

令和元年6月18日現在

機関番号：13501

研究種目：基盤研究(C) (一般)

研究期間：2016～2018

課題番号：16K00329

研究課題名(和文) 海馬から大脳皮質への宣言的記憶の転写モデル

研究課題名(英文) Transcriptional model of declarative memory from hippocampus to neocortex

研究代表者

服部 元信 (HATTORI, Motonobu)

山梨大学・大学院総合研究部・准教授

研究者番号：40293435

交付決定額(研究期間全体)：(直接経費) 2,600,000円

研究成果の概要(和文)：本研究では、出来事や事実に関する記憶を司る海馬と大脳皮質における相補的な学習システムを工学的にモデル化することを目的とし、生物学的に妥当なモデルの構築を行い、その特性を計算機シミュレーションによって調査した。特に、長期的な記憶の座である大脳皮質に重点を置き、新しい記憶によって古い記憶が壊されることを防ぐには、どのような仕組みが必要になるのか、生物学的な妥当性を考慮したモデル化を行った。その結果、誤差の逆伝搬を行わない生物学的に妥当な学習方法と擬似リハーサルを組み合わせることで、忘却の抑制が行えること、また、重みの重要度を考慮することでさらに忘却の抑制を改善できることを明らかにした。

研究成果の学術的意義や社会的意義

意識的に思い出すことのできる様々な出来事に関する記憶(エピソード記憶)や事実に関する記憶(意味記憶)は宣言的記憶と呼ばれ、思考や推論をといった極めて高次の情報処理で用いられている。宣言的記憶は、初めに海馬に蓄えられ、その後徐々に大脳皮質へと転写されていくと考えられているが、その仕組みは未解明である。本研究では、人間のように知的で柔軟な情報処理システムの実現に向けて、その基盤となる宣言的記憶の形成過程を工学的に模倣した。特に、人間のように、古い記憶を破壊することなく、次々と新しい情報を追加的に記憶していく仕組みについて、そのモデル化を行った。

研究成果の概要(英文)：In this study, we were aimed at modeling the complementary learning system in the hippocampus and neocortex responsible for memories about events and facts. In particular, we focused on the neocortex, which is the locus of long-term memory. We considered what kind of mechanism would be required to prevent old memories from being destroyed by new memories, and constructed biologically plausible models. By examining their characteristics by computer simulation, we have revealed that it is possible to reduce catastrophic forgetting by combining a biologically relevant learning method that does not propagate output errors backward through the network and pseudorehearsal. We have also shown that catastrophic forgetting is further reduced by considering the importance of weights.

研究分野：情報工学

キーワード：海馬 大脳皮質 記憶 ニューラルネットワーク 破局的忘却 No-prop CHL

様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

1. 研究開始当初の背景

脳の機能において記憶はとりわけ重要であり、我々人間の知的な活動は記憶を抜きにしては語れない。記憶は、まず毎日の個人的な出来事の経験、すなわち、エピソード記憶として獲得され、そこから時間的文脈が除去された事実に関する記憶(意味記憶)が形成されると考えられている。これら意識的に思い出すことのできる記憶(宣言的記憶)は、思考や推論といった高次な情報処理に利用されているため、人間のように知的で柔軟な情報処理システムを構築するにあたっては、如何にして宣言的記憶の形成過程を模倣するかが大変重要である。宣言的記憶は、初めに海馬に蓄えられた後、徐々に大脳皮質へと転写されていくと考えられている。しかし、(1)海馬にどのように蓄えられ、それがどのような仕組みで取り出され転写されるのか、(2)大脳皮質に既に固定されている記憶を破壊することなく新しい情報をどのように追加するのか、(3)エピソード記憶からどのように時間的文脈が除去されていくのか、などそのメカニズムの全容は未解明である。そのため、特に海馬に関する生理学的な研究が近年盛んに行われている。また、計算機科学の分野においても、海馬の計算モデルや工学モデルが近年いくつか提案されている。しかし、従来の研究では、海馬やその一部のみをモデル化の対象とすることがほとんどであった。それに対し研究代表者らは、海馬のみならず、長期記憶の座である大脳皮質までを考慮し、記憶の形成過程を巨視的にシステムとして捉えた研究を行ってきた。そして、本研究を進めていく上での基礎となる以下の結果を得ている。

- (1) 海馬の有するパターン補完機能とスパイクタイミング依存性シナプス可塑性による学習によって、エピソードから時間的文脈に不変な記憶が形成できることを示した。
- (2) 解剖学並びに生理学的知見に基づいた海馬 CA3 のモデルを構築し、海馬が機能的に自己想起と相互想起の領域に分けられることを示し、これらがエピソード記憶の曖昧性解消に寄与する可能性があることを示した。
- (3) 海馬歯状回における神経新生を採り入れた海馬モデルを構築し、神経新生が互いに類似した情報の記憶を容易にするとともに、追加学習能力の改善、複雑な時系列情報の記憶に寄与することを明らかにした。
- (4) カオスニューラルネットワークを導入した海馬-大脳皮質モデルを構築し、時間変化のない静的なパターンに対して、破局的忘却を抑制しつつ長期記憶の形成を可能にすることを示した。

2. 研究の目的

本研究では、研究代表者のこれまでの研究を発展させ、宣言的記憶を形成する海馬-大脳皮質モデルを構築する。研究期間内には以下のことを明らかにする。

- (1) 宣言的記憶を可能とする海馬モデルとその特性
初期的な記憶の座である海馬では、次々と新しい記憶を蓄えていく必要があるため、高速な学習が不可欠である。一方、比較的小さい部位であるため、効率的な学習によって大記憶容量を実現していると考えられるが、一般に、高速な学習で大記憶容量を実現するのは非常に難しい。同時に、海馬においても、過去の記憶を破壊することなく新規な情報を記憶する機構が必要である。本研究では、これらの機能を実現するには、どのような仕組みが必要となるのかを明らかにする。また、構築したモデルの記憶容量や想起特性を主に計算機実験により明らかにする。
- (2) 海馬モデルからの記憶の抽出の仕組み
睡眠中の脳活動の研究によると、REM 睡眠時に脳幹の神経細胞がカオス的に発火し、それが皮質まで伝搬することや、徐波睡眠時に海馬で高頻度の特徴的な活動が見られることが分かっている。これらは長期記憶の固定に関与している可能性がある。ここでは、こうした知見を採り入れ、海馬モデルに蓄えられた情報を自律的に想起し、長期記憶を担う大脳皮質モデルへ転写する仕組みを計算機実験により明らかにする。
- (3) 長期記憶形成のための大脳皮質モデルと破局的忘却抑制の仕組み
海馬から抽出された情報は、最終的な記憶の貯蔵庫である大脳皮質に蓄えられる。その際、既に大脳皮質に蓄えられている情報を破壊することなく、新規な情報を追加的に学習する必要がある。ここでは、生理学的に妥当な大脳皮質モデルを構築し、破局的忘却を抑制するにはどのような仕組みが必要となるのかを主に計算機実験により明らかにする。
- (4) 神経心理学的な妥当性と工学的な有用性
本研究で構築した宣言的記憶を形成する海馬-大脳皮質モデルの特性について、神経心理学的実験によって得られている人間の記憶特性との整合性について調べ、本モデルの妥当性を検証する。また、記憶容量、重複を含む複雑な時系列群に対する想起能力、耐雑音性などの工学的に重要な特性を主に計算機実験により明らかにする。

3. 研究の方法

本研究課題は以下のように実施した。

- (1) 海馬を中心とした宣言的記憶に関する生理学、神経心理学、認知科学的知見の調査
詳細な研究が行われているげっ歯目を中心に、海馬に関する生理学的研究成果の調査を行った。また、脳に損傷を受け、宣言的記憶に障害が生じた患者に関する神経心理学的知見の調査、並びに、認知科学や認知心理学視点からの宣言的記憶に関する研究の調査を行った。文献よる調査の他、国内外の学会にも参加して情報収集を行った。

(2) No-prop における擬似リハーサルの適用方法の改良

大脳皮質モデルとして、階層型ニューラルネットワークの出力層の重みのみを学習する誤差非逆伝搬学習法(No-prop)の改良を行った。従来、大脳皮質ネットワークとしては、階層型ニューラルネットワークを誤差逆伝搬法で学習するモデルを採用してきたが、この学習法には生物学的な妥当性が乏しいという問題点があった。これまでの研究で、階層型ネットワークの出力層の重みのみを学習する No-prop に、擬似リハーサルを適用し、破局的忘却を抑制できることを明らかにしているが、より忘却を抑制できるように学習方法に改良を加えた。

No-prop は最小二乗法(Least Mean Square: LMS)で学習を行うが、正しく学習できる学習パターン数の上限(LMS Capacity)が存在することがわかっている。一方、睡眠時の記憶の固定法として提唱されている擬似リハーサルでは、ネットワークにランダムな入力(擬似パターン)を与え、それに対応した出力のペアを複数用意し、新規な学習パターンとともに学習させる。一般に、擬似パターンの数が多い方が、ネットワークがそれまでに獲得した入出力関係を細かくサンプリングできるため、破局的忘却がより抑制される。しかし、擬似パターン数と新規に学習させる学習パターン数の合計が LMS Capacity を超えてしまうと、No-prop では正しい学習が行えなくなる可能性がある。そこで、従来の手法では、LMS Capacity を超過しないように擬似パターンの集合を複数の部分集合に分割し、それぞれの部分集合に含まれる擬似パターンを新規な学習パターンとともに学習させる方法を採用していた。この手法でも破局的忘却が抑制できることが明らかになっているが、その抑制度合いは限定的であった。これは、擬似パターン集合を複数に分割したため、結果的に少数の擬似パターンを用いた場合と同様の効果しか得られなかったためと考えられる。

そこで、擬似パターン集合を分割せずに新規な学習データとともに No-prop で学習させる方法を考案した。この方法では、ネットワークの LMS Capacity を超えてしまう場合があるが、そうした状況でも正しく学習が行えるように、2 つの方法を用いた。擬似パターンは、学習済みのネットワークから得られた入出力パターンであるため、学習開始当初から出力における誤差はかなり小さくなっている。それに加えて、一般に、擬似パターンの数は新規に学習させるパターン数よりも多いため、両者の誤差をまとめて評価すると、新規学習パターンが十分に学習できていないにもかかわらず、学習が終了してしまう可能性がある。そのため、擬似パターンと新規な学習パターンの学習の収束の評価を別々に行うようにした。また、擬似パターンと新規学習パターンに同じ学習率を用いると、数の少ない新規学習パターンの学習が十分に行えない可能性がある。これを回避するため、擬似パターンを学習する際の学習率を相対的に小さくするとともに、擬似パターン数に反比例して小さくする方法を導入した。

(3) No-prop の汎化能力の向上と破局的忘却の抑制

No-prop を学習に用いるネットワークでは、最終中間層のニューロン数を増加させることで、LMS Capacity が大きくなり、多数のパターンが学習できるようになるとともに、学習性能が向上する。しかし、過剰な中間ニューロン数の使用は、過学習の原因となり、汎化能力の低下を招くことになる。そこで、全パターンに対する出力の小さいニューロンや絶対値の小さい重みが多く結合しているニューロンを重要でないものとして削減する手法を考案し、No-prop に適用した。この手法に擬似リハーサルも適用できるようにし、汎化能力を向上させつつ、破局的忘却が抑制できるように No-prop に改良を加えた。全体的な構成は、擬似パターンを生成するネットワークと追加学習を行うネットワークからなるデュアルネットワーク構造とした。

(4) より生理学的に妥当な学習方法の導入

ここでは、大脳皮質モデルの別の学習方法として、局所的なニューロン間のみで学習が可能な Contrastive Hebbian Learning (CHL)を採用し、擬似リハーサルによって破局的忘却の抑制ができるかどうかを計算機シミュレーションによって調査した。上述の通り、階層型ニューラルネットワークの学習法として一般的に良く用いられる誤差逆伝搬法は、強力な学習方法ではあるが、出力層の誤差が入力層まで伝搬するという学習のため、その生物学的な妥当性は乏しい。一方、CHL は、生体の脳でも確認されているヘッブ則を用いた局所的なニューロン間の学習を採用しているにもかかわらず、誤差逆伝搬法と等価な学習が行えることが示されている。ここでは、階層型ニューラルネットワークの学習法に CHL を採用し、擬似リハーサルを用いることで破局的忘却を抑制できるかどうかを調査した。

(5) 学習率、中間層数と破局的忘却の抑制度合いの調査

Complimentary Learning Systems (CLS)理論では、海馬と大脳皮質において相補的な学習が行われており、長期記憶の座である大脳皮質では、破局的忘却を抑制するために、非常にゆっくりとした学習が行われているとしている。これは、人工的なニューラルネットワークの学習でいえば、小さい学習率で学習を行うことに相当するが、なぜ小さい学習率だと破局的忘却が抑制できるのかは、まだ良くわかっていない。ここでは、学習に CHL を用いた大脳皮質モデルにおいて、学習率の大きさと破局的忘却の抑制度合いの関係について調査を行った。同時に、中間層数が破局的忘却の抑制度合いに与える影響についても計算機シミュレーションにより調査した。

(6) 重みの重要度による正則化の導入

マウスに関する生理学的な研究によると、あるタスクを学習した後、学習に関係したシナプス後細胞の樹状突起スパインの一部が生涯に渡って保存され、長期記憶の形成に重要な役割を果たしていることが示唆されている。こうした知見に基づき、学習中に重要な重みをオンラインで推定し、重要度が高い重みほど学習を行いにくくする正則手法を CHL に適用した。また、この重みの重要度による正則化手法と擬似リハーサルを用いた手法を併用した場合の破局的忘却の抑制度合いについて、計算機シミュレーションにより調査を行った。

4. 研究成果

(1) No-prop 学習による階層型ニューラルネットワークにおける破局的忘却の抑制

図 1 は、49 次元のアルファベットパターン(A,B,C,D,E)の5つを学習させた後、擬似リハーサルを用いて別のアルファベットパターン(F,G,H,I,J)を追加学習させたときの(A,B,C,D,E)の忘却の抑制度合いを示している。評価指標である平均 goodness は、忘却が全くないときに最大値 1 をとる。ネットワークには、3 層の階層型ニューラルネットワークを用い、中間層のニューロン数は 50 とした。すなわち、このネットワークの LMS Capacity は 50 である。また、擬似リハーサルに用いる擬似パターン数は、0 から 500 まで変化させた。比較対象として、擬似パターン群を分割する従来法、誤差逆伝搬法に擬似リハーサルを用いる方法を用いた。

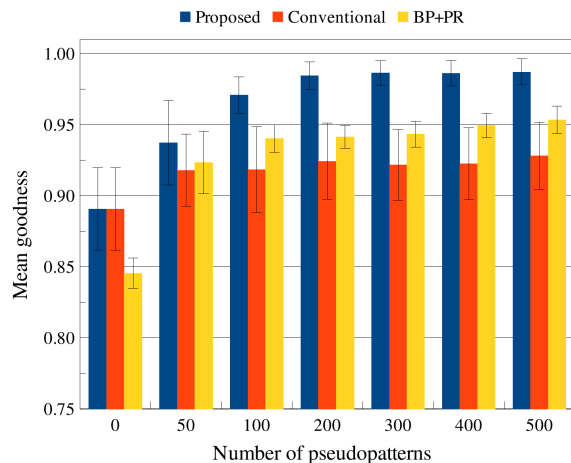


図 1: (F,G,H,I,J)を学習した後の(A,B,C, D,E)に対する平均 goodness.

この図から、提案手法は従来法よりも、大幅に破局的忘却を抑制できることがわかる。また、学習に誤差逆伝搬法を用いた場合よりも、忘却の抑制性能が優れていることがわかる。

擬似パターンが 500 個のとき、新規学習パターンと合計すると、学習パターンの総数は 505 個となり、LMS Capacity の 50 を 10 倍以上上回っているが、提案手法を用いることで学習は毎回収束した。

(2) No-prop 学習におけるニューロン削減と破局的忘却の抑制

表 1 は、3 種類のベンチマーク問題に対して、提案するニューロン削減方法を適用する前と後の認識率の比較である。それぞれのデータセットのうち、半分を学習に用い、残りの半分をテストデータとして認識率を評価した。ネットワークの構造は 3 層とし、各ネットワークの中間ニューロン数は、学習データ数と同じになるようにした。すなわち、LMS Capacity が学習データ数になるように設定した。

表 1: ベンチマーク問題に対する削除前後の認識率(%)

	Iris	Cancer	Hepatitis
削除前	96.21	96.09	72.13
削除後	96.70	96.14	77.18

この実験から、提案する削減方法を用いることで、認識率が有意に向上することがわかった。すなわち、ニューロン数を削減することで、汎化能力が向上することがわかった。

表 2, 3 は、49 次元のアルファベットパターン{(Q,q),(R,r),(S,s),(T,t),(U,u)}の5つを大文字の入力に対して小文字が出力されるように学習させた後、擬似リハーサルを用いて別のアルファベットパターン{(V,v),(W,w),(X,x),(Y,y),(Z,z)}を追加学習させたときの{(Q,q),(R,r),(S,s),(T,t),(U,u)}の忘却の抑制度合いを示している。ネットワークには、3 層の階層型ニューラルネットワークを用い、中間層のニューロン数は 100 とした。結果は、50 回の試行の平均である。

表 2 より、提案手法では 3 割程度の中間ニューロンが削減されており、また、それに伴い忘却の抑制性能が向上していることがわかる。表 3 の結果と比較すると、従来法よりも忘却が抑制され、誤差逆伝搬法に擬似リハーサルを用いた場合と同程度の抑制度合いとなっていることがわかる。

表 2 : 提案手法の平均 goodness と削減中間ニューロン数

	擬似パターン数		
	100	150	200
削減前	0.920	0.926	0.939
削減後	0.924	0.933	0.946
削減ニューロン数	31.30	30.85	30.90

表 3 : 従来法と BP+擬似リハーサルの平均 goodness

	擬似パターン数		
	100	150	200
従来法	0.904	0.919	0.936
BP+擬似リハーサル	0.930	0.937	0.942

(3) Contrastive Hebbian Learning と擬似リハーサルによる破局的忘却の抑制

図 2 は, 49 次元のアルファベットパターン $\{(Q,q), (R,r), (S,s), (T,t), (U,u)\}$ の 5 つを学習させた後, 擬似リハーサルを用いて別のアルファベットパターン $\{(V,v), (W,w), (X,x), (Y,y), (Z,z)\}$ を追加学習させたときの $\{(Q,q), (R,r), (S,s), (T,t), (U,u)\}$ の忘却の抑制度合いを示している. ネットワークには, 3 層の階層型ニューラルネットワークを用い, 中間層のニューロン数は 50 とした.

この図から, 生物学的により妥当性の高い学習法である CHL と, 擬似リハーサルを用いることで, 誤差逆伝搬法に擬似リハーサルを用いたときと遜色のない破局的忘却抑制の性能が得られることがわかる.

この他にも, 擬似パターン数を多くすると忘却の抑制が改善されるが, あまり多くしすぎると追加学習データの学習が成功しなくなることで, 忘却の抑制には中間ニューロン数や中間層数に応じた擬似パターン数が必要であること, 中間層数の多いネットワークほど追加学習時に小さな学習率を用いると忘却が抑制できること, 大きな学習率は行き過ぎた学習を惹き起こし忘却の原因となることなどを実験により明らかにした.

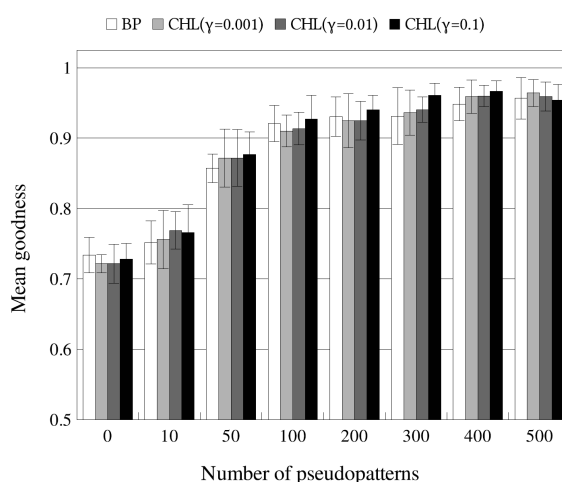


図 2 : $\{(V,v), (W,w), (X,x), (Y,y), (Z,z)\}$ を学習した後の $\{(Q,q), (R,r), (S,s), (T,t), (U,u)\}$ に対する平均 goodness.

(4) 重みの重要度による正則化の導入

図 3 は, 擬似リハーサルと, 重みの重要度による正則化を組み合わせた場合の破局的忘却の抑制度合いを示している. この実験では, 100 次元のランダムパターン 5 つを 1 組として, 10 回追加学習させた後, 1~9 回目に学習したデータの忘却の抑制度合い調べた. ネットワークには, 4 層の階層型ニューラルネットワークを用い, 中間層のニューロン数はすべて 100 とした. 比較対象として, 忘却の抑制を用いなかった場合 (CHL), 擬似リハーサルのみを用いた場合 (pp1 ~ pp30), 重みの重要度による正則化のみを用いた場合 (c0.0025) を用いた. 図中の pp の後の数字は, 擬似リハーサルにスイープリハーサルを用いた際の擬似パターン数, c の後の数字は, 重みの重要度による正則化のパラメ

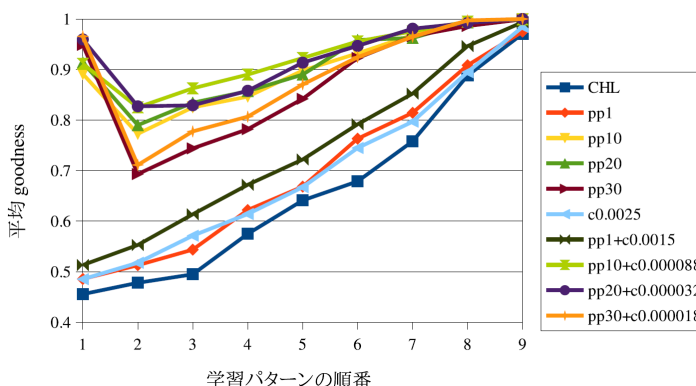


図 3 : 10 回追加学習した場合の平均 goodness.

ータの値を表わしている。

この図から、擬似リハーサルと重みの重要度による正則化の両方を用いることで、高い goodness が得られることがわかる。すなわち、睡眠時の記憶の固定法の仮説として提唱された擬似リハーサルと、生物学的な妥当性のある重みの重要度による正則化手法を用いることで、破局的忘却がより抑制されることがわかった。

5. 主な発表論文等

〔雑誌論文〕(計4件)

Hattori, M. and Nakano, S., Characteristics of contrastive Hebbian learning with pseudorehearsal for multilayer neural networks on reduction of catastrophic forgetting, International Journal of Computational Intelligence Studies, 7, 3&4, 289-331, 2018, 査読有, 10.1504/IJCISTUDIES.2018.096184

Hattori, M. and Tsuboi, H., Reduction of Catastrophic Forgetting for Multilayer Neural Networks Trained by No-Prop Algorithm, Proceedings of IEEE International Conference on Information and Communication Technology, 2018, 査読有, 10.1109/ICOIACT.2018.8350665

Nakano, S and Hattori, M., Reduction of Catastrophic Forgetting in Multilayer Neural Networks Trained by Contrastive Hebbian Learning with Pseudorehearsal, Proceedings of IEEE 10th International Workshop on Computational Intelligence and Applications, 2017, 査読有, 10.1109/IWCIA.2017.8203567

Hattori, M. and Kobayashi, Y., A Hippocampal Model for Episodic Memory using Neurogenesis and Asymmetric STDP, Proceedings of IEEE and INNS International Joint Conference on Neural Networks, 2016, 査読有, 10.1109/IJCNN.2016.7727885

〔学会発表〕(計2件)

中野峻太, 服部元信, 擬似リハーサルと重みの重要度を用いた破局的忘却の抑制, 電子情報通信学会ニューロコンピューティング研究会, 2019年3月

辺見航平, 服部元信, ニューラルネットワークを用いた強化学習における環境状態の分布を考慮した擬似リハーサルの導入, 情報処理学会第80回全国大会, 2018年3月

6. 研究組織

(1)研究分担者

なし

(2)研究協力者

研究協力者氏名: 中野 峻太

ローマ字氏名: (NAKANO, Shunta)

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。