# 行動クローニングによる事前学習を用いた強化学習に基づくタクシー経路推薦手法

甲斐 大雅† 石黒 慎† 三村 知洋† 神山 剛†

†長崎大学 〒852-8521 長崎市文教町 1-14

E-mail: †taiga.kai@k.cis.nagasaki-u.ac.jp, †{ishiguro, mimura}@db.nagasaki-u.ac.jp,

†kami@nagasaki-u.ac.jp

**あらまし** 本研究では，深層強化学習を用いたタクシーの流し営業における道路レベルでの経路推薦において，ドライバーの運行履歴に基づく事前学習を組み合わせた手法を提案する．道路レベルでは経路の選択肢が膨大かつ乗車需要が時空間的に偏在しているため，ランダムな経路探索では報酬を得る機会が乏しく学習が進みづらい．本手法は，乗客獲得につながる経路選択を事前に学習し初期方策とすることで，学習初期の探索効率の向上を図る．名古屋市の道路ネットワークと実際のタクシー運行データを用いた実験の結果，事前学習による学習効率の改善が確認された．しかし，ドライバーが周辺部に滞留し高需要な中心部に帰ることができず，性能が低下することが明らかになった．本稿では，これらの実験結果と分析を示し，今後の課題を述べる．

**キーワード** 強化学習，タクシー経路推薦，行動クローニング

## 1. はじめに

近年，タクシードライバーの高齢化が進み，人手不足が深刻化している[1]．ベテランドライバーが減少する一方で，新人ドライバーの早期戦力化が求められている．特に，路上で乗客を直接発見して獲得する「流し営業」においては，「どの時間帯に，どのエリアの，どの道路を走れば乗客を獲得しやすいか」という暗黙知をベテランドライバーは長年の経験から身につけており，新人ドライバーがこの知識を習得するには長い時間を要する．したがって，経験によらず効率的な乗客獲得を可能にする技術的支援が求められる．これに対し，タクシー乗車需要を予測し，どの時間帯にどのエリアでどの程度の需要が見込まれるかをドライバーに提示するサービスが提供されている[2][3][4]．しかし，エリア毎の需要予測だけでは，具体的にどの道を走るべきかという行動指針を得ることは難しく，道路レベルで走行経路を推薦するアプローチが求められる．

道路レベルでの経路推薦に対し，深層強化学習を用いたアプローチが注目されている．深層強化学習では，シミュレーション環境上で一定期間のシナリオに基づき乗客を発生させ，その状況下でタクシーを走行させる．最初はどの経路が良いかわからないため，ランダムな経路選択から始め，同一シナリオでのシミュレーションを繰り返す中で，乗客獲得の経験をもとに，乗客を獲得しやすい経路選択を学習していく．しかし，道路レベルでは経路の選択肢が膨大であり，かつ流し営業ではその道路に到達して初めて乗客の有無がわかるため，ランダムな経路選択では乗客の獲得が難しく学習が進みづらい．最適な経路選択の学習には非常に多くのシミュレーションを要し，限られた計算資源では実用的な精度の達成が困難となる．

この課題に対処するため，本研究では運行履歴に基づく事前学習による学習効率化を提案する．実際に乗客を獲得できた経路を含むドライバーの運行履歴をシミュレーション上で再現することで，乗客獲得につながる経路選択をあらかじめ学習させる．この事前学習した経路選択から深層強化学習を開始することで，ランダムな経路選択からではなく，有望な経路に絞って学習を開始でき，学習の効率化が期待される．本研究では，名古屋市の道路ネットワークと実際のタクシー運行データを用いて提案手法の有効性を検証した．

実験の結果，提案手法は事前学習なしと比較して学習効率の改善は見られたものの，学習した経路選択による乗客獲得率は低く，十分な性能は得られなかった．原因分析を行ったところ，深層強化学習の設計上の問題により，ドライバーが周辺部に滞留し中心部の学習が進まないことが性能低下の主な要因であることがわかった．本稿では，これらの実験結果と分析を示し，今後の課題を述べる．

## 2. 関連研究

タクシー運行最適化の研究は，経路推薦（route recommendation），車両再配置（repositioning），配車（dispatching），フリート管理（fleet management）など様々な文脈で行われてきた．これらは営業エリアの区分（エリア，グリッド，道路），営業形態（配車アプリによる営業，流し営業），対象とするドライバーの数（単一，複数）に違いはあるものの，本質的には「タクシーをどこへ向かわせるか」という共通の問題を扱っている．そのため，異なる文脈の研究であっても本研究に関連する知見を提供しうる．本章では，この観点か

ら関連研究を概観し，行動クローニングによる事前学習を用いた深層強化学習という本研究のアプローチの位置づけを明らかにする．

### 2.1. ルールベース・最適化・需要予測に基づく手法

タクシー運行最適化に対する初期の研究では，乗客の配車依頼に対して最も近いタクシーを割り当てるなどのルールベース手法[5][6]や，組合せ最適化に基づく配車手法[7][8]が提案されてきた．また，需要予測に基づくアプローチとして，Yuan et al.[9]は GPS 軌跡から乗客獲得確率を予測し経路を推薦するシステムを提案し，DeNA[10]は需要予測と供給量を考慮した道路レベルの経路推薦を商用化した．

しかし，これらの手法は現時点の需給状態に基づいて行動を決定するため，現在の経路選択が将来の車両分布に与える影響を考慮できないという課題があった．

### 2.2. 深層強化学習に基づく手法

2.1 節で述べた課題を解決するため，深層強化学習に基づくアプローチが提案されてきた．深層強化学習は，現在の判断が将来の状況に影響を与えることを考慮しながら行動を学習できるため，現時点の需給状態のみに基づく意思決定が長期的に非効率となるという課題に対して，有効な枠組みである．

Lin et al.[11]は，大規模配車プラットフォームにおけるフリート管理問題に対して，マルチエージェント深層強化学習を適用した．地理的・協調的文脈を考慮することで効率的な車両再配置を実現したが，地図をグリッドに分割し，各グリッドセルを行動空間として定義している．Xu et al.[12]は，学習と計画を組み合わせた配車手法を，Li et al.[13]は，平均場近似によるエージェント間相互作用のモデル化をそれぞれ提案した．これらの研究は，大規模な配車プラットフォームにおいて有効性を示したが，いずれもグリッドベースの環境表現を採用している．グリッドベースのアプローチでは，実際の道路ネットワークの構造が抽象化されるため，「どの道路を走るべきか」という具体的な行動指針を得ることが困難である．

### 2.3. 深層強化学習手法における道路ネットワークの拡張

Kim & Kim[14]は，道路ネットワークをグラフとして表現し，グラフニューラルネットワーク(GNN)とマルチエージェント深層強化学習を組み合わせたフリート管理手法を提案した．各道路の Q 値を GAT で近似することで，道路レベルでの経路選択を実現した．

しかし，1 章で述べたように，道路レベルでは経路の組合せが膨大かつ報酬が疎であるため，ランダムな探索からでは学習が進みづらいという課題がある．

### 2.4. 行動クローニングによる事前学習

1 章で述べたように，この課題に対しては，ドライバーの運行履歴を用いた事前学習により，有望な経路から学習を開始するアプローチが考えられる．Gammelli et al.[15]は，行動クローニングによる事前学習とオンライン強化学習を組み合わせたフリート管理手法を提案し，事前学習により学習ステップ数を大幅に削減できることを示した．ただし，同研究の行動空間は地域レベルの移動にとどまっており，道路レベルでの経路推薦は扱われていない．

### 2.5. 本研究の位置づけ

タクシー運行最適化の研究は，ルールベース手法から深層強化学習へと発展し（2.1 節，2.2 節），さらに道路レベルでの経路推薦が実現されてきた（2.3 節）．しかし，道路レベルでは行動の組合せが膨大かつ報酬が疎であるため，学習初期のランダムな経路選択では学習が進みづらい課題がある．行動クローニングによる事前学習はこの課題に対して有効であるが，既存研究では道路レベルの行動空間には適用されていない（2.4 節）．本研究は，道路レベルでの経路推薦に行動クローニングによる事前学習を適用し，ランダムな経路選択に代えて，運行履歴に基づく事前学習から深層強化学習を開始することで，学習の効率化を実現する．

## 3. 問題設定と提案手法

本章では，まず深層強化学習による経路推薦の枠組みを説明し（3.1 節），次に提案手法である運行履歴に基づく事前学習について述べる（3.2 節）．なお，3.1 節で述べる深層強化学習の枠組みおよびシミュレータの実装は，Kim & Kim[14]が提案した手法および公開しているソースコードに基づく．

### 3.1. 深層強化学習による経路推薦

#### 3.1.1. シミュレーション環境

1 章で述べたように，深層強化学習ではシミュレーション環境上でタクシーを走行させ，乗客獲得の経験をもとに経路選択を学習する．本節では，このシミュレーション環境について述べる．

シミュレーション環境は，実際の道路ネットワークを有向グラフとして表現し，その上で複数のタクシーが同時に走行する．シミュレーションは 1 分を 1 タイムステップとして進行し，各タイムステップにおいて道路の終点に到達したドライバーは交差点で次に進む道路を選択する．この選択が学習の対象である．

乗客の発生は，過去の乗車実績データから作成した 1 週間分のオーダー発生シナリオに基づく．流し営業ではドライバーはその道路に到達するまでオーダーの発生を知ることができない．ドライバーがオーダーのある道路に到達すると，同じ道路上にいる空車ドライバーにランダムに割り当てられる．一定の待ち時間内に割り当てられなかったオーダーは破棄される．1 週間分のシナリオを用いた 1 回のシミュレーションを 1 エピソードと呼ぶ．

### 3.1.2. マルコフゲームとしての定式化

本問題では，複数のドライバーが同じ道路ネットワーク上で同時に乗客を探索する．同じ道路に複数のドライバーがいる場合，オーダーはその中からランダムに 1 台に割り当てられるため，ドライバー同士が乗客を奪い合う状況が生じる．このような複数ドライバー間の競合を扱うため，本問題を

マルコフゲーム$G:=(N,S,A,P,R,\gamma)$として定式化する．ここで，$N$，S，A，P，R，γはそれぞれ，エージェント数，状態空間，行動空間，遷移確率，報酬関数，割引率を表す．

道路ネットワークは有向グラフ$G_R:=(V,E)$表現し，V は交差点の集合，E:={lj|j=1,2,…,Nroad}は道路の集合とする．以下に各構成要素を定義する．

**エージェント**：空車状態のドライバーをエージェントとして定義する．エージェント数はドライバーの出退勤により変動するため，時刻tにおけるエージェント数を $\mathrm{N_t}$とし，エージェントiが時刻tにいる道路をと$\mathrm{l_t^i}$表記する．同じ道路・同じ時刻にいる制御可能なエージェントは同質であると仮定し，同一の方策を共有する．

**状態**$\mathrm{s_t}\in S$：各道路$\mathrm{l_j}$について，エージェント数$\mathrm{N_{j,t}}$，オーダー数$\mathrm{N_{j,t}^{call}}$，速度$\mathrm{speed_{j,t}}$の 3 つの値を観測する．時刻tにおける状態$\mathrm{s_t}$は，これらをすべての道路について連結したものとして定義する:

$$\mathrm{s_t}:=\left[\left(\mathrm{N_{j,t}},\mathrm{N_{j,t}^{call}},\mathrm{speed_{j,t}}\right)\right]_{\mathrm{j=1}}^{\mathrm{N_{road}}}$$

**行動**$\mathrm{a_t}\in\mathbf{A}$：エージェントiの行動$\mathrm{a_t^i}$を，時刻tにおける次の道路の選択として定義する．道路$\mathrm{l_j}$から道路$\mathrm{l_k}$への移動を$\mathrm{l_j}\to\mathrm{l_k}$と表記する．各道路は異なる接続先を持つため，選択可能な行動の集合は道路ごとに異なる．

**報酬**$\mathbf{R_t}\in\mathbf{R}$：エージェントiの報酬 $\mathrm{R_t^i}$は，オーダーが割り当てられた場合に 1，そうでない場合に 0 とする．

**方策**：道路レベルの方策$\pi(\mathrm{l_j}\to\mathrm{l_k}|\mathrm{s_t})$を，状態$\mathrm{s_t}$において道路$\mathrm{l_j}$から道路$\mathrm{l_k}$へ移動する確率として定義する．

**状態遷移確率**P：現在の状態$\mathrm{s_t}$で行動$\mathrm{a_t}$を選択したとき，次の状態$\mathrm{s_{t+1}}$になる確率を表す．本問題では，ドライバーが選択した道路へ移動すること自体は決定論的である．ただし，移動先の道路における相対位置のランダムな設定，オーダーの割り当て，ドライバー数の調整などにより，次のタイムステップにおける各道路のドライバー数やオーダー数には確率的要素が生じる．

**割引率**$\boldsymbol{\gamma}$：将来の報酬をどの程度重視するかを決めるパラメータであり，0 から 1 の間の値をとる．$\gamma$が 1 に近いほど将来の乗客獲得を重視し，0 に近いほど目先の乗客獲得を重視する．

### 3.1.3. DQN による経路選択の学習

本研究では，Deep Q-Network（DQN）を用いて経路選択を学習する．

**Q 値と Q 関数**

Q 値とは，ある状態においてある道路を選択した場合に，その後も走行を続けることで将来得られる報酬の見込みを数値化したものである．Q 値が高い道路ほど，その道路を選ぶことで将来的に乗客を獲得しやすいと評価される．

Q 値をすべての状態と道路の組み合わせに対して求める関数を Q 関数と呼ぶ．Q 関数は，状態$\mathrm{s_t}$において道路$\mathrm{l_j}$から道路$\mathrm{l_k}$へ移動し，以降も走行を続けた場合の割引累積報酬の期待値として定義される：：

$$\mathrm{Q^{\pi}}(\mathrm{s_t},\mathrm{l_j}\to\mathrm{l_k};\theta)$$
$$:=\mathrm{E^{\pi}}\left[\mathrm{G_t^i}\,\middle|\,\text{エージェント i が時刻 t に}\mathrm{l_j}\to\mathrm{l_k}\text{を選択}\right]$$

ここで，$\mathrm{G_t^i}:=\sum_{\mathrm{k=0}}^{\infty}\gamma^{\mathrm{k}}\mathrm{R_{t+k}^i}$は割引累積報酬で,θは Q 関数を近似するニューラルネットワークのパラメータである．道路の数が膨大であるため，すべての状態と道路の組み合わせに対して Q 値を表として保持することは困難であり，ニューラルネットワークを用いて Q 関数を近似する必要がある．本研究では，道路ネットワークがグラフ構造を持つことから，グラフ上で隣接ノードの情報を集約して各ノードの値を推定できる Graph Attention Network（GAT）を用いる．

**Q 値の更新**

学習は同一シナリオでエピソードを繰り返すことで進行する．学習初期ではすべての道路の Q 値がほぼ等しいため，経路選択はランダムに近い状態から始まる．各タイムステップにおいて，

エージェントは状態$\mathrm{s_t}$を観測し，次の道路を選択し，報酬$R_t^i$を得る．この結果に基づいて，Q 値は以下のターゲット値$\mathrm{y_t^i}$に近づくよう更新される：

$$y_t^i=\begin{cases}R_t^i=1 & \textit{（オーダーを獲得した場合）}\\ \gamma\hat{Q}^{\acute{\pi}}(s_{t+1},l_{t+1}^i;\acute{\theta}) & \textit{（それ以外の場合）}\end{cases}$$

ここで，$l_{t+1}^i$はエージェント$i$が時刻$t+1$にいる道路，$\hat{Q}^{\acute{\pi}}$は学習の安定化のために一定間隔でパラメータθからコピーされるターゲットネットワーク，$\acute{\theta}$である.乗客を獲得できた場合，その道路の Q 値は報酬 1 に向かって更新され高くなる．獲得できなかった場合，その道路の Q 値は，次の道路の中で最も Q 値が高い道路の値を$\gamma$で割り引いた値に向かって更新される．これにより，乗客獲得に成功した道路だけでなく，その手前の道路の Q 値も上昇する．

あるエピソードで学習された Q 値は次のエピソードに引き継がれる．次のエピソードでは，更新された Q 値に基づいてより有望な経路が選択されやすくなり，新たな乗客獲得の経験が得られることで，さらに Q 値が更新される．このように，エピソードを繰り返す中で乗客獲得の経験に基づく Q 値が積み重なり，乗客を獲得しやすい経路の学習が進んでいく．

**方策**

学習された Q 値に基づいて，各交差点でどの道路を選択するかを決定する規則を方策と呼ぶ．方策$\pi(l_j \to l_k|s_t)$を，状態$s_t$において道路$l_j$から道路$l_k$へ移動する確率として定義する．

単純に Q 値が最も高い道路だけを選ぶと，すべてのドライバーが同じ道路に集中し，オーダーの奪い合いが激化してかえって乗客獲得が困難になる．そこで，Q 値に応じた確率で道路を選択する確率的方策を用いることで，有望な道路に多くのドライバーを向かわせつつ，複数の道路に分散させる：

$$\pi(l_j \to l_k|s_t) = \frac{Q^{\pi}(s_t, l_k)^{\beta}}{\sum_{l_m \in S(l_j)} Q^{\pi}(s_t, l_m)^{\beta}}$$

ここで，$S(l_j)$は道路$l_j$の後続道路の集合，$\beta$は温度パラメータである．$\beta$が大きいほど Q 値の高い道路に集中し，小さいほど均等に分散する．なお，この設計では各交差点において 1 ステップ先の道路の Q 値のみに基づいて経路を選択する．

### 3.2. 提案手法

DQN による学習初期では Q 値が未学習であるため，経路選択はランダムから始まる．道路レベルでは経路の組合せが膨大であり，ランダムな経路選択で乗客を獲得できる機会は極めて限られるため，Q 値の学習がなかなか進まない．

そこで本研究では，行動クローニングによりドライバーの運行履歴から Q 値を事前に学習し，その Q 値を初期値として深層強化学習を行う手法を提案する．タクシードライバーの運行履歴には，実際に流し営業を行い，結果として乗客を獲得できた経路が含まれている．これらの経路は必ずしも最適ではないが，ランダムな経路選択と比較すれば，乗客獲得につながる可能性が高い経路である．ただし，行動クローニングのみではドライバーが走行した経路しか学習できないため，さらに深層強化学習を行うことで経路選択の最適化を期待する．提案手法は以下の 2 段階で構成される．

**第 1 段階（事前学習）**：ドライバーの運行履歴をシミュレーション上で再現し，その走行結果に基づいて Q 値を事前に学習する．具体的には，3.1.1 節のシミュレーション環境上で，運行履歴に記録された各ドライバーの走行経路をそのまま再現する．事後学習では Q 値に基づいて次の道路を選択するのに対し，事前学習ではドライバーが実際に選択した道路をそのまま使用する点が異なる．シミュレーション上でのオーダーの発生・割り当ておよび Q 値の更新は 3.1.3 節と同一であり，再現した走行の結果オーダーが割り当てられれば，その経路の Q 値が更新される．事前学習には全ドライバーの運行履歴を用いることで，広範囲の道路に対して Q 値を学習させる．これにより，乗客獲得につながる経路に対して高い Q 値が学習された状態の Q 関数が得られる．

**第 2 段階（事後学習）**：事前学習で獲得した Q 値を初期値として，3.1.3 節の DQN による深層強化学習を行う．事前学習により乗客獲得につながる経路の Q 値が既に高いため，Q 値に基づく経路選択がはじめから有望な経路を選択しやすくなる．その結果，乗客獲得の機会が増え，Q 値の更新が速く進むことで，限られたエピソード数でも効率的な学習が期待される．さらに，事後学習ではシミュレーション上で様々な経路を試行するため，ドライバーの経験だけでは気づけなかった有効な経路の発見や，経路選択のさらなる最適化が期待される．

## 4. 実験

本章では，提案手法の有効性を検証する．まず評価方法を述べ（4.1 節），次に結果を示す（4.2 節）．

### 4.1. 評価方法

本研究の目的は，事前学習によって事後学習の学習効率が改善されるかを検証することである．そこで，事後学習のエピソードごとの性能推移（学習曲線）を比較する．事前学習により，より少ないエピソード数で高い性能に到達できれば，学習効率が改善されたと判断できる．

性能指標には，Hit Rate（総乗車発生応答率）を用いる．Hit Rate は，シナリオ内で発生した全乗車のうち，ドライバーが対応できた割合である．学習によって経路選択が改善されれば，より多くの乗車に対応できるようになるため，Hit Rate は経路選択の良さを直接的に反映する指標である．

以上を踏まえ，以下の 2 手法について Hit Rate の学習曲線を比較する．データは，名古屋市を拠点とし約 1,400 台のタクシーを保有するタクシー会社から提供を受けた運行データを使用した．事前学習期間（2023 年 4 月 1 日～4 月 7 日）と事後学習期間（4 月 8 日～4 月 14 日）の各 1 週間に分割した．タクシーの需要には曜日ごとの周期性があるため，1 週間を 1 つのシナリオの単位とした．

**提案手法（事前学習＋事後学習）**：3.2 節で述べた 2 段階の手法である．事前学習期間のシナリオを用いて行動クローニングによる事前学習を 1 エピソード行い（第 1 段階），その後，事後学習期間のシナリオを用いて深層強化学習を 50 エピソード行う（第 2 段階）．

**ベースライン（事後学習のみ）**：事前学習を行わず，Q 値が未学習の状態から事後学習期間のシナリオを用いて深層強化学習を 50 エピソード行う．3.2 節で述べた課題，すなわちランダムな経路選択から学習を開始する手法に相当する．両者の違いは事前学習の有無のみであり，事後学習の条件は同一である．

### 4.2. 実験結果

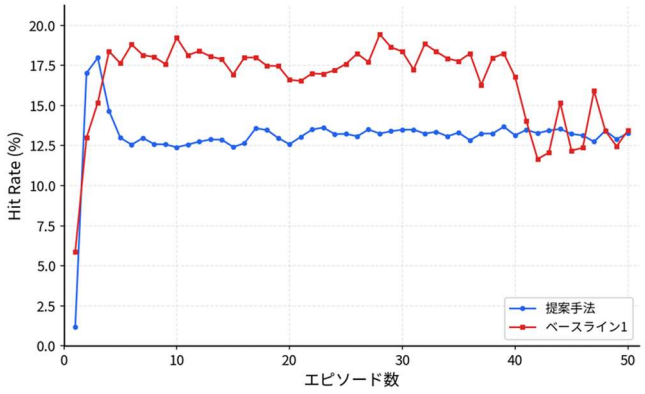


**図 1　提案手法とベースラインの学習曲線**

図 1 から，提案手法はベースラインよりも速く 18%程度のピークに到達しており，事前学習による学習効率の改善効果が確認できる．しかし，ピークに到達した直後から Hit Rate は低下し，13%程度で収束した．

ベースラインは，提案手法に遅れてピークまで上昇したが，同様にピーク到達後に Hit Rate が低下し，収束した．

本来，学習が進むと，Hit Rate はさらに上昇していくことが期待される．しかし，Hit Rate のピークが約 18%と低い水準に留まっており，学習によって十分な性能に到達できていない．さらに，学習を継続するとかえって Hit Rate が低下している．次章では，これらの原因を分析する．

## 5. 分析

4 章では，Hit Rate のピークが低い水準に留まること，および学習を継続すると Hit Rate が低下することが確認された．本章では，これらの原因を分析する．まず，学習後の Q 値の空間分布を比較し（5.1 節），次にシミュレーション中のドライバーの空間分布を分析する（5.2 節）．最後に，これらの分析結果を統合し，Hit Rate が頭打ちとなり低下に転じるメカニズムを明らかにする（5.3 節）．

### 5.1. Q 値の空間分布

提学習によってどのような経路選択が獲得されたかを把握するため，各手法の学習終了後における Q 値の空間分布を比較する．Q 値は各道路における乗客獲得の期待値を表すため，Q 値の分布は学習された方策の特性を反映する．図 2〜4 に，事前学習後，提案手法の学習後，ベースライン（事後学習のみ）の学習後の Q 値の空間分布をそれぞれ示す．赤色に近いほど Q 値が高く乗客獲得の期待が大きい道路を，青色に近いほど Q 値が低い道路を表す．

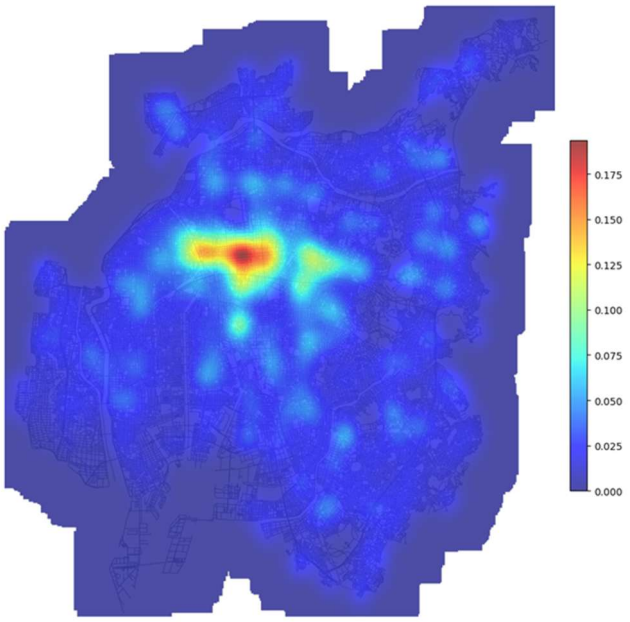

**図 2 提案手法(事前学習後)Q 値空間分布**

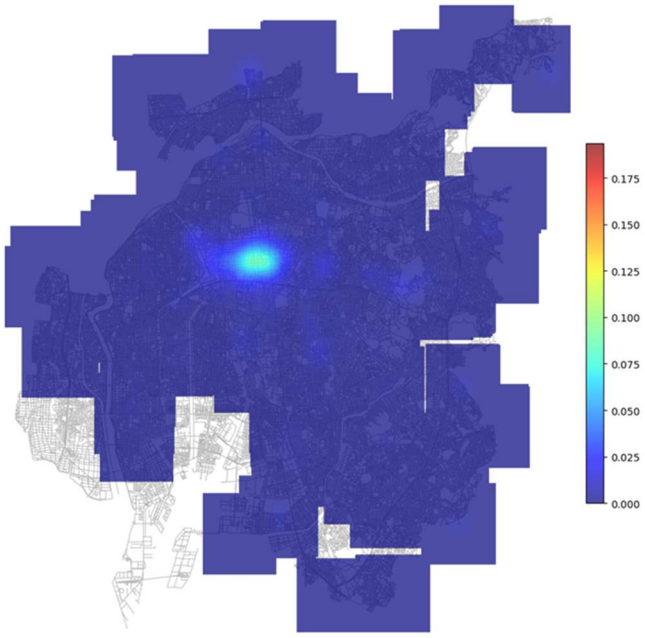

**図 3 提案手法(事後学習後)Q 値空間分布**

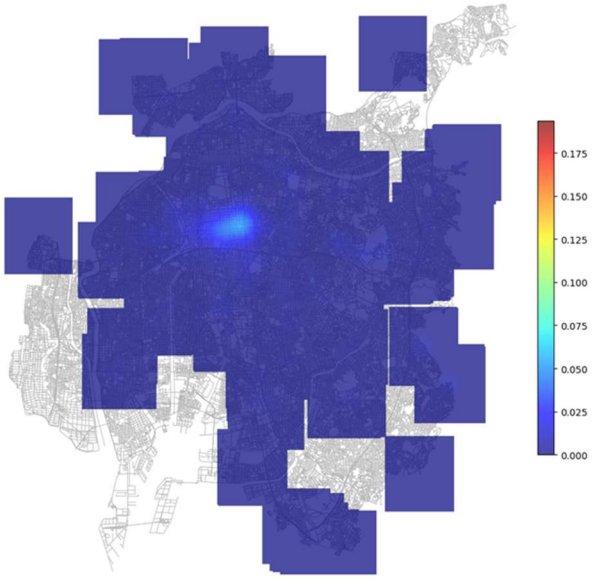


図 4 ベースライン(事後学習後)のQ値分布

事前学習後（図 2）では，高需要な中心部の Q 値が高く，周辺部に向かうにつれて Q 値が低下する分布が形成されている．この分布は，実際の名古屋市における乗客発生の空間分布と一致しており，事前学習によって実データに基づく需要分布が Q 関数に反映されていることを示している．

一方，提案手法の学習後（図 3）では，事前学習で形成されていた広範囲の Q 値と比較して，全体的に Q 値が薄くなっており，特に中心部の Q 値が大幅に低下している．この分布は，ベースラインの学習後（図 4）とほぼ同様の分布となっている．すなわち，事前学習で獲得された中心部の高い Q 値および周辺部の Q 値が，事後学習の過程で失われている．

### 5.2. 学習効率の比較

次に，Q 値の分布がドライバーの行動にどのように影響しているかを確認するため，シミュレーション中のタクシーの空間分布を分析する．

実際の運行データでは，タクシーは主に中心部に集中している．名古屋市では中心部の需要が高いため，中心部で乗客を獲得し，周辺部で降車させた後に中心部へ戻るという移動パターンが多い．事前学習では，この運行履歴に基づいて Q 値を学習している．

図 5 に，事後学習のシミュレーションにおけるタクシーの分布を示す．実際の運行データとは対照的に，タクシーが周辺部に滞留し，中心部へ帰ってこない傾向が確認された．

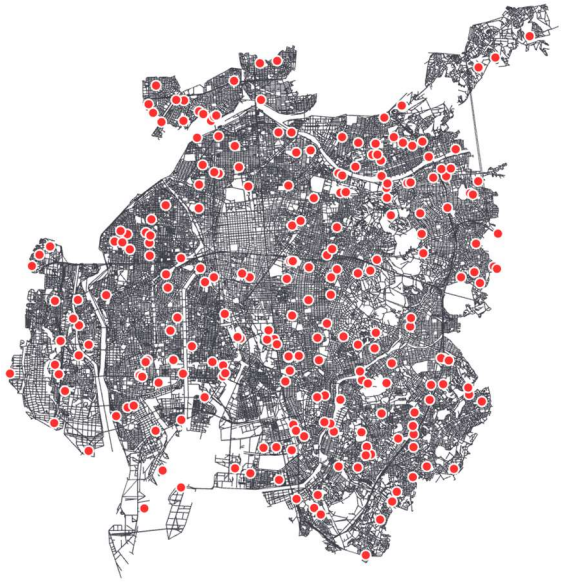

図 5 事後学習のシミュレーション中のドライバー分布

## 6. 考察

本章では，5 章の分析で確認されたドライバーの周辺部滞留（図 5）と Q 値分布の変化（図 2〜4）がなぜ生じるのか，また Hit Rate の低下とどのように関係しているのかを考察する．

### 6.1. 周辺部にドライバーが滞留する原因

5.2 節で，事後学習のシミュレーションにおいてドライバーが周辺部に滞留する傾向を確認した（図 7）．この原因は，エージェントの経路選択の仕組みにある．エージェントは各交差点において，1 ステップ先の道路の Q 値のみに基づいて経路を選択する．中心部で形成された高い Q 値は，隣接する道路へと連鎖的に伝播していくが，名古屋市の道路ネットワークでは中心部から周辺部の距離は長く、多数の交差点を経由する必要があり，その過程で割引率 $\gamma$ の累積により Q 値は減衰する．そのため，周辺部のドライバーから見ると，中心部方向を示す Q 値の勾配が存在しない．加えて，周辺部でも局所的に乗客を獲得する機会があるため，周辺部にも局所的な Q 値が形成される．ドライバーはこの局所的な Q 値に吸い寄せられ，高需要な中心部に帰ることができず，周辺部に滞留する．

### 6.2. Hit Rate が低下する原因

このドライバーの周辺部滞留は，Q 値分布の変化と Hit Rate の低下の双方を引き起こす．中心部のドライバーが減少すると，中心部で乗客を獲得する機会が減り，中心部の道路の Q 値が更新されなくなる．更新されない Q 値は減衰していくため，学習が進むにつれて中心部の Q 値は薄くなっていく．これが，事後学習後の Q 値分布（図 3，図 4）が事前学習後（図 2）と比べ

て中心部のQ値が薄くなっていた原因である.そして,中心部のQ値が薄くなることで,ドライバーが高需要な中心部へ向かう経路選択ができなくなり,中心部で発生する乗車に対応できず,Hit Rateが低下する.

### 6.3. 提案手法のHit Rateがより早く低下する原因

以上で述べたHit Rate低下の原因は提案手法とベースラインに共通するが,図1の学習曲線では提案手法の方がベースラインよりも早くHit Rateが低下に転じている.これは,事前学習によるQ値の存在が周辺部滞留の進行速度に影響を与えているためと考えられる.提案手法は事前学習により学習初期から周辺部にもQ値が形成されているため,エージェントは早い段階から周辺部のQ値に従って経路を選択し,周辺部への滞留が速やかに進行する.一方,ベースラインは学習初期にQ値が未学習であるため,経路選択はランダムに近く,偶然に中心部方向へ移動する機会がある分,周辺部への滞留は相対的に遅く進行する.この低下速度の差は,事前学習で形成されたQ値がエージェントの経路選択に実際に影響を与えていたことを意味する.

## 7. おわりに

本研究では,タクシーの流し営業における道路レベルでの経路推薦を目的として,行動クローニングによる事前学習と深層強化学習を組み合わせた手法を提案し,名古屋市の道路ネットワークと実際のタクシー運行データを用いて有効性を検証した.

実験の結果,提案手法はベースラインよりも速くHit Rateのピークに到達し,事前学習による学習効率の改善が確認された.しかし,Hit Rateのピークは約18%と低い水準に留まり,その後約13%まで低下して収束した.

分析と考察より,この性能低下の原因は,エージェントが1ステップ先の道路のQ値のみに基づいて経路を選択する設計にあることがわかった.名古屋市の道路ネットワークでは中心部から周辺部の距離は長く,多数の交差点を経由する必要があるため,中心部の高いQ値が周辺部まで伝播せず,ドライバーが周辺部に滞留して高需要な中心部に帰ることができない.その結果,中心部の学習が進まず,Hit Rateが低下する.また,提案手法では事前学習で形成された周辺部のQ値がかえって周辺部への滞留を加速させるため,ベースラインよりも早くHit Rateが低下に転じることが確認された.このHit Rate低下速度の差は,事前学習のQ値がエージェントの経路選択に実際に影響を与えていたことを示している.

今後の課題として,1ステップ先のみではなく,長距離の移動を考慮した経路選択手法の検討が必要である.例えば,周辺部で一定時間乗客を獲得できないドライバーを高需要な中心部へ帰還させる仕組みや,道路レベルの経路選択とエリアレベルの移動先決定を組み合わせた階層的な経路選択が考えられる.

## 参 考 文 献

[1] 山越伸浩,“タクシー輸送の担い手の確保とその在り方”, 立法と調査, No.467, pp.18-36, 2024.

[2] 川崎仁嗣,石黒慎,深澤佑介:AIタクシー:交通運行の最適化をめざしたタクシーの乗車需要予測技術, NTT DOCOMOテクニカル・ジャーナル, Vol.26, No.2, pp.15-21 (2018).

[3] DiDiモビリティジャパン:DiDiヒートマップ機能を提供開始,プレスリリース (2019). URL https://didimobility.co.jp/info/20191106731/

[4] みんなのタクシー:事業説明会 移動・交通の最適化に向けて協業を加速,プレスリリース (2019). URL https://www.sride.jp/jp/list/20191105/

[5] D. Lee, H. Wang, R. L. Cheu, and S. H. Teo, "Taxi dispatch system based on current demands and real-time traffic conditions," Transportation Research Record, vol. 1882, pp. 193-200, 2004.

[6] A. Alshamsi, S. Abdallah, and I. Rahwan, "Multiagent self-organization for a taxi dispatch system," in Proc. 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS), pp. 21-28, 2009.

[7] K. T. Seow, N. H. Dang, and D. H. Lee, "A collaborative multiagent taxi-dispatch system," IEEE Transactions on Automation Science and Engineering, vol. 7, no. 3, pp. 607-616, 2010.

[8] L. Zhang, T. Hu, Y. Min, G. Wu, J. Zhang, P. Feng, P. Gong, and J. Ye, "A taxi order dispatch model based on combinatorial optimization," in Proc. 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 2151-2159, 2017.

[9] N. J. Yuan, Y. Zheng, L. Zhang, and X. Xie, "T-Finder: A recommender system for finding passengers and vacant taxis," IEEE Transactions on Knowledge and Data Engineering, vol. 25, no. 10, pp. 2390-2403, 2013.

[10] DeNA, "AIでタクシーの収益性向上と人手不足解消へ 次世代タクシー配車アプリ「MOV」," プレスリリース, 2019年12月10日.URL: https://dena.com/jp/news/4550

[11] K. Lin, R. Zhao, Z. Xu, and J. Zhou, "Efficient large-scale fleet management via multi-agent deep reinforcement learning," in Proc. 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1774-1783, 2018.

[12] Z. Xu, Z. Li, Q. Guan, D. Zhang, Q. Li, J. Nan, C. Liu, W. Bian, and J. Ye, "Large-scale order dispatch in on-demand ride-hailing platforms: A learning and planning approach," in Proc. 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 905-913, 2018.

[13] M. Li, Z. Qin, Y. Jiao, Y. Yang, J. Wang, C. Wang, G. Wu, and J. Ye, "Efficient ridesharing order dispatching with mean field multi-agent reinforcement learning," in Proc. The World Wide Web Conference (WWW), pp. 983-994, 2019.

[14] J. Kim and K. Kim, "Optimizing large-scale fleet management on a road network using multi-agent deep reinforcement learning with graph neural network," in Proc. 2021 IEEE Intelligent Transportation Systems Conference (ITSC), pp. 990-995, 2021.

[15] H. Jayasinghe, T. Jayatilaka, R. Gunawardena, and U. Thayasivam, "Data-driven simulation of ride-hailing services using imitation and reinforcement learning," in Proc. 34th International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems (IEA/AIE), Lecture Notes in Computer Science, vol. 12798, pp.41-52, 2021.