2007－MUS－73 (12)
2007／12／15

確率文脈自由文法による旋律の拍節モデル推定

丹治 信　安藤 大地　伊庭斉志

東京大学大学院新領域創成科学研究科基盤情報学専攻

本稿では, 旋律の拍節構造を扱うための, 確率文脈自由文法 (PCFG) による拍節 PCFG モデルについて述べる. PCFG のような確率モデルを用いることの利点として, 揺らぎや曖昧性を含むモデルを構築できる点, また尤度という定量的な尺度でモデルの推定・評価が可能な点が挙げられる. 我々は問題設定として,PCFG から確率的に旋律が生成されると考え, 逆問題として, 与えられた音列の拍節構造を PCFG の最尤導出として推定する. 評価実験として既存楽曲からモデルの確率パラメータを学習し, 小節認識実験を行った. 結果, 単純な拍節 PCFG モデルと, 音程の情報を加えたモデルで,F-score がそれぞれ約 79% と 87% 程度の認識率となった.

Metrical Model Estimation for Melody using Probabilistic Context-Free Grammar

Makoto Tanji　Daichi Ando　Hitoshi Iba

Dept. of Frontier Informatics, Graduate School of Frontier Sciences, The University of Tokyo.

In this paper, we describe a new model named 'Metrical PCFG Model' to capture metrical structure of music melody. Probabilistic models like PCFG have advantages in that it can include the ambiguity and the varying in model and it can be evaluated by the likelihood criterion. We assumed that melodies are generated from the Metrical PCFG Model stochastically, and estimated the Metrical Structure of maximum likelihood derivation as a reverse-problem. In experiment, the probability parameters of the model were estimated by training data, and then the model was tested to recognition problem. The results for the simple model without pitch and the pitch-annotated model were 79% and 87% in F-score respectively.

1　はじめに

拍節構造は, 楽曲を分析し, 理解する為に欠かすことの出来ない楽曲構造である. 和声や旋律と並んで基本的な楽曲構造であるため, 楽曲の演奏表情付けやジャンル推定などの, より高次の音楽情報処理に必要となる. 人間は意識的に, もしくは無意識に拍節構造を知覚し, この拍節構造は GTTM [7] の理論などで指摘されているように, 階層構造を形成すると言われる. 人間の聴者は事前の経験から, この拍節構造に対する何らかのモデルを持っており, そのため経験を積めば楽譜が与えられなくても, 音列から拍節構造や調に対する情報を知覚することができる. このモデルは音楽的な常識やイディオムなどに例えられ, これをコンピュータ上で表現することで上記のような応用が見込まれる. そのため本稿では, このモデルを確率文脈自由文法で表し, 推定することを目的とする.

音楽は言語と共通の特徴を幾つか持っており, それ故, 言語学の手法を取り入れた研究が古くから行なわれてきた. 隠れマルコフモデル,N グラムモデル等による楽曲の解析, 生成もその一つと見ることができる. 近年, 言語学と確率統計などの手法を使って音楽情報処理を行なう研究が行なわれている.Bod [1] は言語と音楽の両方の視点からフレーズ構造の解析を試みた.Conklin [4] は PCFG を用いた旋律の簡約を評価した.Yamamoto ら [3] は PCFG を用いて演奏音列からの

リズムとテンポの推定を行なっている.

本研究では, 確率文脈自由文法を用いた楽曲の拍節構造のモデル化を目標とし, 評価として, 拍節構造の認識を試みる. 拍節構造は小節レベルで見ると $\frac{4}{4}$ 拍子や $\frac{3}{4}$, $\frac{6}{8}$ 拍子などの構造であるが, アウフタクトや最後の小節が不完全な長さで終わる楽曲なども考慮する. また, 本研究で扱う PCFG は解析モデルでもあり, 同時に生成モデルでもある. よって, 応用として楽曲データの解析と, その楽曲データから推定したモデルによって自動作曲を行なう応用などが考えられる.

本稿では, まず,PCFG について第 2 章で簡単に説明する. 次に, 第 3 章で PCFG を使った, 我々の手法について述べ, 第 4 章で行なった実験について報告する. 最後に, 考察とまとめとする.

2 PCFG(Probabilistic Context-Free Grammar

CFG(Context-Free Grammar, 文脈自由文法) は, 言語学者 Noam Chomsky によって広く知られることになった生成文法 [2] の一つである. 生成文法とは数学的に厳密に定義された形式的なモデルであり, 記号と生成規則 (または書き換え規則) のセットで表される. 生成文法は, 本質的に異なった能力を持つ 0 型から 3 型までの分類 (Chomsky 階層) が知られている (図 1). 図 1 の左下の 3 型 (正規文法) が最も制限があり, 外側の文法はその内側の文法を包含する能力を持っている. 文脈自由文法は, 正規言語より強力であるが文脈依存文法よりは能力の低い 2 型に分類され, 自然言語処理や, プログラム言語などにも使われている.

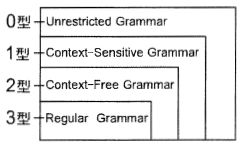

図 1 Chomsky 階層

PCFG(Probabilistic Context-Free Grammar) は, 名前の通り CFG を確率モデルに拡張したものである. 自然言語処理や RNA の解析, 確率モデル GP [9] などで使われており, 確率を付加したモデルを考えることで, 曖昧性や揺らぎのあるデータに対して解析を行うことが広く行われている.

2.1 形式的表現

PCFG は,G をある PCFG とすると次のように表される. ここで, 生成規則は Chomsky 標準型での表記である.

- $G = \{V_T, V_T, P, S\}$
- V_N : 非終端記号の集合
- V_T : 終端記号の集合
- P : 次のような形式の生成規則の集合
  $< A \to B \quad C, \quad p >$
  p : probability)
- S : start symbol ($S \in V_N$)

開始記号 S から始め, 生成規則を繰り返し適用して終端記号の列を得る過程を導出と呼び, 得られた終端記号列 $\mathbf{W}=\{t_1, t_2, t_3, \cdots\}$ を文と呼び,$S \overset{*}{\Longrightarrow} \mathbf{W}$ のように表す. 文 $\mathbf{W}$ に対する導出 $\mathbf{T}$ は, 生成規則が $A \to B \quad C$ のような形を取るため, 図 1 のように木構造で表され, また確率は次式で計算される.

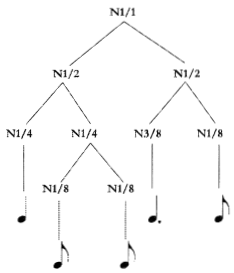

図 2 PCFG の導出木例

$$P(\mathbf{W}, \mathbf{T}) = p(w_1) \cdots p(w_N) = \prod_{i=1}^{N} p(w_i)$$

ここで,w_i は i 番目に適用された生成規則である. 一般に, 一つの文に対する導出が複数ある場合, その文法は曖昧であると言われ, 確率の一番高い導出を, 最尤導出と呼ぶ. このように構文の曖昧性を確率的モデルによって処理することで, 複数の導出から確率的に尤もらしい一つを選ぶことができる.

3 PCFG による拍節モデル

本章では,PCFG による拍節モデルとそれによる分析について述べる. おおまかな流れとして, まず拍節 *PCFG* モデルと呼ばれる確率モデルを定義し, 旋律を拍節構造を伴って生成することを示す. 次に, 本稿で報告する拍節モデルの推定問題について述べるために, 生成方向とは逆の既存楽曲からのモデルの推定方法と, 推定されたモデル上での未知楽曲の拍節構造の推定について述べる.

3.1 モデルの概要

まず, 図 3 のような旋律の生成されるプロセスを考える. パラメータ θ を持った旋律生成モデルから, 確率 $P_\theta(W)$ で旋律 W が生成される. この W とその導出木を併せて, 拍節構造を伴った楽譜のようなものと考えられる. そして, 演奏者がそれを演奏することで演奏情報が得られる.

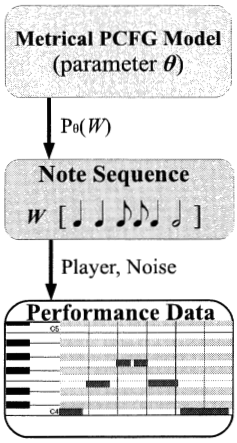

図 3 旋律生成の順方向プロセス

通常, 我々が観測できるのは図中の Note Sequence か Performance Data である. よってこれらから旋律生成モデルを推定することが必要になる. 旋律生成モデルは, 旋律に対する確率を与えると言う意味で, 我々が持つ音楽に対する事前知識と捉えられる. 人間のはこの事前知識によって, 演奏情報が揺らいでいてもある程度音符列を把握し, 耳で聴いた音列から一つの拍節構造を知覚できる.

Nonterminal Symbols	**Terminal Symbols**
S, Beat$x$, Ny ($x \in \mathbf{A}, y \in \mathbf{B}$)	note:x, rest:x ($x \in \mathbf{B}$)
Production Rules	
S $\rightarrow$ Nx Beat$x$ ($x \in \mathbf{A}$)	
S $\rightarrow$ Nx Beat$y$ ($x \in \mathbf{B}, y \in \mathbf{A}, x < y$)	(2)
Beat$x$ $\rightarrow$ Nx Beat$x$ ($x \in \mathbf{A}$)	(1)
Beat$x$ $\rightarrow$ Nx Nx ($x \in \mathbf{A}$)	(1)
Beat$x$ $\rightarrow$ Nx Ny ($x \in \mathbf{A}, y \in \mathbf{B}, y < x$)	(2)
Nx $\rightarrow$ Ns Nt ($x, s, t \in \mathbf{B}$, s + t = x)	
Nx $\rightarrow$ note:x ($x \in \mathbf{B}$)	
Nx $\rightarrow$ rest:x ($x \in \mathbf{B}$)	
where $\mathbf{A}$ = {4/4, 3/4} $\mathbf{B}$ = {1/1, 1/2, 1/4, 1/8, 1/16, 1/32, 3/4, 3/8, 3/16, 3/32}	

表 1 Grammar of Metrical PCFG

3.1.1 拍節 PCFG モデル

我々は, 旋律生成モデルとして表 1 に示されるような拍節構造に対する PCFG を用意し, 拍節 PCFG モデルと呼ぶことにする.

拍節 PCFG モデルでは, 開始記号 S が旋律全体を表す. 導出は S から始まり, 小節を表す Beat$x$(x には長さが入る.) を生成し,Beat$x$ は規則群 (1) によって同じ x を持つ Beat$x$ Beat$x$... と, 連続した小節を生成する. 個々の Beat$x$ からは, 音符の長さを意味する Nx が生成され,Nx はそのまま x の長さの音符である終端記号 $note:x$ が出力されるか, もしくはさらに小さな Nx $\rightarrow$Ns Nt と, 再帰的にリズムが分割される.

また, アウフタクトで始まる楽曲や, 終了小節が小節の長さより短く終わる場合などを考慮し, 規則群 (2) のような生成規則を追加している.

例として, 図 4 と図 5 のような音符列の導出木を考える. 音程については次節で述べるが, 簡単の為ここでは考えない. 図中の音列は, 全音符の長さを 1 として,W={1/4, 1/8, 1/8, 3/8, 1/8, 1/2} と表される音符列を, 二つの拍節の取り方で楽譜に表したものである (小節を埋めるため休符を挿入してある). 上記の音列を聴いたとき, 聴者は自然と図 4 のような拍節構造を知覚すると考えられる. これは, 次のように説明できる. 図 4 に示される導出中の生成規則は $N1/1- > N1/2 \quad N1/2$

のように音楽のリズム構造として「自然」なものが多い. 一方, 図 5 中の生成規則は $N1/1 -> N5/8 \quad N3/8$ のような「ありそうもない」為, 導出確率 $P(W,D)$ が低くなる.「自然な」とか「ありそうにない」などを,PCFG の確率パラメータとして推定する方法については 3.2 節で述べる.

具体的な観測データが与えられた時の $P_\theta(W)$ を θ をパラメータとした尤度, もしくは尤度関数と呼ぶ. 全観測データの尤度を最大にするようなパラメータ θ が, 求める旋律生成モデルである. また, ある θ の下で, 尤度が最大となる導出木を最尤導出として旋律の拍節構造と見なす.

このように, 確率モデルによって尤度と言う単一の尺度で評価が可能になる.

3.1.2 音程情報を含めた拡張

3.1.1 節で述べた拍節 PCFG モデルは音程情報を考えないモデルであった. ここでは, 音程の情報を付加したモデルを考える. 音程情報は次のように付加する.

Nx[Beat, Position] → By[Beat, Position'] Cz[Beat, Position'']

Nx[Beat, Position] → note:x[Degree]

ここで,Beat は $\frac{4}{4}$ のように拍子を表し,Position は $\frac{1}{4}$ のような拍の位置を表す. また,Degree はスケール上での度数を表す.

これらの規則により,「$\frac{4}{4}$ 拍子の一拍目には C が来やすい」と言った知識が N1/4[4/4, 1/4] → note:1/4[1], のように表現可能になる. これによって, アルペジオの続くような入力に対しても音程情報から $P_\theta(W)$ の区別をつけることができると期待できる.

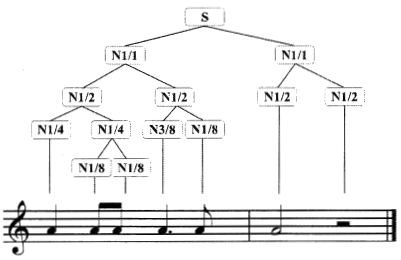

図 4　尤度 大

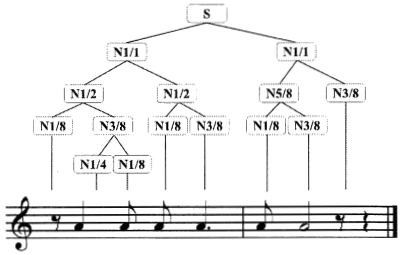

図 5　尤度 小

3.2 モデル推定

既存楽曲からのモデルの確率パラメータ推定は, 次式で表されるように学習データの尤度を最大にする $\hat{\theta}$ を求めることである.

$$\hat{\theta} = \arg\max_{\theta} P_\theta(W)$$

ここで,θ は PCFG における生成規則の確率パラメータの集合,W は学習データである. 学習データの尤度を最大にするパラメータ $\hat{\theta}$ は,PCFG の文法が曖昧で無い場合や導出木が与えられる場合は, 観測データの導出木中で使われる生成規則の数を単純に数え上げることで求められる.

$$P(A \to \alpha) = \#\ (A \to \alpha) / \sum_{\beta} \#\ (A \to \beta)$$

$\#\ (x)$ は, 観測データ中で使われた生成規則 x の回数である. しかし, 入力は音符列であり, 本稿で扱う文法は曖昧性を含む. その為, このような不完全データからのモデル推定には,EM アルゴリズムを使う.

3.2.1 EM アルゴリズム

EM(Expectation Maximization) アルゴリズムは名前の通り, モデルの, 学習データに対する期待値を最大化するようにパラメータを推定する手法である. 具体的には, 初期のパラメータ θ_0 から始め, 現在のパラメータ θ_i での, 観測データに対する尤度が大きくなるようにパラメータを更新する. このステップを複数回繰り返し, 収束したパラメータを解とする. 実際は尤度に対する山登り法なので, 初期値に依存する.

PCFG においては,Inside-Outside アルゴリズム [6] と呼ばれる効率的な EM アルゴリズムが存在する. 確

率パラメータの更新式は,

$$\tilde{P}(A \to \alpha) = \acute{\#}(A \to \alpha) / \sum_{\beta} \acute{\#}(A \to \beta)$$

で与えられ,$\acute{\#}(x)$ は,Inside-Outside アルゴリズムで推定される,現在のパラメータ θ_i での生成規則 x の回数となる. 詳細は参考文献 [10] を参照.

また, 括弧付けされ, 部分的に導出が付加されているデータに関しても, そのデータを使って少ない計算量でパラメータを求める方法が提案されており [8], 本稿では既存楽曲からのパラメータ推定で, 小節の区切りの情報を使う.

4 Experiments

モデルの有効性を評価するため, 小節構造を認識する実験を行なった. 実験の目標は,3.2 節の方法で推定した拍節 PCFG モデルによって未知の楽曲の拍節構造がどの程度認識できるかを評価することである. おおまかな流れは次の通りである.

4.1 Input Data

まず, 学習データの楽曲群は小節構造付きの旋律データに変換される. 入力データは単旋律で, 以下に示すように音価とスケール上での度数をペアにしたリストで与える.

W = [(note:1/4[5]) (note:1/2[1] note:1/4[7] note:1/4[1]) (note:1/2[6] note:1/4[5] note:1/4[1]]

なお, 学習データ中の小節位置を利用するため, 小節線に相当する場所に括弧を挿入している. 学習データには,Essen Folksong Collection [5] を使用した. 主にヨーロッパの民謡を集めたもので, 単旋律であり, 小節とフレーズの情報メタデータが付加されている. 我々が構築した拍節 PCFG モデルでは,$\frac{4}{4}$, 及び $\frac{3}{4}$ の長さの小節のみを扱うため, 拍子が $\frac{3}{2}$,$\frac{4}{2}$ の楽曲などは除いた.

全部で約 3400 曲の学習データから, その中で学習に使う楽曲の割合をランダムに learning Ratio={0.2, 0.4, 0.6, 0.8} と割合を変えて学習データを分割した.

4.2 Training and Test

次に,EM アルゴリズムによってパラメータ推定を行った.EM アルゴリズムの繰り返し回数は 20 回とした. テスト用のデータとして, 同じデータベース中から音列だけを取り出したデータを作り, 得られた拍節 PCFG モデルを使い, 正しい認識が出来ているかを評価した. 拍節 PCFG モデルは,3.1.1 節で述べた拍節だけのモデルと,3.1.2 での音程情報を使ったモデルの両方を推定した. なお, 拍節だけのモデルでは音程情報は無視している. その後, 最尤導出をその音列に対する拍節構造として求め, 実際に付加されている小節線の位置と比較をした.

4.3 Result

実験結果を図 6・7 に示す. 横軸はパラメータ推定に使用した学習データの全体に対する割合であり, 大きいほど大量のデータを使ったと言える. 縦軸はテストデータの F-score で表示してある.F-score は, 以下の式で与えられる評価尺度であり, 精度と再現率の調和平均である.

- 精度 $= \frac{\text{小節線一致数}}{\text{予想した小節線数}}$
- 再現率 $= \frac{\text{小節線一致数}}{\text{実際の小節線数}}$
- $F - score = \frac{2*\text{精度}*\text{再現率}}{\text{精度} + \text{再現率}}$

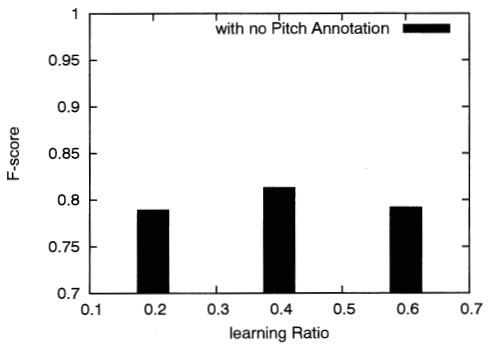

図 6 Pitch 情報を含まない場合

音程情報を使わない拍節 PCFG モデルでは, 最大で Learning Ratio が 0.4 での F-score=0.813, 平均 0.79 程度であった. また, 間違った導出例を見ると 4 分音符の連続のようなリズムに変化の無い楽曲が多く, こう言った楽曲に対してはリズム情報だけでは限界があることが示唆される. 一方, 音程情報を使用した方は, 最大で Learning Ratio=0.6 での 0.882, 平均 0.87 と相対的に高い値が得られた. この結果にから, 音程情報を用いるこ

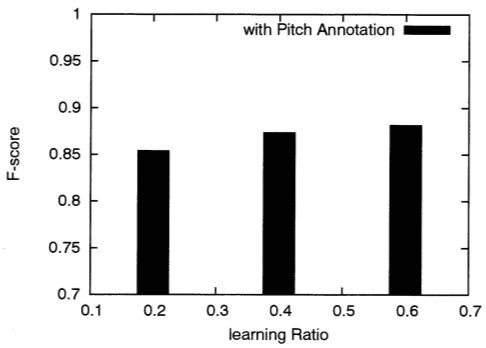

図 7　Pitch 情報を含む場合

とでより精度が上げられることが確認できた.

使用する学習データ数 (Learning Ratio) の変化に関しては, 認識率と正の相関があると予想され, 図 7 ではその傾向が見られるが, 図 6 では 0.4 より 0.6 での成績が低くなる例も見られた. これは, 音程情報のパラメータが必要ないため, 学習データの量が 0.4 程度で充分であった可能性があり, 詳細は今後の研究課題とする. また, 実際に生成規則の確率を見ても 0.4 以上での変化はあまり見られなかった.

5　Conclusion,Future Work

本稿で, 我々は確率文脈自由文法を使った拍節 PCFG モデルによる, 楽曲の拍節構造のモデル化と, その評価を行なった. 実験より, 音程情報を使わないモデルでも平均 0.79 程度の認識率であることから, 最尤導出により実際の小節位置をある程度正しく表すモデルが得られていることが分かる. 音程情報を組み入れたモデルでは,F-score が平均 0.81 という数値が得られ, 拍節以外の要素でより詳細なモデルを構築できることが考えられる. ただし, パラメータ数の増加は, 一般に必要とする学習データを増大させてしまうため, 本質的に必要なパラメータを選ぶ必要がある.

この様な数理的なモデルで, 拍節構造がモデル化できることが示された. 楽曲のモデル化は, その構造が理解できる点や, 自動生成に使えることから, 応用があると考えられる. 今後の目標として, 以下の目標が挙げられる.

- 今回の音程情報のように, 拍節以外の情報をモデルに組み入れる.
- 混合モデルなどを使ってパターンの抽出を行う.
- 入力として演奏音列や音響信号を扱う.

以上の要素を踏まえ, 統合的なモデルを目指して行きたい.

参考文献

[1] R Bod. Probabilistic grammars for music. In *Belgian-Dutch Conference on Artificial Intelligence*, 2001.

[2] Noam Chomsky. *Syntactic Structures*. The Hague: Mouton, 1957.

[3] 山本 et al. 確率文脈自由文法を用いた音楽演奏 midi データのリズム・テンポの推定. In **日本音響学会講演論文集**, pages 571–572, 2006.

[4] E. Gilbert and D. Conklin. A probabilistic context-free grammar for melodic reduction. In *International Workshop on Artificial Intelligence and Music*, 2007.

[5] H.Schaffrath. *The Essen Folksong Collection in the Humdrum Kern Format*. Menlo Park, CA: Center for Computer Assisted Research in the Humanities, 1995.

[6] J. D. Lafferty. A derivation of the inside-outside algorithm from the em algorithm. Technical report, IBM Research Report, 1993.

[7] Fred Lerdahl and Ray Jackendoff. *A Generative Theory of Tonal Music*. The MIT Press, 1983.

[8] F. Pereira and Y. Schebes. Inside-outside reestimation from partially bracketed corpora. *30th Annual Meeting of the Association for Computational Linguistics (ACL-93)*, pages 128–135, 1992.

[9] Hasegawa Yoshihiko and Hitoshi Iba. Estimation of distribution algorithm based on probabilistic grammar with latent annotations. In *IEEE Congress of Evolutionary Computation*, 2007.

[10] 北研二. **確率的言語モデル**. 言語と計算 4. 東京大学出版会, 1999.