

経路検索に現れる移動需要と 駅周辺環境の関連性分析

細江 美欧¹・桑野 将司²・森山 卓³

¹学生会員 鳥取大学 工学研究科社会基盤工学専攻 (〒680-8552鳥取県鳥取市湖山町南四丁目101)
E-mail: d19t4003b@edu.tottori-u.ac.jp

²正会員 鳥取大学教授 工学研究科社会基盤工学専攻 (〒680-8552鳥取県鳥取市湖山町南四丁目101)
E-mail: kuwano@tottori-u.ac.jp

³非会員 鳥取大学助教 工学研究科社会基盤工学専攻 (〒680-8552鳥取県鳥取市湖山町南四丁目101)
E-mail: moriyama@tottori-u.ac.jp

経路検索システムの検索履歴には、検索者の潜在的な移動需要が含まれると考えられている。しかし、検索者の属性や移動目的に関する情報は記録されていないことから、どのような人のどのような移動目的による移動需要であるかは定量的に明らかにされていない。本研究では検索履歴に現れる移動需要と駅周辺環境の関連性から検索者像を推察した。具体的には、香川県で運行される鉄道路線「ことでん」を分析対象として、移動需要と駅周辺環境の関連性を表現するモデルを構築した。複数の関数形と説明変数群を仮定したモデルの中から最適なモデルを選定し、得られた推定結果から検索者の人物像と移動目的を推測した。その結果、検索履歴に現れる移動需要は、住民や観光客による非定常的な移動需要と関連があることを明らかにした。

Key Words : *average number of searches, comparison of functional forms, adaptive lasso*

1. はじめに

公共交通利用者の交通行動を時間的・空間的制約なく把握するために、GPSによる位置情報データ、交通系ICカードの乗降履歴データ、高速道路におけるETCの利用履歴データなどの交通ビッグデータに注目が集まっている。これらの交通ビッグデータを活用することによって、従来の紙面調査では拾いきれなかった詳細な交通行動が捉えられると期待されている。特に、交通系ICカードの乗降履歴データには「カード利用者が、いつ、どこから、どこへ移動したのか」という移動軌跡が記録されていることから、交通行動の実態を把握する上で注目されている。しかしながら、地方都市では交通系ICカードシステムの空白地域が未だ多いのが現状であり、行動データの蓄積が十分ではない。

そこで近年、公共交通利用者の交通行動を把握するための新たなデータとして、経路検索システムに蓄積される検索履歴データが着目されている。経路検索システムとは、目的地や出発地、利用希望日や希望時間帯を入力すると、最適な経路や運賃、所要時間などが提示される

Webサービスである。経路検索システムには、「NAVITIME」や「Yahoo!乗換案内」など全国規模のものから、山陰両県を対象とした「バスネット」、山口県を対象とした「キララバスナビ」など地域限定で導入されているものが存在する。「検索者が、いつ、どこから、どこへ移動したいのか」という利用希望が記録されている検索履歴データを分析することによって、公共交通における潜在的な移動需要が推測できると期待されている。

太田ら¹⁾はNAVITIMEの検索履歴データから、終電を指定した終夜バスの検索、北陸新幹線開通後の金沢・富山着の検索、訪日外国人向けのシステムの検索、人気観光施設を目的地に設定した検索などをそれぞれ集計分析し、日別あるいは出発駅や到着駅毎の移動需要について考察している。見生ら²⁾は鳥取県に導入されているバスネットの検索履歴データから、検索者が、いつ、どこから、どこへ移動する予定であったのかを曜日や時間帯毎に集計し、利用者の移動について分析した。香川ら³⁾は見生らと同様にバスネットの検索履歴データを用いて、地方都市の交通行動の特徴を明らかにしている。具体的には検索履歴データの起終点情報を検索者の潜在的移動

需要と定義した上で、主成分分析を適用することによって、組合せの多い起終点情報をいくつかの特徴的な検索パターンとして集約した。さらに、算出された主成分得点を時系列データと見なし、各検索パターンの月変動および曜日変動を求めることで対象地域の移動需要の特性を解明した。

このように経路検索システムの検索履歴データ（以下、経路検索データ）に基づく移動需要の分析に関する研究の蓄積が進んでいる一方で、移動需要はどのような人のどのような目的によるものであったのかについては明らかにできていない。これは、経路検索データには検索者の個人属性や移動目的に関する情報が記録されていないためである。既往研究においても検索回数から検索者像を特徴付けることや⁴⁾、起終点周辺における特定の施設の有無などから移動目的を推測するに留まっている⁵⁾。

以上のように、経路検索データは移動需要を推測するための新たなデータとして着目されているものの、どのような特性を持つ人が経路検索システムを利用しているかについて不明な点が多いため、公共交通の運行計画や政策立案のための活用には課題が残る。

そこで、本研究では、経路検索データに現れる移動需要と公共交通の駅周辺環境との関係を定量的に明らかにすることで、検索者の人物像とその移動目的を推測することを目的とする。本研究では香川県で運行されている鉄道路線「ことでん」を分析対象とし、ことでんにおける移動需要を出発駅・到着駅別平均検索件数として定義する。駅周辺環境に関する情報は、オープンデータである人口分布データや施設立地データなどから取得する。そして、路線上の駅間で発生する移動需要と駅周辺環境との関係を表現するモデルを構築、推計することで得られたパラメータ、すなわち両者を関係付けるために重要となる変数の特徴から移動需要がどのような人のどのような移動目的であるかを推測する。

2. データ概要と基礎集計分析

(1) ことでんの概要

本研究では香川県で運行されている鉄道路線「ことでん」について分析を行う。ことでんは高松琴平電気鉄道株式会社によって運行されている地方私鉄である。ことでんには、図-1に示すように琴平線、長尾線、志度線の3路線全52駅が存在し、香川県高松市の中心部から放射状に路線が広がっている。また、瓦町駅は3路線の結節点となっており、3路線すべてに乗り換えが可能である。

政府が公開する統計データ⁶⁾によると、ことでんの駅の中で最も1日の平均乗降者数が多い駅は「瓦町駅」であり、その数は約1.4万人である。次いで、「高松築港

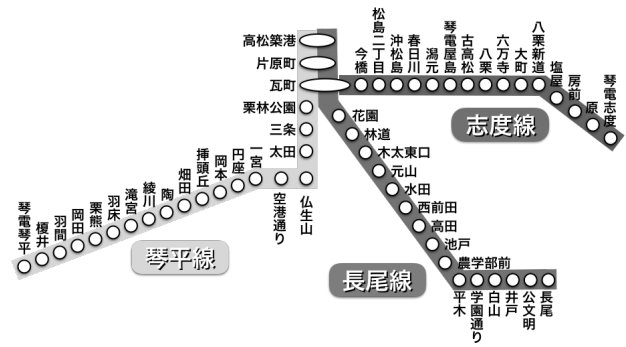


図-1 ことでん路線図⁷⁾

駅」（約1.2万人）、「片原町駅」（約0.5万人）の順に平均乗降者数が多い。高松築港駅から瓦町駅の区間は琴平線と長尾線の2路線が共用しており、ことでんの中で最も運行本数が多い区間である。この区間は高松市の中心市街地内に存在する。高松築港駅の近くにはJR高松駅や高松港が位置しており、他交通機関への乗換えが可能である。また、片原町駅周辺には百貨店や丸亀町商店街などが存在し、高松市の商業的中心地である。中心部から離れた駅の周辺には、住宅地や観光地が広がり、特に、琴平線上の駅周辺には主要な観光地が存在する。例えば、栗林公園駅近くには国の特別名勝に指定された回遊式日本庭園が、琴電琴平駅近くには「こんぴらさん」の呼び名で有名な金刀比羅宮がある。

(2) 経路検索システム「NAVITIME」の検索履歴データ

本研究では、経路検索システム「NAVITIME」の検索履歴データを用いる。NAVITIMEは、ナビタイムジャパン社が開発、提供するWEBサービスシステムで、鉄道、飛行機、新幹線、特急、有料道路、路線バス、高速バス、フェリー、徒歩などのさまざまな移動手段から、最適な移動経路を検索できる。このシステムは全国規模のサービスであり、月間利用者数は約5.1万人である⁸⁾。PC、携帯電話、スマートフォン、タブレット端末向けにサービスは展開されており、それぞれの端末からいつでもどこでも検索が行える。

本研究では、NAVITIMEにおいて出発駅および到着駅の双方でことでんの駅が指定された検索を分析対象とする。表-1にNAVITIMEの検索履歴データに記録される情報を示す。このうち、指定時刻、時刻指定方法、起終点名、起終点コードを用いる。データの収集期間は、2013年3月1日から2015年2月28日までの24ヶ月間で、全データ数は416,882件である。ここで、指定時刻を上記24ヶ月以外とした検索881件および同一起終点を対象とする検索145件は無効データとして除き、415,856件を有効データとした。

表-1 NAVITIME の検索履歴データに記録される情報

項目	備考
アクセス方法	アプリケーション, WEB
利用者ID	
検索実行時刻	
指定時刻	
時刻指定方法	出発時刻指定, 到着時刻指定, 始発指定, 終発指定
指定経由地数	
起終点	起終点名, 起終点コード, 起終点種別 (駅, 空港, 港, 連絡バス停, 路線バス停, 高速バス停), 緯度, 経度, 住所 (市区町村まで), 住所コード, メッシュコード

(3) 駅周辺環境データ

本研究では、ことでん駅周辺の人口および各施設数を駅周辺環境データとして収集する。駅周辺の範囲の定義としては、駅勢圏を用いる。駅勢圏とは、鉄道の駅を中心として、その駅を利用すると期待される範囲を表す。既往研究⁹⁾¹⁴⁾を参考に、半径500m圏内、半径800m圏内、半径1000m圏内を駅勢圏として定義し、各範囲別に人口および施設数を集計した。

人口分布データは e-Stat¹⁵⁾より収集した。e-Stat では、各府省が公表する統計データを検索、閲覧、ダウンロードできる。e-Stat において公表されている 2015 年度国勢調査の結果から、駅周辺の 14 歳以下人口、15 歳以上 64 歳以下人口、65 歳以上人口、外国人人口を収集した。

施設立地データは e-Stat、国土数値情報¹⁶⁾および経済産業省が提供する商業統計¹⁷⁾より収集した。国土数値情報では、土地利用に関する情報や施設の立地に関する情報などの国土に関する基礎的なデータが得られる。経済産業省の商業統計からは、平成 26 年商業統計結果に基づいて集計された卸売・小売業の事業所数に関するデータが取得可能である。e-Stat からは、事業所、国土数値情報から市町村役場、医療機関、福祉施設、文化施設、学校、都市公園、観光資源、集客施設、道の駅、都道府県指定文化財に関するデータを、商業統計からは総合スーパー、専門スーパー、コンビニ、ドラッグストア、その他のスーパー、専門店、家電大型専門店、中心店、その他の小売店、無店舗販売に関するデータを収集し、駅周辺の施設数をそれぞれ集計した。それぞれの施設概要については各収集元のウェブページを参照されたい。

(4) ことでんを対象とした検索状況

ことでんを対象とした検索状況を把握するために月別、曜日別、時間帯別に検索件数を算出した。

図-2に総検索件数の月別推移を示す。図-2より、2013年8月、2013年12月、2014年3月、2014年8月、2014年12月で検索件数が大きく増えているのがわかる。これは、長期休暇における帰省や旅行などに伴う検索が関係してい

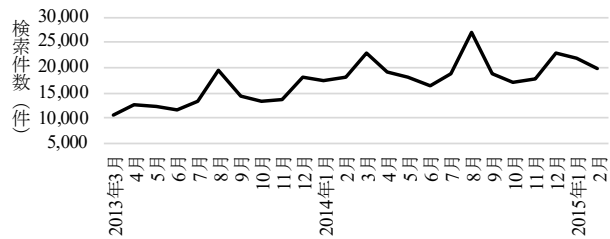


図-2 総検索件数の月別推移

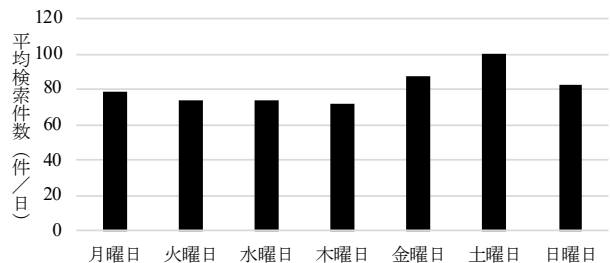


図-3 曜日別検索状況

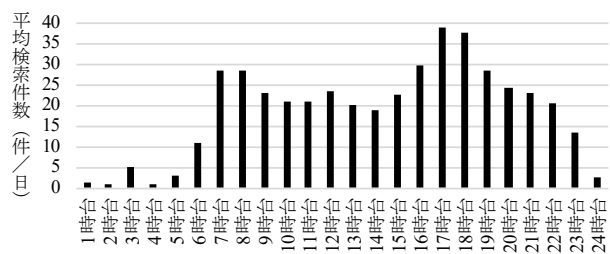


図-4 出発時刻指定の時間帯別検索状況

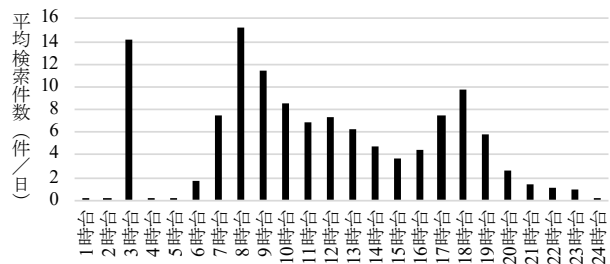


図-5 到着時刻指定の時間帯別検索状況

ると考えられる。

曜日別の平均検索件数を図-3に示す。図-3より、平日においては月曜日から木曜日までの平均検索件数に大差は見られないが、金曜日については他の曜日に比べ多いことがわかった。また、日曜日よりも土曜日の平均検索件数の方が多く、土曜日の平均検索件数が最も多いことが明らかとなった。平日の平均検索件数の平均は77件、休日の平均検索件数の平均は92件であり、休日の方がNAVITIMEは利用されやすいことがわかった。これは観光客や来訪客による検索や、日常的にことでんを利用しない者の検索が関係していると推察できる。

次に、時刻指定方法毎に時間帯別の検索状況を把握する。NAVITIMEの時刻指定方法には表-1に示した通り出発時刻指定、到着時刻指定、始発指定、終発指定の4つがある。総検索件数に占める各時刻指定方法の検索件数はそれぞれ約78%、約19%、約1%、約2%であり、出発

時刻指定による検索が最も多い。時刻指定方法毎の時間帯別平均検索件数の結果を図-4および図-5に示す。ただし、始発指定は指定日当日3時を出発時刻として、終発指定は指定日翌日3時を到着時刻としてシステム内で経路が検索されることから、始発指定は出発時刻指定、終発指定は到着時刻指定として集計した。

図-4に示す出発時刻指定された検索結果の時間帯分布より、1時台から4時台は運行時間外であることから検索件数が少ないことがわかる。一方、7時台から8時台および17時台から18時台にかけて平均検索件数が多い。これらの時間帯は朝夕のラッシュ時刻と一致しており、実際のことで利用者数と検索件数の時間帯分布が類似していることを示唆する。また、図-5より、到着時刻指定では3時台、8時台、18時台に多く検索が行われていることがわかった。8時台、18時台は出発時刻指定された検索と同様に朝夕のラッシュ時の検索が関係すると考えられる。一方、3時台の検索は終発での検索と考えられる。

3. 移動需要と駅周辺環境の関連性分析

(1) モデルの構築

本研究では、駅間で発生する移動需要と駅周辺環境との関係を記述するモデルを構築する。ことでんは計52駅を持つため、出発駅と到着駅の組合せは $52 \times 52 = 2,704$ から同一起終点を除いた2,652組合せ存在する。本研究ではNAVITIMEの検索履歴データ415,856件に基づいて、まず駅同士の組合せ毎に分析対象期間内の出発駅・到着駅別平均検索件数を算出する。そして、得られた平均検索件数を駅間の移動需要と定義する。これは移動需要がある駅間には平均的によく検索されることを想定している。一方、駅周辺の環境として出発駅および到着駅毎の駅勢圏内の人口と施設数を説明変数に利用する。本研究では駅勢圏の定義として各駅からの範囲として3つ（半径500m圏内、半径800m圏内、半径1000m圏内）を設けているが、どの範囲が最も適切であるかは事前に判断できない。そこで、本研究では各範囲内の人口や施設数を変数とするモデルを構築し、推定結果を比較することとする。

モデルの関数形には、線形回帰式、目的変数を対数変換する場合の片対数形回帰式、説明変数を対数変換する場合の片対数形回帰式、両対数形回帰式の4つを候補にする。すべての回帰式について記すことは紙面の都合上困難なため、例として線形回帰式を式(1)のように示す。

$$y_{od} = \sum_{i=1}^I \alpha_i x_{oi} + \sum_{j=1}^J \beta_j x_{dj} + \gamma d_{od} + c + \varepsilon_{od} \quad (1)$$

ここで、 y_{od} は出発駅 o から到着駅 d への平均検索件数、 α_i は出発駅 o 周辺の人口または施設数に関する i 番目のパ

表-2 各関数形と説明変数群の組合せ

関数形	説明変数群
線形	半径 500m 圏内の人口と施設数, 駅間距離
	半径 800m 圏内の人口と施設数, 駅間距離
	半径 1000m 圏内の人口と施設数, 駅間距離
	半径 500m, 800m, 1000m 圏内の人口と施設数, 駅間距離
目的変数を対数変換する片対数形	半径 500m 圏内の人口と施設数, 駅間距離
	半径 800m 圏内の人口と施設数, 駅間距離
	半径 1000m 圏内の人口と施設数, 駅間距離
	半径 500m, 800m, 1000m 圏内の人口と施設数, 駅間距離
説明変数を対数変換する片対数形	半径 500m 圏内の人口と施設数, 駅間距離
	半径 800m 圏内の人口と施設数, 駅間距離
	半径 1000m 圏内の人口と施設数, 駅間距離
	半径 500m, 800m, 1000m 圏内の人口と施設数, 駅間距離
両対数形	半径 500m 圏内の人口と施設数, 駅間距離
	半径 800m 圏内の人口と施設数, 駅間距離
	半径 1000m 圏内の人口と施設数, 駅間距離
	半径 500m, 800m, 1000m 圏内の人口と施設数, 駅間距離

ラメータ、 x_{oi} は出発駅 o 周辺の i 番目の人口または施設数、 β_j は到着駅 d 周辺の人口または施設数に関する j 番目のパラメータ、 x_{dj} は到着駅 d 周辺の j 番目の人口または施設数、 γ は出発駅と到着駅間の距離に関するパラメータ、 d_{od} は出発駅 o から到着駅 d 間の距離、 c は定数項、 ε_{od} は出発駅と到着駅の組毎に独立に正規分布 $N(0, \sigma^2)$ に従う誤差項である。

本研究では、目的変数と説明変数の関係が非線形になる場合も想定し、式(1)をもとに目的変数もしくは説明変数に対数をとる片対数形と、両方に対数をとる両対数形の回帰式も考える。両対数形回帰式は、経済学で生産性を計測する際に用いられるコブ・ダグラス型関数に相当する。また、この関数に駅間の距離を重みとして付与すると、2地点間の交通量推定に用いられる重力モデルとなる。

本研究では、4つの関数形と4つの説明変数群との組合せによって得られた複数の推定結果から、移動需要と駅周辺環境の関連性を記述するモデルについて検討する。表-2に本研究で検討するすべてのモデルに関する関数形と説明変数群の組合せをまとめている。

(2) Adaptive Lassoを用いた変数選択

与えられたデータに対してさまざまなモデルが考えら

れる場合には、包括的なモデル、すなわち多くの説明変数を導入するモデルほど、現象をより模倣できることは明らかである。一方で、包括的なモデルは未知パラメータを多く含むことから、推定誤差が大きくなってしまい、モデルの精度が低くなる可能性がある。また、データに対するあてはまりのよさが同程度である場合は、より単純なモデルのほうが望ましいといわれている。よって、現象を模倣する上での精密さと単純さの2つの観点からモデルを特定した上でパラメータを推定する必要がある。

a) Adaptive Lassoの概要

変数選択の伝統的な方法としては、総当たり法や逐次選択法が挙げられる。総当たり法とは、すべてのモデルの組合せの中から最適な変数の組を選択する方法であるが、考慮する説明変数の数が多い場合にはその組合せ数が爆発し、変数選択の方法として非効率的である。逐次選択法とは、逐次的に説明変数を入れ替えることで適合度の高いモデルを選択する方法で、変数の選択規則によって変数増加法、変数減少法、変数増減法、変数減増法に細分化される。この方法は変数選択の方法として特に多用されているが、選択される結果の不安定性が指摘されている^{18),19)}。

このような背景から、新たな変数選択の方法として Lasso (Least Absolute Shrinkage and Selection Operator) が提案された。Lasso とは説明変数のパラメータの絶対値の和として与えられる L_1 ノルムを制約条件として課した上で、最小二乗法によりパラメータを推定する方法である。回帰モデルに対しては、その制約の性質からパラメータの一部を完全に 0 に推定するという特徴を有しており²⁰⁾、変数選択とパラメータ推定を同時に行える。0 と推定されたパラメータに対応する説明変数は目的変数に寄与しないと解釈できることから、得られたモデルには有意な説明変数のみが残る。このような利点から、Lasso は考慮する説明変数が多い場合の変数選択の方法としてその価値が認められる一方で、サンプルサイズを大きくしても 0 となるパラメータの説明変数が正しく選択されないことが指摘されている^{18),20)}。

そこで、本研究では上述の問題を解決するために改良された Adaptive Lasso²⁰⁾を採用する。Adaptive Lasso では、Lasso の正則化項に重みパラメータ w をかけたものを正則化項としている。線形回帰式における Adaptive Lasso は次の式で表される。

$$\arg \min_{\beta} \left\| \mathbf{y} - \sum_{k=1}^p \mathbf{x}_k \beta_k \right\|^2 + \lambda \sum_{k=1}^p w_k |\beta_k| \quad (2)$$

ここで、 \mathbf{y} を目的変数、 \mathbf{x}_k ($k = 1, \dots, p$) を説明変数、 β_k を k 番目の説明変数に対応するパラメータとする。ただし、 \mathbf{y} に関しては中心化、各 \mathbf{x} に関しては標準化されているものとする。また λ は正則化パラメータと呼ばれ

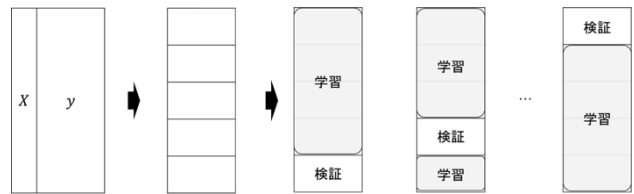


図-6 クロスバリデーションの概略²⁰⁾

る。 λ を大きくすると一般に 0 と推定されるパラメータが多くなり、 λ を小さくすると 0 となるパラメータが少なくなる。 w_k は重みパラメータベクトルである。 $\mathbf{w} = (w_1, \dots, w_p)$ の推定値 $\hat{\mathbf{w}}$ は式(3)により与えられる。

$$\hat{\mathbf{w}} = \frac{1}{|\hat{\beta}|^\gamma} \quad (3)$$

ここで、 $\gamma > 0$ であり、 $\hat{\beta}$ は最小二乗法²¹⁾や Ridge^{22),23)} によって得られた推定値とする。本研究では Ridge による推定量を用いることとした。式(2), (3)より、Adaptive Lasso の推定量は以下のように与えられる。

$$\hat{\beta} = \arg \min_{\beta} \left\| \mathbf{y} - \sum_{k=1}^p \mathbf{x}_k \beta_k \right\|^2 + \lambda \sum_{k=1}^p \hat{w}_k |\beta_k| \quad (4)$$

b) 正則化パラメータの設定基準

L_1 ノルムを正則化項として課す際、モデルは正則化パラメータ λ の値に依存する。そのため、 λ の値の設定は重要な問題となる。

本研究では λ の選択方法としてクロスバリデーション (以下、CV) を用いる。CV は、観測データを学習データと検証データに分割し、仮想的に予測二乗誤差を推定する方法である^{18),20),24)}。CV の概略を図-6 に示す。まず、サンプルサイズ n のデータを K 個 ($K \leq n$) のデータ集合に分割する (図-6 では、 $K = 5$ としている)。次に、 k 番目のデータ集合 D_k のみを除く ($K - 1$) 個のデータ集合を学習データとしてパラメータを推定し、得られる推定量を $\hat{\beta}^{-k}$ とする。そして、パラメータ推定に用いていない残りのデータ集合 D_k を検証データとして、検証誤差 CV_k を計算する。検証誤差 CV_k は式(5)で与えられる。

$$CV_k = \frac{\sum_{i \in D_k} (y_i - \mathbf{x}_i^T \hat{\beta}^{-k})^2}{n_k} \quad (5)$$

ただし、 n_k はデータ集合 D_k の大きさである。この計算を計 K 回繰り返すことで得られる式(6)の検証誤差の平均をモデル選択の基準値として用いる。

$$CV = \frac{1}{K} \sum_{k=1}^K CV_k \quad (6)$$

本研究では、正則化パラメータ λ を変化させながら、 λ と平均検証誤差の関係を見る。その際に、1) 平均検証誤差が最小となる λ を選択する場合と、2) 1 標準誤差ルールに基づいた λ を選択する場合を考える。2) の方法は、平均検証誤差の最小値にその標準誤差を加算した

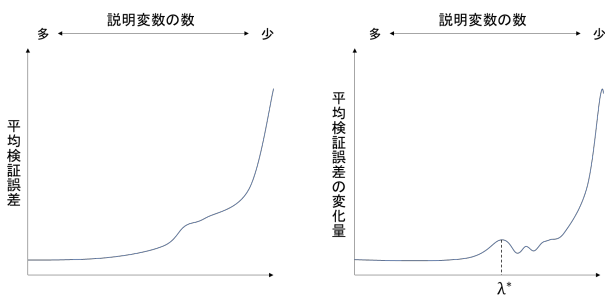


図-7 平均検証誤差の変化量増減のイメージ図

値を上回らない平均検証誤差を持つ λ のうち、最大のものを選択する方法である。これらの方法は、CVによる λ の選択方法としてよく用いられる。

さらに、これら2つの方法に加え、3) 平均検証誤差の変化量から λ を選択する方法を検討する。この方法について図-7を例に説明する。図-7の左図は平均検証誤差と λ の関係を表している。 λ の値が大きいほど0と推定されるパラメータが多くなることから、この値を大きくするにつれて平均検証誤差も一般には大きくなる傾向にある。本研究では、 λ を変化させたときの平均検証誤差の変化量を算出し、その変化の増減傾向に着目した。図-7の右図は平均検証誤差の変化量と λ との関係を表している。図-7の例では破線部分の λ^* において平均検証誤差の変化量が急激に大きくなる。平均検証誤差の変化量が急激に大きくなるまでは、平均検証誤差は λ によってあまり変化せず小さいまま0と推定されない説明変数の数が減っていく状態と考えられる。一方で、変化量が急激に大きくなった後は、平均検証誤差が大きくなったまま0と推定されない説明変数が減っていく状態になっていると考えられる。そこで、平均検証誤差が小さい中でも、より単純なモデルを選択するという考えに基づいて、平均検証誤差の変化量が最初に増加傾向を示したときの λ を設定することとした。

以上、3つの方法を正則化パラメータの設定基準とした上でモデルを推定する。なお、本研究では既往研究²⁹⁾²⁹⁾を参考に、CVの分割数は $K = 10$ とした。

(3) モデルの選定

本研究では表-2に記した16種のモデルに対し、Adaptive Lassoを用いた変数選択とパラメータ推定を行う。その際、正則化パラメータの設定基準に3つの方法を用いることから、得られる推定結果は計48通りある。これらの結果から移動需要と駅周辺環境の関連性を記述するモデルを検討しなければならないが、すべての推定結果を考察し、比較することは困難である。そこで、1) 変数選択後の説明変数の数が多くないこと、2) 15歳以上64歳以下人口に関するパラメータが正で有意であることの2つの条件を設定し、モデルを選定することとした。

1) の「変数選択後の説明変数の数が多くないこと」は、説明変数の多さによる推定結果の解釈の複雑さを避けるために設定した。本研究では、Adaptive Lassoによる変数選択後説明変数の数が、選択前の数の90%以下であることを条件とした。

2) の「15歳以上64歳以下人口に関するパラメータが正で有意であること」は直感に即した推定結果の解釈のための条件である。これは経路検索システムはPCやスマートフォンから利用されることを考えると、検索者は子供や高齢者よりは15歳以上64歳以下の人々と想定するほうが自然と考えたからである。

以上2つの条件をもとにモデルを絞り込んだのち、自由度調整済み決定係数が高いモデルを採用する。

(4) 推定結果

3. (1)に示した4種類の関数形と4つの説明変数群、および3. (2). b)に示した3つの λ の決定基準をもとに48種の結果を導出した。そのうち3. (3)に示した条件を満たすモデルは3つであった。この3つのうち自由度調整済み決定係数が高いモデルとして、目的変数を対数変換した片対数形回帰式かつ説明変数に半径500m圏内の人口と施設数および駅間距離を用いたものが採用された。以下、条件を満たし、最も精度が高かったモデルの結果についてのみ考察する。

推定結果を表-3に示す。なお、この結果では1標準誤差ルールに基づいた正則化パラメータが用いられている。また、目的変数を対数変換する場合の片側対数形のモデルでは、出発駅・到着駅別平均検索件数が0件となった組は事前に除外している。さらに本研究では、どの駅においても半径500m圏内に存在する観光資源が存在しなかったため、この説明変数も分析から除外している。

表-3中の推定値が「-」となっている説明変数はパラメータが0と推定されたものであり、平均検索件数に影響しない説明変数と解釈できる。

次に、0と推定されなかった説明変数について着目すると、出発駅と到着駅に関するパラメータに大きな違いは見られないことがわかった。ただし、出発駅では有意にならなかった「福祉施設」と「その他の小売店」が到着駅では有意なパラメータとして残り、出発駅では負で有意となった「集客施設」と「道の駅」が到着駅では正で有意なパラメータとして残った。

出発駅と到着駅ともに負のパラメータとして推定された説明変数には、「14歳以下人口」、「外国人人口」、「都市公園」、「学校」、「事業所」、「専門スーパー」、「専門店」が挙げられる。これらの説明変数は出発駅・到着駅別平均検索件数の減少に影響し、移動需要が発生しにくいことを表している。例えば、「学校」や「事業所」に関するパラメータの符号が負であることか

表-3 推定結果

説明変数	推定値
切片	0.00
駅間距離	-0.44
出発駅 500m 圏内の 14 歳以下人口	-0.30
出発駅 500m 圏内の 15 歳以上 64 歳以下人口	0.48
出発駅 500m 圏内の 65 歳以上人口	0.27
出発駅 500m 圏内の外国人人口	-0.44
出発駅 500m 圏内の都市公園	-0.46
出発駅 500m 圏内の医療機関	1.06
出発駅 500m 圏内の福祉施設	-
出発駅 500m 圏内の文化施設	0.17
出発駅 500m 圏内の学校	-0.11
出発駅 500m 圏内の市町村役場	-
出発駅 500m 圏内の集客施設	-0.02
出発駅 500m 圏内の道の駅	-0.07
出発駅 500m 圏内の都道府県指定文化財	0.36
出発駅 500m 圏内の事業所	-0.36
出発駅 500m 圏内の総合スーパー	0.10
出発駅 500m 圏内の専門スーパー	-0.04
出発駅 500m 圏内のコンビニ	-
出発駅 500m 圏内のドラッグストア	0.16
出発駅 500m 圏内のその他のスーパー	0.37
出発駅 500m 圏内の専門店	-0.25
出発駅 500m 圏内の家電大型専門店	0.10
出発駅 500m 圏内の中心店	-
出発駅 500m 圏内のその他の小売店	-
出発駅 500m 圏内の無店舗販売	0.17
到着駅 500m 圏内の 14 歳以下人口	-0.25
到着駅 500m 圏内の 15 歳以上 64 歳以下人口	0.51
到着駅 500m 圏内の 65 歳以上人口	0.18
到着駅 500m 圏内の外国人人口	-0.48
到着駅 500m 圏内の都市公園	-0.38
到着駅 500m 圏内の医療機関	1.10
到着駅 500m 圏内の福祉施設	-0.05
到着駅 500m 圏内の文化施設	0.26
到着駅 500m 圏内の学校	-0.07
到着駅 500m 圏内の市町村役場	-
到着駅 500m 圏内の集客施設	0.01
到着駅 500m 圏内の道の駅	0.00
到着駅 500m 圏内の都道府県指定文化財	0.29
到着駅 500m 圏内の事業所	-0.53
到着駅 500m 圏内の総合スーパー	0.10
到着駅 500m 圏内の専門スーパー	-0.09
到着駅 500m 圏内のコンビニ	-
到着駅 500m 圏内のドラッグストア	0.18
到着駅 500m 圏内のその他のスーパー	0.46
到着駅 500m 圏内の専門店	-0.18
到着駅 500m 圏内の家電大型専門店	0.17
到着駅 500m 圏内の中心店	-
到着駅 500m 圏内のその他の小売店	0.15
到着駅 500m 圏内の無店舗販売	0.20
サンプル数	2293
決定係数	0.61
自由度調整済み決定係数	0.60

ら、通学や通勤を移動目的としている人は、経路検索システムは利用しないと考えられる。つまり、毎日決まっ

た経路による移動である通学や通勤においては検索システムを利用しないと解釈でき、この結果は妥当といえる。

出発駅と到着駅ともに正のパラメータとして推定された説明変数には、「15 歳以上 64 歳以下人口」、「65 歳以上人口」、「医療機関」、「文化施設」、「都道府県指定文化財」、「総合スーパー」、「ドラッグストア」、「その他のスーパー」、「家電大型専門店」、「無店舗販売」が挙げられる。これらの説明変数は出発駅・到着駅別平均検索件数の増加に寄与すると考えられ、移動需要の発生に関係のある変数といえる。人口に関する変数として「15 歳以上 64 歳以下人口」だけでなく「65 歳以上人口」も正のパラメータとして推定されたことから、経路検索システムの利用者は生産年齢人口に加え高齢者人口とも関連性があると考えられる。また「医療機関」、「総合スーパー」、「ドラッグストア」、「その他のスーパー」、「家電大型専門店」からは、移動目的は通院あるいは買い物であることが推察される。ただし、買い物を目的とした移動については生鮮食品や加工食品などの食品購入だけでなく、施設の特徴から衣類や化粧品、家電や住宅関連商品の購入も含まれると考えられる。

「文化施設」、「都道府県指定文化財」は、観光客の存在を示唆している。したがって、経路検索システムの利用者は診療や不定期に行う購買行動、居住地域とは異なる場所での観光行動のために経路検索システムを利用していると推察される。

以上の結果を踏まえると、経路検索データに現れる移動需要は、通院や買い物を目的とした住民や観光客との関連性が強いと考えられる。通院や買い物、観光による移動は非定常的なものであり、出発駅・到着駅別平均検索件数と組み合わせることによって、日々発生する総移動需要を表現できると考えられる。

4. おわりに

本研究では経路検索データに現れる移動需要と駅周辺環境との関連性を明らかにし、検索者の属性や移動目的を推察するために、4つの関数形と4つの説明変数群の組合せから16種のモデルを構築した。さらに、移動需要に影響を及ぼす変数のみを抽出するために、Adaptive Lassoを適用し、説明変数の選定とパラメータの推定を行った。Adaptive Lassoにおける正則化項パラメータの設定基準としては3通りの方法を検討した。

異なる関数形、異なる説明変数群、異なる正則化項パラメータの設定基準のもとで推定した複数の推定結果を、2つのモデル選定条件とモデル適合度を表す自由度調整済み決定係数によって比較した。その結果、目的変数を対数変換した片対数形で、説明変数に半径500m圏内の

人口と施設数および駅間距離を用いたモデルが、候補モデルの中で最適と判断された。

採用された片対数形モデルの結果、全説明変数の約 84%を占める41の説明変数が有意なものとして選択された。さらに選択された説明変数は、出発駅と到着駅で大きな違いは見られなかったことから、検索者が往路、復路ともに経路検索システムを利用していることが示唆された。これは妥当な結果であり、適切にモデル推計がなされていることを裏付けるものである。

選択された説明変数のうち、正のパラメータに対応する説明変数から、検索者は生産年齢人口や高齢者人口と関連があり、診療や不定期に行う購買行動、居住地域とは異なる場所での観光行動のために経路検索システムを利用していると推察された。

このように、駅間の移動需要と駅周辺環境の関連性を分析することで、経路検索データに現れる移動需要がどのような人のどのような移動目的によるものであるかを推測できることが示された。非日用品の買い物、通院を目的とした住民、あるいは観光客が経路検索システムを利用しているという結果は、事前に想定できる結果である。しかし、定量的な根拠を持って経路検索システムの利用者像を明らかにできたことは、今後経路検索データを用いた研究のみならず実際の公共交通の運行計画評価・立案手法の開発に大きく寄与するものと期待できる。

さらに、本研究で用いた駅周辺環境に関するデータはすべて無償で公開されているオープンデータである。これらのデータは各府省から定期的に公表されていることから、継続的な分析を可能としている。さらに、分析対象地域のデータを収集すれば、どの地域においても本手法の適用が可能となっている。経路検索システムの利用者像を理解するために多用されてきたアンケート調査を回避できることから、調査実施者と回答者の双方の負担が少ないことも提案手法の利点であると考えられる。

本研究では人口に関する説明変数として、「14歳以下人口」、「15歳以上64歳以下人口」、「65歳以上人口」、「外国人人口」を用いている。これらは駅周辺に居住している人口、すなわち夜間人口に相当するといわれている。夜間人口に加えて、昼間人口に関する情報も取得できれば、人口の流入と移動需要の関連も推定できる可能性がある。さらに、「15歳以上64歳以下人口」を年代別に集計できれば、より詳細に検索者像が推測可能になることから、その集計方法についても検討していきたい。

参考文献

- 1) 太田恒平, 野津直樹: 経路検索条件データを用いた交通・観光行動分析 ~移動需要ビッグデータでわかること~, 土木計画学研究発表会・講演集, Vol. 52, CD-ROM, 8pages, 2015.
- 2) 見生元気, 伊藤昌毅, 川村尚生, 菅原一孔: 公共交通乗換案内サービスを用いた利用者行動解析, 地理情報システム学会講演論文集, Vol. 21, CD-ROM, 4pages, 2012.
- 3) 香川喬之, 桑野将司, 福山敬, 谷本圭志, 川村尚生: バス経路検索履歴データを用いた移動特性の分析, 交通工学論文集, Vol. 2, No. 2, pp. A_115-A_124, 2016.
- 4) 寺部慎太郎, 重里光佑, 内山久雄: 鉄道経路探索 web サイトに関する利用実態の特徴分析, 土木計画学研究・論文集, Vol. 26, No. 3, pp. 569-574, 2009.
- 5) 轟朝幸, 谷口滋一, 高宮則夫: 公共交通乗り換え検索ログの特徴分析と活用可能性について, 土木計画学研究発表会・講演集, Vol. 31, CD-ROM, 4pages, 2005.
- 6) 統計情報リサーチ: 高松琴平電鉄の駅別乗降客数ランキング, http://statsearch.jp/traffic/train/passengers_company_ranking_55.html
- 7) 高松琴平電気鉄道株式会社: ことでん路線図, http://www.kotoden.co.jp/publichtm/kotoden/timetable_revision/time/
- 8) 株式会社ナビタイムジャパン: NAVITIME DATA, <https://recruit.navitime.co.jp/workplace/navitimeedata.html>
- 9) 長尾基哉, 中川大, 松中亮治, 大庭哲治, 望月明彦: 地方都市における鉄道・軌道の運行頻度に着目した駅周辺人口分布の経年変化に関する研究, 土木計画学研究・論文集, Vol. 27, No. 2, pp. 399-407, 2010.
- 10) 大賀淳平, 森本章倫, 大門創: 道路容量の減少からみた構想段階における LRT 導入検討の簡便な枠組みに関する研究, 土木学会論文集 D3 (土木計画学), Vol. 73, No. 5, pp. I_783-I_790, 2017.
- 11) 小瀬木祐二, 戸川卓哉, 鈴木祐大, 加藤博和, 林良嗣: 大都市圏スケールでのインフラ維持管理・更新費用の将来推計手法の開発, 土木計画学研究・論文集, Vol. 27, No. 2, pp. 305-312, 2010.
- 12) 中野卓, 出口敦: 首都圏郊外の駅前商業地区における世帯動向の実態と居住環境の課題に関する研究, 日本建築学会計画系論文集, Vol. 83, No. 745, pp. 435-445, 2018.
- 13) 高瀬達夫, 湧井克明, 小山健: 長野県における地方鉄道の価値を考慮した路線存続に関する研究, 土木学会論文集 F4 (建設マネジメント), Vol. 67, No. 4, pp. I_57-67, 2011.
- 14) 古内大志, 村木美貴: 防災性を考慮した自立・分散型エネルギーシステム構築に関する研究, 都市計画論文集, Vol. 53, No. 3, pp. 986-992, 2018.
- 15) 総務省統計局: e-Stat 政府統計の総合窓口, <https://www.e-stat.go.jp/>
- 16) 国土交通省: 国土数値情報, <http://nlftp.mlit.go.jp/ksj/>
- 17) 経済産業省: 商業統計, <https://www.meti.go.jp/statistics/tyo/syogyo/index.html>
- 18) 川野秀一, 廣瀬慧, 立石正平, 小西貞則: 回帰モデリングと L1 型正則化法の最近の展開, 日本統計学会誌, Vol. 39, No. 2, pp. 211-242, 2010.
- 19) 荒木孝治: 罰則付き回帰とデータ解析環境 R, オペレーションズ・リサーチ, Vol. 58, No. 5, pp. 261-266, 2013.
- 20) 川野秀一, 松井秀俊, 廣瀬慧: スペース推定法による統計モデリング, pp. 1-27, 共立出版, 2018.
- 21) Zou, H.: The adaptive lasso and its oracle properties, *Journal of the American Statistical Association*, Vol. 101, pp. 1418-1429, 2006.
- 22) Basu, T., Einbeck, J. and Troffaes, C. M. M.: A sensitivity analysis of the adaptive lasso, https://www.researchgate.net/publication/335570398_A_sensitivity_analysis_of_the_adaptive_lasso
- 23) Chen, K. and Chan, K.: Subset ARMA selection via the adaptive Lasso,

- Statistics and its interface, Vol. 4, No. 2, pp. 197-205, 2011.
- 24) 野崎俊貴, 木村拓海, 川野秀一: スパース推定に基づく適応正則化オンライン学習の特徴選択問題, 計算機統計学, Vol. 29, No. 2, pp. 117-131, 2016.
- 25) 大浦聡一郎, 藤田隆吾, 松川徹, 西郷浩人, 鈴木英之進: LASSOに基づく運転手表情データからの特徴抽出, 電気関係学会九州支部連合大会講演論文集, Vol. 69, pp. 70-71, 2016.
- 26) 林裕志, 平松達雄, 小出大介, 田中勝弥, 大江和彦: 電子カルテデータベースからのLASSOロジスティック回帰による医薬品副作用シグナルの検出: ケース・コントロール研究, 薬剤疫学, Vol. 21, No. 2, pp. 51-62, 2017. (?)

ANALYSIS OF THE RELATIONSHIP BETWEEN TRANSPORTATION DEMAND BASED ON ROUTE SEARCH RECORDS AND CHARACTERISTICS AROUND RAILWAY STATION

Mio HOSOE, Masashi KUWANO, Taku MORIYAMA

In the public transit navigator system, the user's travel demands, such as the requested departure place and destination place and what time to move, are recorded as log data. However, the log data don't include information of user's attributes and their travel purpose. Therefore, it is not possible to clarify what kind of people travel for what purpose from the log data. This study analyzes the user's attributes and their travel purpose from the relationship between travel demands on the log data and characteristics around railway station. Specifically, this study uses the log data of "NAVITIME", which is a public transit navigator system introduced in Japan, and expresses travel demands. Characteristics around railway station are expressed by population data and facility's location data. Then, this study constructs several models by different functional forms and explanatory variables, and chooses the most suitable model among them. From the estimation results of the best fit model, it is confirmed that inhabitants and tourists use public transit navigator system for nonsteady travel purpose such as shopping, attending a hospital and sightseeing.