

# 制御策に対する運転者の意思決定が交通流に与える影響

Influence of Driver's Decision Making by Traffic Control on Traffic Flow

間瀬 雄太\*<sup>1</sup> 荒井 幸代\*<sup>2</sup>  
Yuta Mabuchi Sachiyo Arai

\*<sup>1,2</sup>千葉大学大学院工学研究科建築都市科学専攻  
Department of Architecture and Urban Science, School of Engineering, Chiba University

We are concerned with the traffic control. For the desirable traffic control, we need to control traffic flow from microscopic view as well as macroscopic view. Though it is essential to consider controlling from microscopic view, most of previous researches about the traffic control mainly focused on macroscopic point of view. We define that the side to control from macroscopic view is the information center, and the side to control from microscopic view is the drivers. In our research, we pursue the effective direction of the information center considering the driver's property of individual decision.

We introduce Q-learning to acquire the strategy of the drivers under the direction of the information center. In this paper, we show the influence of the driver's learned decision under the loose control on traffic flow. In addition, we analyze the influence on traffic flow when we changed the strategy of the information center.

## 1. 本研究の背景と目的

交通渋滞が与える経済的損失や環境への負荷は深刻な問題であり、交通流を制御し渋滞を改善するための方策が重要視されている。近年の道路交通システムの発展により、事故や渋滞情報をリアルタイムで収集し運転者に配信することが可能になったため、各運転者は提供される情報に基づいて、選択可能な各経路の所要時間を推測し、経路を選択できるようになった。しかし、情報が示す最短経路に車両が集中し、逆にその経路が混雑する可能性は十分考えられる。道路のネットワーク構造によっては「提供されない場合」より旅行時間が増大する Braess のパラドクスが生じることが知られている。これは、運転者が局所的な情報下で、利己的かつ近視眼的に経路を選択することが、交通渋滞を発生させる原因の一つになっていることを示している。道路のネットワーク全体を巨視的にとらえ、最適な交通流を実現することが必要であるが、この制御策の下では個々の運転者の旅行時間が必ずしも最短になるわけではない。したがって制御策に運転者が従うとは限らないことから運転者の利己的な特性を考慮した制御策が求められているといえる。

そこで、本研究では、Braess のパラドクスを持つネットワークを対象として、交通流を制御する主体を渋滞情報配信センタ（以下、センタ）と運転者の二つに分類し、それぞれの意思決定を反映した交通流制御問題を対象とする。センタ側の情報提供戦略の獲得方法については、Q 学習を用いた文献 [内田 10] の定式化したものが、実際に運転者に与える情報提供内容は、運転者の意思決定特性を考慮していくつかのケースに分類して実験した。情報提供に対する運転者の意思決定が交通流に与える影響を考察し、この知見に基づいて、運転者の利己的な特性を踏まえた効果的な情報提供法を確立することを目的とする。

## 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を最小化することに帰着する。このとき、ネットワーク全体の交通流が最

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、つまり、運転者が始点ノードで経路を選択する際に、選択可能な経路で必要とする旅行時間を情報として取得し、その時点での最短時間となる経路を選択した状態は利用者均衡配分（以下 UE : User Equilibrium assignment）と呼ばれる\*<sup>1</sup>。

### 2.2 交通ネットワークのパラドクス

運転者に対して交通ネットワークの全体の交通量と各経路の所要時間に関する情報が与えられた場合に、交通流が一定の均衡状態、つまりは UE 状態に至る。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交通容量の増加が、UE と SO の乖離を助長するような現象も知られている。本研究では、この現象の代表的な例として Braess のパラドクス [亀田 05][Braess 09] が生じるネットワークを対象とする。以下で、Braess のパラドクスについて説明する。

Braess のパラドクス 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 route1 ( $e_{AC} e_{CB}$ )、route2 ( $e_{AD} e_{DB}$ ) の 2 本であり、交通量はそれぞれ  $f^1$ 、 $f^2$  である。ネットワーク上の交通量が  $f = f^1 + f^2 = 1$ 、 $e_{AC}$  及び  $e_{DB}$  の旅行時間関数  $t_{AC} = x_{AC}$ 、 $t_{DB} = x_{DB}$ 、かつ  $e_{AD}$  及び  $e_{CB}$  の旅行時間関数  $t_{AD} = t_{CB} = 1$  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  $t_{CD} = 0$ ）を追加したネットワークを考える。このとき新たな経路 route3 ( $e_{AC} e_{CD} e_{DB}$ ) が追加される。 $f = f^1 + f^2 + f^3 = 1$  とすると SO は依然と

\*<sup>1</sup> 均衡配分理論の仮定 [加藤 88]: 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以上である。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: \*<sup>1</sup>間瀬 雄太, y.mabu@chiba-u.jp

\*<sup>2</sup>荒井 幸代, arai@tu.chiba-u.ac.jp

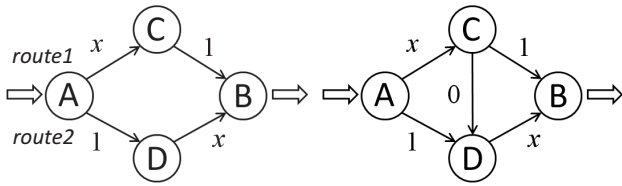


図 1: 1OD4 リンクのネット ワーク

して  $f^1 = 0.5, f^2 = 0.5, f^3 = 0$  となり、平均旅行時間は 1.5 であるが、UE は  $f^1 = 0, f^2 = 0, f^3 = 1$  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 route3 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

図 3 は、Braess のパラドクスが発生する交通ネットワークにおける運転者の学習モデルである。各経路は離散化したセルで表し、図 3 のネットワークにおけるリンクの添え字はリンク上のセル数を示している。文献 [内田 10] がセンタ側の学習と、その制御効果を示したのに対して、本研究では運転者側の意思決定戦略を強化学習によって獲得した場合の効果を示す。以下に、本研究で用いる情報提供戦略の学習アルゴリズムの概略を示す。

**センタの学習** センタはネットワーク全体の状態観測と運転者全体の旅行時間を報酬とした強化学習によって分岐ノード A と C における各リンクの交通量の増減に関する戦略を獲得する。センタは増か減を運転者に知らせてこれに従った行動をしてもらうことによってパラドクスを解消する。

ここでは、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している。エージェントは状態入力として各リンクの交通量を観測し、方策に従い、運転者に情報提供を行う。ここで、情報提供とは、以下に示す行動に相当する。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。なお、運転者の前進は ASEP [西成 02] を拡張させて記述した\*2。

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、報酬は以下のように離散化して表現する。

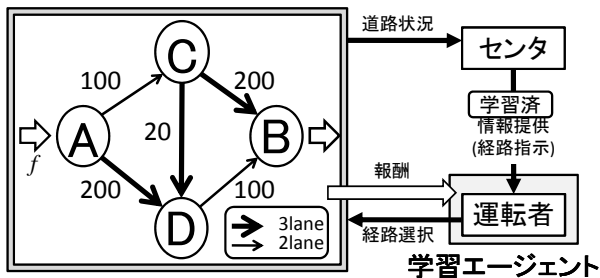


図 3: Braess のパラドクスが生じるネットワーク交通流制御モデル：学習済の情報提供戦略下での運転者の学習段階

\*2 3.2 節 [内田 10] の設定に従い、最高速度を  $4[\text{cell/unit}]$  とし、更新は  $4[\text{split}]$  を  $1[\text{unit}]$  として行って、「スロースタートルール」と呼ばれる、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ルールを用いた。また、車線変更はないとした。

- 減少:  $u_l^{t+1} = u_l^t - 0.01, u_k^{t+1} = u_k^t + 0.01$
  - 一定:  $u_l^{t+1} = u_l^t, u_k^{t+1} = u_k^t$
  - 増加:  $u_l^{t+1} = u_l^t + 0.01, u_k^{t+1} = u_k^t - 0.01$
- (ただし、配分率  $u_l^t = \frac{x_{v,\ell}^t}{x_v^t}, u_k^t = \frac{x_{v,k}^t}{x_v^t}$  とする)

図 4: 行動集合

1.  $e_{AC}, e_{CD}$  の密度を状態観測  $s$  とし、行動  $a$  によりノード A, C における配分率  $u_A, u_C$  を増減させ配分率を決定する。(  $e_{ij}$ : ノード  $i, j$  間を接続するリンク )
2. ノード A, C に到着した運転者は、 $u_A, u_C$  に応じて経路が振り分けられ、センタの指示した経路を選択する。
3. 運転者の平均旅行時間を負の報酬  $r$  として与え、式 (1) に従い、Q 値を更新する ( $\alpha$ : 学習率,  $\gamma$ : 割引率)。

図 5: センタの学習アルゴリズム

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

**状態** 状態  $s \in \mathcal{S}$  はリンクの密度  $d_\ell(\ell$ : リンクのラベル) を離散化したものである。  $0 \leq d_\ell \leq 1$  であり、0.1 刻みで 10 段階の評価値が状態入力として与えられる。

**行動** 行動  $a \in \mathcal{A}$  は、運転者が分岐ノードに到着した際、二つの経路を選び、選んだ二つの経路の間で、選択すべき経路として情報提供される確率を増減させることと表現する。例えば、時刻  $t$  において、ある分岐ノード  $v$  における総流入交通量を  $x_v^t$ 、ノード  $v$  に対応する二つの制御リンク  $\ell, k (\ell \neq k)$  の交通量を  $x_{v,\ell}^t, x_{v,k}^t$  とする。このとき、行動集合  $\mathcal{A} = \{ \text{減少}, \text{一定}, \text{増加} \}$  とし、それぞれの行動を図 4 のように記述する。

**報酬** 運転者の平均旅行時間を報酬  $r$  として与える。ただし、対象問題が平均旅行時間の最小化であるため、報酬  $r$  は平均旅行時間の負の値とした。

上記した、状態、行動、報酬に従うセンタの学習アルゴリズムを図 5 に示す。

### 3. 提案法 運転者の意思決定の導入

センタ側の学習後の戦略に対して運転者側は強化学習による意思決定を導入し、その挙動を計算機実験により観察する。また、情報提供の方法を変えた場合の実験も行い、それぞれの方法が意思決定に与える影響についても検証し評価する。さらに、運転者の利己的な意思決定特性を考慮した場合でも平均旅行時間を SO にする情報提供法の考察を行う。図 3 で示したネットワークを対象として以下に本研究で用いた情報提供法と運転者の意思決定を導入したモデルを示す。

1. 学習を行わずランダムな経路変更をする運転者を各経路にランダムに流入させる操作を十分な時間行い、各経路上に存在する運転者数を一定にし、この状態を初期状態として学習エージェントである運転者の流入を開始する。
2. Case1 運転者は、ノード A 到着時に、状態入力  $s \in S$  として経路候補集合  $\{e_{AC} \rightarrow e_{CB}, e_{AD} \rightarrow e_{DB}, e_{AC} \rightarrow e_{CD} \rightarrow e_{DB}\}$  のうちその時点での最短経路と  $u_A, u_C$  により決まるセンタの指示経路の二つの経路情報を観測し、行動  $a \in A$  として経路候補集合から経路選択をする。
 

Case2 運転者は、ノード A 到着時に、状態入力  $s \in S$  としてリンク候補集合  $\{e_{AC}, e_{AD}\}$  のうちその時点での最短リンクと  $u_A$  により決まるセンタの指示リンクの二つの経路情報を観測し、行動  $a \in A$  としてリンク候補集合からリンク選択をする。

Case3 運転者は、ノード C 到着時に、状態入力  $s \in S$  として経路候補集合  $\{e_{CB}, e_{CD} \rightarrow e_{DB}\}$  のうちその時点での最短経路と  $u_C$  により決まるセンタの指示経路の二つの経路情報を観測し、行動  $a \in A$  として経路候補集合から経路選択をする。

Case4 運転者は、ノード A 到着時に、Case1 で示した経路候補集合のうち最短経路を観測し、観測した経路を選択する。この Case ではセンタの指示はなく学習は行わない。
3. 終点ノードに到着した運転者に旅行時間を負の値として報酬  $r$  を与え  $Q$  値を更新する。

図 6: 運転者の学習アルゴリズム

用語の定義 到着した運転者に対し情報提供を行うノードを指示ノードとよび、運転者のいるノードから終点ノードまでの経路の中から選択すべきリンクまたは経路をセンタが指示する。また、経路指示ノード以外のノードを最短ノードとよび、到着した運転者が利己的な経路選択、すなわち、運転者のいるノードから終点ノードまでの旅行時間が最短の経路を選択するノードとする。ここで、情報提供法を以下に示す 4 つのケースに分類した。

- Case1 : ノード A, C を指示ノード
- Case2 : ノード A を指示ノード, ノード C を最短ノード
- Case3 : ノード A を最短ノード, ノード C を指示ノード
- Case4 : ノード A, C を最短ノード

上記した各 Case に従う運転者の学習アルゴリズムを図 6 に示す。

## 4. 計算機実験と考察

### 4.1 予備実験：センタの情報提供戦略の獲得

予備実験として、流入交通流量  $f = 0.9$  の場合の対象ネットワークをセンタ側で制御し SO を実現するための戦略を、図

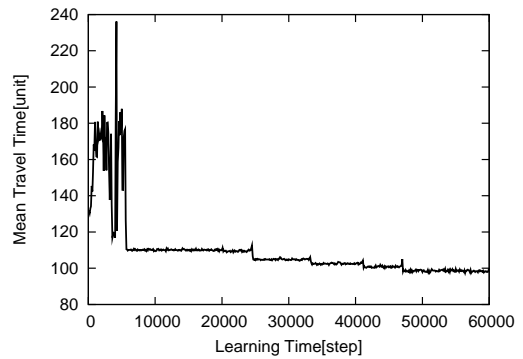


図 7: 学習のプロセス ( $f = 0.9$ )

5 のアルゴリズムを用いて獲得し、学習の収束性を確かめる。ここで、流入確率はポアソン分布に従いとし、ポアソン分布の平均値に流入交通流量を当てはめることで、流量の制御を行った。流入確率に従い発生させた運転者は、まずノード A の手前にある待ち行列に入り、流入しようとする  $e_{AC}$ 、または  $e_{CD}$  に空きがあれば、ネットワークに流入可能であり、空きがなければ、空くまで待ち行列内で待機する。運転者の旅行時間は、ノード AB 間を移動するのに費やした時間に、待ち行列での待機時間を加えたものと表現する。この設定は以降の実験でも同様である。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  $\epsilon$ -greedy 選択 ( $\epsilon = 0.1$ ) を採用する。試行回数は 60,000[step] とし、収束性を確かめるため実験では 50,000[step] 以降は決定的選択 ( $\epsilon = 0$ ) とした。その結果を図 7 に示す。横軸は学習回数、縦軸は学習の更新 100[step] 毎の平均旅行時間である。図 7 より 50,000[step] 以降は一定の配分率に収束しており、学習が収束し、SO を実現する戦略が獲得できた。

### 4.2 実験 1：意思決定導入前の情報提供の効果

実験 1 として、センタの学習により獲得した SO を実現できる情報提供戦略による交通流制御の効果を検証する。前章で述べた各 Case に従って情報提供を行い、運転者に経路を選択させる。なお、ここでは経路指示ノードにおいて情報提供を受けた運転者は必ずセンタの指示経路に従う。こうして得られた各流量における運転者の平均旅行時間を図 8 に示す。情報提供による経路指示がない Case4 に比べ、Case1,2 では短い平均旅行時間が得られており、情報提供が交通流配分に効果的であったといえる。

### 4.3 実験 2：意思決定導入後の情報提供の効果

実験 2 では、運転者に意思決定を導入した場合の運転者の意思決定の挙動を 3 節で述べた各 Case に従って、図 6 のアルゴリズムを用いて実験し検証する。学習のパラメータは学習率  $\alpha = 0.1$ 、割引率  $\gamma = 0.9$  とし、行動選択には  $\epsilon$ -greedy 選択 ( $\epsilon = 0.3$ ) を採用する。また、試行回数は 40,000[step] とし、情報提供を受け学習を行う運転者の数は 300 と設定した。情報提供法ごとの実験結果を図 9 に示す。横軸は学習回数、縦軸は学習の更新 100[step] 毎の平均旅行時間である。図 10、図 11 は横軸を流量、縦軸を学習エージェントである運転者数とし、最短経路と指示経路から成る状態に対し、学習の結果、指示経路を選ぶ行動価値が高くなった運転者数を *Obey*、最短経路を選ぶ行動価値が高くなった運転者数を *Disobey*、*Obey* の内、指示経路と最短経路が一致していたために指示経路を選ぶようになった運転者数を *Match* とし、Case1, Case2 の場合をそれぞれプロットした。

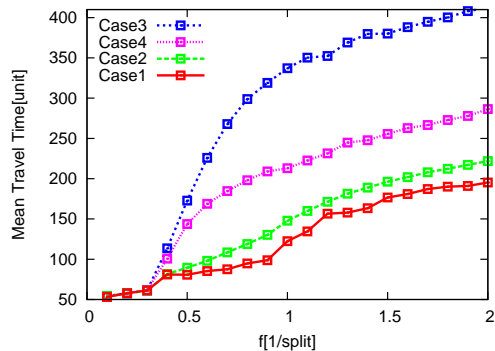


図 8: 情報提供法の効果

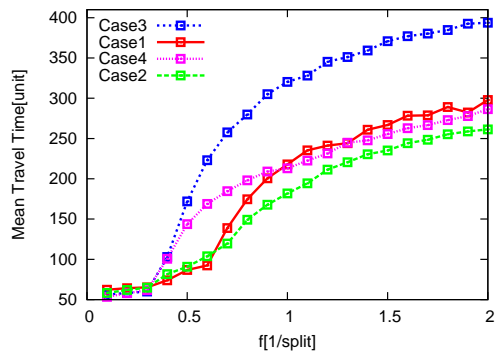


図 9: 意思決定導入時の情報提供法による効果

図 9 より、低い流量では Case 1, 2 は Case 4 よりも短い平均旅行時間を得られているので情報提供として効果があったが、高い流量ではその三つの中で平均旅行時間の差が小さくなっていった。これは、図 10 より Case 1 では、最短経路と指示経路が一致せず、また、ほとんどの運転者が最短経路を選択していた。全ての運転者が指示経路を選択することで SO が実現できるが、一部の運転者が指示経路を選択するだけではその運転者の旅行時間が短くはならず、指示とは関係ない経路を選択する運転者が多くなったと考えられる。図 11 より、Case 2 では、最短経路と指示経路が一致している場合が多く、指示経路を選択する運転者の数に関係なく、最短経路を選ぶことが指示経路を選択することになった。その結果、指示経路を選択する運転者が多くなり Case 4 よりも短い平均旅行時間になったと考えられる。両者とも報酬が旅行時間であるために、最終的には最短経路を選択することを学習した運転者が多くなってしまった。指示経路を選択させるためには運転者が旅行時間以外の評価基準を持つ必要があるといえる。一方、Case 3 では、意思決定規範の有無に関わらず、情報提供のない Case 4 よりも平均旅行時間が大きくなってしまった、つまり、ノード C での情報提供は効果がないと考えられる。このことから、対象ネットワークで情報提供を行う上でノード C に比べて、ノード A における早期の段階での情報提供に効果があるといえる。

## 5. まとめと今後の課題

本研究では、Braess のパラドクスの生じる交通ネットワークを対象に、交通流を ASEP に基づいてモデル化し、センタと運転者の二つの意思決定主体による交通流の制御モデルを提案した。また、情報提供に対する運転者の意思決定戦略を強化学習を用いて獲得させることによって、系全体の交通流への影響を観察した。その結果、自由走行相に相当する交通流量における情報提供には効果が観察されたが、渋滞相で有意な効果

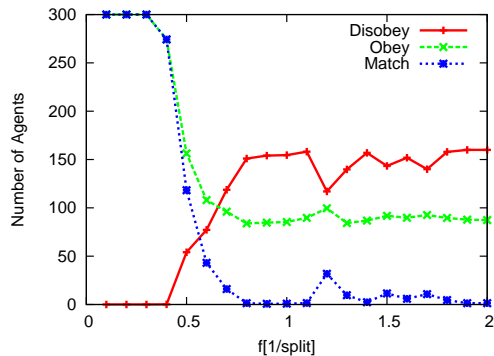


図 10: 学習後の運転者の傾向 (Case1)

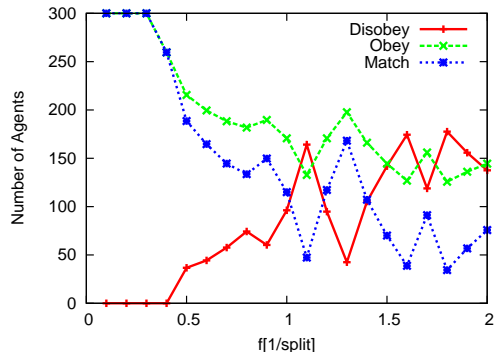


図 11: 学習後の運転者の傾向 (Case2)

が確認できなかった。また、報酬が旅行時間であるために、最終的には最短経路を選択することを学習した運転者が多くなってしまった。一方、経路選択に関する情報提供を行うノードを変化させる実験によって、ネットワークで情報提供を行う上で重要となるノードの存在を確認し、情報の遅れ具合が情報提供ノードの決定に関係することが考えられる。

今後の課題として、システム最適な状態の実現に向けて指示経路を選択する運転者の数を増やすために、旅行時間以外の評価基準として運転者の経路選択のインセンティブを設計し運転者の意思決定を制御する方法の検討も必要であると考えられる。また、今回対象としたのは閉じたネットワークだったのでノードに外部から入力がある場合の制御に拡張することも課題である。

## 参考文献

- [内田 10] 内田英明, 荒井幸代: “ 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 ”, 第 24 回人工知能学会全国大会論文集, 2I1-OS5-7, (2010) .
- [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).
- [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 Traffic 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).