6P-05

RoboCupにおける実数値GAを用いた意思決定の学習

水島　諒[†]　　穴田　一[†]

東京都市大学[†]

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．例えば，チェスや将棋，囲碁といったゲームが挙げられる．そして，これらのゲームにおいては 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，RoboCup と呼ばれる，自律型ロボットによるサッカーの世界大会が毎年行われている[1]．RoboCup とは，西暦 2050 年迄にサッカーの世界チャンピオンチームに勝てる，自律型ロボットのチームを作ることを目標とした大会である．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 つのリーグのうち，各選手がそれぞれ思考し，人間のような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，リアルタイムに計算し判断を下す必要があること，11 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 リーグにおけるこれまでの多くの研究では，チームを構成するエージェントの意思決定を経験則で構築している．そこで，本研究では実数値 GA を用いて意思決定を学習するアルゴリズムを構築し，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 リーグで用いるフィールドは，図 1 のように中央を原点とし，長辺方向を x 軸，短辺方向を y 軸とした直交座標系で，x 軸は自分のゴール側が負の値となっている．フィールドの大きさは長さ 105m，幅 68m である．

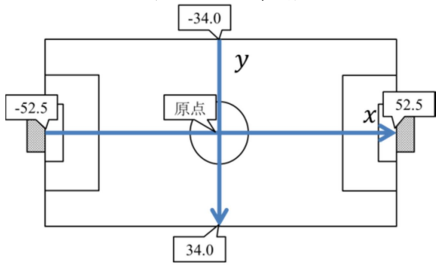

図 1. フィールドの座標系

Decision Making using Real-coded GA for RoboCup
† Ryo Mizushima, Hajime Anada, Tokyo City University

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，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 2 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．○は可能な行動を示し，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．

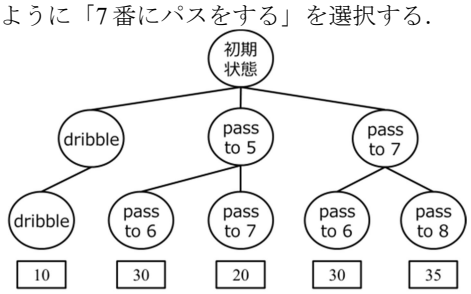

図 2. Chain Action モデルで作成したツリー構造の例

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 を構築している．そのため，評価値を改良する際は(1)式の評価項を検討し，重みや閾値を人為的に調整しなければならないが，それには限界がある．

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実数値 GA を用いて重みや閾値を学習させた．展開を評価するために用いる評価値は

パス，ドリブル，ホールドという 3 つの行動ごとに異なると考えたため，(1)式を変更し，行動 a の評価値 $V(a)$ を次のように定義した．

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

ここで，αは全行動共通の評価項 Y の重み，w_{ap}は行動 a の評価基準 p の評価項 U_{ap}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定義した．

$$Y = \begin{cases} 34.0 - |y_b|/34.0 & if(x_b > th_1) \\ 0.0 & otherwise \end{cases} \ldots x\text{軸からの離れ度合}$$

ここで，y_bは行動後のボールのy座標，th_1はx軸の閾値を表す．2D リーグでは高さの概念が無く，敵陣内のサイドに進んだとしても頭上を通すパスができない．そのため，行動の選択肢が狭まるだけであると考えたため，敵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 の評価基準 p の評価項 U_{ap} には次のようなものを定義した．

$U_{a1} = (x_b + 52.5)/105.0$ …相手陣地への攻め度合
$U_{a2} = \max\{0.0, th_{a2} - dist_{bg}\}/th_{a2}$ …相手ゴールとの近さ
$U_{a3} = dist_{b\_no}/5.0$ …フリー度合

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，$dist_{b_no}$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．そして，全ての評価項がおおよそ 0 以上 1 以下の範囲をとるように調整した．

(3)式の重みと閾値の数は全部で 14 個である．そして，エージェントの意思決定はポジションごとに異なると考え，6 つのポジションごとにパラメータを用意し，総パラメータ数は 84 個とした．

3.2 実数値 GA の適用方法

チームの評価は，過去の RoboCup の日本大会で使用されたチームから 4 チーム選び，その 4 チームとそれぞれ 20 回ずつ対戦させ，次式で表される適合度 G を用いて行う．

$$G = \sum_{i=1}^{4}\sum_{j=1}^{20}(P_{ij} + D_{ij}/100)/(4 \times 20) \quad (3)$$

ここで，P_{ij}はチーム i の j 試合目の試合の勝ち点を表し，勝利なら 3，引き分けなら 1，敗北なら 0 を表す．D_{ij}は得失点差である．チームの評価は計 80 試合の勝ち点の平均を用い，勝ち点の平均が同じ際は得失点差を用いる．

実数値 GA 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 15 チームと選手の基本的な動きを基にパラメータを大まかに決定した 1 チームを作成する．用意した 4 チームと対戦させ，(3)式を用いて適合度 G を求める．

II. 交叉
　ランキング選択[4]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 BLX-a[5]を用いて交叉を行い，新たに 2 チームを作成し，計 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，指定されたパラメータ範囲内のランダムな値に変更する．

IV. 適合度計算
　新たに作成した 16 チームの適合度 G を，それぞれ用意した 4 チームと対戦させ，(3)式を用いて求める．

V. 選択
　エリート戦略[4]に基づき現在の世代における上位半分の 16 チームを次の世代に残す．

VI. 終了条件
　10 世代経過を終了条件とし，II～V を設定した世代まで繰り返す．

4. モデルの評価

本研究では，10 世代になるまで学習を行った．図 3 は各世代における最良チームの適合度を示す．適合度は世代を経るにつれて高くなっていることが分かる．

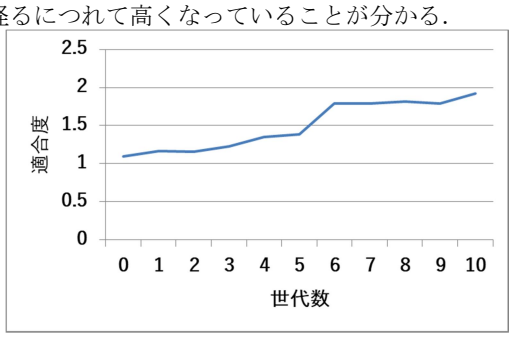

図 3. 各世代における最良チームの適合度

この図から，サッカーの専門的な知識を必要とすることなく，エージェントの意思決定を自己学習することができたと考えられる．しかし，適合度が収束しているとは言いがたい．そのため，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．そして，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，大囿忠親，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.
[4] 佐藤 浩，小野 功，小林 重信，“遺伝的アルゴリズムにおける世代交代モデルの提案と評価”，人工知能学会誌，Vol.12，No.5，pp.734-743，1997.
[5] Eshelman, L.J，“Real-Coded Genetic Algorithm and Interval Schema ta”，Foundations of Genetic Algorithm 2，pp.187-202，1993.