仮想ポテンシャル関数を用いたエージェントの相互制御方法に関する検討

安田 真*・古橋 武**
橋山 智訓***・大熊 繁****

A Study on Mutual Control among Agents by Temporary Potential Functions

Makoto YASUDA*, Takeshi FURUHASHI**,
Tomonori HASHIYAMA*** and Shigeru OKUMA****

One of the major problems of the technique that controls actions of agents by potential functions is to determine the function shapes suitable for the environment. The authors have proposed a method for autonomous acquisition of the parameters of potential functions by the agents with reinforcement learning, especially classifier systems, in multi-agent environment.

In this paper, we study a method to introduce temporary potential functions, which agents can set at desirable positions, so as to improve controllability of other agents. Experiments with a pursuit game are done, and the following are confirmed: acquisition of the temporary potential parameters for capturing the prey agent is possible; acquisition of the temporary potential parameters for the hunter agents to control each other and cooperatively capture the prey agent is also possible; the temporary potential increases the capturing probability of the prey agent.

Key Words: potential function, reinforcement learning, multi-agent system

1. はじめに

複雑、かつ、動的なマルチエージェント環境下では、あらかじめ設計段階で各エージェントに対して、そのセンサ入力情報に応じた適切な行動を与えておくことは難しい。特に協調行動の実現には困難が伴う。そこで、筆者らは、協調行動を含むエージェントの自律的な行動獲得手法について検討を行ない、エージェント自身をポテンシャル関数で記述し、同関数を規定するパラメータを強化学習を用いて個々のエージェントに自律的に獲得させる手法を提案した¹⁵⁾。本論文では、さらに効果的なエージェントの相互制御を可能とするため、エージェントに、その本来の位置以外の場所に仮想的なポテンシャル源を設定させる手法について検討する。

従来、ロボットに代表されるエージェントの行動選択方法としては、センサ入力情報を分節化した入力状態空間に対応した *if-then* 型の行動ルール群を用意し、この中から行動選択をするものが主であった^{7), 14)}。これに対して人工的なポテンシャル関数を行動制御のベースに用いる手法は、エージェントの行動実行時に多大な計算量を要するプランニング量が削減され、通信が不要となるコンパクトで実時間性に優れた制御手法として、主にロボットやマニピュレータの障害物回避軌道制御のための経路探索に用いられてきた^{1), 6), 8), 10), 12), 13), 16), 17)}。しかしながら、ポテンシャル関数を用いる方法においても、関数形が未知の場合はもちろん、たとえ既知であってもその形状が複雑になる場合を想定すれば、環境に応じた適切なポテンシャルパラメータを設定することは依然として困難である。

これに対して筆者らは、強化学習の手法^{3), 4), 7)}を利用し、ポテンシャル関数を定めるパラメータをエージェントに獲得させることでこの問題に対処する方法を試みた。同手法は、エージェント自身がパラメータを獲得する結果、より環境に適した効率的なパラメータが得られる可能性がある。また、*if-then* 型の行動ルールではなくポテンシャル関数を獲得対象とすることで、エージェントの行動記述が簡潔になり、入力状態空間の分節サイズも大きく取れて学習が容易になると期待される。また、環境や行動の複雑度が増して、学習効率向上のためにエージェント間通信が必要になる場合を想

* 沖電気工業(株) 大阪市中央区城見 1-2-27

** 名古屋大学工学研究科電子情報学専攻 名古屋市千種区不老町

*** (財)名古屋産業科学研究所 名古屋市守山区大字下志段味字穴が洞 2271-130

**** 名古屋大学工学研究科電気工学専攻 名古屋市千種区不老町

* Oki Electric Industry Co., Ltd., Osaka

** Dept. of Information Electronics, Nagoya University, Nagoya

*** Nagoya Industrial Science Research Institute, Nagoya

**** Dept. of Electrical Engineering, Nagoya University, Nagoya

(Received December 10, 1999)

定しても、通信データ量が少なくすむと考えられる。さらに、エージェント設計時に要するデータ量を減らし、その分、エージェントの自律性も向上するなどの利点が挙げられる。実験の結果、2体のエージェントが相互に制御し合い、協力して行動するポテンシャルパラメータの獲得が可能なが示された。しかし、一方で、ポテンシャルで表現できるのは基本的には引力と斥力のみであり、エージェント1体に伴うポテンシャルの発生源が1つである限りエージェント間での微妙な制御は難しく、その相互制御性は高いとは言い難いという問題点も明らかとなった。

そこで、本論文では、エージェントに、その本来の位置以外の場所に制御点となる仮想的なポテンシャル源を設置する能力を持たせることで、より直接的かつ積極的に相手に作用を及ぼし得るように、相互制御方法の拡張を計る。具体的には、エージェント自身を表現するスカラーポテンシャル関数に加えて、エージェントが任意の場所に任意の勾配で生成できる一時的に有効な仮想的なスカラーポテンシャル関数を用意し、他のエージェントを自らの望む方向に誘導可能にする。この仮想的なポテンシャルの生成、消去のタイミングは、自己エージェントの要求に応じて決定する。

ポテンシャル関数やポテンシャルパラメータ値の獲得を行なった従来の研究例としては、障害物の形状パターンからポテンシャル場への変換にニューラルネットワークを利用することでポテンシャル場を生成した横山ら¹⁶⁾のもの、ロボットの作業空間をセルに分割し、各セルに与えるポテンシャル値を遺伝的アルゴリズム (GA) により探索することで軌道生成を行なった馬場ら¹⁾のものがあるが、いずれも環境 (障害物) が固定されており、また、マルチエージェントを対象にしたものではない。一方、山城ら¹³⁾は GA を用いて、マルチエージェント環境におけるエージェント間、および、エージェントと環境間のポテンシャルを定めるパラメータの探索を行なっている。しかし、パラメータの探索はグローバルな視点からなされており、個々のエージェントがパラメータの獲得主体とはなっていない。また、エージェントの相対配置と各ポテンシャル関数形は固定されている。

本論文では、マルチエージェント環境において、エージェント自身とエージェント外部の制御点を相互にポテンシャル関数で記述することを目標としている。このため、離散化された入力状態空間に対し、両タイプのスカラーポテンシャル関数の合成勾配によりエージェントの行動を束縛し、かつ、両関数の勾配や仮想的なポテンシャルの設置位置を規定するパラメータを強化学習を用いてエージェント自身に自律的に獲得させる手法について検討を行なう。ただし、ポテンシャル関数の次数は既知とする。追跡ゲーム⁵⁾を用いた本手法の検証実験を行ない、仮想ポテンシャルの導入により獲物の捕獲数が増加し、手法の効果が認められることを示す。

2. 問題設定

本論文で取り扱うエージェント、および、エージェントを

取り巻く環境は以下の通りである。まず、エージェントを、センサ入力情報に応じて行動を決定する行動主体と定める。マルチエージェント環境における各エージェントは能力的には基本的に均一で、組織構造や階層構造を持たない。しかし、エージェントは互いに識別可能な名前や番号を持ち、特定の相手と相互に通信可能である。また、全エージェントの情報を保持して監督となるエージェントの存在は仮定しない。エージェントの行動と行動目的は、並進動作による空間探索に限定する。ただし、選択可能な動作距離と進行方向の角度はあらかじめ離散化されている。一方、エージェントを取り巻く外部環境は、エージェントの1つの動作に応じて状態遷移する。エージェントはある時点における等方的なセンサの感知範囲内の、離散化された局所的な環境情報しか入手できない。すなわち、離散マルコフ決定過程を取り上げる。

以上の条件のもとで、エージェント、もしくはゴールを表すポテンシャル源 (以降、真のポテンシャルと呼ぶ) の勾配、および、エージェントがセンサ感知範囲内の任意の位置に設置可能な制御点を表すポテンシャル源 (以降、仮想ポテンシャルと呼ぶ) の位置と勾配を定めるパラメータを強化学習の手法により同時に獲得する方法を検討する。強化学習手法としては、比較的計算量が少ないことから、経験強化型の学習手法であるクラシファイア・システム^{2), 9)}を使用する。

3. ランダム運動のポテンシャルによる束縛効果

本節では、エージェントをランダムな外力を受けて動く粒子とみなしたとき、ポテンシャルによる束縛の効果について検討する。まず、ランダムな力 $F(t)$ に基づく運動方程式

$$m \frac{d^2 x}{dt^2} + m\gamma \frac{dx}{dt} = F(t) \quad (1)$$

を取り上げる。ただし、左辺第2項は、ゆらぎを生じさせるための抵抗として導入する。 $F(t)$ と $x(t)$ を確率過程として

$$F(t) = \sum_{n=-\infty}^{\infty} F_n e^{i\omega_n t}, \quad x(t) = \sum_{n=-\infty}^{\infty} X_n e^{i\omega_n t}$$

と Fourier 展開すると、(1) 式は、Fourier 成分間の関係として

$$X_n = \frac{1}{-\omega^2 + i\omega\gamma} \frac{F_n}{m}$$

を与える。 $F(t)$ 、 $X(t)$ の強度スペクトルをそれぞれ I_F 、 I_x と記し、また、 $F(t)$ をホワイトノイズと仮定すると、

$$I_x(\omega) = \frac{1}{\omega^2(\omega^2 + \gamma^2)} \frac{I_F}{m^2}$$

という関係が得られる。Wiener-Khinchin の定理¹¹⁾を利用して $x(t)$ 間の相関係数 $\langle x(t_1)x(t_2) \rangle$ を求め、さらに、変位 $x(t)$ の2乗平均 $\langle x(t)^2 \rangle$ との関係

$$\langle x(t)^2 \rangle = \int_0^t dt_1 \int_0^t dt_2 \langle x(t_1)x(t_2) \rangle$$

から、

$$\langle x(t)^2 \rangle = \frac{\pi I_F}{m^2 \gamma^3} t \quad (2)$$

を得る。

Fig. 1 Attracting effect of potential on particles moving randomly

一方、粒子に何らかの力の場合、(1)式は、

$$m \frac{d^2 x}{dt^2} + m \gamma \frac{dx}{dt} = -m \frac{\partial V}{\partial x} + F(t) \quad (3)$$

と拡張される。力の場合（ポテンシャル）を

$$-\frac{\partial V}{\partial x} = -\omega_0^2 x$$

と表すと、上記と同様にして、

$$\langle x(t)^2 \rangle = \frac{\pi I_F}{m^2 \gamma \omega_0^2} \quad (4)$$

を得る。(2)式と(4)式とを比較すれば、その大きさに応じて粒子を固定点に束縛するポテンシャルの効果は明らかである。

Fig. 1には、これら粒子の運動を具体的に示す。時刻 $t=100$ まではランダムに運動し、 $t=100$ 以降は、 $x=100$ の位置に置かれた引力ポテンシャル源の影響を受ける。ポテンシャルは、その形状に応じて、周囲の影響範囲内にある粒子を引き込む。

以上の結果を、通常はランダムに移動し、ポテンシャルを感知するとその影響にしたがって移動するような空間探索エージェントに適用して考える。このとき、ポテンシャルの存在は探索領域を圧縮することに該当する。すなわち、ある程度探索領域の絞り込みを可能とするような情報を持つエージェントが存在する場合、当該エージェントが同領域周辺の他のエージェント近傍に引力勾配を持つポテンシャルを置くことができれば、これを同領域内に引き込むことで同領域の効率的な探索が可能になると期待される。

そこで、本論文では、これを実現するための1つの手法として、仮想ポテンシャルを取り上げる。しかし、問題に応じて仮想ポテンシャルの位置やポテンシャルパラメータを適宜設定することは困難が予想されるため、ポテンシャルの位置やポテンシャルパラメータ値は、クラシファイア・システムを用いてエージェント自身に学習させる。

4. ポテンシャル関数の設定とパラメータ学習方法

ここでは一般的な空間探索問題を取り上げ、局所的な探索を行なう1体のエージェントを対象に、探索地点（固定したゴールや障害物、他のエージェントなど）や仮想的な制御点

をポテンシャル源とした場合のポテンシャル関数の設定方法とポテンシャルパラメータの獲得方法について述べる。

4.1 ポテンシャル関数の設定

エージェントの動作を律するポテンシャル関数（スカラーポテンシャル）を以下のように定める。ポテンシャルの形状としては様々なものが考えられるが、ここでは指数関数型の関数を取り上げる。ただし、有限のセンサの感知範囲内にポテンシャル源があるときのみ、エージェントは当該ポテンシャルの影響を受ける。逆に言えば、ポテンシャルの影響は有限の到達距離を持つ。このとき、ポテンシャル源はエージェントからの作用を受けず、また、ポテンシャルは自己力として作用しないものとする。

● スカラーポテンシャル

N 次元探索空間の各軸方向の単位ベクトル $e_{xi} (1 \leq i \leq N)$ に対し、スカラーポテンシャル $\phi(x_1, x_2, \dots, x_N)$ を

$$\phi(x_1, x_2, \dots, x_N) = a \cdot \exp \sum_{i=1}^N \{-b_i(x_i - x_i^0)^2\} \quad (5)$$

とする。ここで、 a 、 b_i は $x_1, x_2, \dots, x_N$ に依存しない定数値であり、 a はポテンシャルの高さ（深さ）、 b_i は影響範囲を定める。正の a に対して(5)式は引力を、負の a に対しては斥力を表す。また、 x_i^0 はポテンシャルの中心座標のエージェントからの相対位置を表す。

● ポテンシャル勾配

スカラーポテンシャル $\phi(x_1, x_2, \dots, x_N)$ の x_1 方向の勾配 f_1 は

$$f_1 = \frac{\partial \phi}{\partial x_1} = -2ab_1(x_1 - x_1^0)\phi e_{x1} \quad (6)$$

で与えられる。

さらに、ポテンシャル源が複数 (K 個) 存在する場合は、スカラーポテンシャル $\phi(x_1, x_2, \dots, x_N)[j] (1 \leq j \leq K)$ から、

$$f_1 = -2 \sum_{j=1}^K a[j]b_1[j](x_1 - x_1^0[j])\phi[j]e_{x1} \quad (7)$$

と計算される。すなわち、勾配は合成されたポテンシャルにより定まる。

4.2 パラメータ学習方法

適当なポテンシャルパラメータをあらかじめ与えられないような複雑な環境は多々想定されるため、学習によってパラメータを求める意味は大きい。パラメータ学習に際しては、入力状態空間に対応するポテンシャルパラメータ a 、 b_i と（仮想ポテンシャルの場合は）中心座標 x_i^0 をクラシファイア・システムを用いて学習する。エージェントの動作評価はポテンシャルの発見とともに開始し、ポテンシャル源への到達、もしくはポテンシャルの消失とともに終了する。このとき、動作評価はポテンシャル関数からは独立した形（ポテンシャル関数、ないしはパラメータ自体を含まない）で行なう必要がある。これは、例えば、エージェントのポテンシャル源に対

Fig. 2 Rule representation(Classifier)

する距離の変化と到達の可否のみで評価することで実現され、以下の (8) 式として定式化される。

● 評価関数

$$E = \frac{L(0)(1 + RWD)}{\sum_{t=0}^T L(t) \cdot (1 + T)} \quad (8)$$

ここで、 $t(0 \leq t \leq T)$ はエージェントの評価開始から終了までの動作ステップ数、 $L(t)$ は各ステップ時におけるエージェントとポテンシャル源との距離、 RWD はポテンシャル源への到達などのイベントに対する報酬である。評価関数の分母は、移動距離の総和が短く、かつ、全ステップ数が少ない場合に高い評価を与える。 E の最大値は、

$$E_{(T=0)} = 1 + RWD \quad (9)$$

となる。

● パラメータ獲得のための学習ルール表現

エージェントのセンサ入力情報があらかじめ適当に離散化されていると仮定し、前件部 (Condition) を入力状態空間、後件部 (Message) を真のポテンシャルパラメータと仮想ポテンシャルパラメータ、および、その中心座標とした Fig. 2 に示すようなルール表現を行なう。自身を除く他のエージェントを真のポテンシャル源とすれば、前件部と後件部の真のポテンシャルパラメータ部は、他のエージェント個数分の入力状態空間情報とパラメータを並べたものとなる。

仮想ポテンシャルのパラメータ獲得主体はポテンシャルを設置するエージェント自身であるが、他のエージェントをポテンシャル源とする真のポテンシャルのパラメータの獲得主体の選び方には 2 通りある。すなわち、(1) 自身で獲得する場合と、(2) 相手が獲得する場合とである。(1) は相手を固定して自身の動きを制御することを意味し、獲得したパラメータを相手に教える必要はない。そこで、ゴールや障害物ポテンシャルの場合との統一性と通信が不要な利点を考慮して、ここでは (1) の立場を採用する。

クラシファイア・システムを用いた学習システムの処理手順を Fig. 3 に示す。以下、図中の番号にしたがって処理内容を説明する。あらかじめランダムに初期ルールを生成した後、(1) センサから環境情報を取得し、(2) 真のポテンシャル源が存在しない場合は、(4) 他のエージェントに対して仮想ポテンシャルの要求を行なう。(3) 他のエージェントから仮想ポテンシャルパラメータが送信されているかどうかを判別し、

(5) 真もしくは仮想のいずれかのポテンシャル源が存在する場合に、ポテンシャルに基づく動作とその評価処理を開始する。まず、(6) 環境に即した真のポテンシャルパラメータを選択し、(7) 必要に応じて、真のポテンシャルと仮想ポテンシャルを統合した後、(9)(6) 式からエージェントの受ける勾配力を算出して移動する。(10) 報酬イベントが生じた場合、(11)(8) 式に基づき各ルールに対する報酬割り当てを行なう。さらに、(12) 一定回数の試行の後、GA 操作により新たなルールを生成する。一方で、(13) 他のエージェントから仮想ポテンシャルの要求があるかどうかを常時監視し、(14) 要求がある場合は仮想ポテンシャルパラメータを選択して、(15) 要求元のエージェントに送信する。

ただし、すべての入力状態に応じた学習を行なう必要はなく、ポテンシャル配置に対称性があれば、入力状態数の削減が可能である。例えば、1 個のポテンシャル源のみ存在する場合は、等方的な空間分割を行えば十分である。

5. エージェントの動作制御方法

ポテンシャル勾配によるエージェントの動作制御方法は以下の通りである。まず、エージェントの現在位置 $P(x_1, x_2, \dots, x_N)$ に対し、有限のセンサ感知範囲内の空間において、エージェントが 1 回の並進動作で到達可能な地点を $P[k] = P'(x'_1[k], x'_2[k], \dots, x'_N[k]) (1 \leq k \leq C_K)$ とする (Fig. 4(a) に示すように、一般には、 P を中心とする半径が最大並進動作距離の N 次元球面内部の点となる)。さらに、このとき、各 P' 点における勾配 f の値を $f[k] (1 \leq k \leq C_K)$ とする。

一方、エージェントの並進動作の大きさを $M[l] (1 \leq l \leq C_L)$ とする。さらに、各並進動作の進行方向が $u[l] (1 \leq l \leq C_L)$ に離散化されているものとする。

このとき、エージェントの現在位置 $P(x_1, x_2, \dots, x_N)$ における進行方向の単位ベクトル $u[l] = u_{x1}[l]e_{x1} + u_{x2}[l]e_{x2} + \dots + u_{xN}[l]e_{xN}$ について、その各 $f[k]$ の $u[l]$ 方向への方向微分 $f'_u[k][l]$ は

$$f'_u[k][l] = \frac{d\phi[k]}{du[l]} = u[l] \cdot f[k] \quad (10)$$

と計算される (Fig. 4(b) 参照)。 $f[k]$ は等位面 $\phi = \text{const}$ に垂直なベクトルとなる。そこで、ポテンシャル関数とポテンシャルパラメータが与えられたとき、Fig. 3 の手順 (9) におけるポテンシャルに基づく移動は以下のように実行する。

Fig. 3 Flow of processing of system

Fig. 4 (a) Example of reachable positions of agent
(b) Directional differentiation

- (1) センサ入力情報として観測されたポテンシャル源の距離から、エージェントの到達可能な地点 $P'(x_1[k], x_2[k], \dots, x_N[k])$ における同勾配 $f[k]$ を求める。
- (2) すべての k に対して $f[k]$ が同じ場合、ランダムに移動する。
- (3) すべての l に対して各並進動作 $M[l]$ の進行方向単位ベクトル $u[l]$ を求める。
- (4) $f[k]$ と $u[l]$ から方向微分 $f'_u[k][l]$ を求める。
- (5) $f'_u[k][l]$ の中から負の最大値を与える $f'_u[k][l] = l_M$ を選択する。
- (6) 並進動作 $M[l_M]$ を実行する。

6. シミュレーション実験

追跡ゲームを例題として、仮想ポテンシャルを利用したエージェントの相互制御方法のシミュレーションによる検証

実験を行なった。

6.1 追跡ゲームの設定

獲物 (1 体) とハンター (1~4 体) の追跡ゲームを取り上げる。同ゲームには様々なバリエーションがあるが、ここでは以下のような設定を用いた。

- 全体はサイズ (FIELD SIZE) が 9~19 柵の正方形のフィールドからなり、上下端と左右端はトーラス状に繋がっており、無限平面を構成している。
- ハンターは 1 シミュレーションステップに 1 柵分上下左右いずれかの方向に進むか、もしくは、その場に留まる。獲物は指定シミュレーションステップ (PREY MOVE INTERVALS) に 1 柵分上下左右いずれかの方向にランダムに進むか、もしくは、その場に留まる。
- ハンター同士は同じ柵目を占められない。
- エージェントは自身を中心とした固定した視野を持つ。入力状態数を減らし、かつ、より現実的な視野表現を行なうため、視野はボックス状に分割し、遠方になるほど情報量を減少させる。Fig.5(a) に 1 例を示す。実線が視野ボックスを表す。エージェントにはボックス内部の詳細位置はわからない。また、視野のサイズ (SIGHT RANGE) を図中の矢印方向の柵目の数で定める。
- Fig.5(b) で、獲物 P の捕獲条件は、ハンターが 1 体の場合は P に重なること、2 体の場合は (N) と (S) か (E) と (W) の 2 箇所に同時に位置すること、3 体の場合は (N)~(W) 中の 3 箇所に同時に位置すること、4 体の場合は (N)~(W) の 4 箇所に同時に位置することと定めた。

Fig. 5 Setting of agent's sight in pursuit game

Table 1 Values of the rule strength

Initial value of strength (S_{INIT})	100.0
Threshold of strength (S_{TH})	$S_{INIT}/2$
Maximum value of tax (T_{MAX})	0.05
Minimum value of tax (T_{MIN})	0.01
Renewal number of step of tax (T_{STEP})	10
Value of bid (Bid)	0.1

6.2 ポテンシャル関数の設定と動作学習方法

探索空間は2次元平面なので、スカラーポテンシャル $\phi(x, y)$ を以下のように設定した。ただし、各 x, y 座標値は罫目の中心位置の値を用いる。

$$\phi(x, y) = a \cdot \exp\{-b_1(x - x^0)^2 - b_2(y - y^0)^2\} \quad (11)$$

このとき、スカラーポテンシャル $\phi(x, y)$ の勾配 f は

$$f_x = -2ab_1(x - x^0)\phi, \quad f_y = -2ab_2(y - y^0)\phi \quad (12)$$

と計算される。獲物の周囲の捕獲位置に同ポテンシャルを設定し、ハンターにはこれを1組のポテンシャルパラメータとして学習させた。

評価関数 E は(8)式をそのまま用いた。 RWD がエージェントの動作で生じたイベントに対する評価値であり、報酬として発火したルール強度に加算される。 $t = 0$ は視野内にポテンシャル源が入り、エージェントがポテンシャルに基づき動作し始めた時間、 $t = T$ は何らかのイベントが生じてポテンシャルに基づく動作が終了した時間とする。

ルール強度の初期設定はTable 1のように行なった。 tax は1シミュレーションステップごとに全てのルールから一定の強度を減じる税金を表し、これにより、正の報酬を受けないルールの強度は速やかに減少する。また、獲物を捕獲したルールは捕獲数 $COUNT$ に応じて tax を

$$tax = \max(tax', T_{MIN})$$

$$tax' = T_{MAX} - \frac{T_{MAX} - T_{MIN}}{T_{STEP}} COUNT$$

と変化させ、より多く獲物を捕獲したルールは、強度の減少割合を緩やかにすることで発火確率を高めた。また、入力状態に一致するルールが複数存在する場合は、ルール強度に比例したルーレット選択によって発火するルールを定めた。

分割された視野情報を入力状態空間として、ポテンシャルパラメータ a, b_1, b_2 と(仮想ポテンシャルの場合は)中心座標 x^0, y^0 を獲得するための学習ルール表現をFig. 6に示す。前件部はFig.5(a) (SIGHT RANGE=3)の視野を例と

Table 2 Values of the GA parameters

Population	5000
GA Interval	(Experiment(1), (2)) 500 (Experiment(3)) 1000
Probability of mutation	2 %
Probability of crossover	50 %

し、24個に分割された視野を0, 1にdon't care *を加えたスキーマ5文字で表現している。ポテンシャルパラメータは各3ビット、座標値は2ビットとした。また、仮想ポテンシャルはエージェント当たり1個に制限した。

GA処理に関するパラメータをTable 2に示す(各実験の詳細については次節を参照)。GA IntervalはGAの実行間隔(何シミュレーションステップごとに実行するか)を表す。獲物を捕獲したルールは保存して次世代に残し、強度が S_{TH} 以下になったルールは削除して新規に生成する。これら以外のルールについては、強度に比例した確率で生き残るものをルーレット選択し、2点交差と突然変異を施した。

7. 実験条件

以上のような設定のもとで、Table 3に示す各実験を行なった。ただし、HUNTERSはハンターの数である。INITIAL POSITIONがFixedの場合はハンターと獲物の初期配置が固定されていることを、Randomの場合はランダムに設定されることを表す。また、PREY MOVEMENTがYesの場合には獲物はランダムに移動し、Noの場合には移動しない。MAXIMUM STEPは、あるルールが発火後、そのルールにしたがって動作する際の最大シミュレーションステップ数を表す。実験(1), (2), (3)における獲物とハンターの初期配置をFig. 7(1), (2), (3)にそれぞれ示す。

実験(1)では2体のハンターH1, H2により仮想ポテンシャルを用いて獲物Pの捕獲を行なう。ただし、ハンターと獲物の初期配置は固定でPは移動しない。図中の破線は各ハンターの視野を表す。実験(2)ではFIELD SIZE^(注1), SIGHT RANGE, HUNTERS, PREY MOVE INTERVALを変化させて仮想ポテンシャルの効果を調べる。このとき、獲物がランダムに移動する点が実験(1)と異なる。実験(3)ではH1, H2とPの初期配置はランダムであり、PREY MOVE INTERVALを変化させてランダムに移動する獲物を捕獲可能かどうかを検証する。

各実験におけるイベント報酬はTable 4のように定めた。発火して獲物を捕獲したルールは R_{CAPT} の報酬を得る。また、発火して獲物を捕獲できなかった場合、すなわち、獲物が視野外に逃げた場合は R_{OUT} 、ルール発火後MAXIMUM STEPが経過した場合は R_{OVER} の負の報酬を与える。挟み込み動作を獲得するため、発火してハンターが単独で獲物周囲の(N)~(W)のいずれかの位置に到達した場合のルール

(注1) FIELD SIZEを変更した場合のハンターと獲物の初期配置は、H1とPはフィールド右上隅、H2は左下隅からの距離がFig. 7(2)と同じになるように設定した。

Fig. 6 Rule representation(classifier)

Table 3 Experimental Conditions

	FIELD SIZE	SIGHT RANGE	HUNTERS	INITIAL POSITION	PREY MOVEMENT	PREY MOVE INTERVAL	MAXIMUM STEP
Experiment(1)	9	2~3	2	Fixed	No	-	15
Experiment(2)	9~19	1~5	1~4	Fixed	Yes	1~10	15
Experiment(3)	9	1~5	2	Random	Yes	1~10	20

(1)Initial positions of hunters and prey are fixed.

(2)Initial positions of hunters and prey are fixed. Prey moves randomly.

(3)Initial positions of hunters and prey are randomly set. Prey moves randomly.

Fig. 7 Experiments

Table 4 Values of the rewards

Capture reward (R_{CAPT})	$S_{INIT} \times 20$
Side reward (R_{SIDE})	$S_{INIT} \times 5$
Out of sight reward (R_{OUT})	$-S_{INIT} \times 10$
Overtime reward (R_{OVER})	$-S_{INIT} \times 10$
Overlap reward (R_{OLAP})	$-S_{INIT} \times 10$

8. 実験結果と考察

8.1 実験(1)

・仮想ポテンシャルパラメータの獲得

Fig. 8 と Fig. 9 は、実験(1)におけるルール強度の最大値 [RULE STRENGTH], および、獲物の捕獲に成功したルール数 [CAPTURE RULES] の変化を示す。ただし、評価ステップ数 [EVALUATION STEPS] は、獲物がハンターの視野内にあり、ポテンシャルに基づいて動作している状態でカウントしたシミュレーションステップ数を指す。捕獲ルールがかなり初期段階から見つかっていることから、ポテンシャルパラメータの獲得自体は容易であることがわかる。最大強度は様な増加を見せないが、これはルーレット選択により発火するルールを定めた結果、捕獲ルールの発火インターバルが大きくなり、ルール強度が tax の支払い効果によって減少していくためである。

Fig. 10 は、獲物を捕獲するまでに2体のハンターが移動した経路の例を示す。Fig. 7(1) に示したように、H1 は $(x, y) = (5, 3)$, H2 は $(x, y) = (2, 6)$ を初期位置とし、H2 の視野内には P が位置しないため、H1 は Fig. 11 に示すような仮想ポテンシャル T を図中の位置に生成した。T の作る引力勾配を受けて、H2 は $(x, y) = (3, 6) \Rightarrow (4, 6)$ と移動することで P を視野内に捉え、捕獲に至ったことがわかる。

・捕獲におけるポテンシャルの引き込み効果

上記実験で得られたポテンシャルパラメータを用いて繰り返し

には R_{SIDE} という中間報酬を与える。また、ハンター同士が同じ柵目を占めた場合は R_{OLAP} という負の報酬を与える。 R_{CAPT} か R_{OVER} イベントでシミュレーションはリセットされ、ハンターと獲物は初期配置状態に戻される。

Fig. 8 Maximum rule strength(exp.(1))

Fig. 9 Number of capture rules(exp.(1))

Fig. 10 Hunter's routes(exp.(1))

Fig. 11 Temporary potential function(exp.(1))

返しシミュレーションを行ない、ハンターが獲物を捕獲するまでにどのような位置分布を取るかを求めた。Fig. 12(a)はハンターがランダムに動作した場合、Fig. 12(b)は真のポテンシャルのみを用いて動作した場合、Fig. 12(c)は仮想ポテ

ンシヤルも加えて動作した場合の結果であり、それぞれ 20 回の計算により得られた。縦軸 (z 軸) が頻度を表し、実線が H1、破線が H2 の分布である。なお、Fig. 12(b) と Fig. 12(c) は、Fig. 12(a) に対して z 軸を 5 倍に拡大して表示している。この結果、Fig. 12(b) では、P のポテンシャルが最初から H1 の視野内にあるため、H1 に対して大きな引き込み効果を与えている様子が見られる。一方、H2 はランダムに移動して偶然 P を視野内に捉えた後、そのポテンシャルの影響を受けるため、H1 に比べると位置の分散の度合は大きくなる。しかし、Fig. 12(c) のように、H1 が適切な位置に仮想ポテンシャルを生成すれば、それによって H2 も P の方向に引き込むことができ、速やかに獲物が捕えられる。

この引き込み効果を定量的に比較するために、H2 の初期位置からの移動距離の 2 乗平均値のシミュレーションステップごとの変動の様態を Fig. 13~Fig. 15 に示す。各図の点が移動距離の 2 乗値、実線がそれらの平均値を表し、それぞれ Fig. 12(a)~(c) のデータから得られた。各図で値が途中から一定値になるのは、獲物が捕獲されたことを表す。以上の図から、真のポテンシャルが距離の変動幅を狭め、ハンターを獲物に急速に近づけること、仮想ポテンシャルによってその傾向がさらに強められることがわかる。また、Fig. 13 の 1500 ステップ付近までステップ数に比例して距離の 2 乗平均値が増加しているのは (2) 式に対応した結果である。逆に、Fig. 15 では (4) 式が示すように一定値に収束しており、環境に合わせて位置と大きさが適切に設定されたポテンシャルは、エージェントに対して期待された通りの引き込み効果を及ぼすことが示されている。

8.2 実験 (2)

さらに、ポテンシャルによる引き込み効果の傾向を調べるため、獲物がランダムに移動するとして、Table 3 の実験 (2) を行なった。このとき、同効果の評価尺度として、Fig. 12(a)~(c) で示した頻度の総和のハンター 1 体当たりの平均値 [AVERAGE FREQUENCY] を縦軸 (z 軸) に取った。試行回数は 20 回とした。

結果を Fig. 16, Fig. 17 に示す。Fig. 16 の実線はハンターがランダムに移動した場合であり、破線は真のポテンシャルを用いた場合の結果である。[HUNTERS], [PREY MOVE INTERVALS] に関わらず、ポテンシャルの効果により頻度平均値は約 2 桁圧縮されている。Fig. 17 は、実線が真のポテンシャルのみを用いた場合であり、破線が仮想ポテンシャルを追加した場合の効果を示す。この結果、[FIELD SIZE] と [SIGHT RANGE] が共に大きくなるにつれて両者の差が開き、この例では仮想ポテンシャルを導入することで最大 5 倍程度改善されることがわかった。この結果は、獲物を捕獲するためにハンターは仮想ポテンシャルによって他のハンターを視野内に引き込む必要があり、したがって、各ハンターの視野の重なりが小さく、かつ、フィールド面積に占めるハンターの視野面積の総和が大きい場合に、最も仮想ポテンシャルの効果が発揮されるとして理解される。

Fig. 12 Distribution of H1 and H2(exp.(1))

Fig. 13 $\langle x^2 + y^2 \rangle$ from initial position(exp.(1)) Fig. 14 $\langle x^2 + y^2 \rangle$ from initial position(exp.(1)) Fig. 15 $\langle x^2 + y^2 \rangle$ from initial position(exp.(1))

8.3 実験 (3)

Fig. 18 には、実験 (3) における獲物の捕獲数 [NUMBER OF CAPTURES] の増加割合を示す。図中、◇, +, □ は仮想ポテンシャルを用いた結果で、それぞれ PREY MOVE INTERVAL を 1, 5, 10 とした。各結果とも、シミュレーションステップ数に比例して捕獲数は増加している。一方、

Fig. 16 Average frequency(exp.(2))

Fig. 17 Average frequency(exp.(2))

×は仮想ポテンシャルを用いず PREY MOVE INTERVAL を 10 とした結果で、捕獲数の増加割合は◇と同程度であった。□と×の差がそのまま仮想ポテンシャルの有無に起因していることから、捕獲数を増加させる仮想ポテンシャルの効果は明らかである。

9. おわりに

本論文では、エージェントの動作をスカラーポテンシャル関数で制御し、環境に適したポテンシャル関数のパラメータをクラシファイア・システムを用いてエージェント自身に学習させる方法をベースに、さらにエージェントの相互制御性

Fig. 18 Number of captures(exp.(3))

を向上させるために、仮想的なスカラーポテンシャルパラメータを同時に獲得させる手法について検討を行なった。

実験の結果、マルチエージェント環境下で、(1) 複数エージェント間で協調的な空間探索を行なえる仮想ポテンシャルパラメータの自律的な学習・獲得が可能であること、(2) 仮想ポテンシャルの利用により空間探索の効率が向上することを示した。

今後は、仮想ポテンシャルの有効性に関するより定量的な解析を行なうとともに、エージェントの数を増したさらに複雑な問題への適用評価、および、最小限のエージェント間通信によるポテンシャルパラメータ情報の交換を取り入れた効率的な協調動作の学習方法について検討する予定である。

参考文献

- 1) 馬場, 久保田: 遺伝アルゴリズムを用いたロボットマニピュレータの軌道生成及び障害物回避, 日本ロボット学会誌, 11-2, 299/302(1993)
- 2) D.Goldberg: Genetic Algorithms in Search, Optimization, and Machine Learning, Addison Wesley (1989)
- 3) 三上貞芳: 強化学習のマルチエージェント系への適用: 人工知能学会誌, 12-6, 845/849(1997)
- 4) 宮崎, 小林: 離散マルコフ決定過程下での強化学習: 人工知能学会誌, 12-6, 811/821(1997)
- 5) 沼岡, 大沢, 長尾: マルチエージェントシステム, 共立出版, 26/36(1998)
- 6) 奥富, 森: ポテンシャル場を用いたロボットの動作決定, 日本ロボット学会誌, 1-3, 226/232(1983)
- 7) S.Russel and P.Norvig: Artificial Intelligence A Modern Approach, Prentice Hall (1995)
- 8) 佐藤圭祐: 極少点のないポテンシャル場を用いたロボットの動作計画, 日本ロボット学会誌, 11-5, 702/709(1993)
- 9) S.Sen and M.Sekaran: Multiagent coordination with learning classifier systems, in the Working Notes of the IJCAI-95 Workshop on Adaptation and Learning in Multiagent Systems, 84/89(1995)
- 10) 杉山, 秋下: 流体力学的ポテンシャルを用いた自律移動ロボットの経路計画~十字路における障害物回避~, 日本ロボット学会誌, 16-6, 839/844(1998)
- 11) 戸田, 齊藤, 久保, 橋爪: 統計物理学, 岩波書店 (1978)
- 12) 辻, P.G.Morasso, 重橋, 金子: 収束時間を調節可能な人工ポテンシャル法によるマニピュレータの動作計画, 日本ロボット学会誌, 13-2, 285/290(1995)
- 13) 山城, 山田, 遠藤: 組合せ問題に対するマルチエージェントア

プローチ-ポテンシャル構成法の検討-, 電気学会第8回インテリジェント・システム・シンポジウム論文集, 349/352(1998)

- 14) 安田, 橋山, 古橋, 大熊: 決定木を用いたサッカープレイヤー制御のための状態空間の自律的構成法に関する検討, 日本ソフトウェア科学会, コンピュータソフトウェア, 16-5, 14/24(1999)
- 15) 安田, 古橋, 橋山, 大熊: マルチエージェント系のポテンシャル関数による制御と強化学習の適用に関する検討, 計測自動制御学会論文集, 36-8, 698/706(2000)
- 16) 横山, 内田, 福島, 柿沼: 障害物回避問題におけるポテンシャル場の自律生成, 日本ロボット学会誌, 10-4, 510/519(1992)
- 17) R.Volpe and P.Khosla: Manipulator Control with Superquadric Artificial Potential Functions: Theory and Experiments, IEEE Trans. on Systems, Man and Cybernetics, 20-6 (1990)

[著者紹介]

安田 真

1985年関西学院大学理学部物理学科卒業。同年、沖電気工業(株)入社。92年より同社関西総合研究所勤務。97年より(財)名古屋産業科学研究所に出向。主として、創発システムとエージェント技術に関する研究に従事。情報処理学会、人工知能学会、日本ソフトウェア科学会各会員。

古橋 武 (正会員)

名古屋大学工学研究科電子情報学専攻助教授。1985年名古屋大学大学院工学研究科博士課程修了。同年、東芝入社。88年名古屋大学工学部助手。90年同助教授、現在に至る。工学博士。95年からインターネットを利用した国際会議を企画。主として、不完全・不確実情報利用技術、パターン・記号統合に関する研究に従事。

橋山 智訓 (正会員)

1996年名古屋大学大学院工学研究科博士課程後期課程修了。同年名古屋大学大学院工学研究科助手。97年(財)名古屋産業科学研究所主任研究員。2000年名古屋市立大学自然科学研究教育センター講師。博士(工学)。主として、ソフトコンピューティング手法を利用した、知的情報処理技術に関する研究に従事。日本ファジィ学会、電気学会、IEEE、人工知能学会、認知科学会各会員。

大熊 繁 (正会員)

1977年名古屋大学大学院博士課程満了。同年名古屋大学助手。79年同講師、84年同助教授、90年より同教授。工学博士。主として、ロボティクス、パワーエレクトロニクス、創発型ソフトコンピュータに関する研究に従事。電気学会、ロボット学会、IEEE、パワーエレクトロニクス研究会各会員。