

複数の集計データの融合による 非集計交通手段選択モデル

日高 健¹・志賀 孝広²

¹株式会社豊田中央研究所 社会システム研究領域 (〒 480-1192 愛知県長久手市横道 41-1)
E-mail: hidaka@mosk.tytlabs.co.jp

²株式会社豊田中央研究所 社会システム研究領域 (〒 480-1192 愛知県長久手市横道 41-1)
E-mail: t-shiga@mosk.tytlabs.co.jp

移動需要推定のためのモデルの多くは大規模な交通行動調査を必要とするが、多くの地方自治体では、近年、金銭的・時間的制約から、それら調査の実施が困難になってきている。一方、位置情報データを利用した人の移動行動理解の研究も標準的な方法として一定の地位を築きつつもあるものの、プライバシー保護のためにその活用はまだ限定的である。そこで、我々は集計データのみを利用して個々のエージェントの交通手段を付与するための非集計の交通手段選択モデルを提案する。具体的には、通勤・通学利用交通手段の集計データと他の集計データから生成された 1 日の疑似的な非集計データを利用して、通勤・通学交通手段選択モデルを推定する。集計データによる個々の選択結果の推定はデータ拡大法（潜在的な選択を観測された集計データと一致するようにシミュレートする方法）を通じて行われる。さらに、トリップチェーン行動を利用して他の目的に対しても交通手段を付与する。これらのモデルを地区別の交通手段割合の推定に適用した結果、直観と矛盾しない結果を得ることができた。提案モデルを適用することにより、元のデータでは明らかにされない集中地点でのごとの集約化、属性や目的とのクロス集計など様々な視点から人の移動行動の分析が可能となる。

Key Words: transportation mode choice model, aggregate data, human mobility and activity

1. はじめに

人の移動行動の理解は、移動需要予測を主として長年に渡り発展を続けてきた。歴史的には、1950 年代以降、集計型の交通需要推定手法である四段階推定が大きな役割を担い、その後、移動が個人の活動から派生するとの考えのもとにアクティビティベースモデルが登場し、世帯や個人単位での移動行動の理解へと進んでいった^{1),2),3),4),5),6)}。アクティビティベースモデルは、人の移動行動の理解に対して多大な貢献を果たしてきた一方で、パーソントリップ調査のような大規模な交通調査を必要とする。しかしながら、こうした大規模な交通調査は、時間的・金銭的なコストの大きさから実施数が減少してきているのが現状である。

一方で、ICT 技術の進展、スマートフォンの普及により個人の位置情報データを活用した人の移動行動理解の研究も盛んになった。そこには、GPS を利用したものや^{7),8)}、携帯電話の通信履歴を利用したもの^{9),10)}、SNS の情報を利用したもの^{11),12)}などが挙げられる。しかし、個人の位置情報データの活用は個人情報保護の観点から利活用には制約を受けることも多い。

このような背景から、著者ら¹³⁾は、国勢調査や生活時間調査のような個人情報を一切含まない集計データに着目し、これらを融合することにより、疑似的に人

の移動行動データを生成する手法を開発している。この人の移動行動の疑似データは、年齢、性別、居住エリア、1 日の行動スケジュール、1 日の滞在地を情報として持つ多数の疑似的な個人（エージェント）の集合として表現される。既存の研究においても、アクティビティベースモデルの適用により多数のエージェントを生成し、都市圏規模での移動需要を推定する研究も存在するが、モデルの推定にあたり大規模な非集計の交通行動調査を必要とするため、データ収集の観点に課題がある。もちろん、大都市においてはこうした交通行動調査データが揃っており、適用可能な場合も多いが、時間的・金銭的成本などの理由から調査ができない自治体においては実施が困難である。著者らが提案する手法の特徴は、国勢調査などの公的機関が公開する、いわゆるオープンデータや市販の統計データのみを用いて人の移動行動を生成できる点にある。使用するデータは、基本的には日本全国をカバーするものであり、そのため日本においては大規模な調査が困難であった自治体に対しても安価に、かつ簡便に移動行動データを提供することが可能となり、交通政策立案への活用が期待される。

しかしながら、これまでのモデルでは、誰が（どんな属性）、どんな目的で、どこに行くかは生成できるも



図-1: 移動行動生成モデルの概要

の、どうやって、すなわち何の交通手段で移動するかについては生成できておらず、交通計画を行うにあたっては不十分であった。そこで、本研究では、これまでのモデルと同様に集計データのみを利用することを特徴とした非集計の交通手段選択モデルについて新たに提案する。

個別の選択データは、プライバシー保護、データベースの容量制約等に起因して集計化されて公開されることがある。この集計化処理により情報が限定化され、統計的な分析の妨げとなる。このような集計データから個々のレベルの選択モデルを推定する問題への取組みは、マーケティングリサーチ¹⁵⁾、計量生物学¹⁶⁾、政治学^{17),18)}等多岐に渡る分野の問題への適用が既に行われているが、我々の知る限り交通分野への適用例はまだない。特に、交通手段選択は、属性・目的地等に大きく依存するため、非集計データなしにこれを行うことは難しい。このような問題に対して、本研究では、著者ら¹³⁾の提案した人の移動行動生成モデルによって生成される疑似的な非集計データの生成を利用し、データ拡大法(Data augmentation)¹⁹⁾によって欠損した交通手段を推定しつつ、交通手段選択モデルのパラメータ推定を行う手法を提案する。

2. 人の移動行動生成モデル

(1) 全体概要

提案する移動行動生成モデルは3つのステップにより移動行動の生成を行う(図-1)。移動行動生成モデルの詳細に関しては日高ら¹³⁾に譲り、以下では概略を述べる。

(2) 属性

国勢調査の地域メッシュ統計に基づき、性別、年齢(5歳区分)、居住地メッシュ(500m²)を持つ疑似的な個人(エージェント)を生成する。

(3) 行動系列

総務省の社会生活基本調査²⁰⁾やNHK国民生活時間調査²¹⁾等のいわゆる生活時間調査に基づき、エージェントの1日の行動系列(行動スケジュール)を付与す

る。生活時間調査は主に2つのデータから構成される。1つはある時刻において該当行為を行っている人の割合を指す「時間帯別行動者率」であり、もう1つは1日の中で該当行為を少しでもした人の割合を指す「1日の行動者率」である。しかしながら、この2つのデータからは隣接する2つの時間帯間の行動の遷移確率は与えられないため推定を行う必要がある。

そこで、著者ら¹³⁾は行動の間に1次のマルコフ性を仮定し、「時間帯別行動者率」を制約とした以下の(拡張された)エントロピー最大化問題として遷移確率を求める問題を定式化した。

$$\text{Minimize } \sum_{i,j} (1 - \beta_i \delta_{ij}) a_{ij}^{(t)2} \quad (1a)$$

$$\text{subject to } \sum_i a_{ij}^{(t)} = 1, \quad (1b)$$

$$a_{ij}^{(t)} > 0 \quad \forall i, j, \quad (1c)$$

$$y_i^{(t+1)} - \sum_j a_{ij}^{(t)} y_j^{(t)} = 0. \quad (1d)$$

ここで、 β_i は同じ行為の続きやすさを制御するパラメータ、 $a_{ij}^{(t)}$ は時刻 t において行為 j を選んだ人が、時刻 $t+1$ において行為 i を選択する確率、 $y_i^{(t)}$ は時刻 t における行為 i の行為者率を表す。式(1b)と(1c)は遷移確率行列が満たすべき性質を表しており、式(1d)は各時間帯における行為者率を保証するための条件になっている。 β_i は「1日の行為者率」と最も適合するように調整し、そのときの遷移確率行列が解として得られる。「1日の行動者率」は1次のマルコフ性という仮定上、必ずしも一致するとは限らない。解として求められた遷移確率行列を用いることで、確率的に行動の系列を付与することができる。

(4) 目的地系列

次に、1日の滞在場所(目的地)の付与について説明する。移動は行動が変化するときが発生すると仮定し、行動が変化した際にはWilsonのエントロピーモデル²²⁾から求められる選択確率を用いて目的地を選択する。目的地の選択確率は、

$$P(\text{pos}^{(t+1)} = m \mid \text{act}^{(t+1)} = i \neq j, \text{pos}^{(t)} = l) = \frac{A_l B_m O_l D_m e^{-\gamma c_{lm}}}{\sum_{m'} A_l B_{m'} O_l D_{m'} e^{-\gamma c_{lm'}}}. \quad (2)$$

と表すことができる。ここで、 $\text{pos}^{(t)}$ は時刻 t における位置、 $\text{act}^{(t)}$ は時刻 t における行動、 O_l はメッシュ l における発生量、 D_m はメッシュ m における集中量、 c_{lm} は lm 間の距離、 A_l, B_m, γ はパラメータを表す。このモデルでは、あるメッシュ間の移動量は、発生量(移

動需要)と集中量(魅力度)にそれぞれ比例し,また距離が近いほど移動量が大いことを表現している。

(5) 交通手段モデルの位置づけ

前節までは,提案してきた人の移動行動生成モデルの概略について説明してきた。本節では,人の移動行動生成モデルにおける提案手法の位置づけについて説明を行う。

人の移動行動生成モデルにおける因果関係の仮定をさらに拡張する。すなわち,移動(目的地)が交通手段を規定するという仮定である。もちろん,人の移動行動全てがそのような因果関係になっているとは言えないが,目的地が決まった後に交通手段を考えるという仮定は,出発地/目的地をクエリとして経路を探索し,所要時間や費用を比較して交通手段を決める,今日における我々の情報検索行動と照らし合わせてみても不自然な仮定ではないだろう。

3. データ

では,どのような入力データをもとに交通手段選択モデルを推定すればよいか。交通手段選択モデルでは一般的にパーソントリップ調査のような非集計の交通調査データが用いられることが多い。これは,サンプル調査(～数%)であるものの非常に大規模な交通行動のための調査であり,サンプルの実際の交通手段選択結果が得られることに起因している。しかしながら,パーソントリップ調査は提供先・使用目的に制限があり,一般に必ずしも入手は容易ではない。また,主に主要都市圏で行われており,対象エリアが限定されるというデメリットも持つ。大都市交通センサスは三大都市圏における公共交通の利用実態把握のための調査であるため,公共交通に限って言えば最も有用なデータといえるが,その他の交通手段が含まれていないためにかなり限定的なデータといえる。国勢調査は10年に一度の大規模調査とその中間年の簡易調査の2種類に分かれる。その大規模調査のなかで,通勤・通学に関する利用交通手段調査が行われている。このデータは移動目的が通勤・通学に限定されているものの,その強みは全数調査であることであり,また入手も容易である。それぞれの交通調査の種類とその特徴を表-1にまとめた。

データの入手が容易であること,日本全国をカバーするデータであること,地域メッシュ単位での集計もされていて現状の人の移動行動生成モデルとの相性がいいことなどの理由から国勢調査の通勤・通学利用交通手段データを用いることとした。移動目的に関しては,通勤・通学と限定的ではあるものの,トリップチェ

イン行動を考えることで合理的な交通手段の付与が可能である。トリップチェーン行動を利用した移動手段付与の妥当性については(2)節で詳しく議論する。

図-2に国勢調査の通勤・通学利用交通手段調査データから計算した¹岡山市周辺の交通手段の分担率を示す。図の例は,4次メッシュ(500mメッシュ)分担率を示しており,高分解能のデータが得られる様子が分かる。図からは,どの交通手段に対しても比較的きれいに空間的な相関が表れており,特に岡山駅や倉敷駅を中心に市街地では家用車分担率が低い(青色)一方で,徒歩や自転車の分担率が高い(橙色)様子や鉄道路線網が未発達な岡山市南部域では鉄道分担率が低い様子など地域の公共交通の発達の度合いや家用車の利用状況を反映したデータであることが確認された。

4. 方法

(1) 通勤・通学利用交通手段選択モデル

交通手段選択の基本モデルとして,非集計ロジットモデルを用いる。この理由としては,これまでの人の移動行動生成モデルにより,エージェントの属性,ODなどが利用できることにある。他の代表的な選択肢として,集計値をそのまま利用する集計ロジットモデルの適用も考えられるが,集計値ベースのモデリングの場合,生態学的相関(ecological correlation)などの問題が起こることが指摘されている²⁴⁾。他の集計的なモデリングとして,地質学分野で発達してきた組成データに関する統計解析手法を社会経済データへ適用したものが²⁵⁾。その中では,土地利用の割合について推定を行っており,交通手段分担率に関しても同様のデータ(組成データ)として扱えることが示唆されている。しかしながら,土地利用とは違い,交通手段選択の場合,そこに含まれる意思決定者の人数がエリアによって異なるため,組成データとして扱うよりも人数を直接扱うモデルを適用すべきである²⁾。

非集計ロジットモデルにも様々なモデルが存在するが,その中でも最も基本的な多項ロジットモデル(Multinomial Logit Model)を用いた場合の尤度関数 \mathcal{L} は以下の式で表される。

$$\mathcal{L}(\{y_{i,m}\}_{i \in \mathcal{I}, m \in \mathcal{M}}) = \prod_{i \in \mathcal{I}} \prod_{m \in \mathcal{M}} P_{i,m}^{y_{i,m}} \quad (3)$$

$$P_{i,m} = \frac{\exp(-C_{i,m})}{\sum_{m \in \mathcal{M}} \exp(-C_{i,m})} \quad (4)$$

$$C_{i,m} = \beta_m^T \mathbf{x}_{i,m} \quad (5)$$

ここで, \mathcal{I} は個人 i の集合, \mathcal{M} は交通手段 m の集合を表す。 $\mathbf{x}_{i,m}$ は個人 i の交通手段 m に対する説明変数で, $y_{i,m}$ は個人 i の交通手段 m の選択結果の2値変数(選

表-1: 交通調査の種類 (須永²³⁾ を参考に作成)

調査種類	パーソントリップ調査	大都市交通センサス	国勢調査
対象地域	主に主要都市圏	三大都市圏	全国
対象交通機関	全交通手段	鉄道・バスのみ	全交通手段
移動目的	目的別	全目的一括	通勤・通学のみ
空間解像度	ゾーン (数 km ²)	駅/バス停	500m メッシュ
入手性	難	難	易
備考	提供先・使用目的が制限	利用目的の申請, 承認の必要有	—

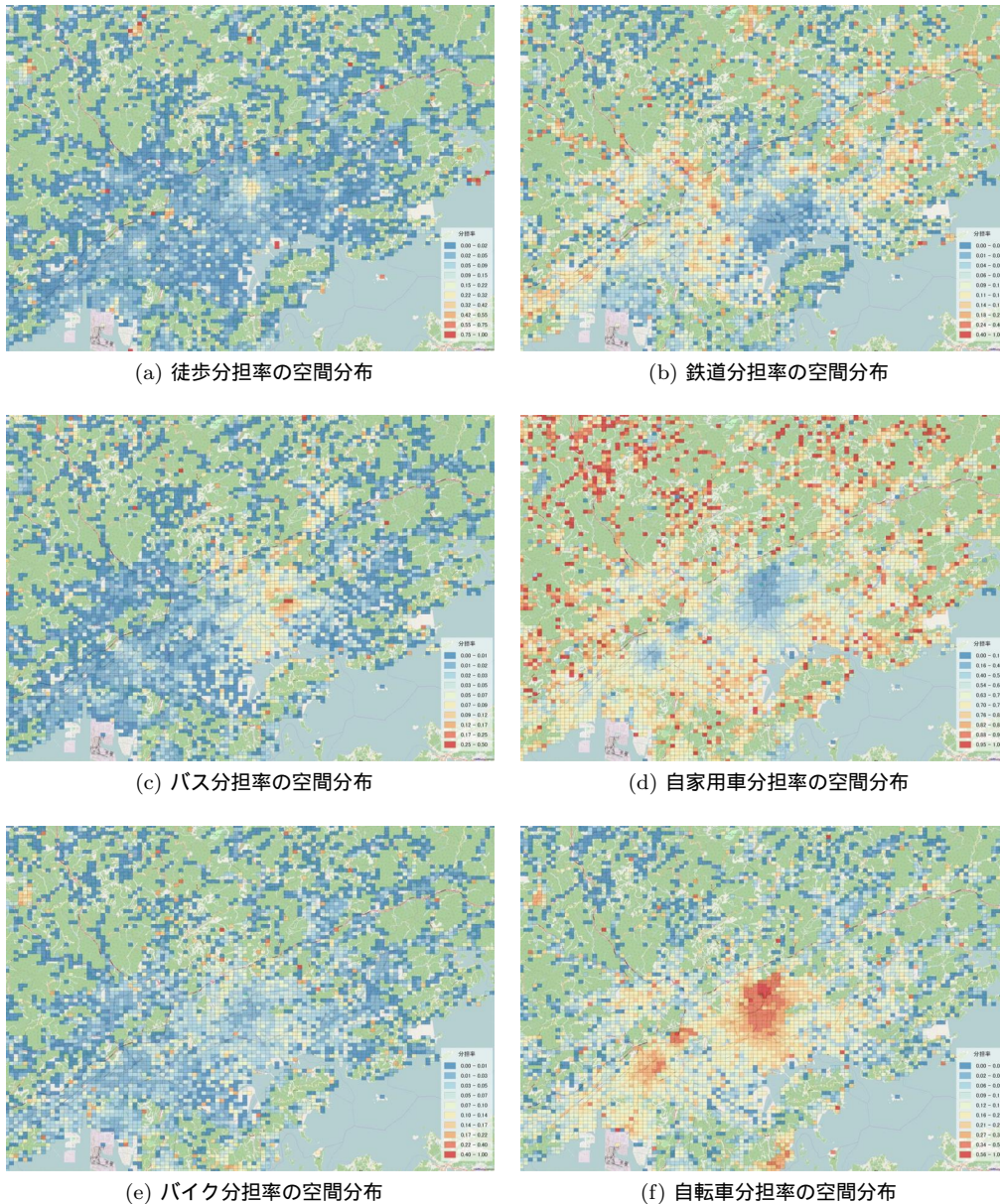


図-2: 国勢調査から計算した岡山市周辺部の各交通手段の分担率

択していれば 1, そうでなければ 0), $C_{i,m}$ は個人 i の交通手段 m のコストである. また, β_m は交通手段 m に関するパラメータである.

これに対して今我々が求めたい問題では個人の選択

結果, すなわち $y_{i,m}$ が与えられない. 代わりに与えられるのはメッシュごとの選択人数である. 個別の選択結果を集計データから推定することはデータ拡大法¹⁹⁾, すなわち, 潜在的な選択結果を観測された集計データ

と矛盾なくシミュレートすることにより求めることができる。今回の問題に対する一つの課題は、観測された集計和をみたく個々の選択の組み合わせを考える必要により生じる計算の困難性である。Chen and Yang¹⁵⁾ は M 個の代表的なクラスを考えることで問題を簡素化しているが、クラス内の個人が同質であると仮定していることや、そもそものクラス数の決定などに恣意性が残る。Choi et al.¹⁶⁾ は与えられた集計値のもと個人の選択の全ての可能性を考えるアプローチを採用しているが、いうまでもなく大規模な問題になると計算量的な困難性の問題が生じる。Park¹⁷⁾ は Gail et al.²⁶⁾ によって提案された再帰計算による方法を 2 項選択の問題に適用、Park and Jeong¹⁸⁾ はこれを多項選択の問題に拡張している。

いま、エリア a における交通手段 m の選択人数を N_m^a とする。また、このエリアにおける総通勤・通学者数 N^a は各交通手段の選択人数の和、すなわち $\sum_{m \in \mathcal{M}} N_m^a$ で与えられる。したがって、このエリアにおける交通手段 m の分担率 S_m^a は N_m^a/N^a と計算される。いま、エリア内の N^a 人の個人の選択確率が仮に $\{P_{i,m}^a\}_{i \in \mathcal{I}, m \in \mathcal{M}}$ と与えられた場合に、エリアの交通手段選択人数 N_m^a が得られる尤度は各個人の選択を独立な事象と考えれば、以下のポアソン二項分布で与えられる。

$$\Pr(X = N_m^a) = \sum_{A \in F_{N_m^a}} \prod_{i \in A} P_{i,m}^a \prod_{j \in A^c} (1 - P_{j,m}^a) \quad (6)$$

ここで、 $F_{N_m^a}$ は $\{1, 2, \dots, N^a\}$ の中から選ぶことのできる全ての N_m^a 要素の部分集合族であり、また A^c は集合 A の補集合である。例えば、 $N^a = 3$ で $N_m^a = 2$ の場合、 $F_{N_m^a} = \{\{1, 2\}, \{1, 3\}, \{2, 3\}\}$ となる。したがって、この例における尤度は

$$\begin{aligned} \Pr(X = 2) &= P_{1,m}^a P_{2,m}^a (1 - P_{3,m}^a) \\ &+ P_{1,m}^a P_{3,m}^a (1 - P_{2,m}^a) + P_{2,m}^a P_{3,m}^a (1 - P_{1,m}^a) \end{aligned} \quad (7)$$

と計算することができる。ポアソン二項分布の直接的な計算は計算困難性を伴うが、Park¹⁷⁾ にも用いられている再帰計算^{27),28)} や離散フーリエ変換を用いた方法など効率的な計算方法が既に提案されている。さらに、計算効率性を重視し、一般的な確率分布をもとにした近似による計算も提案されている²⁹⁾。本研究においても、計算の効率性を重視し、一般的な確率分布による近似方法を採用する。以下では、平均と分散を用いて分布間の比較を試みる。

ポアソン二項分布の平均 μ_{PB} と分散 σ_{PB}^2 は以下の式で与えられる。

$$\mu_{PB} = \sum_i P_i \quad (8)$$

$$\sigma_{PB}^2 = \sum_i P_i (1 - P_i) \quad (9)$$

ただし、ここでは個人 i の選択確率を単純に P_i と置いた。

まず、二項分布を用いた近似を考えてみる。いま、 $P = \sum_i P_i/N$ と置けば、二項分布 $\text{Bin}(N, P)$ の平均 μ_{Bin} と分散 σ_{Bin}^2 は以下の式で与えられる。

$$\mu_{\text{Bin}} = NP = N \cdot \frac{1}{N} \sum_i P_i = \sum_i P_i = \mu_{PB} \quad (10)$$

$$\sigma_{\text{Bin}}^2 = NP(1 - P) \quad (11)$$

式 (10) から分かるように平均値は一致する。一方、分散の大小を評価するために $P_i = P + \alpha_i$ と置く。ここで、 α_i は P_i の平均値からのずれであり、 $\sum_i \alpha_i = 0$ を満たす。このとき、ポアソン二項分布の分散 σ_{PB}^2 は

$$\sigma_{PB}^2 = \sum_i P_i (1 - P_i) \quad (12)$$

$$= \sum_i (P + \alpha_i) \{1 - (P + \alpha_i)\} \quad (13)$$

$$= \sum_i \{P(1 - P) - \alpha_i P + \alpha_i(1 - P) - \alpha_i^2\} \quad (14)$$

$$= NP(1 - P) - (1 - 2P) \sum_i \alpha_i - \sum_i \alpha_i^2 \quad (15)$$

$$= \sigma_{\text{Bin}}^2 - \sum_i \alpha_i^2 \quad (16)$$

となることが分かる。

また、ポアソン二項分布を平均 $\mu_{\text{Pois}} = NP$ のポアソン分布で近似した場合は、平均 = 分散の関係より分散 $\sigma_{\text{Pois}}^2 = NP$ が得られる。

以上のことから、これらの分散の関係は $\sigma_{PB}^2 \leq \sigma_{\text{Bin}}^2 \leq \sigma_{\text{Pois}}^2$ となる。したがって、二項分布の近似はポアソン分布に比べると、近似の精度が良く、また全ての P が等しいときのみ両者は一致する。図-3 に 3 つの分布の比較の例を示した。ポアソン二項分布、二項分布、ポアソン分布の順に分散が小さい様子が分かる。

以下では、この二項分布を用いた近似によって尤度関数を計算することとする。このときの尤度関数は、

$$\begin{aligned} \mathcal{L}(\{N_m^a\}_{a \in \mathcal{A}, m \in \mathcal{M}}) &= \\ &\prod_{a \in \mathcal{A}} \prod_{m \in \mathcal{M}} N^a C_{N_m^a}(\mu_m^a) \mu_m^a^{N_m^a} (1 - \mu_m^a)^{N^a - N_m^a} \end{aligned} \quad (17)$$

$$\mu_m^a = \frac{1}{N^a} \sum_{i=1}^{N^a} P_{i,m}^a \quad (18)$$

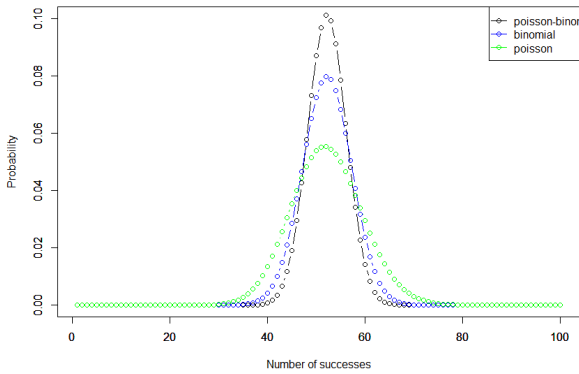


図-3: ポアソン二項分布, 二項分布, ポアソン分布の比較

$$P_{i,m}^a = \frac{\exp(-C_{i,m}^a)}{\sum_{m \in \mathcal{M}} \exp(-C_{i,m}^a)} \quad (19)$$

$$C_{i,m}^a = \beta_m^\top \mathbf{x}_{i,m}^a \quad (20)$$

となる。

実は式 (17) の最尤値を求めることも $P_{i,m}^a$ の和の計算を含むことから困難である。しかしながら、尤度関数に一般的な確率分布である二項分布を用いていることから、標準的な MCMC (Markov Chain Monte Carlo) ソフトウェアを用いて容易に推定ができる。MCMC を用いた推定の方法については次章で詳しく述べることにする。パラメータ β_m が決定すれば、式 (19) の多項ロジットモデルを用いて交通手段選択を行うことができる。

本節の最後に、式 (19) は非集計の確率的な意思決定モデルであれば、多項ロジットモデルである必要はないということを強調しておく。多項ロジットモデルは IIA 特性を持つなど様々な指摘がされており、式 (19) のモデルは問題によって分析者が上手く設定すべきである。

(2) トリップチェーン行動を利用した移動手段付与モデル

前節では、国勢調査の通勤・通学利用交通手段データを用いてエージェントに通勤・通学交通手段を付与する手法、具体的には多項ロジットモデルのパラメータを決定するための尤度関数の設定方法について議論した。本節では、この通勤・通学利用交通手段をベースに残りの目的に対する移動交通手段を付与する方法を説明する。

通勤・通学利用交通手段から他の目的の移動交通手段

を与えることの妥当性は以下の 2 つの理由による。まず、平日の通勤・通学を含むトリップチェーン行動を考えると、帰宅行動や立ち寄り行動は交通手段の観点からいけば従属的な扱いである。すなわち、通勤・通学利用交通手段に依存して他の交通手段が決まると考えられる。分かりやすい例でいえば、通勤・通学に車を利用する人は、帰りに立ち寄る買物も車で移動し、帰宅も車で移動する可能性が極めて高い。2 つ目の理由として、通勤・通学利用交通手段はそのエリアの特性を表しているということである。すなわち、駅が遠くて鉄道の利用がしづらい場所は通勤・通学目的であろうが、買物目的であろうが不便であることに変わりはないということである。

上記の理由から通勤・通学交通手段選択モデル (多項ロジットモデル) とトリップチェーン行動を基にした交通手段付与ルールを組み合わせることによって他の移動目的に対する交通手段の付与を行う。今回設定した交通手段付与ルールは以下の通りである。

Rule1 所有物 (車/バイク/自転車) は一緒に移動しなければならない

Rule2 一度家に帰れば、新たに選択が可能 (交通手段がリセットされる)

Rule3 トリップチェーンにおける主は通勤・通学 (それ以前の交通手段も影響を受ける)

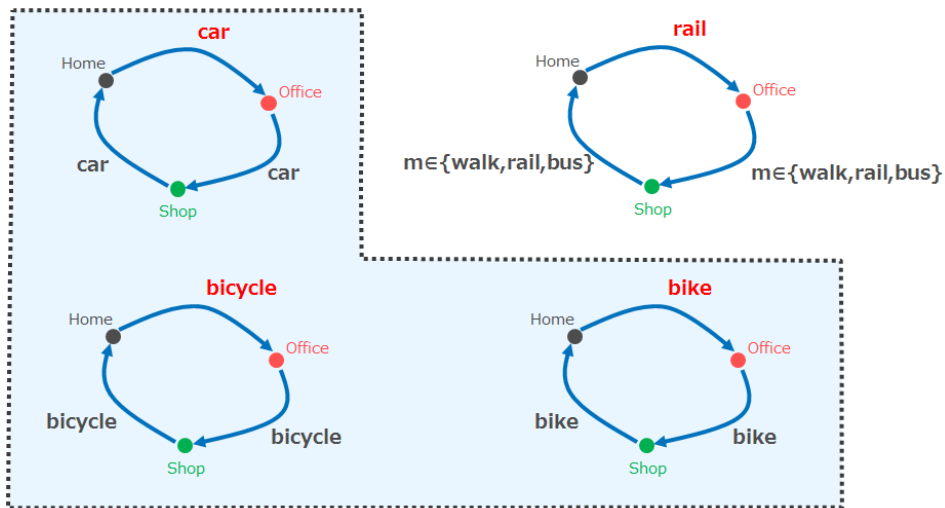
図-4 にそれぞれのルールの適用例を示す。図-4a の例では、所有物 (車/バイク/自転車) が選択された場合はその後の移動についてもその交通手段によって制約を受ける例を示している (点線枠内)。また、右上の例は鉄道で通勤した例で、その後の移動は徒歩もしくは公共交通の中から選ぶという制約を受ける例である。図-4b の例では、通勤は車でやっているが、一旦家に帰ったためにその後の買物のための移動に関してはどの手段を選んでも構わないことを示している。図-4c の例では、通勤前に買物に行く例で、この場合は買物の交通手段が主ではなく、あくまでも通勤の交通手段が主であり、買物が従属していることを示している。

5. 岡山市における分析事例

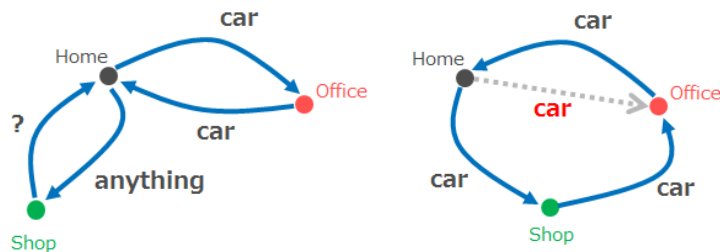
本章では、これまで説明してきた提案手法を岡山市周辺部を例に適用する。

(1) 対象地域

日高・志賀¹⁴⁾によって適用された岡山市中心部を含むエリアのうち、中心の 20km 四方の 4 次メッシュ領域を居住地とする人たちを対象とする。このエリアに含まれるメッシュは 1,600 メッシュであるが、このうち国勢調査データにおいて通勤・通学利用者データを含



(a) Rule1 の適用例 (点線枠内：所有物に制約を受ける例，右上：所有物以外での制約)



(b) Rule2 の適用例

(c) Rule3 の適用例

図-4: ルールの適用例

んでいるメッシュは 1,189 メッシュとなった。また，日中の移動先の候補は日高・志賀¹⁴⁾と同様，周辺の 3 次メッシュを含む合計 2,400 メッシュを対象とする。

(2) 説明変数

本節では，推定に用いる説明変数 $x_{i,m}^a$ について説明する。 $x_{i,m}^a$ は，個人のインデックス i とエリアのインデックス a を含むことから分かるように個人に依存する項とエリアに依存する項から成り立っている。個人に依存する項はこれまでの人の移動行動生成モデルによって生成された個人の属性を説明変数とするものであり，また，エリアに依存する項はその他の GIS データより得られる説明変数で，そのエリアの特徴を表すものである。

a) 移動距離

人の移動行動生成モデルによって得られる目的地系列データから，居住地メッシュと通勤先（通学先）のメッシュを抽出し，2 地点間のメッシュ中心の直線距離を求める。通勤先のメッシュと通学先のメッシュが両方が存在する場合には，学業が主だと仮定（職場はアルバイト先に相当すると仮定）し，通学先のメッシュへの距離を用いる。本来であれば，道路ネットワークを用いて

実際の道路距離を求めるべきであるが，ここでは簡単のため直線距離にルートファクターを乗じることで道路距離の代替指標とする。ルートファクターは道路距離/直線距離の比を表すもので，一般にはトリップ距離の短い徒歩や自転車などに対して道路ネットワークが省略されるときに道路距離が過大に推計されることを避けるために導入されるものであり，通常 1.1~1.3 の値が設定される²⁴⁾。また，近年の日本国内における実証研究では平均的には 1.3 程度になることが報告されているが³⁰⁾，今回は 1.2 と設定した。図-5 にメッシュごとのエージェントの直線距離の平均と分担率の関係をバブルチャートで示した。バブルのサイズはメッシュの通勤者数を示している。図から分かるように徒歩，車，自転車などは移動距離と強い相関を持つことが読み取れる。各エージェントの OD が疑似的に生成されたデータであるにも関わらず，このような相関関係が表れるのは通勤・通学の OD パターンが現実と整合するように生成されているためであると考えられる。

b) 10 代ダミー

人の移動行動生成モデルによって生成される年齢属性のうち，10 代であるか否かのダミー変数を設定する。これは，10 代の車やバイクの通学を抑制するための変

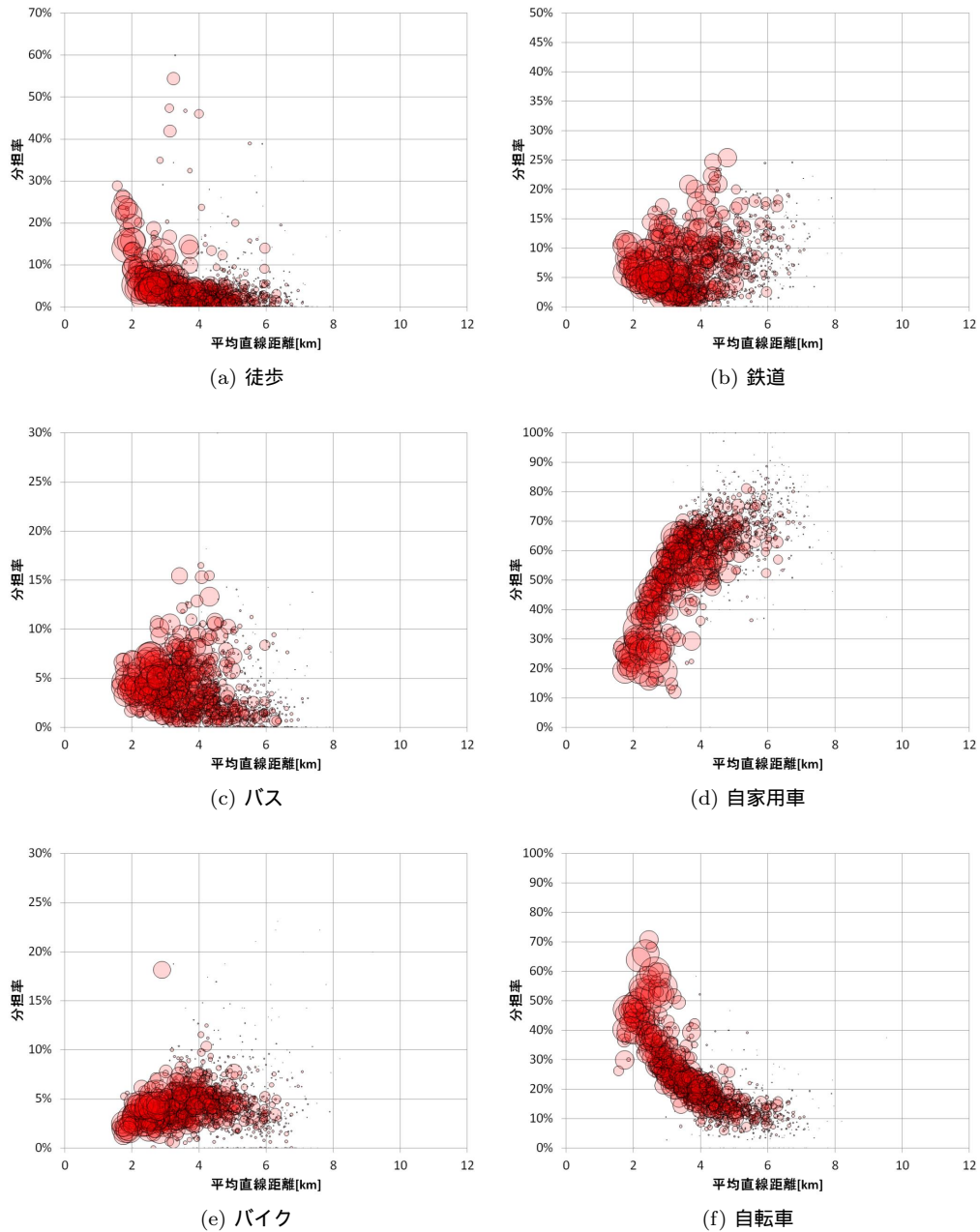


図-5: 平均直線距離と分担率

数である。

c) アクセス/イグレス距離

鉄道やバスは図-5b や図-5c に見て取れるように、移動距離との相関はあまり見られない。これはこれらの交通手段の利用しやすさ、すなわち駅やバス停が近くに存在するか否かに大きく依存するためだと考えられる。したがって、駅やバス停までのアクセス距離およびイグレス距離を変数に設定する。駅やバス停の位置情報（緯度経度情報）に関しては、国土交通省の国土数値情報ダウンロードサービスより入手が可能である^{31),32)}。アクセス/イグレス距離は、各メッシュの中心座標と最寄りの駅およびバス停までの直線距離にルートファクターを

乗じたものとして設定する。図-6 にメッシュ中心、駅、バス停の位置情報を図示した。また、図-7, 8 にそれぞれ最寄りの駅、バス停までの直線距離と分担率の関係を示す。図から駅やバス停までのアクセスが悪くなると分担率が下がる傾向が見て取れる。

d) バス運行頻度

運行頻度も公共交通手段の Level of Service(LOS) を決定する大きな要因の一つである。バスの運行頻度も、国土数値情報ダウンロードサービスより入手が可能である³³⁾。ここでは最寄りのバス停を経路内に含むバス路線の中での運行頻度の最大値を使用する。

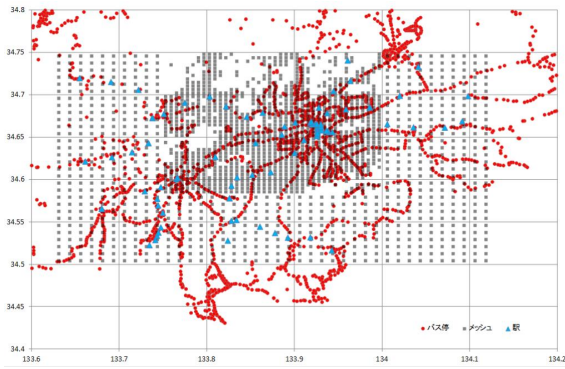


図-6: メッシュ中心, 駅, バス停の位置情報

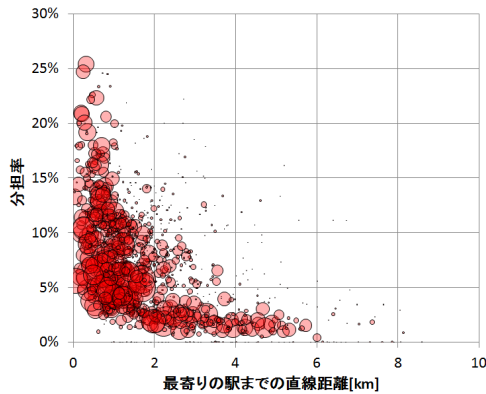


図-7: 最寄りの駅までの直線距離と分担率

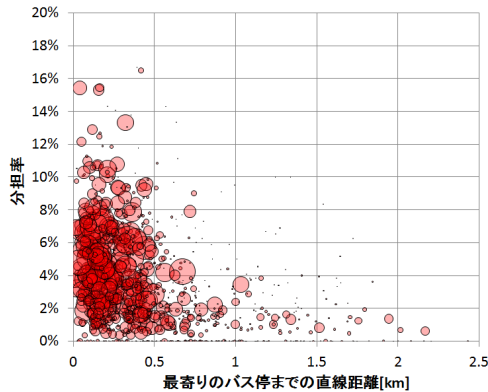


図-8: 最寄りのバス停までの直線距離と分担率

e) 昼間人口

都市中心エリアは、一般に地代（駐車場代）が高く、それゆえに車の所有率や、都市部への通勤に対する交通手段選択率に影響があると考えられる。こうした地代の高いエリアは、居住人口が多いエリアというよりも昼間に人の集まるエリアと考えることができることから人の移動行動生成モデルによって生成される昼間

人口、具体的には 12 時台のメッシュ人口を変数として導入する。

(3) コスト関数

これまで述べてきた説明変数を用いて、各交通手段のコストは以下のように与えられる。

$$C_{i,walk}^a = 0 \quad (21)$$

$$C_{i,rail}^a = \beta_{0,rail} + \beta_{1,rail} \cdot r \cdot x_{i,dist}^a + \beta_2 \cdot r (x_{i,acsb}^a + x_{i,egrbs}^a) \quad (22)$$

$$C_{i,bus}^a = \beta_{0,bus} + \beta_{1,bus} \cdot r \cdot x_{i,dist}^a + \beta_2 \cdot r (x_{i,acsb}^a + x_{i,egrbs}^a) + \beta_3 \cdot x_{freq}^a \quad (23)$$

$$C_{i,car}^a = \beta_{0,car} + \beta_{1,car} \cdot r \cdot x_{i,dist}^a + \beta_4 \cdot x_{dummy10}^a + \beta_5 \cdot x_{dens}^a \quad (24)$$

$$C_{i,bike}^a = \beta_{0,bike} + \beta_{1,bike} \cdot r \cdot x_{i,dist}^a + \beta_4 \cdot x_{dummy10}^a \quad (25)$$

$$C_{i,cycle}^a = \beta_{0,cycle} + \beta_{1,cycle} \cdot r \cdot x_{i,dist}^a \quad (26)$$

ただし、ここで $x_{i,dist}^a$ は通勤・通学直線距離、 $x_{i,acsb}^a$ は駅までの直線アクセス距離、 $x_{i,egrbs}^a$ は駅までの直線イグレス距離、 $x_{i,acsb}^a$ はバス停までの直線アクセス距離、 $x_{i,egrbs}^a$ はバス停までの直線イグレス距離、 x_{freq}^a はバス運行頻度、 $x_{dummy10}^a$ は 10 代ダミー、 x_{dens}^a は昼間人口（千人）、 r はルートファクターを表す。また、 $\beta_0 \sim \beta_5$ は推定するパラメータである。

ここでは、徒歩のコストを基準として設定しているため、徒歩に関するパラメータを推定する必要はない。

(4) パラメータ推定

a) 事前処理

パラメータ推定を行う事前の処理として、メッシュごとの通勤・通学者の合計値を観測側（国勢調査）とモデル側（移動行動推定）で合わせる必要がある。これは人の移動行動生成モデルが生活時間調査データを基に確率的に移動行動を生成しているために起こる問題である。図-9 に比較の結果を示した。全体的に推定値の方が少し過大になっているが、相関は極めて高く通勤・通学者をうまく生成できている様子が分かる。図の赤丸で示された外れ値は刑務所を含むメッシュであり、推定が難しい例外的なメッシュである。

こうした通勤・通学者数のずれに起因する推定精度の低下を防ぐために、パラメータの推定時には国勢調査の通勤・通学者数に合うよう疑似データ側からサンプリングを行った。具体的には、国勢調査の人数の方が多い場合には復元抽出、また、国勢調査の人数の方が少ない場合には非復元抽出によって同数のサンプルを確保した。

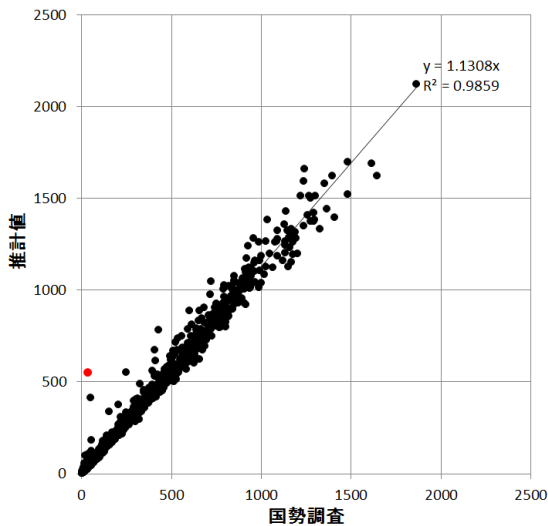


図-9: メッシュ別通勤・通学者数の比較

b) MCMC ソフトウェア

(1) 節で議論したように、式 (17)-(20) の最尤値を求めることは困難である。しかしながら、ベイズ推定、特に MCMC と呼ばれるサンプリング手法を用いることによりパラメータ推定が可能である。今回は、フリーで利用できる MCMC ソフトウェアの一つである Stan³⁴⁾ を利用し、パラメータ推定を行った。

Stan の特徴は、ハミルトニアンモンテカルロ法 (HMC 法) の実装法の一つである NUTS³⁾ を利用している点で、これによりステップ間の相関が低いサンプリングが可能になるため、少ないステップ数で収束し、計算時間が短くなる点にある³⁵⁾。Stan は比較的新しい言語ではあるものの、マニュアルも充実しており、近年では日本語で読める書籍^{35),36)} も登場したことで開発しやすい環境になってきている。

c) 推定結果

対象エリア 1,189 メッシュ全てを用いて推定する場合、約 40 万人と大量のデータとなるため計算時間が莫大になる。そこで、対象エリアの中からランダムに 300 メッシュを抽出し、推定を行った。また、10 代は車とバイクでは通勤・通学をしないとの仮定をおき、10 代ダミーに関しては正の大きな値を与えることで推定を行わないこととした。

バーンイン期間は 100、サンプリング期間は 500 と設定し、8chain を独立に実行し、全部で 4,000 サンプルを得た。バーンイン期間、サンプリング期間ともにメトロポリスヘイスティング法やギブスサンプリング法を用いる場合に比べて非常に少ないが、図-10, 11 を見てわかるようにどのパラメータも 100 回前後のサンプリングで収束しており、良好なサンプリングができ

表-2: パラメータの推定結果

	変数名	mean	C.I. 95%
定数項	$\beta_{0,rail}^*$	3.91	[3.63, 4.25]
	$\beta_{0,bus}^*$	4.22	[3.94, 4.52]
	$\beta_{0,car}^*$	1.20	[0.91, 1.51]
	$\beta_{0,bike}^*$	4.75	[4.46, 5.10]
	$\beta_{0,cycle}^*$	0.731	[0.495, 1.015]
移動距離	$\beta_{1,rail}^*$	-2.33	[-2.58, -2.11]
	$\beta_{1,bus}^*$	-2.03	[-2.27, -1.81]
	$\beta_{1,car}^*$	-2.14	[-2.38, -1.92]
	$\beta_{1,bike}^*$	-2.18	[-2.42, -1.96]
	$\beta_{1,cycle}^*$	-1.51	[-1.74, -1.31]
アクセス/イグレス	β_2^*	0.570	[0.538, 0.602]
バス運行頻度	β_3^*	-0.0192	[-0.0214, -0.0168]
昼間人口	β_5^*	0.208	[0.197, 0.219]

* : 信用区間 (C.I.) に 0 を含んでいないという意味での有意性

ている様子が分かる。推定されたパラメータを表-2 に示す。全てのパラメータに関して有意であるとの結果が得られた。

推定したパラメータを用いて、メッシュ内の各交通手段選択人数の期待値を求める。メッシュ内の交通手段選択人数の期待値は式 (8) の定義より、メッシュ内のエージェントの選択確率の和で求められる。国勢調査と推定値の比較を図-12 に示す。どの交通手段に対しても $R^2 > 0.7$ の精度で推計できていることが分かる。特に車や自転車など分担率の高い交通手段の推計精度が高いことも分かる。

また、図-13 には各エージェントに対する選択確率を多項ロジットモデルによって求めたものをプロットした。非常に興味深い点として、パーソントリップ調査のように個人の結果を陽に与えずに、集計情報のみから推定しているにも関わらず、徒歩の移動距離が 2km 以内であることや、自転車の移動距離のピークが 2-3km であることなど直観に合う結果が得られている。

(5) 空間相関の考慮

これまで全てのメッシュは独立であるとの仮定のもと推定を行ってきたが、「距離が近いメッシュは似たような分担率の傾向を持つ」というような空間的な相関を利用することで推定精度の向上を試みる。その前にまず国勢調査から算出される分担率と今回推定した結果から算出される分担率の差が空間的な相関を持つのかどうか検証を行う。

a) 空間的自己相関の検証

空間的自己相関の検証として、空間統計学の分野で最も用いられる指標の一つである Moran's I 統計量による検証を行う。Moran's I 統計量は以下の式で定義される。

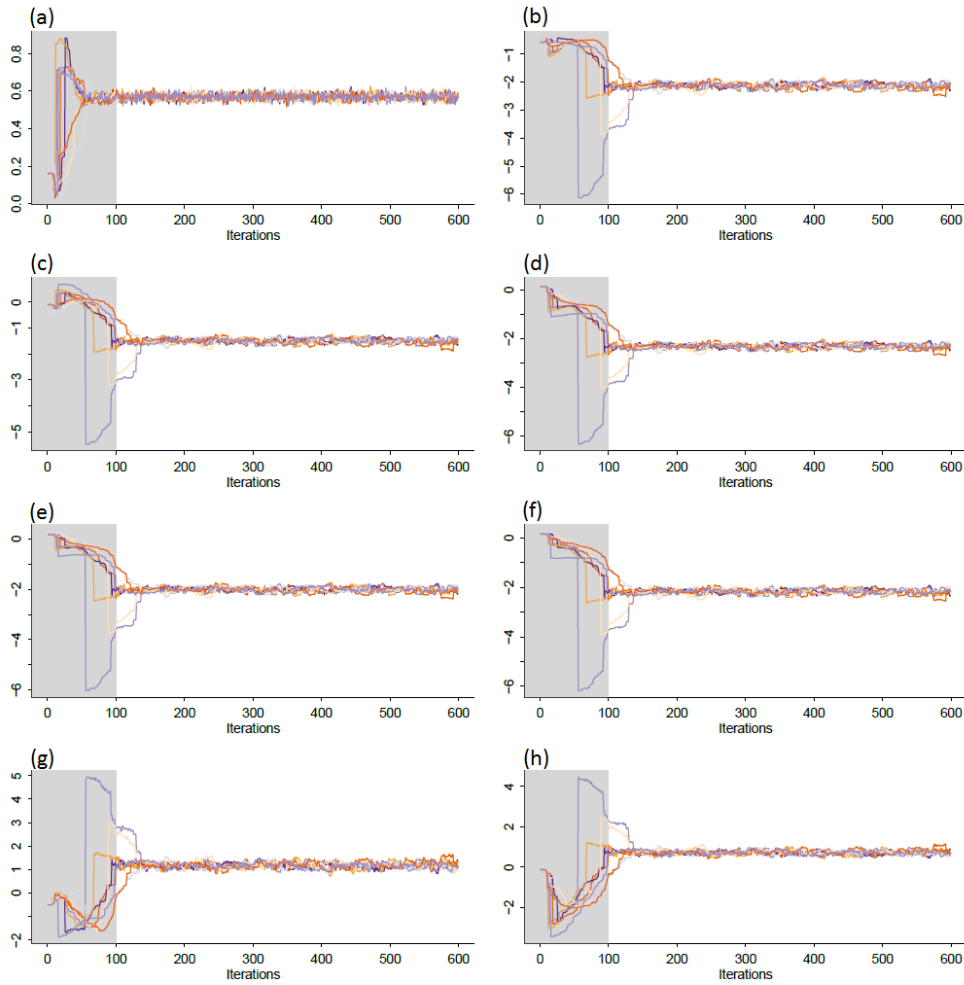


図-10: パラメータのトレースプロット

(a) β_2 (b) $\beta_{0,car}$ (c) $\beta_{0,cycle}$ (d) $\beta_{0,rail}$ (e) $\beta_{0,bus}$ (f) $\beta_{0,bike}$ (g) $\beta_{1,car}$ (h) $\beta_{1,cycle}$

$$I = \frac{n}{\sum_{i,j} w_{i,j}} \frac{\sum_{i,j} w_{i,j} (y_i - \bar{y})(y_j - \bar{y})}{\sum_i (y_i - \bar{y})^2} \quad (27)$$

ここで、 n はサンプル数、 y_i は観測値、 w_{ij} は空間重み行列 W の ij 要素である。空間重み行列はエリア間の関係性を記述するための行列であり、代表的なものとしては隣接行列や距離の逆数を用いたものがある。Moran's I 統計量が 0 より大きい場合は正の自己相関、0 より小さい場合には負の自己相関があることが示唆される。

国勢調査と推計値の分担率の差を求め、Moran's I 統計量を計算した。ただし、空間重み行列は距離の 2 乗の逆数、すなわち $w_{ij} = d_{ij}^{-2}$ と設定した。計算結果を表-3 に示す。どの交通手段においても正の空間自己相関が出ており、バイク以外は 1% 有意であることが分かった。図-14 にそれぞれの交通手段の発生地点集計での分担率の差の空間分布を示す。確かに Moran's I 統計量の大きい鉄道などは過少に推計されている地域が西側に固まっているなど空間的な相関が見て取れる。

表-3: 分担率の差の Moran's I 統計量

交通手段	Moran's I	p-value
徒歩	0.0589	0
鉄道	0.193	0
バス	0.145	0
車	0.102	0
バイク	0.00937	0.0694
自転車	0.110	0

b) 空間相関を考慮したモデリング

多項離散選択モデルの空間相関を考慮したモデルへの拡張の既存研究としては、例えば、ランダム効果の事前分布に空間構造を取り入れる条件付き自己回帰 (Conditional Auto Regressive; CAR) モデルがある。CAR モデルは主に生態学分野で用いられている手法であったが^{37),38)}、近年ではエリアごとの電気自動車の保有台数の予測³⁹⁾ や土地利用割合の予測²⁵⁾ など土木計画

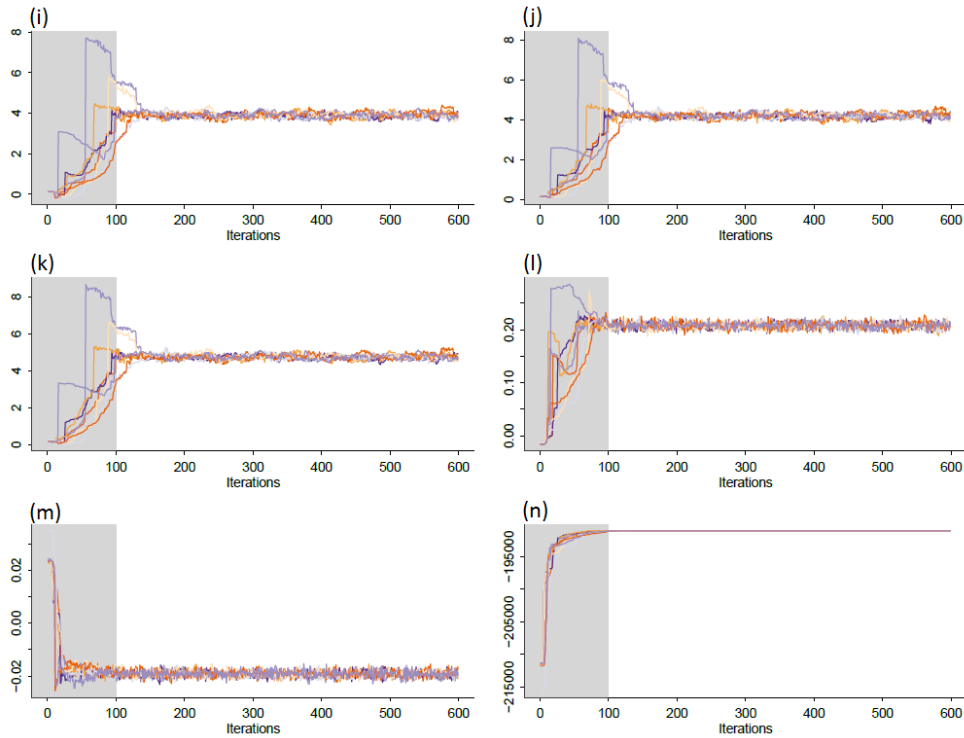


図-11: パラメータのトレースプロットの続き

(i) $\beta_{1,rail}$ (j) $\beta_{1,bus}$ (k) $\beta_{1,bike}$ (l) β_5 (m) β_3 (n) 対数尤度

分野での適用事例も増えてきている。CAR モデルは空間構造のある階層ベイズモデルであり、その推定法としては前述の MCMC 法が用いられることから我々の考えているモデルとの相性がいい。しかしながら、空間構造をランダム効果として導入するため、我々の行っているように一部のメッシュのみをサンプリングして推定するような場合には向かない。

空間相関を考慮するほかのアプローチとして Griffith らによって提案された固有ベクトル空間フィルタリング (Eigenvector-based Spatial Filtering; ESF) がある⁴⁰⁾。これは空間重み行列の固有ベクトルの一部を説明変数として導入する非常にシンプルなモデルであり、非常に実用性が高いことが特徴である。ESF の多項選択モデルへの適用例は現在のところ吉田らのもの⁴¹⁾ などまだ少ないが、そこでは ESF により土地利用の的中率が向上することが示されている。

空間重み行列から固有ベクトルを計算するとサンプル(エリア)の数だけ固有ベクトルが算出される。全ての固有ベクトルを説明変数に追加すると過適合 (overfitting) を引き起こす恐れがあるため、Moran's I 統計量の大きい固有ベクトルを候補にステップワイズ法によって選択する方法⁴⁰⁾ や lasso 推定⁴²⁾ に基づく選択方法⁴³⁾ などが提案されている。

lasso 推定は、ベイズ的な観点からみれば事前分布に

Laplace 分布を設定したときの MAP 推定量と考えられる (ベイズアン lasso)。MCMC 法を用いた推定を行っていることからベイズアン lasso を用いたパラメータ推定が考えられるが、候補となる固有ベクトルの数が多いため計算時間が莫大になる。

そこで、以下の手順によって固有ベクトルを選択し、推定を行う。

1. 空間相関を考慮しない推定を行う (前節)
2. 推定値から求めた分担率と国勢調査データから求めた分担率との差を求める (分担率の誤差分布)
3. lasso 推定を用いて分担率の誤差分布を回帰し、固有ベクトルを選択
4. 選択された固有ベクトルを説明変数に加えて再推定

手順 3 では、一般的な lasso 推定そのものを行うので計算時間は問題にはならない。

lasso 推定は 2 乗和誤差関数に L_1 ノルムの制約項を課した関数

$$E(\beta) = (\mathbf{y} - X\beta)^\top (\mathbf{y} - X\beta) + \lambda \sum_j |\beta_j| \quad (28)$$

を最小化するパラメータ推定法である。ただし、 \mathbf{y} は目的変数ベクトル、 β はパラメータベクトル、 X は計

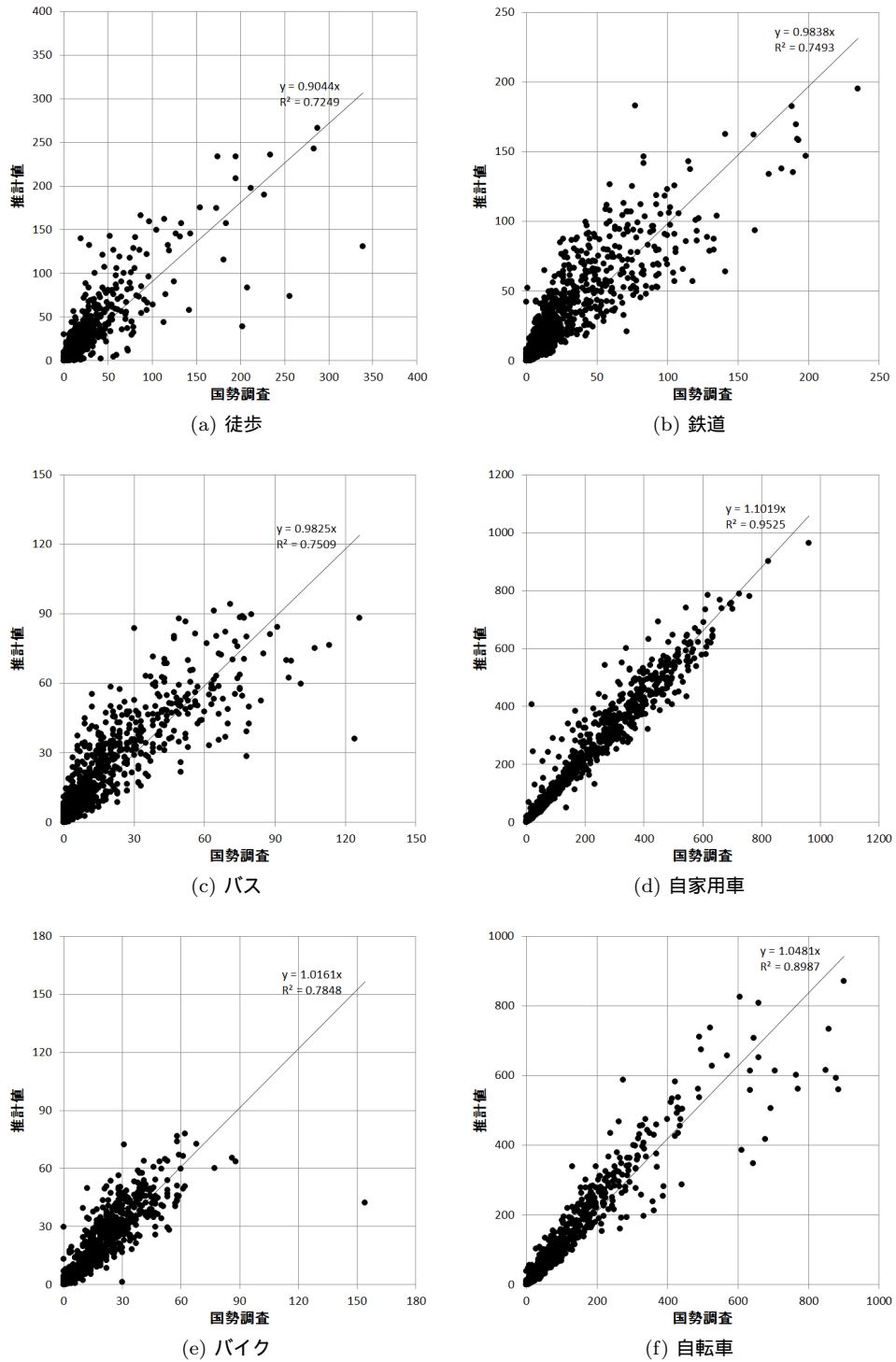


図-12: メッシュ内選択人数の推定結果

画行列である．また， λ は正則化パラメータと呼ばれ，制約項の強さを表し， λ が小さければあてはまりは良いがパラメータ数の多い複雑なモデルになり， λ が大きければあてはまりは良くないがパラメータ数の少ないシンプルなモデルになる． λ の決定は一般に交差検証法 (Cross Validation) によって行われる．

c) 推定結果

推定は統計解析向けのプログラミング言語である R を用いて行った．lasso の推定には R のパッケージの一つである "glmnet"⁴⁴⁾ を利用した．正則化パラメータの決定には，10 分割の交差検証法を用いた．パラメータの決定の基準には一般に平均二乗誤差 (Mean Squared Error; MSE) が用いられる．また，MSE を基準にして

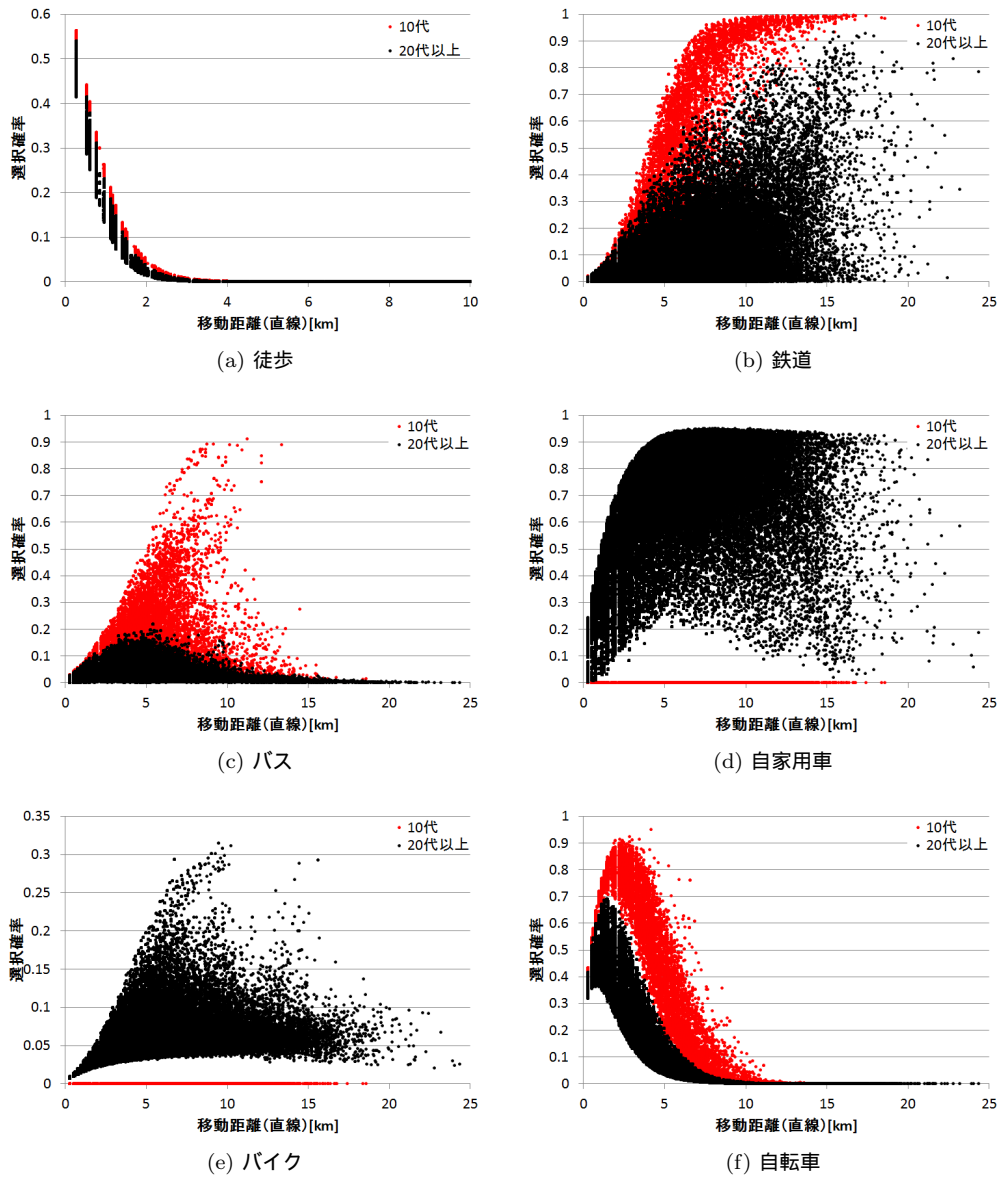


図-13: 個人の手段選択確率の推定結果

もなお標準誤差分の不定性があることから、最小 MSE の標準誤差内の最も大きい λ を選ぶことによって最もシンプルなモデルを選択する基準がある (one-standard-error-rule; 1SE) ⁴⁵⁾。それぞれの基準で決定した λ に基づいて選択した固有ベクトルの数を表-4 にまとめた。また、図-15 に各交通手段における $\ln \lambda$ と MSE の関係を示した。図中横軸上の数字は各 $\ln \lambda$ に対応して選ばれる固有ベクトルの数であり、図中左側の点線は MSE 基準によって決定された $\ln \lambda$ を、右側の点線が 1SE 基準によって決定された $\ln \lambda$ を示している。MSE 基準だと交通手段によっては 100 以上の固有ベクトルが選ばれているのに対して、1SE 基準の場合、鉄道のみ 18 個の固有ベクトルが選ばれる結果になった。今回は、計算量の制約とできるだけシンプルなモデルにするとの

考えから 1SE 基準に基づき、鉄道に関してのみ固有ベクトルを加えることとした。

上記で選ばれた 18 個の固有ベクトルを説明変数に加え、パラメータの再推定を行った。パラメータの推定結果を表-5 に示す。固有ベクトル以外のパラメータについては全てのパラメータで有意、また固有ベクトルは 18 パラメータ中の 11 パラメータで有意であるとの結果が得られた。推定したパラメータから計算されるメッシュ内の交通手段選択人数の決定係数の変化を表-6 に、また、Moran's I 統計量の変化を表-7 に示す。鉄道に ESF を適用した結果、鉄道の推定精度が大幅に向上し、空間相関も小さくなっていることが分かる。また、他の交通手段に関してはほとんど影響を受けていないことが分かった。図-16 に ESF 適用後の分担率の差の空

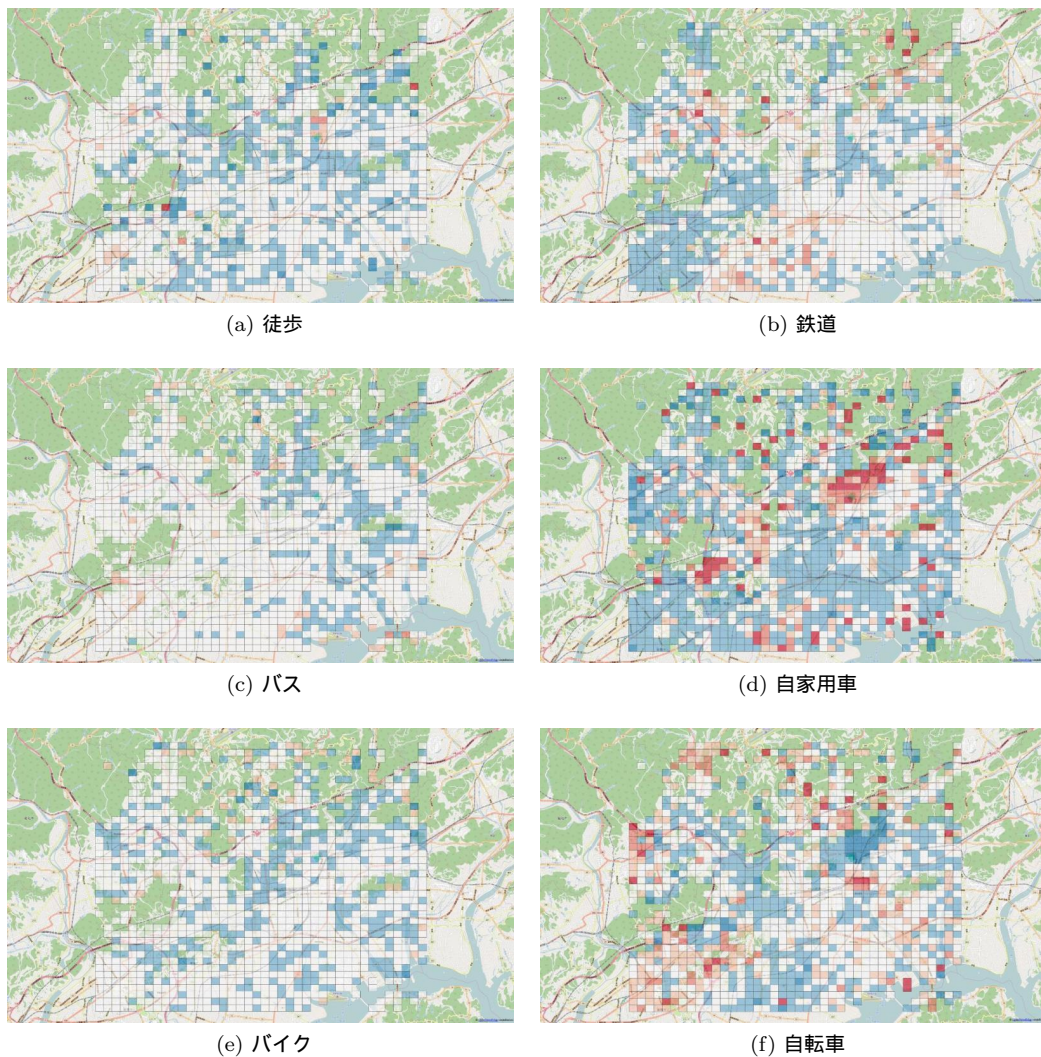


図-14: 分担率の差の空間分布

表-4: 各基準に基づいて選択された固有ベクトルの数

交通手段	MSE 基準	1SE 基準
徒歩	22	0
鉄道	77	18
バス	143	0
車	105	0
バイク	0	0
自転車	54	0

間分布を示す。図-14b と比較すると西側の分担率が過少に推計されていた地域が減り、空間相関が小さくなっている様子が分かる。

(6) 交通手段推定を応用した分析例

本章の最後に今回提案した交通手段推定を応用した分析の例を示す。

1つ目の例は、岡山市内の主要4病院への通院目的で

の交通手段の分析である。図-17に主要4病院への通院手段の分析例を示す。図の交通手段分担率は、各病院を含むメッシュへ通院目的で移動した人を交通手段別に集計することで計算できる。結果、済生会病院や川崎病院のように駅に近い病院は鉄道分担率が高い代わりに車の分担率が低くなっている。また、駅から少し離れた赤十字病院ではバスと車の分担率が高くなっている様子が分かる。

2つ目の例は、岡山市内の主要エリアにおける買物目的での交通手段の分析である。図-18に岡山市内14か所への買物交通手段の分担率を示す。分析に用いた14か所の地図は図-19に示す通りである。図-18, 19からは、鉄道路線へのアクセスのしやすさが分担率に表れている様子が分かる。また、特に興味深い点として津島エリアで徒歩の割合が非常に高いことである。このエリアの特徴としては、岡山大学が近く、学生の居住者が極めて多いこと、公共交通への利便性がよくな

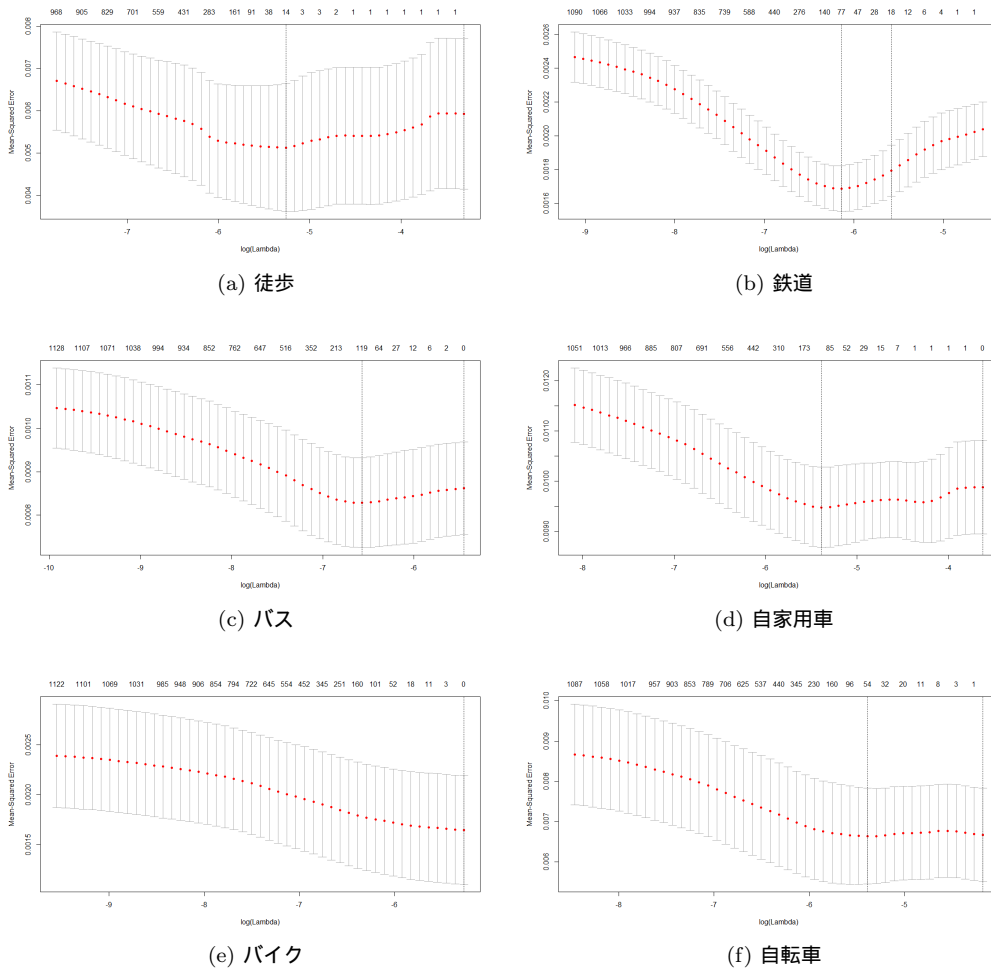


図-15: $\ln \lambda$ と MSE

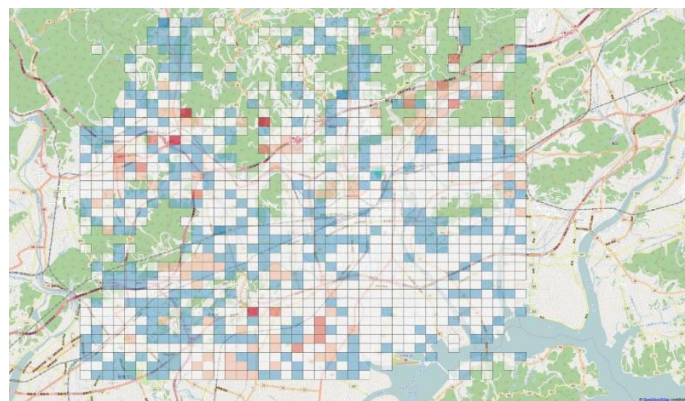


図-16: ESF 適用後の分担率の差の空間分布 (鉄道)

い点にある．こうしたことから徒歩の割合が高くなっていると考えられる．

このように本手法を適用することで、国勢調査では分からなかった集中地点での集計や、目的や属性などのクロスでの集計など様々な切り口で分析が可能と

なる．

6. まとめ

本稿では、人の移動行動推定技術の一つとして交通手段を付与する方法について提案した．入手性が高く、

表-5: パラメータの推定結果

	変数名	mean	C.I. 95%
定数項	$\beta_{0,rail}^*$	3.67	[3.23, 4.17]
	$\beta_{0,bus}^*$	4.23	[3.87, 4.62]
	$\beta_{0,car}^*$	1.48	[1.10, 1.89]
	$\beta_{0,bike}^*$	5.03	[4.66, 5.44]
	$\beta_{0,cycle}^*$	0.928	[0.613, 1.29]
移動距離	$\beta_{1,rail}^*$	-2.47	[-2.82, -2.18]
	$\beta_{1,bus}^*$	-2.15	[-2.50, -1.85]
	$\beta_{1,car}^*$	-2.31	[-2.67, -2.02]
	$\beta_{1,bike}^*$	-2.36	[-2.71, -2.06]
	$\beta_{1,cycle}^*$	-1.64	[-1.99, -1.36]
アクセス/イグレス	β_2^*	0.686	[0.647, 0.725]
バス運行頻度	β_3^*	-0.0162	[-0.0184, -0.0139]
昼間人口	β_5^*	0.211	[0.200, 0.222]
第 1 固有ベクトル	β_{EV1}^*	6.94	[4.49, 9.29]
第 2 固有ベクトル	β_{EV2}^*	-1.19	[-3.13, 0.724]
第 6 固有ベクトル	β_{EV6}^*	0.566	[-0.522, 1.65]
第 7 固有ベクトル	β_{EV7}^*	-4.74	[-5.98, -3.52]
第 8 固有ベクトル	β_{EV8}^*	-3.42	[-4.35, -2.54]
第 10 固有ベクトル	β_{EV10}^*	3.00	[1.90, 4.15]
第 12 固有ベクトル	β_{EV12}^*	5.35	[4.03, 6.67]
第 18 固有ベクトル	β_{EV18}^*	4.53	[3.08, 5.92]
第 31 固有ベクトル	β_{EV31}^*	-3.49	[-4.37, -2.60]
第 49 固有ベクトル	β_{EV49}^*	0.462	[-0.998, 1.91]
第 50 固有ベクトル	β_{EV50}^*	0.758	[-0.881, 2.48]
第 51 固有ベクトル	β_{EV51}^*	-0.0452	[-1.60, 1.49]
第 55 固有ベクトル	β_{EV55}^*	-0.548	[-1.50, 0.394]
第 56 固有ベクトル	β_{EV56}^*	2.58	[1.69, 3.49]
第 78 固有ベクトル	β_{EV78}^*	1.38	[0.608, 2.16]
第 232 固有ベクトル	β_{EV232}^*	3.66	[-10.6, 19.8]
第 251 固有ベクトル	β_{EV251}^*	2.84	[0.445, 5.86]
第 431 固有ベクトル	β_{EV431}^*	-2.17	[-3.94, 0.683]

* : 信用区間 (C.I.) に 0 を含んでいないという意味での有意性

表-6: ESF の導入による決定係数の変化

交通手段	ESF 無	ESF 有
徒歩	0.725	0.727
鉄道	0.749	0.862
バス	0.751	0.773
車	0.953	0.951
バイク	0.785	0.789
自転車	0.899	0.902

表-7: ESF の導入による Moran's I 統計量の変化

交通手段	ESF 無	ESF 有
徒歩	0.0589	0.0589
鉄道	0.193	0.107
バス	0.145	0.141
車	0.102	0.108
バイク	0.00937	0.00877
自転車	0.110	0.108

全国全てを網羅する国勢調査データを入力に利用することで、日本全国どこでも交通手段の推定ができる。また、非集計の確率的な意思決定モデルを含む一般的な枠組みとして交通手段推定を定式化したことにより、集計データの活用の可能性を拡げることができた。さらに、エージェントのトリップチェーン行動に着目し、行動の制約から通勤・通学以外の行動についても交通手段を付与する手法についても提案した。今回提案した手法を適用することで、元のデータでは分からなかった集中地点ベースでの分担率の集計や、目的や属性な

どとのクロスでの集計など様々な切り口で人の移動行動を分析できるようになった。

提案モデルの一つの課題として、目的地選択と交通手段選択において属性を同等にしか扱えないことがある。その結果、高齢者の交通行動に関しても通勤・通学の交通手段モデルを基に与えられてしまっている。これに対しては、利用するデータに依存するため、現状の枠組みの中では改善が難しい。また、今後の展望として、発着地の直線距離をネットワークを考慮した距離に置き換えることで精度の向上が見込める。さらに、

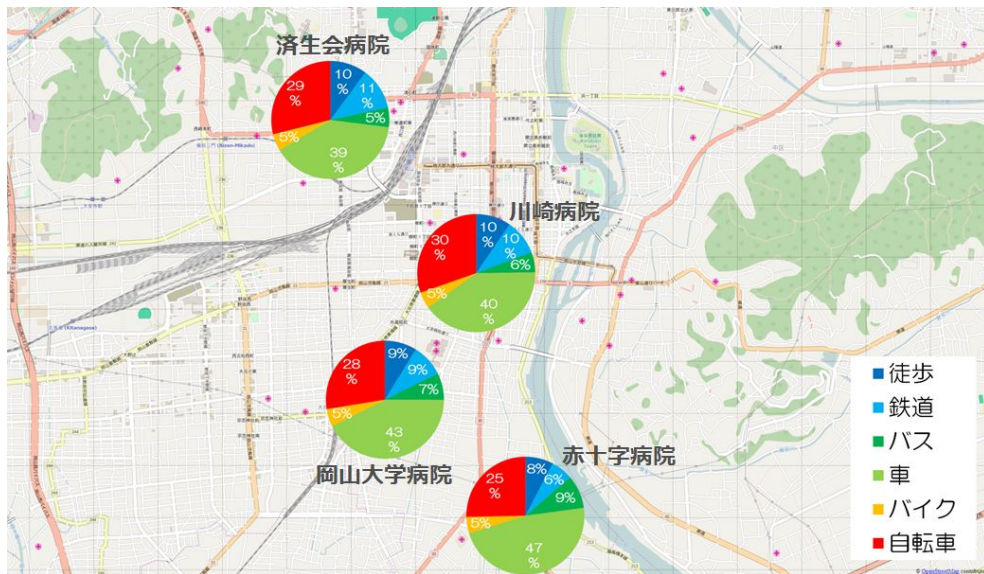


図-17: 岡山市内主要 4 病院への通院手段分析

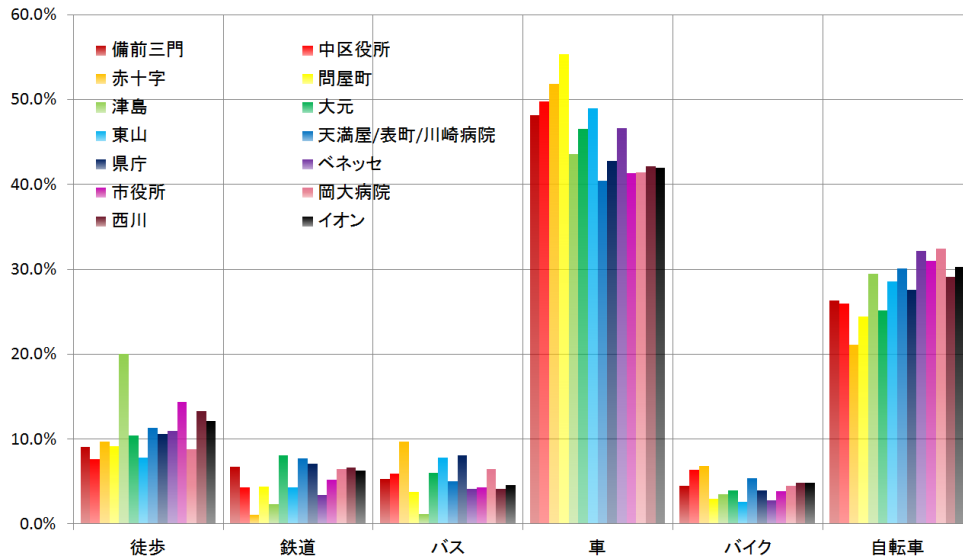


図-18: 岡山市内 14 か所への買物交通手段分析

本稿では代表交通手段のみを考えていたが、多項ロジットモデルを他のモデルに置き換えることで複数の交通手段を考慮したモデルにも取り組んでいく。

注釈

¹ 国勢調査データでは発生地点ごとの通勤・通学利用交通手段の選択人数が与えられるため、各交通手段の選択人数を全交通手段の選択人数で除して分担率を計算した

² 例えば、「1 人中 1 人が車を選択したエリア」と「100 人中 100 人が車を選択したエリア」を同等に扱うべきで

はないということ

³No-U-Turn Sampler の略。それまでの HMC 法では更新回数やステップサイズを分析者が設定する必要があったが、NUTS を利用することでこれらが適応的に決まる。詳しくは豊田³⁶⁾を参照のこと。

参考文献

- 1) 藤井聡, 大塚祐一郎, 北村隆一, 門間俊幸: 時間的空間的制約を考慮した生活行動軌跡を再現するための行動シミュレーションの構築, 土木計画学研究・講演集, Vol. 14, pp. 643-652, 1997.
- 2) Ben-Akiva, M. and Bowman, J.: Activity based disaggregate travel demand model system with daily activity schedules, *Transportation Research Part A*, Vol.35,



図-19: 買物交通手段で分析した 14 か所の地図

- No.1, pp.1-28, 2000.
- 3) Bhat, C. R., Goulias, K. G., Pendyala, R. M., Paleti, R., Sidharthan, R., Schmitt, L., and Hu, H. H.: A household-level activity pattern generation model with an application for Southern California, *Transportation*, Vol.40, No.5, pp.1063-1086, 2013.
 - 4) Miller, E. J. and Roorda, M. J.: A Prototype Model of Household Activity/Travel Scheduling, *Proceedings of the 82nd Annual Meeting of the Transportation Research Board*, Washington D.C., 2003.
 - 5) Arentze, T. A. and Timmermans, H. J.: A learning-based transportation oriented simulation system, *Transportation Research Part B: Methodological*, Vol.38, pp.613-633, 2004.
 - 6) Recker, W. W.: The household activity pattern problem: General formulation and solution, *Transportation Research Part B*, Vol.29, No.1, pp.61-77, 1995.
 - 7) Furletti, B., Cintia, P., Renso, C., and Spinsanti, L.: Inferring human activities from GPS tracks, *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*, Chicago, 2013.
 - 8) Hayano, R. S. and Adachi, R.: Estimation of the total population moving into and out of the 20 km evacuation zone during the Fukushima NPP accident as calculated using "Auto-GPS" mobile phone data., *Proceedings of the Japan Academy. Series B, Physical and biological sciences*, Vol.89, No.5, pp.196-199, 2013.
 - 9) González, M. C., Hidalgo, C. a., and Barabási, A.-L.: Understanding individual human mobility patterns., *Nature*, Vol.453, No.June, pp.779-782, 2008.
 - 10) Kang, C., Gao, S., Lin, X., Xiao, Y., Yuan, Y., Liu, Y., and Ma, X.: Analyzing and geo-visualizing individual human mobility patterns using mobile call records, *Proceedings of the 18th International Conference on Geoinformatics*, Beijing, 2010.
 - 11) Hasan, S., Lafayette, W., and Ukkusuri, S. V.: Understanding Urban Human Activity and Mobility Patterns Using Large-scale Location-based Data from Online Social Media, *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*, Chicago, 2013.
 - 12) Wu, L., Zhi, Y., Sui, Z., and Liu, Y.: Intra-urban human mobility and activity transition: Evidence from social media check-in data, *PLoS ONE*, Vol.9, No.5, pp.1-13, 2014.
 - 13) 日高健, 大野宏司, 志賀孝広: 集計データの統合による都市内の移動行動データ生成, *土木学会論文集 D3*, Vol.72, No.4, pp.324-343, 2016.
 - 14) 日高健, 志賀孝広: 複数の統計データの融合による都市内の人の移動行動推計と時点更新, *土木計画学研究・講演集*, Vol. 55, 松山, 2017.
 - 15) Chen, Y. and Yang, S.: Estimating disaggregate models using aggregate data through augmentation of individual choice, *Journal of Marketing Research*,

- Vol.44, No.4, pp.613–621, 2007.
- 16) Choi, T., Schervish, M. J., Schmitt, K. A., and Small, M. J.: A bayesian approach to logistic regression model with incomplete information, *Biometrics*, Vol.64, No.3, pp.424–430, 2008.
 - 17) Park, T.: Bayesian analysis of individual choice behavior with aggregate data, *Journal of Computational and Graphical Statistics*, Vol.20, No.1, pp.158–173, 2011.
 - 18) Park, T. and Jeong, S.: Efficient bayesian analysis of multivariate aggregate choices, *Journal of Statistical Computation and Simulation*, Vol.85, No.16, pp.3352–3366, 2015.
 - 19) Tanner, M. A. and Wong, W. H.: The calculation of posterior distributions by data augmentation, *Journal of the American Statistical Association*, Vol.82, No.398, pp.528–550, 1987.
 - 20) 総務省統計局: 平成 23 年社会生活基本調査, <http://www.stat.go.jp/data/shakai/2011/index.htm>, 2012, 最終閲覧 2017.04.
 - 21) NHK 放送文化研究所: 2015 年 国民生活時間調査, http://www.nhk.or.jp/bunken/research/yoron/20160217_1.html, 2016, 最終閲覧 2017.04.
 - 22) Wilson, A. G.: A statistical theory of spatial distribution models, *Transportation Research*, Vol.1, No.3, pp.253–269, 1967.
 - 23) 須永孝隆: 都市交通手段分担影響要因の分析とライトレールトランジット導入時の課題-京都市をモデルとして-, Technical Report T03001, (財)電力中央研究所, 2003.
 - 24) 土木学会土木計画学研究委員会 (Editor): 非集計行動モデルの理論と実際, (社)土木学会, 1995.
 - 25) 吉田崇紘, 堤盛人: 空間効果を考慮した組成データ解析, 土木計画学研究・講演集 Vol.49, 2014.
 - 26) Gail, M. H., Lubin, J. H., and Rubinstein, L. V.: Likelihood calculations for matched case-control studies and survival studies with tied death times, *Biometrika*, Vol.68, No.3, pp.703–707, 1981.
 - 27) Chen, X. H., Dempster, A. P., and Liu, J. S.: Weighted finite population sampling to maximize entropy, *Biometrika*, Vol.81, No.3, pp.457–469, 1994.
 - 28) Chen, S. X. and Liu, J. S.: Statistical applications of the poisson-binomial and conditional bernoulli distributions, *Journal of the American Statistical Association*, Vol.7, pp.875–892, 1997.
 - 29) Hong, Y.: On computing the distribution function for the poisson binomial distribution, *Journal of Computational Statistics & Data Analysis*, Vol.59, pp.41–51, 2013.
 - 30) 森田匡俊, 鈴木克哉, 奥貫圭一: 日本の主要都市における直線距離と道路距離との比に関する実証的研究, *GIS-理論と応用*, Vol.22, No.1, pp.1–7, 2014.
 - 31) 国土交通省国土政策局国土情報課: 国土数値情報 バス停留所データ, <http://nlftp.mlit.go.jp/ksj/gml/datalist/KsjTmplt-P11.html>, 2015.
 - 32) 国土交通省国土政策局国土情報課: 国土数値情報 鉄道データ, http://nlftp.mlit.go.jp/ksj/gml/datalist/KsjTmplt-N02-v2_2.html, 2015.
 - 33) 国土交通省国土政策局国土情報課: 国土数値情報 バスルート, <http://nlftp.mlit.go.jp/ksj/gml/datalist/KsjTmplt-N07.html>, 2015.
 - 34) Stan Development Team: Stan home page, <http://mc-stan.org/>.
 - 35) 松浦健太郎: Stan 入門 次世代のベイジアンモデリングツール, 岩波データサイエンス Vol.1 ベイズ推論と MCMC のフリーソフト, (Ed. by 岩波データサイエンス刊行委員会), pp. 63–79, 岩波書店, 2015.
 - 36) 豊田秀樹: 基礎からのベイズ統計学ハミルトニアンモンテカルロ法による実践の入門, 朝倉書店, 2015.
 - 37) 久保拓弥: データ解析のための統計モデリング-一般化線形モデル・階層ベイズモデル・MCMC, 岩波書店, 2012.
 - 38) 深澤圭太, 石濱史子, 小熊宏之, 武田知己, 田中信行, 竹中明夫: 条件付自己回帰モデルによる空間自己相関を考慮した生物の分布データ解析, *日本生態学会誌*, Vol.59, No.2, pp.171–186, 2009.
 - 39) Chen, T. D., Wang, Y., and Kockelman, K. M.: Where are the electric vehicles? a spatial model for vehicle-choice count data, *Journal of Transport Geography*, Vol.43, pp.181–188, 2015.
 - 40) Griffith, D. A.: *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding Through Theory and Scientific Visualization*, Springer, 2003.
 - 41) 吉田崇紘, 堤盛人: 多項離散選択モデルへの固有ベクトル空間フィルタリングの適用, 地理情報システム学会講演論文集, 2013.
 - 42) Tibshirani, R.: Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol.58, No.1, pp.267–288, 1996.
 - 43) Seya, H., Murakami, D., Tsutsumi, M., and Yamagata, Y.: Application of lasso to the eigenvector selection problem in eigenvector-based spatial filtering, *Geographical Analysis*, Vol.47, pp.284–299, 2015.
 - 44) Friedman, J., Hastie, T., Simon, N., and Tibshirani, R.: Package glmnet: Lasso and elastic-net regularized generalized linear models, <https://cran.r-project.org/web/packages/glmnet/index.html>.
 - 45) Hastie, T., Tibshirani, R., and Friedman, J.: *The Elements of Statistical Learning; Data Mining, Inference, and Prediction*, Springer, 2009.

(2017. 7. 28 受付)

DISAGGREGATE MODE CHOICE MODEL BY FUSION OF MULTIPLE AGGREGATE DATA

Ken HIDAKA and Takahiro SHIGA