

目的地選択の高解像度化にむけた 汎化性の高いパラメータ推定アルゴリズム

金田 侑大¹・浦田 淳司²・石井 良治³

¹学生会員 東京大学 工学部社会基盤学科 (〒 113-8654 東京都文京区本郷 7-3-1)

E-mail: kaneda@bin.t.u-tokyo.ac.jp

²正会員 東京大学大学院 工学系研究科社会基盤学専攻 (〒 113-8654 東京都文京区本郷 7-3-1)

E-mail: urata@bin.t.u-tokyo.ac.jp

³正会員 東京大学大学院 工学系研究科社会基盤学専攻 (〒 113-8654 東京都文京区本郷 7-3-1)

E-mail: rishii@bin.t.u-tokyo.ac.jp

本研究では、目的地選択モデルのゾーン細分化にむけて汎化性の高いパラメータ推定アルゴリズムを提案した。深層学習分野では、周辺の誤差関数空間がフラットなパラメータほど汎化性能が高いという仮説が立てられており、フラットなパラメータを探索するアルゴリズム Sharpness-Aware Minimization が開発されている。本研究ではこのアルゴリズムを Network GEV モデルへ適用するため、外点ペナルティ法によって制約条件を考慮し、さらに Momentum SGD と組み合わせ、パラメータの振動を抑えたアルゴリズムを構築した。数値実験では、母集団となる人工データからさまざまなサンプルサイズでサンプルデータを抽出し、特に小サンプル下で提案手法から得られたパラメータが高い汎化性を持つことを確認した。

Key Words: *Network-GEV Model, Destination Choice, Parameter Estimation, Deep Learning, Loss Landscape*

1. はじめに

近年、小規模な山村集落におけるモビリティの整備や滞在・交通拠点の開発、ライドヘイリングサービスにおける点ベースでの配車システム設計、移動弱者のためのマイクロランジットなど、細かいスケールでの交通施策が求められるようになってきている。細かいスケールで確度の高い施策を実行するためには、計画のスケールと交通需要予測のスケールを合わせる必要があり、特に、小さいゾーン設定・細かな時間幅での交通需要予測が求められる。

その端緒として、世界中で 2000 年以降、四段階推定法に変わる手法であるアクティビティシミュレータを用いた交通需要予測が行われ始めている。しかし、前述の施策スケールと比してアクティビティシミュレータのスケールは依然として大きく、詳細化が求められる。また、アクティビティシミュレータの多くは複数のサブモデルの組み合わせからなるが、そのうちのひとつである目的地選択モデルは、選択肢である目的地ゾーンのスケールを柔軟に変化させることができる。目的地選択モデルのゾーン設定の細分化の実現は細かなスケールでの交通需要予測に大きく貢献する。

過去、予測精度の向上にはモデルの複雑化と説明変

数の多彩化が進められてきた。Network Generalized Extreme Value モデル (Network GEV モデル)¹⁾ や Mixed Recursive Logit モデル (Mixed RL モデル)²⁾ に代表されるような複雑なモデルは計算コストは当然増えるが、計算機の能力向上により適用可能性が広げられつつある。動的モデルやゾーン特性記述の緻密化に代表されるような説明変数の多彩化も、オープンデータやデジタル観測の進展により適用範囲が増している。

しかし、モデルとデータの向上だけで真に予測精度が向上するだろうか。統計学的にはモデルの複雑性と未知データへの汎化性能はトレードオフであると考えられている。機械学習や深層学習の分野では億を超える多パラメータのモデルが開発されるとともに、その汎化性を保証するための計算アルゴリズムや目的関数の改善が進められてきた。一方で、離散選択モデルの分野では、理論面の精緻化に合わせてモデルを複雑化するのみで目的関数や計算アルゴリズムを改良しようとする試みはあまり見られない。

ゾーンを細かく設定した目的地選択モデルは、OD ペアごとの選択肢固有定数を推定することで多パラメータ推定問題となり、さらに、各選択あたりのデータ数は減少するため小サンプル問題となる。このような多パラメータ・小サンプルの推定問題を過学習すること

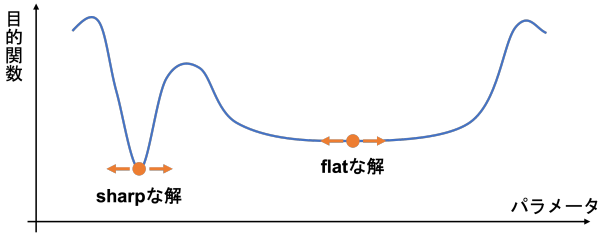


図-1 フラットな解とシャープな解

なく推定するには、汎化性の高いパラメータ推定アルゴリズムが必要である。

そこで本研究では、ゾーン細分化された目的地選択問題を対象問題とし、モデルの汎化性能向上のためのパラメータ推定アルゴリズムを提案する。

2. 汎化性の高いパラメータ推定アルゴリズム

本章では汎化性の高いパラメータを得るパラメータ推定アルゴリズムについて既往研究を整理する。

機械学習、特に深層学習の分野では、データ数を上回るようなパラメータ数を持つモデルでも高い汎化性能を持つ予測が可能となっている。Hochreiter & Schmidhuber(1997)³⁾は、この理由を周辺の損失関数の形状がフラットな解を選んでいるためだとする仮説を示した。図-1のように、損失関数の値は小さいが解周辺で損失関数がシャープな解と、損失関数の値自体は最小ではないが解周辺で損失関数がフラットな解がある場合を考える。母集団における損失関数の形状は、訓練データにおける損失関数の形状と少しズレが生じる。シャープな解はそのズレによって損失関数の値が大きく変化してしまうのに対し、フラットな解では変化が小さい。したがって、フラットな解の方が母集団における損失関数の値が小さくなると考えられる。

Keskar et al.(2016)⁴⁾は、バッチサイズが訓練データのサイズに比べて十分小さい確率的勾配降下法 (SGD: Stochastic Gradient Descent) がフラットな解へと収束しやすいことを実験的に示した。一方で数値実験のみでしか有効性が確認されておらず、実証はされていない。

Foret et al.(2020)⁵⁾はフラットさの指標を定義し、その指標をもとにフラットな解を探索するアルゴリズム Sharpness-aware Minimization (SAM) を開発した。フラットさの定義は次のようである。パラメータ θ における損失関数を $L(\theta)$ とする。定数を ρ ととったとき、

$$\max_{\|\epsilon\|_p \leq \rho} L(\theta + \epsilon) \quad (1)$$

の値を考える (ただし $\|\cdot\|_p$ は p 次ノルム)。この値はパラメータ θ 近傍における損失関数の最大値を表すので、この値が小さいほどそのパラメータはフラットだ

Algorithm 1 Sharpness-aware Minimization

Require: ρ : 探索範囲

Require: p : ノルム次元

Require: η : 学習率

Require: θ_0 : 初期パラメータ

Require: T : 反復回数

1: $t \leftarrow 0$

2: **while** $t < T$ **do**

3: $\mathcal{B} \leftarrow \cup_{i=1}^b \{(x_i, y_i)\}$

4: $\hat{\epsilon}(\theta) = \frac{\rho \cdot \text{sign}(\nabla_{\theta} L(\theta)) \cdot |\nabla_{\theta} L(\theta)|^{q-1}}{(\|\nabla_{\theta} L(\theta)\|_q^q)^{1/p}}$

5: $g = \nabla_{\theta} L(\theta, \mathcal{B})|_{\theta + \hat{\epsilon}}$

6: $\theta_{t+1} = \theta_t - \eta g$

7: $t = t + 1$

8: **end while**

といえる。

ここで、損失関数の最大値をとる ϵ は解析的に

$$\hat{\epsilon}(\theta) = \frac{\rho \cdot \text{sign}(\nabla_{\theta} L(\theta)) \cdot |\nabla_{\theta} L(\theta)|^{q-1}}{(\|\nabla_{\theta} L(\theta)\|_q^q)^{1/p}} \quad (2)$$

(ただし、 $\frac{1}{p} + \frac{1}{q} = 1$)

と解くことができる (ただし $|\cdot|$ はベクトルの各要素の絶対値)。勾配降下法により式 (1) の値を最小化する。 $L(\theta + \hat{\epsilon})$ における勾配ベクトルは、

$$\nabla_{\theta} L(\theta + \hat{\epsilon}) = \nabla_{\theta} L(\theta)|_{\theta + \hat{\epsilon}} + \frac{d\hat{\epsilon}(\theta)}{d\theta} \nabla_{\theta} L(\theta)|_{\theta + \hat{\epsilon}} \quad (3)$$

ここで数値実験上から式 (3) の第二項は無視することができると思われる。第一項のみを勾配とした勾配降下法を解くことでフラットな解を探索することができる。また、 p, q の値は $p=2, q=2$ の場合がもっとも汎化性能が高いことが実験上わかっている。

アルゴリズムは以下となる。訓練データを $\mathcal{S} = \cup_{i=1}^n \{(x_i, y_i)\}$ 、損失関数を $L(\theta, \mathcal{S})$ とおく。毎回のパラメータ更新のたびに、訓練データからバッチサイズ b の分だけバッチデータ \mathcal{B} を選ぶ。その後バッチデータをもとに式 (2) の値を計算する。得られた $\hat{\epsilon}$ を用いて $g = \nabla_{\theta} L(\theta, \mathcal{B})|_{\theta + \hat{\epsilon}}$ を計算し、 $\theta_{t+1} = \theta_t - \eta g$ とパラメータ更新を行う。擬似コードは **Algorithm 1** の通りである。

3. 目的地選択モデル

(1) 採用するモデルと高解像度化の課題

目的地選択モデルには Bekhor & Prashker(2008)⁶⁾ の提案する Generalized Nested Logit モデル (GNL モデル) を採用する。GNL モデルの G 関数は次のように表

される。

$$G(y) = \sum_m \left(\sum_k (\alpha_{mk} y_k)^{\mu_m / \mu} \right)^\mu \quad (4)$$

目的地選択問題においては、調査の便宜上設定されたゾーンと個人の認識するゾーンに違いがあるため隣接するゾーン間に誤差相関が存在する。GNL モデルはこの誤差相関を隣接ゾーンごとにネストを組むことで表現している。一方このモデルは scale parameter の推定が必要であるため、対数尤度関数がパラメータに対して大域的に上に凸にはならず、複数の局所解が存在する。従来の最尤法によるパラメータ推定手法は単に尤度最大となるパラメータを推定結果とするのみであるため過学習の問題が懸念され、汎化性の高いパラメータ推定アルゴリズムの開発が望まれる。さらに、Zhu & Ye(2018)⁷⁾ は目的地選択問題の効用関数における OD ペアごとの選択肢固有定数の必要性を指摘している。本研究でも効用関数に OD ペアごとの選択肢固有定数を組み込み、推定を行う。

本研究のモデルはゾーン細分化すればするほど、多パラメータ・小サンプルな推定問題となる。分析対象地域に計 J 個のゾーンを設定し、ある出発ゾーンにいる個人が自らのいるゾーンも含めた J 個の選択肢の中から目的地ゾーンを選ぶモデルを考える場合、モデル全体で推定すべき選択肢固有定数の個数は $J(J-1)$ 個となる。実際には、他の説明変数に対してもパラメータを設定するため、さらに推定すべきパラメータ数は増加する。

また、得られた選択行動のデータ数を N とすると、ある 1 つのゾーンを出発ゾーンとするデータは平均 N/J 個しか存在しない。実際にはゾーン内人口などによりデータの偏りが生じるため、トリップが観測されない OD ペアや出発データ数が少ないゾーンが存在する。トリップが観測されない OD ペアではその間の効用確定地を負の無限大だと推定されてしまう恐れがあるし、出発データ数の少ないゾーンでは不安定なパラメータ推定となる。

(2) 適用手法との関係性

前章で述べたフラット性の仮説を本問題にも適用可能であることを説明する。あるゾーンから個人 i が目的地ゾーンを選択する行動を考える。個人や発着ゾーンに関する説明変数ベクトル \mathbf{x}_i と選択結果 d_i からなるデータを (\mathbf{x}_i, d_i) とする。母集団 $\mathcal{D} = \cup_{i=1}^N \{(\mathbf{x}_i, d_i)\}$ からサンプルデータ $\mathcal{S} = \cup_{i=1}^M \{(\mathbf{x}_i, d_i)\} \subset \mathcal{D}$ を抽出し、モデル推定を行う。

データ (\mathbf{x}_i, d_i) に対してモデルのパラメータが θ のときの対数尤度を $LL(\theta, \mathbf{x}_i, d_i)$ と定める。モデルの真のパラメータ $\hat{\theta}$ は母集団における最尤推定を行うこと

で求められる。母集団における対数尤度

$$LL_{\mathcal{D}}(\theta) = \sum_{i=1}^N LL(\theta, \mathbf{x}_i, d_i) \quad (5)$$

を最大化する $\theta_{\mathcal{D}}$ は最尤推定量の不偏性・一致性より、

$$P(\|\hat{\theta} - \theta_{\mathcal{D}}\| < \varepsilon) \rightarrow 1 \quad (N \rightarrow \infty) \quad (6)$$

$$E_{\mathcal{D}}(\theta_{\mathcal{D}}) = \hat{\theta} \quad (7)$$

を満たす。母集団のサンプル数は十分大きいいため、式(6)の左辺は十分 1 に近くなる。一方、サンプルデータのみでは $\theta_{\mathcal{D}}$ は得られず、代わりにサンプルデータにおける対数尤度

$$LL_{\mathcal{S}}(\theta) = \sum_{i=1}^M LL(\theta, \mathbf{x}_i, d_i) \quad (8)$$

を用いてパラメータ推定を行い、パラメータ $\theta_{\mathcal{S}}$ を得る。同様に最尤推定量の不偏性より $E_{\mathcal{S}}(\theta_{\mathcal{S}}) = \hat{\theta}$ が成り立つので、平均 $\mathbf{0}$ の誤差ベクトル ε_{θ} を用いて

$$\theta_{\mathcal{S}} = \hat{\theta} + \varepsilon_{\theta} = \theta_{\mathcal{D}} + \varepsilon_{\theta} \quad (9)$$

と表せる。対数尤度値についてもパラメータの不偏性が保たれていれば、

$$LL_{\mathcal{D}}(\theta_{\mathcal{D}}) \approx \frac{N}{M} LL_{\mathcal{S}}(\theta_{\mathcal{D}}) = \frac{N}{M} LL_{\mathcal{S}}(\theta_{\mathcal{S}} + \varepsilon_{\theta}) \quad (10)$$

と近似できる。

従来のパラメータ推定では、サンプルデータからパラメータ推定を行う場合、

$$\arg \min_{\theta} LL_{\mathcal{D}}(\theta) = \theta_{\mathcal{S}} \quad (11)$$

を暗に仮定している。しかしながら、式(9)、式(10)に示した通り

$$\arg \min_{\theta} LL_{\mathcal{D}}(\theta) = \theta_{\mathcal{S}} + \varepsilon_{\theta} \quad (12)$$

である。そのため、交通需要予測シミュレーションにおいて汎化性能を上げる、すなわち $LL_{\mathcal{D}}(\theta_{\mathcal{S}})$ を最大化するには、式(10)より

$$\frac{N}{M} LL_{\mathcal{S}}(\theta_{\mathcal{S}} + \varepsilon_{\theta}) \quad (13)$$

を最大化させる方が望ましい、言い換えると、サンプルデータの対数尤度関数において、パラメータ周囲 $|\varepsilon_{\theta}|$ の範囲で対数尤度が大きく減少しないパラメータを選ぶことが望ましい。これはフラットな解ほど、母集団における良い対数尤度値を達成できるということである。

4. 提案アルゴリズム

SAM のアルゴリズムを前章の目的地選択問題に適用し安定して推定するために、次の 2 つのアルゴリズムの改良を行った。

(1) 外点ペナルティ法

GNL モデルの scale parameter には制約条件が存在し, **Algorithm1** を単純に適用することができない. 本研究ではモデルの対数尤度 LL に以下のような外点ペナルティ法を適用することで, scale parameter の制約条件を緩和する.

$$LL' = LL + \lambda \left(\sum_m \text{penalty}(\mu_m) + \text{penalty}(\mu') \right) \quad (14)$$

ただし, μ' は

$$\mu' = \frac{\mu}{\min_m \mu_m} \quad (15)$$

により求められる値で $0 \leq \mu' \leq 1$ を満たし, penalty 関数は

$$\text{penalty}(\mu) = \begin{cases} |\mu|^2 & (\mu \leq 0) \\ 0 & (0 \leq \mu \leq 1) \\ |\mu - 1|^2 & (\mu \geq 1) \end{cases} \quad (16)$$

によって表わされる関数である. λ はハイパーパラメータであり, 学習ごとに選ぶ必要がある.

(2) Momentum SGD・学習率スケジューリング

scale parameter は他のパラメータと比較して勾配ベクトルの絶対値が大きく, 振動を起ししやすい. その結果 scale parameter を安定して推定することができないため, Momentum SGD と学習率スケジューリングの二つの手法を組み合わせることで scale parameter の振動を抑える.

Momentum SGD⁸⁾ は過去のパラメータの更新方向をパラメータの更新ベクトルに加えることで, パラメータの振動を防ぎ速くかつ安定的に推定問題を収束させる. SAM へ適用すると,

$$v_{t+1} = \alpha v_t - \eta \nabla_{\theta} L(\theta, \mathcal{B})|_{\theta_t + \varepsilon} \quad (17)$$

$$\theta_{t+1} = \theta_t + v_{t+1} \quad (18)$$

のようにパラメータ更新を行う. ただし α はハイパーパラメータで, $v_0 = \mathbf{0}$ である.

学習率スケジューリング⁹⁾ は学習が進むにつれて学習率を減少させていくことで収束の安定性を図る手法である. 本節では, i 回目のパラメータ更新時の学習率 η_i を

$$\eta_i = \frac{\eta_0}{i^p} \quad (19)$$

とすることで, 学習率を学習が進むにつれて減少させていく. η_0 は初回の学習率を指すハイパーパラメータ, p は学習率の減少の割合を表すハイパーパラメータである.

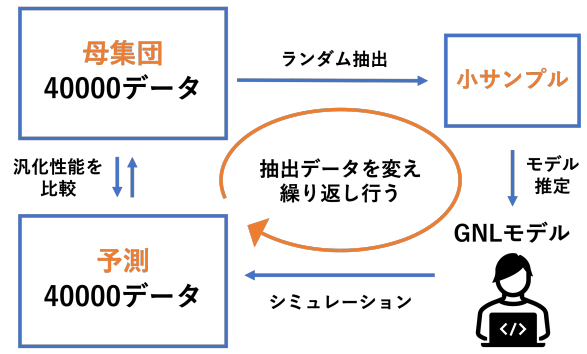


図-2 数値実験の概要

表-1 差分進化法のハイパーパラメータ

戦略	best/1/bin
1 世代の候補数	15
スケール係数	[0.5,1] から世代ごとにランダム抽出
交叉率	0.7

5. 数値実験

本章では, 尤度最大化によって得られるパラメータと比較して, フラット性を考慮したパラメータを用いた方がモデルの汎化性能が高くなることを, 数値実験により示す.

(1) 実験設定

本実験では Bekhor & Prashker(2008)⁶⁾ のモデルに基づいて生成した, 仮想の都市空間における人工トリップデータを用いる. データ数は 40000 である. このトリップデータは以降「母集団データ」と呼ぶ. 次に母集団データからランダムに 40, 400, 4000 データの「サンプルデータ」を抽出する. このサンプルデータを学習用データとし, Bekhor & Prashker(2008)⁶⁾ のモデルを用いて (2) 項で挙げる各推定手法でパラメータ推定を行う. 最後に, 得られたパラメータと「母集団データ」を用いて母集団における対数尤度を計算し, 母集団のトリップの再現性を評価する. この作業を 100 回繰り返した. 実験設定を図解したものを図-2 に示す. なお, 比較のための評価尺度については (3) 項で説明する.

(2) 比較する手法

本実験では, 大域的な尤度最大点を求めるパラメータ推定手法として差分進化法 (DE: Differential Evolution) を採用する. フラット性を考慮したパラメータ推定手法としては提案手法 (SAM) のほかに確率的勾配降下法 (SGD: Stochastic Gradient Descent) でも実験を行い, それぞれの手法で得たパラメータの汎化性能について

表-2 確率的勾配降下法のハイパーパラメータ

パラメータ	N=40	N=400	N=4000
バッチサイズ	4	16	256
学習率	0.01	0.01	0.001
ペナルティに対する係数	100	100	10000
反復回数	30000	20000	15000

表-3 提案手法のハイパーパラメータ

パラメータ	N=40	N=400	N=4000
バッチサイズ	4	16	256
初期の学習率	0.01	0.01	0.001
学習率の減衰率	0.2	0.2	0.2
ノルム次元	2	2	2
探索範囲	0.05	0.05	0.05
Momentum SGD の α	0.9	0.9	0.9
ペナルティに対する係数	100	100	10000
反復回数	40000	20000	15000

も比較する。ハイパーパラメータは表-1, 表-2, 表-3に示す通りである。

(3) 評価尺度

汎化性能の評価尺度には、母集団における対数尤度、パラメータ推定値の分布の2つの指標を用いる。母集団における対数尤度は以下のように求める。母集団を $S = \cup_{i=1}^n \{(x_i, y_i)\}$ とし、サンプルデータから得られた推定パラメータ θ と説明変数 x_i から得られる選択肢 k の選択確率を $P(k, \theta, x_i)$ とする。このとき、

$$\sum_{i=1}^n \delta_{ik} \log P(k, \theta, x_i) \quad (20)$$

が母集団における対数尤度である（ただし δ_{ik} は個人 i が選択肢 k を選択した時 1, そうでないとき 0 をとる変数である）。

さらに、それぞれの手法で大域的な尤度最大点もしくはフラットな点に収束しているかを確認するため、最終尤度・フラットさの指標でも比較する。フラットさの指標は Keskar et al.(2016)⁴⁾ の提案した指標を用いる。解 θ まわりのフラットさは訓練データにおける対数尤度 $LL(\theta)$ と定数 ρ を用いて、

$$\frac{\max_{\|\epsilon\| \leq \rho} LL(\theta) - LL(\theta + \epsilon)}{1 + LL(\theta)} \quad (21)$$

と定義する。この値が小さいほどパラメータはフラットである。

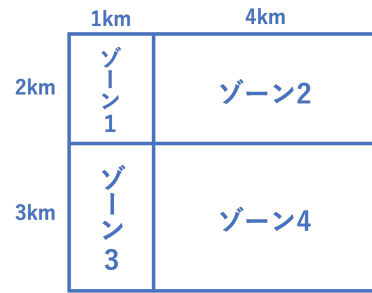


図-3 仮想的な都市空間の構成

(4) 母集団データの作成

母集団データは以下のような目的地選択シミュレータを用いて作成した。

まず、図-3 のような仮想的な都市空間を考える。各ゾーンには 10000 人ずつ人が住んでいて、居住地は一樣分布から生成される乱数によって決定されている。

そして、ゾーン i に住む y 個人 n がゾーン j へ移動するときの効用の確定項を以下のように定める。

$$V_{ij}^n = c_{ij} + \beta_1 d_{ij}^n + \beta_2 \log C_j^n + \log A_j \quad (22)$$

ただし、 c_{ij}, β_1, β_2 は個人間で共通のパラメータで、 d_{ij}^n は個人 n の居住地からゾーン j のセントロイドまでの距離、 C_j^n は個人 n にとってのゾーン j の魅力度、 A_j はゾーン j の面積を指す。 C_j^n はそれぞれ独立な対数正規分布に従うと設定する。

また、効用の確定項 V_{ij}^n である個人 n のゾーン選択は以下の G 関数をもつ GNL モデルによって決定する。

$$G(y) = \sum_{m=1}^4 \left(\sum_{k=1}^4 (\alpha_{mk} y_k)^{\mu_m / \mu} \right)^\mu \quad (23)$$

さらに allocation parameter である α_{mk} は Bekhor & Prashker(2008)⁶⁾ と同様に定める。

$$\alpha_{mk} = \frac{L_{mk}}{L_k} \quad (24)$$

ただし、 L_{mk} はゾーン m とゾーン k の接する長さ、 L_k はゾーン k 全周の長さとする。 L_{mm} はゾーン m の他のゾーンと接していない部分の長さであり、ゾーン m とゾーン k が接していない場合 $L_{mk} = 0$ とする。

(5) 結果

a) 対数尤度の比較

各手法で得たパラメータを用いて計算した、サンプルデータにおける対数尤度を図-4 に、母集団における対数尤度を図-5 に示す。サンプルデータにおける対数尤度はどの手法も変わらないため、十分学習が進行していることがわかる。母集団における対数尤度については、N=40 の場合は SAM と SGD が同程度高い尤度、N=400

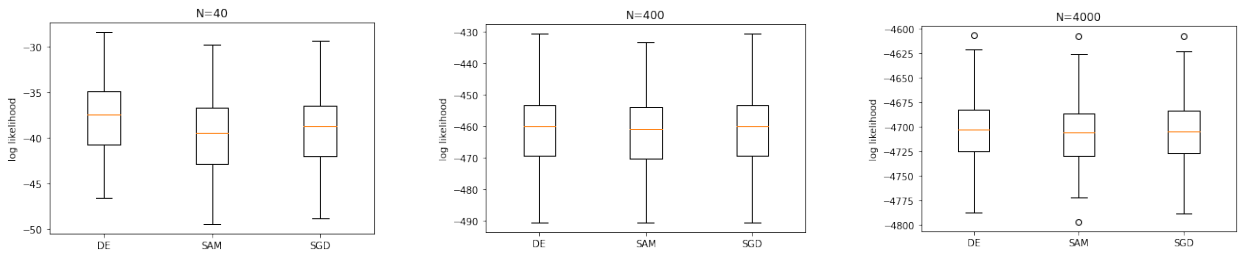


図-4 サンプルデータにおける対数尤度の比較

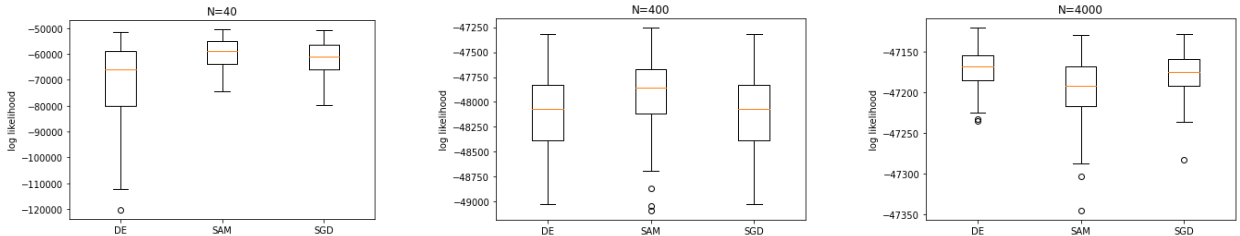


図-5 母集団における対数尤度の比較

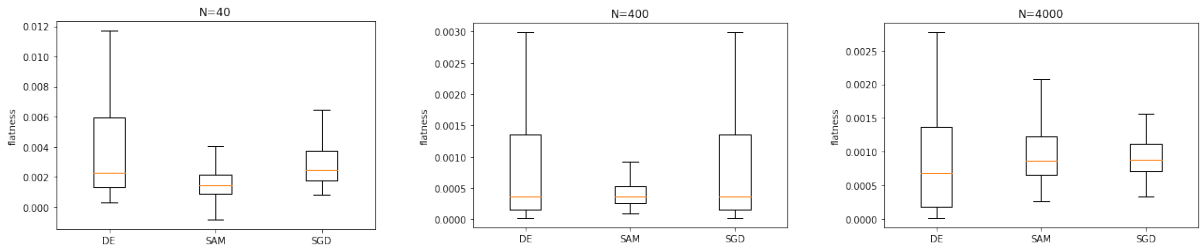


図-6 フラットさの比較

の場合は SAM が高い尤度、N=4000 の場合は DE がやや高い尤度を示した。

提案手法である SAM は、小サンプルの場合に高い汎化性能を示し、サンプル数が増えても他の手法と変わらない汎化性能を示すことがわかった。また、DE は小サンプルの場合に過学習を起こし汎化性能を大きく下げることが確認できた。

b) パラメータ推定値の比較

パラメータ推定値を比較した結果を図-7、図-8、図-9 に示す。青線がパラメータの真値を指す。特に N=400 や N=40 で、提案手法の推定するパラメータ分布に真値が含まれることが確認された。フラットな解に収束していると考えられる SAM や SGD に比べ、DE のパラメータ推定値のばらつきが大きいことから、単純な尤度最大化では得られるパラメータがデータによってばらつくことも確認できた。

c) フラットさの指標を用いた比較

フラットさの指標を比較した結果を図-6 に示す。N=400 と N=40 の場合は SAM や SGD は DE に比べてフラットな解に収束していることがわかる。特に小

サンプルである N=40 の場合に SAM の解がフラットな解に収束していることもわかる。一方、N=4000 の場合に SAM のフラットさが大きくなっているが、これは確率的なパラメータ選択の中でフラットさの大きい状態で学習が終了してしまったためであると考えられる。

6. 結論

(1) 本研究の成果

本研究では、目的地選択モデルの高解像度化にむけて、小サンプル・多パラメータの状況でも汎化性の高いパラメータを推定できるパラメータ推定アルゴリズムを提案した。本研究で提案する SAM というアルゴリズムは、解周辺の目的関数の値がフラットな解を探索するものである。本研究では、SAM を Network GEV モデルのパラメータ推定に適用するため、scale parameter の制約条件を満たすための外点ペナルティ項の導入、探索時のパラメータ更新の安定化のための Momentum 型のアルゴリズムの導入、という 2 つの改良を行った。この改良により、提案アルゴリズムは対数尤度関数上で安定して収束するようになった。

本研究の対象問題は目的地選択モデルのゾーン細分化による高解像度化であったが、提案したアルゴリズムは別の高解像度化、たとえば時間幅の細分化や個人の異質性を考慮したユーザークラスの細分化などにも適用可能である。移動弱者を対象としたマイクロランジットのような施策は、より細かな個人特性・空間・時間の分解能での需要予測の下で検討されるべきであり、小サンプルでも汎化する本研究に推定アルゴリズムの貢献可能性は大きい。さらには、目的地選択モデルに限らず、交通分野・離散選択問題の別の問題で用いられる多峰なモデルに対しても適用可能である。提案アルゴリズムを用いてパラメータ推定をすることで、多パラメータ時の汎化性が保証されるため、さらなる複雑なモデルを実データへ適用する可能性も広がる。

(2) 今後の展望

今後の課題を挙げる。一つに、Momentum SGD と組み合わせて振動を抑えてもなお、scale parameter の推定が不安定な点である。サンプルデータのデータ数が十分に大きい場合でも scale parameter の真値を推定することは難しい。このことは提案アルゴリズムに限らず、どのパラメータ推定手法でも言えることであり、scale parameter を安定的に推定する手法が求められる。定式化上、scale parameter の感度は、効用関数の説明変数のパラメータよりも大きく、その特性を踏まえたアルゴリズム開発が必要だと考える。

また、深層学習分野では、汎化性能の高いパラメータ推定手法を導入することで、パラメータ数がデータ数を上回るような状況下でも過学習を引き起さずに、汎化性能が向上し続ける double descent という現象が確認されている。本研究でフラットな解のアルゴリズムで汎化性能が向上することが確認できており、double descent の現象が離散選択モデル分野でも起こることが確認されることが期待される。この現象が確認できれば、allocation parameter や scale parameter を構造化せずに全パラメータを推定する Network GEV モデルで高い汎化性能が得られることが予測される。

謝辞: 東京大学の羽藤英二教授には、研究の発展にあたり、有意義な助言をいただいた。また、本研究の一部は JST さきがけ特定課題調査の助成を受けたものです。ここに感謝の意を示します。

参考文献

- 1) Daly, A. and Bierlaire, M.: A general and operational representation of generalised extreme value models, *Transportation Research Part B: Methodological*, Vol.40, No.4, pp.285–305, 2006.
- 2) Mai, T., Bastin, F., and Frejinger, E.: A decomposition

method for estimating recursive logit based route choice models, *EURO Journal on Transportation and Logistics*, Vol.7, No.3, pp.253–275, 2018.

- 3) Hochreiter, S. and Schmidhuber, J.: Flat Minima, *Neural Computation*, Vol.9, No.1, pp.1–42, 01 1997.
- 4) Keskar, N. S., Mudigere, D., Nocedal, J., Smelyanskiy, M., and Tang, P. T. P.: On large-batch training for deep learning: Generalization gap and sharp minima, *CoRR*, Vol.abs/1609.04836, 2016.
- 5) Foret, P., Kleiner, A., Mobahi, H., and Neyshabur, B.: Sharpness-aware minimization for efficiently improving generalization, *CoRR*, Vol.abs/2010.01412, 2020.
- 6) Bekhor, S. and Prashker, J. N.: Gev-based destination choice models that account for unobserved similarities among alternatives, *Transportation Research Part B: Methodological*, Vol.42, No.3, pp.243–262, 2008, A Tribute to the Career of Frank Koppelman.
- 7) Zhu, J. and Ye, X.: Development of destination choice model with pairwise district-level constants using taxi gps data, *Transportation Research Part C: Emerging Technologies*, Vol.93, pp.410–424, 2018.
- 8) Qian, N.: On the momentum term in gradient descent learning algorithms, *Neural Networks*, Vol.12, No.1, pp.145–151, 1999.
- 9) Robbins, H. and Monro, S.: A Stochastic Approximation Method, *The Annals of Mathematical Statistics*, Vol.22, No.3, pp.400 – 407, 1951.

(Received ?)

(Accepted ?)

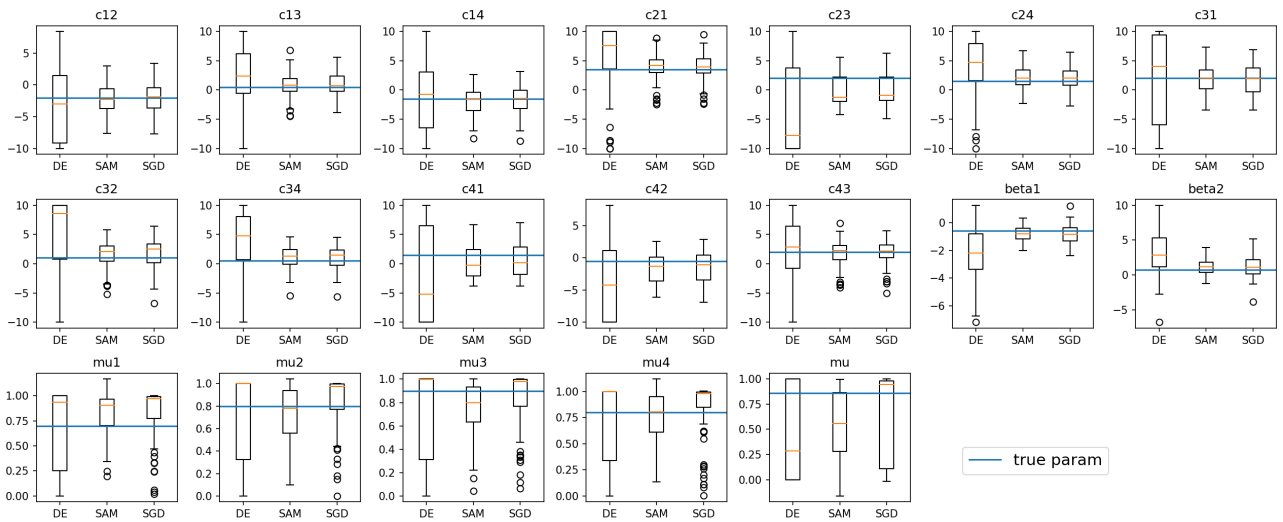


図-7 N=40でのパラメータ推定値と真値

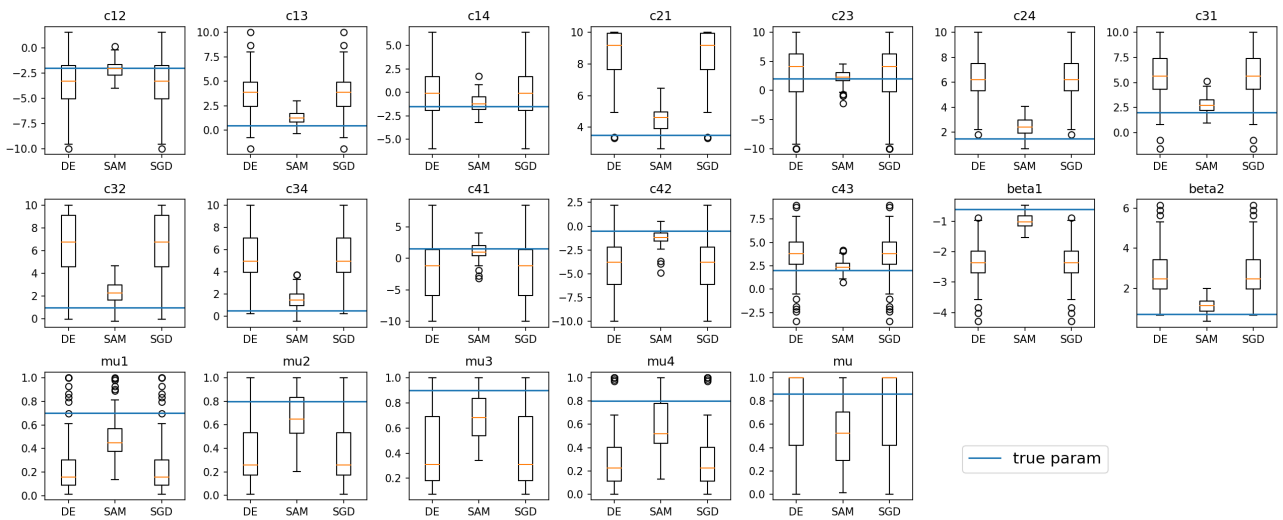


図-8 N=400でのパラメータ推定値と真値

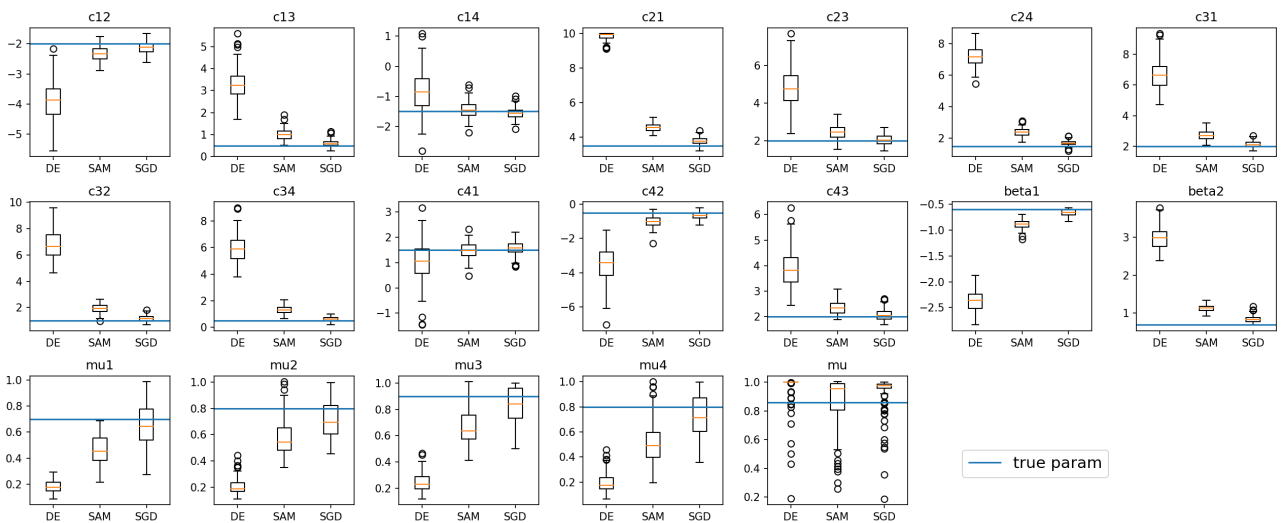


図-9 N=4000でのパラメータ推定値と真値

Parameter Estimation Algorithm with High Generalizability for High Resolution Destination Choice Model

Yudai KANEDA, Junji URATA and Ryoji ISHII

This paper proposes a generalizable parameter estimation algorithm for zone segmentation in a destination selection model. In the field of deep learning, it is hypothesized that parameters with a flat error function space have higher generalization performance, and an algorithm called Sharpness-Aware Minimization has been developed to search for such parameters. To apply this algorithm to the Network GEV model, we used the outlier penalty method to take into account constraint conditions and combined it with Momentum SGD to construct an algorithm that suppresses parameter oscillations. In numerical experiments, sample data were extracted from artificial population data at various sample sizes, and we confirmed that the parameters obtained from the proposed method have high generalizability, especially under small sample sizes.