

和音進行解析のための語彙フリー無限グラムモデル

吉井和佳^{†1} 後藤真孝^{†1}

本稿では、和音系列に対する統計的言語モデルとして、ノンパラメトリックベイズ理論に基づく n グラムモデルについて述べる。従来の経験的なスムージングに基づく n グラムモデルには、主に三つの問題、すなわち、理論的な裏付けがなく改善が困難であること、 n の値を一意に指定しなければならないこと、考慮する和音の種類（語彙）は恣意的に決めざるを得ないこと、が存在した。これらの問題を解決するため、我々は語彙フリー無限グラムモデルを提案する。このモデルは、あらゆる音の組合せを和音として許容するため語彙が不要で、和音系列中の各和音が異なるコンテキスト長（理論上は無限でもよい）を持つことを許容する。ある和音系列が与えられた上で、次の和音を予測するときには、 n の値を一意に決めることなくあらゆる可能性を考慮できる。また、これまで観測したことがない和音が出現したとしても、その和音のゼログラム確率（構成音の同時出現確率）をこれまで観測してきた和音の構成音に基づいて計算することで、適切に n グラム確率を推定することができる。実験の結果、従来の n グラムモデルよりも低いパープレキシティを達成することが分かった。

A Vocabulary-Free Infinity-Gram Model for Chord Progression Analysis

KAZUYOSHI YOSHII^{†1} and MASATAKA GOTO^{†1}

This paper presents a novel nonparametric Bayesian n -gram model as a statistical language model for symbolic chord sequences. Standard n -gram models based on heuristic smoothing have three fundamental problems—that they have no theoretical foundation, that the value of n is fixed uniquely, and that a vocabulary of chord types is defined in an arbitrary way. To solve these problems, we propose a vocabulary-free infinity-gram model. It accepts any combinations of notes as chord types and allows each chord appearing in a sequence to have an unbounded and variable-length context. Our experiments showed that the perplexity obtained by the proposed model is significantly lower than that obtained by the state-of-the-art models.

^{†1} 産業技術総合研究所 (AIST)

1. はじめに

音楽の三大要素のひとつであるハーモニーと密接な関係がある和音は、作曲者のスタイル¹⁾ や楽曲のジャンル²⁾ を特徴づけるなど、重要な役割を担っている。そのため、音楽情報処理分野において、和音進行解析は取り組むべき重要な課題のひとつであり³⁾⁻⁶⁾、研究基盤としての和音系列データベースの整備も進められている^{5),6)}。例えば、音楽音響信号を対象とした和音認識を行う上では、和音スペクトルのダイナミクスを表現する音響モデルだけではなく、和音の連鎖のしやすさ（和音進行）を表現する確率モデル（言語モデル）を利用することで、認識精度を向上させることができる⁷⁾⁻⁹⁾。これは、標準的な音声認識システムにおいて、音声スペクトルのダイナミクスを表現する音響モデルだけではなく、単語の連鎖のしやすさを表現する言語モデルが利用されているのと同様である。さらに、音楽音響信号を対象とした自動採譜を行う上では、和音進行が適切にモデル化されていれば、各和音区間ごとに出現しやすい音高の候補を絞り込むことで採譜精度が向上すると考えられる。

これまで、和音系列に対する言語モデルとして、 n グラムモデルが広く利用されてきた^{1),2),4),6)-9)}。 n グラムとは、ある和音系列中に含まれる連続した n 個の和音からなる部分系列のことであり、 n グラムモデルは $n-1$ 次のマルコフ性の仮定に基づいている。すなわち、ある和音系列中の各和音はその直前の $n-1$ 個の和音（コンテキスト）のみから導出されていると仮定する。和音は時間局所的な依存性を持つことが知られているため、この仮定はある程度妥当であると考えられている。我々の目的は、長さ $n-1$ の任意のコンテキストが与えられたときに、次の和音の予測確率（ n グラム確率）を高精度に計算可能な n グラムモデルを学習することである。しかし、学習データとして利用可能な和音系列の分量は限られており、素朴な学習方法では学習データに現れない n グラムの予測確率は 0（学習データ以外のどのようなデータでも一切出現しない）と推定されてしまう。このようなゼロ確率問題は過学習によって引き起こされ、 n の値が大きくなるにつれて一層深刻になる。この問題を回避するため、これまでさまざまな経験的なスムージング方法が提案されてきた¹⁰⁾。

しかし、和音進行解析に従来の経験的なスムージング方法に基づく n グラムモデルを適用する上では、三つの本質的な問題があった。すなわち、1) n グラムモデルの理論的な裏付けがないこと、2) 各和音は異なるコンテキスト長を持っているにもかかわらず、 n の値を一意に決める必要があること、3) 語彙として考慮すべき和音の種類や粒度（例：major, minor, augmented, diminished, seventh, ninth, …, それらの派生形）が恣意的に決められること、が問題であった。特に、3) はこれまで議論されたことがない問題であった。例

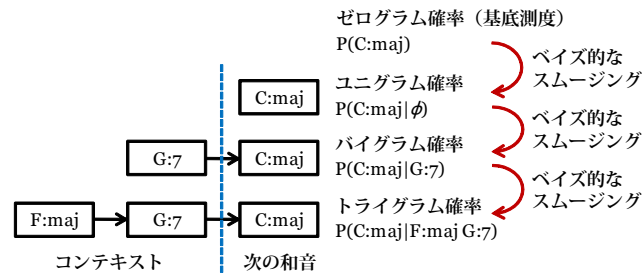


図 1 階層ノンパラメトリックベイズモデルによる n グラム確率のスムージング

例えば、和音認識タスクにおいては、語彙に含まれない和音は「未知の和音」としてひとくくりにするか、語彙に含まれる和音のうちで最も近いものとみなすことが一般的であった。そのため、認識結果や学習される n グラムモデルの間に互換性がなかった。また、自動採譜タスクにおいては、和音系列に対する n グラムモデルを得られた音高の配置から学習し、音高の配置に関する制約として n グラムモデルを利用することで音高を推定するという自己組織的な構造学習法が考えられる。このような場合には、和音の表現として、従来の恣意的なラベルを利用するのではなく、音高の組合せを直接モデル化するのが望ましい。

このような問題意識のもと、自然言語処理分野で提案された最新のノンパラメトリックベイズ n グラムモデル^{11)–14)} を拡張することで、和音系列に対する語彙フリー無限グラムモデルを提案する。図 1 に示すように、このモデルでは n グラム確率が階層的にスムージングされており、以下のような利点がある。1) 観測データの背後にある生成モデル(確率過程)を考えることで、次の和音の予測確率を理論的に定式化できる。2) 与えられた和音系列中の各和音が、それぞれ異なる長さのコンテキスト(理論的には無限の長さであってもよい)を持つことを許容する。モデルを学習する過程で、各和音ごとにコンテキスト長の事後分布を推定できる。3) あらゆる音の組合せ(音楽で通常用いられているものに限らない)を和音の種類として許容できる。新しい音の組合せの和音が出現するたびに語彙を必要に応じて拡張していくことができる。結果的に、提案モデルは優れた予測精度を達成することができるだけでなく、学習データに含まれる、さまざまな長さを持つ統計的に特徴的な和音進行パターン(最も多く出現するものとは限らない)を見つけることができる。

本稿の構成は以下の通りである。まず、第 2 章で和音のシンボル表現について述べる。第 3 章で n グラムモデルに関する関連研究を紹介し、第 4 章で我々の提案モデルを説明する。第 5 章で比較実験について報告し、第 6 章で本研究のまとめを述べる。

表 1 ラベルに基づく表記法と構成音に基づく表記法

和音の種類	ラベルに基づく表記	構成音に基づく表記
Major	maj	100010010000
Minor	min	100100010000
Diminished	dim	100100100000
Augmented	aug	100010001000
Major Seventh	maj7	100010010001
Minor Seventh	min7	100100010010
Seventh	7	100010010010
Dim. Seventh	dim7	100100100100
Half Dim. Seventh	hdim7	100100100010
Min. (Maj. Seventh)	minmaj7	100100010001
Major Sixth	maj6	100010010100
Minor Sixth	min6	100100010100
Ninth	9	101010010010
Major Ninth	maj9	101010010001
Minor Ninth	min9	101100010010
Suspended Second	sus2	101000010000
Suspended Fourth	sus4	100001010000

2. 和音のシンボル表現

本稿では、和音のシンボル表現として表 1 に示す 2 種類の表記法、すなわちラベルに基づく表記法および構成音に基づく表記法を用いる。前者は、音楽で通常用いられているテキストラベル(和音の語彙)に基づく表記法であり、和音認識の研究などで採用されている。後者は、そのようなテキストラベルが存在しなかったと仮定した場合の語彙フリーな表記法であり、予備知識や事前学習なしで自動採譜を行うときなどがこれに相当する。

2.1 ラベルに基づく表記法

ラベルに基づく表記法として、Harte ら⁵⁾ によって提案された直感的で理解しやすい表記法を用いる。各和音は 17 種類のラベルと 12 種類の根音との組み合わせで表現される。本稿では、従来の研究^{1),4)} と同様に、 $C\#$ と $D\flat$ などの同一のピッチクラス^{*1} に属する音高は区別しないものとする。例えば、 C major の和音、 $G\flat$ diminished seventh の和音は、それぞれ $C:maj$ 、 $F\#:\dim7$ と表記する。また、音がない区間や打楽器のみの区間を示すには、特別な記号 N を用いる。したがって、語彙サイズは 205 ($17 \times 12 + 1$) となる。

*1 オクターブ内に存在する 12 個の異なる音高: $\{C, C\#, D, D\#, E, F, F\#, G, G\#, A, A\#, B\}$

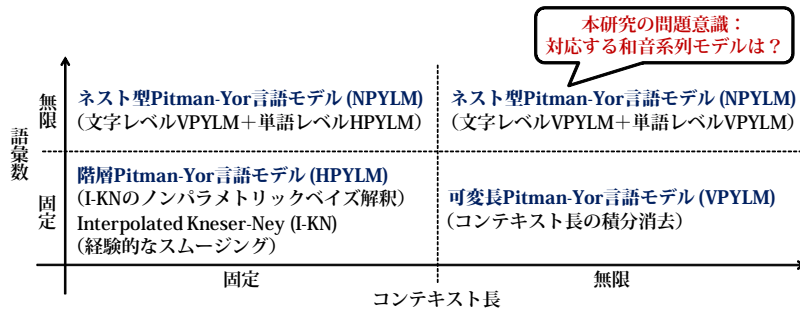


図 2 自然言語処理分野における単語系列に対する n グラムモデルの発展

2.2 構成音に基づく表記法

構成音に基づく表記法として、根音に対する構成音の相対音高を用いた表記法を提案する。各和音は、根音と 12 次元のバイナリベクトルとの組み合わせで表現される。このベクトルの各要素は、対応する相対音高が和音に含まれるか否かを示している。例えば、C major の和音は C,E,G から構成されているため、C:100010010000 と表記する。ただし、D major の和音は D:001000100100 ではなく D:100010010000 と表記する。この方法では、表 1 に含まれないあらゆる音の組合せを表記することができる。例えば、C major に F が追加された和音は、C:100011010000 と表記する。特別な記号 N を加えて、語彙サイズは 49153 ($2^{12} \times 12 + 1$) となる。本研究では語彙数は実質有限であるが、4.1 節で述べるように、楽譜に基づいて構成音のカウントを考慮するような表記法にすれば、語彙数は無限となる。

3. 統計的言語モデル

本章では、まず、 n グラムモデルの学習について説明する。その後、図 2 に示される通り、自然言語処理分野でこれまで提案されてきた経験的なスムージングに基づく n グラムモデルおよび 3 つのノンパラメトリックベイズ n グラムモデルについてそれぞれ紹介する。

3.1 問題設定

いま、語彙 (和音の種類) の集合を W とし、そのサイズを V (本研究では 205 あるいは 49153) とする。ある和音を $w \in W$ とし、長さ $n-1$ のコンテキストを $u \in W^{n-1}$ とする。学習データ X は、長さ M の和音系列 $x_1 \cdots x_M$ であるとする。ここで、 $x_m \in W$ ($1 \leq m \leq M$) であり、簡単のため学習データは 1 つの和音系列のみであるとした (複数の和音系列であっても以降の議論がそのまま適用可能である)。ある n が与えられると、 n グラムモデルでは、

ある和音 x_m は直前の $n-1$ 個の和音 (コンテキスト) に依存していると仮定する。

$$P(x_m | x_1 \cdots x_{m-1}) = P(x_m | x_{m-(n-1)} \cdots x_{m-1}) \quad (1)$$

我々の目的は、学習データ X が与えられたときに、コンテキスト u に続く和音 w の予測確率 $P_u(w|X)$ を求めることである。学習データ X においてコンテキスト u に続いて和音 w が出現した回数を c_{uw} とすると、予測確率 $P_u(w|X)$ の最尤推定値は

$$P_u^{ML}(w|X) = \frac{c_{uw}}{c_u} \quad (2)$$

で与えられる。ここで、ドット (\cdot) はそのインデックスに関して和をとったものを表し、 $c_u = \sum_{w'} c_{uw'}$ である。ただし、この素朴な方法では、 n グラム uw が学習データ X 中に存在しない場合 ($c_{uw} = 0$)、その確率は 0 になってしまう。この問題を解決するため、経験的あるいは理論的な様々なスムージング方法が提案されており、以下で説明する。

3.2 Kneser-Ney スムージングに基づく n グラムモデル

Kneser-Ney スムージング (KN)¹⁵⁾ とその変種は、最も高精度な n グラムモデルを学習できる一連の手法であることが実験的に知られている¹⁰⁾。基本的には、出現回数 c_{uw} が 1 以上であれば、ある一定の値 $d_{|u|}$ をディスカウント (減算) してから $P_u(w|X)$ を推定する。ここで、 $d_{|u|}$ はコンテキスト長 $|u|$ ごとに異なるパラメータである。さらに、Interpolated KN (I-KN) では、和音 w の n グラム確率をより低次の $n-1$ グラム確率で補間する。

$$P_u^{IKN}(w|X) = \frac{\max(c_{uw} - d_{|u|}, 0)}{c_u} + \frac{d_{|u|} t_u}{c_u} P_{\pi(u)}^{IKN}(w|X) \quad (3)$$

ここで、 t_u は、学習データ X においてコンテキスト u に続く和音の異なり数 (語彙 W のうちで何種類の和音が出現するか) であり、 $\pi(u)$ は、コンテキスト u から最初の和音を除去して得られるひとつ短くなったコンテキストである。 $P_{\pi(u)}^{IKN}(w)$ は和音 w の $n-1$ グラム確率であり、式 (3) で $\pi(u)$ を u に代入することで得られる。このような n グラム確率の補間は再帰的に定義されており、最終的にユニグラム確率はゼログラム確率を用いて補間される。和音 w のゼログラム確率としては、一般的に一樣確率として $1/V$ が仮定される。ディスカウントパラメータ d_0, \dots, d_{n-1} は所定の計算式に従って導出するか、交差検定法を用いて最適化することができる¹⁵⁾。

3.3 階層 Pitman-Yor 言語モデル

Teh¹¹⁾ は、ノンパラメトリックベイズ理論を用いて、階層 Pitman-Yor 言語モデル (Hierarchical Pitman-Yor Language Model: HPYLM) と呼ばれる新しい n グラムモデルを提案している。興味深いことに、従来より高精度なスムージングができることが経験的に知

られていた I-KN は、近年 HPYLM の近似であることが明らかとなった。そのため、より精緻なモデルである HPYLM は、I-KN より優れた予測精度を達成可能である (図 2 左下)。

3.3.1 階層 Pitman-Yor 過程とモデルの定式化

まず、ノンパラメトリックベイズモデルを構成する際にしばしば利用される Pitman-Yor 過程 (PY)¹⁶⁾ と呼ばれる確率過程について簡単に説明する。PY はディリクレ過程 (DP) の自由度を増やすように拡張されたものであり、DP と同様に確率分布上の確率分布となっている。すなわち、あるサンプル空間 (例: 和音の語彙 W) 上の確率分布 (例: あるコンテキストが与えられたときの次の和音の確率分布) に対する確率分布である。したがって、PY は語彙 W 上の n グラム確率分布に対する事前分布として利用可能である。いま、 d および θ を正の実数とし、 G_0 をサンプル空間上の分布とすると、PY は

$$G \sim \text{PY}(d, \theta, G_0) \quad (4)$$

と書ける。 d はディスカウントパラメータ、 θ は集中度、 G_0 は基底測度と呼ばれる。PY から生成される G もまたサンプル空間上の確率分布である。 θ の値が大きくなればなるほど、 G_0 と似通った G が生成される確率が高くなる。

HPYLM は PY を階層化することで定式化できる。いま、語彙 W 上のユニグラム分布 G_ϕ があるとする。ここで、 ϕ は長さ 0 のコンテキストを表し、 $G_\phi(w)$ で和音 w のユニグラム確率を表わすものとする。長さ 1 のコンテキスト u が与えられた上でのバイグラム分布 G_u は、ユニグラム分布 G_ϕ とは異なっているもの (特に高頻度の和音に関して) G_ϕ といくらか類似していると考えられる。このとき、バイグラム分布 G_u はユニグラム分布 G_ϕ を基底測度とする PY から生成された、すなわち、 $G_u \sim \text{PY}(d_1, \theta_1, G_\phi)$ と考える。ここで、 d_1 および θ_1 はディスカウントパラメータおよび集中度であり、長さ 1 の任意のコンテキストをもつユニグラム分布の生成過程で共有されている。一般に、長さ $n-1$ のコンテキストが与えられた上での n グラム分布 G_u は、 $G_{\pi(u)}$ を基底測度とする PY から

$$G_u \sim \text{PY}(d_{|u|}, \theta_{|u|}, G_{\pi(u)}) \quad (5)$$

として生成されると考える。ここで、 $d_{|u|}$ および $\theta_{|u|}$ はコンテキスト長 $|u|$ におけるディスカウントパラメータおよび集中度である。ただし、 $n-1$ グラム分布 $G_{\pi(u)}$ もまた未知であるので、 $n-1$ グラム分布 $G_{\pi(u)}$ に対する事前分布として、 $d_{|\pi(u)|}$ および $\theta_{|\pi(u)|}$ をパラメータとし、 $n-2$ グラム分布 $G_{\pi(\pi(u))}$ を基底測度とする PY を考える。このような生成過程は再帰的に定義することができ、最終的にユニグラム分布 G_ϕ は

$$G_\phi \sim \text{PY}(d_0, \theta_0, G_0) \quad (6)$$

として生成されたと考える。ここで、 G_0 はグローバルな基底測度であり (ゼログラム分布)、

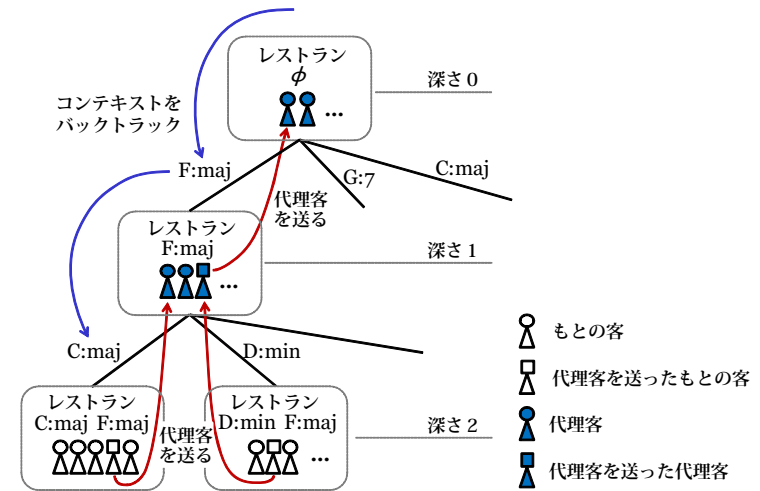


図 3 階層 Pitman-Yor 言語モデルにおける客の着席方法 (CRF: 中華料理店フランチャイズ)

通常は一様分布 $G_0(w) = 1/V$ が仮定される。これにより、 G_0 以外のすべての n グラム分布が未知であることによる不確実性をベイズ的に自然に取り扱うことが可能になる。

最終的に、HPYLM は深さ $n-1$ を持つ接尾語木で表現できる。例として、図 3 に $n=3$ の場合を示す。木の中の各ノードごとに、それぞれ対応するコンテキストが存在する。すなわち、木の根ノードから着目するノードまで枝に沿って木を下ることは、コンテキストに含まれる和音の一つずつ過去に遡ることを意味する。

3.3.2 確率的なデータ生成

HPYLM が定義できたので、観測データ X の確率的な生成過程を考える。すなわち、実際には X は学習データとして与えられているので、我々はすでに各観測変数 x_m の具体的な値を知っているが、それらの値がかつどのような過程を経て決まっていたのかについて考える。これは、中華料理店フランチャイズ (Chinese Restaurant Franchise: CRF) と呼ばれるメタファーを用いて説明できる¹¹⁾。CRF では、コンテキストはレストラン、学習データ X に含まれる M 個の観測変数は客、語彙 W に含まれる V 種類の和音は料理に例えられる。したがって、接尾語木の各ノードごとにレストランが存在する。各レストランには上限のない数のテーブルが存在し、客はそのいずれかのテーブルに着席する。各テーブルごとにひとつの料理が提供されており、そのテーブルに座った客が同じ料理を食べる。

いま、和音系列 $x_1 \cdots x_M$ がこの順番で生成されたと仮定し、 $m-1$ 人の客 $x_1 \cdots x_{m-1}$ がいずれかのレストランのいずれかのテーブルにすでに着席している場合に、 m 番目の客 x_m がどのように振る舞うかについて考える。まず、客 x_m は深さ $n-1$ にあるレストラン $u = x_{m-(n-1)} \cdots x_{m-1}$ (図3における葉ノードの一つ) に入る。ここで、レストラン u において料理(和音) w を提供しているテーブルの数を t_{uw} とし、テーブルの総数を t_u で表わす(複数のテーブルで同じ料理が出されることもある)。また、料理 w を提供しているテーブル k ($1 \leq k \leq t_u$) に着席している客の数を c_{uwk} とする(テーブル k で提供されている料理が w と異なる場合は $c_{uwk} = 0$)。レストラン u において、客 x_m は、(i) $c_{uwk} - d_{|u|}$ に比例する確率で既存テーブル k ($1 \leq k \leq t_u$) に着席するか、(ii) $d_{|u|}t_u + \theta_{|u|}$ に比例する確率で新規テーブル $k' = t_u + 1$ に着席する。(i) の場合、客 x_m は着席したテーブル k で提供されている料理 w を食べることになる。すなわち、観測変数 x_m の値(m 番目の和音)は w と決まり、 c_{uwk} がインクリメントされる。(ii) の場合、新規テーブル k' で提供される料理 w を注文するため、代理客が親レストラン $\pi(u)$ に送られる。代理客は親レストラン $\pi(u)$ において、もとの客 x_m と同様の確率的な振る舞いの結果、いずれかの既存あるいは新規テーブルに着席して、そのテーブルで提供される料理 w' を食べるようになったとする。その結果、レストラン u における新規テーブル k' で提供される料理も w' と決まり、客 x_m は料理 w' を食べることになる。すなわち、観測変数 x_m の値(m 番目の和音)は w' と決まり、 $t_{uw'}$ および $c_{uw'k'}$ がインクリメントされる。ここで、代理客がレストラン $\pi(u)$ において新規テーブルに着席した場合には、新たな代理客がその親レストラン $\pi(\pi(u))$ に送られることに注意する。これが再帰的に行われると、最終的に、代理客が木の根にあるレストラン ϕ に送られることがある。もし、根レストラン ϕ で代理客が新規テーブルに着席した場合は、そのテーブルで提供される料理は基底測度 G_0 に従って確率的に選ばれる。

以上をまとめると、これまで到着した客の着席方法 S (全レストランにおける客とテーブルの配置) が与えられたときに、コンテキスト u に続く和音 w は、条件付き確率分布

$$P_u^{\text{HPY}}(w|S) = \frac{c_{uw} - d_{|u|}t_{uw}}{c_{u..} + \theta_{|u|}} + \frac{d_{|u|}t_u + \theta_{|u|}}{c_{u..} + \theta_{|u|}} P_{\pi(u)}^{\text{HPY}}(w|S) \quad (7)$$

に従って生成される。ここで、式(7)は再帰的な定義になっていることに注意する。例えば、 $P_{\pi(u)}^{\text{HPY}}(w|S)$ は式(7)の $u \leftarrow \pi(u)$ を代入することで得られる。空の接尾後木(あらゆるレストランで $c_{uwk} = 0$ かつ $t_{uw} = 0$) から始めて、 M 人の客を順番に加えていくことで、学習データ X に対する着席方法が求まる。式(7)と式(3)を比較すると、 $\theta_{|u|} = 0$ かつ $t_{uw} = 1$ に固定した HPYLM は I-KN と等価になることが分かる。

```

1 空の接尾語木を準備          5 for i=1:∞
2 for m=1:M (ランダム順)      6   for m=1:M (ランダム順)
3   客  $x_m$  を深さ  $n-1$  に追加  7     客  $x_m$  を削除
4   end                        8     客  $x_m$  を深さ  $n-1$  に追加
                               9   end
                               10 end

```

図4 階層 Pitman-Yor 言語モデルのためのギブスサンプリングアルゴリズム

3.3.3 予測分布とベイズ推定

我々の目的は、コンテキスト u に続く和音 w の予測分布 $P_u(w|X)$ を求めることである。もし、 X に含まれる全ての客の着席方法 S が分かっていたら、予測分布は式(7)で求めることができる。しかし、真の着席方法は未知であるので、着席方法の事後分布 $P(S|X)$ のもとでの予測分布 $P_u^{\text{HPY}}(w|S)$ の期待値

$$P_u^{\text{HPY}}(w|X) = \sum_S P_u^{\text{HPY}}(w|S) P(S|X) \quad (8)$$

を計算する必要がある。この和は解析的に計算することはできないが、マルコフ連鎖モンテカルロ法(MCMC)の一種であるギブスサンプリングを用いると

$$P_u^{\text{HPY}}(w|X) \approx \frac{1}{L} \sum_{i=1}^L P_u^{\text{HPY}}(w|S_i) \quad (9)$$

として近似可能である。ここで、 L は事後分布 $p(S|X)$ からギブスサンプリングによってサンプルされた具体的な着席方法の個数であり、 i はそのインデックスである。

図4に、CRFの重要な性質である交換可能性(Exchangeability)を利用したギブスサンプリングのアルゴリズムを示す。一見すると客の到着順は最終的な着席方法 S に影響を与えようと思えるが、同じ着席方法であれば、客順によらず確率 $p(S|X)$ は一定であることが知られている。具体的には、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ることによって、全ての客に対する着席方法を得る。ここで、「事後的」であるとは、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)うえで、着席方法を決めるということの意味する。したがって、 $x_m = w$ であるならば、客 x_m は料理 w を提供している既存テーブルに着席するか、新規テーブルに座って料理 w を食べるかのいずれかである。両者の確率は、式(7)の第一項および第二項でそれぞれ与えられる。次に、交換可能性の性質を利用することで、現在得られている着席方法のサンプルをもとにして、

新しい着席方法のサンプルを新たに生成することを考える (MCMC) . 具体的には, ある客をランダムを選択し, 木から削除したのちに, 最後に到着した客だとみなして再度木に加えるという操作を行うことで, 新たな着席方法をサンプルすることができる. 客を削除する際には, その客から派生した代理客が存在すれば, それらを再帰的に削除すると同時に, 客がいなくなったテーブルも削除する. 実際には, 連続する着席方法のサンプルには強い相関があるため, 十分な間隔をあけることで独立と見なせる L サンプルを取得する.

また, パラメータ d_0, \dots, d_{n-1} および $\theta_0, \dots, \theta_{n-1}$ も未知であるので, ギブスサンプリングによって推定を行う. 具体的には, 定義域が $0 < d_{|u|} < 1$ および $0 < \theta_{|u|}$ であることから, それぞれに対してベータ事前分布およびガンマ事前分布を設定する. そして, 補助変数を介した事後分布を導出してサンプリングを行う (詳細は文献¹¹⁾ を参照のこと).

3.4 可変長 Pitman-Yor 言語モデル

HPYLM の問題は, 全ての客が深さ $n-1$ のレストランに強制的に入れられることである. これは全ての和音のコンテキスト長を $n-1$ に固定していることを意味し, 適切ではない. この問題を解決するため, 持橋と隅田¹²⁾ は, それぞれの客が異なる深さのレストランに入ることを許容する可変長 Pitman-Yor 言語モデル (Variable-order Pitman-Yor Language Model: VPYLM) を提案している. VPYLM では, それぞれの観測変数 x_m には n グラム長を示す潜在変数 z_m が対応していると仮定する. ただし, z_m の真の値は未知であるので, ベイズ的に z_m の値のあらゆる可能性を考慮に入れて (積分消去することで) 次の和音の予測を行う. すなわち, 異なる n に対応する n グラムモデルを重みつきで加算することで, 無限混合モデルの一種である無限グラムモデルが構成できる (図 2 右下).

3.4.1 モデルの定式化

VPYLM は, 客 x_m が入るべきレストランの深さ (あるいは n グラム長 z_m) を確率的にモデル化することで定義できる. 一般に, 学習データが多くなるほど高次 n グラムモデルを精度よく学習できるようになることから, 多く観測されるコンテキストをもつ客ほど, 深いレストランに入りやすくなるような機構が必要である. このような要求に合致するものとして, 棒折り過程 (Stick-Breaking Process: SBP)¹⁷⁾ を利用することができる.

いま, 潜在変数 z_m の値は SBP からどのように確率的に決まるかについて考える. まず, 客 x_m は経路 $\phi \rightarrow x_{m-1} \rightarrow x_{m-2} \rightarrow \dots$ に沿って木を下ってゆく. その経路において, 客が深さ $i-1$ ($1 \leq i \leq \infty$) のレストラン u_{i-1} に到着するたびに, 確率 $\eta_{u_{i-1}}$ で停止するか, 確率 $1 - \eta_{u_{i-1}}$ で通過するかを選択するものとする. したがって, 客が木を下って行って深さ $n-1$ のレストランで停止する確率, すなわち $z_m = n$ となる確率は,

$$P_u(n|\eta) = \eta_{u_{n-1}} \prod_{i=1}^{n-1} (1 - \eta_{u_{i-1}}) \quad (10)$$

で与えられる. この確率は, n が大きくなるに従って指数的に減衰する.

いま, 木に含まれる各ノード u に対するパラメータ η_u は未知であるので, 事前分布を設定することで不確実性を適切に取り扱うことにする. 具体的には, 定義域が $0 < \eta_u < 1$ であることから, α および β を超パラメータとするベータ事前分布を設定する.

$$p(\eta) = \prod_{u \in \text{tree}} \text{Beta}(\eta_u | \alpha, \beta) \quad (11)$$

3.4.2 確率的なデータ生成

VPYLM が定義できたので, HPYLM と同様に観測データ X の生成過程を考えることができる. m 番目の客 x_m に着目すると, まず, 潜在変数 z_m の値が式 (10) の棒折り過程に従って確率的に決まる. 潜在変数 z_m の値が与えられると, 客 x_m は指定された深さのレストランに入り, 観測変数 x_m の値は 3.3.2 節で説明した CRF に従って確率的に決まる. ここで, HPYLM では深さ $n-1$ のノード (葉ノード) 以外には代理客しか存在しなかったが, VPYLM ではもとの客とそれらに由来する代理客とが混在する可能性があることに注意する. その場合でも, 客の数 c_{uwk} をカウントする際には区別を行わない.

以上をまとめると, これまで到着した客の着席方法 S およびパラメータ η が与えられたときに, コンテキスト u の長さ $n-1$ とそれに続く和音 w は, 条件付き確率分布

$$P_u^{\text{VPY}}(w, n | S, \eta) = P_u^{\text{VPY}}(w | n, S) P_u(n | \eta) \quad (12)$$

に従って生成される. ここで, $P_u^{\text{VPY}}(w | n, S)$ は HPYLM における式 (7) と同様に求めることができ, $P_u(n | \eta)$ は式 (10) で与えられる. さらに, 式 (12) において, 未知の確率変数 n および η を積分消去する (あらゆる可能性を考慮する) ことで,

$$P_u^{\text{VPY}}(w | S) = \sum_n P_u^{\text{VPY}}(w | n, S) P_u(n | S) \quad (13)$$

が得られ, n を指定する必要がない無限グラムモデルを構成できる. ここで, $P_u(n | S) = \int P_u(n | \eta) p(\eta | S) d\eta$ であり, 式 (10) と式 (11) との共役性を利用すると

$$P_u(n | S) = \frac{a_{u_{n-1}} + \alpha}{a_{u_{n-1}} + b_{u_{n-1}} + \alpha + \beta} \prod_{i=1}^{n-1} \frac{b_{u_{i-1}} + \beta}{a_{u_{i-1}} + b_{u_{i-1}} + \alpha + \beta} \quad (14)$$

として簡単に求められる. ここで, $a_{u_{i-1}}$ および $b_{u_{i-1}}$ は, S に含まれる到着済みの客が木を下って行く際に, レストラン u_{i-1} で停止した客の数および通過した客の数である.

3.4.3 予測分布とベイズ推定

コンテキスト u に続く和音 w の予測確率は、図 4 で示される HPYLM に対するギブスサンプリングアルゴリズムに少しの変更を加えることで、3.3.3 節と同様に求めることができる。具体的には、図 4 において客 x_m を接尾後木に追加する前に、潜在変数 z_m の値を事後分布からサンプルすることで客 x_m が入るべきレストランの深さを決める必要がある。いま、与えられた学習データにおいて $x_m = w$ であるならば、 $z_m = n$ となる事後確率はベイズの定理を用いると

$$P_u(n|\mathbf{S}, w) \propto P_u(w, n|\mathbf{S}) = P_u^{\text{VPY}}(w|n, \mathbf{S})P_u(n|\mathbf{S}) \quad (15)$$

として求めることができる。

3.5 ネスト型 Pitman-Yor 言語モデル

自然言語処理分野における従来の n グラムモデルの本質的な問題は、新語や造語が作られるなどして語彙は日々成長していくものであるにもかかわらず、有限の語彙を明示的に定義しなければならないことであった。単語系列をモデル化する上でこの問題を解決するため、持橋ら^{13),14)} は可算無限個の単語上に定義された基底測度 G_0 に基づくネスト型 Pitman-Yor 言語モデル (Nested Pitman-Yor Language Model: NPYLM) を提案している。もし、従来通り一様分布 $G_0(w) = 1/V$ を基底測度に用いると、語彙のサイズが $V \rightarrow \infty$ であるときに、 $G_0(w) \rightarrow 0$ となってしまう。代わりに、文字レベルの VPYLM に基づく「綴りモデル」を単語レベルの VPYLM の基底測度に用いることができる。言い換えれば、各単語は文字の系列であり、それらの生成は文字レベルの CRF に従うと仮定する。単語の長さは、非負整数上に定義されたポアソン分布に従うことが知られている。したがって、あらゆる単語 w のゼログラム確率 $G_0(w)$ を、その単語を構成する文字群の生成確率と単語長の確率との積として定義することで、無限語彙モデルを得ることができる (図 2 上側)。

4. 和音系列に対する語彙フリー無限グラムモデル

本章では、単語系列をモデル化する上で単語の語彙を定義する必要がない NPYLM¹⁴⁾ と同様の問題意識のもと、和音系列をモデル化する上で和音の種類に関する語彙を定義する必要がない語彙フリー無限グラムモデルを提案する。

4.1 定式化

和音系列をモデル化する上では、単語系列のための n グラムモデルである NPYLM を直接適用することはできない。単語は文字の経時的な並びであるのに対し、和音は音符の同時的な組み合わせであるので、NPYLM とは異なる基底測度 G_0 が必要になる。

この問題を解決するため、和音の構成音に基づく確率モデルを和音レベルの VPYLM の基底測度 G_0 として利用する方法を提案する。具体的には、2.2 節で説明した構成音に基づく表記法を利用して、ある和音に 12 種類ある各ピッチクラスの音が「含まれているかどうか」に関するモデル化を行う。この場合、語彙サイズ $V = 49153$ は大きいながらもまだ有限であり、従来の有限語彙の n グラムモデルでも扱うことはできる。しかし、本研究で提案する方式では、楽譜情報が利用可能であれば、ある和音に各ピッチクラスの音が「何個含まれているか」を非負整数上に定義される分布 (ポアソン分布や負の二項分布など) を用いてモデル化することで、容易に無限語彙モデルを定式化できる利点がある。

基底測度 G_0 は、根音のクラスと各構成音の有無とが独立であるという仮定に基づいてモデル化する。2.2 節の表記法に従えば、一般に和音 w は $w_0:w_1 \cdots w_{12}$ と書くことができる。ここで、 w_0 は根音のクラスを示す確率変数であり、他はバイナリ値をとる確率変数である。ただし、 $w = N$ であれば、 $w_0 = N$ として残りの変数は利用しないものとする。このとき、 w_0 は 13 次元の離散分布に、残りの各変数はそれぞれベルヌイ分布に従うと仮定すると、

$$G_0(w) = p(w|\pi, \tau) = \pi_{w_0} \prod_{i=1}^{12} \tau_i^{w_i} (1 - \tau_i)^{1-w_i} \quad (16)$$

とできる。ここで、 $\pi = \{\pi_C, \pi_{C\#}, \dots, \pi_B, \pi_N\}$ は、13 種類のシンボル、すなわち 12 種類の根音のクラスか特別なクラス N の生起する確率であり、 $\tau = \{\tau_1, \dots, \tau_{12}\}$ は対応する相対音高の存在確率を示している。ただし、 $w = N$ であれば、 $G_0(w) = \pi_N$ とする。さらに、パラメータ π および τ の値は未知であるので、離散分布およびベルヌイ分布の共役事前分布としてディリクレ分布およびベータ分布を設定することで、不確実性を適切に取り扱う。

$$p(\pi, \tau) = \text{Dir}(\pi|\mathbf{a}_0) \prod_{i=1}^{12} \text{Beta}(\tau_i|b_0, c_0) \quad (17)$$

ここで、13 次元ベクトル \mathbf{a}_0 、正の実数 b_0 および c_0 は超パラメータである。

4.2 予測分布とベイズ推定

我々が提案するモデルは VPYLM の拡張であるので、コンテキスト u に続く和音 w の予測確率は、3.4.3 節と同様に求めることができる。ただし、階層的なスムージングに必要な基底測度 $G_0(w)$ を式 (16) を用いて求めるためには、パラメータ π および τ の値が必要である。そこで、客の着席方法 S が与えられた上での、パラメータ π および τ の事後分布からサンプルすることを考える。具体的には、式 (16) と式 (17) の共役性を利用すると

$$p(\boldsymbol{\pi}, \boldsymbol{\tau} | \mathbf{S}) = \text{Dir}(\boldsymbol{\pi} | \mathbf{a}_0 + \mathbf{n}) \prod_{i=1}^{12} \text{Beta}(\tau_i | b_0 + n_i, c_0 + \bar{n}_i) \quad (18)$$

で計算することができる。ここで、13次元ベクトル \mathbf{n} の各要素を n_v ($v \in \{C, C\#, \dots, B, N\}$) とすると、 n_v は根レストラン ϕ において根音が v である料理 w ($w_0 = v$) を提供しているテーブルの数、 n_i は i 番目の相対音高を含む料理 w ($w_i = 1$) を提供しているテーブルの数、 \bar{n}_i は i 番目の相対音高を含まない料理 w ($w_i = 0$) を提供しているテーブルの数である。

本モデルに対するギブスサンプリングアルゴリズムは、VPYLM に対するアルゴリズムに少し変更を加えることで得られる。具体的には、客（代理客でもよい）が根レストラン ϕ で新規テーブルに着席したとき、 n_v および n_i あるいは \bar{n}_i の値が、そのテーブルで提供される料理に応じてインクリメントされる。客を木から削除することに伴い、根レストランからテーブルが削除される場合には、 n_v および n_i あるいは \bar{n}_i の値が、そのテーブルで提供されていた料理に応じてデクリメントされる。全ての客に対する着席方法をサンプルとして取得するたびに、パラメータ $\boldsymbol{\pi}$ および $\boldsymbol{\tau}$ の値も式 (18) からサンプルを行う。

5. 評価実験

本章では、提案するノンパラメトリックベイズ n グラムモデルの予測精度を評価するために行った2つの比較実験について報告する。

5.1 実験条件

実験用のデータセットとして、Harte ら⁵⁾ によるビートルズの180曲の楽曲に対する和音アノテーションデータを用いた。和音の使われ方は調に依存するため、途中で転調のない長調の楽曲137曲を選択し、全て八長調に転調を行った。このデータに含まれる和音の総数は10,761であった。まず、VPYLMの基本的な性能を確かめるため、2.1節で説明したラベルに基づく表記法を用いた実験を行った。この実験では、全部で9種類の異なる n グラムモデル、すなわち、Good-Turing法 (GT)、Witten-Bell法 (WB)、I-WB、KN、I-KN、Modified KN (MKN)、I-MKN、HPYLMおよびVPYLMを評価した。ここで、GTおよびWBは、自然言語処理分野や音声認識分野において従来から広く使われていた経験的なスムージング手法である¹⁰⁾。MKNはKNの改良版であり、式(3)で表わされる予測確率において、ディスカウント $d_{|u|}$ の値をカウント c_{uw} に合わせて変化させることができる¹⁰⁾。「I-」は低次 n グラムからの補間を行うモデルであることを示している。最初の7モデルに関しては、The SRI Language Modeling Toolkit (SRILM)¹⁸⁾ による実装を用いた。

表2 ラベルに基づく表記法を用いた比較実験におけるパープレキシティ (語彙サイズ 205)

n	GT	WB	I-WB	KN	I-KN	MKN	I-MKN	HPYLM	VPYLM
1	16.8	15.6	15.7	15.8	16.0	15.6	15.7	15.8 (± 0.03)	n : サンプリング 13.4 (± 0.33)
2	20.3	14.2	15.5	14.5	15.2	14.5	15.8	14.5 (± 0.10)	
3	23.5	15.4	18.5	16.1	16.0	16.1	16.3	16.0 (± 0.18)	
4	25.5	16.8	22.5	16.8	17.7	16.5	15.5	13.9 (± 0.25)	n : MAP 推定値 12.9 (± 0.35)
5	26.3	17.5	25.4	15.5	16.2	16.0	14.1	13.7 (± 0.23)	
6	27.0	17.8	27.2	14.6	15.1	16.2	13.5	13.6 (± 0.23)	
7	27.3	18.0	28.3	14.3	14.5	16.4	13.3	13.6 (± 0.23)	n : 積分消去 (∞) 11.9 (± 0.22)
8	27.3	18.0	28.8	14.1	14.2	16.6	13.2	13.6 (± 0.22)	
9	27.3	18.0	29.0	14.0	14.1	16.8	13.1	13.5 (± 0.23)	
10	27.3	18.0	29.2	14.0	14.0	16.7	13.1	13.5 (± 0.23)	

次に、提案する語彙フリー無限グラムモデルの性能を確かめるため、2.2節で説明した構成音に基づく表記法を用いた実験を行った。この実験では、最初の実験で用いた9種類のモデルに加えて、HPYLMおよびVPYLMの基底測度を第4章で提案した語彙フリーな基底測度に変更したモデルを評価した。これらのモデルを「VF-」で表わすことにすると、VF-VPYLMが我々の提案する語彙フリー無限グラムモデルであり、VF-HPYLMが語彙フリー固定長 n グラムモデルである。各モデルの予測精度を評価するため、データを楽曲単位で10分割して交差検定を行い、パープレキシティを測定した。パープレキシティが低いほど優れた予測精度を持つモデルである。

5.2 実験結果

最初の実験においては、表2に示すように、VPYLMが最も低いパープレキシティを達成した。各和音ごとに n グラム長である n の値を、事後分布からランダムサンプルして一意に決める場合、最大事後確率をとる値に一意に決める場合、積分消去する場合とを比較すると、本来未知である n の値を明示的に一意に決めるのではなく、あらゆる可能性を考慮する方が優れた予測精度が得られることが分かった。また、学習データの量がもっと多ければ、客を木の深いノードに追加することによる過学習の問題が低減され、高次のHPYLMとVPYLMの差は縮まると考えられるが、VPYLMの学習はHPYLMの2倍以上高速であり、消費メモリは5分の1程度であった。これは、VPYLMは必要以上に深いノードに客を追加しないため、展開すべきノードの数が少なくすむからである。I-KNは、HPYLMの近似であることから予想される通り、HPYLMよりやや悪い予測精度であった。一方、I-KNを改良したI-MKNは、ディスカウントの値を変化させる点ではHPYLMと同じであるが、HPYLMよりやや優れた予測精度を達成した。これは文献¹¹⁾の結果と共通している。

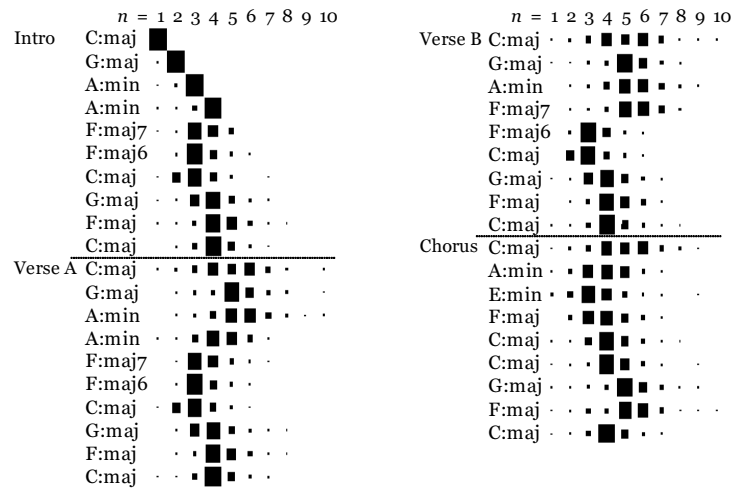


図5 ビートルズ「Let It Be」の和音系列に対する n グラム長の事後分布推定結果

ビートルズの「Let It Be」における和音系列に対して、VPYLMで推定した n の事後分布を図5に示す（別の実験におけるVF-VPYLMでもほぼ同様の結果が得られた）。より多く出現する和音進行パターンほど大きな n が選ばれる確率が高くなっている。また、長さ $n-1$ のコンテキスト u に続く和音 w の確率は、 $P_u(w, n|X) = \sum_S P_u(w, n|S)P(S|X)$ で与えられるので、この確率を大きいもの順に並べると、さまざまな長さを持つ確率的な和音進行パターンを見つけることができる。表3に、実験に用いたビートルズの楽曲137曲から見つかった統計的に特徴的な和音進行パターンとその確率を示す。これらは、必ずしも出現頻度が高いパターンではなく、和音進行そのものが強い結び付きを持っているパターンであることに注意する。例えば、C:7 F:7 C:7は典型的なブルースロック進行であり、ビートルズが多用したことで一般大衆に人気が出たもののひとつである。また、当時としては珍しい、seventhを多用した和音進行パターンを使用する傾向が強いことが分かる。

第二の実験においては、表4に示すように、我々の提案モデルであるVF-VPYLMがもっとも低いパープレキシティを達成した。構成音に基づくより精緻な基底測度を導入することで、HPYLMおよびVPYLMとともに大幅に予測精度が向上している。最初の実験よりも他のモデルとの精度差が大きいのは、提案するモデルが語彙サイズが大規模あるいは無限であるときでも、学習データのスパースさに対して頑健であることを示している。

表3 ビートルズの楽曲から見つかった統計的に特徴的な和音進行パターン

	確率的和音進行パターン ($n \geq 3$ のみを表示)
0.701	C:7 F:7 C:7
0.682	B:maj F:maj G:maj
0.656	A:min C:7 F:maj
0.647	F:min G:maj C:maj
0.645	F:maj F:maj G:maj C:maj
0.632	E:min C:7 F:maj
0.623	C:maj7 D:min7 E:min7
0.627	B:maj F:maj G:maj C:maj
0.622	D:min7 G:sus4 G:maj
0.620	D:min G:maj C:maj F:maj C:maj

表4 構成音に基づく表記法を用いた比較実験におけるパープレキシティ（語彙サイズ 49153）

n	GT	WB	I-WB	KN	I-KN	MKN	I-MKN
10	38.3	24.4	39.1	18.4	18.5	21.7	17.5

n	HPYLM	VF-HPYLM	n	VPYLM	VF-VPYLM
10	18.0 (± 0.29)	16.5 (± 0.60)	∞	15.8 (± 0.29)	14.6 (± 0.55)

6. おわりに

本稿では、和音の種類に関する語彙を事前に与える必要がなく、 n の値を事前に指定する必要もないノンパラメトリックベイズ n グラムモデルを提案した。比較実験の結果、これまで実験的に精度が優れていることで知られていた n グラムモデルよりも、さらに優れた予測精度を達成することが確認できた。

従来は、限られた音の組合せのみが音楽的に有用な響きをもち、それらは人間が定義するものであるとの暗黙の前提があった。しかし、一般的なポピュラー音楽理論において、majorコードや minorコードなどは確立された和音であるが、そのような基本和音にいくつかの音が足された和音をどこまで確立されたコードとしてみなすか（呼び名を与えるか）は自明ではない。マイルス・デイビス・クインテットのピアニストであったハービー・ハンコックが、ネフェルティティ（1967年）の録音の模様を語った1989年のインタビュー記事によれば、「ウェインがこの曲を書いた時、コード・シンボルはまったくなかった」とある。この曲は従来のコードの概念にとらわれない不思議な響きを有しており、新しい和音が生み

出されていることを示す良い例である。これは、自然言語において日々新しい単語や造語が生み出されているのと同じ現象であることに気づけば、自然言語処理分野で提案された NPYLM (単語系列のための無限語彙無限グラムモデル) に着想を得て、和音系列のための和音の構成音に基づく新しい確率モデルを定式化することが可能となる。

本研究は、いくつかの新しい研究テーマを切り開く基礎となるものである。まず、音楽音響信号に対する和音認識や自動採譜の研究において、確率的な言語モデルを確率的な音響モデル (隠れマルコフモデルや確率的非負値行列分解) とベイズ的に統合することによって、両者を確率的な枠組みのもとで同時に最適化・自己組織化することが可能になると考えられる。これまでは、言語モデルと音響モデルは別々のデータを用いて学習されており、本質的な改良の余地が残されていた。別の方向性として、大量の楽譜から、究極的には大量の音楽音響信号から、計算機に「和音」の概念を教師なしで獲得させることが可能かどうか研究することは非常に興味深い。我々は、ある特定の音の組み合わせのみが「和音」として成立することを知っている。これは、経験に基づくものなのであろうか。誰かから教えられる、あるいは専門家によって定義されるものなのであろうか。与えられた和音語彙はどれほど適切なものであろうか。その適切さは数値化可能なのだろうか。この種の疑問に答えるためには、あらゆる音の組合せが「和音」を形成する可能性を考慮せねばならず、無限モデルを構成することが本質的に必要だと思われる。ノンパラメトリックベイズ理論は、このようなメタレベルの問題に立ち向かうための強力な武器となりうるアプローチである。

謝辞 本研究は、科研費 若手 (B) 23700184 の支援を受けて行われた。研究を行う上で貴重なアドバイスをくださった持橋大地氏 (統計数理研究所) および平田圭二氏 (はこだて未来大学) に感謝する。

参 考 文 献

- 1) Ogihara, M. and Li, T.: N-gram Chord Profiles For Composer Style Representation, in *9th International Conference of Music Information Retrieval (ISMIR)*, pp. 671–676 (2008).
- 2) Pérez-Sancho, C., Rizo, D. and Nesta, J.M.I.: Genre Classification Using Chords and Stochastic Language Models, *Connection Science*, Vol.21, No. 2-3, pp. 145–159 (2009).
- 3) Paiement, J.-F., Eck, D. and Bengio, S.: A Probabilistic Model for Chord Progressions, in *6th International Conference of Music Information Retrieval (ISMIR)*, pp. 312–319 (2005).
- 4) Scholz, R., Vincent, E. and Bimbot, F.: Robust Modeling of Musical Chord Sequences Using Probabilistic N-grams, in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 53–56 (2009).

- 5) Harte, C., Sandler, M., Abdallah, S. and Gómez, E.: Symbolic Representation of Musical Chords: A Proposed Syntax for Text Annotations, in *6th International Conference of Music Information Retrieval (ISMIR)*, pp. 66–71 (2005).
- 6) 金子仁美, 川上大輔, 嵯峨山茂樹: 機能と声解析データの作成とその統計解析, 音楽情報科学研究会 研究報告, No. 2010-MUS-85, pp. 1–8, 情報処理学会 (2010).
- 7) Cheng, H.-T., Yang, Y.-H., Lin, Y.-C., Liao, I.-B. and Chen, H.-H.: Automatic Chord Recognition for Music Classification and Retrieval, in *IEEE International Conference on Multimedia & Expo (ICME)*, pp. 1505–1508 (2008).
- 8) 須見康平, 糸山克寿, 吉井和佳, 駒谷和範, 尾形哲也, 奥乃博: ベース音高と和音特徴の統合に基づく和音系列認識, 情報処理学会論文誌, Vol.52, No.4, pp. 1803–1812 (2011).
- 9) Khadkevich, M. and Omologo, M.: Use of Hidden Markov Models and Factored Language Models for Automatic Chord Recognition, in *10th International Conference of Music Information Retrieval (ISMIR)*, pp. 561–566 (2009).
- 10) Chen, S. and Goodman, J.: An Empirical Study of Smoothing Techniques for Language Modeling, Technical Report TR-10-98, Computer Science Group, Harvard University (1998).
- 11) Teh, Y.W.: A Bayesian Interpretation of Interpolated Kneser-Ney, Technical Report TRA2/06, NUS School of Computing (2006).
- 12) 持橋大地, 隅田英一郎: 階層 Pitman-Yor 過程に基づく可変長 n-gram 言語モデル, 情報処理学会論文誌, Vol.48, No.12, pp. 4023–4032 (2007).
- 13) 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, 自然言語処理研究会 研究報告, No. 2009-NL-190, p.49, 情報処理学会 (2009), <http://chasen.org/daiti-m/paper/nl190segment.pdf>.
- 14) Mochihashi, D., Yamada, T. and Ueda, N.: Bayesian Unsupervised Word Segmentation with Nested Pitman-Yor Language Modeling, in *ACL-IJCNLP*, pp. 1000–1008 (2009).
- 15) Kneser, R. and Ney, H.: Improved Backing-Off for M-gram Language Modeling, in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vol.1, pp. 181–184 (1995).
- 16) Pitman, J. and Yor, M.: The Two-Parameter Poisson-Dirichlet Distribution Derived from a Stable Subordinator, *Annals of Probability*, Vol.25, No.2, pp. 855–900 (1997).
- 17) Sethuraman, J.: A Constructive Definition of Dirichlet Priors, *Statistica Sinica*, Vol.4, pp. 639–650 (1994).
- 18) Stolcke, A.: SRILM – An Extensible Language Modeling Toolkit, in *International Conference on Spoken Language Processing (ICSLP)*, pp. 901–904 (2002).