

令和4年12月8日

レーザー光のカオスの遍歴の自発性を用いた高効率な強化学習を実現 ー脳を模倣した意思決定できる AI ハードウェアー

1 ポイント

- ・ レーザ光のカオスの遍歴の自発性を用いて、高効率な強化学習を世界で初めて実現した。
- ・ 脳のダイナミクスを模倣した自立的に意思決定できる機械学習への応用が期待される。

2 概要

埼玉大学大学院理工学研究科数理電子情報部門の内田淳史教授、金沢大学理工研究域機械工学系の砂田哲教授、東京大学大学院情報理工学系研究科システム情報学専攻の成瀬誠教授らの共同研究グループは、マルチモード半導体レーザー(※1)における縦モード(※2)間のカオスの遍歴(※3)を用いて、機械学習(※4)方式の一つである強化学習(※5)における問題例の解決方法を提案し、実験での実証に成功しました。

近年、既存のコンピュータを支える半導体の集積技術には限界が近づいており(ムーアの法則の終焉(※6))、それを打破する情報処理方式を提供するために、光を活用した機械学習に注目が集まっています。その一つの取り組みとして、強化学習の問題例として知られる、当たり確率が未知の複数台のスロットマシン(選択肢)からの報酬の最大化を目的とした多腕バンディット問題(※7)を、光ダイナミクスを活用して解くという研究が行われてきました。しかしながら既存研究では、選択肢の数が多い場合に性能が大幅に劣化するという課題がありました。

本研究では、複数の縦モードを有するマルチモード半導体レーザーにおけるカオスの遍歴現象を利用して、多腕バンディット問題を解く方式を新たに提案しました。カオスの遍歴は脳の自発的機能において重要な役割を担う現象として知られており、本方式に取り入れることで、既存研究で問題となった選択肢が多い場合にも対応でき、従来用いられるソフトウェアのアルゴリズムよりも高効率に意思決定が実現できることを示しました。すなわち、光のカオスの遍歴を利用することで、どんなに選択肢が多くても、既存アルゴリズムよりも少ない試行で、自発的に正しい選択肢を推定できます。そして、本手法の有効性に関して半導体レーザーを用いた実験で検証し、レーザーにおけるカオスの遍歴を用いて強化学習が実現できることを世界で初めて実証しました。

本研究の新規性は、レーザー光のカオスの遍歴を用いて、強化学習の代表的な問題である多腕バンディット問題を解いた点になります。従来はコンピュータ内でソフトウェア的に解く方法が主流でしたが、多腕バンディット問題を効率的に解くための光ハードウェアを開発した点に新規性があります。またカオスの遍歴とは、異なる複数のカオス状態間を遷移する現象ですが、多腕バンディット問題における探索(当たり確率の高いスロットマシンを見つけること)を行う際に、カオスの遍歴と呼ばれる物理現象を用いることで、従来法よりも高効率に探索が行えることを初めて発見しました。

本研究で提案した方式は実験的に実装できるため、今後専用デバイスを開発することで、効率的で高速な強化学習用ハードウェアを実現できる可能性を秘めています。また本研究の結果は、レーザーのみならず、脳のダイナミクスとして知られているカオスの遍歴を強化学習に利用できることを示唆しており、自立的に意思決定する AI 技術の実現が期待されます。

本研究成果は、2022年12月8日午前4時(日本時間)にアメリカ科学振興協会誌『Science Advances』オンライン版に掲載されました。

3 研究の背景

近年、機械学習などにおけるコンピューティングの爆発的需要拡大に対し、既存のコンピュータを支える半導体の集積技術の成長(ムーアの法則)に終焉が近づいており、それを打破するために、GPU(グラフィック・プロセッシング・ユニット)などに挙げられる電子アクセラレータの利用が活発的です。一方で光は電気と比較して高速性や空間波長多重性など高機能化が期待されており、光を活用した機械

学習の研究が世界中で行われています。機械学習の一つである強化学習とは、行動により報酬が得られる環境下において、得られる報酬を最大化するように試行錯誤を通じて行動を学習する方法のことで、強化学習の中でも多腕バンディット問題が知られており、この問題では未知の当たり確率を持つ複数台のスロットマシンの中から、ユーザが1台を選択して報酬を得ることを繰り返し行い、総報酬の最大化を目指します。しかしながら、どのスロットマシンが良いかを探ること(探索)と、探索の結果から良いスロットマシンを選択すること(知識利用)が重要な中、探索と知識利用のどちらに偏ってもならないという、探索と知識利用のジレンマ(※8)が存在しています。これまでに光を活用して多腕バンディット問題を解く研究が行われてきましたが、選択肢の数が多い場合に性能が大幅に劣化するという課題がありました。

研究グループはこの課題を解決するために、カオスの遍歴と呼ばれる不安定な状態間を自発的に遷移する現象を用いました。カオスの遍歴は、脳科学を代表とする多くの研究分野において観測されており、半導体レーザにおいても観測されています。複数の縦モードを有するマルチモード半導体レーザに戻り光を付加することで、光強度が最大となる縦モードが競合して不規則に発振し、カオスの遍歴と見なすことができます。またカオスの遍歴は、脳の自発的な活動を理解するためにも重要視されています。そこで、カオスの遍歴を機械学習のような知能システムに活用することは、自己の意思を有するAI技術の実現が期待できるため、挑戦的な試みであると考えられます。

4 研究内容

本研究では、戻り光を有するマルチモード半導体レーザにおけるカオスの遍歴に着目し、複数のスロットマシンにおける多腕バンディット問題の新たな解決方法を提案しました。半導体レーザでは光ディスクの表面や光ファイバの端面で生じる反射により、レーザ光の一部がレーザ自身に戻り光として注入されて、不規則なカオス出力光が得られます。特にマルチモード半導体レーザを用いた場合には、縦モード間でのエネルギーのやり取りが発生して、カオスの遍歴と呼ばれる不規則なモード競合が生じます。

カオスの遍歴とは、異なるカオス状態の間を自発的に遷移する現象です。マルチモードレーザでは戻り光により一つ一つの縦モードがカオス状態となります。ここで半導体レーザ内部のキャリア密度を介して、縦モード間で結合が生じています。レーザの全エネルギーは一定であるため、一方のモードが強く発振すると他のモードが弱くなるという現象が生じます。戻り光があるとモード間のエネルギーのやり取りが不規則に生じるため、光強度が最大のモードが自発的に遷移します。

ここでカオスの遍歴におけるモード間の光強度の強弱を、スロットマシンの探索に割り当てます。つまり、各モードとスロットマシンを対応させて(例えば1つ目のモードにスロットマシン1を割り当てる)、最も光強度の大きいモードに対応するスロットマシンを選択します(図1)。

さらに外部から強度が一定の光を注入することで、カオスの遍歴の制御を行います。例えば、選択したスロットマシンが当たった場合には、対応するモードに外部光を注入して、注入光量を増加させます。すると、そのモードの光強度が大きくなるため、そのモードの光強度が最大となる確率が増加し、次の選択においてスロットマシンが選ばれやすくなります。反対にスロットマシンがはずれた場合には外部光を弱くすることで、そのモードの光強度が最大となる確率が減少し、スロットマシンが選ばれにくくなります。

このように、スロットマシンの選択 → その結果に応じて外部光の強さを変化 → モードの大きさが変化してスロットマシンの選ばれやすさが変化 → 次のスロットマシンの選択、を繰り返すことにより、当たり確率の高いスロットマシンを自発的に探索できるようになります。

強度が一定の外部光を半導体レーザに注入すると、外部光の波長に最も近い縦モードが引き込まれて、外部光の波長と一致します(インジェクションロッキングと呼ばれています)。ここで外部光は強度が一定のため、外部光の強さを大きくしていくと、その縦モードの発振は安定化します。さらに一つの縦モードが安定化すると、モード間の結合効果により他のモードも安定化します。

外部光を注入された一つの縦モードが安定化する場合、レーザ内の全てのエネルギーがその縦モードに集中し、他のモードの光強度はゼロになります。そのため、外部光を注入されたモードが常に最大の光強度になります。したがって、このモードに対応するスロットマシンが毎回選択されることになり、当たり確率の高いスロットマシンを見つけたことになります。

上述の提案手法の有効性を示すために、レーザの数値モデルを用いて数値計算で調査を行いました。具体的には、Lang-Kobayashi 方程式と呼ばれる、戻り光を有する半導体レーザを記述する微分方程式をコンピュータで解きました。この方程式は、戻り光を有する半導体レーザの実験結果を良く再現

でき、信頼できる数値モデルです。数値計算により、カオスの遍歴によりスロットマシンの探索を自発的に行い、当たり確率が最大のスロットマシンに対応する縦モードの発振が安定化して、多腕バンディット問題を解けることが分かりました(図 2)。またスロットマシンの台数を変化させて、多腕バンディット問題を複数回解き、当たり確率が最大のスロットマシンを選択した割合(平均正答率)を計算しました。その結果、スロットマシンの台数を増加させても平均正答率が 1 に到達しており、多くのスロットマシンにおける多腕バンディット問題を正しく解けることを示しました(図 3)。

さらに、スロットマシンの台数を増加させた時の意思決定性能を調査しました。当たり確率が最大のスロットマシンを選択した割合が 0.95 以上に到達するまでに必要な試行回数と、スロットマシンの台数との関係と調査しました。加えて従来法として、ソフトウェアのアルゴリズムとして知られる UCB1-tuned 法(※9)を用いた場合との比較を行いました(図 4)。スロットマシンの台数が増加した時に、マルチモード半導体レーザ(提案方式)の方が、UCB1-tuned 法よりも少ない試行回数で多腕バンディット問題を解けることを示しています。特にこれらの直線の傾きを表すスケーリング係数が、UCB1-tuned では 1.06 であるのに対し、提案方式では 0.70 に低減しています。これは選択肢の数が増加した場合に、提案手法の方が効率良く解ける事を示しています。例えば、スロットマシンの台数が 5000 台では、マルチモード半導体レーザ(提案方式)の方が、UCB1-tuned 法の約 16 %の試行回数で多腕バンディット問題を解けることを示しています。スロットマシンの台数が増えると、さらに効率的に最良の選択肢を推定することも可能です。

以上のように、選択肢の数が増加した場合に、従来法よりも高効率に多腕バンディット問題を解くことができ、提案手法が優れていることを示しました。本手法はレーザ実験での実証も行われています。

5 今後の展開

本研究で示した方式は、レーザを用いた強化学習という点において、他の光情報処理方式と組み合わせた展開が期待できます。例えば、光通信ネットワークにおいて重要な役割を担う自律的な波長の割り当ての最適化において、本手法を用いた強化学習による高速化や高効率化が期待できます。また今後は、光集積回路により専用デバイスを開発することで、高速な強化学習用ハードウェアを実現できる可能性も秘めています。

さらに本研究の結果は、レーザだけでなく、他のハードウェアにおけるカオスの遍歴を強化学習に利用できる可能性を示唆しています。特に脳のダイナミクスを模倣したカオスの遍歴を用いた強化学習を実現することで、自発的に意思決定を行う AI ハードウェアの実現が可能であり、ヒト型ロボット等への応用が強く期待できます。

6 原論文情報

掲載誌 Science Advances, Vol. 8, eabn8325 (2022).

論文名 Controlling chaotic itinerancy in laser dynamics for reinforcement learning

著者名 Ryugo Iwami, Takatomo Mihana, Kazutaka Kanno, Satoshi Sunada, Makoto Naruse, and Atsushi Uchida

岩見龍吾, 巳鼻孝朋, 菅野円隆, 砂田哲, 成瀬誠, 内田淳史

7 研究助成

本研究は、科学技術振興機構(JST) 戦略的創造研究推進事業(CREST) 研究領域「新たな光機能や光物性の発現・利活用を基軸とする次世代フォトニクス of 基盤技術」(領域総括:北山 研一 大阪大学名誉教授)における研究課題「ナノ光学と光カオスを用いた超高速意思決定メカニズムの創成」(研究代表者:成瀬 誠 東京大学大学院情報理工学系研究科教授、課題番号 JPMJCR17N2)、日本学術振興会科学研究費助成事業(基盤研究 A、課題番号 19H00868)の支援を受けて実施されました。

8 参考図又は写真等

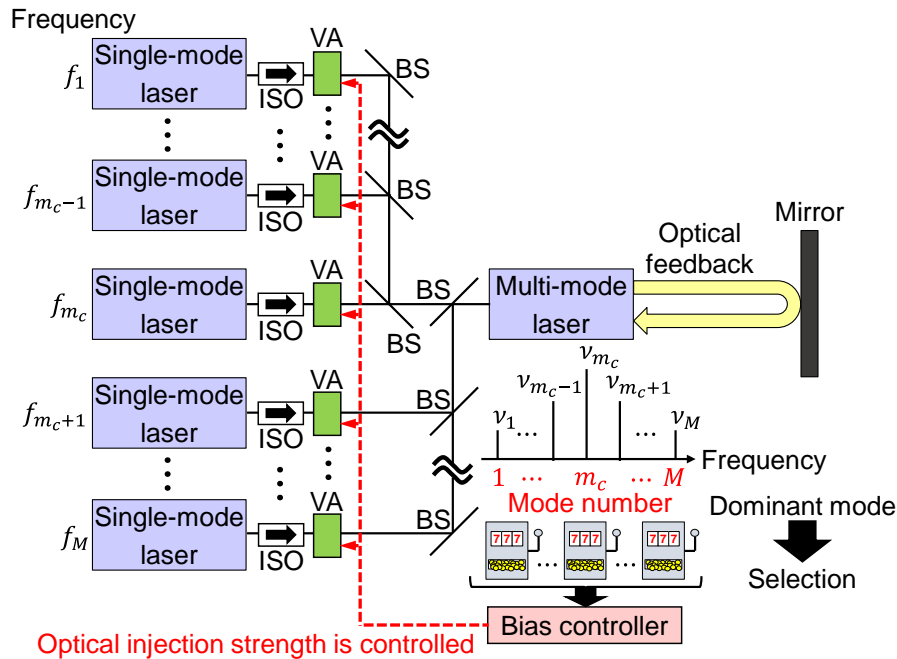


図 1. 本提案方式の概念図を示しています。戻り光を有するマルチモード半導体レーザにより、複数の縦モード間のカオス的遍歴が発生します。半導体レーザ内部でのキャリア密度を介した縦モード間の結合により、モード間でのエネルギーのやりとりが発生します。レーザ内部のエネルギーは一定であるために、あるモードは強くなり、他のモードは弱くなるという現象が生じます。このようなカオス的遍歴は戻り光により発生します。また強度が一定のシングルモードレーザ光を注入することにより、一つの縦モードを強く発振させることができ、カオス的遍歴の制御を行います。外部から一定強度の光を注入して、一つの縦モードが外部光に引き込まれることで、安定化が行われます。一つのモードが安定化すると、結合の効果により他のモードのエネルギーが全て安定化したモードに集中するため、他のモードの発振はゼロになります。このように、強度が一定の外部光をカオス的遍歴状態のレーザに注入して、モードの光強度を安定化させて一つのモードのみ発振させることを、制御と呼んでいます。以上のように、スロットマシンの試行結果に応じて光注入を制御することで、多腕バンディット問題を解くことができます。

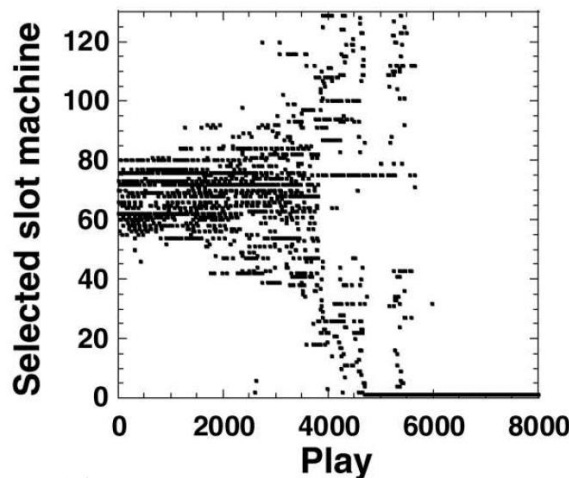


図 2. スロットマシンの試行回数に対する選択されたスロットマシンの番号を示しています。スロットマシンの台数は 129 台に設定しました。スロットマシン 1 の当たり確率を 0.9 に設定し、残りのスロットマシンの当たり確率を 0.7 に設定しました。試行回数が少ない場合には、異なるスロットマシンを選択している様子がわかります。一方で試行回数が増えると、スロットマシン 1 を常に選択し続けています。このよう

に試行回数が多くなるにつれて、当たり確率が最大のスロットマシン 1 が選択され続けていることから、多腕バンディット問題が解けたことが分かります。

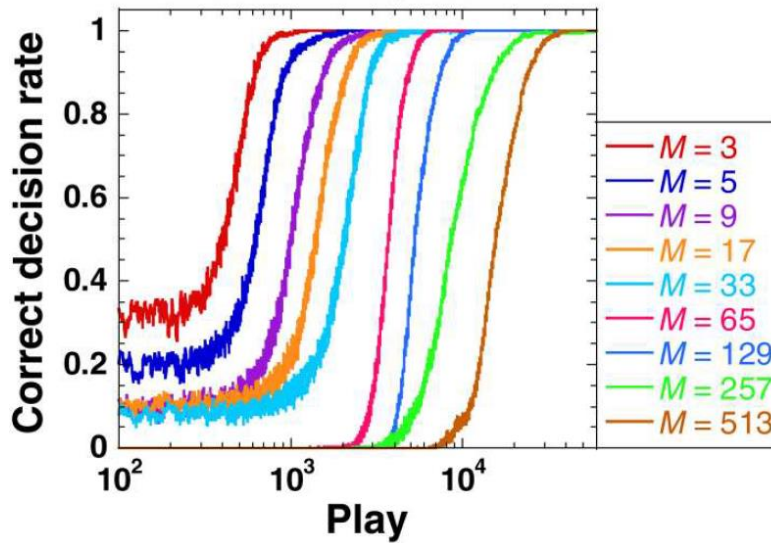


図 3. 試行回数の変化に対する当たり確率が最大のスロットマシンを選択した割合（平均正答率）を示しています。スロットマシンの台数 M も変化させています。図 2 と同様にスロットマシン 1 の当たり確率を 0.9 に設定し、残りのスロットマシンの当たり確率を 0.7 に設定しました。どのスロットマシンの台数の場合でも、試行回数が増えるにつれて、平均正答率は 1 へと収束していることが分かります。

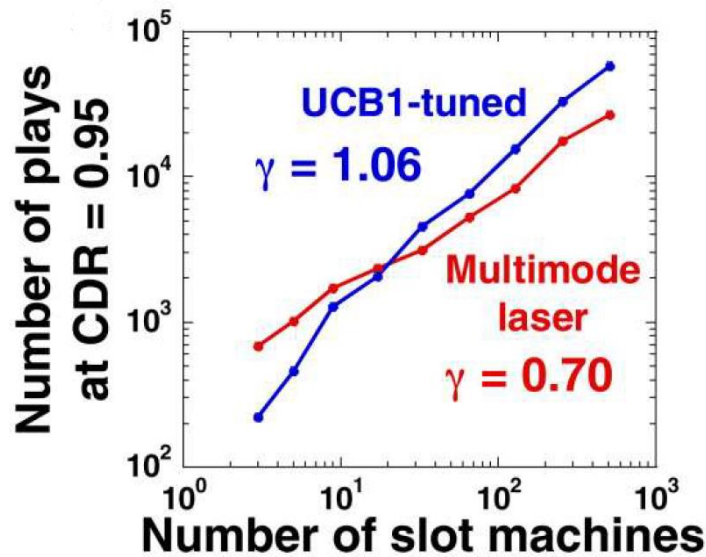


図 4. 当たり確率が最大のスロットマシンを選択した割合が 0.95 に到達するのに要した試行回数と、スロットマシンの台数の関係を示しています（赤色）。ソフトウェアのアルゴリズムとして知られる UCB1-tuned 法と比較しました（青色）。試行回数が低い値ほど良い結果になります。提案手法（マルチモードレーザ、赤色）の方がグラフの傾きが小さく、スロットマシンの台数が増加しても試行回数が少ないため、高効率な意思決定を実現できることが分かります。

9 用語解説

※1 マルチモード半導体レーザ

複数の縦モード（波長）を有する半導体レーザのことです。

※2 縦モード

レーザーにおいて発振する波長を表します。一般的にレーザーでは内部の共振器により光を増幅させており、共振器の長さや屈折率、およびレーザー内部に使用する物質により増幅できる波長が決まります。この時、共振器内で増幅された光の波長のことを縦モードと呼び、複数の縦モードが離散的に分布しています。

※3 カオスの遍歴

数式で記述されるルールに従いますが、その挙動は一見すると不規則である現象の事をカオスと呼びます。カオスの遍歴とは、異なるカオス状態の間を不規則に遷移する現象のことです。

※4 機械学習

人工知能において用いられる主要技術の一つであり、コンピュータがデータから学習して答えを導く方法のことです。

※5 強化学習

行動により報酬が得られる環境下において、得られる報酬を最大化するように試行錯誤を通じて行動を学習する方法のことです。強化学習は、エージェント(ユーザ)と環境の相互作用から成り立ちます。エージェントは、様々な行動のうち環境の状態に合わせて一つの行動を選択し、その行動により環境から報酬を得ます。得られる総報酬の最大化を目指し、エージェントが試行錯誤を通じて行動を学習することで強化学習が達成されます。

※6 ムーアの法則の終焉

半導体の集積密度は、1.5年で2倍になるという経験則がムーアの法則です。この経験則に従って半導体の集積密度が向上してきましたが、その経験則に限界があり、近年、この法則は成り立たなくなっています。

※7 多腕バンディット問題

強化学習における問題例の一つです。当たり確率が未知である複数台のスロットマシンの中から、1台のスロットマシンを繰り返し選択して、スロットマシンから得られる報酬を最大化することを目的とした問題です。多腕バンディット問題は、無線通信におけるチャンネル最適化や、インターネット広告配信、囲碁のAIなどに利用されています。

※8 探索と知識利用のジレンマ

探索とは、多くの報酬を得るために、どの選択が良いかを判断する行動のことです。一方で知識利用とは、探索で得られた結果をもとに、多くの報酬が得られる選択をして報酬を稼ぐ行動のことです。探索を多くすればどの行動が良いかの判断が容易になりますが、知識利用に要する回数が減るため、総報酬が減ります。一方で探索を少なくすると、どの行動が良いかの判断が行えないため、最善の行動を選択できずに総報酬が減ります。このように、探索と知識利用にはジレンマが存在しています。

※9 UCB1-tuned 法

信頼限界上限 (Upper Confidence Bound) を用いた手法であり、スロットマシンの当たり確率の推定値に、その推定値の不確かさを上乗せしてスロットマシンを選択する方法です。この方法はソフトウェアで実装できます。