

論文 / 著書情報  
Article / Book Information

論題(和文)	強化学習を用いた動的混雑課金額決定手法の実験室実験による検証
Title(English)	Evaluation of reinforcement learning-based dynamic congestion pricing method using laboratory experiments
著者(和文)	佐藤 公洋, 瀬尾 亨, 田子 裕亮, 田名部 淳, 布施 孝志
Authors(English)	Kimihiro SATO, Toru SEO, Yusuke TAGO, Jun TANABE, Takashi FUSE
出典(和文)	第63回土木計画学研究発表会・講演集, , ,
Citation(English)	, , ,
発行日 / Pub. date	2021, 6
権利情報 / Copyright	本著作物の著作権は土木学会に帰属します。 Copyright (c) 2021 Japan Society of Civil Engineers.

# 強化学習を用いた動的混雑課金額決定手法の実験室実験による検証

佐藤 公洋<sup>1</sup>・瀬尾 亨<sup>2</sup>・田子 裕亮<sup>3</sup>・田名部 淳<sup>4</sup>・布施 孝志<sup>5</sup>

<sup>1</sup>学生会員 東京大学大学院 工学系研究科社会基盤学専攻 (〒 113-8656 東京都文京区本郷 7-3-1)

E-mail: sato@trip.t.u-tokyo.ac.jp

<sup>2</sup>正会員 東京大学大学院助教 工学系研究科社会基盤学専攻 (〒 113-8656 東京都文京区本郷 7-3-1)

E-mail: seo@civil.t.u-tokyo.ac.jp

<sup>3</sup>正会員 株式会社地域未来研究所 (〒 531-0003 大阪府大阪市北区堂島 1-5-17)

<sup>4</sup>正会員 株式会社地域未来研究所 (〒 531-0003 大阪府大阪市北区堂島 1-5-17)

<sup>5</sup>正会員 東京大学大学院教授 工学系研究科社会基盤学専攻 (〒 113-8656 東京都文京区本郷 7-3-1)

E-mail: fuse@civil.t.u-tokyo.ac.jp

交通渋滞を緩和する施策の1つとして、1日の中での交通需要の変動を考慮する動的混雑課金の有用性が提唱されている。また、課金主体と道路利用者の間にある情報の非対称性に対応するため、Trial-and-error 型の課金額決定手法が提案されている。本研究では、強化学習を用いた Trial-and-error 型の動的混雑課金手法を実験室実験により検証する。その結果、先述の混雑課金手法が社会的余剰を増加させ、被験者の出発時刻を分散させる場合があることが示された。また、獲得効用の最大化ではなく遅着幅の最小化を図る人が一定数いることが分かった。

**Key Words:** dynamic congestion toll, trial-and-error, reinforcement learning, laboratory experiment

## 1. はじめに

自動車の交通渋滞問題解決のために、ソフト面での対策の1つとして混雑課金が注目されている。特に、1日の中での交通需要の変動を考慮する動的混雑課金の有用性が提唱されている (Vickrey<sup>1</sup>, Arnott et al.<sup>2</sup>, 赤松<sup>3</sup>, 桑原・赤松<sup>4</sup>, Zhu and Ukkusuri<sup>5</sup>, Qiu et al.<sup>6</sup>)。しかしながら、混雑課金において利用者の出発時刻選択のための経済計算 (例：個人の時間価値の値) を課金主体の側が具体的に把握することは困難であり、情報の非対称性が存在すると言える。このため、最適課金額の即座の決定は難しい。

情報の非対称性に対処するために、観測可能な交通データを用いた Trial-and-error 手法が提案されている (Li<sup>7</sup>, Ye et al.<sup>8</sup>)。本手法は、何らかの混雑課金を課した際の交通状態を観測し、観測データに基づく課金額の調整を試行錯誤的に繰り返し行い、最適課金を発見するものである。Seo and Yin<sup>9</sup>, Seo<sup>10</sup> は、1日の中での交通需要の変動を考慮する Trial-and-error 型の動的混雑課金手法を提案している。また、佐藤ら<sup>11</sup> は、Seo<sup>10</sup> を基にして、強化学習を用いた Trial-and-error 型の動的混雑課金手法を提案している。強化学習とは、ある環境内において、エージェントが観測した状態に基づき行動を選択する中で報酬を獲得し、その報酬に基づき最適な行動を学習するものであり、詳細が未知の環境

においても学習可能である。強化学習を用いた動的混雑課金手法を適用したシミュレーション実験では、時間価値の値が異なる環境でも一定の混雑低減性能が発揮されることが示されている。

上述の混雑課金手法の検証にはコンピュータ上のエージェントシミュレーションなどが主に用いられる。そのようなシミュレーション内のエージェントは、与えられた全ての情報を正確に評価し合理的な行動を選択するよう仮定されていることが多い。一方、実際の交通システムの利用者である人間はそのような行動はとらない。一部の情報しか評価しなかったり、情報の評価が不正確であったり、選択行動が限定合理的であったりする。そのため、上述の混雑課金手法が人間が利用する交通システムに対しても有効かどうかは明らかではない。

そこで本研究では、佐藤ら<sup>11</sup> の動的混雑課金手法の人間が利用する交通システムに対する有効性を実験室実験により検証する。具体的には、提案手法の有する混雑低減性能の検証、および被験者による出発時刻選択の要因の分析を行う。

本稿の構成は以下の通りである。第2章にて、実験で用いる混雑課金手法の概要を示す。第3章にて、実験の内容を示す。具体的には、実施状況や実験の流れを示す。第4章にて、被験者属性の基礎分析結果を述べる。第5章にて、実験の結果を基にした分析を行う。第6章にて、結論と今後の課題をまとめる。

## 2. 混雑課金手法の概要

本章では、実験で検証する混雑課金手法の概要を示す。本手法は佐藤ら<sup>11)</sup>の手法とほぼ同一である。本手法は、交通需要の変動を考慮する出発時刻問題を対象に、強化学習を用いた Trial-and-error により最適な動的混雑課金を決定するものである。課金の徴収は交通ネットワーク上のボトルネックで行われる。強化学習では、状態を交通データ、行動を課金額の変更とし、ボトルネックでの待ち時間を減少させる行動に対し良い報酬を与えることで学習を進める。

第(1)節では、用いる交通モデルを定義する。第(2)節では、被験者のコストを定義する。第(3)節では、課金額更新の設定を示す。

### (1) 交通モデル

本研究では、交通モデルとして単一ボトルネックモデル<sup>12)</sup>を用いた。具体的には、図-1のように2個のノードと1本のリンクから構成されるネットワークで、リンク上にボトルネックがあり、ボトルネック上でのみ Point Queue の待ち行列が形成されるとする。即ち、待ち行列の物理的な長さは考慮しない。OD ペアは1個であり、OD 交通量は固定と仮定する。

数学記法は以下の通りである。

$j$ : 日数

$t$ : 勤務地への到着時刻

$t^*$ : 被験者の勤務地への希望到着時刻

$\mu$ : ボトルネックの単位時間当たりの容量

$a_j(t)$ :  $j$  日目において勤務地への到着時刻が  $t$  のものの居住地からの出発率

$N_j(t)$ :  $j$  日目、時刻  $t$  におけるボトルネックでの待ち行列台数

$w_j(t)$ :  $j$  日目、時刻  $t$  にボトルネックを流出する被験者のボトルネックでの待ち時間

$\tau_j(t)$ :  $j$  日目、時刻  $t$  にボトルネックを流出する被験者に課される課金額

$F$ : 居住地とボトルネックの間の自由旅行時間

$F_d$ : ボトルネックと勤務地間の自由旅行時間

$t_1(t)$ : 時刻  $t$  に勤務地に到着した被験者のボトルネックからの流出時刻

$t_e(t)$ : 時刻  $t$  に勤務地に到着した旅行者のボトルネックへの流入時刻

今回の実験では、 $t^*$  は午前9時(一定)、 $\mu = 4$  人/分、 $F = 0$  分、 $F_d = 20$  分と設定した。また、被験者の人数が少ないため、 $a_j(t)$  には実際の値を20倍したものをを用いた。被験者が選択可能な出発時刻は、午前6時35分から午前9時50分の間で5分刻みの時刻とする。

$j$  日目、時刻  $t'$  におけるボトルネックでの待ち行列

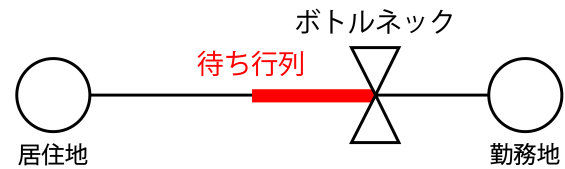


図-1 単一ボトルネックモデルの概要図

台数の変化を式(1)、(2)のように表す。

- $N_j(t') = 0$  かつ  $a_j(t_e^{-1}(t')) < \mu$  の場合

$$\frac{dN_j(t')}{dt'} = 0 \quad (1)$$

- その他の場合

$$\frac{dN_j(t')}{dt'} = a_j(t_e^{-1}(t')) - \mu \quad (2)$$

ここで、 $t'$  はボトルネックへの流入時刻を示す。

$j$  日目、時刻  $t_1(t)$  におけるボトルネックでの待ち時間を式(3)のように表す。

$$w_j(t_1(t)) = \frac{N_j(t_e(t))}{\mu} \quad (3)$$

### (2) 被験者のコスト

$j$  日目、時刻  $t$  における被験者1人当たりの一般化コスト  $c_j(t)$  を式(4)により定義する。

$$c_j(t) = \tau_j(t_1(t)) + \alpha\{F + w_j(t_1(t)) + F_d\} + L(t) \quad (4)$$

ここで、 $t$  は勤務地への到着時刻、 $\alpha$  は1人の旅行者における単位時間当たりの時間価値を示す。また、 $L(t)$  は  $j$  日目、時刻  $t$  における早着コストまたは遅着コストを表し、式(5)、式(6)により定義されるものとする。

- $t < t^*$  の場合

$$L(t) = \beta(t^* - t) \quad (5)$$

- $t \geq t^*$  の場合

$$L(t) = \gamma(t - t^*) \quad (6)$$

ここで、 $\beta$  は希望到着時刻と比べて早く勤務地に到着する場合の単位時間あたりのコスト(早着コスト)、 $\gamma$  は希望到着時刻と比べて遅く勤務地に到着する場合の単位時間あたりのコスト(遅着コスト)を示す。今回の実験では、 $\alpha = 30$  円/分、 $\beta = 13.5$  円/分、 $\gamma = 36$  円/分と設定した。

### (3) 課金額更新

課金額は7日毎に更新する。時間帯ごとに強化学習エージェントが設けられ、それぞれの時間帯の課金額の増減を決定する。動的混雑課金額決定手法の詳細は佐藤ら<sup>11)</sup>における分散制御型手法を参照されたい。各エージェントは事前にシミュレーション実験により学習を完了させたうえで、実験室実験中も学習を進行させるよう設定した。

なお、佐藤ら<sup>11)</sup>の手法では、混雑課金額は強化学習によって決定される項とシミュレーションの day-to-day dynamics を安定化させる項の和として定義されている。しかし、今回の実験では安定化項は省略するものとする。この理由は、通常の場合には安定化項の大きさは強化学習項と比較し小さいこと、および、実験室実験を試行（後述する「1セット目」）した際に被験者の行動がシミュレーションとかけ離れていたため安定化項の値が不適切になったことである。

### 3. 実験室実験の内容

第(1)節では、日時や会場等、実験の実施状況をまとめる。第(2)節では、実験の具体的な流れをまとめる。また、実験システムの構成、および実験で被験者が使用したアプリの画面を示す。

#### (1) 実施状況

実験の実施状況は以下の通り：

日時 2020年11月21日(土)

会場 大阪, 東京

被験者 調査会社登録のモニター

注記 東京大学対面授業基準に則った COVID-19 対策の下で実施

#### (2) 実験の流れ

本節では、実験の流れや実験システムの構成を示す。本実験は、被験者を集め、実験の趣旨・方法・報酬について説明した上で、Webブラウザ上アプリを通じて設問に繰り返し回答させるという形で行った。本実験での設問は、Webブラウザ上アプリを通じて各被験者に交通学習シミュレータ上で集計・計算された前日の交通行動と当日の課金額を示し、当日の交通行動を選択させるというものである。

実験には以下の機器・ソフトウェアを用いる。

- 被験者のスマートフォン・タブレット
- 入出力用の Web ブラウザ上アプリ（以下、Web アプリ）
- 強化学習による混雑課金手法を組み込んだ交通シミュレータ（以下、交通学習シミュレータ）

実験システムの構成図を図-2として示す。また、実験システムの内、被験者が実験中に使用した Web ブラウザ上アプリの画面を図-3として示す。

実験の流れは以下の通りである。

1. 被験者の集合
2. 被験者に実験の概要，操作説明，獲得効用と報酬について説明する。
3. 被験者のスマートフォン・タブレットを使用し，Web アプリにアクセスしてもらう。
4. 実験開始.  $j := 1$  日目とし，0 日目の時間帯別旅行時間と 1 日目の混雑課金額を設定する。

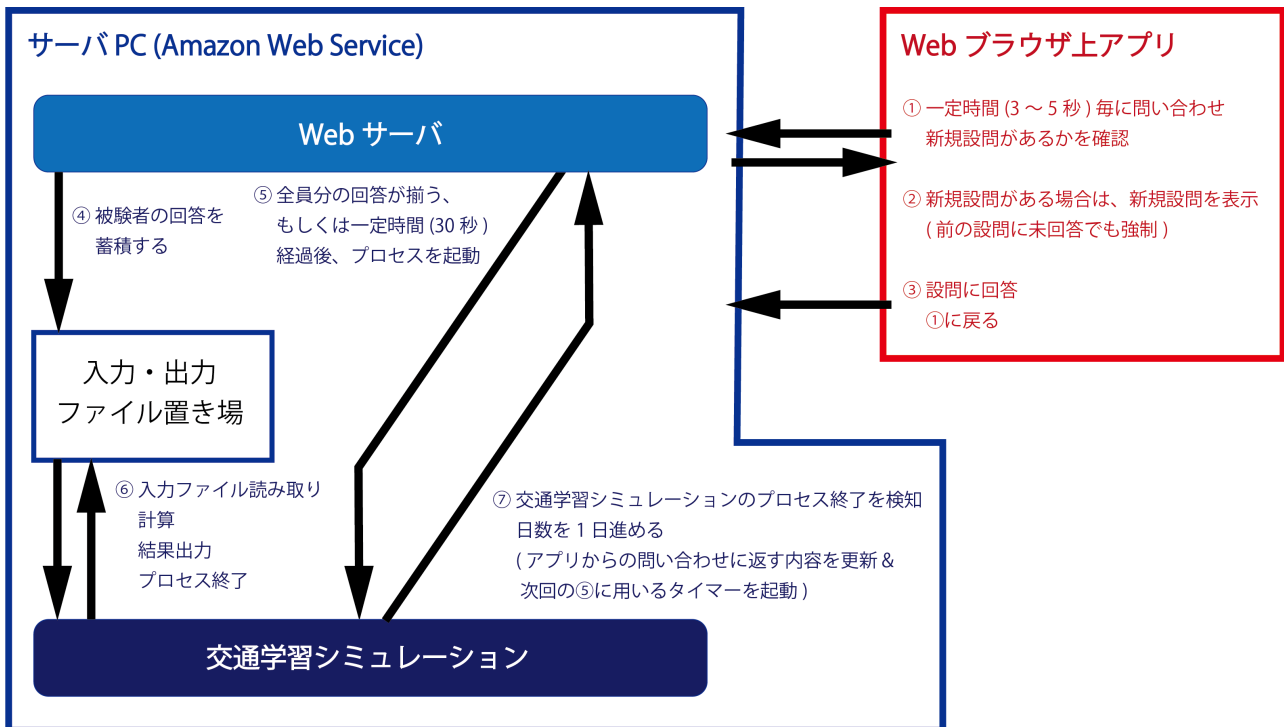


図-2 実験システムの構成図

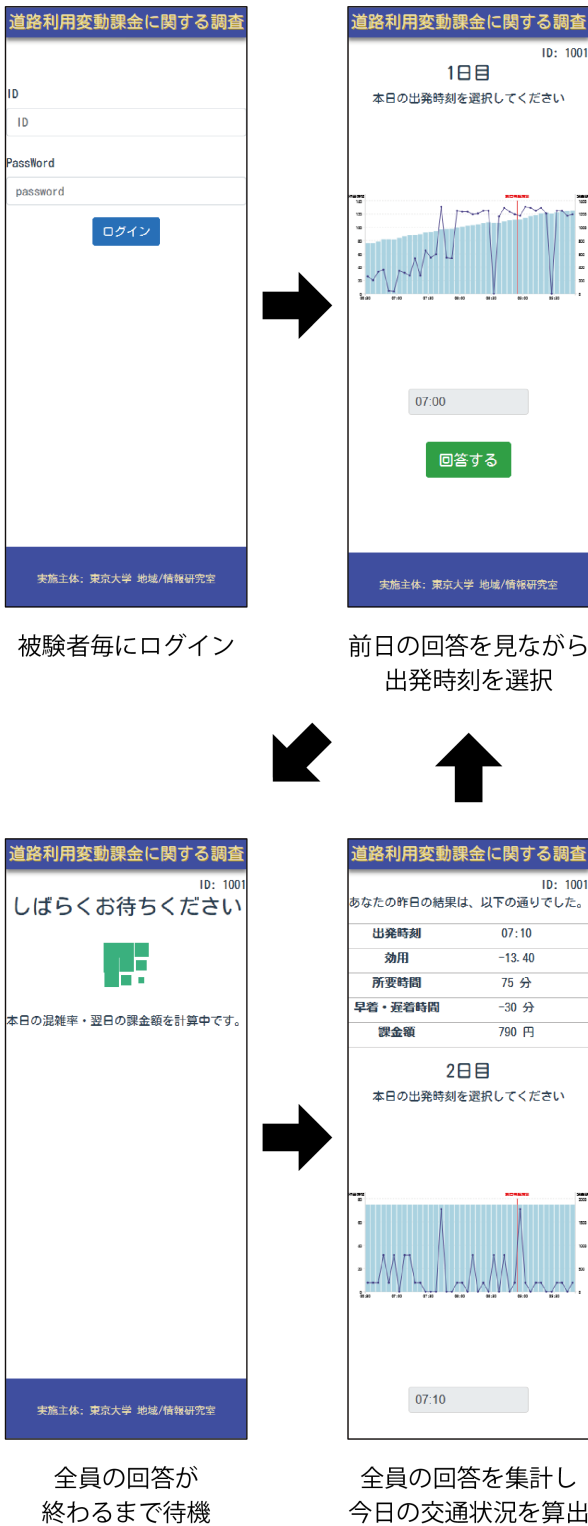


図-3 Web ブラウザ上アプリの画面

- 4.1. 被験者に  $j-1$  日目の時間帯別旅行時間,  $j-1$  日目の各被験者の獲得効用,  $j$  日目の混雑課金額を提示する.
- 4.2. 被験者が  $j$  日目の出発時刻を選択する.

4.3. 全ての被験者の入力終了するか, 一定の待ち時間が経過したのち, 交通学習シミュレータにより  $j$  日目の交通状況を計算し, 混雑課金の学習を進め,  $j+1$  日目の混雑課金額を算出する.

4.4.  $j$  が終了日でなければ  $j := j+1$  としてステップ 4.1. に戻る. 終了日であればステップ 5. に進む.

5. 被験者毎に総獲得効用を算出し, それに応じた金銭的報酬を支払い, 実験終了.

以上の流れを 1 セットとする. 被験者の入力待ち時間は 30 秒とした. 終了日は 102 日目とした (但し, 1 日目と 2 日目は被験者が操作に慣れるための試行日とした). 被験者の報酬は 3000 円の基本報酬と, 総獲得効用に比例した最大 600 円分のポイントとした. 計 3 セットの実験を行い, それぞれの被験者数は 1 セット目: 35 名, 2 セット目: 33 名, 3 セット目: 40 名であった. また, セットごとの所要実時間は 1 時間程度であった. ただし, 1 セット目は前述した安定化項の問題が発生したため, 実験結果の分析からは除外した.

#### 4. 被験者の属性の基礎集計

本章では, 実験に参加した被験者の社会経済属性の集計結果を示す.

被験者の性別は, 男性が 53 名, 女性が 53 名であった. また, 被験者の年齢を図-4, 職業を図-5, 普段の自動車運転頻度を図-6, 普段の高速道路利用頻度を図-7, 年収を図-8, オリンピック混雑課金への賛否を図-9 として示す.

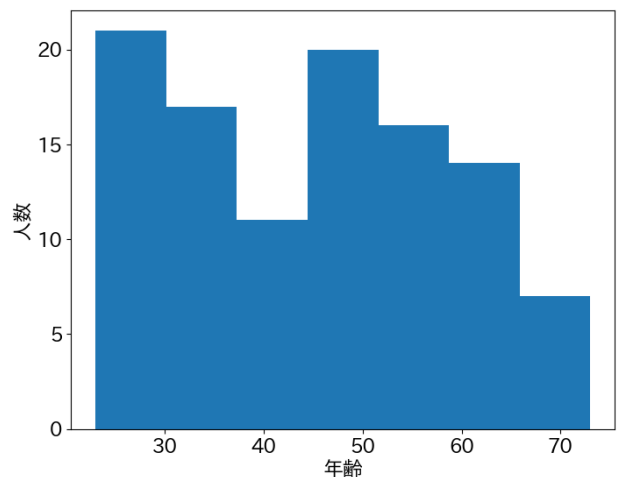


図-4 被験者の年齢

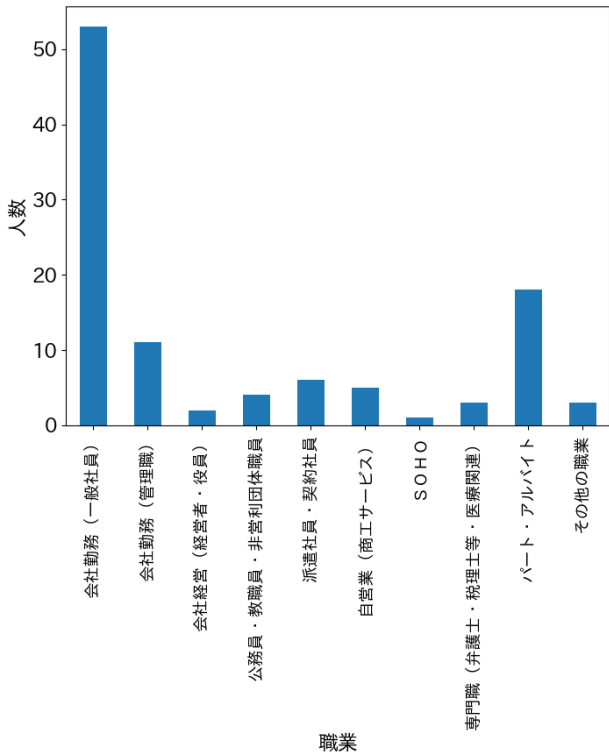


図-5 被験者の職業

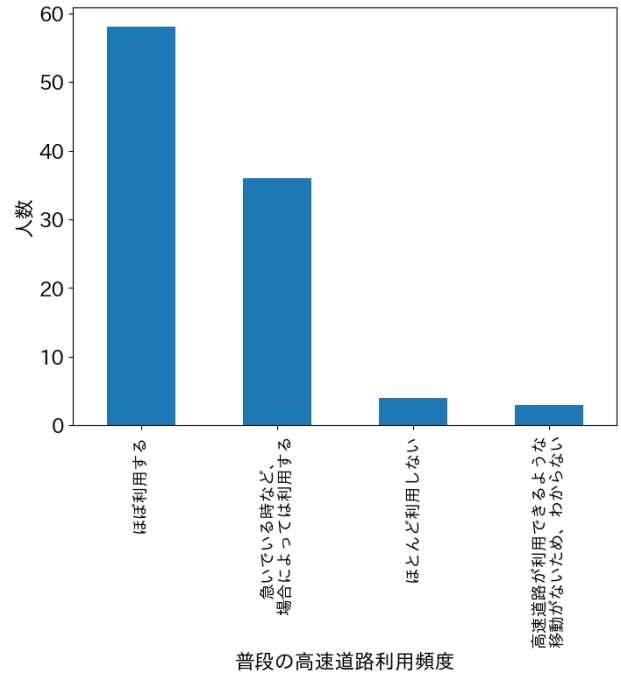


図-7 被験者の普段の高速道路利用頻度

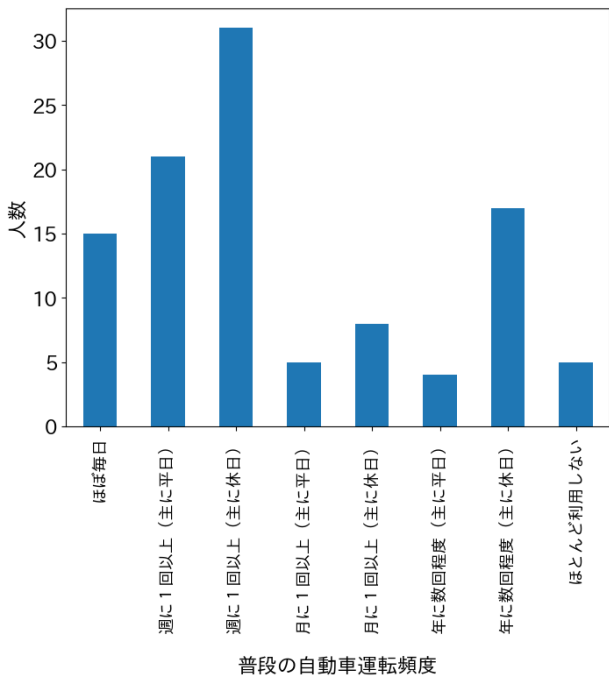


図-6 被験者の普段の自動車運転頻度

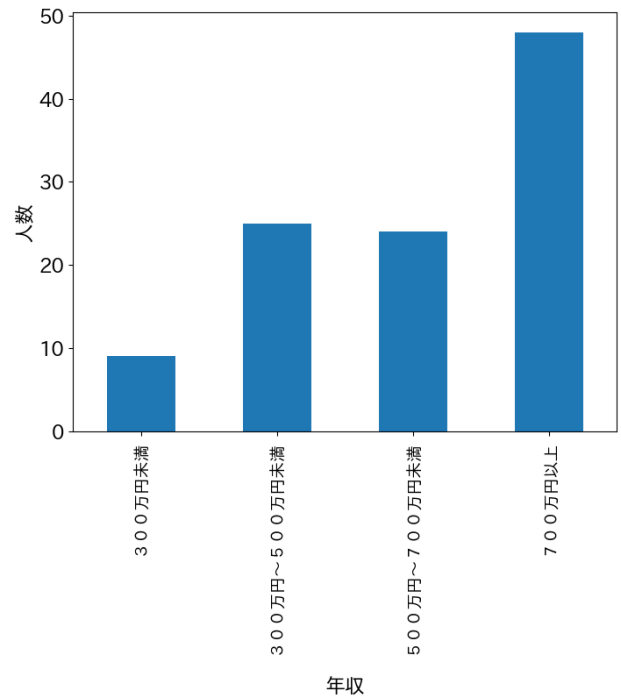


図-8 被験者の年収

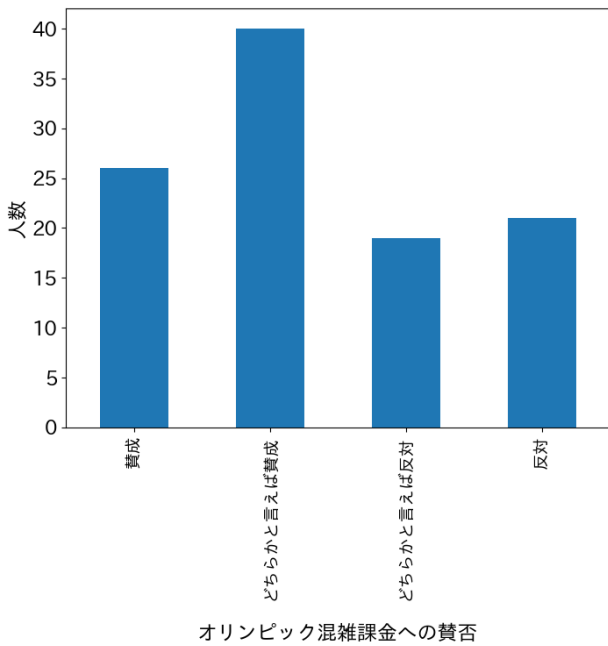


図-9 被験者のオリンピック混雑課金への賛否

なお、普段の自動車運転頻度は同乗する場合も含み、タクシーやバスは除くものとする。普段の高速道路頻度は「勤務地までの移動の途中に高速道路があった場合に利用するか(自分もしくは家族が料金を支払う場合)」を示すものとする。また、オリンピック混雑課金は「選手や大会関係者の円滑な移動のために、東京オリンピック・パラリンピックの期間中に限り、首都高速道路の料金を1000円上乗せすることでマイカー利用を抑制する施策(午前6時から午後10時まで)」を指すものとする。

## 5. 実験室実験の結果と考察

本章では、実験室実験の結果を基に、強化学習を用いた動的混雑課金額決定手法における混雑低減性能の検証、被験者による出発時刻選択の要因の分析等を行う。具体的には以下の内容を確認する。

- 実験の各セットにおける社会的余剰の推移を基に、強化学習を用いた動的混雑課金額決定手法における社会的余剰の増加性能を検証する。
- 実験の各セットにおける出発時刻分布の標準偏差の推移を基に、強化学習を用いた動的混雑課金額決定手法における出発時刻の分散性能を検証する。
- 獲得効用等を基にして、被験者による出発時刻選択の要因を分析する。

なお、実験の各セットでの  $j$  日目の社会的余剰は式 (7) のように表されるとする。

$$S_j = -\frac{1}{n} \sum_i [\alpha \{F + w_j(t_i(t_i)) + F_d\} + L(t_i)] \quad (7)$$

ここで、 $S_j$  は  $j$  日目における社会的余剰、 $n$  は実験の各セットでの被験者数、 $t_i$  は被験者  $i$  が勤務地に到着した時刻を表す。

また、 $j$  日目における被験者  $i$  の獲得効用は式 (8) のように表されるとする。

$$U_{j,i} = -c_j(t_i) \quad (8)$$

ここで、 $U_{j,i}$  は  $j$  日目における被験者  $i$  の獲得効用を示す。

なお、実験初期(1, 2日目)は出発時刻分布を均一とした初期状態の影響が強く、また被験者が操作に慣れるための試行日であったため、本論文の分析では除くものとする。これ以降は3~102日目を1~100日目として扱うものとする。

### (1) 2セット目の実験結果

#### a) 全体の結果

社会的余剰の推移を図-10、出発時刻分布の標準偏差の推移を図-11、課金額平均の推移を図-12として示す。

全体として、社会的余剰が増加傾向にあり、出発時刻選択が分散していることが読み取れる。また、図-11より、40日目付近で出発時刻分布の標準偏差の増加量が大きいことが読み取れる。これは、40日目付近で獲得効用の高い出発時刻を被験者が選択し始めたためであると考えられる。ここから、被験者が獲得効用の高い出発時刻を発見するまでにはある程度の時間を要する場面があると考えられる。

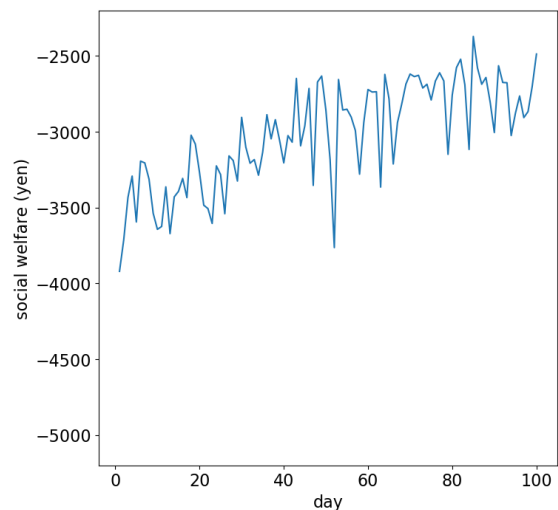


図-10 2セット目の実験における社会的余剰の推移

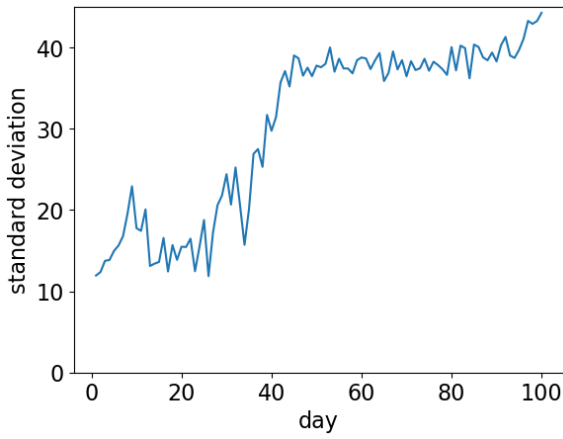


図-11 2セット目の実験における出発時刻分布の標準偏差の推移

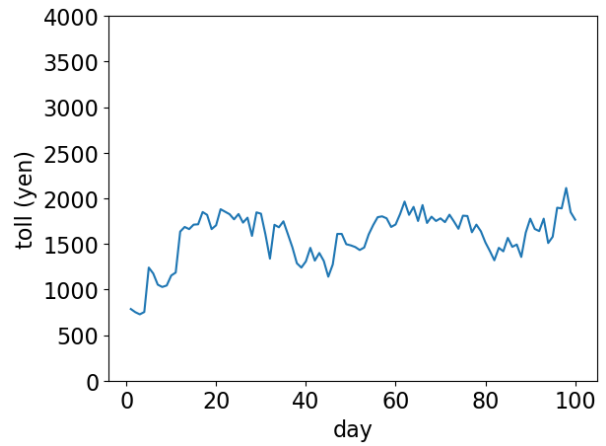


図-12 2セット目の実験における課金額平均の推移

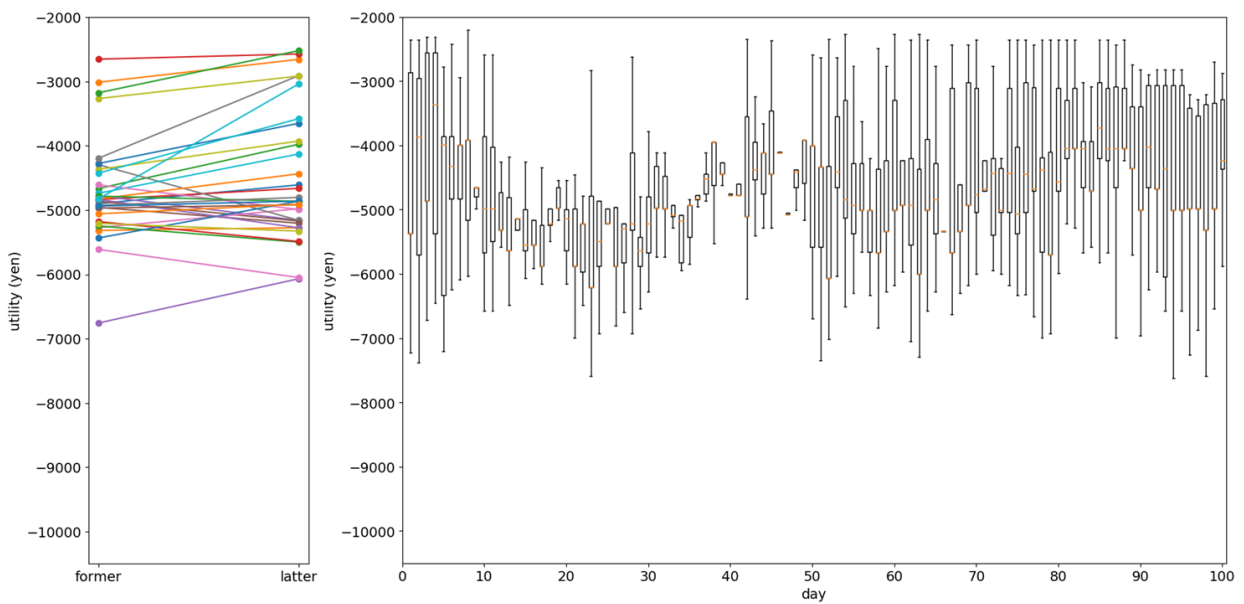


図-13 2セット目の実験における各被験者の前半と後半の獲得効用平均，および各日における獲得効用分布

**b) 各被験者の結果**

各被験者の前半(1日目～50日目)と後半(51日目～100日目)の獲得効用平均，および各日における獲得効用分布を示す図を図-13として示す。

なお，図-13の右側の図において，外れ値は図示しないものとした。ひげの長さは四分位範囲の1.5倍を上限とし，これに入らない値を外れ値とした。図-13より，被験者毎の獲得効用には大きなばらつきがあり，均衡していないことが読み取れる。また，獲得効用を増加させられた被験者と，増加させられていない被験者が存在することが分かる。

**(2) 3セット目の実験結果**

**a) 全体の結果**

社会的余剰の推移を図-14，出発時刻分布の標準偏差の推移を図-15，課金額平均の推移を図-16として示す。

全体として，社会的余剰が増加～変化なしの傾向にあり，出発時刻選択が分散していることが読み取れる。また，図-15より，出発時刻分布の標準偏差の増加率に大きな変化はないことが読み取れる。一方，図-16より，2セット目の実験に比べて課金額平均が最終的に約2倍に増加していることが読み取れる。しかし，図-14において社会的余剰の目立った増加が見られないことから，課金が有効に働いていない可能性があると言える。



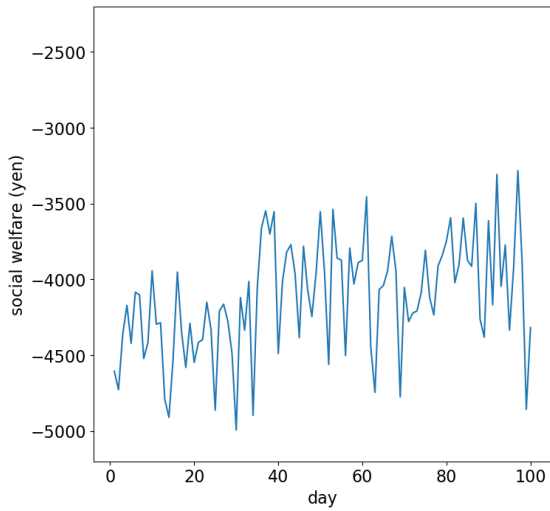


図-14 3セット目の実験における社会的余剰の推移

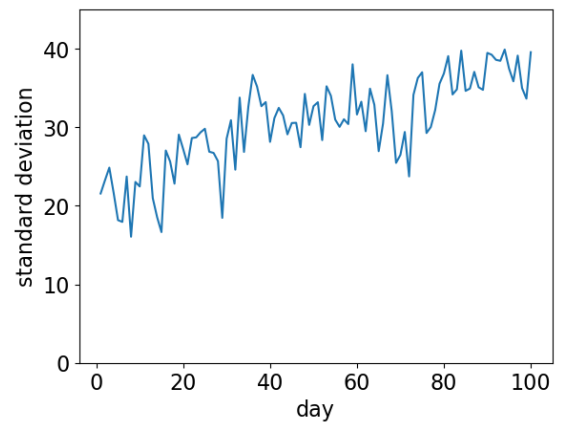


図-15 3セット目の実験における出発時刻分布の標準偏差の推移

**b) 各被験者の結果**

各被験者の前半(1日目~50日目)と後半(51日目~100日目)の獲得効用平均, および各日における獲得効用分布を示す図を図-17として示す.

なお, 図-17の右側の図において, 外れ値は図示しないものとした. ひげの長さは四分位範囲の1.5倍を上限とし, これに入らない値を外れ値とした.

図-17より, 被験者毎の獲得効用にはばらつきがあり, 均衡していないことが読み取れる. また, 被験者個人の獲得効用は減少傾向にある.

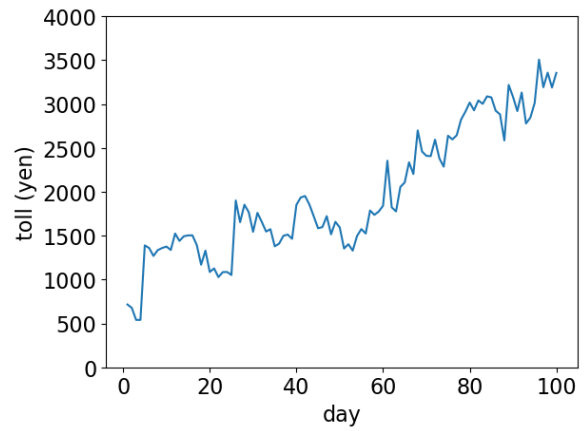


図-16 3セット目の実験における課金額平均の推移

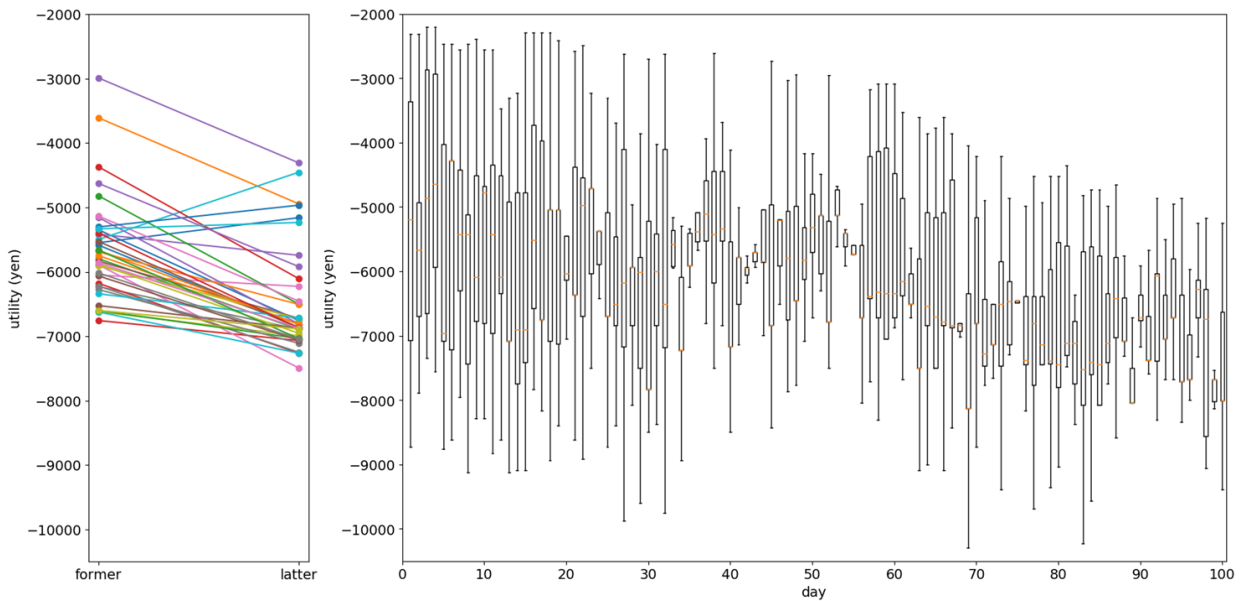


図-17 3セット目の実験における各被験者の前半と後半の獲得効用平均, および各日における獲得効用分布

### (3) 考察

本節では実験結果について考察する。

まず、混雑課金手法は社会的余剰を増加する傾向にあることがわかった。この社会的余剰増加の要因は、出発時刻の分散による渋滞の減少であり、本手法が意図した通りに働いたといえる。これは佐藤ら<sup>11)</sup>のシミュレーションによる検証結果と整合的であり、本手法は人間が利用する交通システムに対しても有効である可能性が示唆された。

一方で、佐藤ら<sup>11)</sup>のシミュレーションによる検証結果と大きく異なる結果も得られた。まず、被験者間の獲得効用は大きく異なり、均衡とは程遠い状態にある。また、被験者毎に獲得効用が増加した場合や、減少した場合も見られた。さらに、2セット目の実験では全被験者の獲得した総効用は日数の経過とともに余り変動しなかったのに対し、3セット目では日数の経過とともに減少する傾向にあった。佐藤ら<sup>11)</sup>のシミュレーションでは旅行者は基本的に獲得効用を増加させ均衡状態を実現するような合理的な行動をとると仮定していたため、大きく異なる結果となったと考えられる。

本実験での被験者の行動原理は明らかではないが、その一端を以下に述べる。各被験者が実験後半(51日目～100日目)に得た効用の平均を横軸、実験後半での遅着幅の平均を縦軸にとった散布図を、実験のセット毎に図-18、図-19として示す。なお、遅着幅は「勤務地への実際の到着時刻と希望到着時刻の差」を示し、式(9)のように表されるものとする。

$$v_i(t_i, t^*) = t_i - t^* \quad (9)$$

ここで、 $v_i(t_i, t^*)$  は被験者  $i$  の遅着幅を示す。

図-18、図-19より、実験の各セット内において、実験後半で得た効用の平均が相対的に低い被験者は遅着幅の平均の絶対値が小さい傾向にあることが読み取れる。一方で、実験後半で得た効用の平均が相対的に高

い被験者は遅着幅の平均の絶対値が相対的に大きいことが読み取れる。よって、「得られる効用の大小ではなく遅着幅を小さくすることを図る被験者」が一定数存在したと考えられる。

## 6. 結論と今後の課題

本研究では、強化学習を用いた Trial-and-error 型の動的混雑課金手法を実験室実験により検証した。本手法は、1日の中での交通需要の変動を考慮し、時間別に課金額を設定するという特徴がある。実験では、被験者を集めて、各被験者に前日の交通状況と当日の課金額を示した上で当日の交通行動を選択させる手順を繰り返して行った。また、前述の実験を3セット行った。1セット目では課金額が過度に増加したが、設定の変更により、2、3セット目では課金額は適切な値をとった。

2、3セット目の実験における全体的な結果としては、日数の経過に伴って社会的余剰が増加、もしくは維持され、出発時刻が分散する傾向が見られた。この点は既往研究でのシミュレーションによる検証結果と同様である。一方、個人毎の獲得効用では2、3セット目の双方で大きなばらつきが見られ、均衡には至らなかった。また、3セット目の実験では個人毎の獲得効用は減少する傾向が見られた。この点は既往研究でのシミュレーションによる検証結果と大きく異なるものであり、シミュレーションで仮定された行動原理と実際の人間のその違いが大きいことを示唆している。

例えば、被験者による出発時刻選択の要因の分析において、獲得効用の最大化ではなく遅着幅の最小幅を図る人が一定数いることが分かった。但し、今回の実験では出発時刻選択を30秒毎に行っていたため、得られる効用や遅着幅について思考する時間が短かったと思われる。これにより、取り急ぎ遅着幅を重視する被

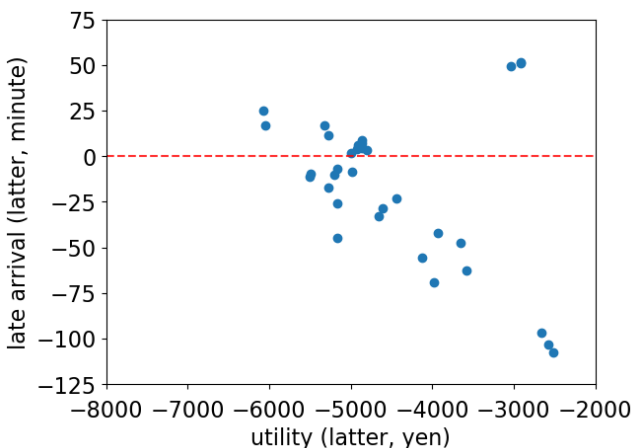


図-18 2セット目の実験の後半における効用-遅着幅の散布図

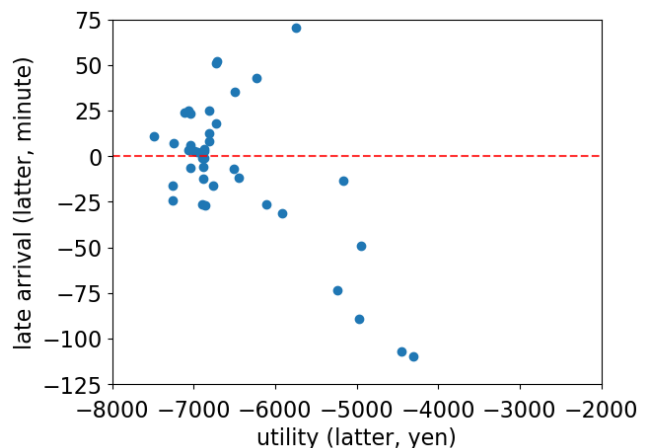


図-19 3セット目の実験の後半における効用-遅着幅の散布図

験者の人数が多くなった可能性がある。

今後の課題には、まず、被験者の社会経済属性と獲得効用の関係の分析がある。また、被験者数や実験内日数を増やした実験の実施、出発時刻選択の制限時間を延長した実験の実施、課金額更新における変動幅や課金額更新頻度の更なる検討が考えられる。

**謝辞：**本研究は国土交通省新道路技術会議の研究課題「学習型モニタリング・交通流動予測に基づく観光渋滞マネジメントについての研究開発」及び科学研究費補助金 20H02267 の助成を受けた。ここに謝意を表します。

#### 参考文献

- 1) Vickrey, W. S.: Congestion theory and transport investment, *The American Economic Review*, Vol. 59, No. 2, pp. 251-260, 1969.
- 2) Arnott, R., de Palma, A. and Lindsey, R.: Economics of a bottleneck, *Journal of Urban Economics*, Vol. 27, No. 1, pp. 111-130, 1990.
- 3) 赤松隆：交通流の予測・誘導・制御と動的なネットワーク配分理論, 土木計画学研究・論文集, Vol. 13, pp. 23-48, 1996.
- 4) 桑原雅夫, 赤松隆：動的ネットワーク解析—これまでの知見とこれからの展望—, 土木学会論文集, No. 653/IV-48, pp. 3-16, 2000.
- 5) Zhu, F. and Ukkusuri, S. V.: A reinforcement learning approach for distance-based dynamic tolling in the stochastic network environment, *Journal of Advanced Transportation*, Vol. 49, No. 2, pp. 247-266, 2015.
- 6) Qiu, W., Chen, H. and An, B.: Dynamic Electronic Toll Collection via Multi-Agent Deep Reinforcement Learning with Edge-Based Graph Convolutional Networks, *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, pp. 4568-4574, 2019.
- 7) Li, M. Z. F.: The role of speed-flow relationship in congestion pricing implementation with an application to Singapore, *Transportation Research Part B: Methodological*, Vol. 36, No. 8, pp. 731-754, 2002.
- 8) Ye, H., Yang, H. and Tan, Z.: Learning marginal-cost pricing via a trial-and-error procedure with day-to-day flow dynamics, *Transportation Research Part B: Methodological*, Vol. 81, pp. 794-807, 2015. ISTTT 21 for the year 2015.
- 9) Seo, T. and Yin, Y.: Optimal pricing for departure time choice problems with unknown preference and demand: Trial-and-error approach, *Transportation Research Board 98th Annual Meeting*, 2019.
- 10) Seo, T.: Trial-and-error congestion pricing scheme for morning commute problem with day-to-day dynamics. *Transportation Research Procedia*, Vol. 47, pp. 561-568, 2020. (The 22nd EURO Working Group on Transportation Meeting, 18-20 September 2019, Barcelona, Spain).
- 11) 佐藤公洋, 瀬尾亨, 布施孝志：強化学習を用いたデータ駆動型の動的混雑課金の最適化手法. 土木学会論文集 D3 (土木計画学), Vol. 76, No. 5, 2021, in press.
- 12) Vickrey, W. S.: Congestion theory and transport investment. *The American Economic Review*, Vol. 59, No. 2, pp. 251-260, 1969.

(2021. 3. 5 受付)

## Evaluation of reinforcement learning-based dynamic congestion pricing method using laboratory experiments

Kimihiko SATO, Toru SEO, Yusuke TAGO, Jun TANABE and Takashi FUSE