

神経回路モデルによる言語とロボット動作の相互連想学習

日下 航*

尾形 哲也*

高橋 徹*

奥乃 博*

* 京都大学大学院 情報学研究科 知能情報学専攻

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。人間のよう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”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。また、人間の認知機構をモデル化・実装し、その振る舞いを調べることで人間の認知に関する知見を得る“構成的手法”が近年その重要性を増している。人間が言語をいかにして獲得し、実世界と関連付けているかという問題は非常に重要であり、この問いに対する構成的なアプローチという意味でも、脳を模した神経モデルによる言語獲得・言語接地の研究は大きな意義を持っている。

2. 従来研究の問題と認知科学の知見

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モデルは、過去にも杉田ら [1] 及び尾形ら [2] によって提案されている。これら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。そのため、この方式による相互連想を成立させるためには、感覚-運動系フローと単語列を表現するパラメータ空間が共通の構造を持ち得るという条件が必要となる。しかし、学習対象文が2~3単語で構成される非常に単純な場合を除き、この条件の成立は困難である。なぜならば、単語列のように有限状態機械の履歴とみなせる系列をRNNに学習させる場合、一般にそのパラメータ空間はフラクタル構造をとるためである。記憶すべき単語列が長くなれば後半の単語は深いフラクタル階層に埋め込まれ、その単語に対応するパラメータは空間全域に細かく分散する。このため、その単語の意味を抽出し感覚-運動系フローに反映するような構造が組織化されることは望めない。

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 [3]。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。こうして、構造と内容語を分離して認知することで、汎化能力を得ていると考えられる。そのため、この仕組みをモデルに反映することが重要となる。

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。これでは、事前に想定した単語のみしか学習す

ることはできず、扱う単語を増やせば対応する入力ノードがそれに伴って際限なく増えてしまう。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)から構成されている。この性質によって、少数の最小単位の組み合わせから無限とも言える表現を作り出すことができる。これは二重分節と呼ばれ自然言語の主要な性質の一つとされる。そこで、入力単位を従来より一段引き下げ、二重分節性の獲得を目指すことは言語獲得研究において本質的課題となる。

3. 言語-運動統合認知モデル

本稿で提案する言語-運動統合認知モデルは、具体的に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。動作パターン、言語ともにトップダウンのモデル化は行わず、データの汎化によって自己組織的に構造を獲得させることを目指す。モデルは行為パターンを自己組織化する感覚-運動系RNNと、言語を自己組織化する言語RNN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。このモデルでは人間の認知モデルに沿って、言語は格からなる構造と内容語に分離して認知される。今回の実験では簡略化のため、格は語順から一意に決まるという仮定をおいた。

我々のモデルの概略を図1に示す。モデルはMultiple Timescale RNN (MTRNN) [4]を元に行っている。モデル中のニューロン群は、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(τ)はそれぞれ2, 5, 70となっており、時定数が大き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。

モデルは、学習・認識・生成の3つの機能を実現する。

- (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。

3.1 言語獲得手順

- (1) 身体モデル (body model) の学習: 動作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。これにより、行為パラメータ空間が動作パターンの特徴空間として組織化され、言語と対応づける基盤となる。また未学習パターンへの汎化も生じる。

Learning of Bidirectional Associative Memory between Language and Robot's Motion using a Neuro-dynamical System: Wataru Hinoshita, Tetsuya Ogata, Toru Takahashi, and Hiroshi G. Okuno (Kyoto Univ.)

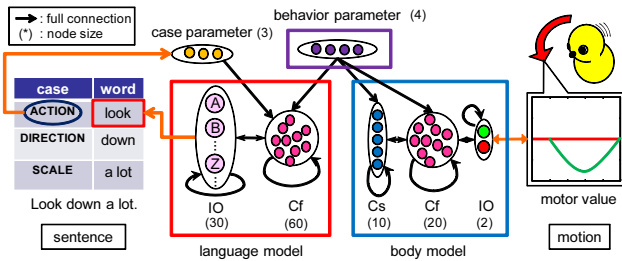


図 1: モデル構成

表 1: Lexicon

case	word
ACTION	look, turn
DIRECTION	right, left, up, down, upper right, upper left, lower right, lower left
SCALE	a bit, some, a lot

(2) 言語モデル (language model) の学習: 指定した行為を表現する文の、指定した格に当て嵌まる単語の綴りを生成できるように、言語モデルの結合重みを更新する。行為は行為パラメータによって、格は格パラメータによって、それぞれ指定される。格パラメータノードはそれぞれが格に対応しており、指定した格に対応するノードのみが発火する。言語モデルの IO ノードはそれぞれが文字に対応しており、それらが順次発火することで綴りを表現する。

3.2 相互連想手順

- (1) 動作 文連想: 身体モデルに動作パターンを認識させ、行為パラメータを得る。得た行為パラメータを用いて、言語モデルにそれぞれの格に当て嵌まる単語の綴りを生成させる。
- (2) 文 動作連想: それぞれの格に当て嵌まる単語の綴りを言語モデルに順番に認識させ、行為パラメータを得る。得た行為パラメータを用いて、身体モデルに動作パターンを生成させる。

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。言語は、[ACTION] [DIRECTION] ([SCALE]) の構造を持ち、表 1 に示す語彙からなる全 64 通りを用意した。ACTION は首振り動作の種類 (顔の向きを戻すかどうか)、DIRECTION は首振りの方向、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。このため、動作パターンは 48 通り存在する。

文とそれに対応する動作の全 64 組のうち、48 組を学習させる (含まれる動作パターンは 40 通り)。その後、動作パターンから文の連想、文から動作パターンの連想を全 64 組に対して行い、能力の評価を行った。

4.1 結果

- (1) 動作 文連想: 全 64 文に含まれる 176 単語に対して、動作パターンと格を与えた時に適切な綴りを生成できるかを評価した。結果を表 2 に示す。また、生成に失敗した 10 単語のうち 7 単語は SCALE であることから、他の特徴に比べ動作の大きさの区別が難しいことが分かる。

表 2: 連想正解率

	既学習	未学習	合計
動作 文	130/132	36/44	166/176
文 動作	43/48	12/16	55/64

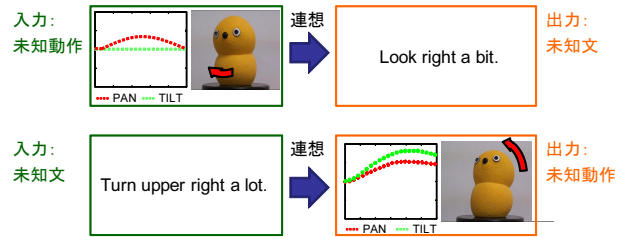


図 2: 未学習動作・文に対する联想結果

- (2) 文 動作連想: 全 64 文に対して、格と単語の綴りの組を与えた時に適切な動作を生成できるかを評価した。結果を表 2 に示す。また、動作パターン自体も未学習である場合の正解数は、5/8 である。

図 2 に联想結果の例を示す。例は、未学習データであり、なお且つ適切な生成を行ったと判定したうちで、最も正解データとの誤差が大きかったものを選択した。この結果から、提案モデルは未学習パターンに対しても適切な相互連想が可能であることが示された。

5. おわりに

本稿では、神経力学モデルを用いたロボットによる言語とロボット動作の相互連想について報告した。実験では、モデルは動作パターン、単語の綴り、格と動作の関係性を構造化し、未学習パターンに対しても適切な相互連想が可能であることを示した。また、格の情報を分離しないモデルでは学習が失敗したことから、構造と内容を分けて認知することの重要性が確認された。

今後の課題として、格の自己組織化がある。今回の実験では、格は語順から一意に決まるという仮定をおいたが、現実にはその仮定は成り立たない。そこで、文の構造を自己組織化し、単語を格によってクラス分けするネットワークの学習が必要となる。これは、文字系列で表現される文の集合を学習させた MTRNN の Cf に対し、som (self-organizing map) を作用させることで実現できることが確認されている。

また、感覚系との統合も重要な課題である。今回の実験では、感覚系の入力を扱わなかったため、状況に関わらず文の解釈は一定であった。現在は視覚と運動を統合した学習を行うことで、外部状況に応じた文の解釈を行うモデルの実現を目指している。

謝辞 本研究の一部は JST さきがけ、科研費基盤 (B)、科研費学術創成の支援を受けた。

参考文献

- [1] Sugita, Y. et al.: “Learning semantic combinatoriality from the interaction between linguistic and behavioral processes,” *Adaptive Behavior*, Vol.13, No.1, pp.33-52, 2005.
- [2] Ogata, T., et al.: “Two-way Translation of Compound Sentences and Arm Motions by Recurrent Neural Networks,” In *Proc. of IROS-2007*, pp.1858-1863, 2007.
- [3] Dominey, P. F., et al.: “A neurolinguistic model of grammatical construction processing,” *Cognitive Neuroscience*, vol.18, no.12, pp.2088-2107, 2006.
- [4] Yamashita, Y. et al.: “Emergence of Functional Hierarchy in a Multiple Timescale Neural Network Model: a Humanoid Robot Experiment,” *PLoS Comput. Biol.*, vol.4, no.11, 2008.