協調促進トークンメカニズムの Trial-and-error 制御

高野 壮稀[1]・原 祐輔[2]

[1]学生会員　東北大学 大学院情報科学研究科（〒 980-8579 仙台市青葉区荒巻字青葉 6-6-06）
E-mail: soki.takano.p3@dc.tohoku.ac.jp
[2]正会員　東北大学准教授　大学院情報科学研究科（〒 980-8579 仙台市青葉区荒巻字青葉 6-6-06）
E-mail: hara@tohoku.ac.jp (Corresponding author)

本研究では，都市交通インフラの効率的な利用を実現するにあたって，1) 支払い不要による公平性，2) 利用者自身に利用タイミングを検討させることを通した簡便な選好表明，3) インフラ容量の効率的な利用の 3 点を実現するための新しい割当メカニズムである，協調促進トークンメカニズムを扱う．本メカニズムでは，利用者が 1 人 1 つ持つトークンを消費することで対象の道路リンクが通行できるという設定のもと，管理者は非対称な情報のもとでトークン消費者に対してより良いトークン再配布タイミングを決定する必要がある．そしてより良い均衡状態へ導く制御のためには動学化を考える必要がある．本研究では，利用者の日々の学習や意思決定に加え，管理者の行動をモデル化することで，まずメカニズム下の日々のダイナミクスを考え，これらを踏まえた上で社会的厚生を最大化する Trial-and-error 制御に関して議論する．

Key Words: *mechanism design, artificial payments, bottleneck congestion, information asymmetry*

1. はじめに

道路空間や都市交通インフラの効率的な利用を実現する仕組み・政策には，渋滞や混雑が生み出す外部性を内部化するためのロードプライシング[1)] や通行権取引制度[2)3)] 等の価格を通した市場メカニズムによるアプローチが一般的である．日本においてもロードプライシングや動的料金の導入が検討され始めているものの，金銭支払いが発生するため，利用者の受容性の点で未だ課題が存在している．本来，希望到着時刻の異質性を含む利用者の交通需要と道路インフラの供給量制約を適切にマッチングさせることができれば，必ずしも価格を通した市場メカニズムを利用しなくても，パレート効率的な道路インフラの割当は実現できる．しかし，希望到着時刻などは私的情報であり，道路管理者と利用者の間で情報の非対称性が存在するため，その非対称性を解消するために価格が用いられている．そこで，高野・原[4)] は金銭的支払いではなく，将来の道路利用機会の消費に基づく自己選択メカニズムを用いた協調促進トークンメカニズムによって，情報の非対称性を解決する新たなメカニズムを提案した．

協調促進トークンメカニズムでは，道路利用者は無料配布されるトークンを用いて，自己選択的に望ましい時間帯枠のボトルネック通過を表明する．道路供給者は各時間帯枠の需要に応じてトークン消費者へのトークン再配布タイミングを変化させ，ボトルネックでの渋滞を発生させないように道路需要をコントロールする．ここでは，金銭的な支払いは発生せず，利用者は当日トークンを消費して希望の時間帯枠にボトルネックを通過するか，別の時間帯枠や別の経路を利用するかを検討する．協調促進トークンメカニズムは上記の手続きを通して，1) 支払い不要による公平性と受容性，2) 利用者自身に利用タイミングを検討させることを通した簡便な選好表明，3) インフラ容量の効率的な利用の 3 点を目指す割当メカニズムである．

高野・原[4)] は協調促進トークンメカニズムの静学解析として，特定のトークン再配布タイミングが与えられた条件下における利用者均衡を仮定し，協調促進トークンメカニズムが最適割当を実現する通行権取引制度を近似する効率性を達成させられること，早い者勝ちと比較すると大幅な改善が見込まれることを示した．このように均衡時には効率的に機能することが確認されたが，現実の状況下においては，利用者・道路供給者ともに情報の非対称性が存在し，任意の初期状態から均衡状態にどのように移行するのか不明確である．また，道路管理者が最適なトークンの再配布タイミングを決定するための方策も未解決である．そこで，本研究では，協調促進トークンメカニズム下における利用者・道路供給者の動学的特性に焦点を当てる．利用者は自身以外の利用者の行動に対して正確な情報を有していないために，自身が取りうる選択肢の想定する期待効用と真の期待効用の間にはギャップが存在する．同様

に，道路管理者も利用者需要を正確に把握していないため，決定したトークンの再配布タイミングと最適な再配布タイミングの間にはギャップが存在する．これらのギャップに対して，双方がどのように信念を更新することで，望ましい割当が実現できるかを動学モデルを用いて検討する．

本研究の貢献は以下の 4 点である．

- 協調促進トークンメカニズム下における日々のトークン保持者の変動を考慮した Day-to-Day ダイナミクスのモデル化．(**3.**)
- 不完全情報下において，協調促進トークンメカニズム下における利用者の日々の選択行動をモデル化するために，他の利用者行動に関する信念を定式化し，利用者の信念更新の学習方策と日々の選択行動をモデル化．(**4. (1), (2), (3)**)
- 不完全情報下において，協調促進トークンメカニズム下における管理者のトークン再配布タイミング決定問題をモデル化するために，利用者効用に関する信念を定式化し，信念更新の学習方策と再配布タイミング決定の意思決定をモデル化．(**5. (1), (2), (3)**)
- 再配布タイミングを固定した利用者行動のみ動学化した状況設定における数値実験を実施し，本モデルの動学特性や他のメカニズムとの比較・検討を行なった．(**6.**)

2. 既往研究のレビュー

(1) 利用者の学習と意思決定

本研究で扱う利用者の学習や意思決定のモデル化に関して，Horowitz[5] は，経路選択の意思決定に着目してネットワーク均衡の安定性に関して議論し，過去の経験の重み付き平均を旅行時間の予測値としてモデル化した．その後他にも旅行時間の学習に関して様々な実験や提案が見られる．小林・藤高[6] は合理的期待形成仮説に基づいて，ベイズ学習を用いるなどのドライバーの学習行動を明示的に考慮した経路選択モデルを提案し，ドライバーが学習行動を通じて合理的期待を形成するメカニズムについて数値実験により考察している．中山ら[7] は長期的学習を考慮するための遺伝的アルゴリズムの導入を提案した．Iida et al.[8] は実験室実験によりドライバーの経路選択行動のメカニズムを考察している．

(2) 管理者の行動・ダイナミクス

Li[9] は、未知の需要関数を持つリンクにおける静的な交通に対する試行錯誤的な価格設定方式を提案した．Sandholm[10] は，静的ネットワーク交通において進化ゲーム理論によるアプローチを行なった．Seo[11] は，出発時刻選択における試行錯誤的な価格設定方式を提案した．

(3) 本研究の位置付け

本研究では利用者の学習に関して，メカニズムの開始初期ほど新たな情報を重視し，時間の経過とともに過去を重視するような重みづけをした重みつき平均を用いることとした．

3. 状況設定とメカニズムの概要

(1) 状況設定

本研究では，協調促進トークンメカニズムを提案した高野・原[4] と同じ状況を設定する．単一ボトルネックと迂回路が存在するネットワークを考え，単一ボトルネックでは，ボトルネック容量以上の交通需要が集中した際には渋滞が発生するとする．一方で，迂回路は旅行時間は大きいが十分な容量が存在する状況を考える．この状況の具体例は，災害復旧期における被災により道路容量が減少し関係車両のみを通行可能とした高速道路と，その迂回路となる一般道である．本研究では，ボトルネック容量は時間によって変化しないものとする．

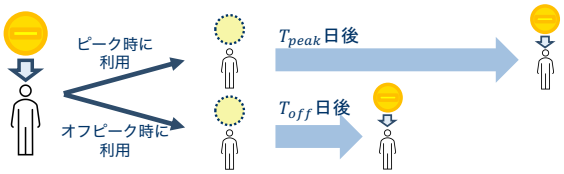
道路管理者は，対象ボトルネックの容量を適切に利用者に割り当てることにより，ボトルネックでの渋滞発生を防ぎ，社会的厚生（＝利用者効用の総和）が最大になることを目指す．

利用者は車両を 1 人 1 台運転し，簡単のため 1 日 1 回のみ通行するものとする．各利用者は日々自身の選好に応じて期待効用を最大化しようと行動する．利用者の当日の効用や明日以降の効用は，各時間帯枠に対する効用の組み合わせであるタイプにより決定し，そのタイプも日々確率的に決定する．ただし，各タイプの効用ベクトルとタイプの確率分布は既知とする．これらが本研究で考える状況設定である．

(2) メカニズムの概要

金銭的支払いを必要とせず，網羅的な選好表明が不要であり，高い効率性を達成できるメカニズムである協調促進トークンメカニズムは大きく下記に示す 3 つの特徴を持つ．

- トークンは供給者から各利用者に最大 1 つ電子的に無料で配布される．
- ボトルネックの通過にはトークン消費による利用申請が必要であり，申請時にトークンを消費する．
- 消費したトークンは事前に設定された再配布タイミングに応じて一定期間後に再配布される．再配



図–1 協調促進トークンの概要

布タイミングは申請した時間帯枠に応じて決められている．

図–1 に示した協調促進トークンメカニズムの一例を説明する．仮に時間帯枠がピーク時，オフピーク時の 2 つとすると，対象のボトルネックを通行したい利用者には，ピーク時に利用する，オフピーク時に利用する，対象の道路の利用をしない (迂回路を利用する) の 3 つの選択肢がある．ボトルネック通行のためにはトークンを消費する必要がある．道路管理者は，ボトルネックでの渋滞を発生させないように，トークンを再配布するタイミングとして，ピーク時には T_{peak} 日後，オフピーク時には T_{off} 日後と設定している．このような条件下で，利用者は再配布タイミングと当日及び明日以降の効用に応じて今期の行動を選択する．

このメカニズムでは，道路管理者はあたかもロードプライシングにおける通行料金を調整するのと同様に，各時間帯枠に対応する再配布タイミングを調整することで交通需要をコントロールする．金銭支払いではなく，トークンの再配布タイミングという時間軸を用いることが本メカニズムの特徴である．

利用者はボトルネックを通行したい場合，事前にトークンを消費して申告をする．その申告を元に，道路管理者は容量に応じて申告者に通行権を確率的に割り当てる．割当者はボトルネックを通過し，非割当者は迂回路を通過する．この場合，需要が容量を超えることはなく，渋滞は発生しない．

(3) 文字の定義

次章以降のモデル化のために必要な文字を定義する．まず，利用者と道路管理者の状況設定に関する文字を定義する．時間によって変動しない利用者数を N，対象ボトルネックの容量を c とする．$t = 1$ からメカニズムによる施策が開始し，以降 1 日ごとに $t = 2, 3, \ldots$ とする．そして時間によって変動する各 t 期 ($=t$ 日目) のトークン保持者数を N_t とする．通過可能な時間帯枠を離散化し，時間帯枠の数を κ，各時間帯枠を $k \in \{1, 2, \ldots, \kappa\}$ とする．利用者がトークンを消費せず，迂回路を選択することを選択肢 $k = 0$ とすると，利用者の選択肢は $k \in \{0, 1, 2, \ldots, \kappa\} = K$ と表記できる．t 期に時間帯 k に対して利用者がトークンを消費した際に，通行権が割り当てられる確率を $\pi_{t,k}$ とする．

将来の不確実性を考慮するために，利用者のタイプ θ が日々，確率的に変動する状況を考える．なお，タイプを規定する確率密度関数 $f(\theta)$ は利用者全体で共通かつ既知とする．利用者のタイプは離散的なタイプ $\theta \in \{1, 2, \ldots\} = \Theta$ によって表現される．t 期にタイプ θ の利用者が選択肢 k を利用した際に得られる効用は $u_{t,\theta,k} \in \mathbb{R}^+$ とする．ただし迂回路を通行する選択肢 $k = 0$ の効用は，全ての期間 t，全てのタイプ θ において，$u_{t,\theta,0} = 0$ と基準化する．t 期のタイプ θ の利用者の効用ベクトルは $\boldsymbol{u}_{t,\theta} = (u_{t,\theta,1}, u_{t,\theta,2}, \ldots)$ で表す．

t 期に選択肢 k を選んだタイプ θ の利用者数は $x_{t,\theta,k}$，その利用者数ベクトルは $\boldsymbol{x}_{t,\theta} = (x_{t,\theta,0}, x_{t,\theta,1}, \ldots)$ で表す．同様に t 期のタイプ θ の選択肢 k の選択確率を $p_{t,\theta,k}$，選択確率ベクトルを $\boldsymbol{p}_{t,\theta} = (p_{t,\theta,0}, p_{t,\theta,1}, \ldots)$ で表す．タイプ θ に関わらず t 期に選択肢 k を選んだ利用者の総和を $x_{t,k} = \sum_{\theta \in \Theta} x_{t,\theta,k}$ と表す．各利用者は自身の期待効用を最大化するように行動すると仮定する．

次に，トークンの配布と消費に関する文字を定義する．t 期の時間帯枠 k の再配布タイミングを $T_{t,k}$，全時間帯枠の再配布タイミングの組合せを $\boldsymbol{T_t} = (T_{t,1}, T_{t,2}, \ldots)$ と表す．

(4) 協調促進トークンの役割

協調促進トークンの役割について説明する．トークンの保有にはオプション価値が存在し，それがトークン果たす役割として大きな意味を持つ．トークン保有のオプション価値とは，ある時点でトークンを保有している場合に得られる期待効用とトークンを保有していない場合に得られる期待効用の差分として表現される．例えば，当日トークンを消費すると，明日は迂回路しか利用できないため，明日の期待効用は 0 であるが，当日トークンを温存すると，明日はボトルネック通過から得られる期待効用が存在する．この 2 つの差がトークンのオプション価値であり，v_{token} として表す．今期にトークンを保持した場合と消費した場合の差は，時間割引率 β を用いて

$$\left(\sum_{t=1}^{T-1} \beta^t \right) v_{token}$$

と表される．この式は将来のオプション価値の割引現在価値の総和が今期のトークン消費によって失われることを意味する．この値を artificial payments と呼ぶ．以上より，当日のトークンの消費は，利用者の効用から上記の artificial payments を支払うことと等しいため，再配布タイミング T の調整により，道路管理者は artificial payments を調整することができる．この支払額を通して，交通需要のマネジメントが可能となる．

(5) 日々変動するトークン保持者数

動学化にあたって考慮すべき，日々変動するトークン保持者の数について説明する．トークンを消費したのちトークンが再配布されるまでには期間があるため，前日までのトークンの消費に対応して日々のトークンの保持者数は変動する．日々のトークン保持者数は，前日にトークンを消費しなかった人数と今日トークンが再配布される人数の合計となる．そのため t 期のトークン保持者数 N_t は式 (1) のように表せる．ここで $\mathbb{I}(\cdot)$ は括弧内を満たすなら 1，そうでなければ 0 を返す指示関数である．

$$N_t = x_{t-1,0} + \sum_{t'=1}^{t-1} \sum_{k \in K \setminus \{0\}} \mathbb{I}(T_{t',k} = t - t') \cdot x_{t',k} \quad (1)$$

4. 利用者の学習と行動

(1) 利用者の選択行動と必要とする情報

利用者は自身の期待効用を最大化するように選択を行うため，まず期待効用に関して述べる．ある時間帯枠に対象の道路を利用したい場合にはトークンの消費によって申告する．そして利用権は申告者の中から確率的に割り当てられる．この時 t 期にタイプ θ である利用者の選択肢 k の期待効用は，t 期の時間帯枠 k の利用可能確率 $\pi_{t,k}$ を用いて式 (2) で表される．なお $k = 0$ の期待効用は全タイプにおいて 0 に基準化する．

$$\begin{cases} EU_{t,\theta,k} = u_{\theta,k} \cdot \pi_{t,k} + u_{\theta,0} \cdot (1 - \pi_{t,k}) - \left(\sum_{t'=1}^{T_{t,k}-1} \beta^{t'} \right) \cdot v_{token\,t} \\ \qquad\qquad = u_{\theta,k} \cdot \pi_{t,k} - \left(\sum_{t'=1}^{T_{t,k}-1} \beta^{t'} \right) \cdot v_{token\,t} \quad \forall k \in K \setminus \{0\} \\ EU_{t,\theta,0} = u_{\theta,0} = 0 \end{cases} \quad (2)$$

式 (2) 中の利用可能確率 $\pi_{t,k}$ とオプション価値 v_{token} は式 (3),(4) で表される．

$$\pi_{t,k} = \begin{cases} \frac{c}{x_{t,k}} & \text{if } x_{t,k} \geq c \\ 1 & \text{otherwise} \end{cases} \quad \forall k \in K \setminus \{0\} \quad (3)$$

$$v_{token\,t} = \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t'=t+1}^{\tau} \sum_{\theta \in \Theta} \sum_{k \in K \setminus \{0\}} p_{t',\theta,k}\, u_{\theta,k}\, \pi_{t',k} \quad (4)$$

式 (2) より期待効用の算出には，利用可能確率 $\pi_{t,k}$ とオプション価値 v_{token} が必要である．しかし当日の各時間帯枠の利用者数を事前に正確に知覚できないため，利用可能確率を正確に計算できない．そのため，利用者は自身の持つ情報から推測する必要がある．さらに自身のオプション価値を正確に知覚するには，今期だけでなく将来の利用者数や利用可能確率も必要である．そこで，自身の選択のために利用者が推測する値を信念と呼ぶ．すなわち，利用者は自身の”信念”に基づき，日々自身にとって最適だと思う選択を取り，さらに新たな情報から学習して信念を更新すると仮定する．

(2) 提供される情報と信念

利用者自身のオプション価値 v_{token} と利用可能確率 $\pi_{t,k}$ を考えるにあたって，利用者が既知とする情報と管理者から提示される情報を整理する．まず，利用者は今期の自身の（タイプの）効用ベクトルを知っており，他の全タイプの効用ベクトルとタイプを規定する確率密度関数 $f(\theta)$ も既知である．利用者は時間によって変動しない総利用者数 N と供給容量 c も知っている．そして管理者は，$t+1$ 期の開始時に t 期の各時間帯枠のトークン消費者数を利用者に提示する．利用者はこれにより $t+1$ 期におけるトークンの保持者数を知る．

ここで，他利用者の選択肢 k の選択確率 $p_{t,k}^{believe}$ を信念とする．信念である選択確率 $p_{t,k}^{believe}$ の値を用いると，利用可能確率とオプション価値が推測できる．このとき利用可能確率は式 (5) のように推測され，オプション価値は，式 (6) のように推測されるとする．なお，自身の将来のオプション価値である式 (6) 中に他利用者の選択確率 $p_{t,k}^{believe}$ が現れるが，時間的に平均すれば将来の自身は翌日の他者と同じ確率で選択をすると考えられるため，自身の将来平均選択確率として $p_{t,k}^{believe}$ を用いている．また，ここでは敢えて $\pi_{t,k}^{believe}$ ではなく $\hat{\pi}_{t,k}^{believe}$ を用いている．理由としては，翌期のトークン保持者数によって値が大きく変動する $\pi_{t,k}^{believe}$ は，長期的にトークンの平均価値を考えるオプション価値には不適当であると考えたためである．ゆえに翌期のトークン保持者数による変動が少ない $\hat{\pi}_{t,k}^{believe}$ を用いる．これは学習率 α を用いて式 (8) で表す．

$$\pi_{t,k}^{believe} = \begin{cases} \frac{c}{N_t \cdot p_{t,k}^{believe}} & \text{if } N_t \cdot p_{t,k}^{believe} \geq c \\ 1 & \text{otherwise} \end{cases} \quad \forall k \in K \setminus \{0\} \quad (5)$$

$$v_{token\,t}^{believe} = \sum_{k \in K \setminus \{0\}} p_{t,k}^{believe} \bar{u}_k \hat{\pi}_{t,k}^{believe} \quad (6)$$

where

$$\bar{u}_k = \sum_{\theta \in \Theta} f(\theta)\, u_{\theta,k}, \quad (7)$$

$$\begin{cases} \hat{\pi}_{t+1,k}^{believe} = \alpha \pi_{t,k}^{believe} + (1-\alpha) \hat{\pi}_{t,k}^{believe} \\ \hat{\pi}_{1,k}^{believe} = \pi_{1,k}^{believe} \end{cases} \quad \forall k \in K \setminus \{0\} \quad (8)$$

(3) 利用者の選択・学習モデル

利用者の選択と学習に関して述べる．まず選択に関して，利用者は式 (2) に倣って，式 (9) にように期待効

用を求める．

$$\begin{cases} EU_{t,\theta,k} = u_{\theta,k} \cdot \pi_{t,k}^{believe} - \left(\sum_{t'=1}^{T_{t,k}-1} \beta^{t'} \right) \cdot v_{token\ t}^{believe} \\ \qquad\qquad\qquad\qquad \forall k \in K \setminus \{0\} \\ EU_{t,\theta,0} = 0 \end{cases} \tag{9}$$

そして，同一のタイプ θ の利用者は，求めた期待効用からロジット選択 (式 (10)) または最大値選択などのルールに基づいて時間帯枠を選択する．

$$x_{t,\theta,k} = \frac{\exp(\mu \cdot EU_{t,\theta,k})}{\sum_{k' \in K} \exp(\mu \cdot EU_{t,\theta,k'})} \tag{10}$$

選択の結果として各時間帯枠のトークン消費者の総数は下式で表せる．

$$x_{t,k} = \sum_{\theta \in \Theta} x_{t,\theta,k} \tag{11}$$

次に学習に関して，利用者は実際の選択結果から学習し，信念を更新する．学習はこれは学習率 α を用いて，式 (12) によって定式化される．なお，式 (5) 中の学習率 α と異なる値を用いても良いが，本研究では同一とする．

$$\boldsymbol{p}_{t+1}^{believe} = \alpha \frac{\boldsymbol{x}_t}{N_t} + (1-\alpha) \boldsymbol{p}_t^{believe} \tag{12}$$

利用者の選択・学習のアルゴリズムをまとめると下記のようになる．

Step.0 信念の初期値を設定.

Step.1 信念から式 (5),(6) を用いて利用可能確率とオプション価値を推測し，それらから各選択肢の期待効用を式 (9) により求める．

Step.2 期待効用からルール (最大値を選択またはロジット選択) に基づいて時間帯枠を選択する．

Step.3 実際の選択結果から式 (12) のような学習を行い，信念を更新する．

Step.4 トークン保持者数を式 (1) により更新．**Step.1** へ

5. 管理者の学習と行動

(1) 管理者の目的と観測可能な情報

管理者の目的と，目的関数の値を求めるのに必要な情報と，管理者が観測可能な情報の制限について述べる．管理者は，社会厚生の最大化を目的として再配布タイミングを制御し，道路の効率的な利用を図る．ここで本研究では社会厚生を利用者の効用の総和とする (式 (13))．つまり管理者が社会厚生を算出するには利用者の効用を知る必要がある．しかし管理者は利用者の効用を直接観測できないという情報の非対称性の問題を抱える．管理者が観測できるのは，利用者数と課した再配布タイミングに対する各時間帯枠のトークン消費者数のみである．すなわち管理者はそれらの情報をもとに何らかの信念を持ち，社会厚生を推測する必要がある．

$$SW_t = \sum_{\theta \in \Theta} \sum_{k \in K \setminus \{0\}} x_{t,\theta,k} u_{t,\theta,k} \pi_{t,k} \tag{13}$$

(2) 管理者の学習モデル

管理者は社会厚生を最大化したい一方でその算出に必要な利用者の効用が観測できないため，得られる情報から形成される信念から，社会厚生を推測する必要がある．まず管理者の学習と信念の更新，そして信念からの社会厚生の算出について述べる．

利用者のロジット選択を仮定すると各スロットの利用者数 $x_{t,k}$ とトークン保持者数 N_t から式 (14) で示される対数尤度関数を最大化することで，各スロットの期待効用 $EU_{t,k}^{believe}$ を推定する．なお $EU_{t,k}^{believe} = 0$ は既知とする．

$$\boldsymbol{EU}_t^{believe*} = \underset{\boldsymbol{EU}_t^{believe}}{\arg\max} \sum_{k \in K} \frac{x_{t,k}}{N_t} \log(\widetilde{P}_{t,k}) \tag{14}$$

where

$$\widetilde{P}_{t,k} = \frac{\exp(EU_{t,k}^{believe})}{\sum_{k' \in K} \exp(EU_{t,k}^{believe})} \quad \forall k \in K \tag{15}$$

これにより全選択肢の $EU_{t,k}^{believe*}$ が推定される．また，式 (3) より利用可能確率 $\pi_{t,k}$ が算出される．

これらを用いると，式 (16) の最適化問題を解くことで各時間帯枠の平均効用 $u_{t,k}^{believe}$ を算出できる．ここで式 (17) 中では敢えて $\pi_{t,k}$ ではなく $\hat{\pi}_{t,k}$ を用いている．今期のトークン保持者数によって値が大きく変動する $\pi_{t,k}$ ではなく，利用者がオプション価値に関して長期的なトークンの平均価値を想定することを管理者が予測すると考えたためである．ゆえに今期のトークン保持者数による変動が少ない $\hat{\pi}_{t,k}$ を用いる．これは学習率 γ を用いて式 (18) で表す．

$$\boldsymbol{u}_t^{believe} = \underset{\boldsymbol{u}_t^{believe}}{\arg\min} \sum_{k \in K} \left(EU_{t,k}^{believe} - \widetilde{EU}_{t,k}^{believe} \right)^2 \tag{16}$$

where

$$\begin{aligned} &\widetilde{EU}_{t,k}^{believe} \\ &= u_{t,k}^{believe} \pi_{t,k} - \left(\sum_{t'=1}^{T_{t,k}-1} \beta^{t'} \right) v_{token}^{believe} \\ &= u_{t,k}^{believe} \pi_{t,k} - \left(\sum_{t'=1}^{T_{t,k}-1} \beta^{t'} \right) \frac{1}{N_t} \sum_{k \in K \setminus \{0\}} x_{t,k} u_{t,k}^{believe} \hat{\pi}_{t,k} \\ &\qquad\qquad\qquad\qquad \forall k \in K \setminus \{0\} \end{aligned} \tag{17}$$

$$\begin{cases} \hat{\pi}_{t+1,k} = \gamma \pi_{t,k} + (1-\gamma)\hat{\pi}_{t,k} \\ \hat{\pi}_{1,k} \quad = \pi_{1,k} \end{cases} \quad \forall k \in K \backslash \{0\} \quad (18)$$

上記の式により各時間帯枠の平均効用 $u_{t,k}^{believe}$ が算出されるため, 式 (19) で定義する社会厚生 (Social Welfare, SW) により，管理者は t 期に利用者に課した再配布タイミングを評価することができる．

$$SW_t^{believe} = \sum_{k \in K \backslash \{0\}} x_{t,k} u_{t,k}^{believe} \pi_{t,k} \quad (19)$$

(3) 管理者の行動モデル

社会厚生の最大化を目指す管理者がそれを叶えるための行動について述べる．社会厚生の最大化を目指すには，$u_{t,\theta,k}$ は定数であるため，式 (13) より効用 $u_{t,\theta,k}$ の大きいタイプ θ，時間帯枠 k に関してトークン消費者数 $x_{t,\theta,k}$ を増やし，同時に効用 $u_{t,\theta,k}$ の小さいタイプ θ，時間帯枠 k に関してトークン消費者数 $x_{t,\theta,k}$ を減らすことで利用可能確率 $\pi_{t,k}$ を上げれば良い．すなわち通行者数が多いほど，そしてトークン消費者の中で効用の高いタイプの利用者の割合が高いほど社会厚生は大きくなる．ここで管理者の操作できる再配布タイミングについて述べると，式 (9),(10) から，ある時間帯枠の再配布タイミングを大きくするほど，低い効用を知覚する人はその時間帯枠を選ばなくなり，トークン消費者数は減って，効用の高いタイプの利用者の割合が高くなることがわかる．よって，各時間帯枠のトークン消費者数が供給容量より僅かに多くなるようにすれば，一般に社会厚生は大きくなる．従って管理者の行動原理は以下の 4 点であるとする．

- 供給容量とトークン消費者数の差が最も大きい時間帯枠に関して，トークン消費者数が大きい場合には再配布タイミングを 1 日遅らせる（T を 1 大きくする）.
- 供給容量とトークン消費者数の差が最も大きい時間帯枠に関して，トークン消費者数が小さい場合には再配布タイミングを 1 日早める（T を 1 小さくする）.
- 更新後の再配布タイミングにおける SW と更新前の SW を比較し，更新前の SW が更新後の SW より大きい場合，再配布タイミングを更新前に戻す.
- 再配布タイミングを更新前に戻した時間帯枠は，供給容量とトークン消費者数の差を考えず，他の時間帯枠の再配布タイミングが変更されない限り再配布タイミングを固定する.

ここで，社会厚生を最大化するには一度に再配布タイミングを更新することも考えられる．しかしその場合，管理者は未実施の再配布タイミングに関して社会厚生を推測する必要があるが，再配布タイミングが変わった場合トークン消費者数が変化するだけでなく，それに伴い平均効用も変化する．よって管理者にとって社会厚生の推測が困難である．また利用者に関しても，再配布タイミングの大きな変動がある場合には信念の形成が難しくなる．従って本研究では再配布タイミングを 1 つの時間帯枠毎に 1 期分ずつ増減させ，実施後に観測された利用者数から推測される社会厚生で再配布タイミングを評価することとする．

これらを踏まえて，管理者の行動のアルゴリズムを以下にまとめる．

Step.0 再配布タイミングの初期値を設定.

Step.1 設定した再配布タイミング下の利用者の選択を観測.

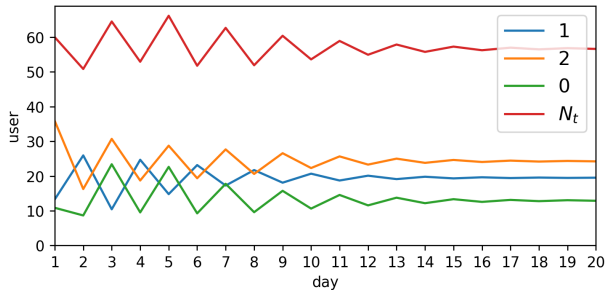
Step.2 利用者の実際の選択結果から，式 (15)〜(19) により設定した再配布タイミング $\boldsymbol{T}_t$ の SW を推測する.

Step.3 利用者の実際の選択結果から上記の行動原理によって再配布タイミングを更新．**Step.1** へ

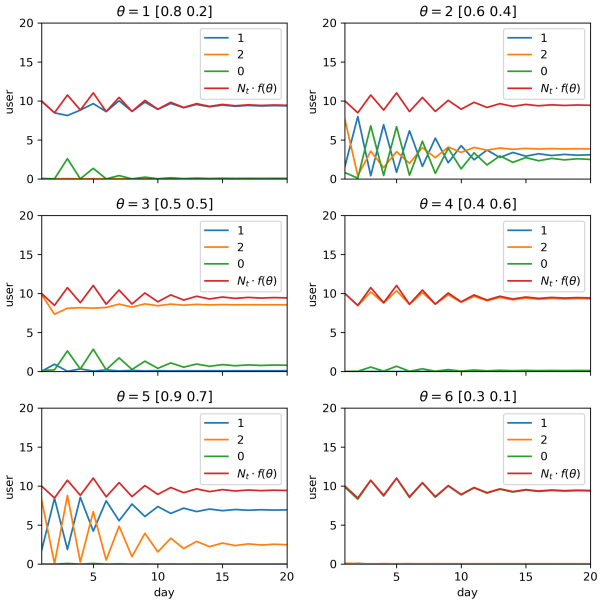
6. 数値実験

ここでは再配布タイミングを固定した場合の利用者の学習の数値実験を行う．パラメータ設定として，総利用者数 $N = 120$，供給容量を $c = 15$，時間帯枠数を 2, $|K| = 3$ とした．利用者のタイプ θ は 6 種類とし, それぞれ $\boldsymbol{u}_1 = (0, 0.8, 0.2)$, $\boldsymbol{u}_2 = (0, 0.6, 0.4)$, $\boldsymbol{u}_3 = (0, 0.5, 0.5)$, $\boldsymbol{u}_4 = (0, 0.4, 0.6)$, $\boldsymbol{u}_5 = (0, 0.9, 0.7)$, $\boldsymbol{u}_6 = (0, 0.3, 0.1)$ の効用ベクトルを持つ．また，各タイプの正規確率分布は $f(1) = f(2) = \cdots = f(6) = 1/6$ の等確率とした．再配布タイミングは時間帯枠ごとにそれぞれ 3 日後 2 日後とした．選択ルールはロジット選択で，サイズパラメータ $\mu = 30$ とした．学習率は $\alpha = 0.4^{(1+t/4)}$ とし，初めは新たな情報を多く取り入れ，時間の経過に伴い過去を重視すると仮定した．初期の信念 $\boldsymbol{p}_t^{believe}$ は $p_{t,k}^{believe} = N/(\kappa + 1)$ $\forall k \in K$ とした．以下では第 1 期から 20 期までの様々な値の推移について表した図について説明する．なお，変動を防ぐために 1 期に N/2，2 期に N/3，3 期に N/6 の利用者にトークンを配布している．

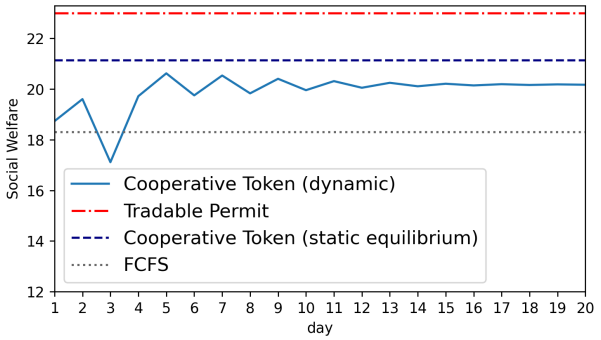
図-**2** は，第 1 期から 20 期までの各時間帯枠の選択者数とトークン保持者数の推移である．この図から，時間の経過とともに各時間帯枠の選択者数やトークン保持者数は収束することがわかる．図-**3** は，各タイプにおける第 1 期から 20 期までの各時間帯枠の選択者数とトークン保持者数の推移である．この図からも，初期は各値が変動するが，時間の経過とともに収束していくことがわかる．また，各タイプの利用者は自身の知



図–2 ロジット選択の場合における各時間帯枠の利用者数の推移



図–3 ロジット選択の場合における各タイプの利用者数の推移



図–4 ロジット選択の場合における他メカニズム等との SW の比較

覚する期待効用に応じて選択しており，例えば $\theta = 1$ では時間帯枠 1 の効用が大きいため，全員が選択肢 1 を選んでいることがわかる．一方で $\theta = 6$ ではいずれの時間帯枠で知覚する効用も小さいため全員が選択肢 0 を選んでいることがわかる．よって本メカニズムは，自身の効用に応じて選択をさせることで道路空間の効率的利用に寄与できていると言える．

図-4 は，第 1 期から 20 期までの SW を，理論最適値である通行権オークション制度 (Tradable Permit)，早い者勝ち (FCFS)，そして $N = N_{30}$ としたときの静的な均衡時における協調促進トークンメカニズム (Cooperative Token (static)) の SW と比較したものである．この図から本研究で考えている動的なトークンメカニズムの SW は静的な均衡時における SW に劣っていることがわかる．その理由は主に 2 つ考えられる．1 つ目にロジット選択を仮定したことにより厳密に均衡していないためである．同一タイプ内において，選ばれた時間帯枠の期待効用は等しいという均衡条件はロジット選択を考える以上成り立たないことがわかるだろう．2 つ目に利用者の推測するオプション価値 v_{token} が過小に推測される点である．式 (6),(7) より利用者は持ちうる効用の平均をオプション価値の計算に用いると仮定しているが，本来オプション価値は選択の結果知覚する効用の平均であり，図-3 で見たように利用者はより効用の大きくなるような選択をするため，仮定した値よりも大きくなるはずである．この誤差により再配布タイミングによる効果が減少して効用の低い利用者がトークンを消費することで，SW が小さくなっていると考えられる．一方でこの図からは，動学を考慮した協調促進トークンメカニズムは前述のように少し小さく算出されている SW でも早い者勝ちと比較して優位なメカニズムとなっていることもわかる．

7. おわりに

本研究の成果をまとめる．まず全体として協調促進トークンメカニズム特有の，日々の利用候補者 (=トークン保持者) の変動を考慮した Day-to-Day ダイナミクスの検討を行なった．具体的には，メカニズム下の利用者が選択行動のために他利用者の行動について信念を持ち学習することをモデル化した．そしてメカニズム下の管理者が観測できない情報を推定し，再配布タイミングを調整し需要を制御する方法を検討した．最後に再配布タイミング固定時における利用者の行動に関する数値実験により，利用者のモデルの性質や動学を考える場合の本メカニズムの有効性と課題を示した．このように，Day-to-Day ダイナミクスを考える場合においても協調促進トークンメカニズムは，利用者の効率的な割り当てに寄与することが明らかとなったが，本研究には様々な課題が存在する．1 つ目に，6 章の中で述べた通り利用者がオプション価値を過小に推定するという仮定を置いてしまっている．この点に関しては改めてよく検討を重ねることとする．2 つ目に 5 章で管理者の学習と行動を定式化した一方で数値実験には至らなかった点である．これはメカニズムの挙動をより詳細に明らかにするためにも早急に行うべきであると考える．3 つ目に，利用者数などが収束するのに長い期間を要する点である．これは管理者が再配布タイミ

ングを調整する過程を含めた数値実験を行いつつ，管理者の行動モデルを再検討していくことにより解決を図る．

謝辞：本研究は JSPS 科研費 (21H01450) の助成を受けた研究の一部です．ここに記し，感謝を表します．

REFERENCES

1) Vickrey, W. S.: Congestion theory and transport investment, The American Economic Review, Vol.59, No.2, pp.251–260, 1969.
2) 赤松 隆，佐藤 慎太郎，Nguyen Xuan Long: 時間帯別ボトルネック通行権取引制度に関する研究, 土木学会論文集D, Vol.62, No.4, pp.605–620, 2006. [Akamatsu, T., Sato, S., Nguyen, X.L.: Tradable time-of-day bottleneck permits for morning commuters, *JSCE Journal of Infrastructure Planning,* Vol.62, No.4, pp.605-620, 2006.]
3) Akamatsu, T. and Wada, K.: Tradable network permits: A new scheme for the most efficient use of network capacity, Transportation Research Part C: Emerging Technologies, Vol.79, pp.178–195, 2017.
4) 高野壮稀，原祐輔: 道路空間の効率的利用に寄与する協調促進トークンメカニズムの提案と解析，土木計画学研究・講演集, Vol. 65, 2022.
5) Horowitz, J. L.: The stability of stochastic equilibrium in a two-link transportation network, Transportation Research Part B: Methodological, Vol.18, No.1, pp.13–28, February 1984.
6) 小林 潔司，藤高 勝己: 合理的期待形成を考慮した経路選択モデルに関する研究, 土木学会論文集, Vol.1993, No.458, pp.17–26, 1993. [Kobayashi, K. and Fujitaka, K.: Route choice models with endogenous rational expectations formation, *Doboku Gakkai Ronbunshu*, Vol.1993, No.458, pp.17–26, 1993.]
7) 中山 晶一朗，藤井 聡，北村 隆一: ドライバーの学習を考慮した道路交通の動的解析, 土木計画学研究・論文集, Vol.16, pp.753–761, 1999. [Nakayama, S., Fujii, S. and Kitamura, R.: Drivers' learning and road transportation system behavior, *INFRASTRUCTURE PLANNING REVIEW,* Vol.16, pp.753–761, 1999.]
8) Iida, Y., Akiyama, T. and Uchida, T.: Experimental analysis of dynamic route choice behavior, Transportation Research Part B: Methodological, Vol.26, No.1, pp.17–32, 1992.
9) Li, M. Z.: 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, September 2002.
10) Sandholm, W. H.: Evolutionary Implementation and Congestion Pricing, The Review of Economic Studies, Vol.69, No.3, pp.667–689, July 2002.
11) 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, 22nd EURO Working Group on Transportation Meeting, EWGT 2019, 18th – 20th September 2019, Barcelona, Spain.

(Received 2023. 10. 13)

Trial-and-Error Control of a Cooperative Token Mechanism

Soki TAKANO and Yusuke HARA