

対話的進化ロボティクスの観測に基づく教示の設計*

片上 大輔[†]・山田 誠二[‡]

Interactive Evolutionary Computation for Real Robot from Different Viewpoints of Observation*

Daisuke KATAGAMI[†] and Seiji YAMADA[‡]

In this paper, we describe a fast learning method for a mobile robot which acquires autonomous behaviors from interaction between an operator and a robot. We develop a behavior learning method ICS (Interactive Classifier System) using interactive evolutionary computation regard for the teaching cost and a mobile robot is able to quickly learn rules so that an operator can directly teach a physical robot. Also the ICS is a novel evolutionary robotics approach using an adaptive classifier system. We classify teaching methods as internal observation (Learner View) and external one (Teacher View) from viewpoints of observation, and investigate relationship between observation methods and the results. We have two experiments based on our teaching methods on a real world.

1. はじめに

人間が活動する環境内において行動する自律ロボットでは、未知環境における行動、予期しない人間からのインタラクションなど、行動に必要な知識を事前に用意しておくことが難しい状況でタスクをこなすことが要求される。そこで、自律行動の獲得のための学習や環境への適用が必要となってくる。

近年においては、学習や適応の枠組みとして強化学習や進化計算手法を用いてロボットに自律的に制御器を獲得させる研究が注目を集めてきた。これらの手法の目的の一つとしては、制御器設計における人間の介在をできるだけ排除することがあげられる。このことは同時に、身体性や環境との相互作用ダイナミクスを制御器の構築に反映させる際に設計者による不適切・不必要なバイアスを排除することにつながるといえる。そのため、従来はできる限り知識を強化学習の枠組みにいれずにエージェントに試行錯誤させて学習することが前提とされて

きた。しかし、実環境問題に適用するにあたってその実行速度が問題となっている。

そこで、環境に存在する人間とのインタラクションを利用するアプローチが行われてきている。特に、アプリアリナ知識を持たないロボットや初期段階の試行錯誤のロボットにおいては、人間からの教示は非常に有効な自律行動の獲得手法であるといえる。しかし、ある程度自律性を持ったロボットにおいては、人間からの教示に完全に従う必要はない。教示が必要でない場合に関しては、人間に負担をかけることなく、人間とのインタラクションにより蓄えられた教示情報を元にして自律的に行動するべきである。このように、人間とロボットがインタラクションを通じて、ロボットの自律性を確立する手法が必要であるといえる。

麻生らは、人間と音声会話によるコミュニケーションを行う事情通ロボットによって、未知環境の地図情報を構築する枠組み [1] を提案しているが、人間とロボットとのインタラクションによりロボットの行動を獲得してはいない。石黒らは、強化学習により移動ロボットの状態空間の構築 [2] を行っているが、始めに人間が教示した行動を見本にして学習を行うが、その後はロボットが内部状態の構築を行うだけで、人間とのインタラクションは行われぬ。堀口らは、人間とロボットのインタラクションの設計として相互主導型インタラクションの概念を用い、力覚フィードバックを利用した移動ロボット

* 原稿受付 2002年10月17日

[†] 東京工業大学 大学院 総合理工学研究科 CISS,IGSSE,Tokyo Institute of Technology; 4259 Nagatsuta, Midori-ku, Yokohama 226-8502, JAPAN[‡] 国立情報学研究所 National Institute of Informatics; National Center of Sciences, 2-1-2 Hitotsubashi, Chiyodaku, Tokyo 101-8430, JAPAN*Key Words:* human-robot interaction, online learning, evolutionary robotics.

の自動化プロセスと人間の操作の協調行動を実現 [4] しているが、その学習結果をロボットの行動獲得には反映してはいない。稲邑らは、ユーザとの対話に基づいて Bayesian Network を用いて確率的にロボットの行動獲得 [5] を行っているが、進化計算手法により段階的に行動獲得を行う我々の手法と方法的に大きく異なる。

これらの研究に対して我々は、ロボットが動作する際に人間から適切な行為としての教示情報を受け取って、タスクを解決しうる状態認識・行為ルールの獲得を実現する手法を提案した。このような枠組みを Interactive Evolutionary Robotics (IER) とよび、今までにいくつかの研究を行ってきた [6]。ここで、教示の効果は一定ではなく、教示者の教示の仕方や環境によって教示の効果が大きく変わってくることがわかってきた。

本論文では、IER の枠組みにおいて、外部からロボットを眺める人間の視点と内部情報をもとに学習するロボットの視点とのずれが生じることによって、教示学習の効果に大きな差が発生することを実環境におけるオンライン学習により実験的に検証する。

2. 教示学習における観測の違い

一般的な知識を表現する上で行われる知識の抽象化においては、ロボットの状況認識の方法およびそれに基づく行動方策の表現方法が必要となるが、これらは身体性を持つロボットにとっては事前に明らかではなく、これを事前に設計する場合、設計の負荷が大きいことと、扱う対象を制限をしてしまうといった問題が生じる。ここで、身体性とは「環境や自身の内部状態を感知できる感覚能力、環境に働きかける多様な運動能力、それらを結ぶ情報処理能力が密に結合しさらに経験を通じて学習する機能」とする。そこでヒューマン・ロボットインタラクションを通して学習を行う教示学習を行うことになる。

そこでは、ロボットが実際に環境と相互作用を行いながらタスクを実行する過程で、ロボット自身は自律的に state-action からなるルールを獲得し、教示者はその実現の上で必要となる行為、すなわち各時点でのロボットの正しい動作をロボットに与えるという方法が考えられる。このようなアプローチにおいては、教示者は自らの外的な視点とロボットの視点との違いについては意識することなく、ロボットの行動に対して外部の視点からの正しい行為を与えるのみによって、ロボットに作業を教示することが可能である。つまり、ロボットが行う認識・行為決定が教示者の行為の情報から自らの視点に基づく状況認識に変換することによって、教示者・ロボット間の違いを吸収しているといえる。これによって、教示情報を少なくしていき、適した状況認識・行為決定を行うことができ、事前にロボットの身体性に依存した設計を行うことが困難である場合にも負荷の少ない教示を行うシステムを構築することができる。

ところが、外部からロボットを眺める教示者の視点か

らは、ロボットの視点から眺められた世界がいかにロボットによって解釈されるかが既知ではない。なぜなら、ロボットが身体性を通して世界と相互作用しているとき、それはロボットの身体性・環境・タスクに依存しているからである。よって教示者の負荷が少ない教示システムを構築することは可能になるがその効率性に問題が残る。つまり、このような問題では、外的な観測者の視点に基づく知識を直接ロボットに教示することは困難である。また、局所情報しか知らないロボットが特定の全域目標の実現に対してどのように自己の行動を修正したらよいのかを知ることは一般に困難である。しかも、ロボットには目標が実現していることさえ明確に認識できないかもしれない。目標の実現を認識できるのは、システム内部のロボットではなく、システムの外部から全体を眺めることのできる観測者である。一方システム内部の情報を知らずには整合性のあるルールを獲得することは困難である。結局、観測者の認知とロボットのそれにはずれがあるためそれを教示によってうまく伝えるのは難しい問題であるといえる。このような問題が教示手法にとって重要となってくると考えられる。

3. 観測者の認知に基づく教示法

前述のような認識のずれの問題はロボットの身体性の問題などで多く取り上げられてきたが、教示を行う際にどのような問題を引き起こすのかについての報告についてはあまり行われてこなかった。本研究では教示を行う際に、その認識のずれがどのような影響を与えるのかについて検証するために、簡単な設定を用意した。本研究では、システム外部から全体を眺めることのできる観測者の認知を用いて教示を行う方法 (Fig.1) を外的観測法とよぶ。逆にシステム内部から状況を判断するロボットの認知を用いて教示を行う方法 (Fig.2) を内的観測法とよぶ。

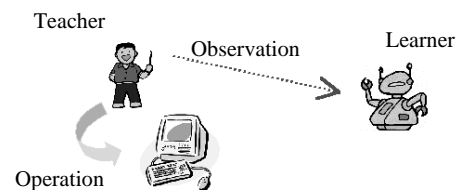


Fig. 1 External observation

ロボットの認知と外的観測者による認知の間には一般に大きな差異がある。例えば実環境上でロボットを動かした場合、センサ情報の獲得によって形成されるロボットの内部モデルは、その状況および観測の履歴に依存して大きく異なったものになってしまう。その結果、ロボットはあらかじめ設計者が設計した行動規則に従うものの、状況の変化に対応して設計者の意図する行動を

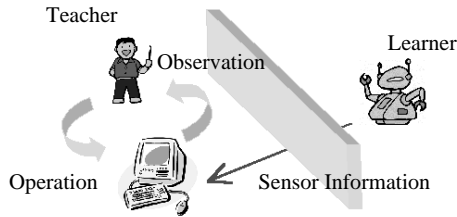


Fig. 2 Internal observation

生成しないことがある。これは、外的観測者（ここでは設計者）の想定するロボットの環境認識とロボットの実際の認識像が異なっているために生じるものである。さらにその環境が複雑で、外的観測が困難であればあるほどそのずれは大きくなることになり重要な問題となる。特に教示を行う場合はこのようなギャップは重要である。

例えば、Fig.3のようにロボットの状況認識の精度が悪い場合、教示者が行為a5を教示した時点でロボットには行為a5を行う条件部の認識、すなわち内部状態の分割が行われていないためにロボットには教示情報がうまく伝わらず、行為a5の教示はできない。また逆に、Fig.4のように教示者側が環境における変化を認識していない場合、教示者の内部状態の中では行為a5というものは存在しない。よって、ロボット側が点線で示した内部状態の分割が可能であったとしても、教示がなされないために教示による学習は出来ないこととなる。このような、認識のずれが環境に多く存在する問題下において、いわゆる教示学習はうまくいかないといえる。

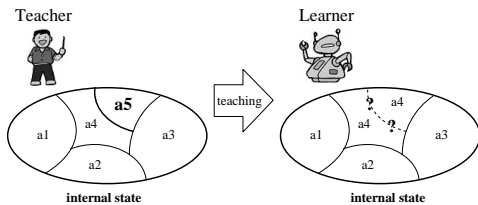


Fig. 3 When robot don't perceive the difference of states

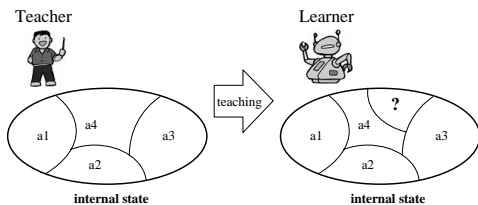


Fig. 4 When teacher don't perceive the difference of states

三島、浅田らは、この違いを Cross Perceptual Aliasing[3] とよび学習者と教示者の間の内部状態の違いを問題にしている。従来の教示における学習がうまくいかなかったのは、この認識のずれが問題であったためであり、従来の対話的進化計算においては、提示された解候補の

評価を逐次行うことで学習を進めて行くため、システムの内部に人間の評価を取り入れているにもかかわらずその観測手法は外的でありその評価能力を生かしきれていないともいえる。

本研究では、外的観測法と内的観測法を用いた教示に基づく対話的計算法により進化ロボティクスにおけるオンライン学習を行いその違いを検証する。我々は、この対話的計算法による進化ロボティクスを実現するために、対話的手法に基づいたロボット学習システム、対話的クラシファイアシステム (ICS) を構築した。Fig.5にICSが想定する教示の環境の例を示す。ここでは、ロボットは行動ごとにセンサ情報をシステムに出力し、教示者はそのセンサ情報やロボットを直接観察しながらジョイスティックを用いて、ロボットを直接操作し、タスクを達成することで教示を行うものとする。

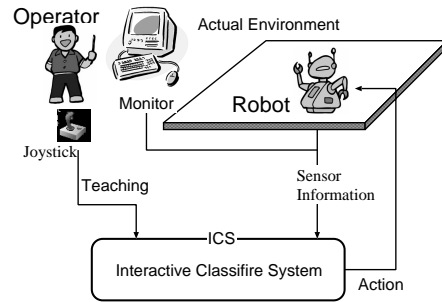


Fig. 5 Teaching environment

4. 対話的クラシファイアシステム

ICSは学習分類子システム (Learning Classifier System: LCS) にIECの対話機能を組み込むことにより、自律的な学習に加え教示による学習も行うことができるロボット学習モデルである。学習アルゴリズムであるLCSにはWilsonが提案したXCS[7]を使用している。XCSはZCS[8]を改良したもので、精度 (accuracy) とよばれるパラメータを追加したものである。クラシファイアは#(don't care symbol) を条件部に含むことによって一般化を行うことができるが、HollandのクラシファイアシステムやZCSはそれを効果的に行うことができなかった。これは、クラシファイアシステム自体が一般化を適切に進めるためのメカニズムを備えていなかったことに起因しており、過度に一般化 (overgeneral) されたクラシファイアによってシステムの性能が悪化するという現象が報告されてきた [9]。XCSでは特に、過度に一般化したクラシファイアを抑制するために、クラシファイアの有効性を、従来の強度だけではなく精度によっても決定している。この精度はクラシファイアを実行した結果受け取った報酬とその予測値との誤差により計算され、これによりルールを過度に一般化せずに適切に一般化できることが報告されている [7]。また、Bookerによって提案された遺伝的アルゴリズムの戦略の一つである制限

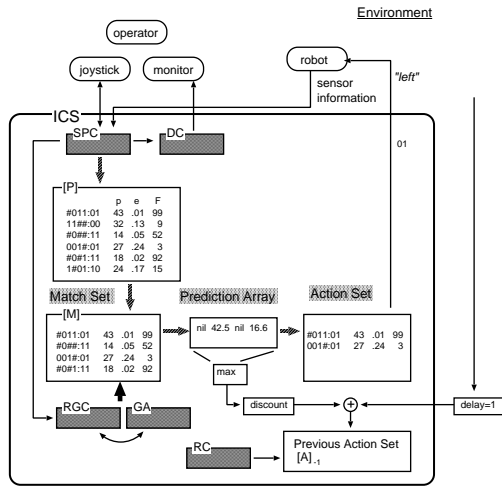


Fig. 6 Overview of interactive classifier system

交配 (restricted mating) の導入により, 探索を安全かつ効率良く行うことができる. 構築したシステムの概要図を Fig.6 に示す.

本研究で開発したシステムは, 操作者の教示情報をもとにクラシフィアを作成するルール生成部 (RGC), ロボットに装備した近接センサと CCD カメラの画像情報を処理するセンサ処理部 (SPC), GUI インタフェース等の表示部 (DC), 学習を行う強化学習部 (RC) からなり, すべて Linux 上で開発されている. C 言語および GTK+ で記述されている. 以下, 各モジュールを説明する.

【RGC】 Rule Generation Component は, 教示によるルールの作成を行う. 教示者はロボットをインタフェースに表示される情報を見ながら, ジョイスティックを用いて操作し, そこでの操作履歴とその時のロボットのセンサ情報をセンサ処理部 (SPC) が受け取り, それより RGC が新しくルールを作成しルールリストに加える. ルールの作成手続きは, 主に XCS[7] を基本に教示情報からルールを作成できるように改良した.

- (1) システムはロボットのセンサ情報 X と教示情報 a_t を SPC から受け取る.
- (2) 集団 $[P]$ から X にマッチしたクラシフィアがマッチセット $[M]$ に移され, システムは $[M]$ で表されたおのこの行為 a_i を支持するクラシフィアの Prediction 値を Fitness 値で正規化して $P(a_i)$ を作成する. $P(a_i)$ の値は Prediction Array に置かれ, 行為が選択される. 行為選択は, 決定論的行為選択もしくは, ルーレットホイール選択により行われる.
- (3) 行為選択により選ばれた行為 a_j と教示により得られた行為 a_t を比較し, $a_j \neq a_t$ ならば, $[M]$ の中で行動部に a_j を持つルールの行動部を a_t に書き換える. $a_j = a_t$ ならば, 変更はしない.
- (4) 選ばれた行為 a_j を支持する $[M]$ の中のクラシフィアからなる行動セット $[A]$ を作成する. 行為 a_j

は効果器に送られ, a_t の入力があった場合は, すぐに報酬 r_{teach} が与えられる. a_t の入力がない場合も報酬 r_{imm} が環境から返される (返されない場合もある).

【RC】 Reinforcement Component は, クラシフィアシステムにおける強化学習部である. 前のステップのクラシフィアのパラメータを更新することで学習を行う. 教示者の操作がないときは, ロボットはそれまでに作成されたルールから自律的に行動を行うことができる.

【DC】 Display Component は, SPC により処理されたデータの表示を受け持つ. CCD カメラの画像処理には Video4Linux の API と Imlib を用いている. 開発したインタフェースを Fig.7 に示す.

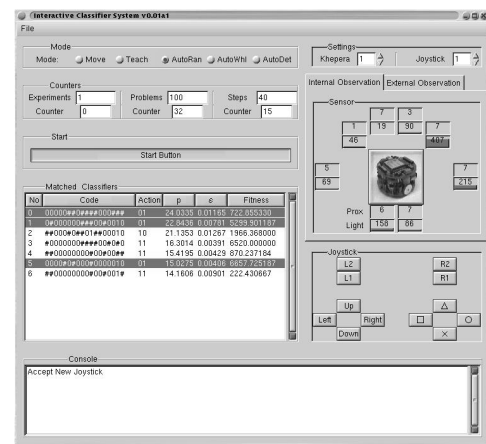


Fig. 7 User interface

【SPC】 Sensor Processing Component は, ロボットの各種センサの処理および教示情報の処理を行う. 処理されたデータは DC および RGC に送られ表示とルールの作成を行う.

5. 実験

5.1 実験 1: 光源到達・障害物なし

外的観測法と内的観測法を用いた教示による効果の違いを調べるために, オンライン実ロボットの比較実験を行った. Fig.8 にその環境を示す.

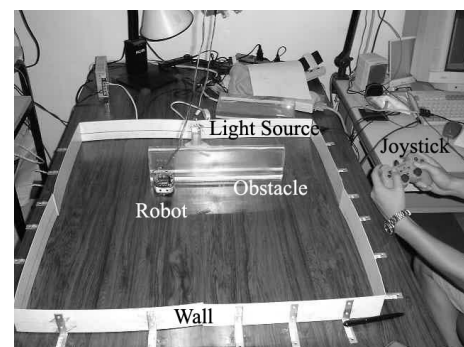


Fig. 8 Experimental environment

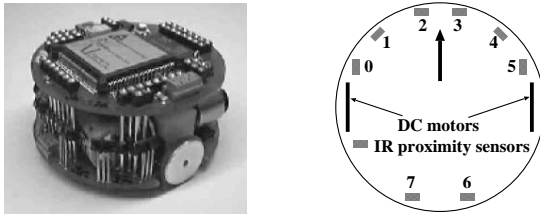


Fig. 9 A mobile robot: Khepera

移動ロボットとして、Kheperaを用いる。Kheperaは、直径55[mm]、高さ32[mm]、重さ70[g]で、モトローラ68331、RAM 256[Kbyte]、ROM 512[Kbyte]を搭載している。また、DCモータ（ロボットの移動1[s]あたり8[mm]）を二つ、赤外線近接センサと光センサが一体になったものをFig.9の右図の位置に八つ装備している。このセンサの値は、0~1023である。また、ロボットは無線タレットを装備しており、システムとの無線による通信が可能である。また、ジョイスティックにはSONY社製アナログコントローラDUALSHOCKを用いている。

タスクは、白いプラスチック版に囲まれた領域のなかに光源を置き、任意の初期位置からその光源にたどり着くまでのステップ数を最小にすることである。外的観測法（ここでは、Teacher Viewとよぶ）と内的観測法（Learner Viewとよぶ）を用いた教示による方法と、従来法である単純にクラシファイアシステムにより学習する方法（Auto）と比較した。

一つのクラシファイアは条件部と行動部からなる20ビットの文字列で表示される。条件部は{0,1,#}から、行動部は{0,1}からなるビット列で表され、「#」はdon't careシンボルとよばれ、ルールの一般化を行うのに用いられる。例えば#000#100000##00#01で表されるクラシファイアは、条件部が左の18ビットの文字列で表されている。最初の8ビットは、移動ロボットKheperaの回りに装備された赤外線近接センサの反応（0~255）を経験的に閾値で2値化した値とし、また光センサの反応を同様に2値化したものを、次の8ビットとする。さらに、環境の知覚騙し問題に対応するために1ステップ前のロボットの行為をその次の2ビットに表現している。行動部は右の2ビットの文字列で「11」が前進、「01」が左旋回、「10」が右旋回、「00」が後進を表している。

適合度関数 F は以下のように定義した。各光センサの値 $light_i$ の合計値を正規化して、シグモイド関数を用いた値を適合度 F とした。Payment Range は係数であり、ここではXCSの文献[7]の値と同じ値を用いた。

$$u = \sum_{i=0}^7 light_i \frac{PaymentRange}{light_{MAX} \times 8}, \quad F = \frac{1}{1 + \exp(-u)}$$

外的観測を行う場合は、教示者が環境全体を見渡せることで、大域目標の実現を簡単なものにするだろう。一方、ロボットには知覚できるが観測者には知覚できない

認知におけるセンサの認識のずれが生じる、または、知覚しているが意識しないような小さな障害物を見落とすことで学習が収束しないことがあるかもしれない。本研究では、この外的観測法を表現するために、環境全体を見渡せるカメラを設置し、教示者はそのカメラの情報をみながら教示を行う。インタフェースに表示されるカメラの情報をFig.10に示す。



Fig. 10 External observation settings

内的観測を行う場合は、ロボットの知覚をそのまま観測者が知覚することができるため、教示から素直に学習が進むといえるが、環境全体を見渡せることができない。この内的観測法を表現するために、ロボットのセンサ値を表示するGUIを用いる。センサ値を数値だけで見ても直感的に分かりにくいいため、同時にグラフで表現できるようにした。教示者はこれを見ながら教示を行う。インタフェースに表示されるセンサ情報をFig.11に示す。

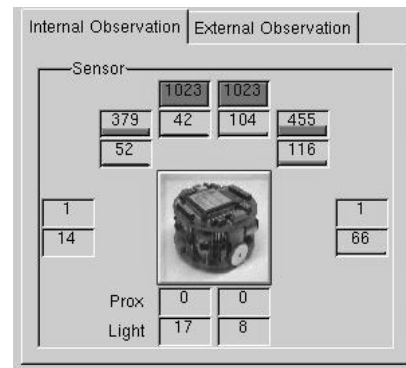


Fig. 11 Internal observation settings

ロボットが40ステップを移動するか、もしくは光源にたどり着くかを1試行として、教示として5試行ランダムな初期位置から始める。さらに検証として獲得されたルールを用いて自律行動によりあらかじめ決められた5ヶ所の初期位置からそれぞれ1試行ずつ評価を行う。この手続きを7回、合計で35試行の学習と35回の検証を行った。Table 1に実験のパラメータを示す。

5.2 実験1結果

実験1において、5ヶ所の初期位置から光源にたどり着いた回数(Fig.12)を求めた。教示なしの場合(Auto)は35問題までにほとんどGoalに到達するような行動を

Table 1 Experimental Parameters

Parameters	Value
number of problems in one experiment	35
number of experiments	1
maximum size of the population	300
probability to do crossover	0.8
probability of mutating one bit	0.04

獲得できなかったが、本手法において15問題でいくつかGoalにたどり着くような行動を獲得し始め、35問題までに8割から10割の確率でGoalにたどり着く行動を学習することができた。かかった学習時間は、1問題の教示に約1分、35問題行うのに4,50分で行うことができる。Goal到達回数をみるとTV, LV共にあまり差はなく、環境の認識にずれが少ない場合観測の仕方によって学習効果にほとんど差は見られなかった。

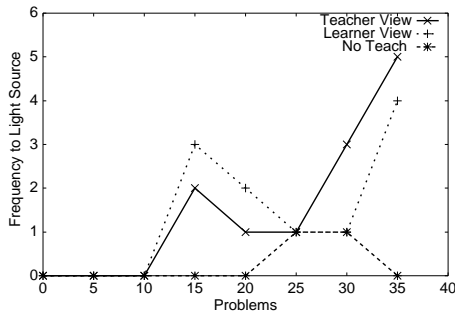


Fig. 12 Frequency to light source

5.3 実験2：光源到達・障害物あり

タスクの違いによる教示の違いを調べるために、実験1の環境にさらに障害物を導入した。これにより障害物を回避しながら光源にたどり着かなければならないため、外的観測を行う場合と、内的観測を行う場合では観測情報に違いが生じやすく、結果教示方法に差が出やすいと想定される。障害物は透明なプラスチック版によって作られており、光源の光を通すことができる。つまり、ロボットは障害物越しに光源方向を知ることができ、LVにおいて光源方向の特定が難しくなることを防いだ。実験のパラメータはTable 1と同じである。以上の設定により外的観測法 (Teacher View) と内的観測法 (Learner View) を用いた教示による方法を比較した。

5.4 実験2結果

ロボットが5ヶ所の初期位置から光源にたどり着いた回数 (Fig.13) を求めた。初めの段階ではあまり差が出ないが、実験1と異なりTVの方は一つの場所ではゴールにたどり着くものの教示からの学習がうまく進まずに、Goal到達回数を増やさない。一方、LVの方は40ステップで6割Goalにたどり着くことができた。実験1の場合は環境が近接センサの情報をほとんど使用せず、光セン

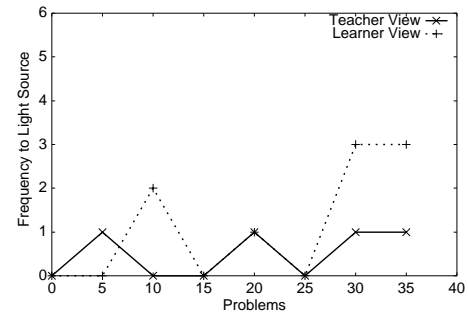


Fig. 13 Frequency to Light Source

サの情報だけで有効なルールを獲得できるため、教示者とロボットの認識のずれが少なく、二つの教示法にあまり差はでなかった。しかし実験2の場合、全体を見渡せた外的観測の方が一見教示がうまく進む気がするが、教示者がロボットを外から観察して予想したロボットの状態と実際のロボットの状態との間にずれがあるため、教示者の教示をうまく学習していない。実際にはロボットの内部状態を確認しながら教示を行った内的観測の方が有効なルールをうまく作成できている。

6. 考察

結果の違いに関する考察と、ICSの効果およびその問題点、今後の課題について議論する。

6.1 一般的な教示の例

本実験のTVとLVについて、それぞれ行われた教示の典型的な例をFig.14とFig.15に示す。TVによる教示については、教示者は環境を俯瞰的に観察するためロボットの詳細なセンサ情報の変化はわからないが、障害物の位置をあらかじめ把握することができる。よって、教示者はロボットを直接光源方向へ進ませることはせずに、あらかじめ障害物をよけてから光源方向に向かおうとする教示が行われている。この時ロボット側では、離散化された内部状態に変化はないのに教示信号が変わるという現象が何度か発生する。そのため、ある内部状態に対応して教示する行為が発散し、うまく学習するのは難しいといえる。

一方、LVによる教示については、教示者はロボットの状態をセンサパネルより詳細に見ることができ、内部状態に変化があった場合すぐに観測することができる。俯瞰的な観測はできないが、光源方向は光センサによりわかるので、まず光の方向に向う。そして障害物に近づき始めて障害物を認識し、迂回しながら光源方向を目指す教示となる。ここではロボットは障害物との衝突もあり、あまりスマートな行為獲得とはいえないが、内部状態の変化に対して行為が割り当てられ学習がうまく進んだといえる。

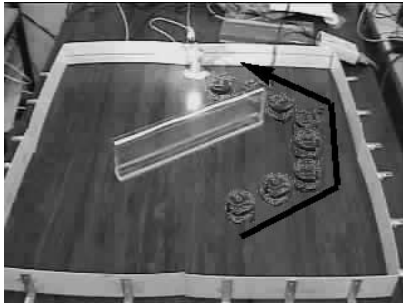


Fig. 14 TV example of teaching

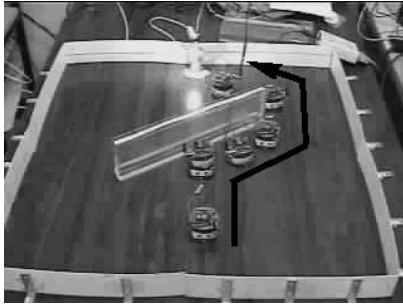


Fig. 15 LV example of teaching

6.2 教示の効果

35 試行後に内的観測によって作成されたルールの内、予測報酬の値が高いものから 4 ルールを Table 2 に示す。それぞれ、前に光があるとき前に進む、右に光があるとき右に曲がるなどの常識的で効率的なルールが作成されている。ここで、Prediction 値はそのクラシファイアの行為 a_i が出力された時に内部または外部の受け取った報酬の平均であり一般の Classifier System の強化値に相当する。また、Teaching は教示から作られたルールであることを示している。この実験においては、集団数 300 の上位 50 のルールの中に教示により作られていないルールは三つしかなく、9 割以上が教示されたルールで占められていた。Prediction の値が高いものが人間の教示により作られたルールであることから教示により有効な行動の獲得がうまく進んでいることがわかる。

Table 2 Created rules by LV method

Condition	Action	Prediction	Teaching
0#0#0####0#10##1#	10	413.7	yes
0010###0##000#1###	01	364.0	yes
00###0#000#1#0#01#	11	292.0	yes
###000000101001011	11	256.0	yes

6.3 教示者の知識による影響

今回の実験では著者によって教示が行われているが、実験環境について詳しく知っている人間が内的観測によって教示しているので良い結果が得られるのではという疑問が考えられる。そこで、事前知識を持つ教示者とその知識を持たない教示者によりその違いがでるのかを

実験により調べた。実験の設定等は前述の実験 2 のものと同様の設定とした。実験に用いるロボット、センサ情報、プログラミング等の知識を持たない教示者グループ ($G1=s-1,s-2$) と、持つ教示者グループ ($G2=s-3,s-4$) の 2 グループを用意し、教示者の知識による影響を検証する。50 試行をそれぞれ 5 試行の繰り返しとし、 $G1$ と $G2$ の光源までの到達回数合計をそれぞれ調べた。Table 3 に示す。

Table 3 The sum total of the number of attainment

Teacher	1	2	3	4	5	6	7	8	9	10
G1(amateur)	1	4	5	0	4	2	2	5	2	3
G2(expert)	0	3	3	1	2	5	4	3	2	3

ここで $G1$ と $G2$ において教示効果に違いがあるのかを調べるために、棄却域の確率 5% で分散分析を行ったところ、 $F(1,18) = 0.7781; p > 0.05$ により $G1$ と $G2$ の間に有意差がなかった。つまり、教示者の事前知識によって教示の効果は変わらないことがわかった。おもしろいことに、 $G2$ の方が到達回数が少ない場合がある。これは、教示の際に内部情報を確認しながらロボットに教示情報を教える LV においては、ロボットのセンサ機能、システムの学習方法を知り、環境を不完全にモデル化しようとするエキスパートと、環境をモデル化できないため内部情報を逐次確認しながら慎重に教示を行う素人の方との間に教示効果において差がなかったものと考えられる。しかし、教示にかかった時間に関しては、 $G1$ が $G2$ に比べ約 1.5 倍ぐらい時間がかかっていた。

6.4 認識の違いと解決方法

前述のように、身体性を持つロボットに対して教示を扱う場合、Fig.3, 4 のようなギャップが生じてしまう。ロボットの身体性の特徴としてロボットの世界観測の能力は限定されており、ロボットは全体的な視点に基づく大域的な状況認識を行うことが不可能である。つまり、ロボットは自らの観測能力の範囲内における局所的な観測入力に基づいて状況認識を行う必要がある。このようにロボット外部の世界における状態空間においてはマルコフ性が成り立つが、ロボットの観測の不完全性によりマルコフ性が失われるという状況における行動決定問題は、部分観測マルコフ決定過程 (POMDP: Partially Observable Markov Decision Process) [10] としてモデル化することができ、近年では POMDP 上での行動獲得を扱った学習手法 [11] が提案されている。POMDP 上でロボットが現在の状態を特定するには、現在の状態における観測値だけでなく動作を行うことで得られる観測情報を蓄積し、これにより状況を判別する機構が必要となる [12]。本論文では、クラシファイアの条件部に 1 ステップ前のロボットの行動を加えているがこの効果について検証してはならず、この点に関しての詳細な考察

は今後の課題である。

ここで、ICSで用いている学習アルゴリズム、XCSはHollandのクラシファイアシステムとは異なりメッセージリストが存在しない。このためXCSは時系列的な情報を扱うことが難しく、従ってマルコフ性を持たない問題においては十分な学習が困難であるといえる。よって、当然本研究のICSも時系列的な学習は難しく、大きな制約となっている。しかしこの問題に対してはクラシファイアシステムの方で時系列な学習を実現する研究[13]が報告されており、それらの手法の適用が考えられる。

7. おわりに

ヒューマン・ロボットインタラクションを通じてロボットの自律行動を獲得していく提案手法を用いて、観測の観点から教示を外的観測による教示と内的観測による教示にクラス分けし、実機を用いた実験においてその効果について調査した。設定した実験環境において認識のずれが少ない単純なタスクにおいてはあまり差はでなかったが、認識のずれが生じるタスクにおいては内的観測における教示が効果を示した。認識のずれを判別し、認識のずれがある場合には内的観測における教示を行い、ない場合には外的観測による教示に戻ることで、最小の負荷で最大の教示効率をあげることができると考える。

また本研究では、ロボットの実環境における高速な学習を可能とし人間には記述困難な複雑なロボットプログラムを簡単な教示をすることで学習、獲得することを示した。本研究の各実験における学習は40~50分(PentiumIII 500MHz, RAM128M)を必要とする。しかし、オンラインで進化的に獲得させようとすると、「光源到達」という単純な行動でさえも、条件部と行動部の組み合わせを変えて移動ロボットを走らせてみるといった試行錯誤を何回も繰り返す必要があり、この作業は相当の労力を必要とする。さらに、障害物がある環境になると適切な行動が判断できなくなり、その設計はより難しくなるといえる。よって、簡単な教示により自動的に適切な行動を獲得できることは、価値があると考えられる。

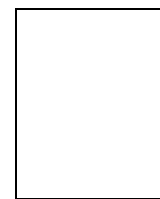
参考文献

- [1] H. Asoh, Y. Motomura, I. Hara, S. Akaho, S. Hayamizu, T. Matsui: Combining probabilistic map and dialog for robust life-long office navigation; *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 807-812 (1996)
- [2] H. Ishiguro, R. Sato, T. Ishida: Robot oriented state space construction; *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1496-1501 (1996)
- [3] C. Mishima, M. Asada: Active learning from cross perceptual aliasing caused by direct teaching; *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1420-1425 (1999)

- [4] Y. Horiguchi, T. Sawaragi, G. Akashi: Naturalistic human-robot collaboration based upon mixed-initiative interactions in teleoperating environment; *IEEE International Conference on Systems, Man, and Cybernetics*, pp. 876-881 (2000)
- [5] 稲邑, 稲葉, 井上: ユーザとの対話に基づく段階的な行動決定モデルの獲得; *日本ロボット学会誌*, Vol. 19, No. 8, pp. 983-990 (2001)
- [6] D. Katagami, S. Yamada: Interactive classifier system for real robot learning; *IEEE International Workshop on Robot and Human Interaction*, pp. 258-263 (2000)
- [7] S. W. Wilson: Classifier fitness based on accuracy; *Evolutionary Computation*, Vol. 3, No. 2, pp. 149-175 (1995)
- [8] S. W. Wilson: ZCS: a zeroth order classifier system; *Evolutionary Computation*, Vol. 2, pp. 1-18 (1994)
- [9] D. Cliff, S. Ross: Adding temporary memory to ZCS; *Adaptive Behavior*, Vol. 3, No. 2, pp. 101-150 (1994)
- [10] A. R. Cassandra, L. P. Kaelbling, M. L. Littman: Acting optimally in partially observable stochastic domains; *12th National Conference on Artificial Intelligence*, Vol. 2, pp. 1023-1028 (1994)
- [11] 木村, L. P. Kaelbling: 部分観測マルコフ決定過程下での強化学習; *人工知能学会誌*, Vol. 12, No. 6, pp. 822-830 (1997)
- [12] 井上, 太田, 小林, 新井: 部分観測環境下における自律的状態分割による強化学習; 第12回自律分散システム・シンポジウム資料, pp. 223-226 (2000)
- [13] S. W. Wilson, D. E. Goldberg: A critical review of classifier systems; *The Third International Conference on Genetic Algorithms*, pp. 244-255 (1989)

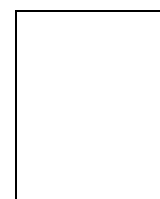
著者略歴

かたがみ だいすけ
片上 大輔 (正会員)



1973年9月3日生。1996年電気通信大学電気通信学部機械制御工学科卒業。1999年東京工業大学大学院博士前期過程(知能システム科学専攻)修了。2002年同大学院博士後期課程修了。同年同大学院総合理工学研究科助手。現在に至る。人工知能学会, 日本ロボット学会, IEEE各会員。

やまだ せいじ
山田 誠二



1960年10月11日生。1984年大阪大学基礎工学部卒業。1989年同大学院博士課程修了。博士(工学)。同年大阪大学基礎工学部助手。1991年同大学産業科学研究科講師。1996年東京工業大学大学院総合理工学研究科助教授。2002年国立情報学研究所教授。現在に至る。工学博士。人工知能, 特に, 知的Web, ヒューマンエージェントインタラクションに興味をもつ。情報処理学会, 日本ロボット学会, 電子情報通信学会, AAAI, IEEE, ACM各会員。