2I1-OS5-7

情報提供戦略のQ学習による交通ネットワーク流の制御

Learning Strategy of Effective Information Services for Indirect Traffic Control

内田英明[*1]　荒井幸代[*2]
Hideaki Uchida　Sachiyo Arai

[*1]東京大学大学院工学系研究科システム創成学専攻
Department of Systems Innovations, School of Engineering, the University of Tokyo

[*2]千葉大学大学院工学研究科
Graduate School of Engineering, Chiba University

We are concerned with Braess's paradox of traffic flow. Braess's paradox is caused by selfish routing of each driver who makes shortsighted decision based on currently available information. Because such a paradox is observed everywhere, there have been numerous researches to solve it; i.e., game or decision theoretic approaches. These previous researches mainly focused on the driver's decision making to resolve this paradox. In other words, the previous researches where input of decision maker is given and they seek a rational strategy to avoid being paradox. While, we focused on information services side to make traffic flow desirable. Thus, in our research, the strategy of each driver is given, and we pursue how to control the driver's input to make traffic flow more easily.

We introduce Q-learning to acquire the strategy for solving the paradox. In this paper, we show the results that information center, which distributes information to the drivers, successfully acquire the control traffic. In addition, we analyze the situation of traffic flow that requires the information service strategy.

1. 本研究の背景と目的

近年, 高度交通システム（Intelligent Transport Systems）普及に伴い，事故や渋滞情報を運転者にリアルタイムで配信することが可能になってきている．しかし，情報提供が渋滞緩和に貢献しているのかについては十分に議論されているとはいえない．一般に，運転者に提供される渋滞情報は,「ある区間における現在所要時間」といった限られた地域の部分情報である．一方，情報を受け取った運転者は，提供された時点での情報に基づいて，選択可能な各経路の所要時間を予測し，経路を選択する．しかし，この限られた情報に基づいた運転者の意思決定には，情報の遅れや，他運転者の行動予測は陽に加味されていないことが多く，その結果，系全体からみた交通流の配分が最適にならない [加藤 88]．この現象は Braess のパラドクスとして説明される．また，この現象を回避するため，情報が交通状況に与える影響の評価についてでも研究がなされている [吉井 00][大口 03].

本研究では，交通流を制御する主体として渋滞情報センタ（以下，情報センタ）と運転者の２つをモデル化し，強化学習を適用する．センタはネットワーク全体の状態入力から経路選択の指針をトップダウンに提供し，運転者はその情報を基に行動することによって，動的に変化する交通流に適応してルーティング政策を変化させる制御モデルを提案する．

以下，２章で対象問題をモデル化し，問題設定とアプローチを示す．３章で提案手法，４章で実験結果とその考察を行い，５章で結論と今後の課題を述べる．

2. 問題設定

2.1 交通流配分問題

交通流制御の目的はネットワーク全体のコスト（旅行時間）を最小化することに帰着する．このとき，ネットワークが最適化された状態を実現する交通流の配分はシステム最適配分 (以下 SO : System Optimum assignment) と呼ばれる．一方，各々の運転者が自己の経路選択行動を最適化した結果到達する均衡状態は利用者均衡配分 (以下 UE : User Equilibrium assignment) と呼ばれる．

均衡配分理論における仮定　Wardrop は 1952 年に，これらを数理的に記述した均衡配分理論の基本概念として Wardrop の原理を提唱した．Wardrop の原理第１原則を説明する上で交通ネットワーク上における運転者の行動，及びシステムの性質として以下の仮定をおく．

1. 運転者の行動
   - 全ての運転者は，ネットワーク上の各リンク（道路）の旅行時間に関する情報を取得可能である（完全情報）
   - 全ての運転者は旅行時間が最短の経路を選択する（合理的選択）
2. システムの性質
   - 各リンクの旅行時間は交通量のみに従属な単調増加関数である
   - システムは定常状態である

このとき，十分な時間経過の結果として次の均衡状態が実現される．

Wardrop の第１原則（等時間原理）　ある OD ペアについて，利用されている経路の旅行時間は全て等しく，利用されない経路の旅行時間はそれよりも大きいか，せいぜい等しい．これが上記の UE 状態の定義であり，どの運転者も経路変更によるインセンティブ（旅行時間の短縮）を持たない状態である．

連絡先: [*1]内田英明：uchida@save.sys.t.u-tokyo.ac.jp
[*2]荒井幸代：arai@tu.chiba-u.ac.jp

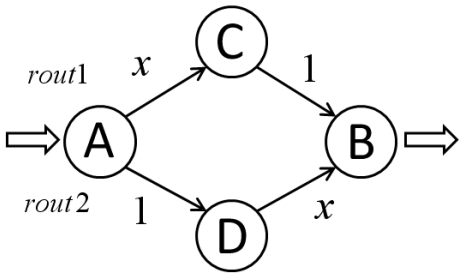

図 1: ショートカット追加前のネットワーク

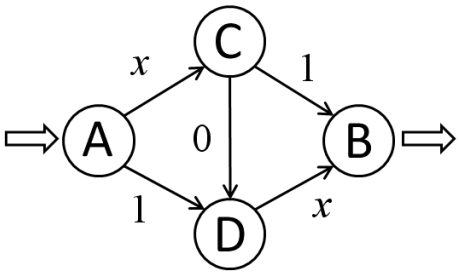

図 2: Braess ネットワーク

2.2 交通ネットワークのパラドクス

前節より，交通ネットワークの全体が与えられた場合，Wardrop の第一原則を満たすように，交通流が一定の均衡状態，つまりは UE 状態に至ることが示された．しかし，この UE 状態とシステムとして最も望ましい状態である SO 状態は必ずしも等価ではない．

また，なかには特定の条件下で道路投資によるリンク容量の増加が，UE と SO の乖離を助長する現象が知られている．特に，Pigou-Knight-Downs のパラドクス [竹内 97]，Downs-Thomson のパラドクス [竹内 97]，及び Braess のパラドクス [亀田 05][Braess 09] は交通工学や経済学の分野においてしばしば議論される問題である．以下では，本論文で扱う Braess のパラドクスについて説明する．

Braess のパラドクス　次に Braess のパラドクスについて説明する．まず，図 1 のような交通ネットワークを考える．ここで，起点ノード A から終点ノード B までの経路は rout1（$e_{\mathrm{AC}}\, e_{\mathrm{CB}}$），rout2（$e_{\mathrm{AD}}\, e_{\mathrm{DB}}$）の 2 本であり，交通量はそれぞれ f^1_{AB}，f^2_{AB} である．ネットワーク上の交通量が $f^{\mathrm{AB}} = f^1_{\mathrm{AB}} + f^2_{\mathrm{AB}} = 1$，$e_{\mathrm{AC}}$ 及び e_{DB} の旅行時間関数 $t_{\mathrm{AC}} = x_{\mathrm{AC}}$，$t_{\mathrm{DB}} = x_{\mathrm{DB}}$，かつ e_{AD} 及び e_{CB} の旅行時間関数 $t_{\mathrm{AD}} = t_{\mathrm{CB}} = 1$ とする．このとき，SO 及び UE は共に $x_1 = 0.5$，$x_2 = 0.5$ となり，平均旅行時間は 1.5 である．

しかし，図 2 のように，リンク容量が大きく旅行時間も短いショートカット e_{CD}（旅行時間関数 $t_{\mathrm{CD}} = 0$）を追加したネットワークを考える．このとき新たな経路 rout3（$e_{\mathrm{AC}}\, e_{\mathrm{CD}}\, e_{\mathrm{DB}}$）が追加される．$f^{\mathrm{AB}} = f^1_{\mathrm{AB}} + f^2_{\mathrm{AB}} + f^3_{\mathrm{AB}} = 1$ とすると SO は依然として $f^1_{\mathrm{AB}} = 0.5$，$f^2_{\mathrm{AB}} = 0.5$，$f^3_{\mathrm{AB}} = 0$ となり，平均旅行時間は 1.5 であるが，UE は $f^1_{\mathrm{AB}} = 0$，$f^2_{\mathrm{AB}} = 0$，$f^3_{\mathrm{AB}} = 1$ で平均旅行時間は 2 である．これは，ノード C，D のどちらを経由する運転者にとっても rout3 が必ず最短経路となるためであり，新たな投資によって既存のネットワークにリンクを付与することが，ネットワークのパフォーマンス低下につながってしまうことから逆説的であるといえる．

2.3 対象問題のモデル化とアプローチ

交通流は「センタ」と「運転者」の 2 種類の意思決定主体によって制御されるものと考え，本研究は図 3 に示す学習モデルを導入し，環境を交通流とした Q 学習によって定式化する．

ここで，センタを学習主体であるエージェント，ネットワーク上の交通流を環境とする．エージェントは環境から状態観測として交通情報を収集し，意思決定の結果としての方策に従い，運転者に情報提供を行う．そして，運転者はこの情報を基に経路選択を行い，その結果として実現される運転者の平均旅行時間を報酬としてエージェントに与える．このとき，エージェントの行動は運転者を介して環境に作用するため，運転者

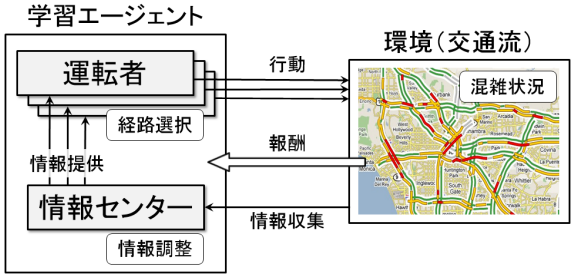

図 3: センタと運転者の学習モデル

の意思決定モデルについても評価することが可能である．また，学習中のネットワークの交通量は不変とした定常状態を考える．これは，交通量を一定とすることで MDP 環境を仮定できるため，Q 学習の最適値への収束性を保証することができるからである．

3. 提案手法 -情報提供戦略の学習-

3.1 情報センタの強化学習

センタは式 (1) に従って Q 値を更新する．このとき，状態，行動は次のように離散化して表現する．状態 $s \in \mathcal{S}$ は交通ネットワークにおけるリンクの密度 $d_l(l:$ リンクのラベル) を離散化した値で表現する．$0 \leq d_n \leq 1$ であり，0.1 刻みで 10 段階の評価とする．行動 $a \in \mathcal{A}$ は運転者が分岐ノードに到着した際，センタが情報提供する配分率を変動させることと表現する．時刻 t において，ある分岐ノード i における総流入交通量を x_i^t，対応する制御リンク $j(j = 1, 2, \cdots, n)$ の交通量を $x_{i,j}^t$ とする．ただし，n はノード i の出次数である．このとき，行動集合 $\mathcal{A} = \{$ 減少, 一定, 増加 $\}$ とし，それぞれの行動を図 4 に記述する．また，旅行時間の最小化が目的であるため，負の報酬として平均旅行時間を与えることとする．

$$Q(s,a) \leftarrow Q(s,a) + \alpha \left\{ r + \gamma \max_{a' \in \mathrm{A}(s')} Q(s',a') - Q(s,a) \right\} \quad (1)$$

3.2 運転者の意思決定

ASEP モデル　運転者の前進は ASEP[西成 02] によって記述する．ただし，ASEP では最高速度が 1[cell/split] に限定されるため，より現実的な設定とするためモデルの拡張を行う．まず，最高速度を 4[cell/unit] とし，更新は 4[split] を 1[unit]

- 減少：$u^{t+1} = u^t - 0.01$
- 一定：$u^{t+1} = u^t$
- 増加：$u^{t+1} = u^t + 0.01$

（ただし，配分率 $u^t = \frac{x_{i,j}^t}{x_i^t}$ とする）

図 4: 行動集合

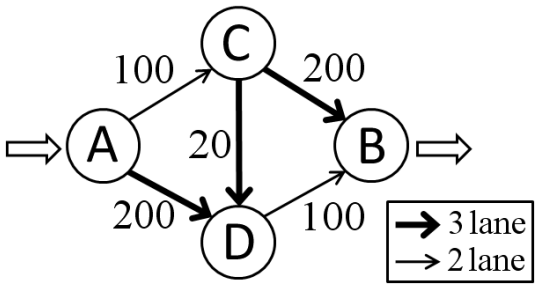

図 5: Braess のパラドクスを持つネットワーク

として行い，1[split] ごとに ASEP モデルと同様のルールを適用する．ASEP との相違は，一度停止した運転者はその unit が終了するまで停止し続けるという点である．これは「スロースタートルール」と呼ばれ，一度停止した車は発進が鈍くなるという慣性力の性質を反映するためのルールである．また，パラメータとして前進確率を 0.99 と設定する．車両密度以外の影響を排除するため，車線変更はおこなわない．

提供情報の利用　均衡配分理論の仮定 [加藤 88] によれば，運転者が各リンクの旅行時間について情報を取得した場合，交通ネットワークは利用者均衡配分 (UE) 状態に陥る．これに対し，センタの学習モデルによって獲得された情報提供戦略は，運転者に対し必ずしも最短経路の提示を行うとは限らない．

そこで，センタの情報提供率 $p(0 \leq p \leq 1)$ を導入する．これによって，確率 p でセンタの指示した経路に従い，確率 $(1-p)$ でその時点での最短経路を選択する，という運転者の意思決定を反映する．これは運転者の中に，センタの指示に従って経路選択をする経路浮動層が p の割合で存在することを表す．本来選択するはずであった経路とは別の経路に指示された場合，経路浮動層は通行料の割引など，何らかのインセンティブによって，指示を受け入れると考えることができる．

4. 計算機実験と考察

4.1 対象ネットワーク

実験対象のネットワークとして図 5 のような有向グラフを考える．この交通ネットワークでは Braess のパラドクス [Braess 09] が発生することが知られている．このパラドクスは，ショートカットリンク CD の存在により，UE の平均旅行時間がネットワークの交通時間が最小化されたシステム最適配分 (以下 SO) での旅行時間に比べ大きくなってしまうというものである．つまり，運転者の自由度が大きくなった状態は，必ずしもネットワーク全体のパフォーマンス向上にはつながらないため，センタによる制御を必要とすると考えられる．

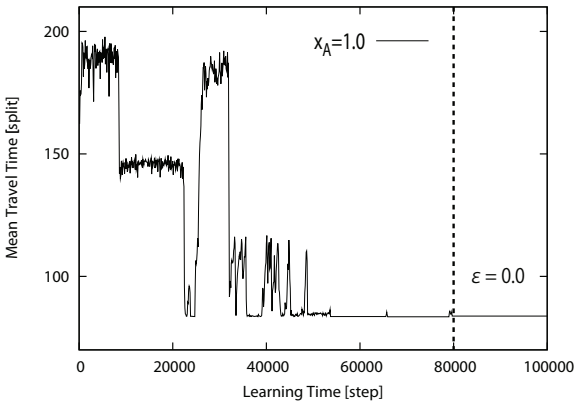

図 6: 学習のプロセス

4.2 予備実験

予備実験として，流入交通量一定の環境（定常状態）について，Q 学習の適用による最適値への収束性を確かめる．交通流は CA モデルにより再現し，500[split] のシミュレーションを学習 1[step] とする．ここでは分岐ノード A, C において経路選択に関する情報提供を行うため，リンクの密度を離散化した値の組み合わせによって状態表現を行う．状態集合 $\mathcal{S} = \{s_{\mathrm{AC}}, s_{\mathrm{CD}}\}$ の状態数は 10^2，行動集合 $\mathcal{A} = \{a_{\mathrm{C}}, a_{\mathrm{D}}\}$ の行動数は 3^2 となる．Q 値は初期値を 0，学習のパラメータは学習率 $\alpha = 0.03$, 割引率 $\gamma = 0.9$ とし，行動選択には ϵ-greedy 選択 ($\epsilon = 0.3$) を採用する．この設定は断りのない限り以降の実験でも同様とする．また，運転者に対する情報提供率 $p = 1$ とする．これはセンタの情報提供に必ず従うという状態であるが，学習の最適性を評価する必要があるためこの設定を用いる．

流入交通量 $x_{\mathrm{A}} = 1.0$[台/split] における学習結果を図 6 に示す．横軸は学習回数，縦軸は学習の更新 100[setp] 毎の平均旅行時間である．$\epsilon > 0$ であるため収束後も平均旅行時間にノイズが発生しているが，収束性を確かめるため実験では 80000[step] 以降は決定的な行動選択 $\epsilon = 0$ とした．このとき，学習終了時のそれぞれの平均旅行時間は最適値には収束しないが，準最適値であることが確認された．これは行動集合の要素数が少ないことが原因であると考えられる．

4.3 情報提供戦略の適応的利用

予備実験から，Q 学習によって獲得された情報提供戦略の有効性を示した．しかし現実の交通流は日々変化するため，流入交通量を観測してから学習を適用することは困難である．そこで，事前にオフラインで学習し，観測された流入交通量に応じた戦略を選択する方法を用いる．このとき，戦略とは学習によって関数近似された Q 値表を指す．

また，流入交通量は連続量であることから，全ての状態に対して学習を行うことは困難である．そこで，事前学習で獲得した離散的な Q 値表を補間することを考える．補間の方法として次の 3 手法を比較する．(1) 0 次補間：観測点から最も近い Q 値表を採用する．(2) 重ね合わせ：隣接する 2 値点の Q 値表の平均をとる．(3) 1 次補間 (線形補間)：隣接する 2 値点の Q 値表を，観測点からの距離によって重み付けする．このとき，補間された Q 値表の再学習は行わず，決定的行動選択による解探索を行う．実験設定は流入交通量が $0 < x_{\mathrm{A}} \leq 2.0$ の範囲において，0.1[台/split] 刻みで事前学習を行い，0.01[台/split] 刻みに補間を行った．

補間手法ごとの実験結果を図 7 に結果を示す．ここで縦軸は，1 次補間の結果に対する他手法の平均旅行時間の割合，横軸は流入交通量である．補間手法としては 1 次補間の精度が

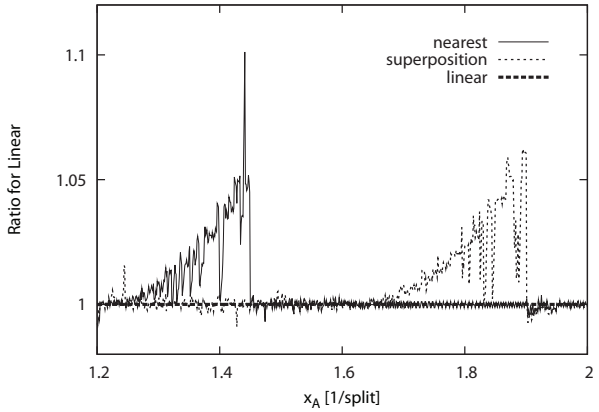

図 7: 補間手法による効果

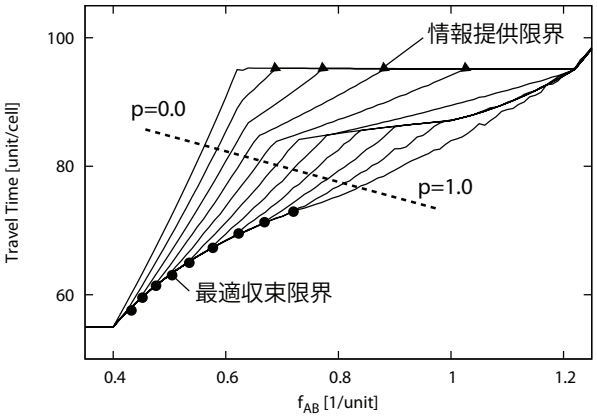

図 8: 情報提供率による平均旅行時間の変化

安定して最も高く，実際に学習した状態入力でなくとも，Q 値表を基に実時間の解探索を行うだけで準最適解が得られることがわかった．これは，各状態間が連続な位相構造であるためと考えられる．

4.4 情報提供の有効性

提案手法では全ての運転者が車載器を搭載している状況を考えている．このとき，$p = 1$ で SO 状態が実現されることはこれまで述べてきたとおりであるが，$p = 0$ では UE 状態が実現される．これは全ての運転者が車載器を持たない状況下で仮定される均衡状態と等価であり，センタの情報提供効果を著しく低下させる．また，社会的に望ましい状態は SO であるため，p の値は大きいほど良い制御であるが，p は同時に運転者の負担に関するパラメータでもある．そこで，情報提供率を $0 \leq p \leq 1.0$ の範囲で 0.1 刻みに変化させた場合の学習結果を図 8 に示す．ここで，縦軸は平均旅行時間，横軸は流入交通量である．

このとき，$x_{\mathrm{A}} \leq 0.4$ 及び $1.2 \leq x_{\mathrm{A}}$ の範囲では p による有意な差はなく，SO と UE が等価であった．一方，$0.4 < x_{\mathrm{A}} < 1.2$ の範囲では p によって平均旅行時間の違いが確認された．これは，この範囲において情報提供による制御が可能であることを示している．そこで，平均旅行時間の変化率が不連続となる特徴的な 2 点を以下に定義し，考察を加えた．

1. 最適収束限界 f^{SO}
   SO と等価な平均旅行時間を実現できる限界点

2. 情報提供限界 f^{UE}
   UE と等価な平均旅行時間に陥ってしまう限界点

最適収束限界 f^{SO} は，本来 $p = 1$ の状況下で実現される SO 状態をより小さい情報提供によって実現する境界である．p の増大に伴い SO 状態を維持する範囲が大きくなることから，f^{SO} は p についての関数として表現することができると考えられる．また，情報提供限界 f^{UE} は，情報提供を行っているにもかかわらず $p = 0$ と同等の UE 状態に陥ってしまう境界である．これは情報提供効果の低下に対する指標となり，やはり p についての関数として表現することができると考えられる．また，情報提供限界に関しては $p = 0.5$ 以上では確認されないことから，低い情報提供率を保っている場合には交通状況の変化によって容易に UE 状態へ陥ってしまう危険性が示唆されている．

5. まとめと今後の課題

本研究では，はじめに，Braess のパラドクスが発生する交通ネットワークにおいて，情報提供による制御方法を提案し，強化学習の特徴を利用した Q 値表の補間によって戦略を連続な状態に対しても拡張可能であることを示した．また，情報提供におけるコストを考慮したネットワークの最適化に向け，センタの情報提供率 p を変化させた場合の平均旅行時間の変化を考察した．

今後は，情報提供率と平均旅行時間との間に成り立つ関係がネットワークの形状に依存すると予想されるため，ショートカットリンクの距離や車線数を変動させた場合の計算機実験を行う必要があると考えられる．また，運転者の意思決定モデルを更に階層化し，環境の多様性を表現することも課題である．

参考文献

[加藤 88] 加藤晃, “ 交通量配分理論の系譜と展望 ”, 土木学会論文集, No. 389, IV-8, pp. 15-27, (1988).

[吉井 00] 吉井稔雄, 桑原雅夫: “ リアルタイム交通情報の提供効果 ”, 土木学会論文集, No. 653/IV-48, pp. 39-48 (2000).

[大口 03] 大口敬, 佐藤貴行, 鹿田成則: “ 渋滞時の代替経路選択行動に与える交通情報提供効果 ”, 土木計画学研究・講演集, 30 巻, No. II, p. 99, (2003).

[竹内 97] 竹内健蔵: “ 中小都市交通ネットワークにおける交通政策の視点 : Downs-Thomson のパラドックスの検証 ”, 経済と社会, 25, pp. 37-53 (1997).

[亀田 05] 亀田壽夫: “ 独立分散最適化によるネットワークにおける性能劣化パラドックス ”, 日本オペレーションズ・リサーチ学会和文論文誌, 48 巻, pp. 26-47 (2005).

[Braess 09] D Braess, A Nagurney, T Wakolbinger: “ On a Paradox of Trafffc Planning ”, Transportation Science, vol. 39, No. 4, pp. 446-450, (2009).

[西成 02] 西成活祐: “ 交通流のセルオートマトンモデルについて ”, 応用数理, vol. 12, No. 2, pp. 26-37, (2002).