

レーザーカオス時系列を用いた Q 学習の提案と実証

Q-Learning based on Chaotic Laser Time Series

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

1. 東大情報理工, 2. 金沢大, 3. 埼玉大

○ T. Urushibara¹, N. Chauvet¹, S. Kochi², S. Sunada², K. Kanno³, A. Uchida³, M. Naruse¹

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

E-mail: takashi_urushibara@ipc.i.u-tokyo.ac.jp

我々は光を用いた知的機能創成を目指した研究を行い、これまでにレーザーカオスを用いてバンディット問題の解決等の意思決定の実証研究を行ってきた¹⁾。バンディット問題は、状態数 1 の簡易な状況を扱う基盤であり、高次の問題への発展可能性に課題があった。そこで本研究では、Q 学習における Q-table の学習を、並列された多数のバンディット問題と捉え、特徴的な時間構造を含んでいるレーザーカオス時系列を用いて解決する原理を提案し実証する。

Q 学習で扱う問題として Cart-Pole Balancing を考える²⁾。これは、図 1 のような倒立振り付きの車を一定間隔ごとに左か右に押す行動を選択し、振子が倒れず、なおかつ車が原点より遠く離れないようにするものである。従来の Q 学習では、離散的に区分けされた状態 s_i で各行動 a_j を選択することの評価値 $Q(s_i, a_j)$ の表である Q-table を学習する。本研究では図 2 のように各状態に対して閾値を Q-table とする。また、各状態に対して同一の不規則時系列を用意し、当該状態において時系列の数字を順次読み出し、それが閾値未満のときは行動 1 を、閾値以上のときは行動 2 を選択するとする。その結果により閾値を更新する。すなわち、ある状態での行動選択を 2 本腕バンディット問題と見なす。本研究ではレーザーカオス (サンプリング周期: 10, 20, ..., 100 ps) を時系列として用いた。また、比較対象としてカオス時系列の順序をランダムに入れ替えたもの (シャッフルカオス)、一様乱数、正規乱数も用いた。

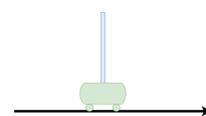


Fig. 1: Cart-Pole

倒立を開始後、ある回で失敗するか、失敗せずに連続 150 回行動するまでを 1 エピソードとする。1000 エピソードに渡って学習を続け、この 1000 エピソードを 1 トライとする。100 個の異なる時系列について異なる 8 個の区間を用い、その各々について 10 トライずつ、計 8000 トライを行った。

学習の推移を成功試行数の平均で評価すると、一様乱数は学習が遅く、その他はほぼ同様の速度で学習が進捗した。ただし、学習の後半部について 145 回アクションをとる前にエピソードを終えた試行数の平均は図 3 のように評価された。この指標は、試行が直ちに失敗してしまうような、際だって性能の悪い事象の生起を大きく反映する。図 3 中の青線、赤線は各々レーザーカオス時系列及び時系列の順序をランダムに入れ替えた時系列に対する結果を示す。多くのサンプリング周期で元のレーザーカオス時系列がよい性能を示している。また、一様乱数 (黄点線)、正規乱数 (緑点線) より優れた性能を示している。この一因はカオス時系列に含まれる時間相関にあると考えられる。

State	s_1	s_2	...	s_i	...	s_N
Action1				$Q(s_i, a_1)$		
Action2				$Q(s_i, a_2)$		

↓ Bandit
↓ Bandit
↓ Bandit
↓ Bandit

State	s_1	s_2	...	s_i	...	s_N
Action-Threshold				$Th(s_i)$		

Fig. 2: Conventional Q-table (upper) and proposed Q-table (lower) designed for chaotic time series.

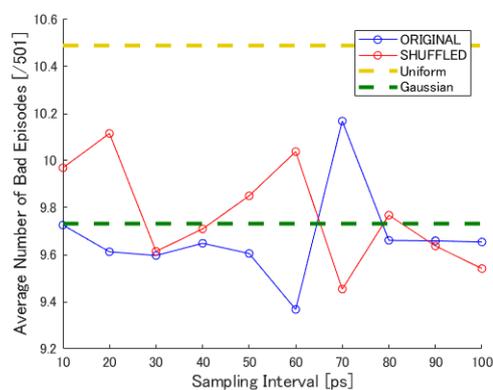


Fig. 3: Average number of bad episodes in one try.

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