord progression in real time by predicting the user's melodies. The purpose of this system is to provide the user with both the congruency- and novelty-based enjoyment through jam sessions. To obtain the congruency-based enjoyment, the input (the user's melody and previous/current chords) and the output (the subsequent chord) should have a causality. For this reason, we do not introduce any randomness into the determination of the output from the input, even though this is a common approach to maintaining novelty[2]. In addition, this causality should be acquired by users through jam sessions and thus should be consistent and musically appropriate. If the causality is completely immobilized, however, users may discover the input/output relationships completely and quickly become bored with the jam session. The causality should therefore always evolve. To summarize, causality between input and output should be (1) deterministic (not involving any random process), (2) musically appropriate, and (3) always evolving, in order to attain both congruency and novelty.

To fulfill these three requirements, in BayesianBand, we use a probabilistic model for melody prediction and chord inference, as follows:

1. The chord having the maximum likelihood, given an input, is always determined.
2. The probabilistic model is trained with existing pieces of music.
3. The probabilistic model is incrementally updated to adapt to the user's melodic tendency.

Various jam session systems, described in previous studies [3, 4], have been developed, but most of these assumed that the chord progression is fixed in advance. Aono et al. [5] developed a jam session system that did not assume a fixed chord progression. When the user plays a chord progression (the system judges so if more than three notes are played simultaneously), the system automatically recognizes it and considers it to be repeated. This system therefore did not aim to determine the chord progression by predicting the user's performance.

Melody predictions have also been widely attempted. Conklin [6] developed a melody prediction system by regarding melodies as Markov chains. Pachet [2] developed a system, called the *Continuator*, that generates a sequence of notes that can follow the melody played by the user. This system learns the user's melodies as a Markov tree structure and, when the user plays a melody, recursively generates the following note, based on this structure. There have thus been various studies of Markov-based melody prediction, but they did not model the horizontal dependency of the chord progressions behind melodies.

Harmonization, which aims to give a chord progression to a melody, is also an important topic and has been widely studied. Kawakami et al. [7] proposed a method for harmonization using a hidden Markov model (HMM) in which the melody and chord progression are modeled as observed and hidden variables, respectively. These studies, however, assume that the whole input melody is referable from beginning to end; they do not aim at harmonization for the future's melody by predicting it.

3   System Overview and Algorithm

The main functions of BayesianBand are (1) the chord determination by prediction of the user's melody and (2) incremental updating of the prediction model. In this section, we describe their algorithms used to generate these functions after providing the system overview.

3.1   Problem Statement

The input is the user's melody; the output is a chord progression. For simplicity, the timings of chord changes are limited to occurrence at the beginning of each measure. The input is a monophonic melody, the key is given, and no key modulation occurs. The first chord is the tonic chord of the given key. Due to a small amount of training data, the target chords are limited to the seven diatonic chords.

3.2   System Overview

The system overview of BayesianBand is shown in Fig. 1. Because the initiative for tempo control is governed by the system, the accompaniment is automatically performed with a constant tempo. When the user presses a key on the MIDI keyboard, the system predicts the next note and infers the next chord. This process is repeated for each keystroke; accordingly, the chord inference result is updated after each keystroke. Immediately before changing the measure, the chord having the maximum likelihood is determined as the next chord. In parallel, the incremental update of the prediction model is performed at each keystroke.
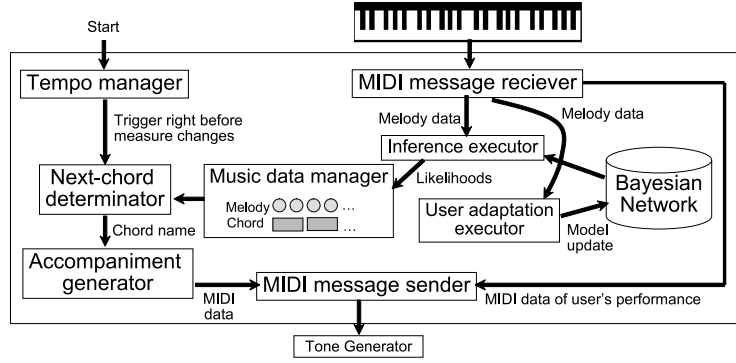
Fig. 1. System overview of BayesianBand.

n_i	12 note names
c_i	7 diatonic chords

Fig. 2. Bayesian network used for melody prediction and chord determination.

3.3 Algorithm for Melody Prediction and Chord Determination

Here we deal with the problem of inferring the most likely subsequent chord c_{t+1} of a chord progression $\boldsymbol{c} = (c_1, \cdots, c_t)$ by predicting the next note n_{t+1} of a given melody $\boldsymbol{n} = (n_1, \cdots, n_t)$. In general, a melody and a chord progression have different sequential causalities, $p(n_{t+1}|\boldsymbol{n})$ and $p(c_{t+1}|\boldsymbol{c})$, and simultaneous elements in the melody and chord progression also have a causality, described as $p(c_t|n_t)$. For simplicity, the sequential causalities $p(n_{t+1}|\boldsymbol{n})$ and $p(c_{t+1}|\boldsymbol{c})$ are approximated by trigram models, described as $p(n_{t+1}|n_{t-1}, n_t)$ and $p(c_{t+1}|c_{t-1}, c_t)$, respectively.

Using these sequential and simultaneous causalities, the relationship between a melody and a chord progression can be described as the Bayesian network shown in Fig. 2. In general, Bayesian networks should be singly connected because when this is the case, a low-complexity algorithm for probability calculation can be applied. To make a singly-connected network, the dependencies unrelated to the nodes n_{t+1} and c_{t+1} are omitted, since the values of the other nodes have already been observed or determined.

The inference process is performed at each keystroke. Once a key is pressed, the observed note names are set to n_{t-1} and n_t and the determined chord names to c_{t-1} and c_t. Then the inference is executed: the probability densities for the nodes n_{t+1} and c_{t+1} are calculated using Pearl's method [8]. After this process is repeated at each keystroke, the chord having the maximum likelihood is determined as the next chord immediately before changing the measure.

3.4 Algorithm for Incremental Model Update

Incremental model update aims not only to retain novelty in chord determination but also to improve the accuracy of melody prediction by adapting the model to the user. The basic method of achieving this is to calculate the conditional probability $p(n_{t+1}|n_{t-1}, n_t)$ as a weighted mean of the probability precalculated from a corpus and the probability calculated online from the user's performance. As the number of notes performed by the user increases, the weights are gradually changed so that the latter's weight becomes larger. Specifically, the conditional probability $p(n_{t+1}|n_{t-1}, n_t)$ is calculated using the following equation:

$$p(n_{t+1}|n_{t-1}, n_t) = \frac{p_0(n_{t+1}|n_{t-1}, n_t) + \alpha \left\{ \log N(n_{t-1}, n_t) \right\} \frac{N(n_{t-1}, n_t, n_{t+1})}{N(n_{t-1}, n_t)}}{1 + \alpha \log N(n_{t-1}, n_t)},$$

where $p_0(n_{t+1}|n_{t-1}, n_t)$ is the probability calculated from a corpus, $N(n_{t-1}, n_t)$ is the frequency that the user played n_{t-1} and n_t in this order, $N(n_{t-1}, n_t, n_{t+1})$ is the frequency that the user played n_{t-1}, n_t, and n_{t+1} in this order, and α is a constant.

4 Implementation and Trial Use

4.1 Implementation

We implemented a prototype system of BayesianBand using Java. We used Crest-MuseXML Toolkit[3] for implementing the overall framework and Weka[4] for learning and using the Bayesian network. For learning the Bayesian network, we used 415 pieces of standard jazz music (pairs of melodies and chord progressions).

4.2 Results of Trial Use

The first author used the implemented prototype system. After repeating a jam session several times, he understood rough trends in chord progressions generated by the system to some extent and reflected his chord prediction in his improvisation. He felt pleasure when the predicted chord was actually played by the system. When his chord prediction failed, his melody and the generated accompaniment often sounded unharmonious, but in some cases they sounded harmonious. In those instances, he felt novelty. Thus, BayesianBand to some extent successfully provided a trial user with the two kinds of enjoyment.

Predicting the next chord while playing was enjoyable, like a game in itself. This type of pleasure cannot be provided by jam sessions in which the chord progression is fixed in advance or is determined at random.

An excerpt of the melodies played and the prediction results are shown in Fig. 3. When the three best candidates for each note prediction were considered, the notes immediately following half of the played notes were successfully predicted.

---

[3] http://www.crestmuse.jp/cmx/
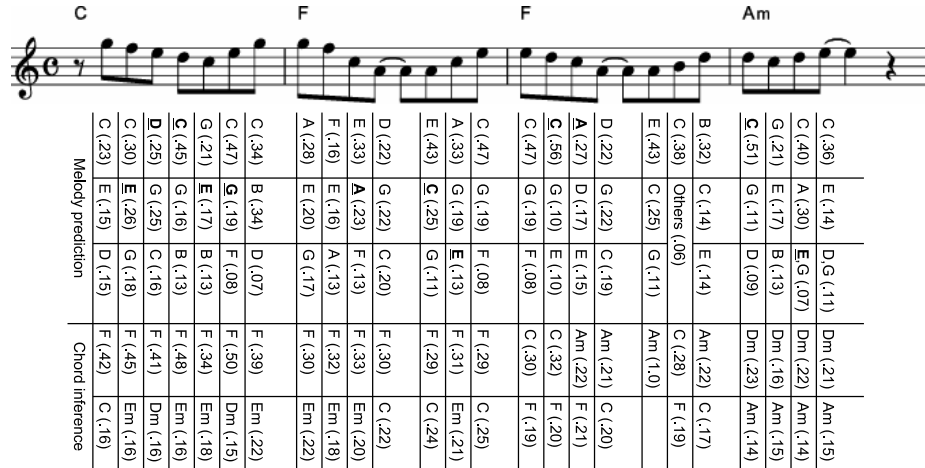[4] http://www.cs.waikato.ac.nz/ml/weka/

Fig. 3. Example of performed melodies and its melody-prediction and chord-inference results obtained using BayesianBand. The values in parentheses are the likelihoods; the boldfaced characters represents the names of the notes that were actually played.

Chord symbols above the staff: C   F   F   Am

Melody prediction				Chord inference	
C (.23)	E (.15)	D (.15)		F (.42)	C (.16)
C (.30)	**E** (.26)	G (.18)		F (.45)	Em (.16)
D (.25)	G (.25)	C (.16)		F (.41)	Dm (.16)
C (.45)	G (.16)	C (.16)		F (.48)	Em (.16)
G (.21)	E (.17)	B (.13)		F (.34)	Em (.18)
C (.47)	**G** (.19)	F (.08)		F (.50)	Dm (.15)
C (.34)	B (.34)	D (.07)		F (.39)	Em (.22)
A (.28)	E (.20)	G (.17)		F (.30)	Em (.22)
F (.16)	E (.16)	A (.13)		F (.32)	Em (.18)
E (.33)	**A** (.23)	F (.13)		F (.33)	Em (.20)
D (.22)	G (.22)	C (.20)		F (.30)	C (.22)
E (.43)	C (.25)	G (.11)		F (.29)	C (.24)
A (.33)	G (.19)	**E** (.13)		F (.31)	Em (.21)
C (.47)	G (.19)	F (.08)		F (.29)	C (.25)
C (.47)	G (.19)	F (.08)		C (.30)	F (.19)
A (.27) C (.56)	D (.17)	E (.15)		F (.32)	F (.20)
D (.22)	G (.22)	C (.19)		Am (.22)	F (.21)
E (.43)	C (.25)	G (.11)		Am (1.0)	
C (.38)	Others (.06)			C (.28)	F (.19)
B (.32)	C (.14)	E (.14)		Am (.22)	C (.17)
C (.51)	G (.11)	D (.09)		Dm (.23)	Am (.14)
G (.21)	E (.17)	B (.13)		Dm (.16)	Am (.15)
C (.40)	A (.30)	E,G (.07)		Dm (.22)	Am (.14)
C (.36)	E (.14)	D,G (.11)		Dm (.21)	Am (.15)

5 Conclusion

In this paper, we proposed a new jam session system, called *BayesianBand*, in which the user and the system mutually predict each other's performance. The topic of human-system jam sessions based on the mutual prediction is interesting as a target domain for research involving man-machine collaboration. In the future, we plan to investigate through long-term experiments how humans and systems collaboratively create music.

References

1. Shimojo, S.:   Research plan for Shimojo implicit brain function project. (http://impbrain.shimojo.jst.go.jp/jpn/about_jpn.html) in Japanese.
2. Pachet, F.: The Continuator: Musical interaction with style. In: Proc. ICMC. (2002)
3. Nishijima, M., Watanabe, K.: Interactive music composer based on neural networks. In: Proc. ICMC. (1992) 53–56
4. Goto, M., Hidaka, I., Matsumoto, H., Kuroda, Y., Muraoka, Y.: A jam session system for interplay among all players. In: Proc. ICMC. (1996) 346–349
5. Aono, Y., Katayose, H., Inokuchi, S.: An improvisational accompaniment system observing performer's musical gesture. In: Proc. ICMC. (1995) 106–107
6. Conklin, D., Witten, I.H.: Multiple viewpoint systems for music prediction. J. New Music Res. **24** (1995) 51–73
7. Kawakami, T., Nakai, M., Shimodaira, H., Sagayama, S.: Hidden markov model applied to automatic harmonization of given melodies. In: IPSJ SIG Notes. 99-MUS-34 (2000) 59–66 in Japanese.
8. Pearl, J.: Probabilistic Reasoning in Intelligent Systems. Morgan Kaufmann (1988)