

# 状態空間モデルによる突発的検索行動の抽出

小田島 輝知<sup>1</sup>・桑野 将司<sup>2</sup>・森山 卓<sup>3</sup>・細江 美欧<sup>4</sup>

<sup>1</sup>学生会員 鳥取大学 持続性社会創生科学研究科工学専攻 (〒680-8552 鳥取市湖山町南四丁目101)

E-mail: m20j6004z@edu.tottori-u.ac.jp

<sup>2</sup>正会員 鳥取大学教授 工学研究科社会基盤工学専攻 (〒680-8552 鳥取市湖山町南四丁目101)

E-mail: kuwano@tottori-u.ac.jp

<sup>3</sup>非会員 鳥取大学助教 工学研究科社会基盤工学専攻 (〒680-8552 鳥取市湖山町南四丁目101)

E-mail: moriyama@tottori-u.ac.jp

<sup>4</sup>学生会員 鳥取大学 工学研究科社会基盤工学専攻 (〒680-8552 鳥取市湖山町南四丁目101)

E-mail: d19t4003b@edu.tottori-u.ac.jp

交通統計の手法は時代とともに改良が加えられているが、現在も代表値の把握に留まっている。交通の安全性と合理性の確保のためには、突発的移動需要の要因を明らかにすることが重要である。本研究では兵庫県南部で路線バスや高速バスを運行する神姫バスグループの経路検索システムのログデータを用いて、状態空間モデルにより突発的検索行動を抽出し、その要因を考察した。具体的には、日別検索件数の推移を説明する最適なモデルを検討し、モデルを以って説明できなかった不規則な変動から突発的検索行動を抽出した。さらに、その要因についてウェブや新聞を利用し考察した。モデル分析の結果、365日の分析対象期間中に、18日突発的検索が発生したことが明らかとなった。また、広範囲に影響を及ぼす天候不良や大規模な催し物などが、突発的移動需要の発生に影響していることが明らかとなった。

**Key Words :** *public transit navigator system, log data, time series analysis, irregular fluctuations, state-space model*

## 1. 序論

交通統計の手法は時代とともに改良が加えられてきたが、交通センサスやパーソントリップ調査などは依然として代表値の把握に留まっている。ここでの代表値とは平均的な交通行動を示すものであり、突発的な交通行動の分析には重点が置かれていない。日常的な状態と対をなす状態を本研究では「突発的」な状態と定義する。突発的な状態は、平常時と比べ混雑した状態、および平常時より閑散した状態のこととする。具体的には、スポーツの大会や有名な展覧会が開かれるため交通行動が多い状態、および自然災害や天候不良により交通行動が少ない状態などが考えられる。このような突発的な状態により交通の安全性と合理性が脅かされる可能性があるため、突発的移動需要の要因を解明することが重要である。

近年、ICT (Information Communication Technology) の進展と普及が目覚ましく、ETC (Electronic Toll Collection System) や携帯電話やスマートフォンの基地局データ、GPSデータ、交通系ICカードなどを通じて交通行動に関

する詳細なデータが得られるようになった。このようなデータを用いて、経路検索システムのログデータと突発的移動需要との関係を分析した研究が近年行われている。伊藤ら<sup>1</sup>は、ログデータと公共交通の需要や利用状況との関係を分析した。鳥取県に導入されている経路検索システム「バスネット」のログデータを解析し、この検索件数と実際のバス利用者数を比較した結果、経路検索された件数と実際の利用者数との間には高い相関があることを指摘している。坪内らの研究<sup>2</sup>では、Yahoo!経路検索のログデータを用いて、異常混雑の予測を行っている。坪内らはまず、天気や曜日、住民や通勤してくる人の属性などさまざまな特徴量を用いて定常時の検索数を明らかにした。次に、事前検索数と当日の検索数との相関関係を分析し、将来の検索件数をモデル化した。そして得られた将来検索数と定常時の検索数を比較し、その差がある場合は異常混雑が発生する可能性が高いと定めた。この結果、大規模混雑につながるようなイベントによる突発的移動需要は高い精度で検出できるものの、平常時の5倍程度の中規模混雑につながるイベントでは精度が

低いことを明らかにした。現在この手法を用いて、一部の鉄道駅における5日先までの混雑を予報するサービスがスマートフォンアプリなどにより一般に提供されている<sup>23,24</sup>。検索行動と実際の交通行動は異なるが、ログデータを用いればこのように実際の交通行動が概ね把握できることが明らかとなっている。

ここで、大雨・台風・積雪などの発生やイベントの開催によって非日常的な利用者が急増し、利用者の積み残しが発生した状況を考える。このとき交通系ICカードデータのような実行データでは、乗車できなかった人のデータが得られないため、どの程度の需要があったかを把握することができない。一方で移動希望が入力される検索ログデータを用いれば、実際の移動需要を推定できる可能性がある。このように突発的移動需要を分析する上では、交通系ICカードデータのような実行データよりも経路検索システムのログデータが適していると考えられる。さらに、近年中小都市ではバス利用者数減少を背景に、運行計画の再考が求められている。しかし、人材や費用などの資源制約から、計画見直しのために必要な調査を実施することは容易ではない。経路検索システムのログデータは導入が比較的容易で、かつ近年全国各地で導入が進められており、より少ない資源での調査を可能とすることが期待できる。以上の理由から、本研究では経路検索システムのログデータに着目する。

突発的移動需要を分析する本研究では、状態空間モデルを用いて、過去の経路検索システムの総検索数から突発的検索行動があった日を抽出する。その後、抽出された日の突発的移動需要の要因について考察を行うことを目的とする。

## 2. 使用データの概要と基礎集計

### (1) 使用データの概要

本研究では、神姫バスグループが提供するバス事業者版経路検索システム「神姫バスNavi」の2018年10月1日から2019年9月30日までの1年間分のログデータを用いる。神姫バスグループは兵庫県南部を中心に路線バスを運行し、神戸市や姫路市などから四国地方、山陽地方、山陰地方などを結ぶ高速バスを運行する企業のグループである。「神姫バスNavi」は運行する路線バスおよび高速バスの経路と時刻、時刻表の検索など、表-1に示す5つの機能を有している。

このシステムには、PCやスマートフォンなどからブラウザを通じて利用する方法とスマートフォン向けアプリから利用する方法がある。PCからブラウザを通じて利用した場合の条件入力画面と経路案内画面を図-1に示す。入力できる項目は、出発地と目的地2か所のバス停

表-1 神姫バスNaviの機能とその説明

機能	内容
経路検索	経路、時刻、運賃を示す
時刻表表示	バス停の時刻表を表示する
乗り場マップ表示	バス停の位置を示す地図を表示する
バス位置表示	バスの現在の位置を表示する
ルート表示	指定した経路のバスが通るバス停を地図上に表示する



図-1 神姫バスNaviの条件入力画面と経路検索案内画面

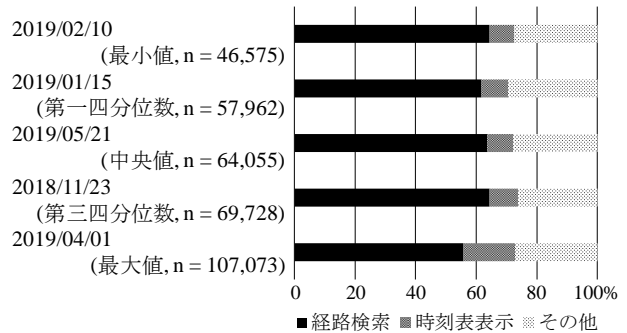


図-2 利用機能の割合

名あるいは特定の1か所のバス停名、出発予定あるいは到着予定の日時、時刻の指定方法である。時刻の指定方法とは、出発時刻、到着時刻、始発バスあるいは最終バスの選択肢から、どの条件を指定することを指す。

神姫バスNaviのログデータには、システム利用年月日、システム利用時刻、バス停名、利用機能(表-1を参照)、使用端末(Android, iPhone, PC, フィーチャーフォンなどの区別)の情報が含まれている。図-2に代表の5日における利用機能の割合を示す。ここで、代表の5日とは日別総検索件数を昇順に並べ替えたときの最小値、最大値および四分位数をとる日を指す。図-2の利用機能の割合を見ると、代表の5日とも似た割合を示しており、経路検索が約65%と最も大きく、時刻表表示が約8%と次いで大きい。利用される機能は、日ごとに大きな違いがないことから、本研究では利用される機能の別には着目せず、日別システム総利用件数(アクセス件数)による分析を行う。

(2) 基礎集計

図-3に1年間の日別システム総利用件数の推移を示す。1年間では延べ23,508,476件の利用があった。1日あたりでは平均約64,407件の利用があり、中央値は2019年5月21日(火)の64,055件、最も利用件数が大きい日は2019年4月1日(月)の107,073件、最も利用件数が小さい日は2019年2月10日(日)の46,575件である。

図-4に日別システム利用件数のヒストグラムを示す。1日当たりの利用件数の分布は、65,000件より多く70,000件以下の度数が最も大きい。90,000件以上の日は2019年の4月1日(月)、9月25日(水)、3月21日(木・祝)、8万5千件以上の日は2019年の6月7日(金)、9月23日(月・祝)、5月6日(月・祝)、1月2日(水)、2018年10月1日(月)に観測されている。

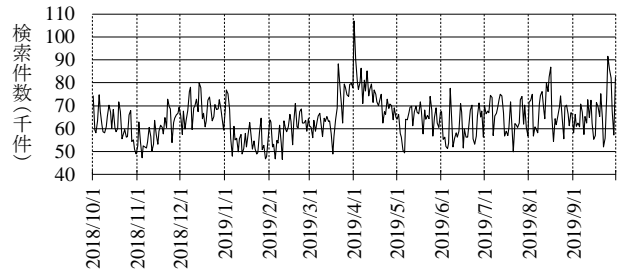


図-3 日別システム総利用件数

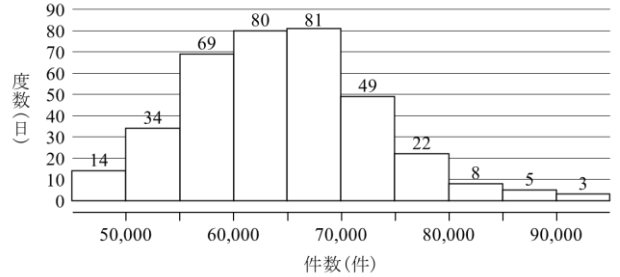


図-4 日別システム利用件数のヒストグラム

3. 突発的検索行動抽出のための統計的手法

本研究では、状態空間モデルを用いて突発的検索行動の抽出を試みる。時系列モデルの1つである状態空間モデルは、データの変動分析に適しているといわれる。

$n = 1, 2, \dots, T$ を時点、 $x_n$ を状態値(ベクトル)、 $y_n$ を観測値としたとき、状態空間モデルの一般式は式(1)の状態方程式と式(2)の観測方程式で表現できる。観測値 $y_n$ に加えて状態値 $x_n$ を仮定することが状態空間モデルの特徴である。

$$x_n = F_n x_{n-1} + G_n v_n \quad (1)$$

$$y_n = H_n x_n + w_n \quad (2)$$

ただし、

$$x_0 \sim N(x_{0|0}, V_{0|0})$$

$$v_n \sim N(0, V_v)$$

$$w_n \sim N(0, \sigma^2)$$

$F_n, G_n, H_n$ : 各時点 $n$ において決まる定数行列

$x_{\cdot|k}$ : 条件付き平均 ( $x_{j|k} = E[x_j | y_{1:k}]$ )

$y_{1:k}$ : 時点 $n = k$ までに得られた観測値全体

$$(y_{1:k} = \{y_1, y_2, \dots, y_k\})$$

$V_{\cdot|k}$ : 分散共分散行列

$$(V_{j|k} = E[(x_j - x_{j|k})(x_j - x_{j|k})^T])$$

$v_n$ : システムノイズ行列

$V_v$ : システムノイズ行列の分布の分散の行列

$w_n$ : 不規則成分

$\sigma^2$ : 不規則成分の分布の分散

状態空間モデルは状態値の遷移と観測値と状態値の関係がマルコフ過程に従うことを仮定している。本研究では、全ての状態値と観測値は正規分布に従うと仮定する。カルマンフィルタアルゴリズムにより、各時点の値を求める。次の時点を予測した後にフィルタリングを行い、観測値を用いて状態値の補正を行う。

[次時点の予測]

$$x_{n|n-1} = F_n x_{n-1|n-1} \quad (3)$$

$$V_{n|n-1} = F_n V_{n-1|n-1} F_n^T + G_n Q_n G_n^T \quad (4)$$

[フィルタリング]

$$K_n = V_{n|n-1} H_n^T (H_n V_{n|n-1} H_n^T + \sigma^2)^{-1} \quad (5)$$

$$x_{n|n} = x_{n|n-1} + K_n (y_n - H_n x_{n|n-1}) \quad (6)$$

$$V_{n|n} = (I - K_n H_n) V_{n|n-1} \quad (7)$$

ただし、

$Q_n$ : 各時点 $n$ において決まる分散共分散行列

$K_n$ : 各時点 $n$ において決まる定数行列

このようにして得られたフィルタ分布は、 $n$ 時点までのすべての観測値を用いた状態の推定となる。

次時点の予測とフィルタリングにより $n$ 時点すべてのパラメータ求まったところで、最終時点から順に時点に戻りながら、改めてパラメータを計算する平滑化を行う。式(8)から式(10)に示す。

$$A_n = V_{n|n} F_n^T V_{n+1|n}^{-1} \quad (8)$$

$$x_{n|T} = x_{n|n} + A_n (x_{n+1|T} - x_{n+1|n}) \quad (9)$$

$$V_{n|T} = V_{n|n} + A_n (V_{n+1|T} - V_{n+1|n}) A_n^T \quad (10)$$

ただし、

$A_n$ : 各時点 $n$ において決まる定数行列

状態空間モデルには多様なバリエーションがあるが、本研究ではローカルモデルとローカル線形トレンドモデルを用いる。ローカルモデルでは、平均を表すレベル成分が常に一定値であると仮定する一方で、ローカル線形トレンドモデルでは、レベル成分に線形トレンドを考慮する。また、レベル成分およびトレンド成分はそれぞれ確定的 ( $v_{n1} \equiv 0, u_n \equiv 0$ ) および確率的 ( $v_{n1} \neq 0, u_n \neq 0$ ) なものが存在する。周期要素のあるモデルでは、さらに元の時系列データに周期変動があることを仮定す

る。周期成分にも確定的または確率的のそれぞれを仮定することができる。本研究では、これらの組合せから考えられる複数のモデルを推計・比較し、赤池情報量基準 AIC によって、最適なモデルを決定する。

以上を定式化すると、本研究で使用する観測方程式は次のように表される。

$$y_n = s_n + d_n + w_n \quad (11)$$

ただし、

$s_n$  : ローカルレベル成分

$d_n$  : 周期変動成分

$w_n$  : 不規則成分

トレンド  $t_n$  を含むレベル成分  $s_n$  は長期的な傾向を表現する。

$$s_n = s_{n-1} + t_{n-1} + v_{n1} \quad (12)$$

$$t_n = t_{n-1} + u_n \quad (13)$$

ただし、

$$v_{n1} \sim N(0, \tau_s^2)$$

$$u_n \sim N(0, \tau_t^2)$$

$t_n$  : ローカルトレンド成分

$v_{n1}$  : ローカルレベル成分の誤差成分

$u_n$  : ローカルトレンド成分の誤差成分

$\tau_s^2$  : ローカルレベル成分の誤差成分の分散

$\tau_t^2$  : ローカルトレンド成分の誤差成分の分散

周期変動成分  $d_n$  は季節や曜日のような繰り返し現れる変動成分を表現する。事前分析として、1 年間の検索件数の推移に対し自己相関係数を算出した結果、ラグが 7 のときに高い正の相関が確認された。使用データ期間が 1 年であることから、月変動や季節変動は考慮できないため、本研究では周期変動は曜日による影響のみを想定し、以下の通り、周期を 7 日とする。

$$\sum_{i=0}^6 d_{n-i} = v_{n2} \quad (14)$$

ただし、

$$v_{n2} \sim N(0, \tau_d^2)$$

$v_{n2}$  : 周期成分の誤差成分

$\tau_d^2$  : 周期成分の誤差成分の分散

このとき、各成分で構成される状態ベクトル  $x_n = (s_n, t_n, d_n, d_{n-1}, \dots, d_{n-5})^T$  には一期前のみに基づく状態方程式が成り立つ。

次に得られる各モデルの AIC を求める。各モデルのパラメータと AIC の結果を表-2 に示す。分析の結果、「C-2 確定的周期要素を持つローカル線形トレンドモデル」の AIC が最小となった。このモデルを用いて得られた不規則変動から、突発的検索行動について分析する。次章ではこの不規則変動から突発需要が発生した日を抽出し、さらにその要因について考察を行う。

#### 4. 突発的検索行動の抽出と要因の考察

##### (1) 状態空間モデルによる突発的検索行動の抽出

図-5 の対数変換した日別システム総利用件数に対して、状態空間モデルを適用した結果得られたレベル成分、周期成分、不規則成分を図-6、図-7、図-8 に示す。

図-6 のレベル成分を見ると、2019 年 7 月ごろから多少の増加傾向は見られるものの、期間全体を通じて大きな変動はみられない。これは、長期的な視点からは経路検索システムの利用者数は大きく増減していないことを意味する。レベル成分を期間に分けて考察すると、2018 年 12 月の後半から 2019 年 1 月まで、2019 年 4 月初旬においてレベル成分値が比較的大きいことがわかる。一方、2019 年 11 月上旬、2019 年 1 月下旬から 2 月中旬までは値が小さい。以上のことから、冬休みや年度始めは利用者が多く、大きなイベントが少ない平常時は利用者が少ない傾向が明らかとなった。また、2019 年 9 月末ごろに増加傾向にあるが、これは 2019 年 10 月からの増税による運賃値上げがあったため、関連する情報を収集する利用者数が増えたものと考えられる。

図-7 に示す周期成分には、確率的要素が含まれていないため、年間を通じて曜日ごとに一定の値を持つ。周期成分の曜日ごとの成分値を示した結果を図-9 に示す。図-9 から、月曜日から木曜日までは目立った特徴は無い一方で、金曜日から日曜日にかけて周期成分は特徴的な変動を示している。金曜日次いで土曜日が正で大きな値を示し、日曜日は負の値を返している。古川ら<sup>5)</sup>は、経路検索システムの利用者は、公共交通を利用する直前に検索する人もいれば、余裕をもって事前に検索する人もいることを指摘しており、特に観光目的と考えられる検索は、数日前に検索される傾向があることを明らかにしている。古川らの考察を参考に、ビジネスによる検索がお

表-2 検討したモデルのパラメータと AIC

モデル	レベルの確率成分	傾き成分	傾きの確率成分	周期成分	周期の確率成分	AIC
A-1	0	0	0	0	0	-414.98
A-2	$v_{n1}$					-582.65
B-1	0	$t_n$	0	0	0	-412.32
B-2	$v_{n1}$					$u_n$
B-3			$u_n$			-565.80
C-1	0	$t_n$	$u_n$	$d_n$	0	-430.01
C-2	$v_{n1}$					-697.58
C-3					$v_{n2}$	-695.58
D-1	0	$t_n$	0	$d_n$	0	-420.84
D-2						$v_{n2}$
D-3					$v_{n1}$	$u_n$

おむね公共交通を利用する直前であると考え、このような利用は平日に多く、休日に少ないと推察できる。一方で行楽などによる検索は公共交通を利用する前日に行われるものと考え、金曜日および土曜日に検索件数が増えることに説明がつく。また日曜日は以上の理由からその他の曜日に比べ検索件数が少ないと解釈できる。このように周期成分には、このようなビジネスを理由に直前に検索する平日利用者と、行楽の前日に検索する休日利用者の検索時期の違いが現れているのではないかと考えられる。

以上、日別システム総利用者数の変動から、レベル成分と周期成分で説明できない不規則成分(図-8)に突発的検索行動が含まれていると考え、次節で考察する。

## (2) 突発的検索行動の要因の考察

図-8の不規則成分から、突発的検索行動が発生したとみられる日を抽出したものを図-10に示す。突発的と判断する基準は、不規則変動の絶対値が0.1以上をとるものとした。この結果、突発的検索行動が発生した日は分析対象の1年間に18日あった。そして当該日およびその前後数日について、天候、催し物、学校行事などを中心にウェブサイトや過去の新聞記事から情報を集めた。この調査の結果、突発的検索行動が発生した18日のうち10日について要因を推察することができた。

2019年9月23日(月)、25日(水)、29日(日)、30日(月)に大きな突発的変動を示している。同26日(木)および30日(月)には神戸市内でラグビーワールドカップが開催されていた。さらに神姫バスは需要の増加を見込んで、ラグビーワールドカップの開催地付近で同9月26日(木)より約1か月間、連接バスの運行を行っていた。すなわち、スポーツの大規模イベント開催、あるいはそれに伴うバス運行計画の変更が、数日前から当日までの検索件数の突発的な増加に影響したと考えられる。

2018年12月31日(月)、2019年1月2日(水)も突発的検索が発生した日として検出されている。一般に大晦日の12月31日と元旦の翌日である1月2日は道路交通において交通渋滞の特異日と認知されている。これは、買出しや初詣に出かける需要が高まるためとされている。この現象は公共交通にもあてはまるため、12月31日と1月2日に不規則成分が大きくなったと考えられる。

次に、突発的変動の最も多い要因として考えられる天候に着目する。分析の結果、大型台風が通過する前後、降雪、および激しい降雨があった日に突発的成分が大きいことが明らかとなった。不規則成分が大きい変動を示した日は、台風が通過した2018年10月1日(月)、2019年7月27日(水)、2019年8月14日(水)の前後、西日本広く降雪があった2019年1月26日(土)、近隣の広島県などで避難勧告が発令されるほど激しい降雨があった

2019年6月7日(金)がある。ここで、天候が悪化した際に検索件数が増える理由は大きく2つ考えられる。1つ目は、公共交通機関のダイヤの乱れを予想し、平常時と異なる時刻での移動を望むためである。2つ目は、神姫バスNaviにある現在のバス位置を調べる機能の利用希望が高まるためである。また2019年6月7日(金)のように、対象地域で天候の悪化がそれほど大きくなくても、検索件数が大きくなる日がある。これは、周辺地域の大きな災害の発生などがメディアで取り上げられるために、公

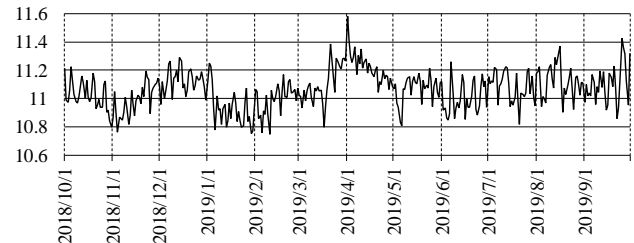


図-5 対数変換した日別システム総利用者数

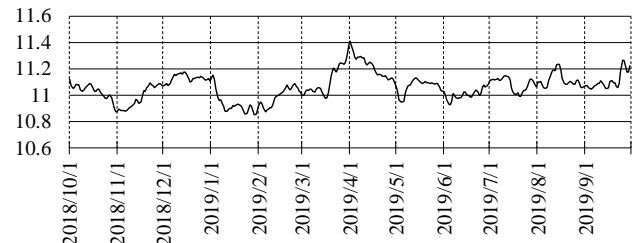


図-6 レベル成分

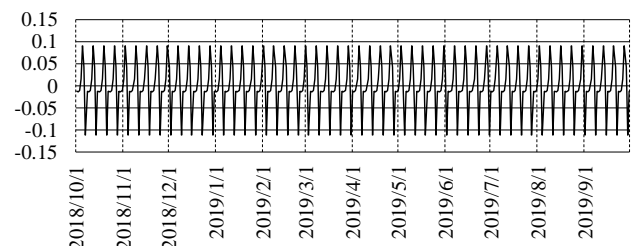


図-7 周期成分

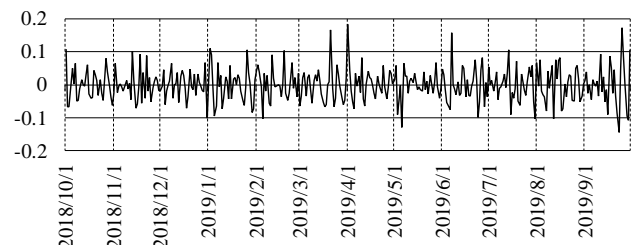


図-8 不規則成分

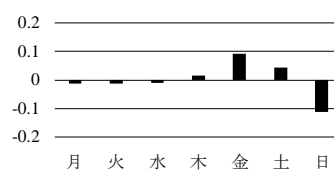


図-9 曜日ごとの成分

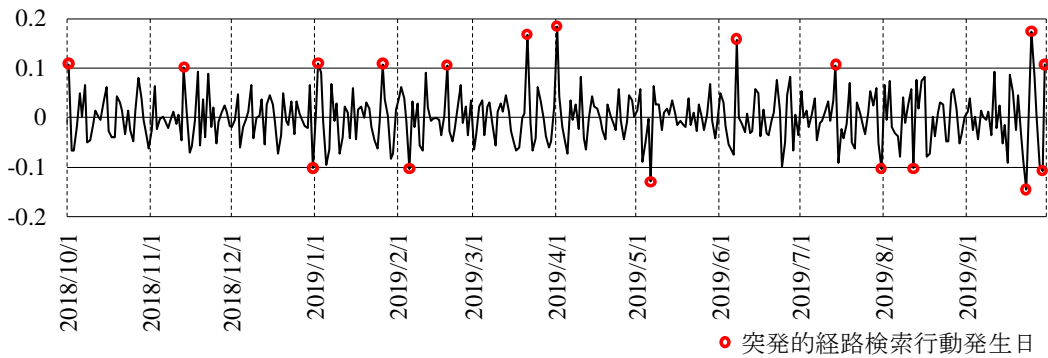


図-10 突発的検索行動の抽出結果

公共交通機関のダイヤの乱れを気にする利用者が増えたからではないかと考えられる。

最後に、要因の特定ができていない日における小規模の催し物などについて調べた。2019年3月21日(木)を調査したところ、県立大学の合格発表があったことがわかった。しかし、最寄りのバス停を目的地とする検索は全検索件数のわずか0.3%である268件に留まったため、対象地域全体に及ぼす影響はわずかといえる。このように明確な要因を特定できない日は8日あった。考えられる要因のさらなる精査と、多方面への調査の展開が今後の研究課題として残っている。

## 5. 結論

本研究では、神姫バスグループが提供する経路検索システム「神姫バスNavi」のログデータから突発的な検索数があった日を抽出し、その検索行動の要因を考察した。交通行動の実態を把握する方法としては現在もさまざまな手法が検討されているが、本研究では状態空間モデルを用いた不規則変動成分の抽出を行った。不規則成分から抽出された突発的検索行動の発生要因を、ウェブと過去の新聞を用いて考察した。状態空間モデルの適用により、突発日は対象の1年間のうち18日抽出された。この18日について要因の調査を行ったところ、総検索件数は広範囲に影響を及ぼす天候の悪化と大規模な催し物から影響を受けていることがわかった。

本研究では突発的検索行動が抽出された18日のうち8日において明確な要因を特定できなかった。これらの課題に対しては以下の3通りの解決方法が考えられる。1

つ目は本研究で用いた状態空間モデルを拡張することである。状態空間モデルには多様なバリエーションを組み合わせることができる特徴があるため、考えられる要因をモデルに多様に組み込み、全てのモデルの組み合わせを試行することが有効である可能性がある。2つ目は、本研究で用いた突発的検索行動の要因の調査方法を精査すると同時に、多方面に展開し多様な方法を用いた調査を行うことである。3つ目は、分析対象範囲の大きさを変えて本分析手法を再度適用することである。そもそも分析対象範囲の大小によって検出できる要因の規模は異なると考えられる。狭域的な要因を明らかにすることが、広域的な突発的移動需要の要因解明につながる可能性がある。

## 参考文献

- 1) 伊藤昌毅, 見生元気, 川村尚生, 菅原一孔: 乗換案内サービス利用履歴から探る公共交通の利用状況, 第9回電子情報通信学会ヒューマンプロブ研究会, CD-ROM, 2013.
- 2) 坪内孝太, 下坂正倫, 小西達也, 丸山三喜也, 山下達雄: 乗換案内データを用いた未来の混雑予測の研究, 人工知能学会全国大会講演集, Vol.31, 4pages, 2017.
- 3) ヤフー株式会社: Yahoo!乗換案内のデータで未来の混雑を予測する, <https://about.yahoo.co.jp/info/bigdata/special/2017/04>
- 4) ヤフー株式会社: Yahoo!乗換案内, 約4,000万人のビッグデータを活用し, 当日から5日分の電車内や駅の混雑度を10分単位で予測する新機能「異常混雑予報の提供を開始」, <https://about.yahoo.co.jp/pr/release/2018/02/19a>
- 5) 古川ゆり, 桑野将司, 秋元美穂奈, 菅原一孔: 経路検索履歴データを用いた事前検索時間の分析, 交通工学論文集, Vol.5, No.2, pp.A\_184-A\_192, 2019.

?

## EXTRACTION OF FLUCTUATIONS IN LOG DATA IN THE PUBLIC TRANSIT NAVIGATOR SYSTEM BY STATE-SPACE MODEL

Terutomo ODAJIMA, Masashi KUWANO, Taku MORIYAMA and Mio HOSOE