

RoboCup サッカーにおける SIRMs ファジィシステムを用いたログからの行動評価

Evaluation of Actions using SIRMs for RoboCup Soccer

三船 哲史^{*1}, 中島 智晴^{*1}, 秋山 英久^{*2}, 関 宏理^{*3}

Satoshi MIFUNE^{*1}, Tomoharu NAKASHIMA^{*1}, Hidehisa AKIYAMA^{*2}, Hirosato SEKI^{*3}

大阪府立大学^{*1}, 福岡大学^{*2}, 関西学院大学^{*3}

Osaka Prefecture University^{*1}, Fukuoka University^{*2}, Kwansai Gakuin University^{*3}

satoshi.mifune@cs.osakafu-u.ac.jp, tomoharu.nakashima@kis.osakafu-u.ac.jp

akym@fukuoka-u.ac.jp, seki@kwansai.ac.jp

Abstract

Evaluation of action chains using Single-Input Rule-Modules (SIRMs) fuzzy models is studied in this paper. An action chain is a series of elemental actions such as pass, dribble, and shoot. The aim of using SIRMs fuzzy models is to learn good strategies from existing strong teams. In the training process, training patterns for the SIRMs fuzzy model are generated from game logs that are produced after the games of the target team. The results by the numerical experiments show that the proposed method improves the performance of our team.

1 はじめに

ロボット工学と人工知能の領域横断型研究プロジェクトとして RoboCup が知られている。RoboCup には様々なリーグが存在しており、それぞれの特徴を生かした研究、開発が行われている。RoboCup サッカーでは、競技で勝利することが重要視され、ただ単に勝利するだけではなく、賢く安定して勝利することが望まれる。そのためには、チーム全体でどのような戦術を取るかが重要である。高度な戦術を取るためには、プレイヤーが状況に応じた的確な行動選択を行うことが必要である。本論文では、RoboCup サッカー 2D シミュレーションにおいて、単一入力ルール群 (Single Input Rule Modules : SIRMs) ファジィシステム [1] を用いたログからの行動評価方法を提案する。

本研究室で開発を進めているチームは、開発者の調整の繰り返しによって定められたパラメータにより行動の良し悪しを評価している。そのため、適切な評価に基づいた行動選択ができていない可能性がある。また、提案手法において、手動による調整ではコストや性能に限界がある。

そこで、世界大会上位チームの試合ログを模範とする行動選択ができるようにする。評価システムとして、SIRMs ファジィシステムを用いる。数値実験では、SIRMs ファジィシステムによる評価を組み込んだチームと組み込まないチームそれぞれで試合を行い、ペナルティエリアへの侵入回数を比較する。

2 RoboCup

RoboCup は、ロボット工学と人工知能の発展が目的の自律移動型ロボットによるサッカーなどを題材とした研究プロジェクトである。RoboCup には「西暦 2050 年までに、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チームを作る」という目標があり、この目標に向けて盛んに研究が行われている。RoboCup にはサッカー以外にも、大規模災害への対応のシミュレーションや災害現場で活躍するロボットの開発を促進するレスキューリーグ、次世代のロボット技術者育成を目的としているジュニアリーグなど、複数のリーグが存在する。本論文では、RoboCup サッカーシミュレーションリーグを研究の対象とする。

シミュレーションリーグはこの研究プロジェクトの立ち上げ当時から存在する最も古いリーグの 1 つである。サッカーシミュレーションでは、実機を使用せずに、コンピュータ内に用意された仮想フィールド上でサッカー競技を行う。サッカーシミュレーションには 2D リーグと 3D リーグがある。Figure 1, 2 に 2D リーグと 3D リーグの試合の様子を示す。

2D リーグでは、基本的な動作 (キックやドリブルなど) はコマンドとして実装されている。そのため 2D リーグでは高レベルな意思決定を主な研究対象としている。一方、3D リーグでは、エージェントはヒューマノイドロボットで形成されているため、基本的な動作を関節から制御する必要があり、基本的な動作が非常に重要である。本論文では 2D リーグを扱う。2D リーグでは、二次元平面を仮想

サッカーフィールドとし、円形のエージェントをプレイヤーとして競技を行う。また、プレイヤーやボールの位置と速度は全て二次元ベクトルとして表される。各プレイヤーはそれぞれ独立したエージェントとしてプログラムされており、制限された視覚情報や聴覚情報からドリブルやパス等の行動を選択する。

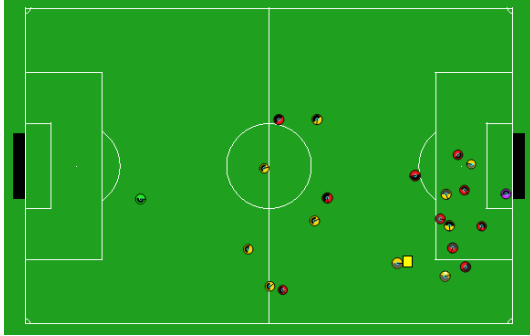


図 1: 2D Simulation League



図 2: 3D Simulation League

3 行動選択

チームを強くするためには、各プレイヤーが的確に行動を選択する必要がある。プレイヤーの行動選択方法について説明する。

3.1 行動探索

本論文で使用するプレイヤーは、行動連鎖と呼ばれる木構造を探索することで行動選択を行う [2]。行動連鎖の例を Figure 3 に示す。探索は以下に示す最良優先探索に基づく。Figure 3 において、丸で囲まれた数値は行動を行ったときの評価値である。初期状態からドリブルを選択した場合、評価値 30 が与えられる。また、ドリブルの後にパスを選択すると、評価値 35 が与えられる。

最良優先探索に基づいて行動連鎖を生成する手順は以下のとおりである。まず、プレイヤーの現在の状態をルート

ノードに入力する。次に、ノードに入力された状態において実行可能な行動の候補 (パス, ドリブル, シュートなど) を生成する。生成された行動に対して評価値を計算し、その行動を実行した場合の予測状態と共に子ノードに追加する。ノードが追加されるたびに評価値が最大であるノードを選択し、そのノードにおける予測状態から再び実行可能な候補の行動を生成する。これを繰り返すことで、ノード数があらかじめ設定された最大値に達するまで探索木を成長させる。ただし、木の深さがあらかじめ設定した値を越える場合や、ノードの予測状態から行動が生成できない場合、行動連鎖の終了条件に設定されている行動 (シュート) が生成された場合は、その葉ノードでの子ノード生成は行わないものとする。構築された木構造の中からノード列をつなげると、行動連鎖が得られる。

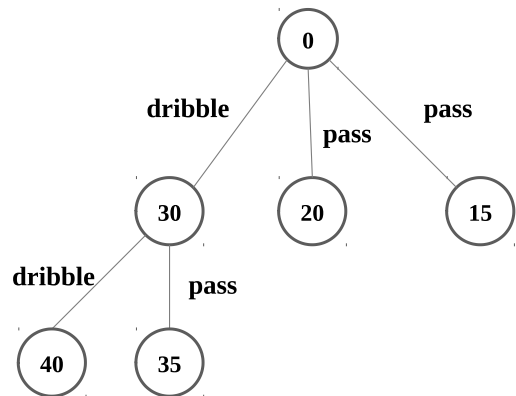


図 3: An example of action chains

3.2 行動評価

本研究室で開発を進めているチームでは、ゴールからの距離、ペナルティエリアへの侵入、サイド攻撃といった評価項目で評価値を算出している。それぞれの評価項目に対する点数は試合での調整の繰り返しによって定められている。そのため、適切な評価ができていないか不明である。また、手動による調整ではコストや性能に限界がある。そこで、行動の適切な評価を行うために、世界大会上位チームの試合ログにおける行動を模範とするような SIRM_s ファジィシステムのパラメータを獲得できるように学習を行う。

4 提案手法

本論文では、SIRM_s ファジィシステムを用いた学習を行う。SIRM_s ファジィシステムの概要と学習方法について説明する。

4.1 SIRM_s ファジィシステム

ファジィ推論の一つとして単一入力ルール群 (Single Input Rule Modules: SIRM_s) ファジィシステムがある。SIRM_s

ファジィシステムの概形を Figure 4 に示す．この手法では，各入力項目に対してルール群を用意する．ルール群は対応する入力項目だけが前件部変数に含まれる，1 入力のファジィルールで構成される．各ルール群には重視度と呼ばれる実数重みが割り当てられている．ルール群のファジィ推論結果の重視度付き総和を最終出力とする．SIRMs ファジィシステムは入力項目数が多い場合に，従来のファジィ推論モデルよりも大幅にファジィルール数とパラメータ数を削減することが可能であり，様々な制御問題へ応用されている．

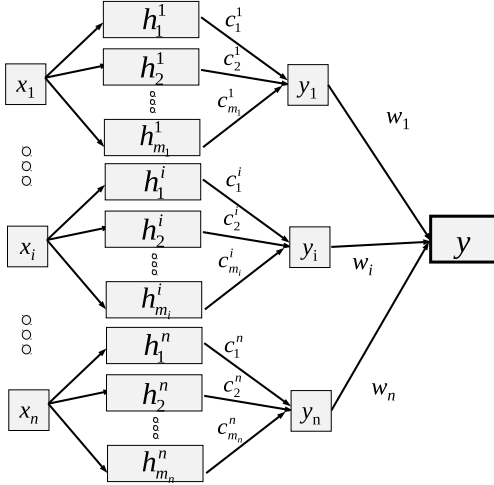


図 4: Overview of an SIRMs fuzzy model

Figure 4 において， $x_i (i = 1, \dots, n)$ は i 番目の入力項目に対応する前件部変数， $y_i (i = 1, \dots, n)$ はそのルール群の推論結果である． $h_j^i (i = 1, \dots, n, j = 1, \dots, m_i)$ は i 番目のルール群における j 番目のルールの前件部適合度， $c_j^i (i = 1, \dots, n, j = 1, \dots, m_i)$ は i 番目のルール群における j 番目のルールの後件部の実数出力値を意味する．また， $w_i (i = 1, \dots, n)$ は各入力項目の重視度， y は SIRMs ファジィシステムの最終出力を表す．入力が与えられたとき， h_j^i, y_i は以下の式で求められる．

$$h_j^i = A_j^i(x_i) \quad (1)$$

$$y_i = \frac{\sum_{j=1}^{m_i} h_j^i \cdot c_j^i}{\sum_{j=1}^{m_i} h_j^i} \quad (2)$$

$A_j^i (i = 1, \dots, n, j = 1, \dots, m_i)$ は i 番目のルール群における j 番目のルールの前件部変数 x_i のメンバシップ関数であり，以下の式のガウス関数とする．

$$A_j^i = \exp\left(-\frac{(x_i - a_j^i)^2}{b_j^i}\right) \quad (3)$$

最終出力 y は以下の式のようにルール群の推論結果の重視度付き総和として求められる．

$$y = \sum_{i=1}^n w_i \cdot y_i \quad (4)$$

本論文では，SIRMs ファジィシステムのパラメータの学習に最急降下法を用いる．最急降下法は，評価関数が最小値に収束するように各パラメータをベクトルの勾配の逆方向に探索していく手法である．学習するパラメータは，前件部変数のメンバシップ関数のパラメータ，後件部の実数出力値，および入力項目の重視度とする． p 番目の入力パターン (x_1^p, \dots, x_n^p) に対して，理想の出力が y^{Tp} ，実際の出力が y^{0p} であったとすると，評価関数 E^p を以下の式で表される．

$$E^p = \frac{1}{2}(y^{Tp} - y^{0p})^2 \quad (5)$$

式 (5) より，重視度 w_i ，後件部の実数出力値 c_j^i ，前件部変数のメンバシップ関数のパラメータ a_j^i と b_j^i の修正量は， t を現在の学習回数， $\alpha, \beta, \gamma, \eta$ を学習係数とすると以下の式で求められる．

$$\Delta w_i(t+1) = \alpha \cdot (y^{Tp} - y^{0p}) \cdot y_i(t) \quad (6)$$

$$\Delta c_j^i(t+1) = \beta \cdot w_i(t) \cdot (y^{Tp} - y^{0p}) \cdot \frac{h_j^i(t)}{\sum_{j=1}^{m_i} h_j^i(t)} \quad (7)$$

$$\Delta a_j^i(t+1) = \gamma \cdot w_i(t) \cdot (y^{Tp} - y^{0p}) \cdot (c_j^i(t) - y_i(t)) \cdot \frac{h_j^i(t)}{\sum_{j=1}^{m_i} h_j^i(t)} \cdot \frac{2 \cdot (x_i^p - a_j^i(t))}{b_j^i(t)} \quad (8)$$

$$\Delta b_j^i(t+1) = \eta \cdot w_i(t) \cdot (y^{Tp} - y^{0p}) \cdot (c_j^i(t) - y_i(t)) \cdot \frac{h_j^i(t)}{\sum_{j=1}^{m_i} h_j^i(t)} \cdot \left(\frac{x_i^p - a_j^i(t)}{b_j^i(t)}\right)^2 \quad (9)$$

4.2 SIRMs ファジィシステムへの学習

提案手法では，ペナルティエリアにボールを持ち込むことができた一連の行動を成功エピソードと定義する．SIRMs ファジィシステムの入力として，現在のプレイヤーの x 座標， y 座標，目標点の x 座標， y 座標，目標点と目標点から最も近い敵プレイヤーとの距離の 5 つを用いる．現在のプレイヤーと目標点の座標はフィールドサイズ+10 が 1.0 となるように $[-1.0, 1.0]$ の範囲に正規化した．目標点と目標点から最も近い敵プレイヤーとの距離は 30 が 1.0 とな

るように $[0.0, 1.0]$ の範囲に正規化した．入力情報の例を Figure 5 に示す．Figure 5 において， x_1, y_1 はパスを出すプレイヤーの x 座標， y 座標， x_2, y_2 はパスを受け取るプレイヤーの x 座標， y 座標， $dist$ はパスを受け取るプレイヤーとそのプレイヤーに最も近い敵プレイヤーとの距離であり，これらが SIRM_s ファジィシステムへの入力となる．

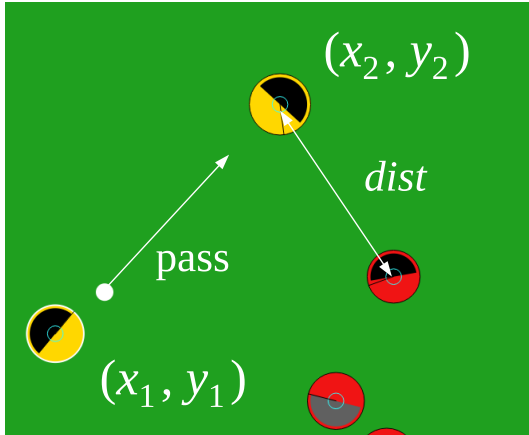


図 5: Elements of an input vector to the SIRM_s fuzzy model

各入力項目に 5 個のガウス型メンバシップ関数を用意し，初期のガウス型メンバシップ関数を隣同士がグレード 0.50 で交差するように設定する．また，後件部実数値の初期値をすべて 0.00 とする．各入力項目の重視度の初期値を 0.25 に設定する．教師信号を成功エピソード中の行動に対しては 1，その他のエピソード中の行動に対しては -1 とする．世界大会上位チームの試合ログからパスやドリブルといった行動を取り出し，成功エピソードとその他のエピソードに分け，入力情報と教師信号を付加した学習用データとする．

4.3 行動の評価

学習を終えた SIRM_s ファジィシステムを用いて行動評価する．プレイヤーの x 座標， y 座標，目標点の x 座標， y 座標，目標点と目標点から最も近い敵プレイヤーとの距離の 5 つを SIRM_s ファジィシステムに入力し，出力値を算出する．試合中に各プレイヤーが実行可能な行動の候補に対して SIRM_s ファジィシステムの出力値を計算し，行動の評価値として使用する．プレイヤーが実行可能な行動の候補に対する SIRM_s ファジィシステムの出力値が 0 より大きいものがない場合，既存の評価関数を用いて行動を評価し，行動を選択する．

4.4 予備実験

学習用データや学習係数 $\alpha, \beta, \gamma, \eta$ について予備実験を行い，決定した．

予備実験では，学習用データとして HELIOS2013 [2] 対 agent2d [3] の 100 試合分のログを用いる．HELIOS2013

は RoboCup2013 準優勝チームである．また，agent2d はオープンソースのチームであり，多くのチームのベースチームとして用いられている．入力情報に対する SIRM_s ファジィシステムの出力値が 0 より大きければ，成功エピソード中の行動，0 より小さければその他のエピソード中の行動と識別する．10-fold cross validation により，識別率を調査した．

4.4.1 学習用データ

予備実験では，SIRM_s ファジィシステムの入力として，現在のプレイヤーの x 座標， y 座標，目標点の x 座標， y 座標の 4 つを用いた．学習係数 $\alpha, \beta, \gamma, \eta$ ，教師信号を変化させながら学習を行った．ほとんどの行動がその他のエピソード中の行動と識別され，成功エピソード中の行動とその他のエピソード中の行動を上手く識別することができなかった．そこで，識別率を高めるために学習用データを変化させながら識別率を調査した．学習用データに以下の変更を加えた．

まず，入力情報に目標点と目標点から最も近い敵プレイヤーとの距離を加えた．これは，現在のプレイヤー位置と目標点と同じであっても，敵の位置によって成功エピソード中の行動，その他のエピソード中の行動の両方が存在するので，上手く識別できなかったのではないかと考えたためである．次に，ボールをキープしているときのキックなどの識別しにくい行動を学習用データから除外した．また，ドリブルは敵を避ける行動などゴール方向に進まないものが多く，プログラムの作り込みが重要であるので，学習用データからドリブルを除外し，パスのみとした．学習用データにこれらの変更を加えたとき最も識別率が高かった．

4.4.2 学習係数

学習係数 $\alpha, \beta, \gamma, \eta$ をそれぞれ 0.10, 0.10, 0.01, 0.01，教師信号を成功エピソード中の行動に対しては 1，その他のエピソード中の行動に対しては -1 とする．全ての学習係数を 1/10 倍，1/100 倍，1/1000 倍して識別率を調査した．識別率を Table 1 に示す．Table 1 より，学習係数を 1/100 倍したとき識別率が最も高いと分かる．したがって，本研究の実験において SIRM_s ファジィシステムに用いる学習係数 $\alpha, \beta, \gamma, \eta$ をそれぞれ 0.001, 0.001, 0.0001, 0.0001 とする．

表 1: Various sets for learning rates
($\alpha=0.10, \beta=0.10, \gamma=0.01, \eta=0.01$ を 1 とする)

学習係数の倍率	成功	その他	全て
1	0.6798	0.3356	0.5709
1/10	0.3640	0.7618	0.4898
1/100	0.8472	0.2524	0.6590
1/1000	0.8545	0.2073	0.6497

5 数値実験

数値実験では、提案手法を組み込んだ opuSCOM と組み込まない opuSCOM を比較する。opuSCOM は本研究室で開発を進めているチームであり、昨年の JapanOpen では5位という結果に終わった。opuSCOM はドリブルによるサイド突破を中心としているチームである。RoboCup サッカー 2D シミュレーションでは、試合における各サイクルのボールとすべてのプレイヤーの位置、実行した行動がログに記録される。ログから行動を抽出するために、本実験におけるパスの定義を示す。連続するキックの中で、次にキックするプレイヤーが、キックしたプレイヤーと同じチームの異なるプレイヤーである場合はパスと定義する。ドリブルは敵から避ける行動などゴール方向に進まないものが多く、プログラムの作り込みが重要である。そのため、学習用データからドリブルを除外し、パスのみとする。予備実験と同様に学習用データとして、HELIOS2013 対 agent2d の 100 試合分のログを用いる。例として、学習用データ 1 試合分のパスの軌跡を Figure 6 に示す。赤い線が成功エピソード中のパス、青い線はその他のエピソード中のパスである。ボールを奪われにくい後ろへのパスを選択し続け、敵フィールドへ侵入しないといった状況避けるために、SIRMs ファジシステムでの行動評価を適用するのは敵フィールド上のみとする。SIRMs ファジシステムによる評価を組み込んだチームと組み込まないチームそれぞれで試合を実行し、ペナルティエリアへの侵入回数、試合中の行動、ボール支配率について調査する。試合は、JapanOpen2013 に出場した 5 チーム、agent2d と 100 試合ずつ行う。

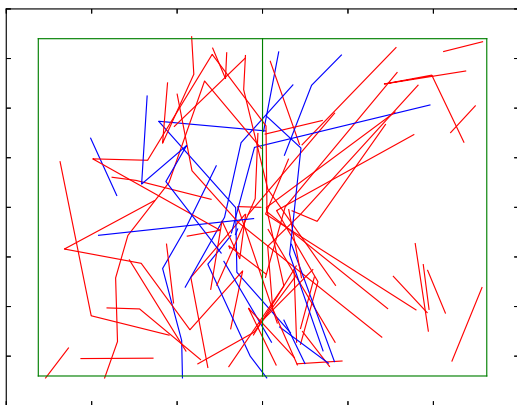


図 6: Successful/unsuccessful passes that are used as the training patterns of the SIRMs learning

6 実験結果

提案手法を組み込んだ opuSCOM と提案手法を組み込まない opuSCOM それぞれを、JapanOpen2013 に出場した 5 チーム、agent2d と 100 試合ずつ対戦させた。提案手法を組み込んだチームの試合と組み込まないチームの試合に

おけるペナルティエリアへの侵入回数を調査した。Table 2 にペナルティエリアへの侵入回数を示す。

表 2: The number of episodes that led the ball to the opponent's penalty area

対戦相手	提案手法なし	提案手法あり
agent2d	1327	989
A_TSU_BI-	712	418
HillStone	973	699
KU_BOST	886	693
ThinkingAnts	833	454
Ri-one2013	978	527

Table 2 より、提案手法を組み込むと、ペナルティエリアへの侵入回数が減少したことが読み取れる。提案手法を組み込んだチームではパスの学習により、敵フィールドでのパス回しが多くなり、攻撃にかかる時間が増加した。また、提案手法を組み込まないチームでは、ペナルティエリアに侵入する行動に非常に大きな評価値が与えられるので、無理矢理ペナルティエリアにボールを入れる行動を選択をすることがある。そのため、提案手法を組み込んだチームでは、提案手法を組み込まないチームに比べてペナルティエリアへの侵入回数が減少したと考えられる。

次に、提案手法を組み込んだチームと組み込まないチームの試合における行動を比較すると、スルーパスの回数に変化が見られた。スルーパスとは、誰もいないスペースへパスを出し、そこに味方プレイヤーが走り込み、受け取るものであり、決定的なチャンスに繋がりがやすい。スルーパスの例を Figure 7 に示す。Figure 7 において、右が攻撃方向である。スルーパスを出すことで 2 人の相手プレイヤーを抜きさることができる。本論文では、敵フィールド上でレシーバがパスを出されたときから、パスを受け取るまでに x 軸方向に 5m 以上走ったものをスルーパスと定義する。Table 3 にスルーパスの回数を示す。

表 3: The number of through passes

対戦相手	提案手法なし	提案手法あり
agent2d	172	461
A_TSU_BI-	327	378
HillStone	682	580
KU_BOST	310	403
ThinkingAnts	174	314
Ri-one2013	1024	1135

Table 3 より、提案手法を組み込むと、5 つのチームに対してスルーパスの回数が増加していることが分かる。特に学習用データの相手チームであった agent2d に対しては大幅に増加していることが読み取れる。学習用データにおけるエピソードの総数、スルーパスを含むエピソードの数を成功エピソードとその他のエピソードに分けて Table 4 に示す。

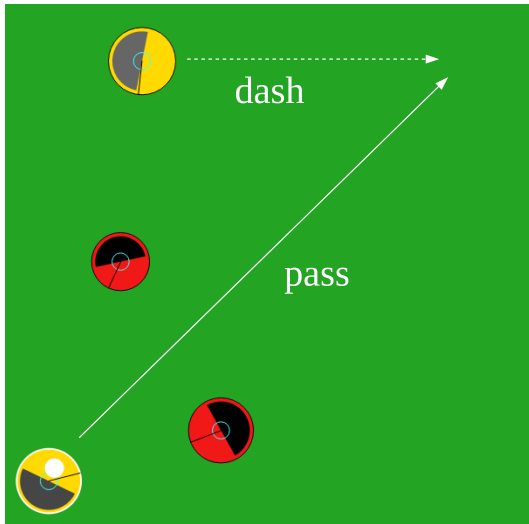


図 7: An example of through passes

表 4: The number of episodes in training patterns

	成功	その他
エピソードの総数	1592	1083
スルーパスを含むもの	1165	107

Table 4 より, 成功エピソードの 73% がスルーパスを含んでいることが読み取れる. また, スルーパスを含むエピソードの 92% が成功エピソードであることがわかる. このことから, スルーパスは成功エピソードの特徴的な行動であることが分かる. また, 提案手法を組み込んだチームにおけるスルーパスの増加は, 提案手法によって成功エピソードを学習できたことを示していると考えられる.

次に, 提案手法を組み込んだチームと組み込まないチームのボール支配率を調査した. ボール支配率とは, ボールを保持している時間の割合である. 本論文では, ドリブルやパスを行っているときボールを保持していると定義する. 試合におけるボール支配率を Table 5 に示す.

表 5: Ball possession

対戦相手	提案手法なし	提案手法あり
agent2d	55.86	66.66
A_TSU_BI-	47.98	58.07
HillStone	48.31	54.87
KU_BOST	56.51	66.71
ThinkingAnts	47.49	58.72
Ri-one2013	61.15	68.91

Table 5 より, 全てのチームに対してボール支配率が向上していることが読み取れる. パスを学習したことにより, ドリブルでの無理な突破をせずにパス回しをすることが多くなり, 敵にボールを奪われる回数が減ったためだと考えられる.

7 おわりに

本論文では SIRM_s ファジィシステムを用いて世界大会上位チームの試合ログを学習し, その出力値による行動評価について調査した. 実際に提案手法を組み込んだチームで試合を行った. 学習用データの成功エピソードを学習できたことを示した. ペナルティエリアへの侵入回数は減少したが, スルーパスの回数, ボール支配率など違う観点から見るとチームの性能が向上した. 今後の課題としては, スルーパス後などの状況に合わせた評価により, ペナルティエリアへの侵入回数を増やすこと, 様々な相手チームに対応できるような評価などが挙げられる.

参考文献

- [1] 湯場崎直義, 易建強, 廣田薫, “複数入力ファジィ制御のための単一入力ルール群結合型ファジィ推論モデルの提案” 日本ファジィ学会誌, Vol.9, No.5, pp.699-709, 1997.
- [2] Hidehisa Akiyama, Tomoharu Nakashima, Katsuhiko Yamashita, HELIOS2013 Team Description Paper, *RoboCup2013*, CD-ROM(6 pages), Eindhoven, The Netherlands, (2013)
- [3] 秋山英久, *RoboCup サッカーシミュレーション 2D 必勝ガイド*, 秀和システム, 2006.