数理解析研究所講究録
第460巻 1982年 17-26

あるsequential assignment problemについて

大阪府立大学総合科学部　中井 達

1. Introduction

Sをstate space，Pをstochastic transition functionとするstationary Markov chainを(S_m, P)とするとき，このMarkov chainの上でのsequential stochastic assignment problem $\{A_n, S, P, r, \{F_s\}_{s\in S}, \beta\}$を考える。但し$A_n$はaction spaceであり，二つのdisjoint set $\{a_1, \ldots, a_n\}$と$\{\phi\}$の和としてあらわせる。また全てのstate $s\in S$に対してnon-negative random variable X_sの値x_sを観測することが出来る。この確率変数は独立で，分布関数$F_s(x)$は既知とする。$r(a, x)$は，観測値xに対しaction aを用いたときのrewardとし$E r(a, X_s) < \infty$であると仮定する。βはdiscount factorであり$0\le\beta\le 1$とする。

現在まで考えられているsequential stochastic assignment problemと同様に，次の様なsequential assignment problemを考える。([1][3][4][5][7])。まず，position (A_n, s)を現在のstateがsであり，そのときのaction spaceがA_nである状態とする。positionが(A_n, s)であるとき，確率変数X_sの

実現値x_sを観測して action a を A_n の中から選ぶ。もしaが$\{a_1, \ldots, a_n\}$のうちのa_iであれば、immediate reward $r(a_i, x_s)$を得る。このとき次の stage での position は(A_n^i, t)である。$A_n^i = \{a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n\}$であり$t$は Markov chain (S, m, P)の next stage である。またaが action "p" に等しければ、immediate reward は0であり、次の stage での position は(A_n, t)である。即ち action "p" は pass することと同じでありその回数には制限は無い。以上の様な場合の、total expected discounted reward を最大とする問題を考える。ここで action "a_i"を取ることは謂ゆる sequential assignment problem での action a_i を観測値xに assign することに等しい。

2. Problem formulation.

Sequential stochastic assignment problem $\{A_n, S, P, r, \{F_s\}_{s \in S}, \beta\}$において、現在の position が$(A_n, s)$であるとき、optimal strategy の下での value を$v(A_n, s)$、また、観測値xを知ったという条件付の、この problem の value を$v(A_n, s \mid x)$とすれば、よく知られた dynamic programming formulation より次の recursive equation が得られる。

$$v(A_n, s) = E_s\, v(A_n, s \mid X_s)$$

(1)

$$v(A_n, s \mid x) = \max\Big\{\max_{1 \le i \le n}\Big\{r(a_i, x) + \beta\int v(A_n^i, t)P(s, dt)\Big\},$$

$$\beta \int v(A_n, t)\, P(s, dt) \}.$$

但し ここにおいて $A_n = \{a_1, \ldots, a_n\} \cup \{p\}$, $A_n^i = \{a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n\} \cup \{p\}$ とする。(1)式に関して よく知られた Markov decision process と同様に optimal strategy の存在とその value の存在が示される。ここでは以下 optimal strategy と value が求められる様に reward function に条件を付ける。即ち $r(a, x) = r_1(a) r_2(x)$ とあらわせる場合を考える。また state が s であるときの確率変数を X_s の代りに $r_2(X_s)$ を考えることにより $r(a, x) = r(a) x$ と考える。また A_n に対し $r(a_1) \geq \cdots \geq r(a_n)$ を仮定する。このとき optimal strategy 及びその value について次の事が成り立つ。

Proposition 1. 全ての $s \in S$ に対して次の様な function の列 $\{g_i(s)\}_{i=1,\ldots}$, $\{h_i(s)\}_{i=1,2,\ldots}$ が存在する。即ち

$$g_1(s) \geq g_2(s) \geq g_3(s) \geq \cdots \geq g_n(s) \geq \cdots \geq 0$$

$$h_1(s) \geq h_2(s) \geq h_3(s) \geq \cdots \geq h_n(s) \geq \cdots \geq 0$$

であり, position (A_n, s) において次の事が成立する。

1) 観測値の値が x であるときの optimal strategy は

$g_i(s) \leq x < g_{i-1}(s)$ のとき i-th action a_i を選び $(i = 1, \ldots, n)$

$0 \leq x < g_n(s)$ のとき action "p" を選ぶ事である。

2) position (A_n, s) のときの value は

$v(A_n, s) = \sum_{i=1}^{n} r(a_i) h_i(s)$ である。

3) $\{g_i(s)\}_{i=1,2,\ldots}$ 及び $\{h_i(s)\}_{i=1,2,\ldots}$ は次の recursive equation によって求めることが出来る。

$$g_i(s) = \beta \int h_i(t)\, P(s, dt),$$

$$h_i(s) = \int_{g_{i-1}(s)}^{\infty} g_{i-1}(s)\, dF_0(x) + \int_{g_i(s)}^{g_{i-1}(s)} x\, dF_0(x) + \int_0^{g_i(s)} g_i(s)\, dF_0(x),$$

かつ $h_1(s) = \int_{g_1(s)}^{\infty} x\, dF_0(x) + \int_0^{g_1(s)} g_1(s)\, dF_0(x)$, $g_1(s) = \beta \int h_1(t) P(s, dt)$

$g_0(s) = \infty$.

さて、次に $\mathcal{T}_s$ を $P(T_s \leq t) = P(s, t)$ であるような確率変数とする。今 s' が s へ収束するとき、$X_{s'}$ 及び $T_{s'}$ がそれぞれ X_s 及び T_s へ確率収束すると仮定する。この仮定の下で次の事が成り立つ。

Corollary 1　$g_i(s)$ ($s \in S$, $i=1,2,\ldots$) は S に関して連続である。

以下では上の事を仮定して話を進める。

3　Special case ($S \subset R^n \times R^1$)

この節で、stochastic transition function P が、現在の state s のみならず、観測値 x の値に depend する場合を考える。即ち前節の definition に従がえば、$S \subset R^n \times R^1$ で、$S \ni (s, x)$ と考える。また、$dF_{(s,x)}(y) = I_x$ (I_x は indicator function) かつ、$P((s,x),(t,y)) = P(s, x, t)\, F_t(y)$ である場合に相当する。この問題では、Proposition 1 における関数 $g_i(s, x)$ が、よ

り簡単な式で求めることが出来る。但し、ここでは次の様な仮定を設ける。即ち $P(s, x, t)$ は x に関して stochastically increasing かつ、$\partial^2 P(s, x, t)/\partial x^2 \geqq 0$、更に、$F_s(x)$ は s に関して stochastically increasing であるとする。

今、$u(A_n, s) = E\, v(A_n, (s, X_s))$ かつ $u(A_n, s \mid x) = v(A_n, (s, x))$ と定義すれば、これらの値は、optimal strategy の下での value 及び、条件付の value であり、(1)式を用いて次の recursive equation を得ることが出来る。

$$u(A_n, s) = \int_0^\infty u(A_n, s \mid x)\, dF_s(x)$$

(2)
$$u(A_n, s \mid x) = \max\left\{ \max_{1 \le i \le n} \left\{ r(a_i)x + \beta \int u(A_n^i, t)\, dP(s, dt) \right\},\ \beta \int u(A_n, t)\, dP(s, dt) \right\},$$

但し $A_n = \{a_1, \cdots, a_n\} \cup \{p\}$ かつ、$A_n^i = \{a_1, \cdots, a_{i-1}, a_{i+1}, \cdots, a_n\} \cup \{p\}$。

このとき次の事が成り立つ。

Proposition 2. 全ての $s \in S$ に対して次の様な関数列 $\{t_i(s)\}_{i=1,2,\cdots}$ が存在する。

$$t_1(s) \geqq t_2(s) \geqq \cdots \geqq t_n(s) \geqq \cdots \geqq 0.$$

このとき、position (A_n, s) における optimal strategy 及び value は次の様にあらわせる。

1) 観測値が x のとき optimal strategy は

$t_i(s) \le x < t_{i-1}(s)$ $(i = 1, \cdots, n)$ のとき i-th action a_i を選び

$0 \le x < t_n(s)$ のとき action "p" を選ぶ事である。

2) value $u(A_n, s)$ は

$u(A_n, s) = \sum_{i=1}^{n} r(a_i)\, t_i(s)$ である。

3) $\{t_i(s)\}_{i=1,2,\ldots}$ は次の recursive equation を満足する。

$$t_n(s) = \int_{d_{n+1,s}}^{\infty} t_{n+1}(s, x)\, dF_0(x) + \int_{d_{n,s}}^{d_{n+1,s}} x\, dF_0(x) + \int_{0}^{d_{n,s}} t_n(s, x)\, dF_0(x)$$

$$t_n(s, x) = \beta \int t_n(t)\, P(s\, x, dt),$$

但し $t_0(s) = \infty$, $-\infty \cdot 0 = 0 \cdot \infty = 0$ かつ $d_{n,s}$ は $t_n(s, x) = x$ の unique root とする。

ここで、前節の Proposition 1 で述べられた $\{g_i(s, x)\}_{i=1,2,\ldots}$ 及び $\{h_i(s, x)\}_{i=1,2,\ldots}$ と Proposition 2 における $\{t_i(s)\}_{i=1,2,\ldots}$ 及び $\{t_i(s, x)\}_{i=1,2,\ldots}$ との関係は、次の様にあらわす事が出来る。

$$t_n(s) = \int h_n(s, x)\, dF_0(x)$$

$$t_n(s, x) = \beta \int t_n(t)\, P(s\, x, dt) = \beta \iint h_n(t, y)\, dF_t(y)\, P(s, x, dt)$$

$$= g_n(s, x)$$

上の Proposition 2 に関連して次の二つの Corollary を得ることが出来る。

Corollary 2. $t_i(s)$ は s に関して増加関数である。

Corollary 3. $t_i(s, x)$ は x に関して増加関数でありまた上に凸な関数である。

4. Variants and example

4.1 2部で考えた問題を N-stage problem として考えれば、Proposition 1 と同様の結果が得られ、recursive equation により逐次計算して optimal strategy 及び value を得ることが出来る。

Corollary 4. 全ての state $s \in S$ に対し、関数列 $\{g_{i,N}(s)\}_{i=1,\cdots,N}$ 及び $\{h_{i,N}(s)\}_{i=1,\cdots,N}$ で次の様なものが存在する。

$$g_{1,N}(s) \geq g_{2,N}(s) \geq \cdots \geq g_{N-1,N}(s) \geq g_{N,N}(s) = 0$$

$$h_{1,N}(s) \geq h_{2,N}(s) \geq \cdots \geq h_{N-1,N}(s) \geq h_{N,N}(s) = 0$$

position が ${}_N(A_n, s)$ であるとき次の事が成り立つ。(但し ${}_N(A_n, s)$ は残り stage の数が N で、position が (A_n, s) である時を言う。またそのときの value を $v_N(A_n, s)$ であらわす。)

1) 観測値の値が x であるときの optimal strategy は

$g_{i,N}(s) \leq x < g_{i-1,N}(s)$ のとき i-th action a_i を選び $(i=1, \cdots, n)$

$0 \leq x < g_{n,N}(s)$ のとき action "p" をとることである。

2) value $v_N(A_n, s)$ は

$v_N(A_n, s) = \sum_{i=1}^{n} r(a_i) h_{i,N}(s)$ である。

3) $\{g_{i,N}(s)\}_{i=1,\cdots,N}$ 及び $\{h_{i,N}(s)\}_{i=1,\cdots,N}$ は次の recursive equation を満足する。

$$g_{i,N}(s) = \beta \int h_{i,N}(t) P(s, dt)$$

$$h_{i,N}(s) = \int_{g_{i-1,N-1}(s)}^{\infty} g_{i-1,N-1}(s)\, dF_s(x) + \int_{g_{i,N-1}(s)}^{g_{i-1,N-1}(s)} x\, dF_s(x) + \int_{0}^{g_{i,N-1}(s)} g_{i,N-1}(s)\, dF_s(x)$$

但し　$h_{1,N}(\omega)=\int_{g_{1,N-1}(\omega)}^{\infty} x\,dF_0(x)+\int_0^{g_{1,N-1}(\omega)} g_{1,N-1}(\omega)\,dF_0(x)$

$g_{1,N}(\omega)=\beta\int h_{1,N}(t)\,P(\omega,dt)$. である。

4.2 Wを space Ωの値をとる parameter, Xを実数値確率変数とし $f(x|W)$ を $W=\omega$ であるときの X の conditional g.p.d.f. とする。$g(\omega)$を Wの prior g.p.d.f. とし, $g(\omega)$と、観測値 $x\in R^1$ に対して $g(\omega|x)$を Wの posterior g.p.d.f とする。

現在まで m個の値 $X_1=x_1,\cdots,X_m=x_m$ を observe したときの posterior distribution を $G_m(\omega|x_1,\cdots,x_m)$として次の仮定を設ける。(但し $G_0(\omega)=G(\omega)$とする。) 1) $\mu_\omega=E[X|\omega]<\infty\ (\omega\in\Omega)$ かつ $\mu=E[\mu_\omega]<\infty$. 2) 与えられた $x_1,\cdots,x_m$ に対して ωに対する充分統計量 $\bar{x}_m$ が存在し 3) $\bar{x}_m$は $\bar{x}_{m-1}$及び x_m により recursively に生成される。 4) $E[W|\bar{x}_m]$ は $\bar{x}_m$の non-decreasing function であり 5) $P(X_m\geq x\,|\,\bar{x}_{m-1})$は $\bar{x}_{m-1}$に関して stochastically increasing とする。

このとき3節と同様に、次の様な sequential stochastic assignment problem を考える。$A=\{a_1,\cdots,a_n\}\cup\{\phi\}$, $S=N\times R^1\ni(m,\bar{x}_m)$, $r(a,x)=r(a)\cdot x\ (r(a_1)\geq\cdots\geq r(a_n))$, $F_{(m,\bar{x}_m)}=F(x|\bar{x}_m)$ かつ $P((m,\bar{x}_m),x,(m',y))=I_{(m+1,\bar{x}_{m+1}(\bar{x}_m,x))}$, 但し $\bar{x}_{m+1}(\bar{x}_m,x)$ は、$\bar{x}_m$及び、$(m+1)$ stage における観測値 x により生成される充分統計量とする。このとき Proposition 2

により optimal strategy 及び value が得られる。この結果は Nakai [3] における解と一致する。

4.3 Random termination である場合の sequential stochastic assignment problem を考える。この場合の problem は $\{A_n, S, P, r, \{F_\sigma\}_{\sigma \in S}, \beta\}$ であらわされ、特に $S = S_1 \cup \{s^*\}$, $P(T = s^* | s^*) = 1$, かつ $P(X_{s^*} = 0) = 1$, 但し T は $P(T \leq t) = P(s^* \, t)$ である確率変数である。また s^* は absorbing state であり、s^* での観測値の値は常に 0 をとると考える。この場合、absorbing state s^* に入れば、reward は常に 0 であり、そのときこの問題は終わると考えられる。

4.4 この sequential stochastic assignment problem を game 的に扱う事も考えられる。即ち、二人の player が、逐次出現する観測値から m 個を選び、それぞれに、action space A_n の m 個の action を取らせる場合を考える。但し player I は、観測値の選択と同時に、どの action をとるかという選択を合わせて行ない、player II は、観測値の選択のみを行なう。但しこのとき、両 player 共に、ここで考えた problem と異なり、pass する回数は制限されているものとする。また player I は maximizing player として player II は minimizing player として行動をとるものとする。このゲームについては Nakai [5] に詳しく述べられていて同様にして optimal strategy

及び value を得ることが出来る。

References

[1]. C. Derman, G.J. Lieberman and S.M. Ross "A sequential stochastic assignment problem" Management Science vol 18, p349 −355, 1972.

[2] G.H. Hardy, J.E. Littlewood and G. Polya "Inequalities" Cambridge University Press, 1934.

[3] T. Nakai "Optimal assignment for a random sequence with an unknown parameter" Journal of Information & Optimization Science, vol 1, p214 ∼ 228, 1980.

[4] T. Nakai "Sequential stochastic assignment problem with rejection" Journal of Information & Optimization Science, vol 2 p169-180, 1981.

[5] T. Nakai "A time sequential game related to the sequential stochastic assignment problem", Journal of Operations Research Society of Japan, vol 25. No. 2. 1982

[6] M. Sakaguchi "A sequential assignment problem for randomly arriving jobs", Rep. Stat. Appl Res. JUSE, vol 19, p99 ∼ 109, 1972.