1997年度日本オペレーションズ・リサーチ学会
春季研究発表会

確率的決定過程上での乗法型関数の最適化

02501655　九州大学　藤田 敏治　FUJITA toshiharu
02302176　九州大学　*津留崎 和義　TSURUSAKI kazuyoshi

1　はじめに

これまでの確率的決定過程の研究においては, 評価系として加法型関数を用いたものがほとんどであった。またその議論において, 最適政策のマルコフ性が暗に認められていた。しかし加法型評価系以外（乗法型, 最小型 ...）の問題では, 最適政策が必ずしもマルコフ政策の中に存在するとは限らない。このことは, **マルコフ過程に適用されてきた通常の動的計画がそのままでは通用しない**ことを物語っている。そこで本論文では, 特に乗法型評価系を採り上げ, 最適政策がマルコフでない場合にも対応しうる解法を与える。

2　乗法型過程

ここではシステム全体の評価として, 各段で得られる利得の積の期待値を考え, これを最大化し, そのときの最適解（最適値・最適政策）を求めていく。具体的には次の問題を考える：

$$\begin{array}{ll}\text{Maximize} & E[r_0(x_0,u_0)r_1(x_1,u_1)\cdots \\ & \quad \cdots r_{N-1}(x_{N-1},u_{N-1})r_G(x_N)] \\ \text{subject to} & \text{(i)}\ \ x_{n+1}\sim p(\cdot|x_n,u_n) \\ & \text{(ii)}\ \ u_n\in U \quad 0\leqq n\leqq N-1.\end{array} \tag{1}$$

ただし, 問題 (1) における期待値は, 初期状態 x_0, マルコフ推移確率 $p(y|x,u)$, および決定列 $\{u_0,\ldots,u_{N-1}\}$ を定める政策に依存した次のような N重和である：

$$\begin{aligned}&E[r_0(x_0,u_0)\cdots r_{N-1}(x_{N-1},u_{N-1})r_G(x_N)] \\ &=\sum_{\substack{(x_1,\ldots,x_N)\\ \in X\times\cdots\times X}}\cdots\sum\{[r_0(x_0,u_0)\cdots r_{N-1}(x_{N-1},u_{N-1})r_G(x_N)] \\ &\qquad \times p(x_1|x_0,u_0)\cdots p(x_N|x_{N-1},u_{N-1})\}.\end{aligned}$$

まず, **利得が非負の場合**を考える。このとき, 一般政策とマルコフ政策による最適値は等しい。すなわち, **最適政策はマルコフ政策の中に存在する**。

次に, **利得が少なくとも1つは負値を含む場合**を考える。このときには, 一般政策とマルコフ政策による最適値は一般に異なる。すなわち, **最適政策がマルコフ政策の中に存在するとは限らない**のである。したがって, 一般政策のクラスで最適解を求めなければならない。

3　両決定過程

両決定過程 (Bidecision Process) においては, 最適化問題を考える際に最大化と最小化の対を用いる。

原問題 (1) に対して, 途中の n 段から始まり最終 N 段で終る一般部分問題群の目的関数を

$$\begin{aligned}&I^n(x_n;\sigma) \\ &=\sum_{\substack{(x_{n+1},\ldots,x_N)\\ \in X\times\cdots\times X}}\cdots\sum\{[r_n(x_n,u_n)\cdots r_{N-1}(x_{N-1},u_{N-1})r_G(x_N)] \\ &\qquad \times p(x_{n+1}|x_n,u_n)\cdots p(x_N|x_{N-1},u_{N-1})\}\end{aligned}$$

とおくと, この期待値を一般政策に関して最大化, および最小化する最大化部分問題群, 最小化部分問題群は次のように定義できる：

$$\begin{aligned}V^N(x_N) &= r_G(x_N) \quad x_N\in X \\ V^n(x_n) &= \max_{\sigma} I^n(x_n;\sigma) \\ &\quad x_n\in X \quad 0\leqq n\leqq N-1,\end{aligned} \tag{2}$$

$$\begin{aligned}W^N(x_N) &= r_G(x_N) \quad x_N\in X \\ W^n(x_n) &= \min_{\sigma} I^n(x_n;\sigma) \\ &\quad x_n\in X \quad 0\leqq n\leqq N-1.\end{aligned} \tag{3}$$

最大化部分問題群 (2), および最小化部分問題群 (3) の最適値関数 V^n, W^n に関し, 次の再帰式が成り立つ。

定理 3.1

$$\begin{aligned}V^N(x) &= r_G(x) \quad x\in X \\ V^n(x) &= \max_{u\in U(n,x,-)}\left[r_n(x,u)\sum_{y\in X}W^{n+1}(y)p(y|x,u)\right] \\ &\quad \vee \max_{u\in U(n,x,+)}\left[r_n(x,u)\sum_{y\in X}V^{n+1}(y)p(y|x,u)\right] \\ W^N(x) &= r_G(x) \quad x\in X \\ W^n(x) &= \min_{u\in U(n,x,-)}\left[r_n(x,u)\sum_{y\in X}V^{n+1}(y)p(y|x,u)\right]\end{aligned}$$

$$\wedge \min_{u\in U(n,x,+)}\left[r_n(x,u)\sum_{y\in X}W^{n+1}(y)p(y|x,u)\right]$$
$$x\in X \quad 0\leqq n\leqq N-1.$$

ただし, $U(n,x,-)=\{u\in U\,|\,r_n(x,u)<0\}, U(n,x,+)=\{u\in U\,|\,r_n(x,u)\geqq 0\}$ である。

この定理から $V^0(x)$ を求めることで最適値は定まるが, 最適一般政策は最大化部分問題群 $\{V^n(x)\}$, 最小化部分問題群 $\{W^n(x)\}$ から得られるそれぞれの最適政策（実はこれらはマルコフ）から構成されることに注意する。

4 不変埋没原理

不変埋没原理では, まず解くべき与問題をそれ自身を含む適度な大きさの問題群に埋め込み, 再帰式を導いてそれを解く。そこで得られた最適解をもとに与問題の最適解を求めるのである。ここで, 問題 (1) を $-1\leqq r_n(x,u), r_G(x)\leqq 1$ 内で考えて, 実パラメータ $\lambda\in[-1,1]$ で埋め込む：

$$\begin{aligned}\text{Maximize}\quad & E[\lambda r_0(x_0,u_0)r_1(x_1,u_1)\cdots \\ & \cdots r_{N-1}(x_{N-1},u_{N-1})r_G(x_N)] \\ \text{subject to}\quad & \text{(i)}\ \ x_{n+1}\sim p(\cdot\,|x_n,u_n) \\ & \text{(ii)}\ \ u_n\in U \qquad 0\leqq n\leqq N-1.\end{aligned} \tag{4}$$

ただし, この埋め込まれた問題 (4) に $\lambda=1$ を代入した問題は原問題 (1) になっていることに注意する。

一般問題 (4) に対して, x_n,λ が初期状態として与えられた一般部分問題群の目的関数を

$$\begin{aligned}&J^n(x_n,\lambda;\sigma)\\ &=\sum_{(x_{n+1},\cdots,x_N)\in X\times\cdots\times X}\cdots\sum\{[\lambda r_n(x_n,u_n)\cdots r_G(x_N)]\\ &\quad\times p(x_{n+1}|x_n,u_n)\cdots p(x_N|x_{N-1},u_{N-1})\}\end{aligned}$$

とおくと, 一般部分問題群

$$\begin{aligned}V^N(x_N,\lambda)&=\lambda r_G(x_N)\quad x_N\in X\quad \lambda\in[-1,1]\\ V^n(x_n,\lambda)&=\operatorname*{Max}_{\sigma}J^n(x_n,\lambda;\sigma)\\ &\qquad x_n\in X\quad \lambda\in[-1,1]\quad 0\leqq n\leqq N-1\end{aligned} \tag{5}$$

に対し, 次の再帰式が成り立つ。

定理 4.1

$$\begin{aligned}V^N(x,\lambda)&=\lambda r_G(x)\qquad x\in X\quad \lambda\in[-1,1]\\ V^n(x,\lambda)&=\operatorname*{Max}_{u\in U}\sum_{y\in X}V^{n+1}(y,\lambda r_n(x,u))p(y|x,u)\\ &\qquad x\in X\quad \lambda\in[-1,1]\quad 0\leqq n\leqq 1.\end{aligned}$$

また, マルコフ政策 π による期待値 $J^n(x_n,\lambda;\pi)$ を用いてマルコフ部分問題群

$$\begin{aligned}v^N(x_N,\lambda)&=\lambda r_G(x_N)\qquad x_N\in X\quad \lambda\in[-1,1]\\ v^n(x_n,\lambda)&=\operatorname*{Max}_{\pi}J^n(x_n,\lambda;\pi)\\ &\qquad x_n\in X\quad \lambda\in[-1,1]\quad 0\leqq n\leqq N-1\end{aligned} \tag{6}$$

を考えても, やはり次の再帰式が成立する。

定理 4.2

$$\begin{aligned}v^N(x,\lambda)&=\lambda r_G(x)\qquad x\in X\quad \lambda\in[-1,1]\\ v^n(x,\lambda)&=\operatorname*{Max}_{u\in U}\sum_{y\in X}v^{n+1}(y,\lambda r_n(x,u))p(y|x,u)\\ &\qquad x\in X\quad \lambda\in[-1,1]\quad 0\leqq n\leqq 1.\end{aligned}$$

定理 4.1, 4.2 を見ると, 全く同じ形の再帰式を満たすことが分かる。このとき, 次の定理が成り立つ。

定理 4.3 (i) マルコフ政策が一般問題の最適値関数 $V^0(\cdot)$ に到達する。すなわち, 一般問題 (5) に最適マルコフ政策 π^* が存在する：

$$V^0(x_0,\lambda)=J^0(x_0,\lambda;\pi^*)\qquad \text{for } \forall(x_0,\lambda)\in X\times[-1,1].$$

(ii) マルコフ部分問題群 (6) の最適値関数は一般部分問題群 (5) の最適値関数に等しい：

$$v^n(x,\lambda)=V^n(x,\lambda)\quad (x,\lambda)\in X\times[-1,1]\quad 0\leqq n\leqq N.$$

(iii) 埋め込まれた問題 (4) の最適マルコフ政策を用いて, 原問題 (1) の最適一般政策が構成される。

5 例題

具体的な数値例を挙げて, 両決定過程, 不変埋没原理, そして全行動を列挙する多段確率決定ツリーの 3 通りで最適解を求める。また, これらが一致し, 最適政策がマルコフ政策でないことも確認する。

参考文献

[1] S. Iwamoto, On bidecision processes, *J. Math. Anal. Appl.* **187**, 676-699, 1994.