E-030

マルコフ確率場を用いた自動作曲
Automatic Composition using Markov Random Fields

住田 浩之 †　　林 朗 †
Sumida Hiroyuki　　Hayashi Akira

1 概要

本論文では，マルコフ確率場を用いた自動作曲を提案する．学習データをある作曲家の MIDI データとし，その作曲家のスタイルをマルコフ確率場の特徴関数と重みで表現することで学習を行う．この確率場から，確率の高い標本をモンテカルロ法により得ることで自動作曲を行う．

2 手法

我々の目的は，モデルの確率場を段々と改善し目標の確率場へ近づけることで，学習データから特徴を学習することである．モデルの改善は，候補特徴を加えたモデルの対数疑似尤度と，加える前の対数疑似尤度の差分から，最もモデルを改善するような特徴を見つけ，新しく特徴集合に加える．続けて，新しい特徴集合の重みを再計算する．このステップを繰り返すことで徐々に目標の確率場に近づけてゆく．

2.1 学習データの表現

本研究であつかう学習データは，MIDI 形式の楽譜データを使用し，縦軸を音階，横軸を時間とし，各位置の音の有無を $\{1,2\}$ で表現する．ここで，音の長さ及び休符の情報を取り除き，更に音階を 1 オクターブ上にまとめ，音階 $i = 1 \ldots 12$, 時間 $t = 2 \ldots T$ の格子とする．また，曲の始まりと終わりを表現するために，$t = 1$ の各 $n_{i,1} (1 \le i \le 12)$ に状態 3 を挿入する．つまり，大きさ $12 \times T$ の格子の各状態が $n_{i,t} \in \{1,2,3\}$ であると考える．

2.2 確率場の表現

2.2.1 マルコフ確率場とは

マルコフ確率場はグラフ G から構成される．グラフの節点は確率変数を表わし，辺は確率変数間の独立性を意味する．すなわち，グラフ中の確率変数はその近傍が観測されたとき，非近傍からは独立である．

2.2.2 特徴関数

各音 $n_{i,t}$ の近傍の音の集合を $N_{i,t}$ とする．特徴関数は近傍 $N_{i,t}$ に含まれる音の配置を表現するものであり，以下のように定義することができる．

$$f_k(n_{i,t}, N_{i,t}) = n_{i,t} \prod_{n_{j,s} \in S_k} n_{j,s}, \quad S_k \subset N_{i,t} \qquad (1)$$

ここで，$k \in K$ は特徴のインデックスである．本来，特徴 f_k は位置 (i,t) に関係するものであり，$f_k^{(i,t)}$ と書くべきであるが，f_k と略記する．なお，特徴関数は t(時刻) に関して不変である $(f_k^{(i,t)} = f_k^{(i,t')})$ とする．

† 広島市立大学情報科学学部　〒731-3194 広島市安佐南区大塚東 3-4-1　Email: sumida@im.hiroshima-cu.ac.jp

表 1 に特徴関数の例を示す．1 列目が t(ただし相対的時刻) を，2 列目が音階 1...12 を，3 列目が状態 $\{1,2,3\}$ を表している．また，2 行以上の特徴は，複数の音から構成される特徴を表している．

- (A) は "ラ (音階 8)" が出る (状態 2) 事を意味する特徴
- (B) は "ド" と "ミ" が同時に出ることを意味する特徴
- (C) は "ド" の 1 時刻後に "ファ" が出ることを意味する特徴
- (D) は "ド" と "ファ" が同時に出た後は "レ" が出ない (状態 1) ことを意味する特徴
- (E) は "曲がミ" で始まる (状態 3) ことを意味する特徴
- (F) は "曲がド" で終わる (状態 3) ことを意味する特徴

(A)

0	8	2

(B)

0	5	2
0	1	2

(C)

0	1	2
1	6	2

(D)

0	1	2
0	6	2
1	3	1

(E)

0	1	3
1	5	2

(F)

0	1	2
1	1	3

表 1 特徴関数の例

また，この各特徴関数には，重み λ_k が割り当てられている．

2.2.3 確率場

各特徴関数にそれぞれ重みが割り当てられているとき，確率場を特徴関数と重みの積をポテンシャルとするギブス分布の形であらわすことができる．

$$\hat{P}(\mathbf{n}) = \frac{1}{Z_{\hat{P}}} \exp \sum_{t=1}^{T} \sum_{i=1}^{12} \sum_{k} \lambda_k f_k(n_{i,t}, N_{i,t}). \qquad (2)$$

$$\mathbf{n} = \{n_{1,1}, n_{2,1}, \ldots, n_{12,T}\}$$

ここで，$Z_{\hat{P}}$ は正規化定数で以下の形となる．

$$Z_{\hat{P}} = \sum_{\mathbf{n}} \exp \sum_{t=1}^{T} \sum_{i=1}^{12} \sum_{k} \lambda_k f_k(n, N_{i,t}) \qquad (3)$$

2.2.4 疑似尤度

最も確率場を改善する特徴を見つけ出すために，最尤推定を行う．しかし，(2) 式を計算するとき，正規化定数である $Z_{\hat{P}}$ の計算は $\{n_{i,t} | 1 \le i \le 12, 1 \le t \le T\}$ の組み合わせの数 (3^{12T}) だけ行わねばならず，莫大な計算量になってしまう．

そこで，近傍の依存関係のみから得た対数疑似尤度を用い，対数尤度を近似する．対数疑似尤度は以下の形となる．

$$\mathcal{PL}_{\hat{p}} = \frac{1}{12T}\log\prod_{t=1}^{T}\prod_{i=1}^{12}\hat{P}(n_{i,t}|N_{i,t}) = \frac{1}{12T}\sum_{t=1}^{T}\sum_{i=1}^{12}\log\hat{P}(n_{i,t}|N_{i,t}) \quad (4)$$

$$\hat{P}(n_{i,t}|N_{i,t}) = \frac{1}{Z_{i,t}}\exp\sum_{k}\lambda_k f_k(n_{i,t}, N_{i,t}). \quad (5)$$

ここで，$Z_{i,t}$ は正規化定数で以下の形となる．

$$Z_{i,t} = \sum_{n=1}^{3}\exp\sum_{k}\lambda_k f_k(n, N_{i,t}). \quad (6)$$

2.3 確率場の構築

2.3.1 特徴選択

単一の音のみで構成される特徴を原子特徴とする．このとき，既存の特徴に $t = \pm 1$ の範囲で原子特徴を加えたものを候補特徴関数 $g \in \mathcal{G}$ とし，候補特徴の集合を $\mathcal{G}$ とする．

各候補特徴を加えた確率場の尤度 $\mathcal{L}_{\hat{p}+\{\alpha g\}}$ と，加える前の確率場の尤度 $\mathcal{L}_{\hat{p}}$ との差をゲインの式 (7) とし，各特徴関数 g についてそれぞれゲインを計算する．

$$Gain(g,\alpha) = \mathcal{L}_{\hat{p}+\{\alpha g\}} - \mathcal{L}_{\hat{p}} = \alpha\tilde{E}[g] - \log\hat{E}[e^{\alpha g}] \quad (7)$$

ここに，$\tilde{E}[g]$ は特徴 g の経験分布に関する期待値，$\hat{E}[g]$ はモデル分布に関する期待値であり，以下のようになる．

$$\tilde{E}[g] = \frac{1}{12T}\sum_{t=1}^{T}\sum_{i=1}^{12}g(n_{i,t}, N_{i,t}) \quad (8)$$

$$\hat{E}[g] = \frac{1}{12T}\sum_{t=1}^{T}\sum_{i=1}^{12}\sum_{n=1}^{3}g(n, N_{i,t})\hat{P}(n|N_{i,t}) \quad (9)$$

各候補特徴について，式 (7) の α に関する偏微分の式 (10) からゲインを最大にする重み α を計算する．

$$\frac{\partial}{\partial\alpha}Gain(g,\alpha) = \tilde{E}[g] - \frac{\hat{E}[ge^{\alpha g}]}{\hat{E}[e^{\alpha g}]} = 0 \quad (10)$$

α に関する勾配が 0 となるときに各特徴は最大のゲインをとり，この中から最も疑似対数尤度を改善する候補特徴 $g_* = \arg\max_{g}\max_{\alpha} Gain(g,\alpha)$ を特徴集合 $\mathcal{F}$ に加える．

2.3.2 重みの再計算

2.3.1 での特徴選択は，候補以外の特徴の重みは固定させたまま行った．よって，新しくなった特徴集合における各重みを再計算する必要がある．各特徴の重みを，疑似対数尤度の式 (4) を最大にするように計算していく．この最大化に必要な (4) 式の勾配は以下の式で表される．

$$\frac{\partial}{\partial\lambda_k}\mathcal{PL}_{\hat{p}} = \tilde{E}[f_k] - \hat{E}[f_k] \quad (11)$$

式 (11) の勾配が 0，すなわち特徴 f_k の経験分布に関する期待値と，モデル分布に関する期待値の差が 0 となるように，重み $\lambda_k(1 \leq k \leq K)$ を設定する．

2.4 確率場誘導アルゴリズム

ここまでのアルゴリズムをまとめると以下のような手順になる．

- *Initial Data:*
  参照分布 $\tilde{P}$ 及び初期モデル $\hat{P}_0$
  学習する特徴の数 N
- *Output:*
  対数疑似尤度 $\mathcal{PL}$ を最大にする $\hat{P}_*$ 及びその時の $f_0, ..., f_N$
- *Algorithm:*
  (0) $\hat{P}^{(0)} = \hat{P}_0$.
  (1) (7) 式で各候補特徴 $g \in \mathcal{G}$ のゲインを計算する．
  (2) 最も高いゲインを得た候補特徴を，特徴集合 f_n とする．
  (3) 式 (11) で各特徴の重みを再計算し，またその時の確率場 $\hat{P}_*$ を計算する．
  (4) $\hat{P}^{(n+1)} = \hat{P}_*$ 及び $n \leftarrow n+1$ とし，ステップ (1) に戻る．

2.5 自動作曲

2.5.1 自動作曲

ここまでで学習した特徴とその重みから，新しい曲の自動作曲を行う．学習パートで得られた特徴集合とその重みから *Simulated Annealing* を用い，音を配置する．

3 実験

- 学習データに用いる MIDI データは，長調と短調とを同時に入力しないようにし，更に入力する学習データの調はすべて統一する．

なお，実験の結果については学会当日に発表する．

参考文献

[1] Stephen Della Pietra, Vincent Della Pietra, and John Lafferty, Member, IEEE, "Inducing Features of Random Fields" IEEE TRANSACTIONS PATTERN ANALYSIS AND MACHINE INTELLIGENCE,VOL. 19, NO. 4, APRIL 1997

[2] Victor Lavrenko, Jeremy Pickens, "Polyphonic Music Modeling with Random Fields" Center for Intelligent Information Retrieval Department of Computer Science University of Massachusetts, Amherst, MA 01003 USA

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