

語彙の分散意味に基づくヒューマノイドロボットの動作生成に向けた基礎的検討

濱園 侑美[†] 小林 一郎[†] 麻生 英樹[‡] 持橋 大地[¶]

Muhammad Attamimi[§] 中村友昭[§] 長井隆行[§]

[†]お茶の水女子大学大学院 [‡]産業総合技術研究所 [¶]統計数理研究所 [§]電気通信大学大学院

[†]{g1020528,koba}@is.ocha.ac.jp, [‡]h.asoh@aist.ne.jp, [¶]daichi@ism.ac.jp,
[§]m_att@apple.ee.uec.ac.jp, [§]tnakamura@uec.ac.jp, [§]tnagai@ee.uec.ac.jp

1 はじめに

近年、日本の少子高齢化社会の到来が問題となっており、問題解決のためにロボットが家庭内で活躍する場面が増えてくると考えられる。ロボットが安価に入手できるようになり、人の感情を推定しコミュニケーションを行うロボットも販売され、ロボットと人のコミュニケーションに大きな進展を見せる可能性がでてきている。このように家庭内にロボットが入ってきた際にロボットと居住者とが協調して暮らせる条件として、言葉や身振りなどを使って居住者の経験をロボットに伝え、同じ動作を学習させることが必要になると考える。

このことを踏まえ、本研究では、ロボットが人の言葉による指示から行動を行うことができるように言葉と動作の対応関係を学習することによって、初めて行う動作であっても言葉の意味から推測し、動作を行なえるようにすることを目的とする。

本研究では先行研究 [3] に対し、理解のための言語資源を増やすことでロボットが言葉の意味をより理解出来るよう改善し、更に動作の学習方法を変えることによるロボットの動作への影響を考察する。

2 提案手法

いくつかの言葉の意味と動作表現の対応関係が既知であるとする。この時、動作と対応関係が分からない未知の言葉が与えられた際に、他の言葉との意味的な関係から対応する動作を推定する手法を提案する。

図 1 に提案手法の概要を示す。言葉を意味空間へ配置する方法は word2vec[2] を用い、動作を動作表現空間へ配置する方法は先行研究 [3] で提案した Activity-Attribute Matrix(AAM)[1] を動作生成に適した形に

改良したもの(「時系列対応 AAM」と呼ぶ)、及びロボットの関節角を用いる 2 通りの方法を採用した。また、言葉と動作の対応関係の学習にはニューラルネットワークを用いる。

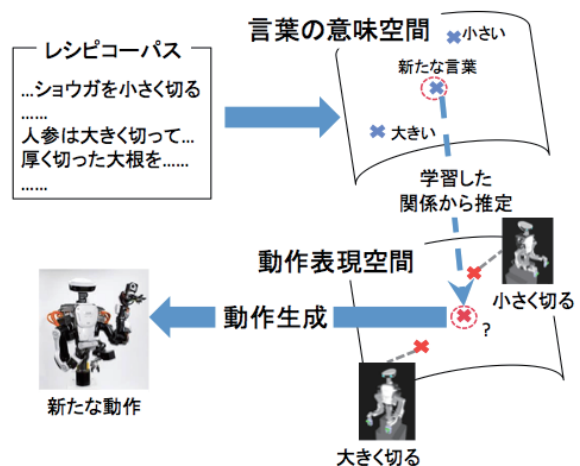


図 1: 提案手法の概要

3 ロボットの動作

3.1 ヒューマノイドロボットの特徴

使用するロボットは(株)川田工業社製双腕ヒューマノイドロボット HIRONXC を用いる。

HIRONXC は首の左右上下運動に 2 つ、腰の回転に 1 つ、左右両腕に各 6 つ、両手に各 4 つの全 24 つの関節を持ち、それぞれの関節角と時間 t を指定することで、 t 秒かけて指定された角度へと関節を動かすことが可能である。

3.2 動作構成

ロボットの調理動作を関節軸の基本動作から構成することを考える。ロボットの24つの関節を全て人手で生成するのは困難であるため、本節で述べる動作構成方法により訓練用データのロボット動作生成を行う。AAMは動作と動作に関連している意味属性を符号化したものである、具体例を図2で示す。

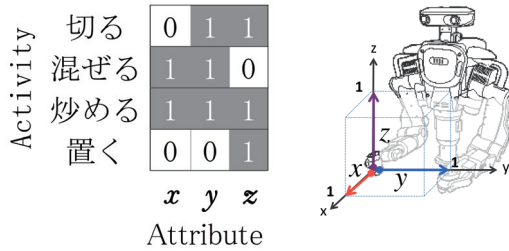


図 2: HIRO を用いた Attribute 作成

M を Activity (活動), N を Attribute (属性) とし、各要素 $a_{ij}(i \in M, j \in N)$ において Attribute の Activity への含有関係について Activity i を構成するのに Attribute j が用いられている場合は1, 用いられていない場合 j は0とする, $M \times N$ 行列により動作を表現する。AAMは、人間に読める Attribute の組み合わせを取ることで、動作の内容が分かりやすいという特徴がある。例えば、図2の右のように、右手の指先を前後に動かす (x), 左右に動かす (y), 上下に動かす (z) を Attribute とした場合、混ぜるという Activity は前後に動かす (x), 左右に動かす (y) の2つの動きからなる、ということが図2の左より分かる。

本研究では Activity を生成する過程において、それぞれの Attribute の度合い、時系列性、速度等が重要となるため、表1に示した、AAMを改良した時系列対応 AAMを用いる。 x, y, z の変動割合と時間 t を組み合わせた $[p_x, p_y, p_z, t]$ を時系列に n 個並べた

$S = s_1, s_2, \dots, s_n = [[p_{x_1}, p_{y_1}, p_{z_1}, t_1], [p_{x_2}, p_{y_2}, p_{z_2}, t_2], \dots, [p_{x_n}, p_{y_n}, p_{z_n}, t_n]]$ と x, y, z の Attribute でのロボット関節情報を与えることにより動作の生成を可能にする。

4 言語の分散意味表現

本研究では Mikolv ら [2] によって提案された word2vec を利用し、単語の意味関係から未知の単語に対する動作の推定を可能にする。コーパスはクック

¹表1中の CY, NN, ..., LH4 はロボットの24つの各関節を表している。

表 1: 時系列対応 AAM の概要

速く切る	0	0	-5	1	0	2	5	1	0	2	5	1
ゆっくり切る	0	0	-5	5	0	2	5	5	0	2	5	5
粗く切る	0	0	-5	2	0	4	5	2	0	4	5	2
細かく切る	0	0	-5	2	0	1	5	2	0	1	5	2
ざっくり切る	0	0	-5	5	0	4	5	5	0	4	5	5

$p_{x_1} p_{y_1} p_{z_1} t_1 p_{x_2} p_{y_2} p_{z_2} t_2 \dots p_{x_n} p_{y_n} p_{z_n} t_n$

CY	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
NY	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
NP	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
RSY	0.1	1.8	0.7	0.1	0.1	1.8	0.7	0.1	0.1	1.8	0.7	0.1
RSP	-2.3	0.7	0.1	0.1	-2.3	0.7	0.1	0.1	-2.3	0.7	0.1	0.1
REP	2.1	-0.5	-2.7	0.1	2.1	-0.5	-2.7	0.1	2.1	-0.5	-2.7	0.1
RWY	0.0	0.0	0.1	0.1	0.0	0.0	0.1	0.1	0.0	0.0	0.1	0.1
RWP	0.2	0.2	2.7	0.1	0.2	0.2	2.7	0.1	0.2	0.2	2.7	0.1
RWR	0.0	-1.8	0.0	0.0	0.0	-1.8	0.0	0.0	0.0	-1.8	0.0	0.0
LSY	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
LSP	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
LEP	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
LWY	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
LWP	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
LWR	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
RH1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
RH2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
RH3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
RH4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
LH1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
LH2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
LH3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
LH4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

パッド²のレシピを用いた。なお、word2vecを用いる際の前処理としてクックパッドデータの材料コーパスを辞書とすることで MeCab による分かちを正確にし、より良い分散意味ベクトルが構築出来るよう目指した。分散意味表現のベクトルを50次元とし、図3にその結果を主成分分析を行い可視化したものを示す。

5 言語と動作の関係学習

言語と動作の関係を学習するために、3層の階層型ニューラルネットワーク (NN) を用い、学習方法としてはバックプロパゲーションを用いた。また活性化関数はシグモイド関数を用いた。

言語は word2vec のうち skip-gram を利用し、50次元の分散意味表現で表した。動作は「切る」の様々なパターンを3.2節に示した時系列対応 AAM で作成し、表2のように s_1 と s_2 の2つの動作を合わせて8次元のベクトルで表した。

表 2: 「切る」動作の時系列対応 AAM

言語表現	s_1				s_2			
	p_{x_1}	p_{y_1}	p_{z_1}	t_1	p_{x_2}	p_{y_2}	p_{z_2}	t_2
基準 (切る)	0	0	-5	3	0	3	5	3
大きい	0	0	-5	3	0	5	5	3
小さい	0	0	-5	3	0	2	5	3

²<http://www.nii.ac.jp/dsc/idr/cookpad/cookpad.html>

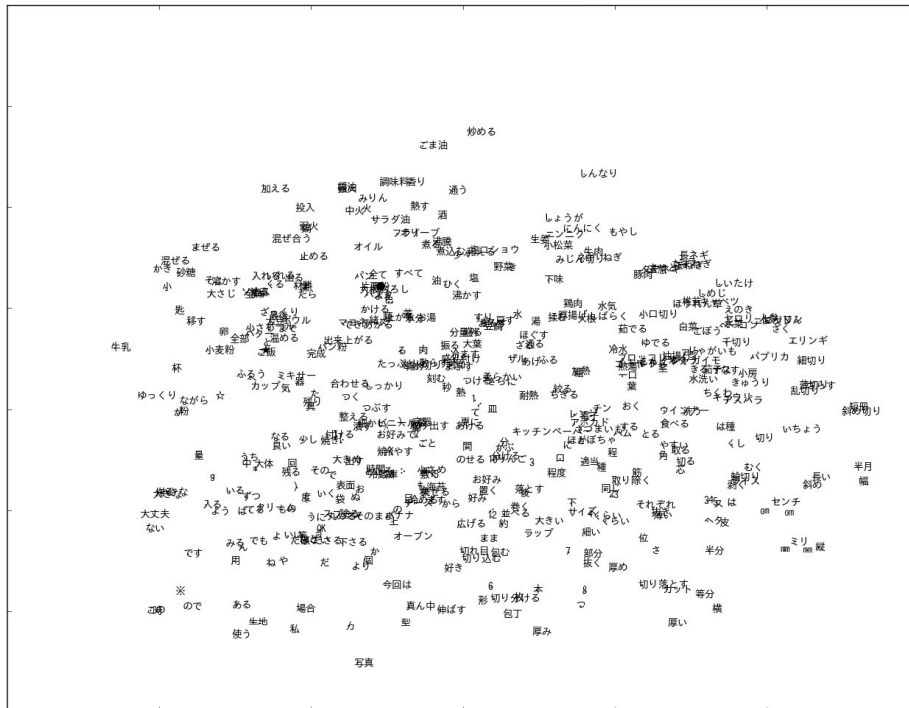


図 3: word2vec によるクックパッドレシピの上位 500 単語の分散意味表現

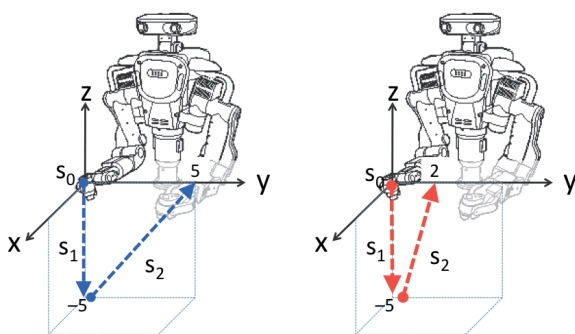


図 4: 表 2 での腕の動き (左: 大きい, 右: 小さい)

表 2 で示した動作は図 4 のように表現出来る。「大きい」の動きを例にとると、 s_1 の動きは、 $s_1 = [p_{x1}, p_{y1}, p_{z1}, t_1] = [0, 0, -5, 3]$ で表され、図 4 左の s_1 でも z への動きを行っている。次の s_2 の動きは、 $s_2 = [p_{x2}, p_{y2}, p_{z2}, t_2] = [0, 5, -5, 3]$ で表され、これも図 4 左の s_2 で y と z への動きを同時に行っていることが確認出来る。

また時系列対応 AAM を用いた動作は相対的な動作

になってしまうため、本研究では絶対的な動作を生成するために、時系列対応 AAM で作成したベクトルからロボットの関節角へ変更した時系列対応ロボット関節情報も用いる。特に本研究ではロボットの右腕を対象とし、以下、これを「時系列対応 RHJ」と呼ぶ。これにより時系列対応 AAM を用いた場合は、入力次元数を 50、出力次元数を 8 とし、中間層を 13 とする NN を構築した。また、時系列対応 RHJ を用いた場合は、入力次元数を 50、出力次元数を 21 とし、中間層を 13 とする NN を構築した。

訓練データとして、図 5 の様に表現される 10 個の言語表現 (大きい, 厚い, 太い, 小さい, 薄い, 細い, ザクザク, ざっくり, しっかり, きちんと) に対して各 100 個の計 1000 個を与えた。なお、各 100 個の訓練データは切断正規分布によって作成した。

学習の後、評価データとして訓練データとして用いた 10 個の言語表現を入力したところ、表 3 に示す結果になった。なお、訓練データとして与えた動作と、評価結果として出てきた動作をそれぞれロボットに動作させた。その時のずれ (単位: cm) の平均二乗誤差を

表 3: 評価結果の比較

言語表現 動作表現空間	誤差									
	大きい	厚い	太い	小さい	薄い	細い	ザクザク	ざっくり	しっかり	きちんと
時系列対応 AAM	0.010	0.003	0.007	0.024	0.010	0.009	0.010	0.003	0.003	0.003
時系列対応 RHJ	1.952	1.241	0.751	0.887	0.923	1.278	6.683	10.18	10.27	10.54

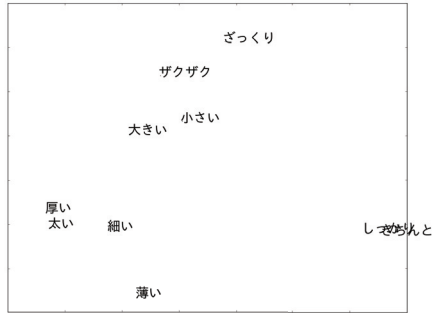


図 5: 訓練データとして与える言語の分散表現

「誤差」としている．時系列対応 AAM が時系列対応 RHJ に比べ、良く学習出来ていることが分かる．しかし、実際にロボットに動作生成させると、図 6 に示すようになり時系列対応 RHJ でも特徴は捉えられていると言える．

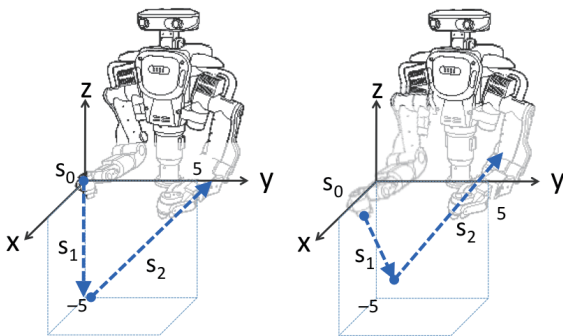


図 6: 「大きい」に対して時系列対応 AAM を用いた場合 (左) と時系列対応 RHJ を用いた場合 (右)

また、未知の言語表現として「速い」「ゆっくり」「粗い」「細かい」を入力とした時、表 4 に示す結果になった．未知語の動作生成に関しても時系列対応 AAM の方が生成が来ている．

既知語、未知語両方の動作生成において、時系列 AAM より時系列対応 RHJ の誤差が大きくなる原因としては、動作の生成に腕の関節情報を与えているた

表 4: 未知語の動作生成比較

言語表現 動作表現空間	誤差			
	速い	ゆっくり	粗い	細かい
時系列対応 AAM	2.609	2.682	0.093	2.114
時系列対応 RHJ	6.014	7.119	3.063	5.526

め、肩における少しの角度のずれが指先の大きなずれに繋がってしまうことが考えられる．動作生成において、切る動作の特徴は捉えられているため、関節情報の渡し方を工夫する方法があると言える．

6 まとめ

本研究では、ロボットの動作生成を行うため調理動作生成を例に動作生成に対応するため、Cheng ら [1] による Activity-Attribute Matrix を参考にして動作要素の時系列変化を捉える関係行列を作成した．また、言語と動作の対応関係を取るため、言語は word2vec を使った分散意味表現をとり、それと作成した時系列対応 AAM と時系列 RHJ の動作のベクトルによるニューラルネットワークを用いた学習を行った．今後の課題としては、言語と動作の対応関係をより明確にし、さらに多種多様な言語から動作生成を可能にするつもりである．

謝辞：本研究ではクックパッドのコーパスを利用させて頂きました．関係各位に感謝申し上げます．

参考文献

- [1] Heng-Tze Cheng, Feng-Tso Sun, Martin Griss, Paul Davis, Jianguo Li, Di You, “NuActiv: Recognizing Unseen New Activities Using Semantic Attribute-Based Learning”, MobiSys’13, Taipei, 2013.
- [2] Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean, “Efficient Estimation of Word Representations in Vector Space”, CoRR, abs/1301.3781, 2013
- [3] 濱園侑美, 小林一郎, 麻生英樹, 長井隆行, 中村友昭, 持橋 大地, “ヒューマノイドロボットの動作生成に向けた基礎的検討”, 人工知能学会 29 回全国大会, 2015