

agent2d のチェーンアクションにおける評価関数の重み調整

Adjustment of weight parameters in a state evaluation function used in "chain action" of Agent2d

田川 諒 谷川 俊策 五十嵐 治一
Ryo Tagawa Shunsaku Tanikawa Harukazu Igarashi

1. はじめに

RoboCup サッカーシミュレーション 2D リーグは、コンピュータ内の仮想フィールド上でエージェント同士がサッカーの試合を行う競技であり、人工知能研究のテストベッドとしても利用されている[1].

この競技では、多くのチームがベースとなるプログラムとして「agent2d」[2]を採用している。agent2d は近年、ボール保持者の行動を、ゲーム探索木と局面評価関数を用いて決定する「チェーンアクション」と呼ばれる手法を実装している。

しかし、agent2d に組み込まれている評価関数は極めて単純であり、局面の優劣の判定には必ずしも十分ではないと考えられる。そこで、我々は複数のヒューリスティクスによって構成された評価関数を考案した[3]。その評価関数はヒューリスティクスによる複数の評価項の線形和で構成されていた。本研究では、この重み係数を合理的に決定する方法を提案する。本手法は、局面に対する評価関数の値が人間の主観評価と相関を持つように、評価関数の各項の値と主観評価値との相関強度から重みを決定する方法である。

また、以前の我々の研究[3]では、全てのプレイヤーが試合の状況に関係なく常に同じ重みを使用していたが、フォワード(FW)やミッドフィールダー(MF)は、得点に繋がりやすい局面を良しとする攻撃用の重みを、ディフェンダー(DF)は失点に繋がりやすい局面を良しとする守備用の重みを持たせた方が、攻守のバランスが取れると考えた。さらに、相手チームよりも得点において勝っている場面では、FW や MF であっても守備用の重みを、逆に負けている場面では DF であっても攻撃用の重みを持たせた方が、勝率が上がるのではないかと考えた。そこで、本研究では攻撃用と守備用の 2 種類の重みを用意し、ポジション(役割)ごとに重みを設定した場合、チームの得点差によって重みを攻撃用と守備用で切り替えた場合について、勝率がどのように変化するかを実験した。

2. サッカーシミュレーション 2D リーグ

RoboCup サッカーシミュレーション 2D リーグの公式シミュレータは、UDP/IP によるクライアント・サーバ方式で構成されている[4]。試合中各エージェントは、サーバからボールやプレイヤーの座標などの環境情報を受け取り、これを基に行動の決定を行う。そして各エージェントは、選択した行動の情報を「kick」や「dash」といった制御コマンドとしてサーバに送信し、サーバはそれを基に環境情報を更新する。エージェント同士は、サーバを介さない直接の通信を禁じられているため、自律分散制御による協調行動の実現が重要となる。

3. agent2d について

3.1 agent2d

本研究で使用した agent2d は、RoboCup2010 世界大会の優勝チーム「HELIOS」を基に作られたサンプルチームプログラムである。オープンソースとして公開されており、基本的な行動戦略などは既実装されている。また、チーム開発を支援するライブラリが豊富に用意されていることから、多くのチームがベースプログラムとして利用している。

3.2 チェーンアクション

ver3.1.0 以降の agent2d では、チェーンアクションと呼ばれる手法により、ボール保持時における行動の決定を行う[5]。ボール保持者はパスやドリブル、シュートといった行動を枝、行動の結果生じる予測状態をノードとする探索木を生成する。各ノードは、評価関数によって優劣が点数化される。点数が最も高いノードを最良優先探索によって探索し、そのノードへ至る行動を選択する(max 戦略)。ただし、点数が最も高くなるノードは必ずしも葉ノードとは限らない。

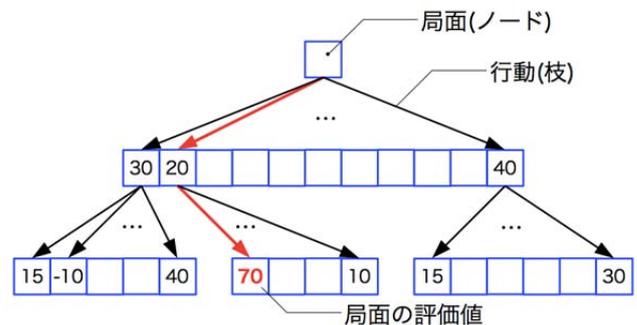


図 1. チェーンアクションで生成される探索木の模式図

4. 評価関数

agent2d における評価関数は、ボールと敵ゴールの距離のみに依存した単純な関数である。そこで、我々は次の(1)のような評価関数を提案した[3].

$$E(s; \omega) = U_0(s) \left[\omega_1 U_1(s) + \omega_2 U_2(s) \right] + \sum_{i=3}^5 \omega_i U_i(s) \quad (1)$$

ここで、 $U_i(s)$ ($i = 0, 1, \dots, 5$) は、局面 s を評価する際に有用だと思われる 6 つのヒューリスティクス(先見の知識)を表した関数であり、 ω_i ($i = 1, 2, \dots, 5$) は重み係数である。(1)の評価関数 $E(s; \omega)$ は、正/負で絶対値が大きいほど自身の所属チームが優勢/劣勢な局面を表すよう定義している。(1)の各項について簡単に述べると、 U_0 ($\in \{-1, 0, 1\}$) は、ボール保持者の所属チームを表している。 U_1 はボール保持者と敵の距離、 U_2 はボール保持者から見て敵ゴール側にいる敵と味方の人数比、 U_3 はボールと両ゴールの距離、 U_4 はボールと両チ

ームのプレイヤーの距離, U_5 はボール周辺のプレイヤーの分布を, それぞれ表している. U_1, U_2 は[0, 10], U_3 から U_5 は[-10, 10]の区間内の値を取るよう正規化されている.

5. 評価関数の重みの決定

5.1 局面の主観評価値と各評価項の相関係数

各評価項 $u \in \{U_0, U_1, U_2, U_3, U_4, U_5\}$ の重みについて, 攻撃用と守備用の2種類の重みを決定するため, 被験者3名にアンケート調査を行った. 試合中の局面を静止画として抽出したものを用意し, 「得点のしやすさ」と「失点のしにくさ」の2つの項目 p について, 左側のチームから見た局面の良し悪しを7段階(-3~+3の値)で評価してもらった.

静止画は, 左チームが(1)の評価関数を実装したチーム, 右チームがオリジナルの agent2d の計7試合から, 左側のチームがボールを持っている局面 50 シーンと, 右側のチームがボールを持っている局面 50 シーンの計 100 シーンを手作業で抽出した. この際, 左チームの評価関数の重みは, 少ない試合数から様々な局面を抽出するため, 適当に重みを調整した. また, アンケートでは, 「得点のしやすさ」は左側のチームがボールを持った 50 シーンを, 「失点のしにくさ」は上記の全 100 シーンを, それぞれ評価の対象とした.

次に, 項目 p ごとに主観評価値 a と, その局面での u の値 b の相関係数 $\rho_{pu} (-1 \leq \rho_{pu} \leq 1)$ を, (2)により被験者ごとに算出した(図2, 3).

$$\rho_{pu} = \frac{\sum_{j=1}^n (a_{jp} - \bar{a}_p)(b_{ju} - \bar{b}_u)}{\sqrt{\sum_{j=1}^n (a_{jp} - \bar{a}_p)^2} \sqrt{\sum_{j=1}^n (b_{ju} - \bar{b}_u)^2}} \quad (2)$$

(2)の式において, $a_{jp} \in \{-3, -2, \dots, 3\}$ は j 枚目の画像の示す局面に対する項目 p での主観評価値, \bar{a}_p は a_{jp} の平均値, $b_{ju} (-10 \leq b_{ju} \leq 10)$ は, j 枚目の画像の示す局面で u が取る値, \bar{b}_u は b_{ju} の平均値, $n \in \{50, 100\}$ は使用した画像の総数を表す.

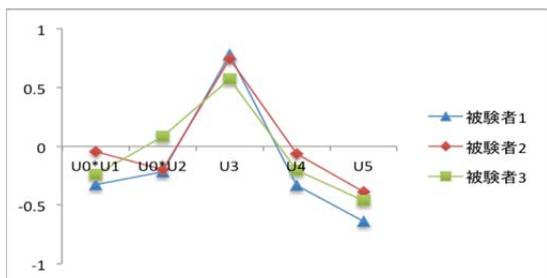


図2. 「得点のしやすさ」と各評価項の相関係数

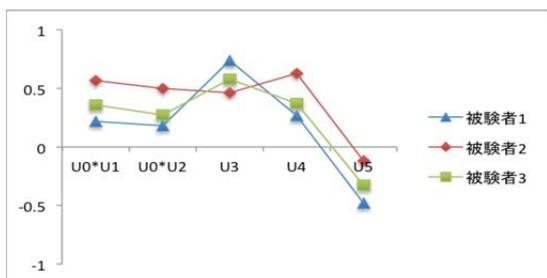


図3. 「失点のしにくさ」と各評価項の相関係数

5.2 相関係数から重みへの変換

評価項 u の係数である重み ω_{pu} は, u が評価関数 $E(s; \omega)$ にどの程度影響力を持つかを表しており, (2)を用いて求めた相関係数 ρ_{pu} (図2-3)から次の(3)のように決定した.

$$\omega_{pu} = \begin{cases} \rho_{pu}^2 & \text{if } \rho_{pu} \geq 0 \\ 0 & \text{if } \rho_{pu} < 0 \end{cases} \quad (3)$$

(3)に示したように, ρ_{pu} が正の値を取った場合は, 重みの差を強調するため, ρ_{pu} の2乗を重み ω_{pu} とした. しかし, ρ_{pu} が負の値を取った場合は, ω_{pu} を0とした. これは, (1)の U_i の設計においては重みが負の値を取ることを想定しておらず, u の取る値が小さくなる程評価値が大きくなるのは不自然と考えたからである.

以上の操作により決定した重みを, 表1に示す.

表1. 主観評価と各評価値の相関係数から決定した重み

評価項目 p	被験者 No.	ω_{p1}	ω_{p2}	ω_{p3}	ω_{p4}	ω_{p5}
得点のしやすさ	1	0	0	0.613	0	0
	2	0	0	0.552	0	0
	3	0	0.007	0.327	0	0
失点のしにくさ	1	0.048	0.033	0.541	0.070	0
	2	0.319	0.248	0.212	0.390	0
	3	0.127	0.073	0.336	0.136	0

6. 評価実験

6.1 全員が同じ重みを用いた場合

表1の重みを用いて, 人間が「得点しやすさ」と感じる局面で大きな値を取る評価関数と, 「失点しにくい」と感じる局面で大きな値を取る評価関数の2種類を作成した. これにより, プレイヤーはそれぞれ攻撃的, 守備的な動きになるのではないかと考えた. そこで, 全員に攻撃用の重みや守備用の重みを持たせたチームと, オリジナルの agent2d(ver3.1.1)とで, 1000回ずつ試合を行った. チェーンアクションにおける探索の深さは2, 探索するノード数の上限は500とした. 表2, 表3に, 各被験者ごとの重みの試合結果を示した. なお, 攻撃用の重みについては, 被験者2の重みと被験者1の重みの比率が一致したので, 表2においては被験者2の重みについての結果を省略している. また, 参考のため agent2d 同士での試合の結果も併記した.

表2. 全員に攻撃用の重みを持たせたチームの対戦結果

被験者 No.	勝率	得点 - 失点	X座標	ボール支配率
1	57.49%	2.64 - 2.29	-5.44	47%
2				
3	58.52%	2.68 - 2.32	-5.37	47%
(agent2d 同士)	49.81%	2.04 - 2.06	+0.32	51%

表3. 全員に守備用の重みを持たせたチームの対戦結果

被験者 No.	勝率	得点 - 失点	X座標	ボール支配率
1	26.45%	0.78 - 1.44	+0.33	67%
2	4.11%	0.05 - 0.81	-3.06	87%
3	7.10%	0.14 - 0.99	+1.38	80%

表2, 表3における「勝率」は(勝った試合数)/(負けた試合数+引き分けた試合数)で求めた割合を, 「得点-失点」は1000試合通しての得点と失点の1試合あたりの平均を, 「X座標」はボールのX座標値の平均(敵ゴール方向をX軸の正の方向, コート中心を原点とする), 「ボール支配率」は試合中にボールを保持していた時間の割合を, それぞれ表している.

攻撃用の重みを持たせたチームの平均得点は, 全ての被験者で agent2d の平均得点 2.04 より高くなっている. また, 守備用の重みを持たせたチームの平均失点は, 全ての被験者で agent2d の平均失点 2.06 より低くなっている. このことから, 人間の主観評価を用いて 5. で定めた重みを用いると, 攻撃的, あるいは守備的なプレーを行うようになったと言える.

表2で被験者3の攻撃用の重みを持ったチームと, オリジナルの agent2d との試合結果において得点力が向上した理由について考察する. この時の重みは, U_3 (ボールと両ゴールの距離)の重み以外はほぼ0で, オリジナルの agent2d の評価関数とほとんど同じであるにも関わらず, 勝率や平均得点において大きな差が見られる. 実際の試合では, agent2d がタッチラインの際に沿って敵ゴールへ向かう場面が多く見られた. これは agent2d が, ゴールとの距離ではなく, X座標の値のみを局面の評価に使用していることによると考えられる. 図4は agent2d の, ボールの距離に関する評価値の分布, 図5は U_3 の値の分布を表す. 右側のゴールが敵ゴールである. これらを見比べると, 敵ゴール正面の評価値が, U_3 の分布ではより高く設定されている事が分かる. これにより, 新しい評価関数を用いたチームの方が正面から攻める機会も増え, 得点力が向上したためと考えられる.

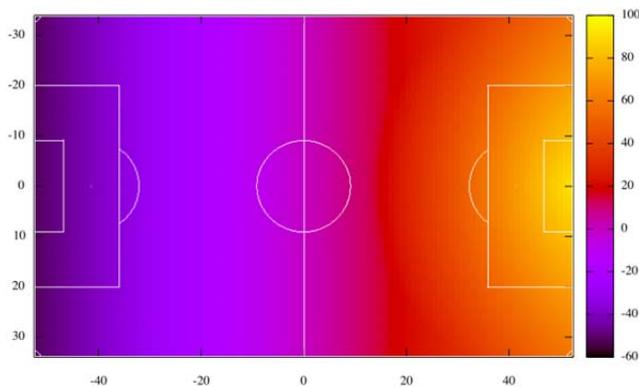


図4. agent2d の評価値の分布

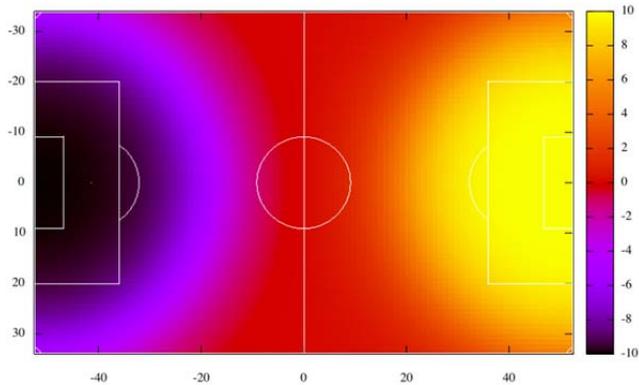


図5. U_3 の値の分布

守備用の重みを持ったチームについては, 被験者2は U_3 (ボールと両ゴールの距離)の重みの比率が, 他の被験者よりも低めになっている. そのため, 無理にゴール方向へ向かわず味方の中でボールを回す機会が増え, 他の被験者よりも平均失点とボール支配率において守備としての効果が表れたと考えられる.

6.2 役割ごとに重みを変えた場合

agent2d のフォーメーションではFWが3名, MFが3名, DFが4名という構成を取っている. 先述の6.1節では, 全てのプレイヤーが同じ攻撃用の重みを使用しており, 結果的に平均得点が増加している反面, 平均失点も増加している. そのため, FWには攻撃用の重みを持たせ, DFには守備用の重みを持たせるというように, それぞれの役割に応じた重みを用いた方が, 全員が同じ重みを用いるよりも攻守のバランスが取れるのではないかと考えた. そこで, 役割ごとに2種類の重みを割り振ったチームと agent2d とで対戦させた.

重みの割り振り方は, (i)攻撃用の重みをFW・MFに, 守備用の重みをDFに割り振る方法と, (ii)攻撃用の重みをFWに, 守備用の重みをMF・DFに割り振る方法の2種類を試した. 攻撃用の重みは被験者3の値を, 守備用の重みは被験者3名の値をそれぞれ使用した(表4).

表4. 役割ごとの重みの割り振り方

		FW	MF	DF
(i)	a	3	3	1'
	b			2'
	c			3'
(ii)	a	3	1'	1'
	b		2'	2'
	c		3'	3'
(iii)		3	3	3

表4では, 被験者kの攻撃用の重みをk, 守備用の重みをk'と表記している. (i)と(ii)の2種類の割り振り方それぞれについて, 最も勝率の高いチームはそれぞれ(i)a, (ii)bであった. これらの重みと, (iii)全員が同じ重みを使用したチームの対戦結果(表2の網掛け部を再掲)を, 表5に示す.

表5. 役割別に異なる重みを使用したチームの対戦結果

重み	勝率	得点-失点	X座標	ボール支配率
(i) a	57.41%	2.60 - 2.31	-5.41	47%
(ii) b	56.83%	2.36 - 2.08	-5.05	55%
(iii)	58.52%	2.68 - 2.32	-5.37	47%

(i)と(iii)を比較すると, DFだけに守備用の重みを持たせてもほとんど変化がない事が分かる. しかし, (ii)と(iii)を比較すると, DFに加えてMFにも守備用の重みを持たせた(ii)は平均失点が減少している. この場合, 実際の試合を観察すると, 近くに敵がいることが多いFWへのパスに対しては慎重な場面が多く見られた. そのためにFWへのパスに手間取り, 敵が守備のために自陣へ下がって守りを固めてしまう場面も見られた. 従って, 平均失点やボール支配率は比較的良い値を取っているが, 平均得点は(iii)と比べて減少してしまっただと考えられる.

6.3 得点差により重みを切り替えた場合

得点でリードしている場面では、積極的に攻め込むより、ボールを奪われないことの方が有益であることが考えられる。そこで、得点差によって攻撃用の重みと守備用の重みを切り替えるチームを作成し、agent2d と対戦させた。具体的には、得点差で 1 点以上リードしている場合は全員に守備用の重みを、逆に 1 点以上リードされている場合は全員に攻撃用の重みを割り振るよう設定した。なお、同点の場合は先述の 6.2 と同様の方法で、役割ごとに異なる重みを割り振った(表 4)。今回、最も勝率の高くなったチームにおける重みの割り振り方を表 6 に示す。

表 6. 点差により切り替える重みの割り振り方

	FW	MF	DF
勝っている局面	1'	1'	1'
同点の局面	3	1'	1'
負けている局面	3	3	3

表 7 に、(i)表 6 に示した重みを用いたチームの結果と、(ii)全員が常に同じ攻撃用の重みを使用したチーム(表 4 の(iii))の結果(表 2 の網掛け部を再掲)を示す。

表 7. 点差で重みの切り替えを行うチームの対戦結果

	勝率	得点 - 失点	X 座標	ボール支配率
(i)	59.61%	2.03 - 2.00	-3.24	55%
(ii)	58.82%	2.68 - 2.32	-5.37	47%

(i)を見ると、平均得点と平均失点の間に差がほとんど無いにも関わらず、勝率はわずかではあるが(ii)よりも優れている。このことから、点差において劣勢である場面では積極的に攻め、優勢である場面では攻撃を抑えて守備に重点を置くという戦略の効果が分かる。

7. まとめと今後の課題

本研究では、従来の研究で考案された複数のヒューリスティクスで構成された評価関数において、各ヒューリスティクスを表す評価項の値と主観評価値との相関係数から重みを決定する方法を提案した。得られた重みをプレイヤーの役割ごとに割り振ったチームを作成し、評価関数を変更する前の agent2d と対戦を行った。その結果、全員が攻撃用の重みを持たせた場合と比べ、DF と MF に守備用の重みを持たせることで、平均失点を減少させることができた。

次に、得点差によって攻撃用の重みと守備用の重みの切り替えを行うと、平均得点と平均失点の差がほとんど無い場合でも勝率を保つことが可能であることが分かった。

今後は、強化学習を用いた役割ごとの重みの決定や、チェーンアクションのゲーム探索木における枝刈りの緩和、評価関数中の項自体の見直しについて、検討する予定である。また、ボールを保持していないプレイヤーの行動決定問題についても、評価関数の利用を考えていきたい。

参考文献

[1] 野田五十樹, “RoboCup におけるマルチエージェントシミュレーション”, セルオートマトン・シンポジウム講演論文集, pp. 63-68, 2011.

[2] RoboCup tools / soccer simulation wiki (<http://sourceforge.jp/projects/rctools/releases/>)

[3] 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”, 電子情報通信学会総合大会, D-8-5, 2013.

[4] 秋山英久, “RoboCup サッカー 2D シミュレーションリーグ解説: 仕組みと環境構築”, 知能と情報, Vol. 23, No. 5, pp. 714-720, 2011.

[5] 秋山秀久, “アクション連鎖探索によるオンライン戦略プランニング”, 人工知能学会研究会資料, SIG-Challenge-B101-6, pp. 23-28, 2011.