

ベイズ文脈自由文法に基づく 和音系列の教師なし構文解析と自動生成

津島 啓晃[†]中村 栄太[‡]糸山 克寿[‡]吉井 和佳[‡][†] 京都大学 工学部情報学科[‡] 京都大学 大学院情報学研究科 知能情報学専攻

1. はじめに

これまで、計算機による自動作曲の研究が盛んに行われてきた。例えば、自動作曲システム Orpheus [1] では、歌詞が入力されると、その韻律を制約としてメロディが自動生成される。一方、コード進行はあらかじめ用意された候補の中から選ばなければならない、自動生成の手法やユーザの編集を支援する手法が望まれていた。

和音系列の自動生成では、ユーザによる和音進行の作成を支援したり、与えられたメロディに対して和音を付与できることが重要である。これらを実現するうえで、これまで、局所的なマルコフ性に基づく n -gram モデルを用いることが一般的であった。しかし、大局的な構造を表現できないため、和音系列の終止構造を再現することは難しく、ユーザがインタラクティブに作曲を行いたい場合でも、ある和音を分割したり、隣り合う和音を簡約するなどといった操作が難しかった。

本研究では、和音系列の生成モデルとして、確率的文脈自由文法 (PCFG)[2] を用いる方法を提案する。生成音楽理論 (GTTM) では、メロディ (音符系列) の背後には木構造が存在すると考え (各音符は葉ノードに配置)、葉から根に向かってたどれば、メロディの簡約化ができることを提唱している。しかし、この種の簡約化規則は音楽理論に基づいて人間が書き下したものであり、各規則の優先度は、音符系列と人手で付与した木構造から推定しなければならなかった [3]。近年、簡約過程とは逆の生成過程を考えることで、PCFG に基づく音符系列の生成モデルが提案され、音符系列だけから生成規則を教師なし学習できるようになってきている [4]。本研究では、音符系列と同様、和音系列の背後には木構造 (導出木) が存在すると仮定し、その推定を試みる。

通常、PCFG の非終端記号は品詞に対応するが、本研究では、非終端記号は和音の「機能」に対応することを想定している。機能と声理論では、和音をその役割によって、C や Am を Tonic, G や Em を Dominant, F や Dm を Sub-dominant に分類できることが知られている。テキスト解析で、HMM を用いて教師なしで品詞が推定できるのと同様に、本研究では、和音系列だけから教師なしで機能を推定することを目指す。さらに、推定した木構造や機能を考慮しながら、与えられたメロディに対して、適切な和音進行を付与する手法を提案する。

2. 提案モデル/手法

本章では、PCFG を用いた和音系列の解析手法とメロディに対する和音系列の生成手法について述べる。本研究では、PCFG に従って和音系列が生成され、各和音に依存して、その区間内の音符系列が生成される階層モデルを考える (図 1)。まず、和音系列を与えて、それらの背後に存在する木構造と、PCFG の導出規則の確率を同時にギブスサンプリングを用いて教師なし学習する。

Unsupervised Syntactic Parsing and Automatic Generation of Chord Sequences Based on Bayesian Context-Free Grammar: Hiroaki Tsushima, Eita Nakamura, Katsutoshi Itoyama, and Kazuyoshi Yoshii (Kyoto Univ.)

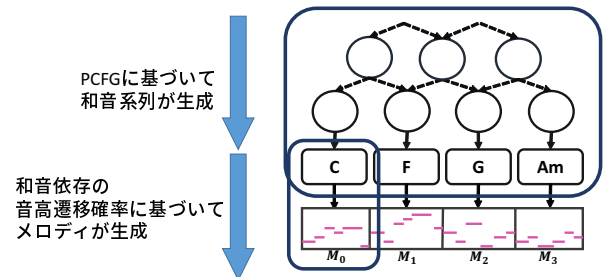


図 1: 和音系列と音符系列の階層的生成モデル

次に、和音系列と音符系列のペアを与えて、各和音下での音高の遷移モデルを学習する。最後に、学習された和音系列生成モデルと音符系列生成モデルを用いて、与えられた音符系列に対する和音進行を確率的に生成する。

本研究では、和音を $\{C, C\#, \dots, B\}$ と和音の種類 $\{\text{Major}, \text{Minor}\}$ を組み合わせて表記する。和音系列はすべてハ長調に移調しておき、楽曲中の A メロ・サビ等のセグメントごとに切り出して、 $\{c_i\}_{i=0}^N$ で表す。ここで、 N は和音系列の長さを表す。また、メロディーについては、和音 c_i の下に音高列 $M_i = \{m_{i,j}\}_{j=0}^{\text{len}(i)}$ が生成される。ここで、 $\text{len}(i)$ は、和音 c_i 区間内での音符数である。

2.1 提案モデル

まず、和音系列に対する PCFG を定式化する。PCFG のモデルパラメータは、非終端記号 A が B, C に分岐する確率 $\theta_{A \rightarrow BC}$ 、非終端記号 A が終端記号 (和音) α を出力する確率 $\theta_{A \rightarrow \alpha}$ 、非終端記号 A が分岐と出力のどちらを選択するかを表す確率 $(\lambda_{A0}, \lambda_{A1}) = \lambda_A$ である。

さらに、PCFG の拡張モデルとして、PCFG と Bigram を統合したモデル (Bigram-PCFG) を定式化する。このモデルでは、通常の PCFG における終端記号への出力確率 $\theta_{A \rightarrow \alpha}$ を、1 つ前の和音 α' を考慮した出力確率 $\theta_{A \rightarrow \alpha | \alpha'}$ に拡張したものを扱う。

次に、和音に基づくメロディの生成過程は、和音中のメロディの音高遷移に着目し、和音 c_i の支配区間で音高列 $\{m_{i,j}\}_{j=0}^{\text{len}(i)}$ が与えられた時の確率 $P(m_{i,j} | c_i, m_{i,j-1})$ によってモデル化する。この学習は最尤推定で行った。

2.2 和音系列の解析

ベイズ学習に基づく本手法での PCFG の学習方法を述べる。まず、モデルパラメータに Dirichlet 分布による事前確率を以下のように導入する。

$$\theta_A \sim \text{Dirichlet}(\xi_A)$$

$$\lambda_A \sim \text{Dirichlet}(\zeta_A)$$

ここで、 $\theta_A = (\cup_{B,C \in N} \theta_{A \rightarrow BC}) \cup (\cup_{\alpha \in T} \theta_{A \rightarrow \alpha}) (N, T$ はそれぞれ非終端記号, 終端記号の集合) であり、 ξ_A, ζ_A はそれぞれ θ_A, λ_A が従う Dirichlet 分布のハイパーパラメータである。推論方法は、推定木とモデルパラメータを交互にサンプリングすることによってパラメータを推定するギブスサンプリング法を用いる。また、Bigram-PCFG モデルに対しても同様に推論を行う。

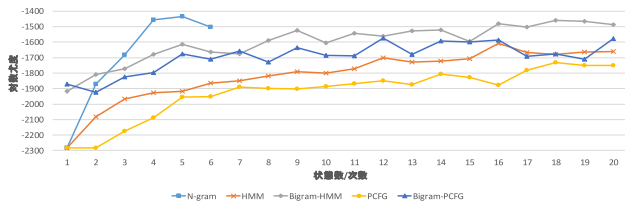


図 2: 各モデルにおけるテストデータ尤度

また、PCFG は解の初期値依存性が大きいため、PCFG の初回の木構造のサンプリングにおける木の最下段の状態系列を、HMM による学習に基づくビタビ探索によって推定される状態系列で固定し、学習の安定性を確保する。また、拡張モデルである Bigram-PCFG についても、HMM の出力確率に観測間の Bigram を統合したモデル (Bigram-HMM と記す) の状態系列推定結果で初期化し、学習の安定性を確保する。

2.3 和音系列の生成

与えられたメロディに対して、学習した PCFG と和音依存の音高遷移確率を用いて和音系列を生成する。まず、音高遷移確率の積によって、 i 小節目の和音 α 上に与えられたメロディ M_i の尤度 $P(M_i|\alpha)$ を求める。次に、非終端記号 A から i 個目から $i+n$ 個目までの音高が生成される周辺尤度 $\beta_{i,i+n}$ を以下で計算する。

$$\beta_{i,i}(A) = \sum_{\alpha} \lambda_{A1} \theta_{A \rightarrow \alpha} P(M_i|\alpha)$$

$$\beta_{i,i+n}(A) = \sum_{B,C} \lambda_{A0} \theta_{A \rightarrow BC} \sum_{j=1}^n \beta_{i,i+j-1}(B) \beta_{i+j,i+n}(C)$$

この値を用いて、Inside-Outside アルゴリズムに基づくランダムサンプリングあるいはビタビ探索を行うことで、メロディに対する和音系列を生成する。

生成された和音系列にユーザが編集を加えたい場合も存在する。例えば、機能を考慮しつつ、系列中のある和音を分割するには、分割したい和音をルートとし、音高系列を葉とする部分木に対して、上記の和音系列生成手法を適用すればよい。

3. 評価

PCFG の言語モデルとしての予測性能と、PCFG に基づく和音系列生成結果について報告する。

3.1 テストデータに対する予測性能

PCFG に基づく和音系列解析手法の性能を評価するため、学習されたパラメータのテストデータに対する尤度を測定した。比較手法として、n-gram, HMM, Bigram-HMM, PCFG, Bigram-PCFG に対して、n-gram の次数、HMM の隠れ状態数、PCFG の非終端記号数を変えてテストデータ尤度を評価した。

実験には、The SALAMI Annotation Data [5] のうち、長調のポピュラー音楽 468 曲から抽出した 8 個以上 32 個以下の和音からなる和音系列 1101 個を用いた。和音は、Major, Minor, Other に分類した。データを 11 分割して交差検証を行い、各手法について、学習時に最尤となったパラメータを用いてテストデータに対する尤度を求めた。学習データとテストデータは、同じ曲の和音系列が両データセット間で共有されないようにした。

実験結果を図 2 に示す。PCFG については、適切な非終端記号数を選べば 2-gram 同程度以上の性能が得られ、

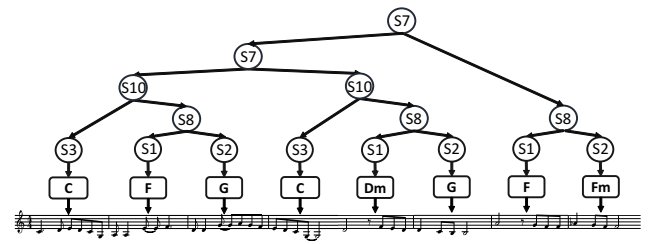


図 3: 提案手法による和声付けの例

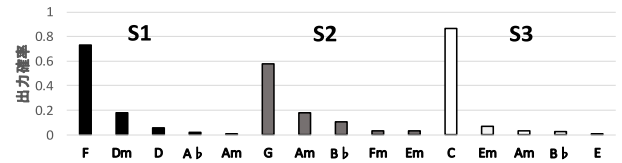


図 4: PCFG の出力確率 (各状態における上位 5 和音)

Bigram-PCFG については、適切な非終端記号数を選べば 3-gram と同程度の性能が得られることが分かった。

3.2 和音生成結果

PCFG を用いたメロディに対する和声付け結果の例を考察する (図 3)。この例では、隠れ状態数 3 の HMM で各和音の潜在状態を推定し、非終端記号数 12 の PCFG における初期値として用いた。この PCFG において終端記号を出力する 3 つの非終端記号の出力確率 (図 4) を見ると、C と Am, G と Em, F と Dm がそれぞれ同じ状態から優勢に出力され、機能的和声による和音分類に類似した性質が見られた。その上で、図 3 の和声付けの結果を見ると、和音の機能を表す状態 S1, S2 に分岐する S8 などの上層の状態が学習されていることにより、和音に対するメロディの尤度と和音の機能的な繰り返し構造の両方を考慮した和音系列が得られていることが分かった。

4. おわりに

本稿では、ベイズ文脈自由文法に基づいて和音系列の生成規則を教師なしで学習する手法を提案した。本手法では、和音の機能という和声学の概念を自動獲得することができた。テストデータ尤度による評価では、2-gram と同程度以上の性能が得られ、ユーザが和音系列をインタラクティブに編集するうえでの有用性を確認した。

今後は、和音系列作成インタフェースを開発し、それを用いたユーザ評価を行いたい。また、和音系列のサンプリングにモンテカルロ木探索 [6] を応用することで、PCFG に基づく木構造の尤度を考慮するだけでなく、各種の音楽的な特徴量に基づいて、生成された和音系列の音楽的な自然さを考慮できる手法を確立したい。

謝辞 本研究の一部は、JSPS 科研費 24220006, 26700020, 26280089, 16J05486, 15K16054, 16H01744, JST CREST, ACCEL の支援を受けた。

参考文献

- [1] 嵯峨山茂樹ほか: “日本語歌詞からの自動作曲,” The Operations Research Society of Japan, pp. 547–553, 2009.
- [2] M. J. Steedman: “A Generative Grammar for Jazz Chord Sequence,” Music Perception, Vol.2, No.1, pp. 52–77, 1984.
- [3] 浜中雅俊ほか: “音楽理論 GTTM に基づくグルーピング構造獲得システム,” 情処論, Vol.48, No.1, pp. 284–299, 2007.
- [4] E. Nakamura et al.: “Tree-Structured Probabilistic Model of Monophonic Written Music Based on the Generative Theory of Tonal Music,” IEEE ICASSP, pp. 276–280, 2016.
- [5] J. B.L.Smith: “Design and Creation of a Large-Scale Database of Structural Annotations,” ISMIR, pp. 555–560, 2011.
- [6] C. Browne et al.: “A Survey of Monte Carlo Tree Search Methods,” IEEE T-CIAIG, Vol. 4, No. 1, pp. 1–43, 2012.