6M-06

動学的意思決定におけるゲームの構造学習の分析

今川　裕貴†　　下川　哲矢‡

一橋大学大学院　経済学研究科†　　東京理科大学　経営学部経営学科‡

研究背景

ゲーム理論における混合戦略均衡に関して、従来から実験室における実験が数多く実施されてきた(O'Neill(1987), Rapoport and Boebel(1992)など)。たとえばO'Neill(1987)では、実験室内で完備2人ゼロ和ゲームをスチューデントに対して行い、混合戦略均衡が人々の行動と整合的であることを報告している。また、現実のゲームを使った実験室外での検証もいくつか行われている。(M. Walker and J.Wooders(2001),Ignacio Palacios-Heurta (2003)など)。

Ignacio Palacios-Heurta (2003)はプロサッカーリーグで行われたペナルティキックのデータを引用し、プロのゲームに対する行動を分析した。ペナルティキックはone-shot 2人ゼロ和ゲームであり、各プレイヤーの行動はほぼ同時に行われ、各プレイヤーの利得も直ちに更新される。この論文で興味深い点は、ペナルティキックではAction に対する成功か失敗かの確率が事前には明らかでないため、被験者は単にゲームの混合戦略だけでなく、ゲームの構造そのものを学習する必要がある。彼らの結果は、このような構造理解が必要なゲームに対しても、プロフェッショナルは非常に合理的にプレーを行っていることを示している。

Ignacio and Oscar(2008)ではさらに、ペナルティキックに準じたゲームを実験室内でモデリングし、プロのサッカープレイヤーとアマチュアプレイヤー（スチューデント）の行動を分析・比較した。彼らは、①プロは一部条件を満たさないものの、フィールド内と同様に合理的にプレーを行うこと,②スチューデントは均衡から乖離し、合理的にプレーを行うことができないことを、明らかにしている。

研究内容

本研究では、これらの結果を受けて、どのように人々が混合戦略均衡を学習するのかを実験データを用いて分析する。特に注目したいのは、ゲームの環境変化が学習に与える影響である。学習理論におけるこれまでの多くの研究では、ペイオフ関数などのゲームの構造が所与であり、その意味で被験者は混合戦略均衡のみを探索している。しかし、これらの結果は、もしゲームの構造そのものが変化した場合、維持できないのではないかと予想している、たとえばゲームの構造が不確実な場合、混合戦略均衡を学習する時間が非常に長く必要になったり、必ずしも均衡に収束しなくなったりするのではないかと予想している。

現実の経済活動では、ゲームの構造(ルールやペイオフ)そのものが動的に変化すると考えるのが自然である。その意味で、ゲームの構造変化に関する学習過程を明らかにし、それをモデル化することは、一定の意義があると考える。以上の背景をもとに、本研究では、構造理解が必要な定和ゲームの実験を行う。実験対象はスチューデントとし、2人一組で行うこととする。

今回の実験では一定の確率で利得表が変化し、プレイヤーがペイオフの構造に対して不確実性をもつゲームを行う。これらのゲームは、甲.サッカーのペナルティキック　乙.テニスのサーブを模したゲームである。具体的には、基準となる利得表は利得が 1，0 となるゼロ和ゲームであり、確率分布に従って変化するように設定する。

以上の実験では、Matlab で作成したアプリケーションを UI(図 1、2)として使用し、16 人分の有用なデータを取得した。

An Augmented reinforcement learning model considering structural changes under uncertainties

† Imagawa Yuki・Graduates School of Economics, Hitotsubashi University

‡ Tetsuya Shimokawa・School of Management, Tokyo University of Science

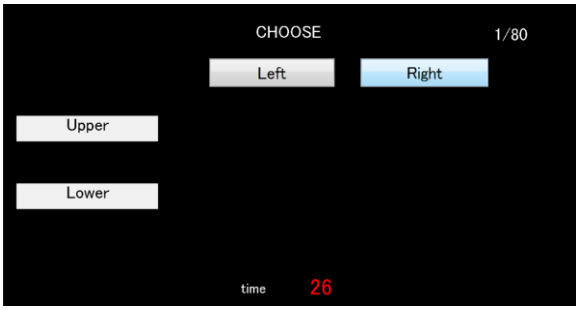
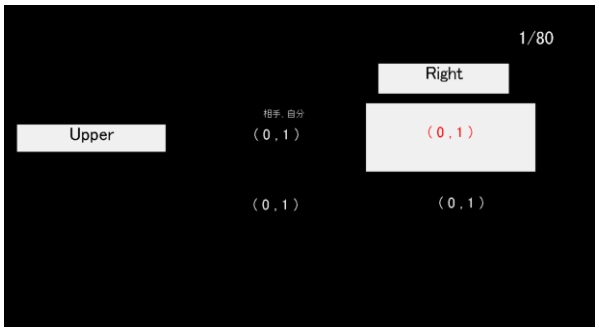


図 1: 実験選択画面(上)　図 2:　実験結果表示画面(下)



導入する学習モデル

当実験では①3 次変数強化学習モデル　②新たに構造理解に対する項を加えた学習モデル、並びに A.通常のモデル　B.Logit モデル、それぞれを組み合わせた 4 つのモデルを用い学習過程の分析を行う。(以下①-A, ①-B,②-A, ②-B とする。) 3 次変数強化学習モデルとは Erev and Roth(1998)で用いられたモデルであり、強化学習モデルは人々の学習プロセスを表す代表的なモデルとして一定の評価を得ているが、今回注目するゲーム構造の学習には十分に対応しているとは予想しづらい。各々の利得表が選ばれる確率を学習する項を加えたものを、強化の増分として定義しなおす。このモデルの意味としては、プレイヤーが試行を経験するごとにプレイヤーの予測する利得表の変化確率がアップデートされていくというものである。

分析

分析は、1. モデルのパラメータ推定並び、モデル化及び適合度の検定を行う。

まず各モデルのパラメータを、MSD 値によって数理計算ソフト「Matlab」を用いて推定する。推定するパラメータは、　初期強化の値 q(1), 忘却のパラメータ φ, 試行錯誤のパラメータ εである。具体的にはグリッドサーチ法を用いてq(1)は、1 刻み、そのほかのφ, εは、0.01 刻みで,それぞれ 0<q(1)<2000,　$0 < \varphi, \varepsilon < 1$の範囲で推定し、以下を推定値とする

$$q(1)\varphi, \varepsilon = \text{argmin MSD value}$$

$$\text{where MSD value} = \sum_{t\in T}\sum_{n\in N}\sum_{j\in M}\frac{\left[p_{nj}(t)-I(s_i(t),j)\right]^2}{T\cdot N\cdot M_n}$$

なお、それぞれ、有限回 T、被験者の数 N、プレイヤー n の行動の数M_n、t 期に被験者 n が選んだ行動$s_n(t)$とする。$I(s_i(t),j)$は指示関数で、以下のようになる。

$$I(s_i(t),j) = \begin{cases} 1 \ if \ \ (s_i(t) = j \\ 0 \ if \ \ (s_i(t) \neq j \end{cases}$$

今回の実験の結果から、それぞれのパラメータの値、MSD 値は以下、図 2 のようになった。

甲.PK	eptheron	phi	q(1)	MSD
①-A	0.25	0.12	1474	0.2461
①-B	0.02	0.08	560	0.2466
②-A	0.09	0.1	1207	0.2477
②-B	0.55	0.17	596	0.2470
乙.Tennis	eptheron	phi	q(1)	MSD
①-A	0.27	0.14	28	0.2405
①-B	0.6	0.9	958	0.2386
②-A	0.29	0.15	13	0.2417
②-B	0.83	0.49	271	0.2400

図 2: 分析結果

考察

実験結果から、MSD 値を比べると、甲では①-A、乙では①-B がモデルとして説明度が高いものであることがわかる。新たに構造理解に対する項を加えたモデルについては、今回どちらの実験においても適合度の高いモデルとは言えなかった。これは、プレイヤーごとに主観確率の形成過程が異なることが一つの要因であると考えられる。(主観確率形成時のバイアスの存在など)。また、ゲームの構造学習が複雑すぎたために、各プレイヤーが強化学習のようなシンプルなプロセスをたどったことも考えられる。

以上をもとに、今後はプレイヤーの属性ごとの分析などを行うことでより精緻なモデルを組み立てる必要がある。

主要参考文献

Ignacio Palacios-Heurta and Oscar Volij, Experientia Docet: Professionals Play Minimax in Laboratory Experiments, *Econometrica*, vol. 76(1), 71-115, 2008

Erev, I. & Roth, A., "Prediction how people play games: Reinforcement learning in games with unique strategy equilibrium". *American Economic Review*, 88, 848-881, 1998