曖昧言語表現に対するロボット動作の対応関係学習

濵園 侑美 † 小林 一郎 † 麻生 英樹 ‡
中村友昭 § 長井隆行 § 持橋 大地 ¶
† お茶の水女子大学大学院 ‡ 産業総合技術研究所 § 電気通信大学大学院 ¶ 統計数理研究所

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

1 はじめに

日本は超高齢化社会に突入しており，人手不足をロボットを利用することで解決をはかる場面が増えると考えられる．家庭内でロボットを用いる場合，ロボットが居住者と協調して暮らせる条件として，言葉や身振りを用いることで居住者の経験をロボットに伝え，ロボットはそれを学習することが必要になると考える．

このことを踏まえ，人の言葉による指示からロボットが動作を行なうことを目標に，言葉と動作の対応関係を学習することによって，初めて行な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している．本研究では先行研究 [1] に対して，多様なロボット動作と曖昧な表現との対応関係をより正しく学習出来る枠組みを検証する．

2 提案手法概要

動作に関する言葉と，動作を変化させる曖昧表現の意味と動作表現の対応関係が既知であるとする．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．

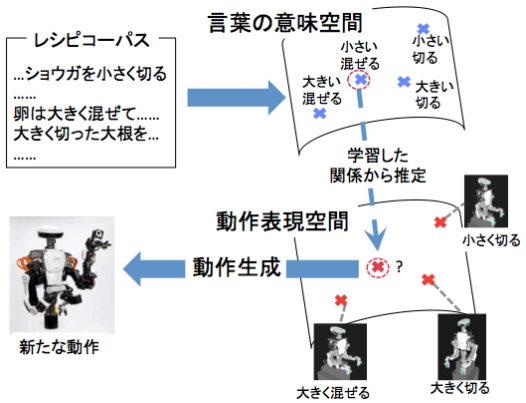

図 1: 提案手法の概要

言葉は，動作に関する言葉と，動作を変化させる曖昧表現の言葉のそれぞれにラベルをつけ，それらを組み合わせて表現する．言葉を意味空間へ配置する方法は word2vec[2] を用い，動作を表現空間へ配置する方法は，先行研究 [1] で提案した時系列対応 AAM を用いる．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．

3 ロボット動作

3.1 ヒューマノイドロボットの概要

（株）川田工業社製ヒューマノイドロボット HIRONXC を用いる．HIRONXC は全 24 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 秒かけて関節を指定された角度へと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ために，時系列対応 AAM[1] を用いる．これは Cheng ら [3] による，動作と動作に関連している意味属性を符号化した Activity-Attribute Matrix(以下 AAM) を参考にして作成したものである．AAM の具体例を図 2 で示す．

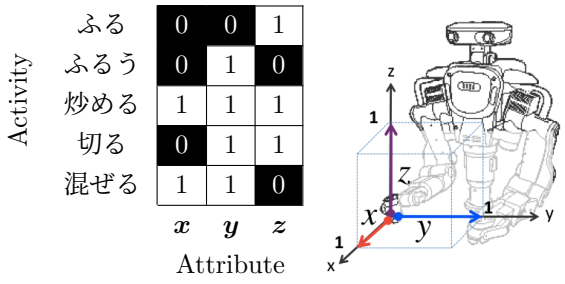

図 2: HIRONXC を用いた AAM 作成

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 Attribute j は 0 とする．図 2 において，Attribute は右手の $\boldsymbol{x}$（前後），$\boldsymbol{y}$（左右），$\boldsymbol{z}$（上下）の動きであり，例えば Activity の「ふる」は，Attribute の $\boldsymbol{z}$ が含まれる動き，つまり右手が上下方向に動く動きであることを示している．

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，$\boldsymbol{x}$，$\boldsymbol{y}$，$\boldsymbol{z}$ の変動割合に時間の要素 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 $[[p_{x_1}, p_{y_1}, p_{z_1}, t_1], [p_{x_2}, p_{y_2}, p_{z_2}, t_2], \cdots, [p_{x_n}, p_{y_n}, p_{z_n}, t_n]]$ を与えることにより動作の生成を可能にする．なお，時間要素 t は本研究では速度とし，$[0, 1]$ で表した．また，ロボットを実際に動かすにはそれぞれの関節角を指定する必要がある．そこで，Attribute の $\boldsymbol{x}$，$\boldsymbol{y}$，$\boldsymbol{z}$ それぞれに対して，係数となるベクトルとの内積をとることにより，ロボットの動作生成を可能にする．表 1 に本研究で提案する時系列対応 AAM を示す．

表 1: 時系列対応 AAM の概要

	s_1				s_2				…	s_n			
速く切る	0	0	-5	0.8	0	3.5	5	0.8		0	3.5	5	0.8
速く混ぜる	0	3.5	0	0.8	3.5	-3.5	0	0.8		-3.5	0	0	0.8
ゆっくり切る	0	0	-5	0.3	0	3.5	5	0.3		0	3.5	5	0.3
細かく切る	0	0	-5	0.5	0	1	5	0.5		0	1	5	0.5
ザクザク混ぜる	0	6	0	0.8	6	-6	0	0.8		-6	0	0	0.8
	p_{x_1}	p_{y_1}	p_{z_1}	t_1	p_{x_2}	p_{y_2}	p_{z_2}	t_2	…	p_{x_n}	p_{y_n}	p_{z_n}	t_n
	↑	↑	↑		↑	↑	↑			↑	↑	↑	
	$\boldsymbol{x}$	$\boldsymbol{y}$	$\boldsymbol{z}$		$\boldsymbol{x}$	$\boldsymbol{y}$	$\boldsymbol{z}$			$\boldsymbol{x}$	$\boldsymbol{y}$	$\boldsymbol{z}$	
RSY	0.1	1.8	0.7		0.1	1.8	0.7			0.1	1.8	0.7	
RSP	-2.3	0.7	0.1		-2.3	0.7	0.1			-2.3	0.7	0.1	
REP	2.1	-0.5	-2.7		2.1	-0.5	-2.7			2.1	-0.5	-2.7	
RWY	0.0	0.0	0.1		0.0	0.0	0.1			0.0	0.0	0.1	
RWP	0.2	0.2	2.7		0.2	0.2	2.7			0.2	0.2	2.7	
RWR	0.0	-1.8	0.0		0.0	-1.8	0.0			0.0	-1.8	0.0	

4 言葉の分散意味表現

Mikolov ら [2] によって提案された word2vec は，単語をベクトルで表現し，同じ文脈の中にある単語はお互いに近い意味を持つように単語をベクトル化して表現する定量化手法である．本研究ではこれを利用し，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．コーパスはクックパッド [1] 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にした．

[1]http://www.nii.ac.jp/dsc/idr/cookpad/cookpad.html

曖昧表現の word2vec による言葉の分散意味表現を主成分分析し可視化した結果が図 3 である．

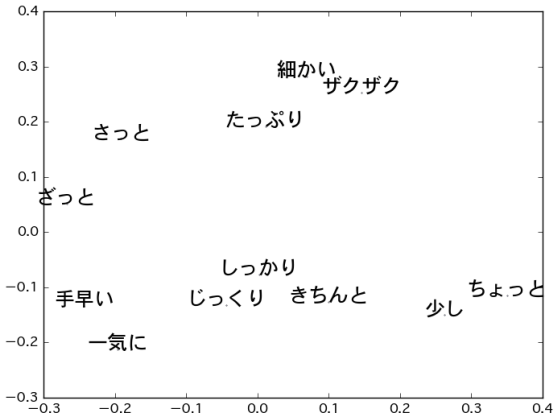

図 3: word2vec による分散意味表現

5 言葉と動作の関係学習

言葉と動作の関係を学習するために，ニューラルネットワーク (NN) を用いる，ネットワークは図 4 に示した Net1, Net2, Net3 の 3 種類を先行研究で用いたネットワークと比較した．なお，Net1, Net2, Net3 の中間層での活性化関数はソフトプラス関数とする．

先行研究： 曖昧表現と動作表現を一つのベクトルとして入力し，ロボット動作を出す 3 層のニューラルネットワーク．中間層での活性化関数はシグモイド関数を用いる．

Net1： 曖昧表現と動作表現からそれぞれの中間層を経た後足し合わせ，ロボット動作を出すネットワーク．

Net2： 曖昧表現と動作表現のそれぞれの中間層を足し合わせ，中間層を経た後ロボット動作を出すネットワーク．

Net3： 動作表現によって曖昧表現が変化すると仮定する．まず，曖昧表現と動作表現のそれぞれの中間層を足し合わせ，出力された動作により変化した曖昧表現の中間層と，動作表現の中間層を足し合わせ，中間層を経た後ロボット動作を出すネットワーク．

6 実験

ネットワークの学習方法は誤差逆伝播法を用い，またロボット動作への出力では活性化関数にシグモイド

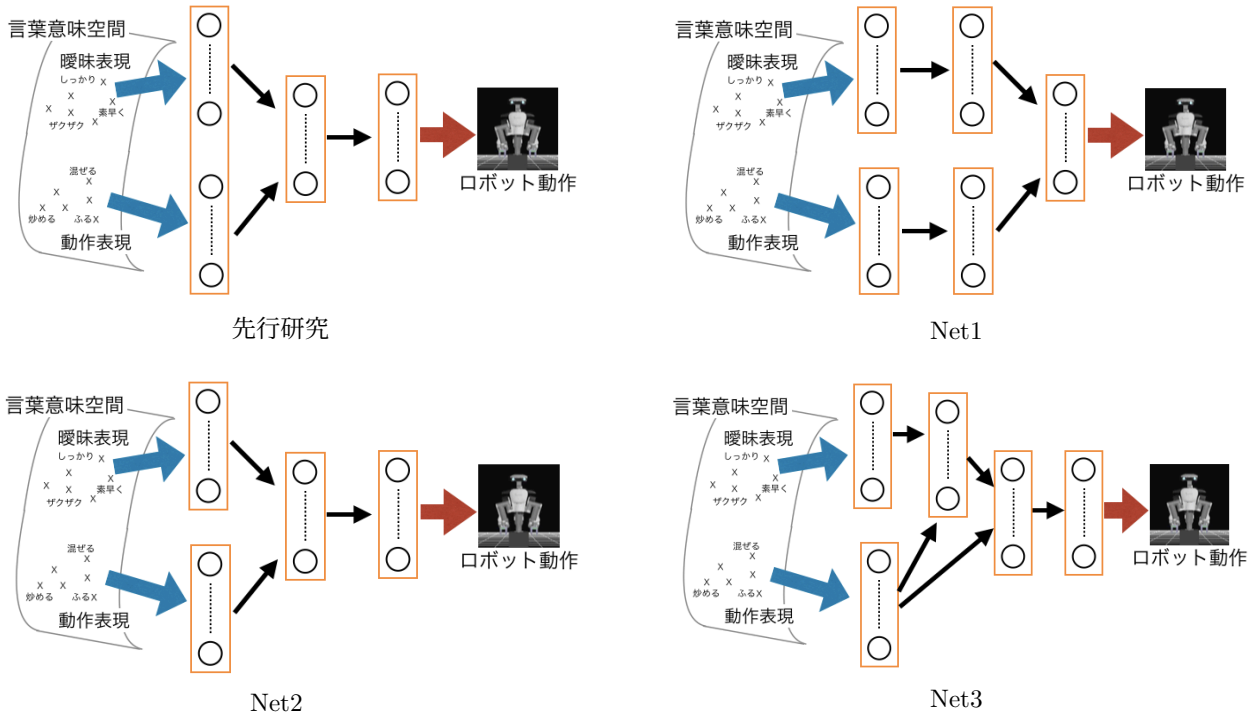

図 4: Network の概要

関数を用いた．言葉は word2vec のうち skip-gram を利用し，1 単語を 100 次元の分散意味表現で表した．動作は様々なパターンを 3.2 節に示した時系列対応 AAM にて作成し，表 2 に示す，時刻 s_1 から s_6 の 6 つの動作を合わせて 24 次元のベクトルで表した．

表 2: 動作の時系列対応 AAM

動作	s_1				s_2				…	s_6			
	p_{x_1}	p_{y_1}	p_{z_1}	t_1	p_{x_2}	p_{y_2}	p_{z_2}	t_2		p_{x_6}	p_{y_6}	p_{z_6}	t_6
ふる	0	0	3.5	0.5	0	0	-3.5	0.5		0	0	3.5	0.5
ふるう	0	3.5	0	0.5	0	-3.5	0	0.5		0	-3.5	0	0.5
炒める	3.5	0	0	0.5	-3.5	0	3	0.5		0	0	-3	0.5
切る	0	0	-5	0.5	0	3.5	5	0.5		0	3.5	5	0.5
混ぜる	0	3.5	0	0.5	3.5	-3.5	0	0.5		-3.5	0	0	0.5
つぶす	0	0	-3.5	0.5	0	0	3.5	0.5		0	0	3.5	0.5

表 2 で示した動作のうち「切る」と「混ぜる」は図 5 のように表現出来る．

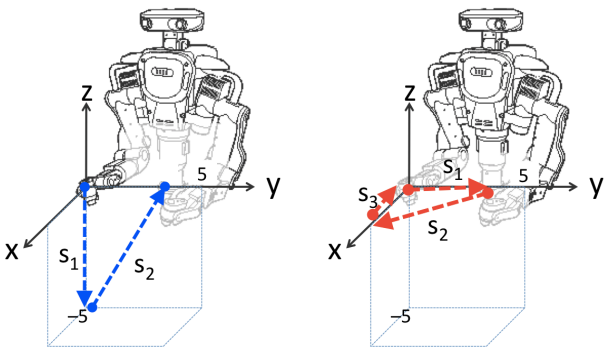

図 5: ロボット動作（左：切る，右：混ぜる）

「切る」の動きを例にとると，s_1 の動きは，s_1=$[p_{x_1}, p_{y_1}, p_{z_1}, t_1]$ = $[0, 0, -5, 0.5]$ 次の s_2 の動きは，s_2=$[p_{x_2}, p_{y_2}, p_{z_2}, t_2]$ = $[0, 5, -3.5, 0.5]$ で表されており，確かに図 5 左でも s_1 で $\boldsymbol{z}$ へ，s_2 で $\boldsymbol{y}$ と $\boldsymbol{z}$ への動きを同時に行っている．なお，曖昧表現による動作の程度は，表 2 の着色部や速さを表す t_n を変化させることで変化させる．

これにより，入力層のノード数は Net1，Net2，Net3 では 100，出力層のノード数を 24 とし，また，中間層のノード数は 9 から 30 を試行し，最も平均二乗誤差が小さくなった 20 とする．訓練データとして，6 個の動作表現と，12 個の曖昧表現のうちいずれか 7 個をそれぞれ組み合わせた全 42 種に対し各 100 個の全 4200 個を与えた．なお，訓練データは切断正規分布により作成した．学習をした後，評価データは訓練データとして与えた動作表現と曖昧表現の組み合わせの 42 種，未知の言語表現として訓練データに用いた動作表現と曖昧表現の組み合わせ以外の 30 種を入力とした．

表 3，表 4 は予測し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，s_1 から s_6 での $\boldsymbol{x}$，$\boldsymbol{y}$，$\boldsymbol{z}$ 軸に関するそれぞれの誤差 (単位：cm) と，速さの誤差の平均二乗誤差を「誤差」として表した．表 3 はネットワーク毎の誤差の平均を，表 4 は全体の誤差平均が最も小さい Net3 について，それぞれの動作についての平均二乗誤差を表しており，着色部は未知の言語表現を与えた場合である．

表 3: ネットワーク比較

Network	誤差		
	全体	既知	未知
先行研究	2.6951	2.6543	2.7522
Net1	2.5506	2.4740	2.6580
Net2	2.2794	2.2499	2.3208
Net3	2.2756	2.2456	2.3177

表 4: Net3 評価結果

誤差		動作表現					
		ふる	ふるう	炒める	切る	混ぜる	つぶす
曖昧表現	さっと	2.63	2.63	2.63	2.63	2.62	2.63
	ざっと	3.08	3.08	3.09	3.08	3.07	3.08
	ザクザク	3.53	3.53	3.53	3.54	3.54	3.54
	たっぷり	2.23	2.23	2.23	2.23	2.23	2.23
	手早い	2.63	2.63	2.63	2.63	2.62	2.63
	一気に	3.08	3.08	3.08	3.08	3.07	3.08
	少し	2.24	2.26	2.24	2.23	2.24	2.25
	細かい	2.24	2.24	2.24	2.23	2.24	2.23
	ちょっと	2.23	2.23	2.24	2.23	2.23	2.23
	しっかり	0.94	0.92	0.91	0.91	0.92	0.91
	きちんと	1.14	1.14	1.14	1.13	1.15	1.14
	じっくり	1.35	1.35	1.36	1.35	1.36	1.36
平均		2.28	2.28	2.28	2.27	2.27	2.28

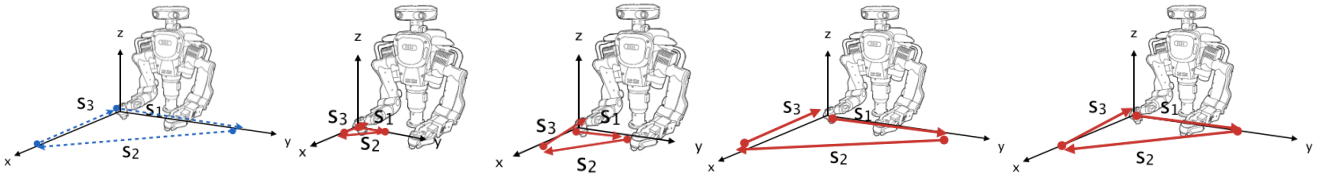

図 6: 「ザクザク」「混ぜる」のロボット動作

7 考察

先行研究 [1] と比較し，提案したネットワーク全てにおいて誤差の平均が小さくなった．また，Net2，Net3 が Net1 よりも誤差の平均は小さくなっており，動作表現と曖昧表現の言葉を足し合わせた後，ロボット動作を出力する前に一度中間層を経たネットワーク構成の学習が有用であると言える．また Net3 は全ての誤差の平均が最小であり，動作表現によって曖昧表現が変化するという仮定が正しいと言えるであろう．

表 4 より，動作表現毎の誤差の平均がほとんど変わらないことから，動作の特徴を捉えることができていると言える．また，1 つの曖昧表現における動作表現毎の誤差について，白色部と着色部に大きな差が見られないことより，曖昧表現が動作毎に変化する学習ができていると言える．

Net3 において最も誤差の大きい「ザクザク」「混ぜる」の全てのネットワークの s_1 から s_3 までの動作は図 6 のようになり，提案した 3 つの手法は「混ぜる」動きができており，Net3 は予想データとより似た動きになることが確認出来る．

8 おわりに

本研究では，ロボットが言葉と動作の対応関係を学習することにより，曖昧表現の動作を行えるようにすることを目指した．ロボット動作と言葉の対応関係を学習する枠組みとして 3 種類のネットワークについて，先行研究 [1] との比較・検証を行なった．提案した 3 種類のネットワークは先行研究の結果を上回ったことを確認した．

今後の課題は，他の未知語を入力として与えた場合の動作の変化や，さらに多種多様な言語表現や複雑なロボット動作の関係学習を行ないたい．

参考文献

[1] 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, “語彙の分散意味表現とロボット動作との対応関係の学習”, 第 30 回人工知能学会全国大会, 2016.

[2] Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean, “Efficient Estimation of Word Representations in Vector Space”, International Conference on Learning Representations, 2013.

[3] Heng-Tze Cheng, Feng-Tso Sun, Martin Griss, Paul Davis, Jianguo Li, Di You, “NuActiv: Recognizing Unseen New Activities Using Semantic Attribute-Based Learning”, Mobile Systems, Applications, and Services, 2013.