

# マルコフ確率場に基づく統計的な音楽情報の解析\*

陳映融, 米田隆一, 西本卓也, 嵯峨山茂樹 (東大情報理工)

## 1 はじめに

MIDI や楽譜などシンボリックな音楽情報から調 (tonality) や和声 (chord) を推定する技術は、自動編曲、自動伴奏、音楽情報検索など多くの場面で重要な役割を果たす。これらのシンボリックな音楽情報を入力とし、調、和声、カデンツ定型を推定する問題はラベル付け問題と同質と考える。また、楽曲においてはある調性に基づく和声が和声学の規則に従って進行するものである。従って、音楽情報処理において楽曲の構文的分析を行う際には、音楽知識を用いた解析手法が有効であると考えられる。本稿では、これらのラベル付け問題において、2 次元的な音楽情報 (縦の和音、横の声部進行等) を表現することができ、かつ素性関数の設計により音楽知識を統計的に還元することができるマルコフ確率場に基づくアプローチを論じる。

## 2 和声進行と調のモデル化

### 2.1 和声進行のモデル化

和声学においては、典型的な和声進行ルールとカデンツ定型が存在し、各和音記号には固有な進行 (例: II は V のみに進行する) がある。しかし、ダイアトニック和音の代わりに他の調から和音を借用する場合 (例: secondary dominant) もあるので、楽曲のある時刻における和音は、和声進行ルールだけでなく、その時刻の各声部の構成音にも依存している (ここでは非和音を考慮しない)。

Fig.1[1] に示すように、和音は周辺の音楽的文脈 (影の部分) に依存する。このような 2 次元的な音楽情報に対する文脈依存性はマルコフ確率場における素性関数により設計可能である。多声部あるいは単声部の旋律を入力系列とすると、和声推定とカデンツ同定は条件付き確率場でのラベル系列を求める問題として扱える。

### 2.2 調のモデル化

従来の調認識手法 [2] では心理学的統計に基づいて調の音高分布の標準パターンを作り、入力楽曲の音高分布と比較し調を判別している。このような標準パターンの有効性はある程度実証されているが、入力楽曲のジャンルやスタイルに応じパターンを調整する柔軟性が欠けている。

ここでは、同じく音高分布に基づき、より柔軟性のある最大エントロピーモデルを考える。具体的には、ひとつの調に対して音高の相対頻度を値とする素性関数を 12 個用意し、学習によりパラメータ (素性関数の重み) を決定する。楽曲の調決定問題は、このモデ

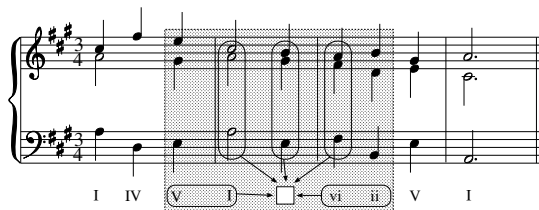


Fig. 1 和音の周辺の譜面に対する依存関係はマルコフ確率場の素性関数に対応している。実線は評価実験で素性関数として利用した部分。

ルにおいて、楽曲が与えられたときの条件付き確率が最大となる調を求める問題として扱える。

## 3 マルコフ確率場

### 3.1 概要

マルコフ確率場 [3] は一般化したマルコフ過程で、周辺の文脈に依存する値をモデル化している。標本空間  $S$  に定義された確率変数  $X = X_1, \dots, X_m$  に対し、各々の  $X_i$  がラベル  $Y_i (Y_i \in L)$  を表すとき、 $X$  を確率場と呼ぶ。 $X$  が以下の二つの性質を満たす場合、マルコフ確率場であると言う ( $P(X_i = Y_i)$  を  $P(Y_i)$  と略す)。

$$P(\mathbf{Y}) > 0, \forall \mathbf{Y} \in \mathbf{X} \quad (\text{positivity}) \quad (1)$$

$$P(\mathbf{Y}_i | \mathbf{Y}_{S-\{i\}}) = P(\mathbf{Y}_i | \mathbf{Y}_{N_i}) \quad (\text{Markovianity}) \quad (2)$$

ただし、 $N_i$  は  $i$  の近隣である。マルコフ確率場をグラフ (確率変数が頂点、近隣同士の関係が辺) で表現すると、確率はクリーク (完全部分グラフ) $c$  に対応するポテンシャル関数  $V_c$  の対数線形モデルになる [3]。

$$P(\mathbf{Y}) = \exp \left( - \sum_{c \in C} V_c(\mathbf{Y}) \right) / Z \quad (3)$$

$C$  は頂点の部分集合から構成したクリークの集合である。 $V_c$  を素性関数  $f_i$  とそれに対応する重み  $\lambda_i$  の積と見なすと、式 (3) は最大エントロピーモデルにおける確率分布関数の式と同一である。

### 3.2 最大エントロピー法

最大エントロピー法 (Maximum Entropy, ME) は、学習データとモデルの間にいくつかの制約を設け、これらの制約を満たす最も一般的な分布を持つモデルを求める手法である。制約はモデルに反映されるべき学習データの特徴を表し [4]、学習データにおける素性 (素性関数) の期待値とモデルによる素性の期待値を一致させることにより実現する。

$$\sum_{x,y} \tilde{P}(x,y) f_i(x,y) = \sum_x P(x) \sum_y P(y|x) f_i(x,y) \quad (4)$$

ただし  $\tilde{P}(x,y)$  は学習データから得た確率分布である。最大エントロピーモデルの条件付き確率は以下のパラメトリック形式で表すことができる。

$$P_\Lambda(y|x) = \exp \left( \sum_i \lambda_i f_i(x,y) \right) / Z_\Lambda(x) \quad (5)$$

$$Z_\Lambda(x) = \sum_y \exp \left( \sum_i \lambda_i f_i(x,y) \right) \quad (6)$$

ここで、 $\Lambda = \{\lambda_1, \dots, \lambda_n\}$  はモデルのパラメータであり、 $\Lambda$  の値が定めればモデルは決定される。また、 $\lambda_i$  は素性  $f_i$  の重み、 $Z_\Lambda$  は正規化項である。

素性関数は通常、ラベルの有無を表す 2 値関数と考えるが、調認識では相対頻度を表す実数関数とす

\*Statistical Analysis of Music Information Based on Markov Random Fields. by CHEN, Ying-Rong, YONEDA, Ryuichi, NISHIMOTO, Takuya, SAGAYAMA, Shigeki (Graduate School of Information Science and Technology, The University of Tokyo)

る。例えばある楽曲の調性がト長調、C の相対頻度が 0.0693 の場合、素性関数の値は以下のように定義する。

$$f(\text{音高} = C, \text{調性}) = \begin{cases} 0.0693 & \text{if 調性} = G \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

### 3.3 条件付き確率場

文献 [5] では、ME の特殊形である条件付き確率場 (Conditional Random Fields, CRF) を提案している。CRF はマルコフ確率場と同質であり [6],  $S$  に定義された  $X$  と、ラベル集合  $L$  に定義された  $Y$  と、二つの確率場がこの Markovianity:  $P(Y_i | X, Y_{S-\{i\}}) = P(Y_i | X, Y_{N_i})$  を満たす場合、 $(Y, X)$  は CRF であると言う。

このように、CRF は観測された入力系列の全体に基づいた確率場であり、観測系列が与えられたときのラベル系列全体の条件付き確率をモデル化する手法である。系列全体は単一の対数モデルで表現するので、ある状態における各素性関数の重みは固定でなく、自動調整される [5]。和声解析とカデンツ同定では、このグローバルなモデルで文脈依存性を表現し、全体の確率の最大化により最適な系列を推定する。

## 4 評価実験

### 4.1 評価方法

提案手法の有効性を実証するため、評価実験を行った。調認識を除き、和声推定とカデンツ同定では各実験データに階名に変換したものを評価実験に用いた。定量評価では一曲とそれ以外に分割して交差検証を行った (Table 1)。和声推定とカデンツ同定における素性関数は音楽知識に基づき、試行錯誤で決定した (詳細は後述)。

### 4.2 和声解析

Humdrum Toolkit に付属する和声ラベル付きのバッハの chorale の 16 曲 (humdrum-kern フォーマット) を用い、CRF で和声を解析する実験を行った。Chorale から (ソプラノ, 和音), (バス, 和音), (ソプラノ + バス, 和音) 三種類のペアを抽出し、各自の素性関数を用いてパラメータを学習させ、和音 (非和声音の種類) の同定はこのタスクに含まない) を決定した。利用した素性関数は現在音, 前後音, 前後和音, 前音 + 次音, 前和音 + 次和音, 前音 + 前和音, 次音 + 次和音, 前音 + 現在音 + 次音, 前々和音 + 前和音, 次和音 + 次々和音, 以上 12 種と現在和音との組合せ (ソプラノ + バスと和音のペアは 21 種) であった。和音は diatonic chord と secondary dominant の三和音を対象とし、転回形と基本形は同じと見なした。八分音符を時間の単位とすると、(ソプラノ, 和音), (バス, 和音), (ソプラノ + バス, 和音) の正解率は、それぞれ 80%, 83%, 85% で、解析結果の誤りは主に正解和音と構成音の類似する和音との混同であった。

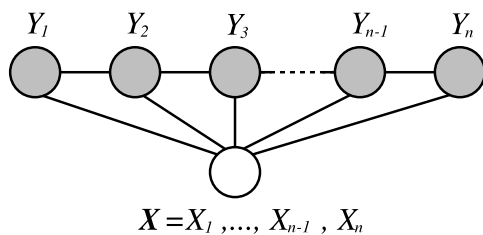


Fig. 2 CRF の概念図。入力系列  $X$  から全体確率の最大化するラベル系列を求める。

Table 1 実験の概観

	手法	入力形式	定量的評価
和声解析	CRF	humdrum	80% ~ 85%
カデンツ同定	CRF	独自形式	99%
調認識	ME	MIDI	85%

### 4.3 カデンツ同定

文献 [7] の pp.104-105 のバス課題において、階名を入力とする各曲のカデンツ形式とドミナント定式のラベルペアの付与を行った。ドミナント定型句では、始まりを B(Begin), 終了までを I(Inside), 他の部分を O(Other) と標記し、カデンツ形式では、K1, K2, K3 と標記した。付与するラベルは {K1,K2,K3} と {B, I} のペア, および O の標記であった。現在音, 前後音, 前々音, 次々音, 前音 + 現在音, 現在音 + 次音, 前音 + 前ラベル, 次音 + 次ラベル, 前ラベル + 次ラベル, 前音 + 現在音 + 次音のそれぞれと現在ラベルとの組合せを素性とする CRF で決定したところ、総音符数 229 中、ラベルペアの誤りは 3 箇所 (精度 99%) で、誤った 3 箇所の正解は全て O の標記であった。

### 4.4 調認識

MIDI の音高 (音名) の相対頻度を素性関数とする ME で調を決定した。1 つの調につき 4 曲ずつ、合計 96 曲 (長調, 短調それぞれ 12 調) の MIDI データを用意し、各曲の旋律が調固有の特徴とみなされるような過学習を防ぐため、各曲を 11 回移調してデータを増やし、パラメータの事前分布を正規分布 (分散を 64 に設定した) と仮定した。96 曲に対して交差検証したところ、正解 82 曲 (精度 85%), 他の 14 曲は正解の近親調と誤認識された。誤認識された 14 曲のうち 5 曲は属調と認識され、下屬調, 平行調, 同主調の誤認識はそれぞれ 3 曲であった。

## 5 まとめ

音楽情報の解析に周辺の文脈に依存する値をモデル化するマルコフ確率場 (最大エントロピーモデル, 条件付き確率場) を適用した。具体的に、和声解析, カデンツ同定に条件付き確率場, 調認識に最大エントロピーの適用を示した。評価実験において、それぞれ 80% ~ 85%, 99%, 85% の精度が得られ、誤った認識結果は正解と類似な性質を持つことも確認した。今後の展開としては、音高のヒストグラム情報のみによる調認識から和声の解析処理を含めるように拡張したい。また、マルコフ確率場による音楽表現のモデル化と応用を検討したいと考える。

## 参考文献

- [1] S. Kostka *et al.*, *Tonal Harmony: With an Introduction to Twentieth-Century Music*, McGraw-Hill, 1999.
- [2] C. L. Krumhansl, *Cognitive Foundations of Musical Pitch*, Oxford Univ. Press, New York, 1990.
- [3] S. Z. Li, *Markov Random Field Modeling in Computer Vision*, Springer-Verlag, 1995.
- [4] K. Nigam *et al.*, "Using Maximum Entropy for Text Classification," *In IJCAI-99 Workshop on Machine Learning for Information Filtering*, 61-67, 1999.
- [5] J. Lafferty *et al.*, "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data," *International Conference on Machine Learning*, 2001.
- [6] H. M. Wallach, "Conditional Random Fields: An Introduction," *University of Pennsylvania CIS Technical Report MS-CIS-04-21*, 2004.
- [7] 島岡ら, *和声理論と実習 I*, 音楽之友社, 1964.