

文脈依存する身体動作-単語間の関係性のインタラクティブ学習

Interactive Learning of Context-depending Relationship between Human Motions and Words

坂戸 達陽 *1 稲邑 哲也 *1*2

Tatsuya Sakato

Tetsuanri Inamura

*1国立情報学研究所

National Institute of Informatics

*2総合研究大学院大学

SOKENDAI(The Graduate University for Advanced Studies)

Human motions which have the same physical movement and different meanings cannot be classify correctly when a robot performs motion recognition considering only physical movement of the motion. In order to classify that motions correctly, the robot should consider not only physical movement but also contexts in which the motions are performed. In this paper, we assume meaning of a motion corresponds to a word, and propose an interactive learning using virtual reality (VR) environment to acquire context-depending relationship between human motions and words. In the interactive learning, motions performed with various tools in various places are shown to a person using VR environment. The person instruct words correspond to the motions.



(a) 釣りの身体動作



(b) 料理の身体動作

図 1: まぎらわしい身体動作の例

1. はじめに

同じような身体動作でも場面によって意味が異なる場合がある。例えば、釣りの動作とフライパンで料理をする動作は、その身体動作だけを見ればよく似ているが、それぞれの動作の持つ意味は全く異なっている(図 1)。ロボットが人の身体動作のみからその動作の意味を推定する場合には、このような動きの似ている身体動作の意味の違いを区別できないという問題がある。しかし一方で、人は日常生活場面において、これらの身体動作を間違えることはない。それは、人がある動作の意味を推定する場合は、その身体動作だけでなく、それが行われ

ている場所やそのときに使っている道具などの、文脈情報を考慮しているからである。日常生活場面では、このように文脈情報を考慮しなければ意味の推定が困難な場面は多い。よって、ロボットが我々の暮らす日常生活場面で活動するためには、ロボットにも文脈情報を考慮して意味を推定すること求められる。ロボットが文脈情報を考慮して、身体動作の意味を推定する場合、何を文脈情報として扱うかという問題がある。例えば、ロボットが扱う文脈情報として、その身体動作が行われている場所を考慮した場合を考える。この場合、海岸で図 1 の身体動作が行われた場合は釣り、キッチンで図 1 の身体動作が行われた場合は料理であると、それぞれの身体動作の意味を区別することができる。しかし一方で、単に場所を文脈として扱うだけでは、海岸で料理を行った場合など、同じ場所で身体動作が同じだが意味が異なる身体動作を行った場合は区別することができない。この場合は、場所だけでなく、その身体動作で用いられている道具も考慮する必要がある。このように、ロボットが文脈を考慮した身体動作の意味の推定を行うためには、場所だけでなく、その動作に用いられている道具など、様々な要素を考慮する必要がある。これに対して小椋らはトピックを考慮した身体動作認識を提案している [1]。

ロボットが文脈依存の身体動作の意味を獲得する方法として、様々な文脈上で行われる身体動作に対して、人にその身体動作に対応する単語が何であるかを答えてもらうという方法が挙げられる。しかし、ただロボットが遭遇する場面を観察するだけでは、必要な場面を網羅することは難しいと考えられるため、ロボットが人に対して積極的に場面を提示して、そこで行われている身体動作に対応する単語を答えてもらうというインタラクティブ学習を行うことが望ましい。Araki らは、ロボットが物体の形や振ったときの音、触覚情報を観測している間に人が発話することによって、単語情報を含めたマルチモーダルな物体概念を獲得する [2]、インタラクティブ学習を提案している。ただし、彼らの研究で提示しているものは、その場に存在する物体に関する情報であり、文脈依存の身体動作の意味は対象としていない。

人に場面を提示する際に、想定されるすべての文脈と身体動作の組み合わせに対して、対応する単語を答えてもらうのは人に対する負担が大きくなってしまいうため、場面の提示は、ただ闇雲に行うのではなく、何らかの方策に基づいて行うべき

連絡先: 坂戸 達陽, 国立情報学研究所, 〒 101-8430 東京都千代田区一ツ橋 2-1-2, sakato@nii.ac.jp

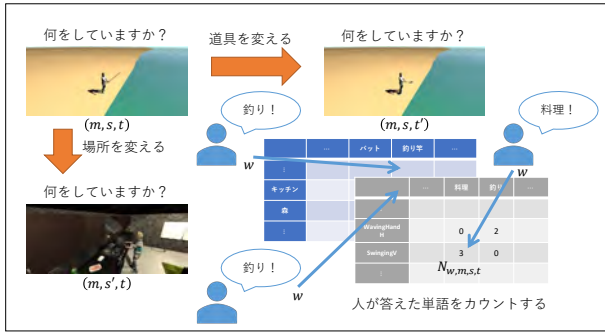


図 2: VR 環境を用いたインタラクティブ学習の概略図

である。

海岸, キッチンなど, 提示する情報に場所に関する情報が含まれる場合, 現実環境で人に次々と場面を提示することは困難である。また, 文脈と身体動作の組み合わせの中には, 現実にはほぼありえないような組み合わせも存在し, 現実環境では, 想定されるすべての文脈および身体動作について対応する単語を人に答えてもらうことはできない。

これらの問題を解決するための方法として, 仮想現実 (VR) 環境を用いて場面を提示することが考えられる。VR 環境では, 場所の移動や道具の変更は, モデルを切り替えるだけよい。また, 現実環境ではほぼありえないような文脈と身体動作の組み合わせも提示することは容易である。

本稿では, 身体動作の意味を単語とし, 文脈依存の身体動作-単語間の関係性を仮想現実環境を用いてインタラクティブに学習する手法を提案する。

2. 提案手法

2.1 インタラクティブ学習

本稿で提案するインタラクティブ学習の概略を図 2 に示す。本稿で提案するインタラクティブ学習では, まず, システムが人に, VR 環境上でアバターが身体動作を行っている場面を見せ, 人にその身体動作に対応する単語を答えてもらう。人が答える単語は予め決められた単語の中から選ばれる。システムは単語ごとに特定の場所, 道具, 身体動作の組に対してその単語を人が答えた回数を記録する。その後, 次の場面に移る。次に人が答える場面は, システムが現在提示している場面の場所, 道具, 身体動作のうち, いずれかが異なるものの中からシステムがランダムに選ぶ。システムは決められた数の場面を答えてもらうと学習を終了する。

2.2 特定の場面における身体動作-単語間の関係性の獲得

場所 s において, 道具 t を用いて身体動作 m を行っている場面を人に見せたときに, 人がその場面に対応する単語が w であると答えた回数を $N_{w,m,s,t}$ とすると, 特定の場面 (m, s, t) における身体動作単語間の関係性は, $\mathbf{N}_{m,s,t}$ をパラメータとするディリクレ分布 $\text{Dir}(\mathbf{N}_{m,s,t})$ によって得る。ここで $\mathbf{N}_{m,s,t}$ は,

$$\mathbf{N}_{m,s,t} = \{N_{w_1,m,s,t}, N_{w_2,m,s,t}, \dots, N_{w_W,m,s,t}\} \quad (1)$$

である。 W は想定される単語 w の数である。

3. 実験設定

評価実験において, 想定する身体動作を表 1 に, 想定する場所, 道具, 単語を表 2 に示す。

表 1: 想定する身体動作

身体動作
右手を垂直平面上で左右に振る (WavingHandV)
右手を水平平面上で左右に振る (WavingHandH)
両手をお腹の前で上下に振る (SwingingV)
両手を右上から袈裟に振り下ろす (SwingingH)

表 2: 想定する場所, 道具, 単語

場所	道具	単語
海岸	(道具なし)	挨拶
キッチン	フライパン	料理
森	バット	掃除
ゲームセンター	釣り竿	釣り
	布	打つ
	マレット	切る
	斧	

評価実験では, 学習によって獲得された身体動作-単語間の関係性の評価だけでなく, 人に提示する場面を完全にランダムに選択するシステムとの比較も行う。

4. おわりに

本稿では, 身体動作の意味を単語とし, 文脈依存の身体動作-単語間の関係性を仮想現実環境を用いてインタラクティブに学習する手法を提案した。提案した選択手法の評価が今後の課題である。

参考文献

- [1] 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.
- [2] T. Araki, T. Nakamura, T. Nagai, K. Funakoshi, M. Nakano, and N. Iwahashi: Online object categorization using multimodal information autonomously acquired by a mobile robot. *Advanced Robotics*, Vol. 26, pp. 1995–2020, 2012.