繰り返し囚人のジレンマゲームに対する 2 種アプローチについての考察

宮島研究室 **S0321007** 漆原知喜

1　研究背景

応用数学の一分野である「ゲーム理論」の中で最も有名な題材の 1 つに,「囚人のジレンマゲーム」がある. これは, 2 人の容疑者がそれぞれ「自白しない」,「自白する」の一方を自身の行動として選択し, 2 人の行動の組み合わせに応じてそれぞれが利得を得るというゲームである. 一般に, 囚人のジレンマゲームでの各戦略に対する利得は $b > a > d > c \ \wedge \ 2a > b + c$ を満たす実数 a, b, c, d に対して表 1 のように整理され, これを**戦略形**と呼ぶ [1]. 表の第 1 列は容疑者 A の取る行動を, 第 1 行は容疑者 B の取る行動を示している. また, 表の要素 (X, Y) は容疑者 A, B がそれぞれ X, Y の利得を得ることを表す.

表 1　一般の囚人のジレンマ 戦略形

容疑者 A / 容疑者 B	自白しない	自白する
自白しない（協力）	(a, a)	(c, b)
自白する（裏切り）	(b, c)	(d, d)

容疑者 A, B が利得（の期待値）を最大にするように行動すると仮定すると, 容疑者 A, B のどちらにとっても（相手の行動に関わらず）「自白する」することが有効である. すなわち, 囚人のジレンマゲームでは「自白する」が**支配戦略**となっている. しかし,「2 人の利得の合計」を最大化したい場合, 最も合理的な戦略は 2 人が共に「自白しない」を選択することである. このように, **各プレイヤーが自分の利得を最大にするように行動しても, 全体としての利得は最大化されない**. これが「**ジレンマ**」と呼ばれる所以である.

なお, 行動の選択肢を表す言葉として「自白しない」,「自白する」を一般化した「協力」,「裏切り」を採用する場合もある. この場合, 囚人のジレンマゲームは, **互いが私的利益を追求することで, 協力した時と比べてそれぞれの利得が下がってしまう**ということを主張する. 本論文では, 行動の選択肢として「協力」,「裏切り」を採用することにする.

2　本研究の目的及び有用性, 新規性

ゲーム理論では, 同一のゲームが繰り返しプレイされる状況を一般に**繰り返しゲーム**という. **本研究の目的は, 繰り返し囚人のジレンマゲームをモデルとした新たな数理モデルを 2 つ提案し, それらのモデルに基づいてシミュレーションを行い, 結果を考察することである**. 特に, プレイヤー数を $N(\gg 2)$ 人とし, 各プレイヤーを 1 セルとみなした二次元格子空間上での繰り返し囚人のジレンマゲームをシミュレーションする. このように多数プレイヤーによる繰り返し囚人のジレンマゲームを二次元多体系モデルによってシミュレーションし考察することは, 現実社会における意思決定個体間の協力現象について理解を深めたり, また予測を立てることに有用である.

書籍 [2] では各プレイヤーの行動選択を行う代表的な手法として, **進化論的アプローチ**と**強化学習アプローチ**の 2 種類が紹介されている. しかし, これらのアプローチの紹介だけにとどまっており, 具体的な数理モデルやシミュレーション結果については言及されていない. そこで本研究では, 2 種類のアプローチそれぞれについて, 筆者が提案したオリジナルの数理モデルを用いてシミュレーションを行う.

以降で説明する 2 種のアプローチについて, 各種数理モデルの詳細及びシミュレーション結果のアニメーション等については, 発表スライドにて示す.

3　進化論的アプローチのシミュレーション結果とその考察

進化論的アプローチは「各行動がもたらす利得が自身及び近傍個体計 9 個体の獲得利得合計に占める割合」を計算し, 自身の行動をより大きな利得が期待される行動に高確率で更新するというものである.

シミュレーションの結果, **時間発展とともに「裏切り」を選択する個体が増加する**ことと, **利得を表すパラメータの値によって, 行動を「裏切り」に更新する個体の増加の勢いに差が見られる**ことがわかった.

考察では, シミュレーション結果を正当化するための数学的解析を行った. 具体的には,

- **全ての個体について, 時刻 $t = n$（$\in \mathbb{Z}_{\geq 0}$）で行動を「協力」に更新する確率は共通である**
- **獲得利得の期待値を真の獲得利得であるとみなす**

の 2 点を仮定したとき, ある 1 個体が時刻 $t = n$（$\in \mathbb{Z}_{\geq 0}$）で行動を「協力」に更新する確率 p_n が漸化式

$$p_0 = \frac{1}{2}, \quad p_{n+1} = \frac{a(p_n)^2 + cp_n(1-p_n)}{a(p_n)^2 + (b+c)p_n(1-p_n) + d(1-p_n)^2} \quad (n \in \mathbb{Z}_{\geq 0})$$

で定まる数列 $\{p_n\}$ に従うことを示した. また, この数列 $\{p_n\}$ の極限値が 0 であること, すなわち

$$\lim_{n\to\infty} p_n = 0$$

を証明し, 上の 2 点の仮定の下で, **行動「協力」を選択する個体数の時間極限が 0 となる**ことを示した.

4　強化学習アプローチのシミュレーション結果とその考察

強化学習アプローチではシミュレーションの前に, ある 1 個体をエージェントとして強化学習を行う. この強化学習モデルは, **評判**の概念を用いることで, エージェントが「行動『協力』を選択することがより大きな収益に繋がる」ということを学習するように構築されている. そして, そのエージェントが強化学習により身に付けた行動指針（最適方策）に全 N 個体が従うとしてシミュレーションを行った.

シミュレーションの結果, **時間発展を通しての協力個体率（各時刻 t において行動「協力」を選択した個体数が全個体数に占める割合）は 50% 超でほぼ一定に保たれる**ということがわかった. また, **学習回数の増加に伴って, 協力個体率が上がる**こともわかった.

考察では,「シミュレーション結果から, **評判**の概念を組み込んで今回構築した強化学習モデルが妥当なモデルであった」ことや「本アプローチで用いた強化学習モデルの場合, 学習回数が 10 回程度を超えてくれば強化学習を用いる意義が生まれる」ことを結論づけた.

5　今後の課題

進化論的アプローチに関する今後の課題は, 考察にて仮定した命題そのものの証明, もしくは, その仮定を多少弱めた場合についても今回と同様の結論を導くことである.

強化学習アプローチに関する今後の課題は, 複数のエージェントに対して同時に強化学習を行う手法である「マルチエージェント強化学習」を本論文の数理モデルに適用し, 新たな発見を探ることである.

参考文献

[1] 神戸伸輔, 入門 ゲーム理論と情報の経済学, 日本評論社, (2004)

[2] 江崎貴裕, データ分析のための数理モデル入門 本質をとらえた分析のために, ソシム株式会社, (2020)