

時空間光ダイナミクスを用いた意思決定実験 Experiment on decision making using spatiotemporal optical dynamics

埼玉大¹, 東大情理²

○森尻 健正¹, 武埴 賢人¹, 菅野 円隆¹, 成瀬 誠², 内田 淳史¹

Saitama Univ.¹, Univ. Tokyo²

○Kensei Morijiri¹, Kento Takehana¹, Kazutaka Kanno¹, Makoto Naruse² and Atsushi Uchida¹

E-mails: k.morijiri.632@ms.saitama-u.ac.jp, auchida@mail.saitama-u.ac.jp

はじめに: 近年, 多くの分野で強化学習が用いられている. 強化学習の問題例の1つとして, 多腕バンディット問題が挙げられる[1]. 多腕バンディット問題とは, ある確率で報酬が得られる複数台のスロットマシンから 1 台のスロットマシンを選択することを繰り返し, 得られた結果から学習を行い, 報酬の最大化を目的とする問題である.

多腕バンディット問題を解決する手法として, 綱引き理論が提唱されている[2]. 綱引き理論を用いた複数の半導体レーザカオス波形比較による意思決定が提案されている[3]. この手法では, 複数の半導体レーザカオス波形に対してバイアスを導入し, 振幅の最大値を示すスロットマシンを選択する. 選択したスロットマシンの結果に応じて, バイアスを変化させることで意思決定を達成する. これにより数値計算において, 最大で 1024 台のスロットマシンの多腕バンディット問題の意思決定に成功している[3]. しかしながら, この手法ではスロットマシンの台数に対応する数の半導体レーザが必要となるため, 物理実装が困難である.

一方で, 空間光変調器を用いた多数のノードを持つリザーバコンピューティングの実験実装が報告されている[4]. 空間光変調器を正方形の領域で分割し, 各領域をノードとすることで多数のノードの実装を行っている. この手法では, 最大で 16384 個のノードを持つリザーバコンピューティングが達成されている.

そこで本研究では, 半導体レーザの時空間光ダイナミクスを用いた意思決定手法を提案し, 意思決定を実験的に達成することを目的とする.

方法: 時空間光ダイナミクスを用いた意思決定の実験装置図を Fig. 1 に示す. 空間光変調器と偏光フィルタを用いてレーザ光の強度変調を達成する. 強度変調された光の空間パターンをカメラで取得し, コンピュータ (PC) に送信する. 空間パターンをスロットマシンの台数で分割し, 対応する領域の光強度にバイアスを PC で可算する. 最大値に対応するスロットマシンを選択し, 結果に応じてバイアスを変化させる. カメラで取得された各領域のデータに対して, フィードバック強度 β を乗算し, SLM にデータを送信する. この操作を繰り返すことにより, 意思決定を達成する.

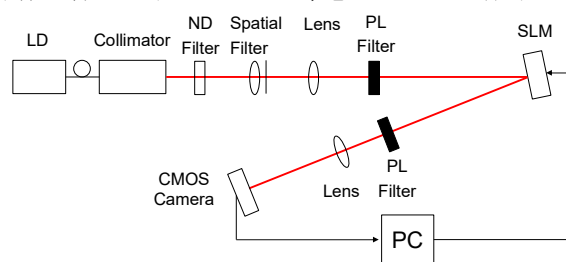


Fig. 1. Experimental setup. LD, laser diode (wavelength 642 nm); SLM, spatial light modulator; PC, personal computer; PL, polarization filter.

結果: スロットマシンの台数が 16 台である多腕バンディット問題の結果を Fig. 2 に示す. スロットマシンの当たり確率は $(P_1, P_2, P_3, P_4, \dots, P_{2m+1}, P_{2m+2}) = (0.7, 0.5, 0.9, 0.1, \dots, 0.7, 0.5)$ と設定した. つまり最も高い当たり確率である 0.9 をスロットマシン 3 に割り当て, 意思決定を行った. Fig. 2(a) はスロットマシンの選択状況を示す. スロットマシン 3 を選択した場合を赤点で示し, その他のスロットマシンを選択した場合を黒点で示す. 全てのスロットマシンを選択した後, 最適なスロットマシンであるスロットマシン 3 を選択していることが分かる.

Fig. 2(b) は平均正答率 (Correct Decision Rate) を示す. 平均正答率とは, 各試行回数に対して報酬確率が最大であるスロットマシンを選択した回数の割合を示す指標である. 約 700 Play で平均正答率は 1 に到達し, スロットマシン 3 を選択していることが分かる. このように, スロットマシンの台数が 16 台である多腕バンディット問題において, 意思決定を実験的に達成した.

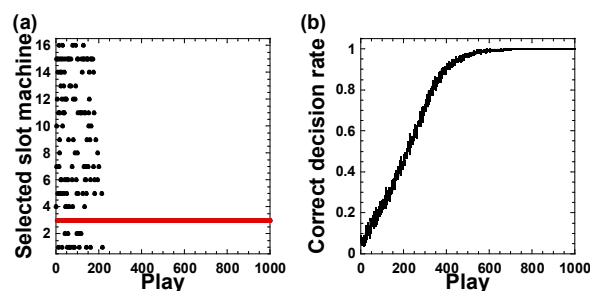


Fig. 2. (a) Slot machine selection. (b) Correct decision rate. The number of slot machine is $N = 16$. The best slot machine is slot machine 3.

まとめ: 本研究では, 半導体レーザの時空間光ダイナミクスを用いた意思決定手法を提案し, 意思決定を実験的に達成した. 空間光変調器と偏光フィルタを用いてレーザ光の強度変調を行い, 複雑な時空間ダイナミクスを生成した. 光の空間領域を分割し, 各スロットマシンに割り当てることで 1 つの半導体レーザから多数の時間波形の生成を行った. それらの時間波形を用いた意思決定実験を行い, 16 台のスロットマシンの多腕バンディット問題において, 意思決定を実験的に達成した.

参考文献

- [1] H. Robbins., et al, Bull. Am. Math. Soc., **58**, 527 (1952).
- [2] S. J. Kim., et al, New J. Phys., **17**, 083023 (2015).
- [3] K. Morijiri, et al, Proc. NOLTA, **1**, 290 (2020).
- [4] P. Antonik, et al., Nature Machine Intelligence, **1**, 530 (2019).