不確実な交通環境下における経路選択モデルとロバストネスについて

Robust Route-Choice Models under Uncertain Traffic Environment

宮城俊彦*
By Toshihiko Miyagi

1.　はじめに

道路ネットワークにおいて目的地にいたる経路選択に伴う不確実性をモデル化するにあたって次のような状況を想定する。今、2 地点間を連絡する十分な道路容量をもつ信号交差点を含む 1 本の道路を一人のドライバーが走行する状況を考える。この場合、混雑は全く存在せずリンク所要時間は一定であり、時間遅れのランダム性は交差点での停止時間に依存する。従って、トリップ回数が増えるにつれて目的地までの所要時間は正規分布に従うと仮定できる。次に、道路容量の増減に伴う影響を考えてみよう。たとえば、片側 4 車線の道路を 3 車線、2 車線と徐々に小さくした場合、目的地までの到達時間の変動はドライバー特性に依存する。ドライバーは希望維持速度を持っていると仮定した場合、1 車線になっても彼の希望速度が維持できるならば、車線の減少の影響はほとんどないと考えよい。逆に言えば、車線の増加の影響による所要時間の変動はドライバーの維持した希望速度の分布に関連している。ドライバーの希望速度分布は対数正規分布を仮定するほうが自然であろう。こうした状況は、道路混雑の影響を加味するとさらに複雑になる。ドライバーは希望速度を維持できず、交通量によっては車線変更ができない場合もある。信号交差点での待ち時間も交通量に依存する。したがって、目的地までの到達時間が正規分布あるいは対数正規分布に従うと仮定するのは十分な根拠を持たない。精緻な調査を繰り返し到達時間の確率法則を決定することも一つの方法であるが、本研究では所要時間分布を仮定しない場合について考える。人々の行動結果としての利得が環境に依存し、その利得の確率分布が未知の場合を不確実な環境と呼ぶ。

*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）

本研究の目的は不確実な環境下での経路選択行動をモデル化することであるが、その場合の重要な概念がロバストネスである。ロバストネスの概念は、もともと統計学や確率制御理論の分野で研究されてきたが、最近では動学的な経済モデル構築の主要な概念としても注目されている（Hansen and Sargent,2007）。意思決定行動の観点では、ロバストネスは安全サイドの行動、逆に言えば、最悪ケースのシナリオを想定した意思決定であり、ゲーム理論の min-max 原理に対応している。

経路選択問題に対応させれば、ドライバーは目的地までの経路所要時間を最小にすることを目的としているが、環境（自然）はドライバーの意思決定に干渉し、目的地までの所要時間を長くするような働きをすると考える。この問題は Patek and Bertsekas(1997)によって確率的最短経路ゲーム (Stochastic Shortest Path Games:SSPG)として定式化され、その特性が明らかにされた。Patek and Bertsekas の議論は Bertsekas and Tsitsiklis（91）によって提案されたマルコフ意思決定問題(MDP)に基づく確率的最短経路(Stochastic Shotest Path;SSP)問題の拡張であるが、しかし、マルコフ連鎖の推移確率が分かっているという仮定は保持されている。この仮定はＭＤＰにおいては現時点での意思決定に次の時点での状態の最適値関数の情報が必要であり、推移確率が分かっていればその期待値が求められるので必要な仮定になる。しかし、交通ネットワークにおいて状態間（たとえばノード間）の推移確率をすべて知ることはほとんど不可能である。そのため、推移確率を仮定しない方法の開発が必要になる。このため、Ｑ学習あるいは価値関数を近似する方法が提案されている。この場合、近似された関数は真の価値関数とは異なるため、学習によって更新される必要がある。許容される範囲に価値関数を更新できればシステムはロバストネスをもつことになる。言い換えれば、真の価値関数が存在するとして、それを近似関数で置き換えたとしてもシステムが安定的に解を持つのであれば、その近似はロバストである。

上記のような最適制御理論に基づくロバストネスとは異なる概念がゲーム理論では提案されている。自然がどのような戦略をとろうともプレイヤーの利得が一定の水準を確保できるような選択ルールを求めれば良いというもので、長期的には合理的な意思決定行動に漸近する行動ルールを与える。このような行動ルールは Hannan 一致的と呼ばれる。Hannan ロバストネスは、不確実な環境下での行動を想定し、意思決定での各ステージでの最適化行動を仮定していない。つまり、best response ではなく、 現状をより改善する better response を仮定する。

本研究は, 経路選択行動に焦点をあて、ロバストなモデルリングの基本フレームを検討したものである。まず始めに確率ゲームとして定式化の基本フレームを示し、次に、ロバストＤＰとの関係を分析する。そして、最後に Hannan 一致性とロバストネスとの関係に言及する。

２．SSPG

２．１　表記法

状態空間を $\mathbf{S}$ とおき、その要素を $i=1,\cdots,n$ とラベルを付す。各 $i\in\mathbf{S}$ に対し、エージェントの行動集合、環境（自然）の行動集合を各々 $A_i=\{a_{1i},\cdots,a_{mi}\}, B_i=\{b_{1i},\cdots,b_{ni}\}$ で表わす。これら２つの集合を制御制約集合と総称する。 $a\in A_i, b\in B_i$ のもとで、$i\in\mathbf{S}$ から $j\in\mathbf{S}$ への推移確率を $p_{ij}(a,b)$ で表わす。またそのときの推移コストの期待値を $c_i(a,b)$ とおく。ある状態から次の状態へ移行するプロセスをステージと呼び、$T=\{0,1,\cdots,N-1\}$ で記述する。

エージェントは起点から終点までの全ステージに渡るコストを最小化しようとする最小化エージェントであるのに対し、環境はエージェントに敵対し、全ステージのコストを最大化するようにシステムに作用する最大化エージェントとして位置づける。最小化エージェントの各ステージでとる戦略の集合を $\Pi=\{\pi\,|\,\mathbf{a}(i)\in A_i, i\in\mathbf{S}\}$ 、最大化エージェントの１ステージ戦略集合を $\mathcal{T}=\{\tau\,|\,\mathbf{b}(i)\in B_i, i\in\mathbf{S}\}$ とおき、それぞれのエージェントの戦略集合を次のように定義する。

$$\boldsymbol{\pi}=\{\pi^0,\cdots,\pi^k,\cdots\}、\boldsymbol{\tau}=\{\tau^0,\cdots,\tau^k,\cdots\}$$

エージェントが常に同じ戦略をとるとき、すなわち、$\pi^k=\pi$ *for any* k のとき、定常戦略と呼び、$\pi_\infty=\{\mu,\cdots,\mu,\cdots,\mu\}$ とおく。最大化エージェントに対しても同様の定義を用いる。与えられた戦略 $\pi\in\Pi,\tau\in T$ の下での推移確率行列を $P(\pi,\tau)$ とするとき、その要素を次のように表わす。

$$[P(\pi,\tau)]_{ij}=p_{ij}(\mathbf{a}(i),\mathbf{b}(i)):=p_{ij}(\pi,\tau)$$

同様に、推移コストに関しても次の表記法を用いる。

$$[c(\pi,\tau)]_{ij}=c_{ij}(\mathbf{a}(i),\mathbf{b}(i)):=c_{ij}(\pi,\tau)$$
$$[c(\pi,\tau)]_{i}=c_{i}(\mathbf{a}(i),\mathbf{b}(i))=\sum_{j\in S}p_{ij}(\pi,\tau)c_{ij}(\pi,\tau)$$

戦略のペア $\{\pi,\tau\}$ によってもたらされる期待費用ベクトルは、次式で与えられる。

$$\begin{aligned}&\phi(\pi,\tau)=\lim_{t\to\infty}\inf h^t(\pi,\tau)\\&\mathit{where}\\&h^t(\pi,\tau):=c(\pi^0,\tau^0)\\&\qquad+\sum_{\tau=1}^{t}\Big[P(\pi^0,\tau^0)\cdots P(\pi^{\tau-1},\tau^{\tau-1})\Big]c(\pi^\tau,\tau^\tau)\end{aligned}\tag{1}$$

ドライバーは目的地に到着すれば、その他の状態を移行しないものと仮定する。これを表現するため、状態 1 を吸収状態と定義し、次の仮定をおく。

$$\begin{aligned}&\mathit{For}\ A_1\ \mathit{and}\ B_1\\&p_{ii}(a,b)=1\\&c_1(a,b)=0\end{aligned}$$

２．２　確率ゲーム

Altman and Hordijk (1995)は、２人ゼロ和確率ゲームが不確実な環境下での最悪ケースの意思決定問題に応用できることを明らかにし、待ち行列が発生する問題に適用した。また、Patek and Bertsekas(1997)は、確率的最短経路問題(SSP)に対し同様なアプローチを適用している。彼らは、SSP に課せられる標準的な仮定に加えて正則性条件を仮定し、そしてベルマンオペレータの単調性および連続性を利用して SSPG のいくつかの特性を明らかにしている。最小費用経路探索問題は、一般的には、ステージ $T=\{0,1,\cdots,N-1\}$ で観測されるダイナミックシステムとして定式化できる。ただし、Patek and Bertsekas(1997)は SSPG を割引率なしの無限計画期間問題として定式化しており、本研究もそのフレームを踏襲する。本研究に関連する主要な結果のみをここに示す。

まず、次のベルマンオペレータを導入する。

$$\mathbb{T}(x) = \inf_{\pi\in\Pi,\tau\in\mathcal{T}} \sup \left[c(\pi,\tau) + P(\pi,\tau)x\right] \qquad (2)$$

あるいは、

$$[\mathbb{T}h]_i \equiv \underset{ab}{val}\left[\sum_{j\in S} p_{ij}(a,b)(c_{ij}(a,b) + h_j)\right]$$

ここに val は minmax 値を表す。

命題１：オペレータ T は唯一の不動点、$\mathbb{T}(x^*) = x^*$、をもつ。

命題２：不動点 $\mathbb{T}(x^*) = x^*$ は、SSPG の均衡費用ベクトルである。その均衡を達成する定常戦略 $\pi^* \in \Pi, \tau^* \in \mathcal{T}$ が存在し、$\{\pi^*, \tau^*\}$ に対し次式が成立する。

1. $\phi^* = \phi(\pi^*, \tau^*)$
2. $\phi(\pi, \tau^*) \geq \phi(\pi^*, \tau^*)$
3. $\phi(\pi^*, \tau) \leq \phi(\pi^*, \tau^*)$

３．ロバストＤＰモデル

Patek and Bertsekas(1997)によって示された SSPG の minmax 定理は、正確な推移確率が与えられていることが前提である。しかし、一般に推移確率を与えることは困難であり、また、「次元の呪い」の問題がある。「次元の呪い」を回避するため、あるいは。計算効率性を上げるために、推移確率あるいは価値関数を別の方法で予測し、近似することがＤＰを応用した実用的なモデル開発において重要である。しかし、近似に伴う誤差は価値関数の更新過程で伝播するので誤差の累積によって解は収束しない恐れも出てくる。このため、収束が保証される価値関数あるいは推移確率の近似が必要になる。それがロバストネスの概念である。この節では Nilim and Ghaoui(2005)に従い、SSPG をロバスト DP として定式化する。SSPG では環境のとる行動あるいは戦略が与えられている必要があった。しかし、SSP において環境の方策を具体的に想定するよりも環境は推移確率そのものに作用すると考えたほうがより自然である。すなわち、

$$p_{ij}(a,\tau) = \sum_{b\in B_i} p_{ij}(a,b)\mathbf{b}(i) := p_{ij}^a \qquad (3)$$

このように考えると、確率ゲームは動的計画問題(DP)で表されることになる。ただし、通常の DP(ODP)と異なる点は、推移確率が未知という点である。この場合、最小化エージェントはどのような推移確率が生じるのかを知らないという意味で不確実な環境に直面している。推移確率が明示的に与えられない場合のＤＰを以下ではロバストＤＰ問題と呼ぶ。(3)に示したようにロバストＤＰでは環境のとる戦略は推移確率に埋没しているので具体的に表示する必要はない。したがって、以下では、最小化エージェントがステージ t でとる行動 $a \in \mathbf{A}$ のみを対象にする。起こりうる推移確率の集合 Θ の部分集合 $\mathcal{P}^a$ とおき、最大化エージェントが $\mathcal{P}^a$ から推移確率を選択する問題を考える。$\mathcal{P}^a$ を信頼性集合と呼ぶ。ロバストＤＰではどのように信頼性集合を特定化していくかが重要になる。

環境が選択する戦略を $\tau = \left(P_t^a\right)_{a\in A, t\in T}$ とおく。P_t^a はステージ t で行動 a のときの $n \times n$ 推移確率行列を表わす。システムがスタートする初期状態は i_0 とおく。方策 π は各ステージでエージェントのとる行動の系列として、$\pi = (\mathbf{a}_0, \cdots, \mathbf{a}_t, \cdots, \mathbf{a}_{N-1})$ で表わす。$\mathbf{a}_t(i)$ はステージでの状態 i でエージェントのとる行動を表わす。このとき、エージェントの戦略空間を $\Pi = \mathbf{A}^{nN}$ で表わす。このとき、ＯＤＰは、与えられた推移確率行列 τ に対し、次のように定義できる。

$$\phi(\Pi, \tau) = \min_{\pi\in\Pi} h(\pi, \tau) \qquad (4)$$

行動環境の取りうる戦略の許容集合は $\mathcal{T} := \left(\otimes_{a\in A}\mathcal{P}^a\right)^N$ である。このとき、定常な環境は次のように定義される。

$$\mathcal{T}_s = \left\{\tau = (P_t^a)_{a\in A, t\in T} \in \mathcal{T} \mid P_t^a = P_s^a \;\; for\ every\ s,t \in T\right\}$$

ロバスト制御問題とは次式を満足する方策を決定する問題である。

定常モデル

$$\phi(\Pi, \mathcal{T}_s) = \min_{\pi\in\Pi} \max_{\tau\in\mathcal{T}_s} h(\pi, \tau) \qquad (5)$$

非定常モデル

$$\phi(\Pi, \mathcal{T}_s) \leq \phi(\Pi, \mathcal{T}) = \min_{\pi\in\Pi} \max_{\tau\in\mathcal{T}} h(\pi, \tau) \qquad (6)$$

Nilim and Ghaoui(2005)は、ＤＰの変形を用いて(6)を解き、また、非定常なマルコフ推移確率行列の定常性を仮定した信頼性集合を求めるために定常な推移確率の仮定から誘導される $\mathcal{P}^a$ による信頼性の上界を求めるというアプローチを取っている。こ

のとき、次のことが成立することを証明した。

定理（ロバスト DP：Nilim and Ghaoui,2005）

(1) 双対性

$$\phi(\Pi,\mathcal{T})=\min_{\pi\in\Pi}\max_{\tau\in\mathcal{T}}h(\pi,\tau)=\max_{\tau\in\mathcal{T}}\min_{\pi\in\Pi}h(\pi,\tau):=\psi(\Pi,\mathcal{T})$$

(2) 最悪ケース価値評価

$$v_t(i)=\min_{a\in\mathbf{A}}\left[c_t(i,a)+\sigma_{P_i^a}(v_{t+1})\right]$$

(3) 最適制御方策 $\pi^*=(\mathbf{a}_0^*,\cdots,\mathbf{a}_{N-1}^*)$

$$a_t^*(i)\in\arg\min\left\{c_t(i,a)+\sigma_{P_i^a}(v_{t+1})\right\},i\in S$$

(４) 最悪ケースの方策評価

$$v_t^\pi(i)=c_t(i,\mathbf{a}_t(i))+\sigma_{P_i^{a_t(i)}}(v_{t+1}^\pi),i\in S$$

ロバストＤＰの核心をなすのは(2)～(4)に現れる $\sigma_{\mathcal{P}}(v)$ である $\sigma_{\mathcal{P}}(v)$ は

$$\sigma_{\mathcal{P}}(v)=\sup\left\{p'v\mid p\in\mathcal{P}\right\}$$

で定義される集合 $\mathcal{P}$ の支持関数である。Nilim and Ghaoui(2005）は $\sigma_{\mathcal{P}}(v)$ が最尤法あるいは相対エントロピー（カルバック情報量規準）などを用いて求めれることを明らかにした。

４．Hannan 一致性とロバストネス

Nilim and Ghaoui(2005)は定常モデル（５）の解を求めるのが困難という理由で２段階に分けた方法をとった。しかし、定常モデルは Hannan 一致性を満足するリグレットマッチングモデルで解くことができる。以降では、表記を簡便にするため利用可能経路集合 ℝ とし、経路のコストを $h_r,r\in\mathbb{R}$ で定義する。まず、時刻Ｔまでのリグレットを次式で定義する。

$$L_T(r)=\sum_{t=1}^{T}h_r(\mathbf{I}_{a_t},\mathbf{I}_{b_t})-\min_{i\in S}\sum_{t=1}^{T}h_r(a_t,\mathbf{I}_{b_t}),r\in\mathbb{R} \tag{7}$$

Hannan 一致性は次のように定義できる（Cesa-Bianchi, and Lugosi 、2006)。

$$\limsup_{T\to\infty}\left(\frac{1}{T}\sum_{t=1}^{T}h_r(\mathbf{I}_{a_t},\mathbf{I}_{b_t})-\min_{r\in\mathbb{R}}\frac{1}{T}\sum_{t=1}^{T}h_r(a_t,\mathbf{I}_{b_t})\right)\le 0,\ a.s. \tag{8}$$

(7)において、 I_{a_t} は時刻ｔでプレイヤーが選択した行動に応じて１－０をとる指示変数であり、ランダム戦略を行使するとき、 I_{a_t} は確率変数になる。

（７）の右辺第２項はＤＰ問題として定式化できる。したがって、Hannan 一致性では第１項を求めることができたら、ＳＳＰの近似解を得ることができる。しかし、(8)はランダム選択における期待利得がＤＰの下限を与えることを述べているに過ぎず、どれくらい乖離しているかは明らかにはしていない。ところが、ゼロ和ゲームを想定すれば、もう少し強いことが言え、次式が成立する。

$$\frac{1}{T}\sum_{t=1}^{T}h_r(\mathbf{I}_{a_t},\mathbf{I}_{b_t})=\min_{\pi\in\Pi,\tau\in T}\max\frac{1}{T}\sum_{t=1}^{T}h_r(\pi,\tau),r\in\mathbb{R}$$

左辺の値を求める方法、すなわち、不確実な環境下でのランダム経路選択モデルについては宮城(2007)によって提案されている。宮城(2007)の方法ではもう少し一般的な状況が想定されているものの収束計算に時間がかかりすぎる難点がある。したがって、ロバストＤＰの計算効率性との比較が重要になる。

参考文献

Altman, E. and A. Hordijk (1995): Zro-sum Markov games and worst-case optimal control of queueing systems, Quueing Systems,21, special issue on optimization of queueing systems,S. Studham (ed.),415-447.

Bertsekas, D.P. and J.N. Tsitsiklis(1991):Analysis of stochastic shortest path problems, Mathmatics of Operations Research, 16, 580-595.

Cesa-Bianchi, N., and G. Lugosi (2006): prediction, learning, and Games. Cambridge University Press,NY.

Hart, S. & A. Mas-Collel : A simple adaptive procedure leading to correlated equilibrium, Econometrica 68(5), pp.1127-1150, 2001.

Hansen L.P. and Sargent T.J.(2007):Robustness, Princeton University Press, USA.

Nilim, A. and L.E. Ghaoui (2005): Robust control of Markov decision processes with uncertain transition matrices, Operations Research, 53(5), 780-798.

Patek, S.D., and D.P. Bertsekas(1999): Stochastic shortest path games, SIAM Journal on Control and Optimization, 37(3),804-824.

宮城俊彦(2007):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経路選択行動のモデル化、土木計画学研究発表会講演集。