

# 言語の認知科学第8回配布資料

## 単純再帰型ネットワーク

浅川伸一

2009年11月25日

エルマン [2] の考案した単純再帰型ネットワーク (通称エルマンネット) によって文章の処理が可能である。このことは言語学者にとってインパクトの強い研究であった。その理由は、エルマンの用いたニューラルネットワークでは、いわゆる「刺激の貧困」「否定証拠の欠如」でも文章理解が可能だからである。

「刺激の貧困」とは、プラトンの考えた問題であるとされている。言語の発達過程にある幼児が耳にする言葉は、多くの言い間違いや不完全な文を含んでおり、限りある言語データしか与えられない。それにもかかわらず、どうしてほとんど無限に近い文を発話したり解釈したりできるようになるのか、という問題である。

「否定証拠の欠如」とは、ゴールド [6] によって定式化された。ある文法規則に従う正しい文と、文法規則に従わない誤った文 (否定証拠) を一つずつ与えていき、元の文法を推定することを文法推論という。正しい文法推論とは、与えられた文の列から元の文法を推定するときに、ある有限の数の例から先では、その推定が同じ正しい結果に収束して変わらないということである。文法規則に従う正しい文だけを提示したときは文法推定が不可能である、すなわち獲得できないことをゴールドは証明した。

エルマンのネットワークでは明示的な教師信号による誤り訂正機構を仮定しない。また、言語の記号処理モデルで用いられるような書き換え規則や木構造の構文木を必要とせず文法構造に関する知識が創発する点も強調される。エルマンの示した系列学習の枠組での言語獲得とは、子どもが (大人による) 言語環境に曝されることから引き起こされる学習の結果であると主張される。

エルマンの研究 [3] のめざしたことは (1) 言語的知識の表象とはどのようなものか、(2) 時制の一致や文章の関係代名詞による文章の再帰的構造などの複雑な文法がどのように表象されているのか、(3) ほぼ無限とも思われる単語の組み合わせが限られた資源のネットワークの中でどのように調整されているのか、という疑問に答えることであった。

# 1 エルマンネット

エルマンネットでは、入力層は入力信号を処理する入力ユニットと、直前の時刻までの中間層の状態を入力とする文脈ユニットとで構成される(図1)。文脈ユニットは以前の間層をコピーするためだけ(すなわち中間層から文脈ユニットへの結合強度は1.0)である。結合強度の学習は順方向の結合についてだけ行われるので、通常の誤差逆伝播法がそのまま適用できる。

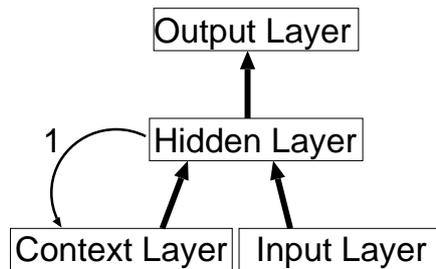


図 1: エルマンネット

ある時刻  $t$  で処理される内容は、その時点での入力信号と、それ以前の時刻  $t-1$  までで処理された回路の状態を表す信号とが同時に処理される。すなわち、文脈層は  $t-1$  時刻までの過去の状態を記憶していることを意味する。この結果、ある時刻  $t$  でのネットワークの状態は現在の入力と過去の入力履歴の集合によって決まることになる。例えば、図2において過去の影響を考えれば時刻  $t$  における中間層の状態は

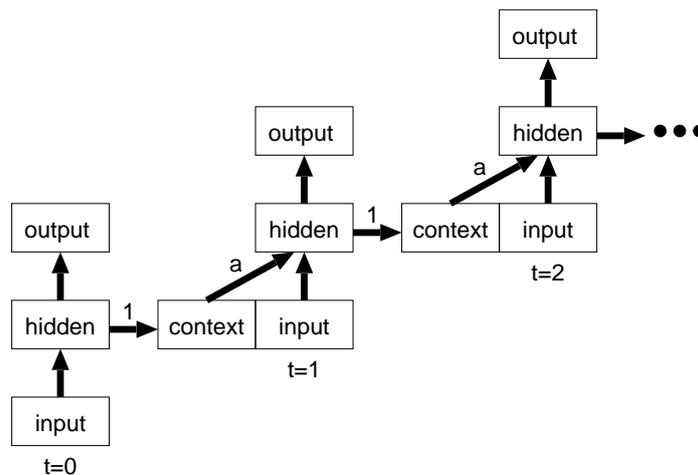


図 2: エルマンネットの時間発展

$$\begin{aligned}
h(t) &= I(t) + ah(t-1) \\
&= I(t) + a(I(t-1) + h(t-2)) \\
&= I(t) + aI(t-1) + a^2I(t-2) + a^2h(t-3) \\
&= \sum_{\tau=0}^T a^\tau I(t-\tau)
\end{aligned} \tag{1}$$

と表すことができる。ここで  $I(t)$  は時刻  $t$  における入力を表している。文脈層からの影響 (図中では  $a$  と表記) が 1 より小さければ過去の入力からの影響が指数関数に従って小さくなることを表している (1 より大きければ過去の状態からの影響は大きくなる)。このことは、エルマンネットが一つ前の状態を保存しておくという単純な構造にもかかわらず、過去の全履歴に依存した出力を生じることを示している。

このような単純回帰型ネットワークの強力な性質を利用して、エルマンネットでは言語情報処理などへの応用が試みられてきた。

## 2 単語予測課題

エルマン [3] は、自身の考案したエルマンネットを用いて文法学習などの複雑な構造を表現できることを示した。文章を構成する単語を逐次入力層に与え、ネットワークは次の単語を予測するように訓練される。この訓練手続きを系列予測課題 (または単語予測課題) という。エルマンは、系列予測課題によって次の単語を予想することを繰り返し学習させた結果、文法構造がネットワークの結合係数として学習されることを示した。エルマンネットによって、埋め込み文の処理、時制の一致、性や数の一致、長距離依存などを正しく予測可能であることが示されている [2, 3, 4]。

表 1: エルマンの用いた文法規則

S	→	NP VP “.”
NP	→	PropN   N   N RC
VP	→	V (NP)
RC	→	who NP VP   who VP (NP)
N	→	boy   girl   cat   dog   boys   girls   cats   dogs
PropN	→	John   Mary
V	→	chase   feed   see   hear   walk   live   chases   feeds   seeds   hears   walks   lives

これらの規則にはさらに 2 つの制約がある。(1) N と V がの数が一致していなければならない。(2) 目的語を取る動詞に制限がある。例えば hit, feed は直接目的語が必ず必要であり、see と hear は目的語をとってもとらなくても良い。walk と live では目的語は不要である。

表 1 にエルマンが用いた文章生成則を示した。文章は 23 個の項目から構成されている。8 個の名詞と 12 個の動詞、関係代名詞 *who*、及び文の終端を表すピリオドである。この文法規則から生成される文 *S* は、名詞句 NP と動詞句 VP と最後にピリオドから成り立っている。名詞句 NP は固有名詞 PropN か名詞 N か名詞に関係節 RC が付加したものの何れかである。動詞句 VP は動詞 V と名詞句 NP から構成されるが名詞句が付加されるか否かは動詞の種類によって定まる。関係節 RC は関係代名詞 *who* で始まり、名詞句 NP と動詞句 VP か、もしくは動詞句だけのどちらかが続く。

入力層においては一ビットが一単語に対応するように単語の数だけユニットが用意された (実際には別の目的もあって 23 ビットではなく 26 ビット、すなわち 26 個のユニットが用意されていた)。出力層のユニットも一ユニットが一単語を表すように入力層と同じ数だけのユニットが用意された。中間層は 70 個のユニットが用意された。エルマンネットの特徴である文脈層ユニットは中間層のユニット数と同数の 70 個である。実際にエルマンが用いたシミュレーションでは入力層と中間層の間にもう一層、中間層と出力層の間にもう一層の計 5 層のネットワークであったのだが中間層の数は、問題の本質でない。エルマンは表 1 に従って生成された文章を一単語ずつ徐々にネットワークに示し、次に来る単語を予測させる訓練を行なった。すなわち入力層にある単語を提示し、出力層における教師信号として次に来る単語を与えたのである。

訓練の結果、ネットワークは次の単語の種類を予測できるようになった。例えば *boy* が提示されるとネットワークは次に来る単語として、関係代名詞 *who* もしくは単数を主語とする動詞 *feeds, seeds, hears, walks, lines* を表すユニットがほぼ等確率で活性化され、複数形を主語とする *s* の付かない動詞や他の名詞を示すユニットは全く活性化されなかった。反対に、複数名詞である *boys* が提示されると *who, chase, feed, see, hear, walk, live* が等確率で活性化された。ネットワークに *boys who Mary chases* まで提示されると文頭の主語 *boys* が複数であるために複数形を主語とする動詞が等しく活性化された。このようにエルマンのネットワークは中央埋め込み文のある、いわゆる長距離依存を正しく予測できたのである。

ここで大切なことは、エルマンのネットワークでは文法知識はネットワークの結合係数の大きさとして表象されていることである。明示的な書き換え規則のようなルールは全く与えられていない。さらに、関係代名詞による文章の再帰的構造は中間層の活性値で表現される状態空間の中に表現されていることである。そして、この文法知識は否定的な証拠を提示されることによって獲得されるのではなく、単純に次の単語を予測するだけしか行っていない点も強調される。

### 3 文字予測課題

エルマンら [5, 3] は

Manyearsagoaboyandgirlivedbytheseattheyplayedhappily.

のような文章の区切りを見つけることをネットワークに要求した。ここでの入力は文中の 1 音素で、出力は次の 1 音素を予測するように訓練された。

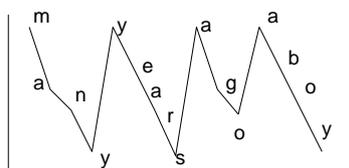


図 3: エルマンによる音素予測課題のエラー曲線

図からエラーは語頭で高く、語末まで減少している様子が分かる。エラー曲線を確信度と解釈すると、単語内の次にくる音素をかなり確信を持って予測していることが分かる。一方、入力が単語の終りに達すると次にどの語がくるか予想できないので、結果としてノコギリ状のエラー曲線になる。エラーの特徴については、実際観察される幼児の言葉の誤りと類似していて a boy を aboy と切り出してしまうオーバーシューティング (overshooting) や、they を the y とするアンダーシューティング (undershooting) のエラーが観察される。

### 4 小さく始めることは本当に重要なのか

エルマンの主張には、さらに 2 点ほど重要な点がある。それらは、「小さく始めることの重要性」と「言語獲得期における記憶容量の制限」と呼ばれる。エルマンの主張によれば、言語獲得期の幼児における記憶容量の制限は言語獲得に対して否定的な要因としてではなく、むしろ記憶容量が制限されている結果として、複雑な文章を処理しないで済むことで言語獲得が可能になるという。文法学習では記憶容量を制限することがむしろ有利に働くと主張している。現生人類が他の種と異なる特徴は、長い成育期間と顕著な学習能力である。進化の過程において、成体に達するまでの発育期間が長いことは自然淘汰から見て不利なはずである。にもかかわらず人類が滅びずにここまで文明社会を発展させたのは、まさにこの学習能力によるものであり、幼児から大人へと成長する過程で記憶容量が徐々に増加することが、我々ホモサピエンスにとって決定的に重要だったというのである。ニューラルネットワークによるシミュレーションから進化の問題を論じてしまう破天荒なところが、良くも悪くもエルマンのすごいところでもあるのだが。

これら「小さく始めることの重要性」と「言語獲得期における記憶容量の制限」と呼ばれる2点については否定的な証拠も提出されていることに言及しておきたい。この2点を仮定せずとも言語入力にある種の意味構造を仮定することでエルマンネットの言語獲得能力が劇的に向上することが示されているからである [8]。ロードとブラウトによれば言語獲得には小さく始めることが重要なのではなく、ソフトな意味論的制約 — 例えば犬は猫を追いかけるが、猫が犬を追いかけることはほとんどない — を付加することで最初から複雑な構文を与えても学習が可能であることが示されている。エルマンの訓練したネットワークでは、「犬が猫を追いかける」と「猫が犬を追いかける」とが等確率で訓練文に含まれていた。さらにエルマンが作った訓練文には、「少年が追いかけた少年が追いかけた少年が歩いた」などというような構文的には正しくても実際にはほとんど用いられないことのない文章が含まれていた。ロードとブラウトはこのような点を改善した文章 — 彼らの用語ではソフトな制約という — を用いて訓練することにより、小さく始める必要は必ずしも必要ではないことを示した。ロードとブラウトの研究によれば第二言語獲得が難しいのは第一言語である母国語の獲得の必然的結果であるとされる。二つの言語を最初から同時に学習する条件のエルマンネットは、単一言語を学習する条件のネットワークと比べて僅かに学習が成立するのが遅れるが、単一言語条件とほぼ同時期に二つの言語を獲得することが可能であった。一方、単一言語を習得したエルマンネットに対して第二言語を習得させた場合学習が進行し難いことが示された。このことはバイリンガルの成立に関する常識的な見解とも合致していると思われる。

また、子どもが人称代名詞の知識を正しく獲得するためには自分以外の大人たちの間で交わされている会話を聞くことが重要だという実験事実とニューラルネットワークによるシミュレーション研究 [12] も関係していると思われる。

## 5 生成文法理論と統計的構造学習モデル

チョムスキーの生成文法理論においては、言語獲得には普遍的で言語固有の生得性が要求される。生成文法を前提とした言語習得理論は、連続仮説に基づき大人の文法と同じ強力な装置 (例えば統語範疇、句構造規則) が幼児の文法にも存在すると仮定するため、幼児の発話に表われる意味的、形態的な制限を説明するために、様々なアドホックな原則に訴えざるを得なかった。しかし、1990年代に入ってニューラルネットワークの分野で開発されたモデルにおいては、このような生得性を仮定せずとも言語知識が学習によって創発し、記号処理的な書き換え規則を仮定せずとも統語規則を学習しようと主張されている [3, 5]。さらに、最近ではこの考え方を先鋭化させ、言語獲得とは言語の持つ多様な統計的確率的性質を学習することであるというアイデアに発展してきている [10, 11]。このような立場を取る研究は、統計的 (または

確率的) 構造学習モデルあるいは多重制約充足仮説と呼ばれる。事実、乳児は明示的なガイダンスや報酬なしに保育者の発話の統計的な側面を自然にかつ自動的に符号化しているという証拠がある [9]。そしてこのような学習は子宮内で始まっていると思われる。なぜなら新生児は母親の話す言語を好んで聞く傾向があることが実験的に確かめられている [7]。

言語の持つ統計的な性質を獲得することが重要であるという多重制約充足仮説のアイデアは、単純なマルコフ連鎖だけを用いた確率的言語モデルでは文法の問題は説明できないとしてチョムスキーの生成文法理論においては長い間無視されてきた。例えば、チョムスキーが考案した文章 “Colorless green ideas sleep furiously” は、英語を母国語とする聞き手には文法的に正しいと判断できるが意味をなさないことが了解されるが、統計的言語モデルでは文法判断ができないとされてきた。統計的構造学習モデルの枠組では、この文章でさえ、Property, Property, Things, Action, Manner という自然な英語の文法構造を反映しているということになる [1]。最近のニューラルネットワーク研究の動向を見ると、子どもは普遍文法の知識を持って産まれて来るといふ生成文法の仮説だけが言語獲得の諸事実を説明する仮説ではない [11] のかも知れない。

エルマンの研究や統計的 (確率的) 構造学習モデルに代表されるニューラルネットワーク理論は、どのように言語知識が学習されて行くのかという問題や、言語能力と言語運用とを区別して考える必要がない、という生成文法理論では説明が難しかった問題に答えることができる。このことは理論上大きなアドバンテージを持つと言えるだろう [10, 11]。加えて「子どもは規則に違反する例文を提示されないのになぜ正しく文法を学習するのか」というベーカーのパラドックス (Baker’s paradox) をも矛盾なく説明できる。

上記のような統計的構造学習モデルの視点で言語獲得を考えれば、言語獲得における子どもの課題は、生成文法理論の主張する普遍文法におけるパラメータ設定問題ではなく、むしろ言語の使用そのもの、および背後にある言語の統計的 (確率的) 構造を学習することであると言えるだろう [10]。

言語が人間という種に特異的で領域固有であるという言語学者の主張も近年疑われ始めている。他の種は確かに我々人間のような言語を持っていないが、同時に我々人間のようにバイオリンを引いたりゴルフを楽しんだりしない。言語が種に特異的で領域固有であるのならバイオリンやゴルフも種に特異的で領域固有の知識だということになってしまうからである。

## 参考文献

- [1] Joseph Allen and Mark S. Seidenberg. The emergence of grammaticality in connectionist networks. In B. MacWhinney, editor, *The Emergence of Language*, pages 115–151. Lawrence Erlbaum, Mahwah, NJ, 1999.

- [2] Jeffrey L. Elman. Finding structure in time. *Cognitive Science*, 14:179–211, 1990.
- [3] Jeffrey L. Elman. Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, pages 195–225, 1991.
- [4] Jeffrey L. Elman. Learning and development in neural networks: The importance of starting small. *Cognition*, pages 71–99, 1993.
- [5] Jeffrey L. Elman, Elizabeth A. Bates, Mark H. Johnson, Annette Karmiloff-Smith, Domenico Parisi, and Kim Plunkett. *Rethinking Innateness: A connectionist perspective on development*. MIT Press, Cambridge, MA, 1996. (邦語認識発達生性性, 性, 性, 性, 性, 性, 性立性).
- [6] E. M. Gold. Language identification in the limit. *Information and control*, 10:447–474, 1967.
- [7] Christine Moon, Robin Panneton-Cooper, and William Fisher. Two-day-old infants prefer their native language. *Infant Behavior and Development*, 16:495–500, 1993.
- [8] Douglas L. T. Rohde and David C. Plaut. Language acquisition in the absence of explicit negative evidence: How important is starting small? *Cognition*, pages 69–109, 1999.
- [9] Jenny R. Saffran and Elissa L. Newport Richard N. Aslin. Statistical learning by 8-month-old infants. *Science*, 274:1926–1928, 1996.
- [10] Mark S. Seidenberg. Language acquisition and use: Learning and applying probabilistic constraints. *Science*, pages 1599–1603, 1997.
- [11] Mark S. Seidenberg and Maryellen C. MacDonald. Constraint satisfaction in language acquisition and processing. In Morten H. Christiansen and Nick Charter, editors, *Connectionist Psycholinguistics*, chapter 9, pages 281–318. Ablex publication, Westport, CT, 2001.
- [12] 大嶋百合子. ことばの意味の学習に関するニューラルネットワークモデル—人称代名詞の場合—. *心理学評論*, 40(3):361–376, 1997.