

量子干渉を用いた協調的意思決定とマルチエージェント強化学習への応用

Collective decision making through quantum interference and its application to multi-agent reinforcement learning

東大情理¹, グルノーブルアルプス大²(M1) 新川 浩彬¹, Nicolas Chauvet¹, Guillaume Bachelier², Serge Huant², André Röhm¹,堀崎 遼一¹, 成瀬 誠¹ Univ. Tokyo¹, Univ. Grenoble Alpes²(M1) H. Shinkawa¹, N. Chauvet¹, G. Bachelier², S. Huant², A. Röhm¹, R. Horisaki¹, M. Naruse¹

E-mail: ahiroshinkawa@gmail.com

異なる報酬確率をもつ複数のスロットマシンが存在し、予め決められた選択回数の中で得られる報酬和を最大化する問題をバンディット問題と呼ぶ。各タイムステップでどのマシンを選択するか意思決定を行う必要があるが、複数のエージェントが存在し同じマシンが選択された場合、報酬が分割される。Amakasu らは軌道角運動量が量子干渉した光子対を用いることで、エージェント同士が直接に意思疎通を行わずとも、選択競合を完全に回避する協調的意思決定が可能であることを示した¹⁾。

バンディット問題には状態遷移が存在しないが、ここでは状態遷移を伴うより一般的な強化学習に対し協調的意思決定を適用する。本研究では手始めに5×5のグリッドワールドを考える。エージェントは各マス(状態)で上下左右いずれかの行動を取り、これにより別状態へ遷移する。本研究では複数のエージェントが同時に環境を探索しながら、全ての状態-行動ペアに関して最適なQ値を求めることを目的とする。ただし、複数のエージェントが同時に同じペアを選択した際は1つの更新のみ反映されたとする。この時、状態-行動ペアをバンディット問題の腕、Q値の更新量を報酬と見ることで、各時刻でどの腕を探索するかという意思決定問題を見出すことができる。

意思決定に関し以下の2軸で分析した。1軸目は腕選択の方式であり、ランダム選択とソフトマックス法の比較考量、2軸目は複数エージェントの選択競合の有無である。選択競合の回避はAmakasu らが提案した量子干渉を用いた手法を拡張することで、エージェントの直接的な意思疎通なしに実現される。図1は50体のエージェントが環境を探索する際の、最適なQ値と学習中のQ値の差の時間発展を表している。ランダムな腕選択よりもソフトマックス法を使用した方が、また選択競合を回避した方が収束が速いことが分かる。図2は、競合回避をしない場合に有効な更新の割合(競合時は1つのみが有効)を示している。エージェント数の増加に伴って競合が多くなるため、量子干渉による競合回避がより重要になると言える。

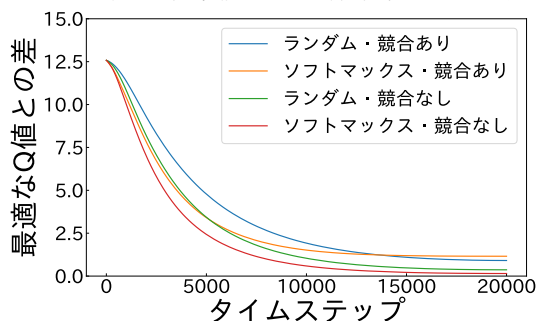


図1: 最適なQ値との差の時間発展

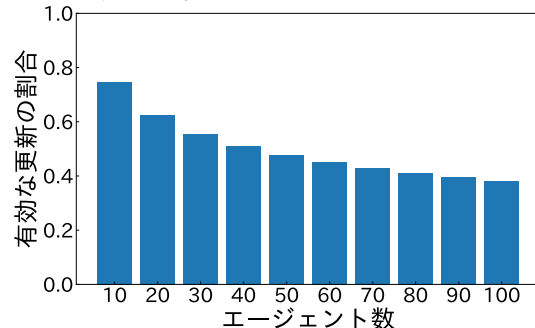


図2: エージェント数と有効な更新の割合の関係

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) T. Amakasu, *et al.* Conflict-free collective stochastic decision making by orbital angular momentum of photons through quantum interference, *Sci. Rep.* **11**, 21117 (2021).