論 文

Technical Papers

予測能力を持つサッカーエージェントによる協調戦術の獲得

Acquisition of Cooperative Tactics by Soccer Agents with Ability of Prediction and Learning

熊田 陽一郎 東京大学 大学院総合文化研究科 広域科学専攻
Yoichiro Kumada Department of Systems Science, The University of Tokyo
kuma@blake.c.u-tokyo.ac.jp

植田 一博 東京大学 大学院情報学環・学際情報学府
Kazuhiro Ueda Interfaculty Initiative in Information Studies, The University of Tokyo
ueda@gould.c.u-tokyo.ac.jp

keywords: soccer agents, cognitive modeling, cooperative tactics, Bayesian prediction, adaptive learning

Summary

Designing soccer agents operating on the Soccer Server has became a standard problem in the multi-agent domain, and this paper describes the soccer agents that can learn to make use of cooperative tactics. Considering the ways actual coaches of soccer enable their players learn to execute the soccer tactics, we developed a method of agents' learning to distinguish good tactics from not-so-good tactics. It is made up mainly of small practical tasks requiring a few agents, of acquisition of appropriate cognitive maps by decomposing the situations into grid information, and of optimization of total play by a kind of adaptive learning. Because the agents perceive the environment as a grid, they have a finite number of condition spaces and are able to predict the behavior of opponents by learning the conditonal probabilities. Each condition has its own utility learned in an evolutionary method.

1. はじめに

サッカーサーバー上で動くサッカーエージェントについては，マルチエージェントシステムの標準問題として多くの研究がなされており，エージェントがボールを扱う上での基本スキルの研究 [Asada 96] や，エージェントをフィールドに配置するフォーメーション [Stone 99] や攻撃パターンなど戦略（strategy）の研究が主となっている．しかし，スキルと戦略とのあいだには本来，戦術（tactics）という階層が存在すると実際のサッカーの専門家は指摘している [Hughes 80]．本研究の目的は，学習によって，戦術の中でも特に重要だと考えられる協調的戦術を獲得するサッカーエージェントの構築である．実際のサッカーの指導 [Hughes 80] で戦術がいかに獲得されるかを参考に，小人数による練習課題の達成，グリッドによる適切な認知地図の獲得，そして，適応学習によるプレーの最適化を軸にした戦術決定の方法を提案する．エージェントは環境をグリッド化して知覚することで有限の状態変数空間を持ち，他エージェントの挙動を条件付き確率によって予測推定する．さらに推定結果と各状態変数の効用に基づき自らの戦術を決定する．練習を通じエージェントは条件付き確率と効用関数を学習することができる．この学習の結果，同じチームのエージェント間で効用関数が共有され，その結果，協調戦術が遂行されるようになる．

2. 従来のサッカーエージェントの主流

2 次元空間という限定された空間内でエージェントがプレーを行なう RoboCup において，強いチームを作るために何をすべきか，という方針は明白である．すなわち，

- 敵のパスを奪うのに良い位置にいる（位置取り），
- 自分の領域にきた敵のパスを確実にカットできる（基本技術），
- 誰にパスを出せば敵に取られにくいか判断できる（判断），
- 敵に取られにくいパスが出せる（基本技術），
- 味方のパスを受けるのに良い位置にいる（位置取り），
- 味方のパスを確実に受けられる（基本技術），
- 敵に取られないようにドリブルできる（基本技術），
- ゴールキーパーに取られないようなシュートが撃てる（基本技術）．

ゲームのログ [RoboCup Official Site] の分析を見る限り，上位チームは上記の条件をほぼ満たしている．現在のエージェントの流行としては，エージェントが行なう学習も，上記のうち「基本技術」「位置取り」「判断」という 3 つをテーマにしたものが多い．その中でも，基本技術と位置取りの学習を対象にし，技術を磨き，フォーメーションを作る，というタイプの研究が多い．なぜなら，単純化された 2 次元空間内では，アクションの選択肢が少ないため，コンフリクトする状況に適切な判断を下す，といった「判断力」もそれほど高いものは要求されていないからである．ゆえに，技術とフォーメーションでほぼ大勢が決まる，と言って過言ではない．

しかし，実際のサッカーはそれだけではない．サッカー指導者 Charles Hughes は,「サッカーで一番重要なのは，プレーのシステム（フォーメーション）だと信じ込んでいる人たちは，“生兵法は大怪我のもと” ということわざを，思い出していただきたい」と，フォーメーションのみを重視する態度を戒めている [Hughes 80]．現時点での RoboCup ではフォーメーションの善し悪しのウェイトは高いのは事実であるが，そのフォーメーションを「本来サッカーの本質的な部分ではないのだ」とする Hughes の意見は,「協調」のあり方を模索するサッカーエージェントの研究においても長い目で見れば考慮に値すると考えられる．そこで本研究では，Hughes の主張するサッカーの指導法，学習法にならい, 以下の事項に着目して，協調的な機械学習への応用を試みる．すなわち，

（1）フィールドのどの部分を使って練習するのか，
（2）練習には何人のプレーヤーが参加するのか，
（3）単純な練習内容の設定，

である．

3. 学習により協調戦術を獲得するサッカーエージェント

3·1 目　　標

前章で触れた Hughes の指導法 [Hughes 80] を参考に，小人数での協調戦術を学習するサッカーエージェントを構成する．方針は以下の通りとする．

（1）小人数による練習課題の達成
（2）グリッドによる適切な認知地図の獲得
（3）他エージェントの行動の予測と，その学習
（4）状態の効用関数の学習

以下に，詳細を説明する．

小人数による練習課題の達成

具体的には，アタッカ, ディフェンダが 3 対 2 での練習を想定する．この練習では，攻守のエージェント数には差がある．というのも，あるローカルな状況が，一方のチームにとって攻撃なのか守備なのか，という区別は，そのチームに数的優位があるかどうかに依存するからである．さらに，壁パスなどの基本的な協調戦術を運用するためには最低 3 エージェントが必要となることから，攻撃側を 3 エージェントとした．フォーメーションを学習の対象とした場合であれば，全員参加の練習が前提となる．しかし，本研究ではフォーメーション等の学習を目的としていない．[Hughes 80] によれば，学習させるべき練習課題に応じた人数とフィールドを適切に設定すべきであり，最初からフルゲームで練習をさせるべきではない，という．本研究では，アタッカ（図 1 中の A1〜A3，以下同様）が図 1 のようなミニフィールドのエンドライン（図 1 中のミニフィールドの一番下の太線，以下同様）にボール（図 1 中の b，以下同様）を通すことができたら，アタッカの勝ち, ディフェンダ（図 1 中の D1〜D2，以下同様）がサイドライン（図 1 中のミニフィールドの左右のライン，以下同様）からボールを割らせたら，ディフェンダの勝ち，という課題を与える．アタッカならば，この課題により, ローカルな状況で確実にゴールに向かってゲインを得る戦術を獲得できる．*1

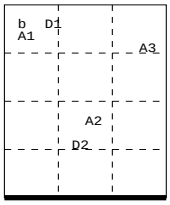

図 1　状態のグリッド化

グリッドによる適切な認知地図の獲得

戦術行動のトリガーとなる状態を有限状態に離散化するため，フィールド上の様子をグリッドで表現する．実際のサッカープレイヤーは 3 次元の視覚情報を 2 次元の鳥観図として持ち，その認知地図上で数秒後の予測や視覚外の状況の推定を行なうと言われている [麓 95]．離散化することで実時間での意思決定や学習を可能にしている．

他エージェントの行動の予測と，その学習

有限に絞り込まれた状態間の条件付き確率に基づき，一定時間後の将来の状態を予測する．ゲームによる経験から条件付き確率表を更新し，他エージェントの振舞いの傾向に関する予測の精度*2を向上させることができる．

状態の効用関数の学習

与えられた課題を達成できた時の有限状態のシークエンスに対し報償を与えることで，状態の効用関数が学習される．

*1　ただし，この戦術をフルゲームで利用するためには，より広域の戦術として，ミニフィールド内でいかに数的優位を保つか，という戦術が必要になるが，本研究では扱わない．
*2　ここでの精度とは他エージェントの行動（パスかドリブルか，等）の予測と実際の行動とが合致する頻度のことであり，精確な位置情報等の予測を意味しない．

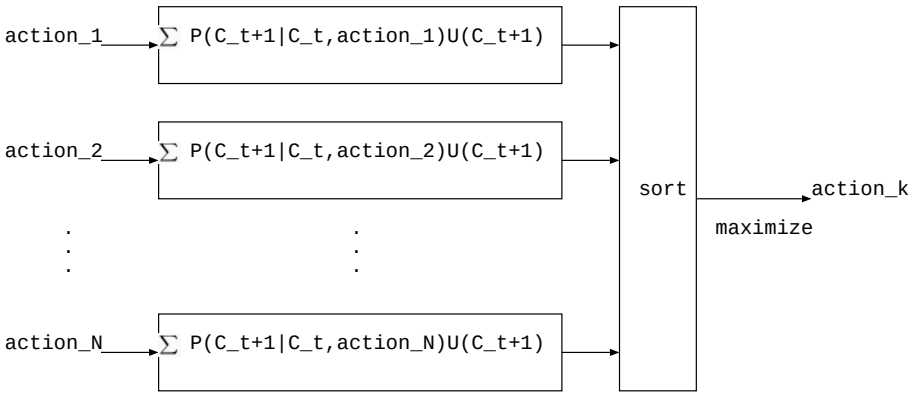

図 2 戦術決定のモデル

3·2 エージェントの構成要素—変数，確率，および効用関数

前節で挙げた方針を実現するエージェントモデルの構成要素を準備する．

状態変数： 3×4 のグリッドフィールド（図 1）上で以下のような典型的な状態を正規状態 C として変数化する．すべての状態を考慮すると，グリッド化しても膨大な変数ができてしまう．それを回避するため，ディフェンダがアタッカをマークしている状態のみを登録する．

- ディフェンダは必ずいずれかのアタッカをマークする．
- 少なくとも 1 人のアタッカはボールをキープしている．
- 少なくとも 1 人のディフェンダはアタッカにチャレンジする．
- 1 グリッドに存在するエージェントは両チームより 1 人ずつとする．

条件付き確率： ある状態下（$C_t=c_t$）でアクション A_t を実行した時，状態（$C_{t+1}=c_{t+1}$）が生起する条件付き確率を $P(C_{t+1}=c_{t+1}|C_t=c_t,A_t)$（小文字の表記は，3·3 節 で summation を取る際のことを考慮している）とおく．これは，エージェントの選択する戦術が，状態遷移に与える効果を記述する動作モデルである．

効用関数： ある状態 C の効用を $U(C)$ とおく．

グリッド化にあたって，ディフェンダにアタッカがぴったりと張りついた状態を正規状態として登録した．この処理は，実時間処理のために必要である．しかし，実際ミニゲームが始まると，正規状態には含まれない状態がしばしば生起する．本エージェントでは，そうした非正規状態になった場合，正規状態が回復するまでの間，標準の意思決定サイクルを用いず，例外処理で各々がアクションを選択する方法も試みた．しかし，この方法で望ましくないのは，非正規状態がしばしば起こるために，実質的に各エージェントが例外処理のみでゲームを進めるようなシークエンスが起こるということである．この時，エージェントには意思決定サイクルにおける学習の機会が与えられないだけでなく，それを用いて戦術決定する機会さえも与えられない．こうした事態を回避するため，非正規状態が生じた時，各々のエージェントが正規状態を一回のアクションで回復するようにプログラムした．具体的には，同一グリッドに味方が複数いる場合，背番号の若いエージェントが，別グリッドに移り正規状態を回復する．敵がフリーになっているとき移動すべきグリッドは，敵のいるグリッドである．アタッカならば正規状態で 1 人はフリーに成り得るので，前進，サイドへの移動の順に，正規状態を保持する配置を取る．これは各エージェントによって自律的に達成される．これにより，規則的に意思決定サイクルを動作させ，学習を効率よく行うことができる．

3·3 エージェントの構成

エージェントは以下のステップを繰り返す意思決定サイクル [Russell 95] として構成される．

Step. 1 期待効用を最大化する戦術の選択

$$\begin{aligned} action \leftarrow argmax_{A_t} \sum_{c_t} [Bel(C_t=c_t) \\ \times \sum_{c_{t+1}} P(C_{t+1}=c_{t+1}|C_t=c_t,A_t) \\ \times\ U(c_{t+1})] \end{aligned}$$

$\sum_{c_{t+1}} P(C_{t+1}=c_{t+1}|C_t=c_t,A_t)U(c_{t+1})$ により，現在状態 $C_t=c_t$ のときのアクション A_t の期待効用が求められ，現在状態に関する確率 $Bel(C_t=c_t)$ の期待値を取ることで総期待効用が算出される．この総期待効用を最大化するアクションを出力する．

Step. 2 期待状態の確率分布の予測

$$\begin{aligned} \widehat{Bel}(C_{t+1}) = \sum_{c_t} P(C_{t+1}|C_t=c_t,A_t) \\ \times\ Bel(C_t=c_t) \end{aligned}$$

上式より，条件付き確率表と現在状態の確率分布から次期状態に関する確率分布の予測 $\widehat{Bel}(C_{t+1})$ を求める．

Step. 3 知覚による確率分布の更新

$$Bel(C_{t+1}) = \alpha P(E_{t+1}|C_{t+1})\widehat{Bel}(C_{t+1})$$

上式では，新しい知覚情報 E_{t+1} を用いてアクション遂行後の確率分布の推定値をベイズの定理により更新している．$P(E_{t+1}|C_{t+1})$ はセンサモデルで，環境がどのようなデータをエージェント内部に生成するか，という確率である．フルゲームのように，知覚情報が不完全である場合には，確率分布の更新が必要になる．ちなみに，本研究のシミュレーションは，ゲームをミニフィールドに限定している．そのため，個々のエージェントはフィールド全体に関して完全情報を持っていると見做しているので，この Step. 3 は必要がない．

3·4 エージェントの学習

前節で述べた意思決定サイクルは一見簡潔に表現されているが，サッカーをするためには諸々の確率や効用をどのようにおいたらよいのかを，プログラマーがあらかじめ知ることは困難である．したがって，このエージェントはいくつかの学習を必要とする．サッカーエージェントが実時間で動作しなければならないことを考慮すれば，以下の 2 種類の学習モードがあることがわかる．

- 標準的な条件付き確率（ベイズ推定）と効用関数をオフラインで学習させておく．
- ゲームの最中にオンラインで学習させる．

ただし，本研究で構築されるエージェントにおいてはオフライン学習とオンライン学習の構造は同じであるから，敵の事前研究としても，試合中の戦術変更としても，以下の学習を用いることができる．

§ 1 条件付き確率表の学習

条件付き確率の更新は知覚情報から得られる推定信念の順序に基づき，増減する．推定信念は視覚外の情報を除けば，知覚に基づいた確実な情報であるから，これによって，推定信念が正である状態を結果とする条件付き確率の値はポイントを得る．ここで，w は学習効率のパラメータである．

$$\begin{aligned} & P_{new}(C_{t+1} = c_{t+1}|C_t) \\ = \ & P_{prev}(C_{t+1} = c_{t+1}|C_t) + wBel(C_{t+1} = c_{t+1}) \end{aligned}$$

§ 2 効用関数の学習

成功したイベントを構成するシークエンスに報償を与える適応学習法を採用する．例を用いて考える．

(1) 状態 C_1 において A1 がドリブル
(2) 状態 C_2 において A2 が移動
(3) 状態 C_3 において A1 から A2 にパス
(4) 状態 C_4 で A1 が移動
(5) 状態 C_5 で A2 から A1 にパス
(6) 状態 C_6 でドリブル $\rightarrow$ 成功

以上の流れは，ワン・ツー・パスとドリブルによるシークエンスを表しているが，この流れを作るために攻撃エージェントが行なったアクションと状態の組み合わせに対して報償を与える．従って，上記のパスやドリブルなどボールにからんだアクションだけでなく，フリースペースに走り込んでパスを待つ，といったアクションにも報償が与えられることになる．

報償の与え方は，

$$U_{new}(C_{N-i}) = U_{prev}(C_{N-i}) + \text{food} \times d^i \quad (1)$$

C_N はゲーム終了状態で，終了から 1 状態遡ると報償は $d(< 1)$ の割合で減衰する．

4. 実　　験

4·1 初期状態に関する仮定

本研究では，3 章で述べたように，ミニフィールドでの小数エージェントによる練習課題の達成を目的としたミニゲームを行い，これを通してエージェントに基本的な協調戦術を 3·4 節で述べた学習法により学習させる．Hughes によれば，ミニゲームの練習で最も重要なことは，指導者が，練習の初期状態を適切に設定することだ，という．ここで，初期状態とは，フィールド上でのボールやプレイヤの配置である．従って，以下の 2 つの条件を満たす状態を初期状態とした．

- ボールのあるグリッドの前方に少なくとも一人のディフェンダ側エージェントがいる，
- 他のエージェント全員を後方に残して単独のエージェントが前方で待ち構えていてはいけない．

例えば，図 3 の左に示されたエージェントとボールの配置は，初期状態としては不適当である．一方，図 3 の右の配置は，上記の条件を満たしている．

全正規状態のうち，324 状態のみがミニゲームの初期状態として適当である．

4·2 エージェントのスキルに関する仮定

本研究のシミュレーションは，サッカーの基本的なスキルはあらかじめ全エージェントに同等に与えられているものとしていた．ボールやエージェントなどの物理的要素は意図的にシミュレーションより除外した．物理的要素に関わるスキル，例えば，精確なパス，およびパスの運動軌道の予測なども考慮しなかった．主な方針を以下に述べる．

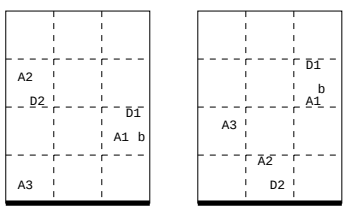

図 3 左：初期状態として不適切な例，右：初期状態として適切な例

- 全エージェントは同期して意思決定を行う．
- 同一グリッド下で，アタッカ，ディフェンダはボールの支配権を 1:1 の確率で得る．
- 支配権を得たエージェントは自分の意思決定に基づきボールを操作できる．
- 支配権を失うことをエージェントのスキル上のミスと見做す．
- 効用学習において，勝ちシークエンスのうちミスした状態を報償の対象から除外する．
- 同様に，負けシークエンスのうちミスした状態を罰の対象から除外する．

以上の方法により，スキルの点では同等な複数のエージェントによる戦術獲得の環境が提供できると考えられる．

4·3 シミュレーションの繰り返し回数

各ゲームでは 10 回の意思決定サイクルを繰り返し，10 回の意思決定で決着が付かなければノー・ゲームとする．各ゲームは，初期状態 324 の中からランダムに選ばれた初期状態からスタートする．324 すべての初期状態が，一回ずつスタート状態として選ばれ，ゲームがすべて終了した時点で，1 シミュレーションサイクルが終了した，と見做す．1 回の実験で，シミュレーションサイクルを 100 回繰り返す．

4·4 賞　罰　比

成功したゲームを構成する状態に対し報償を与えるとともに，失敗した状態には罰が与えられる．報償と罰の比率によりアタッカチームの勝率は変化する．そこで，さまざまな賞罰比によって予備的にシミュレーションを行った．その結果，賞罰をほぼ等率に与えたとき，学習によるアタッカの勝率の上昇が優れていた．よって賞罰比として 1:1 を採用した．

4·5 減　衰　比

ゲームの終端状態から過去の状態に与えられる賞罰は，状態を時間で一つ遡るごとに一定比率（式 (1) の d）で減衰が与えられる．さまざまな減衰比でシミュレーションを行った結果，減衰比を 0.85 としたときが，アタッカチームの勝率が安定して高い値を取った（図 4）．本研究における効用関数の学習方法は，タスクの達成から過去に遡って報償を減衰させる点において Grefenstette の profit sharing の方法 [Grefenstette 80] に類似している*3．一般的な profit sharing は成功における報酬による強化のみであるのに対し，本研究の扱う敵対環境下の学習問題には失敗系列が存在するため罰が与えられる．

*3 最適な減衰比については宮崎の定理 [宮崎 94] があるが，マルチエージェントシステムの強化学習に応用できるかどうかは明らかでない．

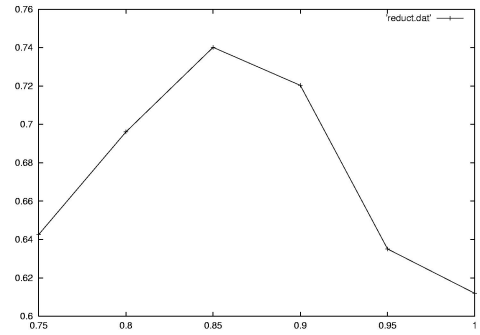

図 4 減衰比とアタッカの勝率（10 シミュレーションの平均値）

5. 結　　果

賞罰比 1:1，減衰比 $d = 0.85$ の下での実験結果について記す．

5·1 学習の有無に関する勝率の比較

学習の効果を，勝率の変化から推定する．勝率は，1 シミュレーションサイクル 324 ゲームのうち勝利した割合である．シミュレーションサイクルの繰り返しで，この勝率がいかに変化するかを見てみよう．本節で示すグラフ（図 5）において，縦軸はアタッカの勝率（50 回の実験の平均値）を，横軸はシミュレーションサイクルを示す．グラフ下（mean00）は，アタッカ，ディフェンダともに学習をしない場合のアタッカの勝率を示す．一貫してチャンスレベルである．グラフ上（mean85）は，アタッカ，ディフェンダともに学習をした場合のアタッカの勝率を示す．学習により，アタッカが勝率を伸ばすことが示されている．ディフェンダも学習しているが，平均してアタッカの勝率がチャンスレベルを有意に上回っているので，学習によって，人数差とそれにより可能になる基本戦術が，勝敗の決定要因の一つとして効いていると考えられる．

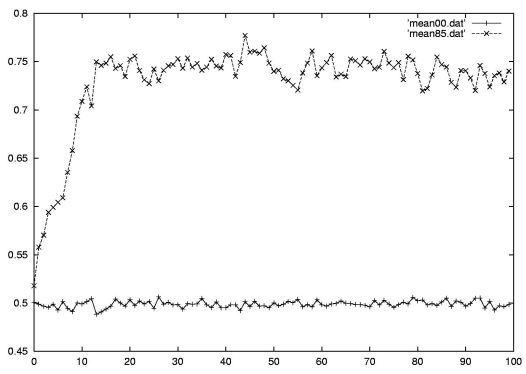

図 5 アタッカの勝率の推移（mean00 は学習なし，mean85 は減衰比 0.85 の学習，50 回の平均値）

5·2 基本的協調戦術の創発

本研究の第一の目的は，学習によるエージェントの協調戦術の獲得である．そこで，アタッカの勝率が最も良かった減衰比 0.85 の場合に，アタッカがどのような協調戦術を獲得しているかを調べた．本エージェントは学習によって，以下のような小人数による基本戦術を獲得した．図 6，図 7，そして 図 9 中の矢印は，プレイヤ，及び，ボールの移動を意味する．これらは，すべてシミュレーションの実例である．

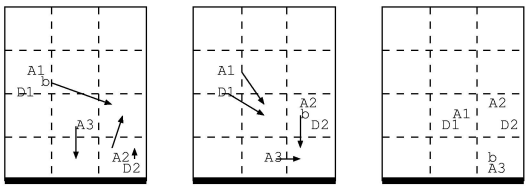

図 6 壁パス

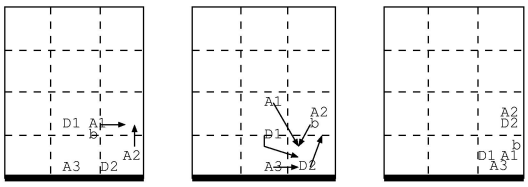

図 7 ワン・ツー・パス

壁パス： エージェント A1 は A2 にパスを出し，パスを受けた A2 は A3 にパスを出している（図 6）．

ワン・ツー・パス： エージェント A1 は A2 にパスを出した後，エンドライン際のグリッドに移動して A2 からのパスを受けている（図 7）．

ここに見られるようなパスは，ボールが味方から出されるのを受け手が現在位置で待つのではなく，スペースに移動してボールを取っている．こうしたパスを高次パスと分類し，シミュレーション毎の高次パスの数が学習により増加している様子を図 8 に示した．アタッカが勝利する場合の典型的なミニゲームのログを図 9 に示す．例えば，図 9 の Scene 2 では，アタッカ側により，「1 人がディフェンダを引き付けてスペースを作り，もう 1 人がパス受けに走り込む」というプレーが行われている．

6. マルチエージェント研究からみた本研究の意義

マルチエージェントの標準問題としてサッカーエージェントを考えた場合，次のようなことが要求されると考えられる．

- 動的環境の下で適応し自律的に振舞えるか．
- 不完全情報を元に最善を尽くせるか．
- 協調動作ができるか．

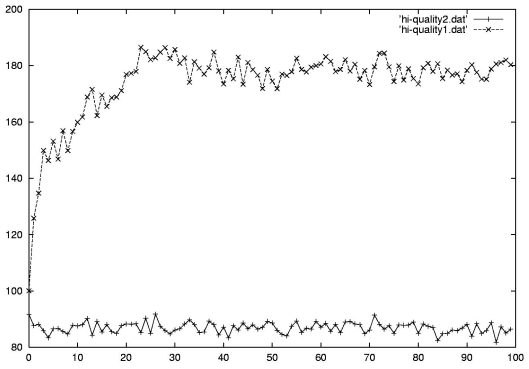

図 8 高次パスの数（high-quality1 は学習あり，high-quality2 は学習なし，50 回の平均値）

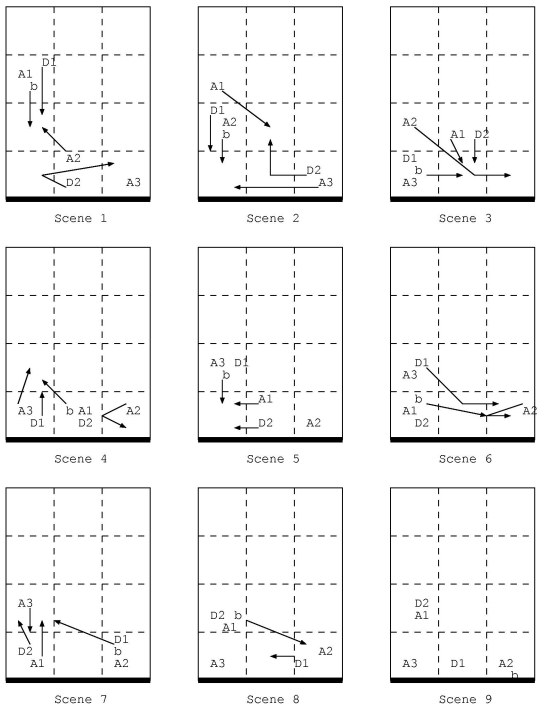

図 9 典型的なミニゲームの推移（アタッカが勝利する場合）

このようなエージェントの条件は，実際のサッカープレイヤと対等にプレイするためには自然で理想的である*4．しかし，2 次元のシミュレーションリーグでは勝利のために以上のような複雑な課題を遂行することの重要性は薄い．事実 RoboCup97 以来の強豪 CMUnited チーム [Stone 99] は，協調という側面からは “locker room agreement” という事前打合せタイプの協調のみで対応し，基本スキルの高さで他のチームを圧倒してきた．しかし，本研究の目的は RoboCup で競争力のあるエージェントを作ることではなく，最終的に実際のサッカープレイヤに類似した能力をエージェントに獲得させることであった．本エージェントは以上の要求を具体化した以下の 2 つの要求をどの程度満たしているのだろうか．

*4 人間のプレイヤとロボットエージェントが同一フィールドでゲームを戦うにはハードウェア上の問題点が多い．

（1）敵が動的に戦術を変更したとしても，本エージェントは条件付き確率を更新し，より正確な予測に修正し得る．
（2）各々の効用関数の学習を通し，協調にもとづく戦術が創発し得る．

本研究のシミュレーションの設定における初期の環境は「自分自身を含むエージェント群の振るまいの傾向（条件付き確率で表現される）をまだ学習していないエージェント群」によって構成されているため，エージェントが状態についての効用を学習し，その振るまいに偏りが現れるにつれ，エージェントにとっての環境は変化する．したがって，各エージェントは動的な環境下で学習を行っている．こうした条件下では，条件付き確率も効用も必ずしも学習によって最適解に収束する保証はないが，環境が動的である場合，つまり，他エージェントの振るまいが定常でない場合，それに対する適応性を持つ，というメリットがある．もちろん学習方法の性質上，環境が静的であれば，学習は収束する．従って，前述の要求 (1) は，ほぼ満たされていると考えられる．

さらに，協調性の獲得はサッカーにとって重要である．従来の研究では協調は導入されていたとしても，エージェントのコミュニケーション機能に依存するものがほとんどである．それに対し本エージェントは現段階ではコミュニケーションによらない協調が目標である．[Hughes 80] は練習段階におけるコミュニケーションの重要性を強調するが，本エージェントは，その段階を事前の学習によってクリアした．ある状況で各々がいかに動くべきか，についての知識を効用関数の形式で獲得するという学習が良い方向に向かっている場合，課題達成というチーム共通の目標のための共通認識が味方エージェント間に生まれる．つまり,「望ましい展開」というものが，学習された効用関数として味方の間に共有され，かつ，他のエージェントの大局的な振舞いについての予測により，チームに有利な状況を作り出すためのアクションを各エージェントが選択することが可能となる．そうして獲得された協調戦術に基づくプレーは，次のような特徴を持つ．

(i) 敵の動きを利用したプレーができる．
(ii) ボールを持たないエージェントによる戦術が明確に存在する．

(i) は，自分の動きに敵がどのように反応するか, という傾向についての知識（本エージェントでは条件付き確率で表現）を利用し，例えば「ディフェンダを引き付ける」ことである．図 9 の Scene 2 で，アタッカ 1 はディフェンダ 2 をアタッカ 3 から引き離す動きに成功している．

(ii) は，パスが出しやすい空間（「望ましい展開」）を積極的に作ることで，ボールを持っているエージェントとボールを持たないエージェントが協調することである．図 9 の Scene 2 で，アタッカ 2 が出したパスをアタッカ 3 は移動して取りに行ったことが確認された．従って，前述の要求 (2) は満たされていると考えられる．

このように本研究の学習方法は，ボールのタッチ回数やパスの成功等に応じて強化をするわけではないので，ボールに関与しないエージェント，および，課題達成に寄与しないエージェントも強化され得る．しかも，チームがタスクを遂行できなかった場合の系列を構成する状態は罰（負の強化）を得るため，学習の繰り返しによって不適切な学習結果は淘汰されていると考えられる．実際，図 8 に見られるように，高次パスのような，より適切なプレーの回数は学習によってほぼ単調に増加しており，この間パスの総数は平均して 600 程度で有意な変化が見られなかったことから，比率としても増加していると言える．この増加はアタッカの勝率の増加（図 5）と比較すると高い相関があると思われ，高次パスがチームの勝利というタスクの達成に貢献していると考えられる．学習途中で一時的に不適切な状態が強化されても，その強化は持続していない．学習初期における効率の問題は考えられるが，それ以上にボールを持たないエージェントの適切な意思決定を強化することが可能になっていることの重要性は高い．

これに対して，チャンピオン･チームの CMUnited[Stone 99] は，他エージェントやボールの位置について，現在の位置や速度に基づいた物理的な内部モデルで予測を行っているが，他エージェントの戦術的意思決定を考慮した予測は行っていない．また，適応学習を行うサッカーエージェントとしては，Andhill [Andou 98] や TIT Ohta [Ohta 98] がある．しかし，本研究がチームとしての連続的な戦術の獲得を目的とした学習であるのに対し，これら先行研究は各々のエージェントのホームポジションの適応などを学習対象にしている点が大きく異なる．

7. 今 後 の 課 題

本研究では，小人数局所フィールドにおけるミニゲームを想定したため，エージェントが運用する技術や獲得する戦術に多様性を欠いた．例えば，エージェントはミニゲームにおいてドリブルを全く用いなかった．これはフィールドが小さいので，ドリブルに適したスペースがないためと推測される．敵の戦術変化に対し，学習によって適応することができるかどうかも，学習の結果得られる戦術がシミュレーション系列によって異なっていないので確認することができなかった．フルゲームへの拡張により，多様性は得られると思われるが，本研究のモデルでは，学習に適したグリッド化をフルゲームのフィールドに対して行うと，計算量爆発を起こす．対処として，

- 様々な大きさのグリッドを階層化して併存させる，
- 関数近似を用いて状態変数を連続化する，

がある．

8. 結　　論

本研究は，現実のサッカー・プレイヤの戦術学習方法をもとにサッカーにおける協調学習のための認知モデルを抽出し，シミュレーションによってモデルの有効性を示した．学習課題は小人数でのミニゲームに限定し，条件付き確率により他エージェントの行動の予測と，適応学習による最適な戦術の獲得が可能なサッカーエージェントを構築した．結果として，前提条件をほとんど必要としない学習方法であっても，局所領域で有効な，壁パスやワン・ツー・パスのような協調戦術が創発することを示すことができた．

謝　　辞

本研究のアイディア段階から貴重なご意見をくださった野田五十樹氏（電子技術総合研究所）ら WAL 研究会の皆様に心より感謝いたします．セミナー等において数え切れないアドバイスをいただいた永野三郎教授，開一夫助教授ならびに植田研究室・開研究室・永野研究室（以上，東京大学）の皆様にも感謝いたします．なお，本研究の一部は，科学技術融合振興財団平成 9 年度研究助成および文部省科学研究費補助金基盤研究 (C)(課題番号：12680369) からの助成を受けています．

◇ 参　考　文　献 ◇

[Asada 96] M. Asada , S. Noda, S. Tawaratsumida, and K. Hosoda : Purposive Behavior Acquisition for a Real Robot by Vision-Based Reinforcement Learning. Machine Learning, Vol.23, 1996.

[Andou 98] T. Andou : Refinement of Soccer Agents' Positions Using Reinforcement Learning. H. Kitano(Ed.). RoboCup-97:Robot Soccer World Cup I, Springer, 1998.

[麓 95] 麓 信義. スポーツ心理学から見たサッカーの理論. 三一書房, 1995.

[Grefenstette 80] J.J. Grefenstette : Credit Assignment in Rule Discovery Systems Based on Genetic Algorithms. Machine Learning, Vol.3, 1998

[Hughes 80] C. Hughes. Soccer Tactics and Skills : British Broadcasting Corporation, 1980.（鈴木泰子訳，サッカーの戦術と技術，日刊スポーツ出版社, 1984）

[Kitano 98a] H. Kitano *et al* : RoboCup: A Challenge Problem for AI and Robotics. H. Kitano(Ed.). RoboCup-97:Robot Soccer World Cup I, Springer, 1998.

[Kitano 98b] H. Kitano *et al* : The RoboCup Synthetic Agent Challenge 97 Robotics. H. Kitano(Ed.). RoboCup-97:Robot Soccer World Cup I,

[Noda 96] I. Noda, *et al* : Soccer Server and researches on multi-agents systems. H. Kitano(Ed.). Proceeding of IROS-96 Workshop on RoboCup. 1996

[Ohta 98] M. Ohta : Learning Cooperative Behaviors in RoboCup Agents. H. Kitano(Ed.). RoboCup-97:Robot Soccer World Cup I, Springer, 1998.

[宮崎 94] 宮崎和光，山村雅幸，小林重信：強化学習における報酬割当ての理論的考察．人工知能学会誌，Vol.9, No.4, 1994.

[RoboCup Official Site] RoboCup Official Site : RoboCup The Robot World Cup Initiative. http://www.robocup.v.kinotrope.co.jp/

[Russell 95] S. Russell, and P. Norvig : Artificial Intelligence A Modern Approach, Prentice-Hall, 1995.（古川康一監訳，エージェントアプローチ人工知能，共立出版，1997）

[Stone 99] P. Stone, M. Veloso, and P. Riley : The CMUnited-98 Champion Simulator Team. M. Asada, H. Kitano(Eds.). RoboCup-98:Robot Soccer World Cup II, Springer, 1998.

〔担当委員：小林重信〕

1999 年 9 月 20 日　受理

著　者　紹　介

熊田　陽一郎(学生会員)

1972 年生．1997 年 東京大学 工学部計数工学科卒業．1999 年 東京大学 大学院総合文化研究科 広域科学専攻 修士課程修了．現在，東京大学 大学院総合文化研究科 広域科学専攻 博士課程在籍中.「窮すれば変じ，変ずれば通ず」という理念に基づき，動的な環境下で柔軟に自らの構造を変化させてゆくシステムを探求している．日本認知科学会学生会員．

植田　一博(正会員)

1963 年生．1993 年 東京大学 大学院総合文化研究科 広域科学専攻 博士課程修了．現在，東京大学 大学院情報学環・学際情報学府 助教授．博士 (学術)．科学的発見 (特に類推と協同)，図的推論，認知的インタフェース，人工社会・経済，マルチエージェント・システムと機械学習，などの研究に従事．著書に『科学を考える：人工知能からカルチュラル・スタディーズまで 14 の視点』(共著，北大路書房),『協同の知を探る：創造的コラボレーションの認知科学』(共編著，共立出版)．日本認知科学会，情報処理学会，Cognitive Science Society，AAAI 各会員．