

木構造モデルに基づくコードとメロディの対話的生成システム

津島 啓晃

糸山 克寿

吉井 和佳

京都大学 工学部情報学科 知能情報学専攻

1. はじめに

計算機的手法による自動作曲の研究は、これまで盛んに行われてきた。例えば、RNN [1] を用いたメロディの自動生成手法が提案されている。一方で、コード進行を扱った自動作曲の研究も行われている。自動作曲システム Orpheus [2] では、コード進行と歌詞が与えられると、コード進行と歌詞の韻律を制約としたメロディが生成される。また、与えられたメロディに対するコード進行の生成(和声付け)に関する研究も多く行われている [3, 4]。これらの問題設定では、コードあるいはメロディがもう一方に基づいて生成されるが、ユーザが生成結果の一部を変更したい場合に、両者を対話的に編集することを可能とするシステムが望まれていた。

また、コード進行は従来、マルコフモデルによって表現されるため、標準的な和声付け手法では、コードを隠れ状態、音高を出力とした隠れマルコフモデル (HMM) が用いられた [3]。しかし、この手法では、コードの機能や繰り返し構造が考慮されないという課題があった。ここで、コードの機能は、音楽理論における機能と声という自然言語における品詞のような概念として捉える。

以上をふまえ、本稿では、数小節からなるメロディあるいはコード進行に対して、もう一方を自動生成することで、ユーザが対話的に生成結果を洗練させていくことができる自動作曲システムを提案する。本研究では、コードとメロディに関する階層的生成モデルを事前学習することで、現在得られているメロディとコード進行から、ユーザが指定した箇所を事後分布に従ってサンプリングする手法を用いる。和声付け手法に関しては、コードの機能や繰り返し構造を考慮するために、コード進行の生成モデルとして木構造モデルである確率的文脈自由文法 (PCFG) を用いた手法 [4] を用いた。コードに対するメロディ生成に関しては、メロディにおける音符の音高と開始位置に関するマルコフ性から、セミマルコフモデルに基づいてメロディの一部を生成する手法を提案する。

2. 提案システム

本システムでユーザに提示される画面の例を図1に示す。画面には、現在のメロディとコード進行及びその背後にある木構造が表示される。まず、はじめにメロディの MIDI ファイルをシステムに入力し、そのメロディに対するコード進行が自動で生成される。ユーザが行う操作は、(1) 現在のメロディに対するコード進行の自動生成、(2) 木構造に基づくコードの分割・統合、(3) 現在のコード進行に対するメロディ生成、(4) コードの開始位置の変更からなる。コードの分割・統合は、画面上のノードを一つ選択したのち、いずれかの操作を選択することで行われる。メロディの生成は、画面上のコードを一つ選択することで行われる。コードの開始位置の変更は、コードの境界をマウス操作によって移動させ行う。以上の動作を繰り返すことによって、コードとメロディを対話的に編集することが可能である。

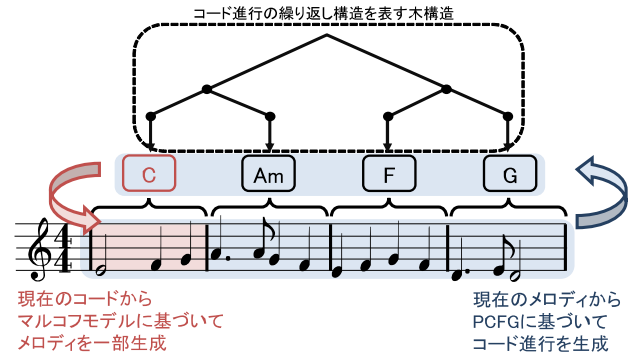


図1: 提案システムにおけるコード進行とメロディの対話的生成

3. 提案モデル

提案システムを実現するため、我々は、(1) PCFG に基づくコード記号系列生成モデル、(2) 拍節マルコフモデルに基づくメロディのリズム生成モデル、(3) コードの条件付きマルコフモデルに基づくメロディの音高系列生成モデルからなる階層的生成モデルを提案する。本章では、各モデルの定式化と学習方法について述べる。

3.1 モデル定式化

PCFG によって生成されたコード記号系列を $z = \{z_n\}_{n=1}^N$ (N はコードの数)、各コードの開始拍位置 (16 分音符単位) を $\phi = \{\phi_n\}_{n=1}^N$ と記す。 n 個目のコードの支配区間 $[\phi_n, \phi_{n+1})$ にある音高系列の部分列を $x_n = \{x_{n,m}\}_{m=1}^{M_n}$ (M_n は、区間中の音符数) と記す。 x_n を繋ぎ合わせて得られる音高系列全体を $x = \{x_n\}_{n=1}^N$ と表す。 $\psi_{n,m}$ は、音高 $x_{n,m}$ の開始拍位置を表し、その集合を $\psi = \{\{\psi_{n,m}\}_{m=1}^{M_n}\}_{n=1}^N$ とする。 ϕ, ψ は、メロディの小節数を L として、0 から $16L - 1$ の整数値を取る。

本稿で用いる確率的文脈自由文法 (PCFG) は、非終端記号の集合 V 、終端記号 (コード記号) の集合 Σ 、及びこれらの記号を用いて木構造を生成する次の3種のルール確率からなる。 $\theta_{A \rightarrow BC}$ は非終端記号 $A (\in V)$ が2つの非終端記号 $B, C (\in V)$ に分岐する確率であり、 $\eta_{A \rightarrow \alpha}$ は $A (\in V)$ が終端記号 $\alpha (\in \Sigma)$ を出力する確率である。 λ_A は非終端記号 A が終端記号を出力する確率である。また、これらのルール確率をそれぞれ θ, η, λ と記す。同様の表記法を本稿を通して用いる。

音符のリズムに関する拍節マルコフモデルは、コードの16分音符単位での相対的な開始拍位置に関する以下の遷移確率によって記述する。

$$p(\psi_{n,m} | \psi_{n,m-1}) = \pi_{\psi_{n,m-1}, \psi_{n,m}} \quad (1)$$

ここで、 $\pi_{a,b}$ は、相対的な拍位置が $a \bmod 16$ から $b \bmod 16 + d$ (d は、拍位置 a, b を隔てている小節線の数) に遷移する確率を表す。

音高系列に関するマルコフモデルは、以下の遷移確率によって記述する。

$$p(x_{n,m} | x_{n,m-1}, z_n) = \tau_{x_{n,m-1}, x_{n,m}}^{z_n} \quad (2)$$

ここで、 $\tau_{a,b}^z$ は、和音 z において、音高が a から b に遷

移する確率を表す。ただし、 $n > 1, m = 1$ の場合には、直前のコードの最後の音符 $x_{n-1, M_{n-1}}$ からの遷移を考える。さらに、音高 x の初期確率 $p(x_{1,1}|z_1)$ も導入する。

3.2 モデルの学習

モデルパラメータ $\Theta = \{\theta, \eta, \lambda, \pi, \tau\}$ を事後分布最大化 (MAP) の枠組みから学習する手法について述べる。PCFG のパラメータ θ, η, λ は、音楽コーパス中のコード記号系列 z を与えて、 z を生成する木 t とパラメータ θ, η, λ を、それぞれ条件付き事後分布 $p(t|\theta, \eta, \lambda, z)$, $p(\theta, \eta, \lambda|t, z)$ によって交互にサンプリングするギブスサンプリングによって教師なし学習する [5]。

マルコフモデルのパラメータ π の事後分布は、 ψ のデータが与えられたとき、ディリクレ事前分布とカテゴリカル分布の共役性から容易に求まる。同様に、 z, ϕ, x のデータの組が与えられたとき、 τ の事後分布も求まる。

4. コードとメロディの対話的生成

本章では、学習された提案モデルを用いて、提案システムの基本動作の基盤となる、現在得られているメロディあるいはコード進行から、もう一方を生成する手法、及びコード進行の選択箇所を分割・統合する手法について述べる。

4.1 メロディに対するコード進行の生成

与えられたメロディ x に対して、コード記号 z と潜在木 t をビタビアルゴリズムを用いて求める。非終端記号 A がメロディ部分列 $\{x_n, \dots, x_m\}$ を生成するような最尤の部分木とコードの尤度を $p_{n,m}^A$ として、葉ノードから頂点に向かって以下のように反復計算することができる。

$$p_{n,n}^A = \lambda_A \max_{z \in \Sigma} \eta_{A \rightarrow z} p(x_n|z) \quad (3)$$

$$p_{n,n+k}^A = (1 - \lambda_A) \max_{B, C \in V} \theta_{A \rightarrow BCP} p_{n,n+1}^B p_{n+1,n+k}^C$$

ここで、 $p(x_n|z)$ は、音高部分列 x_n がコード z のもとで生成する確率であり、以下に与えられる。

$$p(x_n|z) = \tau_{x_{n-1, M_{n-1}}, x_{n,1}}^z \prod_{i=2}^{M_n} \tau_{x_{n, m-1}, x_{n, m}}^z \quad (n > 1) \quad (4)$$

そして、最尤の t と z は、上述の計算におけるビタビパスを木の頂点から順にたどっていくことで得る。

また、生成されたコード進行に対する分割・統合操作も推定された木構造に基づいて行うことができる。系列中のあるコード z_n を分割するには、 z_n を出力する非終端記号をルートとし、音高部分列 x_n を出力する部分木に対して、上記のアルゴリズムを適用すればいい。統合操作を行う場合にも同様である。

4.2 コード進行に対するメロディの生成

n 個目のコード z_n に対して、その支配区間 $[\phi_n, \phi_{n+1})$ で、メロディ (x_n, ψ_n) をサンプリングする手法を述べる。まず、拍位置 $\phi_n + t$ に音高 y_t の音符があり、その直前の音符の長さが d_t であるときの確率 $\alpha_{y_t, d_t} = p(y_t, d_t|z_n)$ をセミマルコフモデルの前向きアルゴリズムに基づいて以下のように算出する。

$$\alpha_{y_t, d_t} = \pi_{\phi_n + t - d_t, \phi_n + t} \cdot \sum_{y_{t-d_t}, d_{t-d_t}} \alpha_{y_{t-d_t}, d_{t-d_t}} \tau_{y_{t-d_t}, y_t}^{z_n} \quad (5)$$

各 t で d_t は、 $d_t \in \{1, \dots, t + \phi_n - \psi_{n-1, M_{n-1}}\}$ の状態空間を持つ。この集合の最後の要素は、メロディを生成したい区間の直前の音符 $\psi_{n-1, M_{n-1}}$ からの遷移を考慮する。

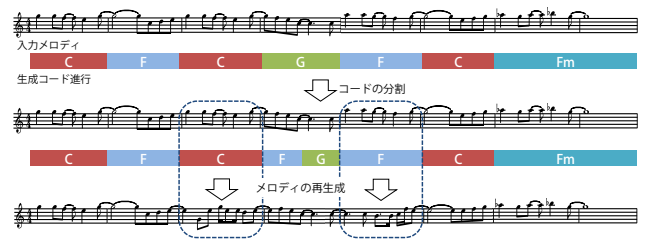


図2: 提案手法を用いたコード進行とメロディの生成例

この周辺尤度を時刻 $t \in \{0, \dots, \phi_{n+1} - \phi_n, \psi_{n+1,1} - \phi_n\}$ に対して算出した結果を用いて、区間の直後の音符 $\psi_{n+1,1}$ から後ろ向きにメロディのサンプルを得る。

5. 生成結果

提案システムによるコード進行とメロディの対話的生成が有効に動作しているかを確認した。PCFG の学習には、The SALAMI Annotation Data [6] 内のポピュラー音楽 468 曲から抽出した、A メロなどの楽節に対応する 8 小節分のコード進行 705 個を用いた。コード記号は、ルート音 $\{C, C\#, \dots, B\}$ と $\{major, minor\}$ の組み合わせと、それ以外の 25 種とした。マルコフモデルの学習には、Rock Corpus [7] 内のポピュラー音楽 194 曲から抽出した、コード進行とメロディのペア 9902 組を用いた。すべての楽曲の調はすべて C に移調した。また、PCFG の非終端記号数は 15 とした。

提案システムを用いたコード進行とメロディの対話的生成の例を図 2 に示す。図中の 1 段目は、はじめに与えたメロディとそれに対して生成されたコード進行を示す。ただし、コードの開始位置は小節線の位置に置いた。2 段目は、4 個目のコード (1 小節) を分割した場合の結果を示す。3 段目は、3, 6 個目のコードのもとでのメロディを再生成した状態を示す。このようなコード進行及びメロディの対話的な生成操作を繰り返すことで、ユーザが生成結果を洗練させていくことができる。

6. おわりに

本稿では、木構造モデルを用いて、コード進行とメロディを互いに対話的に生成する自動作曲システムを提案した。提案システムによって、コード進行の繰り返し構造を考慮した和声付けと、メロディの音高と拍位置のマルコフ性を考慮したメロディ生成が実現された。

現状のメロディのサンプリング手法では、変更したいメロディの前後のメロディとの連結性は考慮しているが、変更後のメロディの大局的な構造は考慮していない。今後は、変更後のメロディ全体の音楽的妥当性を保証するための手法変更を行いたい。また、ユーザインタフェースを開発し、それを用いたユーザテストを行う予定である。

謝辞 本研究の一部は、JSPS 科研費 26700020, 16H01744 および JST ACCEL No. JPMJAC1602 の支援を受けた。

参考文献

- [1] E. Waite *et al.*: "Generating long-term structure in songs and stories," <https://magenta.tensorflow.org/2016/07/15/lookback-rnn-attention-rnn>, 2016.
- [2] S. Fukayama *et al.*: "Orpheus: Automatic composition system considering prosody of Japanese lyrics," *ICEC*, 309-310, 2009.
- [3] M. Allan *et al.*: "Harmonising chorales by probabilistic inference," *NIPS*, 25-32, 2005.
- [4] H. Tsushima *et al.*: "Function- and rhythm-aware melody harmonization based on tree-structured parsing and split-merging sampling of chord sequences," *ISMIR*, 502-508, 2017.
- [5] M. Johnson *et al.*: "Bayesian inference for PCFGs via Markov chain Monte Carlo," *NAACL-HLT*, 139-146, 2007.
- [6] J. B. L. Smith *et al.*: "Design and creation of a large-scale database of structural annotations," *ISMIR*, 555-560, 2011.
- [7] T. D. Clercq *et al.*: "A corpus analysis of rock harmony," *Popular Music*, 145-148, 1992.