

テクニカルノート

BayesianBand : ユーザとシステムが相互に予測し合う ジャムセッションシステム

北原 鉄朗^{†1,†2} 戸谷 直之^{†1}
徳網 亮輔^{†1} 片寄 晴弘^{†1,†2}

本稿では、ユーザとシステムが相互に演奏を予測することで合奏を進行させるジャムセッションシステム「BayesianBand」について述べる。このシステムでは、コード進行をあらかじめ決めておらずに、ユーザの演奏をリアルタイムに予測してコード進行を決定する。ユーザはシステムが決定するコード進行を予測しながら即興演奏をすることで、相互予測に基づくジャムセッションを楽しむことができる。

BayesianBand: Jam Session System Where User and System Mutually Predict Each Other's Performance

TETSURO KITAHARA,^{†1,†2} NAOYUKI TOTANI,^{†1}
RYOSUKE TOKUAMI^{†1} and HARUHIRO KATAYOSE^{†1,†2}

This paper describes a jam session system, called *BayesianBand*, where the user and system mutually predict each other's performance. Not assuming that the chord progression is determined in advance, this system predicts the user's melody and then determines the chord progression in real time. By improvising while predicting the chord progression generated by the system, the user can enjoy a jam session based on the mutual prediction.

^{†1} 関西学院大学理工学部

School of Science and Technology, Kwansai Gakuin University

^{†2} 科学技術振興機構戦略的創造研究推進事業 CrestMuse プロジェクト

CrestMuse Project, CREST, JST

1. はじめに

音楽の楽しさは、予測可能性とそこから逸脱のバランスにあるといわれている¹⁾。音楽がまったく予測できないものであればその音楽は理解不可能であるが、完全に予測可能であればつまらない。音楽は両者の妥協点を実現すべく構成されているという考え方である。下條は、予測可能性を「親近性」、そこから逸脱を「新奇性」と呼び、親近性が快なのは内部モデルが的中したことへの報酬、新奇性が快なのは内部モデルを修正するための情報を検出したことへの報酬であるとする仮説を提唱した¹⁾。

ジャムセッションにおいても、この2種類の快は重要な役割を担う。各奏者は、相手が次にどのような演奏をするのかを予測しながら自らの演奏内容を決定する。予測と実際の演奏内容が一致し、自らと相手の演奏内容が音楽的に調和したとき、親近性による快の感情を覚える。一方、予測と実際の演奏内容が一致しなかったにもかかわらず、音楽的にそれほど不自然ではないとき、新鮮さ（新奇性による快）を覚える。本研究の目的は、この2種類の快によるジャムセッションの楽しさを、人と計算機によるセッションでも実現することにある。

この目的を実現するため、我々は、ユーザとシステムの相互予測に基づくジャムセッションシステム「BayesianBand」の開発を進めている。本システムでは、ユーザが主旋律を、システムが伴奏を担当する。コード進行は事前に決めておらず、システムがユーザの演奏を予測して決定する。ユーザは、システムが次にどのコードを演奏するかを予測して主旋律を演奏し、システムはユーザがどのような旋律を演奏するかを予測して次のコードを決定する。この相互予測が合ったり合わなかったりすることで、ユーザは上述の2種類の快を得て合奏を楽しむことができる。

2. BayesianBand : ユーザとシステムの相互予測に基づくジャムセッションシステム

BayesianBand は、ユーザが主旋律を演奏し、システムがユーザの演奏を予測してコード進行をリアルタイムに決定するジャムセッションシステムである。BayesianBand の目的は、ユーザに、システムとの相互予測に基づくジャムセッションを通じて、上で述べた2種類の快を提供することにある。そのため、原則として、相手の演奏を予測して自分の演奏内容に反映させられるような、一定の即興演奏スキルを持つユーザをターゲットとする。

この2種類の快のうち、親近性を担保するには、入力（ユーザの主旋律と現在までのコード進行）と出力（次のコード）とに因果関係がなくてはならない。音楽システムでは、新

奇性を担保するために出力の決定過程にランダム性を取り入れることがしばしば行われる²⁾が、本研究では入出力間の因果関係を希薄にする可能性があると考え、このようなアプローチは採用しない。また、その因果関係は、即興演奏経験者がジャムセッションを通じて見出し得るものでなくてはならない。そのためには、因果関係に一貫性があり、かつ、音楽的に妥当である必要がある。

しかし、その因果関係が固定化されていると、いずれユーザは因果関係を完全に発見し、飽きてしまう。新奇性が担保されるには、因果関係が固定化されないことが重要である。

以上より、親近性と新奇性の両方を提供するには、入力と出力の因果関係が、(1) ランダム性を持たない、(2) 音楽的に妥当である、(3) つねに変化することが重要であると考えられる。

本研究では、これらの3つの要件を満たすため、確率モデルに基づく旋律予測・コード推論手法を検討する。確率モデルでは、結果(次のコード)が、原因(ユーザの主旋律と過去のコード進行)に依存した確率分布に従って生成されると考える。確率モデルを以下のように用いることで、3つの要件を満たすことができる。

- (1) 与えられた原因に対してつねに最尤の結果を出力することにすることとし、決定過程にランダム性は導入しない。
- (2) 主旋律とコード進行とにおいて音楽的に妥当な関係を得るため、確率分布を既存の楽曲集から学習する。
- (3) 確率分布を入力する度に逐次的に更新することで、旋律とコードの因果関係を徐々に変化させる。なお、確率分布の逐次更新においても、ランダム性は導入しない。

本研究は、ジャムセッションシステム、旋律予測、和声づけといった様々な分野と関連がある。以下、各分野から見た本研究の位置付けを論ずる。

ジャムセッションシステム ジャムセッションシステムは様々な研究がなされてきたが、コード進行は既知とする場合が多かった^{3),4)}。一方、青野らはコード進行をあらかじめ既知としないジャムセッションシステムを開発した⁵⁾。しかし、ユーザの主旋律に合わせてシステムが自らの音楽知識を基にコード進行を決めるのではなく、ユーザがコードを演奏していると判断(同時に3音以上打鍵された場合にそう判断する)されたときに、そのユーザの演奏からコードを認識し、以後のコード進行とするという処理を行っている。

旋律予測 旋律予測も以前より様々な研究がなされてきた。Conklinらは、旋律をマルコフ連鎖と見なして予測するシステムを実現した⁶⁾。松尾らは、直近の音系列パターン(短期記憶に相当)から類似する旋律を探索することにより予測を行う計算モデルを提案した⁷⁾。Pachetは、与えられた旋律に後続する旋律を生成するシステム Continuator を

開発した²⁾。このシステムでは、旋律は木構造を持つマルコフ連鎖として学習され、マルコフ連鎖の状態遷移確率を重みとしてランダムに木構造を探索することで後続旋律を生成する。このように旋律をマルコフ連鎖などとしてモデル化して予測を行う研究は数多く存在するが、主旋律とコード進行の両方の時間的因果関係を扱った例は見られない。和声づけ 与えられた主旋律にコード進行を付与する研究は数多く存在し、たとえば、主旋律を観測変数、コード進行を潜在変数と見なして隠れマルコフモデル(HMM)により最尤のコード進行を推定する研究⁸⁾などがある。しかし、これらの研究は、旋律が最初から最後まで参照可能であることを前提としており、本研究のように未来の演奏内容を予測してそれに対してコードを付与するものではなかった。

上記のいずれにもあてはまらないが本研究と深い関連がある研究に「豊次郎」⁹⁾がある。「豊次郎」は自動伴奏システムの一つで、様々な主旋律と伴奏の組をデータベースとして保持し、ユーザの旋律をデータベースからリアルタイムに探索し、それに合った伴奏を出力する。データベースにユーザの旋律と十分近い旋律がない場合は段階的に抽象化することで、未知の楽曲にもある程度対応できるとしている。しかし、本研究ではユーザは即興演奏を行うことを想定しているため、事例探索型よりも確率モデルに基づく予測の方が適していると考えられる。また、事例探索型ではモデルの逐次更新も困難である。

3. BayesianBand の実現方法

BayesianBand において主たる課題は、確率モデルに基づく旋律の予測とそれに合うコードの推論、およびそのモデルの逐次更新である。以下、その実現方法を述べる。

3.1 問題設定

入力はユーザが弾いた旋律、出力はコード名である。コードは小節の先頭で変わるものとする。簡単のため、入力旋律は単旋律とし、調は既知、転調はないものとする。最初のコードは必ず指定された調におけるトニックコードとする。なお、現実装では学習データの規模を考慮し、対応するコードを7種類のダイアトニックコードに限定しているが、原理的に提案手法の適用範囲がダイアトニックコードに限定されるわけではない。

3.2 処理の流れ

BayesianBand の処理の流れを図1に示す。本システムは主に、ユーザの演奏情報を受け取って次の小節のコードを推論する部分(図中のMIDIメッセージ受信部から音楽データ管理部)と、コード推論結果をもとに伴奏MIDIデータを生成して演奏する部分(図中のテンポ管理部からMIDIメッセージ送信部)の2つに分かれ、これらが別スレッドとし

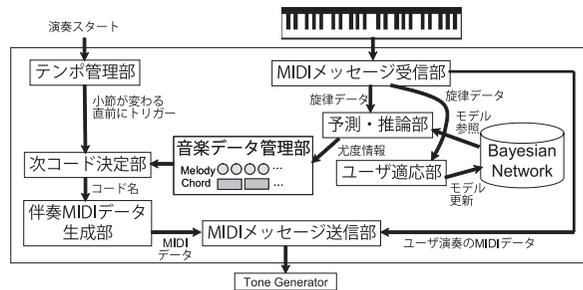


図 1 BayesianBand の処理の流れ
Fig. 1 System overview of BayesianBand.

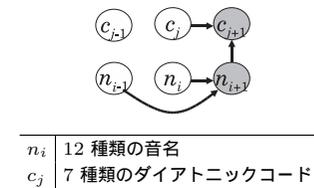


図 2 旋律予測・コード推論のためのベイジアンネットワーク
Fig. 2 Bayesian network for melody prediction and chord inference.

て処理が行われる。これは、前者の推論処理が打鍵ごとに行われるのに対し、推論されたコードが演奏されるのは次の小節の先頭であり、両者は同期しないからである。

現在のシステムでは、テンポ管理の主導権はシステムが握るため、処理が開始されると自動的に一定のテンポで伴奏が開始される。最初のコードは上述のとおりトニックコードである。ユーザが MIDI キーボードで主旋律を演奏すると、その演奏情報が MIDI メッセージ受信部を介してシステムに入力され、次の音（音高）を予測し、それに合うコードを推論する（予測・推論部）。推論結果は音楽データ管理部に書き込まれる。予測・推論処理は打鍵のたびに行われるため、音楽データ管理部における推論結果も打鍵が繰り返されるたびに更新され、小節が変わる直前のタイミングで、最新のコード推論結果が採用されて、1 小節間そのコードが演奏される。具体的には、テンポ管理部が小節が変わる直前のタイミングで、次コード決定部にトリガを送る。次コード決定部は音楽データ管理部から最新のコード推論結果を取得し、最尤のコード名を次のコードと決定する。その後、伴奏 MIDI データ生成部で伴奏 MIDI データが生成され、MIDI メッセージ送信部に送られる。また、以上の処理と並行して、予測モデルの更新も打鍵のたびに行われる（ユーザ適応部）。

3.3 確率モデルに基づく旋律予測・コード推論

本研究では、ユーザの主旋律 $n = (n_1, \dots, n_i)$ と現在までのコード進行 $c = (c_1, \dots, c_j)$ が与えられたときに、最尤の次のコード c_{j+1} を推論するための確率モデルを設計する。一般に、主旋律とコード進行にはそれぞれに時間軸上（横方向）の依存関係があり、各時刻の主旋律とコードにも（縦方向の）依存関係が存在する。そのため、こういった複数の依存関係を確率モデル中で表現する必要がある。しかし、既存の多重マルコフ連鎖（N-gram）に基づく旋律予測^{2),6)} や HMM に基づく和声付けの研究⁸⁾ では、これらのすべての依存関係

はモデル化されていなかった。

(n, c) と c_{j+1} の依存関係を上述の 3 種の依存関係に基づいて考えるため、ユーザが次に弾く音 n_{i+1} を導入する。このとき、これらの同時分布は

$$p(n, c, n_{i+1}, c_{j+1}) = p(c_{j+1}|c, n_{i+1})p(n_{i+1}|n)p(n)p(c)$$

と表される。このとき、 $p(n_{i+1}|n)$ が主旋律の横の依存関係を、 $p(c_{j+1}|c, n_{i+1})$ が縦の依存関係とコード進行の横の依存関係を表す。前者が旋律予測の、後者が旋律予測に基づくコード推論の確率モデルとすることができる。ここで、 n と c は本来独立ではないが、推論時にはどちらも確定されており推論には影響しないため、独立と扱った。このように、旋律予測とコード推論を分割してモデル化することで、たとえば各モデルを異なるコーパスから学習したり、後述のように旋律予測モデルのみを逐次的に更新したりできるようになる。なお、学習データが有限であることを考慮し、過去の主旋律は直近の N_n 個、コード進行も直近の N_c 個（現実装ではどちらも 2）の要素しか考慮しないものとする。

以上で述べた確率モデルは、図 2 に示すベイジアンネットワークとして表すことができる。ベイジアンネットワークでは、少ない計算量で確率計算を収束させられることから単結合（有向グラフにおける向きを取り除いても循環しないネットワーク）が望ましい。上で述べたように n, c は推論時には確定済みであるので、これらの中で閉じた依存関係を省略することで単結合ネットワークが得られる。

実際の推論処理は、前述のとおり打鍵のたびに行われる。 n_{i-1}, n_i に観測された音符情報、 c_{j-1}, c_j に決定済みのコード名がセットされ推論が実行されると、 n_{i+1} と c_{j+1} の確率分布が Pearl の確率伝播法¹⁰⁾ によって計算される。この推論が打鍵のたびに繰り返された後、小節が変わる直前のタイミングで、 c_{j+1} において最も確率の高いコード名が出力される。

3.4 確率モデルの逐次更新

確率モデルの逐次更新の目的は、モデルを固定化させないことで新奇性を担保するだけでなく、ユーザが弾く旋律の傾向を随時学習することで、旋律の予測精度を上げ、より良いコード推論を実現することにある。つまり、旋律予測のモデルを特定のユーザに適応させることが目的であるので、上述の確率モデルのうち $p(n_{i+1}|n)$ のみを逐次的に更新する。モデル更新の基本的なアイデアは、あらかじめコーパスから学習した条件付き確率とユーザの打鍵からオンラインで計算した条件付き確率の重み付き平均を用いることである。重みは、ユーザの打鍵数が増えるに従ってユーザの打鍵から計算した条件付き確率の重みが大きくなるように設計する。具体的には、当該条件付き確率を次式により計算する：

$$p(n_{i+1}|n) = \frac{p_0(n_{i+1}|n) + \alpha \{\log N(n)\} N(n, n_{i+1})/N(n)}{1 + \alpha \log N(n)}$$

ここで、 $p_0(n_{i+1}|n)$ はコーパスから学習した条件付き確率、 $N(n)$ はユーザが音列 n を弾いた回数、 $N(n, n_{i+1})$ はユーザが n の後に n_{i+1} を弾いた回数、 α は定数である。

4. 実装・試用結果

4.1 実装

以上で述べたシステムを Java を用いて実装した。システム全体のフレームワークの実装には CrestMuseXML Toolkit^{*1}を、ベイジアンネットワークの学習・推論には Weka^{*2}を用いた。学習には、文献 11) に掲載されている全 206 曲および文献 12) から抜粋した 209 曲に対する、主旋律とコード進行が組になったデータを用いた。

4.2 試用結果

本システムを第 1 著者が試用した。何度かセッションを繰り返すうちにシステムが生成するコード進行の傾向をある程度つかむことができ、スムーズにジャムセッションをできるようになった。次に来そうなコードを予測しながら自らの演奏を調整し、実際に予測したコードが演奏されたときには、親近性による快の感情を覚えることができた。予測したコードが来なかったとき、不協和な響きになることも少なくなかったが、結果的に音楽的に妥当な響きになる場合もあり、そのときの意外感や新鮮さを楽しむことができた。このように、1 章で述べた 2 種類の快の提供をある程度実現できたと考える。

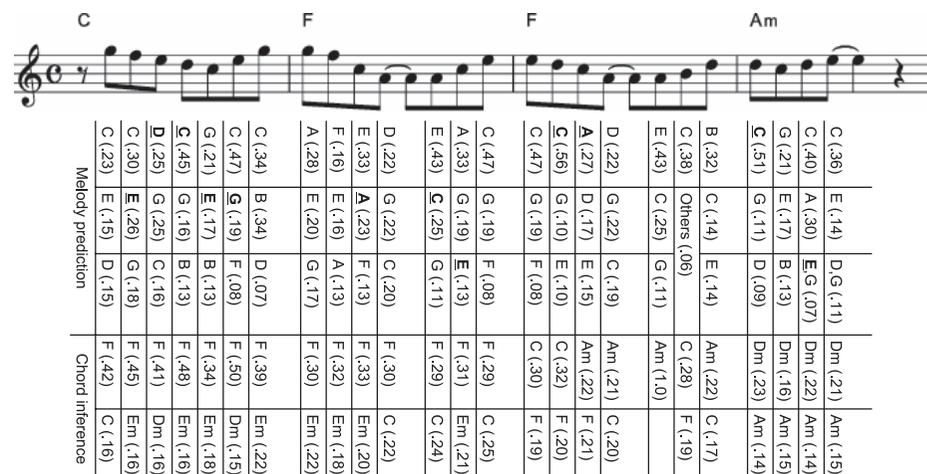


図 3 試用において演奏した旋律の一部とそれに対する旋律・コードの予測・推論結果。括弧内の数字は尤度を、強調された文字は実際に弾かれた音を表す

Fig. 3 Example of performed melodies and its melody-prediction and chord-inference results.

また、演奏しながら次に来そうなコードを「探る」行為は、それ自体がゲームのような楽しいものであった。この楽しみは、あらかじめコード進行が決まっていたり、コード進行がランダムに決まるようなジャムセッションでは味わえないものである。

試用時に演奏した旋律とそれに対する予測結果の一部を図 3 に示す。旋律予測において尤度が最も高かった音だけ見ると予測精度は高くないものの、3-best 候補まで見ると半数の打鍵に対して次の音を予測することができ、結果的に音楽的に自然なコード進行を出力することができた。

しかし、コード進行は比較的パターン化されている場合が多いため、たとえば $c_j = C$ であれば n_i や n_{i+1} がどんな値でも $c_{j+1} = F$ の確率が最も高くなるといったケースが見られた。これを解決するには、ユーザの主旋律に対する次コードの尤度と過去のコード進行に対する次コードの尤度に重み付けを行うといったことが必要となる。

5. おわりに

本稿では、ユーザとシステムが相互に相手の演奏を予測し合うジャムセッションシステム「BayesianBand」を提案した。ベイジアンネットワークを用いることで、ユーザの旋律を

*1 <http://www.crestmuse.jp/cmxml/>

*2 <http://www.cs.waikato.ac.nz/weka/>

ある程度予測して音楽的に自然なコード進行を出力することができた。しかし、新奇性を担保するために導入したモデルの逐次適用によって、どの程度新奇性による快を実現できたかについては評価していない。今後、複数の被験者による評価実験を通じて3つの要件がどの程度満たされているかを明らかにしていく予定である。

参 考 文 献

- 1) 下條信輔：下條「潜在脳機能」プロジェクト研究計画書 (2004).
http://impbrain.shimojo.jst.go.jp/jpn/about_jpn.html
- 2) Pachet, F.: The Continuator: Musical Interaction With Style, *Proc. ICMC* (2002).
- 3) Nishijima, M. and Watanabe, K.: Interactive Music Composer based on Neural Networks, *Proc. ICMC*, pp.53-56 (1992).
- 4) 後藤真孝, 日高伊佐夫, 松本英明, 黒田洋介, 村岡洋一：仮想ジャズセッションシステム：VirJa Session, *情報処理学会論文誌*, Vol.40, No.4, pp.1910-1921 (1999).
- 5) 青野裕司, 片寄晴弘, 井口征士：バンドライクな音楽アシスタントシステムについて, *情報処理学会研究報告*, 94-MUS-8, pp.45-50 (1994).
- 6) Conklin, D. and Witten, I.H.: Multiple Viewpoint Systems for Music Prediction, *J. New Music Res.*, Vol.24, No.1, pp.51-73 (1995).
- 7) 松尾聡子, 片寄晴弘, 井口征士：旋律予測のコンピュータシミュレーションモデルに関する一検討, *情報処理学会論文誌*, Vol.41, No.2, pp.498-508 (2000).
- 8) 川上 隆, 中井 満, 下平 博, 嵯峨山茂樹：隠れマルコフモデルを用いた旋律への自動和声付け, *情報処理学会研究報告*, 99-MUS-34, pp.79-82 (2000).
- 9) 豊田健一, 片寄晴弘：ながしミュージシャンシステム：豊次郎, *インタラクシオン* (2005).
- 10) Pearl, J.: *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann (1988).
- 11) 伊藤伸吾：ザ・プロフェッショナル・スタンダード・ジャズ・ハンドブック, 中央アート出版社 (1992).
- 12) 高島慶司：スタンダードジャズのすべて1, 全音楽譜出版社 (1998).

(平成 21 年 3 月 18 日受付)

(平成 21 年 6 月 4 日採録)



北原 鉄朗 (正会員)

2002 年東京理科大学理工学部卒業。2007 年京都大学大学院情報学研究科博士後期課程修了。博士 (情報学)。日本学術振興会特別研究員 (DC2) を経て、現在、科学技術振興機構 CREST (CrestMuse プロジェクト) 研究員。音楽情報処理, 聴覚的情景分析等に興味を持つ。電気通信普及財団第 19 回テレコムシステム技術学生賞, 京都大学第 2 回総長賞等受賞。電子情報通信学会, 人工知能学会, 日本音響学会, IEEE 各会員。



戸谷 直之

2007 年関西学院高等部卒業。同年関西学院大学理工学部情報科学科に入学。科学技術振興機構 CREST (CrestMuse プロジェクト) において CrestMuseXML Toolkit の開発に従事。音楽情報処理, メディアアートに興味を持つ。



徳網 亮輔

2007 年関西学院高等部卒業。同年関西学院大学理工学部情報科学科に入学。科学技術振興機構 CREST (CrestMuse プロジェクト) において CrestMuseXML Toolkit の開発に従事。実世界センシング, 入力インタフェースに興味を持つ。



片寄 晴弘 (正会員)

1991 年大阪大学大学院基礎工学研究科博士課程修了。工学博士。イメージ情報科学研究所, 和歌山大学を経て、現在、関西学院大学理工学部教授。ヒューマンメディア研究センターセンター長。音楽情報処理, 感性情報処理, HCI の研究に従事。科学技術振興機構さきがけ研究 21「協調と制御」領域研究者。科学技術振興機構 CREST「デジタルメディア (略称)」領域 CrestMuse プロジェクト代表研究者。電子情報通信学会, 人工知能学会各会員。