

海馬-大脳新皮質系に着目した深層学習による記憶モデルの提案

Proposal of a Memory Model using Deep Learning
Based on the HippocampusCerebral Neocortex System.大澤 正彦 萩原 将文
Masahiko Osawa Masafumi Hagiwara

慶應義塾大学 理工学研究科 開放環境科学専攻

Keio University, Graduate School of Science and Technology, School of Science for Open and Environmental Systems

As deep learning is rapidly emerging as a robust machine learning architecture, it is considered to share many features with the cerebral neocortex. Therefore, it is essential to bring the computational insights of deep learning into the field of brain modeling. In this paper, we propose a memory model which focuses on the path from the hippocampus to the cerebral neocortex by implementing deep learning and associative memory to discuss the relationships with human memory and extend the framework of deep learning.

記憶	脳	深層学習 (既存)	深層学習 (提案)
短期記憶	ワーキングメモリ	データ	データ
中期記憶	海馬	×	連想記憶
長期記憶	大脳新皮質	深層学習	深層学習

表 1: 脳と機械学習の記憶に関する比較

1. はじめに

人工知能など様々な分野で記憶に着目した研究は活発に行われており、多くの海馬や大脳新皮質を含む記憶のモデルが古くから提案されてきた [Treves 94, McClelland 95, 大森 94]. また近年では、大脳新皮質に類似するいくつかの特徴を有する深層学習 (Deep Learning) が注目を集めている。

ところが、深層学習の計算論を取り入れた記憶モデルは著者らの知る限り存在しない。深層学習の計算論を応用した記憶モデルは生理学的整合性、工学的応用性の面でも可能性を秘めている。

表 1 に脳と現在の深層学習のそれぞれの記憶の流れを示す。記憶は短期的なものから長期的なものへと移っていくことが知られており、脳にはそれぞれに対応する部位が存在すると考えられる。ところが現在の深層学習を用いた学習は、データが直接長期的な記憶部分へと送られている。

本研究では初めに、Restricted Boltzmann Machine(RBM) と Echo State Network(ESN) を組み合わせた ESN-RBM と、その拡張である Incremental Learning ESN-RBM(IL-ESN-RBM) を提案する。これらはある時刻における RBM の隠れ層の状態から、次時刻の隠れ層の状態及びそれに対応する可視層の状態を予測することができる。これによって連続したデータの記憶や想起が可能である。

次に ESN-RBM や IL-ESN-RBM, 深層学習を組み合わせた記憶モデルを提案する。このモデルでは、記憶はその保存される期間によって 3 種類に分類され、それぞれが 3 つのモジュールに対応する、

1 つ目は短期的な記憶を担う知覚モジュールである。連続的なデータに対して常に次の時刻の状態を予測し、予測しきれな

連絡先: 大澤正彦, 慶應義塾大学 理工学研究科 開放環境科学専攻, 横浜市港北区日吉 3-14-1, TEL:045-563-1141, FAX:045-566-1747, osawa@soft.ics.keio.ac.jp

い情報を選択することで短期的に有用な情報を保持する機構を導入する。

2 つ目は中期的な記憶を保持する海馬モジュールである。これまでも海馬のモデルとして連想記憶が用いられたものや、時系列を処理する機構が取り入れられたものは提案されてきた [森田 89]. しかしながら、RBM は深層学習が成功してから注目された比較的新しい機械学習手法であり、生成モデルとしての性質が海馬モジュールとして有用であると考えられる。

3 つ目は、大規模な深層学習を用いた長期的に記憶を保存する大脳新皮質モジュールである。今回は基本的な深層学習手法を利用した。

以下、2 章で記憶モデルで主に用いる RBM の拡張モデルについて説明し、3 章で提案モデルについて述べる。そして 4 章をまとめとする。

2. RBM の拡張モデル

本章では著者らがすでに提案している既学習情報を失わない追加学習が可能な Incremental Learning RBM(IL-RBM) と、今回新たに提案する Echo State Network RBM(ESN-RBM) 及びそれらを組み合わせた IL-ESN-RBM について説明する。

2.1 IL-RBM

IL-RBM は著者らが提案した RBM の隠れ層素子数の自動決定法 [大澤 14] 並びに追加学習法 [大澤 15] を取り入れて拡張した RBM である。ここではこれらの既存手法の概略を説明する。

文献 [大澤 14] で提案した RBM の隠れ層素子数の自動決定法は、さまざまな実験条件に対して RBM の隠れ層ニューロン数に対するクロスエントロピーの関係を分析し、それらの実験結果から得られた傾向を利用した手法である。比較的少ない素子数を有する RBM に対してデータセットを実際に学習させた際の特性から、適切な素子数の 1 つを予測することができる。

文献 [大澤 15] では、RBM の学習がデータセットに対してエネルギー関数の値が最小化されるように学習することに注目し、エネルギー関数を用いて未学習のデータを検出することができることを示した。また、この未学習データ検出法と、文献 [大澤 14] の隠れ層素子数決定法を利用して、新たな素子を用いた既学習情報を失わない追加学習法を提案した。

実験では提案手法により既学習情報を失わずに新規データを追加学習することが示唆された。また、追加学習機能を有す

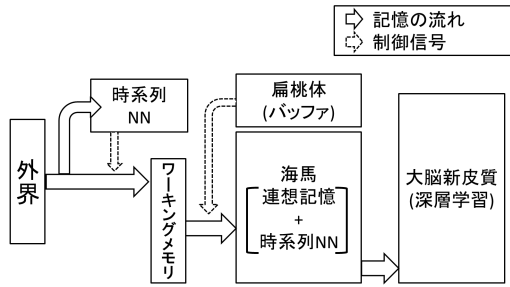


図 1: 海馬-大脳新皮質系に着目した記憶モデル

る RBM を用いた Deep Belief Network (DBN) は、簡単な行動学習 (三目並べ) をする様子が確認された。

2.2 ESN-RBM

RBM を時系列に対応させる研究はこれまでにいくつか発表されている [Sutskever 09, Boulanger-Lewandowski 12]。しかしながら、これらの手法は前節の手法を単純に組み込むことが難しい。そこで著者らの既存手法との親和性が高い ESN-RBM を新たに提案する。

ESN-RBM は RBM の隠れ層と ESN が連結された学習器である。ここで ESN の入力はある時刻 t の RBM の隠れ層の状態、出力は次時刻 $t+1$ における RBM の隠れ層の状態である。このネットワークは、RBM の隠れ層から ESN、RBM の隠れ層という再帰構造を有するため、全体として時系列データを扱うことができる。従って ESN-RBM は時系列データを抽象化し記憶する連想記憶メモリである。

ESN-RBM の学習はまず、入力された時系列データに対して、RBM 部分のみを用いて学習する。その後隠れ層の状態から次時刻の隠れ層の状態を予測するような ESN の出力の重みを決定する。

時系列データの想起を行う場合は、ESN の伝搬と RBM の隠れ層から可視層への伝搬を繰り返す。

2.3 IL-ESN-RBM

IL-ESN-RBM は IL-RBM と IL-ESN を組み合わせたモデルである。新たなエピソードの系列が入力されるたびに、別の隠れ層素子群を用いて学習する。これによって複数のエピソードを分割して記憶することができる。

3. 海馬-大脳新皮質系に着目した記憶モデル

図 1 に提案する記憶モデルの全体図を示す。提案モデルは知覚モジュール、海馬モジュール、大脳新皮質モジュールの 3 つから成る。本章では 3 つのモジュールの詳細について述べる。

3.1 知覚モジュール

図 1 における知覚モジュールに相当する部分を、図 2 に示す。時刻 t における入力ベクトル $\mathbf{x}(t)$ を観測すると RBM の隠れ層に伝搬し、その隠れ層の状態を $\mathbf{h}(t)$ として保存する。次に、RBM の隠れ層の状態を ESN に入力、伝搬させ次時刻の予測された隠れ層の状態 $\mathbf{h}'(t+1)$ を獲得し、保存する。ここで、1 時刻前に予測した $\mathbf{h}'(t)$ と実際に得られた $\mathbf{h}(t)$ を用いて以下の式で予測誤差平均の 2 乗値を算出する。

$$E = \overline{(\mathbf{h}(t) - \mathbf{h}'(t))^2} \quad (1)$$

ここで $\overline{\mathbf{X}}$ はベクトル \mathbf{X} の要素の平均を表す。また RBM の各素子は 0 以上 1 以下の値をとることから、 E も 0 から 1

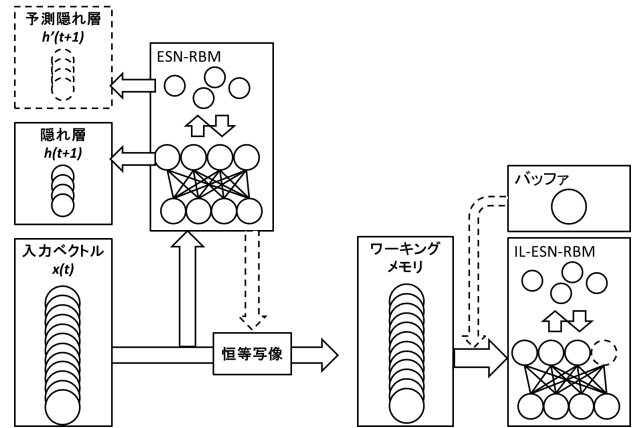


図 2: 知覚モジュール

図 3: 海馬モジュール

の間の実数値をとる。次に、算出された予測誤差平均の 2 乗値を 1 の生起確率として 1bit のバイナリ値 J を算出する。

$$J = \text{binomial}(E^2) \quad (2)$$

$\text{binomial}(p)$ は $p(0 \leq p \leq 1)$ の確率で 1 を、 $1-p$ の確率で 0 を返す関数である。この結果 $J=1$ のとき、観測した入力ベクトル $\mathbf{x}(t)$ をワーキングメモリに伝播する。一方 $J=0$ のときはワーキングメモリに伝播せず、ワーキングメモリは状態を維持する。つまり以下のような定式化が行われる。

$$M(t+1) = \mathbf{x}(t)J + M(t)(1-J) \quad (3)$$

次時刻の隠れ層の状態 $\mathbf{h}(t)$ は、 $J=1$ の時実際に観測した情報を伝搬させることで算出する。一方 $J=0$ の時には観測情報を用いず、予測した $\mathbf{h}'(t+1)$ を次時刻の $\mathbf{h}(t+1)$ として用いる。

これらの一連の定式化によって、次時刻の状態が予測できる情報を削減し、予測しがたい情報を短時間保持する短期記憶を表現する。

知覚モジュールでは以下のような特徴があると考えられる。

可変長のタイムスケール

知的システムを設計する上で、エージェントが外界の情報をどれくらいのタイムスケールで収集するのかというのは重大な問題である。提案システムでは、予測誤差を基にタイムスケールを動的に決定することで、必要に応じた正確性と処理速度を柔軟に調節できる可能性がある。

予測可能性に従った情報圧縮

予測誤差の大きさに従った確率で知覚情報を伝搬させることで、予測可能な情報を削減する。

ここで提案モジュールを用いた実験について示す。32*32 ピクセルの画像 16 フレーム分の簡単な動画を学習した ESN-RBM に対して、同じ動画を 4 回連続で入力した。

RBM の学習回数は 2,000 回、学習率は 0.001、隠れ層数 30、ESN のリザーブは 100 個で行った場合、計 64 フレームの中で伝搬されたのは 28 フレームであった。このような結果からも、提案モジュールによって情報圧縮ができることが示唆されている。

3.2 海馬モジュール

図 1 における海馬モジュールに相当する部分を図 3 に示す。海馬モジュールは IL-ESN-RBM を用いた連想記憶であり、新

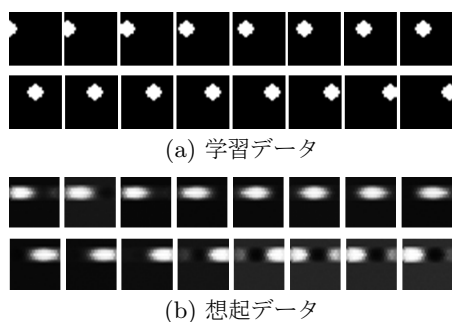


図 4: エピソード記憶の再現実験: 32*32 ピクセルの 2 値画像による時系列データを学習. 左上から右下へ時刻 $t=0\sim 15$ における学習データと, 想起データを示している.

たな入力が与えられた場合追加学習機能を利用しながら学習する. 一方で新たな入力が与えられない場合には, データを想起し大脳新皮質モジュールへ転送する.

また, 記憶は強い感情に結び付けられたものがより定着しやすいと考えられている [阿部 09]. この知見を参考に, 感情をつかさどる扁桃体に見立てたバッファを用意し, その大きさに応じて RBM に与える学習率を調整する.

海馬モジュールは以下のような特徴を持つと考えられる.

エピソード記憶の再現

2 章で述べたように, ESN-RBM は時系列データを記憶する連想記憶メモリとしてみなすことができる. 図 4 に簡単な動画を ESN-RBM に記憶させた例を示す. ここでは, ESN-RBM に対して図 4(a) のような前節の実験で用いたものと同様の 32*32 ピクセルの 2 値画像の時系列データを学習させ, 学習済みの ESN-RBM から時系列データを想起させる実験を行った.

ここで, RBM の隠れ層素子数は 30 個, ESN のリザーバーは 100 個とした. また, RBM の学習率は 0.001 とし, 学習回数は 2000 回である.

図 4(b) は実際に学習した ESN-RBM が想起した時系列データであり, 記憶された時系列データが抽象化されて想起されている様子が示されている.

また, 想起された各フレームを基に RBM 部分のみを用いて伝搬を繰り返すことで, 記憶させたフレームの内の 1 つに収束することが実験からわかっている. これらの特性は ESN-RBM が海馬の重要な機能の 1 つと考えられているエピソード記憶を表現できることを示唆している.

一時的な記憶の蓄積

ヒトの感覚情報は処理を加えられた後に海馬に一度蓄えられ, その後長時間かけて大脳新皮質に送られると考えられている. 提案する記憶モデルはこの知見を参考にしている.

提案モデルで記憶を海馬モジュールから大脳新皮質モジュールに転送させる手順を説明する. 適切に学習された RBM は, 任意の隠れ層の状態から記憶したデータに近いデータを生成することができる. この性質を利用して, 学習を終えた ESN-RBM からデータを複数生成しデータセットを構築する. ここで生成したデータセットを大脳新皮質モジュールである深層学習の学習データとして利用することで, 一度海馬モジュールに蓄えられた学習情報を深層学習に対して転送することができる.

一度学習を終えた RBM を用いて学習データを再構築することでデータセットを再生する手法は, 現在の深層学習の枠組

みにとって親和性が高く, 海馬モジュールを導入することによる深層学習の学習法の変更点が少ないと考えられる.

高速な学習

IL-ESN-RBM は新たな素子を追加することで大きな学習率を用いて学習率を大きく設定しても, 既学習情報は失われない. これにより高速に知覚情報を記憶することができる.

既学習情報を失わない追加学習

IL-ESN-RBM の追加学習機構により, 既学習情報を失わずに追加学習することができる. この性質は, 外界からの情報が次々と入力される海馬にとって重要な機能の 1 つであると考えられる.

行動学習

著者らの研究において, 追加学習機能を有する RBM を用いて簡単な行動学習 (三目並べ) できることを示している [大澤 15]. 海馬モジュールの情報を用いて行動学習が可能であることは, 獲得した記憶を大脳新皮質に移す前に利用可能であることを意味しており, 実践的な知的システムを設計するために重要な性質であると考えられる.

新生ニューロンとの整合性

Seki らの研究によって記憶をつかさどる海馬では成人にも常に新たなニューロンが誕生していることが明らかにされた [Tatsunori 93, Tatsunori 95]. IL-ESN-RBM における新たな素子を追加追加は, このような知見を利用している.

CA3 自己ループとの整合性

海馬は CA1~CA3 の三層構造を構成していることが知られており, その中の CA3 と呼ばれる層には CA3 への再帰的な経路が存在する. RBM の隠れ層から ESN を介した隠れ層へのループは, 海馬の CA3 野における自己ループの知見に対応する.

記憶の定着度

強い感情を伴う記憶はその定着度が高いことが知られている. 扁桃体バッファから学習率を調整することで, 強い感情を伴う記憶を定着しやすいという認知科学的知見との整合性をとることが可能であると考えられる.

3.3 大脳新皮質モジュール

図 1 における大脳新皮質モジュールに相当する部分を, 図 4 に示す. 今回提案する記憶モデルにおいて大脳新皮質モジュールは, 深層学習そのものを用いている. 海馬モジュールへの入力が無い場合に, 海馬モジュールから想起したデータを大脳新皮質モジュールに対して転送する.

深層学習の基本的な枠組み自体は変更していないが, 前段階として海馬モジュールを導入することによる工学的な利点が複数あると考えられる. これらについて詳しく説明する.

多大な学習時間に対するリスク回避

従来の深層学習では, 多大な学習時間が問題視されていた. 深層学習のアルゴリズム自体は改良しないため, 学習時間は既存手法と同程度である.

しかしながら提案記憶モデルにおいては, 外界からの情報がエージェントに入力されていない時間を用いて学習することができるため, 学習時間が長いという欠点が問題になりにくいと考えられる.

ここで海馬モデルを導入した場合の表層的な計算時間の差を簡単な実験によって評価する. web 上から取得した 65 枚の女性の正面を向いた画像に対して, 顔部分の抽出, 目の位置の正規化, 64*64 ピクセルへのリサイズ, グレースケール化などの

	学習率	学習回数	隠れ層数	実行時間 (s)
IL-RBM	0.1	100	可変 (1~65)	$7.9 * 10^{-1}$
dA	0.0002	30,000	不変 (1500)	$3.4 * 10^3$

表 2: 脳と機械学習の記憶に関する比較

前処理を行ったものをデータセットとして、従来法 (denoising Autoencoder:dA) を用いて学習した場合と、提案モデルを学習させた場合で表層的な学習時間を比較する。

ここで、海馬モジュールから大脳新皮質モジュールへ記憶を転送するための計算は、モデルへの新たなデータが入力されていない際に行われることを想定している。従って提案モデルの表層的な計算時間は海馬モジュールの学習時間のみである。

提案モデルでは追加学習を行う際には新たな隠れ層素子を1つ追加するものとし、データが1つ入力されるたびに追加学習を実行する。

あらかじめ行った予備実験で決定した各手法のパラメータと、実際に実行した場合の実行時間を表に示す。

ここで、同じデータセットを扱うために dA では1時間近くかかっているタスクを提案する海馬モジュールは1秒以下で学習が完了している。

データセットが決定していないタスクへの応用

基本的な深層学習のアルゴリズムを適用する場合、学習開始時点でデータセットが完成している必要があった。提案モデルでは、海馬モジュールが十分に追加学習機能を用いてデータセットを収集した後に大脳新皮質モジュールに対する学習を開始することで、データセットがあらかじめ決定していない問題に対しても基本的な深層学習の枠組みで学習が可能であると考えられる。

時系列情報を扱う機構

基本的な深層学習技術では、時系列を扱う機構が存在しない。しかしながらこの記憶モデルにおいては図 4(b) にも示されているように、海馬モジュールから深層学習に対して時系列情報を含んだデータセットを転送することができると考えられる。これは、深層学習の枠組みにおいて明示的に時系列を処理する機構がなくとも、時系列情報を扱うことができることを示唆している。

4. おわりに

本論文では、海馬-大脳新皮質系に着目した記憶モデルを提案した。昨今の深層学習の成功を踏まえ、深層学習の計算論を前提としている。記憶モデルは、知覚モジュール、海馬モジュール、大脳新皮質モジュールの3つから形成され、それぞれ神経科学や認知科学での知見とも整合性が取れている点が複数あると考えられる。

その中でも特に海馬モジュールとして提案した IL-ESN-RBM は、海馬の新生ニューロンや CA3 自己ループなど重要な知見と整合性が取れた構造であるだけでなく、エピソード記憶や大脳新皮質との記憶の仲介などの機能面でも海馬の機能を再現していることが示唆された。

また、提案した海馬モジュールを深層学習の前段階として導入することの工学的な利点として、学習時間が長いという欠点が問題になりにくいことや、データセットがあらかじめ決定される必要がないこと、明示的に時系列を扱う機構がなくとも時系列データを扱えることなどが考えられることを説明した。

今後はより多くの知見を収集し、モデルの更なる精緻化していく予定である。

謝辞

本研究の一部は、科学技術振興機構 (JST) 戦略的想像研究推進事業 (CREST) 「実践知能アプリケーション構築フレームワーク PRINTEPS の開発と社会実践」の支援によって実施した。

参考文献

- [Boulanger-Lewandowski 12] Boulanger-Lewandowski, N., Bengio, Y., and Vincent, P.: Modeling Temporal Dependencies in High-Dimensional Sequences: Application to Polyphonic Music Generation and Transcription, in *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, pp. 1159–1166 (2012)
- [McClelland 95] McClelland, J. L., McNaughton, B. L., and O'Reilly, R. C.: Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory., *Psychological review*, Vol. 102, No. 3, p. 419 (1995)
- [Sutskever 09] Sutskever, I., Hinton, G. E., and Taylor, G. W.: The recurrent temporal restricted boltzmann machine, in *Advances in Neural Information Processing Systems*, pp. 1601–1608 (2009)
- [Tatsunori 93] Tatsunori, S. and Yasumasa, A.: Highly polysialylated neural cell adhesion molecule (NCAM-H) is expressed by newly generated granule cells in the dentate gyrus of the adult rat, *The Journal of Neuroscience*, Vol. 13, No. 6, pp. 2351–2358 (1993)
- [Tatsunori 95] Tatsunori, S. and Yasumasa, A.: Age-related production of new granule cells in the adult dentate gyrus., *Neuroreport*, Vol. 6, No. 18, pp. 2479–82 (1995)
- [Treves 94] Treves, A. and Rolls, E. T.: Computational analysis of the role of the hippocampus in memory, *Hippocampus*, Vol. 4, No. 3, pp. 374–391 (1994)
- [阿部 09] 阿部 和穂: 情動によって記憶が強化されるメカニズム, *生体の科学*, Vol. 60, No. 4, pp. 318–322 (2009)
- [森田 89] 森田 昌彦: 連想記憶の海馬モデル, *電子情報通信学会論文誌 D*, Vol. 72, No. 2, pp. 279–288 (1989)
- [大森 94] 大森 秀樹, 大森 隆司: 学習時定数の違いによっておきる海馬と新皮質の記憶における機能分離のモデル, *電子情報通信学会論文誌 D*, Vol. 77, No. 9, pp. 1882–1890 (1994)
- [大澤 14] 大澤 正彦, 萩原 将文: RBM の学習特性の分析と隠れ層ニューロン数の自動決定法, *信学会信学技報ニューロコンピューティング研究会*, Vol. 114, No. 259, pp. 7–12 (2014)
- [大澤 15] 大澤 正彦, 萩原 将文: RBM における未学習データ検出法の提案と追加学習への応用, *信学会信学技報ニューロコンピューティング研究会*, Vol. 114, No. 515, pp. 283–288 (2015)