数理解析研究所講究録 1373 巻 2004 年 238-246

動的計画問題に対する列挙法について

九州工業大学・工学部　藤田 敏治 (Toshiharu Fujita)
Faculty of Engineering,
Kyushu Institute of Technology

1 はじめに

動的計画法の理論は R.Bellman により創出され ([1]), 非常に幅広い分野において研究・応用がなされている. それは, **最適性の原理**をその基本原理とし, 様々な問題に対する再帰的アプローチを与える理論的枠組みである. 我々も, これまで多くの問題クラスに対して動的計画論を用いて再帰的解法を与えてきた ([4, 6, 7]).

その一連の研究において我々は, 再帰的解法と列挙法 — あるいは言い換えれば, 逐次最適化と同時最適化 — の 2 つの立場からの解法の検証が重要であるとの認識を持つにいたった. そして各種問題に対する, 動的計画法のアプローチは,

1. 同時最適化問題としての厳密な定式化 (目的関数, 推移システム, 実行可能解の明示)

2. 再帰的解法の導出 (同時最適化との同値性の検証)

という流れで進むべきであると考えている.

このことは, 構造が比較的簡単な, そして性質が特に注意を要するものではないような問題を扱う上では, さほど意識せずともよいことのように感じられる. しかし, より複雑な問題を扱う場合, あるいは微妙に性質の異なる要素を含む問題を扱う場合には重要であろう.

ここでは, このような立場で, 動的計画問題をモデル化し, まずは, 同時最適化による解法について考える. なお, ここでの結果は, 現在構築中である動的計画法の計算機用クラスライブラリにも実装し, 主として研究や教育・学習での利用を考えている. このことについても 4 節で簡単に触れる.

2 動的計画問題

2.1 概要と定義

ここでは, 目的関数, 推移システム, 実行可能解の明示を意識した, 一般的な動的計画問題の定式化を与える. まず, 構成要素については

- 状態空間 (離散, 有限)
- 決定空間 (離散, 有限)
- 状態推移システム (確定, 確率など)
- 評価

そして, 解の表現としての

- 政策 (決定関数列)

である．なお，評価に関しては3段階で考える．利得関数による各期の評価，評価関数による履歴の評価，そして最終的な目的関数である．

このとき，問題は「目的関数を最大化（最小化）する政策すべてを求める」と表現される．ただし，

・初期状態は与えられる

・ある期の状態は，そこで取られた決定と推移法則に従って次の期の状態へ移行

・各期において状態と決定に応じた利得が得られる

・評価関数により各期の利得を合成して履歴を評価

・推移法則に応じて履歴の評価関数を政策毎に評価

とする．

次に，以後用いる記号等を定義する．

$N \geq 2$ ： 段数（期数）

$X = \{s_1, s_2, \ldots, s_p\}$ ： 有限状態空間

$U = \{a_1, a_2, \ldots, a_k\}$ ： 有限決定空間

$r_n : X \times U \to \mathbf{R}$ ： 第 n 利得関数 $n = 1, 2, \ldots, N$

$r_G : X \to \mathbf{R}$ ： 終端利得関数

必要に応じて以下を採用してもよい

$X_n(x, u) \subset X$ ： 第 $n-1$ 期の状態 x と決定 u に対し，第 n 期に生じ得る状態の集合

$U_n(x) \subset U$ ： 第 n 期において，状態 x に対し取り得る決定の集合

（より一般には $U_n(x_1, u_1, x_2, u_2, \ldots, x_n)$）

2.2 推移システム

推移システムは，一般に実数値関数 $\phi_n(y|x, u), (y, x, u) \in X \times X \times U$ により次のように表現される：

$$y \sim \phi_n(\cdot|x, u)$$

これは，第 n 期において状態 x に対し決定 u をとった際，第 $n+1$ 期の状態 y への推移が値 $\phi_n(y|x, u)$ により特徴付けられることをあらわす．代表的推移システムとしては，確定的推移システム，確率的推移システム，ファジィ推移システム，そして非決定性推移システムなどがあげられる．

たとえば，確率的推移システムの場合，関数：

$$p_n : X \times X \times U \to [0, 1]$$

$$\left(\sum_{y \in X} p_n(y|x, u) = 1 \ , \ {}^{\forall}(x, u) \in X \times U \right)$$

を考える．これは，第 n 期における状態と決定 (x, u) に対し確率 $p_n(y|x, u)$ で次の状態 y へ推移することを表す．このような確率的推移を $y \sim p_n(\cdot \,|\, x, u)$ と表す．

なお，確定的推移システムについては

$$\tilde{f}_n : X \times X \times U \to \{0, 1\}$$

$$\left(\sum_{y \in X} \tilde{f}_n(y|x,u) = 1 \ , \ {}^{\forall}(x,u) \in X \times U \right)$$

と表されるが，各 (x,u) に対して $\tilde{f}(y|x,u) = 1$ なる y が一意に定まるので，通常，より簡潔に

$$y = f(x,u)$$

と表現される．また，ファジィ推移システムは

$$\mu_n : X \times X \times U \to [0,1]$$

非決定性推移システムは

$$q_n : X \times X \times U \to \{0, 1\}$$

で与えられる．

なお，重み関数を導入した非決定性動的計画問題 ([3], [5]) の推移システムは，上記の 4 推移を包含するものと解釈され，ここでの一般推移システムと同値なものと考えられる．この種の問題に関しては，現在十分な議論を経ていないので，ここでは省略する．

2.3 履歴の評価

システムは，各期における状態と決定（最終期は状態のみ）に応じて

$$r_1(x_1,u_1), r_2(x_2,u_2), \ldots, r_N(x_N,u_N), r_G(x_{N+1})$$

で評価される．そして，履歴 $(x_1, u_1, x_2, u_2, \ldots, x_N, u_N, x_{N+1})$ に対しては，

$$r_1(x_1,u_1) \circ r_2(x_2,u_2) \circ \cdots \circ r_N(x_N,u_N) \circ r_G(x_{N+1})$$

により評価される．なお，$\circ$ は結合演算子（結合律を満たす演算子）であり，この評価を**結合型評価**と呼ぶ．たとえば，$\circ = +$ のときは一般的に用いられる**加法型評価関数**をあらわし，また，$\circ = \wedge$ のときはファジィ環境下で用いられる**最小型評価関数**をあらわす（ただし $\wedge$ は $a \wedge b := \min(a,b)$ で定義される最小演算子）．

2.4 目的関数

確定システム上では，評価関数そのものが目的関数となる．一方，確率システム上においては，評価関数がいわゆる確率変数となるため，通常その期待値を目的関数と考える．

$$E[r_1(x_1,u_1) \circ r_2(x_2,u_2) \circ \cdots \circ r_G(x_{N+1})]$$

$$= \sum_{(x_2,\ldots,x_{N+1}) \in X \times \cdots \times X} \cdots \sum \{[r_1(x_1,u_1) \circ r_2(x_2,u_2) \circ \cdots \circ r_G(x_{N+1})] \times p_1(x_2|x_1,u_1) \cdots p_N(x_{N+1}|x_N,u_N)\}$$

さらには，事前条件付期待値，事後条件付期待値も考えられる．

また，ファジィシステムに対する目的関数は結合型評価の Minimax 期待値：

$$F[r_1(x_1,u_1)\circ r_2(x_2,u_2)\circ\cdots\circ r_G(x_{N+1})]$$

$$=\bigvee_{(x_2,\ldots,x_{N+1})\in X\times\cdots\times X}\cdots\bigvee\{[r_1(x_1,u_1)\circ r_2(x_2,u_2)\circ\cdots\circ r_G(x_{N+1})]\wedge[\mu_1(x_2|x_1,u_1)\wedge\cdots\wedge\mu_N(x_{N+1}|x_N,u_N)]\}$$

で与えられ，非決定性システムに対しては

$$r_1(x_1,u_1)\circ\bigcirc_{x_2}[\{r_2(x_2,u_2)\circ\bigcirc_{x_3}[\{r_3(x_3,u_3)\circ\cdots\circ\bigcirc_{x_{N+1}}[r_G(x_{N+1})\times q_N(x_{N+1}|x_N,u_N)]$$
$$\times q_{N-1}(x_N|x_{N-1},u_{N-1})\times\cdots\}q_2(x_3|x_2,u_2)]\}\times q_1(x_2|x_1,u_1)]$$

を考える．

より一般には，結合型評価の関数（これを g とおく）を評価関数と考え，さらに各期に割引率（これを β とおく）等も考慮した形が考えられる．したがって，期待値の一般的表現として次の 3 つの型を考える：

一般期待値

$$G(r_1,r_2,\cdots,r_N,r_G,\phi_1,\phi_2,\cdots,\phi_N,\circ,\bullet,\oplus,\otimes,\beta,g)$$

$$=\bigoplus_{x_2,\ldots,x_{N+1}}\{g(r_1(x_1,u_1)\circ\beta r_2(x_2,u_2)\circ\cdots\circ\beta^{N-1}r_N(x_N,u_N)\circ\beta^N r_G(x_{N+1}))$$
$$\otimes(\phi_1(x_2|x_1,u_1)\bullet(\phi_2(x_3|x_2,u_2)\bullet\cdots\bullet\phi_N(x_{N+1}|x_N,u_N))\}$$

一般事前条件付期待値

$$G^{\mathrm{pr}}(r_1,r_2,\cdots,r_N,r_G,\phi_1,\phi_2,\cdots,\phi_N,\circ,\oplus,\otimes,\beta)$$

$$=\bigoplus_{x_2}[\{r_1(x_1,u_1)\circ\bigoplus_{x_3}[\{\beta r_2(x_2,u_2)\circ\bigoplus_{x_4}\beta^2 r_3(x_3,u_3)\circ\cdots$$
$$\bigoplus_{x_{N+1}}[\{\beta^{N-1}r_N(x_N,u_N)\circ\beta^N r_{N+1}(x_{N+1})\}\otimes\phi_N(x_{N+1}|x_N,u_N)]\otimes\cdots$$
$$\}\otimes\phi_3(x_4|x_3,u_3)]\}\otimes\phi_2(x_3|x_2,u_2)]\}\otimes\phi_1(x_2|x_1,u_1)]$$

一般事後条件付期待値

$$G^{\mathrm{po}}(r_1,r_2,\cdots,r_N,r_G,\phi_1,\phi_2,\cdots,\phi_N,\circ,\oplus,\otimes,\beta)$$

$$=r_1(x_1,u_1)\circ\bigoplus_{x_2}[\{\beta r_2(x_2,u_2)\circ\bigoplus_{x_3}[\{\beta^2 r_3(x_3,u_3)\circ\cdots$$
$$\circ\bigoplus_{x_{N+1}}[\beta^N r_{N+1}(x_{N+1})\otimes\phi_N(x_{N+1}|x_N,u_N)]\otimes\phi_{N-1}(x_N|x_{N-1},u_{N-1})\otimes\cdots$$
$$\}\otimes\phi_2(x_3|x_2,u_2)]\}\otimes\phi_1(x_2|x_1,u_1)]$$

例 2.1 (通常の期待値)

$$G(r_1, r_2, \cdots, r_N, r_G, p_1, p_2, \cdots, p_N, +, \times, \sum, \times, 1, 1_{\mathbf{R}})$$

$$= \sum_{x2,x3,\ldots,x_{N+1}} \{(r_1(x_1,u_1) + r_2(x_2,u_2) + \cdots + r_{N+1}(x_{N+1},u_{N+1}))$$
$$\times (p_1(x_2|x_1,u_1) \times p_2(x_3|x_2,u_2) \times \cdots \times p_N(x_{N+1}|x_N,u_N))\}$$

ただし，$1_{\mathbf{R}}$ は恒等関数を表すものとする． □

例 2.2 (ファジィ環境下 [2])

$$G(r_1, r_2, \cdots, r_N, r_G, p_1, p_2, \cdots, p_N, \wedge, \times, \sum, \times, 1, 1_{\mathbf{R}})$$

$$= \sum_{x2,x3,\ldots,x_{N+1}} \{(r_1(x_1,u_1) \wedge r_2(x_2,u_2) \wedge \cdots \wedge r_{N+1}(x_{N+1},u_{N+1}))$$
$$\times (p_1(x_2|x_1,u_1) \times p_2(x_3|x_2,u_2) \times \cdots \times p_N(x_{N+1}|x_N,u_N))\}$$ □

例 2.3 (非決定性システム上での加法型評価 [3])

$$G^{\mathrm{po}}(r_1, r_2, \cdots, r_N, r_G, \phi_1, \phi_2, \cdots, \phi_N, +, \sum, \times, 1)$$

$$= r_1(x_1,u_1) + \sum_{x_2}[\{r_2(x_2,u_2) + \sum_{x_3}[\{r_3(x_3,u_3) + \cdots$$
$$+ \sum_{x_{N+1}} [r_{N+1}(x_{N+1}) \times \phi_N(x_{N+1}|x_N,u_N)] \times \phi_{N-1}(x_N|x_{N-1},u_{N-1}) \times \cdots$$
$$\} \times \phi_2(x_3|x_2,u_2)]\} \times \phi_1(x_2|x_1,u_1)]$$ □

2.5 政策クラス

各期において，とるべき決定を与える関数は決定関数と呼ばれる。そして、その決定関数から構成される列が政策である．決定がどのような情報に依存して定まるかにより，以下にあげる 3 種類の政策が定義される．([7], [8])

マルコフ政策

現時刻の状態のみに依存し決定を定める決定関数の列として定義され，$\pi = \{\pi_1, \pi_2, \ldots, \pi_N\}$:

$$\pi_n : X \to U$$

と表される．

一般政策

現時刻までの状態すべてに依存し決定を定める決定関数の列として定義され，$\sigma = \{\sigma_1, \sigma_2, \ldots, \sigma_N\}$:

$$\sigma_n : X^n \to U$$

と表される．以後、一般政策全体を Σ であらわす．

原始政策

部分履歴 (その時刻までの状態と決定の交互列) に依存し決定を定める決定関数からなる列として定義され，$\gamma = \{\gamma_1, \gamma_2, \ldots, \gamma_N\}$：

$$\gamma_n : (X \times U)^{n-1} \times X \to U$$

と表される．

2.6 問題の基本形

以上の議論により，動的計画問題の基本形は次のように与えられる：

一般期待値最適化問題

$$\begin{aligned} \text{Opt}\ \ & G^{\sigma}_{x_1}(r_1, \ldots, r_N, r_G, \phi_1, \ldots, \phi_N, \circ, \bullet, \oplus, \otimes, \beta, g) \\ \text{s.t.}\ \ & \text{(i)}\ x_{n+1} \sim \phi_n(\cdot|x_n, u_n)\ ,\quad n = 1, 2, \ldots, N \\ & \text{(ii)}\ \sigma = \{\sigma_1, \sigma_2, \ldots, \sigma_N\} \in \Sigma \end{aligned}$$

ただし，$G^{\sigma}_{x_1}$ は、初期状態 x_1 および政策 σ に依存して定まる一般期待値を表す。また Opt は最適化子であり，最大化 (Maximize) あるいは最小化 (minimize) のいずれかとする．なお、一般事前条件付期待値最適化問題および一般事後条件付期待値最適化問題に関しても同様に定義される。

例 2.4 (確率最大化問題)

推移システムとして確率システムを考え (すなわち $\phi_n := p_n$ とする)，

$$\circ = +,\ \bullet = \times,\ \oplus = +,\ \otimes = \times,\ \beta = 1,\ g(x) = C_{[a,b]}(x)\ \text{(特性関数)}$$

とおく．この場合，次の確率最大化問題を表す：

$$\begin{aligned} \text{Max}\ \ & P^{\sigma}_{x_1}(a \le r_1(x_1, u_1) + \cdots + r_G(x_{N+1}) \le b) \\ \text{s.t.}\ \ & \text{(i)}\ x_{n+1} \sim p_n(\cdot|x_n, u_n)\ ,\quad n = 1, 2, \ldots, N \\ & \text{(ii)}\ \sigma = \{\sigma_1, \sigma_2, \ldots, \sigma_N\} \in \Sigma \end{aligned}$$

□

2.7 終了集合

問題によっては，終了時刻 N があらかじめ定まらない場合がある．この場合，**終了集合** $T \subset X$ が与えられているものとし，システムは $x_n \in T$ を満たした時点で終了するものとする．なお，確定システム以外では，1つの実行可能解 (政策) に対し，一般に複数の終了時刻が履歴ごとに存在する．

2.8 基本形に関する補足

問題の基本形を与える際，いたずらに複雑化することを避けるため，表現しうる問題クラスを制限している．実際には，より広い問題クラスを想定すべきである．詳細は省略したが，非決定性システムを扱うための一般事前/事後条件付期待値への対応，各期で複数の利得関数をもつ複合型評価問題への対応，多目的最適化問題への対応，そして計算やパズルへ応用可能な非最適化問題への対応，などである．このことにより，モデル自体は非常に広範な問題を表現しうるものが実現できる．

3 列挙法

前節の問題に対し，同時最適化を実現するための列挙法の構成について考える．その前提として初期状態 x_1 は与えらたものとし，以後の議論では固定して考える．

3.1 決定の列挙

動的計画問題を解くに際して必要とされる結論は，各期においていかなる決定を取ればよいか，である．この観点からは，決定空間の直積（あるいはその部分集合）を実行可能解と解釈し，すべての

$$(u_1, u_2, \ldots, u_N) \in \underbrace{U \times U \times \cdots \times U}_{N\text{個}}$$

を列挙し，各々に対し目的関数を計算し比較すればよいと考えられる．実際，確定システム上の問題に対してはこれで十分であることが示される．すなわち，計算機への実装に当たっては，決定空間に関する N 重のループを構成すればよいことがわかる．

3.2 政策の列挙

しかしながら，一般の推移，たとえば確率システムを考えた場合，もはやこの方法は適用できない．第2期以降の状態は確率変数となるため，それに依存する決定もまた確率変数となるためである．この場合，各期において実現可能性のある各状態に対して，それぞれ別個に決定を考える必要がある．したがって，たとえば2期間問題では，$U \times U$ ではなく，一般に直積空間：

$$U \times \underbrace{U \times U \times \cdots \times U}_{p\text{個}} = U^{1+p}$$

を考えなければならない．(p は状態数)

3期間問題ではどうか．それはマルコフ性をもつか否かに依存する．各期の決定にマルコフ性を仮定できるならば，U^{1+p+p} を考えることとなり，そうでなければ，すなわち3期目の決定が2期目の状態にも依存するならば U^{1+p+p^2} となるのである．こういった状況を表現するには政策の概念を用いたほうがわかりやすい．(それゆえ前節では政策クラスに関する最大化という形で定式化した.) 与えられた問題の最適政策がマルコフ政策クラスの中に存在することが証明されている場合にはマルコフ政策全体を列挙し，証明できていない場合には一般政策全体を列挙し，各々に対し目的関数を計算し比較すればよいのである．すなわち，計算機への実装の観点からは，まず，各期におけるマルコフ決定関数あるいは一般決定関数をすべて列挙し，それらに対し，N 期にわたる全ての組み合わせをつくればよい．

3.3 推移ツリーの列挙

政策の列挙による列挙法は，理論上十分と考えられるが，終了時刻未定の問題を考えた場合，実運用上では困難がある．決定関数を列挙する段階で，何期まで列挙すればよいかが定まらないからである．その上限値を最小限にうまく定めることができればよいが，一般には必要十分な上限値を採用することになると思われる．そうした場合，ただでさえ計算量の大きい列挙法は，ちょっとした問題を解く場合でさえも破綻してしまう可能性がある．このことは，実際に不要な決定関数値も含めてすべて列挙せねばならないことに起因する．

そこで，状態と決定の推移をツリー状に表現した**推移ツリー**を導入し，この推移ツリーをもって実行可能解とみなすこととする．推移ツリーとは，

(i) 初期状態とその初期状態に対する決定のペアをその根としてもつ，

(ii) 根を含む各頂点の下には頂点の状態と決定から実現可能な状態およびそれに対応する決定のペア（一般に複数）をもつ，

(ii) 末端の頂点（終端状態を表す）は状態のみからなる

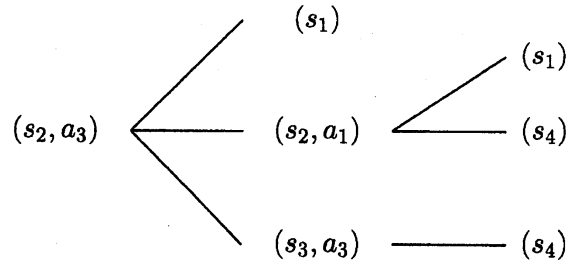
ものとして定義される．

例 3.1

$X=\{s_1,s_2,s_3,s_4\}$, $U=\{a_1,a_2,a_3\}$, $T=\{s_1,\ s_4\}$ とおき，

$$\phi_1(s_1|s_2,a_3),\ \phi_1(s_2|s_2,a_3),\ \phi_1(s_3|s_2,a_3),\ \phi_2(s_1|s_2,a_1),\ \phi_2(s_4|s_2,a_1),\ \phi_2(s_4|s_3,a_3)>0$$

である（上記以外の ϕ_1,ϕ_2 の値は 0 とする）とき，次の推移ツリーが存在する



□

推移システム ϕ_n および，必要ならば決定に関する制約を用いて，実現可能性のある推移ツリーをすべて列挙する．そして，各推移ツリーに対し目的関数を計算し比較すればよい．

一般に推移ツリーと一般政策とは 1 対 1 に対応しないが，次の事実がある．

・任意の一般政策に対し，それにより表現される推移ツリーが一意に存在する．

・任意の推移ツリーに対し，それを表現する一般政策が存在する．

すなわち，一般政策全体にそれが表現する推移ツリーを対応させる写像は全射となる．ゆえに，推移ツリーの列挙による解法が成り立つ．

4 DPフレームワークについて

現在，動的計画法を利用するための計算機用クラスライブラリを構築中で，これを「DPフレームワーク」と呼んでいる．これは，2節で定義された問題を扱うことができるように（実際にはより広い問題クラスに対応，2.8節参照）設計されている．単一の抽象問題を基底クラスとしてもち，その構成要素を具体化した問題を派生させる形で幾層にも階層化が行われている．その結果，ある問題クラスに対し実装された解法は，そこから派生した問題クラスを全て解き得る．（ここでの「解ける」の意味は「理論的に解ける」である．）そして，列挙法は最上位の抽象問題に対し実装され，すなわちあらゆる問題を解き得る．現状，列挙法については若干未実装部分は残るものの，再帰的解法の確立されていない問題クラスに対して，同時最適化と逐次最適化の比較を行う研究補助ツールとしての役割が果たせる段階に近づきつつある．

5 まとめ

ここでは，動的計画論の展開にあたって，同時最適化と逐次最適化の比較の重要性を鑑み，その同時最適化のための列挙法について考えた．

列挙法は，当然ではあるが，完全にすべての実行可能解を網羅していることが必要である．また，その方法により導き出された最適解集合が，与えられた問題の最適解すべてに完全に一致していなければならない．そして，これが重要であるが，それらのことが誰の目からも自明でなければならない．それでいて，はじめて，逐次最適化の理論的正当性を示す土台となりうるのである．

References

[1] R.E. Bellman, *Dynamic Programming*, NJ: Princeton Univ. Press, 1957.

[2] R.E. Bellman and L.A. Zadeh, Decision-making in a fuzzy environment, Management Science, **17**(1970), B141-B164.

[3] 藤田敏治, 分割問題について ～ 動的計画からのアプローチ ～, 第５０回シンポジュウム『ORと数学』, 日本OR学会, 九州大学国際交流プラザ, 平成15年9月9日, pp.1–14.

[4] T. Fujita and K. Tsurusaki, Stochastic optimization of multiplicative functions with negative value, J. Oper. Res. Soc. Japan, **41**(1998), 351-373.

[5] S. Iwamoto and T. Ueno, On Non-deterministic Dynamic Programming, 研究集会「不確実性下での数理決定問題とその関連分野」配布資料, 千葉大学, 平成15年10月17,18日.

[6] S. Iwamoto and T. Fujita, Stochastic decision-making in a fuzzy environment, J. Oper. Res. Soc. Japan, **38**(1995), 467-482.

[7] S. Iwamoto, K. Tsurusaki and T. Fujita, On Markov Policies for Minimax Decision Processes, J. Math. Anal. Appl., **253**(2001), 58-78.

[8] S. Iwamoto, T. Ueno and T. Fujita, Controlled Markov Chains with Utility Functions, Ed. H. Zhenting, J. A. Filar and A. Chen, Markov Processes and Controlled Markov Chains, Chap.8, pp. 135-148, Kluwer, 2002