

動的離隔型 GA (DS-GA) の提案

中山 功^{†,††} 松井 博和^{†††} 野村 由司彦^{†††}

マルチエージェントシステム(以下 MAS と記す)に適用する遺伝的アルゴリズム(以下 GA と記す)の 1 つとして、動的離隔型 GA (Dynamically Separating GA, 以下 DS-GA と記す)を提案する。個体を離隔する GA は島モデル GA などがあるが、DS-GA は、離隔状態が個体数に応じて動的に変化する点が異なる。本論文では、システム最適性と個体最適性が異なる MAS や、環境が動的に変化する MAS への適用を考え、これらの MAS の学習に有効な DS-GA の性質について述べる。また、エージェントの協調作業と分業により組織化が実現した例について述べる。

Dynamically Separating GA (DS-GA)

KOICHI NAKAYAMA,^{†,††} HIROKAZU MATSUI^{†††}
and YOSHIHIKO NOMURA^{†††}

This paper proposes “Dynamically Separating Genetic Algorithm (DS-GA)” that is applied to a multi-agent system (MAS). There have been proposed separating genetic algorithms such as “Island model” GA where agents are ever restricted to contact with each other, and it can be referred to a “statically separating GA”. Differently from the statically separating GA, dynamically separating GA is proposed: the border of restriction changes dynamically according to the number of the agents. Some important characteristics of DS-GA considered to be effective for MAS are examined in such performance to maximize the whole system profit, to attain balance of agents’ profits, and to search the optimum solutions. Furthermore, an experiment is carried out to show that an organization can be realized as a result of cooperative division of work between agents.

1. はじめに

マルチエージェントシステム(以下 MAS と記す)の研究として、MAS に学習アルゴリズムを適用し、未知の環境や動的な環境へのエージェントの適応^{1),2)}や、協調作業や分業による組織化³⁾などによる目的達成能力の向上を目指す研究がある。これらの研究では、個々のエージェントが、システムの大域的目的やその達成度などの一部しか知覚できないため、エージェントが知覚できる範囲に基づく学習を、いかにシステム全体の大域的目的と整合させるかが重要である⁴⁾。本論文では、エージェントが知覚できる範囲の情報である利得や報酬、評価など(以下利得に統一する)を個体最適性、システムの大域的目的として、すべてのエー

ジェントがそれぞれの行動から得る利得の総和をシステム最適性とそれぞれ定義し、エージェント学習によるシステム最適性の向上を考える。

本論文では、MAS に遺伝的アルゴリズム(以下 GA と記す)を適用する研究の 1 つとして、動的離隔型 GA (Dynamically Separating GA, 以下 DS-GA と記す)を提案する。DS-GA は、個体数に応じて個体をコロニーと呼ぶグループに動的に離隔する GA である。個体を離隔する GA として島モデル GA⁵⁾などがあるが、DS-GA は離隔状態が個体数に応じて動的に変化する点で異なる。

ここでは、DS-GA を適用する MAS に、(A) ジレンマ環境と (B) 動的環境と呼ぶ 2 つの環境により、従来の GA で高いシステム最適性を得ることが難しいといわれるものを用いる。

(A) ジレンマ環境：個体最適性が高い行動とシステム最適性が高い行動が、完全には一致しない環境をジレンマ環境と呼ぶ。この環境に GA を適用すると、高いシステム最適性が得られるとは限らない⁶⁾。

† 京都大学大学院情報学研究科

Graduate School of Informatics, Kyoto University

†† ATR 人間情報科学研究所

ATR Human Information Science Laboratories

††† 三重大学工学部

Faculty of Engineering, Mie University

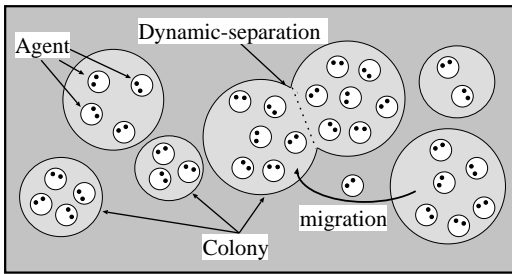


図1 動的離隔型 GA の概念

Fig. 1 Conceptual representation of Dynamically Separating GA.

(B) 動的環境：システムを取り巻く環境の変化により、システム最適性も変化する環境を動的環境と呼ぶ。この環境に GA を適用すると、高い最適性がつねに得られるとは限らない。

これらの環境である MAS に対する DS-GA の性質を、従来の単純な GA (以下 sGA と記す)、島モデル GA、群淘汰 GA の 3 手法との比較により述べる。

本論文の構成は以下のとおりである。2 章で、DS-GA を提案し、DS-GA を用いる MAS の学習アルゴリズムについて述べる。3 章で、DS-GA の性質として、(A) ジレンマ環境において協調作業や分業を実現しやすい性質について述べる。4 章で、(B) 動的環境に適応しやすい性質について述べる。5 章で、より現実的な問題に対する有効性を考察するために、(A)、(B) の複合環境である工学的 MAS に DS-GA を適用し、3 章と 4 章で述べる性質により組織化が実現する例について述べる。6 章で、まとめと今後の課題を述べる。

2. 動的離隔型 GA (DS-GA) の提案

2.1 DS-GA の概念

DS-GA では、各個体は接触できる個体の限界数をもちコロニーと呼ぶグループごとに離隔される。異なるコロニーに存在する個体とは接触できない。同一コロニー内に存在する個体数が増加し限界数以上になる場合、そのコロニーに存在する個体は、さらに半数ずつのコロニーに離隔される。個体数が 0 になる場合、そのコロニーを消滅させる。また、各個体は、移動確率に従い他のコロニーに移動する。DS-GA の概念を図 1 に示す。

2.2 DS-GA を用いる学習アルゴリズム

MAS における学習では、エージェントが自らの知覚に基づく学習手法が不可欠である。強化学習は、エージェント自身で行動を評価し学習できるため、MAS に適用する学習手法として多く用いられる。一方、GA は、進化的学習オペレータに用いる淘汰や選択などが、

MAS の外部に依存する場合が多く、MAS に適用する研究は強化学習に比べ少ない。そこで、本論文では、エージェントの自律性を損なわない GA による学習手法として、資産(累積利得)に基づくエージェントの分裂、または消滅による学習アルゴリズム^{7),8)}を用いる。DS-GA で用いる学習アルゴリズムを以下に述べる。なお、このアルゴリズムによる MAS の進化的学習を、本論文では単に学習と呼ぶ。

- (1) 初期設定：仮想環境内に、 $N_A(t)$ 個 ($t=0$) の個体を作成する。個体 a は、初期資産 $U_A(a, t)$ と無作為に選択される行動決定遺伝子 $Gene_{Act}(a)$ を持つ。各個体を限界数 N_{Lim} ごとのコロニーに離隔する。
- (2) 行動：全個体が、それぞれの遺伝子に従い 1 回ずつ行動(対戦, 仕事, 贈与などの行為の集合)する。このとき、個体は、同一コロニー内にいる相手の行動と自分の行動との相互作用により決定される利得を得る。この利得は、資産として各個体に累積される。ただし、相互作用による利得の値は、本学習アルゴリズムを適用するそれぞれの MAS により異なる。
- (3) 分裂と消滅：分裂は、個体の資産が初期値の倍以上になると起きる。このとき、個体は、分裂前の個体の資産を半分ずつ持つ 2 個体に分裂し、それぞれ遺伝子を引き継ぐ。ただし、遺伝子は突然変異確率 (mutation probability) P_{mut} で変異する。消滅は、個体の資産が 0 以下になると起き、個体が消滅する。
- (4) 移動：個体の移動は、移動確率 (migration probability) P_{mig} で起きる。このとき、個体は、無作為に選ばれるコロニーに移動する。
- (5) コロニーの動的離隔：コロニーの動的離隔は、コロニー内の個体が限界数 N_{Lim} を超えると起きる。このとき、1 つのコロニー内に存在する個体は、2 つのコロニーに離隔される。ただし、2 つのコロニーの個体数の差は 1 以下とする。
- (6) ランダム消去：本手法のアルゴリズムには、MAS に存在できる個体数に制限はないが、計算機の容量には制限がある。そこで、個体総数が初期個体数より増えた場合、観測するコロニー数を限定する意味で、個体総数が初期個体数以下になるまで、コロニーを無作為に消去する。
- (7) 単位時間ループ：(2)~(6) を繰り返す。これを 1 単位時間とし、これに基づく時刻を t で表す。

DS-GA を用いる学習アルゴリズムのメインルーチンを NS チャートで図 2 に示す。

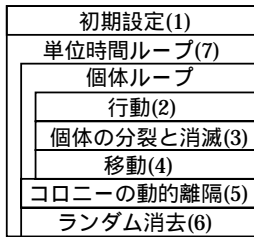


図2 NSチャートで示すDS-GAのメインルーチン
Fig. 2 Main routine of DS-GA shown by NS chart.

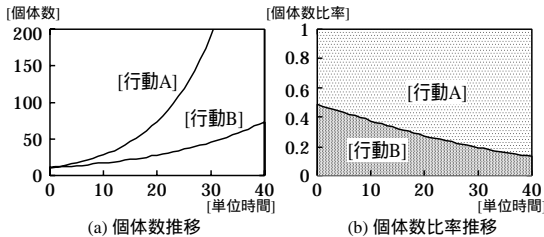


図3 個体数比率推移の概念
Fig. 3 Schematic explanation on the history of population ratio change.

このアルゴリズムは、個体の増加が累積利得に比例し、個体の消去が無作為であるため、累積利得に基づくルーレット選択による進化的オペレータと近い性質を持つ。例として、単位時間ごとの利得の期待値がそれぞれ1.2, 1.1の行動A, Bの個体数の推移を図3(a)に、個体数比率の推移を図3(b)に示す。

利得の期待値が大きい行動が等比数率的に増加し、MAS全体が行動Aを学習することが分かる。

2.3 比較する従来手法のアルゴリズム

以降の章でDS-GAの性質を検証するために比較する従来手法のアルゴリズムについて、上記の学習アルゴリズムと比較して述べる。

- sGA: 隔離のない、従来の単純なGAである。(1) 初期設定で個体を隔離せず、(4) 移動や(5) コロニーの動的隔離がない。(6) ランダム消去では、個体を無作為に消去する。各個体の遺伝子、分裂条件、突然変異などはDS-GAと同じものを用いる。
- 島モデルGA: 隔離状態が変化しない、複数集団GAである。(1) 初期設定では、DS-GAと同じように、個体を限界数ごとの島に隔離する。しかし、(5) 島の動的隔離はなく、(6) ランダム消去では、各島の個体が限界数を超えるとき、その島内の個体を無作為に消去する。各個体の遺伝子、移動確率、分裂条件などはDS-GAと同じものを用いる。
- 群淘汰GA: 群そのものを1つの個体と見なし、群内の全個体の遺伝子を1つの群の遺伝子とし

て扱うGAである。(1) 初期設定では限界数ごとの群を作成し、(2) 行動は群内の個体どうしで行うが、群内の全個体が得た利得の総和の累積を群の資産とする。(3) 分裂は、群の資産が初期値の倍以上になると起きる。このとき、群は、分裂前の群の資産を半分ずつ持つ2群に分裂し、それぞれ遺伝子を引き継ぐ。ただし、それぞれの群の遺伝子は、個体単位の遺伝子ごとに、突然変異確率 P_{mut} で変異する。消滅は、群の資産が0以下になると起き、群そのものが消滅する。(4) 移動と(5) 群の動的隔離はなく、(6) ランダム消去では、群を無作為に消去する。

3. ジレンマ環境への適応

MASでは、個々のエージェントが自らの利得(個体最適性)に基づき、システムの目的達成能力(システム最適性)が高いエージェントの行動の組合せを学習する必要がある。しかし、ジレンマ環境では、学習により高いシステム最適性が得られるとは限らない⁶⁾。また、ジレンマを解消するように利得の最適な割当てをあらかじめ求めること(Credit Assignment Problem, 以下利得割当て問題と記す)は、難しい場合が多い⁹⁾。

本章では、3.1節で、本章および次章で用いる対戦型モデルについて述べる。3.2節で、囚人のジレンマモデルを用いてシステム最適性の獲得について述べる。3.3節で、贈与付きジレンマモデルを用いて利得割当ての獲得について述べる。

3.1 対戦型モデル

MASにおけるタスクの実行や問題解決などの行動は、複数のエージェントの行動選択問題とみることができる¹⁰⁾。そこで、本章および次章では、DS-GAの性質を検証する単純なMASの実験モデルとして、2.2節の学習アルゴリズム(2)行動の部分に、囚人のジレンマなどで用いられる以下のような対戦型モデル(行動選択ゲーム)を適用する。

(2-1) 単位時間ごとに、対戦相手として自己以外の個体を無作為に選択する。ただし、DS-GAのコロニー、sGAの全個体、島モデルGAの島、群淘汰GAの群をそれぞれグループと呼び、同一グループ内に存在する個体を選択する。

(2-2) 対戦する2個体は、それぞれの遺伝子 $Gene_{Act}(a)$ に従い行動を選択する。

(2-3) 互いの選択行動の組合せに応じて利得を得る。遺伝子と利得の詳細はモデルごとに異なるので、各節で述べる。本章では、初期個体数 $N_A(0) = 10,000$,

表 1 囚人のジレンマモデルの利得
Table 1 Profit table of Prisoner's Dilemma Model.

自分 \ 相手	協調C	裏切D
協調C	4	1
裏切D	5	2

限界数 $N_{Lim} = 10$, 突然変異確率 $P_{mut} = 0.001$, 移動確率 $P_{mig} = 0.001, \forall a$, 初期資産 $U_A(a, 0) = 20$ とし, 異なるランダム系列を用いて 100 試行する.

3.2 システム最適性の獲得

ジレンマ環境である MAS に sGA を適用すると, より高い個体最適性を得る行動を学習するため高いシステム最適性が得られない⁶⁾. そこで, 本節では, ジレンマ環境として繰返しのない単純な囚人のジレンマモデルを用いて, DS-GA におけるシステム最適性の獲得について述べる.

3.2.1 囚人のジレンマモデル

MAS におけるジレンマ環境の標準問題の 1 つとして, 囚人のジレンマモデル¹¹⁾がある. 囚人のジレンマモデルは, それぞれの囚人が協調 C と裏切 D のいずれかの行動を選択する対戦モデルである. 本節のモデルの利得を表 1 に示す. 囚人のジレンマモデルの環境は, 互いの利得の総和 (システム最適性) は, 相手の行動にかかわらず協調 C を選択する方が高いが, 自らの利得 (個体最適性) は, 相手の行動にかかわらず裏切 D を選択する方が高いジレンマ環境である. 各個体の遺伝子は, 中間的な戦略が有利な場合があること^{12),13)}を考慮し, 裏切 D を選択する確率を示す遺伝子 $Gene_{Act}(a) \in \{ \frac{0}{10}, \frac{1}{10}, \dots, \frac{10}{10} \}$ とする.

3.2.2 実験と考察

本節の実験では, 100 回の試行すべてにおいて以下のような様相を示した. 各 GA を用いた MAS の学習過程における, 全個体の利得の平均 (平均利得) の推移を図 4 に, 遺伝子別個体数比率の推移を図 5 に示す.

DS-GA では裏切確率 $\frac{0}{10}$ の協調個体が, 群淘汰 GA では裏切確率 $\frac{4}{10}$ 以下の協調的個体がそれぞれ多数を占め, 協調行動を学習した. sGA や島モデル GA では, 裏切確率 $\frac{10}{10}$ の遺伝子を持つ裏切個体が多数を占め, 裏切行動を学習した.

DS-GA において協調個体が増加するときに作用していると考えられるメカニズムを以下に説明する. 説明のために, 協調個体 $\frac{0}{10}$ と裏切個体 $\frac{10}{10}$ の 2 種類に簡略化し, コロニー数を 3 に制限する場合に, 横軸に時間, 縦軸にコロニー内の協調個体比率をとる推移グ

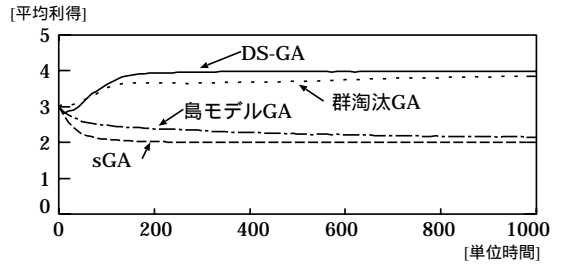


図 4 平均利得の推移
Fig. 4 History of the average profit.

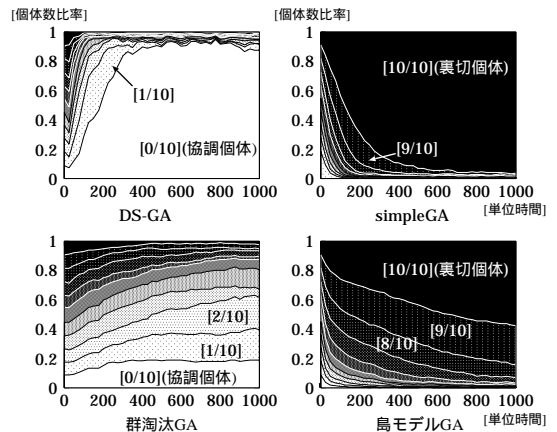


図 5 遺伝子別個体数比率の推移
Fig. 5 History of the population ratios among various genes.

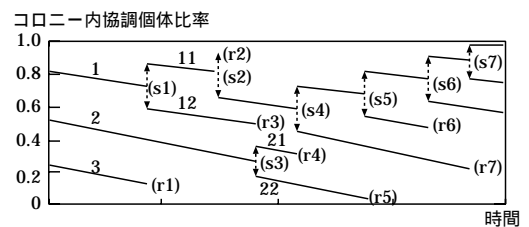


図 6 協調個体比率の推移の概念
Fig. 6 Schematic explanation on the increase in population of the cooperative agents.

ラフを図 6 に示す. 3 本の線が 3 つのコロニーを示し, コロニーの動的隔離を線の分裂 (s1 ~ s7) で, それに対応するコロニーのランダム消去を線の消滅 (r1 ~ r7) で示す. 線 1 からできた線 11 と線 12 が示すように, 同じコロニーから動的隔離によりできるコロニーの協調個体比率の違いは, 図 7 に示すように, 無作為に個体を動的隔離する場合にコロニー内個体数比率がゆらぐ可能性を示す.

各コロニー内部における個体の増加率の差による学習 (グループ内学習) を考えると, 個体最適性が高い

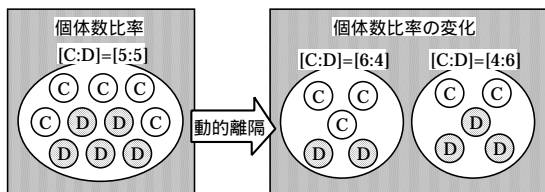


図 7 動的離隔の概念

Fig. 7 Conceptual representation of dynamic separation.

個体ほど増加率が高いため裏切個体が増加し、それぞれのコロニーの協調個体比率を示す線 1, 2, 3 などは離隔あるいは消去まで下がりつづける。一方、コロニー間における個体の増加率の差による学習（グループ間学習）を考えると、システム最適性が高いコロニーに存在する個体ほど増加率が高いため、協調個体比率が高いコロニーに存在する個体ほど早く動的離隔する（図 6 (s1), (s3)）。このため、グループ間学習では、協調個体比率が高いコロニーに存在する個体ほど個体数が増加する。このように、グループ内学習により各コロニーにおける協調個体比率は低下するが、グループ間学習により協調個体比率の高いコロニーに存在する個体が増加するため、MAS 全体としては協調行動を学習したと考えられる。

DS-GA は、上記のようにグループ内学習とグループ間学習を持ち、sGA と鳥モデル GA はグループ内学習のみ、群淘汰 GA はグループ間学習のみを持つ。囚人のジレンマモデルに限らず、システム最適性がより高いグループに存在する個体ほどグループ間学習による増加率が高いため、DS-GA と群淘汰 GA では、MAS 全体としても高いシステム最適性を学習する可能性が高いと考えられる。このメカニズムの結果、高いシステム最適性を学習しやすい DS-GA の性質を、本論文では「性質 1：利得の最大化」と呼ぶ。また、動的離隔による個体数比率のゆらぎにより、より高いシステム最適性を得る個体数比率を得たコロニー内個体は、個体の増加率が高い。このメカニズムの結果、システム最適性の高い個体の組合せ（個体数比率）を探索する DS-GA の性質を、本論文では「性質 2：動的離隔探索」と呼ぶ。

囚人のジレンマに基づき、ジレンマ環境でのシステム最適性を得ようとする研究では、エージェントそのものに協調させるメカニズムを組み込むものが多い^{(10),(14),(15)}。本手法では、システム最適性を得るメカニズムが DS-GA に含まれるため、エージェントに協調させるメカニズムを組み込む必要がない。

3.3 利得割当ての獲得

サッカーのシュートとアシストのように、直接タス

表 2 贈与付きジレンマモデルの利得
Table 2 Profit table of Dilemma model with donation.

自分 \ 相手	補助C	採取D
補助C	0	0
採取D	5	1

クを達成する行動と、それを補助をする行動との分業によりタスクを達成し、高いシステム最適性を得ようとする研究（モンキーバナナ問題⁽⁶⁾、追跡問題^{(4),(11)}など）がある。これらエージェントの役割が不均質なモデルにおける問題の 1 つとして、貢献度に基づく個々のエージェントへの利得の割当てがある。群淘汰 GA のように、タスク達成による利得を全個体に一律に与えると、タスク達成に貢献しない行動まで学習される可能性がある。また、DS-GA、sGA、鳥モデル GA のように、直接タスクを達成した個体だけに利得を与えると、タスク達成に貢献する補助行動は学習されない可能性がある。そこで、本節では、資産の贈与機能を持つ個体による贈与付きジレンマモデルを用いて、エージェント利得割当ての獲得について述べる。

3.3.1 贈与付きジレンマモデル

贈与付きジレンマモデルは、補助 C と採取 D のいずれかを選択する対戦モデルに、自らの資産から対戦相手にいずれかの量（0, 1, 2, 3, 4, 5）を渡す贈与を加えるモデルである。本節の利得を表 2 に示す。

補助 C は、相手の採取を補助しようとするが、自らは利得が得られない。一方、採取 D は、利得を採取し、相手の補助により、より高い利得を得られる。この対戦モデルの環境は、相手の行動にかかわらず採取 D がより高い利得（個体最適性）を得るが、補助 C と採取 D との分業により高いシステム最適性を得るジレンマ環境である。また、贈与は、贈与量がより少ない個体ほど高い個体最適性を得るが、システム最適性は贈与量にかかわらず一定である。各個体の遺伝子は「補助 C か採取 D」と「6 段階の贈与量 0, 1, 2, 3, 4, 5」の組合せを示す遺伝子 $Gene_{Act}(a) \in \{[C:0], [C:1], \dots, [C:5], [D:0], [D:1], \dots, [D:5]\}$ とする。

3.3.2 実験と考察

本節の実験では、100 回の試行すべてにおいて以下のような様相を示した。各 GA における平均利得の推移を図 8 に、遺伝子別個体数比率の推移を図 9 に示す。

本節のモデルではタスク達成に貢献しない行動を用いないため、補助個体と採取個体がほぼ半々の比率で

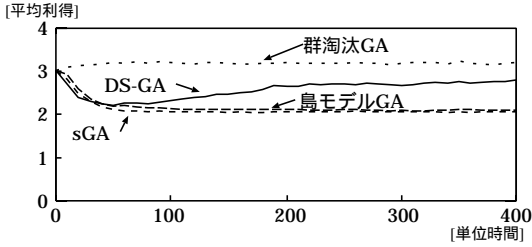


図 8 平均利得の推移
Fig. 8 History of the average profit.

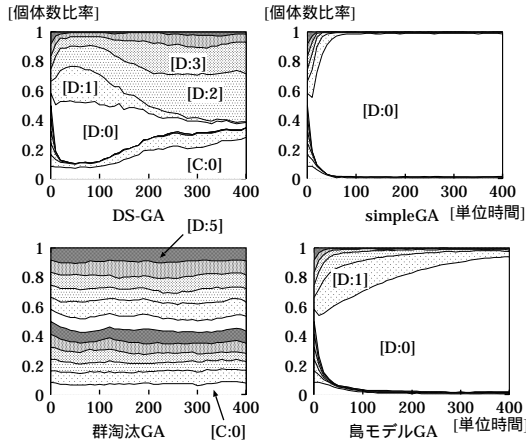


図 9 遺伝子別個体数比率の推移
Fig. 9 History of the population ratios among various genes.

存在する初期状態において、比較的高いシステム最適性を得る単純なモデルである。群淘汰 GA は、贈与量を探索しないため、高いシステム最適性を得る補助個体数比率である群が初期状態に多数存在し、つねに高い平均利得を得た。DS-GA は、贈与しない補助個体 [C:0] と、資産を贈与する採取個体 [D:2], [D:3] が共存した。ここで、最適な補助個体数比率(付録 A.1 参照)は 0.375 であるが、500~1000 単位時間の平均補助個体数比率は 0.346 であった。また、そのときの採取個体から補助個体への利得の最適贈与量(付録 A.2 参照)は 2.50 であるが、500~1000 単位時間の平均贈与量は 2.37 であった。これらは、いずれも最適値の 90%以上の値であり、両者はよく一致したといえる。

DS-GA において、贈与により利得が均衡化されるときに作用していると考えられるメカニズムを、図 10 を用いて以下に説明する。(a) は利得が均衡化されているコロニーを、(b) は均衡化されていないコロニーを示す。また、大きい円はコロニーを、小さい円はエージェントを、内部の文字が行動を、括弧内の数値は利得の期待値を示す。

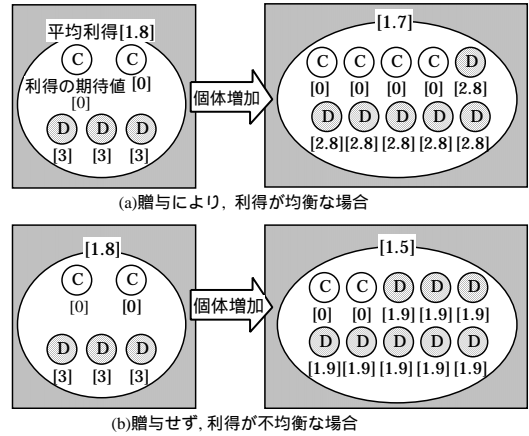


図 10 利得の均衡化の概念

Fig. 10 Schematic explanation on attaining a proper balance through D-agents' distributing profits to C-agents.

DS-GA では、前節で示された <性質 1: 利得の最大化> により、システム最適性がより高いコロニー内個体が増加し、<性質 2: 動的離隔探索> により、より高いシステム最適性を得る遺伝子別個体数比率を獲得する。利得が均衡化されている場合、図 10 (a) のように個体数比率が維持される。利得が均衡化されておらず、個体増加率が遺伝子の種類ごとに異なる場合、図 10 (b) のように個体数比率が変化し、システム最適性も変化する。贈与付きジレンマモデルに限らず、より高いシステム最適性を得ており、かつ利得がより均衡化されているコロニーに存在する個体ほど高い増加率が維持されるため、MAS 全体としても高いシステム最適性を得る利得割当てを学習する可能性が高いと考えられる。このメカニズムの結果、エージェントの利得が均衡化されやすい DS-GA の性質を、本論文では <性質 3: 利得の均衡化> と呼ぶ。

ここで、自らが採取 (D)、相手が補助 (C) のとき得られる利得 $X(D,C)$ がさまざまな値の場合(表 3)における、理論上、最大のシステム最適性を得る最適補助個体数比率(付録 A.1 参照)と、それを維持する採取個体から補助個体への最適贈与量(付録 A.2 参照)、および、DS-GA の実験結果を表 4 に示す。

すべての利得 $X(D,C)$ において、最適個体数比率と最適贈与量の 90%以上を得た。これらの結果から、DS-GA は、利得の贈与と機能を用いることで、利得割当て問題を自律的に解決する可能性が高いと考えられる。また、この性質を利用することで、資源割当て問題に対する協調問題解決^{17),18)} などへの適用が期待される。

表 3 X(D,C) がさまざまな値の利得

Table 3 Profit table for generalized profit value, X(D,C).

自分 \ 相手	補助C	採取D
補助C	0	0
採取D	X(D,C)	1

表 4 X(D,C) がさまざまな値の補助個体数比率と贈与量

Table 4 The ratio of assistant agents and the amount of donation: those by simulation are compared with those by theory.

X(D,C)		5	6	7	8	9	10
補助 個体数 比率(%)	最適値	37.5	40.0	41.7	42.9	43.8	44.4
	実験結果	34.6	38.9	40.7	43.0	41.8	44.7
贈与量	最適値	2.50	3.00	3.50	4.00	4.50	5.00
	実験結果	2.36	2.97	3.37	3.97	4.24	4.94

4. 動的環境への適応

エージェントにとっての動的環境とは、互いの学習や相互作用などが原因でエージェントからみる環境が動的である場合と、エージェントの対戦表そのものの変化などが原因でシステム環境そのものが動的である場合がある。実際の MAS では、後者の動的環境にも適応することが望ましい。本論文では、このような動的環境への適応を、(1) 最適解への収束 (2) 最適解の変化による局所解からの脱出 (1) に戻る... と考え、(1) を環境変化速度と学習による収束速度との競争の問題、(2) を局所解に収束した状態からの脱出問題であると考え。

本章では、4.1 節で、システム環境を静的にした同一行動選択モデルを用いて、(1) の問題について述べる。4.2 節で、変化する利得表により、システム環境を動的にした同一行動選択モデルを用いて、(2) の問題について述べる。本章の実験でも 3.1 節で述べた対戦型モデルを使用する。また、初期個体数 $N_A(0) = 10,000$ 、突然変異確率 $P_{mut} = 0.05$ 、限界数 $N_{Lim} = 10$ 、移動確率 $P_{mig} = 0.001, \forall a$ 、初期資産 $U_A(a, 0) = 20$ とし、異なるランダム系列を用いて 100 試行する。

4.1 収束速度

システム環境が静的である MAS では、最適解(最適個体数比率)へ収束することで学習が終了する。エージェントにとっての環境は、互いの学習による行動の変化や、接触するエージェントの違いなどにより動的な環境(非マルコフ環境)となるため、最適解への収束に時間がかかる。そこで、本節では、同一行動選択モデルを用いて収束速度について述べる。

表 5 同一行動選択モデルの利得

Table 5 Profit table of an identical action model: identical actions return stationary profits to pairing agents.

自分 \ 相手	行動 A	行動 B	行動 C	行動 D
行動 A	1	0	0	0
行動 B	0	2	0	0
行動 C	0	0	3	0
行動 D	0	0	0	4

4.1.1 同一行動選択モデル

同一行動選択モデルは、4つの行動 A, B, C, D の中から対戦相手と同一の行動を選択することにより利得を得るモデルである。本節のモデルの利得を表 5 に示す。このモデルは、システム環境は静的であるが、個体最適性の高い行動が相手の行動により異なるため、エージェントからみる環境は動的である。各個体の遺伝子は選択する行動を示す遺伝子 $Gene_{Act}(a) \in \{A, B, C, D\}$ とする。

4.1.2 実験と考察

本節の実験では、100 回の試行すべてにおいて以下のような様相を示した。各 GA における平均利得の推移を図 11 に、遺伝子別個体数比率の推移を図 12 に示す。

約 30 単位時間までに現れた短期的な性質として、グループごとに同一行動に収束した DS-GA と島モデル GA の平均利得が、sGA や群淘汰 GA より早く増加した。約 30 単位時間以降に現れた長期的な性質として、DS-GA と sGA の平均利得が、島モデル GA や群淘汰 GA に比べ早く最適解付近まで増加した。

DS-GA において、比較的早く収束するとき作用していると考えられるメカニズムをグループ内学習とグループ間学習に分けて以下に説明する。

DS-GA や島モデル GA のグループ内学習では、sGA と比べ接触する個体数が少ない。一般的に、GA では個体数が少ないほど早く収束する。特に MAS では、個体数が多いほど相互作用により環境が動的になる。このため、DS-GA や島モデル GA では、グループ内学習の効率が良いと考えられる。また、群淘汰 GA ではグループ内では学習しない。

DS-GA のグループ間学習では、グループ内学習と <性質 2: 動的隔離探索> により、システム最適性の高いコロニーが現れる可能性があるため、グループ間のシステム最適性に差が付きやすい。群淘汰 GA のグループ間学習では、グループ内学習がないため、グループ間のシステム最適性の差が DS-GA に比べ小さい。このため、DS-GA は、群淘汰 GA よりグループ

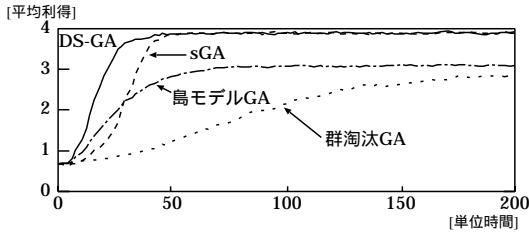


図 11 平均利得の推移
Fig. 11 History of the average profit.

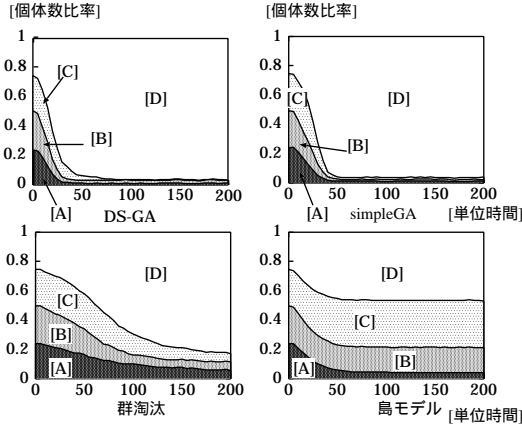


図 12 遺伝子別個体数比率の推移

Fig. 12 History of the population ratios among various genes.

間学習の効率は良いと考えられる。また、sGA、島モデル GA ではグループ間では学習しない。

探索開始時には、グループ内学習による個体の増加率の差がグループ間学習による増加率の差より大きいため、短期的にはグループ内学習の効率が良い DS-GA や島モデル GA の収束が早かった。しかし、接触がグループごとに制限される手法の場合、全体の収束はグループ間学習に依存する。このため、長期的には、グループ間学習の効率が良い DS-GA と、接触の制限がない sGA が全体の収束が早かったと考えられる。

同一行動選択モデルに限らず、DS-GA は、グループ内学習とグループ間学習、および「性質 2：動的離隔探索」により、比較的早く収束すると考えられる。このメカニズムの結果、今回比較した従来手法より早く収束しやすい DS-GA の性質を、本論文では「性質 4：高収束性」と呼ぶ。

4.2 局所解からの脱出

GA を適用する MAS では、局所解からの脱出に複数エージェントの突然変異が必要となる場合があり、収束した局所解からの脱出は難しい。そこで、本節では、システム最適性が 4,000 単位時間周期で変化する利得表を用いる同一行動選択モデルを用いて、収束し

表 6 変化する同一行動選択モデルの利得

Table 6 Profit table of another identical action model: temporary changing profits are returned by identical actions.

自分 \ 相手	行動 A	行動 B	行動 C	行動 D
行動 A	F(t)	0	0	0
行動 B	0	F(t-1000)	0	0
行動 C	0	0	F(t-2000)	0
行動 D	0	0	0	F(t-3000)

$$F(x) = 1 + 2\sin(n\pi / 500) \quad (0 \leq n < 1000 \text{ のとき})$$

$$F(x) = 1 \quad (1000 \leq n < 4000 \text{ のとき})$$

ただし、n は x を 4000 で割った余り ($n = x \bmod 4000$)

た局所解からの脱出について述べる。

4.2.1 変化する利得表を用いる同一行動選択モデル

変化する利得表を用いる同一行動選択モデルは、利得以外は前節のモデルと同じである。本節のモデルの利得を表 6 に示す。このモデルの環境は、システム最適性の高い行動の組合せが時刻により変化するため、システム環境も動的である。各個体の遺伝子は、前節と同じく選択する行動を示す遺伝子 $Gene_{Act}(a) \in \{A, B, C, D\}$ とする。

4.2.2 実験と考察

本節の実験では、100 回の試行すべてにおいて以下のような様相を示した。各 GA における平均利得の推移を図 13 に、遺伝子別個体数比率の推移を図 14 に示す。また、表 7 に、各 GA の平均利得の 12,000 単位時間における時間平均と、つねに 1 つの同一行動を選択する戦略（局所解戦略）、つねにその時刻の最適同一行動を選択する戦略（最適解戦略）で占められた場合の時間平均を示す。

DS-GA では、最適でない同一行動（以下局所解と記す）に収束している状態から、最適同一行動（以下最適解と記す）へと変化する様相を示した。sGA と群淘汰 GA では最初に収束した同一行動から脱出しなかった。島モデル GA では、多様性が維持されているため各時刻における最適解個体は存在するものの、個体数比率に大きな変化はみられず、全体が最適解へと収束することはなかった。このため、平均利得の 12,000 単位時間における時間平均が、DS-GA は局所解戦略より高く、他の 3 手法は局所解戦略より低かった。

DS-GA において、全個体が局所解に収束した状態から脱出するときに作用していると考えられるメカニズムを以下に説明する。グループ内学習が最適解へ向かうためには、最適解個体が、個体最適性に関して局所解個体を上回る必要がある。このとき、グループ内の最適解個体数比率は $r_a > 0.25$ （付録 A.3 式 (12)）を超える必要がある。また、グループ間学習が最適解

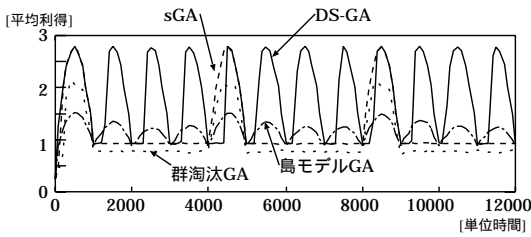


図 13 平均利得の推移

Fig. 13 History of the average profit.

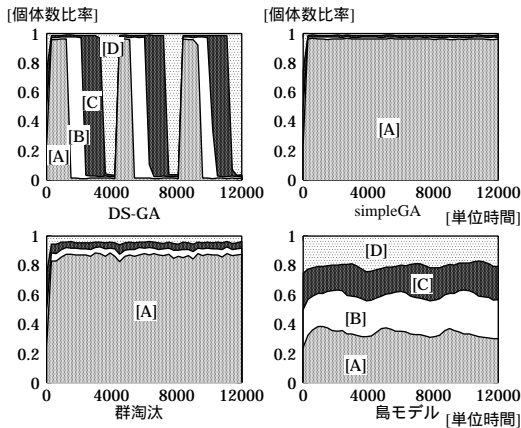


図 14 遺伝子別個体数比率の推移

Fig. 14 History of the population ratios among various genes.

表 7 平均利得の 12,000 単位時間における時間平均

Table 7 The average profit per unit time in 12,000 units.

DS-GA	sGA	群淘汰	島モデル	局所解	最適解
1.82	1.20	0.97	1.16	1.32	2.27

へ向かうためには、最適解を含むグループが、システム最適性に関して全個体が局所解であるグループを上回る必要がある。このとき、グループ内の最適解個体数比率は $r_g > 0.5$ (付録 A.3 式 (13)) を超える必要がある。

DS-GA は、グループ内学習とグループ間学習を持つ。本節のモデルのように $r_a < r_g$ の場合、DS-GA は、突然変異により 1 つのコロニーの最適解個体数比率が r_a を超えれば、グループ内学習によりいずれ r_g を超え、グループ間学習により MAS 全体で最適解を学習できる。sGA は、接触個体数が多く r_a を超える確率が低い。島モデル GA は、グループ間学習がなく、各々の島で r_a を超える必要がある。群淘汰 GA は、群の最適解個体数比率が r_a より大きい r_g を超える必要がある。 $r_a > r_g$ の場合、DS-GA や群淘汰は、グループ間学習がない sGA や島モデル GA に比

べて脱出する確率が高い。さらに、DS-GA では、動的離隔前の最適解個体数比率が r_a や r_g 以下であっても、<性質 2: 動的離隔探索>により r_a や r_g を超える可能性がある。このため、DS-GA では局所解から脱出したと考えられる。

変化する利得を用いる同一行動選択問題に限らず、DS-GA では、グループ内学習とグループ間学習の組合せと<性質 2: 動的離隔探索>により、局所解に収束した状態から MAS 全体が最適解を学習する確率が、他の 3 手法に比べて高いと考えられる。このメカニズムの結果、今回比較した従来手法より局所解から脱出する確率が高い DS-GA の性質を、本論文では<性質 5: 高脱出性>と呼ぶ。

なお、3.1 節のモデルで、突然変異確率 $P_{mut} = 0.05$ とし、全個体が裏切個体 ($Gene_{Act}(a) = \frac{10}{10}$) の状態から実験したところ、DS-GA では平均約 200 単位時間で全行動の $\frac{3}{4}$ が協調 C となったことから、局所解から脱出する性質はジレンマ環境においても有効であると考えられる。

5. 工学的な MAS への適用

これまで、多くの研究が動的環境に適応し組織化する MAS の実現を目指してきた^{(4),(16),(18),(19)}。それらは、エージェントそのものに、組織化やそれによりシステム最適性を高めるメカニズムを組み込もうとしたものが多い。DS-GA では、<性質 1: 利得の最大化>、<性質 2: 動的離隔探索>、<性質 3: 利得の均衡化>、<性質 4: 高収束性>、<性質 5: 高脱出性>などの高いシステム最適性を得やすい性質を、DS-GA そのものが持つ。そのため、エージェントそのものの設計は比較的簡単である。本章では、システムの大域的目的をエージェントの利得に設定する MAS に DS-GA を適用し、エージェントが協調作業と分業により組織化され、システムの目的達成能力を高めた例として、自律ロボットによる生産プロセスを模擬する 3 材料採取・生成 MAS について述べる。

本章では、5.1 節で 3 材料採取・生成 MAS について述べる。5.2 節でジレンマ環境に適応する協調作業や分業による組織化について述べる。5.3 節で動的環境に適応する組織の変革について述べる。

5.1 3 材料採取・生成 MAS

本章では、DS-GA の有効性を検証する MAS の実験モデルとして、2.2 節の学習アルゴリズム (2) 行動の部分に、3 材料採取・生成 MAS を適用する。この MAS は、より少ないコストで材料 A, B, C を採取し、採取した 3 材料からより多くの生成品 P を生成

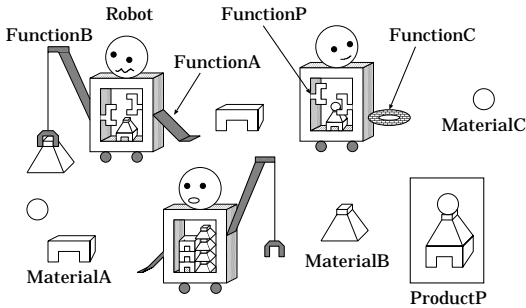


図 15 3 材料採取・生成 MAS の概念

Fig. 15 MAS of three material extraction and one product fabrication (3M-1P MAS).

することをシステムの目的とする．3 材料採取・生成 MAS の概念を図 15 に示す．

ロボット a が時刻 t に保持する材料 A, B, C と生成品 P の 4 要素の量を、それぞれ $E_A(a, t), E_B(a, t), E_C(a, t), E_P(a, t)$ で表す．(2) 行動では、各ロボットは、4 要素のそれぞれの量を操作する単位時間ごとの (2-a) 仕事 (2-b) 贈与 (2-c) 消費を行う．また、(3) 分裂と消滅では、生成品 P を資産として用いる．

(2-a) 仕事 (採取, 生成)

ロボット a は、フィールドから無尽蔵に存在する材料 (A, B, C) のいずれかを採取する機能と、3つの材料から P を生成する機能の 4 つの独立な機能をそれぞれ持ちうる．それぞれの機能を使用してフィールドから単位時間ごとに採取できる量と、3 材料から P を生成できる量を $S_A(a), S_B(a), S_C(a), S_P(a)$ で示し、値が 0 のときは、その機能を持たないものとする．時刻 t におけるロボット a の仕事 $Act(a, t) \in \{A, B, C, P\}$ は、ロボット a が持つ機能の中から単位時間ごとに 1 つの機能が無作為に選択、実行され、いずれかの材料の採取の場合は式 (1) に、 P の生成の場合は式 (2) に従い保持する要素量が変化する (図 16)．また、すべての機能を持たないロボットは、どの仕事もしない．

- 材料の採取 (if $Act(a, t) \in \{A, B, C\}$)

$$E_{Act(a, t)}(a, t) = E_{Act(a, t)}(a, t - 1) + S_{Act(a, t)}(a) \quad (1)$$

- P の生成 (if $Act(a, t) = P$) $x \in \{A, B, C\}, \forall x$

$$E_x(a, t) = E_x(a, t - 1) - S_P(a) \\ E_P(a, t) = E_P(a, t - 1) + S_P(a) \quad (2)$$

ただし、材料を保持する量 $E_A(a, t), E_B(a, t), E_C(a, t)$ の 1 つでも S_P より少ないときに P の生成が選択される場合を“空振り”と呼び、保持する各要素 A, B, C, P の量は変化しない．

(2-b) 贈与

ロボット a は、同一コロニー内で無作為に選択する

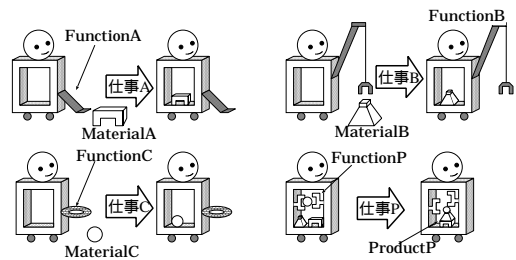


図 16 仕事の概念

Fig. 16 Schematic explanation on works.

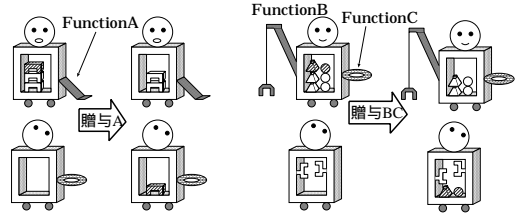


図 17 贈与の概念

Fig. 17 Schematic explanation on donations.

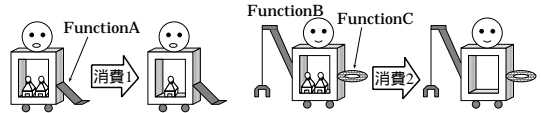


図 18 消費の概念

Fig. 18 Schematic explanation on consumption.

ロボット k に、自らの保持する各要素 A, B, C, P のそれぞれ一定量を同時に一方的に贈与することができる (図 17)．一方的とは、 a が k に贈与する場合に、 k はさらに無作為に選択する l へ贈与することを示す．ただし、ある要素 x について採取する機能や生成する機能がないロボットは、その要素 x を贈与しない．それぞれの要素における単位時間ごとの贈与量を $D_A(a), D_B(a), D_C(a), D_P(a)$ で示す．また、無作為に選択されるロボット k への贈与による、各要素 A, B, C, P の量の変化を式 (4) に示す．

$$x \in \{A, B, C, P\}, \forall x \\ E_x(a, t) = E_x(a, t - 1) - D_x(a) \\ E_x(k, t) = E_x(k, t - 1) + D_x(a) \quad (3) \\ (if E_x(a, t - 1) > D_x)$$

(2-c) 消費

採取・生成機能維持のために消費する機能維持コストとして、維持する機能の数 $\times Cost$ を単位時間ごとに E_P から消費する (図 18, 式 (4))．ここで、 $Cost$ は 1 つの機能の維持に必要なコストを表す．

表 8 3 材料採取・生成 MAS における遺伝子
Table 8 Action-prescribing gene in 3M-1P MAS.

ラベル	内容	能力値, 贈与量
0	能力なし, 贈与なし	$S_x=0, D_x=0$
1	能力あり, 贈与なし	$S_x=1, D_x=0$
2	能力あり, 贈与あり	$S_x=1, D_x \{0.5, 0.2\}$

表 9 3 材料採取・生成 MAS における戦略
Table 9 Strategies-prescribing gene in 3M-1P MAS.

	1ビット目	2ビット目	3ビット目	4ビット目
要素名	材料A	材料B	材料C	生成品P
行動決定遺伝子	0 or 1 or 2	0 or 1 or 2	0 or 1 or 2	0 or 1 or 2

$$x \in \{A, B, C, P\}, \forall x$$

$$E_P(a, t) = E_P(a, t - 1) - \sum_x Cost \times \begin{cases} 0 (S_x(a) = 0) \\ 1 (S_x(a) \neq 0) \end{cases} \quad (4)$$

(3) 分裂と消滅

ロボット a を分裂, 消滅させる資産として, 生成品 P の量 $E_P(a, t)$ を用いる. 各ロボットは, $E_P(a, t)$ が初期値の倍以上になると 2 つに分裂し, 0 以下になると消滅する. 分裂の際, 生成品 P と 3 材料 A, B, C の 4 要素の量は半分ずつになり, 遺伝子は引き継ぐ. このとき, 要素ごとに突然変異確率 P_{mut} で遺伝子が変異する.

本実験では, ロボット a の機能と行動を決定する要因は, 単位時間あたりの採取量 $S_A(a), S_B(a), S_C(a)$, 生成量 $S_P(a)$, 贈与量 $D_A(a), D_B(a), D_C(a), D_P(a)$ の 8 つである. 実験結果の検証を容易にするため, 採取量, 生成量, 贈与量をいずれも 2 種類とし, それぞれ, $S_A(a), S_B(a), S_C(a), S_P(a) \in \{0, 1\}, D_A(a), D_B(a), D_C(a) \in \{0, 0.5\}, D_P(a) \in \{0, 0.2\}$ とする. 値 0 は, 機能を持たない, または, 贈与しないことを意味する. また, 各個体の遺伝子は, (2-b) 贈与で述べたように, 採取や生成機能を持たない要素を贈与する個体は存在しないため, [機能無・贈与有] は削除し, 要素ごとに [機能無・贈与無]:0, [機能有・贈与無]:1, [機能有・贈与有]:2 とラベルづけする遺伝子 $Gene_{Act}(a) \in \{0, 1, 2\}$ とする (表 8).

また, 要素ごとの遺伝子 $Gene_{Act}(a) \in \{0, 1, 2\}$ を $[0, 1, 2, 0]$ のように 3^4 通りで表記し, ロボットの戦略とする (表 9).

この MAS の環境は, 各機能を用いる 4 種類の仕事の分業により機能維持コストをおさえ, 互いの贈与により要素の不足を補うことで高いシステム最適性

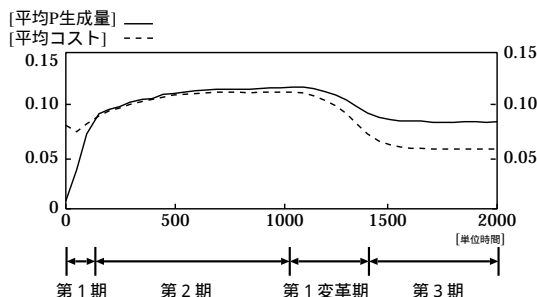


図 19 2,000 単位時間までの平均 P 生成量と平均コストの推移
Fig. 19 2,000 unit times of history of the average product amount and the average cost.

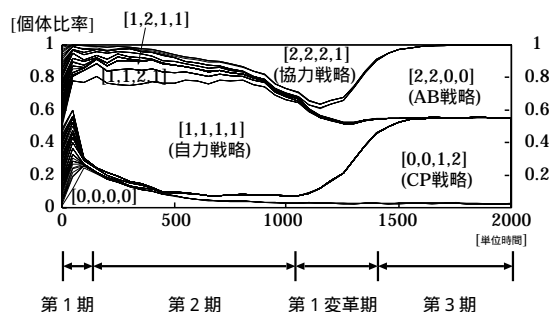


図 20 2,000 単位時間までの戦略別個体数比率の推移
Fig. 20 2,000 unit times of history of the population ratios among various strategic agents.

を得るが, 贈与しない個体がより高い個体最適性を得るジレンマ環境である. また, 5.3 節で述べるように, 2001 単位時間以降, システム環境が変化する動的環境である. ここでは, $N_a(0) = 100,000, P_{mut} = 0.01, N_{Lim} = 10, P_{mig} = 0.001, Cost = 0.03, \forall a, \forall x, E_x(a, 0) = 1$ とし, 異なるランダム系列を用いて 100 試行する.

5.2 実験結果および考察

本節の実験では, 100 回の試行において後述するいくつかの様相が現れた. ここでは, その中で代表的な様相の 1 つについて考察する. 平均 P 生成量と平均コストを図 19 に, 戦略別個体数比率を図 20 に示す. 考察の目安として, P 獲得量 (= P 生成量 - コスト - P 贈与量 + 被 P 贈与量) の平均 (= 平均 P 生成量 - 平均コスト) が正になるまでを第 1 期, 個体数比率の様相が大きく変わる時期を第 n 変革期, その他の期間を第 n 期と呼び, 区別する.

5.2.1 第 1 期, 第 2 期

第 1 期には, 単独で P を生成する場合に最も P 獲得量が多い自力戦略 $[1, 1, 1, 1]$ が多数を占めた. 第 2 期後半には, 自力戦略のほかに協力戦略 $[2, 2, 2, 1]$ が増加した. これは, 自力戦略や協力戦略で占められたコロ

ニーでは、生成には“空振り”の可能性があるが採取にはないため、利用されない材料の余りが生じ、蓄積される。協力戦略は、材料の余りを互いに贈与し合うことで“空振り”の確率が低く、P生成量が高い。このため<性質1：利得の最大化>により増加したと考えられる。ただし、協力戦略が増えるためには、コロニー内が多数の協力戦略で占められる必要がある。初期においては、協力戦略が多数を占めるコロニーがなく、協力戦略の増加率は自力戦略よりも低かった。また戦略 $[0,0,1,2]$ (以下、CP戦略と記す)は、協力個体で占められたコロニーに少ない比率で共存し、“空振り”による材料の余りを利用してPを生成する。<性質3：利得の均衡化>によりこのPを他のロボットに贈与し、システム最適性を高めていると考えられる。

5.2.2 第1変革期、第3期

第1変革期に、戦略 $[2,2,0,0]$ (以下、AB戦略と記す)とCP戦略が急激に増加し、P獲得量が拡大した。第3期には、この2つの戦略がほぼ半々の比率を維持しながら多数を占めて共存した。

これらの2戦略は、自力戦略や協調戦略に比べて機能維持コストが半分である。AB戦略は2種類の材料A,BをCP戦略に贈与し、CP戦略は生成品PをAB戦略に贈与するという協調の結果、共存したと考えられる。実際には、各ロボットが贈与する相手は無作為なので、AB戦略が材料A,BをCP戦略に贈与するとは限らない。しかし、AB戦略とCP戦略のみが存在するコロニー内の材料A,Bについて考えると、AB戦略がAB戦略に贈与したとしても、贈与されたAB戦略も同様に贈与するため、CP戦略に贈与されるまでコロニー内を循環する。同様に生成品Pについても、AB戦略に贈与されるまで循環する。これは、材料A,Bと生成品Pの3要素の循環メカニズムであるといえる。<性質5：高脱出性>によりAB戦略とCP戦略が多数を占めるコロニーが現れ、<性質2：動的離隔探索>により他の戦略を含まず、高いシステム最適性を得る比率で、<性質4：高収束性>によりMAS全体に広まったと考えられる。

複数の自律ロボットからなるこれらの循環メカニズムを組織の形態の1つと考え、協調作業と分業による組織化が実現できたといえる。なお、本節の実験では、100試行中、AB戦略とCP戦略への分業が35回、AC戦略とBP戦略への分業が34回、BC戦略とAP戦略への分業が30回現れ、ABC戦略 $[2,2,2,0]$ とP戦略 $[0,0,0,2]$ への分業が1回現れた。しかし、すべてを3,000単位時間までシミュレートすると、ABC戦略 $[2,2,2,0]$ とP戦略 $[0,0,0,2]$ への分業も、AC戦

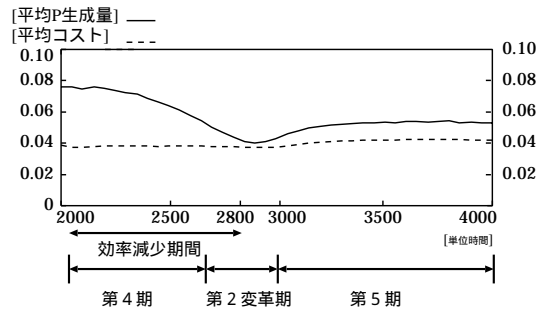


図21 2,000単位時間以降の平均P生成量と平均コストの推移
Fig. 21 History of the average product amount and the average cost after the 2,000 unit time.

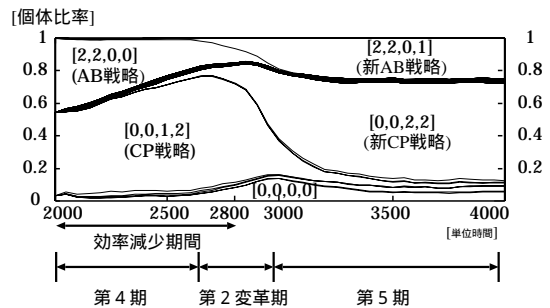


図22 2,000単位時間以降の戦略別個体数比率の推移
Fig. 22 History of the population ratios among various strategic agents after the 2,000 unit time.

略とBP戦略の分業になり、すべての結果がAB戦略とCP戦略、AC戦略とBP戦略、BC戦略とAP戦略の分業になった。これら3つの結果は、数学的にはまったく同じで、100回の試行すべてにおいて同じ結果が得られたといえる。

5.3 動的環境への適応

本節では、環境の動的変化に対する組織の変化をみるため、機能維持コストの減少とPの生成効率、贈与効率の減少を考える。

100試行のうち、前節で考察した実験に関して、それぞれの機能維持コスト $Cost$ を2,001単位時間において、0.03から0.02に削減する。また、Pの生成量 S_P とPの贈与量 D_P を、2,001単位時間から2,800単位時間において単位時間ごとに0.1%、計80%削減する。これにより、 S_P は1から0.2に、 D_P は0.2から0.04になる。これは、単位時間ごとのコストが $\frac{2}{3}$ に、Pの生成効率や贈与効率が $\frac{1}{5}$ になることを意味する。2,000単位時間以降の平均P生成量と平均コストの推移を図21に、戦略別個体数比率の推移を図22に示す。

5.3.1 第4期

第4期では、生成機能を持つ全ロボットの単位時間

ごとの生成量 S_P が徐々に低下する。このため、多くの材料が余り、生成機能に対する需要が増加する。そこで、〈性質 2：動的隔離探索〉などにより、P を生成する CP 戦略 $[0,0,1,2]$ の比率が徐々に増加し、P 生成量の減少を少しでも防ごうとする組合せとなった。これは、第 3 期で得られた循環メカニズムを維持した適応である。また、図 22 において 2001 ~ 約 2600 単位時間の間、 S_P の直線的な低下という環境の変化に対して、CP 戦略の比率は直線的に増加している。これは、〈性質 4：高収束性〉により、MAS が環境の変化に素早く適応しているためと考えられる。

第 4 期の後半には、P 生成効率の悪化により、CP 戦略が保持する材料 C が大量に余ったため、戦略 $[0,0,2,2]$ (以下、新 CP 戦略と記す) へと突然変異を起こした新 CP 戦略がシステム最適性を低下させず、少ない比率で生存している。

5.3.2 第 2 変革期、第 5 期

P 生成の需要が全体的にますます増加した第 2 変革期において、〈性質 5：高脱出性〉により、AB 戦略から突然変異を起こした戦略 $[2,2,0,1]$ (以下、新 AB 戦略と記す) と、先に突然変異した新 CP 戦略が急激に増加する。これにより、新 AB 戦略が材料 A, B を、新 CP 戦略が材料 C と生成品 P をともに贈与する 4 要素循環メカニズムが現れ、贈与された新 AB 戦略も P を生成することで、この環境における全体の P 生成能力が 3 要素を循環する戦略の組合せより向上した。

第 3 期では、材料 A, B と生成品 P の 3 要素循環メカニズムであったが、第 5 期では環境の変化に適応し、材料 C を加えた 4 要素循環メカニズムへと組織を変革したといえる。また、P の生成機能を多くのロボットが持つことで、Cost の減少と P 生成効率、P 贈与効率の悪化という新たな環境における P 獲得量を、3 要素循環メカニズムより高めることができた。これらの結果から、〈性質 1：利得の最大化〉、〈性質 2：動的隔離探索〉、〈性質 3：利得の均衡化〉、〈性質 4：高収束性〉、〈性質 5：高脱出性〉を持つ DS-GA を MAS に適用することで、自律的なエージェントにより動的環境に適応した協調作業や分業による組織化が実現できる可能性がうかがえる。

ここで、採取・生成・贈与をしないが他のロボットから P の贈与を受けて生き延びるという他力戦略 $[0,0,0,0]$ に注目すると、第 3 期では平均約 2.2% であった個体数比率が、第 5 期には平均約 5.3% まで増加した。〈性質 1：利得の最大化〉〈性質 2：動的隔離探索〉により、他力戦略の比率は大きくは増加しないが、協力関係がより進むことによる依存関係の発達が示唆

される。これは、生産性が向上した文明社会において、生産に直接関係しない人が多数存在することに対応しているとも考えられ、興味深い。

6. おわりに

本論文では、エージェントの行動戦略学習により MAS 自体に学習能力を持たせることで、未知の環境や動的な環境への素早い適応や、協調作業や分業による組織化を実現し、システム全体の目的達成能力の向上を実現する手法の 1 つとして、動的隔離型 GA (DS-GA) を提案した。

1 章で、本研究の位置付けを述べた。2 章で、隔離状態が個体数に応じて動的に変化する動的隔離型 GA (DS-GA) を提案した。3 章で、ジレンマ環境である MAS への適用を考え、〈性質 1：利得の最大化〉と〈性質 2：動的隔離探索〉によるシステム最適性の高い行動の獲得や、〈性質 3：利得の均衡化〉による利得割当ての自律的獲得について述べた。システムの目的がエージェントの利得の総和になるように MAS を設計し、DS-GA を適用することで、ジレンマ環境において、システムの目的達成能力が高い協調行動の獲得や、分業と利得割当ての自律的獲得が期待される。4 章で、動的環境である MAS への適用を考え、〈性質 4：高収束性〉による収束の早さと、〈性質 5：高脱出性〉による局所解から脱出する確率の高さについて述べた。動的環境において、解への素早い収束と、局所解からの高い脱出確率が期待される。5 章で、上記の性質が有効に機能し、動的環境に適応した分業や協調作業による組織化を実現した例として、3 材料採取・生成 MAS について述べた。エージェントレベルの学習から、複数のエージェントからなる循環メカニズムが現れたことから、ジレンマ環境と動的環境の複合環境においても、上記の性質のそれぞれが有効に機能し、システム最適性の高い協調作業や分業による動的な組織化が実現できる可能性が示唆される。

DS-GA は、sGA と比較すると計算時間が 5 割ほど長いですが、アルゴリズムは比較的簡単であり、エージェントの設計も簡単になると考えられる。今後は、この利点を生かし、さまざまな MAS に適用することを検討している。

参考文献

- 1) 石田 亨, 桑原和宏: 分散人工知能 (1): 協調問題解決, 人工知能学会誌, Vol.7, No.6, pp.945-954 (1992).
- 2) 村田哲哉, 鈴木恵二, 大内 東: GA によるサツ

- カーエージェントの動的配置探索問題に関する研究, 人工知能学会誌, Vol.14, No.3, pp.446-454 (1999).
- 3) 平山勝敏, 山田誠二, 豊田順一: 山登り法を用いた分散制約充足における組織化, 人工知能学会誌, Vol.10, No.1, pp.80-87 (1995).
 - 4) 松浦賢一, 嘉数侑昇: 非均質エージェント系における協調的行動の形成: 2次元追跡問題における考察, 情報処理学会論文誌, Vol.38, No.6, pp.1083-1093 (1997).
 - 5) Cohoon, J.P., Martin, W.N. and Richards, D.S.: A Multi-population Genetic Algorithm for Solving the K-Partition Problem on Hypercubes, *Proc. 4th ICGA*, pp.244-248 (1991).
 - 6) 鈴木光男: ゲーム理論入門, 共立出版 (1981).
 - 7) 伊藤 昭, 矢野博之: 取引履歴公開下での最適取引戦略—自律的エージェント社会の行動規範, 人工知能学会誌, Vol.10, No.2, pp.271-278 (1995).
 - 8) 伊藤 昭, 矢野博之: 利己的なエージェントの社会におけるつきあい方戦略の進化, 情報処理学会論文誌, Vol.38, No.5, pp.944-952 (1997).
 - 9) 宮崎和光, 荒井幸代, 小林重信: Profit Sharingを用いたマルチエージェント強化学習における報酬配分の理論的考察, 人工知能学会誌, Vol.14, No.6, pp.1156-1164 (1999).
 - 10) 松原繁夫, 横尾 真: 繰り返しゲームにおいて協調行動を生成する先読み型行動選択方法, 人工知能学会誌, Vol.12, No.6, pp.881-890 (1997).
 - 11) 沼岡千里, 大沢英一, 長尾 確: マルチエージェントシステム, 共立出版 (1998).
 - 12) 千葉一博, 平石邦彦: 繰り返し連続化囚人のジレンマゲームの提案, 人工知能学会誌, Vol.13, No.4, pp.560-569 (1998).
 - 13) 千葉一博, 平石邦彦: エージェント侵略の過程と中間的な手をとる戦略の合理性について, 人工知能学会誌, Vol.14, No.4, pp.657-666 (1999).
 - 14) 西岡靖之, 藤田敏之: 利得情報制御による自律分散型エージェントの協調関係の実現, 人工知能学会誌, Vol.15, No.6, pp.1043-1051 (2000).
 - 15) Mor, Y. and Rosenschein, J.S.: Explanation-Based Feature Subset Selection for Decision Tree Induction, *Proc. 1st Int. Conf. on Multiagent Systems*, pp.276-282 (1995).
 - 16) 鈴木恵二, 吉村 潤, 嘉数侑昇: 疲労度パラメータを導入した行動選択ネットワークによるエージェントの創発的組織化に関する考察, 人工知能学会誌, Vol.12, No.1, pp.152-159 (1997).
 - 17) 和氣弘明, 村山隆彦, 服部文夫: ネットワーク資源割当問題向き協調問題解決, 情報処理学会論文誌, Vol.37, No.3, pp.324-332 (1996).
 - 18) 横尾 真: 柔軟で動的なエージェントの組織構造を用いた分散制約充足アルゴリズム, 人工知能学会誌, Vol.11, No.6, pp.933-940 (1996).
 - 19) 太田 順, 横川洋一, 新井民夫: 複数自律ロボット

系における漸進的戦略形成に関する研究, ロボット学会誌, Vol.14, No.3, pp.379-385 (1996).

付 録

A.1 最大のシステム最適性を得る最適個体数比率
戦略 C と戦略 D からなるグループにおいて, 戦略 C 個体の個体数比率を r , 戦略 D が戦略 C との対戦により得る利得を p と置く. 対戦の組合せごとの利得の総和と確率は,

$$(1) C, C: \text{互いの利得の総和 } 0, \text{ 確率 } r^2$$

$$(2) C, D: \text{互いの利得の総和 } p, \text{ 確率 } 2r(1-r)$$

$$(3) D, D: \text{互いの利得の総和は } 2, \text{ 確率 } (1-r)^2$$

であるから, 対戦した 2 個体の利得の総和の期待値 P は下式になる.

$$\begin{aligned} P &= p \times 2r(1-r) + 2 \times (1-r)^2 \\ &= 2(1-p)r^2 + 2(p-2)r + 2 \end{aligned} \quad (5)$$

式 (5) より, P の最大値 P_{max} と, そのときの個体数比率 $r_{P_{max}}$ を求めると, それぞれ下式になる.

$$P_{max} = \frac{p^2}{2(p-1)} \quad (6)$$

$$r_{P_{max}} = \frac{(p-2)}{2(p-1)} \quad (7)$$

A.2 システム最適性を維持する最適贈与量

戦略 C, D の比率を維持する贈与量は, 個体数の増加期待値が戦略 C と戦略 D の間で等しくなり, 最適個体数比率を維持する量である. 戦略 C 個体の贈与量を M_c , 戦略 D 個体の贈与量を M_d とすると, 贈与を含めた戦略 C 個体の利得の期待値 P_c , および戦略 D 個体の利得の期待値 P_d は, それぞれ下式になる.

$$\begin{aligned} P_c &= (0 + M_c - M_c) \times r \\ &\quad + (0 + M_d - M_c) \times (1-r) \\ &= (M_d - M_c)(1-r) \end{aligned} \quad (8)$$

$$\begin{aligned} P_d &= (p + M_c - M_d) \times r \\ &\quad + (1 + M_d - M_d) \times (1-r) \\ &= (M_c - M_d + p - 1)r + 1 \end{aligned} \quad (9)$$

$P_c = P_d$ と式 (9), 式 (10) より, 各戦略の最適贈与量は下式になる.

$$M_d - M_c = (p-1)r + 1 \quad (10)$$

よって, 各戦略の贈与量の差 $M_d - M_c$ が個体数比率 r に依存する. また, 最適個体数比率 $r_{P_{max}}$ のときの $M_d - M_c$ は, 式 (10) に式 (7) を代入し下式になる.

$$M_d - M_c = \frac{p}{2} \quad (11)$$

A.3 最適同一行動への脱出に必要な個体数比率
最適解で得る利得を p , 局所解で得る利得を 1 とす

る．最適同一行動（以下，最適解と記す）を選択する個体数比率を r ，最適解ではない同一行動（以下局所解と記す）の個体数比率を $(1-r)$ で表す．

最適解個体を得る利得の期待値 P_g が，局所解個体を得る利得の期待値 P_l を上回る最適解個体数比率 r_a は， $P_g > P_l$ より下式になる．

$$p \times r_a > 1 \times (1 - r_a) \\ r_a > \frac{1}{p+1} \quad (12)$$

このとき，個体間において最適解に向かうグループ内学習が始まる．

最適解個体を含む個体グループが得る利得の総和の期待値 P_g が，全個体が局所解個体であるグループの利得の総和の期待値 P_l を上回る最適解個体数比率 r_g は，グループ内個体数を N とすると， $P_g > P_l$ および $p > 0$ より下式になる．

$$\{p \times r_g^2 + 1 \times (1 - r_g)^2\} \times N > 1 \times N \\ r_g > \frac{2}{p+1} \quad (13)$$

このとき，グループ間において最適解に向かうグループ間学習が始まる．

(平成 14 年 2 月 4 日受付)

(平成 14 年 4 月 24 日再受付)

(平成 14 年 5 月 15 日採録)



中山 功一

2000 年三重大学工学部機械工学科卒業．2002 年同大学大学院工学研究科博士前期課程修了．2002 年 4 月より京都大学大学院情報学研究所博士後期課程および ATR 人間情報科学研究所にてマルチエージェントによる知的システムの研究に従事．人工知能学会等会員．



松井 博和

1992 年名古屋工業大学工学部電気情報工学科卒業．1994 年名古屋大学大学院工学研究科博士前期課程情報工学修了．1997 年神鋼電機株式会社入社．1998 年三重大学工学部助手，現在に至る．2000 年名古屋大学大学院工学研究科博士後期課程電子機械工学修了．工学博士．



野村由司彦

1976 年名古屋大学工学部機械工学科卒業．1978 年同大学大学院工学研究科修了．同年 NTT 電気通信研究所入所．1990 年名古屋大学工学部助教授．1997 年三重大学工学部教授，現在に至る．工学博士（東工大）．日本機械学会，日本ロボット学会，IEEE 等各会員．土木学会論文奨励賞．電子通信学会学術奨励賞受賞．