

# 神経力学モデルによるロボットの言語・運動の統合的認知 Integration of Linguistic and Behavioral Recognition of Robot with Neuro-dynamical system

尾形哲也<sup>\*1\*2</sup>  
Tetsuya Ogata

日下航<sup>\*1</sup>  
Wataru Hinoshita

奥乃博<sup>\*1</sup>  
Hiroshi G. Okuno

<sup>\*1</sup> 京都大学大学院情報学研究科  
Graduate School of Informatics, Kyoto University

<sup>\*2</sup> 科学技術振興機構  
Japan Science and Technology Agency

This paper proposes an approach for robot to acquire language grounding on the robot's sensory-motor flow using neuro-dynamical models. We trained our neuro-model over a set of sentences represented as sequences of characters. For the integrated recognition, we introduced a cognitive hypothesis that human's brain processes "structure" and "contents" of a sentence separately. Our model is trained with a word spell and the semantic role of the word emerged in the first model. As results of binding the model with sensory-motion pattern, we confirmed that our model could associate a proper word spell with a sensori-motor flow and a semantic role, even if the observed flow had not been learned.

## 1. はじめに

本研究の目的は、神経力学モデルを備えたロボットに、自らの感覚-運動系と接地した言語を獲得させることである。人間の様な柔軟なコミュニケーション能力を実現するには、実世界と言語およびそれらの対応関係を、作り込みでなく、ロボット自らが学習によって構造化する枠組みが必要不可欠となる。これは“記号接地問題[Harnad 1990]”と呼ばれ、人工知能における最重要課題の一つとなっている。

本稿では、特に再帰型神経回路モデルを利用した、この問題についての一つのアプローチ法について紹介をしたい。

## 2. 従来研究と問題点

神経力学モデルを用いてロボットの感覚-運動系と言語を統合的に認知する我々と同様のアプローチは、杉田ら[Sugita 2005]、尾形ら[Ogata 2007]の例がある。これらの研究ではパラメータノードを共有する2つのRecurrent Neural Network (RNN)に、それぞれ感覚-運動系フローと単語列を学習させることで相互連想を実現する。学習対象文が2~3単語で構成される非常に単純な場合であればこの枠組は十分に機能するが、ある程度以上の複雑な文章では、この枠組は十分に働かない。具体的には、単語列が長くなると後半の単語は、RNNパラメータ空間でフラクタル階層にコーディングされてしまい、単語の意味を抽出し感覚-運動系フローに反映させることが困難となる事を我々は明らかにしている。

杉田らや尾形らのモデルの問題は、語順などによる文の“構造”と具体的な意味を担う“内容語”を同一のパラメータ空間に埋め込むことに起因している。一方、人間が言語を認知する場合、文の“構造”と“内容語”は脳の別ルートで処理されると考えられている[Dominey 2000]。例えば、“Marie broke window.”という文は、抽象化された要素(格, case)から構成される“構造”([AGENT][ACTION][OBJECT])と、個々の格のスロットに嵌る“内容語”(例. AGENT = “Marie”)に分けて処理されるのである。そこで、この仕組みをモデルに反映することを考えた。

また従来のRNNによる言語獲得モデルのもう一つの問題としては、入力最小単位が単語であるという点が挙げられる。実

際の言語では、文は単語から構成され、さらに単語はより小さな単位(音素・文字)から構成される。そこで本研究では、入力単位を従来から一段引き下げ文字列から、二重分節性の獲得を目指すこととした。

## 3. 提案する言語-運動統合認知モデル

本研究で提案する言語-運動統合認知モデルでは、モータ値の時系列からなるロボット動作パターンと文字系列からなる文の相互連想を行う。モデルは行為パターンを自己組織化する感覚-運動系RNN (Sensory-Motor RNN)と、言語を自己組織化する言語RNN群(Language RNNs)が少数のニューロンを介して相互作用することで、相互連想を実現する。我々のモデルの概略を図1に示す。

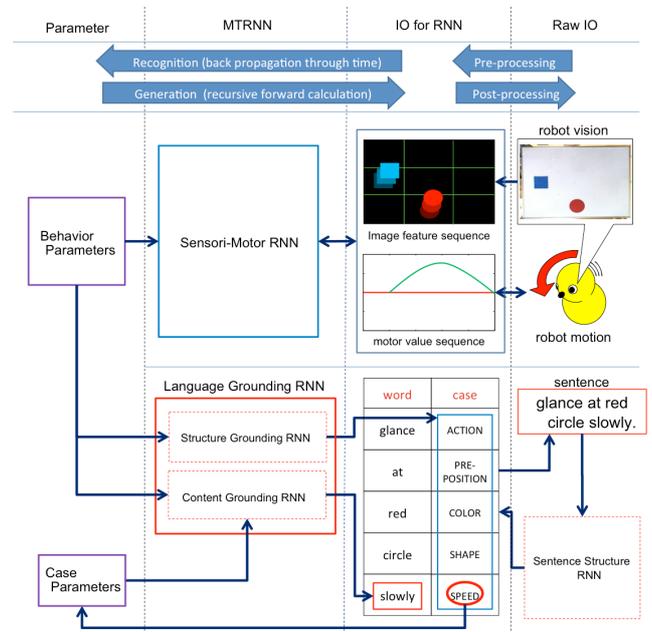


図1 提案する言語-感覚運動統合神経力学モデル

各RNNのブロックはMultiple Timescale RNN (MTRNN) [Yamashita 2008]を利用している。MTRNNのニューロン群は、入出力ノード(IO)と時定数の異なる2種類の文脈ノード(Cf,

Cs), およびパラメータノードの4種類に分類される。IO, Cf, Csの時定数( $\tau$ )はそれぞれ2; 5; 70となっており, 時定数が高いほどニューロン状態の変化が緩やかになる。また, パラメータノードは前向き計算の間は値を一定に保つ。MTRNNは, 学習・認識・生成の3機能を実現する。

(1)学習: IO ノードに教師時系列データを入力し, Back Propagation Through Time (BPTT) によって, 結合重みとパラメータ空間を更新する。

(2)認識: 学習済みモデルの IO ノードに認識したい時系列データを入力し, BPTT によってパラメータノードの値のみを更新する。これにより, 対象データを表現するパラメータを得られる。

(3)生成: パラメータノードに値をセットし, RNN の前向き計算を行うことで, IO ノードの発火状態の時系列データを得る。これが, 与えたパラメータが表現する時系列パターンになっている。

ここで言語 RNN は人間の認知モデルに沿って, 単語品詞種を出力する格 RNN (Structure RNN)と, 単語綴りを出力する内容語 RNN (Content RNN)の2つから構成される。

従来研究[Hinoshita 2011]を基に, 単一 MTRNN を言語(アルファベット列)のみで学習させる。これをそのまま内容語 RNN として用いる。この RNN の Cf ノード発火状態を入力として, 複数 SOM (Self-Organizing Map)で winner-take-all よる学習を行わせる。すると各 SOM に独立した“格(品詞)”が自己組織的に割り当てられる。格 RNN は, アルファベット列に対して, 勝者 SOM 番号列を学習させたものである。

また身体モデル(body model)は, 複数の身体探索(バブリング)のための感覚(視覚)と動作(関節角)パターンデータを与え, 身体モデルの結合重みおよび行為パラメータ(behavior parameter)を更新する。連想時にはこの行為パラメータ空間を介して, 言語-感覚運動の変換が行われる。

#### 4. 言語-運動相互連想実験

小型ロボット Keepon を用いて, 首振り動作とそれを表現する文の相互連想学習を行った。文章は, [ACTION][DIRECTION] or [OBJ] ([SCALE]) の構造を持ち, 全 102 文を用意した。ACTION は首振り動作の種類, DIRECTION は首振りの方向, OBJ はマーカーの色と形状, SCALE は振幅をそれぞれ表現する。SCALE を省略した場合は, “some”と同じ中程度の振幅となる。動作パターンは 48 通り存在する。またマーカーの初期配置は 1080 通りを準備した。よって感覚運動パターンは全てで 48x1080 パターン存在する。その中の 400 の動作パターンを学習に用い, 100 パターンを評価用に用いた。この際, 各動作に対応する文章群(102 文から選択。一動作に複数の文章説明が可能)も用いた。

##### 4.1 結果: 動作から文連想

感覚運動系情報を与えて, 文を連想させた。まず格 RNN における, 非文章(文法的に正しくない文)の生成率は, 既学習データにおいて 0.75% (3/400), 未学習データにおいて 16% (16/100), 全体では 3.8%であった。また内容語 RNN において, 感覚運動情報と格パターンを与えた際の, 単語連想失敗率は既学習データにおいて 0.34% (5/1485), 未学習データで 5.4% (19/355), 全体では 1.3%であった。

全体的に十分な能力を有しているものの, 格 RNN 未学習のパターンへの汎化が比較的難しいことがわかる。これは同一感覚運動状況を表現しうる各シーケンス候補が複数ありうる為で, 何らかの文脈情報による拘束が必要であることを示唆している。

##### 4.2 結果: 文から動作連想

モデルに文と初期視覚情報を与えて, 動作を連想させた。理想的な動作時系列と生成された時系列データの二乗誤差を運動軸とステップ数で平均化したもので, その精度を評価した。その結果, 全体の約 70%程度が, 平均誤差以内にとどまることが確認された。この平均誤差内容の一例をあげると, モデルに “glancelowerleftquickly.” (glance lower left quickly.) という未学習文を与えた場合に, “glance lower left slowly.” に近い動作が生成された。このように第三者からみて “もっともらしい” 文解釈が行われていると判断できる。

#### 5. まとめ

本稿では, 神経力学モデルを備えたロボットによる言語獲得について報告した。本研究の課題は, 単独では意味を成さない文字が二重分節構造を持って文を形成し, 感覚運動系を通して知覚される実世界へと接地されるようなモデルを実現することである。具体的には, 神経力学モデル Multiple Timescale Recurrent Neural Network (MTRNN) を用いて, モータ値および視野画像の時系列からなる行為パターンと文字系列からなる文の相互連想器を学習させた。

一連の実験から, 我々のモデルにおいて単独では意味を成さない文字が二重分節構造を持って文を形成し, 感覚運動系を通して知覚される実世界へと接地される過程が, 限定的ではあるが確認できた。

謝辞: 本研究は, JST さきがけ「情報環境と人」, 科研費学術創成研究(課題番号: 19GS0208), 科研費基盤研究(B)(課題番号: 21300076)の支援を受けた。

#### 参考文献

- [Dominey 2000] Dominey, P. F. and Rumus, F.: Neural network processing of natural language: I. Sensitivity to serial, temporal and abstract structure of language in the infant, *Language and cognitive process*, Vol. 15, pp. 87-127 (2000).
- [Harnad 1990] Harnad, S.: The symbol grounding problem, *Physica D: Nonlinear Phenomena*, Vol. 42, pp. 335-346 (1990).
- [Hinoshita 2011] W. Hinoshita, H. Arie, J. Tani, H. G. Okuno, T. Ogata: Emergence of Hierarchical Structure mirroring Linguistic Composition in a Recurrent Neural Network, *Neural Networks*, Elsevier, available online, Jan. 12. 2011.
- [Ogata 2007] Ogata, T., Murase, M., Tani, J., Komatani, K. and Okuno, H. G.: Two-way Translation of Compound Sentences and Arm Motions by Recurrent Neural Networks, *Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS-2007)*, pp. 1858-1863 (2007).
- [Sugita 2005] 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).
- [Yamashita 2008] Y. Yamashita and J. Tani, “Emergence of Functional Hierarchy in a Multiple Timescale Neural Network Model: a Humanoid Robot Experiment,” *PLoS Comput. Biol.*, vol.4, 2008.