

## サッカークライアントにおける 行動評価の遺伝アルゴリズムによる学習

遠藤和昭, 山口博之, 伊藤禎敏, 小谷善行

(東京農工大学 工学部 電子情報工学科)

### 1 はじめに

マルチエージェント(MA)における行動決定の条件は複数の場合が多く、それらは複雑に関係し合っている。複数の条件が行動の決定に関係があり、その条件が曖昧であるので、行動を一意に決定するのは困難である。

本稿ではMAの具体例としてシミュレーションサッカーのクライアントモデルを取り上げる。本稿のモデルでは、条件により行動を決定するのではなく、環境に対する行動を評価し他の行動と比較することにより行動を決定する。モデルでは、全ての行動を評価する汎用の行動評価関数は存在しない。また、行動評価関数は、行動に対して一対一で存在し内部アルゴリズムも別々である。

本稿では、関数が返す行動評価値により行動を決定するモデルの「行動評価値を遺伝アルゴリズムを用いて調整するモデル」について述べる。

### 2 モデル

#### 2.2 クライアントの行動決定モデル

本クライアントシステムでは、行動評価関数が個々の行動を評価し行動評価値を計算する。そして、その行動評価値を比較することにより行動を決定する。

E……環境

$A_i$ …… $i$ 番目の行動

$V_i(E)$ …… $E$ における $A_i$ の行動評価値

$\text{argmax}(V_i(E))$

……行動評価値が最大のものを選ぶ。

### 2.2 行動評価関数

#### 2.2.1 シュート

行動評価関数はシュートの場合では、ゴールの枠内に対する角度 $\phi$ により決定される。

C……定数

P……パラメータ

V……行動評価値

$V = C \times \phi \times P$

#### 2.2.2 パス

パスにおける行動評価値は、座標空間 $X$ へのパスとして評価する。

I……自分

B……球

EP……敵の選手

AP……味方の選手

EG……敵のゴール

AG……味方のゴール

$d(a,b)$ ……物体 $a$ と物体 $b$ の距離

$P_i$ …… $P_i$ は、パラメータ ( $i=1,2,\dots$ )

$C_i$ …… $C_i$ は、定数 ( $i=1,2,\dots$ )

V……行動評価値

$V = C_1 \times P_1 \times (1/d(B,I)) + C_2 \times P_2 \times (1/d(EP,I))$

$+ C_3 \times P_3 \times (1/d(AP,I)) + C_4 \times P_4 \times (1/d(EG,I))$

$+ C_5 \times P_5 \times (1/d(AG,I))$

Learning of behavior evaluation function in soccer client  
by genetic algorithm ,

Kazuaki Endo, Hiroyuki Yamaguchi, Sadatoshi Ito, Yoshiyuki  
Kotani

Tokyo University of Agriculture and Technology

## 2.3 パラメータ調整システム

### 2.3.1 行動評価値の調整

本モデルの行動評価関数は、行動に対して一対一で存在し内部アルゴリズムも別々である。そのため、各行動評価関数が返す行動評価値の値を調整することが必要である。

関数が返す行動評価値の値は、関数のアルゴリズムと行動評価関数内のパラメータに依存する。本システムでは行動評価値の調整をパラメータの調整により行う。

### 2.3.2 遺伝アルゴリズム

パラメータ調整システムは、遺伝アルゴリズムを用いて行動評価関数のパラメータを調整する。遺伝子はパラメータの列として構成され、パラメータ調整モデルは、遺伝子の淘汰、交叉、突然変異の三つの部分から成り立っている。遺伝子の淘汰は、遺伝子が周りの環境に適応した度合い、適応度を評価の目安として淘汰する。適応度は試合における得失点差をもとに算出される。遺伝子は一様交叉法により交叉される。

## 3 実験法

### 3.1 実験の方法

今回の実験では、東京工業大学の安藤友人氏が作成した andhill を対戦相手とし対戦させる。試合中の得失点をもとに遺伝子を進化させる。第一世代の遺伝子として、パラメータ  $X$  を、 $-1.00 < X < 1.00$  の間にランダムに設定された遺伝子と人間により作成されたパラメータ  $X$  に、 $-0.20 < \alpha < 0.20$  を加えた遺伝子の二つを使用する。遺伝子の総数は 100 個であり、淘汰時に 20 個が選択される。遺伝子 1 個に対する評価時間は、30 秒である。

## 3.2 実験環境

サッカーシミュレータは、Soccer Server Ver3.17 を使用する。シミュレータの設定は、RoboCup 公式設定のパラメータを部分的に変更し、視界を  $270^\circ$  視覚情報の得られる間隔を 0.3 秒とした。

## 3.3 実験の評価法

実験の評価は、収集した得失点をもとに世代ごとの、勝率、最高得点、最低得点、平均得点、最高失点、最低失点、平均失点、最高得失点差、最低得失点差、平均得失点差を求め、世代間で向上しているかどうかをもとに行う。

## 4 結果

表1 得失点の実験結果

得点/失点	第1世代	第4世代	第8世代	第12世代
ランダム	0/117	3/68	2/64	2/66
人間	9/64	14/56	19/35	15/32

## 5 考察

ランダム・人間の場合、共に世代を重ねるごとに失点を減らした。しかし、ランダムの場合と人間の場合とでは試合の内容が大きく異なっている。ランダムの場合、ゴール前に密集したために失点が減少したのに対して、人間の場合には攻撃とポジショニングにより失点を減少させている。

## 参考文献

- [1] 米沢保雄：遺伝的アルゴリズム, 森北出版,(1993).
- [2] 北野宏明 編：遺伝的アルゴリズム, 産業図書,(1993).
- [3] 伊庭斉志：遺伝的プログラミング, 電機大出版局,(1996).