

参照点と座標系を考慮した連続動作の二重分節解析

冠野 仁志 田口 亮

名古屋工業大学 大学院工学研究科

1 まえがき

ロボットによる見まね学習は、人の動作からロボットが新たな動作を学習することを目的としている。人の動作は複数の基本動作の組み合わせから構成されるが、多くの従来手法では、動作毎に教示データを区切って与える必要がある。教示された連続動作をロボットが自動的に分節化し学習できれば、教示コストが軽減される。谷口らは二重分節解析器を提案し、ジェスチャーや音声言語を教師なしで分節化することに成功している[1][2]。しかし、この研究では動作の基準となる参照点や座標系について考慮されていないため、それらに依存した動作は学習できない。そこで本研究では、複数の候補の中から適切な参照点と座標系を推定しながら、教示された連続動作を分節化する手法を提案する。

2 問題設定

本研究の問題設定を説明する。空間上にロボットと複数の物体を配置する。これらの物体の中で教示者が動かした物体を対象物体と呼び、動作の基準となる物体を参照点と呼ぶ。本研究では「(物体1に) 近づける」のような、参照点に依存した動作カテゴリを事前に定義する。また、図1(a)の「(物体1に) 近づける→(物体2に) 乗せる」のような複数の動作カテゴリからなる動作を連続動作と呼ぶ。教示者は、連続動作を見せながら、その連続動作を構成する動作カテゴリのラベルをロボットに教示する。ただし、参照点や座標系、動作カテゴリの分割位置は明示しない。ロボットのそれらを状況に合わせて推定しながら、各動作カテゴリを、図1(b)(c)のような参照点を原点とする相対座標系上での軌跡として学習する。

3 提案手法

本研究では、谷口らが提案した二重分節解析器[2]を拡張し、参照点と座標系の推定を可能とした新たな二重分節解析器を提案する。提案手法のグラフィカルモデルを図2に示す。灰色で示されたノードは観測変数を表し、白色で示されたノードは潜在変数を表している。提案手法では、動作カテゴリ z_s で使用された参照点と座標系

を表す確率変数をそれぞれ O_s, C_s とする。時刻 t における対象物体の絶対位置座標 y_{ts} を、参照点 O_s を原点、座標系の向きを C_s とした相対座標系上に変換したものを y_{O_s, C_s} と記す。観測値 y_{ts} は式(1)に従って生成される。ただし、 $P(y_{O_s, C_s} | \theta_{x_t})$ は θ_{x_t} をパラメータとする二次元正規分布である。

$$y_{ts} \sim P(y | \theta_{x_t}, O_s, C_s) = P(y_{O_s, C_s} | \theta_{x_t}) \quad t = 1, 2, \dots, L_{z_s} \quad (1)$$

提案手法では、従来の二重分節解析器と同様に forward sampling-backward filtering を用いて C_s, O_s , 分割位置 D_s を同時に推定する backward filtering の計算は式(2)(3)のように行う。

$$B_t(i) = P(y_{t+1:T} | z_s(t) = i, F_t = 1) \\ = \sum_c \sum_o \sum_j B_t^*(j, o, c) P(z_s(t+1) = j | z_s(t) = i) \quad (2)$$

$$B_t^*(i, o, c) \\ = P(y_{t+1:T} | z_s(t+1) = i, F_t = 1, O_s(t+1) = o, C_s(t+1) = c) \\ = \sum_{d=1}^{T-t} B_{t+d}(i) P(D_{t+1} = d | z_s(t+1) = i) \times \\ P(y_{t+1:T} | z_s(t+1) = i, D_{t+1} = d, O_s(t+1) = o, C_s(t+1) = c) \quad (3)$$

なお、先行研究[2]のモデルでは z_s も潜在変数であるが、本研究では教示の際に与えている。

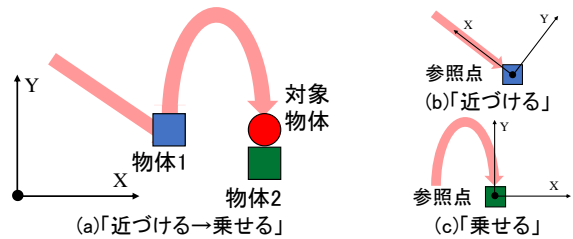


図1 連続動作の例

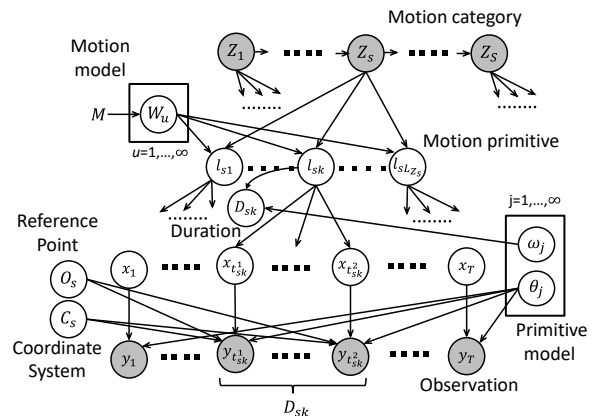


図2 提案手法のグラフィカルモデル

4 実験

4.1 実験条件

実験では、提案手法により教示された連続動作から動作カテゴリの分節化が可能であることを確認する。提案手法では、動作の軌跡を、二次元正規分布を出力分布として持つ HSMM で学習する。本実験では教示内容と学習結果を直接対応させて評価できるように、教示に用いる軌跡も HSMM で生成する。具体的には、図3に示すように3種類の動作カテゴリ（乗せる、近づける、離す）の HSMM を定義した。なお、図中では HSMM の各状態でサンプリングされた位置座標を示しており、横に状態番号を示している。例えば、「乗せる」という動作は青、橙、緑、…の順に位置が遷移していく。これらをもとに、「乗せる→近づける→離す」のような動作カテゴリを3個組み合わせさせた連続動作を20種類生成し、教示データとして用いる。図4は生成したデータの一例を示している。横軸がx座標、縦軸がy座標であり、対象物体の軌道が動作ごとに色別にプロットされている。また、配置された物体を四角で示す。座標系は以下の2種類とした。

- C1: 参照点を原点として、参照点と対象物体を結ぶ軸を x 軸とし、x 軸と直交する軸を y 軸とする座標系。（「近づける」「離す」）
- C2: 参照点が原点となるように絶対座標系を平行移動する座標系。ただし、変換後の座標系において動作の開始位置が正の場合には x 軸を反転させる。（「乗せる」）

図1(b)が座標系C1の、図1(c)が座標系C2の具体例を示している。これらのデータを用いて実験を10回行い、動作カテゴリの推定精度と参照点推定の正解率を評価する。動作カテゴリの推定精度についてはAdjusted Rand Index(ARI)を用いる。

4.2 実験結果と考察

表1に10回の実験における動作カテゴリ推定の ARI と参照点推定の正解率の平均と標準偏差をまとめて示す。図5に動作カテゴリの分節結果の一例を示す。縦軸はイテレーション回数を、横軸は時刻を表しており、各時刻 t における単語カテゴリの推定結果を色別に示している。また、上部には正解の分割位置が示されている。この図から、提案手法によって連続動作の分割位置が正しく推定されていることが確認できる。

図6は学習された単語カテゴリの分布の一例を示す。図から「乗せる」、「近づける」については事前に定義した軌道と同様の軌跡が学習されているが、「離す」については事前の定義と異なっている。「乗せる」や「近づける」は動作の開始位置と参照点の位置で座標系の向きが

決まるが、「離す」は動作の終端位置と参照点の位置で座標系の向きが決まる。そのため「乗せる」や「近づける」と同じようにforward samplingで推定することが困難であり、このような結果になると考えられる。

参照点推定の正解率が0.6程度になっているのは、「離す」という動作カテゴリの参照点推定に失敗しているためだと考えられる。

5 まとめ

本稿では、参照点と座標系を推定しながら動作を分節化する手法を提案した。

今後は、軌道情報だけでなく発話などマルチモーダルな情報を組み込み、より幅広い動作カテゴリに対応できる手法の実現をめざす。

参考文献

- [1] T. Taniguchi, et al: Double articulation analyzer for unsegmented human motion using pitman-yor language model and infinite hidden markov model, In Proceedings of 2011 IEEE/SICE International Symposium on System Integration, (2011).
- [2] T. Taniguchi, et al.: Nonparametric bayesian double articulation analyzer for direct language acquisition from continuous speech signals, IEEE Trans. on Cognitive and Developmental Systems, vol. 8, no. 3, pp. 171–185, (2016).

表1 実験結果

	平均	標準偏差
動作カテゴリ推定の ARI	0.565	0.057
参照点推定の正解率	0.635	0.018

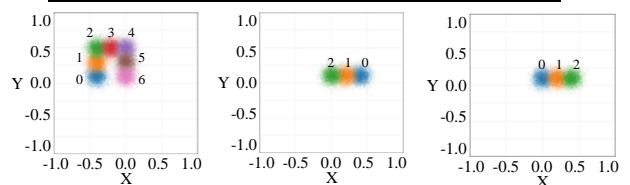


図3 事前に定義した動作カテゴリの分布

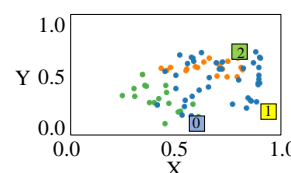


図4 生成したデータの一例

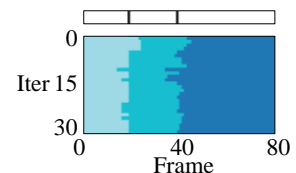


図5 分節結果の一例

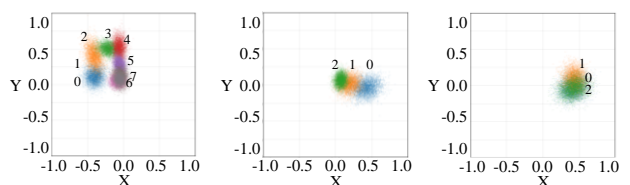


図6 学習された動作カテゴリの分布の一例