3K-05　Profit Sharingを用いたサッカーエージェントのポジショニングの学習

鈴木　健介、入沢　達矢、小谷　善行
(東京農工大学　工学部　電子情報工学科)

1．はじめに

マルチエージェント環境において、複数のエージェントが協調行動を行うにあたり、その動作を事前に人手で記述しておくことは、状態空間が大きいことや状態遷移の不確実性により困難である。

本研究では、サッカーにおけるポジショニング問題を題材にとる。これまで多くとられていたポジショニング方法は、固定ポジション戦略であった。これに対し、本研究では、動的にポジションを変更する戦略をとり、行為－価値関数を用いてポジションを決定する。この行為－価値関数の重みを強化学習法 Profit Sharing により学習する。

また、ポジショニングを学習するにあたり、問題となるのはその評価である。それは、成功または失敗の要因がポジショニングの優劣によるものか、その他の技能によるものか判断が難しいからである。そこで、サッカーをモデル化したゲームを独自に作成し、ここで学習を行った。

2．モデル化したゲームでのポジショニングの学習

2．1　モデル化したゲーム

モデル化したゲームは、15×11に区切られたフィールド上で行う。また、1マスに相当する狭い範囲内では、人数が多い方が有利であると仮定してルールを設定した。主なルールは次の通りである。

- エージェント及びボールの移動は各ステップごとに一斉に実行される
- エージェントは1ステップに1マス、8方向に移動できる
- ボールと同じマスにいるエージェントの数が多いチームがパスの権利を持つ
- パスは8方向に3マスの範囲で、任意の地点へパスできる
- ゴールの隣接したマスではシュートができる

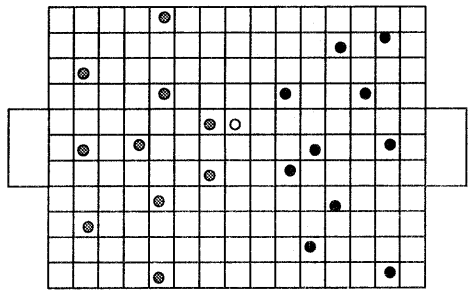

図1．モデル化したゲーム

2．2　行為－価値関数

行為－価値関数は、ある状態における行為の価値を計算する関数である。以下では、状態 X での行動 a に対する重みを $W(X,a)$ と表す。

2．2．1　状態

状態 X は次の通りに定めた。

$X=(x_0,x_1,x_2)$

x_0：自分のいる位置

x_1：ボールのある位置

x_2：パス権の状態

2．2．2　行動

行動 a は、移動する方向に相当し、8方向のいずれか、または移動しないかを示す。

2．3　行動の選択

行動選択にはルーレット選択を用いる。状態 X で行動 a^i を選択する確率 $p(a^i|X)$ は次の式で求める。

$$P(a^i|X)=\frac{W(X,a^i)}{\sum_{a^k\in actions(X)}W(X,a^i)}$$

$actions(X)$：局面 X で取りうる行動の集合

2．4　Profit Sharingによる重みの学習

Profit Sharingは経験強化型の強化学習法で、報酬を得られるまでの状態と行動の対を記憶していき、報酬が得られた時点で、その対に報酬を割り振る。重みの更新は次式で行う

$$W(x_i,a_i)\leftarrow W(x_i,a_i)+f(r,i)$$

Learning positioning behavior of soccer agent using Profit Sharing
Kensuke SUZUKI,Tatsuya IRISAWA,
Yoshiyuki KOTANI
Department of Computer Science,
Tokyo Univercity of Agriculture and Technology
2-24-16 Nakamachi Koganeishi Tokyo Japan

$f(r,i)$:強化関数

i:手番号

r:報酬

報酬をどの程度割り振るかは強化関数により求められ、過去になるにつれて割り振りが減ぜられていく。本研究では、報酬、強化関数及び重みの初期値を次の通りに定めた。

報酬：ゴールを決めた時点で1を与える
強化関数：

$$f(r,i)=r*(0.3)^{N-i}$$

N:ゴールを決めたときの手番号
重みの初期値：10.0

3．学習実験
3．1　実験方法
実験はゴールが決まった時点で1ゲーム終了とし、1,000,000 ゲームを自己対戦により学習を行った。

3．2　実験結果
学習の効果を測るために、対戦による実験を行った。対戦は、学習回数 100000 回ごとに実施し、1回の対戦は 10000 ゲームである。また、対戦エージェントは非学習エージェント及び、ボールを追いかける戦略をとるエージェントの2種類である。図2にその勝率の変化を示す。

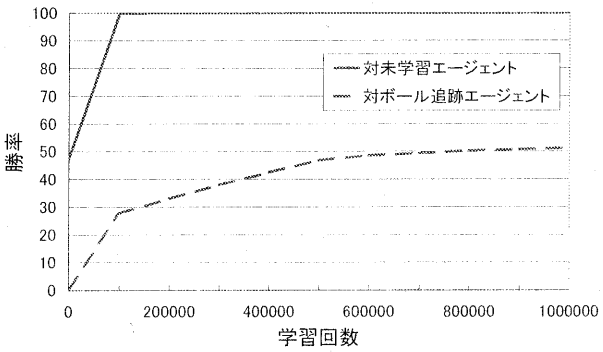

図2．学習回数による勝率の変化

次に図3は、エージェントの1マス前方にボールがある状態での行動選択確率の変化を示したものである。

4．Robocup サッカーでの実験
モデル化したゲームでの学習結果を Robocup サッカーでのポジショニングに適用し、対戦による実験を行った。対戦相手には、固定ポジション戦略を取り、その他の技能は全て同等であるエージェントを用いた。表1に対戦結果を示す。

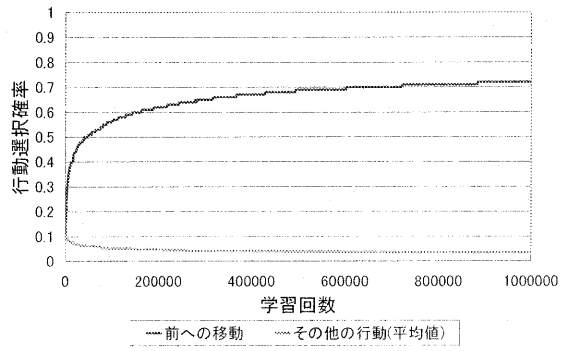

図3．学習回数による行動選択確率の変化

表1．Robocup サッカーでの対戦実験

	学習後	学習前
勝ち	28	24
負け	32	45
引分	40	31
得点	98	55
失点	99	94

5．考察
図2より、学習を重ねるにつれ勝率が上がっていることから、学習がよりゴールを決める方向に進んでいるといえる。また、人手で有効と思われる戦略を設定したエージェントと比べて、同等以上の成績を収めることができた.
図3より、前に進む行動、つまりボールを取りにいく行動が強化されていることが分かる。これは、ゴールを決めるためのより有効な行動が学習されていることを示している。
表1より、ポジショニングの学習の結果、チームの得点能力の向上ができた。

6．おわりに
本稿では、Profit Sharing によるサッカーエージェントのポジショニングの学習について述べた。その結果、次のことがわかった。
・ポジショニングの学習に Profit Sharing を用いることは有効である.
・モデル化したゲームで学習を行うことで Robocup サッカーでの、より有効なポジショニング戦略を獲得することができる。

参考文献
[1] 荒井幸代、宮崎和光、小林重信：「マルチエージェント強化学習の方法論—Q-learning と profit sharing による接近—」,人工知能学会誌, Vol.13, No.4, pp609-617,1998.
[2]安藤友人:サッカーエージェントにおける強化学習を用いたポジショニング RoboCup Workshop'97,1997.