

行動選択ネットワークに基づいた組織的行動の生成に関する研究

吉 村 潤^{†,☆} 鈴 木 恵 二[†] 嘉 数 侑 昇[†]

ここでは、マルチエージェント系の環境における組織的行動の獲得を目標とし、自律的行動選択機構としてダイナミクス機構を導入した行動選択ネットワークを提案する。また、提案法の適用問題として、マルチエージェントによるモンキーバナナ問題を取り上げる。この問題において、個々のエージェントは他のエージェントと協調しながら目標が達成されるような行動を選択し、さらに、組織化の生成過程において生じるデッドロックを自律的に克服しなければならない。すなわち、行動選択ネットワークによる行動戦略に従いながら、観測情報を通じてエージェントどうしの結合・分離といった組織化を通じタスク処理を行わせることを目標とする。本研究では行動選択ネットワークの基本的行動戦略の構築と、組織的行動の実現についての考察を行う。

A Study on Action Strategy for Emergent Organizations Based on Action Selection Network

HIROSHI YOSHIMURA,^{†,☆} KEIJI SUZUKI[†] and YUKINORI KAKAZU[†]

The purpose of this paper is to acquire the action strategy for emergent organizations in multi-agents environment. For this purpose, we propose the action selection network with dynamics mechanism which involves action strategies. Also, as an application of our proposal method, we take up Monkey Banana Problem with multi-agents. In this problem, individual agent must take action to accomplish his aim cooperating with other agents. In addition, the agent must autonomously conquer dead-rocks that are produced in the generation processes of the emergent organizations. Namely, it is our aim to carry out the task management through the emergent organizations which means agents' combinations and separations. First we construct standard action strategies based on action selection network, and discuss about the emergent organizations.

1. はじめに

本論文は行動選択ネットワークに基づくエージェント群の組織的行動の生成を行うことを目的とする。特に複数の目標を与えたリアクティブプランニング問題への適用を通して、提案するネットワーク構造の持つ性質が自律的な組織化能力をどのように変化させるか考察を行う。

ノードを行動モジュールとし複数のノード間によるリンク上での活性伝播を用いた行動選択ネットワーク^{1)~5)}による手法は、自律エージェントが複雑で予測できない環境に対して行動選択を行う手法の1つである。この行動選択アルゴリズムは環境の変化に対応した行動の選択基準をエージェント内部で決定できるこ

とから、さまざまな行動パターンを獲得することができ多目的な行動が要求される環境での行動選択手法として有効であると思われる。近年ではこのアルゴリズムを用いた自律移動ロボットの研究も盛んである^{6)~8)}。

エージェントの行動形成に関して行動選択ネットワークを用いた場合の利点として次の点が考えられる。

(1) 行動の多様性

環境状況によってさまざまに変化する行動のレパートリーをノードに分散することによって包含することができる。強化学習等⁹⁾における一様な行動群の取扱いでは、環境との相互作用で変化する行動パターンを獲得することは困難となろう。

(2) 適応性の分散

行動のレパートリーをノードに分けることができることから、適応性の導入においても分散して行うことができる（モチベーションの導入）。

(3) タスク処理の多様性

行動の多様性が導入可能であるのと同様に、目

† 北海道大学工学部複雑系工学講座

Complex Systems Engineering, Autonomous Systems Engineering, Hokkaido University

☆ 現在、富士通株式会社

Presently with Fujitsu Limited

的とするタスク処理もまたノードレベルの分散性によって導入可能である。

一方、自律システムの集合として定義されるマルチエージェントシステムにおいて、与えられたタスクを協調分散的に処理する方法が考えられ^{10),11)}、特に強化学習による研究が実現されている^{12),13)}。しかし、学習によるエージェントの協調行動の実現に関しては、環境内の目標は一意に設定されていることが多い、複数の目標が存在する場合における自律エージェント群の柔軟な協調行動の実現は困難であると考えられる。

これに対し、行動選択ネットワークによるアプローチでは多数の処理目標を持つ環境に対しては、ネットワークの構造自体に複数の目標を容易に設定することができる特徴を持つ。さらに、その目標の選択基準をエージェントが独立して決定できることから、複数の処理目標が設定された環境におけるエージェントの行動実現手法として有効であると考えられる。しかし、複数ある処理目標が協調的な行動やある種の組織化を必要とする場合において、各エージェントが独立に個々の目標を決定しうる状況から、いかにして協調的な行動を自律的に引き出すかが課題となる。本論文はこのような課題に興味を持ち、各エージェントが個々の目標を独立に決定しながらも、エージェント間および環境との相互作用を通じて、創発的に組織的な行動の生成を行わせるための行動選択ネットワークの構築を目的とする。

よって、ここでは集団形成や組織化によるタスク処理のための行動発現機構の構築に焦点を置き、ダイナミクスを導入した行動選択ネットワークを提案する。そして、この行動選択ネットワークアルゴリズムを用いた自律エージェントの基本的振舞いを示す。また、エージェントどうしの結合や分離といった組織化行動を必要とするタスク処理への適用を行う。具体的なタスク処理としてデッドロックを含んだ問題であるマルチエージェントによるモンキーバナナ問題を用い、行動選択ネットワークの特性と組織的な行動生成能力に関して考察を行う。

以降では、まずダイナミクスを導入した行動選択ネットワークを提案し、次にこのネットワークによる基本的行動戦略に関して簡単な計算機実験を通じた説明を行う。そして、マルチエージェントによるモンキーバナナ問題への適用例を通じて、本提案手法による組織的行動実現の展開を図る。

2. 行動選択ネットワーク

Brooks¹⁴⁾や Maes^{1)~4)}によって提案された Behav-

ior Network Architecture (以下、BNと略す) は従来の伝統的 AI としてのトップダウンアプローチにかかるボトムアップ的アプローチによる行動選択手法である。BN はノイズがあり予測が不可能な環境に対して設計されたものであり、より複雑で特殊な問題領域を扱うことができるアルゴリズムである。

BN は行動を表すノードによって構成される。生物学的には、これらのノードはリズム発生機として働く回路と考えられている。ノードは互いの間で活性値の伝播を行う。この活性値は環境および状態やノード自身の状態によって操作されるリンク上を伝播する。

本研究ではノードのリンク上で活性値伝播に対して次節に述べるような戦略性を導入した伝播方法を用いる。すなわち、本提案手法は Blumberg¹⁵⁾による階層構造化された BN の手法を参考にしているが、上位層から下位層への明示的かつ戦略的な活性伝播制御を取り入れるためにアクション実行権の考え方を導入するとともに、その制御戦略が柔軟に切り替わるように、下層からのフィードバック機構を新たに導入することによって行動選択の柔軟性向上を図っている点に特徴がある。この機構によって、上位層で表される目標に対し、あらかじめ有効な行動が分かっていない場合や、目標選択の適切さが不明確な場合でも、自律的に適切な行動および目標を探し出すことが期待できると考えられる。

2.1 ノードの基本的結合構成

ノードの構成概念を図 1 に示す。エージェントは各ノードのリンク上で活性値伝播（入力と出力）を行うことによって、どのノードが活性化しているかを決定し行動を選択する。

短期記憶的な行動戦略の実現と自律的なデッドロック回避能力の実現のために、ここでは時間変化するダイナミクスを用いる。

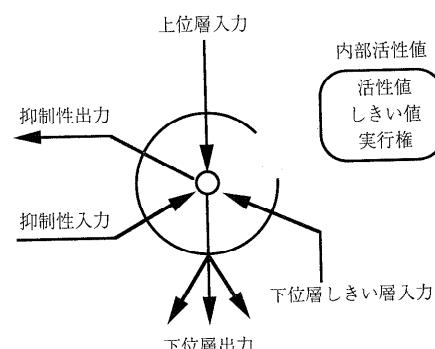


図 1 ノードの構成概念

Fig. 1 Framework of node.

ここで、各ノードの状態および伝播する値（内部活性値）を次のようにする。なお、添え字 i, j はノード番号を表すものとする。

- アクション (Action: Act_i)
- アクション実行権 (Priority: P_i)
- しきい値 (Threshold: Th_i)
- 活性値 (Activation: A_i)
- (興奮性) 入力 (Input: In_{ij})
- 出力 (重み) (Output: W_{ij})
- 抑制性入力 (Inhibit: Ih_i)
- 抑制性出力 (重み) (Inhibit Output: Wh_{ij})

また、興奮性入力は上位ノードあるいは外部からの入力値、抑制性入力は同じ層にある他のノードへの入出力値を意味する。詳しくは以下に説明を行う。

2.2 ノードの基本ダイナミクスとその戦略

行動を選択する際、ここでは各ノード内において戦略性を明確に導入した制御を行う。つまり、ノードの活性値をただ出力するだけでなく、リンクするノードに対する自分との優勢度（アクション実行権、Priority）に依存した重みを出力し、かつ他のノードのアクション実行権の有無によって出力方法を変える方式をとるものとする。このような戦略性を持った制御により、行動の選択が環境からの情報のみに依存した行動ではなく、実際にエージェントが行動を起こすことによって得られる評価、つまりその行動が成功か失敗かを判断しながら行動を選択することができる。すなわち単なる反射行動にとどまらない行動の多様性を実現することが期待できる。

ノードは入力に対して出力結合、または抑制性結合を持つノードへ次のような基本ダイナミクスによって出力値（出力やしきい値）を伝播する。

$$\begin{aligned} f(X_t^{input}, Z_{inhibit}) &= X_{t-1} \\ &= (1 - fatigue)X_{t-1}^{input} + X_t^{input} - Z_{inhibit}. \end{aligned} \quad (1)$$

$$G(X_t^{input}) = Clamp[H(X_t^{input}), 1.0, 0.0]. \quad (2)$$

$$H(x) = 1.0 / \{1.0 + \exp(-(x - 5.0))\}. \quad (3)$$

$Clamp[value, max, min]$ は $value$ を最大 max 、最小 min に領域を限定する。 $fatigue$ はエージェントの疲労度を示す。これら 3 式は各ノードの活性値、しきい値、出力、抑制性出力の計算に用いる。

式 (1) において導入した $fatigue$ によって行動選択の性質を変えることができる。この $fatigue$ パラメータを各入出力において用いることにより、ノードの内部活性値の維持が行われる。すなわち、各ステップごとの入力に対し、過去の入力を割り引くことに

よりダイナミクスが疑似的に起こる。そして結果的にエージェントの行動の継続性が引き起こされる。このパラメータの設定によってエージェント群の組織的行動の性質がいかに変わるかについて後の実験で検証を行う。

以下、各出力の計算とそのときの戦略（出力値などの制御）について説明する。

2.2.1 しきい値と Priority の決定

Priority はしきい値を基準に同じ層にあるノード間での優先順位を決定するのに用いる。なお、しきい値 Th_i は下位層からのしきい値に従う点は注意を要する。

$$Th_i = Clamp \left[\frac{1}{k} \sum_j^k Th_j^{out}, 1.0, 0.1 \right]. \quad (4)$$

$$P_i = \begin{cases} 1, & \text{if } (A_i > Th_i) \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

Th_j^{out} : 出力結合先のノードで、かつその Priority が 1 であるノードの出力しきい値

2.2.2 活性値の更新

ノードは上位層からの出力値と下位層からのしきい値の入力によってそのノード自身の活性度を表す活性値 A_i を計算する。

$$A_i = f(In_i^{in}, Ih_i^{in}). \quad (6)$$

In_i^{in} : 内部興奮性入力

Ih_i^{in} : 内部抑制性入力

$$In_i = \sum_h W_{hi} \times (A_h - Th_h). \quad (7)$$

W_{hi} : ノード h からノード i への興奮性結合重み

$$Ih_i = \sum_q W_{qi} \times (A_q - Th_q). \quad (8)$$

W_{qi} : ノード q からノード i への抑制性結合重み

内部興奮性入力は上位層からの入力、内部抑制性入力は同じ層からの抑制性入力である。

2.2.3 出力の制御

各ノードは他からのノードの入力値を基に出力（重み）を計算し、下位層へ出力する。

$$W_{ij} = f(In_{ij}^W, Ih_{ij}^W) \times \alpha \quad (9)$$

式 (9) を基に出力に関して次のような戦略をとる。

if $\sum_j P_j < 1$ then

$$W_{ij} = f(\alpha \times rand(), 0), \quad (10)$$

else if ($Th_i(t) > Th_i(t-1)$)

$$\& \left(\sum_j P_j > 0 \right) \text{ then}$$

$$W_{ij} = \begin{cases} f(0, \alpha \times rand()), & \text{if } P_j > 0 \\ f(\alpha \times rand(), 0), & \text{if } P_j < 1 \end{cases}. \quad (11)$$

$$\text{else } W_{ij} = f(0, 0). \quad (12)$$

あるノードの自分のしきい値が上昇しての場合、Priorityを持つ下位ノード（群）へ抑制性出力をだし、Priorityを持っていないノードに対する出力を上げる（式（11））。

自分のしきい値が変化しない、あるいは減衰してい場合にはそのままにする（式（12））。

下位ノードのどれもPriorityを有しない場合には、ランダムに出力を与える（式（10））。

各ノードに戦略としてこのような設定をあらかじめ与えることによって、ノードが持つ内部活性値を自律的に制御し、外部興奮性に依存しながら自らの内部活性値を更新・維持されることが期待できる。

2.2.4 抑制性出力の制御

各ノードは同じ層にある他のノードへの抑制を行なう。自分以外のノードを自ら抑制することによって、エージェントはより正しいと信ずる行動を維持するようになる。

$$Wh_{ik} = f(I_{in_{ik}}^{Wh}, I_{ih_{ik}}^{Wh}) \quad (13)$$

抑制性出力に関して次のような戦略をとる。

if ($P_i > 0$)&($Act_i > 0$) then

$$Wh_{ik} = f(0, \beta \times rand()), \quad (14)$$

else if ($P_i > 0$)&($Act_i < 1$) then

$$Wh_{ik} = f(\beta \times rand(), 0), \quad (15)$$

$$\text{else } Wh_{ik} = f(0, 0). \quad (16)$$

Priorityを確保し、かつアクションの実行が有効な場合（ $Act_i > 0$ ）、他のノードへの抑制性出力を上げ（式（14））、実行が無効の場合、出力を下げる（式（15））。

Priorityを確保できていない場合、何もしない（式（16））。

2.2.5 しきい値の制御

ノード i から上位ノード y へしきい値（ O_{iy}^{th} ）を出力する。

$$O_{iy}^{th} = f(I_{in_{iy}}^{th}, I_{ih_{iy}}^{th}). \quad (17)$$

しきい値出力には次のような戦略とする。

if ($P_i > 0$)&($Act_i > 0$) then

$$O_{iy}^{th} = f(0, \gamma \times rand()), \quad (18)$$

else if ($P_i > 0$)&($Act_i < 1$) then

$$O_{iy}^{th} = f(\gamma \times rand(), 0), \quad (19)$$

$$\text{else } O_{iy}^{th} = f(0, 0). \quad (20)$$

Priorityを確保し、アクション実行権の結果が成功ならば、しきい値を下げ（式（18））、失敗ならば増加させる（式（19））。それ以外は何もしない（式（20））。

以上のようなしきい値制御を通じて、ノードの優先順位を決定していく。これにより Priorityを持つノードであれ、その行動結果が成功でなければしきい値が上昇することにより Priorityを失い、また、Priorityを持たないノードでもしきい値を維持することにより、ステップが進むうちに Priorityを持つことが期待される。これはエージェントの行動の維持および中断を意味する。

また、以下の実験において α , β , γ はそれぞれ 10, 10, 3 とする。

3. 基本的行動選択戦略

まず、上で示したような戦略性を導入したノードを用いた行動選択ネットワークの振舞いを確認する。図2に示されるような単純構造の行動選択ネットワークを用いたエージェントによるうろつき行動と、ノード間の活性値伝播の仕方であるダイナミクスについて考察する。

上位ノードを選択ノード、下位ノードをモーターノードとし、モーターノードは上位ノードへのしきい値出力と他のノードへの抑制結合を持っている。モーターノードはそれぞれ2方向の移動に対応している。

3.1 実験

壁で囲まれた2次元（5×5）平面グリッドを環境とし2方向のモーターノードの移動アクションに対するダイナミクスについて計算機実験を行う。ここでは

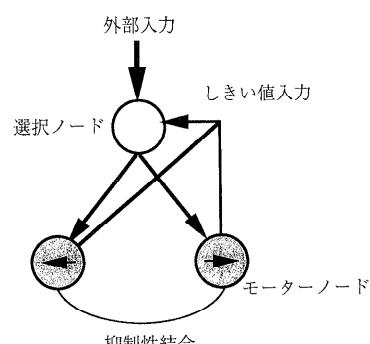


図2 うろつき行動用行動選択ネットワークの構成

Fig. 2 Structure of action selection network for wandering action.

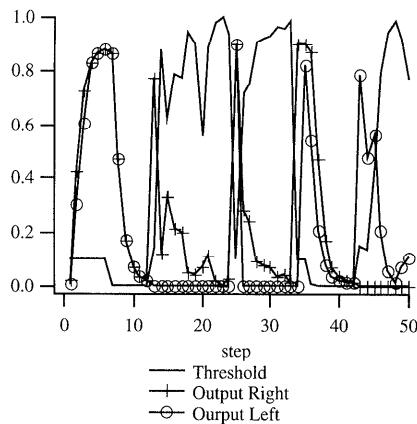


図3 うろつき行動の結果
Fig. 3 Results of wandering action.

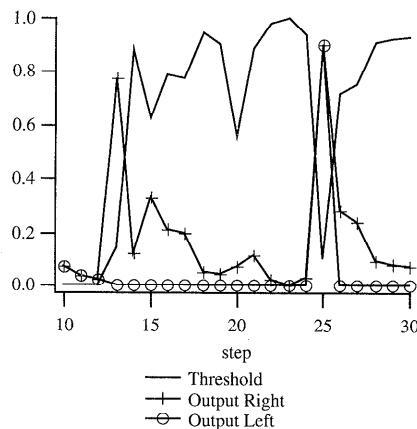


図4 基本的行動戦略の結果
Fig. 4 Results of standard action selection strategy.

外部入力として壁の有無についての情報を与えることによって、行動を開始する。エージェントは1ステップにおいて選択ノードの外部入力の情報からモーターノードのアクション実行権を決定し、選択した行動が成功であると評価されるまでネットワーク内の計算を繰り返す。これによってエージェントは1ステップにおける行動パターンを複雑にすることができます。

ここでは、エージェントが壁にぶつかる行動を選択するとその行動を失敗と評価しそれ以外を成功とした。実験結果を図3に示す。

3.2 考 察

図3はモーターノードから選択ノードへの出力しきい値と選択ノードから左右モーターノードへの出力を表したものである。図3の一部を拡大したものが図4である。

選択ノードへのしきい値が変化することによって、

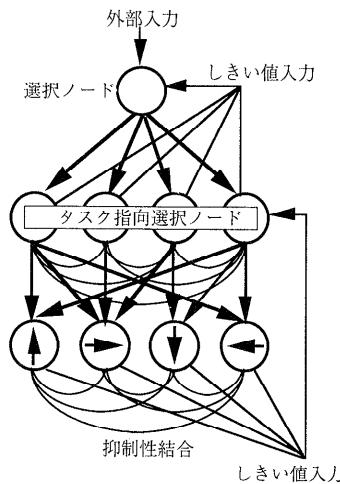


図5 タスク処理向け行動選択ネットワークの構成
Fig. 5 Structure of task oriented action selection network.

2つのモーターノードへの出力が入れ替わり、また、モーターノードの出力による行動の結果によって、しきい値が変化していることが分かる。しきい値が上昇するとそのノードの活性値が減少し、ノードの優勢度(Priority)が変化することにより、他のノードが活性化する。このことを繰り返していることからエージェントがうろつきまわるような行動をしているといえる。

実験結果よりモーターノードの移動失敗によってしきい値戦略を通じて選択ノードのしきい値が上昇し、これにともなう出力結合の制御によって行動が切り替えられること、および外部入力からの入力信号の停止によって時間差を生じながらネットワーク内の活動が停止することが確認された。このようなノードを持つネットワークによって自律的な判断に基づく選択行動の切り替えと、選択した行動の維持が実現できることが確認された。次にこのネットワークを基にタスク処理向けのネットワークを展開していく。

4. タスク処理向け行動選択ネットワーク

ここでは、エージェント群による組織化を通じたタスク処理を実行するための行動選択ネットワークを提案する。ノードは先に示したダイナミクスを基本とし、出力、しきい値出力、抑制性出力結合によって階層的な入出力関係を持ったネットワーク構造である。ネットワークの構成を図5に示す。選択ノードとモーターノード間に中間ノード組み込む。このノードをタスク指向選択ノードと呼ぶことにする。本論文ではこのノードを環境中のエージェントの目標の数と同数のノードを設定する。

このネットワークの特徴としてはモーターノードの上位ノードとしてタスク指向選択ノードを追加している点である。このタスク指向選択ノードからの出力によりモーターノードが持つ内部活性値が変化し、結果的にエージェントの行動を決定することができる。また、このような中間層として存在させることにより、エージェントの目標を多様化することができる。つまり、複雑な環境への適用を実現することができる（モチベーションの導入）。具体的には目標が変化する場合、たとえば目標が増加したり減少する場合でもエージェントの内部レベルで動的にネットワーク構造をエージェントが自律的に制御することができる。図5においては簡略化のためタスク指向選択ノードからモーターノードへの結合状態の一部を省略してある。次にこのタスク処理向け行動ネットワークの適用を説明する。

4.1 計算機実験

エージェント群によるタスク処理の適用例として2次元平面内のマルチエージェント系モンキーバナナ問題を考える。本問題は、天井からぶら下がっているバナナを猿（エージェント）たちが互いに手をつないだり肩の上にのぼったりしながら、すなわちエージェントどうしが分離結合を繰り返して単独ではとることができないバナナを、複数のエージェントが集合することによって獲得を目指すというものである。

ここでの各エージェントの目標は他のエージェントとの結合や分離を繰り返しながらバナナを獲得することであるが、バナナが天井に近いところにあればあるほど多くのエージェントが協力しながら組織化行動をとる必要があること。また同時並行的な組織化も可能であるため、各組織集団が目標物を取得するのに必要なエージェントの数が互いに不足し合うといったデッドロック状況も生まれ得る。このような場合には、現在の目標物の取得を諦め、組織を解体し他の目標物を選択し直すといった行為が望まれる。いずれにせよ、各エージェントはどの目標物に対してどのような組織化を形成するか、またその組織化の過程をどのように行うか、さらにはその組織化自体をどの時点で諦めて、他の目標を選び直すかといった一連の行動計画あるいはその選択のための基準をあらかじめ与えられていない。このような設定下において、各エージェントの自律性として目標物に関する情報と自分の行動結果に対する成功・不成功の情報だけから、行動選択ネットワークの特性を使って、状況に応じた適切な目標物の選択とその取得に必要な行動を自律的に探し出して行く必要がある。

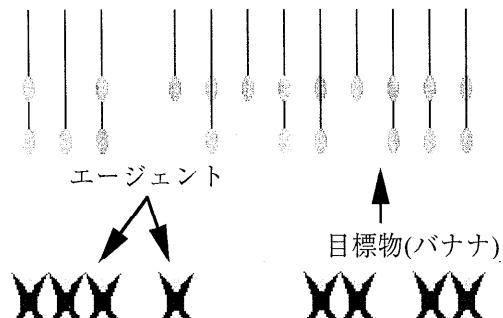


図6 モンキーバナナの世界

Fig. 6 Monkey Banana World.

図6に実験の画面である 14×6 のMonkey Banana Worldを示す。グリッドの下に位置するのがエージェントであり、上から下がっているのがバナナである。エージェントは8近傍に移動可能で、かつ隣り合ったエージェントの上3方向に移動できるものとする。この場合、図5におけるモーターノードは8個となる。また、エージェントのバナナが獲得できる位置は自分の近傍および同じ位置にある場合とする。

タスク指向選択ノードは環境中に生存するバナナの数とうろつき用のノードを設定する。たとえば初期状態で8個のバナナの場合、うろつき用のノードと合わせて9個のノードとした。タスク処理にともない目標としたバナナを獲得し終えたノードは動的に削除されるものとした。最上位の選択ノードにはつねに外部入力（ここでは1.0）を与えておく。

モーターノードの選択は3節と同様に行動が成功するまで行うこととする。ただし、バナナを獲得する行動、あるいはバナナに近づく行動を選択した場合を成功と設定する。

4.2 タスク処理におけるセンシング

ノードへの入力においてタスク指向選択ノードには、つねに自分が担当するバナナの位置に対応した外部入力を選択ノードからの入力とともに与える。

$$A_i = f(I_{in_i}^{out} + I_{in_i}^{in}, I_{h_i}^{in}). \quad (21)$$

$I_{in_i}^{out}$: 外部興奮性入力

外部興奮性入力としてエージェントの現在位置からの各目標バナナへの距離の変化、概略方向を入力値とする。さらに、ここでは次のような手段を用いて協調行動の促進を促すものとした。つまり、重なり合ったエージェント群においてはPriorityを持つタスク指向選択ノードのバナナの位置のセンシングとして、積み重なったエージェント群の中で目標とするバナナに最も近いエージェントからみた距離と方向が伝わるものとする。このような情報伝達を行うことによって、自

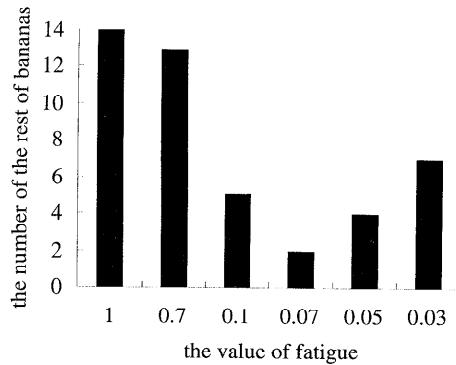


図7 残りバナナ数の結果
Fig. 7 Results of the rest of bananas.

分1人では獲得できないバナナも他のエージェントが自分の上に乗ることによって距離が近くなり、目標物が近くなると見なすことで現在の選択行動の維持をはかり、協調行動の形成を促進させる。結果として協調行動の末に目標とするバナナを獲得されることが期待されることとなる。ただし、情報伝達を受けたタスク指向選択ノードが常にアクション実行権を保持しているとは限らない点については注意を要する。

4.3 実験結果と考察

14×6 の Monkey Banana World, エージェント数を8, バナナの数を20とする。エージェントは最下部に、バナナは高さ3から5のあいだにランダムに配置する。たとえば初期状態は図6のようになる。

まず、ここでは行動選択ネットワーク上で用いられる式(1)におけるすべての疲労度(fatigue)を一様に設定し、それを変化させた場合の組織的様相の変化をバナナの残り数によって評価することを目的とする。方法として200ステップの試行を10回行い、その平均残バナナ数による結果とした。その結果を図7に示す。

これより疲労度を少なくすることによってエージェントの前の状態が維持されることから、回りの状況を判断しエージェントどうしが組織化することによりバナナを獲得すると思われる。また、疲労度が0.1, 0.03, 0.5の場合、0.07に比べ残バナナ数が多くなるのは、エージェントがデットロックの解除や組織化を解除できずに無駄な組織化を行ってしまい結果的に評価値が落ちることになったためと思われる。

また、ある試行におけるエージェント群のバナナの獲得経過を考察する。あるステップでのエージェントがバナナを獲得する一連の経過を図8に示す。図8(a)において重なり合ったエージェント群(左側にある4つのエージェントからなるグループ)に着目する。こ

のとき4つのエージェントに番号1から4を付ける。図8(a)の状態は左から2番目に位置しているバナナ(これをtargetと表す)をこの時点での最も優先すべき目標であると判断してエージェントが組織化している状態である。具体的にはエージェント1がtargetからの距離と方向を観測し、それに対応したタスク指向選択ノードからの出力により行動を選択した結果図8(a)の位置への移動を行う。このときエージェント2はエージェント3から、エージェント3はエージェント1あるいは2より、そしてエージェント4はエージェント3からの情報伝達によりtargetからの観測情報を共有している。つまり、targetに最も近いエージェント1の情報を結合する(重なり合う)ことによって獲得している。そして、次のステップではエージェント1, 3, 4は行動を維持し(静止行動とする)、エージェント2はエージェント1の上に重なる。また、次のステップ(図8(c))も同様にエージェント1, 3, 4は行動を維持しているために、エージェント2がバナナを獲得することができた。このようなことから結合や分離といった組織的行動によって、エージェント単独では行うことができない目標が達成できているといえる。

最後に、本提案ネットワークで導入した上位層のしきい値へのフィードバック制御の効果について行った実験結果を示す。 15×10 のMonkey Banana World, エージェント数12, バナナ2個をそれぞれ高さ10に環境内に等間隔で配置した。このとき、2.2.5項で導入したフィードバック制御を持つエージェント群と持たない場合のエージェント群との組織化能力の比較を行った。つまり、それぞれの場合について、各ステップごとに2個あるバナナの内の片方に対応するタスク指向選択ノードにアクション実行権を保持しているエージェント数とそのエージェント内での最大高さについて200ステップまで試行した結果を図9と図10に示す。

ここで1ステップは全エージェントが1回ずつランダムな順番で行動選択したときを表す。この比較から、導入したフィードバック制御機構を持たせることによって、ある目標へ着目するエージェント数が柔軟に変化していること、また、しきい値の変化に基づいてタスク指向選択ノードがモーターノードへの出力を柔軟に変化せることによってより有効な行動選択がなされ、頻繁に高い高さの組織化を成功させていることが分かる。一方、フィードバック制御機構を持たない場合には、着目するエージェント数の変化はなく、また有効な行動選択が成されにくいために、組織の高さ

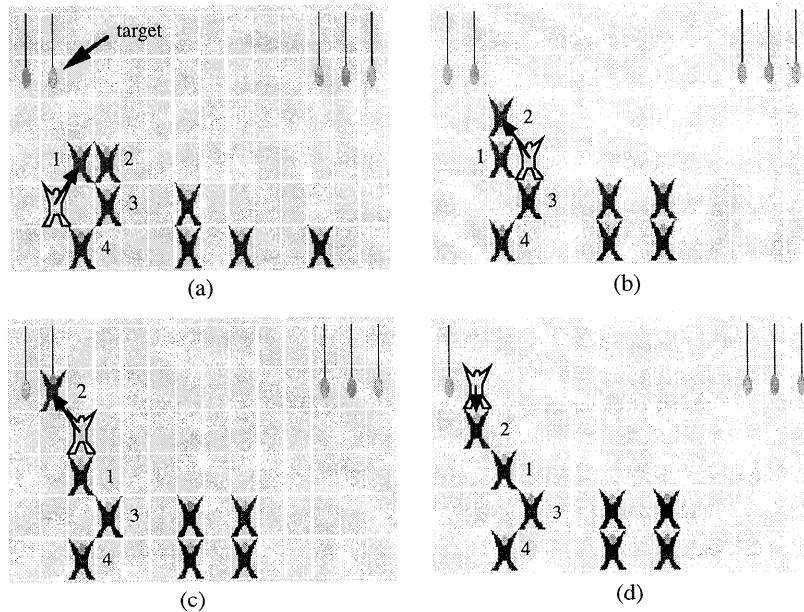


図 8 組織的行動による目標物取得の一例

Fig. 8 An example process of organizing actions for taking a target.

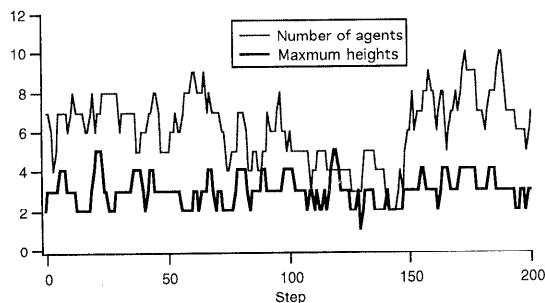


図 9 上位層しきい値へのフィードバックを持つ場合の target 1 へ着目しているエージェント数と最大高さの推移

Fig. 9 Progress of number of agents selecting target 1 with feedback mechanism and the heights of organizations toward target 1.

を上げづらくなっていることが分かる。

以上のことから、各エージェントにおける独立した目標の選択とそれに基づく移動行動の選択における柔軟性が、結果的に組織化の成功をより促進しているといえる。ただし、この実験において、十分なエージェント数(12匹)があるにもかかわらず、高さ10に設定された2つのバナナのどちらも取得できなかった点と、必ずしも着目するエージェント数が多いからといって達成できる組織の高さが高くならない点については問題が残る。これらの問題に関しては、行動選択ネットワークの機構の改良、およびエージェントどうしが作り出す動的環境の影響に関する考察などが必要

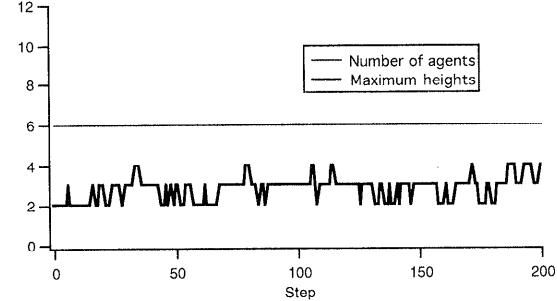


図 10 上位層しきい値へのフィードバックを持たない場合の target 1 へ着目しているエージェント数と最大高さの推移

Fig. 10 Progress of number of agents selecting target 1 without feedback mechanism and the heights of organizations toward target 1.

とされると思われる。

5. 結 言

本研究では組織的な協調行動のための行動発現機構として、ダイナミクス機構をベースに、明示的かつ戦略的な活性伝播制御と下位層から上位層へのフィードバック機構を導入した行動選択ネットワークを提案した。この行動選択ネットワークをマルチエージェントによるモンキーバナナ問題への適用を図り、計算機実験を通して組織的行動の発現を確認した。

今後は、さらに洗練された組織的行動の実現と、多様な組織化として他の問題へのアプローチを試みる。

参考文献

- 1) Maes, P.: How to Do the Right Thing, *Connection Science Journal*, Vol.1, No.3, pp.291-322 (1989).
- 2) Maes, P.: Situated Agents can Have Goals, *Robotics and Autonomous Systems* 6, pp.49-70 (1990).
- 3) Maes, P.: A Bottom-up Mechanism for Behavior Selection in Artificial Creature, *Proc. First International Conference on Simulation of Adaptive Behavior*, Meyer, J.A. and Wilson, S. (Eds.), pp.238-246, MIT Press (1990).
- 4) Maes, P.: Learning behavior networks from experience, *Proc. First European Conference on Artificial Life*, pp.48-57, MIT Press (1992).
- 5) Giszter, S.: Behavior networks and force field for simulation spinal reflex behaviors of the frog, *Proc. Second International Conference on the Simulation of Adaptive Behavior*, Meyer, J. and Wilson, S.W. (Eds), pp.172-181, MIT Press (1992).
- 6) Sahota, M.K.: Action Selection for Robots in Dynamic Environments through Inter-behavior Bidding, *Proc. Third International Conference on Simulation of Adaptive Behavior*, Meyer J.A. and Wilson S. (Eds.), pp.138-142, MIT Press (1994).
- 7) Mataric, M.J.: Integration of Representation into Goal-Driven Behavior-Based Robots, *IEEE Trans. Robotics and Automation*, Vol.8, No.3, pp.304-312 (1992).
- 8) Cherland, S. and Troxell, W.: Interactivism: A Functional Model of Representation for Behavior-Based Systems, *Third European Conference on Artificial Life*, pp.691-703 (Jun. 1995).
- 9) Sutton, R.S.: *Reinforcement Learning*, Kluwer Academic Publisher (1992).
- 10) Mynsky, M.: *The Society of Mind*, Simon & Schuster, New York (1986).
- 11) Isida, T. and Kuwabara, K.: Distributed Artificial Intelligence (1): Cooperative Problem Solving, *Journal of Japanese Society for Artificial Intelligence*, Vol.7, No.6, pp.945-954 (1992).
- 12) Kuwabara, K. and Isida, T.: Distributed Artificial Intelligence (2): Negotiation and Balancing, *Journal of Japanese Society for Artificial Intelligence*, Vol.8, No.1, pp.17-25 (1993).
- 13) Tan, M.: Multi-Agent Reinforcement Learning: Independent vs. Cooperative Agents, *Proc. Tenth International Conference*, pp.330-339, Morgan Kaufman (1993).
- 14) Brooks, R.: A Robust Layered Control Systems for a Mobile Robot, *IEEE Journal of Robotics and Automation*, Vol.RA-2, No.1, pp.14-23 (1986).
- 15) Blumberg, B.: Action-Selection in Hamsterdam: Lessons from Ethology, *Proc. Third International Conference on Simulation of Adaptive Behavior*, Meyer, J. and Wilson, S.W. (Eds.), pp.108-117, MIT Press (1994).

(平成8年1月16日受付)

(平成9年9月10日採録)



吉村 潤（正会員）

1994年北海道大学工学部精密工学科卒業。1996年同大学院工学研究科修士課程精密工学専攻修了。同年、富士通（株）入社。在学中は人工生命、マルチエージェントの研究に従事。



鈴木 恵二（正会員）

1988年北海道大学工学部精密工学科卒業。1990年同大学院工学研究科修士課程精密工学専攻修了。1993年同博士後期課程修了。博士（工学）。同年、北海道大学工学部助手。1996年同大学院工学研究科複雑系工学講座助教授。ニューラルネットワーク、遺伝的アルゴリズム、マルチエージェントなどに関する研究に従事。人工知能学会、日本機械学会各会員。



嘉数 侑昇（正会員）

1973年北海道大学大学院工学研究科精密工学専攻博士後期課程修了。工学博士。1978年北海道大学精密工学科助教授、1987年同教授、1995年同大学院工学研究科複雑系工学講座教授。CAD/CAM、画像処理工学、ロボティクス、知識工学、ニューラルネットワーク、人工生命等の研究に従事。日本機械学会、精密工学会、日本ロボット学会等各会員。