

隠れマルコフモデルを用いた副旋律推定

田村 理遊 池田 剛 小谷 善行 但馬 康宏
東京農工大学 工学教育部 情報コミュニケーション工学専攻

本稿では、隠れマルコフモデル(Hidden Markov Model)を用い、主旋律と和声を持った楽曲に対する副旋律の推定を試みる。背景にある主旋律と和声によって副旋律が定まるという仮定の下で、与えられた主旋律と和声に対してもっともらしい副旋律の進行を推定する問題を扱うことにする。実験は、「副旋律の進行」と「主旋律・和声の出現確率」の近似式をそれぞれ二種類ずつ計四種類用いて行い、それら組み合わせごとの出力結果を比較する。また本稿で提案する同音高連続数に基づいた近似式が、出力結果にどのような影響を与えるかを実験結果から考察した。そこから、近似式ごとの違いが音符の密度や非和声音の生成に関わってくる事が示唆された。

Estimation of Sub-melody Using Hidden Markov Model

Riyu Tamura Takeshi Ikeda Yoshiyuki Kotani Yasuhiro Tajima
Tokyo University of Agriculture and Technology, Graduate School of Technology
Computer Information and Communication Sciences

This paper tries to estimate sub-melody for music that consists of main-melody and harmony applying Hidden Markov Model. We assume that sub-melody depends on main-melody and harmony, and solve the problem of estimating a maximum likelihood sub-melody progression to main-melody and harmony. In our experiment, we use two kinds of approximate expressions of “sub-melody progression” and “appearance probability of main-melody and harmony” respectively, and compare the performances of the each combination. In addition, we consider effect of approximate expressions based on frequent appearance for musical interval. Our result suggests that expressions in each approximation affect the density of notes and appearance of nonharmonic tone.

1. はじめに

本稿では、隠れマルコフモデル(Hidden Markov Model, 以下 HMM)を用い、主旋律と和声を持った楽曲に対する副旋律の推定を試みる。副旋律推定は、楽曲の自動副旋律付

けによる人間の作曲編曲支援として役立つと考えられる。

旋律を自動生成する研究には、構造を持つ自然な音符列の生成、動的計画法による対位法に基づくコストを最小にする対旋律の生成、主旋律の特徴を生かした不干渉な対旋律の生

成などが報告されている[1,2,3]。一方で、HMM を音楽へ適用した研究には、音楽演奏の音符音長系列データに対する音符リズム・テンポ・拍子・小節線の推定、入力旋律に対して自動で和声付けなどがある[4,5]。HMM は、楽譜のように多層の時系列データを確率統計的に処理しやすいため、本稿でも確率モデルとして採用する。

過去の研究では、主旋律とは別の旋律を自動生成する際には、その対象として旋律を主体とした作曲の理論である対位法に基づいた対旋律がしばしば用いられてきた。しかし本稿では、あくまで主旋律を引き立たせる単旋律を副旋律と位置付け、自動生成の対象とする。したがって伴奏か副旋律かがあいまいな単旋律も、全て広義の副旋律として捉えることとする。

本稿では、楽曲から得られる主旋律・副旋律・和声から HMM のパラメータを学習し、副旋律の推定を行う。その際、副旋律の遷移確率と主旋律・和声の出現確率の近似式を、それぞれ二種類ずつ用意し、実験をもとにこれらの性能について考察する。

2 . 副旋律推定のモデル

2.1 副旋律推定の問題

本稿では、背景にある主旋律と和声によって副旋律が定まると仮定し、与えられた主旋律と和声に対してもっともらしい副旋律を推定する問題を扱う。これは、(1)の条件付き確率を最大にする C を探索する問題として捉えることが出来る。

$$P(C|M, H) \quad (1)$$

主旋律 : $M = m_1, m_2, \dots, m_l$

副旋律 : $C = c_1, c_2, \dots, c_l$

和声 : $H = h_1, h_2, \dots, h_l$

記号の添え字はタイムステップ数を表し、 l は終端タイムステップを表す。

(1)は Bayes の定理により以下のように変

形できる。

$$P(C|M, H) = \frac{P(C)P(M, H|C)}{P(M, H)} \quad (2)$$

右辺分母の $P(M, H)$ は、 C に関わらず値が定まるので無視する。よって $P(C)P(M, H|C)$ を最大化することを考えればよい。

2.2 副旋律進行

副旋律 C のように進行する確率 $P(C)$ は以下のように表すことが出来る。

$$\begin{aligned} P(C) &= P(c_1, c_2, \dots, c_l) \\ &= \prod_{i=1}^l P(c_i | c_1, \dots, c_{i-1}) \end{aligned} \quad (3)$$

(3)は計算が進むにつれて l の値が増加し、条件項の種類が膨大な数に膨れ上がるため、限られた学習データから推定を行うのに十分な統計を取るのは困難である。また、副旋律の進行は HMM の状態遷移によって表され、その状態は一つ前の状態にのみ依存することを考慮し、(3)は以下のように近似する。

$$\prod_{i=1}^l P(c_i | c_1, \dots, c_{i-1}) \approx \prod_{i=1}^l P(c_i | c_{i-1}) \quad (4)$$

(4)のような bi-gram を用いることにより条件項の種類は飛躍的に縮小される。この近似の特徴は、副旋律 c_{i-1} と c_i が取り得る音高の組み合わせ全ての遷移確率を持てることである。

しかし、この確率モデルは過去の副旋律から得られる情報が極めて局所的なものに限られるため、音価の情報や c_{i-2} 以前から現在までの副旋律の繋がりが欠落している。

そこで本稿では、(3)に同音高連続数を用いた近似を新たに提案し、(4)とは別の情報から副旋律と捉える。同音高連続数は「現在と同じ音高が過去に連続した回数」と定義する。すなわち $S = s_1, s_2, \dots, s_l$ を同音高連続数としたとき、与えられた副旋律 C に対して以下のような値をとる。

$$C = c, c, d, e, e, e, f, e, d, d, c, c, c, \dots$$

$$S = 0, 1, 0, 0, 1, 2, 0, 0, 0, 1, 0, 1, 2, \dots$$

そして近似式は以下ようになる。

$$\prod_{i=1}^l P(c_i | c_1, \dots, c_{i-1}) \approx \prod_{i=1}^l P(s_i | s_{i-1}) \quad (5)$$

$$s_i = \begin{cases} s_{i-1} + 1 & (c_i = c_{i-1}) \\ 0 & (c_i \neq c_{i-1}) \end{cases}$$

この近似の特徴は、同じ音高が連続した長さ分だけ過去の副旋律の情報を得ることができることである。ただし副旋律の音高情報は、(3)が音高の全種類数を持つことが出来るのに対して、同音程連続数を用いた近似は「過去と同じ音高」「過去と異なる音高」の二種類しか持てないことになる。

つまり、(3)は時間の尺が狭く音価の尺が広い近似で、逆に(4)は音価の尺が狭く時間の尺が広い近似ということになる。

2.3 主旋律と和声の出現確率

副旋律 C のとき主旋律 M と和声 H の組が出現する確率は、以下のように表せる。

$$P(M, H | C)$$

$$= P(m_1, \dots, m_l, h_1, \dots, h_l | c_1, \dots, c_l) \quad (6)$$

(6)は、現在の副旋律の音高 c_i のみから主旋律の音高 m_i と和声の種類 h_i の出現確率が決まるように近似すると以下ようになる。

$$\prod_{i=1}^l P(m_i, h_i | m_1, \dots, m_{i-1}, h_1, \dots, h_{i-1}, c_1, \dots, c_l)$$

$$\approx \prod_{i=1}^l P(m_i, h_i | c_i) \quad (7)$$

(7)は過去の演奏情報を斬り捨てた単純な式である。しかし、副旋律は主旋律や和声との連携によるところが大きく、それらの推移の情報は副旋律推定にとって大きなウェイトと占めていると思われる。よって主旋律と和声を現在の値のみで説明するのは不十分であると考え、新たに二種類のパラメータを加えることにする。一つ目のパラメータは、主

旋律が前タイムステップと比較して音高がどのように変化したかの傾向を表す v_i 、二つ目のパラメータは、現在の和声が「前タイムステップの和声と異なる」か否かを表す t_i である。この二つのパラメータを導入することにより主旋律と和声に関する情報が増し、より多様で調和の取れた副旋律の推定が期待できる。 m_i, h_i, v_i, t_i を用いた(6)の近似は以下のようにになる。

$$P(M, H | C) \approx \prod_{i=1}^l P(m_i, h_i, v_i, t_i | c_i) \quad (8)$$

$$v_i = \{+, -, =\}$$

$$t_i = \{true, false\}$$

v_i の取りうる値は上がった・下がった・変化なし・三種類、 t_i の取りうる値は前述した条件が真か偽かの二種類である。この近似式は、実質(7)に m_{i-1}, h_{i-1} を含めた tri-gram と同じような意味を持つ。しかし tri-gram をそのまま用いた場合、主旋律は音階の数、和声はそれを上回る種類が存在し、副旋律進行と同様に学習データから統計を取るのには困難である。そのため m_{i-1}, h_{i-1} を単純化した v_i, t_i で代用することにした。

2.4 HMM

確率モデルには二種類の HMM を用いる。一つ目の HMM は、(4)に基づき図1のように隠れ状態を副旋律の音高とし、全ての状態から自分を含めた任意の状態へ遷移することが可能なものとする。

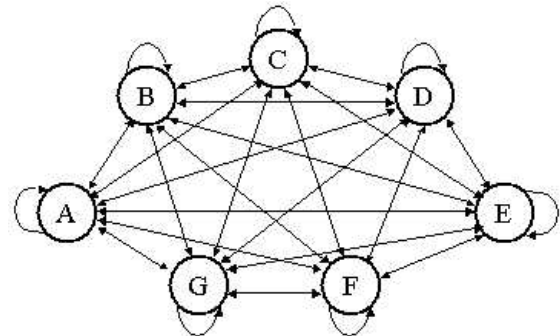


図1：音高を状態とした HMM

オクターブごとの高さの差は考えず、シフトしてひとまとめにする。例えば、音高 C3 と C4 は同じ C として扱う。図 1 の HMM は全音階で構成されているが、半音階で構成した場合は 12 状態となり、さらに休符の状態を含めると全部で 13 状態となる

二つ目の HMM は、(5)に基づき図 2 のように隠れ状態を副旋律と同音高連続数の組とし、全ての状態は同音高連続数を+1 した同じ音高状態か、同音高連続数が 0 の異なる音高の状態にのみ遷移が可能なものとする。この HMM の遷移確率は、音高とは無関係に同音高連続数のみで決定する。異なる音高へ遷移確率は、遷移先の音高に関わらず全て等しくなる。

同音高連続数は、理論上無限に上昇することが可能なため一定の値に達した時点で打ち切る必要があるが、実際には、楽曲が無限に続くことや、副旋律が際限なく同じ音高のみを演奏し続けることはありえないので、本稿ではとくに打ち切りは行わない。

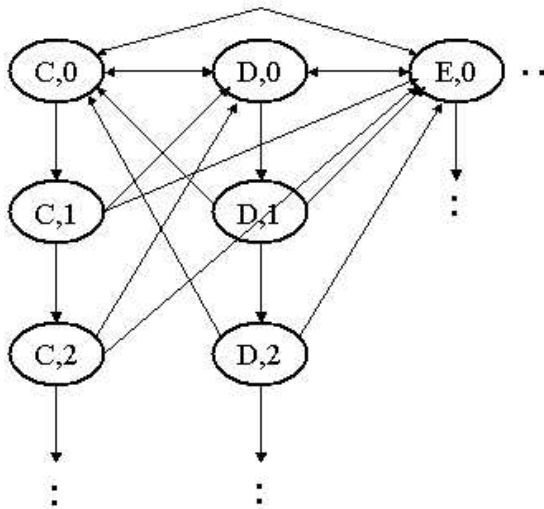


図 2：音高と同音高連続数の組を状態とした HMM

HMM の各パラメータは以下のように設定した。出力 y は主旋律と和声から得られるパラメータのことを意味する。すなわち(7)の近似の場合は m_i, h_i の組を指し、(8)の近似の場合

は m_i, h_i, v_i, t_i の組を指す。

初期確率 π_i 状態 q_i から始まる確率

遷移確率 a_{ij} 状態 q_i から状態 q_j へ遷移確率

出現確率 $b_j(y)$ 状態 q_i における出力 y の出現確率

3. 副旋律推定実験

3.1 実験方法

2.1 で設定した副旋律推定の問題に対して、実験を行った。前述した通り $P(C|M, H)$ は $P(C)P(M, H|C)$ から求め、これらが最大確率を持つときの副旋律 C を探索する。探索には Viterbi アルゴリズムを用いる。副旋律進行 $P(C)$ の近似式には(4)(5)を、主旋律と和声の出現確率 $P(M, H|C)$ を求める近似式には(7)(8)を用い、それぞれの近似による副旋律推定結果の比較を行う。統計を取るための学習サンプルには童謡を中心とした 30 曲を用いる[6]。副旋律の単位長は八分音符とするため、それよりも短い音符の入った曲は学習データとして用いていない。

また、実験結果は音高列の形式で出力される。一つの音符の音価までは確認できないため、音高が変化する箇所を音符の切れ目とする。出力結果を譜面に当てはめると以下のようなになる。

$$C = g, g, d, g, f+, d, d, c, g, g$$



図 3：出力の音高列と譜面の関係

3.2 実験結果

closed データ「大きな古時計 (作：ヘンリー・C・ワーク)」のケースで、出力された副旋律の一部を図 4 と図 5 に示す。楽譜は、一段目が入力として与えられた主旋律と和声で、二段目から五段目が出力された副旋律を表している。どちらの図も、主旋律の最初の音が演奏される小節から 8 小節先までを抜粋したものである。

図 4：近似式別の出力結果

図 4 は、 $P(C)$ と $P(M, H | C)$ の近似式の異なる組み合わせごとに出力された副旋律を示す。A-1 は(4)と(7)、A-2 は(4)と(8)、A-3 は(5)と(7)、A-4 は(5)と(8)を用いた結果である。

これらの結果を見て分かるとおり、(4)と(5)を用いた式では、出力された副旋律の傾向に著しい差が生じている。(4)を用いた場合は、音符の密度が低く、ほとんどが二分音符によって構成されている。一方で、(5)を用いた場合は、音符の密度が濃く、最低単位長の八分音符が多くを占めている。

また、ほとんどが和声音で構成されているため不協和音は生じないが、主観的評価としては単調な旋律という印象であった。ただ、(7)を用いた A-3 が 100%和声音であるのに対し、(8)を用いた A-4 はわずかながら非和声音が無理の無い形で使われている。図 4 の場合は、2 小節目と 6 小節目の最後の音符がそれに相当する。その点を考慮すれば、A-4 は他

図 5：同音高連続数の重み別の出力結果

の三種類に比べて一番自然な副旋律であった。

次に、(9)のように確率の対数を取り重み付けを行った式を用いた実験を行った。

$$w \log_{10} P(C) + \log_{10} P(M, H | C) \quad (9)$$

重み： w

(5)と(8)の組み合わせにおける、異なる重み w ごとに出力された副旋律を示す。B-1 は $w=1.0$ 、B-2 は $w=0.3$ 、B-3 は $w=0.01$ 、B-4 は $w=-0.5$ の結果である。全体を通して見ると、重み w の値が低くなるにつれて音符の密度も低くなっていることがわかる。その分、小刻みなフレーズが少なくなり、聴いた印象もゆったりと間延びしたものになっている。

これら図 4 図 5 の傾向は、他の曲で実験した際にも同様にみられた。

4. 考察

図 4 の結果では、(4)と(5)の違いによって、

全体的な音符の密度に著しい差が生じた。この原因は、二つの近似式の持つ特性にある。(4)は、一つ過去の音高のみを条件としているため過去の情報は狭い範囲でしか扱っていない。しかし、今回のように短い音価の音高を状態とした場合は、同じ音高への遷移確率が高くなってしまい、主旋律や和声に変化しない限り副旋律も同じ音高を生成し続けてしまう。この近似式を用いる場合は、副旋律の最小単位をある程度の長さを持ったフレーズにするか、もしくは(3)の近似式を少なくとも tri-gram 以上にするべきである。一方、(5)は同音高連続数が増加するにつれて、同音高への遷移確率が減少する傾向にある。そのため、異なる音高への遷移が頻発してしまい、生成される副旋律は小刻みに動く密度の高いものになる。この特性により(5)に重みを付加した実験では、重みの値が変動することで生成される副旋律のリズムの傾向が調整されている。

(7)を用いたときに和声音ばかりが生成されたのは、近似によって主旋律と和声の情報を大まかに捉え過ぎていたことが原因と考えられる。学習サンプルにおける非和声音の出現頻度が極めて少なく、(7)の近似ではどの主旋律と和声も観測されたとしても非和声音の出現確率が和声音の出現確率を下回ってしまう。(8)のように主旋律と和声の情報を細かく見ることによって、和声音のときに観測されるものと、非和声音のときに観測されるものを区別できる。

5. おわりに

本稿では、HMM を使い、与えられた主旋律と和声に対して新たな副旋律の生成を試みた。副旋律の遷移確率と主旋律・和声の出現確率を表す近似式を、それぞれ二種類ずつ用意し、実験結果から各近似式ごとの傾向を考察した。そこから近似式の違いが音符の密度や非和声音の生成に影響する事が示唆された。

今後は、フレーズを隠れ状態としたアプローチや、童謡とは別のジャンルの学習サンプルを用いた副旋律推定を行いたい。また主観的評価に偏らず、客観的かつ定量的に評価する方法についても検討したい。

参考文献

- [1] 耿豊, 池田剛, 乾伸雄, 小谷善行: 隠れマルコフモデルを用いた曲構造を持つ音符列の推定, 情報処理学会研究報告, 2003-MUS-49, pp31-36, 2003.
- [2] 中潟昌平, 西本卓也, 嵯峨山茂樹: 動的計画法と音列出現確率を用いた対位法の副旋律の自動生成, 情報処理学会研究報告, 2004-MUS-56, pp65-70, 2004.
- [3] 伊藤響, 小澤玲子, 犬塚信博, 伊藤英則: 主旋律の特徴を生かした不干渉な副旋律の生成手法, 情報処理学会研究報告, 99-MUS-31, pp13-18, 1999.
- [4] 齊藤直樹, 中井満, 下平博, 嵯峨山茂樹: 隠れマルコフモデルによる音楽演奏からの音符列推定, 情報処理学会研究報告, 99-MUS-33, pp27-32, 1999.
- [5] 川上隆, 中井満, 下平博, 嵯峨山茂樹: 隠れマルコフモデルを用いた戦慄への自動和声付け, 2000-MUS-34, pp59-66, 2000.
- [6] たのしいこどものうた大全集 202, 長岡書店, 2005.