数理解析研究所講究録 1457 巻 2005 年 62-68

確率過程ゲームの計算機ネットワーク制御への応用

早稲田大学・商学部　理工学部　数理科学研究所
毛利裕昭(Hiroaki Mohri)
School of Commerce, Waseda Institute of Mathematics
Waseda University
mohri@waseda.jp

1．はじめに

確率過程ゲーム（Stochastic Game） とは，非協力ゲームにおける繰り返しゲームを精緻化したものであり．また，マルコフ決定過程の研究と深い関わりを持っている．

ゲーム理論の歴史においては Shapley(1953)の論文に発端をもつが，今までそれほど手がつけられていないといえよう．代表的なサーベイ論文は Mertens(2002), Vieille(2002), Raghavan and Filar(1991)である．まとまった専門書としては，Filar and Vrieze (1996)がある．この書物のイントロダクションには，確率過程ゲームとマルコフ決定過程の研究が独立に発達したこと，その両者に接点をもたせこの分野の発展の為に執筆された趣旨が書かれている．

確率過程ゲームの応用には様々なものが考えられるが，ここではシンプルな計算機ネットワークの制御をとりあげる．言い換えると並列待ち行列の振り分け問題で，スイッチングの問題である．関連文献として Hassin and Haviv (2003)がある．ここでは，Altman(1996)に沿った考え方でモデルを記述して定常状態および有限ステージでの最適方策（ゲームの戦略）を求める．

2. モデルの記述

2. 1 連続時間モデル

ルータが，k 個の並列待ち行列に対して客を割り当てる状況を考える．

- 非協力ゲームとしてのプレイヤーは，ルータと k 個 queue の集合体のサーバの 2 プレイヤーとする．（queue の集合体をまとめてプレイヤーと考えていることに注意する）
- ルータには客がパラメータ λ でポアソン到着する．
- 各 queue のバッファは無限大であると仮定する．
- 各 queue の処理方式は FCFS であるとする．
- サーバのサービス率は各 queue i に対して指数サービスを行なうものとし，そのパラメータは $a_1(i) \in [\underline{\mu}_i, \overline{\mu}_i]$ とする（下付き添え字の 1 はプレイヤー 1 を示している），これがプレイヤー 1 のアクションである．
- ルータつまりプレイヤー 2 がどの queue に客を割り当てるかというアクションを a_2 で表現して，そのとる値は割り当てられる queue の番号であるとする．

ルータとサーバの間には非協力ゲーム状況があり，各 queue のサービス率がルータには厳密には分からないとする．ここでは以下の情報が分かっていると仮定する．

– 各時点の queue の長さ
– ルータがどういった振り分け履歴があるかは分かっている

この問題をマルコフ決定過程として考えたい時に

- **状態：各 queue の長さ、直近のイベントの同定**

が必要になる．ここで注意したいのは，ルータが「客が到着時に意思決定する」のに対して，サーバは「いつでも意思決定する」ことである．この状況を単純化するため，ルータに関して「すべてのイベントで意思決定」可能と考えるとモデルを単純化できる．つまり，

ルータが時間 t でアクション i を取る
↕ 読替え
時間 t 以降直近のイベントが到着なら客を queue i に振り分ける

・**状態**：**各** queue **の長さ（だけでよい）**

ということになる．この状況で具体的に定式化を試みる．

$A_i(s)$: 状態sでqueue iへの客が到着した直後の状態
$D_i(s)$: 状態sでqueue iへの客が出発した直後の状態
微小時間間隔：Δ　,$\pi\{p\}$: pが真なら１、偽なら０を返す
状態の変化（各$queue$の長さの変化）$s=(s_1,s_2,\ldots,s_k)\to s'$
遷移確率は以下のようになる

$$q_\Delta(s' \mid s,a_1,a_2)=\begin{cases} \lambda'\Delta+o(\Delta) & (a_2=i, s'=A_i(s), i=1,2,\ldots,k) \\ a_1(i)\Delta+o(\Delta) & (s'=D_i(s), s_i\neq 0, i=1,2,\ldots,k) \\ 1-(\lambda'+\sum_{i=1}^{k}a_1(i)\pi\{s_i>0\})+o(\Delta) & (s'=s) \\ o(\Delta) & (\text{otherwise}) \end{cases}$$

２．２　離散時間モデル

　このような連続時間のままでは問題を取り扱いにくいので**モデルの離散化**を以下のようにして行なう．

$$\lambda=\lambda'\Delta,\ \underline{\mu_i}=\underline{\mu_i'}\Delta,\ \overline{\mu_i}=\overline{\mu_i'}\Delta$$

すると離散化のモデルは以下のようになる．

状態空間：$S=\mathbf{N}^k, s\in S$
プレイヤー：$N=\{\text{Player1}:\text{サーバ}, \text{Player2}:\text{ルータ}$
　注意：サーバにqueueは２つだが１人とみなす

アクション：

Player 1 $A_1=\prod_{i=1}^{k}\left[\underline{\mu_i},\overline{\mu_i}\right]$
Player 2 : $A_2=\{1,2,\ldots,k\}$

離散化後の遷移確率は以下のようになる

$$
q(s^{'} \mid s, a_1, a_2) = \begin{cases} \lambda^{'} & (a_2 = i, s^{'} = A_i(s), i = 1,2,\ldots,k) \\ a_1(i) & (s^{'} = D_i(s), s_i \neq 0, i = 1,2,\ldots k) \\ 1 - (\lambda^{'} + \sum_{i=1}^{k} a_1(i)\pi\{s_i > 0\}) & (s^{'} = s) \end{cases}
$$

3．AT-AR 性を利用した最適政策（戦略）

3．1　AT(Additive Transition Property)性

名前が示すように，推移確率がゲームの各プレイヤーの部分に分解されることを示す．つまり，言い換えると各プレイヤー部分の和を考えると全体の推移確率となっていることをいう

$$
q_1(s^{'} \mid s, a_1, a_2) = \begin{cases} a_1(i) & (s^{'} = D_i(s), s_i \neq 0, i = 1,2,\ldots,k) \\ 1 - (\lambda^{'} + \sum_{i=1}^{k} a_1(i)\pi\{s_i > 0\}) & (s^{'} = s) \end{cases}
$$

$$
q_2(s^{'} \mid s, a_1, a_2) = \lambda \qquad (a_2 = i, s^{'} = A_i(s), i = 1,2,\ldots,k)
$$

$q_1(s^{'} \mid s, a_1, a_2)$、$q_2(s^{'} \mid s, a_1, a_2)$を上記の様におくと

$$
q_1(s^{'} \mid s, a_1, a_2) = q_1(s^{'} \mid s, a_1, a_2) + q_2(s^{'} \mid s, a_1, a_2)
$$

が成立し、これをAT性を満たすという。

3．2　AR(Additive Reward Property)性

AT 性が推移確率に関するものであったが，一方，AR 性は利得関数に関する同様な性質で利得関数がプレイヤーごとに分解されることを言う．ここでは，２つのプレイヤーは対立的でゼロサムゲームと考えることにする．

状態 s でアクション(a_1,a_2)が起きた時、ルータが支払う（サーバが得る）利得は以下のようになる.

$$r(s,a_1,a_2)=h(s)+\sum_{i=1}^{k}e_i(a_1)+\sum_{i=1}^{k}d_i\pi\{a_2=i\}$$

ここで
$h(\cdot)$: holding cost
$e_i(\cdot)$: queue i のサービスコスト
d_i : ルータのqueue iへの振り分けコスト
とする

３．３　AR-AT 性を利用した最適戦略

両方の関数（推移率、利得関数）とも

- サーバ部分（Player 1) + ルータ部分(Player 2 ）
  となっている
- 両方の性質を満たすことを AR-AT 性 と呼ぶ

ここで，有限計画期間、無限計画期間利得（コスト）と最適政策（戦略）について考える.

ステージnにおける期待利得 : $r_n(f_1,f_2)(s)$、固定割引率 : $0<\beta<1$

有限計画期間利得 : $\phi_\beta^n(f_1,f_2)=\sum_{k=1}^{n}\beta^{k-1}\cdot r_k(f_1,f_2)(s)$

有限計画期間利得 : $\phi_\beta(f_1,f_2)=\sum_{k=1}^{\infty}\beta^{k-1}\cdot r_k(f_1,f_2)(s)$

これらの割引ゲームの値は

$$v_\beta^n(s)=\sup_{f_1}\inf_{f_2}\phi_\beta^n(f_1,f_2)(s)=\inf_{f_2}\sup_{f_1}\phi_\beta^n(f_1,f_2)(s)$$

$$v_\beta(s)=\sup_{f_1}\inf_{f_2}\phi_\beta(f_1,f_2)(s)=\inf_{f_2}\sup_{f_1}\phi_\beta(f_1,f_2)(s)$$

となる．各 Player の政策(戦略) f_i が与えられた時（初期状態 s）最適政策（戦略）の存在は，Raghavan,Tijs and Vrieze(1985)では，AR-AT 性を満たすなら以下の存在性を証明し，それをもとめるアルゴリズムを示している．

- 有限計画期間の場合、Markov 最適戦略が存在
- 無限計画期間の場合、定常戦略の存在

上記の結果を本モデルに適用できることはあきらかである．

3. まとめと今後の課題

本論では，並列待ち行列の客の割り当て制御に関して確率過程ゲームの枠組みで AR-AT 性を利用して最適戦略が存在しそれを求められることを示した．より，ネットワークの構造を拡張することが第一の今後の課題である．到着ポアソン，指数サービスといった単純なモデルで並列待ち行列が多段に拡張することが第一に挑戦すべきテーマである．

謝辞

本研究は早稲田大学特定課題研究助成費 2004A-122 の成果の一部である．

[参考文献]

[1]E. Altman. "A Markov game approach for optimal routing into a queuing network", 1994, Technical Report 2178, INRIA Sophia Antipolis.

[2]J. Filar and K. Vrieze, "Competitive Markov Decision Processes",1996, Springer.

[3]R. Hassin and M. Haviv, "To Queue or Not To Queue", 2004, Kluwer Academic Press.

[4]J.F. Mertens, "Stochastic Games", Handbook of Game Theory with Economic Application Vol.3, 2002, Chapter 47, pp.1809-1832.

[5]T.E.S. Raghavan and J.A. Filar, "Algorithms for Stochastic Games: -A Survey", ZOR(Mathematical Methods of OR) , 1991, pp.437-472

[6]L.S. Shapley, "Stochastic Games", Proceedings Nat. Acad. Of Science USA, 1953, 39, pp.1095-1100.

[7]S.H. Tijs, T.E.S. Raghavan and O.J. Vrieze, "On Stochastic Games with Additive Reward and Transition Structure", Journal of Optimization Theory and Applications, 1985, 47, pp.451-464.

[8]N. Vieille, "Stochastic Games: Recent Results", Handbook of Game Theory with Economic Application Vol.3, 2002, Chapter 48, pp.1833-1850.