G-001

最大エントロピーモデルに基づく統計的な音楽情報の解析
Using Maximum Entropy for Music Modeling

米田 隆一[†]　西本 卓也[†]　嵯峨山 茂樹[†]
Ryuichi Yoneda　Takuya Nishimoto　Shigeki Sagayama

1. はじめに

本稿では、MIDI, 楽譜等のシンボリックな音楽情報を入力としてラベル (対旋律、和音、調等) を付与する汎用的な手法を提案する。我々の研究グループでは、このようなラベル付け問題を取り扱う際、音声認識における言語モデルと同様、マルコフモデルを適用することが多かった。しかし、MIDI、楽譜などの 2 次元的な情報 (縦の和音、横の声部進行等) は、マルコフ連鎖のようなモデルでは不十分であり、より広いコンテキストの重視が必須である。このような背景のもと、広いコンテキストを柔軟に設定でき、かつ、我々の持つ音楽的知識を素性関数の設計という操作に還元することができるマルコフ確率場 (最大エントロピーモデル) に基いたアプローチを論じる。

2. 音楽の確率定式化

2.1 一般化した数理構造

図 1 は、マルコフ確率場の概念を幾何学的に表現したものである。各ノードは音高、和音等のラベルを表し、黒のノードは現在付与しようとするラベルである。矢印はノード同士が関係を持っていることを表す。これは、MIDI、楽譜が持つ 2 次元的な情報によくマッチする。すなわち、旋律、和音等を作成する際、縦の和音構成、横の声部進行共に考慮にいれなければならない。さらに、音楽の持つ繰返し、模倣等の構造を考えると、より遠くのコンテキストまで影響が及んでいると考えられる。言語と比較して圧倒的に語彙サイズが小さいにもかかわらず、音楽がリッチな表現力を持つ所以である。

2.2 対位法、和声学における文脈依存の例

対位法においては、平行 5 度、同一音程の 4 度以上連続の禁止等の制約がある。これは、旋律の 1 つ、あるいは数個前後の文脈により対旋律の決定が可能である。和声学においては、典型的な和声の終止定型が存在しバス旋律より決定が可能である (例: fa-so-so-do に対する II-I-V-I-曲末)。これは旋律の先読みにより和音の決定が可能である。これらの文脈依存性は、最大エントロピーモデルにおける素性関数により容易に設計可能である。

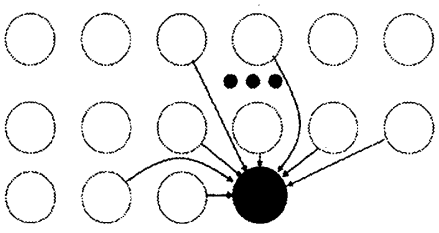

図 1: 音楽を一般化した数理的な構造

[†]東京大学, The University of Tokyo

3. 最大エントロピーモデル

3.1 マルコフ確率場

マルコフ確率場 [1] は、周辺の文脈に依存する値をモデル化するのに適した手法である。F が S において以下の条件を満たすとき、マルコフ確率場であるという。(1) $P(f) > 0(\forall f \in F)$) (positivity), (2) $P(f_i|f_{S-\{i\}}) = P(f_i|\mathcal{N}_i)$ (markovianity). ただし $\mathcal{N}_i$ は i の近隣 ($i \in \mathcal{N}_i$) である。近隣同士の関係をグラフのアークとみなしたとき、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 の対数線型モデルになるといわれている。

$$P(f) = \exp\Big(\sum_{c\in C} V_c(f)\Big)/Z \tag{1}$$

我々の音楽情報解析では、V_c を素性関数 f_i とそれに対応する重み λ_i の積であるとみなす。素性関数 f_i は通常、ラベルの有無を表す 2 値関数を考える。

$$f(\text{前和音}, \text{和音}) \equiv \begin{cases} 1 & \text{if 前和音} = V \ \&\ \text{和音} = I \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

これは、最大エントロピーモデル [2] における確率分布関数の式と本質的に同一である[‡]。

$$p_\lambda(y|x) = \exp\Big(\sum_i \lambda_i f_i(x,y)\Big)/Z_\lambda(x) \tag{3}$$

ただし Z は正規化項 $Z_\lambda(x) = \sum_y \exp\big(\sum_i \lambda_i f_i(x,y)\big)$ である。

3.2 MEMM, CRF

Maximum Entropy Markov Model (以下、MEMM) は最大エントロピー法 (以下、ME) の順次適用により、確率の積が最大となるラベル系列を最適解とみなす手法である。最適解の探索にはビームサーチ等が適用できる。Conditional Random Fields (以下、CRF) は入力系列そのものを入力とするグローバルな最適解を求める手法であり、ME の特殊形、かつ、HMM の一般化となっている [3]。図 2 は、概念を幾何学的に表現したものである。

なお、式 (3) におけるパラメータ λ の推定には反復スケーリング法が適用できる。

4. 評価実験

4.1 評価方法

ドミナント定型同定と調認識の実験を除き、Humdrum Toolkit[4] に付属する和声ラベル付きのバッハのコラール 16 曲 (humdrum-kern フォーマット) をすべて階名に

[‡]式 (3) におけるパラメータ λ の事前分布として、通常、正規分布を仮定するが、本稿では分散を MEMM では 0, CRF では 10 に設定した。

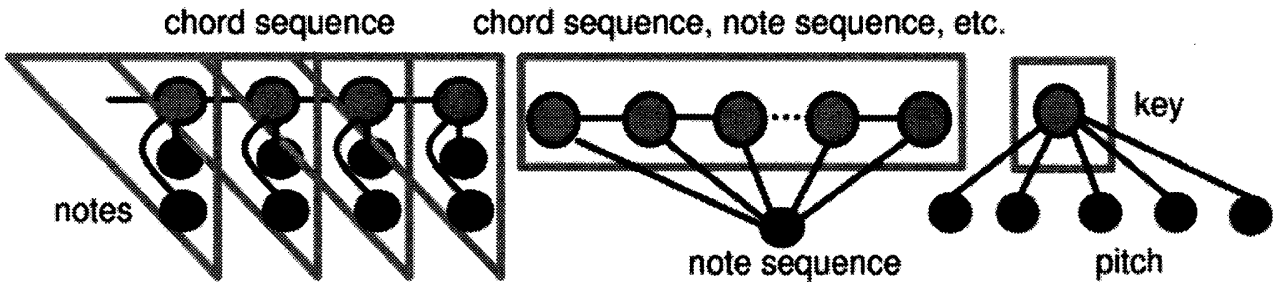

図 2: マルコフ確率場のグラフによる表現: 黒のノードは入力、グレイのノードは出力、グレイの線で囲まれた領域は最適解を表す。左図: 和声解析の概念図 (MEMM)。各声部の音高と前和音から最尤の和音を求める操作を繰返し適用していく。中図: 対旋律付け、和声付け等の概念図 (CRF)。入力系列そのものから、グローバルな最適解を求める。右図: 調認識の概念図。楽曲全体の音高頻度により調を決定する。

変換したものを評価実験に用いた。音符を持たない弱拍部 (2 連続 8 分音符のソプラノ旋律に対応する 4 分音符のバス旋律の弱拍部) 等には、ダミーのラベルを与えた。また、曲頭、曲末にもまた別のダミーのラベルを与えた。なお、対旋律付け、和声付け実験において出力された対旋律、和音進行に対する主観評価実験は行っていない。和声付け、ドミナント定型句同定、和声解析では、定量的評価は、1 曲とそれ以外に分割する交差検定で行った。

4.2 対旋律付け、和声付け、和声解析

ソプラノとバスの旋律のペアを学習する対旋律付け実験、各声部の旋律と和音のペアを学習する和声付け実験共に MEMM で決定したところ、妥当な出力系列が確認された。正解は一意ではないので、定量的評価として、前和音、現在音、前音を素性とする現在和音の決定を ME で行なった。すなわち、式 (2) において ({ 前和音, 現在音, 前音 } の異なり数 × 現在和音の異なり数) の個数の素性関数を考える。ソプラノ、バスの旋律に対する和音の正解率は、それぞれ 61%, 64% であった。

4 声部それぞれの旋律と和音のペアを学習する和声解析 (非和声音の種類の同定はこのタスクに含まれない) 実験において、各声部の前音、現在音、前和音を素性とする ME で現在和音を決定したところ、正解率は 75%であった。

4.3 ドミナント定型句の同定

文献 [5] の pp. 114-115 のバス課題において、ドミナント定型、終止定式の決定にあたり、階名を入力とする Begin, Inside, Outside の 3 ラベルの付与を考える。すなわち、定型句の始まりを B, 終了までを I, その他を O とラベル付けする。これは、出力系列が一意であるラベル付け問題である。現在音、前後 2 音、前音+現在音、現在音+次音を素性とする CRF で決定したところ、総音符数 174 中、誤りは 1 個所 (精度 99%) であった。

4.4 調認識

MIDI の音高 (mod 12) の相対頻度を素性関数とする ME で調を決定する。すなわち、式 (2) において素性関数を 12 個用意し、値を楽曲全体から得られる相対頻度とする。MIREX (Music Information Retrieval Evaluation eXchange) 2005 にて公開されている Symbolic Key Finding の訓練データ 96 曲を、すべての調が 1 セット中に含まれるよう 4 分割し交差検定したところ、誤りは 17 曲 (精度 82%) であった。

表 1: 実験の概観

	手法	入力形式	定量的評価
対旋律付け	MEMM	humdrum	-
和声付け	ME,MEMM	humdrum	61%〜64%
D 定型同定	CRF	独自形式	99%
和声解析	ME	humdrum	75%
調認識	ME	MIDI	82%

5. まとめ

周辺の文脈に依存するラベル列のモデル化に適した最大エントロピーモデル (マルコフ確率場) を音楽情報の解析に適用した。そして、シンボリックな音楽情報を入力とする種々のタスクを統一的な枠組みでできることを示した。具体的には、対旋律付け、和声付け、ドミナント定型の同定、和声解析、調認識への適用を示した。今後の展開としては、音程等、音楽のモデル化に有効な素性を吟味したい。また、ドミナント定型同定から構文解析 (カデンツ同定等の楽曲構造解析) へ拡張したい。また、音高のヒストグラム情報のみによる調認識は妥当とはいいがたい。和声も一連の解析処理へ含めたいと考える。

参考文献

[1] S. Z. Li, *Markov random field modeling in computer vision,* Springer Verlag, 1995.

[2] A. L. Berger, S. A. Della Pietra, and V. J. Della Pietra, "A maximum entropy approach to natural language processing," *Computational Linguistics,* 1996.

[3] J. Lafferty, A. McCallum, and F. Pereira, "Conditional Random Fields: Probabilistic models for segmenting and labeling sequence data," *Proc. of ICML,* 2001.

[4] D. Huron, The Humdrum Toolkit: Software for Music Research, http://dactyl.som.ohio-state.edu/Humdrum/.

[5] 島岡譲, "音楽の理論と実習 I," 音楽之友社, 1982.