

# 有限資源状況における利己的エージェントの調和的獲得競争

Optimum distribution of limited resources by harmonic competition between selfish agents

菅野 勇紀<sup>\*1\*2</sup>

Yuuki Kanno

斉藤 啓介<sup>\*3</sup>

Keisuke Saitou

三上 貞芳<sup>\*4</sup>

Sadayoshi Mikami

長野 章<sup>\*4</sup>

Akira Nagano

<sup>\*1</sup>公立はこだて未来大学大学院システム情報科学研究科

Future University Hakodate, Graduate School of System Information Science

<sup>\*2</sup>日本データサービス株式会社

Nippon Data Service Co.,Ltd.

<sup>\*3</sup>公立はこだて未来大学

Future University Hakodate

<sup>\*4</sup>公立はこだて未来大学情報アーキテクチャ学科

Future University Hakodate, Department of Media Architecture

The problem of the optimizing distribution about limited resources strongly attracts attention in the social situation. We will find the way to the sustainable and inhibitory use of resources with solving this problem. One answer about this problem is the use of the multi selfish agent system. In this paper, we propose an architecture helps for a agent to exchange their experiences between other agents. Agent which received information from other agents decides or changes its own plan to get best benefit. Consumption of resources will be balanced through these motions. We also show results of experiments with this protocol in virtual market.

## 1. はじめに

資源の有効分配に関する問題は昨今の社会情勢において極めて注目される課題のひとつといえるが、こうした問題は、特に開放的な市場モデルにおいてはマルチエージェントシステムによる処理に好適な課題であると考えられる。開放的な市場モデルでの経済主体・取引主体は、自律的かつ知的な行動主体であり、各自が特定の利益を代表するからである。マルチエージェントシステムを利用した資源分配機構の実現手法は二通り考えられる。ひとつはエージェント間の競合調整を一種の分散制約充足問題とする考え方である。この場合、エージェントは探索器に過ぎず、競合調整機能は上位の制御者に期待することになる。もうひとつは、エージェント間の競合調整をエージェント=プレイヤーに委ねる考え方である。この場合、エージェントは効用最大化原理に基づき、各自の合理性の範囲で利己的に行動し、市場は非協力  $n$  人ゲームを構成することになる。すなわち、競合調整は均衡の解可能域でなされる。しかし、開放的な市場モデルにおける各主体は何らかの利益を代表して財等のやり取りを行い、効用の最大化を目指すものであり、各主体は利己的な存在であるから、調整を受けて自己の選択を行うという概念にそぐわない。また、資源があらかじめ明らかではない場合もあり得ること、制御者に期待される能力は膨大であること、そもそも上位者による制御が可能であるならば競合調整機構は不要であるとも考えられること等から、少なくとも開放的な市場を前提として考えた場合、利己のエージェントによる競合調整機構が有効であろう。

一方、こうした利己的なエージェントによる資源分配は前述の通り非協力  $n$  人ゲームを構成するが、ここでは資源を有限と仮定するため、社会的ジレンマに陥ることになる。こうした社会的ジレンマは、特に、 $n$  人型チキンゲームである「共有地の悲劇 (The Tragedy of Commons)」によってモデル化される。すなわち、各個体が自己の効用を最大化しようと目的合理的に行動することで、個体の集合である群の効用が損なわれ、結局は各個体の将来的な効用の低下、とりわけ資源の枯渇による破局を招来することになる。こうした社会的ジレンマ状況を

克服するために、各主体は何らかの操作・制約によりその合理性を制限されるべきことになる。つまり、得られるはずの利益を諦めるという行動を選択することを迫られる場面を設定することになる。こうして利己のエージェントは、利己的本質を維持したままに調和的に振舞うことが求められることになる。開放性を前提としたマルチエージェントシステムによる競合調整機構の中で、このような取り扱いはまったく例外的なものである。そこで、こうした操作・制約をどこに、どのような形で設定すべきかが問題となる。

## 2. 制約の枠組み

### 2.1 既往の研究

マルチエージェント環境における社会的ジレンマの克服を目的として、周辺環境状況に応じて内部評価の生成方法を変化させる学習エージェントについていくつかの提案が行われている [Mikami 94][森山 02] が、報酬情報を開示することを前提としている点等に難がある。また、有限資源に回復能力を設定し、各エージェントの要求の総和が回復能力の範囲を超えた場合に資源の収奪を禁止するというモデル構築に関する提案もある [Akiyama 02][谷本 06] が、資源の総量を正確に知ることは一般に困難であり、もし知ることができたとすれば上位者による適正分配が可能となるから、エージェントによる獲得競争にそぐわない。資源の総量や資源に与える消費者の行動選択の影響を正確に知ることは困難であるとして、不完備情報ゲームを用いてエージェントの意思決定モデルを構築するという研究もある [佐藤 07]。

### 2.2 提案

本稿では、分散した有限資源の獲得機構におけるエージェントが個別には探索者であることに着目し、探索に有用な情報を他エージェントと戦略的に取引する能力を有するエージェントで構成される資源探索空間を提案する。ここで、他エージェントから情報を得るためには資源の利己的な獲得を抑制する必要があるという規制を設定する。これにより、市場に参加しようとするプレイヤーに対してエージェントへの調和的振舞いのインプットを促し、資源獲得のための戦略選択の自由の確保と資源の護持・持続的利用の両立を図るものである。

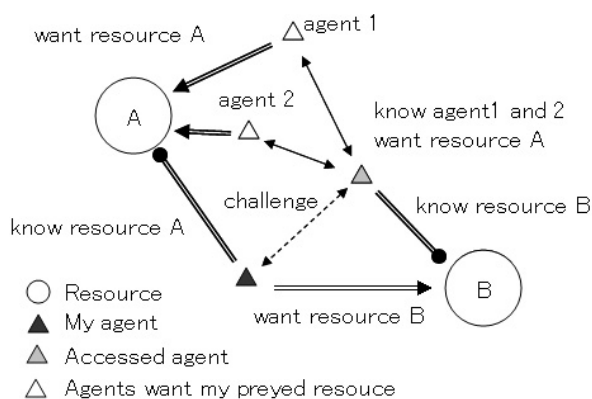


図 1: 提案する資源探索空間のモデル図

### 2.3 資源探索空間の構成

資源探索空間には資源  $E = \{e_1, \dots, e_m\}$  が配置される。エージェント  $A = \{a_1, \dots, a_n\}$  はプレイヤーによって投入される。資源は有限であり、その 1 個の初期状態量は全エージェントの要求の総和より少ない。ただし、資源は複数存在し、かつ、相互に代替性がある場合がある。

エージェントは散漫な hill climbing 探索を行う。このとき、エージェントには獲得すべき資源の属性と量がタスクとして与えられている。獲得すべき資源の属性とは、その資源の主要な性質のことであり、共通する属性を有する資源の間には「代替性がある」と表現する。すなわち、エージェントはタスクとして与えられた資源属性と代替性がある資源を探索すればよいのであって、特定の資源を探索するのではない。このことは、現実世界の石油等のエネルギー財が一般に複数の産出箇所から採集されること、それを備蓄・販売する事業者が複数存在することに対応している。以後、エージェント  $A_i$  にタスクとして与えられた属性を有する資源と代替性がある資源のことを「エージェント  $A_i$  の目的資源」と呼び、

$$E_i^{aimed} (\subseteq E)$$

で表す。

エージェントは有限の燃料を持ち、行動ごとに定率で消費するため、燃料の範囲内でのみ空間内で探索行動を行うことができる。エージェントが空間内で資源に隣接した場合、「資源に到達した」という。到達した資源が目的資源であった場合には燃料が定率で補充される。燃料が尽きた場合には環境から退場する。

エージェントの行動履歴はエージェント内のメモリに蓄積される。メモリの構成を表 1 に示す。

表 1: エージェントが保持するメモリの構成

名称	内容
所有者 (プレイヤー)	{ 識別子 }
目的資源情報	{ 累数, 識別子 }
到達資源情報	{ 累数, 識別子, 位置, 属性 }
獲得資源情報	{ 累数, 識別子 }
接触エージェント情報	{ 累数, 各エージェント情報 { 識別子, 開示情報 {...} }

エージェントは資源に到達した場合、その資源に関する情報をメモリに蓄積する。これは到達した資源が目的資源ではな

い場合も同様とする。すなわち、エージェントはこれまでに到達した資源の情報はすべて保有する。以後、あるエージェントが過去に到達したことがある資源を「そのエージェントの到達資源」と呼び、また、特に到達資源が目的資源であった場合には、その資源を「そのエージェントの獲得資源」と呼び、それぞれ以下のように表す。なお、 $t$  は時刻である。

$$E_{i,t}^{accessed} (\subseteq E) \text{ (到達資源)}$$

$$E_{i,t}^{acquired} (\subseteq E, \subseteq E_i^{aimed}, \subseteq E_{i,t}^{accessed}) \text{ (獲得資源)}$$

エージェントが探索過程で他のエージェントと隣接した場合、「他エージェントと接触した」という。エージェント  $A_i$  がこれまで接触したことがあるエージェントの集合を以下のように表す。

$$N_{i,t} (\subseteq A, \subseteq (A \setminus A_i))$$

他エージェントと接触したエージェントは、相手のメモリ内容の一部を常に参照することができる。ただし、メモリの全部の参照には相手の許諾が必要である。常に参照できる情報は「目的資源の識別子」、「獲得資源の識別子」の 2 つである。これらを「開示情報」と呼ぶ。

開示情報には到達資源の位置情報などは含まれないため、これらを参照するためには相手の全メモリの参照を要求しなくてはならない。その場合、先に自己の全メモリを提供する必要がある。参照を要求されたエージェントは、相手の開示情報を以下の基準に従って評価し、許諾するかを決定する。なお、許諾があった場合を「交流が成立した」と呼ぶ。

まず、自己の目的資源の性質と相手の目的資源の性質がまったく一致しない場合には許諾しない。相互に社会的関係を有しないので、利己的振る舞いの原則に立ち返るためである。

次に、式 (1) により相手エージェント  $A_j$  の調和的態度に関わる指標  $F(A_i, A_j)$  を評価する。

$$F(A_i, A_j) = \begin{cases} 0 & \text{if } O(A_i, A_j) + P(A_i, A_j) > |E| \times v_i \\ 1 & \text{otherwise.} \end{cases} \quad (1)$$

ここで、

$$O(A_i, A_j) = |E_{j,t}^{acquired} \cap \sum_{A_k \in N_{i,t}} E_k^{aimed}| \quad (2)$$

$$P(A_i, A_j) = |(E_j^{aimed} \cap E_{i,t}^{accessed}) \setminus E_{j,t}^{acquired}| \quad (3)$$

(2) 式は相手の獲得済み目的資源が自己の接触済み他エージェントの目的資源としてどの程度の頻度で現れたか、(3) 式は相手の獲得済み目的資源と代替性がある資源が他に存在したかを表している。 $v$  は相手エージェントをどれだけ厳しさを評価するかを表す係数で、 $0 \leq v \leq 1$  である。

$P$  が 1 でない場合には交流は成立しない。すなわち、その場合、相手は多数のエージェントが目的資源として考えている資源から採集を行い、かつ、他に選択しうる資源があったことを示している。このプロセスは、いわばエージェントに以下のような価値観を擬制することになる。すなわち、他にも要求が満たせる方法があったにもかかわらず、結果的に枯渇が予想されるような資源から採集を行った相手には、それが善意であったか悪意であったかに関わらず、便宜を図ることを拒否するような判断である。善意である場合を包含したのは、本稿で提案する環境ではすべてのエージェントは初期において善意であり、したがって初期段階では一部の資源への集中が発生することが予想されるためである。

交渉が成立した場合には、相手のメモリ内容をすべて参照することができる。そして、自己の未獲得の目的資源と相手の到達資源に代替性がある場合、その資源に最小の燃料で到達することができる。つまり、他エージェントと交流することで目的資源に効率的に到達することが期待できることになる。また、交渉が成立した場合には、相手の到達資源は自己の到達資源として記録される。

交流が成立しなかった場合には、エージェントは自己の能力の範囲で通常の探索を継続することになる。その場合にも罰則などはない。

エージェントは他エージェントと交流することを強制されない。すなわち、エージェントには、他エージェントと積極的に交流する戦略 (Cooperative Strategy: 友愛戦略)、他エージェントと交流するかどうかをその都度決定する戦略 (Opportunity Strategy: 日和見戦略)、他エージェントと一切交流しない戦略 (Prideful Strategy: 独善戦略) および 3 つの戦略を強化学習により選択的に実践する戦略 (Learning Strategy: 学習戦略) が用意される。

友愛戦略を採用するエージェントは、到達したすべてのエージェントに対して交流を試みる。日和見戦略を採用するエージェントは、相手エージェントの開示情報に含まれる「接触エージェント数」が多い場合には交流しない。接触エージェントが多い場合には、自己の獲得資源を目的とするエージェントに出会っている確率が高く、自己の全メモリを提供しても無駄に終わる可能性が高いからである。なお、交流が成立しなかった場合にも、相手エージェントは自己の全メモリを参照することができることに注意が必要である。学習戦略を採用するエージェントは Q-learning により交流するか否かを選択する。エージェントは環境の状態  $s$  において行動  $w$  を Greedy 選択する。これらの組み合わせに伴う価値  $Q(s, w)$  を学習サイクルごとに次式に従って更新する。ここで  $t$  は離散時間、 $rt + 1$  は今回の行動で得られた報酬、 $\alpha$ 、 $\gamma$  はそれぞれ学習率と割引率を表す。

$$Q(s_t, w_t) \leftarrow (1 - \alpha)Q(s_t, w_t) + \alpha[r_t + \gamma \max_w Q(s_{t+1}, w)] \quad (4)$$

なお、報酬関数は本来は最終的に得られた資源量と関連づけるべきであるが、ここでは計算を簡単にするために、交流が成立した場合に正の報酬が与えられることとした。また、負の報酬は考えない。

プレイヤーは市場へのエージェント投入にあたり、自己のエージェントにいずれかの戦略を適用すると共に、その戦略に見合った資源採取タスクを教示する。資源採取タスクは、どのような資源をどのようなタイミングでどれだけ採取するかというような行動セットである。例えば、友愛戦略をとるエージェントはできるだけ多く交流を成立させるために、他のエージェントが目的としていない資源を選択することが有利となる。そこで、もし探索の過程で自己が目的とする資源が多数エージェントの目的資源であることが分かった場合は、例えばより遠隔地にある別の資源を改めて目的とするというような行動を教示しておくことが考えられる。逆に、独善戦略をとるエージェントにはそのような配慮は無用なため、知りうる限りの資源の中から最も省資源で到達できるものを目指すことを教示すべきであろう。ただし、本稿で以上にあげたようなタスクはあくまで例示であり、戦略とタスクは明確に関連づかない。これらはあくまでプレイヤーの選択に委ねられる。

本稿では、これらの 4 つの戦略を実践する探索エージェントを用いた比較実験を行い、本稿で提案する資源探索空間においてはいずれの戦略をとることが優位となるのかを実証する。

### 3. 実験

#### 3.1 実験環境

前述のような資源探索空間を  $200 \times 200$  セルのトラス世界として仮構した。本実験では資源の属性についてはすべての資源で共通とした。つまり、すべての資源に代替性を設定した。これにより、エージェントはいずれの資源でも到達さえすれば獲得することができる。資源量はエージェントによる獲得により 1 ずつ減じる。0 になると環境から消滅する。

エージェントは資源を 3 採集することを目的として、各戦略を与え、戦略ごとに以下のタスクを設定した。友愛戦略では、自己の現在の到達資源のうち、それを目的とする他エージェント総数が最も少ないものを目的資源として行動するようにした。独善戦略では現在の到達資源のうち最も近接するものを目的とするようにした。日和見戦略では交流後に目的資源を見直すようにした。すなわち、もし交流が成立すれば友愛戦略のタスクと同様に、成立しなければ独善戦略のものと同様に振舞うようにした。

エージェントは初期値 500 の燃料を持ち、1 時刻に 1 セル移動し、その都度燃料を 1 ずつ消費する。燃料は目的資源に到達したときに 250 補充される。

日和見戦略を用いるエージェントが交流を実施するかを判断する際に利用する閾値は 3 とした。つまり相手エージェントが 3 以上の他エージェントと接触した場合には交流しない。

また、実験空間は資源が豊富な状況 (rich space: 豊穡な空間) と貧弱な状況 (poor space: 貧弱な空間) を想定し、資源数と初期状態量をそれぞれ (10, 3)、(5, 2) とする 2 空間とした。それぞれの空間について、同じ戦略をとるエージェント 4 台による 4 つの戦略についての 4 組のシミュレーション、及びすべてのエージェントを 1 台ずつ計 4 台用いた 1 組のシミュレーションを実施した。シミュレーションはすべてのエージェントがタスクを達成するか退場したときに終了することとした。このシミュレーションを 100 回試行し、各戦略について、平均タスク成功率・シミュレーション終了までの平均ステップ数・ゲーム終了要因について見た。

#### 3.2 実験結果

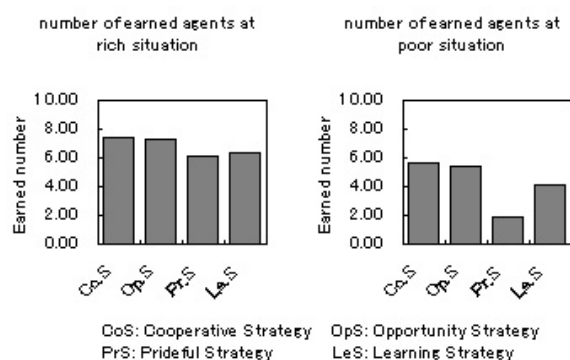


図 2: 平均探索タスク達成率 (上: 豊穡な環境, 下: 貧弱な環境)

図 2 は各戦略をとるエージェント群の平均探索タスク達成数である。豊穡な空間においてはすべての戦略において平均達成数に大きな差が見られない。つまり、他エージェントと交流することは資源獲得の成功にさほど影響を与えていない。一方で貧弱な空間では、友愛的戦略・日和見戦略・学習戦略をとる

エージェントの平均達成率が高く、独善戦略をとるエージェントの平均達成率が極端に低くなっている。これは、豊穡な空間では散漫な探索でも資源にたどり着くことができる場合が多いため、交流を行わないエージェントでもタスクを達成することができているのに対し、貧弱な空間では他エージェントから情報を得て活動しない限り資源に到達することが困難であり、単なる散漫な探索ではすぐに燃料が尽き退場してしまっていると考えられる。

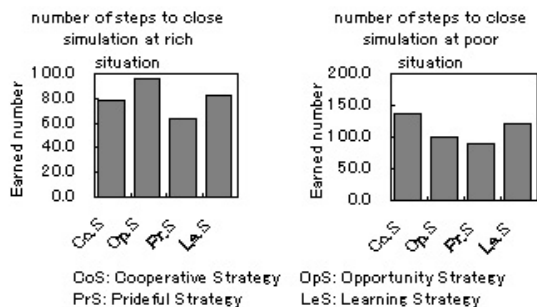


図 3: 平均シミュレーション終了ステップ数 (上:豊穡な環境, 下:貧弱な環境)

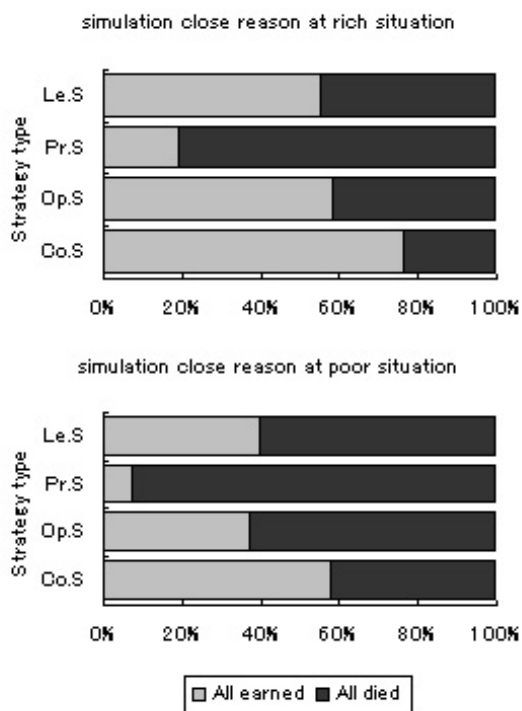


図 4: シミュレーション終了要因

図 3 は各戦略をとるエージェント群のシミュレーション終了までの平均ステップ数, 図 4 はシミュレーション終了要因を示している。ステップ数については友愛戦略をとるエージェント群と独善戦略をとるエージェント群において差が少なく、いずれも比較的早期に終了している。一方で、シミュレーションの終了要因は、友愛戦略群では「全エージェントのタスクの

達成」が多であるのに対し、独善戦略群では「全エージェントの退場」が多である。ここに、上述の通り資源獲得に失敗することが多い独善戦略群においては比較的早期に全エージェントが退場してしまうことが示されている。

#### 4. 考察

実験を通じて、本稿で提案する環境においては、友愛戦略をとるエージェント、すなわち資源の集中的な収奪に繋がる恐れがある選択をできるだけ避けようとする誘因を持つエージェント群がもっとも環境効用を増加させるという結果が得られた。すなわち、そのような市場に参加しようとするプレイヤー間では、友愛戦略をとるエージェントを投入することが均衡となる。従って、資源の持続的利用と乱獲的収奪を防止しようとする制度設計者は、提案環境を採用した資源獲得機構を構築することで、すべての資源を知悉した上で高度な能力を持つ上位制御者を設定するような負担を受けることなく、自由な獲得競争を維持した開放的な市場と持続的な資源利用を期待することができると思われる。

#### 5. まとめ

開放的な有限資源市場モデルにおける資源分配機構について、探索に有用な情報を他エージェントと戦略的に取引する能力を有するエージェントで構成される資源探索環境を提案した。実験により、提案環境においては積極的に他エージェントと交流しようとするエージェント群ほど優位であることがわかった。これにより、資源の持続的利用と乱獲的収奪を防止しようとする制度設計者は、提案環境を採用することで、自由な獲得競争を維持した開放的な市場と持続的な資源利用を期待することができると思われる。

#### 参考文献

[Mikami 94] Mikami, S. and Kakazu, Y.: Co-operation of Multiple Agents Through Filtering Payoff, in Proc. 1st European Workshop on Reinforcement Learning, EWRL-1, pp. 97-107 (1994)

[森山 02] 森山甲一, 沼尾正行: 環境状況に応じて自己の報酬を操作する学習エージェントの構築, 人工知能学会論文誌, Vol. 17, No. 6, pp. 676-683 (2002)

[Akiyama 02] Akiyama, E., Kaneko, K.: Dynamical systems game theory II A new approach to the problem of the social dilemma, Physics D 167, pp. 36-71 (2002)

[谷本 06] 谷本潤, 脇山宗也, 相良博喜: 外生的資源制約を課した環境ジレンマゲームの普遍モデルに関する研究, 日本建築学会環境系論文集, No. 610, pp. 107-112 (2006)

[佐藤 07] 佐藤勇気, 西野成昭, 上田完次: 不完備情報ゲームを用いた資源消費の意思決定に関する研究, 第 21 回人工知能学会全国大会論稿集, <http://www.ai-gakkai.or.jp/jsai/conf/2007/data/paper-225.html> (2007)