

# 人間の教示特性に基づくロボット行動学習アルゴリズムの提案

飯田 史也      原文雄 （東京理科大学工学部）

## Robot behavior learning based on the characteristics of human instruction

Fumiya IIDA, Fumio HARA

Dept. of Mechanical Engineering, Science Univ. of Tokyo

**Abstract:** There is a difficulty in a human-friendly robot learning its behavior by means of human instruction. The difficulty is mainly spotted at the learning algorithm of the robot, which do not take account of characteristics of human instruction and can be improved. This paper suggests an effective learning algorithm constructed by the results obtained through analysing characteristics of human instruction, and evaluates its effectiveness by computer simulation.

**Keywords:** Machine learning, Characteristics of human instruction, OR-reinforcement, Comparative reinforcement

### 1. 背景

近年、人間社会で活躍するロボットの期待が高まっている。中でも福祉ロボットに代表されるように、人間と直接触れあうような場でのロボットの要求は徐々に大きくなり、人間が親和感を感じるロボットの研究も進められている。しかしこのような人間の感性には大きな個人差があり、人間が快く感じるロボットの行動を決定するのは非常に難しい。どのようにロボットの行動に親和性や柔軟性を持たせるかが問題となる。そのひとつの解決方法としてロボットに行動の学習機能を備えることが考えられ、更に人間が望むロボットの行動を人間の教示により学習を進めることが望まれる。しかし人間の教示法には特性があり、学習のアルゴリズムによっては人間の教示による学習が困難になることがある。そこで本研究では人間の教示法の特性を考慮に入れた学習アルゴリズムを提案し、その性能をシミュレーション実験により評価する。

### 2. 人間の教示特性

J.S.Bruner 1) は人間の学習における本質的特徴として、  
検証手続きの定式化、  
検証手続きの実施、  
検証結果と何らかの基準の比較

という手順があることを指摘し、  
が行われるときには修正の知識が必要であり、更に目的の達成に向かっていかどうかという指針があることが望ましく、教授者はこれらを学習者に示すことが望ましいと述べている。

ここで我々はロボットへの人間の教示には目的の達成に向かっていかどうかの指針が含まれていることに注目した。すなわち、人間がロボットにある特定の行動を実行させることを望んだ場合、人間はロボットの行動目標の達成が部分的であってもそれが達成に向かっていれば正の報酬を与える 2) 本研究ではこの特性を利用した行動学習アルゴリズムを提案する。

### 3. ロボットの学習アルゴリズム

人間がロボットに行動の教示を行うことを想定すると、誰にでも直感的に解りやすい方法が望まれ、そのためには口

ボットが人間の自然な教示により学習できる学習アルゴリズムを備えておくことが望ましい。

2章で述べた人間における学習の手順は原理的に強化学習法で実現できる。しかし従来の強化学習法は行動の自由度が大きくなった場合、学習には膨大な時間がかかり、人間の教示による学習は非現実的になってしまう。更に人間の教示特性である、目的の達成に向かって指針となる部分的なタスクの達成に対する正の強化を考慮に入れていないため、その情報を有効に学習に利用できない。これらの問題を考えるにあたって、以下のように問題設定を行う。

ロボットには複数のアクチュエータが備えられている。このアクチュエータの数を行動の自由度と呼ぶ。

ロボットがある状態に対して行う行動は各アクチュエータの出力値の論理積の形で表現できるものとする。

人間の教示はスカラー値で、ロボットが1回の行動を実行する毎に与えられるものとする。

これらの問題設定のもとで人間の教示特性を考慮に入れた学習アルゴリズムとして、「OR強化型学習アルゴリズム」および「比較強化型学習アルゴリズム」を提案する。強化学習の土台となるQ値の更新式は

$$Q(s,a)=(1-\alpha)Q(s,a)+r \quad (1)$$

とする。ここでsはロボットが認識した状態、aは実行した行動、 $\alpha$ は学習率、rは人間の教示による報酬値とする。一回の行動に対する強化でどの行動に対応するQ値を上式で更新するかを議論する。また状態sの数が増えれば学習回数が増えることになるが、ここでは議論の対象とせず、1つの状態に対する学習を取り扱うものとし、以降、Q(s,a)をQ(a)と記す。

以下に各学習アルゴリズムの内容を示す(図1参照)

#### < AND強化型 >

この学習アルゴリズムは以下に提案する2つのアルゴリズムとの比較のために用意する。従来の強化学習と同様、与えられた報酬を実行した要素行動の論理積で表される行動に対応するQ値のみの強化に使う。

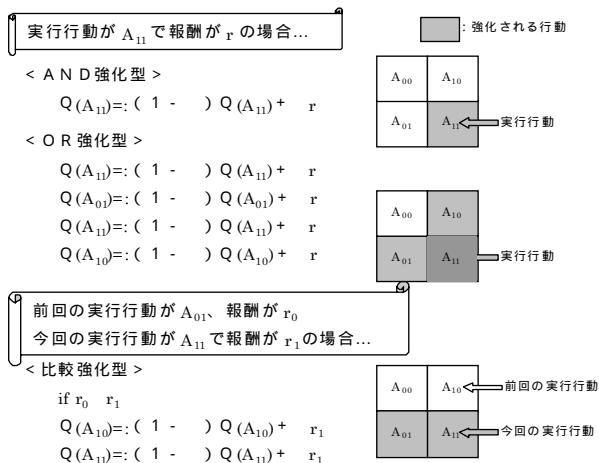


図1 各学習アルゴリズムの強化方法  
(行動自由度2の場合)

< OR 強化型 >

与えられた報酬を実行した行動に対応するQ値だけでなく、実行行動の行動要素を含む行動に対応するQ値も同様に強化を行う。ただし、実行行動に対応するQ値は行動の自由度数分だけ強化される。

< 比較強化型 >

人間が与える報酬が前回の結果との比較評価に基づいているという教示特性を利用して、同じ状態の1STEP前の実行行動および報酬を記憶しておき、今回の実行行動および報酬と比較して強化を行う。前回と比較して報酬が変化した場合、前回の行動と今回の行動で変化した要素行動を含む行動集合を強化する。報酬が前回と変わらなかったときは今回実行した行動集合のみ強化する。この強化型の特徴はOR強化型と違って正解行動を負強化することがない。

4. シミュレーション実験

提案した学習アルゴリズムの評価を行う。本実験では各アクチュエータの出力値は2種類(ON, OFF) 人間がロボットに与えることができる報酬は+または-の2種類(それぞれ報酬値 +1.0, -1.0) であるとする。2)での実験に基づいて、ロボットの行動に対する人間の報酬の決定則として以下の2つを用意する。

行動評価モデルA

同じ状態の前回の行動と比較して、満足できる行動要素の数が増加した場合は+報酬を与え、減少した場合は-報酬を与える。同じ場合は+報酬を与える。ただし、すべての行動要素が満足できる場合には+報酬とし、逆にすべての行動要素が満足できない場合には-報酬とする。

行動評価モデルB

行動目標がすべて達成されたときのみ+報酬を与え、それ以外は-報酬を与える。

学習率 = 0.1、ボルツマン温度 = 0.1、行動自由度 = 2, 6, 10 のもとでシミュレーションを行った。

図2は行動評価モデルAと行動評価モデルBに対して、学習ステップ毎の正解行動に対応するQ値の推移を、行動の

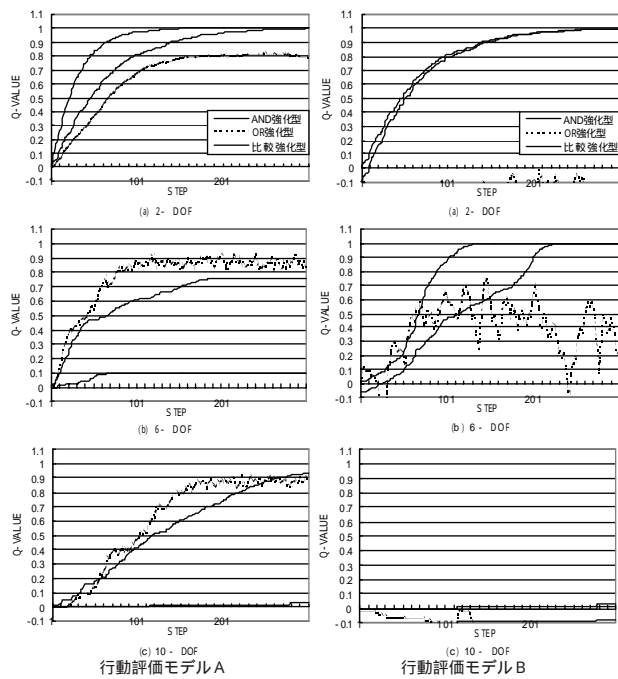


図2 各学習アルゴリズムによるQ値の推移

自由度毎に示している。それぞれのQ値は10個の乱数の種における結果を平均した値である。行動評価モデルAについては自由度が上がるにつれて、従来のAND強化型と比較して、OR強化型と比較強化型の学習アルゴリズムの効果が顕著に現れていることが解る。OR強化型は全体的に比較強化型よりも学習が速く収束しているが、行動評価モデルBに対する学習では正解行動を負強化してしまう性質により安定した強化が行われず、自由度が高くなった場合は正解行動を負強化する行動の方が選択確率が高いため結果的に正解行動を負強化してしまっている。

実際の人間の報酬パターンは行動評価モデルBと行動評価モデルAが折り重なって行われる2)ことを考慮すると、比較強化型の学習が最も優れていると考えられる。

5. 結論

本研究では人間の教示特性に基づいたロボットの行動学習アルゴリズムを提案した。ロボットの行動自由度が増大した場合には従来の強化学習では膨大な学習時間がかかったが、人間の教示特性を反映させた本アルゴリズムにより学習効率が改善され、実際の人間の教示による学習の可能性が示された。また学習率やボルツマン温度を適当に選ぶことによって一層効率が上がることが予想されるが、それは人間との実際のインタラクションによる実験を検討して決定することが望ましいと考えられる。

< 参考文献 >

- 1) J.S. ブルーナー：教授理論の建設, pp.74 - 78, 黎明書房 (1983)
- 2) F. Iida, M. Tabata, F. Hara: Generating Personality Character in a Face Robot through Interaction with Human, Proc. of 7th IEEE International Workshop on Robot and Human Communication (1998) (to appear)