

非定常多腕バンディットゲームの集団実験と ESS ナッシュ戦略の検証

北里大理^A, 弘前大理工^B

中村 隆造^A, 守 真太郎^B

Experimental study of social learning collectives in restless multi-armed bandit

^AKitasato Univ., ^BHirosaki Univ.

R. Nakamura and S. Mori^B

他個体の選択や振る舞いを参考にして選択、行動を行う情報伝達、学習を社会的学習と呼び、自力で情報を探す個人的学習と比較して得られる情報の精度が悪くなくコストも低いことから広く使われている。本講演では非定常多腕バンディット (rMAB) ゲームを用いた社会的学習の2つの集団実験 (実験1, 実験2) の結果について報告する。

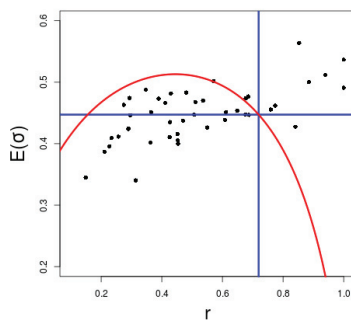


図1 実験1の結果.

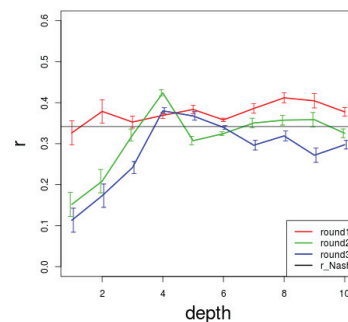


図2 実験2の結果.

バンディットの状態、エージェントの状態は2値 ($\sigma = 0, 1$) であり、 $\sigma = 1$ でリターンが正、 $\sigma = 0$ でリターンが0の状態を表す。 $\sigma = 1$ のバンディットは1台で、確率 q_c で $\sigma = 0$ に変化し、 $\sigma = 1$ のバンディットが追加される。 $\sigma = 1$ のエージェントは $\sigma = 1$ のバンディットが0に変化した時に、 $\sigma = 0$ に初期化される。エージェントのパフォーマンスは σ の時間平均で評価する。エージェントが混合戦略を採用するとき、社会的学習の確率を r で表すと、系は唯一の ESS ナッシュ解を持つ [1]。実験1では、ESS ナッシュ戦略のエージェント9体とヒトが対戦するインタラクティブゲーム、実験2では3人のヒト同士が対戦するゲームで実験を行った。実験1の被験者数は44人、実験2は5チーム15名である。実験1の被験者は約1000ターン、実験2の被験者は30分3ラウンドの実験を行い、被験者3名は順番に選択を行った。1ラウンドあたり約316ターンとなり、ラウンド間に作戦タイムとして10分の休憩時間を与えた。左図は被験者の社会的学習の比率の時間平均 \bar{r} と σ の時間平均 $\bar{\sigma}$ をプロットしたもので、青の実線は ESS ナッシュ解の r の値と σ の期待値 $E(\sigma)$ を、赤の実線は r による $E(\sigma)$ をプロットしたものである。 r の値はばらつきが大きく、被験者平均 ($r = 0.51(3.)$) とナッシュ解 ($r = 0.72$) は一致しなかった。右図は実験2での社会的学習の比率を、バンディットが $\sigma = 0$ に変化してからのターン数 (depth) に対してプロットしたものである。被験者平均を実線で、標準誤差をエラーバーで示している。ラウンドが進むにつれ、Nash 解に近づくことが分かる。

[1] K.Nakayama, M. Hisakado and S.Mori, Sci. Rep.vol.7,No.1937(2017).