

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

An Approach to Learning Robot's Behaviors based on the Reuse of the Knowledge Already Acquired

金子 貴美 小林 一郎
Kimi Kaneko Ichiro Kobayashi

*1 お茶の水女子大学大学院 人間文化創成科学研究科 理学専攻 情報科学コース
Advanced Sciences, Graduate School of Humanities and Sciences, Ochanomizu University

This paper discusses the efficiency of learning robot's behaviors by reusing the knowledge that a robot has already acquired under a particular environment. To reuse the knowledge at different environments, it should be transportable in part to various environments, therefore, the knowledge is represented as a collection of if-then rules of a fuzzy classifier system. We propose a switching system that applies proper behavioral knowledge to an objective environment, recycling the knowledge already acquired; and discuss the efficiency of learning robot's behaviors in comparison with ordinary fuzzy classifier system.

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ィ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ィ推論規則とするファジィクラシファイアシステム（以下 FCS）を適用し、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

実験環境として図 1 に示す 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。今回実験に使ったシステムの構成を図 2 に示す。これは、FCS を 2 つ組み合わせ、環境に合わせて適用するコントローラを切り替えるものとなっている。

このような切り替え学習システムを構築した理由は、FCS が「環境全体における適切なルール群を学習する」仕組みで、

図 3 に示す訓練環境 (1) で学習したルール群を図 5 の学習用環境で学習させると、学習環境に合うルール群を再度、初めから獲得することが必要となり、類似の環境において獲得された知識を利用して学習させたとしても必ずしも、その環境に対する学習効率が上がらなかったことを確認したためである。

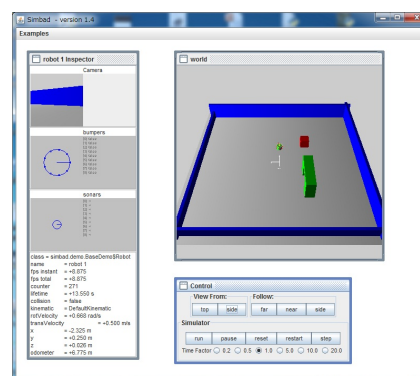


図 1: Simbad 概観

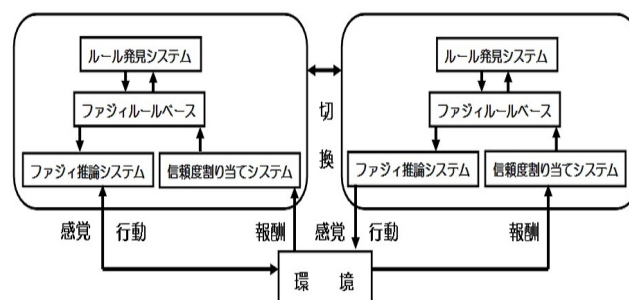


図 2: 拡張 FCS の構成

実験内容として、まず、ランダムに生成したルール群を図 3, 4 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、図 5 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切り替え学習システムを用いて学習させる。また、図 5 の環境で

連絡先: 金子 貴美, お茶の水女子大学大学院人間文化創成科学研究科理学専攻情報科学コース小林研究室, 〒 112-8610 東京都文京区大塚 2-1-1, 03-5978-5708, kaneko.kimi@is.ocha.ac.jp

ランダムに生成したルール群を既得知識を利用せずに最初から学習させ、切り替え学習システムとの学習効率を比較する。同様に、訓練環境と完全には一致しない、より複雑で未知な環境となる図 6 の環境においても学習の可否を含む学習効率の検討を行う。

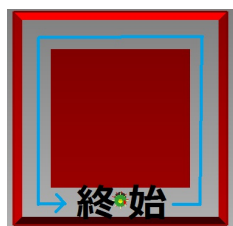


図 3: 訓練環境 (1)

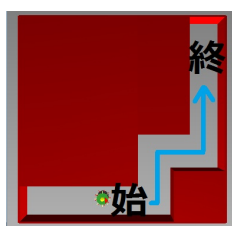


図 4: 訓練環境 (2)

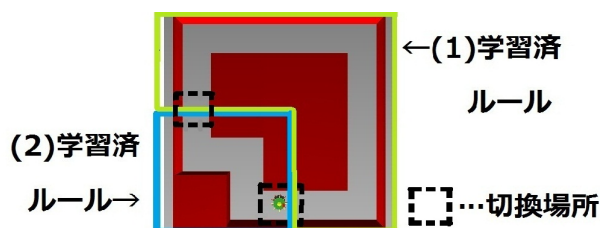


図 5: 学習環境 (1) とルールの切替場所

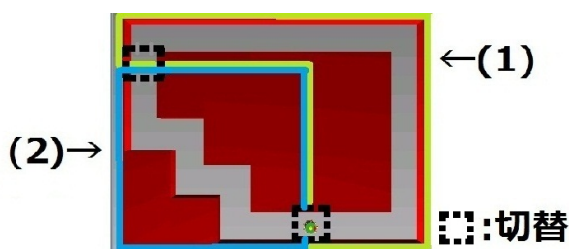


図 6: 学習環境 (2) とルールの切替場所

3.2 FCS による学習

図 2 に沿って、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、5 入力 2 出力 2 評価値の 9 遺伝子座で表現する。左から 5 つの遺伝子座が前件部 (入力) の情報を表わし、その次の 2 つの遺伝子座が後件部 (出力) の情報を表わしている。残りの 2 つの遺伝子座は規則の信頼度および適合度となる。後件部 (出力) である。

メンバーシップ関数の値					角度	速度	信頼度	適合度
S	N	D	S	S	0.0	0.5	1.0	1.5
入力 (前件部)					出力 と 評価値 (後件部)			

図 7: 遺伝子座の設定

5 入力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、障害物との距離の値: 0 ~ 1.5 となる。このセンサーの入力値は 0 ~ 1 の範囲に修正された後、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 8 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、ファジィ推論により得られたルールの適合度の累計となる。

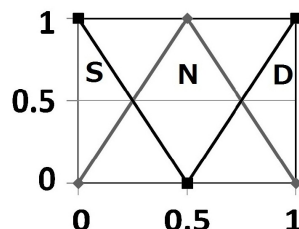


図 8: メンバーシップ関数

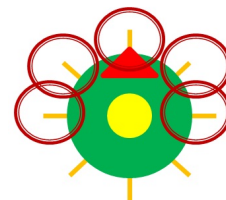


図 9: 使用するセンサ

• ファジィ推論

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ィ推論を行い、値を出力する。推論方法は、文献 [5] を参考に、他の推論法と等価でありながらより簡略的に表現出来るファジィシングルトン型推論法を用いることとし、重みを 1 とした。

• 被覆

推論の際、環境に適用可能なルールが存在していなかった場合、被覆を行い、環境に適用可能なルールを生成するとした。被覆は 1 ステップ毎に 1 個のルールを生成し、適用可能なルールが存在しないと判断される度に行う。このため、初期ルール群よりも学習済みルール群のほうが個体数が増えることがあり得る。

• 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以下のように決めるとする。

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

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

$$\beta = \sqrt{(x - x_0)^2 + (y - y_0)^2} \quad (3)$$

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り、 (x_0, y_0) は始点、もしくは最後に報酬を与えた地点の座標、 (x, y) は衝突地点の座標である。また n は、最後に報酬を与えてから n 番目のサンプリングタイムである。

負の報酬の与え方を (2) と設定したのは、 (x_0, y_0) から (x, y) の距離がより小さい場合に、より大きな負の報酬を与えられるようにするためである (図 10 参照)。

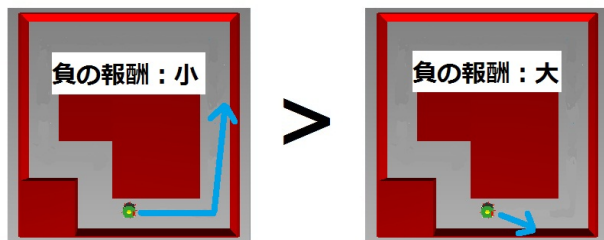


図 10: 負の報酬の与え方

● 遺伝的アルゴリズム

障害物等に衝突した場合、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。

以下に、今回実装した GA の詳細を図 11 の流れに沿って説明する。

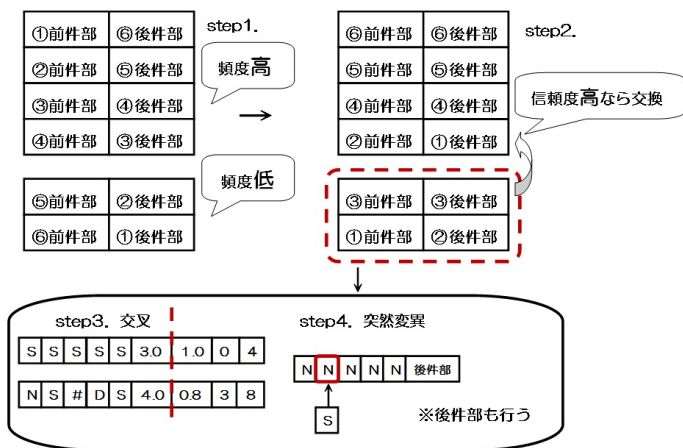


図 11: 遺伝的アルゴリズムの仕組み

step1. 利用頻度に基づく選択

全てのルールを利用頻度累積が高い順に並べ替え、利用頻度の高さが上位 2/3 のルールの集合と、下位 1/3 の集合に分ける。

step2. 信頼度に基づく選択

利用頻度下位 1/3 の集合の中に信頼度が高いルールがあれば、上位 2/3 の集合のルールと入れ替える。今回は入れ替える個体数はルール群の個体数の 10% とした。

これは、殆ど使用されないが、存在していないと困るルールを淘汰しないようにするために行われる。

step3. 交叉

下位 1/3 のルール群で一点交叉を行う。交叉させるルールは下位 1/3 から選択する。ルールの選択、交叉点の選択共にランダムで行われる。

step4. 突然変異

交叉後、突然変異を行う。突然変異は変異率を 0.1 に設定し、要素をランダムに選択して書き換えるとした。初めは前件部が重複するルールがあれば、新しく生成されたルールの前件部を突然変異し重複を

無くす。突然変異後も重複が無くならなかった場合は、新しいルールを削除する。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ず、効率が悪い。そのため、類似した環境に適応した知識を壊すことなく、再利用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、図 12 のように切り替えの整合性が取れず、ルール群とルール群の接続に適切なルールが無く衝突してしまう場合は、衝突したその時点において使用していたルール群のみを FCS で学習させるとした。

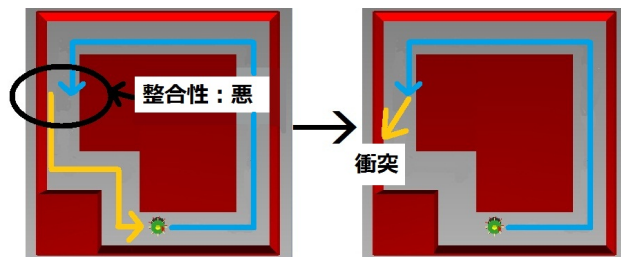


図 12: 切り替えの整合性

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。

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

x, y はロボットの位置の座標 (x, y) の値であり、 $2m$ 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、 $m=0.6$ に設定した。また、報酬もメンバーシップ関数に従い、重み付けをして与えられるよう設定し、衝突時にこの切り替えの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。

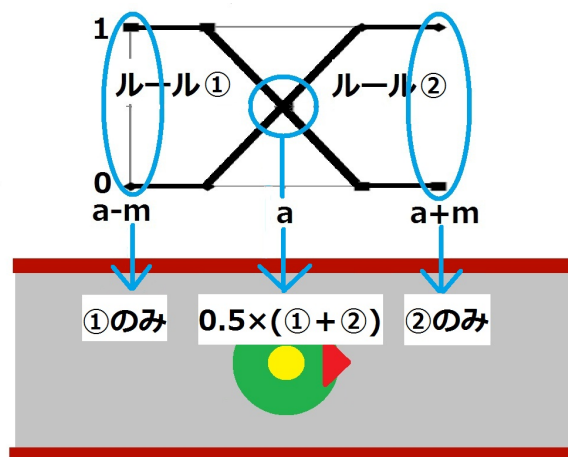


図 13: 切替のイメージ

4. 実験

4.1 実験設定

初期ルール群として個体数 70 個のルール群 A, B と 60 個のルール群 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し得る異なる個体を比較のために示す。ルール群 A, B を訓練環境 (1) で学習させたルールを A', B', ルール群 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のルール群を使った場合は A'+C' 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'	学習環境 (1)	38	65+69
6	B'+D'	学習環境 (1)	23	76+63
7	A	学習環境 (1)	765	73
8	A'	学習環境 (1)	210	60
9	A'+C'	学習環境 (2)	40	65+57
10	B'+D'	学習環境 (2)	81	76+49
11	A	学習環境 (2)	554	70
12	A'	学習環境 (2)	372	65

‘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, また学習環境 (1), (2) 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, ルール A' と C' の組み合わせの場合,

学習環境 (1): 260(事例 1)+26(事例 3)+38(事例 5) = 324 世代

学習環境 (2): 260(事例 1)+26(事例 3)+40(事例 9) = 326 世代

ルール B' と D' の組み合わせの場合,

学習環境 (1): 311(事例 2)+42(事例 4)+23(事例 6) = 376 世代

学習環境 (2): 311(事例 2)+42(事例 4)+81(事例 10) = 434 世代

訓練環境 (1) の学習知識を学習環境 (1), (2) で FCS を用いて再学習した場合の合計世代数は,

学習環境 (1): 260(事例 1) + 210(事例 8) = 470 世代

学習環境 (2): 260(事例 1) + 372(事例 12) = 632 世代

となり, それぞれ以下の関係が成り立っていることが分かる。

学習環境 (1): 切替学習 (324) < 再学習 (470) < 事例 7(765)

学習環境 (2): 切替学習 (326) < 事例 11(554) < 再学習 (632)

よって, 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, 表 1 の学習後のルール数からも分かるように A', B' のルールではルールの学習が殆ど行われず, 切り替え学習中も図 14 や図 15 は経路を維持したままであった。このことから, A', B' のルール使用における一貫性は壊さず, 切り換えの整合性を取るための学習が必要だった C', D' のルール群のみで学習が出来ていたことが分かる。

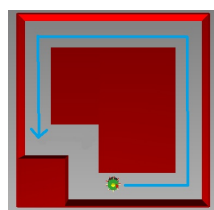


図 14: 学習環境 (1) 経路 (A', B')

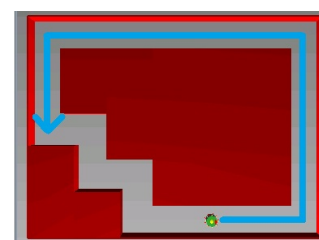


図 15: 学習環境 (2) 経路 (A', B')

5. まとめと今後の課題

本研究では, 特定の環境下において獲得されたロボットの行動知識を再利用可能にすることを目的とし, 知識を移植性が高く表現能力に優れた FCS を用いて表現し, 環境毎に知識を切り替える学習システムを提案した。提案システムと FCS による学習効率を比較し, 実験結果から, 切り替え学習システムを用い, 異なる部分環境を部分的に学習したほうが効率が良いことが分かった。

今後はより精度の高い学習を行えるよう FCS を改良していきたいと思っている。また, 切り替え部分の仕組み自体にも整合性の問題の原因があると考えられるので, 今後はこれらについて更に深く考察をし, より多くの環境, 知識の組み合わせで検証を行い, 既得知識の転用をより良い学習効率向上に繋げていきたい。

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。
- [4] 古橋武, 中岡謙, 森川幸治, 前田宏, 内川嘉樹, ファジィクラシファイアシステムによる知識発見に関する一考察, 日本知能情報ファジィ学会, 1995
- [5] 水本雅晴, 1. 代数積-加算-重心法=ファジィシングルトン型推論法=簡略化推論法, パイオメディカル・ファジィ・システム学会大会, 1990