マルチエージェント逆強化学習による公共空間上の滞在行動分析

松村健太[1]・尾原健太郎[2]・柳沼秀樹[3]・大山雄己[4]

[1]非会員 芝浦工業大学 理工学研究科 社会基盤専攻（〒 135-8548 東京都江東区豊洲３丁目７－５）
E-mail: mh22019@shibaura-it.ac.jp
[2]非会員 東京理科大学大学院 創域理工学研究科 社会基盤工学専攻（〒 278-8510 千葉県野田市山崎２６４１）
E-mail: 7623507@ed.tus.ac.jp
[3]正会員 東京理科大学 創域理工学部 社会基盤工学科（〒 278-8510 千葉県野田市山崎２６４１）
E-mail: yaginuma@rs.tus.ac.jp
[4]正会員 芝浦工業大学 工学部 土木工学科（〒 135-8548 東京都江東区豊洲３丁目７－５）
E-mail: oyama@shibaura-it.ac.jp

公共空間においては空間の質のみならず，他者の存在が生み出す風景が人々を惹きつけ滞在行動を促す．それゆえ，人が認知する空間の魅力は他の滞在者の行動に特徴づけられ，時空間的に変化すると考えられる．こうした公共空間における滞在者の行動・相互作用メカニズムを定量的に評価することで，より持続可能かつ効果的な空間計画を立案できる可能性がある．従前のミクロな空間行動を扱うマルチエージェントシミュレーションは，群衆流再現を目的に規範的な相互作用メカニズムを記述してきた．しかしこれらのアプローチでは正・負の多様な相互作用が存在しうる公共空間上の行動を評価することは難しい．そこで本研究では，実際の公共空間で撮影したビデオデータから画像解析により移動軌跡データを抽出し，逆強化学習によるデータオリエンテッドな報酬推定を行う．高次元な環境に対して頑健な手法である，敵対的学習に基づくマルチエージェント逆強化学習を用いることで，公共空間における滞在者同士の多様な相互作用を学習することを試みる．

Key Words: 敵対的逆強化学習, マルチエージェントシミュレーション, 公共空間分析, 相互作用

1. はじめに

公共空間は人間中心の街づくりや多様化するライフワークといった社会の潮流の中で重要性が高まっている．公共空間の設計においては，各々が好きな使い方で自由に快適に過ごし，自然と滞在したくなる空間作りが求められる．このような人々のアクティビティに着目した公共空間設計は 20 世紀以降，欧米を中心に発展してきた．公共空間を構成する要素の中でも，「人」は重要視されているものの一つである．Jan Gehl[1] は人のアクティビティは空間の魅力の一つであるとしており，William H. Whyte[2] も適度な歩行者数は街に活気を与えると述べている．居心地が良く快適に過ごすことができる公共空間は多くの滞在者を誘引することで，人々の賑わいや温かみといった印象を生み出し，さらなる滞在者の増加につながるだろう．このように公共空間においては，空間自体の質だけでなく滞在者が生み出す風景もまた公共空間の魅力に貢献すると考えらえる．滞在者が公共空間に及ぼす影響を定量的に評価することは，人々が滞在したくなる空間設計に向けての判断材料となるだろう．

公共空間の魅力を明らかにした研究は多く存在し，それらでは主にフィールドワークによる歩行者通行量調査や利用者数調査が一般的におこなわれてきた[3)4)5)]．しかし，歩行者交通量と空間の居心地の良さは必ずしも結びつかないと考えられる．また，多くの滞在者を生むことは混雑や喧騒といった雰囲気を醸成し，公共空間としての魅力が低減する可能性もある．また，空間行動分析の文脈では，所与の行動規範に従ったシミュレーションやモデルベース学習による手法が用いられてきた．このような従来の観測に基づく手法や限られた現象記述に基づく手法では，公共空間における滞在者が生み出す複雑な相互作用メカニズムを正確に評価することは難しい．そのため，行動データに内在する潜在的な空間/行動選好を明らかにすることが求められる．

そこで本研究では，滞在者が公共空間に与える正負の相互作用を捉え，滞在者の行動パターンを明らかにすることで，公共空間設計の一助となることを目的とする．本研究では，長期的な調査から取得した実データを用いて，行動パターンから空間の報酬を推定する逆強化学習を基本とする手法によって従前の課題を解決する．具体的には，敵対的生成ネットワークと逆強化学習を組み合わせたモデルをマルチエージェントに拡張した手法[17)] を用いて，公共空間上の時空間的に変化する正負の相互作用を表現することを目的とする．

二章では，既往研究の整理で本研究の位置付けを明確にした後，三章では，本研究で用いるデータの概要について説明を行う．四章ではモデルの概略について説明し，五章にて分析結果の詳細を論じる．

2. 既往研究と本研究の位置付け

(1) 公共空間の分析

これまで，公共空間の分析においては，フィールドワークによる行動観察を行ったり，公共空間の印象や行動をアンケート調査によって分析する手法が主に用いられてきた[3)4)5)]．行動観察による分析[5)4)] では，人々の行動を目視で記録し，それらの行動特性と周囲環境との関連性を推測することで，公共空間の利用傾向や空間選好を明らかにしてきた．しかし，このようなアプローチは調査期間が限られていることによる信頼性の観点や，調査者の経験に基づく主観的な理解であることから，客観的な評価とはなり得ていない可能性が考えられる．アンケート調査から公共空間の良さを探った研究[3)] では，回答者に提示した公共空間の写真に対する印象から，人々に好まれる公共空間の特徴を明らかにした．しかし，アンケートを用いた手法では，実際の行動とは必ずしも結びつかない点や，利用状況が時空間的に変化する公共空間に対する空間選好を捉えきれない．そのため，公共空間の分析においては，長期観測に基づいて取得されたデータに基づく分析から，公共空間の利用状況を明らかにすることが求められる．

(2) 空間行動分析

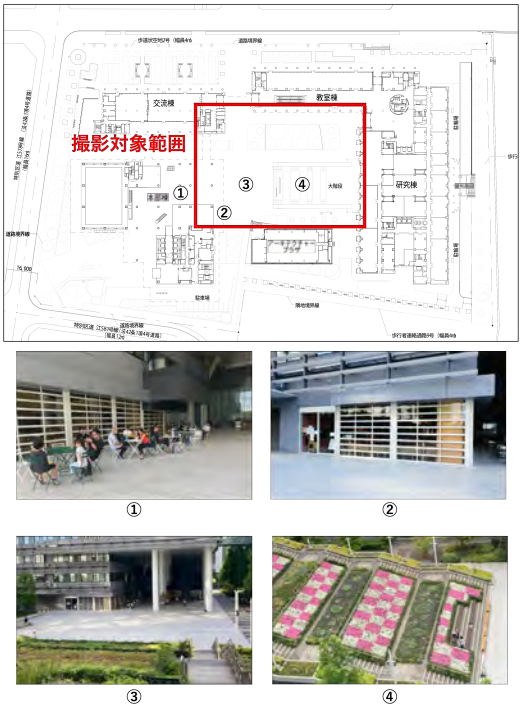
行動分析については，Social Force Model やセルオートマトンといった手法が存在する[6)7)]．しかし，これらは群衆行動表現に重きが置かれていること，意思決定規範が所与であることから，空間評価や複雑多様な行動を表現することは不可能である．それに対し，これらの課題を解決した逆強化学習においては，街中の街路選択に AIRL を適用した研究が存在するものの[9)]，ミクロな空間では，特徴量のスケールや他者との関係性も大きく異なることから，街路スケールで用いられた手法をそのまま適用することは困難であると考えられる．歩行者を対象としたミクロな空間における相互作用の分析では，歩行者同士や歩行者自転車間を扱ったものが多く見られる．しかし，これらは群衆や自転車に対する回避行動という負の作用のみを捉えていることから，人々が生み出す正の側面は捉えられていない[6)10)11)]．

(3) 本研究の位置付け

そこで本研究では，逆強化学習と敵対的生成ネットワークの理論的整合性に着目し，Maximum Entropy IRL では扱えなかった高次元問題への適用が可能となった Adversarial Inverse Reinforcement Learning(AIRL) を下敷きにした手法を用いる．AIRL は実データから観測される行動規範学習することができるかつ，空間評価が可能であることから従前の課題を解決することができる．本研究では，大学内の広場を定点カメラにより長期的に観測したデータを元にした分析を通じて，公共空間における正負の相互作用を明らかにする．

3. 使用データ概要

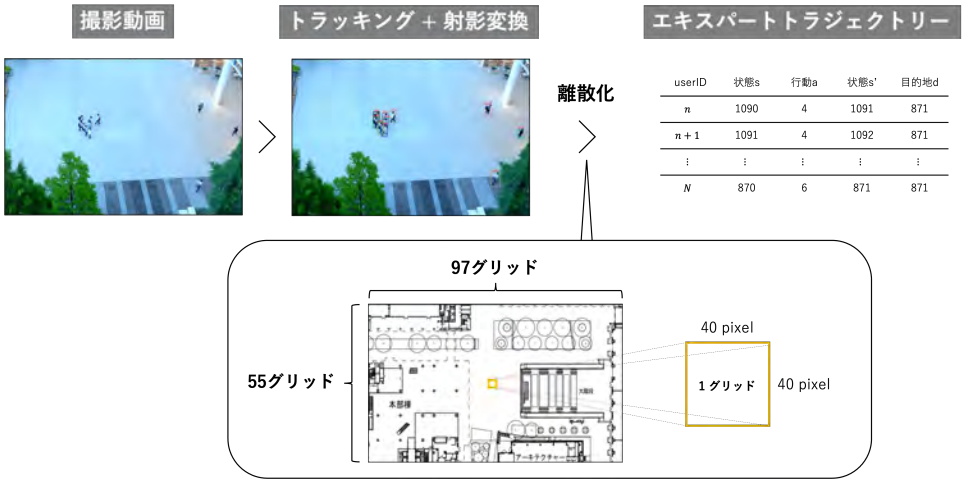
本研究では，現実の公共空間からビデオカメラによって長期観測したデータを使用する．この動画データからトラジェクトリーデータとして人々の動きを抽出し，その後の分析に使用する．



図–1 調査対象範囲図

(1) 調査方法

本研究の対象地は芝浦工業大学 豊洲キャンパスの中央広場である．広場は周囲 3 方がビルになっており，一方は低層施設が立っている (図 1)．広場内にはベンチや花壇が存在しており，レストランとカフェがある．その広場を建物の 7 階から定点カメラを用い，4K(2160 × 3840) 30fps で撮影する．撮影期間は 7 月 8 日〜7 月 26 日のうち気候の良い 6 日分を選定し，日照時間を考慮しながら主に午前 10 時〜午後 6 時に撮影を行った．撮影した動画から軌跡データをトラッキングアルゴリズムを使用し，モデルのインプットとするデータセットを作成する．



図–2 軌跡データの生成フロー

(2) トラジェクトリーデータの作成

本研究では，尾原ら[12] のトラッキングアルゴリズムを使用して，軌跡データを作成する．本アルゴリズムは Multi-object tracking(MOT) が可能な手法であり，学習済みモデルである YOLOX による物体検知と StrongSORT によるフレーム間の動きを追跡するアルゴリズムがベースである．このアルゴリズムに撮影した動画を読み込ませ，トラッキング処理と射影変換を施す (図 2).

この際，トラジェクトリーデータは，連続量 (ピクセル) で生成されるため，本研究で用いるモデルを安定させるために軌跡データの離散化を行う．今回は 40×40 ピクセルを 1 グリッド (約 2 歩分) とした計 5335 グリッド (55×97) の状態空間を使用する．

4. 定式化

本研究では AIRL[13] を用いた学習を行う．AIRL は敵対的生成ネットワーク[14] のフレームワークと同様，実データを模倣する生成ネットワークと実データ/生成データを見分ける識別ネットワークから構成される (図 3).

(1) マルコフ決定過程

公共空間における行動をマルコフ決定過程として表現する．マルコフ決定過程 M は $M = \{S, A, T, R, \gamma\}$ で表現される．S は状態空間集合，A は行動空間，$T(s, a, s')$ は状態 $s \in S$ における $s' \in S$ への遷移確率，R は報酬集合，γ は時間割引率である．以下に状態，行動の定義を示す．

- 状態 s
  公共空間上の空間的位置を示す．40×40 ピクセルを一つの状態として定義する．
- 行動 a
  グリッド間の移動について定義される．行動は「8 方向の移動 (右前・前・左前・左・右・左後・後・右後) + 滞在」の計 9 つの選択肢から構成される．各行動により進むグリッドは一つである．

(2) 方策推定：生成ネットワーク G

方策推定 G ではある OD ペアに対して，現実的なトラジェクトリーを生成できるように学習する．方策推定では，入力として公共空間を構成する特徴量マップをインプットとする (図 4)．G のアウトプットは状態 s を所与として 9 つの行動パターンである．始点グリッドをインプットとして，目的地までのトラジェクトリーを再帰的に出力する．本研究では，状態 s においてその次の状態 s' を考慮した行動表現をするために，畳み込みニューラルネットワーク (CNN)(図 4) を用いる．

方策推定の学習では，ベースライン付き REINFORCE[8] を用いる．REINFORCE は方策勾配法の一種であり，累積報酬から価値関数を引くことで，計算の安定化を目的とする．

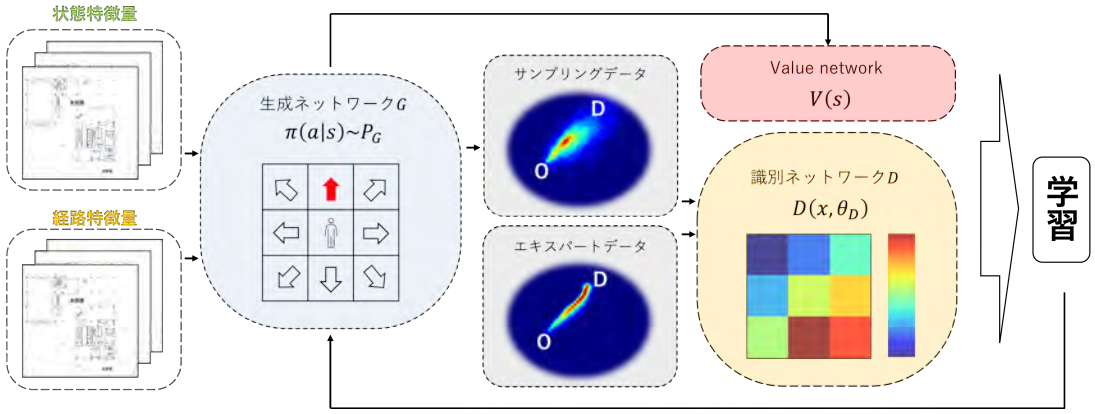
$$\nabla J(\theta) = \frac{1}{N} \sum_{i=1}^{N} \sum_{t=0}^{T} (G_t(\tau^{(i)}) - V(S_t)) \nabla_\theta \log \pi_\theta(A_t^{(i)} \mid S_t^{(i)}) \tag{1}$$

(3) 報酬推定：識別ネットワーク D

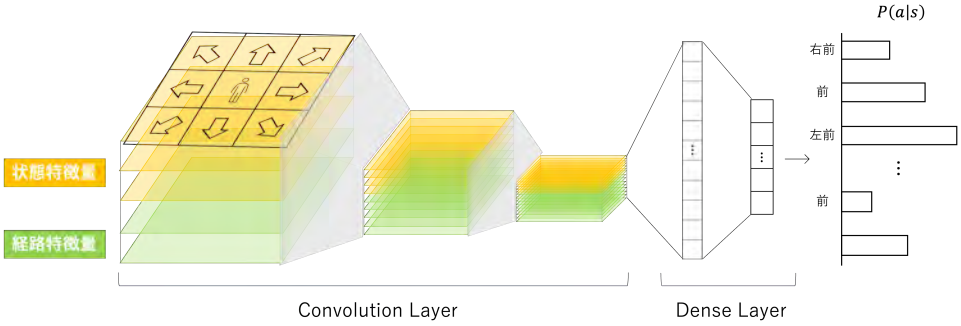
識別器 D では方策推定により生成されたトラジェクトリーデータと実データを見分けることができるように学習する．識別器の確率分布は生成側の確率分布を明示的に扱うことができれば，識別器の確率分布についても同様に明示的に記述することができ，以下のようになる．

$$D_{\theta,\phi}(s, a) = \frac{\exp\left(f_{\theta,\phi}(s, a)\right)}{\exp\left(f_{\theta,\phi}(s, a)\right) + \pi_G(a \mid s)} \tag{2}$$

$f_{\theta,\phi}$ は報酬学習に関する関数であり以下の式で表さ



図–3 AIRL のフレームワーク



図–4 生成ネットワーク G の構成

れる．$h_{phi}(s)$ は修正項であり，報酬推定の曖昧さを緩和する．

$$f_{\theta,\phi}(s,a) = g_\theta(s,a) + \gamma h_{phi}(s') - h_\phi(s) \quad (3)$$

$f_{\theta,\phi}$ の計算では Zhao et.al.[9] と同様に，$g_\theta(s,a)$ については生成ネットワーク同様 CNN を用い，$h_\phi(s)$ については全結合のニューラルネットワークによって近似計算を行う．

識別器の目的関数は以下のような通常の二値分類の式で表される．

$$\min_{\theta,\phi} -E_D[\log(D_{\theta,\phi}(s,a)] - E_{x_G}[\log(1-D_{\theta,\phi}(s,a)] \quad (4)$$

報酬 $R_{\theta,\phi}(s,a)$ は以下の式で定義される．

$$R_{\theta,\phi}(s,a) = \log(D_{\theta,\phi}(s,a) - \log(1 - D_{\theta,\phi}(s,a) \quad (5)$$

$$= f_{\theta,\phi}(s,a) - \log \pi_G(a \mid s) \quad (6)$$

(4) 価値関数推定:Value network

価値関数の計算は，モデルベースの手法では value iteration や一次方程式として計算されていたが，今回のような高次元な問題では適用が難しい[15)16)]．状態 s における価値関数を計算するために 2 層の全結合ネットワークを用いて近似計算を行う．

5. モデル推定結果

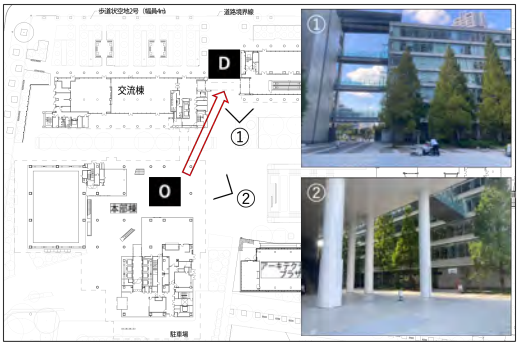
(1) 対象とするトラジェクトリーデータ

今回はモデルの基本的な精度を確認するために，シングルエージェントによる最短経路の学習を行う．対象とするトラジェクトリーデータは，気候条件の良い 7 月 8 日〜7 月 26 日のうちの 6 日分 (7/8，7/11，7/20，7/21，7/23，7/26) に撮影した計 6 時間半の動画を使用する．この動画データをトラッキングアルゴリズムによりトラジェクトリーデータに変換した．その後，userID が変化しているデータの削除し，計 2255 データを取得した．今回はその中でも，本部棟から教室棟へ移動する 876 データを対象として分析を行った．この経路は豊洲駅から大学にくる学生や，教室等を抜けた先にある住宅街へ向かう市民によく利用が見られる．

(2) 使用する空間特徴量

a) 状態特徴量

状態特徴量には各状態 s から，ベンチまでの距離，テーブルまでの距離，柱までの距離，建物 (壁) までの距離，の 4 つの特徴量を用いる．状態特徴量の空間的な位置をグリッド上に吸い付けるのではなく，各特徴量までの距離を用いることで，歩行者が近いものに寄っ



図–5 分析に使用するトラジェクトリーデータ

ていく行動といった現実的な視覚効果を表現している．

b) 経路特徴量

経路特徴量には，各状態 s から目的地 g までの最短経路，目的地の方向を用いる．最短経路の計算では目的地の方向性を一意に定めることを目的に，マルコフ決定過程における斜め方向への移動を除いた距離（=マンハッタン距離）で計算を行った．

(3) 評価指標

a) 編集距離

編集距離は，連続する異なる二つの文字列を比較する指標である．一方の文字列に対し，挿入・削除・置換の操作をおこなってもう一方の文字列に変換できる最小回数として定義される．算出方法は以下の通りである．

$$ED = \frac{1}{N}\sum_{i=1}^{N} \min\left(\frac{Edit(\widehat{x}_i), x_{i,ref}}{T_{i,ref}}, 1\right) \tag{7}$$

ここで，N は検証用データに含まれるトラジェクトリー数，$\widehat{x}_i$ は i 番目の予測トラジェクトリー，$x_{i,ref}$ はエキスパートトラジェクトリーであり，$T_{i,ref}$ はその長さである．

b) Dynamic Time Warping

Dynamic Time Warping(DTW) とは，時系列データの類似度を測るための指標である．二つの時系列データを波形と見做し，各点間の距離を前組み合わせについて計算し，各点同士が最も近くなるようなパスを計算する．この手法の利点は，異なる長さや周期のデータについても適用可能な点である．二つの時系列データ S, T を $S = s_1, s_2, ..., s_M$，$T = t_1, t_2, ..., t_N$ とおくと，ある組み合わせ (i, j) の距離を，

$$\sigma(i, j) = |s_t - t_j| \tag{8}$$

とし，これを各点全てについて求め，その最小値の合計を DTW と定義する．

c) BLEU

自然言語処理の翻訳タスクで用いられる指標であり，生成文と参照文の類似度を求める指標である．連続する n-gram に着目し，それらがどの程度一致するかを計算する．C_j は，予測トラジェクトリーに含まれる連続する j グリッド集合を表し，ω_m はグリッド m の出現回数を表す．

$$BLEU_n = \frac{1}{N}\sum_{i=1}^{N} \min\left(1, \frac{T_i}{T_{i,ref}}\right)\left(\prod_{j=1}^{n} P_j\right)^{\frac{1}{n}} \tag{9}$$

$$P_j = \frac{\sum_{m \in C_j} \min(\omega_m, \omega_{m,max})}{W} \tag{10}$$

(4) 推定結果: 生成ネットワーク G

図 7 に生成ネットワークの推定結果を示す．概ね目的地の方向へ学習が進んでいるものの，目的地に到達できているケースは図 (7)-a よりおよそ 100 トリップほどである．これは今回使用したデータ数が 876 程度と少ないことが挙げられる．そのため，撮影期間を増やし，各 OD データを増加させることが考えらえる．また学習の際，経路特徴量を含まない学習時は，目的地の方向への学習が進まないことから，Zhao et. al.(2023)[9] 同様モデルフリー手法における経路特徴量の重要性が示されたと考える．

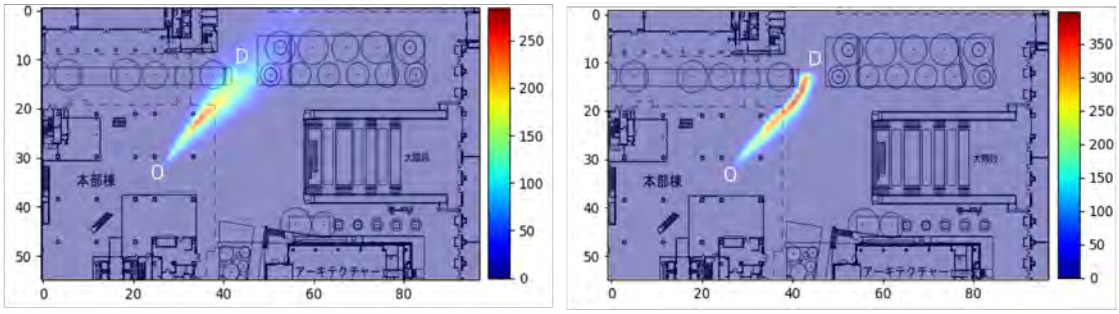
(5) 推定結果: 識別ネットワーク D

図 6 に生成ネットワークの推定結果を示す．可視化にあたっては，OD それぞれ対象エリア内の一グリッドを抽出し，その OD に基づく報酬の計算をおこなった．学習結果より，目的地方向へ近づくような行動の報酬が高くなっており，目的地から遠ざかるような行動の報酬は低くなるように推定された．OD 間の中間において，斜め右前の移動ではなく，上・右方向の報酬が高くなったことは，対面通行者との錯綜が多いことから横に逸れるような行動に高い報酬が割り当てられたと考えられる．

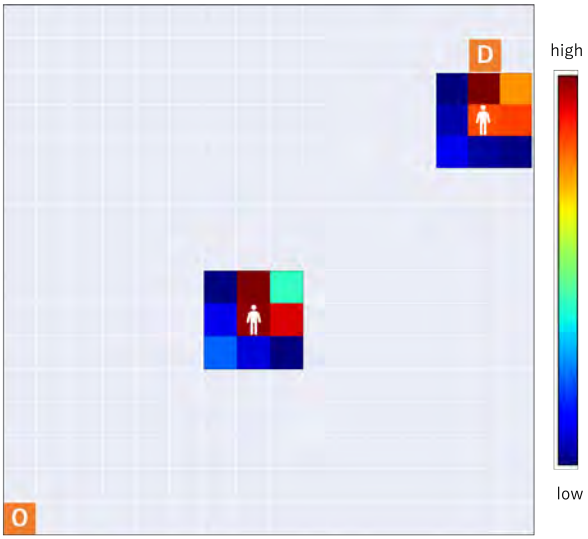
(6) 今後の方針

今回はシングルエージェントの AIRL を用いた最短経路学習を通じて，モデルの基本的な性能を確認することを目的とした．学習結果より，本手法は公共空間における行動特性を適切に捉えられることができることが示されたものの，更なる改善の余地も見られた．今後は，今回使用した AIRL をマルチエージェントに拡張した MA-AIRL[17] を用いた空間相互作用の分析を行う．同時にモデルベースの手法との比較を通じて，モデルフリー手法の特徴を明らかにする．

（a）推定トラジェクトリーデータ　　（b）エキスパートデータ



図–6 生成ネットワーク G の方策比較



図–7 推定報酬結果

参考文献

1) Jan Gehl（2010），Cities for People
2) William H. Whyte（1980），The Social Life of Small Urban Spaces
3) 小林 健治・鈴木 毅・新村 岳広・木多 道宏・舟橋 國男 (2005), 人が居る場面記述にみる人間-環境関係の基礎的分析：都市の中の”いい感じ”に人が居る場面の研究，日本建築学会計画系論文集，Vol.70, No.589，p.77-83
4) 戒田 純也・佐久間 康富 (2021), 公園利用者の滞留行動特性と滞留行動の発生要因に関する研究-大阪市長居公園を事例に-，都市計画報告集，Vol.19, No.4，p.529-533
5) 森一彦・西脇 智子 (2004), 池にある公園におけるベンチの使われ方に関する研究-着座向きと行為内容の関連-，日本建築学会計画系論文集，Vol.69, No.585，p.71-77
6) Bandini，S.,Mondini,M.,Vizzari,G.(2014)Modelling negative interactions among pedestrians in high density situations，Transp.Res.C,Vol.40,pp.251 – 270.
7) Was，J.,Lubas,R.(2014)Towards realistic and effective Agent-based models of crowd dynamics，Neurocomputing,Vol.146,pp.199 – 209.
8) Williams, R, j.（2014), Simple statistical gradient-following algorithms for connectionist reinforcement learning，Machine learning 8.3，pp.229-256.
9) Zhao, Z.Liang,Y.(2023), A deep inverse reinforcement learning approach to route choice modeling with context-dependent rewards，Transp.Res.C,Vol.149，104079
10) Qu,Y.Wu,J.Yin,H.Yang,X.Xiao,Yao.(2021)，Analyzing pedestrian individual and interaction collision avoidance dynamics in traditional scenarios，Transp.Res.C，Vol.133，103445
11) Alsaleh,R.，Sayed,T.(2021)，Markov-game modeling of cyclist-pedestrian interactions in shared spaces: A multi-agent adversarial inverse reinforcement learning approach，Transp.Res.C，Vol.128，103191
12) 尾原 健太郎・柳沼 秀樹・寺部 慎太郎・海野 遥香・鈴木 雄 (2023)，二輪車および歩行者に対応した交通計測 AI の開発および地点別自己学習手法の提案，土木計画学春大会 2023
13) Fu,J.,Luo,K.,Levine,S.(2018)，Learning robust rewards with adversarial inverse reinforcement learning，https://doi.org/10.48550/arXiv.1710.11248
14) Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio(2014)，Generative Adversarial Networks，https://doi.org/10.48550/arXiv.1406.2661
15) Ziebart,B.D.,Maas,A.L.,Bagnell,J.A.,Dey,A.K.,et. al.(2008)，Maximum entropy inverse reinforcement learning，AAAI,Chicago,IL,USA. pp.1433 – 1438.
16) Fosgerau,M.,Frejinger,E.,Karlstrom,A.(2013)A link based network route choice model with unrestricted choice set，Transp.Res.C,Vol.56,pp.70 – 80.
17) Yu,L.,Song,J.Ermon,S.(2019)，Multi-Agent Adversarial Inverse Reinforcement Learning，ICML2019，https://doi.org/10.48550/arXiv.1907.13220

Analysis of Behaviors on Public Spaces Using Multi-Agent Inverse Reinforcement Learning

Kenta Matsumura, Yuki Oyama

In public spaces, the attractiveness of the environment is not only determined by its physical quality but also by the dynamic scenery produced by the presence of others, which let people stay at public space. Therefore, the attractiveness of public space as perceived by individuals is characterized by the actions of fellow occupants and is believed to vary spatiotemporally. By quantitatively evaluating the behavior and interaction mechanisms of occupants in such public spaces, there is potential to propose more sustainable and effective planning.

Conventional micro-level spatial behavior simulations have traditionally described normative interaction mechanisms with the aim of crowd flow replication. However, assessing behaviors on public spaces where a wide range of positive and negative interactions is a challenging task with these approaches. Hence, in this study, we extract trajectory data from video filmed in real public spaces through image analysis and perform data-driven reward estimation using inverse reinforcement learning(IRL).

We will model diverse interactions among occupants in public spaces by employing multi-agent inverse reinforcement learning based on adversarial learning, a robust approach for high-dimensional environments.