

Motion Comparison for Dance Skill Improvement via Sequential Unsupervised Segmentation

Ken Kajiwar^{*}, Issei Saito^{*}, Tomoaki Nakamura^{*}, Daichi Mochihashi[†], and Koki Mimura[‡]

^{*}The University of Electro-Communications, Tokyo, Japan

[†]The Institute of Statistical Mathematics, Tokyo, Japan

[‡]National Center of Neurology and Psychiatry, Tokyo, Japan

E-mail: k2532028@edu.cc.uec.ac.jp

Abstract—本研究の目的は、ダンスやスポーツなど熟練度の異なる複数人の動作を動作単位で定量的・視覚的に比較可能な支援システムを構築することである。従来の Gaussian Process Hidden Semi-Markov Model (GP-HSMM) は、高い表現力を持ち動作分節化に有効であるが、全てのデータを一括で学習して扱う必要があるため、新たな被験者の動作を柔軟に取り入れることが難しく、複数人の動作を比較するのも困難であった。本研究では、この課題に対処するために、ガウス過程のカーネル関数を Random Fourier Features (RFF) によって近似し、その近似精度をマルコフ連鎖モンテカルロ法 (MCMC) により統計的に最適化する手法を提案する。さらに、逐次学習を導入することで、初期モデルで学習した熟練者の動作をベースとしつつ、新たな被験者の動作を段階的に学習可能とし、各動作単位で比較することが可能となる。評価実験では、体操およびダンスの公開モーションキャプチャデータセットを用いて、提案手法が従来手法と同等の分節化精度を維持しながら、従来の手法ではおよそ 87 時間かかった処理がおよそ 3 分に大幅に短縮でき、複数被験者の動作の違いを関節単位で可視化・定量化できる支援システムとして有効であることを示した。

Index Terms—computational efficiency, Gaussian process, motion analysis, random Fourier features, time-series modeling, unsupervised segmentation, sequential learning, motion comparison, Dance skill improvement, MCMC optimization.

I. INTRODUCTION

近年、モーションキャプチャ技術の発展により、人間の複雑な動作を高精度かつ連続的に記録・解析することが可能になってきた [1]–[6]。ダンスやスポーツなどの領域では、従来、動作指導が指導者の主観的な経験や感覚に基づいて行われてきたが、動作を定量的に評価する手法の需要が高まっている。

Gaussian Process Hidden Semi-Markov Mode (GP-HSMM) [7] は、動作の時系列データを意味のある単位に自動的に分割し、各セグメントをガウス過程により柔軟にモデル化できる点で有効なアプローチである。特に、動作の分節化と分析を一貫して処理できる点が有用である。しかしながら、GP-HSMM はその柔軟性と引き換えに、ガウス過程の計算コストが高く、データ規模が増えるにつれて実用的な応用が困難になるという問題を抱えている。

この問題に対処するため、Random Fourier Features (RFF) [8] を用いてカーネル関数を近似し、ガウス過程の計算負荷を大幅に軽減する手法が有効である。RFF を用いることで、計算コストは劇的に削減され、動作解析の可能性が広がった。一方で、RFF は近似的手法であるがゆえに、性能が不安定であり、特に高い精度と安定性が求められる動作分節化タスクにおいては信頼性を損なう要因となる。

本研究では、RFF によるカーネル近似にマルコフ連鎖モンテカルロ法 (MCMC) [9] を導入することにより、近似性能のばらつきを統計的手法により安定化させ、RFF-GP-HSMM の精度の向上を実現する。また、従来の GP-HSMM は、観測点の増加に応じて潜在状態集合と各状態のガウス過程共分散行列が高次元化する非パラメトリック・ベイズモデルであるため、厳密なベイズ推論にはすべての観測データを一括で学習する必要があるが、新たな被験者の動作を柔軟に取り入れることが難しく、複数被験者間の動作比較も困難であった。そこで本研究では、逐次学習 [10] を導入する。これにより、初期データで得た知見を活かしつつ、新たな被験者の動作を段階的に学習することで、熟練度の違う被験者の動作の特徴を、各動作セグメントに基づいて定量的に比較・評価することが可能となる。

体操およびダンスの公開モーションキャプチャデータセットを用いた評価実験により、提案手法の有効性を検証した。体操データセットにおいては、RFF に MCMC を導入することでカーネル近似の精度が向上し、正解ラベルとの正規化ハミング距離の平均値および標準偏差が、MCMC を導入していない RFF-GP-HSMM よりも改善され、GP-HSMM と同等の分節精度を達成した。また、288 万点のデータを含むダンスデータセットに対しては、従来の GP-HSMM では解析におよそ 87 時間を要したのに対し、提案手法ではおよそ 3 分で解析が完了し、実用的な処理時間で解析可能となった。さらに、逐次学習の導入により、熟練度の異なる 2 名のダンサーの動作を振り付け単位で定量的に比較・評価することが可能となり、改善すべき関節動作を可視化する比較支援システムを構築した。

本研究では、RFF によるカーネル近似に MCMC を統合することで高速かつ安定した動作分節化を実現し、さらに逐次学習を導入することで、複数被験者間の動作を柔軟に比較・評価できる新たなフレームワークを提案する。本手法は、従来法では困難であった動作データへの適用と個人差に基づく動作解析の両立を可能にし、実用性と拡張性を兼ね備えた動作理解のアプローチを示すものである。

II. RELATED WORK

近年、モーションキャプチャ技術の進展により、人間の動作を高精度に解析する研究が活発に行われている [1]–[6]。たとえば、Balazsia と Sojka [1] は、CMU MoCap データベースを用いて、31 関節から得られる 3D 時系列データに対し、LDA と Maximum Margin Criterion (MMC) を組み合わせた特徴抽出を行い、歩行者識別において従来手法を上回る

精度を達成した。これは、モーションキャプチャを用いた単純な動作の識別に関する有効な手法の一例である。

また、ダンスのような複雑な動作を対象とした研究も進展している。DanceMVP [6] では、69,300 秒に及ぶダンス動作と音楽を含むデータセット ImperialDance を開発し、5つのジャンル、20 曲の音楽、20 種類の振り付けから構成される大規模マルチモーダルデータを収録している。このデータセットは、語彙認識、パフォーマンススコアリング、リズム整合性評価といった複数タスクを統合的に扱う自己教師あり学習フレームワークのために設計されており、異なる熟練度のダンサーによる繰り返し動作を通じて、技能の違いを捉えるためのものとなっている。

しかしながら、この研究では、動作全体をスコアリングや分類することに主眼が置かれており、分類においては専門家による手動のアノテーション作業が必要不可欠であるため、一般のユーザーや現場レベルで広く受け入れられるような動作解析手法としては不十分である。そのため、一般的に受け入れやすく、かつ自動的に処理可能な解析手法として、振り付けなどの個々の動作単位ごとに習熟度の違いを定量的に分析可能な効率の手法の開発が求められている。

この課題に対して、教師なし分割 [11]–[16] は有望なアプローチとして注目されており、繰り返し現れるパターンを自動的に抽出し、複雑な時系列データの構造解析を可能にする。産業応用を含む様々な実世界のタスクにおいて重要な役割を果たす一方、従来の手法では、セグメントの開始・終了点とカテゴリを同時に推定する必要があるため、組み合わせ爆発による計算コストが問題となり、その回避のために導入されたヒューリスティックな仮定は、タスク依存性が高いという課題を抱えている。

このような背景のもと本研究では、この課題に対処するために Gaussian Process Hidden Semi-Markov Model (GP-HSMM) [7] に着目する。このモデルは、ガウス過程を非パラメトリックモデルとして用いて、繰り返し現れる複雑なパターンを捉え、HSMM (隠れセミマルコフモデル) によってそれらのパターンの持続時間をモデル化する。確率的生成モデルに基づくこの手法は、従来手法よりも複雑な時系列データの高精度なセグメンテーションを可能にし、かつ高い解釈性を有する。実際に、GP-HSMM はその解釈性とセグメンテーション精度のバランスの良さから、多くの研究で活用されている。たとえば、Mimura らは GP-HSMM に階層的ディリクレ過程を組み込むことで、マーマセットの特徴的な行動を教師なしで抽出できることを示した [17]。また、製造現場では作業者の行動解析に応用され、作業パターンだけでなく、オペレーターごとの特性や熟練度の違いも明らかにされている [18]。ロボット学習の分野においては、連続的なタスクを原始的な動作に分割する目的で本手法が利用されている [19]。

高精度な教師なしセグメンテーション手法として有効である一方で、GP-HSMM はガウス過程に起因する計算負荷の高さという課題を抱えている。

この課題に対する有望な解決策として、カーネルを Random Fourier Features (RFF) によって近似する基底関数の導出がある [8]。RFF は、スケーラブルな GPLVM (Gaussian Process Latent Variable Model) [20] [21] を可能にし、精度を損なうことなくハイブリッド HMM の高速化を実現してきた [22]。これらの結果に動機づけられ、我々は GP-HSMM に RFF を

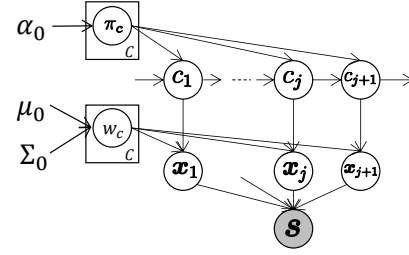


Fig. 1. RFF-GP-HSMM のグラフィカルモデル。

導入した。ただし、RFF は近似手法であるため、性能が不安定であるという問題がある。

そこでさらに、本研究では Metropolis-Hastings 法を用いた MCMC (Markov Chain Monte Carlo) 法 [9] を導入し、RFF で使用される周波数ベクトルおよび位相パラメータを最適化することで、近似のばらつきを抑え、より安定した性能を実現した。

加えて、ベイズ線形回帰における事後分布の逐次更新則に基づき、逐次学習 [10] を導入することで、一度学習した熟練者モデルを保持したまま、新たな中級者のデータを段階的に統合し、複数人の動作比較・評価を効率的に行うことが可能となった。

III. METHOD

A. Generative process

Fig. 1 に提案する RFF-GP-HSMM のグラフィカルモデルを示す。本モデルでは、時系列データ S が次の生成過程を経て得られると仮定する。

まず、遷移確率 π_c をディリクレ分布 (パラメータ α_0) から生成する。

$$\pi_c \sim P(\pi | \alpha_0). \quad (1)$$

次に、第 j セグメントのクラス c_j を、直前のクラス c_{j-1} と遷移確率 π_c に基づいて生成する。

$$c_j \sim P(c | c_{j-1}, \pi_c). \quad (2)$$

従来の GP-HSMM では、各セグメント x_j はガウス過程によって生成されると仮定する。一方、RFF-GP-HSMM では、各セグメントを線形回帰で生成する。回帰重み w_c は平均 μ_0 、共分散行列 Σ_0 のガウス分布から生成する。

$$w_c \sim \mathcal{N}(w | \mu_0, \Sigma_0). \quad (3)$$

クラス c_j に対応するセグメント x_j は、重み w_{c_j} を用いた線形回帰によって生成される。

$$x_j \sim p(x | w_{c_j}). \quad (4)$$

観測された時系列データ S は、これらのセグメント x_j を連結することで得られる。ただし、単純な線形回帰だけでは複雑な時系列パターンを十分に捉えられない。そこで RFF-GP-HSMM では、Random Fourier Features (RFF) により導出される基底関数を用いた線形回帰 [8] を採用し、モデル表現力を高めている。

B. Gaussian Processes

GP-HSMM では、各クラスの時刻 t' の出力値 x' の予測にガウス過程を用いることで複雑に変化する時系列データを表現している。ガウス過程は、 ϕ を基底関数とした線形回帰モデル $x' = \mathbf{w}^T \phi(t')$ に対して、重み \mathbf{w} の周辺化とカーネルトリックを適用した確率モデルである。ガウス過程では、クラス c に分類された分節の集合 \mathbf{X}_c と、対応するタイムステップの集合 t_c が与えられたとき、時刻 t' の出力 x' の予測分布は次式のガウス分布となる。

$$p(x'|t', \mathbf{X}_c, t_c) = \mathcal{N}(\mathbf{k}^T \mathbf{K}^{-1} \mathbf{X}_c, k(\hat{t}', \hat{t}') - \mathbf{k}^T \mathbf{K}^{-1} \mathbf{k}), \quad (5)$$

$k(\cdot, \cdot)$ はカーネル関数であり、 \mathbf{K} は q 行 p 列の要素が次式となる行列である。

$$K(p, q) = k(t_p, t_q) + \beta^{-1} \delta_{pq}, \quad (6)$$

t_p と t_q はそれぞれ、 t_c の p 番目の要素と q 番目の要素であり、 β と δ_{pq} はそれぞれノイズ成分を表すパラメータとデルタ関数である。 \mathbf{k} は、 p 番目の要素が $k(t_p, t')$ となるベクトルである。本稿では、カーネル関数として次式の RBF カーネルを用いる。

$$k(t_p, t_q) = \exp(-\frac{1}{2} \|t_p - t_q\|^2), \quad (7)$$

カーネルを用いることで、基底関数 $\phi(\cdot)$ を直接計算せずに、その内積である $k(t_p, t_q) = \phi(t_p)^T \phi(t_q)$ を計算することで、非線形高次元空間で柔軟に入出力の関係を捉えることができる。

GP-HSMM では、出力が多次元ベクトル \mathbf{x}' の場合は、各次元が独立に生成されると仮定し、その生成確率を次式のように計算する。

$$GP(\mathbf{x}'|t', \mathbf{X}_c, t) = \prod_d p(\mathbf{x}'^{(d)}|t', \mathbf{X}_c^{(d)}). \quad (8)$$

ただし、上付きの添字 d は、それぞれの変数の d 次元目を表している。

ガウス過程では、式 (5) の計算に、 \mathbf{K} の逆行列演算を含んでいる。 \mathbf{X}_c に含まれるデータ点数を N_c とすると、 \mathbf{K} のサイズは $N_c \times N_c$ となる。すなわち、処理する時系列データ点の数 N_c が多くなれば行列 \mathbf{K} のサイズも大きくなり、その逆行列の演算に多くの計算時間が必要になる。これが現状の GP-HSMM の計算時間のボトルネックとなっている。

C. Gaussian process approximation using RFF

ガウス過程では、カーネルトリックにより、基底関数 ϕ を直接計算することなく非線形高次元空間で入出力の対応関係を捉えている。一方で、もし $k(t_p, t_q) \approx \phi(t_p)^T \phi(t_q)$ となる、内積をとると RBF カーネルの近似になるような基底関数 ϕ を導出できれば、 \mathbf{K} を計算する必要のない線形回帰モデル $x' = \mathbf{w}^T \phi(t')$ を適用することができる。このような、カーネルからそのカーネルを近似可能な基底関数を求める方法が Random Fourier Feature (RFF) [8] である。

RFF では、まずカーネル関数のフーリエ変換を考える。式 (7) の RBF カーネルを $\Delta = t_p - t_q$ を入力とする関数

$k'(\Delta) = \exp(-\|\Delta\|^2/2)$ とし、 $k'(\Delta)$ をフーリエ変換すると次式のガウス分布となる。

$$p(\omega) = (2\pi)^{-\frac{D}{2}} \exp(-\frac{1}{2} \|\omega\|^2) \quad (9)$$

次に、この $p(\omega)$ をフーリエ逆変換し、もとのカーネル関数を復元する。

$$k(t_p, t_q) = k'(\Delta) \quad (10)$$

$$= \int_{\mathbb{R}} p(\omega) \exp(i\omega(t_p - t_q)) d\omega \quad (11)$$

$$= \int_{\mathbb{R}} p(\omega) \cos(\omega(t_p - t_q)) + i \sin(\omega(t_p - t_q)) d\omega \quad (12)$$

この式変形にはオイラーの公式 $\exp(iz) = \cos(z) + i \sin(z)$ を用いた。さらに、カーネル関数の出力値は実数となるため

$$k(t_p, t_q) = \int_{\mathbb{R}} p(\omega) \cos(\omega(t_p - t_q)) d\omega \quad (13)$$

となる。次に、 $b \sim \text{Uniform}(0, 2\pi)$ となる乱数 ($p(b) = \frac{1}{2\pi}$) を導入し、 b での周辺化を考えると

$$\int_0^{2\pi} p(b) p(\omega) \cos(\omega(t_p - t_q)) db = p(\omega) \cos(\omega(t_p - t_q)) \quad (14)$$

$$\int_0^{2\pi} p(b) p(\omega) \cos(\omega(t_p + t_q) + 2b) db \quad (15)$$

$$= \left[\frac{1}{4\pi} p(\omega) \sin(\omega(t_p + t_q) + 2b) \right]_0^{2\pi} = 0 \quad (16)$$

が成り立つため、 $k(t_p, t_q)$ は次式のように変形することができる。

$$k(t_p, t_q) = \int_{\mathbb{R}} \int_0^{2\pi} p(b) p(\omega) \{ \cos(\omega(t_p + t_q) + 2b) + \cos(\omega(t_p - t_q)) \} db d\omega \quad (17)$$

ここで、 $A = \omega t_p + b$, $B = \omega t_q + b$ とすると、

$$\cos(A + B) + \cos(A - B) = 2 \cos(A) \cos(B) \quad (18)$$

となり、次式のように変形することができる。

$$k(t_p, t_q) = \int_{\mathbb{R}} \int_0^{2\pi} p(b) p(\omega) \sqrt{2} \cos(\omega t_p + b) \times \sqrt{2} \cos(\omega t_q + b) db d\omega \quad (19)$$

この積分を M 個のサンプル $\omega^{(m)}, b^{(m)}$ でモンテカルロ近似する。

$$\omega^{(m)} \sim p(\omega), \quad (20)$$

$$b^{(m)} \sim \text{Uniform}(0, 2\pi), \quad (21)$$

$$k(t_p, t_q) = \frac{1}{M} \sum_m \sqrt{2} \cos(\omega^{(m)} t_p + b^{(m)}) \times \sqrt{2} \cos(\omega^{(m)} t_q + b^{(m)}) \quad (22)$$

式 (22) は、次式の基底関数で t_p, t_q を M 次元空間に写像したベクトル同士の内積 $k(t_p, t_q) = \phi(t_p)^T \phi(t_q)$ とみなすことができる。

$$\phi(t) = \begin{bmatrix} \sqrt{\frac{2}{M}} \cos(\omega^{(1)}t + b^{(1)}) \\ \sqrt{\frac{2}{M}} \cos(\omega^{(2)}t + b^{(2)}) \\ \vdots \\ \sqrt{\frac{2}{M}} \cos(\omega^{(M)}t + b^{(M)}) \end{bmatrix} \quad (23)$$

すなわち、この $\phi(t)$ はガウスカーネルを近似する基底関数となっている。

D. MCMC-Optimized RFF Kernel

RFF によって近似された RBF カーネルは、次式で示すように定義する。

$$\hat{k}(t_p, t_q) = \frac{1}{M} \sum_{m=1}^M \sqrt{2} \cos(\omega^{(m)}t_p + b^{(m)}) \times \sqrt{2} \cos(\omega^{(m)}t_q + b^{(m)}) \quad (24)$$

ここで、式 (24) における周波数ベクトル ω と位相ベクトル b を、Metropolis-Hastings (MH) アルゴリズムに基づくマルコフ連鎖モンテカルロ (MCMC) 法 [9] によって最適化する。

まず、近似誤差を測るエネルギー関数を次式で定義する。

$$E(\omega, b) = \sum_{p=1}^{N_c} (k(t_p, 0) - \hat{k}(t_p, 0))^2 \quad (25)$$

このエネルギーを指数写像することで、 (ω, b) 上の目標分布を次式のように導出できる。

$$\begin{aligned} p(\omega, b) &\propto \exp[-E(\omega, b)] \\ &= \prod_{p=1}^{N_c} \exp[-(k(t_p, 0) - \hat{k}(t_p, 0))^2] \\ &\propto \prod_{p=1}^{N_c} \frac{1}{\sqrt{2\pi}} \exp[-(k(t_p, 0) - \hat{k}(t_p, 0))^2] \\ &= \prod_{p=1}^{N_c} \mathcal{N}(k(t_p, 0) | \hat{k}(t_p, 0), 1) \end{aligned} \quad (26)$$

MH 法では、現在値 (ω, b) から多変量正規分布に従う提案点 (ω', b') を生成し、次式の受理確率で遷移を決定する。

$$A = \min\left\{1, \frac{p(\omega', b')}{p(\omega, b)}\right\} \quad (27)$$

エネルギーが低下する候補 (比が 1 を超える場合) は必ず受理され、高エネルギー側への遷移は確率 A で受理される。

最終的に得られた最適化パラメータ ω^*, b^* を用いて構築される近似 RBF カーネルは、次式で示すとおりである。

$$\hat{k}^*(t_p, t_q) = \frac{1}{M} \sum_{m=1}^M \sqrt{2} \cos(\omega^{*(m)}t_p + b^{*(m)}) \times \sqrt{2} \cos(\omega^{*(m)}t_q + b^{*(m)}) \quad (28)$$

ここで、 \hat{k}^* は MCMC によって最適化された RFF 近似カーネルであり、学習全体を通して利用される。

E. Sequential Learning of RFF Posterior Computation

RFF を利用することで、RBF カーネルを構成可能な基底関数 $\phi(t) \in \mathbb{R}^M$ が明示的に得られるため、この基底関数を用いた線形回帰で式 (5) のガウス過程回帰を近似しつつ、データセット単位で順次モデルを更新する。クラス c に属する r 番目のデータセット $\mathbf{X}_c^{(r)}$ と対応するタイムステップ $t_c^{(r)}$ が与えられているとき、時刻 t' の出力 x' の予測分布は

$$p(x' | t', \mathbf{X}_c^{(r)}, t_c^{(r)}) = \mathcal{N}(x' | \mu^{(r)}(t'), \sigma^{2(r)}(t')), \quad (29)$$

$$\mu^{(r)}(t') = (\mathbf{m}^{(r)})^\top \phi(t'), \quad (30)$$

$$\sigma^{2(r)}(t') = \beta^{-1} + \phi(t')^\top \mathbf{S}^{(r)} \phi(t'), \quad (31)$$

ただし、重みの事前分布は $\mathcal{N}(\mathbf{w} | \mathbf{m}^{(0)}, \mathbf{S}^{(0)})$, $\mathbf{m}^{(0)} = \mathbf{0}$, $\mathbf{S}^{(0)} = \alpha^{-1} \mathbf{I}$ とし、 β は観測ノイズの精度である。

式 (5) のガウス過程の予測分布では、 $N_c \times N_c$ 行列 \mathbf{K} の逆行列計算が必要になるのに対し、RFF を用いる場合は式 (30) に含まれる $M \times M$ 行列の逆行列だけで済む。 N_c はクラス c に含まれるデータ点数であり、データ規模が大きくなるほど \mathbf{K}^{-1} の計算負荷は増大するが、 M は事前に固定されているため、RFF を用いた線形回帰の逆行列計算はデータ点数に依存せず一定時間で実行できる。

新たに $r+1$ 番目のデータセット $\mathbf{X}_c^{(r+1)}$ とタイムステップ $t_c^{(r+1)}$ が追加されたとき、事後分布は

$$\mathbf{S}^{(r+1)} = \left\{ (\mathbf{S}^{(r)})^{-1} + \beta \sum_{p=1}^{N_c} \phi(t_p) \phi(t_p)^\top \right\}^{-1}, \quad (32)$$

$$\mathbf{m}^{(r+1)} = \mathbf{S}^{(r+1)} \left\{ (\mathbf{S}^{(r)})^{-1} \mathbf{m}^{(r)} + \beta \sum_{p=1}^{N_c} x_p \phi(t_p) \right\}, \quad (33)$$

と更新される。初めのデータセット ($r = 0$) では

$$\mathbf{S}^{(1)} = \left(\alpha \mathbf{I} + \beta \sum_{p=1}^{N_c} \phi(t_p) \phi(t_p)^\top \right)^{-1}, \quad (34)$$

$$\mathbf{m}^{(1)} = \beta \mathbf{S}^{(1)} \sum_{p=1}^{N_c} x_p \phi(t_p). \quad (35)$$

この逐次学習 [10] により、一度取得したモデルデータセットの知見を失うことなく、追加データセットを効率的に取り込みながらモデルを適応的に更新できる。

F. Parameter inference

1) *Blocked Gibbs sampler*: 観測系列におけるセグメントとそれに対応するクラスは、Blocked Gibbs sampler を用いて推論された。Blocked Gibbs sampler の推論手順は以下の通りである。

- 1) すべての観測系列 $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n, \dots$ をランダムに分割し、モデルパラメータを初期化する。
- 2) n 番目の観測系列 \mathbf{s}_n に対して、セグメント $x_{n,j}$ をそれぞれ割り当てられていたクラスから除去し、各クラスの線形回帰パラメータ (Σ_c, \mathbf{m}_c) と遷移確率 $P(c | c')$ を更新する。

- 3) 後述の forward filtering – backward sampling により、新たなセグメント x_{nj} ($j = 1, 2, \dots, J$) および対応するクラス c_{nj} ($j = 1, 2, \dots, J$) をサンプリングし、線形回帰パラメータ Σ_c, m_c と遷移確率 $P(c | c')$ を更新する。
- 4) すべての系列に対してステップ 2 と 3 を繰り返し、収束するまで実行する。

この手順により、RFF-GP-HSMM のパラメータが最適化され、観測系列が分節化される。

2) *Forward filtering-backward sampling*: セグメントとそれに対応するクラスのサンプリングには、forward filtering – backward sampling (FFBS) が用いられた。FFBS は、動的計画法を用いてすべての可能な分節化およびクラス分類パターンの確率を効率的に計算し、それらの確率に基づいてセグメントおよびクラスをサンプリングする手法である。forward filtering ステップでは、観測系列 s における時刻 t で終わる長さ k の部分系列がクラス c に属するセグメントである確率を表す α を計算する。

$$\alpha[t][k][c] = \left(\prod_{t'=t-k}^t p(s_{n,t'} | t', \mathbf{X}_c, t_c) \right) p(k | \lambda) \times \sum_{k'=K_{\min}}^{K_{\max}} \sum_{c'=1}^C p(c | c') \alpha[t-k][k'][c'], \quad (36)$$

ここで、 $s_{n,t'}$ は系列 s_n における時刻 t' のデータ点を表す。 C はセグメントクラスの数であり、 K_{\min} および K_{\max} は許容される最小・最大のセグメント長をそれぞれ表す。項 $p(k | \lambda)$ は、平均セグメント長 λ を持つポアソン分布でモデル化された、セグメント長が k である確率である。

$$P(k | \lambda) = \frac{\lambda^k e^{-\lambda}}{k!}. \quad (37)$$

$p(c | c')$ はクラス間の遷移確率を表し、次のディリクレ多項分布モデルで与えられる。

$$p(c | c') = \frac{N_{c'c} + \alpha}{N_{c'} + C\alpha}, \quad (38)$$

ここで、 $N_{c'}$ はクラス c' に割り当てられたセグメント数を、 $N_{c'c}$ はクラス c' からクラス c への遷移回数を表す。 k' および c' は、 $s_{t-k:t}$ に先行するセグメントの長さおよびクラスを表す。これらの値はこの段階では未知であるため、式 (36) はすべての可能な長さとして周辺化を行っている。

次に、確率 α に基づいて backward sampling を行う。backward sampling では、時系列データの末尾 ($t = T$) から始めて、クラスとセグメント長を逆順にサンプリングする。

$$k, c \sim p(c_{j-1} | c) \alpha[t][k][c]. \quad (39)$$

この手順により、観測系列におけるセグメントおよびそれに対応するクラスを決定することができる。このアルゴリズム全体の手順は Algorithm 1 にまとめられている。

Algorithm 1 Forward filtering-backward sampling.

```

1: // Forward filtering
2: for  $t = 1$  to  $T$  do
3:   for  $k = 1$  to  $K$  do
4:     for  $c = 1$  to  $C$  do
5:       Compute  $\alpha[t][k][c]$ 
6:     end for
7:   end for
8: end for
9:
10: // Backward sampling
11:  $t = T, j = 1$ 
12: while  $t > 0$  do
13:    $k, c \sim p(c_{j+1} | c) \alpha[t][k][c]$ 
14:    $x_j = s_{t-k:t}$ 
15:    $c_j = c$ 
16:    $t = t - k$ 
17:    $j = j + 1$ 
18: end while
19: return  $(x_{J_n}, x_{J_n-1}, \dots, x_1), (c_{J_n}, c_{J_n-1}, \dots, c_1)$ 

```

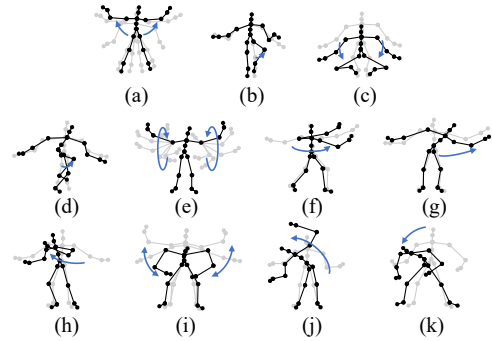


Fig. 2. Motions included in exercise motion: (a) jumping jack, (b) jogging, (c) squatting, (d) knee raise, (e) arm circle, (f) twist, (g) side reach, (h) boxing, (i) arm wave, (j) side bend, and (k) toe touch.

IV. EXPERIMENT

A. Experimented setup

本研究では、2つのモーションキャプチャデータセットを用いて提案手法の有効性を検証した。

1つ目の実験では、Carnegie Mellon University (CMU) モーションキャプチャデータセット¹の Subject 14 に含まれる 3 つの動作系列を用いた。これらの動作には、走行やジャンプなどが含まれており、Fig. 2 に示す。データは 4 fps にダウンサンプリングし、左手・右手・左足・右足の 2 次元位置座標からなる 8 次元データを使用した。各座標は min-max 正規化により、値の範囲を $[-1, 1]$ に正規化した。

本実験では、従来手法である GP-HSMM に加え、ランダムフーリエ特徴 (Random Fourier Features, RFF) を導入した 2 種類の RFF-GP-HSMM モデルを用いて学習を行った。1 つは MCMC を導入していない RFF-GP-HSMM、もう 1 つはカーネル近似の精度向上のために MCMC を導入した RFF-GP-HSMM である。いずれの RFF-GP-HSMM においても、ランダム特徴数は $M = 20$ に設定した。すべての手法において、セグメント長パラメータは $\lambda = 20$ 、最小セグメント長 $K_{\min} = 15$ 、最大セグメント長 $K_{\max} = 30$ とし、ブロック化ギブスサンプラーの繰り返し回数は 5 回とした。

¹<http://mocap.cs.cmu.edu>

TABLE I
COMPUTER USED IN THE EXPERIMENTS.

CPU	Apple M4
Memory	16GB

2つ目の実験では, AAAI 2024 にて提案された DanceMVP [6] に基づく公式ダンスデータセット²を用いた. 本データセットでは, TWICE の楽曲「What is Love?」の 10 秒間の振り付けを踊った動作が記録されており, 21 個の関節に対する 3 次元 (x, y, z) 座標からなる 63 次元の座標データを含む. 本実験ではこの動作データを使用する. 対象関節は, 右膝, 右つま先, 右太もも, 左太もも, 左足, 右肩, 左手, 右手, 左前腕, 左肩, 右足, 首, 左膝, 胸, 頭, 左つま先, 腹, 右上腕, 左上腕, 腰, 右前腕である. すべての関節の座標値から胸の座標を減算し, 胸を原点とする相対座標に変換した. この処理により, 63 次元から胸を基準とした 60 次元の座標データに変換した.

上級者によるダンス動作データとして, 100 fps で記録された 1 回あたり 1,200 フレームのデータの 20 回分を初期モデルデータセットとして学習させた. さらに, 中級者による同様のデータの 20 回分を追加データセットとして, 初期学習で得られた事後分布を事前分布とみなして逐次学習を行った. 本実験で用いた全データは, 2 名の被験者に対して, 各 20 試行, 1 試行あたり 1,200 フレーム, 1 フレームあたり 60 次元の座標データから構成されており, 総計 2,880,000 点の時系列データとなる.

この実験では, MCMC および逐次学習の両方を導入した RFF-GP-HSMM と, 従来の GP-HSMM との比較を行った. RFF-GP-HSMM ではランダム特徴数を $M = 20$ に設定し, 両手法においてセグメント長パラメータは $\lambda = 120$, 最小セグメント長 $K_{\min} = 100$, 最大セグメント長 $K_{\max} = 140$ とし, ブロック化ギブスサンプラーの繰り返し回数は 3 回とした. 両実験において使用した PC の仕様は Table I に示す.

B. Experiment 1: Segmentation Accuracy on the Exercise Dataset

本実験では, 従来手法である GP-HSMM に加え, Random Fourier Features (RFF) を導入した 2 種類の RFF-GP-HSMM モデルを用いて体操動作の分節化を行った. 1 つは MCMC を導入していない RFF-GP-HSMM, もう 1 つはカーネル近似の精度向上を目的として MCMC を導入した RFF-GP-HSMM である.

分節精度の評価には, セクション IV-A で述べた 3 つの体操動作系列を 10 回繰り返すことで得られた合計 30 系列のデータセットを用いた. すべての手法において初期値に対する感度が高いため, 各手法について 10 回の試行を行い, それぞれの試行における分節結果と正解ラベルとの正規化ハミング距離 (Normalized Hamming Distance, NHD) を算出した. NHD は $[0, 1]$ の範囲をとり, 0 に近いほど分節結果と正解ラベルの一致度が高いことを示す.

Table II には, 各手法において 10 回の試行から得られた分節系列と正解ラベルとの NHD の平均値および標準偏差を示す. その結果, MCMC を導入していない RFF-GP-HSMM

TABLE II
NORMALIZED HAMMING DISTANCE (MEAN \pm STANDARD DEVIATION)
FROM 10 TRIALS FOR EACH METHOD.

Method	NHD (Mean \pm SD)
GP-HSMM	0.384 \pm 0.0589
RFF-GP-HSMM	0.445 \pm 0.0759
RFF-MCMC-GP-HSMM	0.408 \pm 0.0427

TABLE III
COMPARISON OF TRAINING TIMES

Trial	GP-HSMM (s)	RFF-MCMC-GP-HSMM (s)
1	2.94×10^5	1.96×10^2
2	3.35×10^5	1.99×10^2
3	3.06×10^5	1.98×10^2
4	3.02×10^5	1.90×10^2
5	3.28×10^5	1.95×10^2
Mean	3.13×10^5	1.95×10^2

は, RFF による近似の導入により高速化を実現しつつも, GP-HSMM に比べてやや精度が劣る結果となった. 一方, MCMC を導入した RFF-GP-HSMM ではカーネル近似の精度が向上し, NHD の平均および標準偏差が改善され, 従来手法である GP-HSMM と同程度の分節精度を達成した.

C. Experiment 2: Efficient Analysis and Comparative Evaluation on the Dance Dataset

本実験では, 288 万点の時系列データを含む大規模なダンスデータセットを用いて, 提案手法と従来手法である GP-HSMM の学習時間を比較した. それぞれの手法について 5 回の試行を行い, 各々の試行で学習時間を計測した. その結果を Table III に示す.

5 回試行の平均値を比較すると, 従来手法の GP-HSMM では学習完了におよそ 87 時間を要したのに対し, 提案手法ではわずか約 3 分で学習が完了し, 実用的な処理時間での解析が可能であることが確認された.

また本実験では, 提案手法における逐次学習の枠組みに基づき, 上級者のダンスデータを初期モデルデータセット $\mathbf{X}_c^{(1)}$ として用い, 式 (30) に基づいて得られる $\mathbf{S}^{(1)}$, $\mathbf{m}^{(1)}$ を初期の事後分布パラメータとした. その後, 中級者のデータ $\mathbf{X}_c^{(2)}$ を追加データセットとして与え, 式 (32), (33) に従って事後分布パラメータ $\mathbf{S}^{(2)}$, $\mathbf{m}^{(2)}$ を更新した. このようにして得られた $\mathbf{m}^{(1)}$, $\mathbf{m}^{(2)}$ はそれぞれ, 上級者および中級者の動作に基づく関節の出力分布の平均ベクトルに対応しており, これらを用いることで各振り付け単位における熟練度の違いを関節ごとに定量的に可視化した.

提案手法により得られた分節結果に基づいて, ダンス動作は自動的に 6 種類の振り付けに分類された. 抽出された 6 つの振り付けの一例を Fig. 3 に示す. さらに, 各振り付けに対して, 上級者および中級者の動作を比較するために, 各関節ごとの 60 次元の $\mathbf{m}^{(1)}$, $\mathbf{m}^{(2)}$ を主成分分析 (PCA) により次元削減し, その結果を振り付けごとに可視化した. 分析結果として, Fig. 3 に示す 6 つの振り付けについて, 上級者の関節ごとの事後分布の平均を青い丸印で示し, 中級者のものをオレンジ色のバツ印で示した. 特に動作の差異が顕著に現れた関節として, 左膝・左手・右前腕の 3 箇所に注目した.

²<https://github.com/YunZhongNikki/ImperialDance-Dataset>

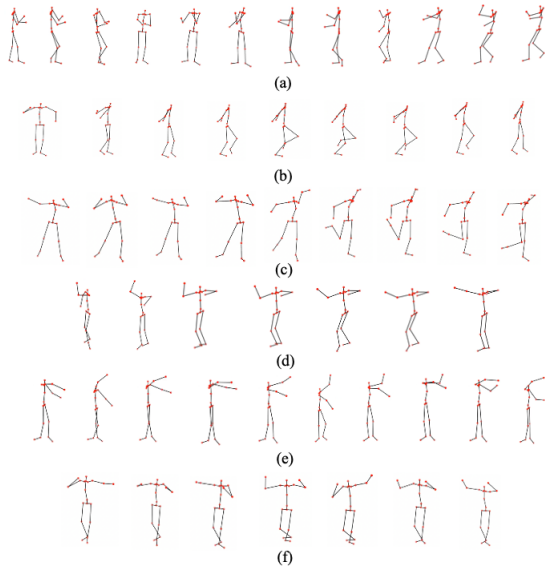


Fig. 3. Choreographies: (a) Choreography 1: Swing both arms to the sides, (b) Choreography 2: Raise and lower the left leg while bending the knee, (c) Choreography 3: Swing the right arm back and forth while lifting the right heel, (d) Choreography 4: Extend the left arm forward, (e) Choreography 6: Swing the arm to the left side, (f) Choreography 7: Stretch both arms to the sides.

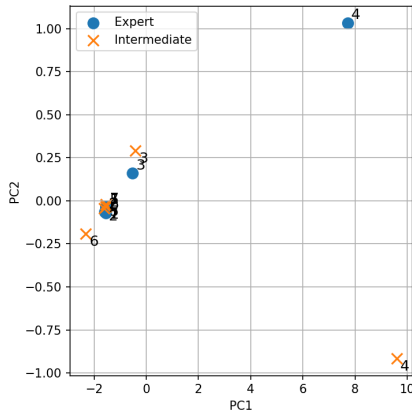


Fig. 4. PCA result of the left knee

まず, Fig. 4 に示す左膝の PCA 分析結果では, 振り付け 4 において, 上級者と中級者の分布が乖離しており, 特に膝を深く曲げる動作に対応する上級者の特徴が明確に現れていた. 次に, Fig. 5 に示す左手の PCA 分析結果では, 振り付け 4 および 5 において, 上級者と中級者の分布に明確な差が見られ, 両者の動作に違いがあることが示唆された. 最後に, Fig. 6 に示す右前腕の PCA 分析結果では, 振り付け 5 において, 上級者と中級者の分布に明瞭な差異が確認された.

これらの PCA 分析結果で示された差異をより具体的に検証するため, 分節化して得られた振り付け 4 および 5 の動作を Fig. 7 (a), (b) にそれぞれ示す. この図では, 各振り付けにおいて上段が上級者, 下段が中級者による動作を示し, 時系列順にスライドとして並べている. また, 解析対象部位を強調するために, 左手は水色の円, 左膝はマゼンタ色の円, 右前腕は黄緑色の円で囲って表示している.

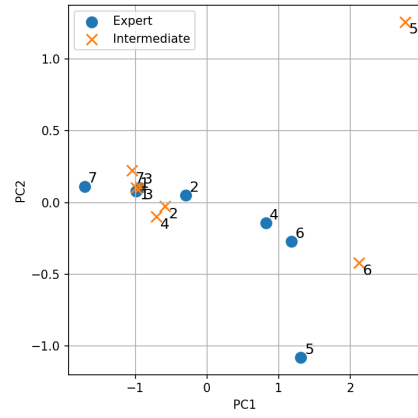


Fig. 5. PCA result of the left hand

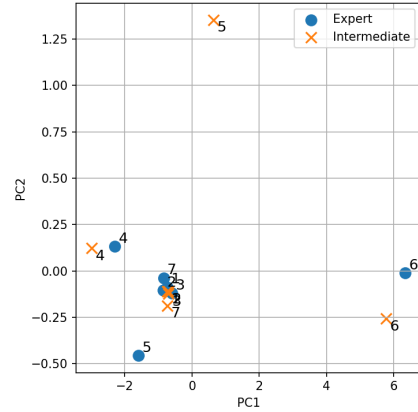


Fig. 6. PCA result of the right forearm

Fig. 7 (a) に示す振り付け 4 の動作スライドを用いて, 各部位ごとに動作の違いを考察する. まず, 左膝に関しては, 上級者がしっかりと膝を曲げているのに対し, 中級者は膝の屈曲がほとんどないという様子が確認された. 次に, 左手の動きに注目すると, 上級者は左手を大きく前方に振り出すことで, 動作にリズム感と力強さを与えているのに対し, 中級者は手の前方への動きが不十分で, 手が徐々に下がっていく傾向が見られた.

Fig. 7 (b) に示す振り付け 5 の動作では, 左手に関して, 上級者は中級者よりも大きく振る動きを行っており, PCA 結果の差異と一致する傾向が確認できた. さらに, 右前腕に関しても, 上級者は右前腕を大きく使ってダイナミックな動作をしているのに対し, 中級者は右前腕の動きが小さく, 動作が抑制されている様子が視覚的に確認できた.

これらの結果から, 逐次学習を導入した提案手法により, 熟練度の異なる複数の被験者の動作を振り付け単位で定量的かつ視覚的に比較可能な支援システムを構築できることが示された. また, 各振り付けにおいて, どの関節の動作を改善すれば上級者に近づけるかを定量的に可視化・評価できることが明らかとなり, 運動指導やフィードバックへの応用可能性が示唆された.

V. CONCLUSION

本研究では, ダンスやスポーツなど熟練度の異なる複数人の動作を振り付け単位で定量的かつ視覚的に比較可能な

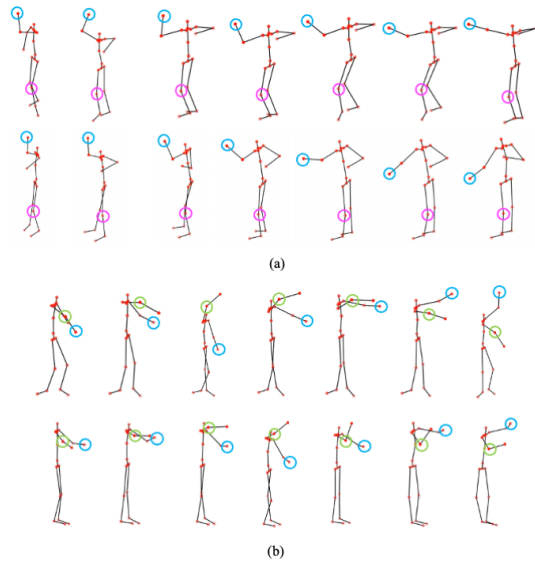


Fig. 7. (a) Visualized movement of Choreography 4 segmented for detailed comparison. (b) Visualized movement of Choreography 5 segmented for detailed comparison.

支援システムの開発を目的とし、Gaussian Process Hidden Semi-Markov Model (GP-HSMM) を拡張した新たなフレームワークを提案した。提案手法では、ガウス過程のカーネル近似に Random Fourier Features (RFF) を用い、さらにマルコフ連鎖モンテカルロ法 (MCMC) によりその近似精度を統計的に最適化することで、従来の GP-HSMM と同等の分節化精度を維持しながら、計算時間の大幅な短縮を実現した。加えて、逐次学習を導入することで、初期モデルデータセットで学習した熟練者の知見を活かしつつ、新たな被験者の動作を段階的に取り込むことが可能となり、個人ごとの動作特性を柔軟にモデル化できるようになった。

体操およびダンスの公開モーションキャプチャデータセットに基づく実験を通じて、MCMC によるカーネル近似の安定化と逐次学習による適応的なモデル更新の有効性が示され、動作単位での関節ごとの違いを定量的に評価し、動作改善のための具体的なフィードバックを提供可能な支援システムを実現した。

今後は、本システムをスポーツやリハビリなど動作の質や熟練度比較が求められる分野へ応用し、熟練者との比較に基づく個別支援を実現することで、実社会の課題解決に貢献できる基盤技術として発展させていく。

ACKNOWLEDGMENT

This research was supported by AMED under Grant Number JP24wm0625124. The authors acknowledge the use of the CMU Motion Capture Database and the ImperialDance dataset in this study.

REFERENCES

- [1] M. Balazian and P. Sojka, "Walker-independent gait recognition with CNN," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 4163–4167.
- [2] F. von Marcard, B. Rosenhahn, M. J. Black, and G. Pons-Moll, "3D people in the wild: From videos to accurate 3D human pose and shape," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 2670–2679.

- [3] F. M. Thoker, H. Doughty, and C. G. M. Snoek, "Skeleton - Contrastive 3D Action Representation Learning," in *Proc. 29th ACM International Conference on Multimedia (MM '21)*, Oct. 20–24, 2021, Virtual Event, China, pp. 1655–1663.
- [4] W. T. Lam, Y. M. Tang, and K. N. K. Fong, "A systematic review of the applications of markerless motion capture technology for clinical measurement in rehabilitation," *Journal of NeuroEngineering and Rehabilitation*, vol. 20, no. 1, pp. 1–26, 2023.
- [5] X. Suo, W. Tang, and Z. Li, "Motion capture technology in sports scenarios: A survey," *Sensors*, vol. 24, no. 9, pp. 2947, 2024.
- [6] X. Li, S. Wu, Z. Yang, and L. Yi, "DanceMVP: Multi-task dance performance assessment via text prompting," in *Proc. AAAI Conf. Artificial Intelligence*, vol. 38, no. 9, 2024, pp. 10270–10278.
- [7] T. Nakamura, T. Nagai, D. Mochihashi, I. Kobayashi, H. Asoh, and M. Kaneko, "Segmenting continuous motions with hidden semi-Markov models and Gaussian processes," *Frontiers in Neurorobotics*, vol. 11, p. 67, 2017.
- [8] A. Rahimi and B. Recht, "Random features for large-scale kernel machines," in *Advances in Neural Information Processing Systems*, J. Platt, D. Koller, Y. Singer, and S. Roweis, Eds., vol. 20. *Curran Associates, Inc.*, 2007.
- [9] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.
- [10] T. Broderick, N. Boyd, A. Wibisono, A. C. Wilson, and M. I. Jordan, "Streaming variational Bayes," *arXiv preprint arXiv:1307.6769*, 2013.
- [11] E. B. Fox, E. B. Sudderth, M. I. Jordan, and A. S. Willsky, "Joint modeling of multiple related time series via the beta process," *arXiv preprint arXiv:1112.1414*, 2011.
- [12] Y. Matsubara, Y. Sakurai, and C. Faloutsos, "Autoplaits: Automatic mining of co-evolving time sequences," in *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data*, 2014, pp. 193–204.
- [13] F. Sener and A. Yao, "Unsupervised learning and segmentation of complex activities from video," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8368–8376.
- [14] P. Bojanowski, R. Lajugie, F. Bach, I. Laptev, J. Ponce, C. Schmid, and J. Sivic, "Weakly supervised action labeling in videos under ordering constraints," in *Computer Vision - ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part V*, Springer, 2014, pp. 628–643.
- [15] D.-A. Huang, L. Fei-Fei, and J. C. Niebles, "Connectionist temporal modeling for weakly supervised action labeling," in *Computer Vision - ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV*, Springer, 2016, pp. 137–153.
- [16] A. Richard, H. Kuehne, and J. Gall, "Weakly supervised action learning with RNN based fine-to-coarse modeling," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 754–763.
- [17] K. Mimura, J. Matsumoto, D. Mochihashi, T. Nakamura, H. Nishijo, M. Higuchi, T. Hirabayashi, and T. Minamimoto, "Unsupervised decomposition of natural monkey behavior into a sequence of motion motifs," *Communications Biology*, vol. 7, no. 1, p. 1080, 2024.
- [18] I. Saito, T. Nakamura, T. Hatta, W. Fujita, S. Watanabe, and S. Miwa, "Unsupervised work behavior analysis using hierarchical probabilistic segmentation," in *IECON 2023-49th Annual Conference of the IEEE Industrial Electronics Society*, 2023, pp. 1–6.
- [19] Y. Mo, H. Sasaki, T. Matsubara, and K. Yamazaki, "Multi-step motion learning by combining learning-from-demonstration and policy-search," *Advanced Robotics*, vol. 37, no. 9, pp. 560–575, 2023. [Online]. Available: <https://doi.org/10.1080/01691864.2022.2163187>
- [20] M. M. Zhang, G. W. Gundersen, and B. E. Engelhardt, "Bayesian non-linear latent variable modeling via random Fourier features," *arXiv preprint arXiv:2306.08352*, 2023.
- [21] Y. Li, Z. Lin, F. Yin, and M. M. Zhang, "Preventing model collapse in Gaussian process latent variable models," *arXiv preprint arXiv:2404.01697*, 2024.
- [22] Y. Jung and J. Park, "Scalable hybrid HMM with Gaussian process emission for sequential time-series data clustering," *arXiv preprint arXiv:2001.01917*, 2020.