

# 二重分節解析に基づいた物体操作に関する動作の教師なし学習

堀 碓惟† 田口 亮†

名古屋工業大学 大学院工学研究科†

## 1 まえがき

近年、人の動作を学習し、認識することの需要が高まっており、その研究が盛んに行われている。それらの中でも、教師なし学習は事前のラベル付けが不要なため、開発コストが低いことや、事前に動作を規定できない場合にも適用できることから注目されている。教師なし学習の手法の一つとして、時系列データを分節化可能な二重分節解析器がある。この手法は、身体動作や車の運転挙動の学習に適用されている[1, 2]。しかし、物体操作の学習は行われていない。人の動作には物体操作が多く含まれるため、人の動作の認識・解析には、物体操作の学習が必要となる。そこで本研究では、物体操作に関する人の動作を二重分節解析器により教師なしで学習することを目的とする。

## 2 提案手法

提案手法のフローを図1に示す。提案手法では、YOLO[3]で物体検出、OpenPose[4]で人の骨格を抽出する。そして、頭、物体の重心位置を基準とした関節の軌跡を二重分節解析器で学習する。

### 2.1 軌跡データ作成

本研究では、頭を原点とした関節の軌跡と、物体を原点とした関節の軌跡の両方をモデル化する。これにより各関節の動作だけでなく、物体の把持や、物体へ手を近づけるなど物体基準での動作を表現することができる。関節の軌跡データは次のように作成する。まず、YOLOで物体の二次元座標を取得する。ただし、検出結果は矩形であるため、その重心を物体の座標とする。次に、OpenPoseで関節の二次元座標を取得する。こうして得られた座標から頭原点、物体原点の関節の軌跡をそれぞれ作成する。

### 2.2 二重分節解析

本研究では、谷口らが提案した二重分節解析器(Hierarchical dirichlet process hidden language model, HDP-HLM)[5]を用いる。HDP-HLMのグラフィカルモデルを図2に示す。人間の言語は、音素の列が単語を構成し、単語の列が文を構成するという、二重分節構造を持つ。HDP-HLMはこの構造を模したモデルである。図2中の $z_s$ が単語、

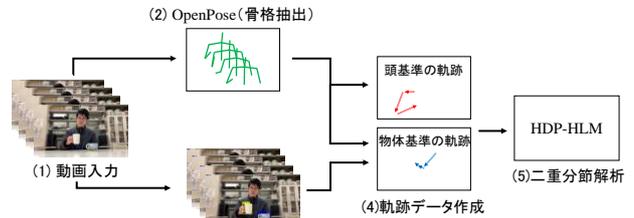


図1 提案手法のフロー

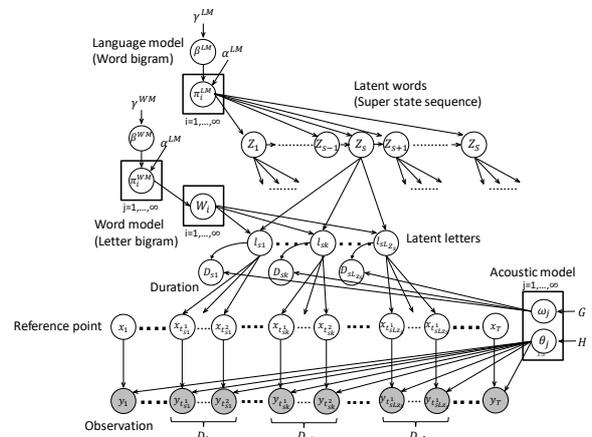


図2 HDP-HLMのグラフィカルモデル

$l_{sk}$ が単語 $z_s$ を構成する音素の一つであり、音素列を表すHSMM (Hidden Semi-Markov Model) を、単語列を表すHSMMが内包することで、二重分節構造を表現している。本研究ではHDP-HLMを用いて物体操作に関する動作の教師なし学習を行う。なお、本稿ではHDP-HLMの単語を動作カテゴリ、音素を動作プリミティブと呼ぶ。動作プリミティブには、関節数×座標系×2次元の多次元正規分布を用い、各関節の位置を表現する。動作カテゴリは動作プリミティブの遷移を表すことで、各関節の軌跡を表現する。

## 3 評価実験

### 3.1 実験条件

実験は、図3左のような、コップA(白)とコップB(柄付)を右手で操作する動画を12[本]用いて10[回]行う。各動画には「コップA/Bに手を近づける」、「コップA/Bから手を離す」、「コップA/Bで飲む」、「コップA/Bを右に動かす」、「コップA/Bを左に動かす」の5動作が含まれる。本実験では、右手の動作を学習するため、右手首と右肘の関節の軌跡を利用する。頭原点、コップA原点の座標系は図3右に示す通りである。なお、計算量削減のため、動作カテゴリの上限を15、動作プリミティブの上限を30とする。

Unsupervised Learning of Actions Related to Object Manipulation Based on Double Articulation Analysis  
Rui Hori†, Ryo Taguchi†  
† Nagoya Institute of Technology



図3 コップを操作する動画と座標系

### 3.2 実験結果と考察

実験の結果、コップ A で飲む動作(図 4,5), コップ A を左に動かす動作(図 6,7)などのコップ A を操作する動作が学習された. 図 4,6 は獲得された動作カテゴリの確率分布の一例である. 例えば, 図 4 の動作カテゴリ 5 は動作プリミティブ (2次元正規分布) が 1 (青), 9 (橙), 1 (青), 25 (緑), 13 (赤) の順番で遷移することを表している. グラフの原点が頭, または, コップ A の位置を示すことから, 頭座標系では, 手首が頭に近づいた後, 下に降ろされる動作を表し, コップ A 座標系では, 手首とコップが同じ位置にあることを表している.

一方で, コップ B を操作する動作では, 手を近づける動作と左に動かす動作が同じ動作カテゴリに分類された(図 8,9). これは, これらの動作の軌跡は, 頭座標系とコップ A 座標系では区別できないことを意味している. 実験で獲得された動作カテゴリは平均 14.5[個]であり, 理想的な動作カテゴリ数 10[個]よりも多い. これは, 図 10 のように, コップ A の初期位置の違いで異なる動作カテゴリとなる場合があることや, コップ B を操作する動作の分節化が適切にできていないことなどが起因している.

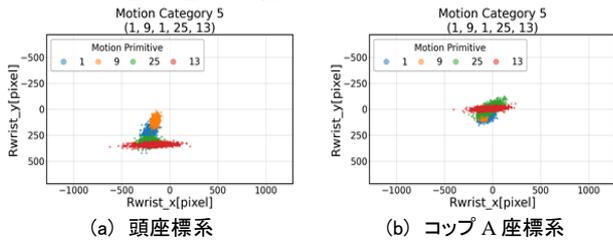


図4 獲得された飲む動作(右手首の位置分布)



図5 動作カテゴリ5として分割された動画の範囲

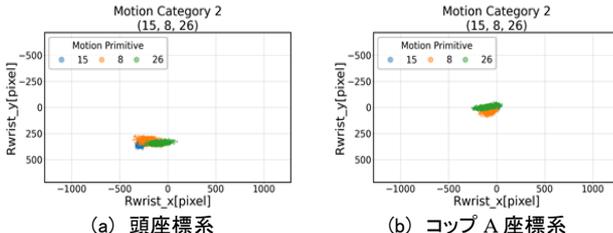


図6 獲得されたコップ A を左に動かす動作(右手首の位置分布)



図7 動作カテゴリ2として分割された動画の範囲

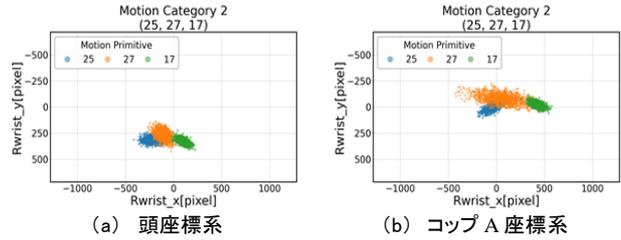


図8 コップ B の操作で区別つかない例(右手首の位置分布)



(a)コップ B に手を近づける動作



(b)コップ B を左に動かす動作

図9 コップ B の操作で区別つかない例の一部フレーム



図10 初期位置の違いにより異なる動作として獲得された例(上段と下段は異なる動作カテゴリとして獲得された)

## 4 まとめ

本稿では, 物体操作に関する動作のカテゴリを, 動画から教師なしで学習する手法を提案した. 提案手法では, 動画から頭基準の関節の軌跡と, 物体基準の関節の軌跡を抽出した後, それらの軌跡を既存手法である HDP-HLM で学習する. 実験の結果, 事前に設定した対象に対する動作のカテゴリを学習できることを確認した. 一方で, 同一動画中に存在する別の対象に対する操作は学習できていない. 本来は, 異なる対象に対する動作から, 共通の動作を学習できることが望ましい. 今後は, 座標系の原点とすべき対象を自動的に選定しながら, 動作を学習する手法の実現を目指す.

### 参考文献

- [1] T.Taniguchi et al.:Proc. IEEE/SICE Int. Symp. Syst. Integr., pp.250-255 (2011)
- [2] T.Taniguchi et al.:Proc. IEEE/RSJ Int. Conf. Intell. Robots syst., pp.849-854 (2012)
- [3] J.Redmon et al :arXiv preprint arXiv:1804.02767 (2018)
- [4] Z.Cao et al :arXiv preprint arXiv:1812.08008(2018)
- [5] T.Taniguchi et al :arXiv preprint arXiv:1506.06646 (2016)