

運動記号と単語の統計的相互参照関係に基づく 言語の位相構造形成

Formation of Topological Linguistic Space based on Stochastic Reference between Motion Symbols and Words

高野 渉 (東大) クリッチ・ダナ (東大) 中村仁彦 (東大)

*Wataru TAKANO (Univ. of Tokyo)

Dana KULIC (Univ. of Tokyo)

Yoshihiko NAKAMURA (Univ. of Tokyo)

We propose a hierarchical model incorporating motion time series data, motion symbols and words. This paper describes a linguistic space which represents a network of the words linked to full body motions. The linguistic space is constructed by using the dissimilarity among words, which can be computed from the association probability of the words and motion symbols. In the linguistic space, words that are semantically similar are located close to one another and included in the same cluster. We validate our approach by constructing a motion symbol space based on the dissimilarity between words, which improves discrimination ability when compared to a motion space constructed based on dissimilarities between motions alone.

Key Words :Symbolization/Recognition/Generation of Motion Patterns, Linguistic Space, Hidden Markov Model

1. はじめに

環境と接するロボットの身体運動を基盤とした知能の研究が数多くなされている。近年、他者の行動を観察した時および自身で同じ運動を行なった時に発火するミラーニューロンが発見され注目を浴びている [1]。このような脳科学の発見にヒントを得て、様々なロボットの運動の記号化・認識および生成が結び付いたモデルが提案されている [2][3][4][5]。

人間の脳において明らかに他の動物と異なる点は言語を有することである。言語は記号情報処理であり古くから人工知能分野において記号と記号の演算により言語をモデル化する研究がなされてきた。この記号処理では記号演算に重きが置かれていたため、その記号が現実世界に接地しておらずロボットの知能に適用することが難しい問題がある。

また、ミラーニューロンが人間の脳の運動性言語野に相当する部位に近い事実や運動パターンの記号処理が言語の記号処理へと発展するというパラダイムから運動パターンの学習モデルを基礎としたロボットの言語情報処理に関する研究がなされている [6]。

筆者らも、図 1 に示すような運動パターンの記号群と運動パターンを表す単語ラベル群との間の連想関係を統計モデルで表現したロボットの言語情報処理に関する研究を行ってきた [7]。これまでの研究では、運動パターンを表す記号と単語ラベル間の関係性に注目したものであり、単語ラベル間の関係性については議論していない。Deacon は記号の階層構造として記号を「アイコン」、「インデックス」、「シンボル」と分類し、シンボルへ発展するにつれて記号と事象間の因果関係が希薄化し、記号同士の結びつきが強まり、これが強力な記号に基づく推論を可能にすると提唱している [8]。このように単語ラベルと運動記号との結びつきのみならず、単語ラベル同士の連合も重要である。

そこで、本稿では単語ラベル間の非類似度に基づき単語ラベル群の位相構造を表現した単語空間を提案する。さらに、その単語間の関係性が運動記号群の構造変化をもたらす、運動記号群・単語群が全て結び付くような階層構造モデルを構築し、観察した運動の認識処理や運動記号群の分類性能の観点

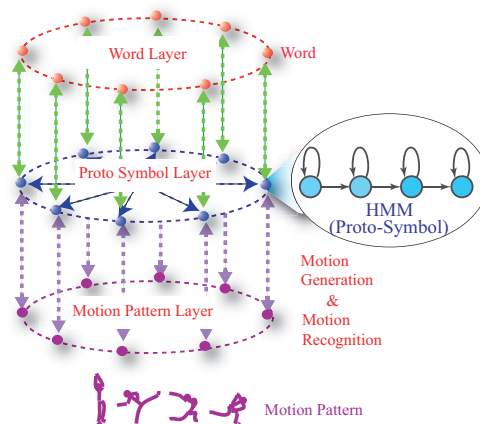


Fig.1 The motion patterns are symbolized by Hidden Markov Models, which are called *proto-symbol*. The proto-symbols are associated with word labels, which are manually given to motion data.

から提案したモデルの有効性を示す。

2. 単語の位相空間と運動記号空間

2-1 単語の位相空間設計

ある運動時系列に対するロボットの運動パターンの記号化モデルに基づく認識結果と人間の単語ラベルに基づく認識結果を確率統計モデルで結び付ける。ロボットはある運動 $O^{(i)}$ を観察し、それを運動パターン毎 $o^{(i)}[k]$ に分節化する。分節化した運動パターンを HMM (Hidden Markov Model) $\lambda_{\mathcal{R}}^{(i)}[j]$ として認識する。

$$O^{(i)} = \{o^{(i)}[1], o^{(i)}[2], \dots, o^{(i)}[l]\} \quad (1)$$

$$\Lambda^{(i)} = \{\lambda_{\mathcal{R}}^{(i)}[1], \lambda_{\mathcal{R}}^{(i)}[2], \dots, \lambda_{\mathcal{R}}^{(i)}[l]\} \quad (2)$$

$$\lambda_{\mathcal{R}}^{(i)}[k] = \arg \max_{\lambda_j} P(o^{(i)}[k] | \lambda_j) \quad (3)$$

ここで、ロボットは予め観察した運動を自律的に運動パターンに分節化し、運動パターンの特徴を HMM(λ) のパラメータとして記憶している [9]。この HMM を運動パターンを記号化した原始シンボルと呼ぶ。人間も同じ運動時系列を観察したとき、主観の評価に基づき運動を分節化し、運動を表す単語ラベル ω として離散的に認識する。

$$\Omega^{(i)} = \{\omega^{(i)}[1], \omega^{(i)}[2], \dots, \omega^{(i)}[m]\} \quad (4)$$

ただし、ロボットと人間の分節基準が異なるため、同じ運動時系列から抽出される分節化された運動パターンの数は必ずしも一致するとは限らないものとする ($l \neq m$)。

観察された運動時系列 $O^{(i)}$ 、ロボットによるその運動時系列の原始シンボルに基づく認識結果 $\Lambda^{(i)}$ および人間による単語ラベルに基づく認識結果 $\Omega^{(i)}$ の三項関係から単語ラベル ω から原始シンボル λ が連想される想起確率 $t(\lambda|\omega)$ を求める。

求められた想起確率を用いて単語間の非類似度を定義する。非類似度として式 (5) で示される Kullback Leibler 情報量を用いる。

$$d(\omega_i, \omega_j) = \sum_{k=1}^{N_\lambda} t(\lambda_k|\omega_i) \{\ln t(\lambda_k|\omega_i) - \ln t(\lambda_k|\omega_j)\}$$

$$D(\omega_i, \omega_j) = \frac{d(\omega_i, \omega_j) + d(\omega_j, \omega_i)}{2} \quad (5)$$

ここで、 N_λ は原始シンボルの数を表す。

この非類似度を単語ラベル間の距離とみなし、この距離を満たすように単語ラベルを多次元空間上の点として布置することによって単語空間を構築する。単語ラベルの空間上への布置は多次元尺度法を用いる。

2.2 単語空間による運動認識

観察された運動時系列は分節化されて原始シンボル列として認識される。この原始シンボルを単語空間に射影することによって観察された運動を単語空間に基づいて認識することができる。原始シンボル $\hat{\lambda}$ を単語空間に射影することは、擬似単語ラベル $\hat{\omega}$ を単語空間に布置することに相当する。擬似単語ラベル $\hat{\omega}$ は原始シンボル $\hat{\lambda}$ を必ず生成するものとする。

$$t(\lambda_i|\hat{\omega}) = \begin{cases} 0, & \lambda_i \neq \hat{\lambda} \\ 1, & \lambda_i = \hat{\lambda} \end{cases} \quad (6)$$

上式の想起確率に基づいて擬似単語ラベルと単語ラベルの距離を求めることができる。この距離を満たすように擬似単語ラベルを単語空間上に布置することによって、観察された運動を単語空間に基づいて認識することができる。

2.3 単語空間と運動データの双方向に基づく運動記号空間の再構築

これまでの原始シンボル空間は、式 (7) で表されるように原始シンボルから生成される運動時系列に基づき計算される非類似度に対称性を考慮した距離測度を用いて原始シンボルを多次元空間上に配置することによって構築されていた。

$$d(\lambda_i, \lambda_j) = \ln P(O_{G,\lambda_i}|\lambda_i) - \ln P(O_{G,\lambda_i}|\lambda_j) \quad (7)$$

ここで、 O_{G,λ_i} は原始シンボル λ_i から生成される運動時系列である。

本稿では、原始シンボルから生成される運動時系列および単語ラベルから計算される原始シンボル間の非類似度に基づ

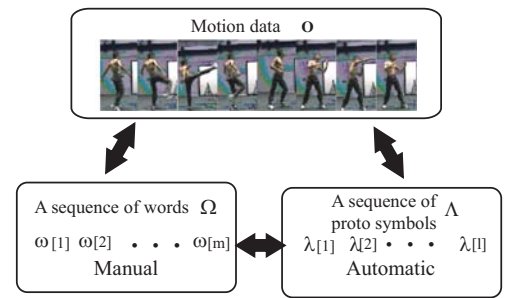


Fig.2 A robot recognizes observed motion as a sequence of proto-symbols. A human also recognizes the motion as a sequence of word labels. The proto-symbols are associated with the word labels through the motion by interaction between the robot and the human.

き原始シンボル空間を構築する。式 (8) に示される原始シンボル間の非類似度に対称性を考慮した距離測度として定義する。

$$\tilde{d}(\lambda_i, \lambda_j) = \ln P(O_{G,i}, \omega_{G,i}|\lambda_i) - \ln P(O_{G,i}, \omega_{G,i}|\lambda_j) \quad (8)$$

$$P(O_{G,i}, \omega_{G,i}|\lambda_i) = P(O_{G,\lambda_i}|\lambda_i)t(\omega_{G,i}|\lambda_i) \quad (9)$$

原始シンボル λ_i は運動時系列 $O_{G,i}$ と単語ラベル $\omega_{G,i}$ を独立に生成すると仮定する。 $t(\omega|\lambda)$ は原始シンボル λ から単語ラベル ω が想起される確率である。

3. 単語位相空間と運動記号空間の構築

3.1 単語位相空間

34 個のマーカを付けた人間の動きを光学式モーションキャプチャにて計測する。計測されたマーカ位置の時系列データを 20 自由度のロボットモデルを用いて逆運動学により関節角時系列に変換する。本稿では、537 個の運動データ ($O^{(i)} : i = 1, 2, 3, \dots, 537$) を用いた。ロボットはこれらの運動データから 50 個の原始シンボル ($\lambda_i : i = 1, 2, 3, \dots, 50$) を自律的に獲得した。また、人間によって与えられた単語ラベルは 61 個であった ($\omega_i : i = 1, 2, 3, \dots, 61$)。単語ラベル間の非類似度を計算し、10 次元の空間に単語ラベルを布置した。ここで、実際の単語ラベル間の距離と空間上での単語ラベル間の距離の誤差の和は多次元空間の次元数が増加するごとに減少し、10 次元空間で誤差が十分に収束していることを確認した。

図 3 は、10 次元の単語空間を第 1 次元および第 2 次元目の軸が張る平面に投影した図を表している。各点が単語ラベルを表しており、各点の色および形はその点が属するクラスタを表している。クラスタ数を 20 と設定してクラスタ分析は行なわれている。したがって、たとえ図 3 では離れて見える単語ラベル同士でも、同じクラスタに属する場合はそれら単語ラベルは類似していることを意味する。表 1 は単語空間をクラスタ分析した結果を示している。第 6 クラスタには、“raise_right_hand_held”、“raise_right_hand_jumping”、“wave_right_hand” という「右手を挙げる」共通の特徴を有した単語ラベルが含まれる。第 11 クラスタは、“catch_ball”、“catch”、“catch_fly_ball” と「捕球」動作に関する単語ラベルを包含している。このように類似した意味を有する単語ラベルが単語空間上で近接に配置され、単語空間が意味を反映した単語の位相構造を表していることを確認できる。

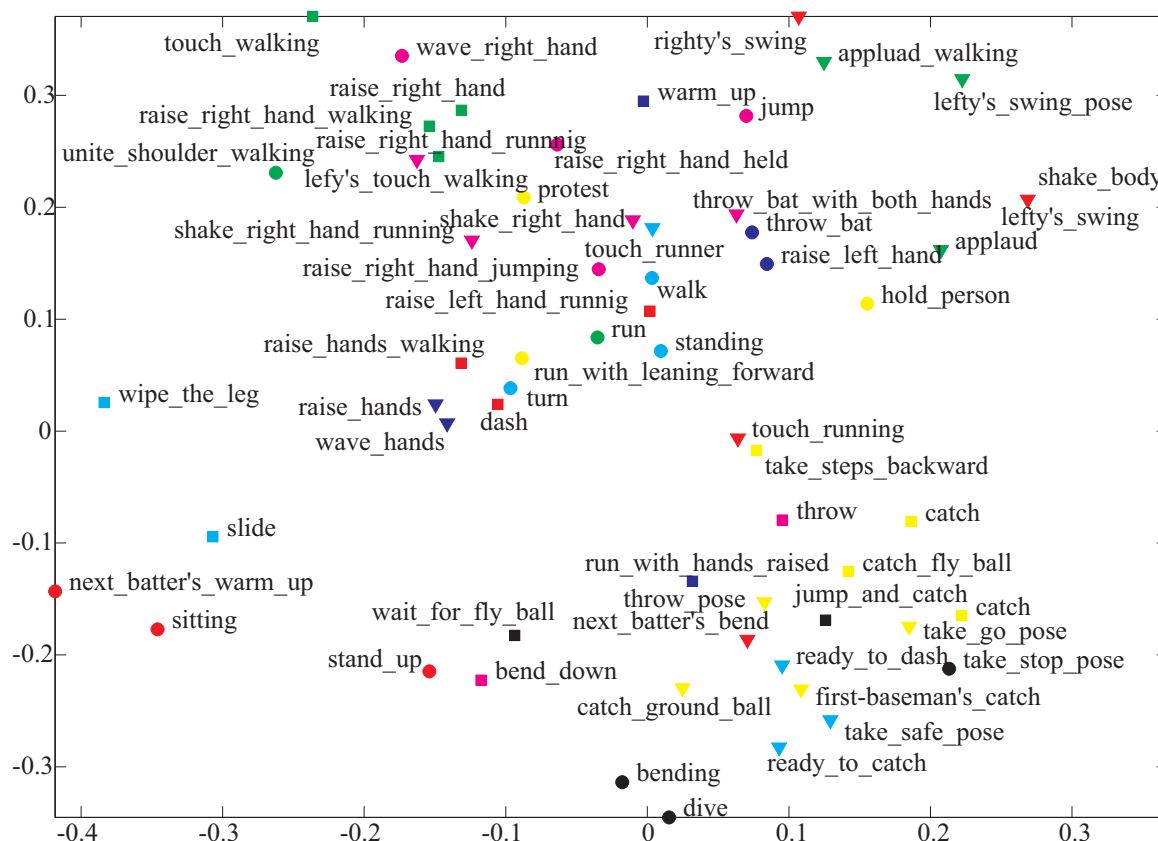


Fig.3 The word space can be constructed based on dissimilarities among the word labels. Plotted objects signify the word labels.

3.2 単語空間による運動認識実験結果

ロボットは人間の運動を観察し、その運動を分節化する。分節化された運動パターンを原始シンボルとして認識し、その原始シンボルに対応する擬似単語ラベルを単語空間に布置する。この一連の情報処理によって観察した運動時系列を単語空間に基づいて認識することができる。図4は、ロボットが「左打者がスウィングしてから走る」運動を観察したときの単語空間に基づいて認識結果である。認識結果は“lefty's_swing_pose, lefty's_swing, turn, run”と単語空間上を遷移している。このことから、ロボットは観察した運動時系列を単語空間に基づいて適切に認識できていることが確認できる。

3.3 運動記号空間

運動時系列と単語ラベルに基づく原始シンボル空間を構築した。図5は運動時系列のみに基づく原始シンボル空間と単語ラベルも用いた場合の原始シンボル空間の比較を示している。運動時系列のみに基づく原始シンボル空間では、原点付近に原始シンボルが密に集まっている。これは原始シンボルが互いに類似しており、分類が難しいことを意味する。一方、運動時系列と単語ラベルを用いて構築した原始シンボル空間では、原始シンボルが空間上に広がった構造が現れる。これは、原始シンボル間の非類似度が増加し分類するに適した空間が構成されたことを示している。このように、単語ラベル情報

が原始シンボルの分類に有益であることを確認した。

4. おわりに

本稿では、単語ラベルと運動パターンの記号モデルとの統計的關係に基づく単語間の距離を定義し、その距離を満たすような単語ラベルの位相空間を構築した。構築した位相空間において類似した意味を持つ単語が互いに近い位置に配置されていることを示し、提案した単語ラベルの位相空間の妥当性を確認した。また、単語ラベル情報を用いた運動パターンの記号間距離を導入し、その距離に基づく運動記号空間を構築した。その運動記号空間を通じて単語ラベル情報が運動記号の分類度を向上するのに有効であることを確認した。

なお、本研究は科学研究費補助金特定領域研究（課題番号：19024024, 研究課題名：膨大な運動データの抽象化・構造化に基づく運動言語の獲得、研究代表者：中村仁彦）および日本学術振興会の支援を受けた。共著者である Dana Kulic は日本学術振興会外国人特別研究員である。また、(株)バンダイナムコゲームスからモーションキャプチャデータの提供を受けた。

文献

- [1] G. Rizzolatti, L. Fogassi, and V. Gallese. Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews*, pages 661–670, 2001.
- [2] M. Haruno, D. Wolpert, and M. Kawato. Mosaic model for sensorimotor learning and control. *Neural Computation*, 13:2201–2220, 2001.

Table 1 Result of clustering analysis

cluster	word
1	stand_up , sitting next_batter's_warm_up
2	raise_left_hand , throw_bat
3	standing, turn , walk
4	run_with_leaning_forward protest, hold_person
5	run,unite_shoulder_walking
6	jump, raise_right_hand_held raise_right_hand_jumping wave_right_hand
7	dive, bending, take_stop_pose
8	run_with_hands_raised, warm_up
9	dash, raise_left_hand_running raise_hands_walking
10	slide, wipe_the_leg
11	catch_ball , take_steps_backward catch , catch_fly_ball
12	touch_walking , raise_right_hand raise_right_hand_running raise_right_hand_walking
13	bend_down , throw
14	jump_and_catch , wait_for_fly_ball
15	lefty's_swing , righty's_swing touch_running , next_batter's_bend
16	raise_hands , wave_hands
17	ready_to_dash , ready_to_catch take_safe_pose , touch_runner
18	first-baseman's_catch ,throw_pose take_go_pose , catch_ground_ball
19	shake_body ,applaud applaud_walking , lefty's_swing_pose
20	lefty's_touch_walking shake_right_hand shake_right_hand_running throw_bat_with_both_hands

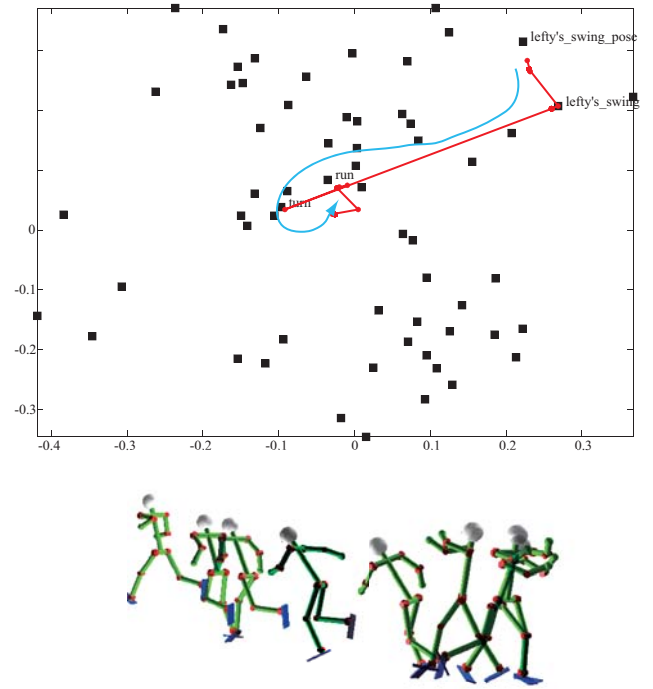


Fig.4

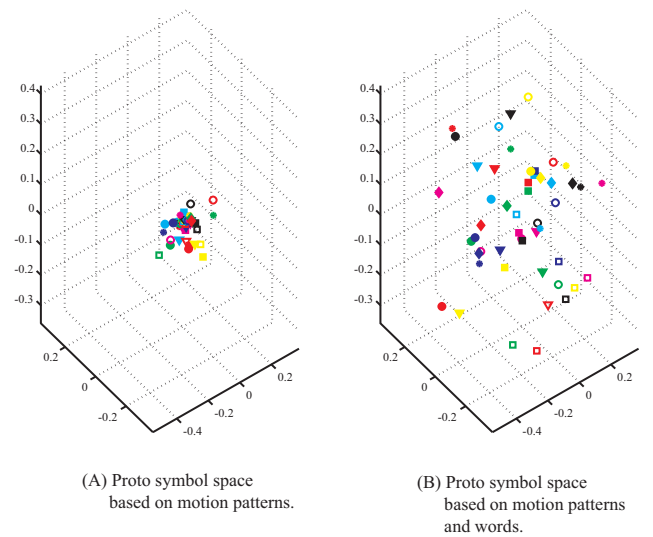


Fig.5 Comparison between two proto-symbol spaces. One is constructed based on motion patterns generated by proto symbols. The other is constructed based on the motion patterns and words generated by the proto symbols.

[3] J. Tani and M. Ito. Self-organization of behavioral primitives as multiple attractor dynamics: A robot experiment. *IEEE Transaction on Systems, Man and Cybernetics Part A: Systems and Humans*, 33(4):481–488, 2003.

[4] A. Billard and R. Siegwart. Robot learning from demonstration. *Robotics and Autonomous Systems*, 47:65–57, 2004.

[5] T. Inamura, I. Toshima, H. Tanie, and Y. Nakamura. Embodied symbol emergence based on mimesis theory. *International Journal of Robotics Research*, 23(4):363–377, 2004.

[6] Y. Sugita and J. Tani. A connectionist approach to learn association between sentences and behavioral patterns of a robot. In *Proc. of*

8th International Conference on Simulation of Adaptive Behavior, pages 467–476, 2004.

[7] W. Takano, K. Yamane, and Y. Nakamura. Capture database through symbolization, recognition and generation of motion patterns. In *Proc. of IEEE International Conference on Robotics and Automation*, pages 3092–3097, 2007.

[8] T. Deacon. Symbolic species : The co-evolution of language and the brain. In *W. W. Norton and Company Inc.*, 1997.

[9] W. Takano and Y. Nakamura. Humanoid robot's autonomous acquisition of proto-symbols through motion segmentation. In *Proc. of IEEE-RAS International Conference on Humanoid Robots*, pages 425–431, 2006.