報酬駆動型システム：
自律分散メカニズム・デザイン実現のための研究

研究ノート

奥　原　浩　之*

Reward Derived System:
Approach for Autonomous Decentralized Mechanism Design

Key Words：Reward Structures, Autonomous Decentralized, Design, Optimization

1　はじめに

20世紀後半頃からの地球の温暖化が意識され，温室効果ガスの排出抑制が国際的に議論されて久しい．地球温暖化が気候に影響するのなら，生態系などの自然環境が変化し，農業や漁業を通じて社会環境へも影響が及ぶ可能性が考えられている．実効性がある温暖化抑制策あるいは緩和策の実施の遅れは，さらなる負担やコストの増加を招くと指摘されている．にもかかわらず，利害の関係がまちまちであるため制度の設計は困難しているように見受けられる．

これに類似する問題解決の重要性は，なにも世界規模の事例を取り上げるまでもなく，国内，地域社会といった規模を問わず，我々の身近な問題で普段感じられるものであり，制度の設計の困難さを示していると言える．

常に個体の最適な応答の結果として，集団にとって効率的な選択が実現されるとは考えられない．そのような場合には，環境を人為的に設計することで，個体の最適な応答により，集団として望ましい効率的な選択を自律分散的に実現する仕組みの解明や設計法が求められる．

本研究ノートでは，特定の戦略の選択を促進・抑制する自律分散メカニズム・デザインを実現するための研究 [1] の一部を簡単に紹介する．文献 [1] では遺伝的アルゴリズム，神経回路網モデル，進化ゲームに関する力学は，意思決定の主体が戦略を選び行動した結果，報酬を原動力として駆動している報酬系と見なすことができ，ペイオフにもとづく効用関数の変化を報酬として入力している報酬駆動型システムと捉えることができることが示されているが，ここでは，遺伝的アルゴリズムについて述べる．

* Koji OKUHARA
1968年9月生
広島大学　大学院工学研究科　システム工学専攻　博士後期課程（1996年）
現在、大阪大学　大学院情報科学研究科　准教授　博士（工学）
ソフトコンピューティング，数理モデリングと最適化・制御
TEL：06-6879-7877
FAX：06-6879-7877
E-mail：okuhara@ist.osaka-u.ac.jp

2　遺伝的アルゴリズム

淘汰と突然変異，さらに交叉と確率を考慮した遺伝子のモデルである簡易な遺伝的アルゴリズム (Simple Genetic Algorithm:SGA) [2] は次のようなものである．個体の最適な応答により，集団として望ましい効率的な選択を自律分散的に実現する仕組みとなっている．

[1] l ビットの0または1で遺伝子を表現する．N 個の個体 $\{\mathrm{x}_1, \mathrm{x}_2\, \mathrm{x}_3, \cdots, \mathrm{x}_N\}$ を作成．$N \ll 2^l - 1 = n$ とする．
[2] 適合度関数 $\mathrm{f}(\cdot)$，突然変異率，交叉率の値と逆温度 β の初期値を与える．
[3] 突然変異する部分と交叉する部分を与える．
[4] 選択，突然変異，交叉を繰り返し世代交代する．

ここで，$\mathrm{f(x)}$ は遺伝子 x に対する適応度関数であり，$\beta > 0$ は Boltzmann の逆温度である．

淘汰は，現世代において存在する遺伝子群 $\mathbf{p} = \{\mathrm{x}_1, \mathrm{x}_2\, \mathrm{x}_3, \cdots, \mathrm{x}_N\}$ のうち，遺伝子が選択される確率で表現される．突然変異は，ビットが反転する確率で表現される．

交叉は，遺伝子 x_i と x_j が交叉率で交叉することで，遺伝子 x' となる確率で表現される．

3　報酬系としての遺伝的アルゴリズム

遺伝子 x が淘汰，突然変異，交叉を経て遺伝子

x'となる遷移確率$M(\mathrm{x}'|\mathrm{x})$が

$$M(\mathrm{x}'|\mathrm{x}) = p(\mathrm{x}'|\mathrm{x})\exp(\phi_{\mathrm{x}';\mathrm{x}})$$

によりモデル化され，遺伝的アルゴリズムが累積無限期間平均コスト関数

$$R^{\pi}(\mathrm{x}) = \lim_{t_f\to\infty} \mathrm{E}\left[\sum_{\tau=0}^{t_f-1} l(\mathrm{x}_{(\tau)}, \pi(\mathrm{x}_{(\tau)}))\right]$$

の最小化$R(\mathrm{x}) = \min_{\pi} R^{\pi}(\mathrm{x})$を評価していると考える．ここで，$\phi_{\mathrm{x}';\mathrm{x}}$は遺伝子$\mathrm{x}$から遺伝子$\mathrm{x}'$へ切替えるために必要な制御入力であり，$p(\mathrm{x}'|\mathrm{x})$は受動的遷移であり，$p(\mathrm{x}'|\mathrm{x}) = 0$なら$M(\mathrm{x}'|\mathrm{x}) = 0$である．また，$\sum_{\mathrm{x}'} M(\mathrm{x}'|\mathrm{x}) = \sum_{\mathrm{x}'} p(\mathrm{x}'|\mathrm{x}) = 1$，$\pi(\mathrm{x}_{(\tau)}) = \phi_{\mathrm{x}'_{(\tau)};\mathrm{x}_{(\tau)}}$である．

コスト関数$l(\mathrm{x};\mathrm{u})$は報酬$r(\mathrm{x}';\mathrm{x})$と制御入力$\phi_{\mathrm{x}';\mathrm{x}}$を用いて

$$l(\mathrm{x},\phi) = \mathrm{E}_{\mathrm{x}'\sim M(\cdot|\mathrm{x})}[r(\mathrm{x}';\mathrm{x}) + \phi_{\mathrm{x}';\mathrm{x}}]$$

で定義されるものとする．

このとき，Bellman方程式は

$$R(\mathrm{x}) = \min_{\phi\in\phi_{\mathrm{x}';\mathrm{x}}}\{l(\mathrm{x},\mathrm{u}) + \mathrm{E}_{\mathrm{x}'\sim M(\cdot|\mathrm{x})}[R(\mathrm{x}')]\} - c$$

となる．ここで，cは平均コストである．

最適制御入力は，$\psi(\mathrm{x}';\mathrm{x}) = -r(\mathrm{x}';\mathrm{x}) - R(\mathrm{x}')$として

$$\phi^{*}_{\mathrm{x}';\mathrm{x}} = \psi(\mathrm{x}';\mathrm{x}) - \log \mathrm{E}_{\mathrm{x}'\sim p(\cdot|\mathrm{x})}[\exp(\psi(\mathrm{x}';\mathrm{x}))]$$

となる[3].

このことは，簡易な遺伝的アルゴリズムでは，報酬$r(\mathrm{x}';\mathrm{x})$を遺伝子$\mathrm{x}$から遺伝子$\mathrm{x}'$へ戦略を切替えたときの適応度の変化

$$r(\mathrm{x}';\mathrm{x}) = \log p(\mathrm{x}'|\mathrm{x})\left(\frac{f(\mathrm{x})}{f(\mathrm{x}')}\right)^{\beta}$$

で与えて駆動することで，累積無限期間平均コスト関数

$$R^{\pi}(\mathrm{x}) = \lim_{t_f\to\infty} \mathrm{E}\left[\sum_{\tau=0}^{t_f-1} \mathrm{E}_{\mathrm{x}'\sim M(\cdot|\mathrm{x})}\left[\log\left(\frac{f(\mathrm{x}_{(\tau)})}{f(\mathrm{x}'_{(\tau)})}\right)\right]\right]$$

の最小化を行っていることを示している．

報酬駆動型システムとしての遺伝的アルゴリズムは，戦略を選び行動した結果，逐次，ペイオフである適応度が判明しプレイヤー間で共有され，報酬にもとづき入力が得られる報酬系であることがわかる．つまり，個体の最適な応答により，集団として望ましい効率的な選択を自律分散的に実現する仕組みを，報酬系による最適化を行うことで実現していると言える．

4　おわりに

本研究ノートでは，意思決定の主体が戦略を選び行動した結果，ペイオフにもとづく効用関数の変化を報酬と考え，それを入力として動作する報酬駆動型システムによる報酬系を考えた．その枠組みの中で，遺伝的アルゴリズムが報酬系による最適化を行っており，自律分散メカニズム・デザインを実現していることを明らかにした．

参考文献

[1] 奥原浩之，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，システム / 制御 / 情報，Vol. 58, No. 11, pp. 462-467 (2014)

[2] M. D. Vose, *The Simple Genetic Algorithm: Foundations and Theory*, MIT Press (1999)

[3] E. Todorov, Efficient Computation of Optimal Actions; *PNAS*, Vol. 106, No. 28, pp.11478-11483 (2009)