情報論的学習理論テクニカルレポート 2009
Technical Report on Information-Based Induction Sciences 2009 (IBIS2009)

個性を考慮した周期的全身運動のオンライン予測
On-line Stylistic Prediction for Human Periodic Motions

松原 崇充* Takamitsu Matsubara　玄 相昊† Sang-Ho Hyon　森本　淳‡ Jun Morimoto

Abstract: This paper describes a novel approach for on-line prediction of human periodic motions. A human periodic motion is modeled by a novel state-space model which is comprised of a low-dimensional phase dynamics and a two-factor (style/state factorial) observation model. Off-line and on-line EM-like algorithms are derived for efficient inference of both corresponding states and style parameters of the state-space model from unlabeled sequential observations. Based on the estimated state and style, future sequences of state and observation can be predicted. The effectiveness of our stylistic prediction is demonstrated for human walking (running) behaviors captured by mocap system.

Keywords: Human Motion Prediction, Style-Content Separation, State-Space Model.

1　はじめに

近年，人とロボットが共存する社会の実現に向けて，人間型・装着型・環境型など様々な形態を取るロボット及びその必要基盤技術の研究開発が活発に行われている．特に，人の動作を認識した上でその未来の状態を実時間で予測する機能の実現は，ロボットが人とインタラクションするために不可欠な要素となる．

人の動作認識・予測問題に対して，Hidden Markov Models(HMMs)[7]，Switching Linear Dynamical Models(SLDMs) [12] や Gaussian Process Dynamical Models (GPDMs)[20] に代表される統計的学習手法に基づくアプローチが取られ，動作の生成モデル化，認識と予測などに一定の成果を上げている．しかしながら，人の動作には多様性（個性）が存在する．歩行動作を例に取れば，個人によって歩き方の癖は異なり，また，歩幅や歩行速度などによっても歩容は大きく異なり得る．このような動作の多様性を考慮しない従来のモデル化法では，多様性を有するデータセットに対して，その平均的なモデルを学習することになる．しかし，それでは個々の動作の予測に適したモデル化が行えたことにはならない．多様なデータを平均化したモデルが，非現実的な動作を表現する可能性もある．

近年，生成モデルにおける潜在変数である状態変数とは独立に，動作の空間的な特徴を大域的に捉えるスタイル変数を導入することで，データの個性を陽に取り扱う考え方が提案されてきている．画像処理 (Computer Vision:CV) 分野においては，アピアランスベースドの認識の文脈においていくつかの研究がある [16, 5, 4, 10]．Tenenbaum らは，スタイル及び（離散的）状態変数が既知なデータセットを対象に，スタイル変数と状態変数が観測に及ぼす影響を分離する双線形モデルの学習法と，テストデータに対するクラス分類や，状態変数に関する内挿や外挿による観測の予測法を提案している．また，その有効性を顔画像，発話音声，手書き文字の 3 種類のベンチマークデータに対して検証している [16]．Elgammal らは，Tenenbaum らによって提案された方法を基に，連続かつ低次元な状態変数を対象に，状態変数から観測への写像を非線形なものに拡張している．各スタイルにおけるデータセットが特定の低次元多様体を成すように，非線形関数のパラメータを最小二乗規範によって学習する．各観測に対する状態変数は一般に未知となるため，Locally Linear Embedding(LLE) による多様体学習によって特定する．連続的に観測された歩容画像や表情画像に対して，個人認証や表情合成に適用している [5, 4, 10].

*奈良先端科学技術大学院大学/ATR 脳情報研究所, 630-0101 奈良県生駒市高山町 8916-5, e-mail takam-m@is.naist.jp
Nara Institute of Science and Technology, 8916-5 Takayama-cho Ikoma Nara 630-0192 JAPAN
†ATR 脳情報研究所, 619-0288 京都府相楽郡精華町 2-2-2, e-mail sangho@atr.jp
ATR Computational Neuroscience Laboratories, 2-2-2 Hikaridai Seika-cho Souraku-gun Kyoto 619-0288 JAPAN
‡ATR 脳情報研究所, 619-0288 京都府相楽郡精華町 2-2-2, e-mail xmorimo@atr.jp
ATR Computational Neuroscience Laboratories, 2-2-2 Hikaridai Seika-cho Souraku-gun Kyoto 619-0288 JAPAN

一方，コンピュータグラフィックス（Computer Graphics:CG）分野においても，スタイル変数を陽に取り扱う方法が発展してきている [3, 1, 6, 18, 19, 15]．CV 分野での主目的が認識や補間であるのに対して，CG 分野での目的は主に直感的な操作における人動作グラフィックスの合成である．つまり，ある人の動作グラフィックスを，他人の個性を考慮したものに合成するなどの処理である．これには主に，データベースなどを利用して入力された動作を望ましいスタイルに変換する方法と [1, 6, 18]，多様な動作データを動的システムとしてモデル化し，そのパラメータ空間における補間によって望ましい合成を実現する方法がある [3, 19, 15]．

データベースに基づく手法の一例として，Arikan らは，スタイルラベルの付いた動作データベースから，任意の動作を合成する手法を提案している．人手によるラベル付きデータを SVM により汎化し，合成処理は動的計画法により効率的に計算される [1]．ただし，データベースにないスタイルの動作を生成することはできない．Hsu らは，同じ状態を持つ異なるスタイル動作間を比較・解析することで，入力された動作を望ましいスタイルに変換する方法を提案している．変換には線形動的システムに基づくシステム同定法を利用する．また，異なるスタイル動作データの時空間的な対応関係を推定する反復計算法も合わせて提案している [6]．Torresani らは，Laban Movement Analysis(LMA) と呼ばれる特徴量を用いてスタイル変数を定義することで，入力動作を任意のスタイルに合成する方法を提案している [18]．

上記のスタイル合成写像を学習する方法と比較して，以下に述べる多様な動作の生成モデルを取り扱う研究は，本研究が目指す人動作の予測の目的により関連が強い．Brand らは Parametric HMM[21] に基づく多様な動作データのための生成モデルとして Style Machine を提案している [3]．さらに，HMM よりも表現能力や計算効率に優れたモデルとして，Boltzman Machine[15] やガウス過程に基づく Multi-Factor Gaussian Process Dynamics Models(MFGPDM)[19] などが提案されている．これらはすべてモーションキャプチャシステムによって計測されたデータに対して，多様性を有するデータを学習し，かつ指定するスタイルの動作を動的システムによって予測することが可能であることを実証している．しかしながら，HMM[3] や Boltzman Machine[15] では，対象動作の複雑さに伴い膨大なパラメータを扱う必要があり，MFGPDM[19] では予測や適応の際に，学習データ数に応じた計算量が必要となるため，本研究が目指すオンラインの動作予測には適さない．

以上の議論から，CV 及び CG 分野で提案されてきた

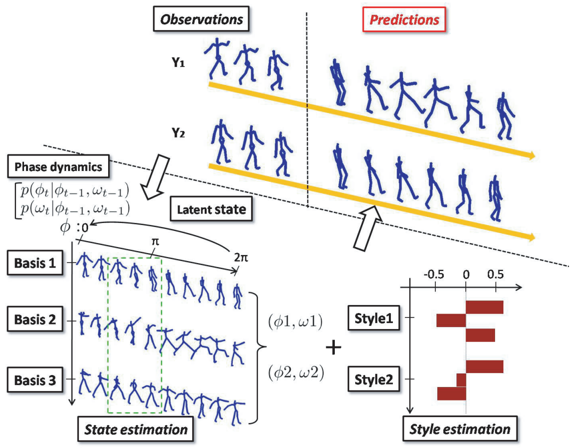


図 1: 提案する全身周期運動予測法の概要図．位相に基づく潜在変数と，観測データの個性を表現するスタイルパラメータを観測データから同時に推定．

関連手法では，スタイルの認識や合成といった目的に合わせた設計がなされている．認識を目的とする方法では，個性を低次元のパラメータで捉えるが，現在の状態から未来の状態を予測する機能を持たない．また，合成法では，人の直感的な操作で任意のスタイル動作を生成することを目的として，多数のパラメータを有する複雑なモデルが用いられることから，わずかなデータからスタイルや状態を推定することは過学習の観点から困難である．本研究の目的であるオンライン予測を実現するためには，逐次的に観測されるテストデータに適合するスタイル/状態変数を効率的に推定可能，かつ実時間で予測処理を行うことも考慮した，シンプルな生成モデルを設計する必要があると考えられる．また，生成モデルに合った効率的な適応・予測アルゴリズムも必要となる．

本稿では，歩行や走行などに代表される人の全身周期運動データのモデル化に焦点を当て，オンライン予測の目的に適した生成モデルと，その適応・予測アルゴリズムの導出を行う．提案手法の概要を図 1 に示す．提案する潜在変数モデルは状態空間モデルの一種であり，多様な動作は共通して定常周期運動であることに基づく位相を潜在変数とする低次元動的システムと，状態変数とは独立して動作の個性を捉える低次元のスタイルパラメータを有する観測モデルで構成される．このモデルに基づいて個人の癖などのいわゆる個性を捉えた正確な動作予測を実現するため，以下の二つの処理を考える．一つ目は，多様性を含んだデータ集合から，潜在変数モデルの学習を行う処理である．二つ目は，スタイル及び状態変数が未知のテストデータから両変数を同時推定し，その結果に基づいて未来の観測を予測する処理である．観測モデルの学習に必要な状態依存基底の推定問題

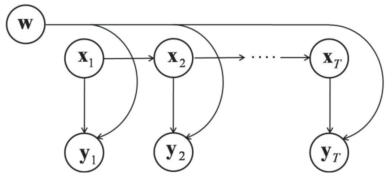


図 2: 全身周期運動の生成モデル．時刻 t の状態 $\mathbf{x}_t$，観測 $\mathbf{y}_t$ に加え，観測データの個性を特定するスタイルパラメータ $\mathbf{w}$ を持つ．

はベイズ推定として定式化される．状態の変化に対して観測が滑らかに変化することを事前分布の形で導入する．テストデータに対する適応は，拡張カルマンフィルタを利用した近似的な Expectation-Maximization(EM) アルゴリズムにより，スタイルパラメータ及び状態変数を効率的に推定する．さらに，Sato らの方法 [14] に基づき，推定アルゴリズムをオンライン化する．導出されるオンラインアルゴリズムでは，過去のデータを効率的に忘却しながら，逐次的に観測されるデータの性質に追従するため，動作の性質が突然切り替わる動的な状況にも対応可能である．

本論文の構成について述べる．2 節では，多様なデータからの生成モデルの学習法について述べる．3 節では，テストデータに対する生成モデルの適応及び，観測データの予測アルゴリズムについて示す．4 節では，提案法をモーションキャプチャで計測された人の歩行及び走行データに対する予測結果について報告し，5 節において結論と今後の課題について議論する．

2　全身周期運動の生成モデル

提案する生成モデルは，潜在変数 $\mathbf{x} \in \mathcal{R}^d$ を有し，観測 $\mathbf{y} \in \mathcal{R}^D$ についての条件付き分布 $p(\mathbf{y}_t|\mathbf{x}_t;\mathbf{w})$，潜在変数の一次マルコフ連鎖の条件付き分布 $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ によって構成される潜在変数モデルである．ここで，$\mathbf{w} \in \mathcal{R}^J$ は付加的に導入された潜在変数であり，これにより動作の個性を特定する．以後 $\mathbf{w}$ をスタイルパラメータと呼ぶ．そのグラフィカルモデルを図 2 に示す．定常周期運動を対象とするため，各動作の状態 $\mathbf{x}$ は位相と角速度によって表現するのが自然である．本稿では，位相は低次元多様体 $\mathcal{S}^n \subset \mathcal{R}^m (n < m)$ 上の点であると仮定し，さらに多様体上の各点と観測データ点とを確率的回帰モデルによって関係付ける多様体埋め込み法を用いる [5, 4]．これらを考慮して，状態遷移モデル，観測モデル共にガウス分布であると仮定して次のように定義する．

$$p(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \mu_x(\mathbf{x}_{t-1}), \Sigma_x(\mathbf{x}_{t-1})) \quad (1)$$

$$p(\mathbf{y}_t|\tilde{\mathbf{x}}_t;\mathbf{w}) = \mathcal{N}(\mathbf{y}_t; \mu_y(\tilde{\mathbf{x}}_t,\mathbf{w}), \Sigma_y(\tilde{\mathbf{x}}_t)) \quad (2)$$

ここで $\mathbf{x}_t = [\omega_t, \phi_t]^T$, $\tilde{\mathbf{x}}_t = g(\mathbf{x}_t) = [\cos(\phi_t), \sin(\phi_t)]^T$, $g(\cdot)$ は写像 $\mathcal{S}^1 \mapsto \mathcal{R}^2$ である．つまり，観測モデルは位相から全身関節姿勢への関数であり，角速度は時系列データの動的な特性を表現するためにのみ用いられる．ϕ は周期運動の位相であり，ω は ϕ の時間微分である．$\tilde{\mathbf{x}}_t$ は同径 $r = 1$，偏角 ϕ とする 1 次元球面 $\mathcal{S}^1$ の点を $\mathcal{R}^2$ で表した変数である．この変換により，観測モデルでは入力の計量として $\mathcal{S}^1$ 上の測地線距離を $\mathcal{R}^2$ 上でのユークリッド距離で近似的に扱う．

全身周期運動の生成モデル学習法の概要を示す．具体的には次の 3 つの手順によって実現される．

1. 周期性に基づくデータアラインメント (2.1 節)
2. スタイル/状態変数のベイズ推定 (2.2 節)
3. 多様体埋込み法による生成モデル学習 (2.3 節)

1. では，対象動作の周期性に基づいて，各データを潜在変数である位相を基軸として配列させる．2. では，配列されたデータセットから低次元の状態依存基底をスタイルから滑らかに分離・推定する．この基底とスタイルパラメータとの内積により，個性を特定した観測データを復元することができる．3. では，推定された状態依存基底をそれぞれガウス過程としてモデル化する．この際，位相値 $\tilde{\mathbf{x}}$ に基づいて共分散行列を決定することで，低次元多様体を基底に埋め込む．以上の処理により，位相，角速度及び，個性を特定する低次元のスタイルパラメータを有する多様な全身周期運動のシンプルかつコンパクトな生成モデルが学習される．

2.1　周期性に基づくデータアラインメント

各動作データがすべて周期性を有すると仮定して，相関解析に基づいて各データを位相値に関して配列するためのアルゴリズムについて述べる．

$\mathbf{Y}^s = [\mathbf{y}_1^s \cdots \mathbf{y}_{C(s)}^s]^T \in \mathcal{R}^{C(s)\times D}$ を特定の人が特定の意図を持って行った定常周期運動からの観測時系列データとする．ここで，$s \in \{1, 2, \cdots, S\}$ はスタイル，$c \in \{1, 2, \cdots, C(s)\}$ は状態のインデックス，$\mathbf{y}_c^s \in \mathcal{R}^D$ はスタイル s 及び状態 c に対応する観測（全身関節角度や体幹の位置及び姿勢など）をそれぞれ表す．状態のインデックスは固有の位相値 ϕ^c を持つとする．つまり，学習データとして $\mathcal{D} = \left\{\mathbf{Y}^1, \cdots, \mathbf{Y}^S\right\}$ が与えられていると前提する．これをすべて位相を軸として配列させるため，自己相関係数と相互相関係数を利用する．まず，以下の式で与えられる自己相関係数最大化によって，各動作データの周期 T を推定する．

$$T^s \leftarrow \arg\max_j A^s(j) \quad (3)$$

ここで，$A^s(j)=\sum_n^N {\mathbf{y}_n^s}^T\mathbf{y}_{n+j}^s$ である．Makiharaらも，自己相関係数を歩行画像からの周期算出のための自己相関を利用している [11]．さらに，相互相関係数最大化により，周期的データとしての類似度を最大にする位相軸上でのシフト量を推定する．

$$h^s \leftarrow \arg\max_j C^s(j) \tag{4}$$

ここで，$C^s(j)=\sum_n^N {\mathbf{y}_n^b}^T\mathbf{y}^s_{j+\mathrm{rd}\left(\frac{nT^s}{T^b}\right)}$ であり，インデックス b は周期が最も短いデータセットのスタイルインデックスとする[1]．これらの結果により，以下の式で表されるように，各動作データを同一位相値毎に配列させることができる．

$$\mathbf{Y}^{all}=\begin{bmatrix} \mathbf{y}^1_{\mathrm{rd}\left(h^1+\frac{T^1}{T^b}\right)} & \cdots & \mathbf{y}^1_{\mathrm{rd}\left(h^1+T^1\right)} \\ \vdots & & \vdots \\ \mathbf{y}^b_1 & \ddots & \mathbf{y}^b_{T^b} \\ \vdots & & \vdots \\ \mathbf{y}^S_{\mathrm{rd}\left(h^S+\frac{T^S}{T^b}\right)} & \cdots & \mathbf{y}^S_{\mathrm{rd}\left(h^S+T^S\right)} \end{bmatrix} \tag{5}$$

ここで，$\mathbf{Y}^{all}=[\bar{\mathbf{Y}}^1_a\cdots\bar{\mathbf{Y}}^S_a]^T\in\mathcal{R}^{DS\times C}$ である．また，$\mathbf{Y}^s_a\in\mathcal{R}^{C\times D}$ は，アラインメント後のスタイル s の時系列データである．

2.2 ベイズ推定によるスタイル/状態分解

本節では，多様なデータをそれぞれ位相を軸に配列させたデータセット $\mathbf{Y}^{all}$ から，スタイル及び状態基底を分離・推定する方法について述べる．

データ集合がスタイルと状態に対して等しく分布している場合，特異値分解によりスタイルと状態を分離表現する双線形モデルを得ることができる [16]．はじめに，$DS\times C$ 行列 $\mathbf{Y}^{all}$ を $S\times DC$ 行列に並べ替えたものを ${\mathbf{Y}^{all}}^{\mathrm{VT}}$ と定義する．これに対して特異値分解を適用すると，以下の式で与えられるように二つの行列に分解できる．

$${\mathbf{Y}^{all}}^{\mathrm{VT}}=\mathbf{USV}^T\approx\mathbf{W}\tilde{\mathbf{Y}} \tag{6}$$

ここで，$\mathbf{W}=[\mathbf{w}^1\cdots\mathbf{w}^S]^T\in\mathcal{R}^{S\times J}$ は $\mathbf{US}$ の $J(\le S)$ 列までの部分行列，$\tilde{\mathbf{Y}}=\left([\bar{\mathbf{Y}}^1\cdots\bar{\mathbf{Y}}^J]^T\right)^{\mathrm{VT}}\in\mathcal{R}^{J\times DC}$ は $\mathbf{V}^T$ の J 行までの部分行列である．つまり，分解に用いたデータ集合は $\mathbf{Y}^s_a\approx\sum_{j=1}^J w^s_j\bar{\mathbf{Y}}^j$ として近似的に表現される．ここで，$\bar{\mathbf{Y}}^j\in\mathcal{R}^{C\times D}$ はスタイルに独立な第 j 状態依存基底，$\mathbf{w}^s\in\mathcal{R}^J$ は第 s スタイルパラメータである．観測ノイズにガウスノイズを仮定したとき，上記の方法が最尤推定解を与える [16]．一般に，観測データ間に相関が高いときには，有効な値を持つ特異値の数は少ない．そのため，特異値の分布に応じて適切な J を設定することで，観測データを十分な精度で説明するコンパクトな基底セットを抽出することができる．

しかしながら，観測データがノイズを含み，さらにデータ集合が十分に与えられない場合には過学習の危険がある．この過学習の問題を回避するため，状態依存基底及びスタイルパラメータに事前分布を導入し，ベイズ推定の枠組みによってこの推論問題を取り扱う．まず，各状態依存基底の事前分布を以下のように定義する．

$$p(\bar{\mathbf{Y}}^j|\bar{\alpha})=p(\mathbf{y}^j_1)\cdot\frac{1}{\sqrt{(2\pi)^{(C-1)D}|\mathbf{K}_s|^D}} \cdot\exp\left(-\frac{1}{2}\mathrm{Tr}\left(\mathbf{K}_s^{-1}\mathbf{Y}^j_{out}{\mathbf{Y}^j_{out}}^T\right)\right) \tag{7}$$

ここで，$\bar{\mathbf{y}}^j_c$ は第 j 状態基底の第 c 行ベクトル，$\bar{\mathbf{Y}}^j=[\bar{\mathbf{y}}^j_1\cdots\bar{\mathbf{y}}^j_C]^T\in\mathcal{R}^{C\times D}$，$\mathbf{Y}^j_{out}=[\mathbf{y}^j_2\cdots\mathbf{y}^j_C]^T\in\mathcal{R}^{(C-1)\times D}$ と定義される．$\mathbf{K}_s\in\mathcal{R}^{(C-1)\times(C-1)}$ はグラム行列，その $(c1,c2)$ 要素は $k_s(\mathbf{y}^j_{c1},\mathbf{y}^j_{c2})$ である．ここで，$k_s(\mathbf{x},\mathbf{x}')=\alpha_1\exp\left(-\frac{\alpha_2}{2}||\mathbf{x}-\mathbf{x}'||^2\right)+\alpha_3\mathbf{x}^T\mathbf{x}'+\alpha_4^{-1}\delta_{\mathbf{x},\mathbf{x}'}$ とし，$\bar{\alpha}=\{\alpha_1,\alpha_2,\alpha_3,\alpha_4\}$ は超パラメータである．

この事前分布は状態基底が位相の変化，つまり時間変化に対して観測の変化が滑らかであることを意味している．スタイルパラメータに関する事前分布 $p(\mathbf{w}^s)$ は，それぞれ期待値が0の等方性ガウス分布 $p(\mathbf{w}^s|\gamma)=\mathcal{N}(\mathbf{w}^s|0,\gamma^{-1}\mathbf{I})$ と置く．これは $\mathbf{w}^s$ の正則化項として働くため，過学習を防ぐ効果が期待できる [2]．このような事前分布及び，加法性ガウスノイズを仮定した尤度関数を用いることで，様々なスタイルを含むデータ集合 $\mathcal{D}$ が与えられた場合の状態基底 $\bar{\mathbf{Y}}^{1:J}$ とスタイルパラメータ $\mathbf{w}^{1:J}$ の推定問題を，事後分布の最大化問題 (Maximum A Posterior: MAP 推定) として取り扱うことができる．本稿では，このMAP解を以下の式で与えられる拘束条件付き局所最適化問題を解くことによって求める．

$$\mathrm{minimize}_{\bar{\mathbf{Y}},\mathbf{w}} \quad p(\bar{\mathbf{Y}}^{1:J},\mathbf{w}^{1:J}|\mathcal{D},\bar{\alpha},\gamma) \tag{8}$$

$$\text{subject to} \quad \mathbf{y}^s_1=\mathbf{y}^s_C;\ \forall s\in\{1,\cdots,S\} \tag{9}$$

ここで，

$$p(\bar{\mathbf{Y}}^{1:J},\mathbf{w}^{1:J}|\mathcal{D},\bar{\alpha},\gamma) \propto p(\mathcal{D}|\mathbf{Y}^{1:J},\mathbf{w}^{1:J})p(\bar{\mathbf{Y}}^{1:J}|\bar{\alpha})p(\mathbf{w}^{1:J}|\gamma) \tag{10}$$

である．式 (9) で与えられる拘束条件は，各基底に対して周期性を拘束するためのものである．これによって，滑らかな周期運動の生成が期待できる．このようなMAP解法は [9, 20] でも用いられている．また，式 (6) で示した特異値分解の解をそれぞれの初期値とすることで，効率的に解を探索できると考えられる．

[1] $\mathrm{rd}(\cdot)$ は少数点以下を四捨五入により切り上げる関数である

2.3 多様体埋込み法による生成モデルの学習

本節では，前節で述べた方法により推定される状態依存基底を用いて，多様なデータを生成するコンパクトなモデルの学習法について述べる．観測モデルに関しては，推定された状態依存基底の各状態に対して 1 次元球面上の対応点と回帰モデルにより結びつけることで，各基底に多様体を潜在させる．さらに，位相の従う動的システムを仮定することで生成モデルを得る．回帰モデルには，周辺尤度や予測分布などの扱いが容易なガウス過程回帰 [13] を利用する．

各基底 $\bar{\mathbf{Y}}^j$ は対応する位相値 $\tilde{\mathbf{X}}$ を入力として，以下の式のように独立にガウス過程としてそれぞれモデル化される．

$$p(\bar{\mathbf{Y}}^j|\tilde{\mathbf{X}},\bar{\beta}) = \frac{1}{\sqrt{(2\pi)^{CD}|\mathbf{K}_y|^D}} \cdot \exp\left(-\frac{1}{2}\mathrm{Tr}\left(\mathbf{K}_y^{-1}\bar{\mathbf{Y}}\bar{\mathbf{Y}}^T\right)\right) \qquad (11)$$

ここで $\mathbf{K}_y \in \mathcal{R}^{C\times C}$ はグラム行列であり，(i,j) 要素は $k_y(\tilde{\mathbf{x}}_i,\tilde{\mathbf{x}}'_j)$，$k_y(\mathbf{x},\mathbf{x}') = \beta_1 \exp\left(-\frac{\beta_2}{2}||\mathbf{x}-\mathbf{x}'||^2\right)+\beta_3\mathbf{x}^T\mathbf{x}'+\beta_4^{-1}\delta_{\mathbf{x},\mathbf{x}'}$，超パラメータは $\bar{\beta} = \{\beta_1,\beta_2,\beta_3,\beta_4\}$ とする．未知入力 $\tilde{\mathbf{x}}^*$ が与えられた場合の予測分布は，$p(\bar{\mathbf{y}}^{j*}|\tilde{\mathbf{x}}^*,\bar{\mathbf{Y}}^j,\tilde{\mathbf{X}}) = \mathcal{N}(\mu^j(\tilde{\mathbf{x}}),\Sigma^j(\tilde{\mathbf{x}}))$ と導出される [13]．$\mu^j(\tilde{\mathbf{x}})$，$\Sigma^j(\tilde{\mathbf{x}})$ 共に解析的に計算される．

特定のスタイルを持つ動作データに関する予測分布は，ガウス分布の加法性及び線形変換に関する分布不変性に基づいて，スタイルパラメータ $\mathbf{w}^s$ を用いて次のように導出される．

$$p(\mathbf{y}|\tilde{\mathbf{x}},\bar{\mathbf{Y}}^{1:J},\tilde{\mathbf{X}};\mathbf{w}^s) = \mathcal{N}(\mu_y(\tilde{\mathbf{x}}),\Sigma_y(\tilde{\mathbf{x}})) \qquad (12)$$

ここで，$\mu_y(\tilde{\mathbf{x}}) = \sum_{j=1}^J w_j^s \mu^j(\tilde{\mathbf{x}})$，$\Sigma_y(\tilde{\mathbf{x}}) = \sum_{j=1}^J (w_j^s)^2\Sigma^j(\tilde{\mathbf{x}})$ である．これが式 (2) に示した生成モデルにおける観測モデルとなる．

次に，状態遷移モデル $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ はすべてのデータが固定の時間刻みで観測されるとして，$p(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{A}\mathbf{x}_{t-1},\mathbf{Q})$ とモデル化される．ここで，状態遷移行列は $\mathbf{A} = [0\ 1; 1\ 1]$，プロセスノイズ行列 $\mathbf{Q}$ はデータに応じて設計される．状態変数は $\mathbf{x}_t = [\omega_t,\phi_t]^T$ と定義し，これにより各観測に対応する位相だけでなく，角速度も潜在変数として扱うことで，多様な速度の動作を捉えることが可能となる．これが式 (1) に示した生成モデルにおける状態遷移モデルとなる．

Wang らの MFGPDMs[20] では，状態遷移モデルと観測モデルの両方の学習を，学習データ数の次元の非線形局所最適化問題として扱うため，その解は初期値に大きく依存する．また，テストデータへの適応にも同様に，非線形局所最適化問題を解く必要がある．これらは，MFGPDMs ではすべてのデータセットを保持する必要があることに起因する．

本稿で提案する手法では，動作の周期性に基づいて単一の状態遷移モデルを特定する．これによって，観測データのアラインメント (2.1 節) が容易となり，結果少数の基底で構成されるコンパクトな観測モデルの学習を因子分解 (2.2 節) と回帰問題 (2.3 節) として分離して取り扱える．さらに，生成モデルの構造がシンプルであるため，次節で示すように，オンライン予測を実現するために不可欠な，状態変数及びスタイルパラメータの効率的な推定アルゴリズムを EM に基づいて導出することができる．

3 EM による状態変数/スタイルパラメータの同時推定

本節では，学習された生成モデルを用いて，与えられたテストデータ $\hat{\mathbf{Y}} = \{\hat{\mathbf{y}}_1\cdots\hat{\mathbf{y}}_T\}^T$ に対応する状態 $\hat{\mathbf{X}} = \{\hat{\mathbf{x}}_1\cdots\hat{\mathbf{x}}_T\}^T$ 及びスタイルパラメータ $\hat{\mathbf{w}}$ を同時推定するための近似的な EM アルゴリズムを導出する．3.1 節では，効率的な処理を実現するため，E，M ステップともに解析解を持つように近似を導入したオフラインアルゴリズムについて述べる．さらに 3.2 節において，尤度関数に忘却率を導入することで，E，M ステップともに逐次的な計算で実現されるオンラインアルゴリズムについて述べる．3.3 節では，スタイルパラメータと状態変数が推定された後，テストデータの予測を行うアルゴリズムについて述べる．

3.1 近似的 EM によるスタイル/状態推定

EM アルゴリズムは，潜在変数を持つ確率モデルにおけるパラメータの最尤解を見出すアルゴリズムである．具体的には，E ステップと M ステップと呼ばれる処理を交互に繰り返しながら，被最適化パラメータの最尤解を求める．しかしながら，式 (11)，(12) で与えられる生成モデルに対して EM アルゴリズムを適用する場合，観測モデルの非線形性のため，E ステップ及び M ステップの解析解は与えられない．そこで本稿では近似的な手法を導入する．前節で述べたように，観測モデルにおける各状態依存基底は非線形性を有するものの，事前分布として滑らかな変化を好むように推定されている．この滑らかさに基づいて，E ステップの近似的解法に拡張カルマンフィルタを利用する [17]．拡張カルマンフィルタでは，観測モデルの期待値に対する局所線形近似モデルが必要となるが，$\mathbf{x}_q$ における係数行列は

$\mathbf{H}_q = \frac{\partial \mu}{\partial \mathbf{x}}|_{\mathbf{x}_q} = \hat{\mathbf{Y}}(\hat{\mathbf{w}})\mathbf{K}_y^{-1}\frac{\partial \mathbf{k}_y(\mathbf{x}_q)}{\partial \mathbf{x}}^T$ と求められる [8].

M ステップに関しては，以下のような近似を導入する．

$$\begin{aligned} \hat{\mathbf{w}}_{k+1} &\leftarrow \arg\max_{\hat{\mathbf{w}}} \int_{\hat{\mathbf{X}}} p(\hat{\mathbf{X}}|\hat{\mathbf{Y}}, \hat{\mathbf{w}}_k) \log p(\hat{\mathbf{X}}, \hat{\mathbf{Y}}|\hat{\mathbf{w}}) d\hat{\mathbf{X}} \\ &\approx \arg\max_{\hat{\mathbf{w}}} \log p(\bar{\mathbf{X}}, \hat{\mathbf{Y}}|\hat{\mathbf{w}}) \end{aligned} \tag{13}$$

ここで $\bar{\mathbf{X}} = \int_{\hat{\mathbf{X}}} \hat{\mathbf{X}} p(\hat{\mathbf{X}}|\hat{\mathbf{Y}}, \hat{\mathbf{w}}_k) d\hat{\mathbf{X}}$ である．これは，観測 $\hat{\mathbf{Y}}$ が与えられたときの，状態変数 $\hat{\mathbf{X}}$ の事後分布が単峰かつ，分散の小さいときに有効な近似である．この近似によって M ステップも解析解を持つ．

上記の近似 EM アルゴリズムは次のように実行される．まずスタイルパラメータの初期値 $\hat{\mathbf{w}}_0$ を設定する．次に，E ステップからアルゴリズムを展開し，尤度が収束するまで繰り返す．これよって，テストデータを説明する $\hat{\mathbf{w}}$ と $\hat{\mathbf{X}}$ が同時に推定される．アルゴリズムの詳細は付録 A に記す．

3.2 オンライン EM への拡張

前節で導出した EM アルゴリズムでは，E ステップは逐次的な計算で実現されるのに対して，M ステップではすべての観測データを用いて計算される，いわゆるバッチ型学習則である．これに対して，スタイルパラメータが観測モデルの期待値における線形パラメータであることに注目すると，Sato ら [14] の方法を用いて導出された EM アルゴリズムをオンライン化することができる．つまり，E ステップと同様に M ステップも逐次的にデータが与えられるたびにパラメータを更新できるオンライン学習法に修正できる．

オンライン EM[14] では，過去の情報ほど効率的に忘れていくように，忘却率 $\lambda(t)(0 \le \lambda(t) \le 1)$ を導入した修正版平均値を関数 $f(x)$ に関して次のように定義する．

$$\ll f(x) \gg_T \equiv \eta_T \sum_{t=1}^{T} \left\{ \prod_{s=t+1}^{T} \lambda_s \right\} f(x_t) \tag{14}$$

ここで，$\eta_T \equiv \left\{ \sum_{t=1}^{T} \left\{ \prod_{s=t+1}^{T} \lambda_s \right\} \right\}^{-1}$ は正規化係数である．さらに，この修正版平均値は次のような逐次計算式を持つ．

$$\ll f(x) \gg_T = (1 - \eta_T) \ll f(x) \gg_{T-1} + \eta_T f(x_T) \tag{15}$$

$$\eta_T = \left\{ 1 + \frac{\lambda_T}{\eta_{T-1}} \right\}^{-1} \tag{16}$$

これらの逐次的な期待値計算を M ステップに利用することで M ステップも逐次計算可能になり，状態変数，スタイルパラメータ共に逐次的に更新可能となる．詳細は付録 A に記す．また，忘却係数を $\lambda < 1.0$ と設定することで，非定常的なデータ（例えば，歩行から走行に突然動作が変化した場合）に対しても追従することが期待できる．この性質は，オンラインで動作予測をする際に，どの時刻からテストデータとして観測するかという動作の開始点を推定する問題を回避できるため，実用上非常に有益である．

3.3 平均値伝搬による動作予測

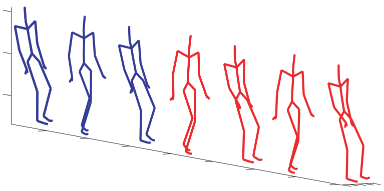
テストデータに対して推定されたスタイルパラメータ及び状態変数に基づいて，動作を予測するアルゴリズムについて述べる．時刻 t までのテストデータが観測されているとき，EM アルゴリズムによって対応する状態の期待値 $\hat{\mathbf{x}}_t$ が推定されているとする．その期待値の 1 時刻先の予測値 $\hat{\mathbf{x}}_{t+1|t}$ は，式 (12) に従った線形変換によって与えられる．さらに，対応する観測の期待値 $\hat{\mathbf{y}}_{t+1|t}$ は，式 (11) により解析的に計算される．ここでは，推定されたスタイルパラメータ $\hat{\mathbf{w}}_t$ に基づいた個性を考慮した予測が行われる．本稿では，このような方法を任意の時刻まで繰り返すことで，個性を考慮した動作予測を実現する．

4 個性を考慮した全身周期運動予測

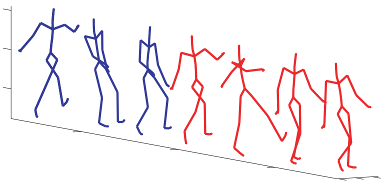
本節では，モーションキャプチャシステムで計測された人の歩行及び走行動作データを対象として，提案手法の有効性を検証する．

4.1 オフライン EM による適応・予測結果

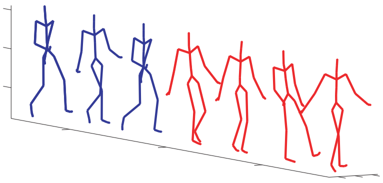
EM アルゴリズムにより個性を考慮した歩行動作の高精度な予測を試みる．人の動作データとして，モーションキャプチャシステムで計測された全身関節データを用いる（CMU:database を利用，http://mocap.cs.cmu.edu）．各時刻の姿勢データ $\mathbf{y}_t \in \mathcal{R}^{62}$ は 120Hz で計測されており，関節角度情報 $\mathbf{q}_t \in \mathcal{R}^{56}$ と体節の絶対位置 $\mathbf{v}_t \in \mathcal{R}^3$ 及び姿勢情報 $\mathbf{r}_t \in \mathcal{R}^3$ で構成されている．本実験では，これらのデータを 60Hz にダウンサンプルし，さらに関節角度の 56 次元のデータのみを用いる．周期運動として，歩行運動を選択し，同一被験者から計測された 8 種類の歩行運動のデータ集合 (08_02.amc ∼ 08_09.amc) を学習データ，3 種類のデータ集合 (08_01.amc, 08_07.amc, 08_10.amc) をテストデータとして用いた．スタイルパラメータの次元は，特異値の分布から $J = 2$ と設定した．ガウス過程回帰の超パラメータは周辺尤度に基づく学習が可能 [13] であるが，本実験では経験的に設定した．実験結果を図 3 に示す．実験では，$0 \sim 1.67$ sec 間の観測データを用いて生成モデルをテストデータに適応



(a)



(b)



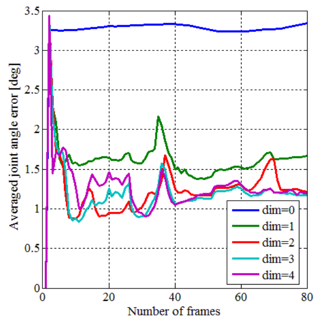
(c)

図 3: 提案法による歩行運動のオフライン予測の結果．(a):08_01.amc, (b):08_07.amc, (c):08_10.amc，青線は観測データ，赤線は予測された結果を表す．(a) は手足の振りが小さく，(b) 及び (c) では手足を大きく振っているそれぞれの特徴が予測結果にも現れており，個性を捉えた予測が行えていることがわかる．すべての姿勢は 2Hz でサンプリングされている．

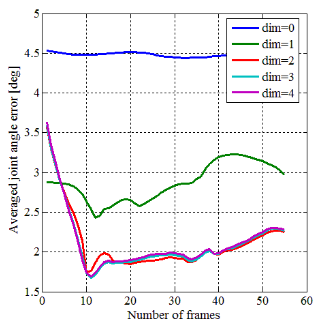
させ，その結果に基づいて 1.67 ∼ 4 sec 間の運動を予測した．図 3 より，(a) は歩幅や腕の振りが小さい歩行動作のデータであり，一方 (b) 及び (c) は，腕を大きく振り，歩幅も大きい歩行動作のデータであることが分かる．人の関節動作は高次元であるにも関わらず，提案手法では少ないデータから，動作の個性と状態を同時推定し，個性を捉えた高精度な動作予測が可能であることがわかる．

4.2 オンライン EM による適応・予測結果

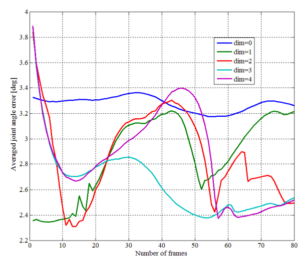
オンライン EM アルゴリズムによるテストデータへの適応及び予測を行う．ここでは，前節よりも学習データの多様性を増加させ，走行データも含めた 10 種類のデータ集合 (歩行:08_02.amc ∼ 08_08.amc, 走行:35_20.amc ∼ 35_22.amc) を学習データ，3 種類のデータ集合 (08_09.amc, 35_18.amc, 08_11.amc) を順に並べたものをテストデータとして用意した．スタイルパラメータの次元は，特異値の分布から $J = 4$ と設定した．実験結果を図 4 に示す．実験では，各観測毎に，スタイルパラメータ及び状態変数の逐次的な更新を行い，さらに，各時刻から 0.5 sec 間の観測データを予測し，その予測誤差を解析することでオンライン予測の有効性を検証した．図 4 より，どの動作に対しても 0.2 sec 間程度の観測を得ることでデータのスタイルや状態を推定し，高い予測精度を実現していることがわかる．さらに，スタイルパラメータの次元を増加させることによって，動作の多様性を捉えることが可能となることも結果より見て取れる．ここで，dim = 0 は観測データのノルムを表している．提案法のテストデータへの適応性は，生成モデルのシンプルな構造や，状態変数及びスタイルパラメータの低次元性に基づくと考えられる．



(a)



(b)



(c)

図 4: 提案法によるオンライン予測誤差．(a):08_09.amc, (b):35_18.amc, (c):08_11.amc．各観測毎に，スタイル/状態変数の逐次的な更新を行い，0.5 sec 間の観測データに対する予測誤差．

5　おわりに

本稿では，個性を捉える全身周期運動のオンライン予測を可能にする潜在変数モデルとその適応・予測アルゴリズムについて提案した．さらに，提案法によりモーションキャプチャデータに対して，高精度な予測が行えることを示し，その有効性を実証した．今後は従来法(例えば [19, 15]) との定量的な比較を行う．また，非周期運動への拡張を目指す．

参考文献

[1] Okan Arikan, David A. Forsyth, and James F. O'Brien. Motion synthesis from annotations. *ACM Transactions on Graphics*, 22(3):402–408, 2003.

[2] Christopher M. Bishop. *Neural Networks for Pattern Recognition.* Oxford University Press, 1995.

[3] Matthew Brand and Aaron Hertzmann. Style machines. In *Proceedings of the 2000 SIGGRAPH*, pages 183–192, 2000.

[4] Ahmed Elgammal. Learning to track: conceptual manifold map for closed-form tracking. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 724–730, 2005.

[5] Ahmed Elgammal and Chan-Su Lee. Separating style and content on a nonlinear manifold. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 478–485, 2004.

[6] Eugene Hsu, Kari Pulli, and Jovan Popovic. Style translation for human motion. *ACM Transactions on Graphics*, 24(3):1082–1089, 2005.

[7] Tetsunari Inamura, Iwaki Toshima, and Nakamura Yoshihiko. Acquisition and embodiment of motion elements in closed mimesis loop. In *IEEE International Conference on Robotics and Automation*, pages 1539–1544, 2002.

[8] Jonathan Ko and Dieter Fox. Gp-bayesfilters: Bayesian filtering using gaussian process prediction and observation models. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3471–3476, 2008.

[9] Neil Lawrence. Probabilistic non-linear principal component analysis with gaussian process latent variable models. *The Journal of Machine Learning Research*, 6:1783 – 1816, 2005.

[10] Chan-Su Lee and Ahmed Elgammal. Nonlinear shape and appearance models for facial expression analysis and synthesis. In *Proceedings of the 18th International Conference on Pattern Recognition*, pages 497–502, 2006.

[11] Yasushi Makihara, Ryusuke Sagawa, Yasuhiro Mukaigawa, Tomio Echigo, and Yasushi Yagi. Gait recognition using a view transformation model in the frequency domain. In *Proceedings of the 9th European Conf. on Computer Vision*, pages 151–163, 2006.

[12] Vladimir Pavlovic, James M. Rehg, and John MacCormick. Learning switching linear models of human motion. In *Advances in Neural Information Processing Systems*, pages 981–987, 2000.

[13] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning.* MIT Press, 2006.

[14] Masaaki Sato and Shin Ishii. On-line em algorithm for the normalized gaussian network. *Neural Computation*, 12(2):407–432, 2000.

[15] Graham W. Taylor and Geoffrey E. Hinton. Factored conditional restricted bolzmann machines for modeling motion style. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 1025–1032, 2009.

[16] Joshua B. Tenenbaum and Williams T. Freeman. Separating style and content with bilinear models. *Neural Computation*, 12:1247–1283, 2000.

[17] Sebastian Thrun, Wolfram Burgard, and Dieter Fox. *Probabilistic Robotics.* MIT Press, 2005.

[18] Lorenzo Torresani, Peggy Hackney, and Christoph Bregler. Learning motion style synthesis from perceptual observations. In *Proceeding of the Advances in Neural Information Processing Systems*, pages 1393–1400, 2006.

[19] Jack M. Wang, David J Fleet, and Aaron Hertzmann. Multifactor gaussian process models for style-content separation. In *Proceedings of the 24th Annual International Conference on Machine Learning*, pages 975–982, 2007.

[20] Jack M. Wang, David J. Fleet, and Aaron Hertzmann. Gaussian process dynamical models for human motion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(2):283–298, 2008.

[21] Andrew D. Wilson and Aaron F. Bobick. Parametric hidden markov models for gesture recognition. *IEEE Transaction on pattern analysis and machine intelligence*, 21(9):884–900, 1999.

A　導出された EM アルゴリズム

E ステップ:

$$
\begin{aligned}
\mathbf{x}_{t|t-1} &= \mathbf{A}\hat{\mathbf{x}}_{t-1} && (17)\\
\Sigma_{t|t-1} &= \mathbf{A}\hat{\Sigma}_{t-1}\mathbf{A}^T + \mathbf{Q} && (18)\\
\mathbf{H}_{t|t-1} &= \frac{\partial \boldsymbol{\mu}(\mathbf{x})\hat{\mathbf{w}}_{t-1}}{\partial \mathbf{x}}|_{\mathbf{x}_{t|t-1}} && (19)\\
\mathbf{K}_t &= \Sigma_{t|t-1}\mathbf{H}_{t|t-1}^T \cdot \\
& \quad \left(\mathbf{H}_{t|t-1}\Sigma_{t|t-1}\mathbf{H}_{t|t-1}^T + \mathbf{R}\right) && (20)\\
& && (21)\\
\hat{\mathbf{x}}_t &= \mathbf{x}_{t|t-1} + \mathbf{K}_t\left(\mathbf{y}_t - \boldsymbol{\mu}(\mathbf{x}_{t|t-1})\mathbf{w}_{t-1}\right) && (22)\\
\hat{\Sigma}_t &= \left(\mathbf{I} - \mathbf{K}_t\mathbf{H}_{t|t-1}\right)\Sigma_{t|t-1} && (23)
\end{aligned}
$$

M ステップ:

$$
\hat{\mathbf{w}}_t = \langle \boldsymbol{\mu}^T\boldsymbol{\mu}\rangle_t^{-1}\langle \boldsymbol{\mu}^T\mathbf{y}\rangle_t \qquad (24)
$$

オンライン M ステップ:

$$
\begin{aligned}
\hat{\mathbf{w}}_t &= \ll \boldsymbol{\mu}^T\boldsymbol{\mu} \gg_t^{-1} \ll \boldsymbol{\mu}^T\mathbf{y} \gg_t && (25)\\
\ll \boldsymbol{\mu}^T\boldsymbol{\mu} \gg_t &= (1-\eta_t) \ll \boldsymbol{\mu}^T\boldsymbol{\mu} \gg_{t-1} \\
&+ \eta_t\boldsymbol{\mu}(\hat{\mathbf{x}}_t)^T\boldsymbol{\mu}(\hat{\mathbf{x}}_t) && (26)\\
\ll \boldsymbol{\mu}^T\mathbf{y} \gg_t &= (1-\eta_t) \ll \boldsymbol{\mu}^T\mathbf{y} \gg_{t-1} \\
&+ \eta_t\boldsymbol{\mu}(\hat{\mathbf{x}}_t)^T\mathbf{y}_t && (27)\\
\eta_t &= \left\{1 + \frac{\lambda_t}{\eta_{t-1}}\right\}^{-1} && (28)
\end{aligned}
$$

ここで，

$$
\begin{aligned}
\boldsymbol{\mu}(\mathbf{x}_{t|t-1}) &= [\mu^1(\mathbf{x}_{t|t-1})\cdots\mu^J(\mathbf{x}_{t|t-1})] && (29)\\
\boldsymbol{\mu}(\hat{\mathbf{x}}_t) &= [\mu^1(\hat{\mathbf{x}}_t)\cdots\mu^J(\hat{\mathbf{x}}_t)] && (30)\\
\langle\cdot\rangle_T &= \frac{1}{T}\sum_{t=1}^{T}(\cdot_t) && (31)\\
\ll \cdot \gg_T &= \frac{1}{T}\sum_{t=1}^{T}(\cdot_t) && (32)
\end{aligned}
$$

である．