数理解析研究所講究録
第1548巻 2007年 178-185

一般化マッチングペニーゲームに関するエージェントベース・シミュレーション分析

広島大学大学院工学研究科　西崎 一郎 (Ichiro Nishizaki),
中倉 剛 (Tsuyoshi Nakakura), 林田 智弘 (Tomohiro Hayashida)
Graduate School of Engineering, Hiroshima University

Abstract—In this paper, to investigate the long-run behavior of players in the generalized matching pennies game, we employ an approach based on adaptive behavioral models and construct an agent-based simulation system in which artificial adaptive agents have mechanisms of decision making and learning based on neural networks and genetic algorithms. We examine the strategy choices of agents and the obtained payoffs in the simulations, and compare the predictions of the Nash equilibria, the experimental data, and the results of the simulations with artificial adaptive agents. Moreover, we also seek similarities between the behaviors of the human subjects in the experiments and those of the artificial adaptive agents in the simulations.

1. はじめに

行プレイヤーと列プレイヤーで利得構造が非対称となる一般化マッチングペニーゲームには，唯一の混合戦略によるNash均衡が存在するが，実験での被験者の行動は一般にNash均衡予測と異なることが報告されている (Ochs, 1995; McKelvey and Palfrey, 1995; McKelvey *et al.*, 2000; Goeree and Holt, 2001)．一般化マッチングペニーゲームに対する被験者実験のみならず，ゲームの理論の予測の是非について多くの実験的研究が積み重ねられてきている．実験的研究では，理論モデルに沿った環境を実験室の中で構築し，被験者には金銭的動機付けをして実験がなされるが，繰返し実験の回数，被験者の人数，さまざまなパラメータ設定などに対して，被験者を伴う実験では，制約があり，実験の再現性に関しても困難さが指摘できる．

本論文では，計算機上で一般化マッチングペニーゲームを繰返し実施する複数の人工適応型エージェントによるシミュレーションによって，被験者を用いた実験では困難であった多くの繰返しやパラメータの細かい変動を伴う分析を試みる．人工エージェントは，自己や相手の過去の行動を入力とし，次の戦略選択を出力とするニューラルネットワークによる意思決定機構をもち，ニューラルネットワークの出力を左右するシナプス結合荷重と各ノードの出力に関するパラメータである閾値を遺伝子とした遺伝的アルゴリズムによって，より高い利得を獲得するエージェントを進化的に優位とする学習機構を有する．このようなマルチエージェントシステムにおいて，エージェントの長期的学習をシミュレーションによって明らかにし，シミュレーション結果を被験者実験の結果，ナッシュ均衡と比較する．

2. 一般化マッチングペニーゲームと被験者実験

混合戦略の唯一のNash均衡をもつ2×2ゲームである一般化マッチングペニーゲームに焦点をあてる．表1にはOchs (1995)によって実施された実験で用いられた一般化マッチングペニーゲームが示される．このゲームでは，行プレイヤーは上(U)か下(D)かを選択し，列プレイヤーは左(L)か右(R)かを選択する．このゲームの利得は表1に示されるように，行プレイヤーは，(U,L)または(D,R)で正の利得を得て，列プレイヤーは，(U,R)または(D,L)で正の利得を得る．このことから明らかに純戦略のNash均衡は存在せず，唯一の混合戦略のNash均衡をもつことが知られている．戦略の組(U,L)での行プレイヤーの利得aを$a \geq 1$とすれば，$a=1$のとき，行および列プレイヤーは対称であり，$a>1$ならば，非対称となる．

表 1: A generalized matching pennies game

		column player	
		L	R
row player	U	$(a,0)$	$(0,1)$
	D	$(0,1)$	$(1,0)$

行プレイヤーが戦略Uを選択する確率をp，列プレイヤーが戦略Lを選択する確率をqとすると行プレイヤーの期待利得π_Rと列プレイヤーの期待利得π_Cはそれぞれ

$$\pi_R = apq + (1-p)(1-q) \tag{1}$$

$$\pi_C = (1-p)q + p(1-q) \tag{2}$$

となり，Nash均衡は$(p^*,q^*) = (1/2, 1/(a+1))$となる．また，ゼロ和ゲームの場合と異なり，maximin戦略はNash均衡戦略と異なる．すなわち，行プレイヤーのmaximin戦略は$\arg\max_p \min_q \pi_R(p,q)$より得られ，列プレイヤーのmaximin戦略も同様にえられ，maximin戦略の対は$(p^S,q^S) = (1/(a+1), 1/2)$となる．

Ochs (1995)は戦略の組(U,L)での行プレイヤーの利得を変化させ，$a=1, a=4, a=9$の3種類のゲームを実験で用意した．被験者のペアが10回のゲームを続け，戦略の選択には2種類の純戦略か混合戦略を選択するかの3つの選択肢がある．混合戦略については戦略Uを10回のゲームで何回選択するかを答え，計算機システムによってその回数だけランダムにUが選択される．ゲームの結果によって被験者は，リスクに対する選好を制御するために，直接の金銭ではなく，くじ券が与えられた．実験の結果は，以下のように要約される．

(1) 戦略の組(U,L)における行プレイヤーの利得aが増加すれば，行プレイヤーが行動Uを選択する頻度が増加した．

(2) 行プレイヤーの利得aが増加すれば，列プレイヤーは行動Lを選択した頻度を減少させた．

上記の結果は，(1) において，a が変化しても Nash 均衡では行プレイヤーの戦略は $p^* = 1/2$ で不変であることに整合しないし，(2) において，a が変化しても maximin 戦略では列プレイヤーの戦略は $q^S = 1/2$ で不変であることに整合しない．

表 2 の第 2 列には，Ochs (1995) によって実施された実験の長期的定常状態相対頻度の推定値が示され，第 3 列には相対頻度が示され (McKelvey and Palfrey, 1995)，第 4 列には McKelvey *et al.* (2000) によって実施された別の実験結果が示される．また，第 5 列と第 6 列はそれぞれ Nash 均衡戦略と maximin 戦略が示される．

表 2: The results of the experiments

	Ochs (1995)	McKelvey and Palfrey (1995)	McKelvey *et al.* (2000)	Nash	maximin
$a=1$	(0.5015,0.4819)	—	—	(0.5,0.5)	(0.5,0.5)
$a=4$	(0.5336, —)	(0.542,0.336)	(0.550,0.328)	(0.5,0.2)	(0.2,0.5)
$a=9$	(0.6309,0.3497)	(0.595,0.258)	(0.643,0.241)	(0.5,0.1)	(0.1,0.5)

Ochs (1995) の実験データに対して，集約したレベルでは，Nash 均衡も maximin 戦略も観測された長期の傾向を説明できなかったが，McKelvey and Palfrey (1995) の量子応答モデル (quantal response model) や Roth and Erev (1995) および Erev and Roth (1998) の学習モデルが観測データをうまく説明できることが示されている．

3. シミュレーションモデル

ゲームの実験結果を説明するために，プレイヤーの合理性を基礎とした Nash 均衡モデルとは異なり，プレイヤーの意思決定における誤りを導入することによってプレイヤーの合理性を緩和するモデル (McKelvey and Palfrey, 1995) や，強化学習によるモデル (Roth and Erev, 1995; Erev and Roth, 1998) が開発されてきた．とくに，強化学習モデルにおいてプレイヤーの合理性の代案として導入された行動形式は適応的行動であった．本論文でも，プレイヤーの行動形式の基礎を適応行動とし，適応行動を実装するために自然な枠組みとしてシミュレーションモデルでの分析を試みる．シミュレーションモデルでは，プレイヤーは複数のエージェントに対応し，各人工適応型エージェントは，ニューラルネットワーク (たとえば，Hassoum (1995)) と遺伝的アルゴリズム (たとえば，Goldberg (1989)) を基礎とした意思決定および学習機構を有する．

3.1 ニューラルネットワークによる意思決定

エージェントは 1 つのニューラルネットワークに対応し，ニューラルネットワークは主として，各ノード間のシナプス結合荷重，各ノードの出力に関するパラメータである閾値によって特徴付けられる．ネットワークの構造を，3 層階層型ネットワークとし，各層のユニット数を固定すれば，エージェントは定めれらた個数のパラメータによって規定されるので，われわれのモデルではエージェントの染色体をこれらのパラメータが記録された

文字列で表現する．エージェント群はマッチングペニーゲームを繰返しプレイし，遺伝的アルゴリズムの枠組みで，得られた利得に関連する適合度にしたがって進化する．

ニューラルネットワークは複数の入力値からシナプス結合荷重及び閾値に基づき出力値を得る．エージェントは入力値として過去の自分の戦略だけでなく，過去の対戦相手の結果も取り入れ，以下の5つの入力値をもつ．(1) 過去の全ゲームで獲得した自己の総利得：x_i^{total}，(2) 直前のゲームで獲得した自己の利得：x_i^{last}，(3) 過去の全ゲームで獲得した対戦相手の総利得：y_i^{total}，(4) 直前のゲームで獲得した対戦相手の利得：y_i^{last}，(5) 総利得を次のゲームに持ち越す割合：ϕ_i．

ニューラルネットの出力は行プレイヤーの場合，戦略 U の選択確率であり，列プレイヤーの場合は，戦略 L の選択確率である．各エージェントに対応するニューラルネットワークは，入力層のユニットが5つで，出力層ユニットが1つであるので，隠れ層のユニット数を m とすると，シナプス結合荷重 $w_l, l = 1, \ldots, 6m$ および閾値 $\theta_l, l = 1, \ldots, m+1$ で表現できる．次に示すようにシミュレーションで高い利得を獲得するエージェントを遺伝的に進化させることで，これらのシナプス結合荷重と閾値を調整する．

3.2 エージェントによるゲームのプレイ方式

マッチングペニーゲームでは行プレイヤーと列プレイヤーでは学習の方法が異なると考えられるので，本シミュレーションでは行プレイヤーとなるエージェントと列プレイヤーとなるエージェントを同数用意し，各エージェント群から，1エージェントずつ選出することによってゲームのペアを決定する．各ペアはあらかじめ規定された回数だけゲームを行う．各エージェントはニューラルネットワークの出力値である戦略選択確率にしたがって2つの戦略を選択する．規定された回数だけゲームを行うことにより，混合戦略が適切に表現される．したがって，この戦略の決定はOchs (1995) の実験において被験者が混合戦略を選択した場合に対応する．

3.3 遺伝的アルゴリズムによる進化的学習

われわれのシミュレーションでは，1組のエージェントによってゲームが行われ，各エージェントは2種類の選択肢のうちどちらかを選ぶ．その結果，各エージェントは利得表に従った利得を獲得する．このゲームを繰返し行う過程で，エージェントは得られた利得に従う適合度によって進化していく．

シミュレーションの流れは次のように要約される．初期個体群として行プレイヤーと列プレイヤーのエージェントを別々に生成する．行プレイヤーの集団と列プレイヤーの集団のから1エージェントずつランダムに抜き出しペアを作る．各エージェントはニューラルネットワークからの出力により，戦略選択確率を定め，各ペアごとに，規定した回数のマッチングペニーゲームを繰返す．各エージェントは当該世代でプレイしたゲームで得られた利得の総計を計算する．個体群のすべてのエージェントに対して，遺伝的アルゴリズムにおける再生，交叉，突然変異の操作を行い，次世代の個体群を生成する．再生はルーレット選択とエリート保存選択を併用する．交叉は1点交叉で，突然変異が適用される遺

伝子，すなわちシナプス結合荷重と閾値に対して $[-1,1]$ の範囲にある実数をランダムに発生させ，その値に書換える．

3.4 誤差とリスク回避に関するパラメータ

McKelvey and Palfrey (1995) は誤差の概念を導入し，量子反応モデルを提案したように，現実のプレイヤーの意思決定には誤差が含まれると考えることは適切である．また，Goeree *et al.* (2003) はリスク回避の概念を導入したモデルが実験データをうまく説明することを示した．そこで，われわれのシミュレーションモデルにおけるエージェントの意思決定に誤差およびリスク回避のパラメータとして導入し，これらのパラメータを変化させることによって得られるシミュレーション結果を系統的に分析することを試みる．

4. シミュレーション結果

4.1 シミュレーションのトリートメント

シミュレーションでは，表1に示される利得表の (U,L) における行プレイヤーの利得 a を $a=1$ と $a=9$ にした2種類のマッチングペニーゲームを取上げて，対称ゲームと非対称ゲームの結果を分析したあと，非対称性のパラメータ a を細かく変動させて，a の変動によるエージェントの行動に対する影響を分析する．次に，リスクに対する態度に関するパラメータおよび誤差に関するパラメータの変動によるエージェントの行動に対する影響を分析する．さらに，初期世代のエージェントの戦略選択確率の分布や，ゲームでのプレイヤーの対戦方式が，最終的なエージェントの戦略選択確率の分布にどのような影響を与えるかについて調査する．

以下のトリートメントを用意する．(1) 対称ゲームの特徴，(2) 非対称ゲームの特徴，(3) 非対称性のパラメータ a の影響，(4) リスクに対する態度に関するパラメータの影響，(5) 誤差に関するパラメータの影響，(6) 初期世代のエージェントの戦略選択確率の分布の影響，(7) プレイヤーの対戦方式の影響．

4.2 シミュレーション結果

本節では紙面の制約から，(3) 非対称性のパラメータ a の影響に関するトリートメントの結果についてのみ報告する．

エージェントが初期世代でランダムにプレイするように設定し，非対称性のパラメータ a の変化に対する戦略選択確率と獲得利得の関係について分析する．シミュレーションでは，各エージェントのリスクに対する態度のパラメータはリスク中立となるように設定され，誤差に関するパラメータは誤差のない意思決定を行うように設定される．シミュレーションは $a=1,1.5,2,3,4,5,7,9,10,20,30$ に対して実施され，横軸に非対称性のパラメータ a をとり，縦軸に対して，戦略選択確率の変動を取った図が図1，図2に，縦軸に対して，獲得利得の変動を取った図が図3，図4に示される．プロットされる点はシミュレーションの700世代から1000世代の平均の獲得利得および選択確率である．また図中の太

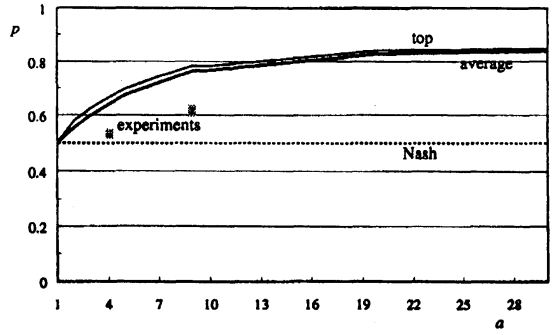

図 1: 行プレイヤー戦略選択確率 p

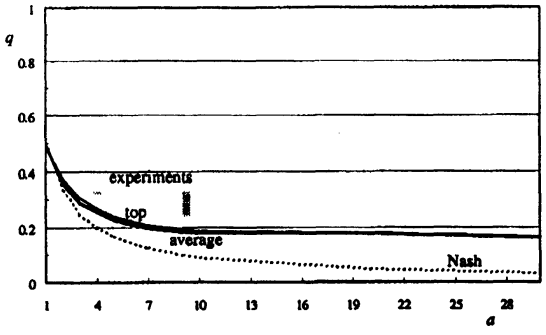

図 2: 列プレイヤー戦略選択確率 q

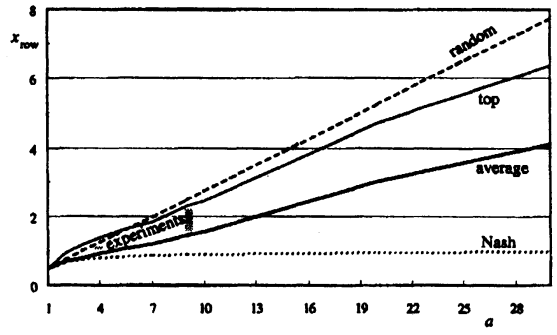

図 3: 行プレイヤー獲得利得 x_{row}

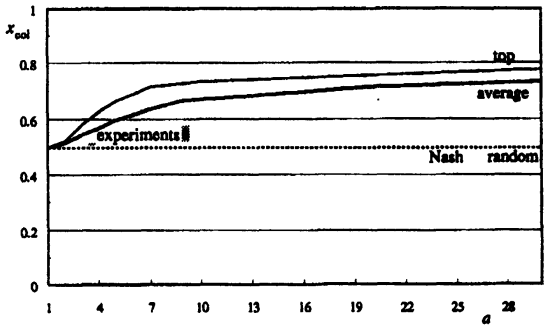

図 4: 列プレイヤー獲得利得 x_{col}

線は全エージェントの平均（average と表記）で，実線は行プレイヤーと列プレイヤーの正規化した効用の和の大きいペア 10 組の平均（top と表記）である．Nash 均衡や対応する利得は破線（Nash と表記）で示され，被験者による実験結果は最大値と最小値の灰色の区間（experiments と表記）で示される．図 5 には，a の変化に対する獲得利得対の変動を示している．

図 1 を見ると，a が増大するにつれて，シミュレーションでの行プレイヤーの戦略選択確率 p は増加するが，$a \leq 10$ の範囲では増分が大きく，$a = 20$ では $p = 0.83$ ぐらいでそれ以降ではあまり増加しない．Nash 均衡の行プレイヤーの戦略選択確率は $p^* = 0.5$ なので，エージェントの確率は明らかに異なった結果である．Nash 均衡の確率 p^* とエージェントの確率は，a が増大するにつれて，隔たっていくが，$a \geq 20$ では，エージェントの確率と Nash の確率の差分はほぼ一定になる．列プレイヤーの戦略選択確率 q は，図 2 を見ると，a が増大するにつれて，減少する．$a \leq 10$ の範囲では減少率が大きく，それ以降ではあまり減少しないが，少しずつではあるが減少している．図 2 を見ると，Nash 均衡の列プレイヤーの戦略選択確率は $q^* = 1/(a+1)$ で示され，エージェントの確率は q^* に追従した遷移を示しているが，エージェントの確率が Nash 均衡の確率よりも明確に大きな値をとっている．さらに，$a \geq 20$ では，エージェントの確率と Nash の確率の差分はほぼ一定となっている．被験者実験と比較すると，エージェントの確率は行プレイヤーの p の場

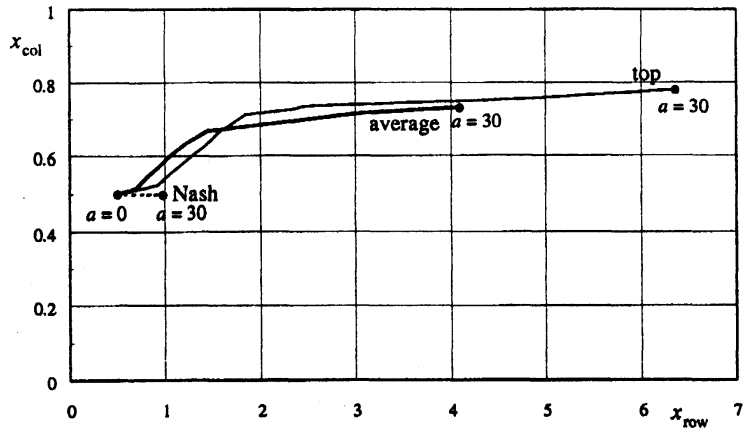

図 5: 行および列プレイヤー獲得利得 (x_{row}, x_{col})

合被験者実験に比べて大きく，列プレイヤーの q は被験者実験に比べてより小さく，より明確な傾向が示されている．

獲得利得に関しては，Nash 均衡の場合，行プレイヤーの利得は $a/(a+1)$ なので，1 に収束するが，行エージェントの利得は図 3 を見ると a が増加するにつれて，$1 \leq a \leq 30$ のシミュレーションの範囲ではほぼ線形に増加していることがわかる．Nash 均衡の場合の列プレイヤーの利得は a の大きさに関わらず 0.5 である．これに対して，列エージェントの利得は $a \leq 10$ の範囲では増分が大きく，$1 \leq a \leq 30$ の範囲ではそれ以降も少しずつ増加しつづけている．被験者実験と比較すると，行プレイヤーの場合の利得の top と average の間，すなわちエージェントの利得分布の中に被験者の利得がある．列プレイヤーの場合は，被験者の利得よりも明確に大きい利得を獲得している．

列プレイヤーの最大利得は a に依存せず常に 1 であるが，図 5 から，a が増大するにつれて，$a = 1$ の $x_{col} = 0.5$ から始まり，$a = 30$ では $x_{col} = 0.7$ となり，0.2 増加している．行プレイヤーの場合は，$a = 1$ の $x_{row} = 0.5$ から始まり，$a = 30$ では $x_{col} = 4.1$ となり，3.6 増加し，top の場合には，$a = 30$ では $x_{col} = 6.4$ となり，6.1 も増加している．この結果は明らかに，Nash 均衡の場合の利得と明白に異なる結果であり，より大きい利得を獲得したエージェントはより強くこの傾向を示している..

5. 結果の要約

シミュレーション結果は以下のように要約できる．長期的に収束した結果は，エージェントの獲得利得は Nash 均衡の利得よりも行プレイヤー列プレイヤーとも明らかに大きく，利得を最大化しようとする適応的なエージェントを仮定すると，Nash 均衡とは明らかに異なる行動をとり，この意味で被験者実験を支持する結果を得た．被験者実験と比べると，行エージェントの獲得利得は被験者よりも若干小さいが，より大きい利得を獲得した上位のエージェントはほぼ同じ利得を獲得し，列エージェントは被験者より明らかに大き

い利得を獲得した．このことから，被験者がさらに経験を積むと，行プレイヤーの利得が若干犠牲になるが，列プレイヤーの利得は増加すると考えられる．

非対称性のパラメータ a が大きくなると，行プレイヤーの利得はほぼ線形に増大するが，列プレイヤーの利得は $a \leq 10$ の範囲では増大するが，$a > 10$ ではほとんど増加しない．よって，非対称性は列プレイヤーに比べて行プレイヤーの利得をより増大させる．

被験者実験と比較するために，リスク態度に関するパラメータと誤差に関するパラメータを導入した．エージェントをリスク回避に設定すると，利得は被験者に比べパレート優位になるが，リスク受容や中立の場合に比べて，被験者の結果に明らかに近い．また，誤差を導入した場合のエージェントは，そうでない場合に比べて，被験者の結果に近い．よって，プレイヤーのリスク態度がリスク回避的であることや意思決定に誤差があることが，被験者の行動を説明する要因となりうると考えられる．

参考文献

Erev, I., and A. E. Roth, Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria, *The American Economic Review* 88, 848–881, 1998.

Goeree, J. K., and C. A. Holt, Ten little treasures of game theory and ten intuitive contradictions, *The American Economic Review* 91, 1402–1422, 2001.

Goeree, J. K., C. A. Holt, and T. R. Palfrey, Risk averse behavior in generalized matching pennies games, *Games and Economic Behavior* 45, 97–113, 2003.

Goldberg, D. E., Genetic Algorithms in Search, Optimization, and Machine Learning, Addison Wesley, Massachusetts, 1989.

Hassoum, M. H., Fundamentals of Artificial Neural Networks, The MIT Press, Cambridge, 1995.

McKelvey, R. D. and T. R. Palfrey, Quantal response equilibria for normal form games, *Games and Economic Behavior* 10, 6–38, 1995.

McKelvey, R. D., T. R. Palfrey, and R. A. Weber, The effects of payoff magnitude and heterogeneity on behavior in 2×2 games with unique mixed strategy equilibria, *Journal of Economic Behavior & Organization* 42, 523–548, 2000.

Ochs, J., Games with unique, mixed strategy equilibria: an experimental study, *Games and Economic Behavior* 10, 202–217, 1995.

Roth, A. E., and I. Erev, Learning in extensive form games: experimental data and simple dynamic models in the intermediate term, *Games and Economic Behavior* 8, 163–212, 1995.