学習オートマトンによる冗長マニピュレータのパスプランニングの戦略獲得

7G−5

成瀬継太郎　嘉数侑昇

北海道大学

1. はじめに

本研究では，複数の障害物を含む作業空間における冗長マニピュレータのパスプランニングのための戦略を獲得することを目的とする．

確率的学習オートマトンは，その内部状態に応じて，出力集合の中から確率的に出力を選択する．その出力に対して評価者である環境からの反応に基づき，各出力確率を更新することによって，試行錯誤的に学習を行なう(再強化学習)．その結果として，与えられた環境に適応することが知られている[NARE89]．

冗長マニピュレータのパスプランニング問題に対しては，手先位置がその目標位置に近づいているかどうか，近くに障害物があるかどうかに基づいて環境からの反応を決定することによって，障害物を回避しながら初期状態から目標点に到達するパスを学習(適応)することが期待される．

このとき，マニピュレータの状態を学習オートマトンの内部状態とすることによって，各内部状態のときにどの出力を選択すれば良いかの戦略を獲得することが可能になる．

ここでは，マニピュレータの各関節に学習オートマトンを付加し，各学習オートマトンが独立に学習することによってパスプランニングのための戦略を獲得する手法を提案する．具体的には，学習オートマトンの内部状態を各関節の角度と角速度に基づいて決定し，出力を各関節の出力トルクとし，手先位置と目標位置との距離の変化と周囲の障害物の状態によって環境からの反応を決定する．最後に計算機実験によりその動作を確認する．

2. 学習オートマトン

ここでは，有限出力の確率的学習オートマトンを扱う．学習オートマトン LA_i は，時刻 t において出力集合 $\{a_j^i \mid j=1,\cdots,N_i\}$ から，確率分布 $\mathbf{p}^i(t)$ にしたがいながら，出力 $a^i(t)$ を選択する．ここで，

$$\mathbf{p}^i(t) = (p_1^i(t),\cdots,p_{N_i}^i(t)), \tag{1}$$

$$p_j^i(t) = Pr(a^i(t) = a_j^i), \tag{2}$$

$$\sum_{j=1}^{N_i} p_j^i(t) = 1. \tag{3}$$

また，初期状態 ($t=0$)での確率分布 $\mathbf{p}^i(0)$ は等確率，すなわち $\mathbf{p}^i(0) = (\frac{1}{N_i},\cdots,\frac{1}{N_i})$ であるとする．

選択されたオートマトンの出力は環境へ入力され，環境は時刻 t での反応 $\psi(t) \in \{0,1\}$ を出力する．ここで，0 は Reward を，1 は Penalty を表す．

学習オートマトンは，この環境からの反応に基づき出力確率分布を更新することによって学習を行なう．出力確率の更新法には様々な方法[OOMM91][NARE89][NARU92]があるが，その一つに L_{R-P} と呼ばれるものがある．これは，$a^i(t) = a_j^i$ のとき，以下のような出力確率分布の更新を行なう．

$$p_j^i(t+1) = \begin{cases} p_j^i(t) + c^i(1-p_j^i(t)) & b(t)=0 \\ (1-d^i)p_j^i(t) & b(t)=1 \end{cases} \tag{4}$$

$$p_{k\,(\neq j)}^i(t+1) = \begin{cases} (1-c^i)p_k^i(t) & b(t)=0 \\ \frac{d^i}{N_i-1} + (1-d^i)p_k^i(t) & b(t)=1 \end{cases} \tag{5}$$

ここで，c^i は Reward に関する，d^i は Penalty に関する[0,1] の学習係数である．

3. 提案手法

パスプランニングには，関節空間におけるものと作業空間におけるもの大別できる．提案する手法においては，作業空間におけるパスプランニングを扱う．そして，各学習オートマトンが学習(適応)した結果として，関節空間におけるトルク出力の戦略が獲得される．

ここで扱う問題は m 関節の冗長マニピュレータに対して，マニピュレータの初期状態と手先位置の目標位置 P_f が与えられたときに，その間の各関節の出力トルク系列を学習により求める問題である．

まず，マニピュレータの各関節 i に学習オートマトン LA_i を付加する．LA_i の内部状態は，各関節の角度と角速度に基づいて決定する．すなわち，LA_i 状態遷移関数を f_i，内部状態の集合を S_i，マニピュレータの関節 i の変位を θ_i として，

$$f_i \mid \theta_i \times \dot{\theta}_i \rightarrow S_i. \tag{6}$$

次に，LA_i の出力は関節 i の出力トルクとする．ここでは，出力トルクは N_i 個に離散化されているものとする．例えば，$N_i = 5$ のとき τ をトルクを表す定数として，$(-2\tau, -\tau, 0, \tau, 2\tau)$ 等が考えられる．LA_i は，(1) 式から (3) 式の確率にしたがって出力を

Strategy Acquisition of Redundant Manipulator Path Planning by Learning Automata
Keitarou NARUSE Yukinori KAKAZU
Hokkaido Univ.

選択する．

マニピュレータの動力学方程式は，次の式で与えられる．

$$\tau = J(\Theta)\ddot{\Theta} + C(\Theta,\dot{\Theta}) + D\dot{\Theta} + G(\Theta). \quad (7)$$

ここで，τは関節トルクベクトル，Θは関節変位ベクトル，$J(\Theta)$は慣性行列，$C(\Theta,\dot{\Theta})$は遠心力，コリオリ力に関する項，Dは粘性摩擦係数，$G(\Theta)$は重力項である．

各関節からの出力トルクと (7) 式をもとに順動力学問題を計算することによって，時刻 t での手先位置 $P(t)$ が決まる．また，障害物 O_k は作業空間における点として表現され，その点を中心として半径 r_k の禁止領域設ける．

LA_i が環境に適応するためには時刻 t 環境からの反応 $\psi(t)$ を定めなければならないが，ここではその要因として手先位置の目標点への接近に関する基準 $\psi_1(t)$ と，周りの障害物の状態に関する基準 $\psi_2(t)$ の二つを考え，$\psi(t)$ はこの両者の結合により定める．

(1) 手先位置に関する基準 $\psi_1(t)$

手先位置 $P(t)$ の評価として目標位置 P_f とのユークリッド距離 d を採用する．$P(t)$ が P_f に近づいていれば Reward，それ以外の場合は Penalty とする．すなわち，

$$\psi_1(t) = \begin{cases} 0, & \text{if } d(P_f,P(t)) < d(P_f,P(t\text{-}1)) \\ 1, & \text{otherwise} \end{cases}. \quad (8)$$

(2) 障害物の状態に関する基準 $\psi_2(t)$

マニピュレータと各障害物 O_k までの距離 D_k に基づいて $\psi_2(t)$ を決定する．

$$\psi_2(t) = \begin{cases} 0, & \text{if } r_k < d(O_k,D_k) \text{ , for all } k \\ 1, & \text{otherwise} \end{cases}. \quad (9)$$

(3) $\psi(t)$ の決定

$\psi_1(t)$ と $\psi_2(t)$ の結合方法に様々なものが考えられるが，ここでは以下の式により結合する．

$$\psi(t) = \psi_1(t) \text{ Or } \psi_2(t). \quad (10)$$

すなわち，手先位置と目標位置の距離が減少しつつ，かつ周囲に障害物がないときのみ Reward が与えられ，それ以外は Penalty が与えられる．

$\psi(t)$ の値に基づいて，(4)(5)式 にしたがい，LA_i は初期状態から目標位置までのパス，すなわち各関節の各状態における出力トルクの学習を独立に行なう．その結果として，各状態における出力トルク集合上の確率分布，すなわち各状態での出力に関する確率的な戦略が獲得される．

4. 計算機実験

前節で提案した手法の動作を確認するために計算機実験を行なった．実験には 2 関節の平面マニピュレータを用いた．マニピュレータのリンクの長さ，質量，粘性抵抗は各者とも1.0 であり，各関節の出力トルクは両者とも(-0.5, 0.0, 0.5) の3出力である．学習オートマトンの内部状態は，関節角に関して 8 状態(等分割)，角速度に関して 6 状態(しきい値 -0.10,-0.05, 0.0, 0.05, 0.10)の 48 状態とする．学習係数は，c^i=0.05, d^i = 0.10 であり，2000 回の試行を行なった．

図1 は，2000 試行でのマニピュレータの軌跡である．図中，黒点は目標位置，円は障害物とその禁止領域である．この図より，初期状態から目標位置へ障害物を避けながら到達していることが確認できる．

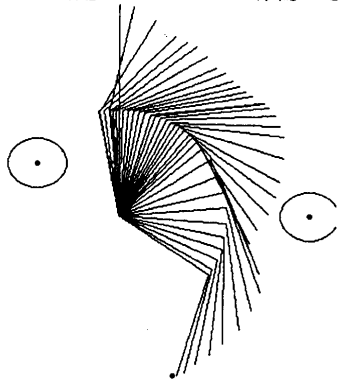

図1 2000 試行でのマニピュレータの軌跡

図2 は，2000 試行の後に獲得された関節 1 に関する戦略の一部をルール型(条件部とトルクの出力確率分布)に表記している．これより，与えられた問題に対する戦略が獲得されていることが解かる．

if $\frac{7}{4}\pi < \theta < 2\pi$, $\dot{\theta} \le$ -1.0, then (0.043, 0.343, 0.614)

if $\frac{7}{4}\pi < \theta < 2\pi$, -1.0 < $\dot{\theta} \le$ -0.5, then (0.012, 0.656, 0.332)

if $\frac{7}{4}\pi < \theta < 2\pi$, -0.5 < $\dot{\theta} \le$ -0.0, then (0.000, 0.946, 0.054)

図2 獲得された戦略(一部)

5. おわりに

冗長マニピュレータのパスプランニング問題に対して，マニピュレータの各関節に学習オートマトンを付加し，手先位置と目標位置との距離および周囲の障害物の状態に基づく再強化学習を行なうことによって，パスプランニングのための戦略を獲得する手法を提案した．そして，計算機実験によってその動作を確認した．

参考文献

[OOMM91] Oommen,B.J. and Iyengar,S.S.: Trajectory Plannning of Robot Manipulator in Noisy Work Space Using Stochastic Automata, The International Journal of Robotics Research, vol.10, No.2, 1991, pp135-148.

[NARE89] Narendra,K.S. and Thathachar,M.A.L: Learning Automaton -An Introduction -, Prentice Hall, 1991.

[NARU92] 成瀬,嘉数:オートマタ表現によるタスク系列生成に関する研究 - 複数タスク集合のクラスタリング -,第44回情報処理学会全国大会講演論文集.