

レーザカオスの時分割多重によるスケーラブルな強化学習

Photonic Reinforcement Learning by Time-division Multiplexing of Laser Chaos

○成瀬 誠¹, 巳鼻 孝朋², 堀 裕和³, 西郷甲矢人⁴, 岡村和弥⁵, 内田 淳史²

1. 情通機構, 2 埼玉大, 3. 山梨大, 4 長浜バイオ大, 5. 名古屋大

○M. Naruse¹, T. Mihana², H. Hori³, H. Saigo⁴, K. Okamura⁵, A. Uchida²

1. NICT, 2. Saitama Univ., 3. Univ. Yamanashi, 4. Nagahama Inst., 5. Nagoya Univ.

E-mail: naruse@nict.go.jp

人工知能の重要課題である強化学習とは、動的に変化する環境における速やかで的確な意思決定であり、ネットワークの瞬時最適化等の多くの重要な応用の基礎をなす¹。我々は、レーザカオスの高速性と乱雑性を活用し、当たり確率の未知なスロットマシンからの獲得報酬を最大化する問題（多本腕バンディット問題（Multi-Armed Bandit problem: MAB））の解決に成功した²。しかし、マシンの台数は2台であり実際的な応用で求められる問題規模に対応するためのスケーラビリティは未解決であった。本研究では、レーザカオスからのサンプリングを時間多重化することによるスケーラブルな原理を示した上で、実際のカオス光を用いた実験結果を示す。また、コンピュータ上で生成した擬似乱数及びカラーノイズとの比較等も議論する。

2^N 個のマシンから当たり確率の最も高い台を選択する問題を考え、各マシンのIDを2進数で表す。時刻 t_1 でサンプリングした光強度が閾値 TH_1 より小さいときは、選択する台の最上位ビットを0とする。次に、時刻 t_2 での光強度が閾値 $TH_{2,0}$ より大きいときは、選択する台の上位から2ビット目を1とする。これを必要なビット数だけ繰り返すことにより最終的に選択する台が決まる(Fig. 1a)。選んだ台から得られた報酬に基づいて閾値を更新する。この原理は単一光子を用いた階層型意思決定³を時間領域で多重化したものでもある。

実験では64台までの動作を確認した(Fig. 1b)。擬似乱数、カラーノイズよりも性能に優れることも確認された。さらに、同一課題に対して異なるカオス光を用いたとき、特定のカオス光が他より優れる傾向が見られ、カオスの拡散性によって一定程度説明できることを示す。

謝辞 本研究の一部は科学技術振興機構 CREST

(JPMJCR17N2), 日本学術振興会科研費基盤 A (JP17H01277), 基盤 B (JP16H03878), 研究拠点形成事業 (A.先端拠点形成型) による。

参考文献

- 1) D. Silver, *et al.*, Nature **550**, 354 (2017).
- 2) M. Naruse, *et al.*, Sci. Rep. **7**, 8772 (2017).
- 3) M. Naruse, *et al.*, ACS Photonics **3**, 2505 (2016).

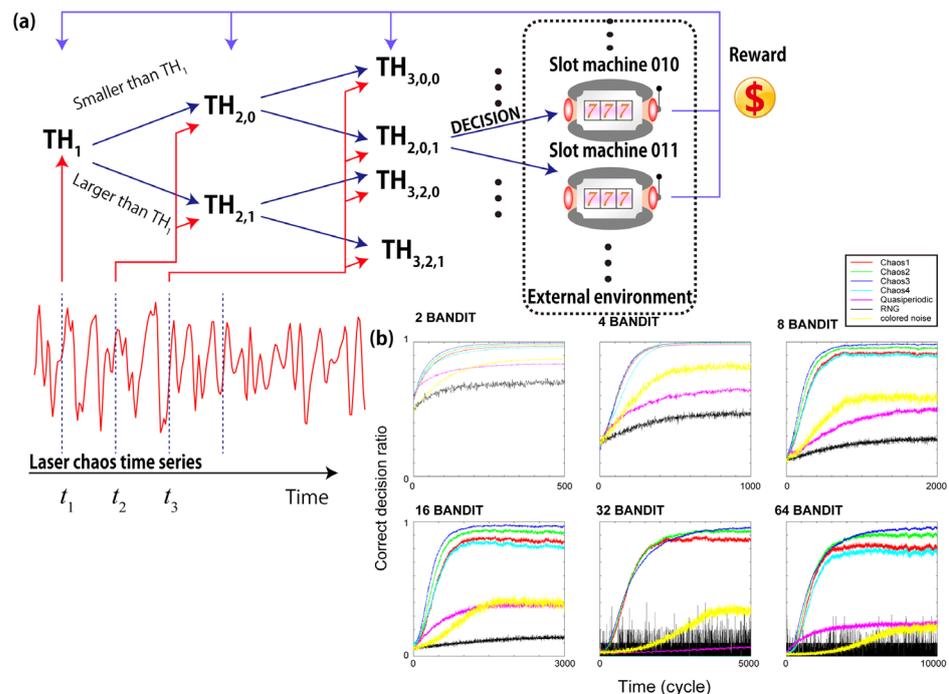


図 1(a) レーザカオスの時分割多重によるスケーラブルな強化学習の原理 (b) 性能評価