

# MTRNN を用いた単語と文法の階層的自己組織化による文の認識・生成

日下 航\* 有江 浩明† 谷 淳† 尾形 哲也\* 高橋 徹\* 駒谷 和範\* 奥乃 博\*

\* 京都大学大学院 情報学研究科 知能情報学専攻 † 理化学研究所 脳科学総合研究センター

## 1. はじめに

人間の言語獲得機構を明らかにするため、神経回路網モデルによる言語学習の研究が盛んに行われてきた [1]. Elman [2] らは, Recurrent Neural Network (RNN) を時系列予測器として用いることで, 文集合のみから品詞などの文法構造が自己組織化されることを示した. しかし, このモデルは文脈に則して次の単語候補を正しく予測することはできるが, 決定論的に特定の文を生み出す生成能力を持たない. また各入出力ノードが単語に対応しているため, 既知の語彙からなる文しか学習できなかった. 語彙の事前知識なしに文を学習するには, 文字から単語, 単語から文といった階層的な合成能力が必要になる. このような階層的合成能力は言語表現の多様性の中核を担っており, 文の生成能力と共に言語の創造的側面を実現する上で必要不可欠な能力であると言える.

本研究の目的は, 階層的言語構造 (文字 → 単語 → 文) を, 文集合のみを用いて, 神経回路網モデルに自己組織化させ, 未知文の認識・生成を行わせることである. 我々は, 時定数の異なるニューロン群から構成される Multiple Timescale RNN (MTRNN) [3] を用いて学習を行い, この階層性の自己組織化を目指す. 学習済み MTRNN を認識・生成器として用いることで, 獲得された構造を基に, 未知の文が正しく生成されるかを検証する. また, 学習時に誤った文を与えることで, 逆に誤りに対してロバーストになるという仮説を立て, その検証も同時に行う.

## 2. 言語学習モデル

MTRNN は, 現在の状態を入力とし次状態を出力する予測器として用いられ, 複数の非線形時系列パターンを学習・汎化することができる. 我々が実験に用いた言語学習用 MTRNN は A から Z の各アルファベットと, スペース, カンマ, ピリオド, クエスチョンの 4 つの記号に対応する計 30 の入出力ノード (IO) を持ち, 文脈ノードには, Fast Context (Cf, 40 ノード) と Slow Context (Cs, 11 ノード) と呼ばれる 2 種類が存在する (図 1), IO, Cf, Cs の順に時定数 ( $\tau$ ) は大きくなり, ニューロン状態の変化が緩やかになる. このモデルでは, 文章は文字に対応した IO ノードが順次発火することで表現される (図 2).

MTRNN の学習は Back Propagation Through Time (BPTT) によって行われる. また Cs 初期値のベクトル ( $Cs_0$ ) から, 特定のパターンを決定論的に生成することが出来, この  $Cs_0$  のパラメータ空間はデータ間の相関から学習時において自己組織的に獲得される. 逆に結合重みを固定した BPTT によって, 時系列パターンからそのパターンを生成する  $Cs_0$  を得ることもできる (パターンの認識). また, MTRNN の認識フェーズでは, 不安定な入力軌道は別の安定軌道に引き込まれる. これにより, 誤りを含んだ文も正しい文章に訂正することができる.

Recognition and Generation of Sentences through Self-organizing Lexicon and Grammar Hierarchically using MTRNN : Wataru Hinoshita (Kyoto Univ.), Hiroaki Arie, Jun Tani (RIKEN), Tetsumiya Ogata, Toru Takahashi, Kazunori Komatani, and Hiroshi G. Okuno (Kyoto Univ.)

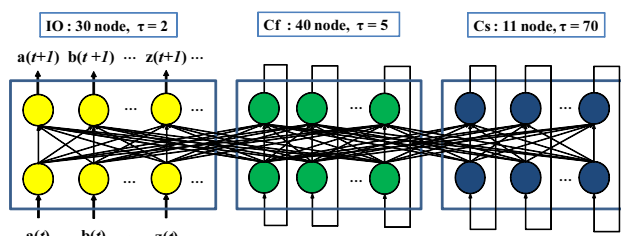


図 1: MTRNN 構成

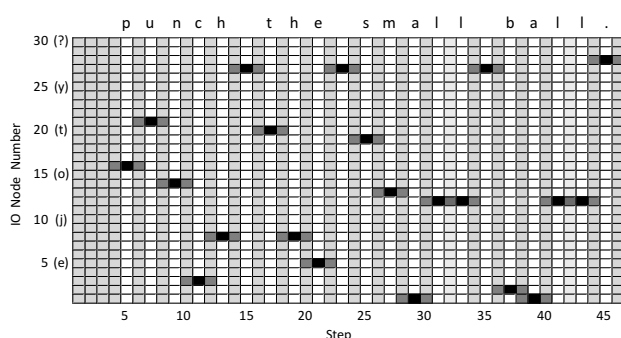


図 2: 入力データ例: “punch the small ball.”

## 3. 言語学習手順

言語の学習は以下の手順で行った.

- (1) 1 つの正規文法から 100 種類の文を生成する.
- (2) そのうち 80 文を MTRNN に学習させる.
- (3) 未学習の 20 文を含む 100 文を用いて, MTRNN の能力を評価する. 評価は以下の手順で行う.
  - (a) 認識: 文を入力し, 重み固定 BPTT により,  $Cs_0$  を計算する.
  - (b) 生成: 得られた  $Cs_0$  から文を生成する.
  - (c) 比較: 元の文と, 生成された文を比較する.

上記の方法で評価した際, もし元の文と生成された文が同じものであったならば, 学習によって獲得された力学系は, その文を表現する安定軌道を持ち, その安定軌道を生成する  $Cs$  初期値が,  $Cs_0$  空間内に埋め込まれていることになる. 実験に用いた文は, 7 つのカテゴリに分類される 17 単語と 9 個のルールからなる正規文法によって生成される 2~6 単語文である.

## 4. 実験

本研究では, MTRNN による言語の獲得実験と, 言語の誤り訂正能力の検証実験の 2 つの実験を行った.

### 4.1 実験 1: 言語獲得実験

3 節に示した手順で, 言語の学習と MTRNN の評価を行った結果, 100 文中 95 文を正しく生成することが確認された. 正解データの内訳は, 既学習文 75/80, 未学習文 20/20 である. 比較実験として, 文法的に誤った語順を持つ文を 20 文認識させたところ, 元の通りに生成された文

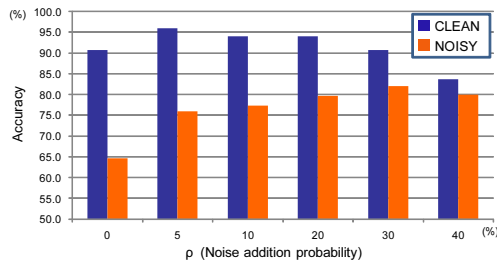


図 3: ノイズ生成率  $\rho$  に対する文の認識・生成精度

は 0 であった。文法的に正しければ未知の文であっても生成可能であることから、MTRNN は学習文を汎化して言語構造を自己組織的に獲得していることが分かる。また既学習にも関わらず正しい文が生成できないことがあるのは、その文を表す軌道の引き込み領域が狭く、うまく  $C_{s_0}$  が見つからないためだと考えられる。

#### 4.2 実験 2: 誤り訂正能力検証実験

1~2 文字が別の文字に置き換えられた不正確な文を MTRNN に認識・生成させ、元の正しい文に訂正する能力を検証した。また、毎学習サイクルにおいて確率  $\rho$  で文に誤りを付与しながら MTRNN の学習を行い、誤り混入確率  $\rho$  が訂正能力に及ばず影響を調べた。実験では、 $\rho = 0, 5, 10, 20, 30, 40(\%)$  の各設定に対して、20 通りの結合重みの初期値から学習・評価を行った。

全体の中で最も優れた学習結果は  $\rho = 30$  の設定で得られ、誤り文を 86/100、正常文を 98/100 という高い精度で正しく認識・生成した。また各  $\rho$  に対して、最も成績の良かった学習結果 3 パターンの平均正解数を図 3 に示す。図 3 より、 $\rho = 30$  までは、学習時の誤り混入率を上げるほど訂正能力が向上し、正しい文に対する認識精度も  $\rho = 0$  と同等かそれ以上を示すことが明らかになった。

以上から、データに誤りを付与して学習することで、得られる軌道の引き込み領域が広がり、力学系の安定性が向上することが明らかになった。

### 5. 解析

学習済み MTRNN が文章を生成する時の、各ニューロン群の発火パターンを解析した結果、IO が文字を、Cf が単語を、Cs が文をそれぞれ表現していることが明らかになった。その論拠を解析結果と共に以下に示す。

- IO: IO ノードはそれぞれが文字に対応しているの、その発火パターンは明らかに文字を表現している。
- Cf: Cf ノードの発火には以下の特徴が見られた。
  - (1) 同一文字に対し、異なる発火パターンを示す。
  - (2) 同一単語に対し、同一の発火パターンを示す。
  - (3) 同一品詞に対し、似た発火パターンを示す。

図 4 に、単語開始時点での Cf 発火パターン (40 次元) の第一、第二主成分を示す。図において、1 文字目が同じであっても異なる発火をしている単語がある (例. “run” と “red”) ことから上記 (1) が確認できる。また、同一単語のクラスタが形成され、さらに品詞による一回り大きなクラスタが形成されていることから、(2), (3) も確認できる。これらの事実から、Cf は自己組織的に獲得した品詞情報を含む形で、単語を表現していることが分かる。

- Cs: Cs ノードの発火には以下の特徴が見られた。
  - (1) 同一単語に対し、異なる発火パターンを示す。

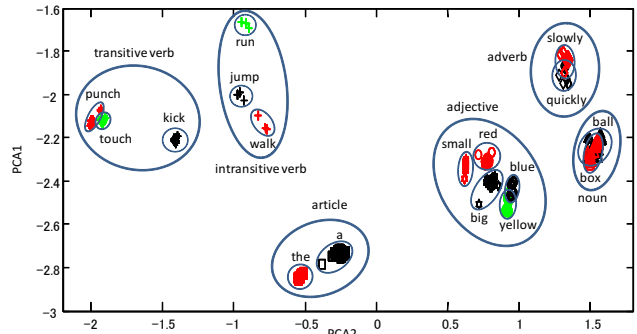


図 4: 単語開始時の Cf 発火: 品詞のクラスタが創発

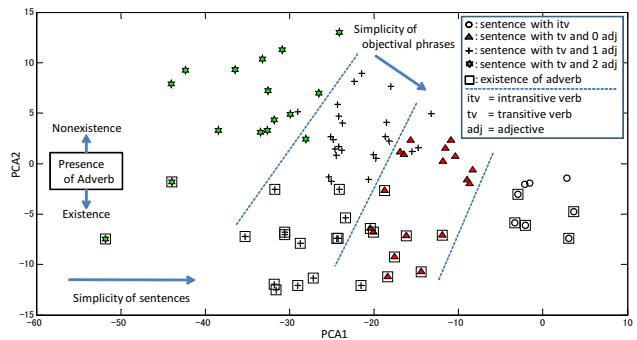


図 5: 文開始時の Cs 発火: 文構造のクラスタが創発

- (2) 同一文に対してのみ、同一発火パターンを示す。
- (3) 類似構造文に対し、似た発火パターンを示す。

図 5 に、文開始時点での Cs 発火パターン (11 次元) の第一、第二主成分を示す。図において、1 単語目と同じ文が多数存在するにも関わらず、全て異なる発火をしていることから上記 (1), (2) が確認できる。また、目的語の有無、目的語を修飾する形容詞の数、副詞の有無といった文構造上の特徴が状態空間に組織化されていることから、(3) が確認できる。これらの事実から、Cs は自己組織的に獲得した文構造の特徴をもとに、文全体を表現していることが分かる。

### 6. おわりに

本稿では、MTRNN によって、文の集合のみから言語構造が階層的に自己組織化され、未知の文でも認識・生成可能であることを示した。また学習時に誤った文を与えることが逆にロバストな構造を生み、誤った文の訂正能力が向上することも確かめられた。

今後、ロボット実験において感覚運動系と言語系の MTRNN が相互作用を行うモデルを構築することで、より認知言語学的な観点からの言語獲得を扱う予定である。

謝辞 本研究の一部は、科研費、GCOE、さきがけの支援を受けた。

### 参考文献

- [1] S. Lawrence and C. L. Giles: “Natural Language Grammatical Inference with Recurrent Neural Networks,” *IEEE Trans. on Knowledge and Data Engineering*, vol.12, no.1, 2000.
- [2] J. L. Elman: “Language as a dynamical system,” In *Mind as Motion: Explorations in the Dynamics Cognition*, MIT Press, pp. 195–223, 1995.
- [3] 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, no.11, 2008.