

## RoboCup における遺伝的アルゴリズムを用いた AI の構築 Artificial Intelligence using Genetic Algorithm for RoboCup

水島 諒<sup>†</sup>      穴田 一<sup>‡</sup>  
Ryo Mizushima   Hajime Anada

### 1. 研究背景及び目的

近年、「ゲーム AI」の開発が盛んに行われている。そして、チェスや将棋、囲碁といったゲームにおいて AI が人間のチャンピオンに勝利するといった事も起きている。また、サッカーゲームでは RoboCup と呼ばれる世界大会が毎年行われている[1]。RoboCup とは 2050 年までにサッカーの世界一のチームに勝てる自律型ロボットチームを作ること为目标とした大会である。この RoboCup には 5 つのリーグがあり、リーグごとに異なる特徴がある。本研究では 5 つのリーグの中で各選手がそれぞれ思考し、人間のように戦術的なサッカーが行われている 2D リーグを扱う。

2D リーグは、移動可能範囲が広いことや、計算時間が制限されていること、複数人同士の対戦であることから、前述のチェスや将棋、囲碁といったゲームより難しいと考えられている。そして、2D リーグではチームを構成するエージェントの AI を経験則で構築することが多い。そこで、本研究では遺伝的アルゴリズムを用いて AI を構築し、その有効性を確認した。

### 2. 既存研究

秋山は RoboCup の 2D リーグ (高さの概念がない) で使用可能な agent2d(Ver 3.1.1) というチームモデルを公開している[2]。agent2d は、ボールを保持しているエージェントの意思決定を、Chain Action モデルを用いて行う。

#### 2.1 フィールドの座標軸

2D リーグでは、図 1 のようにフィールドの中央を原点とし、長辺方向を  $x$  軸、短辺方向を  $y$  軸とした直交座標系で表す。 $x$  軸は自分のゴール側が負の値とするため、図 1 の左のゴール自分のゴールだった場合は  $x$  軸の原点より左が負となる。

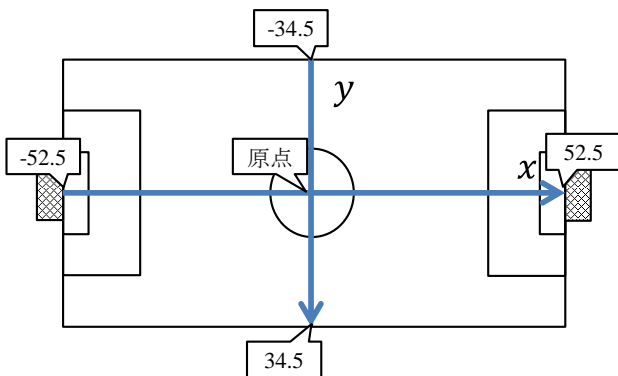


図 1 フィールドの座標軸

#### 2.2 Chain Action モデル

秋山は、自分も敵も同時に動くサッカーのエージェントの意思決定に、ゲーム木探索[3]を参考に Chain Action モデルを提案した。

Chain Action モデルでは、ボールを保持しているエージェントが、現在の全エージェントの配置を初期状態とし、パスとドリブルの行動によって今後の局面がどのように展開するかをツリー構造で表す。そして、その全ての展開を評価し、展開の中で最も評価が高い展開を選択する。選択のための評価値  $V$  は次式で表される。

$$V = x_b + \max\{0.0, 40.0 - dist_{bg}\} \quad (1)$$

ここで、 $x_b$  はボールの  $x$  座標、 $dist_{bg}$  はボールとゴールの距離を表す。第 1 項は敵のフィールドに行くほど値が高くなり、第 2 項はゴールに近いほど値が高くなる。このため、 $V$  の値は敵のゴールに向かうような展開ほど高くなり、選択されやすくなる。図 1 はある局面から考えられる可能な展開をツリー構造で表し、 $V$  の値を(1)式を用いて評価した結果である。○は行動の選択肢であり、中には行動の種類が示されている。末端の行動後の評価値  $V$  の値を□の中に表す。図の全ての展開の中で最良の展開は  $V$  の値が最も高い「ボール保持者は 7 番にパスをし、7 番が 8 番にパスをする」という展開である。そのため、ボールを保持しているエージェントは最良の展開になるように「7 番にパスをする」を選択する。

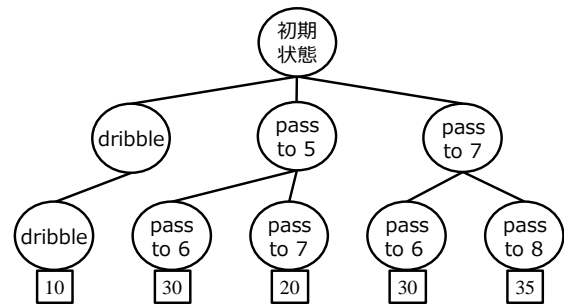


図 2 Chain Action モデルで作成したツリー構造の例

#### 2.3 既存研究の問題点

秋山の研究を含め、これまでの研究ではサッカーの専門的な知識を用いて、経験則で評価値  $V$  を構築していたという問題点があった。

### 3. 提案モデル

#### 3.1 意思決定を学習するための評価値

前述の問題を避けるため遺伝的アルゴリズムを用いる。評価値はパスやドリブルといった行動の種類によって異なると考え、行動  $a$  の評価値  $V(a)$  を次のように定義した。

$$V(a) = \alpha \times Y + \sum_{p=1}^3 (w_{a,p} \times U_p) \quad (2)$$

<sup>†</sup> 東京都市大学大学院 工学研究科

<sup>‡</sup> 東京都市大学 知識工学部

ここで、 $\alpha$ は全行動共通の評価項  $Y$  の重み、 $w_{a,p}$ は評価項  $U_p$ の重みを表す。

全行動共通の評価項  $Y$ には次のようなものを用意した。

$$Y = \begin{cases} 34.5 - |y_b|/34.5 & \text{if } (x_b > th_1) \\ 0.0 & \text{otherwise} \end{cases} \dots x \text{ 軸からの離れ度合い}$$

ここで、 $y_b$ は行動後のボールの $y$ 座標、 $th_1$ は $x$ 軸の閾値を表し、 $Y$ は 0 以上 1 以下の範囲をとる。2D リーグでは相手陣内のサイドに進む意味があまりないと考えたため、相手陣内のサイドより中央の方が高評価になるように調整した。

行動  $a$  の評価項  $U_p$ には次のようなものを用意した。

$$U_1 = (x_b + 54.5)/54.5 \quad \dots \text{相手陣地への攻め度合い}$$

$$U_2 = \max\{0.0, th_2 - dist_{bg}\}/th_2 \quad \dots \text{相手ゴールとの近さ}$$

$$U_3 = dist_{bno}/2.5 \quad \dots \text{フリー度合い}$$

ここで、 $x_b$ は行動後のボールの $x$ 座標、 $th_2$ は $dist_{bg}$ の閾値、 $dist_{bg}$ は行動後のボールとゴールとの距離、 $dist_{bno}$ は行動後のボールとボールから一番近い敵との距離を表す。 $U_3$ の分母にある 2.5 は試合の際によく見かける自選手と自選手から一番近い敵との距離である。そして $U_1, U_2$ は 0 以上 1 以下の範囲をとり、 $U_3$ はおおよそ 0 以上 1 以下の範囲をとるようにした。

今回はパス、ドリブル、ホールドという 3 つの行動ごとに評価の方法を変えるため、それぞれ重みを用意した。そのため (1) 式の第 1 項のパラメータは 2 個、第 2 項のパラメータ数は 12 個ある。よって、総パラメータ数は 14 個になる。

また、エージェントの意思決定方法はポジションごとに異なると考えたため、ポジションごとにパラメータを用意した。今回は 6 つのポジションに分けたため、総パラメータ数は 84 個になる。この 84 個のパラメータそれぞれを遺伝子とし、遺伝的アルゴリズムを用いてパラメータを変化させる。

### 3.2 遺伝的アルゴリズムの適用方法

チームの評価は agent2d と 20 回対戦させ、次式で表される適合度  $G$  を用いて行う。

$$G = \sum_{i=1}^{20} (P_i \times 100 + GD_i) \quad (3)$$

ここで、 $P_i$ は試合  $i$  の勝ち点を表し、勝利なら 3、引き分けなら 1、敗北なら 0 を表す。そして  $GD_i$  は得失点差である。

遺伝的アルゴリズムの流れは以下の通りである。

#### I. 初期世代

84 個のパラメータをランダムに決定したチームを 16 チーム作成。

#### II. 交叉

ランキング選択によって 2 チームを選択することを繰り返し 8 組作成する。そして、1 組ごとに 2 チームで一様交叉を行い、新たに 2 チームを作成することを繰り返し、16 チームを作成する。

#### III. 突然変異

新たに作成したチームのパラメータ全てに 5% の確率で突然変異を起こす。

#### IV. 選択

エリート戦略に基づき現在の世代における上位半分のチームを次の世代に残す。

#### V. 終了条件

10 世代経過を終了条件とし、II~IVを指定した世代まで繰り返す。

### 4. モデルの評価

本実験では agent2d と対戦させ、遺伝的アルゴリズムを用いて学習させた。使用したチーム数は 16、1 チームあたりのパラメータ数は 84、1 世代あたり agent2d と 20 回対戦、終了世代は 10 である。各パラメータの範囲と刻み幅は表 1 に記す。

表 1 各パラメータの範囲と刻み幅

パラメータ名	範囲	刻み幅
重み	[- 5.0, 5.0]	0.2
閾値	[-50.0, 50.0]	2.0
閾値	[ 0.0, 100.0]	2.0
重み	[- 5.0, 5.0]	0.2

図 3 は各世代における最良チームの適合度を示す。適合度は世代数が増すと高くなっていることが分かる。

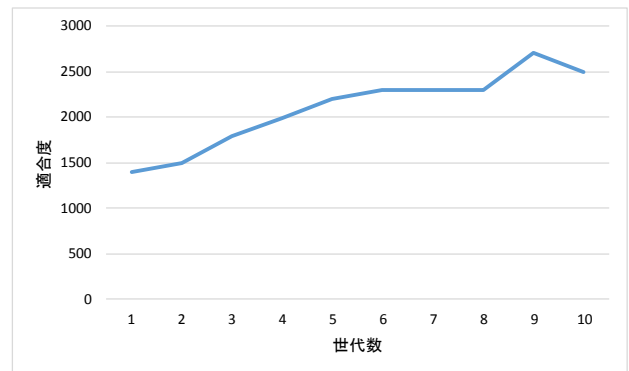


図 3 最良チームの適合度

このことから、サッカーの専門的な知識を必要とすることなく自己学習することによりエージェントの AI を強化することができたと考えられる。

### 5. 今後の方針

今回は終了世代を 10 世代としたが、適合度が収束するまで世代を繰り返さなければならない。また、本研究では (3) 式のような適合度を用いたが、よりよい適合度について検討していかなければならない。

この研究の方法では、対戦相手である agent2d には強くなるが、それ以外のチームには強くなるとは考えられない。そして、対戦相手をいくつか用意しランダムに対戦させたとしても、全てのチームに勝利できるチームが構築できるとは考えにくい。そのため、学習方法についても検討していかなければならない。

#### 参考文献

- [1] “ロボカップ日本委員会 RoboCup Japanese National Committee”, <http://www.robocup.or.jp/original/about.html>
- [2] 秋山 英久, “アクション連鎖探索によるオンライン戦術プランニング”, 人工知能学会研究会資料, SIG-Challenge-B101-6, pp.23-28 (2011).
- [3] 新谷虎松, 大園忠親, 白松俊: “知識システムの実相基礎-スライドで理解する人工知能技術-”, コロナ社, (2012).