

# トピックモデルを用いた認識対象の選択制御とその動作認識への応用

Selection Control of Recognition Target and Motion Recognition based on Topic Model

小椋 忠志\*<sup>1</sup>  
Tadashi Ogura

稲邑 哲也\*<sup>1\*2</sup>  
Tetsunari Inamura

\*<sup>1</sup>総合研究大学院大学  
SOKENDAI(The Graduate University for Advanced Studies)

\*<sup>2</sup>国立情報学研究所  
National Institute of Informatics

Almost all conventional research on recognition of human activities used whole body motion patterns. However, since several motion primitives are related to certain body parts such as right hand, legs, head and so on, recognition might be a failure. The aim of this paper is to prevent recognition error due to mismatch of the dominant body parts of the motion and the motion pattern used in the recognition process. The key idea to realize the selection of appropriate body parts is using a relationship between a situation and the body parts. We adopt the topic model for the selection with an assumption that each situation has an implicit relation to the selection of body parts. In an experiment, we confirmed that the proposed method showed more correct recognition performance than whole body recognition.

## 1. はじめに

人の生活空間内で人の支援を行うロボットの実現のためには、人が今どのような状況で何をしているかという、観測対象である人の身体動作認識の技術発展が必要不可欠である。一方で多くの身体動作認識手法は、全身の身体動作信号を対象としたパターン認識によって実現されている。しかしながら、多くの身体動作は身体部位の一部のみの動作によって動作の意味と結びつき、その部位以外の動作は、認識すべき動作の意味とは関係の持たない。このような一部の部位によって実施される動作に対し、全身の身体動作信号を用いると、認識すべき動作の意味と関係の持たない部位による信号により、誤認識する恐れがある。そこで、人が観測対象の人の動作を認識する際には、その人の状況を考慮することで注目すべき身体部位を選択していると仮定し、本研究ではこの認識過程を実現する。ここでいう状況とは、どのような環境でこれまでどんなことを行っていたか、といった観測対象の人の置かれている状況を示す。

Goutsuらは、4つの関節を網羅的に複数組み合わせその関節群に対し重みを与えることで、スケルトンベースの認識において高い認識率を達成している [Goutsu 15]。Weiらは、関節角ごとの重みづけと2つの動作ラベル間の時間的前後関係性を用いて、複数動作の同時認識を可能にしている [Wei 13]。しかしながらこれらの手法では、動作のラベルと重みが紐づいており、状況に応じて身体部位の選択を行うといった操作はされていない。

小椋らは、トピックモデルを用いることで、状況を考慮した動作認識を提案している [小椋 16]。本研究はこの考え方を応用し、トピックモデルを用いた状況による認識部位の選択制御を行う手法を提案する。提案手法では、状況によって部位を選択することで、与えられた状況での選択された部位以外の動作はその状況では無関係の動作として分離できる。検証実験では、認識すべき動作に対しその動作とは無関係の動作を組み合わせた複合動作を認識対象とし、提案手法を用いることで誤認識を防ぐ有効性を検証する。

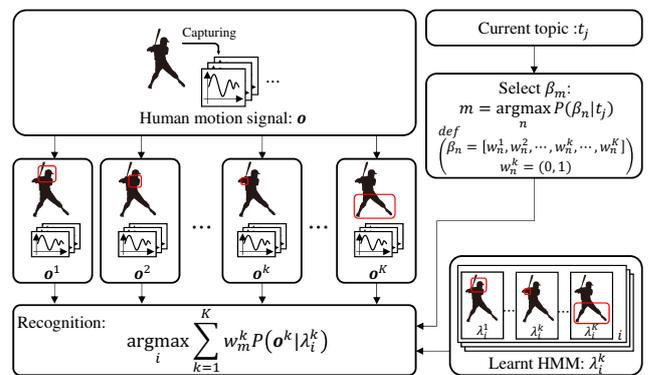


図 1: 現在トピックに基づく身体部位選択と認識手順

## 2. 提案手法

図 1 に提案する認識手法の概要図を示す。観測された身体動作信号  $o$  は身体部位ごとに複数の関節を持つジョイント群として分割する。  $P(o^k | \lambda_i^k)$  は、身体動作信号  $o^k$  と、同様に  $k$  ごとに学習した  $i$  番目の動作の隠れマルコフモデル (Hidden Markov Model; HMM) である  $\lambda_i^k$  との尤度である。その各身体部位に対して、部位の選択制御を行うベクトル  $\beta_n$  を以下のように導入する。

$$\beta_n = [w_n^1, w_n^2, \dots, w_n^k, \dots, w_n^K] \quad (w_n^k \in (0, 1)) \quad (1)$$

ここで、 $k$  は分割された身体部位のインデックスを指し、 $K$  はその分割数である。また  $n$  はベクトル  $\beta_n$  のインデックスである。  $w_n^k$  の値が 1 であればその部位を採択し、0 であれば捨てることとなる。

認識を行う際は、今どんな状況であるかということを表すトピック  $t_j$  に従い、状況に応じた部位選択ベクトル  $\beta_m$  のインデックスである  $m$  を、以下の式のように選択する。

$$m = \operatorname{argmax}_n P(\beta_n | t_j) \quad (2)$$

ここで  $P(\beta_n | t_j)$  はトピック  $t_j$  での  $\beta_n$  の出現確率を表し、トピックモデルの手法である Latent Dirichlet Allocation (LDA)

連絡先: 小椋忠志, 総合研究大学院大学, 東京都千代田区一ツ橋 2-1-2, t-ogura@nii.ac.jp

表 2: LDA による  $\beta_n$  の出現確率  $P(\beta_n|t_j)$ 

topic $t_0$ :	0.748*[0,1,0,1,0,0]	0.172*[0,0,1,1,0,0]	0.034*[0,1,0,0,0,0]	0.031*[1,0,0,0,0,0]	...
topic $t_1$ :	0.508*[1,0,0,0,1,0]	0.253*[0,1,0,0,0,0]	0.107*[0,0,0,0,1,0]	0.067*[0,0,0,1,0,0]	...
⋮	⋮	⋮	⋮	⋮	⋮
topic $t_{29}$ :	0.926*[0,0,1,1,0,0]	0.049*[1,0,1,0,0,0]	0.020*[1,0,0,1,0,0]	0.003*[0,0,1,0,1,0]	...

表 1: 実験に用いる各部位における単一動作

Body part	Motion label
head	head nod
	head turn
	head waggle
left arm	left arm punch
	left arm rotate
	left arm wave
left hand	left hand countup
	left hand pickup
	left hand rps
right arm	right arm punch
	right arm rotate
	right arm wave
right hand	right hand countup
	right hand pickup
	right hand rps
lower body	lower body squat
	lower body step
	lower body walk

表 3: 認識率

Method	Recognition ratio
Whole body	0.083333
Proposed method	0.222222

信号を部位ごとに切り離し、その後再接続することで生成する。その動作は、右手 3 種類・左手 3 種類からそれぞれ 1 つを選択した動作を正解ラベルとし、さらにその他 12 種類から 1 つの動作をノイズのようなものとして正解ラベルの動作とは異なる動作として選択した複合動作である。

その 108 動作に対し、事前に学習した 128 つの HMM による認識結果を表 3 に示す。全身動作による信号を用いた場合と比較し、提案手法の認識率が高いことが確認できる。この結果から提案手法が、複合動作の認識において有効であることが示された。

#### 4. おわりに

本稿では、トピックモデルを用いた状況における身体部位の選択制御と、その動作認識への応用について述べた。トピックモデルによる状況に従い身体部位を選択し認識を行う手法を提案した。有効性を検証する実験では、提案手法を用いることで動作の意味とは関係の持たない動作による誤認識を防ぎ、認識率が向上したことを確認した。実験では比較的類似度が高い動作に対し、128 つという多数の HMM をもって認識を行ったため、認識の条件としては非常に厳しいと言える。一方で、冒頭で述べたように、状況ごとに観測しうる動作はある程度絞ることができるはずである。本稿では、トピックモデルによる学習に用いたコーパスに、直接身体動作部位の選択を意味するベクトルを与えたが、本来であれば実際の動作系列やライフログのようなものから取得することが好ましい。そのため、今後の取り組みとして、[小椋 16]における状況ごとの動作出現確率と、本稿の提案手法における状況ごとの身体部位選択の過程を組み合わせることを予定している。

#### 参考文献

- [小椋 16] 小椋 忠志, 坂戸 達陽, 稲邑 哲也: トピックモデルを考慮した身体動作認識, 第 34 回日本ロボット学会学術講演会, 2Z2-02, (2016).
- [Goutsu 15] Goutsu, Y., Takano, W., and Nakamura, Y.: Motion Recognition Employing Multiple Kernel Learning of Fisher Vectors Using Local Skeleton Features. In Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 79–86, (2015).
- [Wei 13] Wei, P., Zheng, N., Zhao, Y., and Zhu, S. C.: Concurrent action detection with structural prediction. In Proceedings of the IEEE International Conference on Computer Vision, pp. 3136–3143, (2013).

によって学習される。トピックモデルは一般に文章中の単語を対象とした生成モデルであるが、本研究ではトピックモデルにおける単語を、部位の選択制御を行う  $\beta_n$  として扱う。各部位における尤度  $P(\beta_n|t_j)$  に  $w_n^k$  を以下の式によって適用することで、認識結果  $\hat{i}$  を決定する。

$$\hat{i} = \arg \max_i \sum_{k=1}^K w_n^k P(\beta_n^k | \lambda_i^k) \quad (3)$$

#### 3. 検証実験

提案手法の有効性を検証するため、身体動作認識実験を行った。本実験では身体部位の分割を手動で行い、表 1 の左側に示す 6 つの部位へと分割した。表 1 は本実験での各部位における動作のラベルを示し、各部位に対する動作は 3 種である。学習の対象となる動作は、18 種の動作から 2 つ選択された複合動作であり、その合計の 128 つの動作をモーションキャプチャデバイスで収録し、HMM で学習した。また式 (1) における  $\beta_n$  を、 $w_n^k$  のうち 1 つまたは 2 つの値が 1 となるように設定し、その選択部位数  $M$  は 21 となる。 $\beta_n$  を決定するための式 (2) における  $P(\beta_n|t_j)$  は LDA によって学習される。LDA のコーパスは、ある  $\beta_n$  を単語として、その単語が現れやすいよう恣意的に 15 個並べた文書を、1000 文書用意したものを対象とした。LDA によるトピックごとの出現確率  $P(\beta_n|t_j)$  の一部を表 2 に示す。トピック  $t_0$  を例とすると、出現確率  $P(\beta_n|t_j)$  の高い  $\beta_n$  から順番に、出現確率  $P(\beta_n|t_j)$  と  $\beta_n$  の配列が示されている。式 (2) によって選択される  $\beta_n$  は、左腕と右腕の部位を選択するベクトル  $([0, 1, 0, 1, 0, 0])$  である。

認識対象とする身体動作は、現在のトピックが  $t_0$  と与えられた場合を想定した、表 1 に示される 18 種の動作のうち右手 3 種類・左手 3 種類・その他 12 種類を網羅する 108 の生成動作を用意した。ただし、同一身体部位において同時に 3 つの動作を実演することは困難であるため、単一に取得した身体動作