B5-6 自律移動ロボットの行動獲得のためのファジィ状態分割型 Shaping 強化学習

福井大学 工学部 知能システム工学科 進化ロボット研究室
長谷川 大樹　(指導教員：前田 陽一郎)

1. 緒 言

未知環境を実ロボットに学習させる研究として強化学習がよく利用されている。しかしながら、他の最適化手法と比較すると非常に多くの学習時間を必要とすることも指摘されている。

また、一般にロボットに効率よく学習させるために生物の学習メカニズムから工学的模倣を行うのは有効であり、動物の調教やトレーニングなどに用いられる Shaping の概念をロボットの行動学習に応用する研究も行われている。田渕ら [1] は模倣学習の後に強化学習を行うことによって模倣学習の持つ欠点を補う方法を提案している。

そこで本研究では、当研究室で既に提案されている複雑なタスクをいくつかのサブタスクに分け、個々に行動決定ファジィルールを作成し、これらの選択ルールを強化学習により獲得する手法 [2] と、Shaping の概念を強化学習に取り入れた Shaping 強化学習手法 [3] を基に、Shaping の概念を取り入れ調教者が適宜報酬を与えることでロボットに効率的な学習をさせるファジィ状態分割型 Shaping Profit Sharing 法 (ShapingFPS) を提案する。またその有効性を検証するためにサッカーロボットシミュレータを作成し、従来の Profit Sharing 法と比較したので、その結果についても報告する。

2. Shaping について

Shaping とは犬やイルカなどの動物の調教に使われる用語であり、行動分析学の分野でも効率的に行動を強化するための有効な概念として注目されている。Shaping とは行動を形成するという意味であるが、やらせたい行動に行動を少しずつ強化しながら目標行動に近づけていく概念である。

Shaping についての実例として犬のフライングディスクを例にとって説明する。まず初めに最終目標を決める。次に最終目標を達成するために一連の行動を大きく単純化する。ここでの例は、犬が口にくわえるまでディスクを動かす、1m ほど投げて空中でくわえさせる、くわえてからディスクを持ってこさせるという 3 段階の難易度に相当する行動に分割する。

そして、Shaping を開始する。最初の間は毎回必ず餌などの報酬を与える。餌を与えられることが犬にとって好子の出現になり、その行動が強化される。ここで注意するのが好子を与えるタイミングと量である。次に、犬が全体の行動を間違いなくできるようになったら餌を与えるのをときどきにして「よしよし」などの声や身振りで報酬を与えるようにし、サブ目標行動を達成した時には大きな報酬を与える。こうして調教者を介して徐々に目標行動に行動を分化強化していく方法が Shaping である。

3. ファジィ状態分割型 Shaping 強化学習

図 1 に本手法で提案した階層型ファジィ制御手法の概念図を示す。本手法ではまず、学習に用いる入力情報をセンサなどから得られる環境情報をそのまま用いるのではなく、その前段階でマクロ環境状態ファジィルールを用いてマクロ環境状態にすることによって状態の次元数を減らしている [2]。これによって減少された状態を基に ShapingFPS では、獲得した報酬を入力とし、あらかじめチューニングされた下位の基本行動を選択する際に用いられる行動の重みを導出することを目的として、強化学習により重みに付加する行動重み ω_{ik} を学習し、正の報酬を獲得したときは重みを増加させ、また負の報酬を獲得したときは減少させる。

図 2 に本手法のフローチャートを示す。学習開始時には、すべての後件部シングルトンの初期値はある一定値を与えているのでどのサブタスクも同一の確率で選択され、ルーレット選択を行い探索する。そして、報酬 r_t が得られた場合、報酬が得られた時点を 0 ステップと考え、式 (1),(2) より h ステップ前の行動 i を取った場合の行動重みを更新する。

さらにここで調教者が適宜 Shaping 報酬 $S(t)$ を与え、人為的に報酬を変化させることにより、報酬 r_t が得られた行動に加え Shaping 報酬が与えられた行動も重要と認識し、別途報酬を割り引いて与えることにより、連続した行動系列を効率よく学習させる。

$$f(h) = \gamma^h (r_t + S(t)) \tag{1}$$

$$w_{ik} \leftarrow w_{ik} + \alpha f(h) \tag{2}$$

ω_{ik} : 行動重み　$f(h)$: 強化関数
α : 学習率　γ : 割引率　r_t : 報酬
$S(t)$: Shaping 報酬　$S(t) = c$ (c は定数)

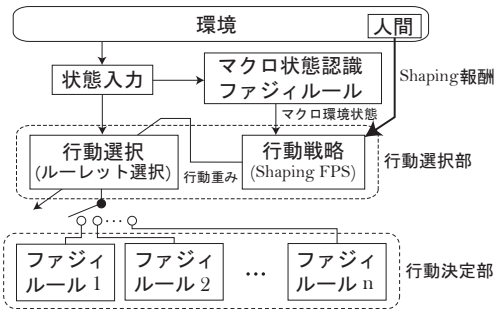

図 1 階層型ファジィ制御手法の概念図

4. シミュレーション実験

今回提案手法の有効性を検証するために、Linux 上で物理計算エンジン Open Dynamics Engine(ODE) を用いてサッカーロボットシミュレータを開発した。図 3 にシミュレーション画面を示す。

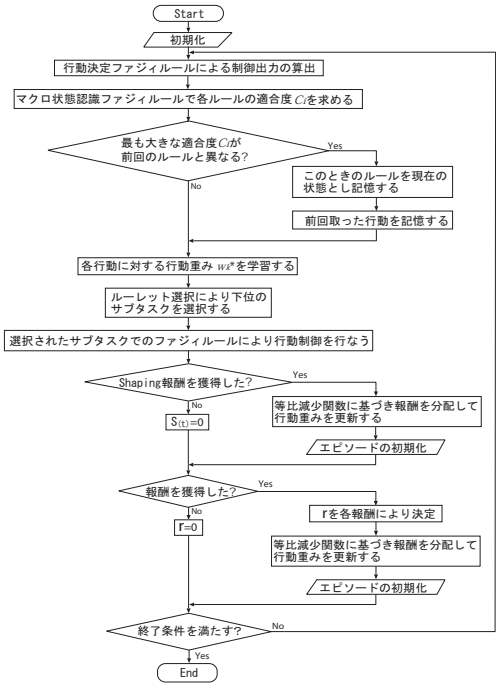

図 2 ShapingFPS のフローチャート

問題設定として、ロボットにシュート行動をさせた。ロボットはフィールド上にあるボールを障害物を回避しながら追跡し、ボールを捕獲したら同じく障害物を回避しながらゴールにシュートする行動をする。ロボットはロボカップ中型ロボットリーグ規格のサッカーロボットを想定し、移動機構にはオムニホイールを用いた全方向移動機構をもち、入力センサーには全方位カメラを搭載し、ロボット前方にはキック機構を搭載していると仮定している。

4·1 シミュレーション条件

障害物は半径 30cm 高さ 50cm の円柱をフィールド (12m × 18m) 上に等間隔に 20 個配置し、ロボットは初期位置からボールを捕獲し、右のブルーゴールを目指しシュートを行う。観測できる情報としてはボール、ゴール、障害物のそれぞれの相対距離と方位であり、ロボットの取り得る基本行動として回り込み、ドリブル、シュート、障害物回避を想定した。それぞれの行動は簡略化ファジィ推論を用いた行動獲得ファジィルールにより行動制御される。また、表 1 に示すような観測された環境情報を一旦マクロ環境状態に落とし、これと獲得した報酬や、Shaping 報酬を入力として ShapingFPS により学習する。学習後には、あるマクロ環境状態における行動選択ファジィルールの行動重みが獲得される。

表 1 の追跡度は、ボールが捕獲しやすい状態であるかどうかを認知させ、追跡行動を促す度合を表し、回避度は障害物に接近していてどれくらい危険な状態であるかを認知させ、回避行動を促す度合を表し、決定度は現時点でシュートすればゴールできるかを認知させ、シュート行動を促す度合を表す。

表 1 マクロ状態認識ファジィルールに用いた入出力変数

観測される環境情報 (入力)	マクロ環境状態 (出力)
ボールの相対距離、相対方位	追跡度 (Dc)
障害物の相対距離、相対方位	回避度 (De)
ゴールの相対距離、相対方位	決定度 (Dd)

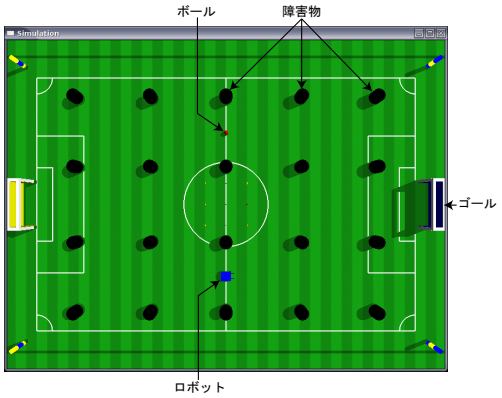

図 3 シミュレーション画面 (初期位置)

4·2 シミュレーション結果

図 3 に示す環境で提案手法である ShapingFPS と従来の FPS を比較するために、200 トライアルを 1 試行として、5 試行の平均をとった結果を図 4 に示す。両者を比較すると、提案手法の方がより速く学習が収束したことがわかる。このことから Shaping の概念を取り入れることによって効率よく学習が行われていることが確認された。

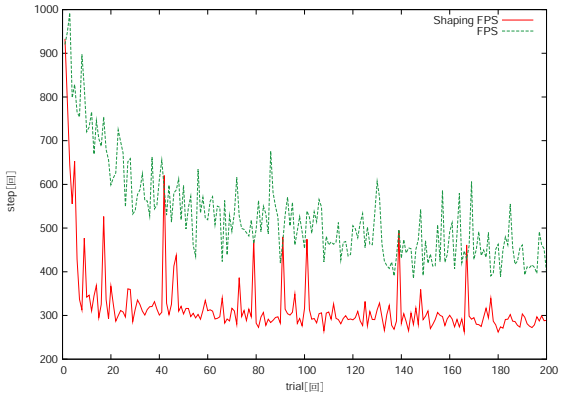

図 4 各試行におけるステップ数の推移

5. 結 言

本研究では、FPS に Shaping の概念を取り入れた ShapingFPS を提案した。Shaping を用いることでインタラクティブに効率的な学習が可能となり、短時間で効果的な行動戦略を獲得できることがわかった。今後の課題として、Shaping は複雑な環境になると報酬を与える回数が増え、調教者の負担が大きくなるといった問題があるので Shaping 操作自体を自動で行うシステムを目指していく必要があると考えられる。

参考文献

[1] 田渕一真, 谷口忠大, 椹木哲夫, “模倣学習と強化学習の調和による効率的行動獲得,” *The 20th Annual Conference of the Japanese Society for Artificial Intelligence*, pp.212-215 (2005)

[2] 花香敏, 前田陽一郎, “自律移動ロボットの戦略獲得のためのファジィ状態分割型強化学習,” 第 *23* 回日本ロボット学会学術講演会, CD-ROM, 3E12 (2005)

[3] 前田陽一郎, 花香敏, “調教の概念に基づいた自律移動ロボットの行動獲得支援手法,” 第 *23* 回ファジィシステムシンポジウム, CD-ROM, WD3-5 (2007)