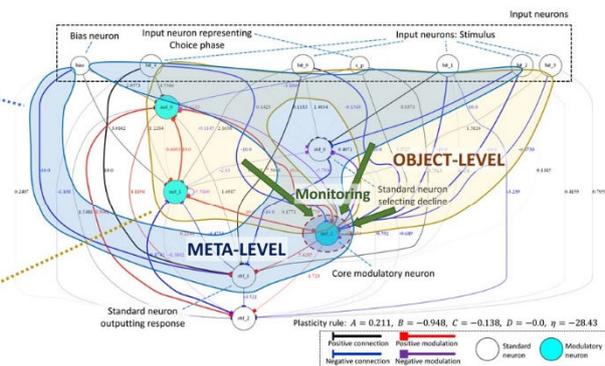


メタ記憶の例



進化したニューラルネットワーク

メタ記憶のコンピュータ内進化に成功！ ～忘れたことに気づく人工知能～

国立大学法人東海国立大学機構 名古屋大学大学院情報学研究科の有田 隆也 教授、鈴木 麗瑩 准教授、大和 祐介 博士後期課程学生の研究グループは、コンピュータ内進化実験により、メタ記憶^{注1)}の機能を持つニューラルネットワーク^{注2)}を創り出すことに成功しました。

近年、脳の回路を模したニューラルネットワークを用いた人工知能技術が急速に進展しています。本研究の対象はメタ記憶の進化です。メタ記憶とは、例えば「昨日の朝食は何を食べたか覚えているかな？」と考えるときに頭の中で働くような、記憶の存在の有無を調べたりコントロールしたりするような認知機能を意味します。

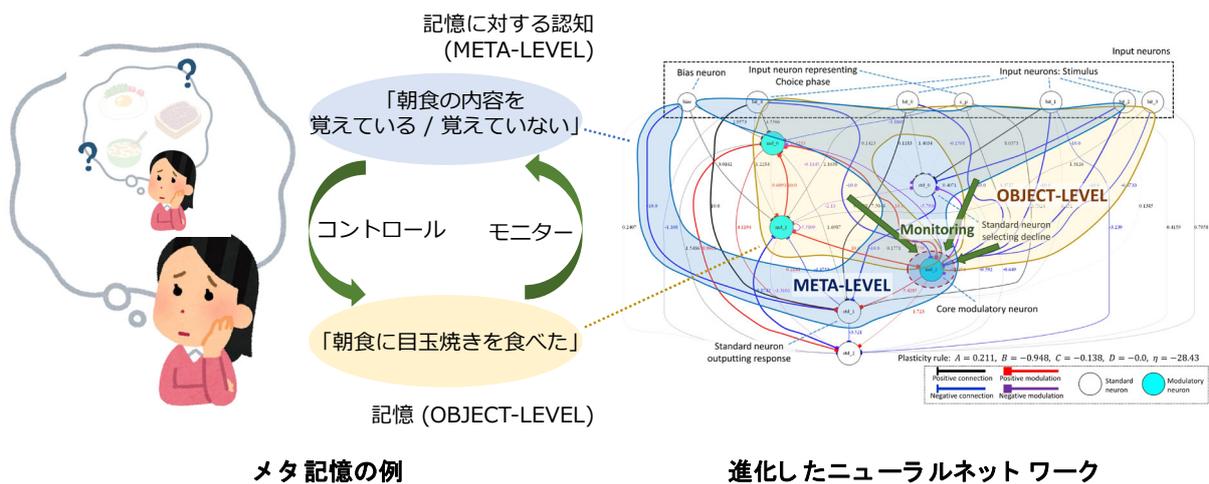
本研究では「遅延見本合わせ課題」を採用しました。最初に赤い丸などのパターンを見た後、複数パターンから最初のパターンを指し示すテストを受けます。正解だと報酬を得ますが、テストを回避しても小さな報酬を得られます。メタ記憶がある場合、記憶に自信があればテストを受け、自信がなければ回避するはずです。

本研究では、この課題を単純化した上で、最初はランダムな構造を持つニューラルネットワーク集団が世代交代を経てメタ記憶を持つように、つまり記憶が壊れがちな場合にはテストを回避するように進化しました。ニューラルネットワークの構造だけでなく、ニューロン間結合の強さの調整、つまり学習の方法についても人手によらずに進化によって決まるところがポイントです。この成果は「ヒトらしい心」、さらには「意識」を持つ人工知能の実現の手がかりになることが期待されます。

本研究成果は、2022年4月26日午後6時（日本時間）付国際科学誌「Scientific Reports」のオンライン版に掲載されました。

【ポイント】

- ・近年、脳の回路を模した人工ニューラルネットワークを用いた機械学習によって、画像や言語などの認識・生成を行う人工知能技術が急速に進展している。
- ・この進展の先には、真に「ヒトらしい心」を持つ人工知能が作れるのかという難問がある。
- ・本研究では、我々の心がそうであるように、人工生命の研究手法を用いて、進化によって、人間の認知機能の一つであるメタ記憶を創り出すことが可能か検討した。
- ・計算機内進化実験の結果、記憶を調べてその記憶の保持状態を、高次の心的表象として保持して出力を切り分けるようなニューラルネットワークが、人手には一切頼らずに進化した。
- ・動物のメタ記憶を調べる実験が近年行われてきており、その存在の有無や実験方法に関する活発な議論に対して、実際に進化できるかという新しい観点を提供する。
- ・応用的な側面については、「ヒトらしい心」、さらには「意識」を持つ人工知能の実現の手がかりになることが期待される。



【研究背景と内容】

近年、脳の回路を模したニューラルネットワークを用いた機械学習によって、画像や言語などの認識・生成を行う人工知能技術が急速に進展しています。本研究では、ヒトのような心を持った人工知能の実現に向けて、メタ記憶に焦点を合わせました。メタ記憶とは、例えば「昨日の朝ごはんは何を食べたか覚えているだろうか？」と考えるときに頭の中で働くような、自分の記憶の存在の有無を調べるような認知機能のことを意味します。本研究グループでは、ヒトの心が世代交代の繰り返しによって進化したのと同様に、計算機の中で進化的なプロセスを実行する人工生命手法を採用しました。

タスクとして、動物のメタ記憶の有無を調べるときによく用いられる「遅延見本合わせ問題」を採用しました。これは、視覚刺激一つ（例えば赤い丸）を動物に提示した後、テスト段階ではそれ以外の視覚刺激（例えば青い四角）を含む複数の視覚刺激を提示し、最初に提示した視覚刺激を動物が選択するとご褒美を与えるというものです。さらに、メタ記憶を調べるために、テスト段階の前にテストに進むかどうかの2択から選ぶ段階も用意します。テスト回避の選択肢を選んだ場合は、ちょっとした褒美を与えます。もし、メタ記憶が働かならば、自分の記憶に自信がある場合はテストに進み、自信がない場合にはテストを回避するでしょう。そうすると、テストの正解率は、テスト前の選択肢がない場合の正解率よりも、不正解になるはずだったテストを受けなかった分だけ向上するはずですが。

本研究では、この問題を単純化したタスク（図1）を用いて進化実験を行いました。各ニューロンにノイズを少しずつ加えることで忘却を表現しました。最初はランダムな構造を持ち記憶能力さえ持たないニューラルネットワーク集団が、世代交代を繰り返すうちに、テストに進むと正解できない、つまり記憶がノイズで壊れがちな場合にはテストを回避するように進化しました（図2, 3）。この研究のポイントは、ニューラルネットワークの構造だけでなく、ニューロン間結合の強さをどういう条件でどのように変えるか、つまり学習方法自体も人手によらずに進化によって決まるといったところです。

本研究の成果は「ヒトらしい心」、さらには意識を持った人工知能の実現の手がかりになることが期待されます。

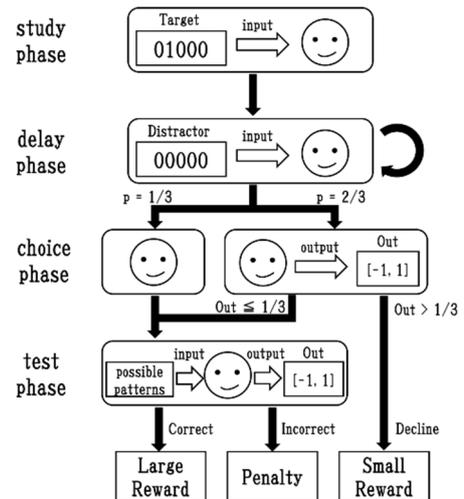


図1. メタ記憶の存在を調べるタスクの手順。01000というパターンが示された後、ノイズが加えられる。その後、確率2/3でテストに進むか選択し、確率1/3でテストに強制的に進む。報酬量は、テストで正解>回避>テストで失敗、の大小関係がある。

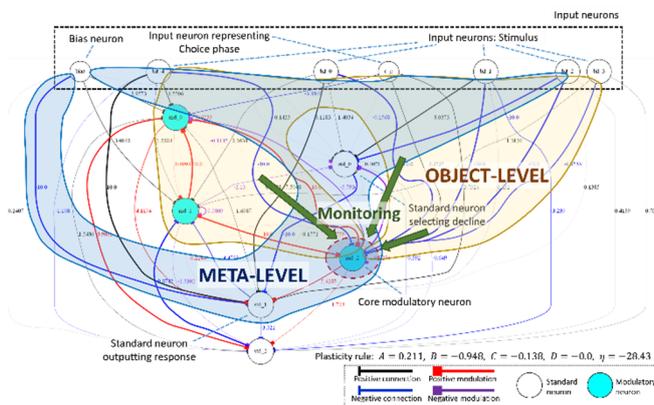


図2. 進化したニューラルネットワーク。最初に示されたパターンを忘れていた場合にはテストを回避する。

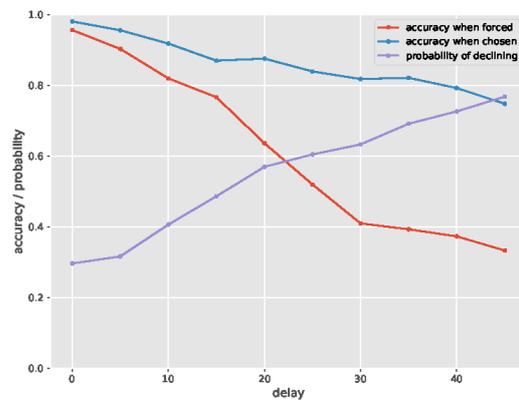


図3. 進化したネットワークのテスト正答率。ノイズを増やす（横軸）につれて、テスト回避の頻度（紫）は増えるとともに、選択して回答した場合の正答率（青）が強制的に回答した場合の正答率（赤）に差をつけるようになる。

【成果の意義】

メタ記憶という高度な認知能力を実現するニューラルネットワークのメカニズムを人手に頼らずに複数発見したことは、「ヒトらしい心」を持つ人工知能の実現に向けた重要な一歩とみなせます。また、動物のメタ記憶の存在を調べる実験の結果に関し、巻き起こってきた議論に対しては、計算機内進化実験による進化妥当性の検証という新たな観点を提供しました。

【用語説明】

注 1) メタ記憶 :

記憶に対する認知。記憶を読み取るモニター機能と記憶内容の変更に関わるコントロール機能に2分できる。

注 2) ニューラルネットワーク :

脳の神経活動から着想を得た数理モデル。ニューロン群とそれらのニューロン間の重み付き有向結合の集合として表現される。

【論文情報】

雑誌名 : Scientific Reports

論文タイトル : Evolution of metamemory based on self-reference to own memory in artificial neural network with neuromodulation

著者 : Yusuke Yamato, Reiji Suzuki, and Takaya Arita (名古屋大学)

DOI : 10.1038/s41598-022-10173-4

URL : <https://www.nature.com/articles/s41598-022-10173-4>