

既得知識の利用に基づくロボットの行動学習への取り組み

金子 貴美 (指導教員：小林 一郎)

1 はじめに

本研究では、ロボットがある環境において獲得した行動知識を他の環境においても利用し、学習効率を向上させることを目的とする。ロボットが獲得した知識を他の環境に転用するにあたって、獲得された知識が人に可読であり、かつ様々な状況に対して部分的に適用可能であることが望ましいため、学習メカニズムにファジクラシファイアシステムを用いて、if-then 規則で表わされるロボットの行動知識を獲得する。そして、様々な環境で既に得た知識を、他の類似した環境で効率良く学習するよう適用する手法を検討する。

2 学習環境

環境に合わせて行動を学習させる手段としては、ニューラルネットワークと遺伝的アルゴリズム(以下 GA)を組み合わせるなど様々な方法が考えられるが、人にも可読性が高く、プロダクション規則でコントローラが構成されるクラシファイアシステムで学習させる。また、プロダクション規則をファジ推論規則とするファジクラシファイアシステム(以下 FCS)を適用することで、環境に柔軟に反応でき、かつ、少ない規則数でコントローラを構成することを目指す。

3 実験環境

3.1 実験仕様

実験環境として Simbad ロボットシミュレータ [1] を用い、FCS を構築した。また、環境に合わせて既得知識を利用することができるようにするため、FCS を拡張し、複数のルール群を部分毎に切り替えて適用、学習するシステムも構築した。

実験内容として、まず、ランダムに生成したルール群を図 1, 2 の訓練環境 (1), (2) でそれぞれ始点から終点まで学習させる。学習して得られた 2 つのルールを、切り替え学習システムを用いて図 3 のように適用し、目的を達成するために一貫性を持ったルール群として学習させる。また、図 3 の環境でランダムに生成したルール群を切り替えずに学習させ、切り替え学習システムとの学習効率を比較する。

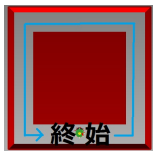


図 1: 訓練環境 (1)

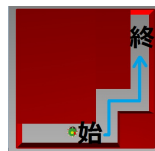


図 2: 訓練環境 (2)

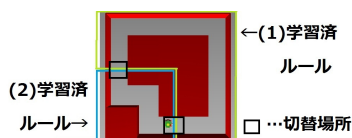


図 3: 学習用環境と、ルールの切替場所

3.2 FCS による学習

実験に用いる拡張 FCS を図 4 に示す。

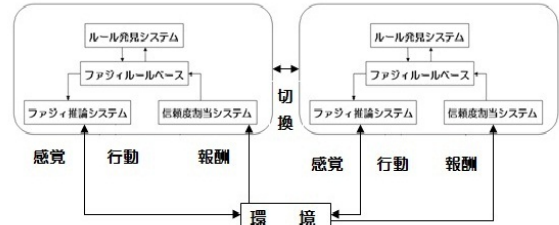


図 4: 拡張 FCS の構成

以下に、FCS の各部の内容について説明する。

● ルールベース

ルールは、ルールは、5 入力 2 出力 2 評価値の 9 遺伝子座で表現する。尚、前の 5 遺伝子が前件部(入力)、後件部(出力)である。

5 入力値はロボットに備え付けられた 5 つのセンサーから得られる値となり、センサーの入力値は 3 種類のメンバーシップ関数：Safe(S), Neutral(N), Danger(D) のいずれかで表現される(図 5 参照)。2 出力はそれぞれロボットの回転角と進行速度となり、2 評価値はルールの信頼度および、ファジ推論により得られたルールの適合度の累計となる。

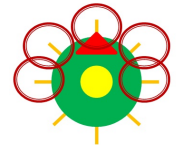
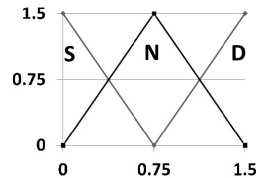


図 5: メンバーシップ関数 図 6: 使用するセンサ

● ファジ推論

センサーの入力を取りこみ、ルールベースの各ルールごとにメンバーシップ関数を用いてファジ推論を行い、値を出力する。推論方法は、重み 1 のファジシングルトン型推論法を用いる。

● 信頼度割り当てシステム

目的の地点、および中継点まで到達したら正の報酬を、到達する前に壁に衝突したら負の報酬を与えるとする。また、ルールの適合度に応じて与える報酬の量を決めるとする。

$$\text{正の報酬 (ルール } i) = \alpha \sum_x \text{ルール } i \text{ の適合度} \quad (1)$$

$$\text{負の報酬 (ルール } i) = (\alpha - \beta) \sum_x \text{ルール } i \text{ の適合度} \quad (2)$$

ここで α は各地点で設定した値, β は最後に報酬を与えた地点からのユークリッド距離である. x は, 最後に報酬を与えてから x 番目のサンプリングタイムである.

● 遺伝的アルゴリズム

目的の地点に辿り着いた場合, もしくは衝突した場合, ルール群に報酬を付与した後に GA を適用し, ルール群を学習させる.

GA においては, 信頼度の高さ上位 2/3 のルールは交叉対象とはせず保存しておき, 下位 1/3 のルールに対して一点交叉・突然変異を施すとした. 突然変異は変異率を 0.1 とした. 始めは前件部で行い, 前件部が重複するルールが生成された場合は新しく生成されたルールを削除する. その後, 後件部のみで突然変異を再度行うとした.

3.3 切り替え学習システム

FCS は環境全体における適切なルール群を学習するため, 学習した環境と類似する他環境でロボットを動作させる際に異なる部分のみ学習させることが出来ず, 再度その環境に適応するように初めから学習させねばならなくなり, 無駄が多い. そのため, 類似した環境に適応した知識を壊すことなく, 異なる部分の知識のみ再学習することを可能にするために切り替え学習システムを導入した. 切り替えはスムーズに行えるよう切り替え点におけるルール群の適用にメンバーシップ関数による重み付けを行い, 切り替えの整合性が取れずに衝突してしまう場合は衝突時に使用したルール群を FCS で学習させるとした.

具体的には, 座標を (a, b) と設定すると, 下記の範囲でメンバーシップ関数に従ってファジィに切り替えるとした.

$$a - n \leq x \leq a + n, \quad b - n \leq y \leq b + n$$

x, y はロボットの位置の座標であり, $2n$ の範囲でメンバーシップ値が 0~1 の値でスムーズに切り替わるようにした. 尚, 今回は環境の大きさが 10 前後と小さいので, $n=1$ に設定した. また, 報酬もメンバーシップ関数に従って与えられるよう設定し, 衝突時にメンバーシップ関数が 0 より大きかったルールのみ GA を適用するとした.

4 実験

4.1 実験設定

初期ルール群として個体数 70 個のルール A, B と 60 個のルール C, D を用意する. ルールはランダムに生成し, 2 つのルール群を使用する際の不均衡さを想定して個体数を意図的に変える. また, それぞれ 2 組のルール群を設定しているのは, 学習過程の相違による同じ目的を達成する異なる個体を比較のために示す. ルール A, B を訓練環境 (1) で学習させたルールを A', B', ルール C, D を訓練環境 (2) で学習させたルールを C', D' と表記し, 切り替え学習システムで複数ルールを使った場合は A+B と表記する.

4.2 実験結果と考察

実験結果を表 1 に示す.

表 1: 実験結果

事例	使用ルール	環境	世代数	ルール数
1	A	訓練環境 (1)	260	65
2	B	訓練環境 (1)	311	76
3	C	訓練環境 (2)	26	63
4	D	訓練環境 (2)	42	82
5	A'+C'	学習用環境	12	65+63
6	B'+D'	学習用環境	14	76+82
7	A	学習用環境	765	73
8	A'	学習用環境	210	60

‘世代数’は, 訓練環境 (1), (2) では始点から終点まで移動出来るようになるまでの, また学習用環境では 1 周出来るようになるまでの学習にかかった世代の数, ‘ルール数’は学習後の個体数である.

表 1 より, 切り替え学習システムを用いた場合の合計世代数は,

$$260(\text{事例 1}) + 26(\text{事例 3}) + 12(\text{事例 5}) = 298 \text{ 世代}$$

訓練環境 (1) の学習知識を学習用環境で FCS を用いて再学習した場合の合計世代数は,

$$260(\text{事例 1}) + 210(\text{事例 3}) = 470 \text{ 世代}$$

となり, 以下の関係が成り立っていることが分かる.

$$\text{切替学習 (298)} < \text{再学習 (470)} < \text{事例 7(765)}$$

切り替え学習システムを用いて学習したほうが学習効率が良いことが分かる. また, 切り替え学習システムを用いた場合は, A', C' のルール使用における一貫性は壊さず, 学習が必要だった C', D' のルール群のみで学習が出来ていた.

5 まとめと今後の課題

本研究では既得知識を利用し, FCS と切り替え学習システムの学習効率を比較した. 実験結果から, 切り替え学習システムを用いたほうが効率が良いことが分かった.

今後は, 既得知識の転用を学習効率向上に繋げるために, 切り替えの整合性の問題についてや, 学習方法についてより深い考察をし, より多くの環境, 知識の組み合わせで検証を行っていきたい.

参考文献

- [1] <http://simbad.sourceforge.net/>
- [2] 高野敏明, 高瀬治彦, 川中普晴, 鶴岡信治: 強化学習における異目的タスク間での知識の転移に関する一考察, 第 27 回ファジィシステムシンポジウム, 9 月, 2011.
- [3] 井上寛康, 高玉圭樹, 下原勝憲, 片井修: 知識を再利用するクラシファイアシステムの複数の環境への適応, 第 3 回 MYCOM, 2002.