氏　名（本籍）	小　野　謙　二（大分県）
学位の種類	工　学　博　士
学位記番号	工　博　第２１９号
学位授与年月日	昭和４５年３月２５日
学位授与の要件	学位規則第５条第１項該当
研究科専門課程	東北大学大学院工学研究科 （博士課程）電気及通信工学専攻
学位論文題目	マルコフ決定過程に関する研究
論文審査委員	（主査） 教　授　本多　波雄　教　授　福島　弘毅 教　授　大泉　充郎　助教授　木村　正行

論　文　内　容　要　旨

1　ま　え　が　き

近年，自動制御理論の１つの分野である確率論的（ｓｔｏｃｈａｓｔｉｃ）制御理論は急速に発達し，多くの新しい制御理論や制御方式が研究されている。しかし，これらの理論を実用化の立場から眺めてみると，応用の実例は甚だしく少ない。その１つの理由として，制御対象における未知のパラメータ（例えば確率分布の平均値，分散など）の推定と確率論的制御理論が別々の立場で論じられていたことがあげられる。

本論文の目的は，未知パラメータの推定と確率論的制御理論を有機的に結合する新らしい制御方式を開発することである。すなわち，観測により得られたデータから未知パラメータの推定値を逐次求めていく段階においても，システムの評価関数（損失とする）が最小となるように逐次制御し

ていく方式を与えた。

考察の対象とするシステムとしては，離散的で有限な状態がマルコフ連鎖をなすシステムの制御，すなわちマルコフ決定過程を論じた。この問題は大別して，推移確率行列が既知である場合と未知である場合に分類される。また，状態を観測する立場から，真の状態を観測できる場合と雑音などが付加され真の状態が観測できない（状態観測が不完全な）場合に別けられる。本論文では，離散型有限マルコフ決定過程において，状態の推移確率行列は既知であるが，状態観測が不完全な場合，および状態の推移確率行列が未知の場合（真の状態が観測可能と不能なとき）について考察した。

2　推移確率行列が既知の決定過程

定常，非周期的，既約なマルコフ連鎖の推移確率行列を

$$P(k)=\| p_{ij}^{k} \| \tag{1}$$

$$\sum_{j=1}^{C} p_{ij}^{k}=1 \qquad \begin{matrix} i=1,2,\cdots,C \\ k=1,2,\cdots,A \end{matrix}$$

であらわす。ここに p_{ij}^{k} は，状態 x_i のとき制御行動 u_k をとることにより状態 x_j に推移する確率をあらわす。また，状態と観測値の対応をあらわす確率行列を

$$Q=\| q_{ij} \| \tag{2}$$

$$\sum_{j=1}^{D} q_{ij}=1 \qquad i=1,2,\cdots,C$$

とあらわす。ここに q_{ij} は，状態が x_i のとき観測値 y_j が得られる確率である。このように，推移確率行列が既知で，観測値が状態の確率的関数であらわされる場合，状態と観測値の対を1つの状態を考えて新しいマルコフ連鎖を構成する。この推移確率は

$$P'(k)=\| p_{(i,j)(i',j')}^{k} \| \tag{3}$$

$$p_{(i,j)(i',j')}^{k} = p_{ii'}^{k}\ q_{i'j'}$$

となる。以上の結果から，状態 (x_i, y_i) で制御行動 u_k を決定したとき状態 $(x_{i'}, y_{j'})$ への推移にともなう損失を $r_{(i,j)(i',j')}^{k}$ と定義する。この損失と新しく定義された推移確率行列 $P'(k)$ に，ダイナミック・プログラミング法を適用した解法を示した。

3　推移確率行列が未知の決定過程

真の状態は観測できるが，状態の推移確率行列が未知な場合として，次のような制御を考えることができる。すなわち，状態 x と観測値 y の対を事象 z と定義する。この事象の系列が r 重マルコ

フ連鎖をなし，rが既知であるとき確率的近似法を適用して直接にシステムの評価関数θの期待値を計算し逐次制御する。またrが未知の場合の制御方法，及び電子計算機によるシミュレーション結果を示した。

3.1　評価関数の期待値

システムの評価関数θの期待値の推定値は，確率的近似法を適用して

$$\hat{E}_{n(i,k)}[\theta \mid \eta_i, u_k] = \hat{E}_{n(i,b)-1}[\theta \mid \eta_i, u_k] + \gamma_{n(i,k)}\{\theta(t+1) - \hat{E}_{n(i,k)-1}[\theta \mid \eta_i, u_k]\}$$

$$\gamma_{n(i,k)} = 1 / n(i,k) \qquad (4)$$

ここでn(i,k)は，時刻t+1までに事象の系列η_iが観測され，かつ制御行動u_kを行なった回数である。また，

$$\theta(t+1) = g(\eta_i, u_k, y_j) \qquad (5)$$

である。ここでy_jは時刻t+1における観測値をあらわす。(4)式は，θの平均値と分散が有限であればn(i,k)→∞で期待値に確率収束することが証明される。

3.2　制御行動の決定

長さrなる任意の事象の系列η_iに対して，システムの評価関数θの期待値を最小とする制御行動uを求めたい。すなわち，

$$E[\theta(t+1) \mid \eta_i, u] = \min_{u'}\{E[\theta(t+1) \mid \eta_i, u']\} \qquad (6)$$

を満足する最適な制御uを求める。このような最適制御行動uを決定する関数は，任意の事象の系列η_iに対する条件付確率$P_r[u \mid \eta_i]$で与えられる。この確率の推定値は，θの期待値の推定値$\hat{E}_n[\theta \mid \eta, u]$を利用し，次式で逐次修正し推定できる。

$$\hat{P}_{n(i)}[u_k \mid \eta_i] = \alpha \hat{P}_{n(i)-1}[u_k \mid \eta_i] + (1-\alpha)\psi_{n(i)}(u_k, \eta_i) \qquad (7)$$

$$\psi_{n(i)}(u_k, \eta_i) = \begin{cases} 1 & \hat{E}_{n(i,k)}[\theta \mid \eta_i, u_k] = \min_{u_l}\{\hat{E}_{n(i,k)}[\theta \mid \eta_i, u_l]\} \\ 0 & \text{その他の場合} \end{cases}$$

ここでn(i)は，時刻t+1までに事象の系列η_iが観測された回数をあらわす。αは$0 < \alpha < 1$なる定数とする。上式はn(i)→∞で最適な制御行動を決定する確率$Pr[u_k \mid \eta_i]$に確率収束することが証明される。

3.3　制　御　手　順

事象の系列がｒ重マルコフ過程をなす場合の制御手順を示す。

(1) 初期セット

任意の制御u_kと事象の系列η_iに対して，初期値をそれぞれ

$$\hat{P}_o[u_k \mid \eta_i] = 1/A \tag{8}$$

$$\hat{E}_o[\theta \mid \eta_i, u_k] = 0 \tag{9}$$

とする。ここでAは制御行動の数。

(2) 制御の決定

事象の系列（長さｒ）をη_iとすると，時刻ｔ＋１における制御行動u_kを確率$\hat{P}_{n(i)}[u_k \mid \eta_i]$にしたがってランダムに決定する。

(3) 評価関数 θ （ｔ＋１）の計算

時刻ｔ＋１における制御行動u_kと観測値y_jと事象の系列η_iに対する評価関数を(5)式によって計算する。

(4) θの条件付期待値

θの条件付期待値の推定値を(4)式によって求める。

(5) 最適な制御行動の決定

事象の系列η_iについて，全ての制御行動に対するθの期待値の推定値を比較して$\psi_{n(i)}(u, \eta_i)$を決める。

(6) 決定関数の修正

全ての制御行動ｕに対して，行動の決定確率を(7)式により修正する。

(7) ステップを進める。

事象の系列を新しくし(2)へとぶ。

3.4 多重マルコフ過程の制御

これまで事象の系列がｒ重マルコフ連鎖をなすと仮定したが，このｒが未知である場合について考察した。この場合，事象の長さ０から始めて，マルコフ性の検定結果にもとづきｒを逐次大きくしていく方法を考えた。また検定として，確率密度をもっているということの他は何も規定しない連（ｒｕｎ）の数を用いるのが有効である。このようなｒが未知のシステムに対して，電子計算機を利用し簡単なシミュレーションを行なった。

4 マルコフ連鎖の関数

状態の集合を$\Omega_x = \{x_1, x_2, \cdots, x_C\}$，観測値の集合を$\Omega_y = \{y_1, y_2, \cdots, y_D\}$，ｆを$\Omega_x$から$\Omega_y$への関数とする。ここでC≧Dとする。また任意の観測値νに対応する$f^{-1}(\nu)$

で表わされる状態の集合を考え，その要素の数をN_ν，任意の長さの観測値の系列を$v_1, v_2, \cdots, v_n, w_1, w_2, \cdots, w_m$であらわす。観測値の系列$v_i \cdot \nu \cdot w_j$が得られる確率を$p(v_i \nu w_j)$であらわす。次に，

$$P_\nu(v_1, v_2, \cdots, v_n ; w_1, w_2, \cdots, w_m) = \prod_{\substack{i=1,2,\cdots,n \\ j=1,2,\cdots,m}} p(v_i \nu w_j) \tag{10}$$

が正則であるような最大の整数nをn_νとすると，次の関係がある。

(1) 任意の観測値νに対して$n_\nu \leqq N_\nu \leqq C$

(2) $C \geqq D$ならば$\Sigma_\nu n_\nu \leqq C$

(3) $\Sigma_\nu n_\nu = C \Leftrightarrow n_\nu = N_\nu$

確率的関数の場合，(10)式が正則であるような系列$v_1, v_2, \cdots, v_c, w_1, w_2, \cdots, w_c$の集合のなかで長さが最少な系列を$v_1^*, v_2^*, \cdots, v_c^*, w_1^*, w_2^*, \cdots, w_c^*$とすると

(4) $\Sigma_\nu n_\nu \leqq C \times D$

(5) $\max_i \{ lg(v_i^*) \} \leqq C$

$\max_j \{ lg(w_j^*) \} \leqq C$

ここで$lg(v)$は，系列vの長さを示す。

観測値の系列から，状態の推移確率行列を同定することは一般に不可能である。また，状態の関数としてあらわされる観測値の系列の分布から，次の観測値を推定することも一般には不可能である。

5 むすび

マルコフ決定過程において，推移確率行列の既知と未知，状態観測が完全と不完全な場合を論じた。本論文の決定過程は，その構造の一般性により，最適制御のみならず，ORなど応用分野も広いものである。

審 査 結 果 の 要 旨

マルコフ決定過程の理論は，確率的制御のみならず，広く情報科学，計画理論，オペレーションズ・リサーチなどの分野においても数学的基礎として欠くことのできないものである。しかし，実用化の立場から眺めてみると，この理論を応用した例は甚だしく少ない。その主な理由として，制御の対象となる系の未知パラメータの確率分布や推移確率行列などの推定と制御理論を関連づける研究がなかったことがあげられる。

本論文は，系の状態の推移がマルコフ連鎖で記述できる場合の制御，すなわち，マルコフ決定過程について，未知パラメータの推定と制御理論とを有機的に結合した新しい制御方式について，著者が研究した成果をまとめたもので全編5章よりなる。

第1章は緒論であって，本研究の目的とともに，マルコフ決定過程についての従来の研究と本研究との関連が述べられている。

第2章では，推移確率行列が既知で観測値が状態の確率的関数として表わされる場合について，状態とその観測値を対とする新しいマルコフ過程を定義することにより，ダイナミック・プログラミングの手法を用いて最適決定を行なう方法を与えている。

第3章では，推移確率行列は未知であるが状態の観測が完全な場合のマルコフ過程の制御について新しい制御方式を与えている。この制御方式は，観測可能な量と制御可能な量の対で表わされる事象がn重マルコフ過程をなす一般的な場合を扱っており，確率的近似法により直接には推移確率行列を推定することなく評価関数を逐次最小にする制御を行なうもので，推定と制御とを有機的に結合した優れた制御方式であり，本研究の重要な成果である。また，本章では〃連〃によるマルコフ連鎖の次数nを検定する方法を与え，これを上記の制御方式に結合して，その実用性を一層高めている。なお，シミュレーションにより本方式の有効性が確かめられている。

第4章では，推移確率行列が未知で状態観測が不完全な系の制御を行なう手がかりとして，マルコフ連鎖の関数について2，3の有用な性質を導いている。

第5章は結論である。

以上要するに，本論文は確率的近似法と〃連〃によるマルコフ性の検定法を導入して，統計的推定の理論と制御理論とを有機的に結合することにより，高度な制御機能をもつ実用性の高い制御方式を開発したもので，その成果はシステム工学および制御工学の発展に寄与するところが少なくない。

よって，本論文は工学博士の学位論文として合格と認める。