Regret Matchingを用いた経路選択行動分析*

―不完全交通情報を仮定した粒子モデル―

AN ANALYSIS OF ROUTE CHOICE BEHAVIOR BY A REGRET MATCHING MODEL *
-ATOMIC MODEL WITH IMPERFECT TRAVEL INFORMATION-

宮城俊彦**・石黒雅彦***
By Toshihiko MIYAGI** and Masahiko ISHIGURO***

１．はじめに

本研究では個々のエージェントを自律的なエージェントとして定義し，それぞれがその直面する環境下で利得最大化行動を学習していく繰り返しゲームとして交通ネットワークフローを求めるモデルを提案する．個々のエージェントは個別のODペアを持ち，利用可能な情報に応じて経路選択肢の集合を決定し，その選択肢の中から経路を選択する．交通環境はそれぞれのエージェントの非協力的な選択によって変化するので，個々のエージェントは交通環境を正確には知りえないが，経験した所要時間を通して交通環境を学習し，適応的に経路選択を行なう．本研究は個々のエージェントを独立した意志決定主体として扱う（粒子モデル(注1)）と同時に、経路選択を各ノードでのリンク選択問題の系列として定式化し，リグレットマッチングによって環境学習をさせることによってNash均衡が実現できることを示したものである．

マッチングとは，観測された情報に基づくエージェントの戦略を経験分布に一致させるメカニズムあるいは行動ルールを指す．ベイズ戦略も主観的分布を観測分布に一致させていくルールであるが，事前確率を事後確率に変換するベイズ確率公式のような合理性のメカニズムを介入させている。これに対し，この論文で扱うマッチング理論は、実現値あるいは観測値のみに基づいている点で、ベイズ理論よりも限定された合理性を扱っている。観測値のみに基づき、最適化のメカニズムを含まない行動学習ルールは適応学習モデルと呼ぶことができ，強化学習あるいはリグレットマッチングが知られている．

これまでの経路選択モデルではエージェント集団を同質で，完全情報，完全予見を持つ無限に分解可能な意思決定集団（非粒子モデル(注1)）として扱ってきた．非集計ロジットモデルはこのエージェント集団に選好関係や認知誤差のばらつき等の不確実性を考慮することで上記のエージェントの同質性の仮定を緩和することはできる.しかし，この場合の不確実性はあくまでエージェント集団に対する外部観察者（あるいは計画者）の不確実性であり，個々のエージェントは最適な行動をとることを仮定している.また，ロジットモデルを前提にした確率的ネットワーク均衡問題では個々のエージェントではなく集団としてのエージェントを扱う必要があり，事実上は集計モデルと同じであり，したがって，交通情報提供あるいは交通環境変化と個別のエージェントの意思決定の相互作用を扱うことはできない.

これらの問題を解決すべく，個々のエージェントを単一の意思決定主体として扱い，経路に関する情報を持たない不完全情報から学習行動によって交通環境を学習し，最適な選択を探索する新たな経路選択モデルが宮城[1)-3)]によって提案された.宮城あるいはMiyagi and Ishiguro[4)]では，エージェントのday-to-dayの経路選択行動をN人プレイヤーを対象とした繰り返しゲームとして記述しており，経路選択モデルとして交通ネットワーク解析に適用できる. Nash均衡を求めるための反復法的アルゴリズム(例えば，fictitious playやsmooth factious play)の妥当性は２人ゲーム場合に限定され、N人ゲームのような一般ゲームに対しては，プレイヤーの利得構造が同一で，相手の行動が観測できるポテンシャルゲーム（混雑ゲーム）以外は知られていない[5)]．Nash均衡の存在は知られていてもそれを求めるための計算手法はまだ開発途上の段階にある[6)]．一方、Hart and Mas-Collel[7)]によって提案されたリグレット・マッチング・アルゴリズムは，N人ゲームを対象に，収束の保証された手法である。実現する均衡はNash均衡を端点として含む相関均衡そして粗い相関均衡[8)](宮城[3)]ではこれをHannan均衡と呼んでいる)である．従って、2人ゼロ和ゲームとして扱う場合にはNash均衡を求めるアルゴリズムとして機能させることができる（ただし，２人非ゼロ和の場合には保証の限り

*キーワーズ：経路選択，交通行動分析，交通ネットワーク分析
**正会員，工博，東北大学大学院情報科学研究科，
***工修，東北大学大学院情報科学研究科学生員
（宮城県仙台市青葉区荒巻字青葉6-6，
TEL：022-795-7495，
E-mail:toshi_miyagi@plan.civil.tohoku.ac.jp）

ではない）．

宮城の学習モデル [3]では大規模な集団を扱い，ポテンシャルゲームの利点を生かすため，特殊な利得関数が用いられている．また，経路集合を既知として経路選択確率を求めるなどの簡略化行われている．これらの仮定のうち，後者の問題を緩和するのがこの研究の目的である．

本研究ではゲーム理論に沿った形でエージェントの経路集合を限定するためにk最短経路法を導入する．エージェントの利用可能経路集合を限定するには Dial 法の利用も考えられる．しかし，経路選択行動を展開ゲームとしてモデル化するには，k 最短経路法のほうがループを含まない，ゲーム木と類似のグラフが得られる点で適している．この場合，ネットワーク上でのエージェントの意思決定は経路ベースではなく，ノードベースになる．宮城によるリグレットマッチングの経路選択行動への応用は経路選択に対する混合戦略を用いており，ノードベースの意思決定の結果としての経路選択確率ではない．展開ゲームではノードベースの行動戦略を行動確率と呼び，混合戦略（本研究の文脈では，履歴の生起確率）と区別しており，実際，両者は必ずしも一致しない．しかし，完全記憶のゲーム(perfect recall games)の場合には両者が一致することが Khun によって明らかにされており [9]，本研究でもエージェントは起点からあるノードまでの経路を記憶しているという完全記憶の仮定をおく．展開ゲームにおけるノードベースでの行動確率をリグレット・マッチングで求めるアイデアは Zinkevich[10]によって提案された．Zinkevich によって提案されたリグレットは仮想リグレット(counterfactual regret; CFR)と呼ばれる．

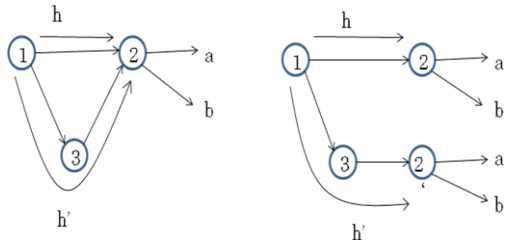
ところで，k 最短経路法で展開されるグラフと展開ゲームのグラフは類似しているが，ゲーム木としての意味付けは大きく異なっている．展開ゲームにおける決定木には他のエージェントの戦略を表すノードとアークが追加されているからである．本研究では展開ゲームにおける決定木から他のエージェントの関与する部分を取り除き，エージェントの自己完結型の展開ゲームとして取り扱っている．その意味でk最短経路法で展開されるグラフはゲーム木ではなく、単一の意志決定者の決定木である．

本研究では，展開ゲームにおけるグラフをk最短最短経路グラフの整合性を図ると同時に，Zinkevich の仮想リグレットモデルをk最短経路樹として展開されるグラフに適用することによって，規模の大きなネットワークに適用可能なリグレット・マッチング・アルゴリズムを提案する．ただし、Zinkevich[10]のモデルも２人ゼロ和ゲームを対象にしており、非ゼロ和ゲームの場合には Nash 均衡への収束は保証されない。そのため、数値計算例を通して Nash 均衡への収束性を検証する．

２．基礎的な概念と表記法

（1）展開ゲームの木とk最短経路樹

展開ゲーム $G=\{N,H,Z,A,I,u\}$ を次のように定義する．N はエージェントの集合，H は履歴(history)の集合である．履歴集合はエージェントのすべての可能な行動系列を表わし，展開ゲームを決定木で表記したときのノードになる．言い換えれば，履歴とは起点からあるノードに至るノード系列で構成される経路である．エージェント $i\in N$ の利得 u_i は決定木における端末ノード $Z\subseteq H$ で決定される．A をエージェントの行動集合おく．このとき，履歴 $h\in H$ での選択可能な行動 $A(h)=\{a:(h,a)\in H\}$ とは、履歴（意思決定ノード）h の後続ノード集合であり，選択が新たな履歴を形成するような行動である． このように行動選択は履歴依存性をもつ．エージェント $i\in N$ の情報分割 $\mathcal{I}_i$ とは， $\mathcal{I}_i$ に含まれる履歴 h 、h' に対し、$A(h)=A(h')$ を満たすような履歴の分割である．情報分割された集合の要素 $I_i\in\mathcal{I}_i$ を情報集合と呼ぶ．$I_i\in\mathcal{I}_i$ における行動集合を $A(I_i)$ で表す．この定義にしたがえば，ループを構成する経路は異なるノード対で分割しなければならない．例えば，図-1 においてノード 2 に至る 2 つの経路 h および h’は異なる履歴を持つ．しかし，(a)図のままではこれらの異なる履歴は同じ行動集合{a，b}をもつため履歴依存性を表現できない．h と h’が無差別ならば，これらの履歴は同じ情報集合に含まれるとし，情報集合 2 の要素を構成することになるが，完全記憶の仮定より，h と h’は識別できるので，(b)図のように分割するが必要である．このとき、履歴 h での行動 a の選択と履歴 h’での行動 a の選択は独立とみなされる．



(a) (b)

図-1 履歴と選択行動

ゲーム木においては，他のエージェントの純粋戦略を表すグラフが混在する．例えば，上図(b)において 1→2 がエージェント１（例えばエージェント）の純粋戦略の１つならば，a または b に至るエッジはエージェント２（例えば信号制御）の純粋戦略を表している．これによって他のエージェントの意思決定の影響を表すが，本研究ではk最短経路法で展開されるグラフは，あるエージェントの意思決定グラフであり，各ノードはすべて情報集合であると仮定する．例えば，上図（b）の場合，

各ノードでの分岐はすべてエージェントのみの意思決定に関わっており，各ノードは信号制御の決定に依存した情報集合である．しかし，上記の仮定の下では信号制御の決定グラフは各ノードに埋め込まれ，表示されない．このように他のエージェントの決定に依存した分岐が起きないため，本来のゲーム木とは異なっている．この簡略化は「リグレット・マッチングは他のプレイヤーがいかなる戦略をとろうとも保証される行動ルールを導く」という保証原理[7, 11]を利用したものである．なお，ループを含まないk経路探索法は，Yenによって提案され，その数多くのアルゴリズムが提案されている．k 経路探索法に関する詳細は巻末の文献[12]に譲る．以降では，k最短経路法で展開されるグラフをゲーム木と区別するため単に決定木と呼ぶ.

（2）行動確率と経路選択確率

展開ゲーム G におけるエージェント i の戦略集合を Σ_i とおく．このとき，行動戦略 $\sigma_i \in \Sigma_i$ とは情報集合 $I_i \in \mathcal{I}_i$ における行動集合 $A(I_i)$ 上で定義される確率分布である.エージェント i の戦略を σ_i，i 以外の全エージェントの戦略を σ_{-i} で表わす．各エージェント $i \in N$ に対し，終端ノードZで定義されるベクトル関数 $u_i(\sigma_i, \sigma_{-i})$ を利得関数と呼ぶ.

$\pi_i^\sigma(h)$ をエージェント i が行動戦略 σ に従ったときの履歴 h の生起確率とおく.履歴 h の生起確率は，ノード h に至るすべての行動系列であり，決定木における経路選択確率とみなせる．決定木の各ノードで行動選択確率と経路選択率の関係は，$\sigma_i(s)$ を履歴 $s \subseteq h$ での行動確率とおくとき，以下のように定義できる.

$$\pi_i^\sigma(h) = \Pi_{s \in h} \sigma_i(s) \tag{1}$$

エージェント i 以外のエージェントの履歴生起確率 $\pi_{-i}^\sigma(h)$ についても同様に定義できる．すべてのエージェントが行動戦略 σ に従ったときの履歴 h の生起確率は，エージェントの独立性を仮定すれば（非協力ゲーム），

$$\pi^\sigma(h) = \Pi_{i \in N} \pi_i^\sigma(h) \tag{2}$$

このとき，エージェント i の期待利得は次式で定義できる.

$$u_i(\sigma) = \sum_{h \in Z} u_i(h) \pi^\sigma(h) \tag{3}$$

Nash 均衡は戦略 σ_i を用いて以下のように定義される.

$$\forall i \qquad u_i(\sigma) \geq \max_{\sigma_i' \in \Sigma_i} u_i(\sigma_i', \sigma_{-i}) \tag{4}$$

また Nash 均衡の近似である ε Nash 均衡も同様に定義できる.

$$\forall i \qquad u_i(\sigma) + \varepsilon \geq \max_{\sigma_i' \in \Sigma_i} u_i(\sigma_i', \sigma_{-i}) \tag{5}$$

（2）Counterfactual Regret(CFR)

エージェントが戦略 σ にしたがって行動しているときのエージェント i の履歴 h での期待利得を $u_i(\sigma, h)$ とする．また，ステージ t で全てのエージェントが戦略 σ^t にしたがって行動しているとき，エージェント i が到達した情報集合 I において得られるステージ t での期待利得を仮想利得 $u_i(\sigma^t, I)$ とおく．.

最後に情報集合 I で選択可能なすべての行動 $a \in A(I)$ に対して戦略 $\sigma|_{I \to a}$ を定義する.これは情報集合 I でエージェント i 以外の全てのエージェントが戦略 σ にしたがって行動するとき，エージェント i のみが戦略 σ ではなく，行動 a を選択するという戦略である.

今，$t = 1, 2, \ldots, T$ の各ステージで，各ノードにおける行動戦略を σ^t とするときの CFR は以下のように定義できる[10].

$$R_i^T(I,a) = \frac{1}{T} \sum_{t=1}^{T} \pi_{-i}^{\sigma^t}(I) \left(u_i(\sigma^t|_{I \to a}, I) - u_i(\sigma^t, I) \right) \tag{6}$$

このリグレットは情報集合 I においてエージェント i が行動戦略 σ によって決定される行動確率分布よりも行動 a を選択したときの効用が大きいときに発生し，行動 a を選択していればよかったというリグレットである.

式(6)のリグレット最小化するための情報集合 I での行動 a を選択する確率は，次式で与えられることがBlackwellの接近性定理に基づいて証明されている[6].

$$\sigma_i^{T+1}(I)(a) = \begin{cases} \dfrac{R_i^{T,+}(I,a)}{\sum_{a \in A(I)} R_i^{T,+}(I,a)} & if \sum_{a \in A(I)} R_i^{T,+}(I,a) > 0 \\ \dfrac{1}{|A(I)|} & otherwise. \end{cases} \tag{7}$$

ただし，$\sigma_i(I)(a)$ は、I での行動確率ベクトルの要素 a の選択確率を表す．また、$R_i^{T,+}(I,a) = \max\left(R_i^T(I,a), 0\right)$ とおいている.この戦略の下での次期の行動は正のリグレットの大きさに応じた確率選択によって決定される.もし情報集合 I で正のリグレットが存在しない場合はランダム選択を取るものと仮定する.

Zinkevich[10]は行動確率式(7)を用いた行動選択を行うことにより，ε Nash 均衡が求められることを示した.

定理（Zinkevich[10]）

エージェントがCFRを用いたリグレットマッチングを行なう場合，次式が成立する.

$$\max_{a \in A(I)} R_i^T(I,a) \leq \frac{\Delta_{u,i} \sqrt{|A_i|}}{\sqrt{T}} \tag{9}$$

$$\frac{1}{T} \max_{\sigma_i' \in \Sigma_i} \sum_{t=1}^{T} (u_i(\sigma_i', \sigma_{-i}^t) - u_i(\sigma^t)) \leq \frac{\Delta_{u,i} |\mathcal{I}_i| \sqrt{|A_i|}}{\sqrt{T}} \tag{10}$$

ただし．$|A_i| = \max_{h:P(h)=i} |A(h)|$．また，$\Delta_{u,i}$ は，次式で与えられる $i \in N$ の効用のレンジを表す.

$$\Delta_{u,i} = \max_z u_i(z) - \min_z u_i(z)$$

式(9)において$T \to \infty$とおけば，リグレットはゼロに漸近すること，また，式(10)より戦略はNash均衡に限りなく近づく．ただし，その収束の速度は，効用のレンジ，情報分割に含まれる情報集合の要素の数，また，情報集合での行動集合の要素数などが影響するため，実現するのは式(6)で定義されるε Nash均衡である．

３．ノードベースのリグレット・マッチングモデル

(1) 仮想リグレットの簡略化

複数のエージェントの意思決定によって変化する交通環境を展開ゲームで表現し，各ノードでの行動選択確率を CFR に従って決定する経路選択問題を考える.エージェントはそれぞれ独自の OD ペアを持ち，OD ペア間の旅行時間が最小になるように行動することを望んでいる．なお，それぞれのエージェントの利用経路集合の要素数は異なっていてもかまわない．言い換えれば，OD ペアごとの利用可能経路数は異なっていてもかまわない.

ループを含まない k 最短経路樹は展開ゲームの決定木と同じ構造をもち，起点からあるノードに至るまでのノードの系列が履歴を構成し，各ノードでの分岐リンクを選択する確率が行動確率で決定される．ノードでのリンク選択（あるいは次のノードの選択）が確率的に決定されるのは，他のエージェントの行動を知らず，そのためリンク選択による最終的なコスト（負の利得）を知ることができないためである．単純な選択行動（いつも決まったリンクを選択する）あるいはランダムにリンクを選択する行動は合理的な行動とは言えず，すべてのエージェントがそのように行動すればネットワークのフローは安定しないか，あるいは達成可能なコストをはるかに上回るコストを強いられることになる．２章に示した仮想リグレットルールは限定的ではあるが，より効率的に最少費用経路を探索する学習法を提供する．

ところで，エージェントの行動集合は道路の幾何的構造によってのみ定義され，他のエージェントの行動によって自分の行動集合が影響を受けることはない.従って，意思決定ツリーは自分の選択ノードのみで完結する.また他のエージェントの行動によって自分の行動選択肢が変わることがないため CFR の簡略化が可能になる.各エージェントの情報集合はエージェントの意思のみによってどの情報集合にもアクセス可能であり，よって

$\pi_{-i}^{\sigma}(I) = 1,\ u_i(\sigma|_{I \to a}, I) = u_i(\sigma, I),\quad \sigma_i(I)(a) = \sigma_i(I, I)$

とおける. このとき情報集合 I での行動 a の選択に伴うリグレットは次式のように置き換えることができる.

ここで情報集合 I' は情報集合 I で a を選択した場合に移行する次の情報集合（ノード）である. この操作によって CFR を後続ノードの期待利得と先行ノードの期待利得の差分のみで表すことができる.

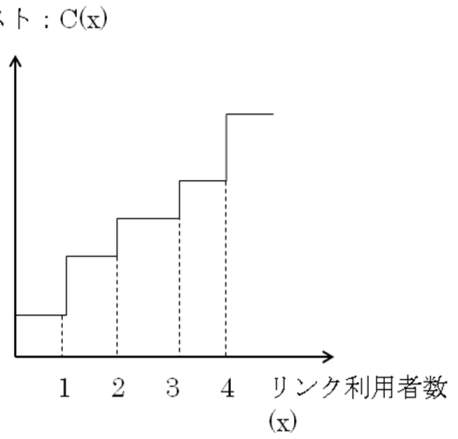


図-2. リンクコスト関数の例

$$\begin{aligned} R_i^T(I,a) &= \frac{1}{T}\sum_{t=1}^{T} \pi_{-i}^{\sigma^t}(I)\left(u_i(\sigma^t|_{I\to a}, I) - u_i(\sigma^t, I)\right) \\ &= \frac{1}{T}\sum_{t=1}^{T}\left(u_i(\sigma^t, I') - u_i(\sigma^t, I)\right) \\ &= \frac{1}{T}\sum_{t=1}^{T-1}\left(u_i(\sigma^t, I') - u_i(\sigma^t, I)\right) + \frac{1}{T}\left\{u_i(\sigma^T, I') - u_i(\sigma^T, I)\right\} \\ &= \frac{T-1}{T} R_i^{T-1}(I,a) + \frac{1}{T}\left\{u_i(\sigma^T, I') - u_i(\sigma^T, I)\right\} \quad (11) \end{aligned}$$

本研究で扱うリンクコスト関数は図-2 に示すようにリンクフローに関する増加関数を仮定する.ただし連続性や微分可能性を仮定する必要はない.具体的にはリンク l において $x \le x'$ を満たす交通量 x， x' に対して（x，x' は正の整数），$c_l(x) \le c_l(x')$ を満たすようなコスト関数であればよい.エージェントは k 経路のうちのどれかを選択する純粋戦略をとるため，リンクフローは整数値である. 起こりうる整数値に対してランダムな値を割り振り，ソーティングすることによって $c_l(x) \le c_l(x')$ を満たす非連続な関数を得ることができる．したがって，エージェントが知覚するリンクコストは $[c_{\min}, c_{\max}]$ の範囲で実現する跳び跳びの数値である．現実の交通環境に照らし合わせて考えれば，同じリンクでもエージェントは非常に空いている状況，渋滞している状況，そうでない状況と様々経験することを表しており，エージェントはどの状況に遭遇するかを前もっては知ることができないことを意味している.

(2) アルゴリズム

CFRを用いたリグレットマッチングの計算アルゴリズムについて，完全情報のケースと不完全情報のケースについて以下に示す．完全情報とは，エージェントはトリップの終了後，事後的に自分が利用しようと考えていたk経路の所要コストを（例えば，交通管制センター等からの情報を利用して）知り得る状況を表す．不完全情報とは自分の経験した経路のコストしか知りえない状況

を指す[3]．いづれの場合もエージェントは最少費用経路を事前には知らない．したがって，k経路探索アルゴリズムはあくまで利用する経路の事前情報に基ずくノードでの分岐を知るためだけに利用される．まず，初めに完全情報のケースを示す．

アルゴリズム 1:　完全情報リグレットマッチング

ステップ１：各エージェントの利得行列を計算する.エージェントの全ての選択経路の組み合わせについてリンク交通量を計算し，経路所要時間として利得行列の要素に格納する.

ステップ２：各エージェントの意思決定ツリーの各情報集合ノードにおいて行動選択確率を等分布に設定する.

ステップ３：式(1)にしたがって，経路選択確率を算出する.

ステップ４：各ノードでの期待利得を計算する.まず端末ノードでの期待利得を全エージェントの選択確率と利得行列から計算する.次に端末ノードから後ろ向きに先行ノードの期待利得を行動選択確率を用いて計算する.

ステップ５：各ノードでの各行動について(11)式で CFR を計算する.

ステップ６：CFR から式(8)を用いて行動戦略を求める.

ステップ７：行動戦略を新たな行動選択確率としてステップ３に戻る.

不完全情報を仮定するケースではエージェントは自分の経験した利得しか知りえないため，エージェントは 1 期の経験によってその経験した経路の情報しか更新することができない.その他の経路については過去の経験を参考に経路選択学習を行なっていく.この非同期的更新による CFR を以下のように定義する.

$$R_i^T(I,a)=\frac{T-1}{T}R_i^{T-1}(I,a)+\frac{1}{T}\left\{u_i(\sigma^T,I')-u_i(\sigma^T,I)\right\}\Big/\sigma_i^{T+1}(I)(a) \quad (12)$$

事後的完全情報下のモデルと比べて，右辺第 2 項を行動戦略で割り増ししているのがわかる.このことは，更新頻度が少ない，すなわち行動戦略で選ばれえる確率が少ないノードが選択され更新されるときには，あたかもその経路がずっと選ばれていた経路となるように更新することを意味する.

アルゴリズム 2:　不完全情報リグレットマッチング

ステップ１：各エージェントの意思決定ツリーの各情報集合ノードにおいて行動選択確率を等分布に設定する.

ステップ２：行動選択確率を経路に沿ってかけ合わせ，経路選択確率を算出する.

ステップ３：　得られた経路選択確率に基づいて各エージェントの選択経路を決定する.

ステップ４：全エージェントの選択経路に基づく実現リンクフローから各エージェントの所要時間を計算する.

ステップ５：各ノードでの期待利得を計算する.まず端末ノードでの期待利得を選択経路に限り所要時間を用いて更新する.次に端末ノードから後ろ向きに先行ノードの期待利得を行動選択確率を用いて計算する.

ステップ６：各ノードでの各行動について式(12)を用いて CFR を計算する.

ステップ７：式(8)を用いて CFR から行動戦略を求める.

ステップ８：行動戦略を新たな行動選択確率としてステップ 2 に戻る.

４．計算例

Zinkevich[10]の CFR モデルは，2 人ゼロ和ゲームにおける Nash 均衡への収束証明であり，非ゼロ和あるいは N 人ゲームに対するものではない．したがって，一般の交通ネットワーク問題にどこまで適用可能かは現段階では数値計算で確認するしか方法がない．この確認のため数値計算を行った.

図-3 に示す格子型の道路ネットワークを想定する.リンクは全て双方向とする.ドライバーごとの OD ペアはネットワーク内に始点ノードと終点ノードをランダムに設定した. 対象ネットワークを図-3 に示す．図中の番号はノード番号であり，トリップの起点または終点となるノードは○印で示す.

5 ドライバー，6×6 ネットワークにおいて，各ドライバーの経路選択肢数を$k=3$，繰り返し計算回数を 50 回とした場合の結果を示す.

OD ペア

	O	D
エージェント 1	12	20
エージェント 2	34	32
エージェント 3	13	8
エージェント 4	3	28
エージェント 5	22	13

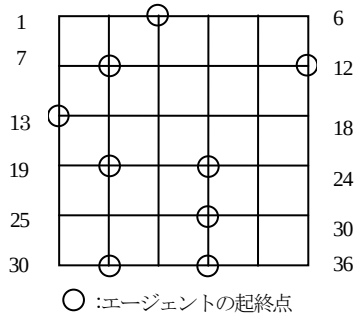


図-3 エージェントのODペア

この交通環境に対してドライバーの情報の利用可能性について2つの場合を想定し，数値実験を行なった.1つ目はドライバーが事後的に完全情報をもつと仮定する場合である.つまり，自身の意志決定ツリーの全ての末端ノードの自己の期待利得を知ることができ，その情報を

次期の行動選択に反映させることができる.

これに対し，不完全情報のケースではドライバーは自分の経験した利得しか知りえないと仮定する.ドライバーに毎期の繰り返し学習で自身の行動戦略に基づき選択を行なうが，結果として1つの経路を選択する.エージェントが得る道路情報はその経路に関する情報のみである.

（1）事後的完全情報のケース

k経路のうちリグレットを最小化するように選択された各ドライバーの最終的な選択経路を表-2に示す.

表-2 エージェントの選択経路結果

エージェント	1	2	3	4	5
選択経路	1	3	2	1	1

この計算結果についてNash均衡の確認を行なう.それぞれのドライバーについて他のドライバーの選択経路を固定し，その上で自身が各経路を選択した場合の所要時間を示したものが表３である.ハイライトが最短所要時間になっておりドライバーの最適応答を示している.

ドライバーの経路選択確率の推移を示したのが図-4である．ドライバー2については17期の学習，その他のドライバーについては5期前後の学習で経路選択確率が1または0に収束している．表２に示した経路が確率１で選択され，Nash均衡を得たことが分かる.

表-3 エージェントの最適応答

エージェント	経路１	経路２	経路３
1	100.6273	170.9668	146.5998
2	31.9489	65.9642	30.5801
3	18.7871	6.7496	34.9801
4	70.0281	137.8463	107.6996
5	40.3893	166.8323	165.8431

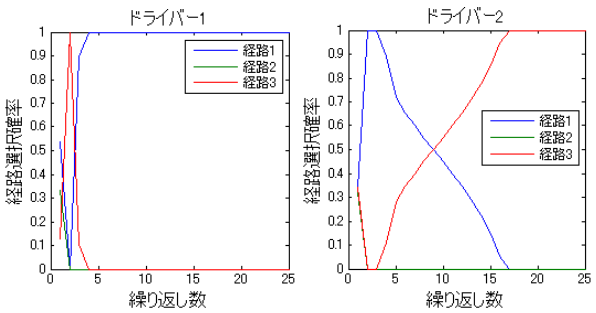


図-4　ドライバーの経路選択確率の推移

全ドライバーの全 CFR を足し合わせたものを総リグレットとして，その繰り返し計算の回数による変化をグラフに表わしたものが図-5 である．繰り返し計算の回数にしたがってリグレットはゼロに漸近する．繰り返し数と Nash 均衡に到達する確率の関係を調べたものを図-6 に示す．グラフを見ると，約 5 割の確率で 10 回という非常に少ない繰り返し計算回数で収束していることがわかる．また計算の繰り返し回数を増やしていけば Nash 均衡に到達する確率は 100%に漸近していく．

この他，ドライバーの選択肢数を変化させた場合，リンクコスト関数に増加関数の仮定を置かない場合，ドライバーに認知誤差がある場合について計算を行なったが，同様に Nash 均衡を得ることができた.

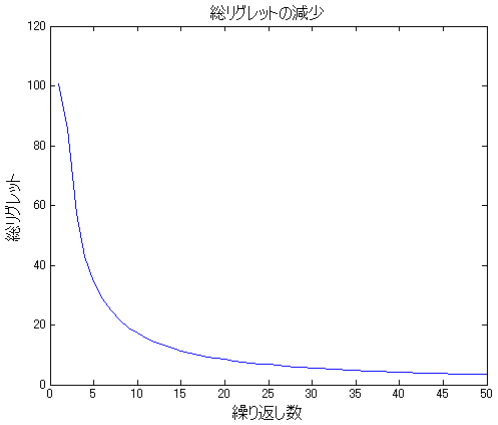


図-5　リグレットの推移

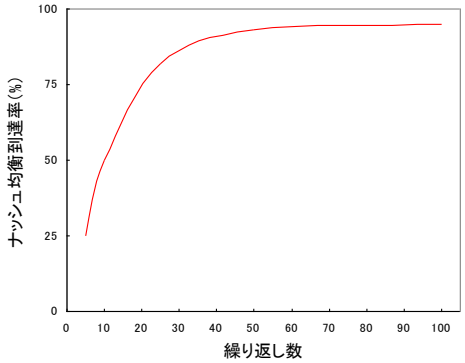


図-6　Nash 均衡到達率と計算回数（完全情報）

（2）不完全情報のケース

ドライバーの選択経路の結果については完全情報のケースと同様の結果が得られた．ドライバーの経路選択確率の推移を図-7に示す.

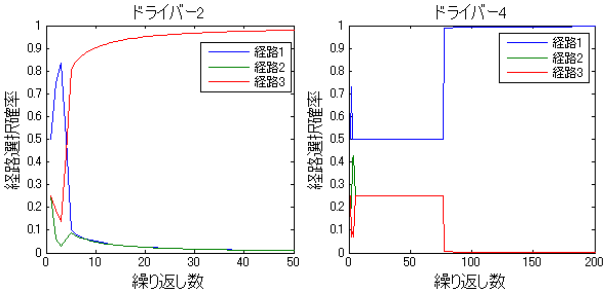


図-7　エージェントの経路選択確率の推移(不完全情報)

ドライバー1 と 4 は約 70 回の繰り返し計算で選択経路が収束し，その他のドライバーはおよそ 10 回の繰り返し計算で収束した．ドライバー4 の選択経路の推移のグラフの波形が角張っているが，これは CFR の不定期更新によるものである.またこの CFR の非定期更新によって 1 回 1 回の繰り返し計算の負荷を劇的に緩和することができる.完全情報のケースでは意志決定ツリー上の全てのノードについて CFR を計算したが，不完全情報のケースでは経験したノードの情報のみを更新するためである.これによって 1 回の繰り返し計算に要する時間が大幅に改善されるため，計算効率の点では望ましいが，収斂した交通状況がどのような均衡を表現するのかは明確ではない.

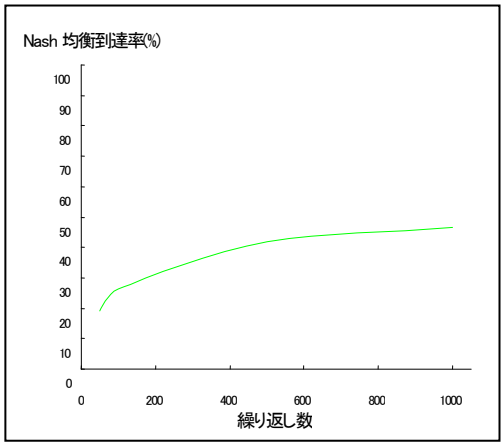


図-8　Nash 均衡到達率と計算回数 (不完全情報)

なぜならば，不完全情報のケースでは常に Nash 均衡に至るとは限らないからである.図-8 に繰り返し計算回数と Nash 均衡に到達する確率の関係を示す.完全情報のケースと比較して Nash 均衡を得られる確率が落ちているのがわかる.実際，この研究で対象にしているような不完全情報のケースでの Nash 均衡の存在についてはほとんどわかっていないので理論的に不整合な結果を得ているわけではない.この点の解明が今後の課題になる.

４．おわりに

本研究ではエージェントの経路選択行動をノードでのリンク選択行動を基準に展開ゲームとして再構築し，CFR を用いたリグレットマッチングによって Nash 均衡を求めることができることを示した.エージェントを個別に扱うことによって個々のエージェントの経験（リグレット）と意思決定の関係をノードでのリンク選択としてモデル化しただけでなく，エージェントの OD ペア，利用可能な経路選択肢の数など，個々のエージェントを差別化できることを示した.また非連続な関数を含む様々なリンクコスト関数に対しても適応可能であることから観測値を直接利用したリンクパフォーマンス関数などにも対応できるものと考えている.

エージェントは自分の選択した経路の情報しか知りえず，その他の経路については過去の情報を参考に経路選択学習を行なっていくという不完全情報下でのモデルについても Nash 均衡を求めることができたが，数値計算例からもわかるように，不完全情報下では常に Nash 均衡に収斂するとは限らない.完全情報でのモデルに比べて必要な計算回数は増加するものの，1 回ごとの計算負荷は劇的に緩和されるという点で今後の実用化に向けての有益な情報を提供している.

注１）非粒子 (non-atomic) モデルという用語は，Schmeidler[13]によって導入された．単一のプレイヤーでは何ら影響を及ぼさないが，各プレイヤーが集計化された集団になることによって影響力を持つとする考え方である．本研究では，その対語として atomic （粒子） model という用語を用いており，ゲーム理論のプレイヤーあるいはセルラーオートマトンのように個々の粒子の相互作用が次第にシステム全体にも影響を及ぼすダイナミクスを考えるモデルとして位置付けている．この発想のもと開発されたのがロスアラモス研究所の Transims と言えよう.

参考文献

1) T. Miyagi: A modeling of route choice behavior in transportation networks: an approach from reinforcement learning, WIT Press, Southampton, UK, pp.235-244, 2004.

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

3)宮城俊彦：経路選択のための知識・学習アルゴリズムの開発とその実用性に関する研究，平成18年度～19年度　科学研究補助金，基盤研究（C）研究成果報告書，2008.

4) T. Miyagi and M. Ishiguro: Modelling of route choice behaviours of car-drivers under imperfect travel information, Proc, of 14th International Conference on Urban Transport and Environment, WIT Press, pp.551-560, 2008.

5) Shoham, Y. and K. Leyton-Brown: Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations, Cambridge University Press, NY, US, 2009.

6) Papadimitriou, C.H.: The complexity of Nash squilibria, In: Algorithmic game Theory (eds: N. Nisan , T.

Roughgarden , E. Tardos , and V. Vazirani) , Cambridge University Press, USA, pp. 29-52.2007.
7) Hart, S. and A. Mas-Colell: A simple adaptive procedure leading to correlated equilibrium , Econometrica, 68(5), 1127-1150, 2000.
8) Peyton Y.H.: Strategic Learning and Its Limit, Oxford University Press, 2004.
9) Fudenberg, D. and J. Tirole: Game Theory, The MIT Press, London, England, 1991.
10) Zinkevich, M. et al.: Regret minimization in games with incomplete information, In Proceeding of the Annual Conference on Neural Information Processing Systems, 2007
11) Fudenberg, D. and D.K. Levine: The Theory of Learning in Games, The MIT Press, USA, 1998.
12) Martins, E.Q.V. and Pascoal, M.M.B: A new implementation of Yen's ranking loopless paths algorithm, A Quarterly Journal of Operations Research, Springer Berlin, Heidelberg, pp.121-133, 2003.
13) Schmeidler D. : Equilibrium points of nonatomic games, Journal of Statistical Physics, Vol.7, No.4, pp.295-300, 1973.

Regret Matchingを用いた経路選択行動分析

―不完全交通情報を仮定した粒子モデル―

宮城俊彦，石黒雅彦

本研究では個々のエージェントを自律的なエージェントとして定義し，それぞれがその直面する環境下で利得最大化行動を学習していく繰り返しゲームとして交通ネットワークフローを求めるモデルを提案する. 個々のエージェントは個別のODペアを持ち，利用可能な情報に応じて経路選択肢の集合を決定し，その選択肢の中から経路を選択する. 交通環境はそれぞれのエージェントの非協力的な選択によって変化するので，個々のエージェントは交通環境を正確には知りえないが，経験した所要時間を通して交通環境を学習し適応的に経路選択を行なう. 本研究はエージェントの経路選択をノードでのリンク選択問題のシークエンスとして定式化し，リグレットマッチングによって環境学習をさせることによってNash均衡が実現できることを示したものである.

AN ANALYSIS OF ROUTE CHOICE BEHAVIOR BY A REGRET MATCHING MODEL -ATOMIC MODEL WITH IMPERFECT TRAVEL INFORMATION-

By Toshihiko MIYAGI and Masahiko ISHIGURO

In classical traffic assignment procedures, drivers have been treated as a collection of an infinite number of particles that have perfect information and perfect foresight. On the contrary, in this paper, each driver was modeled as an individual decision-maker with imperfect information on his travel environment. We have adopted extensive games as the basic framework to describe the process of route choice behavior of naive drivers, and analyzed the game by using "Counterfactual Regret". Each driver's decision-tree was constructed by K-shortest path algorithm. For link costs, monotone increasing, but non-differentiable functions were assumed. We have shown through numerical experiments that the algorithm developed here achieved Nash equilibrium.