対話型分類子システムによる実環境ロボット学習
～記述困難なプログラムを人間の教示から自動抽出する～

＊ 片上大輔　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
〒 226-8502 横浜市緑区長津田町 4259
TEL&FAX:045-924-5218
katagami@ymd.dis.titech.ac.jp

Abstract – 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．本研究では，このような枠組みを**Interactive Evolutionary Robotics(IER)** と呼び，その枠組みについて説明する．
また IER の枠組みにおいて，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 に基づく学習システム，対話型分類子システムを提案しその評価方法について述べる．

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．また，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，教示は任意の時に行われ，ユーザが毎回教示を行うことは必要ではない．

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ク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ク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 Evolutionary 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す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，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．

2　分類子システム

分類子システム ((Learning)Classfier System:CS,LCS) は，遺伝的アルゴリズムによる機械学習 (Genetic Algoritm-Based Machine Learning:GBML) の手法の一つである．

分類子システムは，プロダクションシステムに基づく実行機能，信頼度割り当て (Credit Assignment) に基づく強化学習機能 (Reinforcement Learning)，遺伝的アルゴリズムに基づくルール生成機能の三つの機能から構成されている．

プロダクションシステム　プロダクションシステムは，命題形式もしくはそれに相当するビット列で記述したルールベースを持つ．そして，環境からの入力情報にマッチしたルールについて，メッセージが出力されて実行が繰り返される．この際の競合解消は各ルールの持つ信頼度に従って確率的に行われる．

信頼度割り当て法　信頼度割り当ては，各ルールに報酬を与えることによって信頼度を調整する機能である．競合解消対象のルールはその時点で一定の税金を支払う．実行ルールにはその行為の有効性に応じた報酬が与えられる．しかし，行為の結果がすぐに評価できない場合は，強化学習一般の議論と同様，信頼度割り当ての計算は難しい．

信頼度割り当て法については，各実行ステップごとに評価を行うバケットブリゲード法 (Bucket Brigade Algorithm:BBA) と，ルールの実行履歴を保存しておき，報酬獲得時に過去に実行したルールの信頼度を計算するプロフィットシェアリング法 (Profit Sharing Plan:PSP)，ならびに両者の中間的な性質を持つ階層的チャンキング法 (Hierarchical Chunking Algorithm:HCA) などが知られている．BBA は必要な記憶領域は少ないが学習速度が遅く，PSP はその逆の性質を持つ．HCA は一連のルールを組み合わせる点に特徴がある．一般には PSP 法が優れているとされている．

ルール生成機能　ルール生成機能は，実行の適当なタイミングで起動される．ルールの信頼度を適応度，ルール群を個体群とみなして，遺伝的操作を適用する．そして生成されたルールでルールベースの一部を変更する．

Classfire System は，学習コストが状態数の多項式オーダーの計算量で済み，少ない試行数で学習でき，しかも状況の多様性や動的状況の変化に適応可能な強漸次性を備えている [3].

3　ICS の概要

ICS では，操作者の教示より Classifier を作成する Classifier 生成部（CFGP）, 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（SPP）がある．システムの概要図を Fig.1 に示す.

3.1　クラシファイア生成部 (CFGP)

従来ランダムに作成していた個体を，人間が操作することにより自動生成することで，試行錯誤で学習していた初期段階の学習を進めることができる．また，人間が操作した行動とおなじ個体に対して強化値をあげることで，人間の評価をシステムに組み入れる対話型 EC と同じ主観的評価を行うことができる．またこれは，強化学習の報酬の概念と同じである.

また，**ICS** では教示モードと推論行動モードの 2 つのモードを交互に行うことで学習を進める．2 つのモードの手続きを以下に示す.

教示モード

Step1　操作者がタスクを実行し教示を行う.

Step2　操作者の教示とその時の環境情報によりルールを作成する.

Step3　同じクラスタに属するルールがなければ，新しくルールとして追加する.

Step4　同じクラスタに属するルールがあれば，報酬として強化値をあげる.

自律行動モード

Step1　Rule List に蓄えられたルールに従って行動する.

Step2　ある一定の時間が経過したら，GA によりある一定のルールを置換する.

教示モードと，自律行動モードでの概要図を Fig.2 に示す.

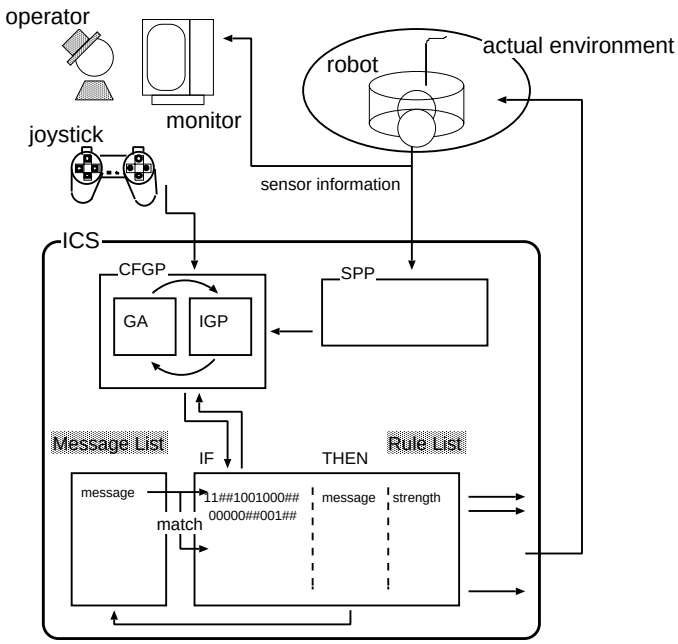

Fig.1 Overview of Interactive Classifier System

4 ロボットプログラムの抽出

クラシファイアシステムが複雑なタスクを扱うためには，CS 内部にデフォルト階層や chain などのルール同士の関連を構築していく必要がある．しかしながら，このような構造を通常の CS 内部に創発させることは難しいということがわかっている．これは，(1) 通常の GA ではルールがいくつかの似たようなものに収束してしまうこと，(2) デフォルト階層や chain を効率よく生成・維持するためのメカニズムが CS には備わってないことなどが原因としてあげられる．(1) の問題については，ニッチ GA を適用する方法がある．ニッチ[1]を形成することによってルールの多様性を維持することができる．一方 (2) の問題については，チャンキングという chain となっているルールを一つのメタルールとしてまとめていく方法がある．

しかし，**IER** の枠組みでは人間が評価関数の役割を担うため，評価関数は与えられた 1 つものではない．よって，(1) の問題は解消できる．一方 (2) の問題については，**IER** の目的である「人間の意図を抽出する」ことから考えても依然重要な問題であり，今後の大きな課題となる．

5 実験の想定

5.1 教示方法

ロボットの情報に基づく教示　操作者は，ロボットを第 3 者の視点からではなく，インタフェースに示された，ロボットのセンサ情報，カメラ情報などのロボットが獲得できる情報をみながら（つまり，ロボットの視点から情報を獲得し）教示を行う．これにより，操作者はよりロボットの獲得できる情報の範囲内で教示を行うことになり，適切なロボットプログラムの作成が容易になる．

人間の情報に基づく教示　操作者は，ロボットを直接外から見て判断し教示を行う．これにより，細かい教示が必要なタスクなどを無駄な動きがなく教示することができる．

5.2 想定タスク

ロボットの学習タスクとして，壁沿いタスクとマルチロボットによるサッカータスクを設定する．

[1]生態的地位（日本語訳）.「完全に同じニッチ関係をもつ 2 種の生物は，同一の生息地で共存することはできない」[J.Grinnell,1917]

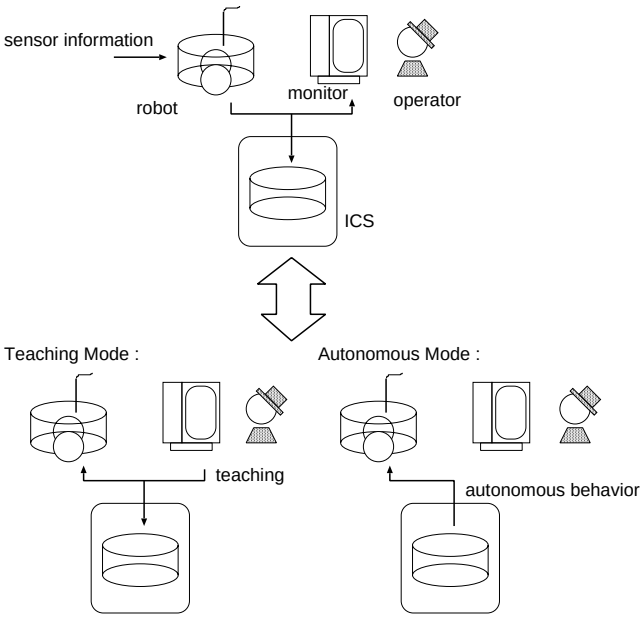

Fig.2 Teaching Mode and Autonomous Mode

壁沿いタスク　壁沿いタスクは壁に沿って動く行動を獲得するタスクである．複数の人間によってそれぞれ教示を行い，同じようなルールが作成されたならば，それは壁沿いタスクの解のプログラムであると考えられる．ここでは，教示の違いと，従来の手法との比較を行う．

マルチロボットによるサッカータスク　マルチロボットによるサッカータスクを用いて，協調作業におけるルールや機能分担によるルールの違い等を抽出する．

例えば従来法では，パス＆シュート等の協調行動をそれぞれのタスクとして分割して解を探索してきた．しかし，その方法では，全ての協調行動の抽出は不可能であり，また抽出された解を使ってロボットにやらせてみてもうまくいかない場合が多い．しかし，人間が操作するといとも簡単にこれらの協調作業をこなすことができる．それも，操作している人間は具体的にそれらの行動の理由を聞いてもうまく説明できないだろう．おそらくそれはその場の一瞬の判断の積み重なった結果であり，それを始めから説明することは難しいからだと考える．

このタスクで作成されたルール群を分析し考察を行う．

6　おわりに

ロボットの実環境における高速な学習が可能となるとともに，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，自動抽出することができると考える．また，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が可能である．

参考文献

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) 山口 智浩, 増渕 元臣, 藤原 一継, 谷内田 正彦. 抽象化副報酬の自動生成による実ロボット強化学習の高速化. 人工知能学会誌, 12(5):60–71, 1997.