

## 対話型進化ロボティクスにおける視点切り替えによる教示

片上大輔<sup>†</sup> 山田誠二<sup>††</sup>

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、行動に必要な知識を事前に用意しておく事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、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注目されている。ここでは、システム外部から全体を眺めることのできる観測者と、システム内部の観測者との認識のずれが学習の効果において大きな問題となる。本論文では、ヒューマン・ロボットインタラクションを通じてその経験からロボットの自律行動を進化計算により獲得していく対話型進化ロボティクスにおいて発生する教示者と学習者の認識のずれを改善し、最小の負担で最大の教示効果を得る教示学習手法を目指す。

### Interactive Evolutionary Robotics by Switching Viewpoints of Observation for Real Robot Learning

DAISUKE KATAGAMI<sup>†</sup> and SEIJI YAMADA<sup>††</sup>

In this paper, we describe a fast learning method for a mobile robot which regard for perceptual aliasing problem between an operator and a robot. It is a new problem which caused by the difference in the internal representation of the world between an operator and a robot. We develop a behavior learning method ICS (Interactive Classifier System) which acquires autonomous behaviors from interaction between an operator and a robot. 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 proposed a new teaching method regard for the difference of perception between teacher and learner by the different viewpoints of observation.

#### 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事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入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 priori な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言える。しかし、ある程度の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。教示が必要でない段階に関しては、人間に負担をかけることなく、過去に蓄積された教示情報を元にして自律性を発揮すべきである。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言える。

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<sup>1)</sup>を提案しているが、人間とロボットとの対話によりロボットの行動を獲得している

<sup>†</sup> 東京工業大学

Tokyo Institute of Technology

<sup>††</sup> 国立情報学研究所

National Institute of Informatics

わけではない。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<sup>2)</sup>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っており、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ない。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の自動化プロセスと人間の操作の協調行動<sup>3)</sup>を実現しているが、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<sup>4)</sup>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、タスクを解決する状態認識・行為ルールの獲得を進化的に実現する手法を提案した。このような枠組みを Interactive Evolutionary Robotics (IER) と呼び、今までにいくつかの研究を行ってきた<sup>5)6)</sup>。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のがわかってきた。

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し学習が進まない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言える。これによって、教示情報を少なくしていると共に、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どの様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、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

前述の様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

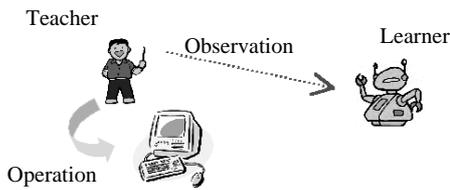


図 1 教示環境における外的観測

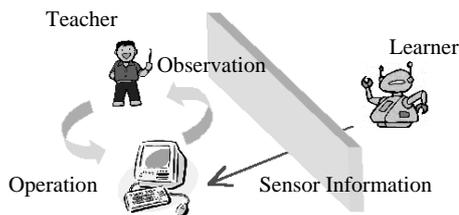


図 2 教示環境における内的観測

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 (Teacher View (TV)) 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 (Fig.2) を内的観測法 (Learner View (LV)) と呼ぶ。

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 の様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、即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 の様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

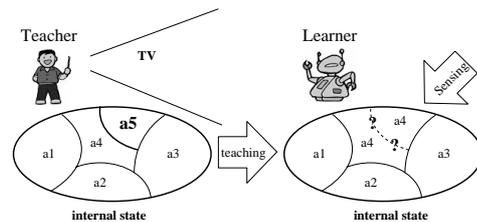


図 3 状態の違いをロボットが認識できない場合

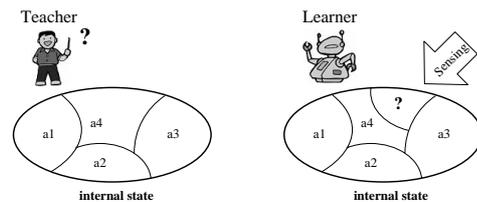


図 4 状態の違いを教師が認識できない場合

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。この様な、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と言える。

三島、浅田らは、この違いを Cross Perceptual Aliasing<sup>7)</sup> と呼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、このような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とも言える。

#### 4. 視点切り替えによる教示

我々は、この認識のずれがおこりうる事を実環境における教示実験によって確認してきた<sup>8)</sup>。実際の教示環境においては、この様な認識のずれは環境によって少なからず存在し、教示効果に大きく影響してくるために、これを考慮する教示方法を構築する必要がある。我々は、教示者の認識がロボットの認識を上回る場合に教示の視点を教示者からロボットに切り替え (Figure 5)、ロボットの認識が教示者の認識を上回る場合にロボットから教示者に教示の要求を行う (Figure 6) ことで、最小の教示者の負荷で最大の教示効果を得ることを提案する。ここでは、内的観測と外的観測の両方の場合 (環境の状態や人間とロボットの認識のずれ) によってうまく切り替えることが重要である。

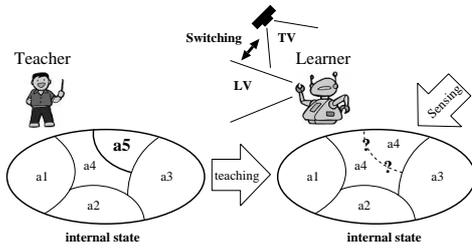


図 5 ロボットの認識が悪い場合

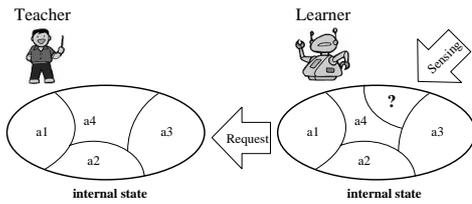


図 6 教師の認識が悪い場合

提案する 2 つの手続きを以下に示す。

- ロボット側の認識が悪い場合

**Step1** 以前と条件部が同じ環境情報が入力される。

**Step2** IF 教示情報が以前と異なる  
**THEN** モードが TV の場合 TV から LV に変更する  
**ELSE** モードが LV の場合 LV から TV に変更する

- 教示者が状況の違いを認識できない場合

**Step1** 条件部に環境情報が入力される。

**Step2** もし、その条件部にマッチするルールがない、もしくは閾値以上の強化値を持つルールが存在しない。

**Step2** 教師に対し教示を要求する。

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を構築した。

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クラシファイア<sup>9)</sup>に基づく学習システム、対話的クラシファイアシステム Interactive Classifier System

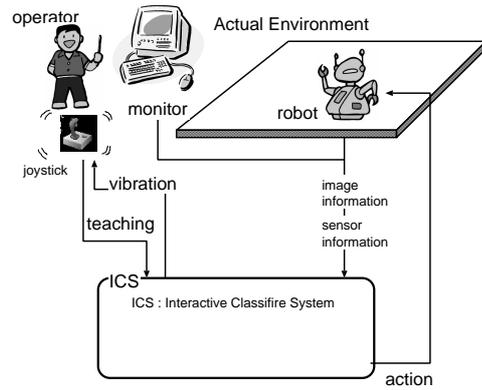


図 7 教示環境の例

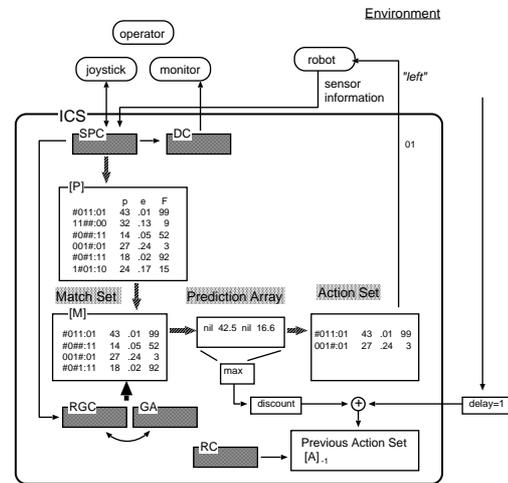


図 8 Interactive Classifier System

(ICS) を構築する。Fig.7 に ICS が想定する教示の環境の例を示す。

### 5. 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<sup>10)</sup> を使用している。XCS は ZCS<sup>11)</sup> を改良したもので、精度 (accuracy) と呼ばれるパラメータを追加したものである。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.8 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ァイアを作成するルール生成部 (RGC)、

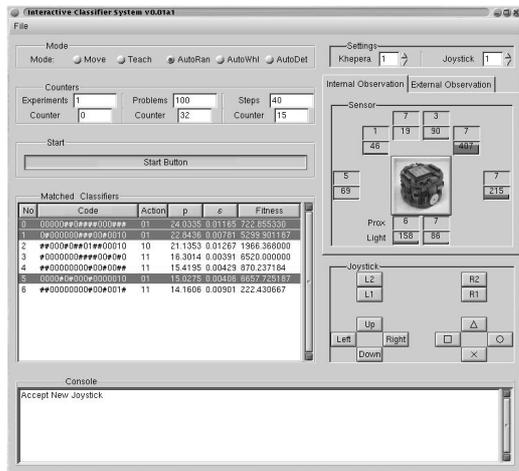


図 9 ユーザインタフェース

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, 全て Linux 上で開発されている。C 言語および GTK+ で記述されている。以下, 各モジュールを説明する。

- (1) [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 を基本に教示情報からルールを作成できる様に改良した。
  - (a) 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。
  - (b) 集団  $[P]$  から  $X$  にマッチしたクラシファイアがマッチセット  $[M]$  に移され, システムは  $[M]$  で表された各々の行為  $a_i$  に関する prediction  $P(a_i)$  を  $a_i$  を支持するクラシファイアの予測の fitness で重みづけされた平均値により作成する。 $P(a_i)$  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。
  - (c) 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。

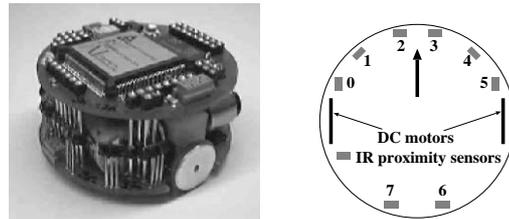


図 10 移動ロボット: Khepera

$a_j = a_t$  ならば, 変更はしない。

- (d) 選ばれた行為  $a_j$  を支持する  $[M]$  の中のクラシファイ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)。

- (2) [RC] Reinforcement Component は, クラシファイ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ァイ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。
- (3) [DC] Display Component は, インタフェースにより 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.9 に示す。
- (4) [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。

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。これらの各モジュール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。

## 6. 実験の設定

移動ロボットとして, Khepera (図 10) 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 10 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。

タスクは障害物回避を行う。障害物回避は、1 状態に対して 1 行動を実行する反射行動により実現できる比較的簡単なタスクであり、認識のずれを改善する本手法の効果を単純に比較するのに適している。実験環境として、110[cm]×90[cm] の平坦な長方形のテーブルの上に直方体の障害物を 2 つ置き、周りには白いプラスチック板で作った壁を設置する。適合度関数  $F_j$  は、以下のように定義する。

$$f_j = \sum_{i=1}^k \{ \alpha s_{max}^i + \beta (|10 - m_{left}^i| + |10 - m_{right}^i| + |m_{left}^i - m_{right}^i|) \}$$

$$g_j = \frac{1}{1 + f_j} \quad \text{Fitness : } F_j = \frac{g_j}{\sum_{j=0}^s g_j}$$

ロボットの行動ステップごとに、センサ値  $s_{max}^i$ ,  $m_{left}^i$  及び  $m_{right}^i$  を計測する。ここで、 $s_{max}^i$  は、 $i$  step 目の実行直後に得られた 8 個の近接センサの最大値、 $m_{left}^i$  及び  $m_{right}^i$  は、それぞれ  $i$  回目の実行直後の左車輪と右車輪の速度でありエンコードにより計測され、 $-10 \sim 10$  の値を返す。で選択された  $s$  個の個体について、一個体につき  $k$  回の実行の結果として得られる  $f_j$  を計算し、トーナメント選択により選択された個体集合で正規化して最良値を 0.0 とした結果を個体  $j$  の適合度  $F_j$  とした。この関数は、ロボットが障害物に近いほど、またロボットが定点旋回を多く行なうほど評価を下げる。

## 7. まとめと今後の課題

本稿では、教示学習に存在する教示者と学習者の認識のずれを改善し、最小の負担で最大の教示効果を得る対話型進化学習の枠組みを提案した。現在システムの実装が完了していないので、今後は、本モデルを開発した ICS に組み込み、被験者を用いた実験を行なう予定である。

## 参 考 文 献

1) H. Asoh and Y. Motomura and I. Hara and S. Akaho and S. Hayamizu and T. Matsui: Com-

binning 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 and R. Sato and T. Ishida: Robot Oriented State Space Construction; *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1496-1501 (1996)

3) Y. Horiguchi and T. Sawaragi and 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)

4) 稲邑 哲也, 稲葉 雅幸, 井上 博允: 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)

5) D. Katagami and S. Yamada: Interactive Classifier System for Real Robot Learning; *IEEE International Workshop on Robot and Human Interaction*, pp. 258-263 (2000)

6) D. Katagami and S. Yamada: Real Robot Learning with Human Teaching; *The Fourth Japan-Australia Joint Workshop on Intelligent and Evolutionary Systems*, pp. 263-270 (2000)

7) C. Mishima and M. Asada: Active Learning from Cross Perceptual Aliasing; *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1420-1425 (1999)

8) D. Katagami and S. Yamada: Interactive Evolutionary Robotics from Different Viewpoints of Observation; *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1420-1425 (2002)

9) J.H. Holland and J.S. Reitman: Cognitive Systems Based on Adaptive Algorithms, *Pattern-Directed Inference Systems* (D. A. Waterman and F. Hayes-Roth), Academic Press, pp. 313-329 (1978)

10) S. W. Wilson: Classifier fitness based on accuracy; *Evolutionary Computation*, Vol. 3, No. 2, pp. 149-175 (1995)

11) S.W. Wilson: ZCS: a zeroth order classifier system; *Evolutionary Computation*, Vol.2, pp.1-18 (1994)

12) S. W. Wilson and D. E. Goldberg: A critical review of classifier systems; *The Third International Conference on Genetic Algorithms*, pp. 244-255 (1989)