

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
東京工業大学大学院
総合理工学研究科知能システム科学専攻

山田 誠二
東京工業大学大学院
総合理工学研究科知能システム科学専攻

Abstract- 従来、ロボットの学習分野ではパラメータ調整が問題となり、人間が意図する行動を学習しない場合が多かった。さらに実環境強化学習では、学習の収束に時間がかかり、特に報酬を得るまでの初期学習に大きな時間コストがかかってしまう。そこで本研究では、対話型進化計算 (IEC) の評価能力を用いた効率のよい実環境ロボット学習を提案する。初期学習の効率化を図り、人間の意図を学習するクラシファイアシステム (CS) に基づく対話型分類子システムを提案し評価する。本研究では、対話型進化ロボティクスにおいて、教示主体における認知を教師におく場合 (外的観測法) と学習者におく場合 (内的観測法) にわけて考え、タスクの違いによってそれぞれの効果を確かめる。

Interactive Evolutionary Computation for Real Robot from Viewpoint of Observation

*Daisuke Katagami
CISS, IGSSE, Tokyo Institute of Technology
4259 Nagatsuta, Midori-ku Yokohama 226-8502, JAPAN
katagami@ymd.dis.titech.ac.jp

Seiji Yamada
CISS, IGSSE, Tokyo Institute of Technology
4259 Nagatsuta, Midori-ku Yokohama 226-8502, JAPAN
yamada@ymd.dis.titech.ac.jp

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

1 はじめに

従来、ロボットの学習分野では、その環境において最適な行動を獲得するために、これをパラメータの最適化問題と考え、その調整を行ってきた。評価においても、人間の評価系の代替モデルを作り、これを最適化システムに組み込んで探索する方法が行われてきた。しかし、評価関数やその他のパラメータ調整がうまくいかず、人間が意図するような行動を学習しない場合が多かった。

そこで、人間を評価系に組み込むというアプローチで進化的に探索を行う、対話型進化計算法 (Interactive Evolutionary Computation (IEC))[1] が行われてきた。

ここでは、人間と機械との相互作用によって主観的評価を行うことができるが、毎回評価を行わなくてはならない操作者の肉体的および心理的疲労が問題となっている。

また一方、実環境強化学習では、学習の収束に時間がかかり、特に報酬を得るまでの初期学習の立ち上がりには大きな時間コストがかかる。しかも1回の行動に必要な大半の時間は、ロボットの感覚、行動系の処理時間に費やさ

れるので、高速に学習するためには、学習試行数の削減自体が必要である。

そこで本研究では、人間を評価系に組み込む IEC の評価能力を用いて効率のよい実環境ロボット学習を行うことを目的とする。これによりロボット学習分野における初期学習の効率化を図り、また人間が意図するような行動を学習することを目的とする。ここでは、教示者の負荷を軽減するために教示は任意の時に行われ、オペレータが毎回教示を行うことは必要ではないこととする。

従来では試行錯誤で学習をしていたため何千何万回の試行が必要であったが、本手法では、人間の教示によりルールを作成することで、初期段階の簡単なコンテキストに基づいたデフォルトルールの作成が容易に行える。また、より環境に適したデフォルトルールの作成が行えるため、もっと詳細なコンテキスト情報に基づいた、より例外的なルールベースの階層をうまく作るようになるといえる。

本研究では、このような枠組みを **Interactive Evolution-**

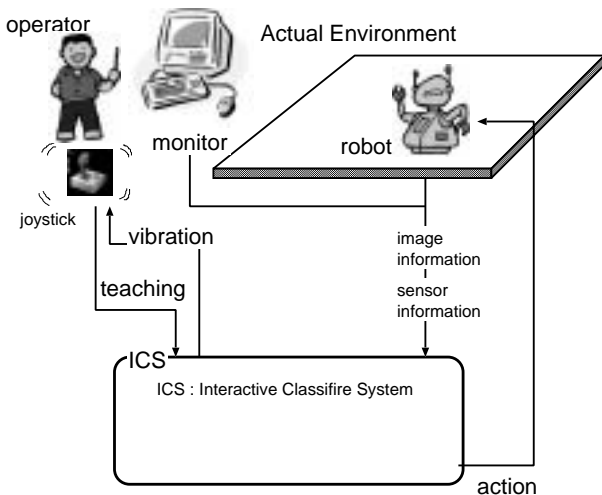


Figure 1 Teaching Environment

ary Robotics (IER) と呼ぶ。IER においては、従来非常に重要視してきた多目的なタスクや動的な環境に適応させることは言うまでもなく、それらの複雑なルールを自動的に抽出し解析することも目的とする。

本研究では、IER の枠組みの上で、少ない試行数で学習でき環境の多様性や動的状況の変化に適応可能な Classifier System[2] に基づく学習システム、対話型分類子システムを構築する。ここでは、Interactive Classifier System (ICS) と呼ぶ。Fig.1 に ICS が想定する教示の環境の例を示す。

ICS が従来の LCS (Learning Classifier System) と異なる点は、機械と人間とのインタラクションを用いた対話型手法の枠組を導入したところにある。これにより、実環境における初期学習の効率化をはかり、集中的な追加学習が可能となると考える。しかし、従来の対話型手法においては教示の仕方による学習効率について、あまり研究されてこなかった。本研究では、観測者の認知に基づく教示法による対話の設計を提案する。

2 観測者の認知に基づく教示法

局所情報しか知らないロボットが、特定の領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良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、むしろ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。しかし、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。

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を、外的観測法 (Fig.2) と呼ぶ。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を内的観測法 (Fig.3) と呼ぶ。

ロボットの認知と外的観測者による認知と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

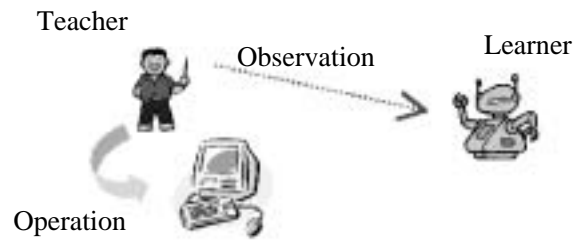


Figure 2 External Observation

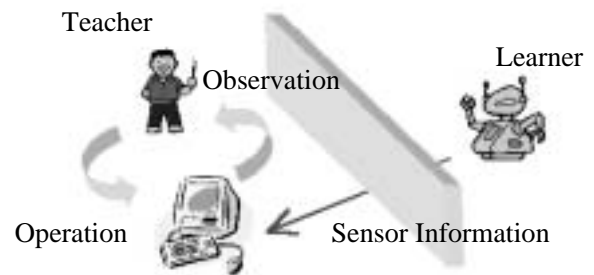


Figure 3 Internal Observation

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。

従来の評価関数の設定やパラメータの調整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ると考えられる。また、従来の対話型進化計算手法においては、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関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。

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型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、この対話型計算法による進化ロボティクスを実現するために、対話型手法に基づいたロボット学習システム、対話型分類子システムを構築した。

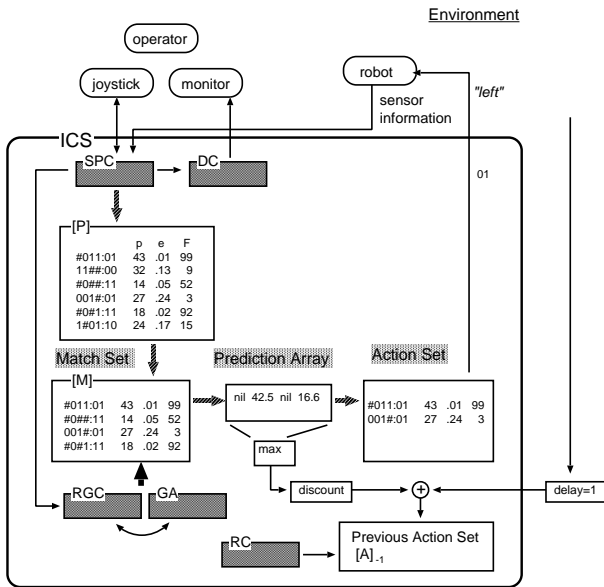


Figure 4 Overview of Interactive Classifier System

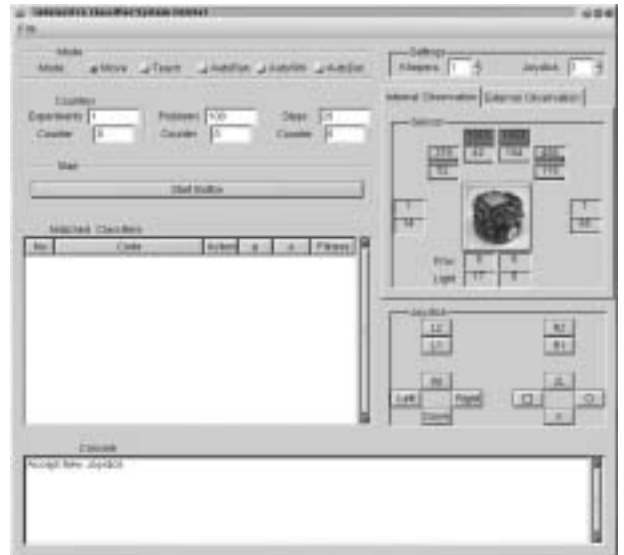


Figure 5 User Interface

3 対話型分類子システム

3.1 概要

対話型分類子システム (Interactive Classifier System: ICS) は学習分類子システム (Learning Classifier System: LCS) の一つである XCS 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。システムの概要図を Fig.4 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 Classifier を作成するルール生成部、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、GUI インタフェース等の表示部からなり、全て Linux 上で開発されている。C 言語および GTK+ で記述されており、CCD カメラの画像処理には Video4Linux を用いている。開発中のインタフェースを Fig.5 に示す。

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[sec] あたり 8[mm]) を 2 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.6 の右図の位置に 8 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。それぞれ、Fig.6 に示す。

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、表示部がそれを処理する。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が受け取り、それよりルール生成部が新しくルールを作成しルールリストに加える。これをルール生成機能とする。

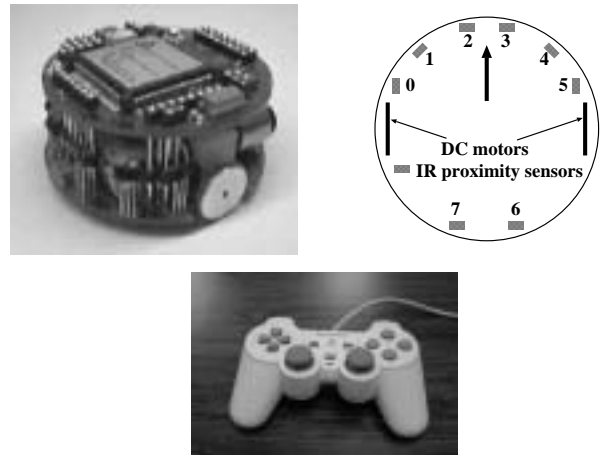


Figure 6 A mobile robot: Khepera and Joystick

また教示者の操作がないときは、ロボットは過去の履歴から、自律的に行動を行う。また前のステップのクラシファイアのパラメータを更新することで強化学習部が学習を行う。これを強化学習機能とする。

従来の実環境におけるロボット学習では、試行錯誤で初期個体の作成を行っていたために、学習の収束に多くの時間がかかった。そこで、本研究では、実環境の学習においてはこの試行錯誤の学習が問題であると仮定した。また、あらかじめその環境に適応した先見知識を作成することは非常に難しいといえる。そこで、ICS においては初期個体の作成を人間の教示から作成することを考える。これにより、初期学習の効率化をおこなうことができる。これを初期学習機能とする。

これらの機能により進化ロボティクスのオンライン学習を行うことを可能としている。

Table 1 Experimental Parameters

Parameters	Value
number of problems in one experiment	30
number of experiments	1
maximum size of the population	300
probability to do crossover	0.8
probability of mutating one bit	0.04

4 実験

4.1 観測に基づく教示による実験 A

認知的観測による教示の違いを調べるために、外的観測法と内的観測法を用いた教示によるオンライン実ロボットと比較実験を行った。Fig.7 にその環境を示す。

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 (Teacher View) と内的観測法 (Learner View) を用いた教示による方法と、従来法である単純に進化計算により学習する方法 (Auto) と比較した。40step を 1 試行とし、教示をもしくは自動探索をランダムな初期位置から 5 試行を行い、検定として 5ヶ所の初期位置からそれぞれ 1 試行評価を行う。これを 6 回、合計 30 試行の実験を行った。実験のパラメータを Table 1 に示す。

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.8 に示す。

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.9 に示す。

4.2 認知観測による教示実験 A 結果

本実験では、ロボットが初期位置から光源にたどり着くまでの期間、あるいは設定した最大ステップ数を消費するまでの期間 (Step to Light Source) と、将来の報酬予測と現在の報酬との誤差 (System Error), GA に用いる評価値 (Fitness) を求めた。

光源までの Step 数を Fig.10 に示す。また、未来の報酬予測の誤差を Fig.11, GA に用いる評価値を Fig.12 に示す。教示は外的観測の方が環境全体を観測できるためうま

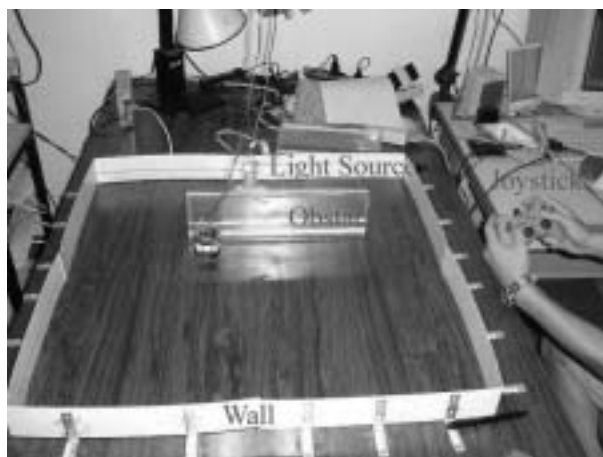


Figure 7 Experimental Environment



Figure 8 External Observation Settings

く行うことができている。あまり差がないが、Step 数において外的観測を行った場合において早く学習が進んでいることがわかる。しかし、解の予測精度および評価値には差がなく、2種類の教示法を用いることにより、予測制度および評価値に影響を与えることなく学習を進めていることがわかる。

4.3 認知観測による教示実験 B

さらにタスクの違いによる認知的観測による教示の違いを調べるために、外的観測法と内的観測法を用いた教示によるオンライン実ロボットの比較実験を行った。Fig.7 にその環境を示す。

タスクは、実験 A の環境のなかに、透明なプラスチック版によって作られた障害物を置いた。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。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。20 step を 1 試行とし、教示をもしくは自動探索をランダムな初期位置から 20 試行を行い、毎試行ごとに検定として 1ヶ所の初期位置からそ

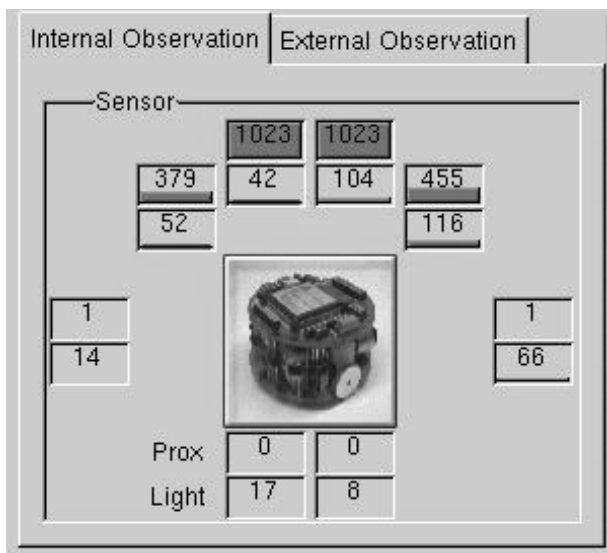


Figure 9 Internal Observation Settings

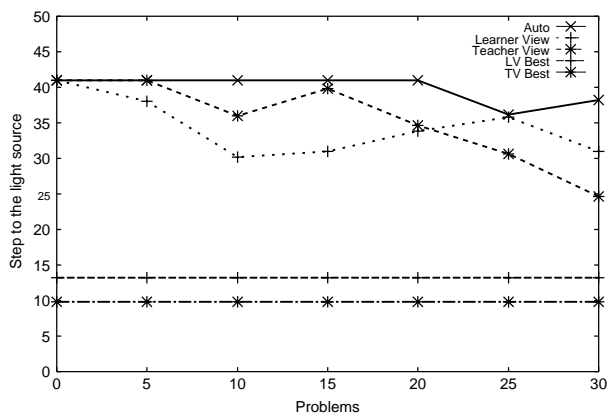


Figure 10 Step to Light Source

れぞれ1 試行評価を行う．実験のパラメータは Table 1 と同じである．

4.4 認知観測による教示実験 B 結果

実験 B では，ロボットが初期位置から光源にたどり着くまでの期間，あるいは設定した最大ステップ数を消費するまでの期間 (Step to Light Source) を求めた．光源までの Step 数を Fig.13 に示す．

教示の最良値の平均 (TV Best, LV Best) を見ると，教示は実験 A と同様に外的観測の方が環境全体を観測できるためうまく行うことができている．しかし，Step 数において実験 A と異なり，内的観測を行った場合の方が早く学習が進んでいることがわかる．実験 A の場合は環境が近接センサの情報ほとんど使用せず，光セン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，2つの教示法にあまり差はでなかった．しかし実験 B 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

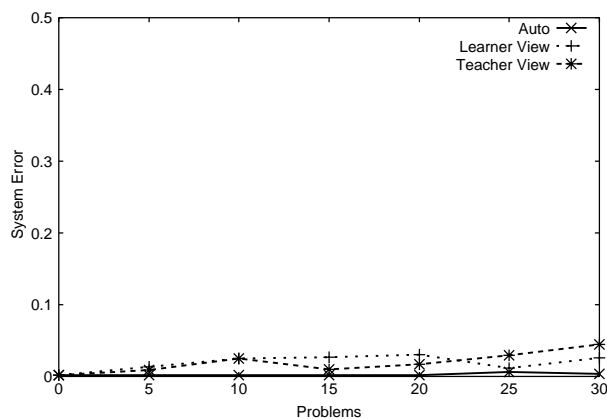


Figure 11 System Error

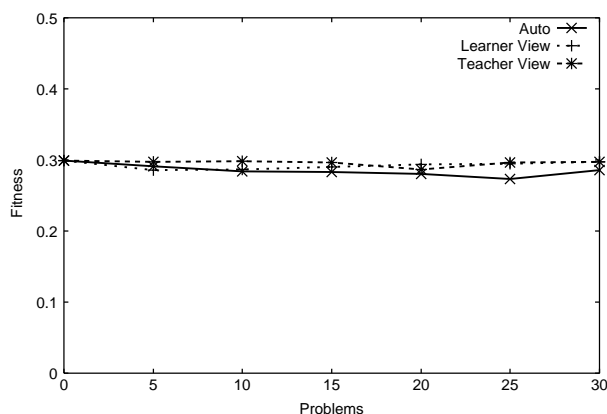


Figure 12 Fitness

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．

外的観測と内的観測の実験を 20 試行したあと，それぞれ作成されたルールを用いて，検定として 5ヶ所の初期位置からそれぞれ 1 試行評価を行う．Table 2 にその結果を示す．

外的観測は教示がうまく進んでいないため検定では 5ヶ所のどの位置からも光源にたどり着いていない．一方，内的観測を行った方は 3ヶ所の場所では光源にたどり着かなかったものの，2ヶ所の位置で 6 ステップ，9 ステップとかなり早いステップでたどり着いている．これは，内的観測を用いた教示がにより，教示から学習がうまく進んだことによりより有効で初期位置に依存しないルールが作られている結果だと言える．

20 試行後に内的観測によって作成されたルールの内，予測報酬の値が最も高いもの 10 ルールを Table.3 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．予測報酬の値が高いものは全て人間の教示により作られたまたはそれらの子供であることから教示がうまく

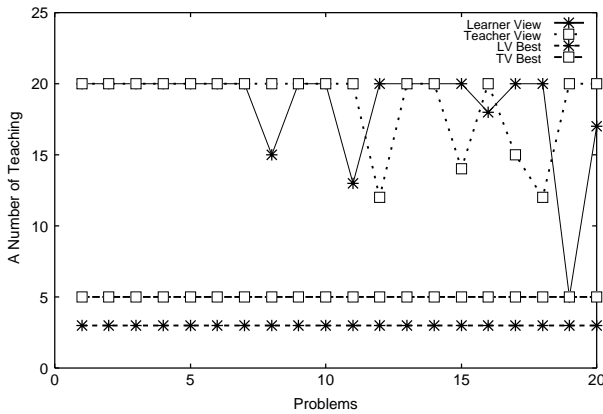


Figure 13 Step to Light Source

Table 2 Experimental Results of Exploit

	Step to Light Source	Reach to Goal
times	1 2 3 4 5	total
Teacher View	-----	0
Learner View	6 9 ---	2

く進んでいることがわかる。

また、後ろ向きに進む教示は行っていないのだが、検定の時に壁にぶつかった時に後ろ向きに進み、壁を回避しつつ光源にたどり着いていた。人間とロボットが協調してルールを作成しているためこのようなルールが作成された結果である。

さらに、センサ値はロボットの生データを使っていたのだが、後で調べたところ左後ろの光センサが他のセンサよりも反応しやすいことがわかった。実際にロボットの右後ろに光源があるときもロボット(内的観測)には左後ろにあるものと見えるため左回転の教示を行っていた。そのため、ロボットも左回転のルールが作成された。しかし、外的観測によって作成されたルールは教示とロボットの状況が逆になるため整合性のとれたルールの作成が難しかったと言える。この点でも認識のずれが問題になっていることがわかる。

5 おわりに

従来のロボット行動学習がとられて来た行動還元主義的なアプローチではなく、人間の内部に焦点をあてた認知主義的な新しいアプローチである ICS の提案を行い、観測の観点から教示を外的観測におけ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。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すことがわかった。

本研究は、観測法の違いにより教示の効果の違いを示すと共に、ロボットの実環境における高速な学習が可能とし、人間には記述困難な複雑なロボットプログラムを

Table 3 Created Rules by Learner View Method

Experiment 20	Condition	Action	Predediction	Teach
	0#0#0#####0#10##1#	10	413.7	2
	0010###0##000#1###	01	364.0	1
	00###0#000#1#0#01#	11	292.0	4
	###000000101001011	11	256.0	1
	0#0#0#####0#10#01#	10	280.3	3
	00#000001001##0##1	10	244.5	1
	00###0#000#1#0##11	11	215.0	1
	#000#100000###100#	01	101.5	3
	#000##001000##0001	11	100.8	6
	#000#0001000##0001	11	99.4	10
	00###0#01#0#10#001	11	82.6	2
	##0#0#0010###0#011	11	83.1	8
	##0#0#0010###00011	11	79.5	4
	#0000###11#0010011	01	81.7	1
	#00#1#0000#00##0#1	11	77.5	27
	00#0###00##0#0#0#1	01	50.5	9

簡単な教示をすることで学習、自動抽出することができることを示した。今後は、人間が教示するときに意図していない要素を持つ情報、例えば人間の反射的な行動や、操作者の選好、またユーザの間で暗黙のうちにできた役割分担によるルールの違い等の情報を獲得することを目指す。

Bibliography

- [1] 高木 英行, 畝見 達夫, 寺野 隆雄. 対話型進化計算法の研究動向. 人工知能学会誌, 13(5):24-35, 1998.
- [2] John H. Holland and Judith S. Reitman. Cognitive systems based on adaptive algorithms. In Donald A. Waterman and Frederick Hayes-Roth, editors, *Pattern-Directed Inference Systems*, pages 313-329, Orlando, 1978. Academic Press.
- [3] S.W. Wilson. Classifier fitness based on accuracy. *Evolutionary Computation*, 3(2):149-175, 1995.