

ヒューマノイドロボットにおける二項動詞獲得手法の性能評価

Evaluation of Body Movement Verbs Acquisition Method for a Humanoid Robot

長谷川 大*¹ ジェプカ ラファウ*² 荒木 健治*³
Dai Hasegawa Rafal Rzepka Kenji Araki

*¹北海道大学大学院情報科学研究科

Graduate School of Information Science and Technology, Hokkaido University

We propose a language grounding method that relates verbs which imply body manipulation to motor angle patterns for a humanoid robot. In established methods, verbs are represented by statistical models based on trajectories or motor patterns of a trajector. In our method we use a novel representation model that has six features including both trajector-reference point relationships and the trajector's trajectory. By using this model, some verbs which do not depend on a trajectory, e.g. "move the right hand close to the left hand" are represented more adequately. In our method a robot generates abstract verb meanings independent of context. As input it uses sets of a user input textual command and a motor pattern. The motor pattern is taught using direct physical feedback. We implemented the algorithm and conducted a verb acquisition experiment. As a result, four problematic verbs, "place-on", "move-close-to", "move-away-from", and "touch-with" were acquired correctly.

1. はじめに

近年の自然言語処理では、言語を言語が意味する物理世界や概念から切り離された単なる記号または記号列と見做し、記号を統計的に処理することによって様々なタスクを実現している。意味処理においても、三つ組や格フレーム、シソーラス、類義語辞典など代表的な手法やツールは記号を記号によって定義する循環的定義が行われ、記号が実世界と結びつくことはない。我々は、深い意味処理を実現するためには、実世界に対応した状態での記号に基づくアプローチが必要であると考えている。しかしそのためには、記号をどのように実世界と結び付けるべきかという記号接地問題 [Harnad 90] に取り組まなくてはならない。ここで考慮すべきは身体の重要性である。Brooks [Brooks 87] は知覚する主体における身体の重要性を主張しているが、これは言語の記号接地においても同様であると言える。このような理由から、我々は、ロボットが自身の知覚を通して言語を獲得するアルゴリズムを実現することが自然言語処理技術の発展に寄与すると考えている。

ロボットにおける言語獲得は既に多くの研究が行われている [Roy 05, Siskind 94, Siskind 01, Steels 08, Steels 01b, Steels 01a]。Roy [Roy 02] や岩橋 [Iwahashi 03] の画像と音声から自動的に物の名前を学習するモデルなど特に名詞に関する研究は成功を収めている。しかし、動詞の獲得に関する研究はまだ十分とは言えない。杉田ら [Sugita 05] はセンテンスとアーム付移動ロボットのモータパターンのペアをリカレントニューラルネットワークにより学習し、シンボル操作可能な動詞の獲得を実現している。杉浦ら [Sugiura 08] はセンテンスとアームロボットの軌道のペアを隠れマルコフモデルを用いて学習し、物体操作を表現するための動詞の獲得を行っている。しかし、「近づける」のような動詞は動作主体と対象物の距離を縮めることを意図しており、軌道やセンサパターンには依存しないため適切な表現モデルとは言い難い。

そこで本稿では、6つの素性による意味表現モデルを使用し、

ヒューマノイドロボットにおける身体操作を表現するための二項動詞の獲得手法を提案する。

2. 提案手法

本稿では、文を形態素を分割する能力と「右手」「左手」「頭」の3つの名詞は既に獲得されていることを前提としている。さらに、ロボットの胸を原点とし、頭上方向をz軸、正面をx軸とする座標系を基準座標系とする。

2.1 システム概要

まずシステムの全体図を図1に示す。ヒューマノイドロボットにはKHR2-HV*¹を使用している。

学習フェーズではロボットはユーザから「右手を頭に置く」などの二項動詞を含む文を入力される。つぎにユーザから適切な動作を直接教示により入力される。教示された動作はモータ角度のパターンとして知覚される。モータパターンは動作認知モジュールによって適切な表現系に変換され、文と動作表現のペアとして用例データベースに格納される。ロボットが実際に経験した事例である用例は汎化モジュールで動詞毎に汎用性のある文脈に依存しない表現に抽象化され、ルールとしてルールデータベースに格納される。

テストフェーズではロボットはユーザから未知の文を入力される。未知の文は、ルールデータベース内に存在する動詞ならば、ルールを用いて行動を生成する。

2.2 二項動詞のための表現モデル

本稿では身体操作を表現するための動詞の表現モデルとして、以下の6つの素性を提案する。

1. 動作主体
2. 動作客体
3. 動作主点と対象点との初期距離と最終距離の変化量
4. 動作主点と対象点との最終距離

連絡先: 長谷川 大, 言語メディア学研究室, 札幌市北区北14条西9丁目, 011-706-7389, hasegawadai@media.eng.hokudai.ac.jp

*¹ Kondo Kagaku Co. Ltd, <http://www.kondo-robot.com/>

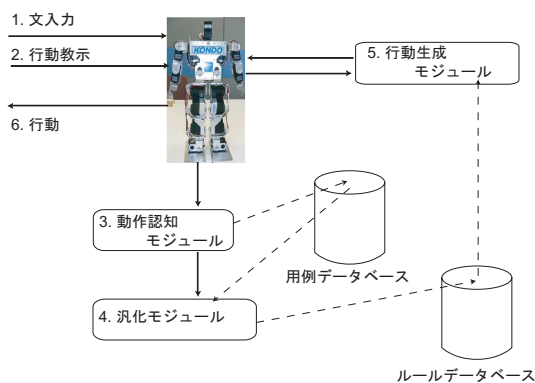


図 1: システム全体図

5. 動作主点と対象点の上下関係
6. 対象点を原点とし、z 軸をロボット頭上方向に固定し、動作主点の初期位置方向に x 軸を向けた座標系における動作主点の軌道

「右手を頭に置く」を表現した場合の例を図 2 に示す。

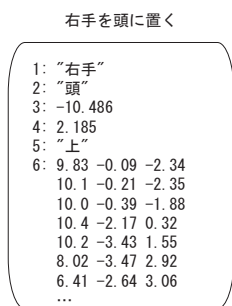


図 2: 表現モデルの例

2.3 動作認知モジュール

動作認知モジュールには入力文と教示されたモータ角度パターンが入力される。「右手を頭に置く」を例に取ると、このモジュールでは、動作主体が「右手」であること、動作の客体が「頭」であることを理解したうえで、その他の素性について値を計算し、それらを入力文とペアにし用例データベースに追加する(図 3)。動作主体と動作客体の認知では、教示されたモータパターンから基準座標系での「右手」と「左手」の軌道を計算する。この軌道から、動作量の大きいものを動作主体とし、他方を動作客体としている。

2.4 汎化モジュール

汎化モジュールでは、ロボットが経験した二項動詞と動作表現のペアから汎用性の高い動作表現を生成しルールとしてルールデータベースに追加する(図 4)。重要素性の決定処理では、各用例間で素性の値の類似性が高ければ、その動詞を表現する素性として重要であるという仮定をし、類似判定を行って重要素性を決定している。これは、様々な文脈で経験した用例にも関わらず値が類似している素性は、その動詞を表現する文脈に依存しない一般的な素性であるという考えに基づいている。さらに、重要素性および軌道について、より汎用性の高い値または軌道を得るために、値を平均化する。

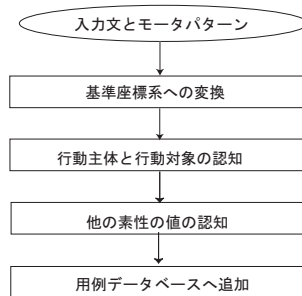


図 3: 動作の認知

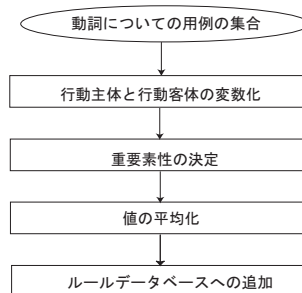


図 4: 用例の汎化

2.5 行動生成モジュール

テストフェーズでは入力文は行動生成モジュールに入力される。行動生成モジュールでは、データベース内にルールが存在する動詞ならば行動を生成する。行動を生成する場合、まず動作主体と動作客体を決定し、ルールから動作の最終位置を決定する。その後、平均化された軌道を最終位置に合わせて平行移動した軌道を生成する。そして最後に、軌道をモータ角度パターンに変換し行動を生成する。

3. 実験

上記のアルゴリズムを実装し、「置く」「近づける」「離す」「触る」「なでる」「突っつく」の 6 種類の二項動詞の獲得実験を行った。2 人の被験者による実験の概要を以下に示す。まず各動詞についてトレーニングセンテンスとテストセンテンスを用意する。トレーニングセンテンスは動作主が右手で対象点が頭である文と、動作主が左手で対象点が右手である文の 2 文とする(「置く」の場合は「右手を頭に置く」と「左手を右手に置く」)。次にテストセンテンスは動作主が左手で対象点が頭である文(「左手を頭に置く」)とする。2 人の被験者がそれぞれトレーニングセンテンスを 3 回ずつ教示を行った場合のテストセンテンスの行動出力結果を表 1 に示す。この際、トレーニング、テストセンテンスともに、ロボットの初期姿勢を変化させた後、入力を行った。出力結果の評価は 3 段階(1: 適切, 2: 適切に近い, 3: 誤り)で行った。

4. 考察

実験の結果、本手法により「置く」「近づける」「離す」「触る」の 4 種類の動詞の評価は、1 (適切) に近く、それらが正しく獲得されたことを示している。獲得された「置く」「近づける」のルールを図 5、図 6 に示すが、動詞毎に適切に素性の

表 1: 実験結果 (1:適切, 2:適切に近い, 2:誤り)

動詞	被験者 A の平均	被験者 B の平均	平均
置く	2	1.6	1.8
近づける	2	1.3	1.6
離す	1	2	1.5
触る	1	1.6	1.3
なでる	2.3	2.6	2.5
突っつく	2.3	3	2.6

選択がなされている。このように我々の表現モデルを使用することにより、動詞がどのように表現されているかが設計者に容易に理解可能である。さらに、本手法は明示的に動作主点と対象点の関係を素性としているため、軌道に依存しない動詞表現が可能である。本実験で獲得された動詞表現は、障害物を想定した状況においても適用可能である。我々は、障害物を迂回するシンプルなプログラムを使用し、障害物を想定した場合の「置く」の軌道生成を行った。図 7 に軌道の内部表現を示すが、ロボットは適切に障害物を迂回して目的の動作を達成した。

一方「なでる」「突っつく」という双方向の軌道を生成する動作の評価は、3 (誤り) に近く、それらの獲得が困難であることを示している。この原因として、本手法では、抽象化の際に軌道をシンプルに平均化していることが挙げられる。これらの動作を適切に表現する素性を導入することは今後の課題である。さらに、本実験で対象とした動詞は、使用したロボットの構造・知覚機構に依存する再現可能な動作を考慮して選択している。今後、使用するロボットを変更することで、対象とする動詞の種類を増加し、本手法の有効性を確認したい。

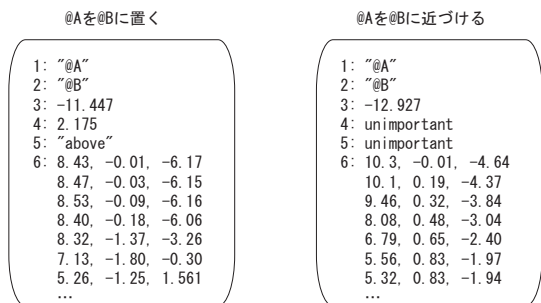


図 5: 「置く」の表現 図 6: 「近づける」の表現

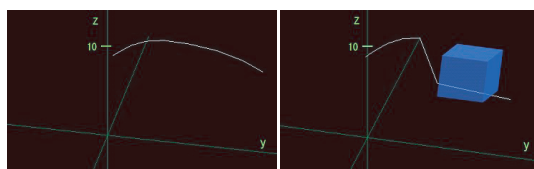


図 7: 障害物を想定した「置く」の軌道の内部表現

5. まとめと今後の予定

本稿では、6つの素性からなる動詞の表現モデルを使用して、身体操作を表現するための二項動詞を獲得する手法の提案を行った。ヒューマノイドロボット上に実験システムを実装し、

実験を行った結果、「置く」「近づける」「離す」「触る」の4種類の動詞を適切に獲得することを確認した。今後、ロボットの変更、素性の追加を行い、さらに多くの種類の動詞について評価実験を行う予定である。

参考文献

[Brooks 87] Brooks, R. A.: Intelligence without representation, *Artificial Intelligence*, Vol. 47, pp. 139–159 (1987)

[Harnad 90] Harnad, S.: The symbol grounding problem, *Physica D*, Vol. 42, pp. 335–346 (1990)

[Iwahashi 03] Iwahashi, N.: Language acquisition by robots? Towards a new paradigm of language processing, *Journal of Japanese Society for Artificial Intelligence, Special Issue on Language Acquisition*, Vol. 48, No. 1, pp. 49–58 (2003)

[Roy 02] Roy, D.: Learning Words from Sights and Sounds: A Computational Model, *Cognitive Science: A Multidisciplinary Journal*, Vol. 26, No. 1, pp. 335–346 (2002)

[Roy 05] Roy, D.: Semiotic schemas: A framework for grounding language in action and perception, *Artificial Intelligence*, Vol. 167, pp. 170–205 (2005)

[Siskind 94] Siskind, J. M.: Grounding language in perception, *Artificial Intelligence Review*, Vol. 8, No. 5-6, pp. 371–391 (1994)

[Siskind 01] Siskind, J. M.: Grounding the lexical semantics of verbs in visual perception using force dynamics and event logic, *Journal of Artificial Intelligence Research*, Vol. 15, pp. 31–90 (2001)

[Steels 01a] Steels, L.: Language games for autonomous robots, *IEEE Intelligent Systems*, Vol. Sep/Oct issue, pp. 16–22 (2001)

[Steels 01b] Steels, L. and Kaplan, F.: AIBO's first words, The social learning of language and meaning, *Evolution of Communication*, Vol. 4, No. 1, pp. 3–32 (2001)

[Steels 08] Steels, L. and Spranger, M.: The robot in the mirror, *Connection Science*, Vol. 20, No. 4, pp. 337–358 (2008)

[Sugita 05] Sugita, Y. and Tani, J.: Learning Semantic Combinatoriality from the Interaction between Linguistic and Behavioral Processes, *Adaptive Behavior*, Vol. 13, No. 1, pp. 33–52 (2005)

[Sugiura 08] Sugiura, K. and Iwahashi, N.: Motion Recognition and Generation by Combining Reference-Point-Dependent Probabilistic Models, in *Proceedings of IEEE/RSJ 2008 International Conference on Intelligent Robots and Systems (IROS 2008)*, pp. 852–857, Nice, France (2008)