A-002

パス上の確率ゲームに対する効率的なアルゴリズム

Efficient Algorithms for Stochastic Games on Paths

清藤駿成† 塩浦昭義‡ 徳山豪†

Takanari Seito Akiyoshi Shioura Takeshi Tokuyama

1. はじめに

確率ゲーム(stochastic game)は，有向グラフの各頂点に付随した利得行列に従ったゲームを行い，グラフ上を確率的に遷移する 2 人ゼロ和ゲームである．ゲームの状態は頂点に置かれる 1 つのトークンで表される．各ステップでは，トークンの置かれた頂点に付随する利得行列に従って二人ゼロ和ゲームが実行され，2 人のプレイヤー1，2 にはそれぞれの行動に応じて利得が与えられる．ゲームの終了後，トークンはある確率分布にしたがって，現在の頂点から隣接する頂点に移動するが，この確率分布は両プレイヤーの行動に依存して定まる．ゲームは無限回のステップに渡って繰り返され，プレイヤー1 は自身の割引総利得の最大化，プレイヤー2 はプレイヤー1 の割引総利得の最小化を目的とする．確率ゲームは Shapley [1] によって提案されたゲームであり，均衡解の存在が示されている．これ以降，確率ゲームの様々な拡張や変種が考えられ，それぞれのゲームにおける均衡解の存在性の証明や，均衡解を求めるためのアルゴリズムおよび計算困難性に関する研究が行なわれてきた．

確率ゲームの特殊ケースの一つとして，トークンの移動が 1 人のプレイヤーの行動のみに依存する，単独支配者確率ゲーム(single controller stochastic game)が知られている．一般の確率ゲームに対しては，均衡解を求める効率的なアルゴリズムが知られていないが，単独支配者確率ゲームは線形計画問題として定式化できることが知られており [2]，したがって内点法などの多項式時間アルゴリズムを用いて効率的に均衡解を求めることができる．また，単独支配者確率ゲームは巡回検査官問題などの実問題をモデル化することができることから [3]，この特殊なゲームに対してさらに効率的なアルゴリズムを構築することは有用である．

本研究では，単独支配者確率ゲームに対するさらに効率的なアルゴリズムの構築を目指し，その手がかりとして与えられるグラフ構造がパスの場合について考える．また，各頂点での行列ゲームおける行動数は 2 つに限定する．本研究では，このような仮定の下で，均衡解がある種の線形方程式系の解として与えられ，その解が漸化式を解くことで得られることを示す．その結果，$O(n)$時間で均衡解を求められることを証明する．

2. パス上の確率ゲーム

n個の頂点$S=\{1,\cdots,n\}$，および枝集合

$\{(s,s-1)|s=2,3,\ldots,n\}$

$\cup\{(s,s+1)|s=1,2,\ldots,n-1\}\cup\{(1,1),(n,n)\}$

からなる有向グラフを考える（図 1 参照）．各頂点$s\in S$に大きさ2×2の利得行列

$$M_s=\begin{pmatrix} m_s^1 & m_s^2 \\ m_s^3 & m_s^4 \end{pmatrix}$$

が与えられているとする．各頂点において各プレイヤーが選ぶことのできる行動を $1,2$ の 2 つとし，それぞれ利得行列M_sの行番号および列番号に対応する．頂点sでプレイヤー$1, 2$ が行動 1 を選ぶ確率をそれぞれ$x_s, y_s\in[0,1]$とする．したがって，各プレイヤーが行動 2 を取る確率はそれぞれ$1-x_s, 1-y_s$となる．トークンの移動はプレイヤー2 の行動のみに依存すると仮定する．プレイヤー2 が行動 1 を取った場合は現在の頂点sから 1 つ値の小さい頂点$s-1$に，行動 2 を取った場合は現在の頂点よりも 1 つ値の大きい頂点$s+1$に移動する．ただし，頂点$1,n$においてそれぞれ行動 1，2 を取った場合は，その頂点にとどまるとする．

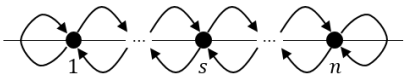

図 1 ゲームで扱うグラフ

各ステップでは，両プレイヤーの行動の組に基づいて利得が配分される．ステップtにおいてプレイヤー1 の得る利得をm_tとおく．すると，所与の割引率$\beta\in[0,1]$に対し，プレイヤー1 の割引総利得は$\sum_{t=0}^{\infty}\beta^t m_t$と与えられる．プレイヤー$1, 2$ はそれぞれ，この値を最大化および最小化するように確率x_s, y_sを決定する．

本研究では，各頂点における利得行列M_sとして，任意の実行列で与えられる場合と，その特殊ケースとして，M_sの各成分m_s^i $(i\in\{1,\cdots,4\})$が頂点番号sをパラメータとする線形関数$f_i(s)$により表現される場合を扱う．

3. アルゴリズムの方針

一般の確率ゲームの均衡解は，動的計画法により計算できることが知られている（[4]などを参照）．最初のステップにおけるゲームの後に，その後のゲームにおいてお互いに最適な戦略がわかっている，つまり均衡解が既知であると仮定する．この仮定は，最初のステップで両プレイヤーが頂点sで純行動の組(z,w) $(z,w\in\{1,2\})$をとった場合，そのときの利得$m_{z,w}$に加えて，移動後の頂点s'での均衡における割引利得$\beta v_{s'}$も得られることと同値である．したがって，その純行動の組を用いた際の割引総利得は$m_{z,w}+\beta v_{s'}$となる．各頂点における利得行列の各成分に対して，同様の操作を行なったものを補助ゲームとよび，その利得行列M_s'は以下の式で与えられる．

$$M_s'=\begin{pmatrix} m_s^1+\beta v_{s-1} & m_s^2+\beta v_{s+1} \\ m_s^3+\beta v_{s-1} & m_s^4+\beta v_{s+1} \end{pmatrix}$$

元のゲームの均衡解は各補助ゲームの均衡解であり，逆に補助ゲームの均衡解は元のゲームの均衡解であることが知られている [1].

本研究で扱うパス上の確率ゲームの場合には，頂点sに

† 東北大学大学院情報科学研究科, Graduate School of Information Sciences, Tohoku University

‡ 東京工業大学工学院経営工学系, Department of Industrial Engineering and Economics, School of Engineering, Tokyo Institute of Technoloy

における補助ゲームの均衡解(x_s, y_s)を，隣接する頂点における補助ゲームの期待利得v_{s-1}, v_{s+1}を用いて表すことができる．それにより得られた均衡解を用いて，期待利得v_sを計算すると，$v_1, \cdots, v_n$を変数とする線形方程式系が得られる．したがって，線形方程式系を解くことで均衡解が容易に得られる．詳細については次節で説明する．

4. パス上の確率ゲームに対するアルゴリズム

利得行列が

$$M = \begin{pmatrix} m^1 & m^2 \\ m^3 & m^4 \end{pmatrix}$$

の場合の行列ゲームの均衡解(x, y)は，$m^1 - m^2 - m^3 + m^4 \neq 0$ならば次の式により与えられる：

$$x = \frac{-m^3 + m^4}{m^1 - m^2 - m^3 + m^4}, \quad y = \frac{-m^2 + m^4}{m^1 - m^2 - m^3 + m^4}.$$

ここで，x, yはプレイヤー1, 2 がそれぞれ行動 1 を選ぶ確率を表す．また，均衡解におけるプレイヤー1 の期待利得vは，次の式により与えられる：

$$v = m^1 xy + m^2 x(1-y) + m^3(1-x)y + m^4(1-x)(1-y).$$

これらの事実を用いて，パス上の確率ゲームの均衡解を計算する．

まず，各補助ゲームの均衡解(x_s, y_s)は，隣接する頂点における補助ゲームの期待利得v_{s-1}, v_{s+1}を用いて次の式で与えられる：

$$x_1 = \frac{q_1 + \beta(-v_1 + v_2)}{p_1}, \quad 1 - x_1 = \frac{r_1 + \beta(v_1 - v_2)}{p_1},$$
$$y_1 = \frac{-m_1^2 + m_1^4}{p_1}, \quad 1 - y_1 = \frac{m_1^1 - m_1^3}{p_1},$$
$$x_s = \frac{q_s + \beta(-v_{s-1} + v_{s+1})}{p_s}, \quad 1 - x_s = \frac{r_1 + \beta(v_{s-1} - v_{s+1})}{p_s},$$
$$y_s = \frac{-m_s^2 + m_s^4}{p_s}, \quad 1 - y_s = \frac{m_s^1 - m_s^3}{p_s},$$
$$x_n = \frac{q_n + \beta(-v_{n-1} + v_n)}{p_n}, \quad 1 - x_n = \frac{r_n + \beta(v_{n-1} - v_n)}{p_n},$$
$$y_n = \frac{-m_n^2 + m_n^4}{p_n}, \quad 1 - y_n = \frac{m_n^1 - m_n^3}{p_n}.$$

ここで，p_s, q_s, r_sは次の式で与えられる：

$$p_s = m_s^1 - m_s^2 - m_s^3 + m_s^4,$$
$$q_s = -m_s^3 + m_s^4,$$
$$r_s = m_s^1 - m_s^2.$$

次に，各補助ゲームの均衡解(x_s, y_s)を用いて，期待利得v_sを求め，整理すると，次の式を得る：

$$(1 + A_1)v_1 - A_1 v_2 = B_1,$$
$$A_s v_{s-1} + v_s - A_s v_{s+1} = B_s,$$
$$A_n v_{n-1} + (1 - A_n)v_n = B_n.$$

ここで，A_sおよびB_sは次の式で与えられる：

$$A_s = \frac{\beta}{p_s}\left(m_s^1 y_s + m_s^2(1 - y_s) - m_s^3 y_s - m_s^4(1 - y_s)\right),$$
$$B_s = \frac{1}{p_s}\left(m_s^1 q_s y_s + m_s^2 q_s(1 - y_s) + m_s^3 r_s y_s + m_s^4 r_s(1 - y_s)\right).$$

これは$v_1, \cdots, v_n$を変数とする線形方程式系である．これらの式から，頂点nでの期待利得を初項とする漸化式が得られる：

$$v_s = \frac{A_s v_{s+1} + D_s}{C_s} \quad (1 \leq s \leq n-1),$$
$$v_n = \frac{D_n}{C_n - A_n}.$$

ここで，C_sおよびD_sは次の式で与えられる：

$$C_1 = 1 + A_1, \quad C_s = 1 + \frac{A_{s-1} A_s}{C_{s-1}},$$
$$D_1 = B_1, \quad D_s = B_s - \frac{A_s D_{s-1}}{C_{s-1}}.$$

したがって，まず値$p_s, q_s, r_s, y_s, A_s, B_s, C_s, D_s$を$s = 1, 2, \ldots, n$の順に計算し，次に値$v_s, x_s$を$s = n, n-1, \ldots, 2, 1$の順に計算すると，均衡解を$O(n)$時間で求めることができる．

5. 利得が頂点番号に依存する場合のアルゴリズム

本節では，各頂点の利得行列における各成分が，頂点番号sをパラメータとした線形関数$f_i(s) = g_i s + h_i$で与えられている場合を考える．前節で述べたように，均衡解を求めるためにはまず値A_s, B_s, C_s, D_sを求める必要があったが，その必要がなくなり，計算が簡略化される．

各頂点の利得行列の各成分に$f_i(s) = g_i s + h_i$を代入すると，p_s, q_s, r_s, y_sは以下のように表すことができる：

$$p_s = g_* s + h_*, \quad q_s = g_{34} s + h_{34}, \quad r_s = g_{12} s + h_{12},$$
$$y_s = \frac{1}{p_s}(g_{24} s + h_{24}), \quad 1 - y_s = \frac{1}{p_s}(g_{13} s + h_{13}).$$

ここで，$g_*, \cdots, h_{13}$は次の式で与えられる：

$$g_* = g_1 - g_2 - g_3 + g_4, \quad h_* = h_1 - h_2 - h_3 + h_4,$$
$$g_{34} = -g_3 + g_4, \quad h_{34} = -h_3 + h_4,$$
$$g_{12} = g_1 - g_2, \quad h_{12} = h_1 - h_2,$$
$$g_{24} = -g_2 + g_4, \quad h_{24} = -h_2 + h_4,$$
$$g_{13} = g_1 - g_3, \quad h_{13} = h_1 - h_3.$$

このときA_s, B_s, C_s, D_sは次の式で与えられる：

$$A_s = 0, \quad C_s = 1, \quad D_s = B_s,$$
$$B_s = \frac{1}{p_s}\{f_1(s) q_s y_s + f_2(s) q_s(1 - y_s) + f_3(s) r_s y_s + f_4(s)(1 - y_s)\}.$$

したがって，v_sは以下のように求められる：

$$v_s = B_s.$$

6. まとめ

一般に，確率ゲームの均衡解は非線型方程式系の解として与えられるが，その解を解析的に求めることは困難である．本研究では，単独支配者確率ゲームにおいて，グラフ構造をパスに限定することにより，均衡解が線型方程式系の解として与えられることを示した．これにより求解が容易になり，$O(n)$時間のアルゴリズムを得ることができた．しかし，グラフをパスに制限するという仮定は非常に強いので，より弱い仮定の下で均衡解を効率的に計算することが今後の課題である．

謝辞

本研究は実社会ビッグデータ利活用のためのデータ統合・解析技術の研究開発プロジェクト（文部科学省）から研究費を受けている．

参考文献

[1] Shapley, L. S. (1953). Stochastic games. Proceedings of the National Academy of Sciences, 39(10), 1095-1100.
[2] Parthasarathy, T., & Raghavan, T. E. S. (1981). An orderfield property for stochastic games when one player controls transition probabilities. Journal of Optimization Theory and Applications, 33(3), 375-392.
[3] Filar, J. A. (1985). Player aggregation in the traveling inspector model. Automatic Control, IEEE Transactions on Automatic Control, 30(8), 723-729.
[4] Raghavan, T. E. S. (2003). Finite-step algorithms for single-controller and perfect information stochastic games. In Stochastic games and applications (pp. 227-251). Springer, Berlin.