5T-8

A Hidden Markov Model Approach for Onset Detection in Music Signals

Ai Takamatsu,Yukihiro Koizumi,Takashi Matsumoto
Department of Electrical Engineering and Bioscience, Waseda university, Tokyo, JAPAN
takamatsu12@matsumoto.eb.waseda.ac.jp
koizumi10@matsumoto.eb.waseda.ac.jp
takashi@matsumoto.elec.waseda.ac.jp

概要

本稿では、音楽情報処理の中でも基本的な問題の一つである発音時刻の検出問題に対して、隠れマルコフモデルを用いた手法の提案及び評価を行う。発音時刻の検出には、音楽データから得られる特徴を用いて、検出関数を作成し、そのピークを検出することで発音時刻を推定するのが一般的である。本稿では音楽特徴量を隠れマルコフモデルの出力とし、短時間フーリエ変換を用いて作成した複数の特徴量を考慮した予測を行う。実装はマルコフ連鎖モンテカルロ (Markov Chain Monte Carlo : MCMC) で行う。性能評価実験を行い、代表的な検出関数との比較を行う。

1　導入

近年、音楽データのデジタル化が進んできている。流れた音楽を自動でコンピュータ上に楽譜にする「音楽自動採譜システム」の実現も夢ではない時代になってきた。このシステムの実現には大きく分けて 3 つの課題がある。音を判別する周波数分析、曲調を判別するリズム分析、そしてそれらに基づき楽譜を作る採譜問題である。この中でも二つ目に述べたリズム分析に注目し、発音時刻の検出関数について考える。

2　隠れマルコフモデル

2.1　隠れマルコフモデル

隠れマルコフモデル (Hidden Markov Model:HMM) とは、時系列データをモデリングするのに有効な確率モデルである。近年では音声認識の他に、文字認識、タンパク質構造の予測、イベント検出などに応用されている。HMM は、状態列 $z := (z_1, \cdots, z_T)$、観測列 $y := (y_1, \cdots, y_T)$、及びパラメータ $\theta = (a, b, \pi)$ によって表現される。ここで T は状態列及び観測列の長さ、a は状態遷移確率、b は観測列の出力確率、π は観測列の初期状態確率を表す。本稿における観測列は、時刻 t で発生したイベント $e := (e_1, \cdots, e_T)$ 及び特徴量ベクトル $f := (f_1, \cdots, f_T)$ で構成され、時刻 t における発音の有無をイベントとして扱う。HMM のグラフィカルモデルを図 1 に表す。

3　アルゴリズム

本稿のアルゴリズムは学習フェーズと予測フェーズから構成される。アルゴリズムのイメージを図 2 に表す。学習フェーズでは与えられた観測列 $y = (e, f)$ を用いて HMM のパラメータ θ の学習を行う。予測フェーズでは学習済みパラメータ及び特徴量ベクトル f を用いて、イベント e を確率値として予測する。なお、HMM のパラメータの学習には MCMC を用いる。学習済みパラメータを用いた、イベント e の予測確率は以下の式で表される。

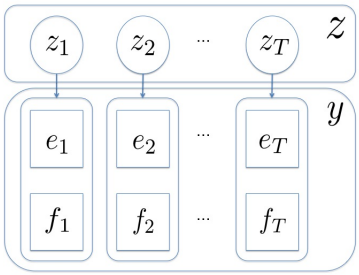

図 1　HMM のグラフィカルモデル

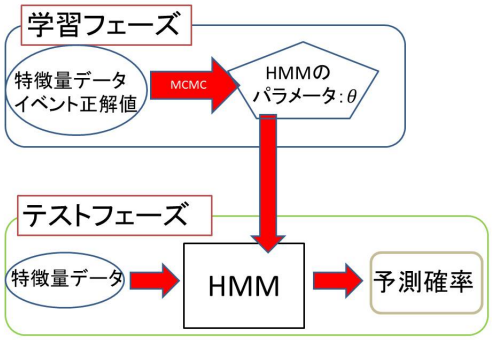

図 2　アルゴリズムのイメージ

$$P(e_t|f,\theta) = \sum_{z_t} P(e_t|z_t,\theta)P(z_t|f,\theta) \qquad (1)$$

4　特徴量

音楽データから短時間フーリエ変換 (Short-Time Fourier Transform : STFT) を用いて、特徴量データを作成した。STFT の式を (2) に示す。

$$STFT_{x,w}(t,\omega) = \int_{-\infty}^{\infty} x(\tau)w(\tau - t)e^{-i\omega\tau}d\tau \qquad (2)$$

ここで $x(t)$ は変換される関数、$w(t)$ は窓関数を表している。

4.1　Spectral Difference 特徴量

各サンプリング周波数、時系列ごとに得られたパワースペクトルを一時刻前のものと差をとる。Spectral Difference は検

出関数としても用いられており、実験では提案手法により作成した検出関数との比較対象としても用いる。参考文献 [1] より

$$SD(t) = \sum_{k=\frac{N}{2}}^{\frac{N}{2}-1} H(|X_k(t)| - |X_k(t-1)|)^2 \qquad (3)$$

ただし $H(x) = (x + |x|)/2$ である。また、N は窓関数の大きさである。そして時系列ごとに足し合わせる。

4.2 Powerspectram を用いた特徴量

STFT により得られたパワースペクトルを使う。

$$POWER_{x,w}[n,\omega] = |STFT_{x,w}(t,\omega)|^2 \qquad (4)$$

これらを時系列ごとに足し合わせる。

4.3 Phase 特徴量

STFT により得られた位相から、次の特徴量を作成する。参考文献 [2] より

$$d_\phi = princarg[\phi(t) - 2\phi(t-1) + \phi(t-2)] \qquad (5)$$

式 (5) 中の $princarg$ とは範囲を $[-\pi, \pi]$ にする関数である。また、d_ϕ は特徴量となる位相差であり、ϕ は位相である。

4.1〜4.3 を特徴量ベクトル f とする。

5 実験

提案モデルと作成した特徴量を用いて、検出関数を作成する。

5.1 実験データ

本稿の実験で用いたデータは、アップル社の音楽作成ツールである *garageBand* を用いて作成した。*garageBand* に組み込まれているピアノ音のサンプルを 2 秒毎に区切り、特徴量加工後に各データ長が 200 となるようデータ作成を行った。また、発音時刻の正解値は作成したデータの楽譜とテンポから算出したものである。

5.2 実験条件

表 1 に実験条件を示す。

表 1 実験条件

特徴量	Powerspectram 特徴量
	Spectral Difference 特徴量
	Phase 特徴量
離散数	10
状態数	10
MCMC サンプリング回数	10000
棄却数	9000
データ長	200
窓関数	hamming 窓
	98-leave-1-out cross validation

5.3 実験結果

Spectral Difference により得られた検出関数と HMM の特徴量ベクトル f を 1 次元の各特徴量とした結果、及び f を 3 次元ベクトルとしたときの結果の一例を図 3 に示す。図の実線は検出関数を表し、影の部分は実際の発音時刻を表す。

6 考察

図 3 より、一般的な検出関数である Spectral Difference に対して、1 次元の特徴量のみを用いた結果では及ばないが、複数の特徴量を用いることで、遜色のない結果を得ることができた。これは、一つの特徴量のみでは検出が難しいデータに対しても、複数の特徴量が互いに補い合って検出をしているためと考えられる。このように提案モデルは複数特徴量を同時に考慮できることが強みである。

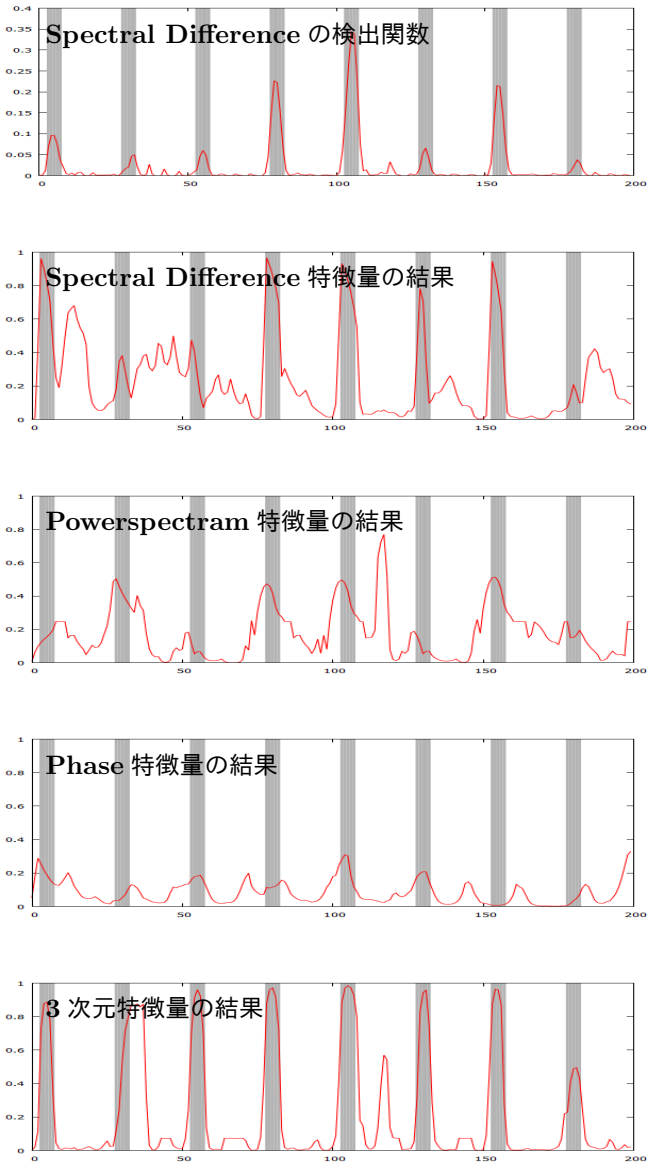

図 3 実験結果の一例

参考文献

[1] Juan Pablo Bello, Laurent Daudet, Samer Abdallah, Chris Duxbury, Mike Davies, and Mark B. Sandler, "A Tutorial on Onset Detection in Music Signals",IEEE Transaction on Speech and Audio Processing,2005.

[2] Juan P. Bello, Chris Duxbury, Mike Davies, and Mark Sandler, "On the Use of Phase and Energy for Musical Onset Detection in the Complex Domain",IEEE Signal Processing Letter, Vol. 11, No. 6, June 2004.

[3] Chris Duxbury, Juan Pablo Bello, Mike Davies, and Mark Sandler,"Complex Domain Onset Detestion For Musical Signals",Proc. of the 6th Int. Conference on Digital Audio Effects,2003.

[4] Ruohua Zhou, Marco Mattavelli, and Giorgio Zoia, "Music Onset Detection Based on Resonator Time Frequency Image",IEEE Transaction on Audio, Speech, and Language Processing, Vol. 16, No. 8, November 2008.