

報酬駆動型システムにおける報酬の設計と 報酬による最適化

奥原 浩之

武藤 克弥 (Katsuya Mutoh)
u255018@st.pu-toyama.ac.jp

富山県立大学 大学院 電子・情報工学専攻 情報基盤工学部門

January 21, 2023

背景

- 集団での効率的な効用を得るため、各々が意思をもった個体が戦略選択するシステムでは、個体での最適な選択が、集団にとっての最適にならない場合がある。
- 個体の最適な選択が、集団の最適選択となるようなルールを設計する必要がある

目的

- ペイオフによるシステム制御を非協力ゲームだけでなく、遺伝子や神経回路などのモデルにも適用できるようにする。
- ペイオフによる安定的なシステム設計と最適なペイオフ探索を行うことで、最適戦略の導出を行う

非協力ゲーム

- 集団の中で個人 1 人が自分の利得のために行動選択する
- 各個人が行った行動 (選択の組み合わせ) によって得点が与えられる

A	B	
	非協力	協力
非協力	(3, 3)	(10, 0)
協力	(0, 10)	(7, 7)

収束(Nash均衡)

図 1: 非協力ゲームにおけるペイオフ表

遺伝子・神経回路モデルにおける均衡

- 生態系の食物連鎖：捕食側・非捕食側
→捕食増加、獲物減少、捕食者減少、獲物増加を繰り返して一定の数に収束
- 結合振動子：バラバラに動いていたものが、やがて全てシンクロした動きをする (同期=収束)
- 神経回路網：各シナプスに情報伝達
→やがて全シナプスに同じ信号がいきわたる (全て同じ値に制御 (収束))

→別々に表現されるシステムのモデル式を全て1つで表す
→これらすべてのモデルをペイオフで制御して収束させる

報酬駆動型システム

$$\dot{\mathbf{x}}_i = \sum_{j=1}^N G_{ij}(\mathbf{x}) \left\{ \mathbf{f}(\mathbf{x}_j) + \sum_{k=i}^N c_{jk} a_{jk} (\mathbf{\Gamma}_{jk} \mathbf{x}_k - \mathbf{\Gamma}_{jj} \mathbf{x}_j) \right\} \quad (1)$$

i, j, k : エージェント $(1, 2, \dots, N)$ \mathbf{x}_i : エージェントの状態
 $\dot{\mathbf{x}}_i$: 状態微分

$G_{ij}(\mathbf{x})$: 状態集合 \mathbf{x} を与えた際のペイオフ双行列

c_{ij} : エージェント間の結合強度 a_{jk} : エージェント間の隣接行列

$\mathbf{\Gamma}_{jj}$: エージェント間の相互作用

得られる解

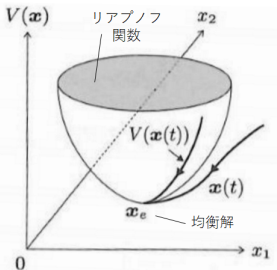
- $G_{ij}(\mathbf{x})$ からペイオフが得られる
- ペイオフで均衡となるとき→エージェント i の均衡解 \mathbf{x}_i^*
- この均衡時にシステムが安定しているかを調べる

漸近安定: 過渡時間で各エージェントの状態が収束していること
→ 均衡状態 $\mathbf{x}_i^*(1, 2, \dots, N)$ で収束しているか確認

漸近安定かどうかの確認

以下二つを満たすときシステムは漸近安定 (リアプノフの安定性理論)

- 均衡解 \mathbf{x}^* に対して、ある関数 $V(\mathbf{x})$ (=リアプノフ関数) が存在
→ $V(\mathbf{x})$ は $V(\mathbf{x}^*) = 0$ かつ $\mathbf{x} \neq \mathbf{x}^*$ の性質がある
- $\mathbf{x} \neq \mathbf{x}^*$ において $\dot{V}(\mathbf{x}) < 0$



ペイオフ変更によるシステム制御

均衡時にシステムが不安定→ペイオフ行列 A を変更する

- (1) 式から \bar{K} と \bar{P}^{-1} を導出する
- $A + \bar{K}\bar{P}^{-1}$ を新たなペイオフとしてシステムにフィードバックする

A			$A + \bar{K}\bar{P}^{-1}$	
(3, 3)	(10, 0)	→	?	?
(0, 10)	(7, 7)		?	?

図 3: ペイオフ変更のイメージ

最適戦略の探索

均衡状態を漸近安定させることで探索所の安定性を保障？

→次は全体の利得が最大となるように探索

- 遺伝的アルゴリズムで最適化
戦略変更時の適応度の変化を参照+コスト関数の最小化

結論

- 非協力ゲーム・遺伝子・神経回路モデルを全て1つの式で表現
- ペイオフを変更することで均衡時に漸近安定化させる方法を提案
- ペイオフ(全体の利得)が最大となるような戦略探索法を提案

修論の研究テーマ (案)

9/12

メカニズムデザイン

- マルチエージェントの一種 (自身の利益を得るため意思決定する個体 (エージェント) が寄り集まった際、集団・社会がどのようなふるまいをするのかを研究)
- (社会などが) 所望する結果を得るために、個人のふるまいや集団のルールを設計する
- 経済活動、計算機ネットワーク、人間行動などに応用

はじめに

概要

提案手法

今後の予定

まとめ

修論の研究テーマ (案)

10/12

ソーシャルネットワークのメカニズムデザイン

- SNS で影響力のあるアカウントのふるまいが、他のユーザにどのような影響を与えるのかの分析

研究例

- Facebook のアカウントを 4 つのクラスにクラスタリングし、メッセージを送信・受信する際のアカウント間のふるまい、アイデア (文章) の拡散具合を分析¹
- フェイクニュースの影響力測定²
→ Twitter のニュース文章 (タイトル・本文)、リツイート情報からなる、真・偽ニュースラベルのついたデータセットを用い、深層学習を実行
- メタ報酬ゲームを用いたソーシャルメディア流行メカニズムの分析³
→ コメント投稿に対する反応を報酬ゲームと捉えて、SNS の流行理由を分析

¹ Amer G. Ghanem et al., 2012.

² Cheng-Lin Wu et al., 2022.

³ Yuki Hirahara et al., 2014.

修論の研究テーマ (案)

11/12

考えられる方向性

- Twitter のテキストマイニングなどを組み合わせて、ユーザ間の相互作用をモデリング
- 1 式に適用できるか調べる
- ペイオフ制御で安定化の確認
- ある目的を設定して、その通りにユーザ (エージェント) がふるまってくれるかシミュレーション

はじめに

概要

提案手法

今後の予定

まとめ

まとめ

- ・メカニズムデザインとしても目的設定ができていない
- ・ソーシャルネットワークのモデルがそもそも 1 式に適用できるかわからない

今後の予定

- 研究テーマの深掘り