経路情報が利用可能な場合における
リグレット最小化基準に基づく交通ネットワーク均衡モデル

Transportation Network Equilibrium models based on the regret minimization criteria given a full travel information

宮城俊彦*
By Toshihiko Miyagi

1. はじめに

リグレットとは、エージェントの行動集合 A に含まれる行動 $k \in A$ がもたらす利得 $V(k)$ とエージェントの行動選択の結果の利得 U の差分、$R(k) = V(k) - U$ 、で定義され、$R(k) > 0$ ならば、行動 k を選択しなかった後悔であり、行動を修正することを要求する。したがって、リグレット最小化基準に基づく行動選択は、$V(k) - U \le 0$ となるような選択行動ルールをどのように求めればよいかが問題になる。この行動ルールはダイナミックに変化する環境において反復して行われる行動すべてに適用可能であり、交通行動を記述する新たなモデルの枠組みを与えるが、本研究では経路選択行動に限定したモデル提案を行っている。

ＩＴＳの進展に伴い、交通情報と交通行動の関係を分析できる新たな枠組みが必要な状況にある。従来の経路選択モデルは均質のドライバー集団を仮定している場合が多い。また、エージェントの得られる情報に関し、過度の条件を仮定する場合が多い。たとえば、確率均衡モデルでは、ドライバーの異質性を仮定し、情報の不確実性を仮定している点では現実的であるが、実際のアルゴリズムでは、ドライバーは自分が実際には選択しなかった他の経路の利用者数を知っており、かつ、それに基づくそれらの経路の利得が計算できることを前提にしている。行動ルールとしてはかなり強い仮定を置いていることになる。すなわち、ここで言う行動ルールは、アルゴリズムとほぼ同様な意味で用いられているが、均衡を前提にしたアルゴリズム構築とは異なっている。ある行動ルールがシステムをどのような均衡に導くのかは行動ルールに依存する。行動仮説に矛盾しないアルゴリズムの開発は、環境変化に伴うドライバー行動をシミュレーションする場合に要請されるもっとも重要な条件の１つである。

Miyagi(2006)は day-to-day の交通環境における交通情報と経路選択行動ルールの関係を完全情報と不完全情報の場合の２つの経路選択問題として定式化している。この定式化はコンピュータ科学あるいはゲーム理論で研究されているリグレット最小化基準の考え方に基づいている。本研究では、ドライバーが経路に関する旅行時間情報を利用できる完全情報のケースについて、より詳細に議論したものである。この場合、リグレットをゼロにする、いわゆる Hannan 一致性を満足する経験分布集合は、不確実な環境を含むダイナミックな環境においても Wardrop 均衡に至る行動ルールが存在することを意味し、経路情報の不確実性のみをもって確率均衡配分を正当化することに否定的な見解を示すものである。

2. 交通情報と経路選択行動

以下においては交通情報を受信し、操作できるプロセッサーをナビゲータと呼ぶ。また、ドライバー、エージェントおよびプレイヤーは同義語として扱う。利得はマイナスの旅行費用と置く。

完全情報下での経路選択行動：各ドライバーはナビゲータを装備しており、それによって交通センターから利用可能経路の旅行時間情報を収集することができると仮定する。各々のドライバーは、自分の利得関数は知っているが他のドライバーの利得関数は知らない。ルート情報は旅行を終えた後に更新されるので、ドライバーは他の経路を選んだならば得たでだろう利得ベクトルを知っている。

不完全情報下での経路選択行動：ドライバーはナビゲータを保有していないか、あるいは、それに従わ

*正会員 工博 東北大学教授 大学院情報科学研究科
（〒980-8579 宮城県仙台市青葉区荒巻字青葉 6-6-06）

ないと仮定する。彼が利用できる交通情報は彼の走行経験によって得られる旅行時間のみである。彼は他のドライバーの利得関数を知らないし、自分の利得関数も知らない。

上記の完全情報仮説は一見すると従来の Wardrop 均衡の経路選択行動仮説と同じに見える。しかし、ここでの完全情報とは利用可能な経路の旅行時間がすべて与えられるという意味の a full information として定義しており、他のドライバーの利得関数が未知という意味では完備（a complete）ではない。すなわち、各ドライバーは他のドライバーの利得構造が分らないので「最適な行動」を選択することはできず、自分の利得を最大にするような「最良の行動」を選択するしかない。

リグレット基準による行動ルールの存在は、Hannan (1957) によって確立されたといわれている。この論文は久しく埋もれていたが、近年になってゲーム論の研究者やコンピュータ科学の分野で再発見され、見直されている。リグレット最小化基準の魅力的な特性の 1 つは、プレイヤーの長期的な平均的利得が、他のプレイヤーの戦略に関わらず、結果として得られる最大利得を保証する行動ルールを与えるという点である。この特性は、Hannan 一致性と呼ばれる。また、利得の定義を若干変更すれば Aumman（1974）によって提案された相関均衡を得ることができる。Hannan 一致性を得るアルゴリズムは Foster and Vohra(1997,1999) によって初めて提案されたが、これは連立方程式を解く必要があり、手間がかかる。これに対し、Hart and Mas-Colell (2000) は、Blackwell の接近性定理(1956)を用いて Hannan 一致性の条件が誘導でき、しかもより簡便なアルゴリズムを得ることができることを明らかにした。また、Blackwell の接近性定理を拡張するためにポテンシャル関数法を提案している(Hart and Mas-Colell, 2001）。本研究は、基本的に Hart and Mas-Colell（2000）の成果に依存しており、それをネットワーク均衡問題に置き換えて解釈し直したものである。

3. 表記法

プレイヤー（エージェント）の集合を $\mathbf{N}=\{1,2,\cdots,N\}$、プレイヤー $i\in\mathbf{N}$ の純粋戦略集合を $\mathbf{S}^i$、利得 $r^i:\mathbf{S}\to\Re$（$\mathbf{S}=\times_{i\in\mathbf{N}}\mathbf{S}^i$）で定義されるワンショット・ゲーム $\Gamma[\mathbf{N},(\mathbf{S}^i)_i,(\mathbf{r}^i)_i]$ を考える。プレイヤー i 以外のプレイヤーを $-i$ と書く。このとき、純粋行動プロフィール $\mathbf{s}\in(\mathbf{s}^i,\mathbf{s}^{-i})\equiv S^i\times S^{-i}$ と置く。プレイヤー i の混合戦略集合を $\Pi^i\equiv\Delta(\mathbf{S}^i)$ そとおき、混合行動プロフィールを $\boldsymbol{\pi}=(\boldsymbol{\pi}^i,\boldsymbol{\pi}^{-i})\in\Pi^i\times\Pi^{-i}$ とおく。$\Delta(\cdot)$ は確率分布が定義される単体。期待利得は、次式で定義される

$$r^i(\boldsymbol{\pi})=\sum_{\mathbf{s}\in\mathbf{S}}\pi(\mathbf{s})r^i(\mathbf{s}) \qquad (1)$$

ワンショット・ゲーム Γ が時間 $t=1,2,\cdots$ を通して繰り返しされる繰り返しゲームを考え、時間 t での行動プロフィール、利得ベクトルを s_t、$r_t:=r(s_t)$ と置く。$\overline{r}_t:=(1/t)\sum_{\tau\le t}r_t$ は、t までの時間平均利得ベクトル。プレイヤー i のプレイの履歴 $\rho_{t-1}=(s_\tau)_{\tau=1}^{t-1}\in\times_{\tau=1}^{t-1}\mathbf{S}_\tau$、そして混合戦略を $\boldsymbol{\pi}_t^i\equiv\boldsymbol{\pi}_t^i(\rho_{t-1})\in\Delta(\mathbf{S}^i)$ と定義するとき、学習アルゴリズムとは写像 $\boldsymbol{\pi}_t:\rho_{t-1}\to\Pi\,\pi_t^i(s^i\,|\,\rho_{t-1})$ であり、過去の履歴に依存するシグマ集合体 $\rho_{t-1}=\{(\mathbf{s}_1,\mathbf{r}_1),\cdots,(\mathbf{s}_{t-1},\mathbf{r}_{t-1})\}$ で定義される

交通に関連した用語とゲーム用語を結びつけるため、以下のような記号を導入する。なお、文脈を分かり易くするため、以下では単一の起終点間でのトリップのみを対象にしている。今、経路集合を $\mathbf{P}=\{1,2,\cdots,M\}$、代表的な経路を $p\in\mathbf{P}$ とおく。プレイヤー i の利用可能経路集合を $\mathbf{P}^i,i\in\mathbf{N}$ とおくとき $\mathbf{P}=\cup_{i\in\mathbf{N}}\mathbf{P}^i$。経路フローを $\mathbf{h}=(h_1,\cdots,h_p,\cdots h_M)$ で表す。リンク集合を $\mathbf{L}$ 、リンクフロー, リンク－経路接続行列要素を $f_\ell,\delta_{\ell p}$ $\ell\in\mathbf{L}$ とおく。また、次のような簡略的な表記法を使用する。すなわち、π_p^i は $\pi^i(p,s^{-i}),p=s^i\in\mathbf{S}^i$ と同等の意味で用いられる。同じ規則は、同様に報酬と経験的分布にも適用される。時間 t までにプレイレー i が経路 p を訪問する頻度は、次式で与えられる。

$$x_{p,t}^i=\frac{1}{t}\sum_{\tau=1}^{t}\mathbf{I}_{\{\mathbf{s}_\tau^i=p\}}=\frac{1}{t}\left|1\le\tau\le t\,|\,\mathbf{s}_\tau^i=p\right|, \qquad (2)$$

ここに、

$$\mathbf{I}_{\{s_\tau^i=s\}}=\begin{cases}1 & if\ s\ is\ played\ at\ time\ \tau.\\ 0 & if\ s\ is\ not\ played\ at\ time\ \tau.\end{cases}$$

このとき、ネットワーク上のフローは経路の経験分

布を用いて次のように定義できる。

$$\sum_{p\in\mathbf{P}^i} x_{p,t}^i = 1$$
$$\sum_{i\in\mathbf{N}} x_{p,t}^i = h_{p,t},\ \forall p\in\mathbf{P} \tag{3}$$
$$\sum_{p\in\mathbf{P}} \delta_{\ell,p}^t h_{p,t} = f_{\ell,t},\ \forall \ell\in\mathbf{L}$$

また、時間平均利得は

$$r^i(\mathbf{x}_t^i,\mathbf{x}_t^{-i}) = \frac{1}{t}\sum_{\tau=1}^{t} r^i(s_\tau^i, s_\tau^{-i}) \tag{4}$$

で与えられ、以下のように平均パス・コストに関係付けられる。まず、リンク $\ell\in L$ の時間 t でのコストを $c_\ell(\mathbf{f}_t)$、経路 $p\in P$ の旅行時間を次のように定義する。

$$u_p(\mathbf{h}_t) = \sum_{\ell\in\mathbf{L}} \delta_{\ell p}^t c_\ell(\mathbf{f}(\mathbf{h}_t))\cdot \tag{5}$$

各々のドライバーが知覚する経路費用を u_p^i とおき、$r_p^i(\mathbf{x}) := -u_p^i(\mathbf{x})$ の関係を仮定する。知覚費用の簡便な定義の１つは平均費用と時間価値を結合させた次式で与えることができる。

$$u_p^i(\mathbf{h}) = w^i u_p(\mathbf{h}) + u_{p0}$$

ここに、w^i ドライバーi の時間価値。u_{p0} ：経路 p の金銭的なコスト

3. リグレット･マッチング･アルゴリズム

3.1　条件なしリグレット

Hart and Mas-Colell（2000）に従い、プレイヤー i の条件なしリグレット(外部リグレット)を定義する。

$$R_t^i(k) := \frac{1}{t}\sum_{\tau=1}^{t}[r^i(k, s_\tau^{-i}) - r^i(s_\tau)]. \tag{6}$$

すなわち、行動 $k\in S^i$ を常に選択するときの平均利得からの差分を表す。これを用いて Hannan 一致性を定義するならば、次のようである。

$$\limsup_{t\to\infty} R_t^i(k) \le 0 \quad for\ all\ k\in S^i \tag{7}$$

式(7)の右辺を $\varepsilon > 0$ に置き換えると、

$$\limsup_{t\to\infty} R_t^i(k) \le \varepsilon \quad for\ all\ k\in S^i \tag{8}$$

Fudenberg and levine（1998）によって提案され ε 一致性である。Hannan 一致性は Nash 均衡や相関を内包する広い概念であるため、ゲーム論の分野ではとくに均衡と結びついた概念とは捉えられていない。しかし、定常的な環境では交通情報をすべてのプレイヤーが共有するため Hannan 一致性は Wardrop 均衡と深く結びついていると考えられる。これを見るため、経験分布を用いて Hannan 一致性の定義し直すことによって次式を得る。

$$R_k^i(\mathbf{x}) = \sum_{\mathbf{s}\in\mathbf{S}}[r^i(k,\mathbf{s}^{-i}) - r^i(\mathbf{s})]\mathbf{x}(\mathbf{s})$$
$$= r^i(k,\mathbf{x}^{-i}) - r^i(\mathbf{x}),\ for\ each\ k\in\mathbf{S}^i \tag{9}$$

すなわち、Hannan 一致性は次式を満足する頻度分布 $\mathbf{x}\ \in\Delta(\mathbf{S})$ の集合を定義している。

$$\bar{r}^i = r^i(\mathbf{x}) \ge \max_{k\in\mathbf{S}^i} r^i(k,\mathbf{x}^{-i}),\ for\ all\ i\in\mathbf{N} \tag{10}$$

この関係は、Williams(1977)が、確率的経路選択において経路を選択する母集団と選択しなかった母集団の異質性を考慮して誘導した関係式と同じである。また、プレイヤーの同時確率分布 $\mathbf{x}\ \in\Delta(\mathbf{S})$ が互いに独立ならば、式（10）は、Nash 均衡に一致する。

次に、式（10）を満足する経験分を得るにはどのような行動ルールを採用すべきであるかを考える。式(7)に見るように Hannan 一致性を満足する経験分布集合は負の象限にあり、閉凸部分集合 $\mathbf{C}$ を構成することができる（図１）。したがって、$\mathbf{C}$ に接する半空間が存在し、Blackwell の定理より、集合の外部から接近できる行動ルールが確実に存在し、また、それは他のプレイヤーの混合戦略に依存しない。Foster and Vohra および Hart and Mas-Colell によって提案された行動ルールは非常にシンプルなものであり、すべての $\mathrm{s}^i\in\mathbf{S}^i, \mathrm{i}\in\mathbf{N}$ に対し、次の混合戦略をとればよいというものである。

$$\pi_t^i(s^i) = \frac{[R_t^i(s^i)]_+}{[\sum_{s^i\in S^i} R_t^i(s^i)]_+},\ where\ [R]_+ = \max(R,0) \tag{11}$$

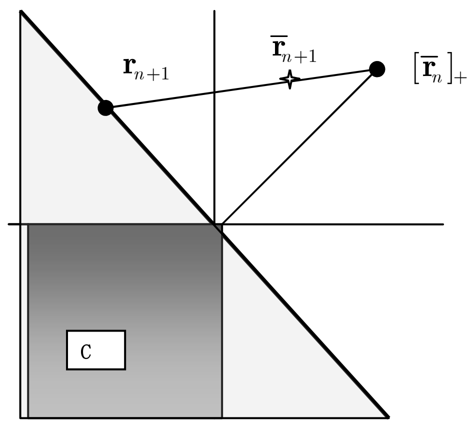

図１　Blackwell の接近性定理

さらに、次の条件が成立すれば

$$\pi_t^i(k) > 0 \quad \text{only if} \quad \bar{r}_{t-1}^i \le r^i(k,\mathbf{x}_{t-1}^{-i})\cdot \tag{12}$$

(11)を満たす戦略はどんな有限ゲームにおいも、プレイヤーの最大リグレットは 0 に収束する。

3.2　条件つきリグレット

式(9)は、以下のように書き改めることができる。

$$R_k^i(\mathbf{x}) = \sum_{\mathbf{s}\in\mathbf{S}} [r^i(k,\mathbf{s}^{-i}) - r^i(\mathbf{s})]\mathbf{x}(\mathbf{s})$$
$$= \sum_{j\in S^i} \sum_{s^{-i}\in S^{-i}} \left[r^i(k,\mathbf{x}^{-i}) - r^i(j,\mathbf{x}^{-i}) \right] x(j,s^{-i})$$

これより、プレイヤー i が行動 j を選択した場合に限定した条件付リグレット（内部リグレット）を定義できる。

$$D^i(j,k) = \sum_{s^{-i}\in S^{-i}} \left[r^i(k,\mathbf{x}^{-i}) - r^i(j,\mathbf{x}^{-i}) \right] x(j,s^{-i}),$$

条件付リグレットは、すべての異なる行動ペア $(j,k)\in S^i$ に対し、行動 j を選択したときに行動 k に置き換えていたならば得られたであろう利得を表している。Foster and Vohra（1997）は条件つきリグレット最小化基準を達成する行動ルールはマルコフ連鎖の定常確率の求める場合と同様な方法で求めることができること、そして、それが相関均衡に至ることを発見した。Hart and Mas-Colell は Blackwell の定理を用いて同じ結論を得ているが、彼らの提案する行動ルールは方程式体系を解く必要がなく、より簡単である。

$$\pi_{t+1}^i(k) = \begin{cases} \gamma \left[D_t^i(j,k) \right]_+ & if \;\; k \neq j, \\ 1 - \sum_{k\neq j} \pi_{t+1}^i(k) & if \;\; k = j. \end{cases}$$

ここに、定数 $\gamma > 0$ は前の時点で選択された行動に再び正の確率になることを保証するように定める。

3.3 アルゴリズム

● 初期化：初期確率分布 $\boldsymbol{\pi}_0^i$ に基づき、純粋行動 $\{s_0^i\}$ を選択する. このときの経験的分布 $\mathbf{x}_0^i$ を求め、結果として生じる経路フロー、リンクフロー、パス・コスト・ベクトル、$(\mathbf{h}_0,\mathbf{f}_0,\mathbf{u}_0)$ を求める。

● For $t = 1,2,\cdots$

$k\in\mathbf{S}^i$ に対する選択性向を決定する。

$$q_t^i(k) = q_{t-1}^i(k) + \frac{1}{t}[R_t^i(k) - q_{t-1}^i(k)],$$
$$where \;\; R_t^i(k) = -(u_k^i(\mathbf{x}_{t-1}) - \overline{u}^i(\mathbf{x}_{t-1}))$$

混合戦略を求める。

$$\pi_t^i(k) = \begin{cases} \dfrac{\left[q_t^i(k)\right]_+}{\sum_{k\in S^i}\left[q_t^i(k)\right]_+}, & \text{if } \sum_{k\in S^i}\left[q_t^i(k)\right]_+ > 0 \\ 0, \text{otherwise} \end{cases}$$

$\boldsymbol{\pi}_t^i$ によって生成される純粋行動 $\{s_t^i\}$ を決定し、経験的分布 $\mathbf{x}_t^i$ を更新する。結果として生じる経路流動量、関連流動量とパス・コスト・ベクトル、$(\mathbf{h}_t,\mathbf{f}_t,\mathbf{u}_t)$ を求める。

条件つきリグレットアルゴリズムもほぼ同様に行うことができる。

参考文献

Aumann, R.: Subjectivity and correlation in randomized strategies, Journal of Mathematical Economics, 1, 67-96, 1974.

Blackwell, D.: An analog of the minimax theorem for vector payoffs, Pacific Journal of Mathematics, 6, 1-8, 1956.

Foster, D. and R. Vohra, Regret in the on-line decision problem, Games and Economic Behavior, 29, 7-35, 1999.

Foster, D. and R. Vohra : Calibrated learning and correlated equilibrium, Games and Economic Behavior 21, 40-55, 1997.

Fudenberg, D. and Levine, D.K.: The Theory of Learning in Games.The MIT Press, Cambridge, MA, USA, 1998.

Hannan, J.: Approximation to Bayes risk in repeated play. In Contribution to the Theory of Games, Vol. III, Annals of Mathematical Studies 39, ed. by M.Dresher, A.W. Tucker, and P. Wolfe. Princeton, Princeton University Press, 97-139, 1557.

Hart, S. and A. Mas-Colell: A simple adaptive procedure leading to correlated Equilibrium, Econometrica, 68(5), 1127-1150, 2000.

Hart, S. and A. Mas-Colell: A general class of adaptive strategies. J. Econ. Theory 98, 26-54,2001.

Miyagi, T.: Multiagent learning models for route choice in transportation networks: An integrated approach of regret-based strategy and reinforcement learning, 11th International Conference on Travel Behaviour Research, Kyoto, 2006.

Williams, H.C.W.L.: On the formulation of travel demand models and economic evaluation measures of user benefit, Environment and Planning A, 9,285-344, 1997.