敵対的逆強化学習と行動的均衡

2024/9/11

行動モデル夏の学校

行動モデルの概要と発展

東京大学　小川大智

交通ネットワークの設計

需要分析に基づく都市空間計画の実現に向けて

- 政策・設計の現場ニーズ
  - 自転車→駅まち 改変，車→歩行者空間整備
  - 歩行者と自動走行の混合流空間の実現
- 数理モデルの導入
  - 空間用途遷移のシナリオ比較
  - 多元的評価指標を用いた多目的関数最適化

交通分野での相互作用・相関関係推定

選択肢間の**相関関係**：経路の重複，赤バス・青バス問題

→構造化プロビット，n-GEV，正則化項導入

内生・相互作用効果：混雑外部性，ライドシェア

→**構造推定，MPEC，マッチング**

本日の発表の流れ

- はじめに
- 行動的均衡について
- 敵対的逆強化学習による内生的な相互作用の推定
- 空間特徴量との取り込み
- ケーススタディ

行動的均衡

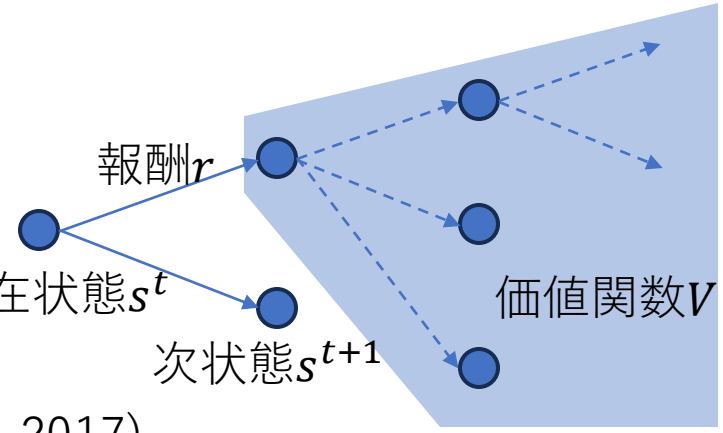
Markov Gameとしての経路選択モデル

マルコフゲームによる基礎づけ

- 現在状態s，行動a，次状態s'
- 価値関数 $V(s) = E\left[\max_a r(a, s'|s) + \gamma V(s') + \varepsilon(s')\right]$

即時報酬　状態価値関数→**将来効用**

時間割引率

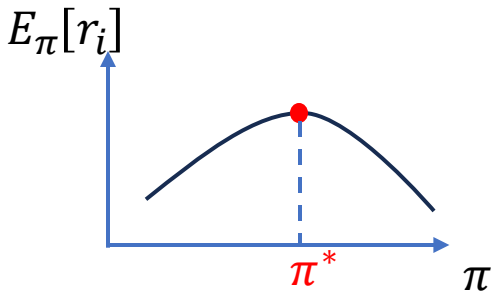


- Recursive Logit モデル(Fosgerau et al., 2013; Oyama & Hato, 2017)
  - 誤差項$\varepsilon(s')$にガンベル分布を仮定
    - Bellman方程式 $V(s) = \mu \ln \sum_a \exp \frac{1}{\mu}\{r(a, s'|s) + \gamma V(s')\}$
  - 再帰的計算により**経路列挙なし**でリンク選択確率を記述

$$\text{選択確率 } \pi(a, s'|s) = \frac{\exp\frac{1}{\mu}\left\{r(a, s'|s)+\gamma V(s')\right\}}{\sum_{a'} \exp\frac{1}{\mu}\left\{r(a', s'|s)+\gamma V(s')\right\}}$$

均衡状態の定義

N人の非協力マルコフゲームの均衡状態

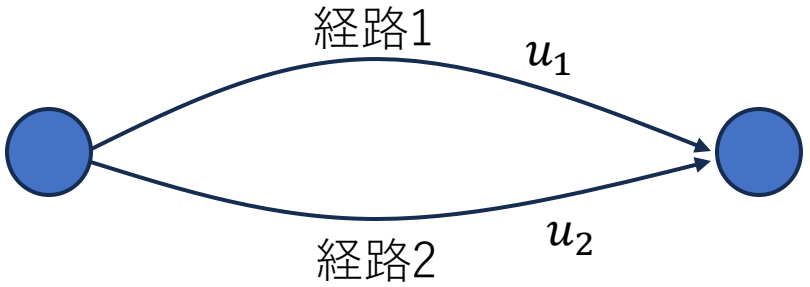


- ナッシュ均衡
  - エージェントiのナッシュ均衡方策$\pi_i^*(a_i, \boldsymbol{a_{-i}}, s_i)$
  - **完全情報の仮定**
  - $\forall \pi_i \ \underbrace{\mathrm{E}_{(a_i, a_{-i})\sim(\pi_i^*, \pi_{-i})}[r_i(a_i, a_{-i}, s_i)]}_{\pi^*\text{の元での報酬期待値}} \geq \underbrace{\mathrm{E}_{(a_i, a_{-i})\sim(\pi_i, \pi_{-i})}[r_i(a_i, a_{-i}, s_i)]}_{\pi\text{の元での報酬期待値}}$
- 相関均衡
  - エージェントiの相関均衡方策$\pi_i^*(a_i, s)$
  - **エージェント共通の状態s**を観測
  - $\forall \pi_i \ \mathrm{E}_{a_i\sim\pi_i^*}[r_i(a_i, s)] \geq \mathrm{E}_{a_i\sim\pi_i}[r_i(a_i, s)]$

マルコフゲームにおける方策間の**相関構造**を定義

単一種類の主体の均衡状態

- 例）2つの経路の選択問題



効用関数

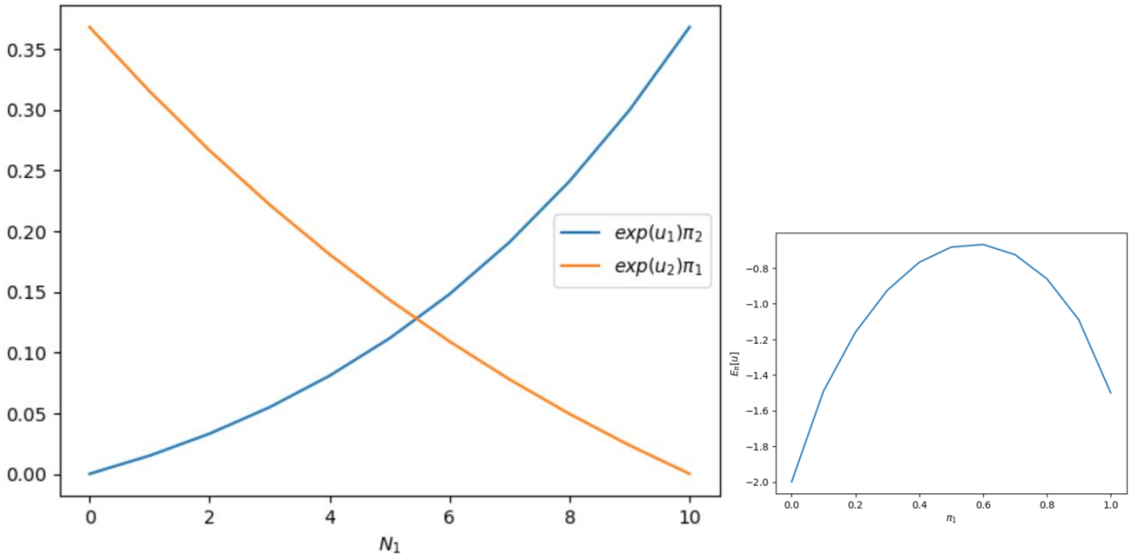
$$u_1 = -(1 + \pi_1)$$
$$u_2 = -(1 + 0.5\pi_2)$$

相互作用

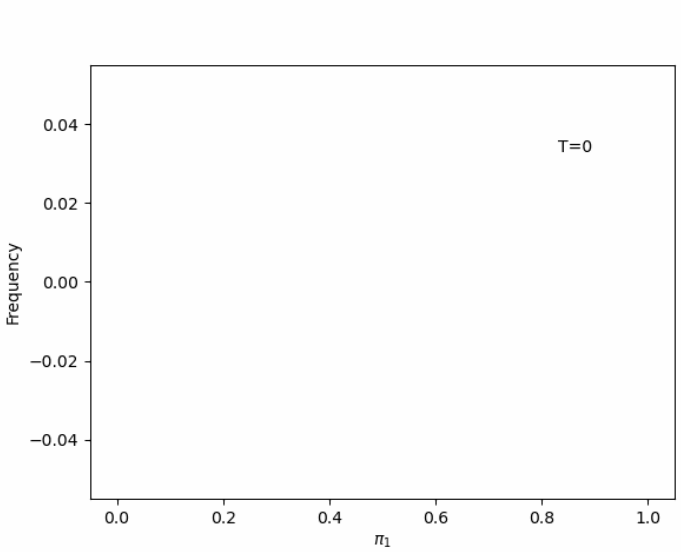
選択確率

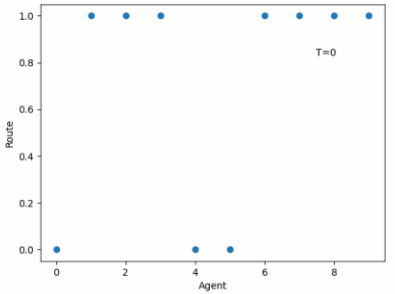
$$\pi_1 = \frac{\exp(u_1)}{\exp(u_1) + \exp(u_2)}$$
$$\pi_2 = \frac{\exp(u_2)}{\exp(u_1) + \exp(u_2)}$$

均衡条件：
効用関数と選択確率の連立方程式の解



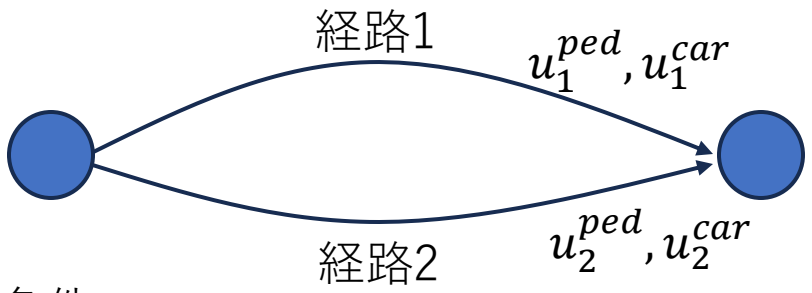
動的なシミュレーション$(N = 10)$





複数種類の主体の均衡状態

- 例）2つの経路の選択問題



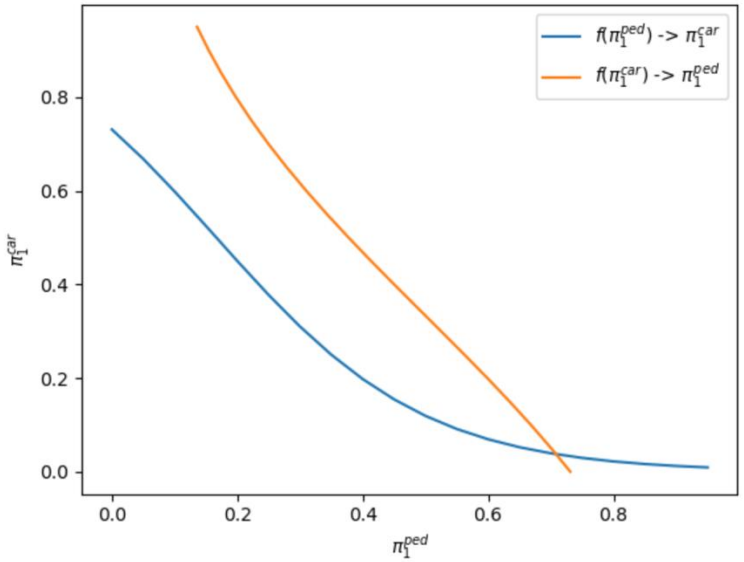
効用関数

$$u_1^{ped} = -(1 + \pi_1^{car})$$
$$u_2^{ped} = -(1 + 0.5\pi_2^{car})$$
$$u_1^{car} = -(1 + \pi_1^{ped})$$
$$u_2^{car} = -(1 + 2.0\pi_2^{ped})$$

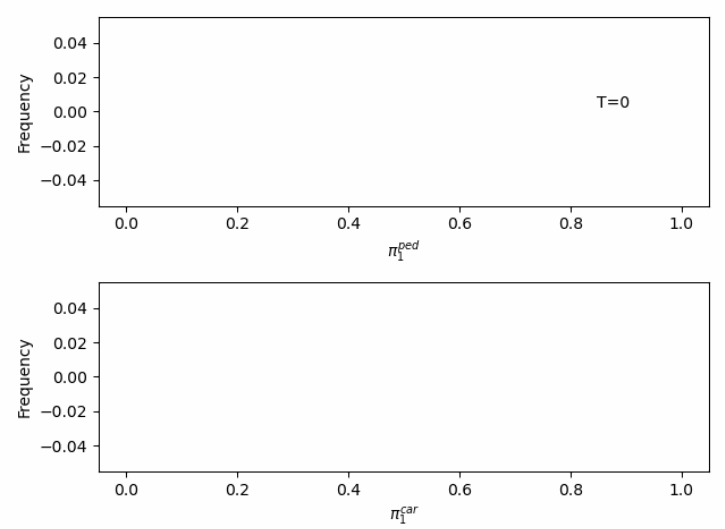
選択確率

$$\pi_1^{ped} = \frac{\exp(u_1^{ped})}{\exp(u_1^{ped}) + \exp(u_2^{ped})}$$

均衡条件：
効用関数と選択確率の連立方程式の解



動的なシミュレーション($N^{ped} = N^{car} = 20$)



- 相互作用の存在により**選択確率が極端**になる場合がある.
- 均衡状態は必ずしも**安定**とは限らない.

AIRLと相互作用の学習

内生効果存在下での推定

内生変数存在下での選択の表現・推定

- 個人iの行動q_i, 説明変数x_i, 誤差項ε_i, パラメータα, β, γ
  - 構造型 $q_i = \alpha + \beta E[q_i] + \gamma x_i + \varepsilon_i$

    内生効果
  - 誘導型 $(1-\beta)E[q_i] = \alpha + \gamma E[x_i]$ →しかし，**識別性** $\frac{\alpha}{1-\beta}, \frac{\gamma}{1-\beta}$ しかわからない
- 最尤推定法では説明変数$E[q_i]$と誤差項ε_iが独立であることが必要

  →**構造推定手法**（経済学分野）
  - 二段階最小二乗法 or 擬似最尤法 (Aguirregabiria & Mira, 2007)

    交通分野での適用
    - 社会的相互作用（福田 et al., 2004）
    - **擬似最尤法**による混雑外部性の内生化（柳沼&福田, 2008）

  →**MPEC**による均衡分析手法（浦田 & 柳沼., 2016）

敵対的逆強化学習（AIRL）

- エントロピー最大化逆強化学習
  - ソフトQ関数 $Q_{soft}^{\pi[t+1:T]} := r(s,a,s') + \gamma \left[\sum_{t'=t+1}^{T} \gamma^{t'-t} r(s,a) + \mathcal{H}\left(\pi^{t'}(\cdot|s)\right) \right]$

    即時報酬　時間割引率　状態価値関数V　**エントロピー項**

    →エネルギーベースのモデルと等価 $\pi(a|s) \propto \exp\{Q^{\pi}(s,a)\}$　　$\mathcal{H}(\pi^{t'}) = -\sum \log \pi^{t'}$
- 敵対的逆強化学習（AIRL）
  - エントロピー最大化逆強化学習の効率的な実装
  - 判別関数
    - 報酬関数推定器 $g_\theta(s,a,s')$
    - 状態価値関数推定器 $h_\phi(s)$
  - 行動関数
    - 方策関数推定器 $\pi(a|s,s')$

学習される判別関数と行動関数

$$\pi(a|s,s') = \frac{\exp\left(g_\theta(s,a,s') + \gamma h_\phi(s')\right)}{\sum_{a'} \exp\left(g_\theta(s,a',s') + \gamma h_\phi(s')\right)}$$

複数交通手段の経路選択モデル

歩車の相関均衡

歩行者の**相関均衡方策** $\pi^*_{ped}(s, a_{ped}, a_{car}) = \frac{\exp\left(r_{ped}(s,a_{ped},a_{car})+\gamma V^d_{ped}(s')\right)}{\sum_{a'_{ped}} \exp\left(r_{ped}\left(s,a'_{ped},a_{car}\right)+\gamma V^d_{ped}(s')\right)}$

- 構造推定による均衡状態推定
  - 相互作用説明変数 $m_{car} = f(r_{car}), m_{ped} = f(r_{ped})$
  - 効用関数
    - 歩行者 $r_{ped} = \theta_{ped} \cdot x_{ped} + \theta^m_{ped} \cdot m_{car}$
    - 車両 $r_{car} = \theta_{car} \cdot x_{car} + \theta^m_{car} \cdot m_{ped}$

π^*の定義に基づき選択確率の計算
→**擬似最尤法**

- AIRLによる均衡状態推定(Yu et al., 2019)
  目的関数
  - $\max_{\pi} E_{\pi}\left[\sum_i \log D_{\theta_i,\phi_i}(s,a) - \log\left(1 - D_{\theta_i,\phi_i}(s,a)\right)\right]$
  - $\min_{\theta_i,\phi_i} E_{\pi}\left[\sum_i \log D_{\theta_i,\phi_i}(s,a)\right] + E_{p_{data}}\left[\log\left(1 - D_{\theta_i,\phi_i}(s,a)\right)\right]$

学習の結果，以下が満たされる．

$$\pi(s, a_{ped}, a_{car}) = \frac{\exp\left(Q_{ped}(s, a_{ped}, a_{car}, s')\right)}{\sum_{a'_{ped}} \exp\left(Q_{ped}(s, a'_{ped}, a_{car}, s')\right)}$$

→**均衡制約条件付き最適化問題（MPEC）**

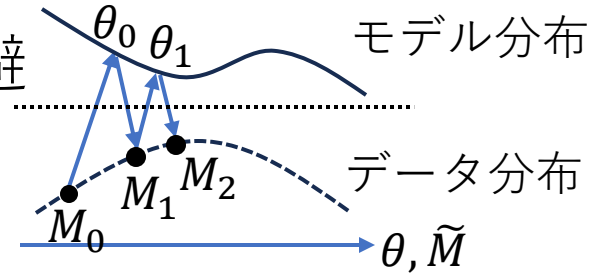
生成系モデルによる内生変数存在下の学習

説明変数と誤差項の相関

- データX，内生変数M，誤差項$\mathbf{E}$
- 最尤推定法の**一致性**の必要条件 $KL[p(X|\theta_0)||p(X|\theta)] \geq 0$
- 内生変数と誤差項の**相関 $\boldsymbol{p(X|\theta) \neq p(X|M,\theta)}$**

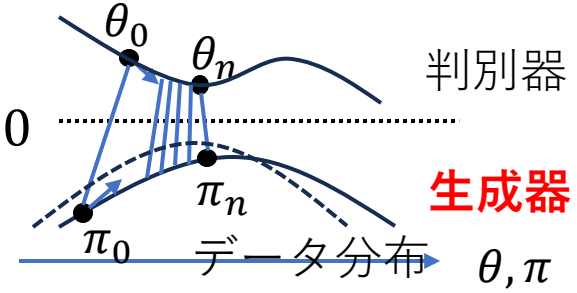
$$KL[p(X|\theta_0)||p(X|M,\theta)] \geq -\log \int p(X|M,\theta)\, dX \neq -\log \int p(X|\theta)\, dX = 0 \ (\because \text{Jensenの不等式})$$

- 擬似尤度関数法(NPL)
  - 内生変数Mに**暫定値$\tilde{M}$**を使うことで，誤差項との相関を回避



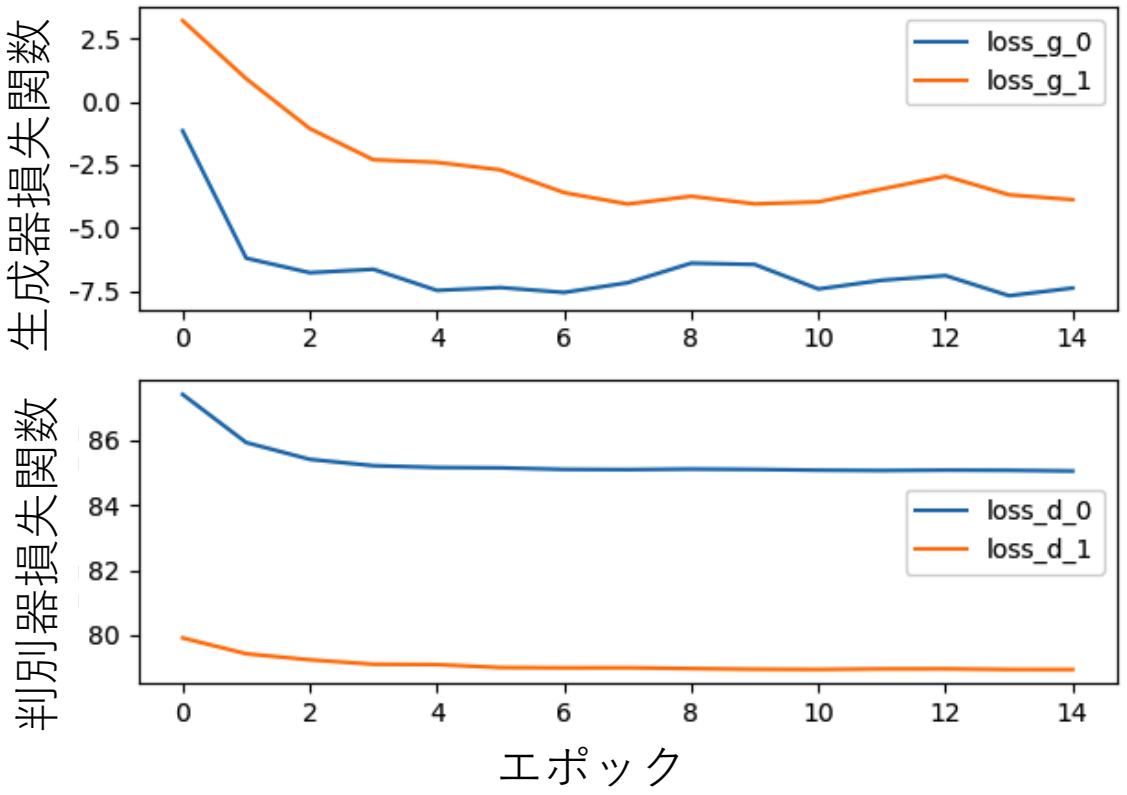
- 敵対的学習
  - 行動関数の学習：$\min_{\pi} KL[p_\pi||p_{data}]$
  - $KL[p_\pi||p_{data}] \geq -\log \int p_\pi(X|M) \frac{p_{data}(X)}{p_\pi(X|M)} dX = -\int \log p_{data}(X)\, dX = 0$

    →**生成データ**を使うことで内生変数の影響を回避

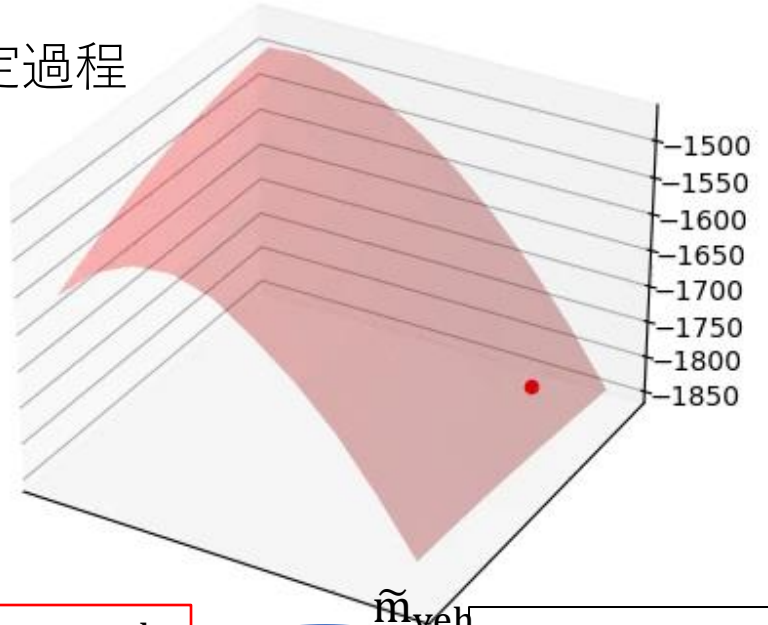


学習過程

AIRLモデル推定過程



RLモデル推定過程

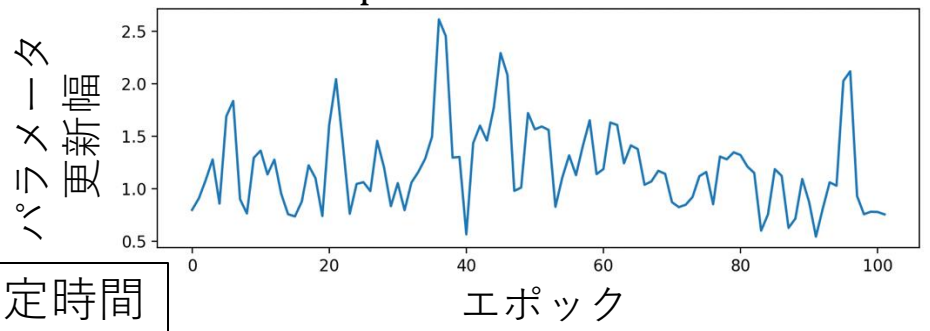


NPL - ($\tilde{\theta}^{ped}$, $\tilde{\theta}^{ped}_{veh}$)

$\tilde{m}_{veh}$

NPL - ($\tilde{\theta}^{veh}$, $\tilde{\theta}^{veh}_{ped}$)

$\tilde{m}_{ped}$

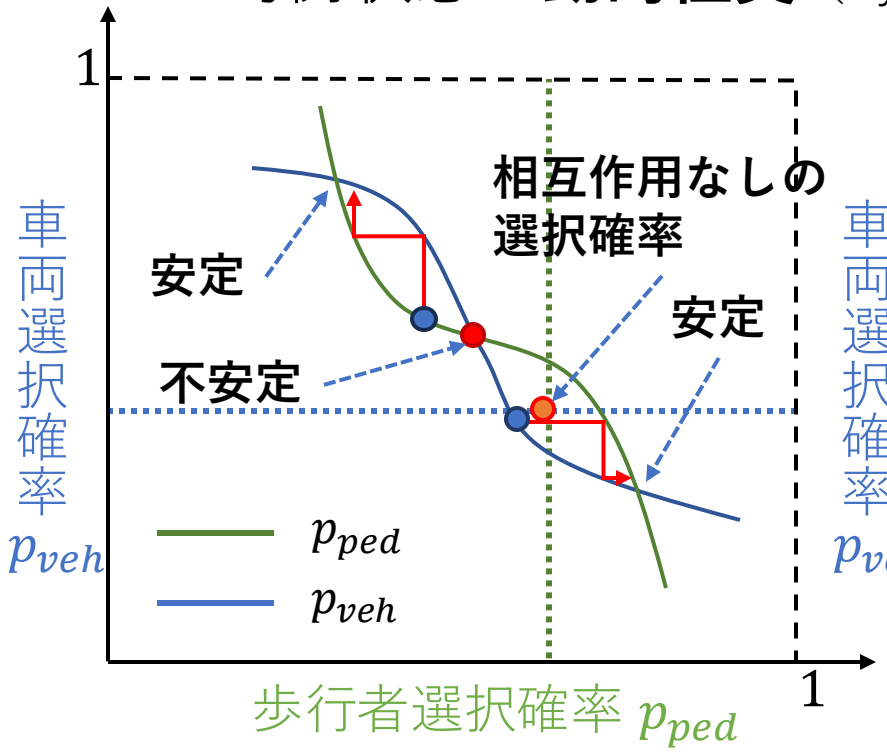


- **RLモデル**では**24時間**ほどの推定時間
- **AIRL**では**30分**ほど.

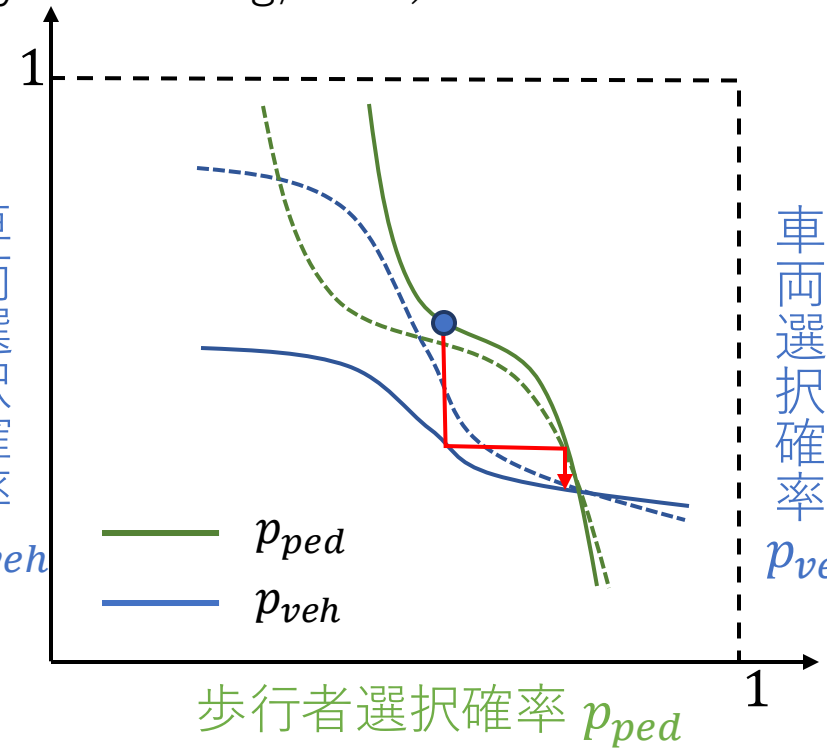
均衡状態の安定性

2エージェント2選択肢のモデルによる定性的分析

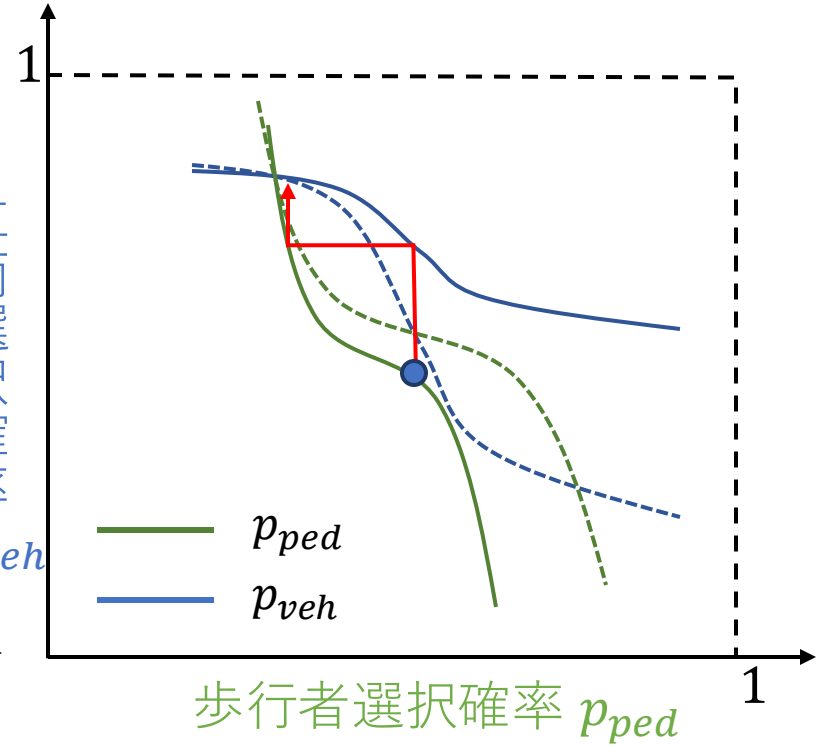
- 負の外部性存在下：選択確率は他者の選択確率に対し単調減少
- 均衡状態の**動的性質**（Iryo & Watling, 2019）



2つの安定均衡と1つの不安定均衡



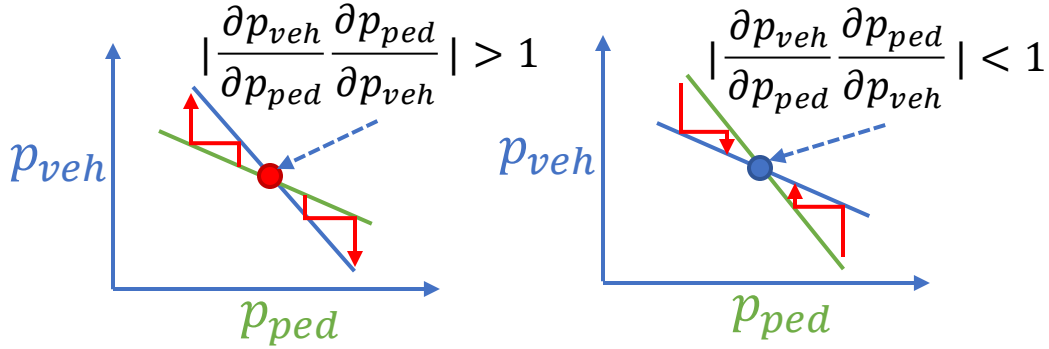
歩行者が卓越する1つの安定均衡



車両が卓越する1つの安定均衡

相互作用の滑らかさによる学習手法

- 相互作用の導入
  - 報酬関数 $g_{\theta_i}(s, a_i, \pi_{-i}) \leftarrow \underbrace{g_{\theta_i}(s, a_i)}_{\text{外生効果}} + \underbrace{l_\psi^i(s, a_i, \pi_{-i})}_{\text{相互作用}=\text{内生効果}}$

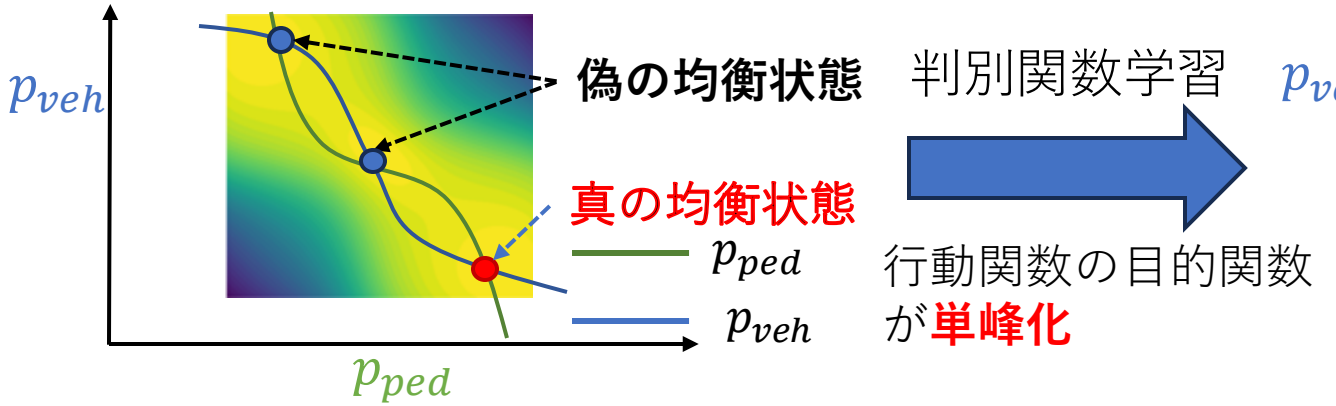


- 推定の安定性と構造化
  - 相互作用の“滑らかさ” $\frac{\partial \pi_{i,a_i}}{\partial \pi_j} = \pi_{i,a_i}\sum_{a_i'}(\delta_{a_i,a_i'} - \pi_{i,a_i})\frac{\partial \tilde{Q}^{\pi_i}_{i,a_i'}}{\partial \pi_j}$ ← $\frac{\partial l_{\psi_i}}{\partial \pi_j}$から計算

    →l_{ψ_i}**の偏微分**に対して制限をかけることで，**安定な均衡解**として判別関数を推定

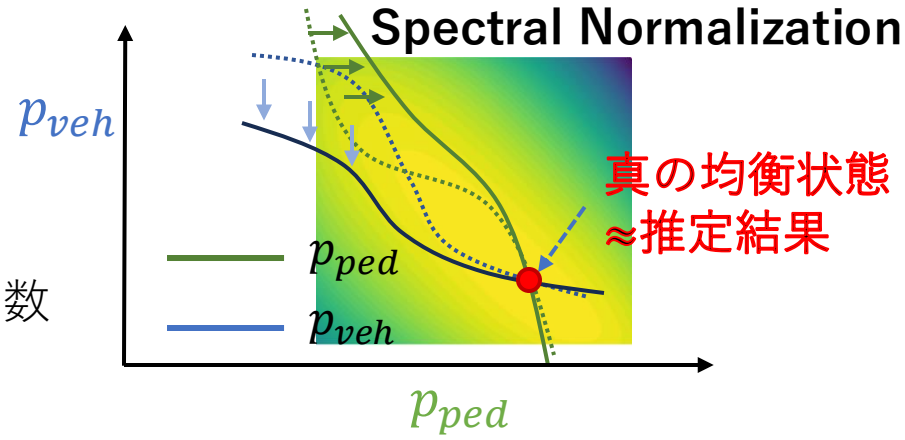
    →Neural Networkでは**Spectral Normalization**により制限可能（Miyato et al., 2018）

さらに，**行動関数の推定の安定性**への寄与



判別関数学習

行動関数の目的関数が**単峰化**



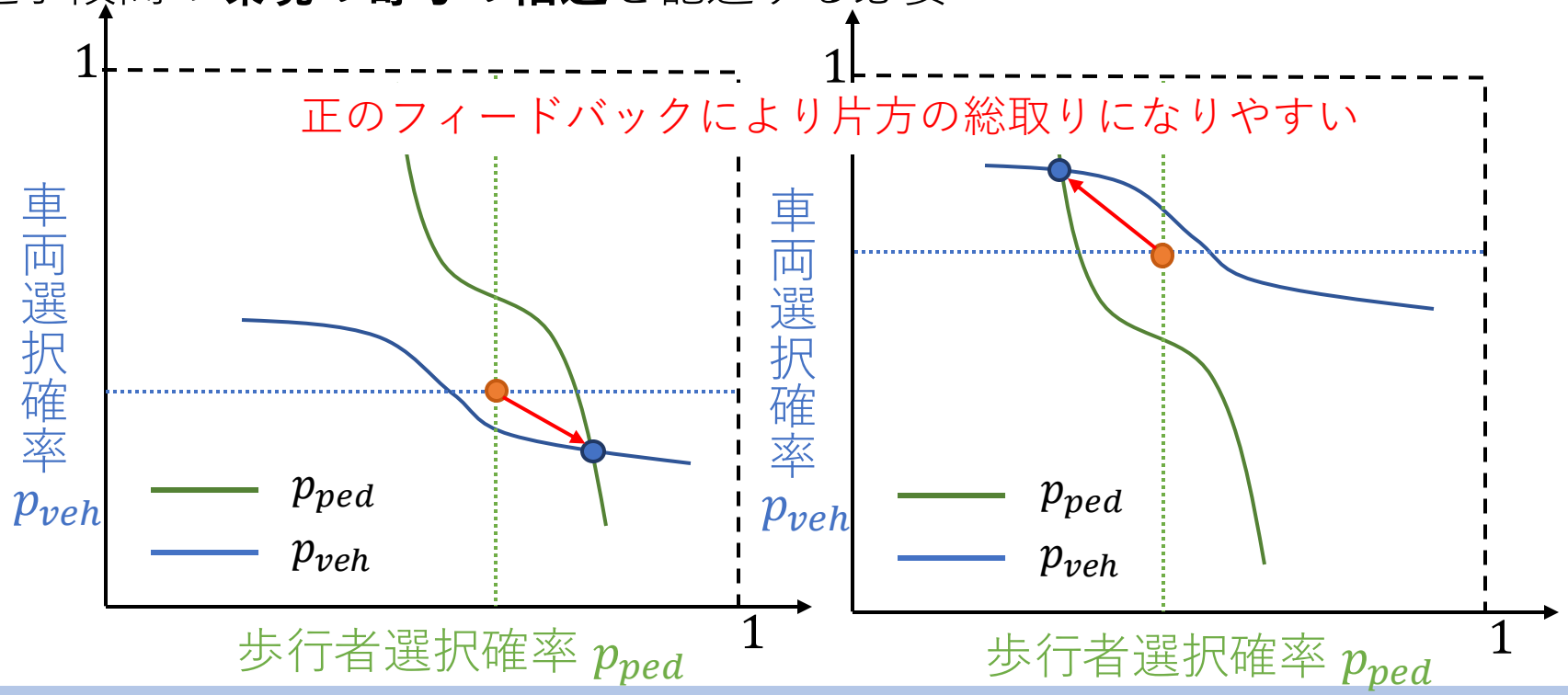
空間的特徴量の同時学習

認識の不確実性

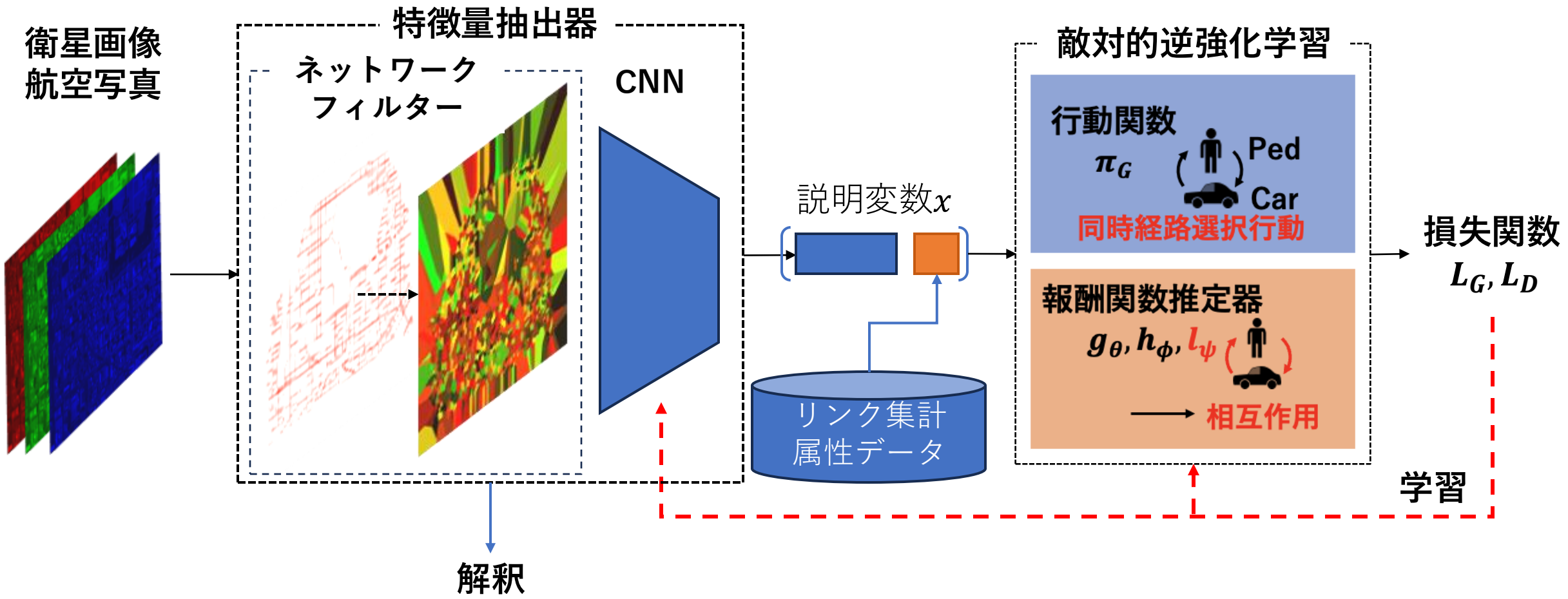
空間構造と相互作用

- 複数交通手段が空間をどのように捉えているかは不明
  →モデル化時に観測可能なのは”客観的な”説明変数のみ
- 社会的増幅をはじめとする複数交通手段の**非対称な相互作用**の表現

→交通手段間の**環境の寄与の相違**を記述する必要



モデル全体図


衛星画像
航空写真
特徴量抽出器
ネットワーク
フィルター
CNN
説明変数x
リンク集計
属性データ
敵対的逆強化学習
行動関数
π_G
Ped
Car
同時経路選択行動
報酬関数推定器
g_θ, h_ϕ, l_ψ
相互作用
損失関数
L_G, L_D
学習
解釈

ケーススタディ

データ

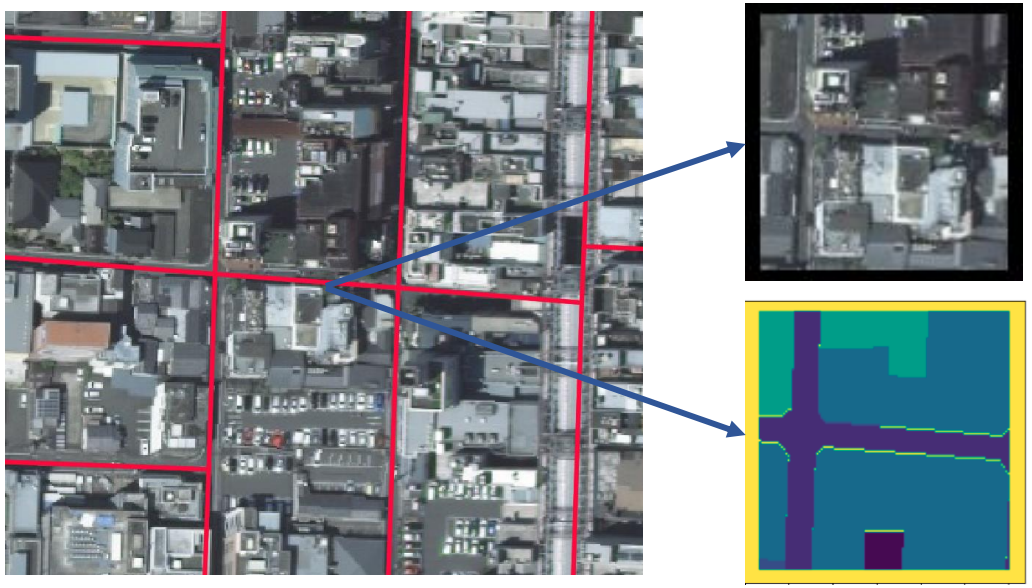
対象敷地：愛媛県松山市

ネットワークデータ

- OpenStreetMap
- 2019年松山市都市計画基礎調査
- 航空写真データ

  地理院地図　全国最新写真（解像度0.5m）

航空写真・衛星画像



都市計画基礎調査 住宅用地，道路，商業用地，…

行動データ：松山街中PP調査

- GPSデータ

  2007年2月-3月

  対象者 108名

  マップマッチングにより経路選択データを作成

  教師データ

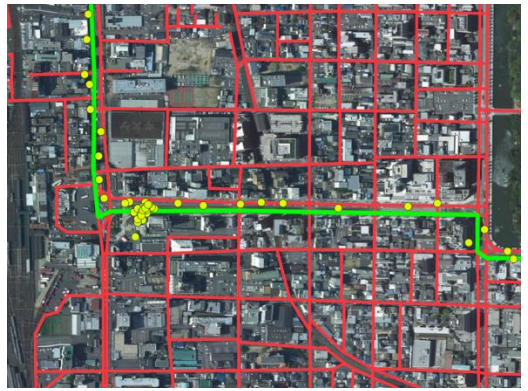
  歩行者 1038行

  自動車 1014行

  テストデータ

  歩行者 221行

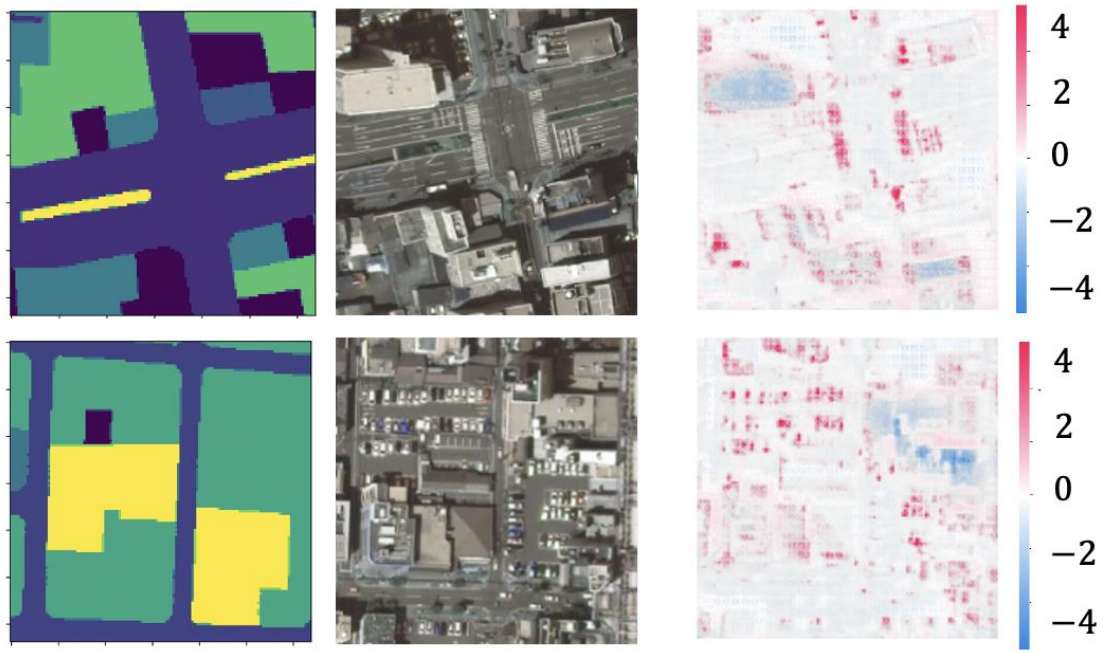
  自動車 212行



：GPS点群

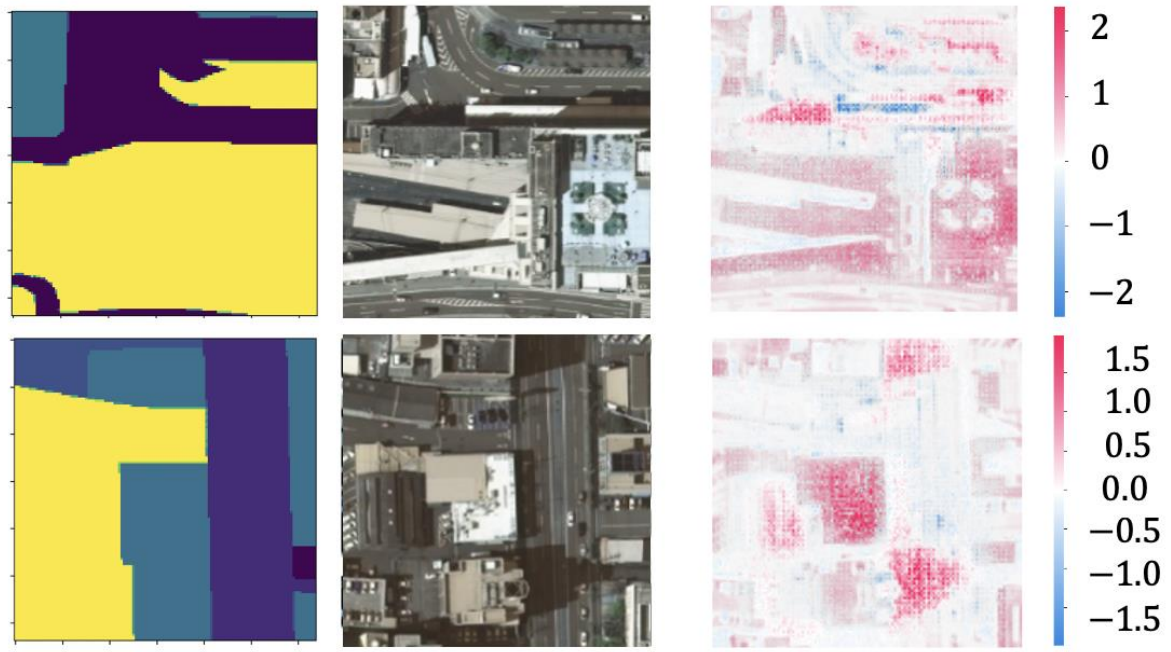
ID	上流リンクID	下流リンクID	目的地ID
1	1	2	1
2	2	3	1
3	3	5	2
⋮	⋮	⋮	⋮

推定結果の解釈



Land use　Image　SHAP

歩行者行動への影響の大きな空間特徴量



Land use　Image　SHAP

車両の行動への影響の大きな空間特徴量

尤度関数による評価

歩車の最終対数尤度

初期対数尤度：歩行者 −2811.40 , 車両 −2025.38

ケース	ノルム制約	画像	最終対数尤度（歩行者）	最終対数尤度（車両）
1			−2707.01	−1776.24
2	○		−2703.66	−1766.62
3		○	−2700.47	−1773.86
4	○	○	**−2694.94**	**−1760.09**

- 車両に対しては正則化と画像特徴量の両方が効いている.
- 歩行者は両方ともあまり効いていない.

→うまく学習できていない.

→データがに単一均衡の仮定を満たしていない可能性

まとめ

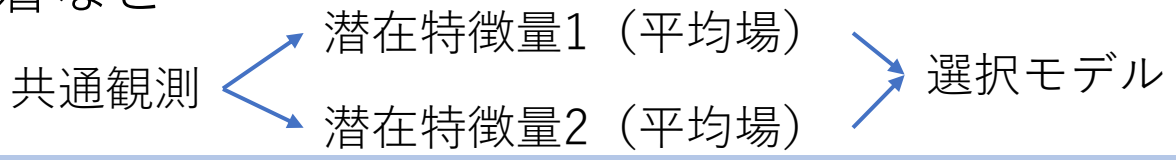
まとめ

- 複数交通手段の**内生的相互作用**に対する安定的・効率的な推定手法
  - 内生変数存在下での最尤推定法と機械学習モデルによる推定の比較
  - **マルコフゲーム**での均衡解の**動的特性と推定過程の安定性**に注目した推定手法

- **非対称な空間的特徴**を取り込んだ街路空間評価手法
  - **一般状態空間モデル**としてのマルチモーダル学習の解釈
  - 環境の空間的特徴の**非対称な寄与**を取り扱うことのできる同時学習手法

今後の課題

- 経路選択行動における空間的特徴の定量的な解釈
- 複数の均衡状態，同一交通手段内の異質性の学習→動学に基づくパラメータのサンプリング
- 空間的特徴を用いたネットワークの制御や設計理論の構築→平均場による一交通手段の問題への帰着など



参考文献

- 福田大輔, 上野博義, & 森地茂. (2004年). 社会的相互作用存在下での交通行動とミクロ計量分析. 土木学会論文集, *765*, 49–64.
- 柳沼秀樹 & 福田大輔. (2008年). 混雑外部性を内生化した離散選択モデルと構造推定. 土木計画学研究・講演集, *37*(225).
- Iryo, T., & Watling, D. (2019年). Properties of equilibria in transport problems with complex interactions between users. *Transportation Research Part B: Methodological*, *126*, 87–114.
- 浦田淳司, 羽藤英二, & 柳沼秀樹. (2016年). 将来効用の動学的異質性を考慮した避難開始選択モデルの構築. *土木学会論文集d3（土木計画学）*, *72*(4), 261–277.
- Aguirregabiria, V., & Mira, P. (2007年). Sequential Estimation of Dynamic Discrete Games. *Econometrica*, *75*(1), 1–53.
- Arjovsky, M., & Bottou, L. (2017年). *Towards Principled Methods for Training Generative Adversarial Networks* (arXiv:1701.04862). arXiv.
- Arjovsky, M., Chintala, S., & Bottou, L. (2017年). *Wasserstein GAN* (arXiv:1701.07875). arXiv.
- Kim, Eui-Jin, and Prateek Bansal. *A deep generative model for feasible and diverse population synthesis.* Transportation Research Part C: Emerging Technologies 148 (2023): 104053.
- Fosgerau, M., Frejinger, E., & Karlstrom, A. (2013年). A link based network route choice model with unrestricted choice set. Transportation Research Part B: Methodological, 56, 70-80.
- Oyama, Y., & Hato, E. (2017年). A discounted recursive logit model for dynamic gridlock network analysis. *Transportation Research Part C: Emerging Technologies*, *85*, 509–527.
- Yu, L., Song, J., & Ermon, S. (2019年). *Multi-Agent Adversarial Inverse Reinforcement Learning* (arXiv:1907.13220). arXiv.
- Miyato, T., Kataoka, T., Koyama, M., & Yoshida, Y. (2018年). *Spectral Normalization for Generative Adversarial Networks* (arXiv:1802.05957). arXiv.