サッカーエージェントによる敵の行動モデルの模倣学習

山岸拓海[1]　五十嵐治一[1]

概要：RoboCup サッカーシミュレーション 2D リーグでよく用いられるサンプルチーム agent2d は，行動決定の際にチェーンアクションを用いる．チェーンアクションはアクション連鎖探索フレームワークであり，探索木と評価関数によって次の行動を決定する．本研究では敵プレイヤがチェーンアクションを使用していると仮定したうえで，敵プレイヤが選択した行動を正解行動とする教師あり学習を行い，敵プレイヤの行動モデルを学習した．agent2d の行動モデルを学習したところ，ボール保持時のパスやドリブルなどの行動一致率は 63.4%~95.8%と役割によっては高い一致率を示したが，ボール非保持時の移動行動の一致率は 26.0%~49.5%にとどまった．

Imitation Learning of Opponents' Action Models by Soccer Agents

TAKUMI YAMAGISHI[1]　HARUKAZU IGARASHI[1]

Abstract: Agent2d is a sample team that is often used in the RoboCup Soccer Simulation 2D League. The players of agent2d use "chain action" algorithm to decide a next action. It is a framework to search action sequences using search trees and state evaluation functions. In this paper, we made players learn their opponents' action models by imitation learning assuming that the opponents use the framework. After the experiments in games against agent2d, the rates of agreement with the actions like passes or dribbles were improved to 63.4%~95.8%. By contrast, those rates in receivers' moving actions remained at low rates of 26.0%~49.5%.

1. はじめに

「RoboCup サッカーシミュレーション 2D リーグ」はコンピュータ上の仮想フィールドにおいて，敵味方合わせて 22 人のプレイヤがサッカーを行うリーグである．各プレイヤが自律していることから，この競技はマルチエージェントシステムのテストベッドとして利用されている[1].

このリーグでよく用いられるサンプルチームとして「agent2d」[2]がある．agent2d のボール保持者は，探索木と評価関数を用いた「チェーンアクション」[3]によって行動を決定する．しかし，agent2d にデフォルトで備わっている評価関数は，ボールと敵ゴールの距離のみを評価するシンプルな関数に過ぎない．そこで谷川らは複数の評価項からなる評価関数を考案し，ボールの位置の変化量を報酬とする強化学習を行った[4][5]．その後，田川ら[6][7]や大内ら[8]も複数の評価項からなる評価関数を用いて，人間の主観評価による強化学習を行った．田川ら[7]や大内ら[8]の研究では，どちらも agent2d に対して谷川ら[5]の結果を上回る勝率を記録したが，プレイヤが観戦者の望む行動を取らなければ正の報酬を与えて方策を強化することができない．そこで，著者らは強化学習ではなく，人間が与えた正解行動を教師とする教師あり学習を適用することで，対 agent2d の勝率において 85.6%を記録した[9]．この数字は上記の強化学習による対戦成績を上回っている[a].

しかし，文献[9]で行ったような多数の局面に対して正解行動を与える作業は人間の負担がかなり大きい．そこで，本研究では敵プレイヤがチェーンアクションを用いて行動決定を行っていると仮定し，その枠組みで敵プレイヤの行動モデルを学習するシステムを開発した．もし，強いチームの行動モデルを精度良く近似できれば，その後に強化学習を行うことで模倣対象のチームを超える強さのチームになる可能性もある．それは本研究の最終的な目的ではあるが，まず前段階として対戦中の敵プレイヤの行動モデルを模倣する模倣学習の方式を考案し，実際に agent2d を対戦相手として学習実験を行った．

2. RoboCup について

2.1 RoboCup サッカーシミュレーション 2D リーグ

RoboCup は，ロボット工学と人工知能の融合発展のために提唱された，自律型移動ロボットによる競技会である．サッカー，インダストリアル，レスキュー，@ホーム，ジュニアの 5 つの部門から構成され，さらにそれぞれが複数のリーグから構成されている．本研究で扱うサッカーシミュレーション 2D リーグは，コンピュータ上の仮想二次元フィールドでサッカーエージェントがサッカーの試合を行うリーグである．シミュレーション 2D リーグのシミュレータはサーバ・クライアント方式を用いている．各エージェントはサーバから環境情報を受け取り，それに応じて行動決定し，選択した行動を制御コマンドとしてサーバへ送信する．

2.2 agent2d

本研究で使用した agent2d は秋山英久氏らによって提供された， シミュレーション 2D リーグ用のサンプルチーム

1 芝浦工業大学
Shibaura Institute of Technology

a) 谷川ら[5]では約 22%，田川ら[7]では約 43%， 大内ら[8]では約 73%であった．

プログラムである．基本的な戦術や行動のみが実装されている．最新バージョンである agent2d-3.1.1 では行動決定の際に，チェーンアクションと呼ばれるアクション連鎖探索フレームワークを用いる[3].

3. チェーンアクション

3.1 ボール保持者の行動決定

シミュレーション 2D リーグでは，プレイヤ間のコミュニケーションはサーバを介さなければ行えない．それにより協調行動の実現が難しく，プレイヤ同士が連携するためには複数のプレイヤによる行動の連鎖をプランニングする必要がある．チェーンアクションは複数プレイヤによる行動連鎖をオンラインで生成・評価するフレームワークであり，本研究でも有用であると考えたためこれを使用する．

チェーンアクションでは，パスやドリブルなどのある程度抽象的な行動を枝，行動後の予測局面をノードとする探索木を生成する．各ノードは評価関数によって点数が与えられる．その後，ボール保持者は点数が最高のノード（局面）を最良優先探索し，そのノードへ至る枝（行動）を選択する．図 1 にチェーンアクションの例を示す．図 1 ではa，b，cが行動，s_0~s_8が局面（s_0は現局面），数値がノードの評価値を表している．この場合はs_7の評価値が最も高いため，プレイヤは次の行動としてbを選択する．なお，今回は計算を簡単にするために一段の探索木のみを考えた．

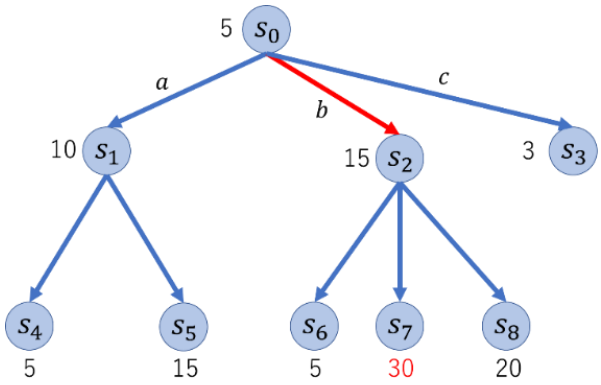

図 1　チェーンアクションを用いた探索木の例

Figure 1　A search tree with chain actions.

3.2 ボール非保持者の行動決定

agent2d のボール非保持者はチェーンアクションを用いず，ボールの位置のみに依存して予め定められた場所へ移動する．したがって，敵プレイヤの位置や行動に応じた柔軟な位置取りをすることができない．

それに対し，大内らはボール非保持者の行動決定（移動先の決定）にチェーンアクションを適用した[8]．本研究でもこの決定方式を用いる．

b) もし，学習エージェントが正解行動を観測やコーチエージェントからの通信などで入手できれば，試合中にオンラインで学習することも可能である．

4. 評価関数の学習

4.1 本研究における基本方針

マルチエージェントシステムではエージェントは自律的に行動を決定する必要がある．本研究のエージェントは 3 章で述べたチェーンアクションモデルに従うことを前提としている．したがって，行動モデルの学習は探索木で用いる評価関数の学習に帰着する．次節で述べるように，本研究ではヒューリスティクスの線形和からなる評価関数を，方策勾配を用いた教師あり学習[10]によって学習する．行動や状態に関する特徴量の線形和を用いる理由は，ニューラルネットワークやディープニューラルネットワークと比べると精度は劣るが，計算が速く 0.1 秒で局面が変化するシミュレーション 2D リーグには適しているからである．

また，学習エージェントは対戦中に敵チームの特定のプレイヤの行動を予測し，敵プレイヤが実際に行った行動（正解行動）との誤差を減少させるように評価関数を学習する．パスの始点や終点など，正解行動は試合後にログファイルから判定できるので，それらを使用した[b].

4.2 評価関数

agent2d における評価関数は，ボールと敵ゴールとの距離のみを評価する関数であった．それに対し田川ら[6][7]や大内ら[8]は，ヒューリスティクスに基づく評価関数を使用した．また，著者らは行動そのものを評価するために，行動の評価項を追加した[9]．本研究では，ボール保持者は文献[9]で提案された(1)の評価関数を，ホール非保持者は大内ら[8]により提案された(2)の評価関数を用いる．ただし，後述する各評価項の評価内容は一部変更している．

$$E(a,s;\omega)=\sum_{i=1}^{3}\omega_i U_i(a,s)+\omega_4 U_4(s) \quad (1)$$

$$E(s;\omega)=\sum_{j=1}^{4}\omega_j U_j(s) \quad (2)$$

aは行動（パス，ドリブルなど），sは行動後の予測局面，ωは重み係数，Uはヒューリスティクスである．(1)の第一項は行動の評価項，第二項は局面の評価項である．なお，Uは$[0,10]$の区間内の値を取るよう正規化されている．各評価項の概要を以下に示す．詳細は付録に記した．

表 1　ボール保持者の評価項

Table 1　Terms in the evaluation functions of passers.

評価項	評価内容
$U_1(a,s)$	現在地からどれだけ前に進むか
$U_2(a,s)$	現在地からどれだけ外に移動するか
$U_3(a,s)$	現在地からどれだけ内に移動するか
$U_4(s)$	敵ゴールとの距離

表 2　ボール非保持者の評価項

Table 2　Terms in the evaluation functions of receivers.

評価項	評価内容
$U_1(s)$	一番近い味方プレイヤとの距離
$U_2(s)$	一番近い敵プレイヤとの距離
$U_3(s)$	オフサイドラインとの距離
$U_4(s)$	敵ゴールとの距離

4.3 学習則

本研究では評価関数の学習に方策勾配を用いた教師あり学習[10]を使用した．この学習法は，方策勾配法[11]と同様に方策の勾配を用いる学習法であり，コンピュータ将棋で提案された学習法である．まず，学習システムの方策をボルツマン分布で定義する．

$$\pi(a|s;\omega) \equiv \frac{e^{E(a,s;\omega)/T}}{\sum_{x\in A(s)} e^{E(x,s;\omega)/T}} \tag{3}$$

ただし，Tは温度パラメータである．次に，正解の方策$\pi^*(a|s)$と学習システムの方策$\pi(a|s;\omega)$の誤差を表す関数を次のように定義する．

$$\delta_{KLD}(\sigma;\pi^*,\pi) \equiv \sum_{s\in\sigma}\sum_{a\in A(s)} \pi^*(a|s)\, ln\frac{\pi^*(a|s)}{\pi(a|s;\omega)} \tag{4}$$

σは学習に使用する局面の集合，$A(s)$は局面sにおいて推測される行動候補の集合，$\delta_{KLD}(\sigma;\pi^*,\pi)$（≥0）は正解の方策と学習システムの方策の距離を表すカルバック・ライブラー情報量（Kullback-Leibler divergence）である．$\delta_{KLD}(\sigma;\pi^*,\pi)$を勾配法で最小化すると学習則は，

$$\Delta\omega = \varepsilon\sum_{s\in\sigma}\sum_{a\in A(s)} \pi^*(a|s)\nabla_\omega\, ln\,\pi(a|s;\omega) \tag{5}$$

と表される．(5)に(3)を代入すると，

$$\Delta\omega = \frac{\varepsilon}{T}\sum_{s\in\sigma}\sum_{a\in A(s)} \pi^*(a|s)\cdot \left[\nabla_\omega E(a,s;\omega) - \sum_{x\in A(s)} \pi(x|s;\omega)\nabla_\omega E(x,s;\omega)\right] \tag{6}$$

$$= \frac{\varepsilon}{T}\cdot\sum_{s\in\sigma}\sum_{a\in A(s)} [\pi^*(a|s) - \pi(a|s;\omega)]\,\nabla_\omega E(a,s;\omega) \tag{7}$$

となる．さらに，本研究では予測局面のボールの座標によって評価関数を使い分けた[9]．5 章の実験ではフィールドの x 座標[c]により 2 つに使い分けている（x>47.0 か x≤47.0 か）．その場合の学習則は，

$$\Delta\omega^{(k)} = \frac{\varepsilon}{T}\sum_{s\in\sigma}\sum_{a\in A^{(k)}(s)} \left[\pi^*(a|s) - \pi\left(a|s;\omega^{(k)}\right)\right] \cdot\nabla_{\omega^{(k)}}E^{(k)}(a,s;\omega^{(k)}) \tag{8}$$

となる．$A^{(k)}(s)$は，局面sにおいて行動するとボール位置が領域kに入るような行動の集合である．また，本研究では 1 試合終了後にミニバッチ学習を行う．したがって，1 試合終了するごとに以下の学習則で重みを更新する．

$$\Delta\omega^{(k)} = \frac{\varepsilon}{nT}\sum_{s\in\sigma}\sum_{a\in A^{(k)}(s)} \left[\pi^*(a|s) - \pi\left(a|s;\omega^{(k)}\right)\right] \cdot\nabla_{\omega^{(k)}}E^{(k)}(a,s;\omega^{(k)}) \tag{9}$$

ただし，nは敵プレイヤの 1 試合中の行動回数である．これはσの要素数に等しい．

4.4 学習用の訓練データの作成

入倉らは agent2d のチェーンアクションを敵ボール保持者に適用し，敵ボール保持者の行動候補を守備プレイヤに推測させた[12]．本研究でも学習エージェントにこの推測機能を持たせて，敵プレイヤのパス先などの行動を推測させる．一般に敵チームの行動モデルは未知であるが，チェーンアクションの枠組みで敵プレイヤの行動モデルを近似，解釈し，評価関数に反映させる．加えて，本研究では敵ボール非保持者の行動候補も文献[8]，[12]を参考に，チェーンアクションの枠組みで推測できるようにした．これらを実装した学習チームと模倣対象のチームを対戦させ，学習エージェントが推測した敵プレイヤの行動選択の確率分布$\pi(a|s;\omega)$と局面sにおける敵プレイヤの行動候補のデータを大量に集めた．

一方，局面sにおける敵プレイヤの正解行動a^*はログファイルから抽出する．ログファイルには RCG ファイルと RCL ファイルの二種類がある．RCG ファイルにはある時刻のボールや選手の位置等が書かれている．また，RCL ファイルには各選手がある時刻において使用した行動コマンド等が書かれている．これらのファイルを照らし合わせ，正解行動a^*を抽出する．なお，ボール保持者の場合は kick コマンドから次の kick コマンドまでを，ボール非保持者の場合は turn コマンドから次の turn コマンドまでを 1 行動と定義した．

上記のように対戦により収集した$\pi(a|s;\omega)$，(s,a^*)と行動候補のデータ集合を，以下では「訓練データ」と称する．これらのデータを用いて，エージェントが敵プレイヤの行動について予測した確率分布$\pi(a|s;\omega)$において，正解行動a^*の確率が最も高くなるように評価関数を学習する[d].

c) 公式シミュレータで用いられているサッカーフィールドについては付録に示した．

d) 学習エージェントは模倣すべき敵プレイヤを予め指定されており，その敵プレイヤに関する訓練データを用いて学習する．

4.5 学習システム

文献[9]の学習システムを一部利用し，敵プレイヤの行動モデルを学習するシステムを開発した（図 2，図 3）．

対戦を行うとサッカーサーバはログファイルを生成する．学習エージェントは敵プレイヤを観測し，敵プレイヤのチェーンアクションデータを生成する．対戦後に教師あり学習プログラムを実行すると，ログファイルと敵プレイヤのチェーンアクションデータから訓練データが作られ，式(9)によって学習エージェントの重みを更新する．

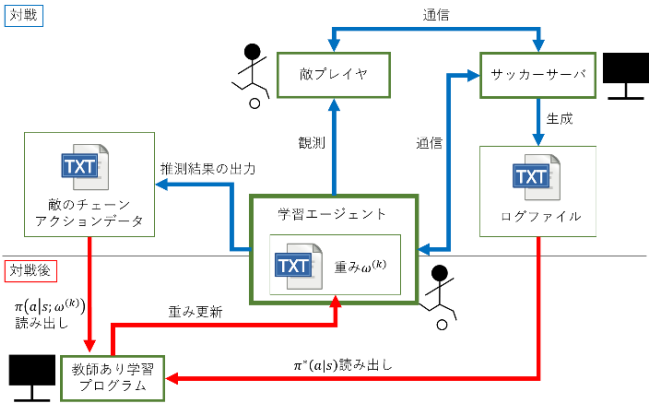

図 2　学習システム

Figure 2　Learning system.

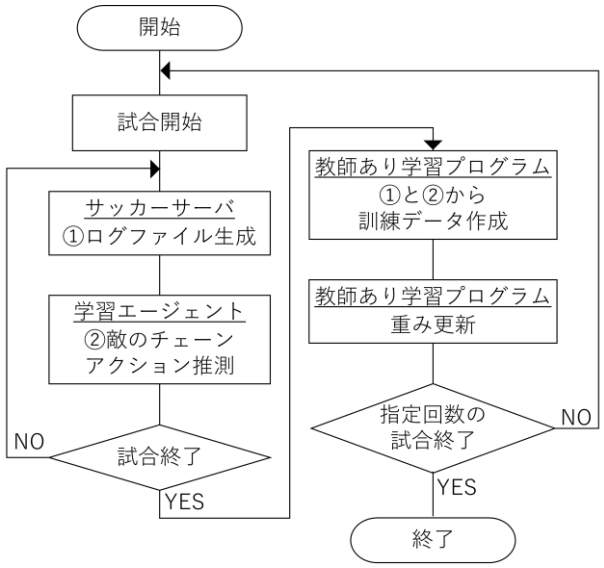

図 3　学習の流れ

Figure 3　Flow of the learning.

5. 実験

5.1 学習実験と評価実験の設定条件

学習実験の条件を以下に示す．

- 対戦相手（教師となるチーム）：agent2d-3.1.1
- 学習試合数：100 試合（延長戦・PK 戦はなし）
- 学習率$\varepsilon = 0.3$
- 温度$T = 10$
- 重みの初期値：すべて 1
- 学習エージェント：ディフェンシブハーフ（DH）1 名，オフェンシブハーフ（OH）2 名，サイドフォワード（SF）2 名，センターフォワード（CF）1 名，計 6 名．それぞれ予め指定した同役割の敵プレイヤの行動を学習する．
- 重み：以下の条件によって 16 種類の重みセットを使い分ける[9]．
  - 役割（DH,OH,SF,CF の 4 パターン）
  - ボール保持者か非保持者か（2 パターン）
  - 予測局面のボールの座標が x>47.0 か x≤47.0 か（2 パターン）

また，評価実験では，agent2d-3.1.1 を対戦相手として 500 試合（延長戦・PK 戦はなし）行い，温度$T \to 0$とした．

5.2 評価実験における一致率の結果

まず，評価実験 500 試合の中から 10 試合を無作為に抽出し，敵プレイヤの行動を推測，正解行動との一致率を計測した．具体的には，試合のログファイルから読み出した敵プレイヤの行動と学習システムが推測した行動の終点（パス先や移動先など）が一致しているかを調べた．ただし，パサーが狙った場所でレシーバがボールを受け取るとは限らないなどの「ずれ」を考慮し，誤差を 5m まで許容するものとした．表 3 と表 4 に 10 試合中のすべての行動について，agent2d の行動との一致率を示した．

表 3　agent2d の行動との一致率（ボール保持者）

Table 3　Rates of agreement with agent2d (passer).

チーム	DH	OH	SF	CF
学習前	4.9%	22.6%	57.0%	17.0%
学習後	89.6%	63.4%	95.8%	84.8%

表 4　agent2d の行動との一致率（ボール非保持者）

Table 4　Rates of agreement with agent2d (receiver).

チーム	DH	OH	SF	CF
学習前	49.1%	20.7%	3.9%	18.1%
学習後	46.5%	26.0%	27.3%	49.5%

表 3 では，どの役割でも 63.4%~95.8%と学習後の行動の一致率が大幅に上昇しており，agent2d のボール保持者の行動をかなり学習できていることがわかる．一方で，表 4 では DH と OH の行動の一致率は変化が少なく，SF，CF では上昇しているものの，いずれも 50%を切っている．原因としては，agent2d のボール非保持者はボールの位置のみに依存して予め定められた場所へ移動するが，ボールが動くと移動の途中でも方向転換をするため，最初に意図した移動行動とは異なる行動を学習したためと考えられる．

5.3 評価実験における対戦結果

評価実験 500 試合の対戦成績を表 5 に示す.

表 5　agent2d との対戦成績
Table 5　Play results against agent2d.

チーム	勝率	平均得点	平均失点
学習前	14.7%	0.88	2.32
学習後	64.1%	3.34	2.54

agent2d の行動を学習したチームの勝率は 64.1%であった. しかしながら，教師を忠実に模倣すれば勝率は 50%に近づくはずである. このような結果になった理由の一つは，前節でも述べたようにボール非保持者が正しく学習できていなかったためだと考えた.

そこで，学習後のチームにおいて，ボール非保持者の行動モデルをデフォルト（agent2d）のモデルに置き換えて agent2d と 500 試合対戦させたところ，勝率は 52.4%となり，予想通り勝率が 50%に近づいた（表 6）.

表 6　ボール非保持者を agent2d に置き換えた場合の結果
Table 6　Results which receivers are replaced with agent2d.

チーム	勝率	平均得点	平均失点
学習後	52.4%	2.29	2.14

6. 今後の展望

本研究では，チェーンアクションを利用し，対戦中の敵プレイヤの行動決定モデルを模倣する学習方式を提案した. 実際に，agent2d を対戦相手として学習実験を行った結果，ボール保持者のパスやドリブルなどの行動一致率はどの役割のプレイヤでも大きく上昇したものの，非保持者の移動行動の一致率はいずれも 50%に届かなかった.

さらなる模倣の精度向上のためには，評価関数の評価項の追加やディープニューラルネットワークなどの大量のパラメータを含む非線形関数の利用が考えられる. また，今回はチェーンアクションの探索木を一段にとどめたが，二段にして連続した行動の組（協調的プレー）を学習することも考えられる.

また，agent2d ではなく，RoboCup 世界大会の上位強豪チームを模倣学習することや，模倣学習の後に強化学習を行うことで教師チーム以上の強さになるかどうかを実験する予定である.

参考文献

[1] Noda Itsuki, Matsubara Hitoshi: Soccer Server and Researches on Multi-Agent Systems, Proceedings of the IROS-96 Workshop on RoboCup, pp.1-7, 1996.

[2] Akiyama Hidehisa, Nakashima Tomoharu: Helios Base: An Open Source Package for the Robocup Soccer 2D Simulation, RocoCup 2013: Robot World Cup XVII, pp.528-535, 2013.

[3] 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.

[4] 谷川俊策，五十嵐治一，石原聖司: RoboCup サッカーシミュレーションリーグ 2D における局面評価関数の学習，ゲームプログラミングワークショップ 2013 論文集, pp106-109, 2013.

[5] 谷川俊策: RoboCup サッカーシミュレーションリーグ 2D における局面評価関数の設計と学習，芝浦工業大学修士論文, 2014.

[6] 田川諒，五十嵐治一: サッカーエージェントにおける局面評価関数の強化学習，ゲームプログラミングワークショップ 2015 論文集, pp.77-83, 2015.

[7] 田川諒: サッカーエージェントにおける局面評価関数の強化学習，芝浦工業大学修士論文, 2016.

[8] 大内斉，五十嵐治一: 局面評価関数を用いたエージェントの移動先決定，ゲームプログラミングワークショップ 2016 論文集, pp.49-56, 2016.

[9] 山岸拓海，五十嵐治一，山岸準，入倉雅春: サッカーエージェントの攻撃時における評価関数:方策勾配を用いた教師あり学習，第 34 回ファジイシステムシンポジウム講演論文集, pp682-687, 2018.

[10] 五十嵐治一，森岡祐一，山本一将: プロ棋士の棋譜データベースを用いない局面評価関数の学習法についての考察，情報処理学会研究報告, Vol.2015-GI-34, No4, pp.1-8, 2015.

[11] Williams, Ronald J: Simple Statistical Gradient-Following Algorithms for Connectionist Reinforcement Learning, Machine learning, Vol.8, pp.229-256, 1992.

[12] 入倉雅春，五十嵐治一，山岸準，山岸拓海: RoboCup サッカーシミュレーションリーグ 2D における守備力の向上，情報処理学会研究報告, Vol.2019-GI-41, No.17, pp.1-5, 2019.

付録

付録 A.1 公式シミュレータにおけるフィールド座標系

シミュレーション 2D リーグの公式シミュレータ（rcssserver）で使用されるフィールドの座標系を図 4 に示す. 左サイドのチームのプレイヤが持つ座標系はフィールドの座標系と同一であるが，右サイドのチームのプレイヤが持つ座標系はすべて反転して使用する.

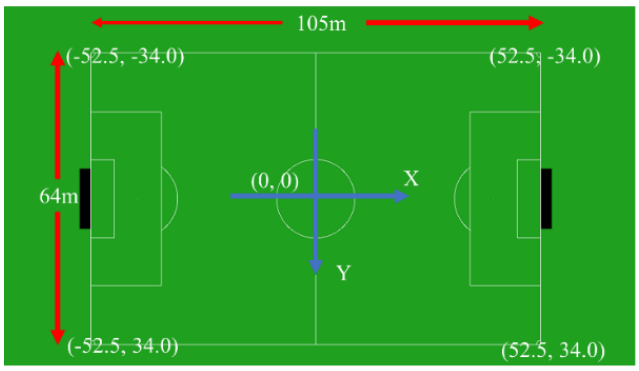

図 4　フィールドの座標系
Figure 4　The coordinate system of the field.

付録 A.2 ボール保持者の評価関数

ボール保持者の評価関数の各評価項について説明する．各評価項では，行動aとボールやプレイヤの配置の状態sから定まる特徴量αを計算し，各評価値を計算する．以下では，その特徴量αと評価項の具体的な関数を示す．

- $U_1(a,s)$：ボールが移動した後の x 座標と現在の x 座標の差を評価する．θは 10.0，τは 2.0 である．

$$U_1(a,s) = 10 \cdot \frac{1}{1+e^{-(\alpha-\theta)/\tau}} \quad (10)$$

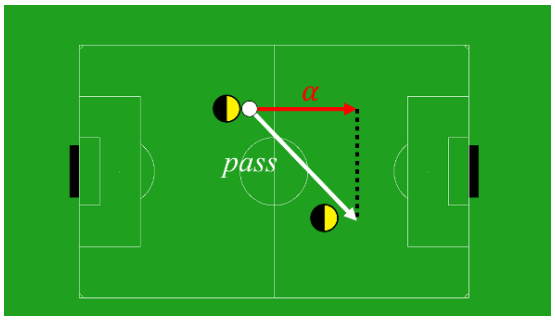

図 5　$U_1(a,s)$におけるα
Figure 5　α for $U_1(a,s)$.

- $U_2(a,s)$：現在地から y 座標の絶対値が大きくなる方向へ，どれだけボールが移動するかを評価する．θは 10.0，τは 2.0 である．

$$U_2(a,s) = 10 \cdot \frac{1}{1+e^{-(\alpha-\theta)/\tau}} \quad (11)$$

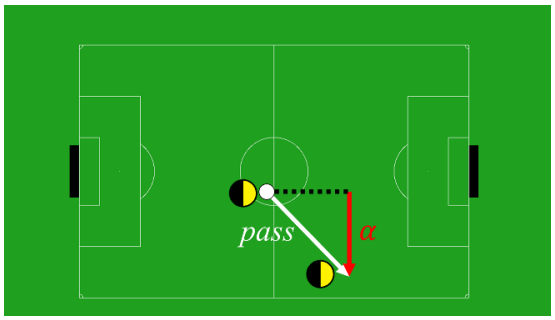

図 6　$U_2(a,s)$におけるα
Figure 6　α for $U_2(a,s)$.

- $U_3(a,s)$：現在地から y 座標の絶対値が小さくなる方向へ，どれだけボールが移動するかを評価する．θは 10.0，τは 2.0 である．

$$U_3(a,s) = 10 \cdot \frac{1}{1+e^{-(\alpha-\theta)/\tau}} \quad (12)$$

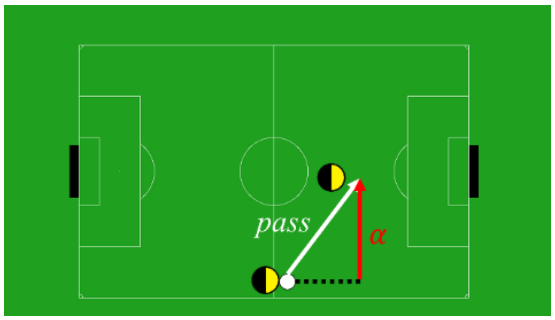

図 7　$U_3(a,s)$におけるα
Figure 7　α for $U_3(a,s)$.

- $U_4(s)$：ボールが移動した後の位置から敵ゴール中心までの距離を評価する．θは 43.0，τは 10.0 である．

$$U_4(s) = 10 \cdot \frac{1}{1+e^{(\alpha-\theta)/\tau}} \quad (13)$$

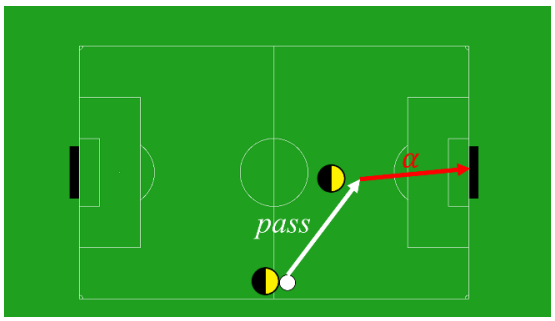

図 8　$U_4(s)$におけるα
Figure 8　α for $U_4(s)$.

付録 A.3 ボール非保持者の評価関数

ボール非保持者の評価関数の各評価項を説明する．

- $U_1(s)$：移動後の自身とそこに一番近い味方プレイヤとの距離を評価する．θは 7.0，τは 2.0 である．

$$U_1(s) = 10 \cdot \frac{1}{1+e^{-(\alpha-\theta)/\tau}} \quad (14)$$

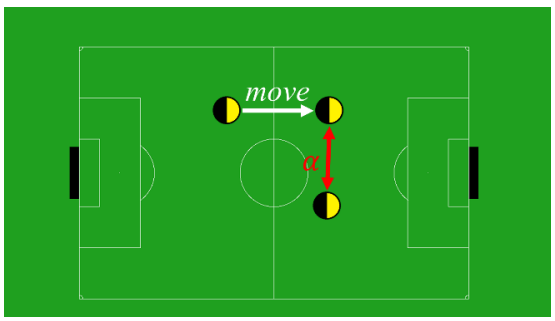

図 9　$U_1(s)$におけるα
Figure 9　α for $U_1(s)$.

- $U_2(s)$：移動後の自身とそこに一番近い敵プレイヤとの距離を評価する．θは 7.0，τは 2.0 である．

$$U_2(s) = 10 \cdot \frac{1}{1 + e^{-(\alpha-\theta)/\tau}} \quad (15)$$

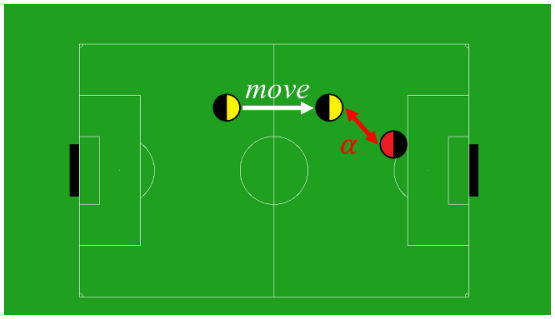

図 10　$U_2(s)$におけるα

Figure 10　α for $U_2(s)$.

- $U_3(s)$：移動後の自身からオフサイドラインまでの距離を評価する．θは 8.0，τは 2.0 である．

$$U_3(s) = 10 \cdot \frac{1}{1 + e^{(\alpha-\theta)/\tau}} \quad (16)$$

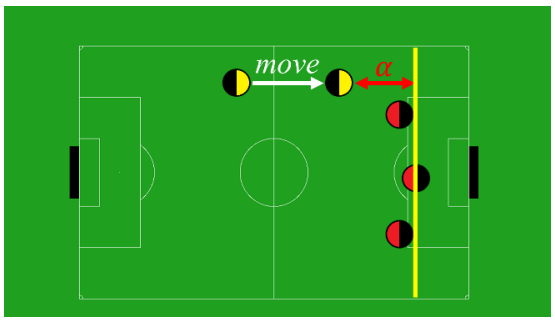

図 11　$U_3(s)$におけるα

Figure 11　α for $U_3(s)$.

- $U_4(s)$：移動後の自身から敵ゴール中心までの距離を評価する．θは 43.0，τは 10.0 である．

$$U_4(s) = 10 \cdot \frac{1}{1 + e^{(\alpha-\theta)/\tau}} \quad (17)$$

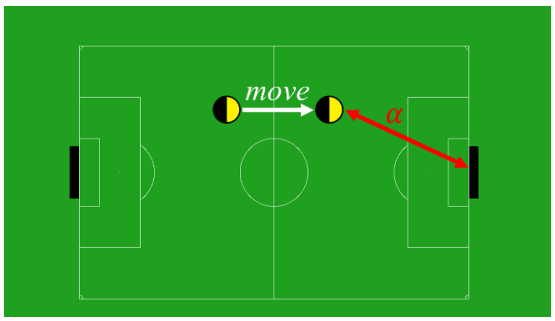

図 12　$U_4(s)$におけるα

Figure 12　α for $U_4(s)$.