

フォトニック Q 学習における状態探索の解析 Analysis of state exploration in Photonic Q-learning

○ 漆原 昂¹, ニコラ ショヴェ¹, 砂田 哲², 菅野 円隆³, 内田 淳史³, 堀崎 遼一¹, 成瀬 誠¹

1. 東大情報理工, 2. 金沢大, 3. 埼玉大

○ T. Urushibara¹, N. Chauvet¹, S. Sunada², K. Kanno³, A. Uchida³, R. Horisaki¹, M. Naruse¹

1. Univ. Tokyo, 2. Kanazawa Univ., 3. Saitama Univ.

E-mail: t-urushibara@g.ecc.u-tokyo.ac.jp

我々は光を用いた知的機能創生を目指した研究を行っている。1) では、強化学習の基礎問題であるバンディット問題を、レーザーカオスから得た乱数列である光カオス時系列を用いて解決した。2) では、バンディット問題とは異なり、複数の状態間を遷移する一般の強化学習を扱った。その提案手法をフォトニック Q 学習と呼ぶ (図 2)。実験では Cart-Pole Balancing のシミュレーションを取り上げた³⁾。ここで、状態は Cart の位置と速さ及び Pole の角度と角速度の 4 変数で記述されるが、この研究では各変数ごとに 6 つに離散化し、計 1296 状態を扱った。実験の結果、提案手法は従来手法 (ϵ -貪欲法を用いた Q 学習) よりも高速であることが示された。しかしながら、実験結果に光カオス時系列の影響を見出すことは困難であった。特に学習の速さに関しては、提案手法は、提案手法で光カオス時系列の代わりに正規乱数列を用いた場合と同程度だった。本研究では、2) の実験の更なる解析により、光カオス時系列の自己相関 (図 1) が学習過程への影響を与えることを確認した。

まず、図 2 に示されるフォトニック Q 学習について説明する。本手法では、閾値を学習の対象とする。各状態に同一の光カオス時系列を設定する。学習においては、各状態に至る度に、その状態に設定された光カオス時系列の乱数を順番に 1 つずつ取り出す。その値と現在の状態の持つ閾値の比較から行動選択を行う。行動の結果として起こる状態遷移と、状態と行動選択の履歴から閾値の表を更新する。以上を繰り返す。

上述の通り、この手法は高速な学習を実現する一方で、学習の速さには光カオス時系列の特徴は陽には現れない。そこで本研究では、学習過程を一定の反復回数ごとにグリッドし、その各々のグリッドの中で実際に遷移した状態の種類数 (Number of States Explored) をカウントした。ただし、Number of States Explored が大きい値であることは、より多くの状態の探索の実現という肯定的な考察も、望ましい状態への絞り込みが甘いという否定的な考察も可能なことに注意する。光カオス時系列の特徴として、サンプリング周期により図 1 のように自己相関が変化するという点に着目し、異なるサンプリング周期間での比較を行った。

10 種類のサンプリング周期ごとに、8000 回の学習を行い、各学習段階での Number of States Explored の平均を計算した。その結果のうち、サンプリング周期 10, 30, 50, 90 ps を抜粋したものが図 3 であり、横軸が学習の進行、縦軸が Number of States Explored に対応する。各グラフのピーク値、即ち横軸の値が 20 のときの Number of States Explored を、図 3 に載っていない 20, 40, 60, 80, 100 ps も含めた 10 通りのサンプリング周期間で比較したものが図 3 中の小図の青線である。このグラフは図 1 の自己相関のグラフと同じ傾向を示している。ゆえに、光カオス時系列の自己相関が、フォトニック Q 学習の探索に強く影響していることが推察される。

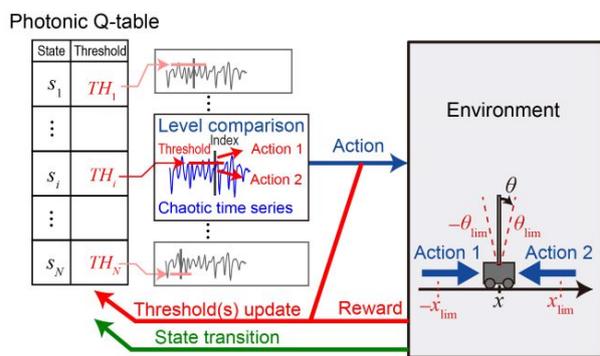


Fig. 2: Photonic Q-learning.

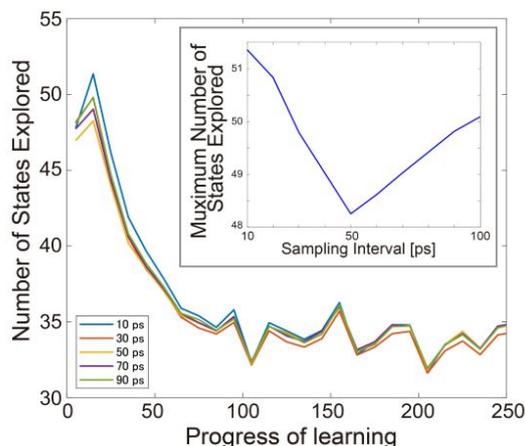


Fig. 3: Reached State Number.

謝辞: 本研究の一部は JST CREST(JPMJCR17N2)、JSPS 科研費 (JP20H00233) の支援を受けた。

参考文献: 1) M. Naruse, et al. Sci. Rep. 7, 8772 (2017).

2) T. Urushibara, et al. JSAP, 9p-Z28-18, Sept., 2020.

3) A.G. Barto, et al. IEEE Transactions on Systems, Man, and Cybernetics, vol. SMC-13, pp. 834-846, Sept./Oct. 1983.