相対的和音進行に基づく 和音進行解析のための語彙フリー無限 Gram モデル

田中 一生[1,a) 井上 真郷[1,b)

概要：本発表では，和音進行に特化した言語モデルを提案する．英語等の自然言語文章を扱う言語モデルの研究は盛んに行われており，和音進行に対しても応用することができる．しかし，和音進行には「調によって各語彙 (和音) の持つ役割が異なる」という性質があるため，多義語だらけの文章であると解釈できる．既存の研究では，1) 転調しない曲のみを対象とする，2) ハ長調/イ短調になるようにあらかじめ移調させておく，という方法でこの問題を解決している．しかし，この方法には「転調する曲」や「調が未知の曲」に対応できないという問題がある．これらを解決するために「相対的和音進行」という，調を明示的に考慮しないモデルを提案する．実験の結果，頻繁に転調するジャズの和音進行データに対して，従来のモデルよりも低い perplexity を達成した．

キーワード：ノンパラメトリック Bayes，Pitman-Yor process，言語モデル，和音進行

A Vocabulary-Free Infinity-Gram Model Based on Relative Chord Progression for Chord Progression Analysis

Tanaka Kazuki[1,a) Inoue Masato[1,b)

Abstract: This paper proposes a specialized language model for chord progression. The models for natural languages are actively studied, and we try to use them for a chord progression model. There is one clear problem; chord progression has a feature that each chord has different meaning depending on the key in music, so chord progression is contrasted as ”text containing many multisense words.” Existing methods solve this problem by 1) limiting compositions without modulation, and 2) transposing the key into the default key (C major or A minor), but these methods are cannot be applied for ”modulation-containing” or ”key-unknown” compositions. To solve this problem, we propose the model which deals with keys not directly but indirectly, named ”relative chord progression.” We validated our method by using jazz compositions, which contain many modulations, and found that our method overcomes existing methods in the sense of achieving the lower perplexity for chord progression.

Keywords: nonparametric Bayes, Pitman-Yor process, language modeling, chord progression

1. はじめに

和音進行は音楽の重要な要素の一つであるハーモニーを記号で表現したものであり，音楽を理解する上で重要な役割を担っている．そのため，音楽情報処理の分野において楽曲の録音データから和音進行を認識する「和音進行認識」は重要なテーマであり様々な研究が行われてきた [1].

和音進行認識では楽曲の音響的特徴のみではなく和音同士の前後関係や文脈を表現した確率モデル（言語モデル）を考慮することで認識性能が向上する [1]．このため，和音進行に適した言語モデルは音楽情報処理において有用であり，和音進行を自動生成することができるため自動作曲やアレンジ等にも役立つと思われる．

一方，音楽には「調との相対的な音高関係によって，各

1 早稲田大学 先進理工学研究科 電気・情報生命専攻
a) kazuki-ta@akane.waseda.jp
b) masato.inoue@eb.waseda.ac.jp

和音の役割が異なる」という性質がある．これに対して，既存研究ではデータを事前に処理し調をそろえることで対応している．しかし，この方法では「転調を含む曲」や「調が未知の曲」に対して対応できない．この問題を解決するため，和音進行自体を相対的に解釈するモデルを提案する．

本論文では，言語モデルの良さを評価するため，予測精度を perplexity で評価した．

2. 統計的言語モデル

2.1 問題設定

単語 (和音進行であれば和音) の集合である語彙を

$$\mathbf{W} \equiv \{1, \cdots, V\} \tag{1}$$

とし, 単語を $w \in \mathbf{W}$ とする．単語列の生成モデル $P(\mathbf{x} \mid \mathbf{\Theta})$ および事前分布 $P(\mathbf{\Theta})$ によって，学習データである単語列の集合

$$\mathbf{X} \equiv \left\{ \mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(L)} \right\} \tag{2}$$

$$\mathbf{x}^{(l)} \equiv \left\{ x_1^{(l)}, \cdots, x_{M_l}^{(l)} \right\} \tag{3}$$

が得られたとする．ただし，$x_m^{(l)} \in \mathbf{W}$ である．このとき，新たな単語列を $\mathbf{x}^{(L+1)}$ として同時分布を

$$P\left(\mathbf{x}^{(L+1)}, \mathbf{X}, \mathbf{\Theta}\right) \equiv \left[\prod_{l=1}^{L+1} P\left(\mathbf{x}^{(l)} \mid \mathbf{\Theta}\right) \right] P(\mathbf{\Theta}) \tag{4}$$

と定義する．本研究の目的は，$\mathbf{X}$ が与えられた際に新たな単語列 $\mathbf{x}^{(L+1)}$ の予測分布

$$P\left(\mathbf{x}^{(L+1)} \mid \mathbf{X}\right) = \left\langle P\left(\mathbf{x}^{(L+1)} \mid \mathbf{\Theta}\right) \right\rangle_{P(\mathbf{\Theta} \mid \mathbf{X})} \tag{5}$$

を求めることである．ただし，$P(\mathbf{\Theta} \mid \mathbf{X})$ は解析的に計算できないため，サンプリングによって近似して求めることとする．また，表記を単純化するため $L = 1$ とするが，容易に複数の単語列に拡張することができる．

2.2 n-Gram モデル

n-gram モデルでは，単語 x_m が直前の $n-1$ 個の単語列 (コンテキスト) $\mathbf{u}_n \equiv \left\{ x_{m-(n-1)}, \ldots, x_{m-1} \right\}$ のみに依存した確率で生成されると仮定する．

$$P(\mathbf{x} \mid \mathbf{\Theta}) \equiv \prod_{m=1}^{M} P(x_m | \mathbf{u}_n, \mathbf{\Theta}) \tag{6}$$

つまり，学習データ $\mathbf{X}$ が与えられた際にコンテキスト $\mathbf{u}$ に続く単語 w の確率 $P(w \mid \mathbf{u}_n, \mathbf{X})$ を求めることができればよい．

2.3 Hierarchical Pitman-Yor Language Model (HPYLM)

HYPLM[2] は n-gram モデルを階層的に生成することができる言語モデルである．

2.3.1 Pitman-Yor Process (PY)

まず，HPYLM の基となる PY について説明する．PY は Dirichlet Process (DP) を自由度が増すように拡張した確率過程であり，確率分布を生成する．任意の確率分布 G_0 が与えられたとき，PY によって生成される確率分布 G を以下の式で定義する．

$$G \sim \mathrm{PY}(d, \theta, G_0) \tag{7}$$

ここで，$d \in [0, 1]$ はディスカウントパラメータ，$\theta > 0$ は集中度，G_0 を基底測度と呼ぶ．PY によって生成される確率分布 G は G_0 と同じ定義域を持っており，θ の値が大きければ大きいほど G_0 に似た分布が生成されやすくなる．

2.3.2 定式化

HPYLM は PY を階層化することで定義できる．1-gram 分布 $G(w \mid \mathbf{u}_1)$ が得られたとする．このとき，$\mathbf{u}_2$ が与えられた際の 2-gram 分布 $G(w \mid \mathbf{u}_2)$ は $G(w \mid \mathbf{u}_1)$ とは異なるが, 似通った分布であると考えられる．そのため, $G(w \mid \mathbf{u}_2)$ が $G(w \mid \mathbf{u}_1)$ を基底測度とした PY によって生成されたと仮定する．

$$G(w \mid \mathbf{u}_2) \sim \mathrm{PY}(d_1, \theta_1, G(w \mid \mathbf{u}_1)) \tag{8}$$

一般化すると, 長さ $n-1$ のコンテキスト $\mathbf{u}_n$ からの n-gram 分布 $G(w \mid \mathbf{u}_n)$ が $G(w \mid \mathbf{u}_{n-1})$ を基底測度とする PY から生成されたものと考える．

$$G(w \mid \mathbf{u}_n) \sim \mathrm{PY}(d_n, \theta_n, G(w \mid \mathbf{u}_{n-1})) \tag{9}$$

この生成過程は再帰的に定義されており，1-gram 分布は

$$G(w|\mathbf{u}_1) \sim \mathrm{PY}(d_1, \theta_1, G(w \mid \mathbf{u}_0)) \tag{10}$$

として生成されたものと考える．ただし，$\mathbf{u}_0, \mathbf{u}_1$ は空集合とする．ここで，$G(w \mid \mathbf{u}_0)$ は 1-gram 分布の基底測度であり，すべての n-gram 分布が $G(w \mid \mathbf{u}_0)$ を間接的に参照して生成される．以上より，HYPLM は深さ n の木構造で表現することができる．根ノードから注目しているノードへ木をたどることは，コンテキストに含まれる単語を 1 つずつ後ろにさかのぼって行くことを意味する．

2.3.3 生成モデル

観測データ $\mathbf{X}$ の生成過程を考える．その際，HPY と等価な確率過程である Chinese restaurant franchise (CRF) を用いる．CRF では，コンテキスト $\mathbf{u}$ はレストラン (ノード) であり，観測データ $\mathbf{X}$ は客の集合，単語 w は料理に例えられる．各レストランには上限のない数のテーブルが存在し，到着した客はそのいずれかに着席する．各テーブルには一種類の料理が置かれており，そのテーブルに座る客は同じ料理を食べることとする．

$m-1$ 人の客が既に到着しているとき，m 人目の客がどのようにふるまうかについて考える．まず，客 x_m が深さ

n のノード $\mathbf{u}_n$ に到着し，(1)$c_{\mathbf{u}_n wk} - d_n$ に比例する確率で既存テーブル k に着席するか，(2)$d_n t_{\mathbf{u}_n \cdot} + \theta_n$ に比例する確率で新規テーブル $k' = t_{\mathbf{u}_n \cdot} + 1$ に着席する．ここで，$c_{\mathbf{u}_n wk}$ はノード $\mathbf{u}_n$ のテーブル k において料理 w を食べている客の人数であり，$t_{\mathbf{u}_n w}$ はノード $\mathbf{u}_n$ における料理 w が置いてあるテーブルの数である．新規テーブルに着席した場合，客 x_m は注文する料理を決定するため，自身のコピーである代理客を親ノード $\mathbf{u}_{n-1}$ へと送る．代理客は親ノードで，通常の客と同様に着席する．(1) 既存テーブルに着席した場合，そのテーブルに置かれている料理 w を食べることになり，本来の客 x_m の座るテーブルに置かれる料理も w と決定する．(2) 新規テーブルに着席した場合，さらに親ノードに代理客を送り，同様の動作を料理が決定するまで再帰的に行う．その際，もし根ノードでも新規テーブルに着席した場合，そのテーブルに置かれる料理 w は基底測度 $G(w \mid \mathbf{u}_0)$ によって決定される．

以上より，これまでに到着した客の着席状態を $\mathbf{S}$ とし，$\mathbf{d} \equiv \{d_i\}_{i=1}^n$, $\boldsymbol{\theta} \equiv \{\theta_i\}_{i=1}^n$ および $G(w \mid \mathbf{u}_0)$ が与えられたとき，$\mathbf{S}$ の事前分布は $P(\mathbf{S} \mid \mathbf{d}, \boldsymbol{\theta}, G(w \mid \mathbf{u}_0))$ と書くことができる．また，パラメータ $\boldsymbol{\Theta} \equiv \{\mathbf{d}, \boldsymbol{\theta}, \mathbf{S}\}$ とコンテキスト $\mathbf{u}_n$ が与えられたとき，単語 w の生成モデルを以下の式で定義する．

$$P(x_m{=}w|\mathbf{u}_n, \boldsymbol{\Theta}) \equiv \frac{c_{\mathbf{u}_n w} - d_n t_{\mathbf{u}_n w}}{c_{\mathbf{u}_n \cdot\cdot} + \theta_n} + \frac{d_n t_{\mathbf{u}_n \cdot} + \theta_n}{c_{\mathbf{u}_n \cdot\cdot} + \theta_n} P(x_m{=}w|\mathbf{u}_{n-1}, \boldsymbol{\Theta}) \tag{11}$$

ここで，上式は再帰的な定義となっていることに注意する．

2.3.4 予測分布と Beyes 推定

尤度関数と事前分布を用いて，予測分布を求めると以下の通りとなる．

$$P(x_m{=}w|\mathbf{u}_n, \mathbf{X}) = \langle P(x_m{=}w|\mathbf{u}_n, \boldsymbol{\Theta}) \rangle_{P(\boldsymbol{\Theta}|\mathbf{u}_m, \mathbf{X})} \tag{12}$$

ただし，$P(\boldsymbol{\Theta}|\mathbf{u}_m, \mathbf{X})$ を計算するのは計算量の問題で難しい．$\mathbf{X}$ が $\mathbf{u}_m$ より十分大きければ以下の式で近似できる．

$$P(x_m{=}w \mid \mathbf{u}_n, \mathbf{X}) \simeq \langle P(x_m{=}w \mid \mathbf{u}_n, \boldsymbol{\Theta}) \rangle_{P(\boldsymbol{\Theta}|\mathbf{X})} \tag{13}$$

この和は解析的に計算することはできないが，Gibbs sampling を用いると以下の式で近似することができる．

$$P(x_m{=}w|\mathbf{u}_n, \mathbf{X}) \simeq \frac{1}{L_{\mathrm{sam}}} \sum_{l=1}^{L_{\mathrm{sam}}} P(x_m{=}w|\mathbf{u}_n, \boldsymbol{\Theta}_l) \tag{14}$$

$$\boldsymbol{\Theta}_l \sim P(\boldsymbol{\Theta} \mid \mathbf{X}) \tag{15}$$

ここで，L_{sam} は事後分布からサンプリングされたサンプルの数である．

具体的な Gibbs sampling のアルゴリズムについて述べる．まず，空の木に全ての客を「事後的」な CRF に従って一人ずつ加えていくことで着席方法 $\mathbf{S}$ の初期値を得る．「事後的」とは，客の食べる料理はすでに決まっている (学習データとして与えられている) 上で着席方法のみを決定するという意味である．したがって，$x_m = w$ であるとき，客 x_m は (1)$c_{\mathbf{u}_n wk} - d_n$ に比例する確率で料理 w が置かれている既存テーブルに着席するか，(2)$(d_n t_{\mathbf{u}_n \cdot} + \theta_n) P(w \mid \mathbf{u}_{n-1}, \boldsymbol{\Theta})$ に比例する確率で新規テーブルに着席し，料理 w を食べるかの二通りとなる．次に，ある客 x_m を木から削除し，最後に到着した客とみなして再度木に加えることで新たなサンプルを生成する．客を削除する際には，その代理客も全て削除することに注意する．全ての客に対してランダムな順番で同様の操作を繰り返すことで Gibbs sampling を行うことができる [2].

また，$\mathbf{d}$ および $\boldsymbol{\theta}$ も事前分布を設定し Bayes 推定する．d_i は Beta 分布を，θ_i は Gamma 分布を事前分布とする．

$$P(\mathbf{d}, \boldsymbol{\theta}) \equiv \prod_{i=1}^{n} \mathrm{Beta}(d_i|a_i, b_i)\, \mathrm{Gamma}(\theta_i|\alpha_i, \beta_i) \tag{16}$$

補助変数を介した事後分布 [3] を導出し，サンプリングする．

2.4 Variable-order Pitman-Yor Language Model (VPYLM)

VPYLM[4] は HPYLM を改良しそれぞれの客が到着するノードの深さを可変としたモデルである．HPYLM の問題点はすべての客が深さ n のノードに必ず到着する点であり，これは単語 w が必ず $n-1$ の長さのコンテキストに従って生成されるということを意味する．しかし，ある単語を予測するとき必要なコンテキストの長さは文脈によって異なる．VPYLM ではこの問題を解決するため，観測値 x_m に対して n-gram の長さを表す潜在変数 z_m が存在すると仮定する．z_m に関して周辺化することであらゆる z_m の可能性を考慮した単語の予測確率を求めることができる．

2.4.1 定式化

VPYLM は，客 x_m が入るノードの深さ z_m の生成過程をモデル化することで定義できる．潜在変数 z_m の事前分布として，VPYLM では stick-breaking process (SBP)[5] を利用する．

潜在変数 z_m の生成過程を考える．客 x_m がある深さのノードに到着する際，x_m は根ノードから x_{m-1}，x_{m-2} に対応するノードへと木を下っていく．このとき，客 x_m は各ノード $\mathbf{u}_i$ において確率 $\eta_{\mathbf{u}_i}$ にしたがってその場で停止するか，さらに下のノードへと下るかを決定する．この過程を経た結果，客が深さ n のノードで停止する確率は以下の式で定義できる．

$$P(z_m = n \mid \mathbf{u}_m, \boldsymbol{\eta}) \equiv \eta_{\mathbf{u}_n} \prod_{i=1}^{n-1} (1 - \eta_{\mathbf{u}_i}) \tag{17}$$

ここで，$\boldsymbol{\eta}$ は Beta 分布を事前分布とし Bayes 推定する．

$$P(\boldsymbol{\eta}) \equiv \prod_{\mathbf{u} \in \mathrm{tree}} \mathrm{Beta}(\eta_{\mathbf{u}} \mid \alpha, \beta) \tag{18}$$

2.4.2 生成モデル

データ $\mathbf{X}$ の生成過程を考える．m 番目の客 x_m に対して，潜在変数 z_m の値が SBP によって確率的に決まる．z_m の値が決まると，それに応じた深さのノードに到着し，CRF によって x_m が決まる．その際，HPYLM では深さ n のノード以外には代理客しか存在しなかったが，VPYLM では本来の客と代理客が混在することになる．しかし，$c_{\mathbf{u}w}$ をカウントする際にはこれらを区別しないものとする．

以上より，パラメータ $\mathbf{\Theta}$ および $\boldsymbol{\eta}$ が与えられたときに，m 番目の単語 x_m と潜在変数 z_m の同時確率は以下の式で表される．

$$\begin{aligned} &P\ (x_m{=}w, z_m{=}n|\mathbf{u}_m, \mathbf{\Theta}, \boldsymbol{\eta}) \\ &= P\left(x_m{=}w|z_m{=}n, \mathbf{u}_m, \mathbf{\Theta}\right) P\left(z_m{=}n|\mathbf{u}_m, \mathbf{\Theta}, \boldsymbol{\eta}\right) \end{aligned} \tag{19}$$

ただし，$P\left(x_m{=}w|z_m{=}n, \mathbf{u}_m, \mathbf{\Theta}\right)$ は式 11 の n に z_m を代入することで得られる．n と $\boldsymbol{\eta}$ に関して周辺化すると以下の式が得られる．

$$\begin{aligned} &P\ (x_m{=}w \mid \mathbf{u}_m, \mathbf{\Theta}) \\ &= \sum_n P\left(x_m{=}w|z_m{=}n, \mathbf{u}_m, \mathbf{\Theta}\right) P\left(z_m{=}n|\mathbf{u}_m, \mathbf{\Theta}\right) \end{aligned} \tag{20}$$

ここで，$P\left(z_m = n \mid \mathbf{u}_m, \mathbf{\Theta}\right)$ は以下の式で求められる．

$$\begin{aligned} P\left(z_m{=}n|\mathbf{u}_m, \mathbf{\Theta}\right) &= \int P\left(z_m{=}n|\mathbf{u}_m, \mathbf{\Theta}, \boldsymbol{\eta}\right) P\left(\boldsymbol{\eta} \mid \mathbf{\Theta}\right) d\boldsymbol{\eta} \\ &= \frac{a_{\mathbf{u}_n} + \alpha}{a_{\mathbf{u}_n} + b_{\mathbf{u}_n} + \alpha + \beta} \times \\ &\quad \prod_{i=1}^{n-1} \frac{b_{\mathbf{u}_i} + \beta}{a_{\mathbf{u}_i} + b_{\mathbf{u}_i} + \alpha + \beta} \end{aligned} \tag{21}$$

ここで，$a_{\mathbf{u}_i}$ は，既に到着している客が木を下って行った際にノード $\mathbf{u}_i$ で停止した客の数であり，$b_{\mathbf{u}_i}$ はノード $\mathbf{u}_i$ を通過した客の数である．

2.4.3 予測分布と Bayes 推定

コンテキスト $\mathbf{u}$ の次に来る単語 w の確率は HPYLM の Gibbs sampling アルゴリズムを少し変更することで求めることができる [4]．まず，客 $x_m = w$ を代理客ごと木から削除し，潜在変数 z_m を以下の式からサンプリングする．

$$\begin{aligned} &P\ (z_m{=}n|x_m{=}w, \mathbf{u}_m, \mathbf{\Theta}) \propto P\left(x_m{=}w, z_m{=}n|\mathbf{u}_m, \mathbf{\Theta}\right) \\ &= P\left(x_m{=}w|z_m{=}n, \mathbf{u}_m, \mathbf{\Theta}\right) P\left(z_m{=}n|\mathbf{u}_m, \mathbf{\Theta}\right) \end{aligned} \tag{22}$$

求まった z_m の深さに客 x_m を追加し，HPYLM と同様に事後的な CRF によって着席させることで新たなサンプル $\mathbf{\Theta}$ を求めることができる．式 13,14 と同様に $\mathbf{\Theta}$ に関して周辺化し，予測確率 $P\left(x_m = w \mid \mathbf{u}_m, \mathbf{X}\right)$ を求めればよい．

2.5 語彙フリー無限 gram モデル (既存モデル)

吉井らは和音進行に対して，以下のような定式化と基底測度を定義することで和音進行に対する語彙フリー無限 gram モデル [1] を提案した．

2.5.1 定式化

各和音 w を根音 r と構成音の相対的な音高関係を表す 12 次元のベクトル $\mathbf{t}$ を使って以下の式で定義する．なお，N は無音の区間などを表現するための，どの和音にも属さない特殊な記号である．

$$w \equiv [r, \mathbf{t}] \tag{23}$$

$$r \in \{\mathrm{C}, \mathrm{C}\#, \cdots, \mathrm{B}, \mathrm{N}\} \tag{24}$$

$$\mathbf{t} \in \{0, 1\}^{12} \tag{25}$$

この定式化によって，12 半音の組み合わせをすべて表現することができる．例えば，C:maj の和音は根音が C で構成音は C,E,G であるため，C:100010010000 と表記する．ここで $\mathbf{t}$ は根音から見た相対音高を考えるため，F:maj(根音:F, 構成音:F,A,C) の和音は F:100001000100 ではなく，F:100010010000 であることに注意する．

2.5.2 基底測度

基底測度 G_0 は根音と各構成音の有無がすべて独立であるという仮定に基づいて定義する．r は categorical 分布に，t_i は Bernoulli 分布に従うと仮定すると，基底測度 $G\left(w \mid \mathbf{u}_0\right)$ は以下の式で定義できる．

$$G\left(w \mid \mathbf{u}_0\right) \equiv P\left(w \mid \boldsymbol{\pi}, \boldsymbol{\tau}\right) \equiv \pi_r \prod_{i=1}^{12} \tau_i^{t_i} \left(1 - \tau_i\right)^{1-t_i} \tag{26}$$

ただし，$r = \mathrm{N}$ であれば $G\left(w \mid \mathbf{u}_0\right) = \pi_{\mathrm{N}}$ とする．また，$\boldsymbol{\pi}$ と $\boldsymbol{\tau}$ の値は Dirichlet 分布と Beta 分布を事前分布とし Bayes 推定する．

$$P\left(\boldsymbol{\pi}, \boldsymbol{\tau}\right) \equiv \mathrm{Dir}\left(\boldsymbol{\pi} \mid \mathbf{a}_0\right) \prod_{i=1}^{12} \mathrm{Beta}\left(\tau_i \mid b_0, c_0\right) \tag{27}$$

この事前分布から事後分布を導出すると以下の式となる．

$$P\left(\boldsymbol{\pi}, \boldsymbol{\tau} \mid \mathbf{S}\right) = \mathrm{Dir}\left(\boldsymbol{\pi} \mid \mathbf{a}_0 + \mathbf{h}\right) \prod_{i=1}^{12} \mathrm{Beta}\left(\tau_i \mid b_0 + n_i, c_0 + \overline{n}_i\right) \tag{28}$$

ここで，h_r は根レストラン $\mathbf{u}_1$ において根音が r であるテーブルの数であり，n_i は $t_i = 1$ であるテーブルの数，$\overline{n}_i$ は $t_i = 0$ であるテーブルの数である．各ステップ毎にこの事後分布からサンプリングすることで $\boldsymbol{\pi}$ と $\boldsymbol{\tau}$ を求める [1]．以上より，パラメータは $\mathbf{\Theta} \equiv \{\mathbf{d}, \boldsymbol{\theta}, \mathbf{S}, \boldsymbol{\pi}, \boldsymbol{\tau}\}$ である．

3. 提案モデル

3.1 相対的和音進行

転調を含む場合や調が未知の場合でも対応できるモデルを定義するためには，調によって回転させられた同じ和音進行 (例：調が C major であるときの C:maj F:maj G:maj と，調が F major であるときの F:maj Bb:maj C:maj) を調の情報なしで同様に扱う枠組みが必要である．そのためには，注目している和音に対しそのコンテキストを相対的

にとらえればよい．例えば，二つの和音進行の 3 番目の和音に注目しているとして考える．一つ前の和音 x_{m-1} が根音 C となるように全体を回転させると，どちらの和音進行も G:maj C:maj D:maj となる．このように和音 x_m に対して，その一つ前の和音 x_{m-1} の根音がそろうように x_m 自身とコンテキストを回転させることで調の情報なしで同一の和音進行とみなすことができる．これを「相対的和音進行」と呼ぶことにする．相対的和音進行に基づき言語モデルを構成することで，調未知の曲や転調を含む曲に対応可能となる．

3.2 定式化

和音 w を r と $\mathbf{t}$ に加えて N であるか否かを表す変数 q の組み合わせであると定義する．

$$w \equiv [r, \mathbf{t}, q] \tag{29}$$

$$r \in \{\mathrm{C}, \mathrm{C}\#, \cdots, \mathrm{B}\} \tag{30}$$

$$\mathbf{t} \in \{0,1\}^{12} \tag{31}$$

$$q \in \{0,1\} \tag{32}$$

また，相対的和音進行では，注目している x_m の一つ前の和音 x_{m-1} の根音 x^r_{m-1} が C となるようにコンテキスト $\mathbf{u}$ と x_m を回転させる．ある和音 w をマイナス方向に f 分回転させる関数 R_{M} とプラス方向に回転させる関数 R_{P} を以下の式で定義する．

$$R_{\mathrm{M}}(w,f) \equiv \begin{cases} [(r-f+12) \bmod 12, \mathbf{t}, q] & (q=0) \\ w & (q=1) \end{cases} \tag{33}$$

$$R_{\mathrm{P}}(w,f) \equiv \begin{cases} [(r+f) \bmod 12, \mathbf{t}, q] & (q=0) \\ w & (q=1) \end{cases} \tag{34}$$

以上より，相対的和音進行におけるコンテキスト $\mathbf{u}'_n$ と注目している和音 x'_m はそれぞれ以下の式で定義できる．

$$\mathbf{u}'_n \equiv \left[R_{\mathrm{M}}\left(x_{m-(n-1)}, x^r_{m-1}\right), \cdots, R_{\mathrm{M}}\left(x_{m-1}, x^r_{m-1}\right)\right] \tag{35}$$

$$x'_m \equiv R_{\mathrm{M}}\left(x_m, x^r_{m-1}\right) \tag{36}$$

結果，和音 w の生成過程は以下のように定義できる．まず，コンテキスト $\mathbf{u}_n$ を回転させた $\mathbf{u}'_n$ に依存した確率によって w' が生成される．

$$w' \sim P(w' \mid \mathbf{u}'_n, \mathbf{X}) \tag{37}$$

生成された w' を一つ前の和音の根音 x^r_{m-1} でプラス方向に回転させることで，w が生成される．

$$w = R_{\mathrm{P}}\left(w', x^r_{m-1}\right) \tag{38}$$

以上の生成モデルをもとに VPYLM を構成する．

3.3 基底測度と 1-gram 分布

以上の定式化をもとに，VPYLM を定義する．実際には，$n=2$ 以降は従来の VPYLM と同様であるため，1-gram 分布と基底測度を定義すればよい．1-gram 分布 $P(w \mid \mathbf{u}_1)$ を以下の式で定義する．

$$P(w \mid \mathbf{u}_1) \equiv \frac{1}{12} \times G_1(\mathbf{t}, q) \tag{39}$$

$$G_1 \sim \mathrm{PY}\left(d_1, \theta_1, P(w \mid \mathbf{u}_0)\right) \tag{40}$$

ただし，$q=1$ であれば $P(w \mid \mathbf{u}_1) = G_1(\mathbf{t}, q)$ とする．ここで，根音 r に関する分布は $\frac{1}{12}$ で固定とし，G_1 は $\mathbf{t}$ と q のみに関する分布とする．そのため，基底測度 G_0 は Bernoulli 分布を使って以下の式で定義する．

$$\begin{aligned} G_0(\mathbf{t}, q) &\equiv P(\mathbf{t}, q \mid \zeta, \boldsymbol{\tau}) \\ &= \zeta^q (1-\zeta)^{(1-q)} \prod_{i=1}^{12} \tau_i^{t_i} (1-\tau_i)^{1-t_i} \end{aligned} \tag{41}$$

ただし，$q=1$ であれば $G_0(\mathbf{t}, q) = \zeta$ とする．ここで，ζ と $\boldsymbol{\tau}$ は未知であるため，Beta 分布を事前分布とし，Bayes 推定する．

$$P(\zeta, \boldsymbol{\tau}) \equiv \mathrm{Beta}(\zeta \mid a_0, b_0) \prod_{i=1}^{12} \mathrm{Beta}(\tau_i \mid a_0, b_0) \tag{42}$$

この事前分布から事後分布を導出すると以下の式となる．

$$\begin{aligned} P(\zeta, \boldsymbol{\tau} \mid \mathbf{S}) &= \mathrm{Beta}(\zeta \mid a_0 + n_q, b_0 + \overline{n}_q) \\ & \prod_{i=1}^{12} \mathrm{Beta}(\tau_i \mid a_0 + n_i, b_0 + \overline{n}_i) \end{aligned} \tag{43}$$

ここで，n_q は根レストラン $\mathbf{u}_1$ において N であるテーブルの数であり，$\overline{n}_q$ は N 以外であるテーブルの数である．また，n_i および $\overline{n}_i$ は式 28 と同様である．既存手法と同様に各ステップ毎に事後分布からサンプリングする [1]．以上より，パラメータは $\boldsymbol{\Theta} \equiv [\mathbf{d}, \boldsymbol{\theta}, \mathbf{S}, \boldsymbol{\pi}, \boldsymbol{\tau}, \zeta]$ となる．

3.4 Gibbs Sampling

提案モデルに対する Gibbs sampling アルゴリズムを考える．注目している単語 x_m に対して，回転されたコンテキスト $\mathbf{u}'$ と客 x'_m を求める．そのうえで VPYLM 同様 z_m をサンプリングし，客 x'_m が深さ z_m のノードに到着し CRF にしたがって着席する．その際，根ノード $\mathbf{u}_1$ に客が到着した場合には特殊な操作が行われる．根ノードでは，$\mathbf{t}$ と q のみの PY を構成する．具体的には，到着した客 x'_m を根音が C となるように再度回転させてから CRF によって着席させる．以上により，ノード $\mathbf{u}_1$ には根音が C の和音および N の和音のみが到着するため，擬似的に $\mathbf{t}$ と q のみの CRF を構成することができる．また，コンテキスト $\mathbf{u}'$ は x_{m-1} の根音が C となるように回転させているため $n=2$ の階層には根音が C の和音または N の和音であるノードしか生成されないことに注意する．以上のアルゴリズムで提案モデルの Gibbs sampling を行う．

4. 実験

4.1 実験条件

4.1.1 使用するデータ

実験用のデータセットとして，リットーミュージック出版「Jazz Standard Bible」[6] からジャズスタンダード 100 曲のコード進行データ (1 コーラス分) を用いた．これらの曲の和音進行はほとんどが転調を含んでいる．調が未知であるという条件で実験を行うため，調がそろうように移調する等の前処理は行わない．

4.1.2 学習設定

それぞれのモデルについて 200 回の Gibbs sampling を行い，$L_{\mathrm{sam}} = 50$ 個の事後サンプルを用いて評価を行った．ノードの伸ばしやすさを決める式 21 の Beta 分布のハイパーパラメータを $(\alpha, \beta) = (4, 1)$ とした [4]．また，n に関しては，計算量の都合から最大値を $N_{\mathrm{MAX}} = 11$ とした．

4.1.3 評価指標

評価指標には，言語モデルの予測精度の評価で用いられる perplexity を用いた．Perplexity は言語モデルの複雑さを表す指標であり，次の単語を予測する際に平均いくつの答えの候補があるかを表す．この値は小さいほど予測精度が高いモデルであるといえる．実験では，8-fold cross validation によってテストセット perplexity を測定し，10 回の相乗平均をとることで評価した．

4.2 結果・考察

既存モデルと提案モデルの比較実験の結果と所要時間を表 1 に示す．転調を含むジャズの和音進行に対し，提案モデルは低い perplexity を達成した．この結果から提案モデルはジャズなど転調を含む和音進行に対して有効であるといえる．これは，相対的に和音進行を解釈することで調の異なる同じコンテキストを同一とみなせるからであると考えられる．（学習データの量がもっと多ければ，既存モデルでも同様のモデルが学習されると考えられる．）

表 1 実験結果
Table 1 Results

モデル	perplexity	計算時間 [h]
既存モデル	19.77	29.38
提案モデル	12.40	20.89

また，データ量が同じであれば学習に必要な計算時間が提案モデルの方が少ない．これは，提案モデルの方が構成される木構造のサイズが小さいからである．n ごとのノード数の平均値を表 2 に示す．表 2 をみると，特に $n = 2$ においてノード数が少ないことがわかる．これは，提案モデルでは $n = 2$ のノードに根音が C である和音をコンテキストとしてもつものしか追加されないためである．また，$n = 5$ や $n = 6$ のノード数が既存モデルより多いことから，提案モデルはより高次の特徴を捉えていることがわかる．以上より，提案モデルは「転調を含む曲」「調未知の曲」に対して有効なモデルであると考えられる．

表 2 n ごとのノード数の平均
Table 2 Average number of nodes (each n)

n	既存モデル	提案モデル	n	既存モデル	提案モデル
1	1	1	7	20	30
2	142	21	8	6	10
3	554	294	9	2	3
4	417	369	10	1	1
5	184	238	11	0	1
6	63	100	-	-	-

5. 結論

本論文では和音進行に対する言語モデルとして相対的和音進行という概念を導入した語彙フリー無限 gram モデルを提案した．このモデルは和音進行の「調によって各語彙の持つ役割が異なる」という性質を考慮しており調が未知の曲や転調を含む曲であっても対応できる．比較実験の結果，転調を繰り返すジャズの和音進行データに対して従来のモデルより低い perplexity を達成した．また，木構造が圧縮され，計算時間の点でも優れていることが分かった．

参考文献

[1] 吉井 和佳，後藤 正孝：和音進行解析のための語彙フリー無限グラムモデル, 情報処理学会研究報告 Vol.2011-MUS-91，No.2，pp.1-10 (2011).

[2] Teh,Y.W.: A Hierarchical Bayesian Languange Model based on Pitman-Yor Processes, in *Proc. of COLING/ACL 2006*, pp.985-992 (2006).

[3] Teh,Y.W.: A Bayesian Interpretation of Interpolated Kneser-Ney，Technical Report TRA2/06, NUS School of Computing (2006).

[4] 持橋 大地，隅田 英一郎：階層 Pitman-Yor 過程に基づく可変長 n-gram モデル，情報処理学会論文誌，Vol.48，No.12，pp.4023-4032 (2007).

[5] Sethuraman,J.: A Constructive Definition of Dirichlet Priors, *Statistica Sinica*, Vol.4, pp.639-650 (1994).

[6] 納 浩一：Jazz Standard Bible セッションに役立つ不朽の 227 曲，株式会社リットーミュージック (2010).

[7] Teh,Y.W., Jordan,M., Beal,M. and Blei,D.: Hierarchical Dirichlet Process, Technical Report 653, Department of Statistics, University of California at Berkeley (2004).

[8] 持橋 大地：最近のベイズ理論の進展と応用 (III) ノンパラメトリックベイズ，電子情報通信学会論文誌，Vol.93，No.1，pp.73-79 (2010).

[9] Bishop,C.M., 元田 浩, 栗田 多喜夫, 樋口 知之, 松本 裕治, 村田 昇 (監訳)：パターン認識と機械学習 (上)(下) ベイズ理論による統計的予測，Springer (2007,2008).

[10] 持橋 大地，山田 武士，上田 修功：ベイズ階層言語モデルによる教師なし形態素解析，自然言語処理学会研究報告，No.2009-NL-190，pp.49，情報処理学会 (2009).