進化と学習における実世界の性質の役割 : 粘性と弾性を有する2リンクマニピュレータを用いた検証

吉岡, 康貴 / YOSHIOKA, Yasutaka

(発行年 / Year)
2009-03-24
(学位授与年月日 / Date of Granted)
2009-03-24
(学位名 / Degree Name)
修士(工学)
(学位授与機関 / Degree Grantor)
法政大学 (Hosei University)

２００８年度　修士論文

進化と学習における実世界の性質の役割

ー粘性と弾性を有する 2 リンクマニピュレータを用いた検証ー

Role of properties of real world in evolution and learning

- Examination using two-link manipulator that has viscosity and elasticity -

指導教員　伊藤　一之専任講師

法政大学大学院　工学研究科

システム工学専攻　修士課程

07R6137

ヨシオカ　ヤスタカ

吉岡　康貴

Abstract

In this paper, we consider abstraction of information using properties of the real world**.** We employ 2-link manipulator as an example, and we design the manipulator by genetic algorithm whose fitness is ease of learning**.** Simulations have been conducted and as a result, a manipulator which has viscosity and elasticity has been obtained, and state-action space has been reduced extremely**.**

もくじ

１．はじめに

近年，ロボットは複雑かつ未知な環境で自律的に振舞うことが期待されている．これを実現するために，ロボットに学習機能を持たせる研究が盛んに行われ，その一手法として強化学習が注目されている．しかし，自由度が増加するに連れて状態行動空間が指数的に増加し，事実上学習が不可能になるという問題がある．

しかし，生物は複雑な実環境において実時間で学習することができ，適応的に振舞うことができる．近年，これら生物の適応的な振る舞いに関する研究が行われ，実世界に存在する様々な性質が，これに寄与していると考えられている[1]-[7].

伊藤らは従来研究において，実世界の性質を利用することで情報の抽象化を行い，実時間で学習が可能な枠組みを提案した[8][9].

本研究では,粘性と弾性を考慮した 2 リンクマニピュレータの制御を例に,「進化」,「学習」,「実世界の性質」のそれぞれの役割について検討する．具体的には，学習しやすさを適応度として，マニピュレータおよび学習のパラメータをシミュレーション上で進化させ検証する.

２．従来の問題点

２．１　強化学習の概要

強化学習とは未知の環境に置かれた知的エージェントが環境との間の相互作用を通して，環境において目標を達成するための適切な行動規則を学習できるようにする手法である．本シミュレーションでは強化学習法の１つである Q 学習を用いる．この学習法の特色として，学習エージェントが環境に関する先験的な知識を持たないときに適用できることである．そのため，前もって対象世界のシミュレーションモデルを得ることが困難であり，学習エージェントを不完全なモデルしか与えられていない対象世界で直接訓練せざるを得ないとき，Q 学習は最も効果を発揮する．

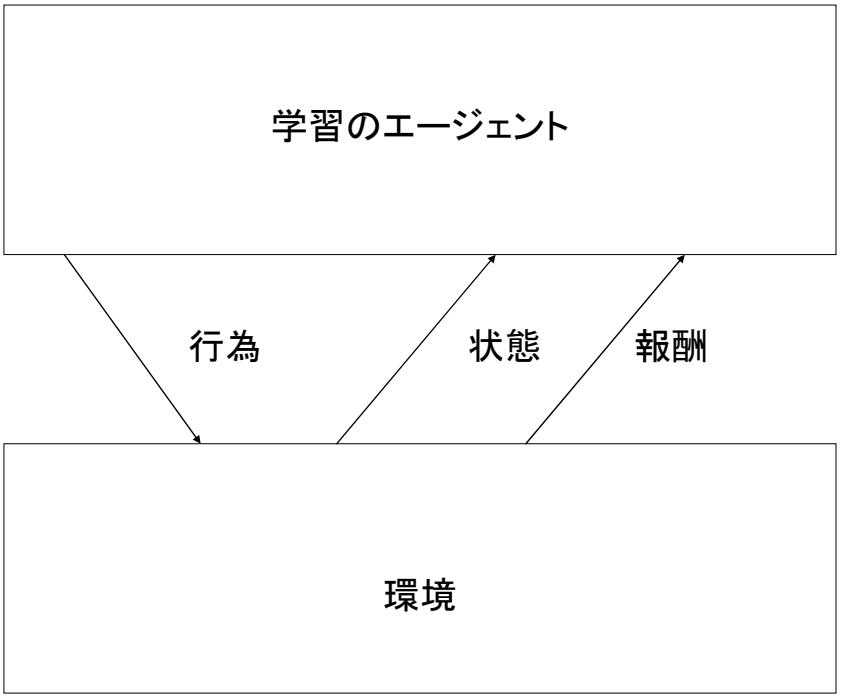

図１　強化学習

２．２　強化学習の問題点

ロボットが未知環境で，自律的に振舞うための一手法として，強化学習が注目されている．しかし，ロボット自身が複雑になることと環境が複雑なため，

強化学習に必要な状態行動空間が膨大になり，実時間内に学習を完了させることは事実上不可能になる．一方，生物は実時間で学習を完了させることができる．

２．３　制御対象とコントローラの問題点

制御系を安定させるために，コントローラを用いることが知られている．通常，コントローラはコンピュータをはじめとする計算機によって実現されている．コントローラを計算機内で実現するためには，実世界の情報を多数の物理量に分解し，センサを用いてコンピュータ内に取り込み，これらの物理量を基に計算を行ってアクチュエータを駆動させるための情報を生成する．従来の研究では，情報処理を制御対象側に負担させても，コントローラ側に負担させても数学的に同じ意味から，ソフトウェアで行うことのできるコントローラ側に情報処理の多くを行わせてきた．

本研究では，数学的に同じ意味でも本質的な違いがあると考え，実現方法により学習速度が大きく変わることを示す．

３　タスク

３．１　タスク１

水平面内で運動する 2 リンクマニピュレータの各リンクをある一定の目標値に追従させることを目標とする.

３．２　タスク２

水平面内で運動する 2 リンクマニピュレータの各リンクの初期値をランダムに設定し，変化する目標値に対して追従させるタスクを行う.

３．３　2 リンクマニピュレータのモデル

タスク１およびタスク２で用いる２リンクマニピュレータのモデルを以下に示す.

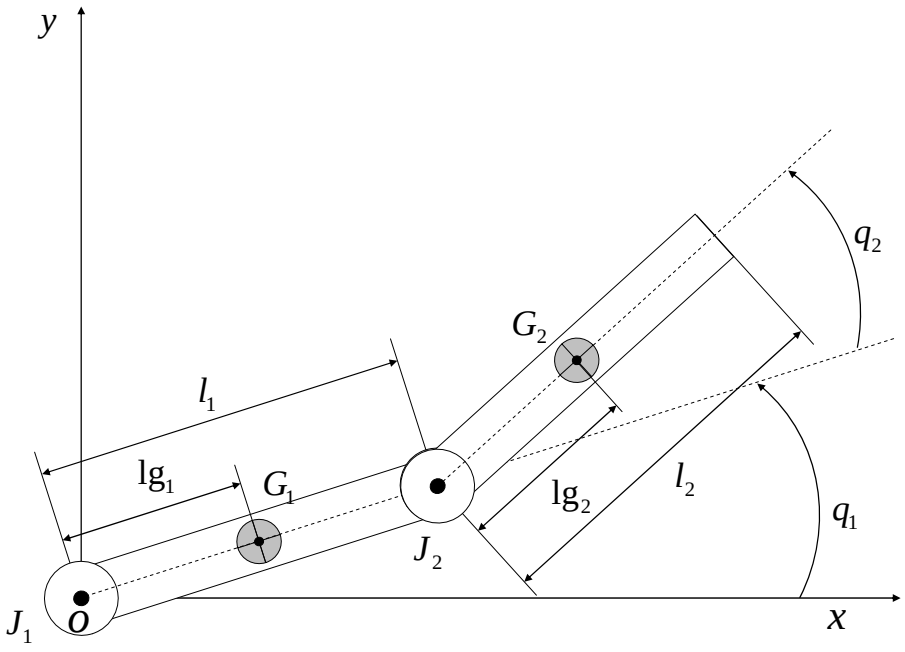

図 2　平面 2 関節ロボットのモデル

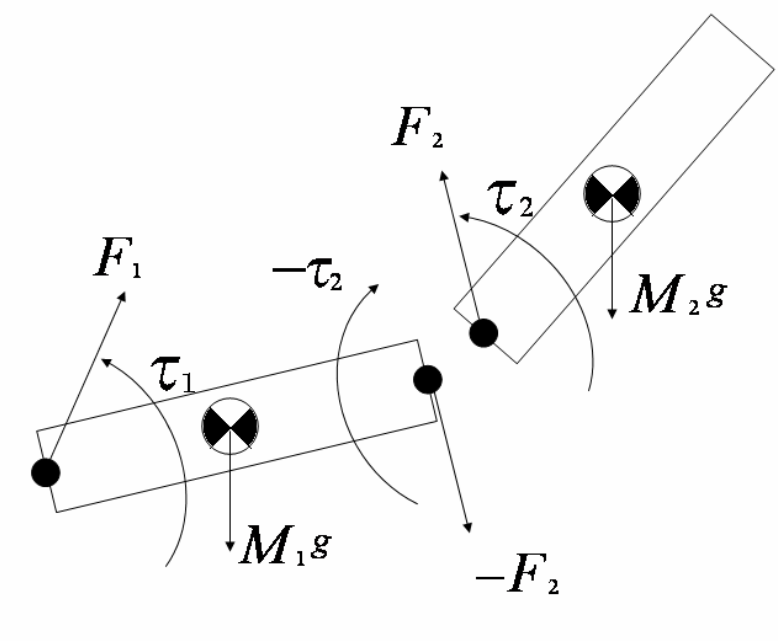

図３　平面２関節ロボットの力学

３．３．１　平面２関節ロボットの運動方程式

図 2，図 3 に示すような 2 つの回転関節の平面 2 関節ロボットアームを考える．第１リンクの回転関節 J_1 は地面に固定され，第１リンク先端の回転関節 J_2 を介して第２リンクが連結されている．　第１リンクは回転関節 J_1 に取り付けられたアクチュエータから τ_1 の駆動トルクを受け，第２リンクは回転関節 J_2 に取り付けられたアクチュエータから τ_2 の駆動トルクを受ける場合を考える．図２のように基本直行座標(J_1-xy)を設定したとき，x 軸からの第１リンクの相対回転角を q_1，J_1 と J_2 を結ぶ直線からの第 2 リンクの相対回転角を q_2 とする．また，J_1 と J_2 を結ぶ距離を l_1 とする．

さらに，第１リンクの質量を M_1，第１リンクの質量中心まわりの慣性モーメントを I_1，J_1 から第 1 リンク質量中心 G_1 までの距離を l_{g1} とする（G_1 は J_1 と J_2 を結ぶ線分の中点とする）．また，第 2 リンクの質量を M_1，第 2 リンクの質量中心まわりの慣性モーメントを I_2，J_2 から第 2 リンク質量中心 G_2 までの距離を l_{g2} とする（G_2 はエンドエフェクターと J_2 を結ぶ線分の中点とする）．さらに y 軸の負方向に作用している重力加速度を g とする．

このとき，第 1 リンクと第 2 リンクのそれぞれについて，拘束力と駆動トルクの作用・反作用の関係に注意して，並進と回転の運動方程式系を記述する．

まず，第 1 リンクの質量中心 G_1 の並進速度を v_1，第 1 リンクの角速度 ω_1 とすると，並進と回転に関する第 1 リンクの運動は以下で記述される．

$$M_1 \frac{dv_1}{dt} = F_1 - F_2 + M_1 g \tag{1}$$

$$I_1 \frac{d\omega_1}{dt} + \omega_1 \times (I_1 \omega_1) = \tau_1 - \tau_2 - l_{g1} \times F_1 + (l_1 - l_{g1}) \times (-F_2) \tag{2}$$

本例の場合，式 (2)の z 成分が回転運動に相当しており，平面運動の特殊性から，$I_2 \times I_1 \omega_1$ の項が削減することに注意する．

まったく同様にして，第 2 リンクの質量中心 G_2 の並進速度を v_2，第 2 リンクの角速度 ω_2 とすると，並進と回転に関する第 2 リンクの運動は以下で記述される．

$$M_2 \frac{dv_2}{dt} = F_2 + M_2 g \tag{3}$$

$$I_2 \frac{d\omega_2}{dt} = \tau_2 - l_{g2} F_2 \tag{4}$$

以上，式(1)～(4)が，例題を記述する運動方程式系全体である．これら 4 つの方程式に対して，先端のリンクから順に，拘束力と駆動トルクを左辺に記述する形に整理すると，以下の表現が得られる．

$$F_2 = M_2 \left(\frac{dv_2}{dt} - g\right) \tag{5}$$

$$\tau_2 = I_2 \frac{d\omega_2}{dt} + l_{g2} \times F_2 \tag{6}$$

$$F_1 = F_2 + M_1 \left(\frac{dv_1}{dt} - g\right) \tag{7}$$

$$\tau_1 = I_1 \frac{d\omega_1}{dt} + \tau_2 + l_{g1} \times F_1 + (l_1 - l_{g1}) \times F_2 \tag{8}$$

式(5)～(8)に現れる時間微分に関する項 $\frac{d\omega_1}{dt}$，$\frac{d\omega_2}{dt}$，$\frac{dv_1}{dt}$，$\frac{dv_2}{dt}$ が q_1，$\dot{q}_1$，$\ddot{q}_1$，q_2，$\dot{q}_2$，$\ddot{q}_2$ を用いて表現もする．

さらに，式(5)～(8)の拘束力 F_1，F_2 を消去することにより，式(6)と式(8)だけを残して，駆動トルク τ_1 と τ_2 に関するスカラーの運動方程式が 2 つ得られる．

式(5)～(8)の構造について説明を加える．式(5)～(8)の時間微分に冠する項は，幾何学的条件（長さと角度の関係）を用いて，運動方程式とは関係なく q_1，$\dot{q}_1$，$\ddot{q}_1$，q_2，$\dot{q}_2$，$\ddot{q}_2$ で表現できる．また，式(5)～(8)に現れる拘束力については，式(5)の F_2 が計算されれば，これを式(7)に代入して F_1 が計算される．さらに，駆動トルクについては，F_2 が計算されればこれを式(6)に代入してして τ_2 が計算され，この τ_2 と F_1，F_2 を式(8)に代入すれば τ_1 が計算される．この力とトルクの計算過程は，先端のリンクから順に関節拘束力が求まり，これにともなって関節トルクも順に求まっていくという，ドミノ倒しのような性質を有している．

また時間微分に関する項 $\frac{d\omega_1}{dt}$，$\frac{d\omega_2}{dt}$，$\frac{dv_1}{dt}$，$\frac{dv_2}{dt}$ に着目すると，速度の関係が第 1 リンクから第 2 リンクに向かって算出される性質があるため，時間微分に関する項も速度の算出と同じ性質を有している．これらの性質は空間運動を行う N 関節ロボットアームにおいても成り立つため，各関節に作用する力とトルクを少ない演算量で拘束に計算する手法として詳しく検討されている．

並進運動（ニュートンの運動方程式）と回転運動（オイラーの運動方程式）を連立させてロボットアームの運動法手指揮を導く場合，各リンクごとに並進運動と回転運動の式を必要とする．このとき，各関節に作用する力とモーメントをもれなく記述する必要がある．最終的に駆動トルク（あるいは駆動力）に関する運動方程式を算出するには，拘束力と拘束モーメントを消去しなければならないこの手順はロボットアームの関節数が増えると面倒になる．このため，上で触れたような少ない演算量の高速計算法が研究されたともいえる．各リン

クごとに並進運動と回転運動の式を立てるメリットは，単に駆動トルク（あるいは駆動力）に関する運動方程式を導出するだけでなく，その計算過程においてすべての関節拘束力と拘束モーメントが求められることにある．ロボットアームの機械設計を念頭においた場合，アームの運動時に作用する動的拘束力や動的モーメントを考慮したリンク構造の設計や関節軸受の選択が基本的な重要課題となるからである．

次項においては，拘束力や拘束モーメントを陽に記述せずに駆動トルク（あるいは，駆動力）に関する運動方程式が求められる．ラグランジュ（Lagrange）力学による導出法について述べる．本手法は，ロボットアーム全体の運動エネルギーと位置エネルギーを使って導出する手法であり，解析力学のハミルトン（Hamilton）の原理と密接な関係がある．ハミルトンの原理とは，「力学系がある特定の時間内に１つの状態から別の状態へ移動する場合，実際にその系が通る道筋は運動エネルギーと位置エネルギーの差の時間積分が最小となるような道筋となる」ことを主張する．

３．３．２．ラグランジュの運動方程式

ラグランジュの運動方程式そのものがニュートン運動方程式と等価であることは，よく書かれた力学の成書に譲る．本項では，これを利用してロボットアームの運動方程式を導出する手法を述べる．

ラグランジュの運動方程式とは，以下で表される方程式である．

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right) - \frac{\partial L}{\partial q_i} = Q_i \qquad (i = 1,2,\cdots,N) \tag{9}$$

ここで，Lはラグラジュアン（Lagrangean）と呼ばれ，Tを対象とするシステムの総運動エネルギー，Uを総位置エネルギーとすると，以下で定義される．

$$U = T - L \tag{10}$$

式(9)の q_iはシステムの一般化座標と呼ばれ，システムの状態を決定できる座標であれば関節変位座標や基準となる直交座標など，どのような座標を選択して

もよい．また Q_iは一般化力と呼ばれ，一般化座標に対応した外力（トルクも含む）を表している．一般化座標と一般化力はペアになっており，座標が並進変位であれば力は並進力となり，座標が回転変位であれば力はトルクとなる．なお，式(9)の Nは一般化座標の個数である．

ラグランジュの運動方程式を利用するメリットは，システムの運動エネルギーと位置エネルギーを表現すれば，ラグランジュアン L を作成した後，機械的に式(9)を計算すれば，所望の運動方程式が得られることにある．

取り扱うロボットアームは，図 2，図 3 に示した平面 2 関節ロボットアームである．

運動エネルギーと位置エネルギーを求めるためには，各リンクの質量中心の並進速度と各リンクの質量中心位置を記述する必要があるが，直接基準直交座標(J_1-xy)の座標表現から計算することにする．

まず，第 1 リンクの質量中心 G_1 の位置座標（x_1, y_1）は以下となる．

$$x_1 = l_{g1}\cos q_1 \tag{11}$$

$$y_1 = l_{g1}\sin q_1 \tag{12}$$

式(11) (12)の両辺を時間 t で微分して次式を得る．

$$\dot{x}_1 = -l_{g1}\sin q_1\dot{q}_1 \tag{13}$$

$$\dot{y}_1 = l_{g1}\cos q_1\dot{q}_1 \tag{14}$$

これより，

$$\|v_1\|^2 = \dot{x}_1{}^2 + \dot{y}_1{}^2 = l_{g1}{}^2\dot{q}_1{}^2 \tag{15}$$

となる．

また，第 1 リンクの角速度 ω_1 は $\dot{q}_1$ である．したがって，第 1 リンクの有する運動エネルギーを T_1，位置エネルギーを U_1 とすると，以下のようになる．

$$T_1 = \frac{1}{2} M_1 \|v_1\|^2 + \frac{1}{2} \omega_1{}^T (I_1 \omega_1)$$

$$= \frac{1}{2} M_1 l_{g1}{}^2 \dot{q}_1{}^2 + \frac{1}{2} I_1 \dot{q}_1{}^2$$

$$= \frac{1}{2} (M_1 l_{g1}{}^2 + I_1) \dot{q}_1{}^2 \tag{16}$$

$$U_1 = M_1 g y_1 = M_1 g l_{g1} \sin q_1 \tag{17}$$

次に，第 2 リンクの質量中心 G_2（x_1 ， y_2）は以下となる．

$$x_2 = l_{g1} \cos q_1 + l_{g2} \cos(q_1 + q_2)$$

(18)

$$y_2 = l_{g1} \sin q_1 + l_{g2} \sin(q_1 + q_2)$$

(19)

式(18) (19)の両辺を時間 t で微分して次式を得る．

$$\dot{x}_2 = -l_{g1} \sin q_1 \dot{q}_1 - l_{g2} \sin(q_1 + q_2)(\dot{q}_1 + \dot{q}_2) \tag{20}$$

$$\dot{y}_2 = l_{g1} \cos q_1 \dot{q}_1 + l_{g2} \cos(q_1 + q_2)(\dot{q}_1 + \dot{q}_2) \tag{21}$$

これより

$$\|v_2\|^2 = \dot{x}_2{}^2 + \dot{y}_2{}^2$$

$$= l_{g1}{}^2 \dot{q}_1{}^2 + l_{g2}{}^2 (\dot{q}_1 + \dot{q}_2)^2 + 2 l_1 l_{g2} \cos q_2 \dot{q} (\dot{q}_1 + \dot{q}_2) \tag{22}$$

となる．また，第 2 リンクの角速度 ω_2 は $(\dot{q}_1 + \dot{q}_2)$ である．したがって，第 2 リンクの有する運動エネルギーを T_1，位置エネルギーを U_2 とすると，以下のようになる．

$$T_2 = \frac{1}{2} M_2 \|v_2\|^2 + \frac{1}{2} \omega_2{}^T (I_2 \omega_2)$$

$$= M_2 (l_1{}^2 \dot{q}_1{}^2 + l_{g2}{}^2 (\dot{q}_1 + \dot{q}_2)^2 + 2 l_1 l_{g2} \cos q_2 \dot{q}_1 (\dot{q}_1 + \dot{q}_2)) + \frac{1}{2} I_2 (\dot{q}_1 + \dot{q}_2)^2 \quad (23)$$

$$U_1 = M_2 g y_2$$

$$= M_2 g (l_1 \sin q_1 + l_{g2} \sin(q_1 + q_2)) \quad (24)$$

以上から，このシステムのラグランジュアン $L=T_1+T_2-U_1-U_2$ を求めると，以下のようになる.

$$L = \frac{1}{2} (M_1 l_{g1}{}^2 + I_1) \dot{q}_1{}^2 + \frac{1}{2} M_2 (l_1{}^2 \dot{q}_1{}^2 + l_{g2}{}^2 (\dot{q}_1 + \dot{q}_2)^2 + 2 l_1 l_{g2} \cos q_2 \dot{q}_1 (\dot{q}_1 + \dot{q}_2))$$

$$+ \frac{1}{2} I_2 (\dot{q}_1 + \dot{q}_2)^2 - M_1 g l_{g1} \sin q_1 - M_2 g (l_1 \sin q_1 + l_{g2} \sin(q_1 + q_2)) \quad (25)$$

式(25)に対するラグランジュの運動方程式は以下の 2 組の方程式である.

$$\frac{d}{dt} \left(\frac{\delta L}{\delta \dot{q}_1} \right) - \frac{\delta L}{\delta q_1} = \tau_1 \quad (26)$$

$$\frac{d}{dt} \left(\frac{\delta L}{\delta \dot{q}_2} \right) - \frac{\delta L}{\delta q_2} = \tau_2 \quad (27)$$

式(26) (27)の 2 つのラグランジュ方程式を計算すると，最終的に次の 2 組の駆動トルクに関する運動方程式を得る.

$$(I_1 + M_1 l_{g1}{}^2 + I_2 + M_2 (l_1{}^2 + l_{g2}{}^2 + 2 l_1 l_{g2} \cos q_2)) \ddot{q}_1 + (I_2 + M_2 (l_{g2}{}^2 + l_1 l_{g2} \cos q_2)) \ddot{q}_2$$

$$- M_2 l_1 l_{g2} \sin q_2 (2 \dot{q}_1 \dot{q}_2 + \dot{q}_2{}^2) + M_1 g l_{g1} \cos q_1 + M_2 g (l_1 \cos q_1 + l_{g2} \cos(q_1 + q_2)) = \tau_1 \quad (28)$$

$$(I_1 + M_2 (l_{g2}{}^2 + l_1 l_{g2} \cos q_2)) \ddot{q}_1 + (I_2 + M_2 l_{g2}{}^2) \ddot{q}_2$$

$$+ M_2 l_1 l_{g2} \sin q_2 \dot{q}_1{}^2 + M_2 g l_{g2} \cos(q_1 + q_2) = \tau_2 \quad (29)$$

本研究では，粘性摩擦と弾性を用いることから 1 リンク，2 リンクにそれぞれ粘性摩擦 c_1，c_2 バネ定数 k_1，k_2 自然長 q_{o1}，q_{o2} を用いた運動方程式を以下に示す．

$$(I_1+M_1l_{g1}{}^2+I_2+M_2(l_1{}^2+l_{g2}{}^2+2l_1l_{g2}\cos q_2))\ddot{q}_1+(I_2+M_2(l_{g2}{}^2+l_1l_{g2}\cos q_2))\ddot{q}_2$$

$$-M_2l_1l_{g2}\sin q_2(2\dot{q}_1\dot{q}_2+\dot{q}_2{}^2)+M_1gl_{g1}\cos q_1+M_2g(l_1\cos q_1+l_{g2}\cos(q_1+q_2)) \quad (30)$$

$$+k_1(q_1-q_{o1})+c_1\dot{q}_1=\tau_1$$

$$(I_1+M_2(l_{g2}{}^2+l_1l_{g2}\cos q_2))\ddot{q}_1+(I_2+M_2l_{g2}{}^2)\ddot{q}_2+M_2l_1l_{g2}\sin q_2\dot{q}_1{}^2 \quad (31)$$

$$+M_2gl_{g2}\cos(q_1+q_2)+k_2(q_2-q_{o2})+c_2\dot{q}_2=\tau_2$$

４．提案手法

本研究では，学習しやすさを適応度として進化を行わせることで，マニピュレータの制御を例に，実世界の性質の有用性を示す．実世界の性質には粘性と弾性を用いる．

４．１　進化計算

本研究は進化計算のアルゴリズムに遺伝的アルゴリズム(GA)を用いる．GA とは生物の進化における遺伝子のメカニズムを模擬したアルゴリズムのことである．生物の遺伝子を数字列に置き換えて表現する．数字列はロボットの形態や性質を表す．初期個体を生成して適応度を計算し，淘汰，選択，交叉，突然変異を行い進化させていく．交叉はある個体の遺伝子とある個体の遺伝子を交換することで，突然変異はある遺伝子を無作為に変化させることにより実現する．交叉には一点交叉を用いて，選択にはルーレット選択を用いる．

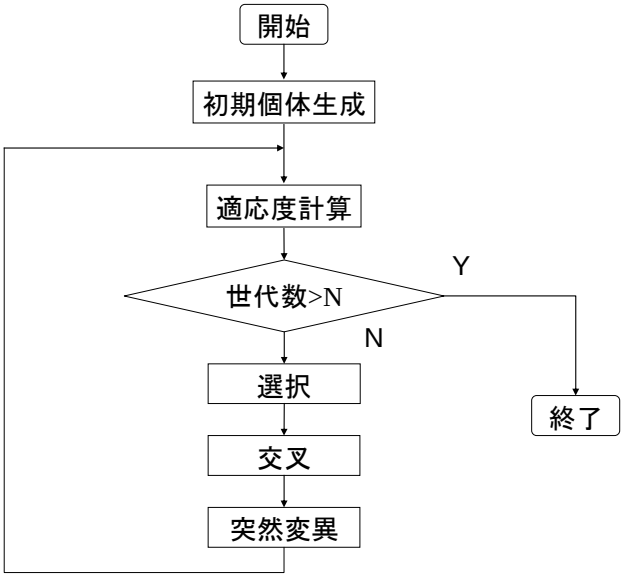

図４　進化計算のフローチャート

４．２　学習

本研究では，強化学習の中でも，Q 学習を用いる．Q 学習の 1 試行の流れを①から⑥に示す．

① エージェントは状態 s を観測する．

② エージェントは任意の行動選択方法に従って行動 a を実行する．

③ 環境から報酬 r を受け取る．

④ 状態遷移の状態 sを観測する．

⑤ 以下の式(32)により Q 値を更新する．

$$Q(s,a)\leftarrow(1-\alpha)\,Q(s,a)+\alpha\{r(s,a)+\gamma\max_{a'\in a}Q(s',a')\} \tag{32}$$

α は学習率($0<\alpha\leqq 1$)，γ は割引率($0\leqq\gamma<1$)

⑥ 報酬が得られるまで以上を繰り返す．

５．シミュレーション

５．１　タスク 1

５．１．１　2 リンクマニピュレータの設定

質量とリンク長の設定は M_1= M_2=3.0kg，l_1= l_2=0.1m とする．マニピュレータの運動方程式はルンゲ･クッタ法を用いて解き，サンプリングタイムは 0.001s に設定する．

５．１．２　Q 学習の設定

表 1 および表 2 は Q 学習の各種パラメータを示す．状態および行動は，GA の遺伝子によって決められる．

表 1　Q 学習の設定

学習回数	20000
Q 値の更新間隔	0.001 s
q_1 の初期値	$\frac{\pi}{6}$ rad
q_2 の初期値	$\frac{\pi}{6}$ rad
正の報酬	100
負の報酬	-100
α	0.5
γ	0.9999
epsilon greedy	0.01

表 2　報酬の条件

報酬(100)	$(-0.1 \leqq q_1 < 0.1)$ $\wedge$ $(-0.1 \leqq q_2 < 0.1)$ $\wedge$ $(-0.1 \leqq \dot{q}_1 < 0.1)$ $\wedge$ $(-0.1 \leqq \dot{q}_2 < 0.1)$
報酬(-100)	$q_1 < -0.9$
報酬(-100)	$q_2 < -0.9$
報酬(-100)	$q_1 \geqq 0.7$
報酬(-100)	$q_2 \geqq 0.7$
報酬(-100)	$\dot{q}_1 < -0.9$
報酬(-100)	$\dot{q}_2 < -0.9$
報酬(-100)	$\dot{q}_1 \geqq 0.7$
報酬(-100)	$\dot{q}_2 \geqq 0.7$

５．１．３　GA の設定

図 5 に遺伝子の設定を示す．遺伝子長を 7 とし，1 番目の遺伝子は，q_1 および q_2 を状態として認識するか否かを決める遺伝子であり，2 番目の遺伝子は，$\dot{q}_1$および$\dot{q}_2$を状態として認識するか否かを決める遺伝子である．3 番目の遺伝子は τ_1 および τ_2 を行動とするか否かを決める遺伝子である．これらの遺伝子の値が全て 0 の場合には，状態・行動空間は構成されないため，学習は行われず，マニピュレータは自身の動力学特性のみに従って運動する．4 番目から 7 番目の遺伝子は，粘性摩擦係数およびバネ定数の大きさを決める遺伝子である．

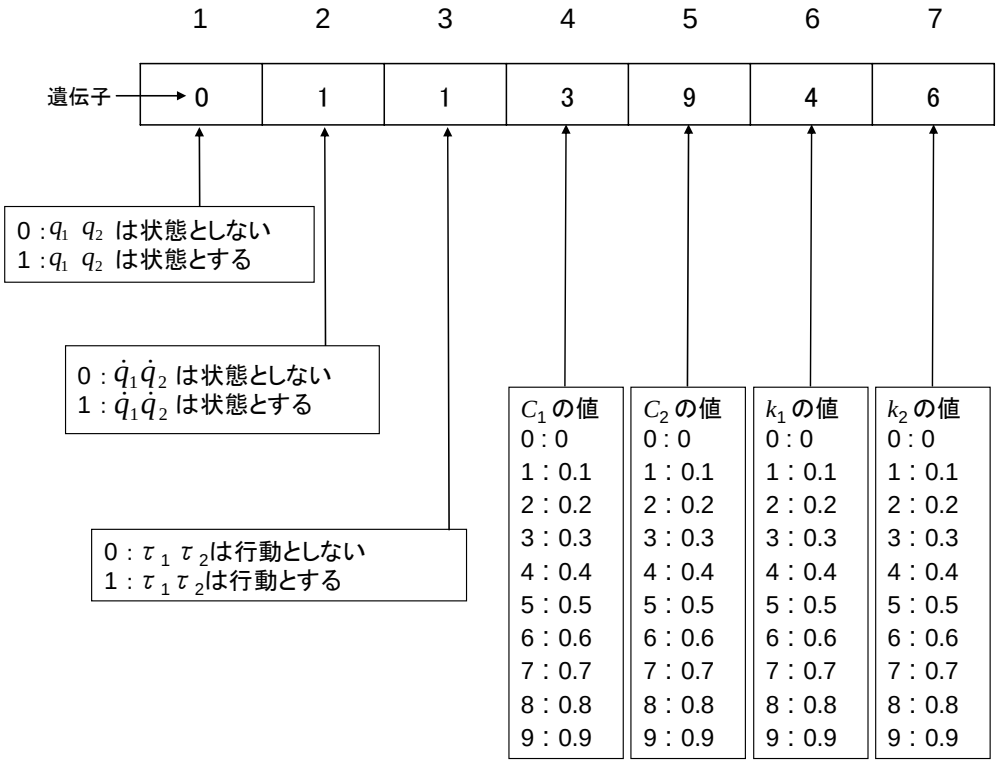

図 5　遺伝子の設定

遺伝子の値により，状態として認識されることとなった場合には，表 3 に示すように，各状態は 10 の領域に分割されて表されるものとする．したがって，状態の大きさは，最大で 10^4 通りとなる．

行動は τ_1，τ_2 の出力をそれぞれ表 4 のように 10 分割して考え，合計 100 通りの行動をとるものとする．

世代数，個体数，交叉率，突然変異率の設定を表 5 に示す．適応度は式(6)とする．ここで，r_n は n 回目の試行において獲得された報酬である．

表 3　状態の構築

q_1　q_2 [rad]	$\dot{q}_1$　$\dot{q}_2$ [rad/s]
q＜-0.9	$\dot{q}$＜-0.9
-0.9≦q＜-0.7	-0.9≦$\dot{q}$＜-0.7
-0.7≦q＜-0.5	-0.7≦$\dot{q}$＜-0.5
-0.5≦q＜-0.3	-0.5≦$\dot{q}$＜-0.3
-0.3≦q＜-0.1	-0.3≦$\dot{q}$＜-0.1
-0.1≦q＜0.1	-0.1≦$\dot{q}$＜0.1
0.1≦q＜0.3	0.1≦$\dot{q}$0.3
0.3≦q＜0.5	0.3≦$\dot{q}$＜0.5
0.5≦q＜0.7	0.5≦$\dot{q}$＜0.7
0.7≦q	0.7≦$\dot{q}$

表 4　行動の構築

τ_1　τ_2[N・m]
-0.5
-0.4
-0.3
-0.2
-0.1
0
0.1
0.2
0.3
0.4

表 5　GA の設定

世代数	100
個体数	100
交叉率	0.3
突然変異率	0.1

$$fitness = \sum_{n=1}^{N} r_n \,(N\text{は試行数}) \tag{33}$$

５．１．４　タスク 1 シミュレーション結果

図 6 に 1 世代毎の適応度の平均値を示す．また，図 7 に 100 世代目の最も適応度の高い遺伝子を，表 6 にその遺伝子の意味を示す．また，図 8 にそのときの応答の様子を示す．これらから，学習を必要とせず，マニピュレータ自身のダイナミクスによって制御を実現する機構が得られたことがわかる．

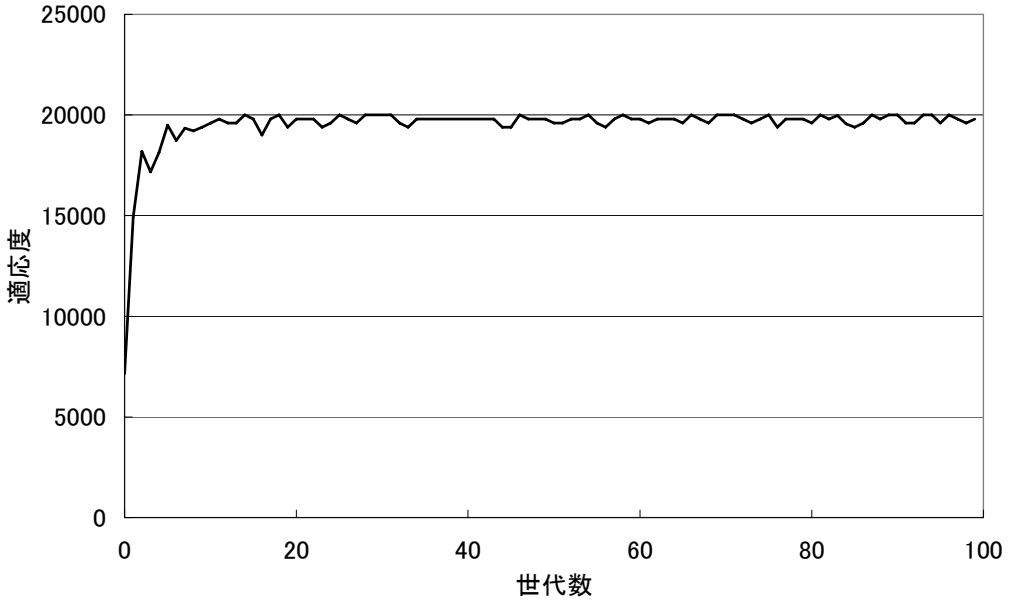

図 6　平均適応度

0	0	0	7	8	4	5

図 7　100 世代目に獲得された遺伝子

表 6　100 世代目に獲得された遺伝子の意味

q_1　q_2	状態としない
$\dot{q}_1$　$\dot{q}_2$	状態としない
τ_1　τ_2	行動としない
c_1	0.7
c_2	0.8
k_1	0.4
k_2	0.5

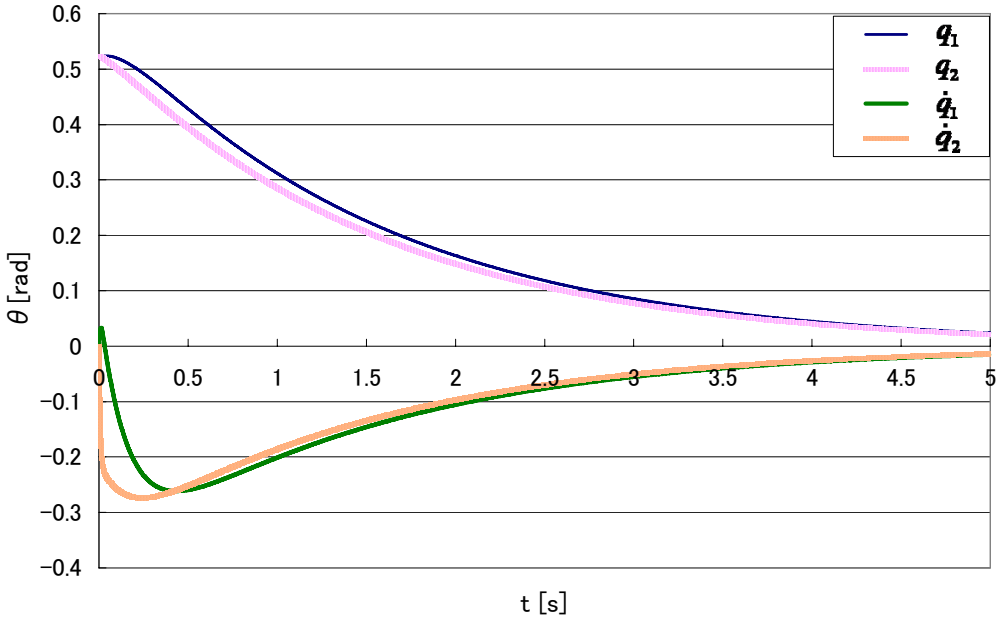

図 8　100 世代における応答の一例

進化の有用性を検証するため，進化を行わず学習のみによって制御を実現する．実世界の性質は使用せず，粘性摩擦係数およびバネ定数はともに 0 とする．q_1，q_2，$\dot{q}_1$，$\dot{q}_2$を状態として τ_1，τ_2 を行動とする．状態空間の構築は表 3 と同様に設定し，行動空間の設定は表 7 に示すものとした．ここで得られた 100 試行毎の成功率を図 9 に示し，図 10 に 20000 回試行後の結果を示す．

表 7　行動空間の設定

行動	τ_1	τ_2
0	0	0
1	0	0.3
2	0	-0.3
3	0.3	0
4	0.3	0.3
5	0.3	-0.3
6	-0.3	0
7	-0.3	0.3
8	-0.3	-0.3

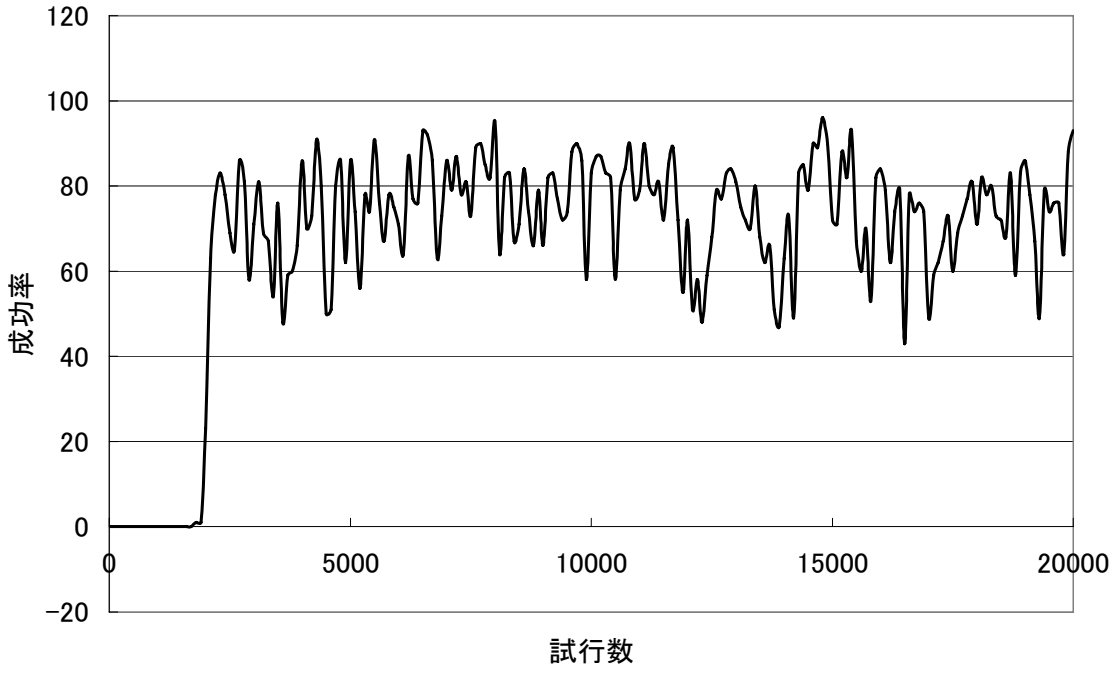

図 9　100 試行毎の成功率

この結果から，学習のみによって制御系を実現することは可能であるものの，学習が収束するまでに多くの試行を必要とすることが分かる．

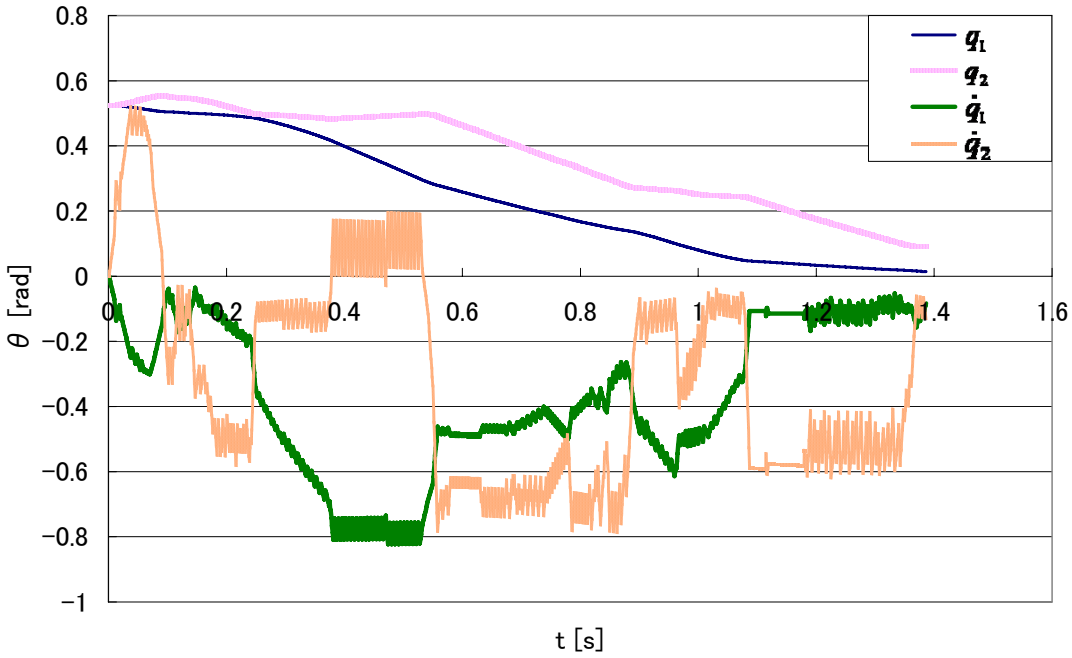

図 10　20000 試行後の結果

以上の結果から，実世界の性質である粘性と弾性をうまく利用することで学習を必要としない機構が進化的に獲得されていることが分かる．

５．２　タスク 2

５．２．１　2 リンクマニピュレータの設定

タスク 1 と同様の設定とする．

５．２．２　Q 学習の設定

Q 値の更新式を式(6)に，Q 学習の各パラメータを表 8 に示す．状態および行動は，GA の遺伝子によって決められる．各関節角度の初期値は 1 試行毎に表 2 のいずれかに設定される．目標値(q_{d1} および q_{d2})は，1 試行毎に表 9 の中からランダムに選択される．報酬は，関節の角度が選択された目標値に対して，±0.1rad の領域に入り，かつ角速度が±0.1rad/s の領域にリンクが入った場合，正の報酬を与える．また，リンクの角度が q<-0.5 または 0.5≦q または角速度が$\dot{q}$<-0.5 または 0.5≦$\dot{q}$の領域となった場合，負の報酬を与える．1 試行につき 10 秒間 2 リンクマニピュレータを運動させる．また，報酬にたどり着いた場合，

そこで試行を終了させる.

表 8　Q 学習の設定

学習回数	5000
Q 値の更新間隔	0.001
正の報酬	100
負の報酬	-100
α	0.5
γ	0.9999
Epsilon greedy	0.1

表 9　初期値と目標値の設定

初期値(q_1 および q_2)	目標値(q_{d1} および q_{d2})
-0.4	-0.4
-0.2	-0.2
0	0
0.2	0.2
0.4	0.4

５．２．３　GA の設定

図 11 に遺伝子の設定を示す．遺伝子長を 7 とし，1 番目の遺伝子は，q_1 および q_2 を状態として認識するか否かを決める遺伝子であり，2 番目の遺伝子は，$\dot{q}_1$ および $\dot{q}_2$ を状態として認識するか否かを決める遺伝子である．3 番目の遺伝子は τ_1 および τ_2 を行動とするか否かを決める遺伝子である．これらの遺伝子の値が全て 0 の場合には，状態・行動空間は構成されないため，学習は行われず，マニピュレータは自身の動力学特性のみに従って運動する．4 番目から 7 番目の遺伝子は，粘性摩擦係数およびバネ定数の大きさを決める遺伝子である．

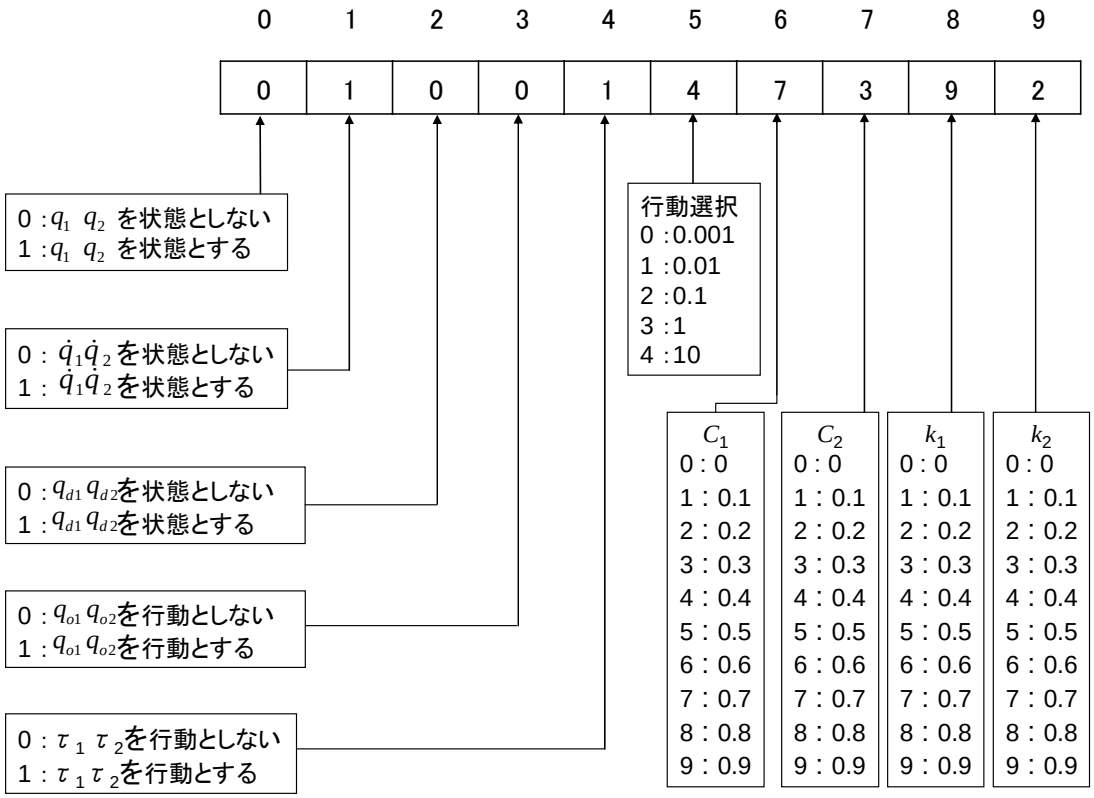

図 11　遺伝子の設定

1 番目の遺伝子が 1 を選択した場合，0.2rad の分割幅で-0.5≦q＜0.5 の領域を分割した 5 の領域と q＜-0.5 の領域および 0.5≦q の領域のそれぞれ 7 の領域を状態とする．2 番目の遺伝子が 1 を選択した場合，0.2rad/s の分割幅で-0.5≦$\dot{q}$＜0.5 の領域を分割した 5 の領域と$\dot{q}$＜-0.5 の領域および 0.5≦$\dot{q}$の領域のそれぞれ 7 の領域を状態とする．3 番目の遺伝子が 1 を選択した場合，q_{d1} および q_{d2} は-0.4rad，-0.2rad，0.0rad，0.2rad，0.4rad の 5 分割した状態とする．これらを表 10 に示す．4 番目の遺伝子が 1 を選択した場合，q_{o1} および q_{o2} は-0.4rad，-0.2rad，0.0rad，0.2rad，0.4rad の中から行動選択し，合計 25 の行動をとる．5 番目の遺伝子が 1 を選択した場合，τ_1およびτ_2は-0.4 Nm～0.4Nm の間を 0.1 ずつ分割し，合計 81 の行動をとる．これらを表 11 に示す．

世代数，個体数，交叉率，突然変異率の設定を表 12 に示す．適応度は式(4)とする．ここで，第 n 回目の試行において 1 度も報酬にたどり着かなかった場合 r_n

は 0 とし，第 n 回目の試行において 1 度でも報酬にたどり着いた場合 r_n は 1 とする．

表 10　状態の構築

q_1　q_2 [rad]	$\dot{q}_1$　$\dot{q}_2$ [rad/s]	q_{o1}　q_{o2} [rad]
q＜-0.5	$\dot{q}$＜-0.5	-0.4
-0.5≦q＜-0.3	-0.5≦$\dot{q}$＜-0.3	-0.2
-0.3≦q＜-0.1	-0.3≦$\dot{q}$＜-0.1	0.0
-0.1≦q＜0.1	-0.1≦$\dot{q}$＜0.1	0.2
0.1≦q＜0.3	0.1≦$\dot{q}$＜0.3	0.4
0.3≦q＜0.5	0.3≦$\dot{q}$＜0.5	
0.5≦q	0.5≦$\dot{q}$	

表 11　行動の構築

τ_1　τ_2[N・m]	q_{d1}　q_{d2} [rad]
-0.4	-0.4
-0.3	-0.2
-0.2	0
-0.1	0.2
0	0.4
0.1	
0.2	
0.3	
0.4	

表 12　GA の設定

世代数	100
個体数	100
交叉率	0.3
突然変異率	0.1

５．２．４　タスク 2 のシミュレーション結果

図 12 に 1 世代毎の適応度の平均値を示す．また，図 13 に 100 世代目の最も適応度の高い遺伝子を，表 13 にその遺伝子の意味を示す．また，図 14 にそのときの成功率を示す．これらから，学習を必要とせず，マニピュレータ自身のダイナミクスによって制御を実現する機構が得られたことがわかる．図 15 に 100 世代目に獲得された遺伝子の学習後の応答を示す．

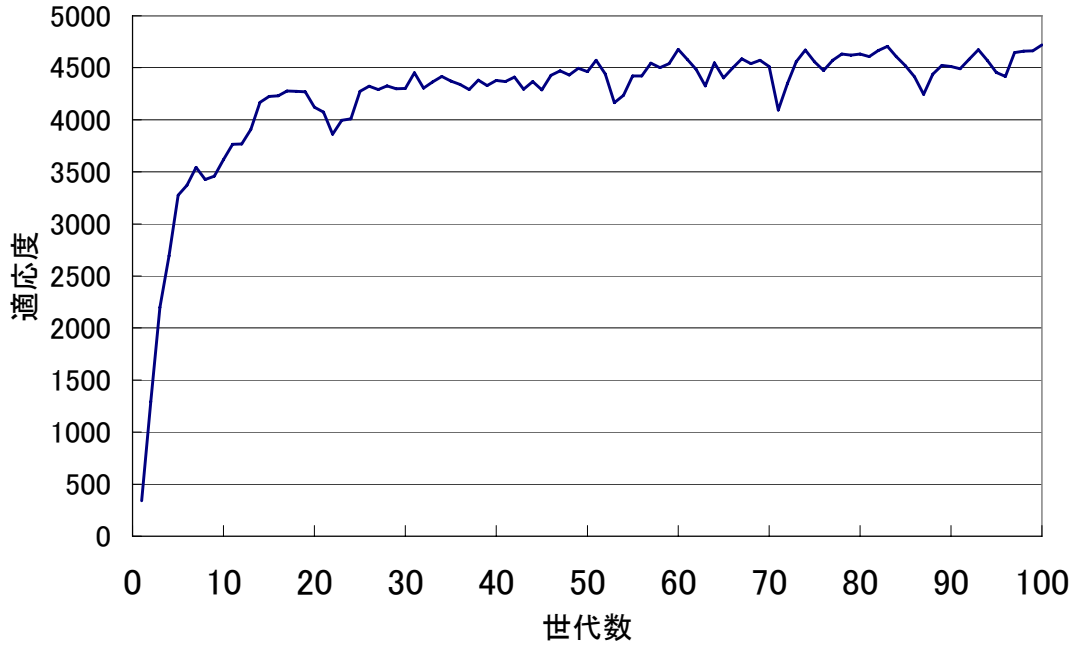

図 12　平均適応度

0	0	1	1	0	4	7	2	9	8

図. 13　100 世代目に獲得された遺伝子

表 13　100 世代目に獲得された遺伝子の意味

q_1　q_2	状態としない
$\dot{q}_1$　$\dot{q}_2$	状態としない
q_{o1}　q_{o2}	状態とする
q_{d1}　q_{d2}	行動とする
τ_1　τ_2	行動としない
行動選択	10.0
c_1	0.7
c_2	0.2
k_1	0.9
k_2	0.8

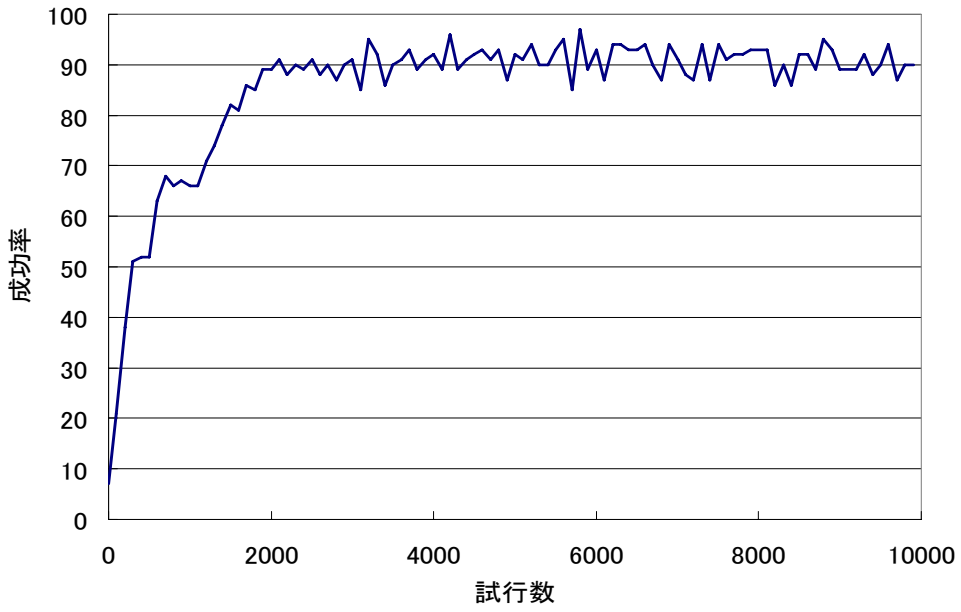

図 14　100 世代目に獲得された遺伝子の成功率

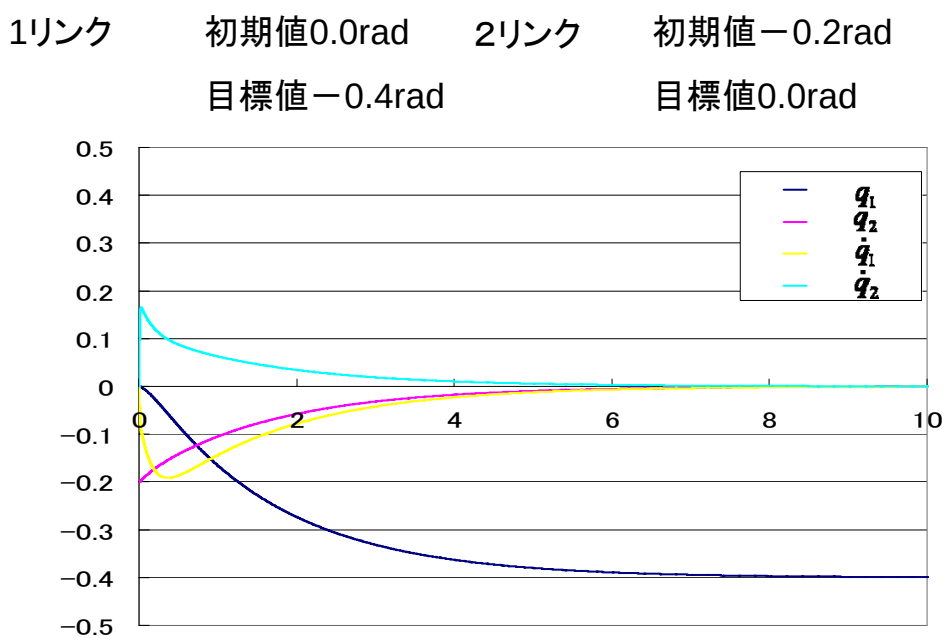

図 15　100 世代における応答の一例

獲得された遺伝子から，状態として角度の目標値のみが使われており，実際のマニピュレータの角度および角速度は学習に用いられていないことが分かる．行動についても，バネの自然長を変化させる行動のみが使われており，各関節のトルクは学習によって調整されていないことが分かる．これは，与えられた目標値に対して適切な物理パラメータ（バネの自然長）を設定することが学習の役割となっており，実際の制御は実世界の力学的性質を利用して，マニピュレータ自身によって行われる構成が獲得されたためである．

進化の有用性を検証するため，進化を行わず粘性と弾性のみによって学習を行い，制御を実現する．実世界の性質は使用せず，粘性摩擦係数およびバネ定数はともに 0 とする．q_1，q_2，$\dot{q}_1$，$\dot{q}_2$を状態としてτ_1，τ_2を行動とする．状態空間の構築は表 14 と同様に設定し，行動空間の設定は表 15 に示すものとした．ここで得られた 100 試行毎の成功率を図 16 に示し，図 17 に 100000 回試行後の結果を示す．

表 14　状態空間の設定

q_1　q_2 [rad]	$\dot{q}_1$　$\dot{q}_2$ [rad/s]
q＜-0.5	$\dot{q}$＜-0.5
-0.5≦q＜-0.3	-0.5≦$\dot{q}$＜-0.3
-0.3≦q＜-0.1	-0.3≦$\dot{q}$＜-0.1
-0.1≦q＜0.1	-0.1≦$\dot{q}$＜0.1
0.1≦q＜0.3	0.1≦$\dot{q}$＜0.3
0.3≦q＜0.5	0.3≦$\dot{q}$＜0.5
0.5≦q	0.5≦$\dot{q}$

表 15　行動空間の設定

行動	τ_1	τ_2
0	0	0
1	0	0.3
2	0	-0.3
3	0.3	0
4	0.3	0.3
5	0.3	-0.3
6	-0.3	0
7	-0.3	0.3
8	-0.3	-0.3

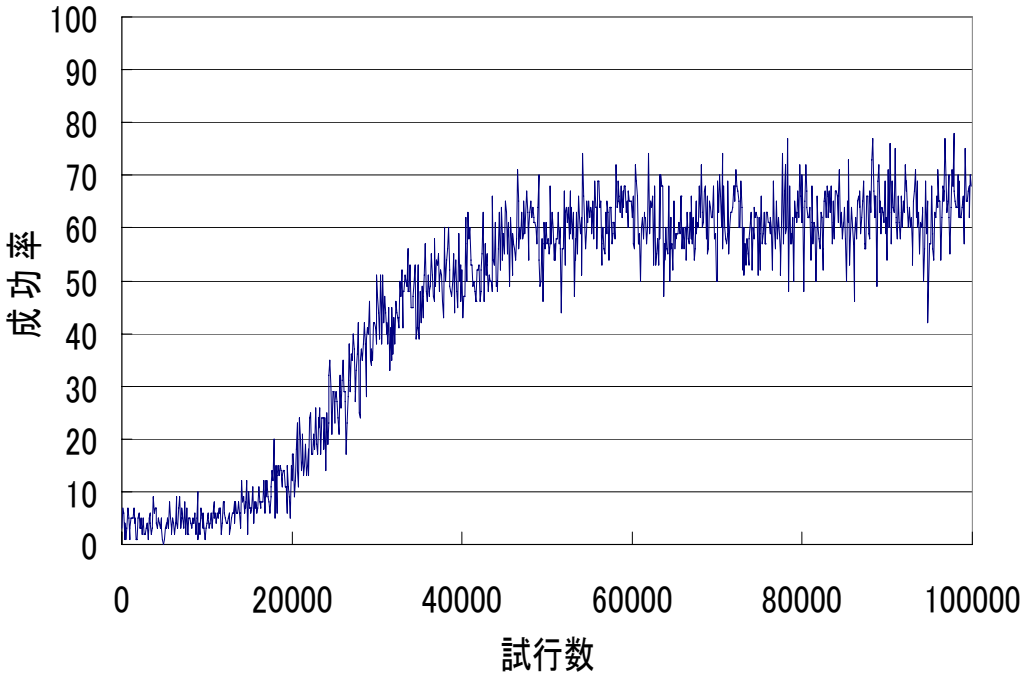

図 16　100 試行毎の成功率（粘性と弾性なし）

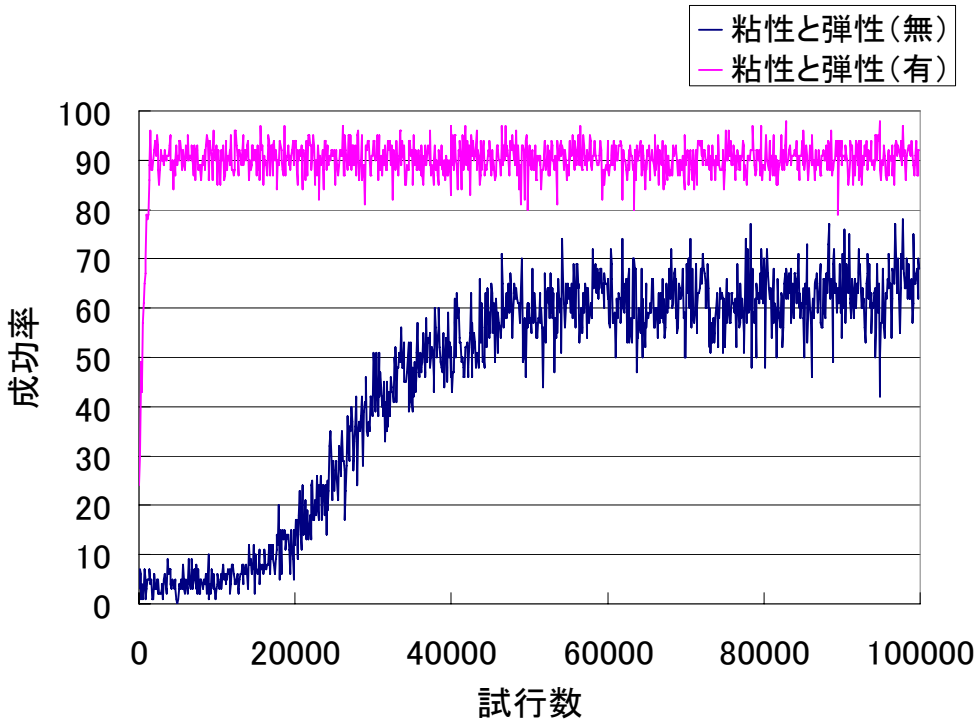

図 17　図 14 と図 16 の結果

この結果から，粘性と弾性が存在しなくても学習を行い，制御系を実現することは可能であるが，粘性と弾性が有る時と無い時では学習の立ち上がりの速さと収束する成功率の違いが大きく違いが現れることが図 17 より分かる．これより，粘性と弾性を学習に適応した際，有用性があることが証明された．

６．考察

獲得された遺伝子から，状態として角度の目標値のみが使われており，実際のマニピュレータの角度および角速度は学習に用いられていないことが分かる．行動についても，バネの自然長を変化させる行動のみが使われており，各関節のトルクは学習によって調整されていないことが分かる．これは，与えられた目標値に対して適切な物理パラメータ（バネの自然長）を設定することが学習の役割となっており，実際の制御は実世界の力学的性質を利用して，マニピュレータ自身によって行われる構成が獲得されたためである．

７．おわりに

本研究では，2 リンクマニピュレータの制御を例に，「進化」，「学習」，「実世界の性質」のそれぞれの役割について検討した.

その結果，「進化」は学習を容易にする身体を獲得できること，「学習」は進化の方向を決める指標になり，実世界の物理パラメータを適切な値に調節するために使われること，「実世界の性質」は状態行動空間を縮退し，学習を容易にすることを確認した.

参 考 文 献

[1] 佐々木正人:“アフォーダンス入門－若きロボット研究者との Q&A－”，日本ロボット学会誌，Vol.24，No.7，pp.791-796，2006.

[2] R. Pfeifer，F.Iida and G.Gomez: “Designing Intelligent Robots-On the Implication of Embodiment-”，日本ロボット学会誌，Vol.24，No.7，pp.791-796，2006.

[3] 有本卓:“巧みさの演出 : ダイナミクスベースト制御”，日本ロボット学会誌，Vol.24，No.7，pp.791-796，2006.

[4] 大須賀公一:“ダイナミクスベースト制御の「こころ」”，日本ロボット学会誌，Vol.24，No.7，pp.791-796，2006.

[5] 石黒章夫:“知の基盤としてのしぶとさの創成”, 日本ロボット学会誌, Vol.24, No.7，pp.791-796，2006.

[6] 古山宣洋:“運動を導く知覚システム”，日本ロボット学会誌，Vol.24，No.7，pp.791-796，2006.

[7] 伊藤一之，大須賀公一，石黒章夫，古山宣洋:“実世界の性質を利用した知覚と制御”，日本ロボット学会誌，Vol.24，No.7，pp.791-796，2006.

[8] 伊藤一之，福森嘉孝: “受動知能に基づく強化学習の汎化能力に関する研究”-蛇型ロボットへの適用-，日本ロボット学会学術講演会，3E11，2005

[9] 伊藤一之，福森嘉孝“知覚量に基づく制御系設計-蛇型ロボットの方向の知覚量を用いたフィードバック制御-”，計測自動制御学会論文集，Vol.42，No.4，pp，436-445(2006)