

経路検索データを用いた公共交通の需要予測手法

2022年1月

細江 美欧

目次

第 1 章 序論	1
1.1 研究の背景	2
1.1.1 ICT 化がもたらした移動時における情報收受方法の変化	2
1.1.2 交通需要予測におけるビッグデータ時代の到来	5
1.1.3 地方部における需要応答型移動サービスの実態	7
1.1.4 新たな交通システムの進展	8
1.1.5 交通需要予測における経路検索データの活用可能性	10
1.1.6 SP データとしての経路検索データによる交通需要予測の可能性	12
1.2 研究の目的	13
1.3 経路検索行動と実行動を関連づけるための考え方	13
1.4 本論文の構成	15
第 1 章 参考文献	19
第 2 章 既往研究のレビューと研究の位置付け	23
2.1 RP ビッグデータを用いた需要予測に関する既往研究	24
2.2 経路検索データに関する既往研究	25
2.2.1 経路検索データの特性分析に関する既往研究	26
2.2.2 経路検索データに基づく突発的需要の検知に関する既往研究	27
2.2.3 経路検索者数と公共交通利用者数の関連性に関する既往研究	28
2.3 研究の位置付け	28
第 2 章 参考文献	31
第 3 章 経路検索データからの検索パターンの抽出	35
3.1 はじめに	36
3.2 分析対象とする地域と交通機関	38
3.3 経路検索システム「NAVITIME」の検索履歴データ	39
3.4 非負値テンソル因子分解	45
3.5 ことばを対象とした検索パターンの抽出	47
3.5.1 属性別の特徴行列の結果と考察	48
3.5.2 コアテンソルの結果と考察	51
3.6 まとめ	53

3.6.1	3章の研究成果と次章への展開	53
3.6.2	3章における今後の課題	54
第3章	参考文献	56
第4章	経路検索システム利用者の移動目的の推察	59
4.1	はじめに	60
4.2	経路検索数の分布の特徴を踏まえたモデル構築	61
4.3	平均経路検索数に影響を及ぼす駅周辺環境要因の候補	62
4.3	平均経路検索数に影響を及ぼす駅周辺環境要因の候補	62
4.3.1	駅周辺環境に関するデータ	62
4.3.2	Adaptive Lasso を用いた変数選択	64
4.4	Bayesian Adaptive Lasso Tobit 分位点回帰モデル	65
4.5	モデルの推定結果と考察	70
4.6	まとめ	72
4.6.1	4章の研究成果と次章への展開	72
4.6.2	4章における今後の課題	73
第4章	参考文献	74
第5章	経路検索数と交通系 IC カード利用数の関係性分析	75
5.1	はじめに	76
5.2	使用データ	78
5.2.1	経路検索システム「NAVITIME」の検索履歴データ	78
5.2.2	交通系 IC カード「IruCa」の乗降履歴データ	79
5.3	経路検索数と交通系 IC カード利用数の相関分析	821
5.4	本研究のアプローチ	82
5.5	経路検索数と交通系 IC カード利用数の成分分解モデルの開発	84
5.5.1	多変量状態空間モデルの一般式	84
5.5.2	計算アルゴリズム	86
5.5.3	モデルの設計	88
5.6	2変量状態空間モデルによる成分分解結果	91
5.6.1	経路検索数と交通系 IC カード利用数の各成分の特徴	91
5.6.2	経路検索数と交通系 IC カード利用数の誤差成分と突発的需要の同等性	97
5.6.3	分散共分散行列の考察	99

5.6.4	経路検索数と交通系 IC カード利用数の関係性のまとめ	99
5.7	経路検索数と交通系 IC カード利用数の誤差成分の関係性	100
5.8	経路検索数の予測値に基づく交通系 IC カード利用数の予測モデルの構築	102
5.9	現況再現性の確認	104
5.10	他モデルと提案手法による現況再現性の比較	106
5.11	まとめ	111
5.11.1	5章の研究成果と次章への展開	111
5.11.2	5章における今後の課題	112
第5章	参考文献	114
第6章	経路検索数に基づく交通系 IC カード利用数予測モデルの開発	115
6.1	はじめに	116
6.2	使用データ	118
6.2.1	経路検索システム「NAVITIME」の検索履歴データ	118
6.2.2	交通系 IC カード「IruCa」の乗降履歴データ	119
6.2.3	経路検索数と交通系 IC カード利用数の特徴把握	121
6.3	経路検索数の予測値に基づく交通系 IC カード利用数の予測モデルの再構築	122
6.4	回帰係数の時間的・空間的安定性の検証	124
6.4.1	回帰係数の時間的安定性の検証	124
6.4.2	回帰係数の空間的安定性の検証	129
6.5	事前収集時間別予測精度の検証	133
6.5.1	事前収集時間別の予測精度の確認方法	133
6.5.2	事前収集時間別経路検索数の推移	135
6.5.3	路線ごとの事前収集時間別予測精度の検証	137
6.6	まとめ	138
6.6.1	6章の研究成果	133
6.6.2	6章における今後の課題	135
第6章	参考文献	140
第7章	結論	141
7.1	本研究のまとめ	142
7.2	本研究の成果	145
7.2.1	交通計画における経路検索データの有用性の実証	135

7.2.2	経路検索データを用いた公共交通の需要変動の予見	137
7.2.3	データ間の潜在的関係性の抽出	135
7.2.4	地方部のビッグデータの活用可能性	137
7.3	今後の展望	149

謝辞

目次

図 1.1	本論文の構成	18
図 3.1	第 3 章の分析フロー	39
図 3.2	3 次数データの場合のスライス順序の候補	39
図 3.3	ことでん路線図 (参照: ことでんグループ)	39
図 3.4	検索条件入力画面	40
図 3.5	検索結果表示画面	40
図 3.6	時刻指定方法別検索割合	41
図 3.7	月別総経路検索数の推移	42
図 3.8	曜日別平均経路検索数	42
図 3.9	乗車駅別平均経路検索数 (上位 10 駅)	43
図 3.10	降車駅別平均経路検索数 (上位 10 駅)	43
図 3.11	CP 分解とタッカー分解の違い (左: CP 分解, 右: タッカー分解)	46
図 3.12	タッカー分解概念図 (3 次テンソル)	47
図 3.13	曜日の特徴行列の結果	48
図 3.14	乗車駅の特徴行列の結果	49
図 3.15	各乗車駅因子を構成する駅の位置関係	50
図 3.16	降車駅の特徴行列の結果	51
図 3.17	各降車駅因子を構成する駅の位置関係	51
図 3.18	コアテンソルの結果	52
図 4.1	第 4 章の分析フロー	60
図 4.2	Tobit 分位点回帰モデルの損失関数	66
図 4.3	曜日別平均経路検索数 (3.3 節図 3.8 の再掲)	72
図 5.1	第 5 章の分析フロー	77
図 5.2	日別経路検索数の推移	78
図 5.3	曜日別平均経路検索数	79
図 5.4	経路検索数のコレログラム	79
図 5.5	日別交通系 IC カード利用数の推移	80
図 5.6	曜日別平均カード利用数	80
図 5.7	カード利用数のコレログラム	81
図 5.8	平日と休日における経路検索数と交通系 IC カード利用数の散布図	82
図 5.9	経路検索データと交通系 IC カードデータの関係	82

図 5.10	本研究のアプローチ	84
図 5.11	日別経路検索数（左）と交通系 IC カード利用数（右）の各成分の推移	92
図 5.12	誤差成分の散布図	101
図 5.13	提案モデルの予測フロー	103
図 5.14	経路検索数の予測値に基づく交通系 IC カード利用数の予測値と観測値	105
図 5.15	交通系 IC カード利用数の予測値と観測値（モデル 1）	107
図 5.16	交通系 IC カード利用数の予測値と観測値（モデル 2）	108
図 5.17	交通系 IC カード利用数の予測値と観測値（モデル 3）	108
図 5.18	交通系 IC カード利用数の予測値と観測値（モデル 4）	108
図 5.19	交通系 IC カード利用数の予測値と観測値（提案手法）	109
図 5.20	経路検索数に基づく交通系 IC カード利用数の予測値の推移	110
図 5.21	交通系 IC カード利用数のみによる予測値の推移	111
図 6.1	第 6 章の分析フロー	117
図 6.2	日別経路検索数の推移	118
図 6.3	曜日別平均経路検索数	119
図 6.4	経路検索数のコレログラム	119
図 6.5	日別交通系 IC カード利用数の推移	120
図 6.6	曜日別平均カード利用数	121
図 6.7	カード利用数のコレログラム	121
図 6.8	平日と休日における経路検索数と交通系 IC カード利用数の散布図	122
図 6.9	回帰係数の推移	126
図 6.10	琴平線の回帰係数の推移	131
図 6.11	長尾線の回帰係数の推移	131
図 6.12	志度線の回帰係数の推移	131
図 6.13	各路線の回帰係数の時間的安全性	132
図 6.14	事前収集時間別経路検索数の概念図	134
図 6.15	提案モデルの予測フロー（再掲）	135
図 6.16	琴平線の事前収集時間別経路検索数の推移	136
図 6.17	長尾線の事前収集時間別経路検索数の推移	136
図 6.18	志度線の事前収集時間別経路検索数の推移	136
図 7.1	提案モデルの予測フロー（再掲）	145

表目次

表 1.1	経路検索データと交通系 IC カードデータが表現可能な「行動」と「需要」	15
表 3.1	経路検索データに記録される情報	44
表 3.2	乗車駅・降車駅別平均経路検索数 (2,652 組合せ中の上位 40 組合せ)	44
表 4.1	各施設の内容	63
表 4.2	各施設の内容	64
表 4.3	推定結果	71
表 5.1	経路検索データと交通系 IC カードデータが表現可能な「行動」と「需要」(再掲)	77
表 5.2	2 変量状態空間モデルのパラメータ推定結果	91
表 5.3	経路検索数の曜日成分のパラメータ推定結果 (表 5.2 より抜粋)	94
表 5.4	交通系カード利用数の曜日成分のパラメータ推定結果 (表 5.2 より抜粋)	96
表 5.5	経路検索数の誤差成分に基づく特異日とその関連事象	98
表 5.6	交通系 IC カード利用数の誤差成分に基づく特異日とその関連事象	98
表 5.7	観測誤差と過程誤差の分散共分散 (表 5.2 より誤差分散のみ抜粋)	99
表 5.8	重み付き回帰分析の推定結果	101
表 5.9	各モデルの RMSE	109
表 6.1	回帰係数の時間的安定性	128
表 6.2	回帰係数の値の差による交通系 IC カード利用数予測への影響	129
表 6.3	路線別重み付き回帰モデルの推定結果と系全体の重み付き回帰モデルの推定結果	130
表 6.4	各路線の事前収集時間別相関係数	137
表 6.5	路線ごとの事前収集時間別 RMSE	137
表 7.1	経路検索データと交通系 IC カードデータが表現可能な「行動」と「需要」(再掲)	143

序論

〈第 1 章の概要〉

人口減少社会における効率的かつ効果的な交通サービスの提供には、日々変動する交通需要の把握が必要である。これまで、交通需要は平常時の予測に注力され、突発的な交通需要の予測に関する研究は少ない。また、交通需要の変動に関する研究は、天候やイベントといった変動要因の解明が目的とされてきた。しかし、交通需要の変動要因が解明できたとしても、予測においては変動要因となる天候変化やイベント発生を事前に把握する必要があり、実用的ではない。

そこで、本研究では突発的な交通需要の変動要因を追求するのではなく、さまざまな要因で発生する需要の変動を経路検索システムに蓄積される検索履歴データで捕捉する。これまで交通需要予測に利用されているビッグデータのほとんどが RP データであるのに対し、経路検索データは将来の移動意向に関する SP データであり、未来の交通需要変動を予見できる可能性がある。RP データである過去の公共交通利用者数と、さまざまな要因による交通需要変動の代理変数として SP データである経路検索数のみを用いて、突発的需要を含めた日々変動する交通需要を予測可能とする方法論の開発が本論文の目的である。

1.1 研究の背景

1.1.1 ICT 化がもたらした移動時における情報收受方法の変化

自動車マイカーとして普及し始めた頃、初めての場所に行くには、予め地図帳で経路を確認し、どこで何を目印にして目的地まで辿り着くかをシミュレーションする必要があった。そして、助手席に座った人はまさにドライバーの助手となり、地図帳や道路標識を見ながらナビゲーションしていた。そんな時代に終止符を打ったのがカーナビゲーションシステム（以降、カーナビ）の登場である。1980年代から徐々に普及をはじめたカーナビは、1990年にGPS（Global Positioning System）が搭載され、自分の位置がわからない場合でも地図上に自車位置を正確に表示し、現在地から目的地までの経路を検索できるようになった（総務省，2014）。さらには、見やすい情報表示と音声案内によって、車線変更や交差点での分岐も安心して快適に行えるようになり、初めて通る道でもまるで通り慣れた道のように自動車での走行を実現させた。その後、1996年にはVICS（道路交通情報通信システム：Vehicle Information and Communication System）と呼ばれるサービスによって、道路の混雑状況が車内でも把握できるようになり、カーナビは単なる最短経路案内ではなく、渋滞を考慮した経路検索や到着時間の算出も可能とした。2002年には通信モジュールを搭載したカーナビが登場したことで、地図データや店舗データの更新が可能となり、常に最新の情報をカーナビで活用できるようになった（総務省，2014）。このような自動車におけるICT（情報通信技術：Information and Communication Technology）の活用によって、運転者の状況や目的に応じた経路情報を收受できるようになったことで、自動車移動にまつわる負担が軽減し、時間的・空間的制約のない移動手段として自動車の利便性は高まった。

自動車が利用しやすい環境は、日本のモータリゼーション化を加速させ、これに付随して交通渋滞を激化させた。交通渋滞の増大は道路交通情報に対する需要を高めるとともに、正確な情報を可及的速やかに提供することを要求させた。ここでも、ICTは大いに活用され、道路交通状況について、それまでの警察官の目視による測定から、車両感知器や光ビーコン、交通流監視カメラを用いた測定に転換された。そして、それら装置から収集される情報を分析し、伝達する交通管制システムが体系化した（福井，1999；織田，2001）。道路交通におけるICTの進展は情報の収集方法だけでなく、提供方法にも寄与し、先に述べたVICSや所要時間表示板などの新たなサービスを生んだ。しかし、新たな情報提供サービスが複数創出され、それらが個別に発信するようになると、それらの相違がしばしば運転者

の混乱を招いた。特に、高速道路では各サービスが提供する情報のずれによって、運転者はどの道路情報が正しいのか判断できず、情報の信頼性が欠如した。そこで、全国的高速道路管理会社が道路交通情報の収集・集約を一元化し、情報の整合性の確保が図られた（安達ら，2002）。このような交通管制システムの高度化に向けた動きにより、ITS（高度道路交通システム：Intelligent Transport Systems）という新たな道路交通システムが策定された。ITS とは ICT を活用し、人と道路と車両とを一体のシステムとして情報で結び、道路交通が抱える交通事故や渋滞などのさまざまな課題を解決することを目的としたシステムである。これまでに触れてきたカーナビや VICS は ITS を構成する主要なシステムといえる。その他に ETC（電子料金収受システム：Electronic Toll Collection System）も代表的システムとして包含される。ETC は 2014 年に ETC2.0 サービスとして進化し、料金収受のほかに時々刻々と変化する道路状況を、運転者の必要なタイミングで提供することを可能とした。例えば、高速道路や有料道路において渋滞回避支援、安全運転支援、災害時の支援が受けられる（国土交通省，2014）。今後、ITS の推進によって道路交通を取り巻く課題が解決され、これまで以上に運転者のニーズに即した信頼性の高い情報が提供されるようになれば、自動車による移動の利便性がより一層高まると期待される。

一方で、公共交通は、自動車に比べて使い勝手の悪さが問題であった。自動車はいつでも好きなときに外出できるという「即時性」、大きな荷物を運べるという「収容性」、Door to door で移動できるという「利便性」、車内では自分だけというプライベート空間の「専有性」など、公共交通に比べて優れている。一方で、20 世紀の公共交通の移動環境は、人々は紙ベースの時刻表で出発時刻と到着時刻を確認し、切符を購入して、改札口で駅員に切符を見せなくてはならないなど、公共交通利用者にとって面倒な手続きが多い「煩雑性」がその利用の障害となっていた。

近年は、駅員の人員削減の観点から、駅務の自動化を目的に自動改札機が導入されるようになったことで、乗降車時の手続きがスムーズになった。さらに、交通系 IC カードの導入によって、切符を事前に購入する必要もなくなり、煩わしさが軽減された。交通系 IC カードは 2001 年に JR 東日本によって「Suica」が販売されたのを皮切りに、都市圏を中心に急速に普及することになる（総務省，2015）。現在では、全国の公共交通機関で相互利用可能な交通系 IC カード（10 カード）が 10 種類存在する（国土交通省，2019）。また、地域の鉄道会社やバス会社が提供する地域独自の交通系 IC カードも存在し、全国で約 40 種類ある（国土交通省，2015）。地域独自の交通系 IC カードは使用できる地域が限定されている不便さがあるものの、地元の商業施設と連携した割引サービスが充実しており、地域住民の保有率は高い。さらに、近年ではモバイル Suica やモバイル PASMO のように交通系 IC カードをスマートフォンに取り込むことができようになり、カードタイプと同様に公共交通を利用可能である。

交通系 IC カードと並んで、公共交通の利便性を飛躍的に高めたのが経路検索システムである。経路検索システムとは出発地や目的地、公共交通を利用したい日時を入力することで、最適な経路や運賃、所要時間などが掲示される Web サービスである。例えば、民間企業が提供する「NAVITIME」や「ジョルダン」など全国の主公共交通網を網羅したものや、交通事業者が提供する「神姫バス Navi」など事業者が運営する交通機関に関して経路情報を提供するものがある。また、鳥取県の「BUSNET」や高知県の「アクセスこうち」などの地域独自で運行するコミュニティバスも網羅した地域特化型の経路検索システムがある。経路検索システムは、カーナビが自動車の利便性を向上させたことに匹敵するほど、公共交通の利便性を高め、情報収受の手段として一般的になっている。

その利用の定着化の背景には、インターネットの普及がある。我が国のインターネットの利用は 1990 年代前半から始まったといわれている。その後、インターネットは一般家庭においても急速に普及し、日常生活の在り方をさまざまな形で変えてきたが、とりわけ、情報収集における手段の変化は大きい。かつて情報は新聞や雑誌、書籍などの紙媒体、またはテレビやラジオから主に入手されることが多く、インターネットを利用する者は少なかった。しかし、今やインターネットを利用し、情報を収集することは当たり前となっている。また、インターネットを利用する際の端末は、普及初期段階においてはパソコンが中心であったが、近年はインターネットに接続する端末はパソコンからスマートフォンに利用の中心がシフトしている。スマートフォンはパソコンとは異なり、簡単に持ち歩くことが可能である。そのため、「何か知りたい」とき、手持ちの端末からインターネットを利用し、知りたいことを瞬時に検索することはありふれた日常の一部となってきた。これはパソコンを使い自宅でゆっくり調べる情報収集とは全く異なる。

特に、経路検索システムの恩恵が大きい状況は、行き先や時間が普段とは異なるときである。どのような経路で行けば効率良く目的地に辿り着けるのか、どの時間帯に公共交通が運行されているのかを移動前あるいは移動中に調べられることは公共交通の利便性を大きく高めた。経路検索システムを利用することで、紙ベースの時刻表で出発時刻を確認することも、路線図で経路を確認する必要もなくなり、公共交通利用時の煩雑性は軽減された。特に、長距離移動では乗換えが頻繁に起こるため、その検索を紙ベースの媒体で揃えてはしばしば厄介である。一方、経路検索システムではこれを一瞬で行うことができる。

近年の経路検索システムには車両の走行位置が確認できるロケーションシステムと連携したものや、遅延・運休などの運行状況も同時に掲示されるものがある。自動車での移動時に VICS や ETC2.0 によってリアルタイムの渋滞情報が提供されるのと同様に、経路検索システムでは予め定められた時刻表などの静的情報だけでなく、時間経過とともに更新される動的情報も提供されるようになった。

1.1.2 交通需要予測におけるビッグデータ時代の到来

ICT の目覚ましい進展は、学術研究においても大きな変化をもたらした。従前、交通分野の研究の難しさには交通行動の観測の問題があった。交通行動は人が何らかの意図を持って行う活動に付随して、時空間上を移動することで発生する。つまり、人々の行動を捉えるためには、活動の発生と活動目的達成までの時空間上での移動を観測することが不可欠である。そのため、地域全体を対象とした長期間に及ぶ観測が必要となるが、ICT 発展以前の技術では都市や国土規模での長期観測は現実には困難であった。したがって、時間的あるいは空間的に観測範囲を限定することが求められ、その上で、分析目的に応じたデータの収集とモデル構築が行われてきた。特に、生産年齢人口が社会の大勢を占めていた時代は、増加する交通需要に対応するために、時空間的範囲を限定した観測に基づく需要予測モデルが開発された。その最も標準的なモデルに四段階推定法がある。四段階推定法は 1950 年代にアメリカで開発され、その実用性の高さから日本はもとより世界中で利用されてきた。交通行動の背後にある複雑な意思決定メカニズムを簡略に表現し、トリップベースの交通需要の全体的動向を捉える手法としての有効性は高い。その一方で、四段階推定法に内在する諸問題への指摘は多く、これまでさまざまな改良案が検討、提案されている（例えば、森地ら、1985；北村、1996）。四段階推定法に代替する予測モデルの開発も注力されてきたが（例えば、飯田ら、2000；金森ら、2009）、分析概念の単純さから、実務上この手法に代替する予測モデルは浸透しておらず、いまだ活用され続けているのが現状である。しかし、四段階推定法は時空間的に観測範囲を制限したパーソントリップ調査に基づいて解析を行うため、代表的な 1 日の平均的な交通需要しか予測できない。つまり、何らかの要因によって交通需要が変動したときには役に立たない。

公共交通の交通需要は交通事故や天候変化などの外的要因と、急な予定変更や混雑回避などの人的内的要因によって、交通需要はそもそも経時的に変化する。交通行動の主体である人々を取り巻く環境はあらゆる面で常に変化しており、交通需要において変動の要素は切り離すことはできない。ここで、「変動」とは確率過程的な現象の移ろいと定義し、偶発的な事象による現象の移ろいや影響する事象の特定が困難な現象の移ろいに加え、周期をもって繰り返す移ろいも含むとする（北村、2003）。鉄道や路線バスのダイヤや経路などの運行計画は、日々変更されるものではないため、定常性を仮定した交通需要予測モデルに基づいて運用を計画することが望ましい。しかし、後述するように、地方都市や都市郊外部では少子高齢化、人口減少による公共交通利用者数の減少とそれに伴う事業者の収益低下によって、路線バスは廃止され、需要に応じて運行が決定するデマンド型公共交通や相乗り交通の

導入が進んでいる（1.1.3 項で説明）。また、運用効率性の向上だけでなく、利用者ニーズに応じた新たな複合型交通サービスとして MaaS（Mobility as a Service）の導入も進んでいる（1.1.4 項で説明）。常に同じ経路、同じダイヤでは運行しない交通サービスの運用のためには、交通需要の変動の考慮が不可欠である。そのような変動を広範囲かつ連続的に捉えたデータの観測を可能としたのが ICT である。

ICT の発展を節目に、交通分野に限らずあらゆる研究分野において、それまでに入手できなかった時間的、空間的に詳細で、大量なデータ、いわゆるビッグデータが入手できるようになった。ビッグデータの定義には多様な見解があり（Milne and Watling, 2019）、それらには標準的で確立された見解はないが、多くの既往研究では Volume（データ量）、Variety（データの種類）、Velocity（データの発生、処理速度）の 3 要素を持つデータがビッグデータとされている（Torrecilla and Romo, 2018; Emani et al., 2015）。2000 年代はビッグデータ時代といわれ、リアルタイムで生成・収集・蓄積されるビッグデータを分析することで、これまで捉えることができなかった事象を認識し、新たな知見を創出できると期待されている。交通分野においては、プローブカーデータ（例えば、Jenelius and Koutsopoulos, 2014; Yang et al., 2017）、GPS による位置情報データ（例えば、Hofleitner et al., 2012; Uij and Shankari, 2015; Simoncini et al., 2018）、交通系 IC カードの乗降履歴データ（例えば、Alsker et al., 2013; Zhong et al., 2015; Wang et al., 2018）などの交通ビッグデータがある。

しかし、前述のような交通ビッグデータは、収集されたデータの利用目的と分析目的が相違していることが多い。そのため、分析者は自身の意図とは異なる形式で収集された莫大なデータを整理し、分析に用いるデータ項目を選択する必要がある。すなわち、分析者はデータの種類や特徴を理解した上で、分析目的に応じた必要なデータを取捨選択する能力が求められた。また、交通ビッグデータが蓄積され始めた当初は、それらが表現する交通行動の特性が明らかになっていないことが多かったため、ビッグデータに適した分析手法を検討するとともに、交通ビッグデータから基礎的知見を得ることに注力されていた（例えば、永井ら, 2011 ; 西内と轟, 2012 ; 北野ら, 2008 ; Ma et al., 2013）。その後、交通ビッグデータに関する初歩的な研究の蓄積が進むと、交通ビッグデータの時系列性を活かし、交通需要の変動に関する研究の蓄積が始まった（例えば、中沢と佐々木, 2011 ; 鳥海ら, 2020）。交通ビッグデータには大勢の人々の移動実態が長期的かつ継続的に記録されるため、交通需要の日々の変動を捉えられる可能性があり、従来のアンケート調査による交通需要予測の代替や補完に役立つ点に注目が高まっている。

1.1.3 地方部における需要応答型移動サービスの実態

前述したように、路線バスを中心に公共交通は決められたルートを決められた時間に走行し、公共交通利用者がそれに合わせるという定時定路線型の運行が一般的であった。しかし、我が国の地方部においては小規模需要によって限界を迎えた従来型の公共交通に対して新たなサービスが求められている。そのひとつとして、注目を集めているのが需要応答型交通のオンデマンド交通である。国土交通省の交通政策基本計画（2014年度～2020年度）ではオンデマンド交通を2020年度までの導入自治体数を700市町村まで増やすことを目標に、全国で導入が推進されてきた（国土交通省，2013）。

オンデマンド交通は利用者が希望する時間帯、乗降場所に応える運行スタイルで、利用者の需要を集約した形で運行する乗合交通手段である。オンデマンド交通には、

- (1) 固定路線は定期運行し、路線の途中に設定された迂回ルート部分は利用者の要求があった場合にのみ迂回運行するタイプ、
- (2) 定時定路線の一部がデマンド区間となり、その区間は利用者の要求があった場合にのみ運行するタイプ、
- (3) 設定された路線のダイヤが利用者の要求によって変動するタイプ、
- (4) 一定の区域をデマンドエリアとし、運行区間そのものが利用者の要求によって、その都度設定されるタイプ。

などがある（鈴木，2012）。導入メリットとしては、個々人の要求に合わせて運行されることから、利用者にとっては利便性が高い移動手段な点である。また、利用者がいなければ運行されないため、交通事業者にとっては経費削減に繋がり、財政負担を軽減できる可能性が高い。一方、デメリットは、要求がなければ運行されない反面、1人でも予約があれば運行せざるを得ず、かえってコストがかかってしまう点が挙げられる。また、利用者の複数の要求を同時に満たすために、経路や所要時間が一定とはならず、定時性について利用者にはある程度の許容が求められる。これらオンデマンド交通の得失は、「いつ」「どこから」「どこへ」「どれぐらい」の需要が発生するかという、交通需要の変動に起因する。

需要応答型交通の本格的な導入に向けて、法整備が進むとともに、全国各地で実証実験が行われているが、構想当初ほどの成果を得られている自治体はほとんどない。現実には、一人当たりの輸送費が定時定路線型交通よりも上昇すること、予約が面倒で結局利用されないこと、一定数の利用者がいなければコスト面から実際には運行できないことなどから、採算性がとれず失敗に終わっている。地方

自治体がオンデマンド交通を採用する最大の理由は小規模需要への対応にあるが、これが足枷となっているケースがほとんどであり、未だ成熟した交通サービスにはいたっていない。

効率的な運行の検討や需要に合わせた車両配置の最適化のためには、

- (1) 低乗合率：利用者が一人の場合や片道回送の場合に非効率的運行になる、
- (2) 不定時運行の場合の車両数の確保：散在的に複数発生する需要に対応するために相当数の車両確保が必要である、
- (3) 時間制約が大きい活動の場合の利用率の低下：予約時間に制限がある場合は利用できないことがある、
- (4) 利用者側の心理的抵抗：予約が面倒、利用者が自分だけだと利用しづらい、
- (5) 財政的負担の増加：予約システムなどの運営費用を含むコストが大きい、
- (6) 限定された利用者：利用登録制の場合は観光客や外国人は使えない

のような課題を解決しなければならない（鈴木，2012）。ここで、(1)，(2)は、日々変動する交通需要に関連する問題であり、高い精度で交通需要の変動を予測することが解決の糸口となる。一方、(3)，(4)，(5)，(6)は予約システムの問題である。経路、出発・到着時刻、所要時間、混雑状況などを確認した上で、予約を行うという一連の行動を誰でも簡単に利用できるシステムを1つのプラットフォームで実現しようとするれば、すでに普及が進んでいる経路検索システムが基盤となると考えることは自然であろう。すなわち、小規模需要における公共交通体系を議論するためのキーワードは、「交通需要の変動の把握」と「経路検索システム」と考えることができる。

1.1.4 新たな交通システムの進展

地方部を中心に、公共交通機関では交通需要が大幅に減少し、交通サービスの水平的な展開に勤しんできた交通事業は変革の局面に立たされ、公共交通のあり方が問われている。一方で、各地の公共交通事業は存続が危ぶまれるような深刻な経営悪化に陥っており、行政の補助金がなければ事業が成立していない。さらに、COVID-19の拡大によって在宅勤務が定着化しつつあることで、感染拡大以前の交通需要には戻らない可能性も指摘されている。

人々のライフスタイルや社会構造が変化する中で、運行本数を確保し、持続的に運営していくために、交通事業者はどのような手立てを講ずるべきかを考えなければならず、従前より交通政策が目指してきた人々のニーズに即した交通サービスの提供は、これまでになく重要度が増している。

1.1.3 項で示したオンデマンド交通の延長に、MaaS がある。MaaS は持続可能な生活ができる地域を実現するために、移動に関するさまざまなサービスを組み合わせて、個人のニーズに合わせた移動サービスを提案しようとする概念である（楠田，2021）。MaaS のプラットフォームは究極の経路検索システムといっても過言ではなく、移動の希望条件を入力し、経路を検索する機能を根幹に、車両の運行状況を確認する機能、交通機関周辺施設の利用状況や混雑状況を確認する機能、一連の活動にかかる費用の決済機能などあらゆる機能や情報を付加したものと想定される。

我が国では 2018 年に「未来投資戦略 2018」において MaaS が国家の主力プロジェクトとして重点施策に位置付けられた。これを受け、2019 年には経済産業省と国土交通省が連携し、新しいモビリティサービスの社会実装を通じた移動課題の解決、および地域活性化に挑戦する地域や企業を応援するプロジェクトとして「スマートモビリティチャレンジ」を始めた（日高ら，2020）。その試験事業として全国 28 都市を対象地域に選定し、MaaS のモデル構築を図っている。さらには、MaaS を普及・推進していく団体も続々と設立されており、MaaS 導入に向けた動きが活発である。

例えば、香川県高松市では *scheme verge* 株式会社やことでんグループを中心に「高松 MaaS」導入に向けて取り組んでいる。高松市ではことでんグループが運営する高松琴平電気鉄道やことでんバスが公共交通の基幹的役割を担っており、そのネットワークは高松都市圏に広がっている。さらに、高松市が独自で運営する放置自転車を活用したレンタサイクルが市民の移動手段として定着しており、MaaS の素地ある先進都市として注目度が高い。2020 年度には先に述べた試験事業の対象地域に選出され、実証実験（国土交通省，2020；国土交通省，2021）の結果をもとに地域経済への影響や制度的課題を整理し、ビジネス環境整備が進められている。

MaaS の本格的展開に向けては、まちづくりと新しい移動サービスとの連携がこれまで以上に必要になると考えられる（牧村，2021）。特に、交通以外の産業と連携を図ることで、移動に付加価値を付帯させることが重要である。そのために、移動目的となる産業との連携が見直され始めている。公共交通やタクシーなどを利用する乗客の多くは、移動サービスそのものを利用することを目的に乗車しているわけではない。交通行動のほとんどは派生需要であり、人々は何らかの移動目的に沿って、自身に合った移動手段を選択している。現在、移動サービスと移動の目的地となるような娯楽施設や商業施設との連携は、顧客管理や分析が難しく、十分とはいえない。しかし、ICT のさらなる進展と活用が進んでいけば、これまで以上にあらゆる情報がビッグデータとして蓄積され、データの共有・統合、そして解析によってシームレスなサービスが検討可能になることが期待される。

1.1.5 交通需要予測における経路検索データの活用可能性

現存の公共交通は定常性を仮定した交通需要予測に基づいて運行計画が設計されていることから、突発的に交通需要が変化したときに、その変化に対応しきれない。例えば、ある駅近くに立地する施設でイベントが開催されたときに、ある駅に向かう交通需要が平常時より増大することが考えられる。このような、いつもとは違う交通需要の発生を突発的交通需要と呼び、これが需要変動である。突発的な交通需要によって利用者が増加した場合、ある駅との間を結ぶ別の駅で電車に乗れなかった人が溢れかえるような事態が起こる。突発的な交通需要の変化に対応できないことは、車内の混雑やダイヤの乱れ、定員超過による積み残しを引き起こし、公共交通利用者の利便性や満足度の低下に影響する負の大きな要因となる。これを回避するためには、突発的な交通需要の発生を事前に予測することが必要である。

突発的な交通需要の予測は、先に述べたオンデマンド交通や相乗り交通の運行においても、また、近い将来訪れる MaaS による統合的な交通サービスの提供においても必須といえる。オンデマンド交通や MaaS のような需要に応じた供給を目指す交通サービスにおいては、将来、どの程度、交通需要が発生するかを予測し、適切な供給量の車両確保や配置を事前に行う必要がある。特に、急激な需要の増減が生じた場合に、効率的な運用ができなければ、利用者の利便性は大きく低下し、交通サービスとしての信頼性を失うとともに、事業者にとっては収益機会の損失に繋がる。

定常的な交通需要であれば伝統的な交通需要予測モデルや交通事業者の経験則によって予測可能であるが、突発的な交通需要は従前モデルでは予測できない。そのため、突発的な交通需要の発生に影響する変数に関するデータを収集し、それらを考慮した交通需要予測モデルの構築が必要であるが、そもそもどのような事象が突発的な交通需要の発生に影響するかを前もって知ることは難しい。既往研究においては、天候悪化 (Khattak and De Palma, 1997; Cools et al., 2010; Ohler et al., 2017)、自然災害 (Abad and Fillone, 2000)、イベント開催 (Pereira et al, 2015) などが原因となり、普段とは異なる交通需要が発生することが指摘されている。しかし、例えば、イベントであれば、お祭り、コンサート、講演会、フリーマーケットなど各地で大小さまざまな規模で開催されており、分析対象とする地域全体のイベント情報をすべて得ることはできない。また、どのくらいの規模、どのような種類のイベント情報をモデルに投入すべきかの検討も難しい。天候悪化も同様に、どの範囲の、どの程度悪化が見込まれる天候情報を収集すべきかを予め設定することは困難である。したがって、イベント情報や天候情報によって、事象の発生後に突発的な交通需要と関連づけることはできても、それらを用いた事前予測は極めて難しく、実用的ではない。

イベント発生による急な乗客増加に伴う運行遅れや、天候悪化による運休が生じた場合あるいは生じる可能性のある場合に、公共交通は定時性が保持されない。このとき、公共交通利用者はいつ車両が来るのか、そもそも自身の移動手段は確保されるかを不安に思う。この不安を解消させる手段に、先に述べた経路検索システムがある。経路検索システムを用いることによって、活動目的を達成するための経路を検索し、普段とは違う状況における移動に対する不安を解消できる。つまり、経路検索システムに蓄積される検索履歴データ（以降、経路検索データ）には平常時とは異なる事態で突発的に発生しうる交通需要が反映されていると考えられる。言い換えると、経路検索データはあらゆる要因が関係しあった結果に現れる突発的な交通需要の代理変数になる可能性がある。

ここで、マーケティング分野で考案された購買行動モデル「AISAS」（近藤，2009；森岡ら，2006）に基づいて派生的な交通行動の発生から達成までを段階的に整理する。AISASとはインターネットの普及に焦点を当てた購買行動モデルであり、Attention（認知）、Interest（興味）、Search（検索）、Action（行動）、Share（共有）から成り立つ。例として、

- (1) Attention：Web サイトや SNS を通じて商品を認知する、
- (2) Interest：商品に興味を持つ、
- (3) Search：検索エンジンを利用して商品について検索する、
- (4) Action：商品を購入する、
- (5) Share：商品について SNS で共有する

がある。派生的な交通行動についても、購買行動と同じように、

- (1) Attention：移動発生要因となる施設や場所を認知する、
- (2) Interest：移動発生要因となる施設や場所に興味を持ち、行動意図が形成される、
- (3) Search：目的地に行くための経路を検索する、
- (4) Action：選択した経路および手段で移動する、
- (5) Share：目的地での体験を SNS で共有する

と考えられる。ただし、5段階目は公共交通の需要予測上は必要不可欠な段階ではない。

上に示すように、経路検索システムは移動の前段階で利用されるものである。経路検索システムにおける経路検索行動が実際の交通行動の予兆であると考えれば、突発的な交通需要の発生に影響する要因（例えば、イベントや天候悪化）について議論せずとも、経路検索データを活用し、将来発生しうる突発的な交通需要の変動を事前に予測できる可能性がある。そして、突発的な変動が予測できれば、検索者にとってより移動利便性の高い経路と手段を掲示可能になる。さらには、需要に応じた移動サ

ービスが前もって準備できる。

1.1.6 SP データとしての経路検索データによる交通需要予測の可能性

紙面調査データ時代においても、ビッグデータ時代においても、交通需要の予測には、顕示選好 (Revealed Preference : 以降, RP) データが主に活用されている。顕在化した行動のみがモデルの根拠となるべき情報であるという考え方は当然であり、実際の交通行動を反映したデータとして信頼性や妥当性は高い。ただし、RP データに基づく交通需要予測は行動主体を取り巻く環境は常に変化しないことを仮定している。そのため、将来何らかの要因によって交通需要が変動した場合は、その前提条件から外れることとなり、適切な需要予測ができない。

マーケティング分野を中心に需要予測において表明選好 (Stated Preference : 以降, SP) データも使われている。新商品の需要を予測する際に、市場に存在しない商品に関する RP データを入手することは不可能なため、仮想的な状況における個人の好みを表す SP データは有効な情報源となる。交通分野においても、SP データに基づく交通需要予測に関する研究が数多く取り組まれている (例えば、森川, 1990 ; 藤井と Garling, 2003, 尾形ら, 2008)。さらに、交通需要予測では SP データと RP データそれぞれを単体で用いた分析よりも、互いに補完的な性質を持つ両データを用いた分析 (SP/RP モデル) の方が予測精度は高いといわれており、従前より SP データと RP データを用いた方法論が提案されている (例えば、Dissanayake and Morikawa, 2000 ; 森川と山田, 1993)。SP データと RP データを同時に用いる意義は、

- (1) RP データだけでは同定できない交通需要に対するパラメータを推定すること、
- (2) SP データから得られる信頼性の高い、交通需要の変動に影響する情報を用いたパラメータの信頼性を上げること、
- (3) SP データに含まれるさまざまなバイアスを修正すること

の3点に集約される (森川と山田, 1993)。既往研究において、SP/RP モデルの有効性は十分に実証されているが、多くはアンケート調査に基づくものであり、交通ビッグデータを活用したモデルは提案されていない。

これは、交通ビッグデータにおいて将来の移動意向を表現したデータがほとんど存在しないことに起因する。そのような中で、唯一、交通ビッグデータの中で SP データに位置付けられるものが経路検索データである。しかし、既往研究において経路検索データを活用した交通需要予測手法の議論は筆

者の知る限り存在しない。

経路検索データには「いつ」「どこから」「どこへ」行きたいという検索者の将来の移動希望が反映されており、RP データでは観測できない未来の行動を表現する交通ビッグデータといえる。経路検索データは潜在的な将来の交通需要を表す唯一の SP データであることを踏まえると、重要な情報源としてその活用可能性を検討することの意義は高い。そして、過去の行動実績である交通系 IC カードの乗降履歴データ (RP データ) に、将来の行動を表す経路検索データ (SP データ) を加味することによって、互いの欠点を補完し、交通ビッグデータを活用した交通需要予測においても予測精度の向上が期待される。

1.2 研究の目的

日々変動する交通需要の予測には交通ビッグデータの活用が期待される。一方、従前より、交通需要予測において RP データと SP データの両方に基づく統合モデルのほうが、それぞれの単体モデルより予測精度が高いといわれている。そこで、本研究では、近年導入が進む需要量応答型の交通システムの運用計画立案に資する情報を提供するために、RP データである交通系 IC カードの乗降履歴データに加え、SP データである経路検索データを活用した新たな交通需要予測手法を提案し、突発的需要も含めた日々の交通需要を予測可能とする方法論の開発とその有用性の検証を目的とする。

1.3 経路検索行動と実行動を関連づけるための考え方

経路検索データはあくまで Web 上の行動を表すものであり、同一個人の経路検索行動と実行動の関係をデータとして収集することは現時点ではできない。つまり、経路検索データだけでは誰が、どのような目的や理由で検索に至ったかの要因を把握することはできない。また、検索者がシステムで提示された経路通りに実際に移動したかどうかは不明であり、経路検索行動がどの程度実行動に反映されたのかも把握できない。さらに、経路検索システムが平常時とは異なる移動の際に利用されることを踏まえると、経路検索データですべての交通行動を記述できるわけではない。

そこで、経路検索データで予測可能な交通需要を表 1.1 に基づいて整理する。ここで、本研究では便宜上、個人レベルでの交通の発生を「行動」と呼び、個人行動を集計した集団レベルでの交通の発生を「需要」と呼ぶ。公共交通を移動手段とする主な活動に (1) 通勤・通学、(2) 定期的な買い物や通院、(3) 非定期的な買い物や通院、(4) 観光や出張、(5) イベント参加や悪天候回避を挙げ、それら活動が活動主体個人で考えたときにはどのような行動となり、集計レベルで考えたときにはどのような交通需要として現れるのかをまとめる。

(1) 通勤・通学

通勤・通学は個人レベルの観点からは日々行う活動であり、定期的行動といえる。個人レベルでの定期的な行動は、集計レベルにおいても、日々安定して発生する需要として顕在化する。そのため、集計レベルでの定常的需要である。

(2) 定期的な買い物や通院

例えば、毎週決まった曜日や 2~3 日に 1 回など一定の間隔で行われる食料品や日用品の買い物、リハビリの定期診察などへの通院は、個人レベルでは定期的な行動である。そのため、集計レベルにおいては周期性として現れ、定常的需要として観測される。

(3) 非定期的な買い物や通院

衣類の購入、友人との外食、風邪を引いたときなどは、普段とは違う場所、時間帯に外出する。これらは、個人レベルで見れば、毎日あるいは週に数回、月に数回も行わない非定期的行動である。しかし、店舗や病院の立場からは、施設の曜日平均来訪者数は安定している。個人レベルでの非定期的な行動も、集計レベルで考えれば、一定数発生している需要と考えられ、定常的需要と定義できる。

(4) 観光や出張

観光や出張についても、不定期の買い物や通院と同様の考え方が当てはまる。

以上 (1) ~ (4) に示すように、公共交通を移動手段とする活動の多くは、個人レベルで見れば定期的行動・非定期的行動に分類できるが、集計レベルで見れば、安定的に発生する定常的需要といえる。

一方、集団レベルで見たときに安定的に発生しない需要があり、これを本研究では「突発的需要」と定義する。この突発的需要は、以下のようなときに平常時と比べて需要が増減すると考えられる。

(5) イベント参加や悪天候回避

イベント参加や悪天候回避などによる行動変化は、個人レベルでの非定期的行動である。イベント開催や悪天候発生に周期性がなければ、集団で集計された需要にも規則性がなく、突発的需要となる。

表 1.1 経路検索データと交通系 IC カードデータが表現可能な「行動」と「需要」

経路検索データ		交通系 IC カードデータ	
個人レベル	集計レベル	個人レベル	集計レベル
<ul style="list-style-type: none"> ・ 非定期的行動 (例) 非定期的な買い物や通院, 観光や出張 	<ul style="list-style-type: none"> ・ 定常的需要 	<ul style="list-style-type: none"> ・ 定期的行動 (例) 通勤・通学, 定期的な買い物や通院 ・ 非定期的行動 (例) 非定期的な買い物や通院, 観光や出張 	<ul style="list-style-type: none"> ・ 定常的需要
<ul style="list-style-type: none"> ・ 非定期的行動 (例) イベント参加や悪天候回避 	<ul style="list-style-type: none"> ・ 突発的需要 	<ul style="list-style-type: none"> ・ 非定期的行動 (例) イベント参加や悪天候回避 	<ul style="list-style-type: none"> ・ 突発的需要

経路検索システムは普段とは異なる行動をとるときに利用されるサービスであり、個人レベルの非定期的行動に関係する。しかし、個人レベルでの非定期的な行動も、周期性があれば集計レベルでの定常的な需要として現れることを踏まえると、経路検索データを用いることによって予測精度の向上が期待されるのは、突発的需要である。そこで、本研究では突発的需要の予測に特化したモデルを構築することで、交通需要予測の精度向上を試みる。

現存の公共交通システムは、定常的な需要の発生には対応できているものの、何らかの事象の発生によって突発的に交通需要が変化した場合には対応できない。交通需要はあらゆる要因によって日々変動するものであり、そのような需要の変化に対応し、需要に即した供給を行うためにも、突発的需要の事前予測は重要である。突発的な需要の発生にはイベントや天候悪化などさまざまな要因が考えられ得るが、それら要因を解明するのではなく、すべての要因が集約された代理変数が経路検索データであると仮定し、これによる新たな交通需要予測手法を提案するのが本研究の新規性・独自性である。本提案手法によって交通需要の変動を捉えることが可能となれば、変容する交通需要に応じたサービスが事前に準備でき、実用的意義は大きいといえる。

1.4 本論文の構成

実証分析に際して、香川県高松市を中心に運行されている鉄道路線「ことでん」を対象に、

- (1) 経路検索データに現れる需要の特性把握（主に 3 章, 4 章）,

(2) 経路検索数に基づく交通需要予測モデルの開発と有用性の検証（主に 5 章, 6 章）
の手順で, 研究を遂行する.

本論文は 7 章で構成される. 論文の構成を図 1.1 に示す.

第 1 章では, 本研究の背景と目的について述べた. そして, 経路検索データで予測可能な交通需要を仮説に基づき整理した.

第 2 章では, RP 交通ビッグデータを用いた交通需要予測に関する既往研究と経路検索データに関する既往研究を整理し, それら研究成果を踏まえた本研究の位置付けを示す.

第 3 章では, 経路検索データで予測可能な交通需要について検討する第一段階として, 経路検索行動の特徴を明らかにする. 経路検索データに関する研究は, 筆者の知る限りでは海外事例, 国内事例ともに極めて少なく, 少しずつ研究が蓄積されているものの, 他交通ビッグデータの研究蓄積と比較し, 十分とはいえない. そのため, 経路検索データが表す行動特性について未解明な点が多い. そこで, 経路検索データに対して情報量縮約を行い, 代表的な経路検索行動を抽出することによって, 経路検索データに反映される行動を概観する.

第 4 章では, 第 3 章で概観した経路検索行動の特性を踏まえ, 検索の多寡に影響する要因を明らかにする. 本研究では検索者は何らかの移動目的を持って, 経路検索システムを利用し, その移動目的の対象が検索要因となると考える. そこで, 駅周辺環境の違いが経路検索数の多少に影響すると仮定し, 駅間の経路検索数と駅周辺の人口分布や施設立地数との関係を明らかにする. これにより, どのような場合に経路検索システムが利用されるかを推察し, 第 3 章の結果を定量的な根拠によって裏付ける.

以上, 第 3 章, 第 4 章によって, 「(1) 経路検索データに現れる需要の特性把握」を行い, 第 1 章で述べた経路検索データで予測可能な交通需要に関する仮説の妥当性を検討する.

続く, 第 5 章と第 6 章では, 「(2) 経路検索数に基づく交通需要予測モデルの開発と有用性の検証」を試みる.

第 5 章では, 第 3 章, 第 4 章の結果を踏まえ, 経路検索データに内在する定常的需要と突発的需要を分解する. 同様に, 交通系 IC カードデータについても定常的需要と突発的需要の分解が必要である. ここで, 両データに含まれる需要は相互に影響を及ぼしながら, 日々変動していると想定される. そこで, 両者の相互関係を考慮した成分分解手法を開発する. そして, 分解した経路検索数の突発的需要と交通系 IC カード利用数の突発的需要の関係を明らかにすることで, 経路検索データによる交通需要

の予測可能性を検討する。また、経路検索データを用いることによる予測精度の向上についても検証する。

第6章では、第5章で明らかとなった突発的需要の関係をもとに、突発的需要を含む日々の交通需要を予測するためのモデルを開発する。その上で、構築した交通需要予測モデルの安定性を、時間的・空間的観点から確認する。さらに、需要予測の観点からは、当日の経路検索数を用いて当日の実行動者数を予測することの意義は少なく、事前情報で当日の実行動者数を予測しなければならない。そこで、どれくらい前に収集された経路検索データで、当日の実行動者数を予測できるかを検討する。具体的には、例えば、予測日当日朝までに収集した経路検索数で実行動者数を予測した場合、予測日前日夕方までに収集した経路検索数で実行動者数を予測した場合など、予測したい日のどれくらい前に収集された経路検索データであれば、交通需要の変動の事前予測が可能かを検証する。

最後に、第7章では、本研究で得られた成果を総括し、今後の研究課題を取りまとめる。

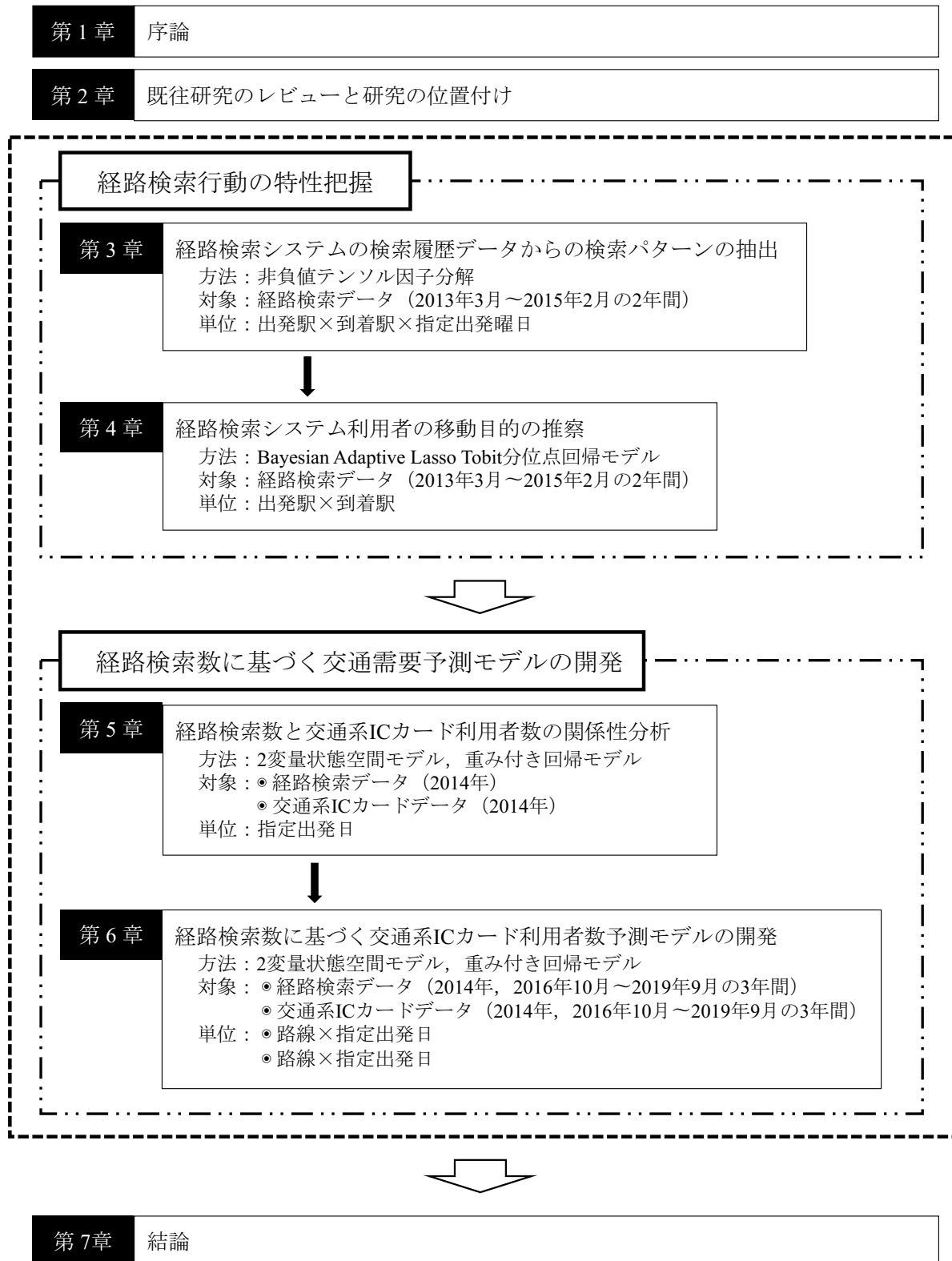


図 1.1 本論文の構成

第1章 参考文献

- Abad, P.R.B., Fillone, A.M. (2020), Changes in travel behavior during flood events in relation to transport modes: the case of Metro Manila, Philippines, *Transportation Research Procedia*, Vol.48, pp.1592-1604.
- Alsger, A., Tavassoli, A., Mesbah, M., Ferreira, L., Hickman, M. (2018), Public transport trip purpose inference using smart card fare data, *Transportation Research Part C*, Vol.87, pp.123-137.
- Cools, M., Moons, E., Creemers, L., Wets, G. (2010), Changes in travel behavior in response to weather conditions: do type of weather and trip purpose matter?, *Transportation Research Record*, Vol.2157(1), pp.22-28.
- Dissanayake, D. and Morikawa, T. (2000), Travel Demand Models with the RP/SP Combining Technique for the Developing Countries, *The International Conference CODATU IX*, pp.103-107.
- Emani, C.K., Cullot, N., Nicolle, C. (2015), Understandable big data: a survey, *Computer Science Review*, Vol.17, pp.70-81.
- Hofleitner, A., Harring, R., Bayen, A. (2012), Arterial travel time forecast with streaming data: A hybrid approach of flow modeling and machine learning, *Transportation Research Part B*, Vol.46, pp.1097-1122.
- Jenelius, E., Koutsopoulos, H. N. (2013), Travel time estimation for urban road networks using low frequency probe vehicle data, *Transportation Research Part B*, Vol.53, pp.64-81.
- Khattak, A.J., De Palma, A. (1997), The impact of adverse weather conditions on the propensity to change travel decisions: a survey of Brussels commuters, *Transportation Research Part A: Policy and Practice*, Vol.31(3), pp.181-203.
- Ma, X., Wu, Y.J., Wang, Y., Chen, F., Liu, J. (2013), Mining smart card data for transit riders' travel patterns, *Transportation Research Part C*, Vol.36, pp.1-12.
- Milne, D., Watling, D. (2019), Big data and understanding change in the context of planning transport systems, *Journal of Transport Geography*, Vol.76, pp.235-244.
- Ohler, F., Krempels, K.H., Möbus, S. (2017), Forecasting public transportation capacity utilisation considering external factors, *VEHITS*, pp.300-311.
- Pereira, F.C., Rodrigues, F., Ben-Akiva, M. (2015), Using data from the web to predict public transport arrivals under special events scenarios, *Journal of Intelligent Transportation Systems*, Vol.19(3), pp.273-288.
- Simoncini, M., Taccari, L., Sambo, F., Bravi, L., Salti, S., Lori, A. (2018), Vehicle classification from low-frequency GPS data with recurrent neural networks, *Transportation Research Part C*, Vol.91, pp.176-191.
- Torrecilla, J.L., Romo, J. (2018), Data learning from big data, *Statistics and Probability Letters*, Vol.136, pp.15-19.
- Vij, A., Shankari, K. (2015), When is big data enough? Implications of using GPS based surveys for travel demand analysis, *Transportation Research Part C*, Vol.56, pp.446-462.
- Wang, Z., Hu, Y., Zhu, P., Qin, Y., Jia, L. (2018), Ring aggregation pattern of metro passenger trips: A study using smart card data, *Physica A*, Vol.491, pp.471-479.
- Yang, H., Wang, Z., Xie, K., Dai, D. (2017), Use of ubiquitous probe vehicle data for identifying secondary crashes, *Transportation Research Part C*, Vol.82, pp.138-160.
- Zhong, C., Manley, E., Arisona, S. M., Batty, M., Schmitt, G. (2015), Measuring variability of mobility patterns from multiday smart-card data, *Journal of Computational Science*, Vol.9, pp.125-130.
- 安達俊朗, 渡辺康男, 川見篤史 (2002), 高速道路交通管制システムの現状とこれから, *東芝レビュー*, Vol.57,

No.12, pp.15-18.

飯田祐三, 岩辺路由, 菊池輝, 北村隆一, 佐々木邦明, 白水靖郎, 中川大, 波床正敏, 藤井聡, 森川高行, 山本俊行 (2000), マイクロシミュレーションアプローチによる都市交通計画のための交通需要予測システムの提案, 土木計画学研究・論文集, No.17, pp.841-847.

尾形信一, 倉内慎也, 森川高行 (2008), RP/SP 融合法に基づく新都市鉄道路線の需要予測における実務的課題, 土木計画学研究・論文集, Vol.25, No.3, pp.701-708.

織田利彦 (2001), 交通管制システムと ITS の活用による今後の展望, 計測と制御, Vol.40, No.3, pp.226-231.

金森亮, 森川高行, 山本俊行, 三輪富生 (2009), 総合交通戦略の策定に向けた統合型交通需要予測モデルの開発, 土木学会論文集 D, Vol.65, No.4, pp.503-518.

北野誠一, 中島良樹, 井料隆雅, 朝倉康夫 (2008), 交通系 IC カードデータを用いた長期間の鉄道利用履歴の分析, 第 37 回土木計画学研究・講演集, CD-ROM, 4 pages.

北村隆一 (1996), 交通需要予測の課題: 次世代手法の構築に向けて, 土木学会論文集, Vol.1996, No.530, pp.17-30.

北村隆一 (2003), 変動についての試行的考察, 土木計画学研究・論文集, Vol.20-1, pp.1-15.

楠田悦子 (2021), 60 分でわかる! MaaS モビリティ革命, 技術評論社, pp.18-21.

国土交通省, 平成 26 年度国土交通白書, <https://www.mlit.go.jp/hakusyo/mlit/h26/hakusho/h27/index.html> (2021 年 12 月 9 日確認).

国土交通省, 令和元年度交通政策白書, <https://www.mlit.go.jp/sogoseisaku/transport/content/001313039.pdf> (2021 年 12 月 9 日確認).

国土交通省, 交通系 IC カードの普及・利便性拡大に向けた検討会, <https://www.mlit.go.jp/common/001097000.pdf> (2021 年 12 月 9 日確認).

国土交通省 (2013), 交通政策基本計画, <https://www.mlit.go.jp/common/001069407.pdf> (2021 年 12 月 10 日確認).

国土交通省 (2020), 令和 2 年度日本版 MaaS 推進・支援事業 38 事業について, <https://www.meti.go.jp/press/2020/07/20200731005/20200731005-2.pdf> (2021 年 10 月 13 日確認).

国土交通省 (2021), 需要に合わせた供給最適化とポストコロナ需要変容を踏まえたデータ分析・施策のための高松 MaaS 実証実験, <https://www.mlit.go.jp/sogoseisaku/transport/content/001410283.pdf> (2021 年 10 月 13 日確認).

近藤史人 (2009), AISAS マーケティング・プロセスのモデル化, JSD 学会誌システムダイナミックス, No.8, pp.95-102.

鈴木文彦 (2012), 地方におけるオンデマンド交通の可能性と課題, オペレーションズ・リサーチ, Vol.57, No.3, pp.124-129.

島海梓, 大口敬, 岡野孝司, 粕谷暁, 糸島史浩, 甲斐穂高, 邢健 (2020), ETC データを用いた OD 交通量の時間変動パターン解析, 生産研究, Vol.72, No.2, pp.173-178.

永井政伸, 日比野直彦, 森地茂 (2011), ETC-OD データを用いた都市高速道路における経路選択行動に関する研究, 土木学会論文集 D3 (土木計画学), Vol.67, No.5, pp.I_589-I_597.

中沢航太, 佐々木邦明 (2011), 高速道路利用データを活用した個人の利用パターン抽出と需要構造の解析, 第 44 回土木計画学研究・講演集, CD-ROM, 7 pages.

西内裕晶, 轟朝幸 (2012), 交通マーケティング手法検討のための IC カードデータを活用した利用者行動特性の

- 把握, 土木学会論文集 F3 (土木情報学), Vol.68, No.2, pp.II_8-II_17.
- 日高洋祐, 牧村和彦, 井上岳一, 井上佳三 (2020), Beyond MaaS 日本から始まる新モビリティ革命-移動と都市の未来-, 日経 BP, pp.19-23.
- 福井良太郎 (1999), ITS 通信技術, 日本ロボット学会誌, Vol.17, No.3, pp.334-338.
- 藤井聡, Garling, T. (2003), 交通需要予測における SP データの新しい役割, 土木学会論文集, Vol.2003, No.723, pp.1-14.
- 牧村和彦 (2021), MaaS が都市を変える 移動×都市 DX の最前線, 学芸出版社, pp.172-196.
- 森岡慎司, 長谷川想, 山川茂孝 (2006), AISAS モデルにみるロコミの形成過程におけるプランニング作法の提案, マーケティングジャーナル, Vol.26, No.1, pp.29-39.
- 森川高行 (1990), ステイティッド・プリファレンス・データの交通需要予測モデルへの適用に関する整理と展望, 土木学会論文集, Vol.1990, No.413, pp.9-18.
- 森川高行, 山田菊子 (1993), 系列相関を持つ RP データと SP データを同時に用いた離散型選択モデルの推定, 土木学会論文集, Vol.1993, No.476, pp.11-18.
- 森地茂, 屋井鉄雄, 田村亨 (1985), 非集計行動モデルによる OD 交通量推計方法, 土木計画学研究・論文集, No.2, pp.45-52.

既往研究のレビューと研究の位置付け

〈第 2 章の概要〉

ビッグデータの本質は時々刻々と変化する現象の移ろいを記録した点にあり，これによって「変動」を分析できる．交通分野においては，RP 交通ビッグデータを用いて，定常的な交通需要に着目した現象理解や需要予測がなされてきたが，近年は，非定常状態，すなわち突発的な交通需要の予測を主眼とする研究も蓄積され始めている．本章では，突発的需要の変動要因を外生変数として導入した交通需要予測モデルに関する既往研究を整理し，需要変動を予見する上での問題点を挙げる．

さらに，公共交通における需要の変動を予見できる可能性のある経路検索データを分析した既往研究を整理する．経路検索データに関する研究は，国内外で研究蓄積が少なく，その特性や活用可能性について十分に議論されていない．そのため，経路検索データの交通需要予測への有用性が疑問視されている．

そこで，本章では，Web 上の潜在需要の特性に関する仮説と Web 上の潜在需要と実空間上の顕在需要を関連づける方法に関する仮説の 2 つの仮説を設定し，本研究の位置づけを示す．

2.1 RP ビッグデータを用いた需要予測に関する既往研究

1.1.2 項で述べたように、人々の交通行動を把握するためのデータとして、パーソントリップデータに代表されるような時系列性を有しない断面データには、当然、交通行動の変化に関する情報は含まれていない。一方で、ビッグデータは半永久的に継続して収集できることから、行動の時間的変化が反映されている。つまり、ビッグデータは状態の遷移過程の情報を持つデータ、いわゆる時系列データといえる。データ量が膨大であることはビッグデータの最大の特徴であるが、データ量の多さに限れば、その特性を持つデータはビッグデータ時代以前から存在する。例えば、先に挙げたパーソントリップデータも何千人、何万人の回答にもなれば、膨大なデータ量を持つビッグデータとなる。ビッグデータの本質はデータ量の多さではなく、時々刻々と変化する現象が記録されている点、すなわち、変動性を有するところにある。

ビッグデータを用いれば、現象の変動が分析できる。医療分野（例えば、早野ら、2017；吉田ら、2018；湯田ら、2018）、気候分野（例えば、Rajput et al., 2021; Imteaz et al., 2021）、金融分野（例えば、佐藤ら、2015；羽柴、2020；Ronaghi et al., 2021）、マーケティング分野（例えば、二田、2015；野口、2015）などあらゆる分野で変動分析が行われるようになった。そして、得られる分析結果から、現象を理解し、例えば、クレジットカードの不正利用検出、渋滞予測、電力の需要予測、風邪の流行予測など、実社会のあらゆる場面でビッグデータは活用されている（城田、2012）。

分析されるビッグデータのほとんどは過去の実績に基づく RP ビッグデータである。交通分野においても、1.1.2 項で挙げたような RP 交通ビッグデータを用いて、交通現象の変動分析がなされている。既往研究では、RP 交通ビッグデータから公共交通利用者の平均的な変動パターンを抽出し、現存サービスの利用のされ方を把握する「定常的現象の理解」（例えば、野上ら、2011；嶋本ら、2014）や、その現象理解に基づいて将来発生する需要を予測する「定常的需要の予測」（例えば、角野ら、2014）という「定常状態」の記述・予測に着目した研究が多かった。

近年は、「非定常状態」、すなわち突発的に変動する需要の予測を主眼とする研究も蓄積されている。定常的需要予測モデルでは説明のできない交通需要の急激な増減を予測するために、既往研究では、時間的あるいは空間的に予測範囲を限定し、絞り込むことが多い。例えば、時間的に予測範囲を絞り込むとは、1 時点先の変動を予測する場合に現時点と 1 時点先の時間間隔を短くすることである。既往研究では、データの過去の変動傾向から将来の変動を予測するモデルが提案されている。この場合、将来の変動は直前の変動に最も影響を受け、変化すると考えられるので、時間間隔は短い方が予測精度は

高くなる。既往研究においても、15分先を予測するモデル (Xue et al., 2015) や 30分先を予測するモデル (Li et al., 2017) と、時間間隔が短いモデルが提案されており、予測時点までの間隔が短いモデルの方が予測精度は高いことが示されている (Yang et al., 2021)。しかし、突発的需要を正確に予測できたとしても、予測時点までの時間間隔が短いほど、交通事業者は需要の変化に対応するための時間を持たず、実用的な交通需要予測モデルとはいえない。

空間的に予測範囲を絞り込むとは、分析対象を対象地域全体ではなく、ある特定の範囲内やある路線というように分析対象を細分化し、限定することである。空間的に範囲を絞り込めば変動要因となりうる事象 (天候やイベント) に関するデータの収集可能性は高まる。特に、天候情報を考慮する場合には有効であり、既往研究においても、分析対象範囲を絞り、天候変化による突発的需要の発生を予測するモデルが提案されている (Guo et al., 2007; Kalkstein et al., 2009; Li et al., 2014; Zhou et al., 2017; 青地と白石, 2020)。また、イベントの影響を考慮する際にも、分析対象範囲を絞った方が、影響を与えたと考えるイベントの候補の抽出が容易となる。ただし、イベントについては空間的範囲を絞り込んだとしても、分析対象範囲内のすべてのイベント情報を入手することは困難である。そのため、特定のイベントと突発的需要の関係にのみ着目することになる (Rodrigues et al., 2017; Pereira et al., 2015; Ding et al., 2016)。しかし、イベントの種類や規模は多種多様であり、同種類のイベントでも毎回内容が全く同じであるとは考えられない。そのため、過去の特定のイベントと突発的需要の関係性を記述できても、今後行われるイベントと過去のイベントの関係性を別途モデル化して説明しなければ、将来のイベントによる突発的需要の予測はできない。

1.1.5 項でも述べたように、天候やイベントなど変動要因の事後的な特定とその影響力は把握できても、突発的需要の事前予測は困難である。既往研究において、RP 交通ビッグデータを用いた突発的需要予測のための有効な分析手法は確立されているとはいえない。

2.2 経路検索データに関する既往研究

RP 交通ビッグデータに対して、交通需要の変動を予測できる可能性のある経路検索データを分析した既往研究は、

- (1) 経路検索データの特性分析,
- (2) 経路検索データに基づく突発的需要の検知

に大別される。ここでは、2タイプの研究についてそれぞれ整理し、既往研究の問題点をまとめる。

2.2.1 経路検索データの特性分析に関する既往研究

経路検索データは公共交通利用者の将来の移動意向、すなわち実空間上には顕在化していない潜在需要を表すデータである。そのため、RP 交通ビッグデータでは捉えられない、公共交通利用者が本来持つニーズを表す SP 交通ビッグデータであり、交通計画への活用可能性が議論されている。既往研究の多くは潜在需要の把握を目的とし、検索者が指定した地点の経路検索数（伊藤ら，2013；太田，2015；太田ら，2015；小竹ら，2016）や時間情報と地点の組合せ別の経路検索数（助友ら，2007；見生，2012；香川ら，2016；桑野ら，2017）から、対象交通機関全体のうちどこにニーズがあるのかを分析している。しかし、あくまで地点ごと、地点間の把握に留まっており、何が起因して検索が行われるのか、ニーズがあるのかは定量的に明らかにされていない。

これは、経路検索データに検索者の個人属性情報が記録されていないことに起因する。経路検索数の多寡によって潜在需要の存在は確認できるものの、どのような人の、どのような移動目的による検索かは、経路検索データだけでは明らかにできない。そのため、経路検索システムの利用者を対象としたアンケート調査の実施（轟ら，2005）や、検索者が指定した地点周辺の施設立地状況を調査すること（太田ら，2015；見生ら，2012；香川ら，2016；寺部ら，2009）で、経路検索データが表現する潜在需要が推測されている。しかし、確証ある結論にはいたっていないのが現状である。

経路検索データと実行動の関係性の解明に注力するのではなく、経路検索データが公共交通利用者の本来のニーズを反映しているという特性を活用して、公共交通の運行計画を見直す研究もある（桑野ら，2020）。また、経路検索システム自体の利便性を高めるために、システム上で掲示する付加情報を検索者の選好に合わせて提供する方法も検討されている（古川ら，2019；石村ら，2014；梶原ら，2014）。これらの研究は、検索者に必要な情報、例えば、検索した駅周辺の観光スポットを経路情報と共に掲示することによって、観光行動や回遊行動を促し、地域活性化に繋げようとするものである。経路検索システムを提供する立場から、サービス改善を目的とした研究はいくつかあり、そのような研究が今後も蓄積され、経路検索システムが ETC2.0 のような高度な情報提供サービスとして進化していくことで、経路検索システムはこれまで以上に浸透し、ひいては公共交通利用時の利便性もより高まっていくだろう。反して、将来の移動意向を表すデータとして経路検索データは活用可能性を秘めているのにも関わらず、実際の公共交通利用との関連について不明点が多いために、このまま活用できなければ、経路検索システムは無駄なデータを蓄積し続けるシステムと化してしまう。

2.2.2 経路検索データに基づく突発的需要の検知に関する既往研究

先に大別したように、(1) 経路検索データの特性分析は比較的研究されているものの、(2) 経路検索データに基づく突発的需要の検知に関する研究はわずかしがなく、しかもそれらは突発的需要を予測しているわけではない。石村ら (2013) は検索者が経路検索システムにアクセスした日時（検索実行日時）と公共交通を利用したい日時（検索指定日時）の差を事前検索時間と定義し、検索者が実行動に移すまでの状態を段階的に想定し、設定した事前検索時間ごとに、駅・日種（平日、土日祝日）・時間帯・発着区分別事前検索数を算出した。そして、検索指定日時（検知対象日時）までの累積事前検索数が分析者の定めた閾値を超えた場合に、突発的需要があったとみなした。さらに、検出した突発的需要の発生に影響したと考えられるイベントを Web 検索し、どのようなイベントであれば、どのくらい前から突発的需要の検出が可能かを考察した。石村らが提案する手法は、経路検索データを用いて突発的需要の発生を予測できる可能性は示されているものの、実際の顕在需要との関連は明らかになっていない。坪内ら (2017) は突発的需要を事前に予測するために、予測対象日を指定した検索で、かつ予測対象日前日までに収集した経路検索数を用いて、予測対象日の経路検索数を予測するモデルを構築した。そして、経路検索数の予測値と分析者が設定した基準から、突発的需要の発生を判断した。その上で、石村らと同様に、突発的需要が発生したとみなした日に開催されたイベントを Web 検索し、大規模なイベントが開催される日ほど突発的需要の発生が検出可能と結論づけている。しかし、坪内らの研究は日々の経路検索数を予測しているに過ぎず、実際の需要との関連は明らかにされていないため、突発的に公共交通利用者数が増える可能性を予測しているだけである。坪内らの提案手法をベースにして、Twitter でつぶやかれた Tweet から未来のイベント日時を抽出し、その日時における突発的需要を、経路検索データを用いて予測した山下らの研究もあるが、この研究もまた経路検索数を予測しているに過ぎない (山下ら, 2018)。また、国外の事例としては、Roosmalen (2019) が経路検索データを用いて、短期的な交通需要の予測手法を提案しているが、やはり実際の需要との関連は議論されていない。

前述した経路検索データを用いて突発的需要を分析した既往研究は、経路検索データで突発的需要を予測できる可能性は示しているものの、実際の公共交通利用者数との関連が明らかにされていないため、あくまで突発的需要の検知に留まっており、交通需要予測手法への適用を試みる研究は筆者の知る限り存在しない。

2.2.3 経路検索者数と公共交通利用者数の関連性に関する既往研究

2.1.1 項および 2.2.2 項で述べた研究において、議論すべき点は経路検索と実際の公共交通利用者数の関連性である。これまで、経路検索データと実際の公共交通利用の相関関係はアンケートデータと比較した研究、国土交通省が行う大規模交通センサスデータなど統計データと比較した研究、駅の乗降者数データと比較した研究において議論されてきた。桑野ら（2020）、伊藤ら（2013）は、ある1年間における1日あたりの平均経路検索数に着目し、実際のバス利用者へのアンケートデータと比較することで、それぞれバス路線別、起終点の組合せ別に平均経路検索数と平均バス利用者数の間に高い相関関係（伊藤ら：0.91、桑野ら：0.84）があることを示した。石村ら（2013）は国土交通省が行う大規模交通センサスデータを用いて、経路検索数との比較分析を行なった。目的変数に各駅を指定した経路検索数、説明変数に各駅の定期総発着人数と定期外総発着人数をそれぞれ投入した単回帰分析の結果から、決定係数がいずれも0.82以上となり、経路検索数と公共交通利用者数との間に関連性があると結論づけている。轟ら（2005）は経路検索データと駅の乗降者数データを用いて、交通事業者別の経路検索割合と実際の乗降者数割合の間に、どの交通事業者においても0.97以上と高い相関があることを示している。

いずれの研究も経路検索データと実際の公共交通利用者数は相関が高いことを示しているが、データ取得期間や対象起終点など、経路検索データとは質的に異なるデータが使用されており、実際の公共交通利用者数との比較は十分とはいえない。また、前述の研究においては平均値や合計値を用いた関係性分析に留まっている。

2.3 研究の位置付け

前節で示した通り、需要変動の予兆検知が期待できるSP交通ビッグデータである経路検索データの分析は国内外で十数件程度であり、そもそも経路検索データがどのような特性を持つかさえ十分に把握できていない。経路検索データを利用するにあたり、まずは、経路検索データの特徴を明らかにする必要がある。自動車トリップや公共交通トリップなどの実行動者数の特性を把握するために、トリップのパターン抽出やトリップの多寡に影響を及ぼす都市要因の解明が既往研究で多く議論されてきた。本研究の3章、4章では、実行動者数の分析と同じように、経路検索行動のパターンや検索に影響を及ぼす要因を明らかにし、経路検索データの交通需要予測への活用可能性を検討する。

一方、数少ない経路検索データ分析では、曜日や月、年を集計単位とした平均的な経路検索行動と平均的な実行動との間に正の相関があることを示している。しかし、これは平均的に経路検索数が多いところは平均的に実行動者数も多いことを主張しているに過ぎず、単純に人口が多いところでは平均経路検索数が多いという当然の傾向を示しているだけで、論を俟たない結果である。都市部では人口が多いため、当然、平均公共交通利用者数も平均経路検索数も多くなる。一方、地方部や都市郊外部では人口が少ないため、両者は少なくなる。したがって、既往研究で示される平均経路検索行動と平均実行動の相関は人口を共通因子とした相関を見ているのみである。

本研究では、平均経路検索数と平均公共交通利用者数の相関を見るだけではなく、日々変動する経路検索数が公共交通利用者数にどれくらい反映されるのかという、真の意味での関係性を明らかにすることを目的とする。そもそも両者の関係性が明らかにできなければ、経路検索数データが交通需要予測に適用できるかを検討できず、既往研究においても問題であった経路検索データが表現する移動意向の不確実性を解決できない。

さらに、将来予測の観点からは、現象を記述するだけでなく、変数間の因果関係を考え、それらに関連づけるモデルの構築が必要となる。経路検索数が公共交通利用者数については、検索者が実際にシステム上で掲示された経路通りに行動したかはわからないものの、集計レベルでの経路検索と公共交通利用の関係における因果は、経路検索が先行して、公共交通利用が追随する、つまり「原因」が経路検索で、「結果」が公共交通利用であることに間違いない。因果関係の解明では、原因の時間的先行性が必要といわれているが、経路検索と公共交通利用における時間的先行性は自明である。したがって、両者の関係性が記述できれば、経路検索数から公共交通利用者数を予測できると期待される。先に述べた通り、ビッグデータの本質は変動性を表現する時系列データであることを踏まえ、本研究では因果関係をもとに時系列データから将来予測をする最も代表的な方法である時系列分析の適用を試みる。

モデルの構築に際して、

仮説 1：「経路検索システムと交通系 IC カードは利用されやすい状況が異なる」

ため、両者の関係性を直接的あるいは単純な方法で結びつけることができない。

そのため、

仮説 2：「両者に包含される突発的事象に起因する需要部分を適切にモデル化できれば、経路検索数は交通系 IC カード利用数の予測精度向上に寄与する」

というのが本研究の研究仮説の核心部である。なぜ 2 つが異なるかについては、1.3 節で詳しく述べた

ため、ここでの再説明は省略する。本研究では、2つを直接関連づけてはならず、突発的需要部分に注目して関連づけるべきという研究仮説に基づき原系列を複数の変動成分に分解する。そのために、時系列モデルの1つである2変量状態空間モデルを用いる。

状態空間モデルは、時系列の変動をいくつかの変動に分解するモデルであり、あらゆる時系列データに対応可能な柔軟性を持つモデルである。交通分野においては、単変量状態空間モデルを用いた成分分解はなされているものの（例えば、川崎ら、2016；中西ら、2020）、2変量間の相互関係を考慮した研究事例は極めて少ない。2変量状態空間モデルは、状態空間モデルを拡張した手法であり、経路検索数と実行者数の2変量間の関係を考慮できるという特長がある。データの成分分解が可能なARIMAモデルやSARIMAモデルのような時系列モデルは自己回帰モデルを発展化させたものもあるが、2変量状態空間モデルのように複数変数間の関係を考慮可能なモデルが少ないことから、本研究では多変量分析への拡張性が高い状態空間モデルを採用する。

取り扱うデータ間のある部分だけに因果関係があるという研究仮説が本研究の独自性であり、この研究仮説を検証するために、2つのデータが時系列データであることを踏まえ、2変量状態空間モデルを用いる点に特色がある。本研究では、5章にて2変量状態空間モデルを用いた経路検索数と交通系ICカード利用数の成分分解手法を開発、推定し、その結果を詳しく考察すること、そして、他モデルとの現況再現性を比較することで、研究仮説を検証する。そして、提案する交通需要予測モデルの実用性観点から、6章にて経路検索データが交通需要の変動の先行指標になりうるのかを検証し、提案モデルの有用性を示す。

第2章 参考文献

- Ding, C., Wang, D., Ma, X., Li, H. (2016), Predicting short-term subway ridership and prioritizing its influential factors using gradient boosting decision trees, *Sustainability*, Vol.8(11), 16 pages.
- Guo, Z., Wilson, N., Rahbee, A. (2007). Impact of weather on transit ridership in Chicago, Illinois, *Transportation Research Record: Journal of the Transportation Research Board*, (2034), pp.3-10.
- Imteaz, M. A., Paudel, U., Santos, C. (2021), Impacts of climate on weather and variabilities of potential water savings from rainwater tanks, *Journal of Cleaner Production*, Vol.311, 8 pages.
- Kalkstein, A. J., Kuby, M., Gerrity, D., Clancy, J. J. (2009), An analysis of air mass effects on rail ridership in three US cities, *Journal of transport geography*, Vol.17(3), pp.198-207.
- Li, Y., Wang, X., Sun, S., Ma, X., Lu, G. (2017), Forecasting short-term subway passenger flow under special events scenarios using multiscale radial basis function networks, *Transportation Research Part C: Emerging Technologies*, Vol.77, pp.306-328.
- Pereira, F. C., Rodrigues, F., Ben-Akiva, M. (2015), Using data from the web to predict public transport arrivals under special events scenarios, *Journal of Intelligent Transportation Systems*, Vol.19(3), pp.273-288.
- Rajput, M., Gahrooei, R., Augenbroe, G. (2021), A statistical model of the spatial variability of weather for use in building simulation practice, *Building and Environment*, Vol.206, 11 pages.
- Rodrigues, F., Borysov, S. S., Ribeiro, B., Pereira, F. C. (2017), A bayesian additive model for understanding public transport usage in special events, *IEEE transactions on pattern analysis and machine intelligence*, Vol.39(11), pp.2113-2126.
- Ronaghi, F., Salimibeni, M., Naderkhani, F., Mohammadi, A. (2021), COVID 19-HPSP: COVID-19 adopted hybrid and parallel deep information fusion framework for stock price movement, *Expert Systems with Applications*, Vol.187, 14 pages.
- Roosmalen, J.J. van (2019), Forecasting bus ridership with trip planner usage data: a machine learning application, Master thesis, *University of Twente*, 210 pages.
- Xue, R., Sun, D.J., Chen, S. (2015), Short-term bus passenger demand prediction based on time series model and interactive multiple model approach, *Discrete Dynamics in Nature and Society*, 13 pages.
- Yang, X., Xue, Q., Ding, M., Wu, J., Gao, Z. (2021), Short-term prediction of passenger volume for urban rail systems: A deep learning approach based on smart-card data, *International Journal of Production Economics*, Vol.231, 12 pages.
- Zhou, M., Wang, D., Li, Q., Yue, Y., Tu, W., Cao, R. (2017), Impacts of weather on public transport ridership: Results from mining data from different sources, *Transportation research part C: emerging technologies*, Vol.75, pp.17-29.
- 青地美桜, 白石陽 (2020), 交通系 IC カードの乗降履歴データを用いた路線バスの乗客人数予測手法の提案, 情報処理学会第 82 回全国講演論文集, Vol.2020, No.1, pp.353-354.
- 石村怜美, 太田恒平, 富井規雄 (2013), 経路検索サービスの実績データに基づく近未来の突発的移動需要の検出, 第 47 回土木計画学会・講演集, CD-ROM, 10 pages.
- 石村怜美, 梶原康至, 太田恒平 (2014), 乗換検索サービスの経路選択データを用いた公共交通の経路選択行動分析, 第 49 回土木計画学研究発表会・講演集, CD-ROM, 5 pages.
- 伊藤昌毅, 見生元気, 川村尚生, 菅原一孔 (2013), 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ンプ研究会, CD-ROM, 3 pages.

- 太田恒平 (2015), 経路検索ビッグデータが明かす鉄道利用の実態, オペレーションズ・リサーチ, Vol.60, No.10, pp.593-599.
- 太田恒平, 野津直樹 (2015), 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, 8 pages.
- 香川喬之, 桑野将司, 福山敬, 谷本圭志, 川村尚生 (2016), 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.
- 梶原康至, 太田恒平, 石村怜美 (2014), 乗換検索サービスの経路選択データを用いた公共交通の経路選択行動分析, https://consulting.navitime.biz/pdf/monograph_20141217.pdf (2021年12月15日確認).
- 角野惇, 猪井博登, 土井健司 (2014), 交通系 IC カードデータを用いた路線バス需要予測, 第 50 回土木計画学研究・講演集, CD-ROM, 5 pages.
- 川崎洋輔, 原祐輔, 桑原雅夫 (2016), 状態空間モデルとプローブ軌跡データを用いた区間途中の流出台数の推定手法の構築, 土木学会論文集 D3 (土木計画学), Vol.72, No.5, pp.I_1123-I_1132.
- 桑野将司, 秋元美穂奈, 細江美欧, 古川ゆり, 菅原一孔 (2020), 経路検索履歴データを用いたバス時刻表の設計に関する研究, 土木学会論文集 D3 (土木計画学), Vol.75, No.6, pp.I_501-I_511.
- 桑野将司, 木下礼央, 福山敬, 谷本圭志, 菅原一孔, 川村尚生 (2017), 非負値テンソル因子分解を用いた検索ログデータからのバス利用特性の分析, 土木学会論文集 D3 (土木計画学), Vol.73, No.5, pp.I_569-I_578.
- 見生元気, 伊藤昌毅, 川村尚生, 菅原一孔 (2012), 公共交通乗換案内サービスを用いた利用者行動解析, 第 21 回地理情報システム学会講演論文集, CD-ROM, 4 pages.
- 小竹輝幸, 梶原康至, 望月優, 野津直樹 (2016), 経路検索条件データを用いた全国観光アソシエーション分析～周遊圏から見出す地域の観光戦略～, 第 54 回土木計画学会・講演集, CD-ROM, 11 pages.
- 佐藤謙太, 小高知宏, 黒岩丈介, 白井治彦 (2015), ネガポジ解析による Web データと株価変動の相関関係評価, 福井大学大学院工学研究科研究報告, Vol.63, pp.75-86.
- 嶋本寛, 北脇徹, 宇野伸宏, 中村俊之 (2014), IC カード利用履歴データを用いた公共交通需要変動分析, 土木学会論文集 D3 (土木計画学), Vol.70, No.5, pp.I_605-I_610.
- 城田真琴 (2012), ビッグデータの衝撃, 東洋経済新報社, pp.152-157.
- 助友久志, 小谷通泰, 秋田直也 (2007), 携帯電話による交通情報提供サイトへのアクセスログ解析, 第 37 回土木計画学研究・講演集, CD-ROM, 4 pages.
- 坪内孝太, 下坂正倫, 小西達也, 丸山三喜也, 山下達雄 (2017), 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, 4 pages.
- 寺部慎太郎, 重里光佑, 内山久雄 (2009), 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.
- 轟朝幸, 谷口滋一, 高宮則夫 (2005), 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, 第 31 回土木計画学研究発表会・講演集, CD-ROM, 4 pages.
- 中西雅一, 前田雅人, 兒玉崇, 佐々木邦明 (2020), 高速道路の施策効果把握のための状態空間モデルの適用可能性に関する考察, 交通工学論文集, Vol.6, No.2 (特集号 A), pp.A_226-A_234.
- 二田総一郎 (2015), 食品メーカーの商品需要予測へのビッグデータ技術活用, NEC 技報, Vol.68, No.1, pp.90-93.
- 野上祐人, 片岡源宗, 熊谷靖彦 (2011), IC カード「ですか」を活用した高知中央地域の公共交通利用実態の基礎

- 分析, 第 44 回土木計画学研究・講演集, CD-ROM, 4 pages.
- 野口博司 (2015), マーケティングにおけるビッグデータの活用について, 流通科学大学論集-流通・経営編-, Vol.27, No.2, pp.219-235.
- 羽柴和摩 (2020), ビッグデータ解析を用いた為替予測に関する一考察, 法政大学大学院紀要. 理工学・工学研究科編, Vol.61, pp.1-8.
- 早野順一郎, 吉田豊, 古川由己, 湯田恵美 (2017), 心拍時系列ビッグデータに潜在するプライバシーリスク, 研究報告マルチメディア通信と分散処理 (DPS), Vol.2017-DPS-172, No.1, pp.1-4.
- 古川ゆり, 桑野将司, 秋元美穂奈, 菅原一孔 (2019), 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 (特集号 A), pp.A_184-A_192.
- 湯田恵美, 吉田豊, 早野順一郎 (2018), 日本人の身体活動の週内変動ウェアラブル加速度センサから得たビッグデータの解析, 情報処理学会研究報告, Vol.2018-EIP-79, No.12, pp.1-2.
- 吉田豊, 古川由己, 湯田恵美, 早野順一郎 (2018), ビッグデータから見た身体活動の曜日による変動, 生体医工学, Vol.56, pp.S223.

経路検索データからの検索パターンの抽出

〈第 3 章の概要〉

経路検索データは、検索者が「いつ」「どこから」「どこへ」移動したいかという将来の移動に対する希望が反映された SP データである。これを分析することによって、経路検索行動の主挙動、すなわち検索者の外出ニーズが把握可能であると期待される。

分析に際し、「いつ」「どこから」「どこへ」の 3 次数で構成されるデータは、各次数の次元数が多くなると、次元の呪いと呼ばれる情報の組合せ爆発が発生し、直接的にデータの特徴を把握しようとするとは複雑になる。一方、分析や結果の解釈を容易にするために次数と次元の解像度を減少させると、どの順序で情報縮約するかによって、異なる結果が得られる危険性がある。

本章では、組合せ爆発の問題と分析者の恣意性を回避するために、データが持つ本来の高次数・高次元構造を保持したまま、データの特徴を抽出する方法として非負値テンソル因子分解の適用を提案する。そして、提案手法によって検索者の代表的な検索パターンを明らかにすることで、経路検索行動の特徴を把握する。

3.1 はじめに

本章では、2.3 節で仮定した仮説 1 : 「経路検索システムと交通系 IC カードは利用されやすい状況が異なる」の考え方は妥当か否かを検討するために、経路検索システムがどのようなときに利用されるのかを定量的に明らかにする。

経路検索システムは、通勤や通学、定期的な買い物や通院など日常的な外出で利用されることは少なく、普段とは異なる非日常的な外出を行うときに利用されると考えられる。しかし、2.2 節で整理したように、経路検索データを分析した既往研究は少なく、経路検索行動の特性さえ十分に理解できていないのが現状であり、経路検索システムが非日常的な外出時に利用されるという推察は想定の外を出ない。

そこで、本章では、分析対象交通機関を対象とした検索が、いつ、どの駅間の移動に関して行われやすいのかを明らかにする。図 3.1 に本章の分析フローを示す。本章では、経路検索行動を概観するために、経路検索データからのパターン抽出を行う。検索パターンは、検索者がシステム上で指定する「乗車曜日（以降、曜日）」「乗車駅」「降車駅」の組合せで構成され、「どの曜日」に「どの駅」で乗車し「どの駅」で降車したいのかを表すものと定義する。

仮説1: 「経路検索システムと交通系 IC カードは利用されやすい状況が異なる」の考え方は妥当なのか？

→ 経路検索システムがどのようなときに利用されるのかを把握する

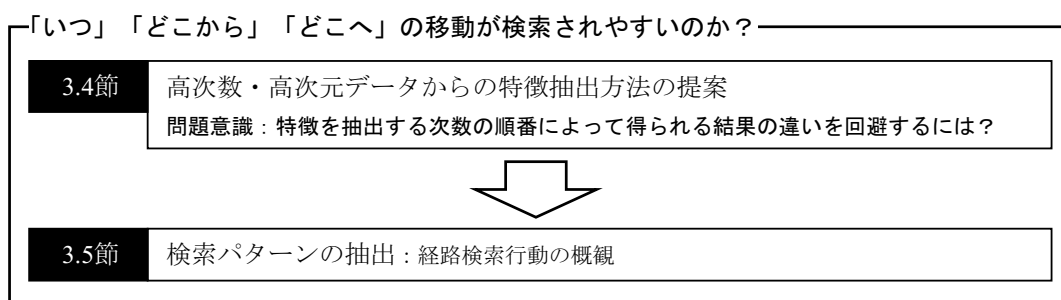


図 3.1 第 3 章の分析フロー

2.2 節で示した、経路検索データから経路検索行動の特徴の把握を目的とした既往研究の多くは、集計分析に留まっており、出発地や目的地を独立に扱い、それぞれの経路検索数や時間帯を固定した経路検索数を分析するものである。例えば、乗車駅ごとの経路検索数の時間的推移（1 乗車駅×曜日）や

特定の曜日における乗車駅・降車駅別経路検索数の比較（1 曜日×乗車駅×降車駅）というように、データを任意にスライスした分析となっている。つまり、本来は高次数・高次元で表されるべきデータを分析者の判断によって低次数・低次元のデータに変換し、分析している。分析者が恣意的に低次数・低次元にデータを変換した場合、データが持つ特徴のすべてを捕捉できていない (Simsekli et al., 2015)。

一方、扱うデータがビッグデータの場合には、各次数の次元数が多くなるにつれて、組合せ数が爆発し、データからの特徴抽出が非常に複雑になる。いわゆる次元の呪いである。例えば、曜日ごとの乗車駅・降車駅別経路検索数（3 次数データ）を把握する場合には、曜日ごとに OD 表を作成することになるが、一般的にこれら进行分析するとき、2 次数（2 軸：例えば、乗車駅と降車駅）に着目し、その特徴を把握した上で、次に残る次数（軸：曜日）を見て、3 つの次数（軸）の特徴を捉えることになる。このとき、図 3.2 に示すように、

- (A) 乗車駅と降車駅の関係把握してから曜日との関係を見る、
- (B) 乗車駅と曜日の関係把握してから降車駅との関係を見る、
- (C) 降車駅と曜日の関係把握してから乗車駅との関係を見る

などの、どの順序で情報を落とし込むかによって複数の次数の組合せが考えられる。

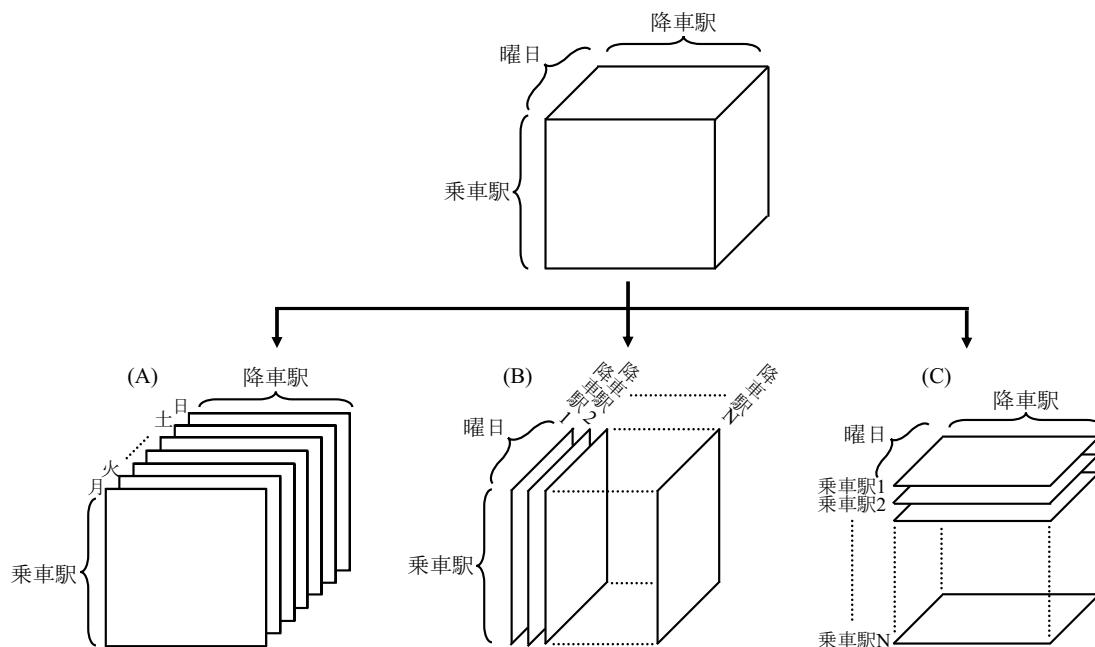


図 3.2 3 次数データの場合のスライス順序の候補

ここで、(A) 乗車駅と降車駅の関係把握してから曜日把握する場合、乗車駅と降車駅の関係を

分析したら必ず誤差が生じる。その後、時間帯との関係性を見ると、乗車駅と降車駅の関係で生じた誤差が持ち越されることになる。この誤差の影響により、(A) 乗車駅と降車駅の関係のみから時間帯との関係を把握するのと、(B) 乗車駅と時間帯の関係を見てから降車駅との関係を把握するのでは、誤差の生じ方が異なるために最終的な結果に差が生まれる。したがって、分析者がどの次数あるいは次元でデータをスライスするかの順序によって、得られる結果が異なる。

そこで、本来の高次数・高次元のデータ構造を保持したまま、データの特徴を抽出するために、非負値テンソル因子分解の適用を提案する。データ構造を保持した分析とはすべての次数(軸)を同時に解析することであり、非負値テンソル因子分解の適用により分析の順序によって生じる誤差をなくし、分析者に依存しない結果を得ることができる。

非負値テンソル因子分解は2次テンソルである行列データの次元を圧縮する際に因子分析が用いられるように、3次数以上のテンソルデータの各要素軸の次元を圧縮する際に用いられる分析方法である。非負値テンソル因子分解の適用事例は比較的新しく、例えば、ソーシャルネットワーク分析(Gauvin et al., 2014; Hashimoto et al., 2013)、購買行動分析(松本ら, 2015)、メディア分析(細江ら, 2018)で用いられている。また、土木計画分野においても近年になって適用例が増えており、例えば、観光行動分析(久保ら, 2019; 熊谷ら, 2015; Yanaguchi and Nakayama, 2020)、交通行動分析(細江ら, 2018; Hosoe and Kuwano, 2020)、都市構造分析(桐村, 2020)での研究報告がある。

本章では、曜日×乗車駅×降車駅で構成される3次テンソルを分析することにより、経路検索行動の特性を把握する。具体的には、「曜日」「乗車駅」「降車駅」の3次数で構成されるデータに対して、非負値テンソル因子分解を適用することによって、経路検索システムで分析対象交通機関を検索対象としたときに指定される代表的な検索パターンを抽出する。そして、得られた検索パターンから経路検索行動の特徴を明らかにする。

3.2 分析対象とする地域と交通機関

分析対象地域は香川県とし、県内を運行する「高松琴平電鉄(ことでん)」を分析対象交通機関とする。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。図3.3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。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、他交通機関への乗り換えが可能である。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。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可能である。中心部から離れた駅周辺には、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「こんびらさん」の呼び名で有名な金刀比羅宮がある。

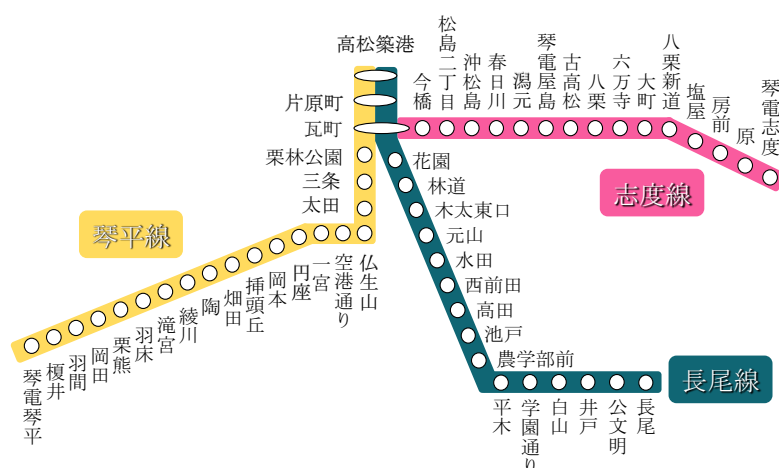


図 3.3 ことでん路線図 (参照：ことでんグループ)

3.3 経路検索システム「NAVITIME」の検索履歴データ

本章では、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。

図 3.4 に NAVITIME の検索条件入力画面を、図 3.5 に検索結果表示画面を示す。図 3.4 に示した検索条件入力画面のように、出発駅や到着駅、希望する利用日時などのいくつかの項目を入力することで、検索者の希望に応じた検索結果が返される。PC 以外に、携帯電話、スマートフォン、タブレット端末向けにも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を行える。

乗換案内

現在の電車運行状況 | 電車混雑状況

出発駅 高松築港 ⇄ 到着駅 瓦町

経由駅 駅を入力 +さらに追加

日付 2021年10月1日 カレンダー

時刻 8 : 時 0 : 分 現在時刻

出発 到着 始発 終電

詳細条件設定+

検索

検索履歴 ログインすると過去の検索履歴を表示できます

例) 2020年1月1日 12:00発 東京 ⇒ 大阪

図 3.4 検索条件入力画面

1 08:00 発 → 08:04 着 通常 ¥ 190円 定期代/内訳

所要時間 4分 乗換 0 回

印刷 メール送信 カレンダー登録 ルート指摘

出発 08:00 発 高松築港

周辺地図 時刻表 混雑予報

4分 1.7km <当駅始発> ことでん琴平線 琴電琴平行 途中の停車駅

運賃 190円

08:04 着 瓦町

周辺地図 時刻表 混雑予報

周辺情報 ホテル グルメ 住宅情報 バス停 駅 駐車場 レンタカー予約

瓦町を出てからの徒歩ルート 周辺にあるスポットまでのルートを検索します。

前3本の発着時刻 後3本の発着時刻

図 3.5 検索結果表示画面

表 3.1 に NAVITIME の検索履歴データに記録される情報を示す。時刻に関連する項目に着目すると、検索実行時刻と指定時刻に関する情報がある。検索実行時刻とは検索者が NAVITIME にアクセスした日時、つまり経路の検索を行なった日時を指す。一方、指定時刻とは検索者が検索条件入力画面上で指定した移動希望日時を指す。具体的には、検索者が指定する起点を出発したい日時あるいは終点到着したい日時のことであり、出発指定か到着指定かは時刻指定方法から判断できる。その他、時刻指定方法には始発指定と終発指定もある。始発指定の場合は移動希望日の 3 時を出発指定時刻として、終発指定の場合は移動希望日翌日の 3 時を到着時刻指定としてシステム内で経路が検索される。また、起終点に関する情報は起終点名だけでなく起終点種別、緯度・経度、住所など詳細に記録されている。本章では指定時刻、起終点名、起終点コードの情報をを用いる。

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2年間で、乗車駅および降車駅の双方でことでの駅が指定された検索は416,882件（571件/日）である。ここで、有効条件として、

- (1) 同一起終点を検索対象としていないこと、
- (2) 指定日時が上記2年以内であること

の2つを設定した結果、415,856件（570件/日）の有効データを得た。

表 3.1 経路検索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, 終発
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

ことでのんを対象とした検索状況を把握するために基礎集計分析を行なった。図 3.6 に時刻指定方法別検索割合を示す。最も選択された指定方法は出発時刻指定であり、検索の大半がこの指定方法によるものであることがわかった。次いで、到着時刻指定、終発指定、始発指定の順に検索割合が高い。ただし、終発指定と始発指定はほとんど選択されていないことがわかる。

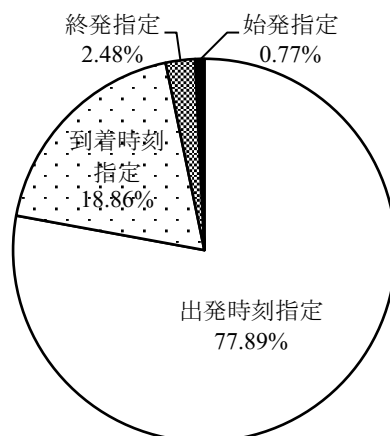


図 3.6 時刻指定方法別検索割合

図 3.7 には経路検索数の推移を示す。なお、ここでは経路検索数を月ごとに集計し、分析期間の経路検索数の増減傾向を把握する。日別経路検索数の推移については、第 5 章以降の分析で詳しく説明する。図 3.7 より、2013 年 8 月、2013 年 12 月、2014 年 3 月、2014 年 8 月、2014 年 12 月で経路検索数が大きく増えている。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ると考えられる。続いて、図 3.8 に曜日別平均経路検索数を示す。平日においては月曜日から木曜日までの平均経路検索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やや高いことがわかった。また、日曜日よりも土曜日の平均経路検索数の方が高く、7 曜日の中で土曜日の平均経路検索数が最も高い。さらに、平日の平均経路検索数の平均は約 77.3 件、休日の平均経路検索数の平均は約 91.6 件となり、休日の方が経路検索システムは利用されやすいといえる。

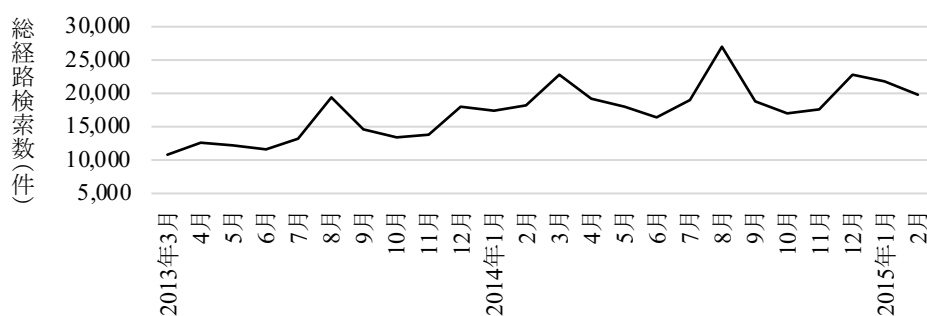


図 3.7 月別総経路検索数の推移

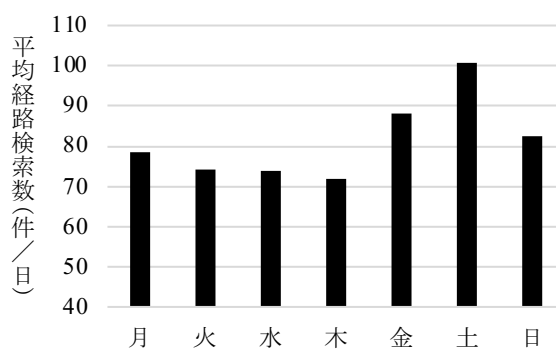


図 3.8 曜日別平均経路検索数

図 3.9 と図 3.10 にはそれぞれ乗車駅別平均経路検索数、降車駅別平均経路検索数を示す。なお、それぞれの図には平均経路検索数の高い上位 10 駅のみを表示している。両図より、乗車駅と降車駅ともに瓦町駅を指定する検索が最も多いことがわかった。瓦町駅はことでん 3 路線の結節駅であることから、この結果は自然といえる。次に、乗車駅と降車駅で挙げられている駅の違いに着目すると、上位 3

駅（瓦町駅，高松築港駅，片原町駅）は乗車駅と降車駅で違いはなく，4 位以下の駅では異なることがわかった。ただし，順位が異なるだけで，上位に挙げた駅は乗車駅と降車駅でほとんど同じである。また，各駅の平均経路検索数に大きな差はないことから，乗車駅と降車駅で検索されやすい駅に大きな違いはないといえる。これは検索者が往路，復路ともに経路検索システムを利用することに関連すると考えられる。さらに，仏生山駅は乗車駅のみで上位 10 駅に，綾川駅は降車駅のみで上位 10 駅に含まれることがわかった。

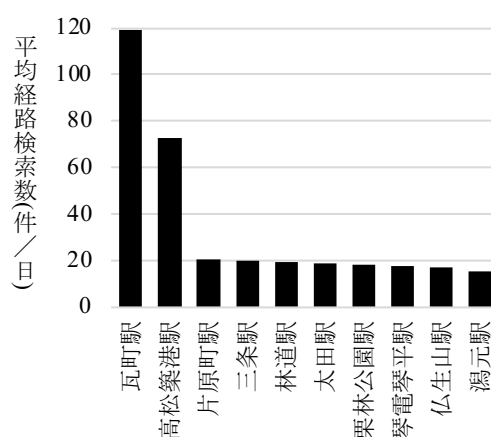


図 3.9 乗車駅別平均経路検索数（上位 10 駅）

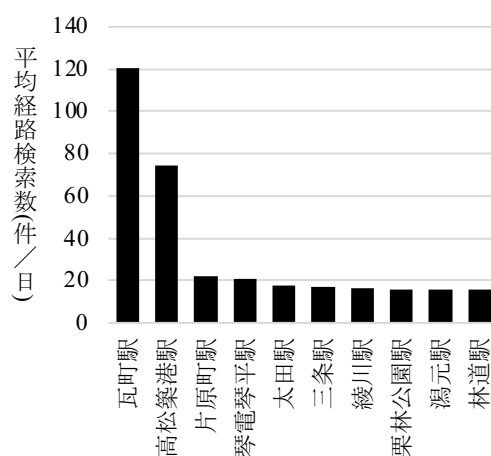


図 3.10 降車駅別平均経路検索数（上位 10 駅）

乗車駅・降車駅別に平均経路検索数が多い上位 40 組合せを表 3.2 に示す。表 3.2 より，「瓦町駅→高松築港駅」の検索が多い。次いで，「高松築港駅→瓦町駅」の検索が多いことから，高松築港駅と瓦町駅間は往路復路ともに検索されるといえる。また，上位 40 組合せのうち，25 組合せは乗車駅または降車駅のどちらかに瓦町駅を指定していた。これは瓦町駅が 3 路線の結節駅であることに起因すると考

えられる。さらに、上位 40 組合せでの経路検索数は総経路検索数の 32.03%を占めることがわかった。

ここでは、上位 40 組合せを示したが、実際には $52 \times 52 - 52 = 2,652$ の組合せが存在する。さらに、曜日ごとの特徴を捉えようとする、 $(52 \times 52 - 52) \times 7 = 18,564$ となり、組合せ爆発が発生する。これを回避し、データ構造を保持したまま分析するために、次節で説明する非負値テンソル因子分解を用いる。

表 3.2 乗車駅・降車駅別平均経路検索数 (2,652 組合せ中の上位 40 組合せ)

順位	乗車駅→降車駅	平均経路検索数 (件/日)	累積経路検索数 (件)	累積経路検索割合 (%)
1	瓦町駅 → 高松築港駅	10.33	7,538	1.81
2	高松築港駅 → 瓦町駅	9.70	14,621	3.52
3	林道駅 → 瓦町駅	7.20	19,879	4.78
4	瓦町駅 → 湊元駅	5.70	24,040	5.78
5	瓦町駅 → 林道駅	5.65	28,162	6.77
6	仏生山駅 → 瓦町駅	5.52	32,188	7.74
7	湊元駅 → 瓦町駅	5.17	35,961	8.65
8	太田駅 → 瓦町駅	5.13	39,707	9.55
9	瓦町駅 → 太田駅	5.01	43,364	10.43
10	三条駅 → 瓦町駅	4.97	46,991	11.30
11	瓦町駅 → 学園通り駅	4.94	50,599	12.17
12	瓦町駅 → 仏生山駅	4.90	54,173	13.03
13	高松築港駅 → 琴電琴平駅	4.83	57,700	13.87
14	高田駅 → 瓦町駅	4.83	61,227	14.72
15	瓦町駅 → 琴電志度駅	4.76	64,704	15.56
16	林道駅 → 高松築港駅	4.60	68,060	16.37
17	瓦町駅 → 高田駅	4.59	71,408	17.17
18	水田駅 → 瓦町駅	4.31	74,552	17.93
19	瓦町駅 → 琴電琴平駅	4.27	77,669	18.68
20	瓦町駅 → 水田駅	4.27	80,783	19.43
21	瓦町駅 → 長尾駅	4.14	83,806	20.15
22	琴電志度駅 → 瓦町駅	4.07	86,776	20.87
23	学園通り駅 → 瓦町駅	4.06	89,742	21.58
24	栗林公園駅 → 高松築港駅	3.86	92,562	22.26
25	高松築港駅 → 林道駅	3.80	95,333	22.92
26	三条駅 → 高松築港駅	3.78	98,095	23.59
27	瓦町駅 → 三条駅	3.77	100,847	24.25
28	農学部前駅 → 瓦町駅	3.77	103,596	24.91
29	太田駅 → 高松築港駅	3.67	106,277	25.56
30	瓦町駅 → 農学部前駅	3.65	108,942	26.20
31	琴電琴平駅 → 高松築港駅	3.58	111,555	26.83
32	高松築港駅 → 湊元駅	3.57	114,161	27.45
33	琴電屋島駅 → 瓦町駅	3.38	116,629	28.05
34	八栗駅 → 瓦町駅	3.36	119,084	28.64
35	瓦町駅 → 八栗駅	3.32	121,511	29.22
36	仏生山駅 → 高松築港駅	3.29	123,914	29.80
37	湊元駅 → 高松築港駅	3.19	126,242	30.36
38	高松築港駅 → 太田駅	3.19	128,569	30.92
39	瓦町駅 → 琴電屋島駅	3.18	130,889	31.47
40	高松築港駅 → 栗林公園駅	3.17	133,203	32.03

3.4 非負値テンソル因子分解

非負値テンソル因子分解 (NTF : Non-negative Tensor Factorization) は非負値テンソルデータ, つまりすべての要素が負値でないテンソルデータからパターンを抽出する手法である. 非負値テンソル因子分解は非負値行列因子分解 (NMF : Non-negative Matrix Factorization) (Lee et al., 1999) の拡張手法である. テンソルとは高次数の配列を表すもので, テンソルの概念ではスカラーは 0 次テンソル, ベクトルは 1 次テンソル, 行列は 2 次テンソルという. 次数 n のテンソルは n 次テンソルと表現でき, 例えば次数 3 のテンソルは 3 次テンソルと呼ぶ.

テンソルには要素数が増えやすい性質があり, 次数が 1 つ高くなるだけで要素数は膨大になる. すべての要素軸の次元 ($= I$) が同じである場合でも, 行列 (2 次テンソル) は要素数が $I \times I = I^2$ となるのに対して, 3 次テンソルは I^3 , n 次テンソルは I^n となる. I が小さい場合でも, 次数が上がるにつれて要素数は急激に増えていく. このような大規模性を持つデータを解釈するために, 行列の次元圧縮を応用したテンソル分解という手法がある (Vazifehdan et al., 2018 ; 松林ら, 2015). テンソル分解は 3 次数以上の高次数データを低次数データの積で近似し, 複雑な構造を持つデータを解釈のしやすい形に展開できる (Fan et al., 2017 ; Taneja and Arora, 2018). ここで, 分解後の結果は元のテンソルデータの要素軸の組合せに依存することなく等価であることが望ましい. これを分解の対称性という. 例えば, $I \times J \times K$ のテンソルと $J \times I \times K$ のテンソルにおいて, 本質的には同じデータを表すにも関わらず, 両者の次元圧縮が異なる結果を返すことは好ましくなく, 分解の対称性を満たす必要がある (石黒ら, 2016). また, パラメータ数の線形性も満たさなければならない (石黒ら, 2016). これは元のテンソルデータの要素数に対してパラメータの総数が大きくなることで, 推定精度がノイズに左右されやすくなることを避けるためである. 分解の対称性とパラメータ数の線形性を同時に満たすテンソル分解には CP 分解とタッカー分解の 2 つが主手法として挙げられる (石黒ら, 2016 ; Yao et al., 2015).

CP 分解 (Han et al., 2014) とタッカー分解 (Sun et al., 2016) は分析の目的に応じて使い分けが必要であり, CP 分解は各要素軸の関係がある程度独立すると仮定する場合に, タッカー分解は各要素軸の関係が相互に依存すると仮定する場合に適している.

図 3.11 を例に, 観光場所とそこを訪れる時点の関係を考える. 「冬はウィンタースポーツをするために寒冷地に行きたくなる」, 「夏は涼しさを求めて避暑地に行きたくなる」というように季節ごとに観光場所が固定されるとして両者の関係を求めたい場合は CP 分解を用いる. 一方で, ウィンタースポーツを楽しめる避暑地もあることから「冬にウィンタースポーツをするために避暑地に行きたくなる」

というように寒冷地と冬の関係だけでなく避暑地と冬の関係も求めたい場合はタッカー分解を用いる。タッカー分解はCP分解に比べ、各要素軸の関係を柔軟に表現できる利点を持つ。

経路検索データへの非負値テンソル分解の適用においては、曜日、乗車駅、降車駅の3属性の関係を事前に想定することは難しい。よって、3属性の相互関係に対して柔軟な表現が可能なタッカー分解を採用する。

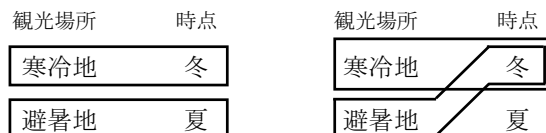


図 3.11 CP 分解とタッカー分解の違い (左 : CP 分解, 右 : タッカー分解)

3 次テンソルを例に、タッカー分解の概念図を図 3.12 に示す。次数が 3 で $I \times J \times K$ の大きさをもつ 3 次テンソル X を考える。タッカー分解は、それぞれ次元が I, J, K で与えられる 3 つの属性を持つ X について、 I 軸に関して次元を Q ($Q \leq I$) に、 J 軸に関して次元を R ($R \leq J$) に、 K 軸に関して次元を S ($S \leq K$) に落とす手法である。すなわち、 I および Q の軸に関する情報をもつ $I \times Q$ の行列 U 、 J および R の軸に関する情報をもつ $J \times R$ の行列 V 、 K および S の軸に関する情報を持つ $K \times S$ の行列 W に情報を分離する。この操作をまとめると、

$$X \cong C \times U^T \times V^T \times W^T \quad (3.1)$$

この分解をランク (Q, R, S) タッカー分解 (rank- (Q, R, S) Tucker decomposition) (石黒ら, 2016) と呼び、行列 U, V, W は特徴行列、 C はコアテンソルと呼ばれる。コアテンソル C は $Q \times R \times S$ の 3 次数で構成され、 Q, R, S は各属性に対する因子数となる。式(3.1)を要素毎に書くと、式(3.2)のようになる。

$$x_{ijk} \cong \sum_{q=1}^Q \sum_{r=1}^R \sum_{s=1}^S c_{qrs} u_{iq} v_{jr} w_{ks} \quad (3.2)$$

なお、タッカー分解では属性毎に異なる次元 (Q, R, S) に圧縮可能である。因子数 Q, R, S の同定方法には、モデル適合度を指標に決定する方法 (Goutte et al., 2010) と結果の解釈のしやすさによって決定する方法 (松林ら, 2015) が提案されているが、後者の方法が採用されることが多い。本研究では、非負値テンソル因子分解を将来予測値の算出や施策の評価のためではなく、経路検索パターンの特徴

を理解するために適用する。そのため、厳密な最適解を求めることよりも、計画論的解釈性が重要であり、結果を比較しながら試行錯誤的に探索し、各特徴行列がデータの特徴を表現する因子数を決定する。

タッカー分解はパラメータとして C, U, V, W を持っており、2乗誤差を最小にするように、非負値制約を与えてパラメータを推計する。つまり、3次テンソル X のランクタッカー分解は次のような問題として定式化される。

$$\begin{aligned} \min_{C,U,V,W} f(C,U,V,W) \\ f(C,U,V,W) = \|X - C \times U^T \times V^T \times W^T\|_F^2 \\ \text{s.t. } c_{qrs} \geq 0, u_{iq} \geq 0, v_{jr} \geq 0, w_{ks} \geq 0 \end{aligned} \quad (3.3)$$

ここで、 $\|\cdot\|_F$ はフロベニウスノルムを表す。非負値テンソル因子分解では、式(3.3)によって最適化された特徴行列 U, V, W とコアテンソル C の組合せにより、データを持つ特徴を把握する。また、特徴行列およびコアテンソルはそれぞれの要素が非負であることを条件とするため、推定される結果から直感的にデータの特徴を捉えることができる。

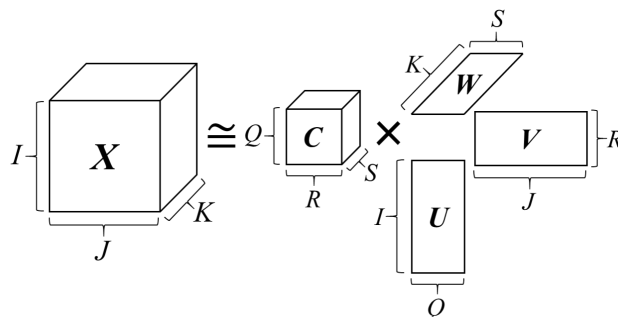


図 3.12 タッカー分解概念図 (3次テンソル)

3.5 ことでのんを対象とした検索パターンの抽出

本章では、検索者が指定した「曜日」「乗車駅」「降車駅」により3次数のテンソルデータを構築し、ことでのん検索パターンを把握する。具体的には、月曜日から日曜日の曜日7区分、乗車駅52駅、降車駅52駅の3属性を用いて、検索者が「どの曜日」に、「どの駅」から「どの駅」への移動を希望しているのかを明らかにする。「曜日」「乗車駅」「降車駅」によって構成される3次テンソルは、要素数が

$7 \times (52 \times 52 - 52) = 18,928$ の要素を持つ。この 3 次テンソルデータに非負値テンソル因子分解を適用させ、曜日、乗車駅、降車駅の各属性の特徴行列とコアテンソルを抽出する。各属性の因子数は、曜日は 3 因子、乗車駅は 4 因子、降車駅は 3 因子とした。コアテンソルのデータサイズは各属性の因子数と対応するため、 $3 \times 4 \times 3 = 36$ の要素数を持つ 3 次テンソルとなる。よって、18,928 の要素数を持つデータを 36 要素のデータに圧縮できたといえる。各属性の特徴行列とコアテンソルは同時に推定されるが、便宜上、3.5.1 項で特徴行列を、3.5.2 項でコアテンソルをそれぞれ分けて考察する。

3.5.1 属性別の特徴行列の結果と考察

(I) 曜日の特徴行列

曜日の特徴行列の結果を図 3.13 に示す。第 1 因子の各曜日に対する特徴行列の数値は「日曜日」が最も高く、次いで「火曜日」、「月曜日」の順に高い。「月曜日」と「火曜日」は第 1 因子以外では特徴行列の数値が低く、第 1 因子において相対的に数値が高い。一方で、第 1 因子の「水曜日」、「木曜日」、「金曜日」の特徴行列の数値は第 2 因子より低く、第 1 因子の「土曜日」の特徴行列の数値は 3 つの因子の中で最も低い。第 1 因子の各曜日に対する特徴行列の数値の相対的な高さに着目すると、第 1 因子は「月曜日」、「火曜日」、「日曜日」を表す因子といえる。週始めは月曜日とすることが多いが、カレンダー上では日曜日始まりが多いことから、第 1 因子は週の前半を表す因子とも考えられる。同様に、第 2 因子、第 3 因子についても特徴行列の数値を相対的に比較すると、第 2 因子は「水曜日」、「木曜日」、「金曜日」、第 3 因子は「土曜日」、「日曜日」を表す。各因子を表す曜日から、第 2 因子は週の後半、第 3 因子は休日を表す因子と言い換えられる。

したがって、曜日 7 区分は「週の前半」、「週の後半」、「休日」で表せられ、曜日によって異なる検索パターンが存在することが示唆された。

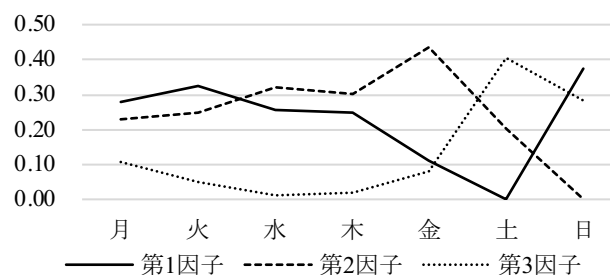


図 3.13 曜日の特徴行列の結果

(II) 乗車駅の特徴行列

乗車駅は駅数が 52 駅と数が多いため、すべての特徴行列の数値を比較し、考察することは複雑である。そこで、駅間の類似性を判断し、各因子の特徴を捉えるために、各因子の特徴行列の最大値に対して 50%以上の数値を示す駅のみを対象に考察する。

図 3.14 に考察の対象となる 16 駅の特徴行列の数値を示す。第 1 因子においては「高松築港駅」の特徴行列の数値が 27.803 で最大となるため、特徴行列の数値が 13.902 以上の駅を第 1 因子を説明する駅と考える。つまり、第 1 因子は「高松築港駅」、「仏生山駅」、「綾川駅」、「琴電琴平駅」、「林道駅」、「水田駅」、「池戸駅」、「学園通り駅」、「長尾駅」、「琴電屋島駅」、「八栗駅」、「琴電志度駅」の 12 駅が対象となり、これらによって第 1 因子は特徴付けられる。同様に、第 2 因子は「瓦町駅」、第 3 因子は「高松築港駅」、「三条駅」、「太田駅」、「仏生山駅」、「林道駅」、「瀧元駅」、第 4 因子は「高松築港駅」、「瓦町駅」で表せられると考える。

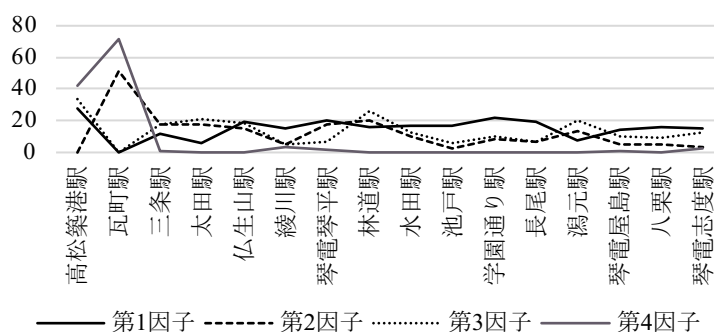


図 3.14 乗車駅の特徴行列の結果

図 3.14 より明らかになった各因子を表す駅をことでん路線図上で対応させると図 3.15 のようになる。図 3.15 において、各因子を表す駅の路線上の位置を黒丸で示しており、それぞれの路線図はその因子と関連性の高い駅のみを記している。例えば、左上の路線図は乗車駅第 1 因子を表す 12 駅のみを路線図上に示している。乗車駅第 1 因子を表す駅は路線図上の位置関係として空間的に広がりがあり、3 路線の終着駅を含んでいるのが特徴といえる。また、第 1 因子は乗車駅因子の中で最も多くの駅で構成されている。これらの駅周辺に、どのような施設が立地しているのか Web 検索を行なったところ、商業施設（スーパーやショッピングモール）や学校が共通して分布していることがわかった。このことから、第 1 因子を構成する駅を乗車駅とする検索は買い物や通学のような活動と関連する可能性が高い。乗車駅第 2 因子を表す「瓦町駅」は第 2 因子以外にも第 1 因子および第 4 因子にも関係しており、特定の検索パターンではなく、複数の検索パターンと関連する駅であることが示唆された。乗車駅第 3

因子を表す駅は高松市の中心市街地に近く、琴平線上の駅が多いことが特徴である。第3因子に属する駅周辺には住宅地の他に、金融機関や郵便局が共通して立地していることから、これらの施設の利用を目的とした移動のための検索との関連が考えられる。乗車駅第4因子を表す駅は高松市の中心市街地内に位置する駅であり、両駅の空間的近接性は高い。また、「高松築港駅」は第4因子以外にも第1因子および第3因子にも関係していることから、「瓦町駅」と同様に複数の検索パターンとの関連が示唆された。

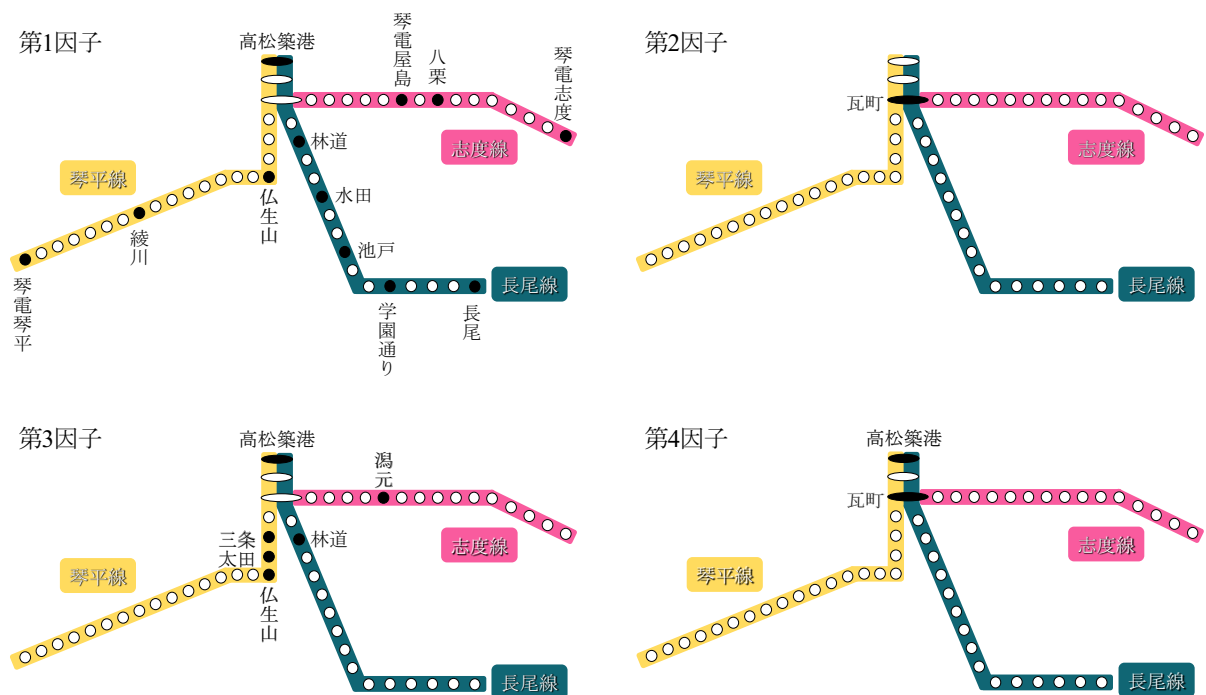


図 3.15 各乗車駅因子を構成する駅の位置関係

(III) 降車駅の特徴行列

降車駅の特徴行列についても乗車駅の特徴行列と同様に考察を行う。降車駅の特徴行列は、図 3.16 に示すように 14 駅が抽出され、第1因子は「高松築港駅」、第2因子は「瓦町駅」、第3因子は「三条駅」、「太田駅」、「仏生山駅」、「琴電琴平駅」、「林道駅」、「水田駅」、「高田駅」、「学園通り駅」、「長尾駅」、「沖松島駅」、「湯元駅」、「琴電志度駅」を表す因子であることがわかった。図 3.16 より明らかになった各因子を表す駅をことごとく路線図上で対応させると図 3.17 のようになる。「高松築港駅」と「瓦町駅」はそれぞれ単独で 1 つの因子を表し、乗車駅としては複数の検索パターンとの関連が示唆されたが、降車駅としては固有の検索パターンを持つことがわかった。これは乗車駅は異なるものの、降車

駅を「高松築港駅」あるいは「瓦町駅」とする検索が多いことが影響したと考えられる。一方，降車駅第3因子を表す駅は乗車駅第1因子と構成が類似しており，乗降駅と対応している可能性がある。

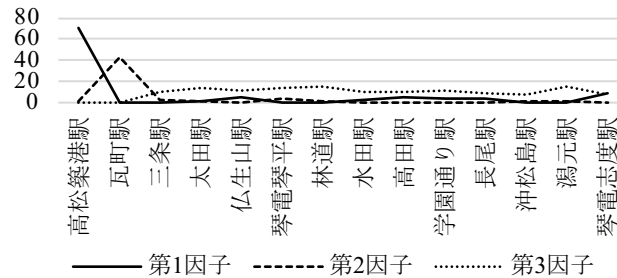


図 3.16 降車駅の特徴行列の結果

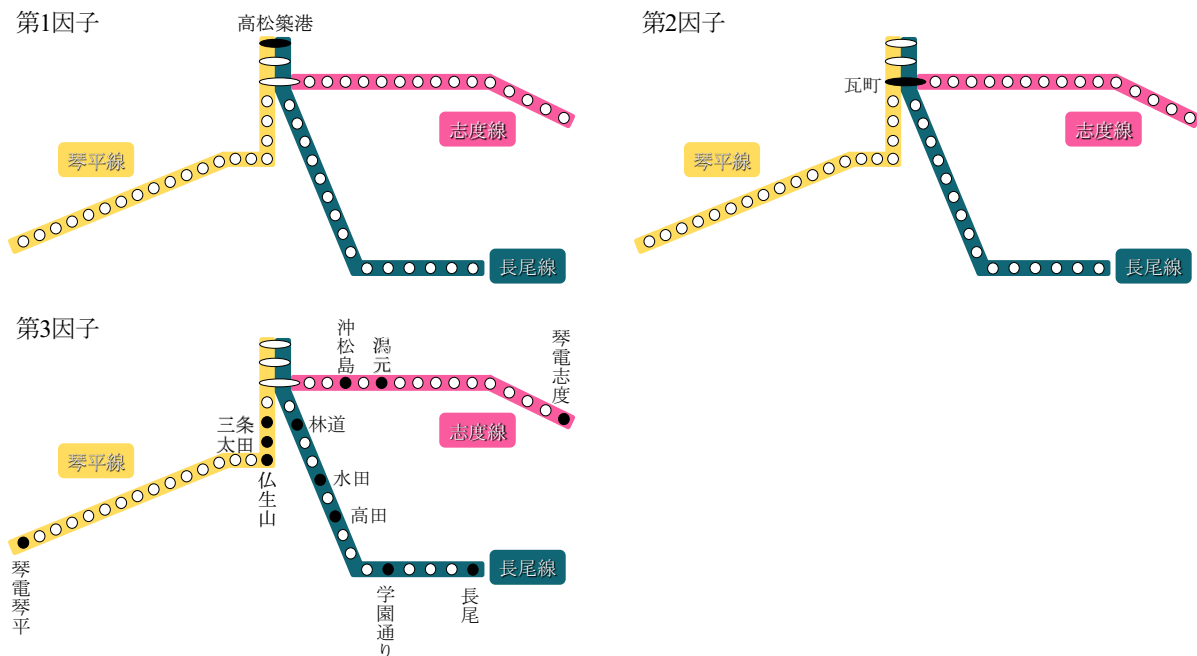


図 3.17 各降車駅因子を構成する駅の位置関係

3.5.2 コアテンソルの結果と考察

各属性の特徴行列の結果を踏まえて，コアテンソルの結果を考察する。コアテンソルは曜日，乗車駅，降車駅の因子の組合せを表し，どの曜日に，どの駅からどの駅への経路を指定した検索があったのかを把握できる。コアテンソルの要素数は因子の組合せの数だけ存在し，36 要素（3 曜日因子×4 乗車駅因子×3 降車駅因子）の検索パターンがある。図 3.18 に 36 要素のコアテンソルの結果を示す。図中

の縦軸の数字は因子の組合せであり，左から曜日因子，乗車駅因子，降車駅因子の順になっている．例えば，コアテンソル値の最も高い $(2, 3, 2)$ は曜日第 2 因子，乗車駅第 3 因子，降車駅第 2 因子の組合せであり，「水曜日，木曜日，金曜日」に「高松築港駅，三条駅，太田駅，仏生山駅，林道駅，湊元駅」から「瓦町駅」へ移動する経路の検索を表している．

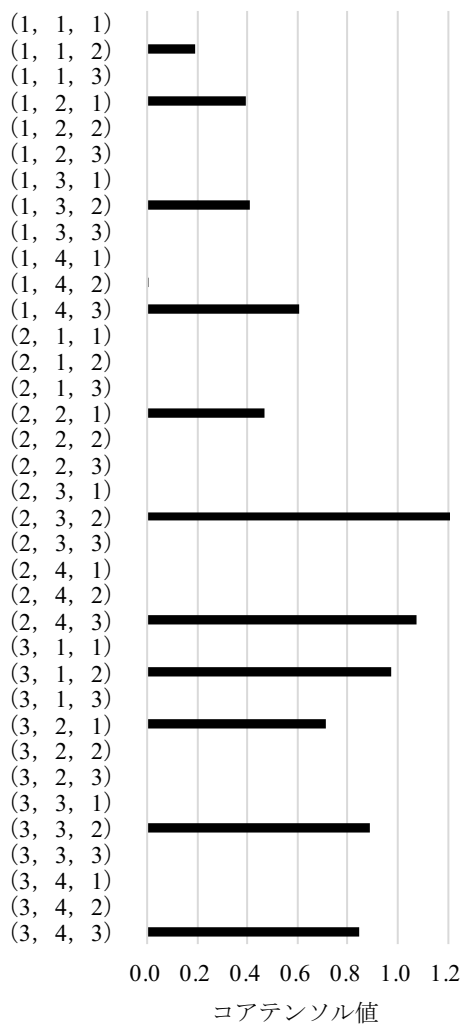


図 3.18 コアテンソルの結果

(注意：縦軸の括弧内の値は左から順に曜日，乗車駅，降車駅の因子番号を示す)

コアテンソル値が 0 となる要素に着目する．全 36 要素のうち 24 要素がコアテンソル値が 0 となり，これらの要素は検索者が指定することのない検索パターンを意味する．以降は，コアテンソル値が 0 を示す組合せは考察の対象外とし，正值をとる 12 要素について考察する．曜日因子ごとの非ゼロの組合せ数に着目すると， $(1, 1, 2)$ ， $(1, 2, 1)$ ， $(1, 3, 2)$ ， $(1, 4, 2)$ ， $(1, 4, 3)$ と曜日第 1 因子である

「週の前半（日曜日，月曜日，火曜日）」での組合せが最も多い．ただし，他の曜日因子の組合せのコアテンソル値と比較すると，これら 5 組合せのコアテンソル値は相対的に低い．したがって，週の前半では他の曜日因子に比べて，さまざまな乗降駅の組合せを指定した経路が検索されるが，その経路検索数は多くはないと考えられる．一方，週の後半（曜日第 2 因子）や休日（曜日第 3 因子）は非ゼロの組合せ数が週の前半よりは少ないものの，コアテンソル値が全体的に高いことから，特定の経路での検索が行われ，その経路検索数は多いといえる．

曜日因子ごとの乗降駅の組合せの違いに着目すると，どの曜日因子においても乗車駅第 2 因子と降車駅第 1 因子，乗車駅第 3 因子と降車駅第 2 因子，乗車駅第 4 因子と降車駅第 3 因子の組合せは非ゼロであった．つまり，これらの乗車駅因子と降車駅因子の組合せでの検索は一週間を通して日々行われている検索パターンと考えられる．ただし，コアテンソル値の高さから，週の前半では乗車駅第 4 因子と降車駅第 3 因子での組合せでの経路，週の後半および休日では乗車駅第 3 因子と降車駅第 2 因子での組合せでの経路が最もよく指定されると考えられる．乗車駅第 4 因子と降車駅第 3 因子は「高松築港駅，瓦町駅」から「三条駅，太田駅，仏生山駅，琴電琴平駅，林道駅，水田駅，高田駅，学園通り駅，長尾駅，沖松島駅，瀧元駅，琴電志度駅」を指定する組合せを表し，高松市中心部から郊外部への移動意向といえる．ことでん沿線上は郊外にいくほど住宅地が広がっていることを踏まえると，帰宅行動に関連した検索と考えられる．一方，乗車駅第 3 因子と降車駅第 2 因子は「高松築港駅，三条駅，太田駅，仏生山駅，林道駅，瀧元駅」から「瓦町駅」を指定する組合せを表し，瓦町駅周辺には商店街や病院，行政施設があることから，これら施設の利用を目的とした移動との関連が示唆された．

3.6 まとめ

3.6.1 3章の研究成果と次章への展開

本章では，2.3 節で仮定した仮説 1：「経路検索システムと交通系 IC カードは利用されやすい状況が異なる」の考え方は妥当性を検討する第 1 段階として，経路検索行動の特性を分析し，経路検索システムが利用される状況についての示唆を得ることを試みた．

経路検索システム「NAVITIME」の起終点にことでんを指定した経路検索データから，経路検索行動の特徴を捉えるために，非負値テンソル分解を用いた特徴抽出方法を提案した．ただし，分析期間にお

ける時点間の特徴の推移は捨象し、分析期間全体の代表的経路検索行動の抽出に注力した。曜日、乗車駅、降車駅で構成される3次元テンソルに対して、非負値テンソル因子分解を適用した結果、18,928要素（曜日 $7 \times$ 乗車駅 $52 \times$ 降車駅 52 ）を持つデータの情報量圧縮に成功し、最終的に12個の代表的な検索パターンを抽出できた。得られた検索パターンのうち、最も経路検索数の多い検索は、「水曜日、木曜日、金曜日（週の後半）」に、「高松築港駅、三条駅、太田駅、仏生山駅、林道駅、湊元駅」から「瓦町駅」を指定する検索、つまり、高松市の中心市街地近くに位置する駅から高松市中心部への移動に関する検索であることが明らかとなった。また、曜日因子ごとに抽出された検索パターンを考察すると、平日の経路検索数は多くはないものの、複数の乗降駅の組合せが検索されることがわかった。一方、休日は平日に比べ指定される乗降駅の組合せが少なく、特定の検索パターンが存在し、その経路検索数も多いことが明らかとなった。

このように、非負値テンソル因子分解を用いることによって、経路検索行動の主挙動として、3つの異なる属性情報を同時に考慮した検索パターンの把握が可能であることが示された。さらに、サンプル数や要素数などの情報量が大きなデータでも、比較的小さな因子数で元のデータの特徴を表現でき、非負値テンソル因子分解によるパターン抽出が交通ビッグデータにも有効であることが示された。非負値テンソル因子分解は画像処理の分野で分析手法として開発されたものであり、交通分野での適応事例は少ない。そのため、交通ビッグデータに対しても特徴抽出方法として有効性を示せたことは学術的に意義があると考えられる。

非負値テンソル因子分解はサンプル量に比例する代表的な経路検索行動を抽出する特性から、本事例においては高松市中心部に位置する駅、郊外部にある特定の駅で一定数検索が行われることが確認できた。しかし、本章では、分析期間において多く検索がなされる乗車駅や降車駅の組合せの概観に留まり、経路検索数が多い駅に共通する要因については駅周辺の施設立地状況をWeb検索し、定性的に把握したが、定量的には明らかにできていない。この経路検索数の多少に影響する要因の解明は、次の4章で行う。

3.6.2 3章における今後の課題

本章では、非負値テンソル因子分解を適用し、特徴行列を抽出した後、値が大きい要素にだけ注目し

て、考察している。特に、乗車駅因子と降車駅因子については、特徴行列の数値が小さい駅は無視しているため、すべての駅についてどの因子に属するか判断できていない。各因子に影響する要素のみを残した特徴行列、コアテンソルの抽出を行う方法としては、正規化項付きのテンソル分解 (Madrid-Padilla and Scott, 2017; Cichocki, 2011 ; 原田ら, 2021 ; Kim and Park, 2007; Eggert and Korner, 2004) が考案されている。しかし、正規化項付きテンソル分解は、結果が解釈しやすくなる一方で、得られる推定結果は元データから多くの情報を失うことになり、誤差が大きくなる。そのため、情報の損失量と結果の解釈のしやすさの関係から、どのくらいの情報を捨てても、妥当な結果が得られるのか検討の追加が必要である。

結果の計画論的解釈性を重視する場合、情報縮約を行なった後にサンプルをクラスタリングし、情報縮約で得られた結果の解釈や妥当性の検証がなされることもある。例えば、因子分析では因子が得られた後に、サンプルをクラスタリングし、グループごとに各因子の因子得点の平均値を算出し比較することで、サンプルグループと因子の関係を推測し、因子の解釈が妥当であるか否かの検討が行われる。しかし、多数の駅を同時に分析しようとする、全体としての平均的な特性のみが抽出されることになり、例えば駅周辺に住宅地が多い、あるいは事業所が多いなどで異なると考えられる駅間の異質性が考慮できない。そこで、他の情報量圧縮方法としてグラフ研磨の適用が考えられる。これは、全体としての特徴を保持しながら異質性を持つ駅ごとにグルーピングするクラスタリング手法であり、非負値テンソル因子分解における因子の解釈問題を解決できる可能性がある。

このように、影響力が小さいと考えられる要素をゼロとして算出する正規化項付きのテンソル分解、異質性に着目してサンプルをクラスタリングするグラフ研磨など、高次元・高次数データの情報量縮約手法はいくつか提案されており、それら分析方法の特質と分析目的に応じた方法論の使い分けを整理することは今後の重要な課題である。さらに、それらの分析方法が本分析で扱ったようなサンプル数が多いとはいえないビッグデータ分析の場合に、どの程度のサンプル数の少なさと特徴が抽出できなくなるのか、すなわちデータの薄さの限界も明確にする必要がある。

第3章 参考文献

- Cichocki, A. (2011), Tensor decompositions: new concepts in brain data analysis?, *Journal of the Society of Instrument and Control Engineers*, Vol.50, No.7, pp.507-516.
- Eggert, J. and Korner, E. (2004), Sparse coding and NMF, 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541), Vol.4, pp.2529-2533.
- Fan, H., Kuang, G., Qiao, L. (2017), Fast Tensor Principal Component Analysis via Proximal Alternating Direction Method with Vectorized Technique, *Applied Mathematics*, Vol.8, pp.77-86.
- Gauvin L., Panisson A., Cattuto C. (2014), Detecting the community structure and activity patterns of temporal networks: a non-negative tensor factorization approach, *PLOS ONE*, Vol.9, pp.e86028.
- Goutte, C. and Amini, M. (2010), Probabilistic tensor factorization and model selection, NIPS2010 Workshop on Tensors, *Kernels and Machine Learning*.
- Han, Y. and Moutarde, F. (2014), Analysis of large-scale traffic dynamics in an urban transportation network using non-negative tensor factorization, *International Journal of Intelligent Transportation Systems Research*, Vol.14, No.1, pp.36-49.
- Hashimoto, K., Iwasaki, T., Furukawa, T. (2013), Tensor Decomposition using Self-Organizing Map and Missing Data Estimation, *IEICE Technical Report*, Vol.112, No.390, NC2012-100, pp.37-42.
- Hosoe, M. and Kuwano, M. (2020), Public Transport Smart Card Data Analysis using Tucker Decomposition, *Proceedings of the 23rd Asia Pacific Symposium on Intelligent and Evolutionary Systems*, pp.52-66.
- Kim, H. and Park, H. (2007), Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis, *Bioinformatics*, Vol.23, Issue 12, pp.1495-1502.
- Lee, D. D. and Seung, H. S. (1999), Learning the parts of objects by non-negative matrix factorization, *Nature*, Vol.401, pp.788-791.
- Madrid-Padilla, OH. and Scott, J. (2017), Tensor decomposition with generalized lasso penalties, *Journal of Computational and Graphical Statistics*, Vol.26, No.3, pp.537-546.
- Simsekli, U., Virtanen, T., Cemgil, A. (2015), Non-negative tensor factorization models for Bayesian audio processing, *Digital Signal Processing*, Vol.47, pp.178-191.
- Sun, L. and Axhausen, K. (2016), Understanding urban mobility patterns with a probabilistic tensor factorization framework, *Transportation Research Part B*, Vol.91, pp.511-524.
- Taneja, A., Arora, A. (2018), Cross domain recommendation using multidimensional tensor factorization, *Expert Systems With Applications*, Vol.92, pp.304-316.
- Vazifehdan, M., Moattar, M.H., Jalali, M. (2018), A hybrid Bayesian network and tensor factorization approach for missing value imputation to improve breast cancer recurrence prediction, *Journal of King Saud University, Computer and Information Sciences*, 10 pages.
- Yamaguchi, H. and Nakayama, S. (2020), Detection of base travel groups with different sensitivities to new high-speed rail services: Non-negative tensor decomposition approach, *Transport Policy*, Vol.97, pp.37-46.
- Yao, D., Yu, C., Jin, H., Ding, Q. (2015), Human mobility synthesis using matrix and tensor factorizations, *Information Fusion*, Vol.23, pp.25-32.

- 石黒勝彦, 林浩平 (2016), 関係学習データ, 講談社, pp.133-154.
- 桐村喬 (2020), 非負値行列因子分解・非負値テンソル因子分解を用いた都市内部構造分析の試み, 人文地理学会大会研究発表要旨, pp.24-25.
- 久保基, 田中宏季, 中村哲 (2019), 非負値テンソル因子分解を用いた観光行動データからの情報抽出, マルチメディア, 分散, 協調とモバイルシンポジウム 2019 論文集, pp.1259-1263.
- 熊谷雄介, 今井良太, 松林達史, 佐藤吉秀, 堀岡力 (2015), 非負値複合テンソル因子分解を用いた訪日外国人観光客の回遊行動分析, 信学技報研究報告バイオ情報学 (BIO), Vol.2015-BIO-42, No.3, pp.1-5.
- ことでんグループ, ことでん路線図, <http://www.kotoden.co.jp/publichtm/kotoden/time/index.htm>, (2021年12月15日確認).
- 原田魁成, 山口裕通, 寒河江雅彦 (2021), スパース非負値行列因子分解を用いた COVID-19 流行期の県間旅行行動の変容分析, 土木学会論文集 D3 (土木計画学), Vol.77, No.2, pp.160-173.
- 細江美欧, 桑野将司, 長曾我部まどか, 福山敬, 石井晃 (2018), 災害時におけるマスメディアとソーシャルメディアの比較分析, 木学会論文集 D3 (土木計画学), Vol.74, No.5, pp.I_65-I_77.
- 細江美欧, 桑野将司, 谷本圭志 (2018), 非負値テンソル因子分解を用いた交通系 IC カードデータからの移動パターンの抽出に関する研究, 都市計画論文集, Vol.53, No.3, pp.1320-1326.
- 松林達史, 幸島匡宏, 林亜紀, 澤田宏 (2015), 非負値テンソル因子分解を用いた購買行動におけるブランド選択行動, 人工知能学会論文誌, Vol.30, No.6, SPI-C, pp.713-720.

経路検索システム利用者の移動目的の推察

〈第 4 章の概要〉

住宅が多い地区と事業所が多い地区では、通勤によって交通量（実行動者数）が多くなるというように、交通行動は都市構成要因である駅周辺環境と密接に関係していることが古くから指摘されている。しかし、経路検索数が多くなる要因については不明な点が多く、既往研究においては定量分析がなされていない。本章では、経路検索行動と駅周辺環境との関連性から、経路検索が多く行われる都市構成要因を明らかにする。

ここで、経路検索数は、(1) 乗車駅と降車駅の組合せによっては 0 件となる、(2) 中心部などの一部の駅を起終点とした経路検索数が多く、郊外部を起終点とした検索数が少ないという偏在性（スパース性）がある、という特徴がある。また、実行動とは異なり経路検索行動に関する研究蓄積がほとんどないため、モデルの構築においては、(3) 経路検索行動に影響を及ぼす駅周辺環境の要因に多数の候補が挙げられる。

これら 3 つの課題を同時に解決するために、本章では、Bayesian Adaptive Lasso Tobit 分位点回帰モデルを構築する。そして、提案手法を用いて経路検索数の多寡に影響を及ぼす要因を明らかにする。

4.1 はじめに

本章では、前章に続き、2.3節で仮定した仮説1の考え方は妥当か否かを検討するために、経路検索システムがどのようなときに利用されるのかを定量的に明らかにする。前章では、検索されやすい駅は明らかにできたが、検索の多さに影響する要因は把握できていない。そこで、本章では、経路検索が行われやすい場所の共通要因を分析し、経路検索システムが利用される状況を把握する。

本章の分析フローを図4.1に示す。本章では、どのような場所が検索されやすいのかを定量的に明らかにすることによって、経路検索システムの利用者がどのような移動目的を持って、検索にいたるのかを推察する。そのために、経路検索数の多寡は移動需要の多寡を表すと仮定し、駅周辺環境との関連性を分析する。ここで、駅周辺環境とはことでん路線上の駅周辺に立地する施設の数と人口分布を指す。乗降車駅周辺の特定の施設の有無から検索者の移動目的を考察した既往研究は存在するが、定性的な分析に留まり、統計的手法を用いて定量的に分析した研究はない。

仮説1:「経路検索システムと交通系ICカードは利用されやすい状況が異なる」の考え方は妥当なのか？

→ 経路検索システムがどのようなときに利用されるのかを把握する

第3章「いつ」「どこから」「どこへ」の移動が検索されやすいのか？

どのようなところが検索されやすいのか？

4.4節

経路検索数の多寡に影響する要因の分析方法の提案

問題意識：検索者の移動目的（経路検索システムの利用理由）は何か？

データの性質上の問題：(1)曜日ごとの乗車駅・降車駅別経路検索数の0件問題

データの性質上の問題：(2)曜日ごとの乗車駅・降車駅別経路検索数の分布の偏在性問題

モデルの推定上の問題：(3)大量の説明変数の選定問題

4.5節

検索者の移動目的からの経路検索システムの利用状況の解明

図4.1 第4章の分析フロー

本章では、経路検索データから曜日ごとに乗車駅・降車駅別平均経路検索数を算出し、これを平均的移動需要と定義する。そして、路線上の駅間で発生する平均的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，推定する．

ここで，本研究で分析対象とすることでん経路検索データは，地方都市におけるデータであることに起因し，

- (1) 一部乗車駅・降車駅区間では経路検索数が 0 件となる問題
- (2) 中心部に位置する駅と郊外部に位置する駅の経路検索数に大きな差が見られるというデータ分布の偏在性問題

がある．さらに，モデル構築の観点からは，

- (3) モデルに投入する説明変数の候補が大量にあるという説明変数選定問題

がある．

これら 3 つの問題点を同時に回避するために，Bayesian Adaptive Lasso Tobit 分位点回帰モデルを構築する点が，本章の方法論的特徴である．そして，モデル推定により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経路検索データが表す平均的移動需要がどのような移動目的によるものかを推察する．

4.2 経路検索数の分布の特徴を踏まえたモデル構築

前章で用いた経路検索データに基づいて，平均的移動需要と駅周辺環境の関係を記述するモデルを構築する．経路検索データ 415,856 件により，まず，分析期間内の曜日ごとの乗車駅・降車駅別平均経路検索数（以降，平均経路検索数）を算出する．

ことでんは計 52 駅を持つため，乗車駅と降車駅の組合せは $52 \times 52 = 2,704$ から同一起終点を除いた 2,652 組合せ存在する．さらに，曜日ごとに組合せを考えると $2,652 \times 7 = 18,564$ 組合せある．このうち，平均経路検索数が 0 の組合せは 6,612 あり，全 18,564 組合せの約 36% を占める．

一般に目的変数を複数の説明変数で説明する際には重回帰モデルが用いられるが，目的変数に 0 が多く含まれる場合，最小二乗法に基づく重回帰モデルでは予測値が負値となる可能性がある（村上ら，2015；水落，2009）．これが，「(1) 一部乗車駅・降車駅区間では経路検索数が 0 件となる問題」である．予測値に負値が多くなると，重回帰モデルにおける誤差項が正規分布するという仮定が満たされなくなる．誤差項の正規性を満たすためには，目的変数を対数変換することが考えられる．しかし，この方法では，0 は対数変換できないために，0 データをあらかじめ分析から除外しなければならない．これでは，0 の情報，つまり値が発生しない（本分析では，検索が発生しない）という状態を無視する

ことになり、分析方法として不適切である。0 が対数変換できないことを回避するために、目的変数のすべての観測値に 1 を足す方法もあるが、この方法が適切であるという根拠はない（山村，2002）。そこで、本章では Type I Tobit モデルに着目する。Type I Tobit モデルでは未観測の潜在変数を仮定し、潜在変数が一定水準を上回るときはそのまま観測されるものの、一定水準以下のときには特定の値に変化され観測されるような目的変数を扱える（山本，2015）。Type I Tobit モデルを用いることによって、一定水準以上では目的変数と説明変数に関係性があり、それ以下では目的変数の値が 0 となることを表現するモデルを構築できる。

次に、平均経路検索数は、主要駅間の経路検索数は多いが、郊外部の利用者の少ない駅間、特に異なる路線の郊外部の駅間の経路検索数は少なくなるため、経路検索数が多い少数の駅の組合せと、経路検索数が少ない多数の駅の組合せが発生し、平均経路検索数の分布は右に裾が長くなると考えられる。この特徴は、前節の表 3.2 で示した通りであり、乗車駅・降車駅の全組合せ 2,652 のうち、上位 40 組合せを構成する限られた駅に関する経路検索数で、全体の経路検索数の約 30% を占める。

これが、「(2) 中心部に位置する駅と郊外部に位置する駅の経路検索数に大きな差が見られるというデータ分布の偏在性問題」である。右に裾が長い分布に対して平均値を予測する回帰モデルを適用すると外れ値の影響を受ける。そこで、分位点回帰モデルを用いる。分位点回帰モデルは任意の分位点を予測する回帰モデルであり、中央値を分位点にすることで、外れ値の影響を回避しながら分布の中心を推定できる（石黒，2013）。

以上、(1) 0 件問題と (2) 偏在性問題を回避するために、Type I Tobit モデルと分位点回帰モデルを組み合わせた Tobit 分位点回帰モデルを用いて、平均的移動需要と駅周辺環境の関係を記述するモデルを構築する。

4.3 平均経路検索数に影響を及ぼす駅周辺環境要因の候補

4.3.1 駅周辺環境に関するデータ

本章ではことでの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（長尾ら，2010；大賀ら，2017）を参考に、駅から半径 500m 圏内を駅勢圏と

して定義し、人口および施設数を集計した。

人口分布データは e-Stat（総務省統計局）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（国土交通省）および経済産業省が提供する商業統計（経済産業省）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、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、駅ごとに駅勢圏内の施設数をそれぞれ集計した。それぞれの概要については表 4.1 および表 4.2 に示す。

表 4.1 各施設の内容

施設	内容
事業所	経済活動が単一の経営主体のもとで一定の場所を占めて行われており、物の生産や販売およびサービスの提供が、従業者と設備を有して継続的に行われている場所
市町村役場	市役所、町役場、村役場、およびそれらの支所、出張所、連絡所
医療機関	医療法に基づく病院、診療所、歯科診療所
福祉施設	高齢者福祉施設、障害者福祉施設、児童福祉施設
文化施設	美術館、資料館、博物館、動物園、水族館、スポーツ施設、図書館など
学校	学校教育法に基づく小学校、中学校、中等教育学校、高等学校、高等専門学校、短期大学、大学、特別支援学校
都市公園	都市公園法で定められている都市公園
観光資源	公益財団法人日本交通公社が設置した観光資源評価委員会が検討・選定し作成した観光資源台帳に掲載されている観光資源のうち評価ランクが A 級以上のもの (例:金刀比羅宮、栗林公園、瀬戸大橋など)
集客施設	アトラクションや展示会などが催事できる空間を有する施設および興業、スポーツなどが観覧できる「観客席」を有する施設など (例:イオンシネマ綾川、牟礼コミュニティセンターなど)

表 4.2 各施設の内容

施設	内容
道の駅	道路利用者のための休憩機能，道路利用者や地域住民のための情報発信機能，町と町が手を結び活力ある地域づくりを共に行うための地域の連携機能の 3 つの機能を併せ持つ休憩施設として設置されている道の駅
都道府県指定文化財	都道府県が文化財保護法第 182 条第 2 項に基づき，国指定などの文化財以外の重要な文化財 (例:金剛誕生釈迦仏立像，絹本着色西谷藤兵衛肖像など)
総合スーパー	衣食住にわたる各種商品を小売するスーパーで，そのいずれかが主たる販売商品であるか判別できない中・大型総合スーパー
専門スーパー	衣料品スーパー，食料品スーパー，ホームセンターを含む住関連スーパーのいずれかに該当する事業所
コンビニ	飲食物品を扱い，売場面積が 30 m ² 以上 250 m ² 未満で，営業時間が 14 時間以上の事業所
ドラッグストア	主として衣料品，化粧品を中心とした健康および美容に関する各種商品を中心として，家庭用品，加工食品などの最寄品をセルフサービス方式によって小売する事業所
その他のスーパー	総合スーパー，専門スーパー，コンビニ，ドラッグストア以外の事業所で，売場面積の 50%以上についてセルフサービス方式を採用している事業所
専門店	取扱商品の 90%以上が衣料品である衣料品専門店，取扱商品の 90%以上が食料品である食料品専門店，取扱商品の 90%以上が住関連商品（自動車，家具，花・植木など）である住関連商品専門店のいずれかに該当する事業所
家電大型専門店	売場面積 500 m ² 以上で，電気機械器具や電気事務機械器具を取り扱う事業所
中心店	百貨店，専門店，家電大型専門店，無店舗販売以外の事業所で，取扱商品の 50%以上が衣料品である衣料品中心店，取扱商品の 50%以上が食料品である食料品中心店，取扱商品の 50%以上が住関連商品（自動車，家具，花・植木など）である住関連商品店のいずれかに該当する事業所
その他の小売店	百貨店，専門店，家電大型専門店，中心店，無店舗販売以外の事業所で，セルフサービス方式を採用していない事業所
無店舗販売	訪問販売，通信・カタログ販売，インターネット販売，自動販売機によって商品を販売する事業所

4.3.2 Adaptive Lasso を用いた変数選択

平均経路検索数を記述するモデルの構築においては，経路検索数の多寡に影響を及ぼす要因に関する既往研究がないため，説明変数は，できるだけ多くを候補とするべきであり，前項に示したような合計 55 種類の駅周辺環境データを収集した。しかし，ここで「(3) モデルに投入する説明変数の候補が大量にあるという説明変数選定問題」が発生する。

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、現象を模倣するうえでの精密さと単純さの2つの観点からモデルを特定し、パラメータの推計が必要である。

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。逐次選択法は既往研究において多用されているが、選択される結果の不安定性が指摘されている（川野ら，2010；荒木，2013）。この問題を解決する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 L1 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（川野ら，2018）、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、得られたモデルに寄与する説明変数のみが残る。しかし、Lasso には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という問題が指摘されており（川野ら，2010；Zou, 2006）、この問題に対応するために、さらに Lasso を改良した Adaptive Lasso が提案されている（Zou, 2006）。この Adaptive Lasso を用いることで (3) 説明変数選定問題を解決する。

4.4 Bayesian Adaptive Lasso Tobit 分位点回帰モデル

(1) 0 件問題，(2) 偏在生問題，(3) 説明変数選定問題を同時に回避するために、Adaptive Lasso 正則化項付き Tobit 分位点回帰モデルを用いる。パラメータ推計に際しては、Alhamzawi (2013) が提案したベイズ推定を用いた方法を適用する。Alhamzawi (2013) の研究では Tobit 分位点回帰モデルのパラメータ推計に最尤推定とベイズ推定を適用し、各推定方法による予測精度を比較した結果、ベイズ推定の方が高精度であると示している。

Tobit 分位点回帰モデルは式(4.1)のように表すことができる。 y_i は目的変数、 y_i^* は i 番目のサンプルに対応する潜在変数、 \mathbf{x}_i は i 番目のサンプルに対応する $k \times 1$ の説明変数ベクトルとし、その転置ベクトルとして $\mathbf{x}'_i = (x_{1i}, x_{2i}, \dots, x_{ki})$ とする。 $\boldsymbol{\beta}_p = (\beta_{1p}, \beta_{2p}, \dots, \beta_{kp})'$ はすべてのサンプルに共通する $k \times 1$ のパラメータベクトルであり、分位点 $p \in (0,1)$ に依存する。本研究では、分位点には中央値 ($p = 0.5$) を設定する。また、 ε_i は i 番目のサンプルに対応する誤差項であり、 $P(\varepsilon_i \leq 0 | \mathbf{x}_i) = p$ を満たすものとする。

$$\begin{aligned} y_i^* &= \mathbf{x}'_i \boldsymbol{\beta}_p + \varepsilon_i \quad (i = 1, \dots, n) \\ y_i &= C(y_i^*) \\ &= \max\{y^0, \mathbf{x}'_i \boldsymbol{\beta}_p + \varepsilon_i\} \end{aligned} \tag{4.1}$$

ここで、 C はリンク関数であり、 y^0 は潜在変数が観測される閾値である。閾値 y^0 は 0 と仮定されることが多いが、任意の定数を仮定してもよい。

パラメータベクトル $\boldsymbol{\beta}_p$ の推定値は以下の損失関数の最小化解とする。図 4.2 に損失関数を示す。

$$\arg \min_{\boldsymbol{\beta}_p} \sum_{i=1}^n \rho_p(y_i - C(\mathbf{x}'_i \boldsymbol{\beta}_p)) \tag{4.2}$$

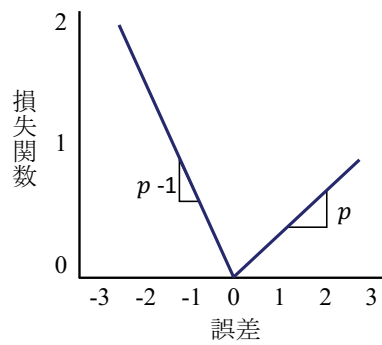


図 4.2 Tobit 分位点回帰モデルの損失関数

ここで、関数 $\rho_p(u)$ は、

$$\rho_p(u) = (p - \delta(u \leq 0))u \tag{4.3}$$

で定義される。 $\delta(\cdot)$ はデルタ関数である。

次に、式(4.2)に正則化項を付与した最小化問題を考える。式(4.2)に基づいて、Tobit 分位点回帰モデルの Adaptive Lasso 推定量は次式で与えられる。

$$\arg \min_{\boldsymbol{\beta}_p} \sum_{i=1}^n \rho_p(y_i - C(\mathbf{x}_i' \boldsymbol{\beta}_p)) + \sum_{j=1}^k \lambda_j |\beta_{jp}| \quad (4.4)$$

$\lambda_j (j = 1, \dots, k)$ は非負の正則化パラメータであり、式(4.4)の第2項が変数選択とパラメータ推定のための正則化項となる。一般に正則化パラメータ λ_j を大きくすると 0 と推定されるパラメータが多くなり、小さくすると 0 と推定されるパラメータが少なくなる。

式(4.4)の最小化問題を考えるにあたり、Alhamzawi (2013) は式(4.4)を最小化することと、誤差項に非対称ラプラス分布 (Asymmetric Laplace Distribution ; 以降, ALD) を仮定して最尤推定を行うことは同値であることに着目し、ALD に基づいたベイズ分析を提案した。ここで、誤差項が従う ALD は、

$$f(\varepsilon|p) = \frac{p(1-p)}{\sigma} \exp\left\{-\rho_p\left(\frac{\varepsilon}{\sigma}\right)\right\} \quad (4.5)$$

で与えられ、 $\rho_p(\cdot)$ は式(3)で定義される関数と同じである。 $\sigma > 0$ は尺度パラメータを表す。また、パラメータ $\boldsymbol{\beta}_p$ の事前分布として、以下のラプラス分布を仮定する。

$$\pi(\boldsymbol{\beta}_p | \lambda_1, \lambda_2, \dots, \lambda_k) = \prod_{j=1}^k \frac{\lambda_j}{2} \exp\{-\lambda_j |\beta_{jp}|\} \quad (4.6)$$

よって、式(4.5)と式(4.6)からパラメータ $\boldsymbol{\beta}_p$ の条件付分布は

$$\begin{aligned} P(\boldsymbol{\beta}_p | \mathbf{y}, \mathbf{X}, \boldsymbol{\lambda}) & \propto \prod_{i=1}^n \exp\left\{-\rho_p\left(\frac{\varepsilon_i}{\sigma}\right)\right\} \times \prod_{j=1}^k \exp\{-\lambda_j |\beta_{jp}|\} \\ & \propto \exp\left\{-\sum_{i=1}^n \rho_p(y_i - C(\mathbf{x}_i' \boldsymbol{\beta}_p)) - \sum_{j=1}^k \lambda_j |\beta_{jp}|\right\} \end{aligned} \quad (4.7)$$

となる。 $\mathbf{y} = (y_1, y_2, \dots, y_n)'$, $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)'$, $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_n)'$ である。この設定の下、式(4.7)の $\boldsymbol{\beta}_p$

に関する最大化と、式(4.4)の最小化は同等のものになる。

さらに、式(4.5)の ALD は正規分布の尺度混合分布として表すことができる (Alhamzawi, 2013; Kozumi and Kobayashi, 2011). すなわち、 $y_i|v_i, \sigma \sim N(\mathbf{x}'_i\boldsymbol{\beta} + (1-2p)v_i, 2\sigma v_i)$ を仮定すると、 v_i が比率パラメータ $p(1-p)/\sigma$ を持つ指数分布に従うとき、 y_i は ALD に従う。ここで、指数分布の確率密度関数は次式で与えられる。

$$P(v_i; \theta) = \theta \exp\{-\theta v_i\} \quad (4.8)$$

θ は $p(1-p)/\sigma$ で表現される比率パラメータである。これより、 \mathbf{y} の尤度は以下の式で表せられる。

$$P(\mathbf{y}|\boldsymbol{\beta}, \sigma, \mathbf{X}) \\ \propto \prod_{i=1}^n \int_0^{\infty} P(\mathbf{y}|\boldsymbol{\beta}, \sigma, \mathbf{X}, v_i) P\left(v_i; \frac{p(1-p)}{\sigma}\right) dv_i \quad (4.9)$$

さらに、式(4.9)は次のように展開できる。

$$P(\mathbf{y}|\boldsymbol{\beta}, \sigma, \mathbf{X}) \\ = \prod_{i=1}^n \int_0^{\infty} \frac{1}{\sqrt{4\sigma\pi v_i}} \times \exp\left\{-\frac{(y_i - C(\mathbf{x}'_i\boldsymbol{\beta}_p) - (1-2p)v_i)^2}{4\sigma v_i}\right\} \\ \times \frac{p(1-p)}{\sigma} \exp\left(\frac{p(1-p)}{\sigma} v_i\right) dv_i \quad (4.10) \\ = \prod_{i=1}^n \int_0^{\infty} \frac{p(1-p)}{\sigma\sqrt{4\sigma\pi v_i}} \times \exp\left\{-\frac{(y_i - C(\mathbf{x}'_i\boldsymbol{\beta}_p) - (1-2p)v_i)^2}{4\sigma v_i} - \frac{p(1-p)}{\sigma} v_i\right\} dv_i$$

また、式(4.6)のラプラス分布型事前分布も正規分布の尺度混合分布として表すことができる (Alhamzawi, 2013; Andrews and Mallows, 1974). $\beta_{jp}|s_j \sim N(0, s_j)$ を仮定すると、 s_j が $\text{Exp}(\lambda_j^2/2)$ に従うときに、 β_{jp} の事前分布はラプラス分布に従う。Alhamzawi (2013) の提案した手法ではベイズ推論の枠組みで考えるために、 λ_j^2 の事前分布を式(4.11)のガンマ分布で与え、

$$\pi(\lambda_j^2) \propto (\lambda_j^2)^{a-1} \exp\{-b\lambda_j^2\} \quad (4.11)$$

さらに、 σ の事前分布には

$$\pi(\sigma) \propto (\sigma)^{-c-1} \exp\left\{-\frac{d}{\sigma}\right\} \quad (4.12)$$

の逆ガンマ分布を設定する.

以上を整理すると, Tobit分位点回帰の階層ベイズモデルが以下のように与えられる.

$$\begin{aligned} y_i &= \max\{y^0, \mathbf{x}'_i \boldsymbol{\beta}_p + \varepsilon_i\}, \quad (i = 1, \dots, n) \\ y_i^* | \boldsymbol{\beta}_p, v_i &\sim N(\mathbf{x}'_i \boldsymbol{\beta} + (1 - 2p)v_i, 2\sigma v_i) \\ v_i &\sim \exp(p(1 - p)) \\ \boldsymbol{\beta}_{jp} | s_j &\sim N(0, s_j) \\ s_j &\sim \exp\left(\frac{\lambda_j^2}{2}\right) \\ \lambda_j^2 &\propto (\lambda_j^2)^{\alpha-1} \exp\{-b\lambda_j^2\} \\ \sigma &\propto (\sigma)^{-c-1} \exp\left\{-\frac{d}{\sigma}\right\} \end{aligned} \quad (4.13)$$

この階層ベイズモデルではMCMC法を用いてすべてのパラメータを更新する. 式(4.13)の下で, 潜在変数 y_i^* の完全事後分布は打ち切り正規分布になる.

$$y_i^* | y_i, \boldsymbol{\beta}_p, \sigma, \mathbf{x}_i, v_i \sim \begin{cases} F(y_i), & \text{if } y_i > y^0 \\ M, & \text{otherwise} \end{cases} \quad (4.14)$$

ここで, $F(y_i)$ は退化分布, $M = N(\mathbf{x}'_i \boldsymbol{\beta}_p + (1 - 2p)v_i, 2\sigma v_i) \delta(y_i^* \leq y_i^0)$ である. また, $\boldsymbol{\beta}_p$ の完全事後分布は平均 $\boldsymbol{\mu}_{\boldsymbol{\beta}_p} = \boldsymbol{\Sigma}_{\boldsymbol{\beta}_p} \mathbf{X}' \mathbf{V} (\mathbf{y} - (1 - 2p)\mathbf{v})$, 分散 $\boldsymbol{\Sigma}_{\boldsymbol{\beta}_p} = (\mathbf{X}' \mathbf{V} \mathbf{X} + \mathbf{S})^{-1}$ の多変量正規分布になる. ここで, $\mathbf{V} = \text{diag}((2\sigma v_1)^{-1}, (2\sigma v_2)^{-1}, \dots, (2\sigma v_n)^{-1})$, $\mathbf{S} = \text{diag}(s_1^{-1}, s_2^{-1}, \dots, s_k^{-1})$ である. 潜在変数 v_i^{-1} の完全事後分布は平均 $\mu = |y_i^* - \mathbf{x}'_i \boldsymbol{\beta}|^{-1}$, 形状パラメータ $\nu = (2\sigma)^{-1}$ の逆ガウス分布である. ここで, 逆ガウス分布の確率密度関数は以下の式で与えられる.

$$P(x | \nu, \mu) = \sqrt{\frac{\nu}{2\pi}} x^{-3/2} \exp\left\{\frac{\nu(x - \mu)^2}{2(\mu)^2 x}\right\} \quad (4.15)$$

ただし、 $x > 0$ とする。そして、 s_j^{-1} の完全事後分布は平均 $\mu = \lambda_j^2/\beta_j^2$ 、形状パラメータ $v = \lambda_j^2$ の逆ガウス分布になる。正規化パラメータ λ_j^2 は形状パラメータ $1 + a$ 、比率パラメータ $s_j/2 + b$ のガンマ分布に従う。最後に、尺度パラメータ σ は形状パラメータ $3n/2 + c$ 、比率パラメータ $\sum_{i=1}^n ((y_i - \mathbf{x}'_i\boldsymbol{\beta} - \xi v_i)^2/(4v_i) + p(1-p)v_i) + d$ の逆ガンマ分布に従う。

4.5 モデルの推定結果と考察

目的変数に曜日ごとの乗車駅・降車駅別平均経路検索数を、説明変数に乗車駅と降車駅間の距離、月曜日を基準とした曜日ダミー変数、駅勢圏内の人口および施設数を投入し、Bayesian Adaptive Lasso Tobit 分位点回帰モデルを推定した結果を表 4.3 に示す。4.3.1 項で記した通り、サンプル数 (=7 曜日×(52 駅×52 駅 - 同一起終点 52 駅)) は 18,564 である。

表 4.3 より、投入した説明変数 55 個のうち、推定値が「0」となった説明変数は 14 個あった。これらは目的変数に影響しない説明変数と解釈できる。一方、全説明変数の約 75%を占める 44 個の説明変数がモデルに寄与するものとして選択された。一般に誤差項に正規分布を仮定したモデルを用いて推定されたパラメータの統計的有意性を測る指標として t 値が用いられるが、Adaptive Lasso のような正規化法では推定値が 0 でない説明変数は統計的有意性があると判断される。

0 でない説明変数に着目すると、駅間距離は負のパラメータとして推定された。乗車駅と降車駅の距離が離れるほど、平均経路検索数は減少するといえる。経済学で空間的な位置関係によって人の移動と施設の分布を把握するために用いられる重力モデルやハフモデルでは、施設間の距離が離れるほど人は移動しないと仮定される(松本と北村, 1991; 金子ら, 2011)。したがって、この結果は自然である。曜日ダミーについては火曜日と水曜日が負のパラメータとなった。これは月曜日の平均経路検索数に対して、火曜日と水曜日の平均経路検索数が減少することを表す。図 4.3 に示す曜日別平均経路検索数の結果(3.3 節の図 3.8 再掲)より、この結果も妥当といえる。次に、乗車駅と降車駅ともに負のパラメータとして推定された人口と施設に着目すると、「14 歳以下人口」、「外国人人口」、「都市公園」、「学校」、「専門スーパー」が挙げられる。これらの変数は平均経路検索数の減少に寄与する施設といえる。特に、「学校」が負のパラメータとなったのは、通学を移動目的とする人は毎日決まった経路を通るため、乗車駅あるいは降車駅周辺に学校がある経路の検索は少なくなる傾向にあるといえる。乗車駅のみで負のパラメータとして推定された施設に着目すると、「市町村役場」、「集客施設」、「道の駅」、「コンビニ」、「その他の小売店」が挙げられる。一方、降車駅のみで負のパラメータとして推定された

施設には、「事業所」、「専門店」がある。

表 4.3 推定結果

説明変数	推定値	説明変数	推定値
定数項	-1.561		
駅間距離	-0.419		
<曜日ダミー>			
火曜日	-0.086	金曜日	0
水曜日	-0.081	土曜日	0.187
木曜日	0	日曜日	0
<乗車駅半径 500m 圏内>		<降車駅半径 500m 圏内>	
都市公園	-0.466	都市公園	-0.564
医療機関	0.616	医療機関	1.095
福祉施設	0	福祉施設	0
文化施設	0.218	文化施設	0.233
学校	-0.183	学校	-0.191
市町村役場	-0.050	市町村役場	0
14 歳以下人口	-0.685	14 歳以下人口	-1.174
15 歳以上 64 歳以下人口	0.992	15 歳以上 64 歳以下人口	1.595
65 歳以上人口	0	65 歳以上人口	0
外国人口	-0.466	外国人口	-0.560
集客施設	-0.084	集客施設	0.099
道の駅	-0.081	道の駅	0.078
都道府県指定文化財	0.278	都道府県指定文化財	0.229
事業所	0	事業所	-0.748
総合スーパー	0	総合スーパー	0
専門スーパー	-0.065	専門スーパー	-0.073
コンビニ	-0.072	コンビニ	0.098
ドラッグストア	0.139	ドラッグストア	0.098
その他のスーパー	0.322	その他のスーパー	0.489
専門店	0	専門店	-0.268
家電大型専門店	0.102	家電大型専門店	0.110
中心店	0.230	中心店	0
その他の小売店	-0.487	その他の小売店	0
無店舗販売	0.155	無店舗販売	0.069
サンプル数		18,564	

続いて、正のパラメータとして推定された曜日ダミーには「土曜日」が挙げられる。この結果から、月曜日の平均経路検索数に対して、土曜日の平均経路検索数が増加することが明らかとなり、非日常的な活動が発生しやすいと考えられる休日と経路検索行動との関連が裏付けられた。乗車駅と降車駅ともに正のパラメータとなった人口施設には、「15 歳以上 64 歳以下人口」、「医療機関」、「文化施設」、「都道府県指定文化財」、「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。さらに、乗車駅のみで「中心店」、降車駅のみで「集客施設」、「道の駅」、「コンビニ」が正のパラメータとして推定された。これらの説明変数は平均経路検索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」が正で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と関連性があると考えられる。また、「医療機関」、「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」からは、移動目的は通院あるいは買い物であることが推測される。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、家電などの購入も含まれる。「文化施設」および「都道府県指定文化財」は、観光客の存在を示唆している。

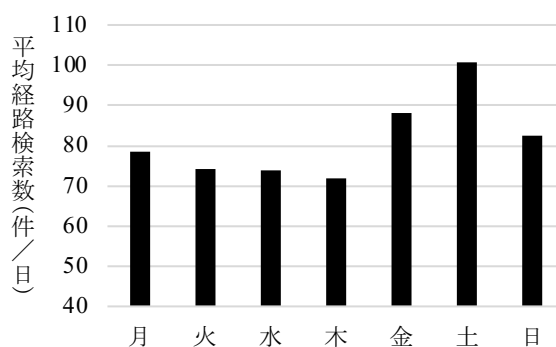


図 4.3 曜日別平均経路検索数 (3.3 節図 3.8 の再掲)

4.6 まとめ

4.6.1 4章の研究成果と次章への展開

本章では、前章の結果を踏まえて、検索の多さに影響する要因を分析し、経路検索システムが利用される状況を推察した。その結果から、仮説1の考え方の妥当性を検討する。

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から、検索者の移動目的を推測するために、Bayesian Adaptive Lasso Tobit 分位点回帰モデルを用いてモデルを構築した。モデル推定の結果から、選択された説明変数は乗車駅と降車駅で大きな違いは見られなかった。このことから、検索者は往路、復路ともに経路検索システムを利用する傾向にあることが示唆された。経路検索システムの利用者は非日常時に検索すると想定されるため、行きだけでなく帰りの移動においても利用すると考えられる。したがって、乗車駅と降車駅とで選択された説明変数に差異がないことは妥当な結果と考えられ、適切にモデル推計がなされていることを裏付ける。さらに、推定された説明変数から、曜日ダミー

については土曜日が有意になったことから、非日常的な活動が発生しやすいと考えられる休日の活動と経路検索行動の関連が示唆された。また、医療機関が有意になったことから「通院」、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「買い物」、文化施設、都道府県指定文化財からは「観光」を移動目的とした経路検索行動と関連性があると推察した。

したがって、経路検索システムの利用者の多くは普段とは異なる非日常的な移動の際に、システムを利用することに相違なく、日常的に利用されやすい交通系 IC カードとは、利用されやすい状況は異なる。つまり、仮説 1 の考え方は妥当といえる。

これを踏まえて、次章では仮説 1 を数値的に検証するとともに、**仮説 2「経路検索データと交通系 IC カードデータに包含される突発的事象に起因する需要部分を適切にモデル化できれば、経路検索数は交通系 IC カード利用数の予測精度向上に寄与する」**ことを仮定し、公共交通の需要予測モデルを提案する。

4.6.2 4 章における今後の課題

4.5 節で述べた検索者像と移動目的はあくまで推測の域に留まるため、検索者へアンケート調査を実施し、個人属性や移動目的を調査することによって、本章における推測の検証が必要と考えられる。また、検索者が指定した時間情報について、曜日だけでなく時間帯によっても経路検索数の多寡は異なり、これに影響する変数にも違いがあることが想定される。そのため、分析単位をより詳細にした分析が必要と考える。さらに、他地域においても同様の分析を行い、得られる推定結果の比較によって、推測される検索者像の一般性の議論が可能になるといえる。

駅周辺環境に関するデータについて、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」を用いているが、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。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、人口の流出入と平均的移動需要の関連も推測できる可能性がある。また、「15 歳以上 64 歳以下人口」を年代別に集計し、より詳細に検索者像について検討する必要がある。

第4章 参考文献

- Alhamzawi, R. (2013), Tobit quantile regression with the adaptive lasso penalty, *The 4th International Scientific Conference of Arab Statistic*, 450 ISSN (1681 6870), 19 pages.
- Andrews, D. F. and Mallows, C. L. (1974), Scale mixtures of normal distributions, *Journal of the Royal Statistical Society : Series B*, Vol.36, pp.99-102.
- Kozumi, H. and Kobayashi, G. (2011), Gibbs sampling methods for Bayesian quantile regression, *Journal of Statistical Computation and Simulation*, Vol.81, pp.1565-1578.
- Zou, H. (2006), The adaptive lasso and its oracle properties, *Journal of the American Statistical Association*, Vol.101, pp.1418-1429.
- 荒木孝治 (2013), 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.
- 石黒格 (2013), 社会心理学データに対する分位点回帰分析の適用: ネットワーク・サイズを例として, 社会心理学研究, Vol.29, No.1, pp.11-20.
- 金子貴誉史, 森本章倫, 亀井三郎, 小川貴裕 (2011), 大規模小売店舗の集客人口算出モデル構築に関する研究, 第44回土木計画学研究・講演集, CD-ROM, 6 pages.
- 川野秀一, 廣瀬慧, 立石正平, 小西貞則 (2010), 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.
- 川野秀一, 松井秀俊, 廣瀬慧 (2018), スパース推定法による統計モデリング, 共立出版, pp.1-27.
- 川野秀一, 松井秀俊, 廣瀬慧 (2018), スパース推定法による統計モデリング, 共立出版, pp.20-26.
- 経済産業省, 商業統計, <https://www.meti.go.jp/statistics/tyo/syougyo/index.html> (2021年10月13日確認).
- 国土交通省, 国土数値情報, <http://nlftp.mlit.go.jp/ksj/> (2021年10月13日確認).
- 大賀惇平, 森本章倫, 大門創 (2017), 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.
- 水落正明 (2009), 打ち切り・切断データの分析, 理論と方法, Vol.24, No.1, pp.129-138.
- 総務省統計局, e-Stat 政府統計の総合窓口, <https://www.e-stat.go.jp> (2021年10月13日確認).
- 長尾基哉, 中川大, 松中亮治, 大庭哲治, 望月明彦 (2010), 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.
- 松本久司, 北村貞太郎 (1991), 人口の転出入における異方性を考慮した重力モデル-土地利用モデルに関する研究 (I)-, 農業土木学会論文集, Vol.1991, No.153, pp.23-30.
- 村上優佳紗, 角田雅照, 戸田航史 (2015), トービットモデルに基づくソフトウェア欠陥数予測の試み, ソフトウェアエンジニアリングシンポジウム 2015 論文集, pp.77-82.
- 山村光司 (2002), 正しい分散分析結果を導くための変数変換法, 植物防疫, Vol.56, No. 10, pp. 22-27.
- 山本勲 (2015), 実証分析のための計量経済学—正しい手法と結果の読み方, 中央経済社, pp.134-136.

経路検索数と交通系 IC カード利用数の関係性分析

〈第 5 章の概要〉

経路検索数と交通系 IC カード利用数では表現する需要は異なる。しかし、天候悪化やイベント時の需要変化、すなわち突発的な需要の発生には、両者の共起性が存在すると考えられる。本章では、(1) 経路検索数と交通系 IC カード利用数の突発的な需要を抽出し、(2) 突発的な需要間の関係性を定量化するという 2 段階分析方法による経路検索数に基づく新たな交通需要予測モデルの開発を行う。

1 段階目では、交通系 IC カードデータと経路検索データに内在する定常的需要と突発的な需要を分解する。ここで、両データに含まれる需要は相互に影響を及ぼしながら、日々変動していると想定される。そこで、両者の相互関係を考慮するために、交通分野では適応事例が極めて少ない 2 変量状態空間モデルを用いた成分分解手法を開発する。

2 段階目では分解した経路検索数の突発的な需要と交通系 IC カード利用数の突発的な需要の関係を重み付き回帰モデルによって明らかにする。

この 2 段階の分析により、継続的・即時的に収集可能な経路検索データから突発的な検索需要を検知し、その情報を用いて突発的な交通需要を予測する。そして突発的な交通需要と定常的な交通需要を合算することで、全体の交通需要を予測するための分析システムを開発する。

5.1 はじめに

本章では、Web 上に表明される移動意向と実空間上に顕在化する交通需要の関係性を明らかにしたうえで、経路検索データを用いた交通需要予測モデルを提案し、公共交通の需要予測において、従前より不明確であった経路検索データの有用性を示す。ここで、1.3 節で述べた通り、同一個人の経路検索データと実行データ（本研究では、交通系 IC カードデータ）は収集することができないため、両者の関係性を個人レベルでは分析できない。そこで、日単位の集計レベルの分析を行い、分析対象とする交通機関における日々の交通需要を予測する。

図 5.1 に本章の分析フローを示す。本章では、まず、仮説 1「経路検索システムはと交通系 IC カードは利用されやすい状況が異なる」ことを検証し、その上で、経路検索数と交通系 IC カード利用数を結びつける有効な対処方法を提案する。

第 3 章、第 4 章より、個人レベルで考えたときに、経路検索データに反映される行動は、非日常時の買い物や通院、観光などに伴う非定期的行動と推察された。この推察と表 5.1（1.3 節の表 1.1 再掲）に基づいて、経路検索データと交通系 IC カードデータが表現可能な「行動」と「需要」を再整理する。表 5.1 に示すように、経路検索データが表現する定常的需要と交通系 IC カードデータが表現する定常的需要はその性質が異なる。これは経路検索システムが非定期的な買い物や通院、観光や出張などで日々利用されることに対して、交通系 IC カードはそのような活動時に加え、通勤・通学、定期的な買い物や通院などでも利用されることが原因と考えられる。つまり、両者は普段安定的に利用される状況が異なるため、経路検索データと交通系 IC カードデータそれぞれが表現する需要は直接的には結びつけられない。このことを日単位の経路検索数と交通系 IC カード利用数の相関分析によって検証する。

しかし、表 5.1 に示す通り、突発的需要については、経路検索データも交通系 IC カードデータも、イベント参加や悪天候回避などが共通要因となる非定期的行動を集計したものと考えられる。このことを踏まえて、経路検索数と交通系 IC カード利用数を関連づける方法論を提案する。

経路検索データと交通系 IC カードデータには定常的需要と突発的需要が異なる割合で混在する。そこで、交通系 IC カード利用数と経路検索数のそれぞれについて定常的需要と突発的需要を統計的に分解する。本章では、2 変量状態空間モデルを用いることによって、2 つの事象間の相関関係を考慮した時系列データの変動成分分解を試みる。そして、行動論的に関連があると考えられる突発的需要の関係性を明らかにする。ここで、交通系 IC カード利用数予測モデルを構築するにあたり、突発的需要の変動が特に大きい特異日における予測精度を向上させるために、重み付き回帰モデルを用いて両者の関

係性を分析する。

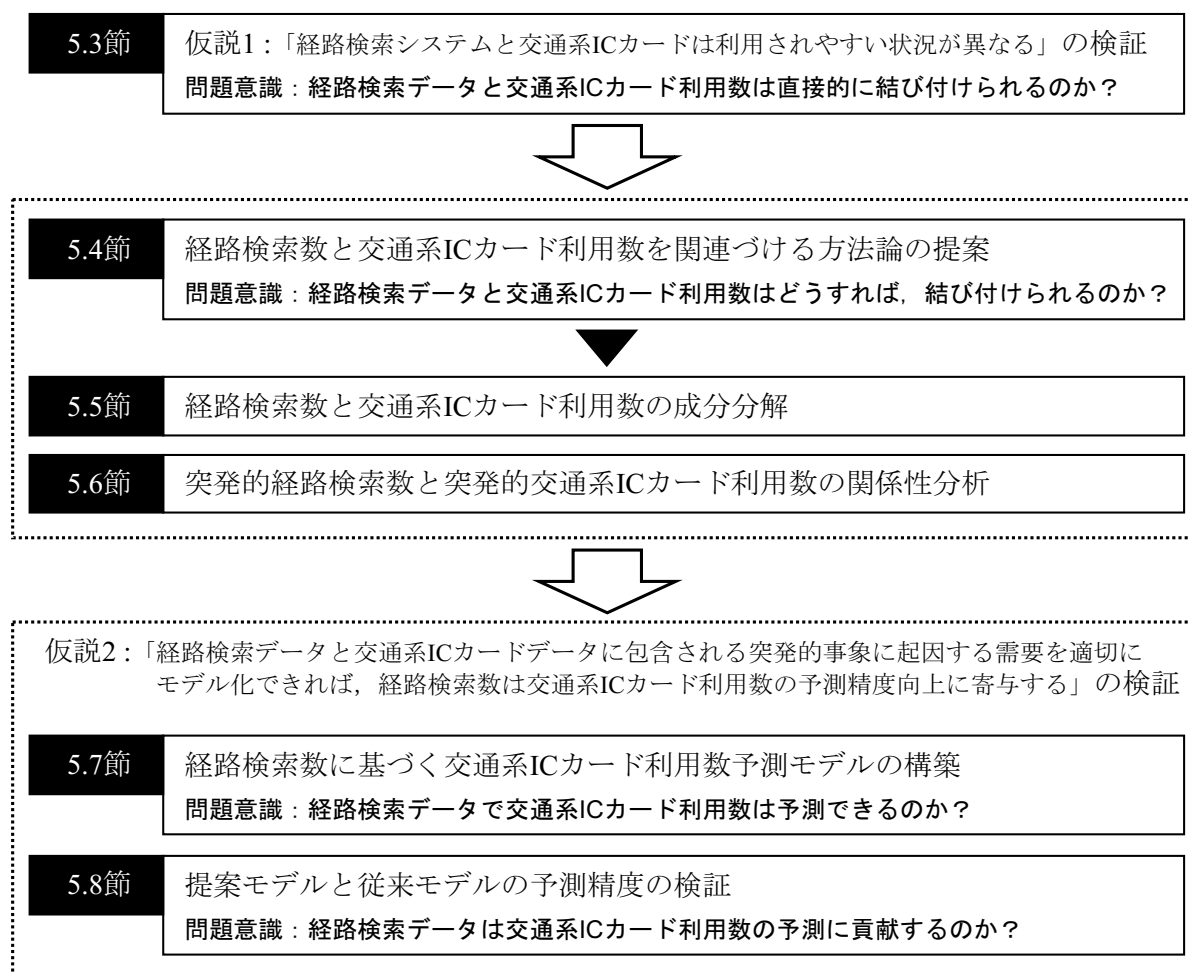


図 5.1 第 5 章の分析フロー

表 5.1 経路検索データと交通系 IC カードデータが表現可能な「行動」と「需要」（再掲）

経路検索データ		交通系 IC カードデータ	
個人レベル	集計レベル	個人レベル	集計レベル
<ul style="list-style-type: none"> ・ 非定期的行動 (例) 非定期的な買い物や通院, 観光や出張 	<ul style="list-style-type: none"> ・ 定常的需要 	<ul style="list-style-type: none"> ・ 定期的行動 (例) 通勤・通学, 定期的な買い物や通院 ・ 非定期的行動 (例) 非定期的な買い物や通院, 観光や出張 	<ul style="list-style-type: none"> ・ 定常的需要
<ul style="list-style-type: none"> ・ 非定期的行動 (例) イベント参加や悪天候回避 	<ul style="list-style-type: none"> ・ 突発的需要 	<ul style="list-style-type: none"> ・ 非定期的行動 (例) イベント参加や悪天候回避 	<ul style="list-style-type: none"> ・ 突発的需要

次に、仮説 2「経路検索データと交通系 IC カードデータに含まれる突発的事象に起因する需要部分を適切にモデル化できれば、経路検索数は交通系 IC カード利用数の予測精度向上に寄与する」ことを検証する。そのために、Web 上と実空間上の異なる空間上に現れる需要の関係性を考慮した交通需要予測モデルを提案し、提案モデルと従前モデルの予測精度を比較する。これによって、「経路検索データは公共交通の需要予測に役立つのか」という問いに答える。

5.2 使用データ

5.2.1 経路検索システム「NAVITIME」の検索履歴データ

本章では、経路検索システム「NAVITIME」を用いて、起終点にことでの駅を指定し、利用指定日を 2014 年 1 月 1 日から 2014 年 12 月 31 日までの 12 ヶ月間内とした経路検索データを用いる。収集されたデータ数は 234,615 件であり、このうち、

- (1) 検索者が経路検索を行った日（NAVITIME にアクセスした日）が利用指定日以前であること
- (2) ことでんが運行する便に該当する検索であること
- (3) 同一起終点を検索していないこと

を満たす 195,448 件を有効データとして用いる。

図 5.2 に 2014 年における日別経路検索数の推移を示す。1 日あたりの平均経路検索数は約 535 件であり、最も経路検索数が多かったのは 8 月 13 日の 1,503 件、最も経路検索数が少なかったのは 10 月 13 日の 188 件であった。

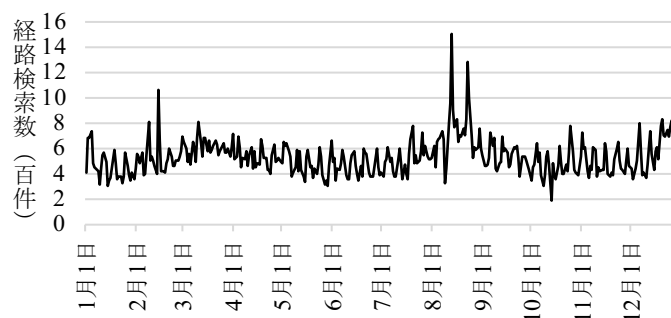


図 5.2 日別経路検索数の推移

図 5.3 には曜日別平均経路検索数を示す。曜日別平均経路検索数は月曜日から木曜日にかけて減少し、木曜日に最も少ない。そして、金曜日と土曜日に増加し、土曜日に最も多いが、日曜日に減少する。また、週末の金曜日から日曜日の平均経路検索数は他の曜日よりも多い。さらに、図 5.4 の経路検索数のコレログラムからは、7 日の周期で強い自己相関があるとわかり、経路検索数には曜日による周期性の存在が示唆された。

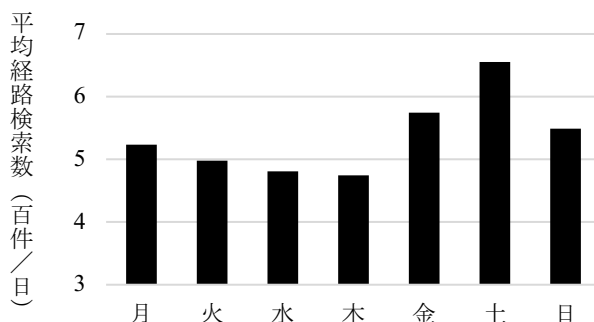


図 5.3 曜日別平均経路検索数

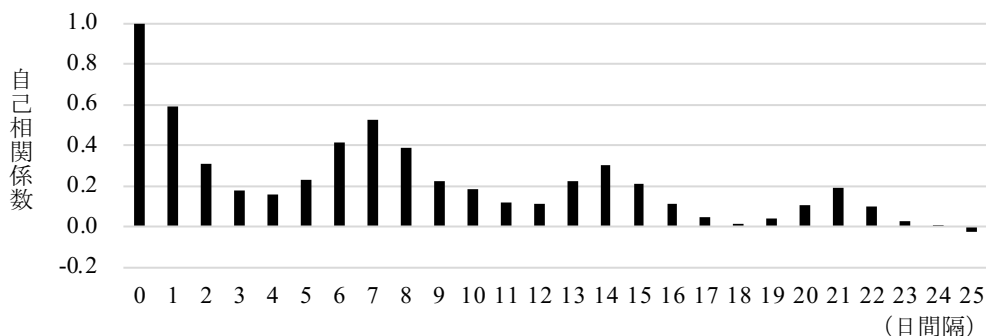


図 5.4 経路検索数のコレログラム

5.2.2 交通系 IC カード「IruCa」の乗降履歴データ

本章では、ことであんに導入されている交通系 IC カード「IruCa」の乗降履歴データを用いる。データの収集期間は経路検索データと同様に 2014 年 1 月 1 日からの 12 ヶ月間であり、この期間に収集されたデータ数は 7,285,256 件である。このうち、

- (1) ことであんの運行時間帯（5 時台から 24 時台まで）に利用されたこと
- (2) 乗車から降車までの時間差が 60 秒以上であったこと

を満たす 7,265,029 件を有効データとして用いる。

図 5.5 に 2014 年における日別交通系 IC カード利用数を示す。1 日あたりの平均交通系 IC カード利用数は約 19,904 件であり、最もカード利用数が多かったのは 4 月 18 日の 28,378 件、最も交通系 IC カード利用数が少なかったのは 10 月 13 日の 1,623 件であった。交通系 IC カード利用数が最も少なかった日は経路検索数の最少日と一致する。

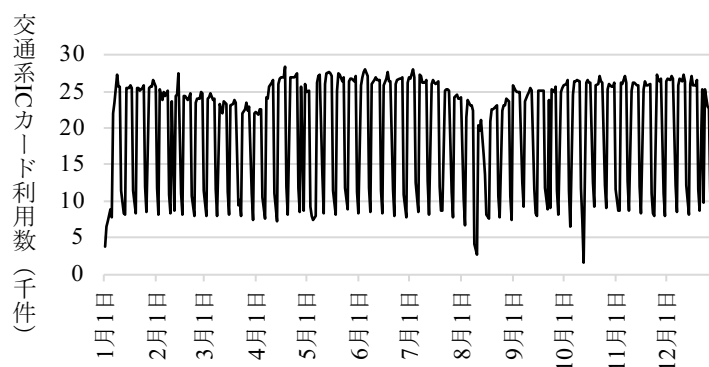


図 5.5 日別交通系 IC カード利用数の推移

図 5.6 には曜日別平均交通系 IC カード利用数を示す。曜日別交通系 IC 平均カード利用数は月曜日から木曜日にかけて増加し、金曜日から日曜日にかけて減少する。また、平日の平均交通系 IC カード利用数 23,988 件に対して休日の平均交通系 IC カード利用数は 9,649 件であることから、平日と休日では交通系 IC カード利用数に大きな差があることがわかった。さらに、図 5.7 の交通系 IC カード利用数のコレログラムからは、7 日の周期で強い自己相関があるとわかり、経路検索数と同様に交通系 IC カード利用数にも曜日による周期性があると考えられる。

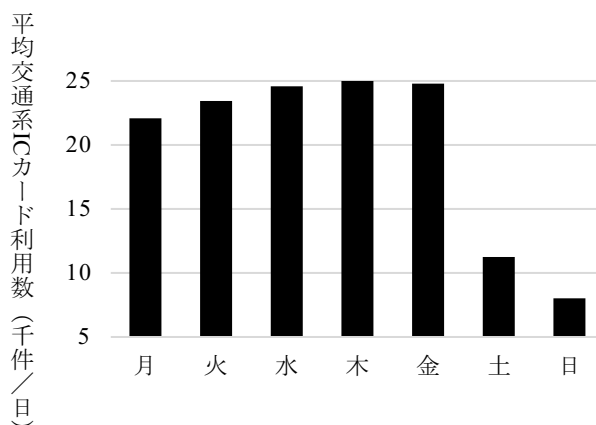


図 5.6 曜日別平均カード利用数

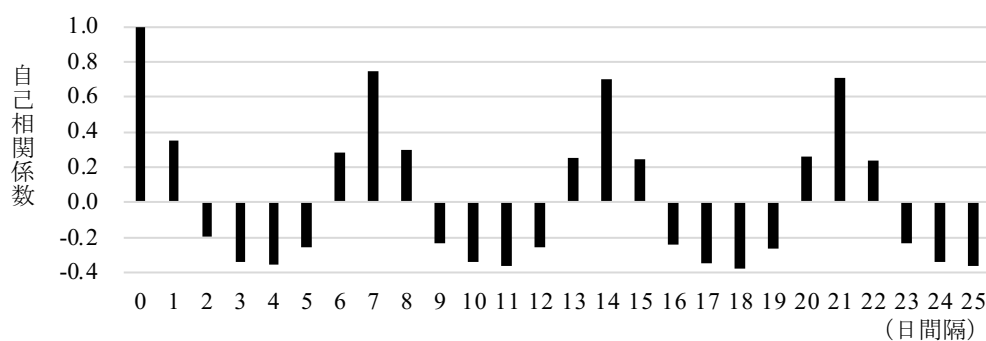


図 5.7 交通系 IC カード利用数のコレログラム

5.3 経路検索数と交通系 IC カード利用数の相関分析

Web 上の潜在需要と実空間上の顕在需要の相関分析を行う。図 5.8 に平日と休日（土日祝日）における日別経路検索数と日別交通系 IC カード利用数の散布図を示す。

両者の全体の関係性を明らかにするために、日別経路検索数と日別交通系 IC カード利用数の相関係数を算出したところ、 -0.35 と負の相関があることがわかった。次に、両データを平日と休日に分割し、相関係数を算出すると、平日では -0.52 と負の相関に、休日では 0.50 と正の相関があることが明らかとなった。

平日においては、通勤・通学、定期的の買い物や通院に関連する定期的行動が多く行われ、ことでん利用者は普段行う移動のために、経路検索システムは利用しないといえる。そのため、交通系 IC カード利用数は多くなるが経路検索数は少なくなり、両者の相関は負となると推察できる。一方、休日においては、定期的行動はあまり行われず、非定期的の買い物や通院、観光や出張といった非定期的行動が多くなる。普段とは異なる移動の際に経路検索システムが利用されることから、休日は経路検索数が多くなり、非定期的行動による交通系 IC カード利用数も多くなることで、経路検索数と交通系 IC カード利用数の関係は正の相関になると考えられる。平日と休日を区別しない全体の傾向は、全日数のうち平日の日数が多いことに影響を受け、負の相関になったといえる。

つまり、データ全体に見られる関係とデータの各層（平日と休日）に見られる関係は一致しないというシンプソンのパラドックスの状況が確認され、これは、

仮説 1：「経路検索システムと交通系 IC カードは利用されやすい状況が異なる」

ことを裏付ける。このことから、経路検索数と交通系 IC カード利用数は直接的には結びつけられないと考えられる。

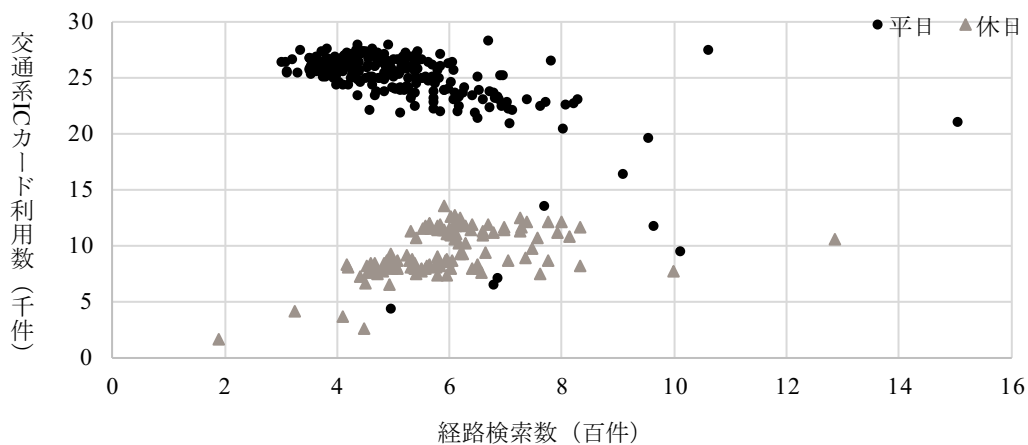


図 5.8 平日と休日における経路検索数と交通系 IC カード利用数の散布図

5.4 本研究のアプローチ

図 5.9 に経路検索データと交通系 IC カードデータの関係を示す。これまで述べてきているように、個人レベルでの交通系 IC カードの利用には定期的行動と非定期的行動によるものがあり、集計レベルでの需要の大部分は利用者の定期的行動の積み重ねである定常的需要と考えられる。さらに、個人レベルでの交通系 IC カードの利用の非定期的行動も集計レベルでは定常的需要として一定数観測される。一方、経路検索システムは個人レベルでの非定期的行動が行われるときに利用されるものの、これもまた集計レベルでは定常的需要として一定数観測され、交通系 IC カードデータと同様に、需要の大部分は定常的需要が占めると考えられる。

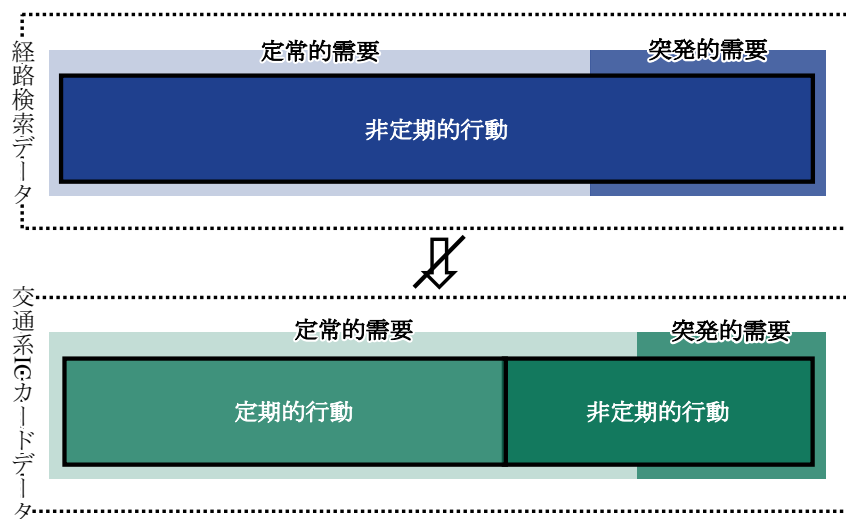


図 5.9 経路検索データと交通系 IC カードデータの関係

ただし、仮説1:「経路検索システムと交通系 IC カードは利用されやすい状況が異なる」ため、両データに内在する定常的需要は、集計した個人レベルの行動に相違がある。しかし、突発的需要であれば、共通して非定期的行動に起因することから関連づけられる可能性がある。

本節以降では、交通系 IC カード利用数と経路検索数のそれぞれについて定常的需要と突発的需要を統計的に分解し、共通性のある突発的需要に関する経路検索数と交通系 IC カード利用数の関係性を明らかにする。そして、経路検索数に基づく交通系 IC カード利用数の予測モデルを構築する。

これにより、本研究の2つ目の仮説である、

「両者に包含される突発的事象に起因する需要部分を適切にモデル化できれば、経路検索数は交通系 IC カード利用数の予測精度向上に寄与する」

の検証を試みる。

そのための手順を図 5.10 に示す。第1段階として、線形・ガウス型の2変量状態空間モデルを用いて、経路検索データと交通系 IC カードデータの両時系列データに含まれる変動成分を分解する。ここで、経路検索数と交通系 IC カード利用数には相互依存性が存在すると仮定し、2変量状態空間モデルを用いることによって、2つの事象間の相関関係を考慮した時系列データの変動成分分解を試みる点に方法論的な新規性がある。

step1 では、考慮する変動成分には時系列の水準を表すレベル成分、曜日や祝日による周期的な変動を表す回帰成分、レベル成分と回帰成分でも表せられない時系列の変動を表す誤差成分を仮定する。なお、レベル成分は1時点前からの影響、すなわち1日前のレベル成分値が次の時点のレベル成分値に影響すると考える。

step2, step3 では、step1 で選定した変動成分に日別経路検索数と日別交通系 IC カード利用数を分解し、誤差成分を突発的需要に相当する部分として抽出する。

次に、第2段階として、step4 で抽出した突発的な経路検索数と交通系 IC カード利用件数の関係を重み付き回帰分析によって明らかにし、step5 で経路検索数から交通系 IC カードの利用者数を予測するモデルを構築する。

最後に、提案する2段階の予測モデルの精度評価のために、step6 で提案モデルの現況再現性の確認、step7 で従前モデルと予測精度を比較する。

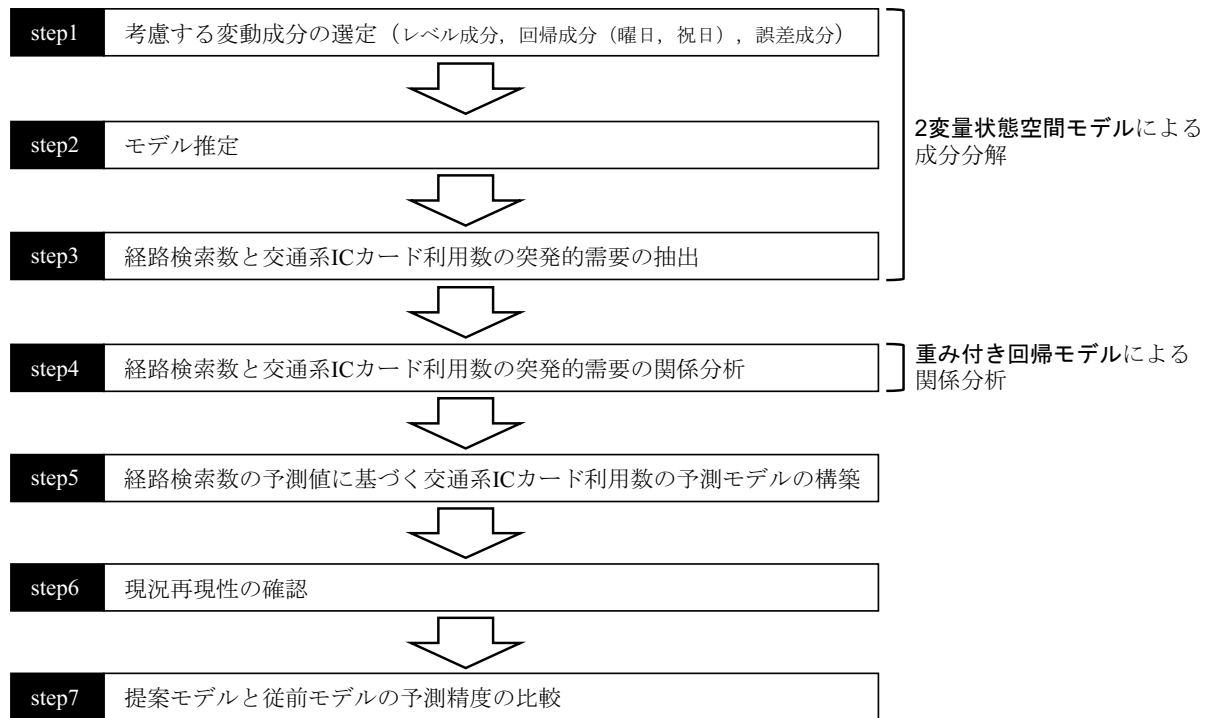


図 5.10 本研究のアプローチ

5.5 経路検索数と交通系 IC カード利用数の成分分解

5.5.1 多変量状態空間モデルの一般式

$t = 1, 2, \dots, T$ を時点, \mathbf{x}_t を m 次元の状態ベクトル, \mathbf{y}_t を n 次元の観測値ベクトルとしたとき, 線形・ガウス型の多変量状態空間モデル (コマンダーとクープマン, 2008 ; Holmes et al., 2020 ; 石田ら, 2018 ; 野村, 2016 ; Holmes et al., 2012) は一般に式(5.1)の状態方程式と式(5.2)の観測方程式で表される.

$$\mathbf{x}_t = \mathbf{B}_t \mathbf{x}_{t-1} + \mathbf{u}_t + \mathbf{C}_t \mathbf{c}_t + \mathbf{G}_t \mathbf{w}_t, \quad \mathbf{w}_t \sim \text{MVN}(\mathbf{0}, \mathbf{Q}_t) \quad (5.1)$$

$$\mathbf{y}_t = \mathbf{A}_t \mathbf{x}_t + \mathbf{a}_t + \mathbf{D}_t \mathbf{d}_t + \mathbf{v}_t, \quad \mathbf{v}_t \sim \text{MVN}(\mathbf{0}, \mathbf{R}_t) \quad (5.2)$$

ただし,

$\mathbf{B}_t, \mathbf{G}_t, \mathbf{A}_t$: パラメータ行列

$\mathbf{u}_t, \mathbf{a}_t$: 平均的な傾向を表すパラメータベクトル

$\mathbf{c}_t, \mathbf{d}_t$: 外生変数

$\mathbf{C}_t, \mathbf{D}_t$: 外生変数にかかるパラメータ行列

$\mathbf{w}_t, \mathbf{v}_t$: 過程誤差ベクトルと観測誤差ベクトル

とする.

パラメータ行列 $\mathbf{B}_t, \mathbf{G}_t, \mathbf{A}_t$ はそれぞれ式(5.3)で表される.

$$\mathbf{B}_t = \begin{bmatrix} B_{11} & \cdots & B_{1m} \\ \vdots & \ddots & \vdots \\ B_{m1} & \cdots & B_{mm} \end{bmatrix}_t, \quad \mathbf{G}_t = \begin{bmatrix} G_{11} & \cdots & G_{1g} \\ \vdots & \ddots & \vdots \\ G_{m1} & \cdots & G_{mg} \end{bmatrix}_t, \quad \mathbf{A}_t = \begin{bmatrix} A_{11} & \cdots & A_{1m} \\ \vdots & \ddots & \vdots \\ A_{n1} & \cdots & A_{nm} \end{bmatrix}_t \quad (5.3)$$

\mathbf{B}_t は $m \times m$ の遷移行列であり, m 次元の状態ベクトルを \mathbf{x}_{t-1} から \mathbf{x}_t に更新するパラメータである. \mathbf{G}_t は $m \times g$ の選択行列であり, 攪乱項である過程誤差ベクトル \mathbf{w}_t の各成分を状態ベクトル \mathbf{x}_t のどの成分に加えるかを定めるパラメータである. 多くの標準的な場合において, $g = m$ かつ \mathbf{G}_t は単位行列とされる.

\mathbf{A}_t は $n \times m$ の観測行列であり, m 次元の状態ベクトル \mathbf{x}_t を n 次元の観測値ベクトル \mathbf{y}_t に射影するパラメータである.

\mathbf{u}_t と \mathbf{a}_t は, それぞれ $m \times 1$ の $n \times 1$ のパラメータベクトルであり, 式(5.4)のように表される.

$$\mathbf{u}_t = \begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix}_t, \quad \mathbf{a}_t = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}_t \quad (5.4)$$

\mathbf{c}_t と \mathbf{d}_t は, それぞれ p 種類と q 種類の外生変数であり, 時点 $t = 1, 2, \dots, T$ のとき \mathbf{c} と \mathbf{d} は $p \times T$ と $q \times T$ の行列で式(5.5)のように表される.

$$\mathbf{c} = \begin{bmatrix} c_{11} & \cdots & c_{1T} \\ \vdots & \ddots & \vdots \\ c_{p1} & \cdots & c_{pT} \end{bmatrix}, \quad \mathbf{d} = \begin{bmatrix} d_{11} & \cdots & d_{1T} \\ \vdots & \ddots & \vdots \\ d_{q1} & \cdots & d_{qT} \end{bmatrix} \quad (5.5)$$

\mathbf{C}_t と \mathbf{D}_t は, 外生変数にかかるパラメータ行列であり, それぞれ $m \times p$ と $n \times q$ の行列で式(5.6)のように表される.

$$\mathbf{C}_t = \begin{bmatrix} C_{11} & \cdots & C_{1p} \\ \vdots & \ddots & \vdots \\ C_{m1} & \cdots & C_{mp} \end{bmatrix}_t, \quad \mathbf{D}_t = \begin{bmatrix} D_{11} & \cdots & D_{1q} \\ \vdots & \ddots & \vdots \\ D_{nq} & \cdots & D_{nq} \end{bmatrix}_t \quad (5.6)$$

$g \times 1$ の過程誤差ベクトル \mathbf{w}_t は、平均が $\mathbf{0}$ で $g \times g$ の分散共分散行列 \mathbf{Q}_t の多変量正規分布に従う。 $n \times 1$ の観測誤差ベクトル \mathbf{v}_t は、平均が $\mathbf{0}$ で $n \times n$ の分散共分散行列 \mathbf{R}_t の多変量正規分布に従う。 \mathbf{Q}_t と \mathbf{R}_t はそれぞれ未知の分散共分散構造を持ち、式(5.7)のように表される。

$$\mathbf{Q}_t = \begin{bmatrix} Q_{11} & \cdots & Q_{1g} \\ \vdots & \ddots & \vdots \\ Q_{g1} & \cdots & Q_{gg} \end{bmatrix}_t, \quad \mathbf{R}_t = \begin{bmatrix} R_{11} & \cdots & R_{1n} \\ \vdots & \ddots & \vdots \\ R_{n1} & \cdots & R_{nn} \end{bmatrix}_t \quad (5.7)$$

\mathbf{Q}_t と \mathbf{R}_t を式(5.7)のように定めることによって、成分間に相関を持たせることができる。同時に得られる観測値が複数あり系列間に相関がある多変量時系列では、成分間の関係をモデルに仮定することで、それらを独立に扱うよりもモデルの精度を向上させることができる。また、状態ベクトル \mathbf{x}_t の次元と観測値ベクトル \mathbf{y}_t の次元が同じ場合 ($m = n$) に、 \mathbf{Q}_t と \mathbf{R}_t を式(5.8)のような対角行列に制約し、係数行列 \mathbf{B}_t , \mathbf{G}_t , \mathbf{A}_t も同様に対角行列と定めると、系列間は独立であることを仮定するモデルとなる。したがって、 $m = n$ 個の単変量状態空間モデルを1つのモデルとして扱うことになる。

$$\mathbf{Q}_t = \begin{bmatrix} Q_{11} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & Q_{gg} \end{bmatrix}_t, \quad \mathbf{R}_t = \begin{bmatrix} R_{11} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & R_{nn} \end{bmatrix}_t \quad (5.8)$$

5.5.2 計算アルゴリズム

状態空間モデルでは、カルマンフィルタアルゴリズム (北川, 2005) の適用によって各時点の状態 \mathbf{x}_t を推定する。 $t = 1, 2, \dots, T$ を時点、 \mathbf{x}_t を状態ベクトル、 \mathbf{y}_t を観測値ベクトルとしたとき、一般的な線形・ガウス型の多変量状態空間モデルを式(5.9)と式(5.10)のように表す。

$$\mathbf{x}_t = \mathbf{F}_t \mathbf{x}_{t-1} + \mathbf{G}_t \mathbf{w}_t, \quad \mathbf{w}_t \sim \text{MVN}(\mathbf{0}, \mathbf{Q}_t) \quad (5.9)$$

$$\mathbf{y}_t = \mathbf{H}_t \mathbf{x}_t + \mathbf{v}_t, \quad \mathbf{v}_t \sim \text{MVN}(\mathbf{0}, \mathbf{R}_t) \quad (5.10)$$

状態 \mathbf{x}_t を推定するために、条件付き分布を規定する平均ベクトルと分散共分散行列を求める。ここ

で、状態 \mathbf{x}_t の条件付き平均と分散共分散行列をそれぞれ式(5.11)と式(5.12)のように定義する.

$$\mathbf{x}_{t|j} := E(\mathbf{x}_t | \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_j) \quad (5.11)$$

$$\mathbf{V}_{t|j} := E[(\mathbf{x}_t - \mathbf{x}_{t|j})(\mathbf{x}_t - \mathbf{x}_{t|j})'] \quad (5.12)$$

カルマンフィルタでは、観測値が得られていない場合に将来の状態を推定する1期先予測(式(5.13), 式(5.14)), 観測値が得られた際に現在の状態を推定するフィルタリング(式(5.15), 式(5.16)), 現在までの観測値に基づいて過去の状態を推定する平滑化(式(5.18), 式(5.19))を繰り返し行い、状態 \mathbf{x}_t を逐次的に推定する. 以下にカルマンフィルタアルゴリズムを示す.

状態 \mathbf{x}_t の1期先予測 ($j = t - 1$) は式(5.13)で与えられる.

$$\mathbf{x}_{t|t-1} = \mathbf{F}_t \mathbf{x}_{t-1|t-1} \quad (5.13)$$

1期先予測の分散は式(5.14)で与えられる.

$$\mathbf{V}_{t|t-1} = \mathbf{F}_t \mathbf{V}_{t-1|t-1} \mathbf{F}'_t + \mathbf{G}_t \mathbf{Q}_t \mathbf{G}'_t \quad (5.14)$$

状態 \mathbf{x}_t のフィルタリング ($j = t$) は式(5.15)で与えられる.

$$\mathbf{x}_{t|t} = \mathbf{x}_{t|t-1} + \mathbf{K}_t (\mathbf{y}_t - \mathbf{H}_t \mathbf{x}_{t|t-1}) \quad (5.15)$$

フィルタリングの分散は式(5.16)で与えられる.

$$\mathbf{V}_{t|t} = (\mathbf{I} - \mathbf{K}_t \mathbf{H}_t) \mathbf{V}_{t|t-1} \quad (5.16)$$

\mathbf{K}_t はカルマンゲインと呼ばれ、式(5.17)のように定義される.

$$\mathbf{K}_t := \mathbf{V}_{t|t-1} \mathbf{H}'_t (\mathbf{H}_t \mathbf{V}_{t|t-1} \mathbf{H}'_t + \mathbf{R}_t)^{-1} \quad (5.17)$$

状態 \mathbf{x}_t の平滑化 ($j = T$) は式(5.18)で与えられる.

$$\mathbf{x}_{t|T} = \mathbf{x}_{t|t} + \mathbf{A}_t(\mathbf{x}_{t+1|T} - \mathbf{x}_{t+1|t}) \quad (5.18)$$

平滑化の分散は式(5.19)で与えられる.

$$\mathbf{V}_{t|T} = \mathbf{V}_{t|t} + \mathbf{A}_t(\mathbf{V}_{t+1|T} - \mathbf{V}_{t+1|t})\mathbf{A}'_t \quad (5.19)$$

ここで, \mathbf{A}_t は式(5.20)のように定義される.

$$\mathbf{A}_t := \mathbf{V}_{t|t}\mathbf{F}'_{t+1}\mathbf{V}_{t+1|t}^{-1} \quad (5.20)$$

5.5.3 モデルの設計

状態空間モデルでは時系列をいくつかの成分に分解できる. 本研究では, 分析対象の時系列データに確率的レベル成分, 回帰成分 (確定的周期成分), 誤差成分を仮定し, モデルを構築する (図 5.10 の step1). このとき, 式(5.1)と式(5.2)において \mathbf{x}_t と \mathbf{y}_t を2変量ベクトルとする. パラメータ行列 \mathbf{B}_t と \mathbf{Z}_t を 2×2 の単位行列とし, 共変量は \mathbf{d}_t のみに仮定する. \mathbf{u}_t と \mathbf{a}_t は $\mathbf{0}$ とする. 本研究で用いる2変量状態空間モデルの観測方程式は式(5.21), 状態方程式は式(5.22)で表現される.

$$\mathbf{y}_t = \mathbf{x}_t + \mathbf{D}\mathbf{d}_t + \mathbf{v}_t, \quad \mathbf{v}_t \sim \text{MVN}(\mathbf{0}, \mathbf{R}) \quad (5.21)$$

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim \text{MVN}(\mathbf{0}, \mathbf{Q}) \quad (5.21)$$

ただし,

\mathbf{y}_t : 観測値

\mathbf{x}_t : レベル成分

$\mathbf{D}\mathbf{d}_t$: 回帰係数 \mathbf{D} とダミー変数 \mathbf{d}_t による回帰成分

\mathbf{v}_t : 誤差成分

とする。

本研究で用いる線形・ガウス型状態空間モデルは、観測値 \mathbf{y}_t と状態 \mathbf{x}_t を線形な関数で表現し、観測誤差 \mathbf{v}_t と過程誤差 \mathbf{w}_t は正規分布に従うことを仮定する。ここで、状態空間モデルの適用において、目的変数は対数変換されることが多いが（石田，2018；沖本，2010；馬場 2018），本研究では分析結果の解釈のしやすさを優先し、このデータ加工は行わない。また、経路検索数と交通系 IC カード利用数の間の相関を考慮するために、式(5.9)の観測誤差の分散 \mathbf{R} と式(5.10)の過程誤差の分散 \mathbf{Q} は分散共分散行列とする。

ここで、各変動成分について以下に説明する。

(I) レベル成分 \mathbf{x}_t （コマンダーとクープマン，2008）

状態 \mathbf{x}_t はレベル成分とも呼ばれる。レベル成分 \mathbf{x}_t は時系列変動の水準を表す。本研究では、経路検索数と交通系 IC カード利用数のレベル成分をそれぞれ $x_{S,t}$ と $x_{I,t}$ と表す。それぞれレベル成分の仮定には、すべての時点で一定である確定的な場合と、値が時点ごとに変化する確率的な場合がある。式(5.7)において、確定的なレベル成分を仮定すると過程誤差ベクトル \mathbf{w}_t の値がすべての時点 $t = 1, 2, \dots, T$ で0に固定され、レベル成分 \mathbf{x}_t の値は一定となる。一方、確率的なレベル成分を仮定すると、過程誤差ベクトル \mathbf{w}_t の値が時点に変化することを認めるため、レベル成分 \mathbf{x}_t の値も時点により変化する。経路検索数と交通系 IC カード利用件数については、日単位の平均値も変化することを想定し、レベル成分は確率的に変動すると仮定する。なお、 \mathbf{x}_t の値は1時点前の \mathbf{x}_{t-1} の値の影響を受け、変動すると考える。

(II) 回帰成分 $\mathbf{D}\mathbf{d}_t$

5.2 節で確認した通り、経路検索数と交通系 IC カード利用数には曜日ごとの周期的な変動がみられる。さらに、平日の経路検索数と交通系 IC カード利用数は、その日が祝日かどうかにも依存する。そのため、回帰成分には周期的な変動を表現する曜日成分と、振替休日を含む月曜日から金曜日までの祝日の影響を表現する祝日成分を仮定する。曜日と祝日による影響をダミー変数 \mathbf{d}_t で表現するとき、ダミー変数 \mathbf{d}_t とダミー変数のパラメータ行列 \mathbf{D} は式(5.22)のように表される。

$$\mathbf{d}_t = \begin{bmatrix} d_{\text{Mon}} \\ d_{\text{Tue}} \\ d_{\text{Wed}} \\ d_{\text{Thu}} \\ d_{\text{Fri}} \\ d_{\text{Sat}} \\ d_{\text{Sun}} \\ d_{\text{PH}} \end{bmatrix}_t, \quad (5.22)$$

$$\mathbf{D} = \begin{bmatrix} D_{S,\text{Mon}} & D_{S,\text{Tue}} & D_{S,\text{Wed}} & D_{S,\text{Thu}} & D_{S,\text{Fri}} & D_{S,\text{Sat}} & D_{S,\text{Sun}} & D_{S,\text{PH}} \\ D_{I,\text{Mon}} & D_{I,\text{Tue}} & D_{I,\text{Wed}} & D_{I,\text{Thu}} & D_{I,\text{Fri}} & D_{I,\text{Sat}} & D_{I,\text{Sun}} & D_{I,\text{PH}} \end{bmatrix}$$

ここで、ダミー変数 \mathbf{d}_t の要素は、それぞれ

d_{Mon} : 月曜日,

d_{Tue} : 火曜日,

d_{Wed} : 水曜日,

d_{Thu} : 木曜日,

d_{Fri} : 金曜日,

d_{Sat} : 土曜日,

d_{Sun} : 日曜日,

d_{PH} : 祝日

とする。

また、ダミー変数のパラメータ行列 \mathbf{D} の要素は、それぞれ

D_{Mon} : 月曜日,

D_{Tue} : 火曜日,

D_{Wed} : 水曜日,

D_{Thu} : 木曜日,

D_{Fri} : 金曜日,

D_{Sat} : 土曜日,

D_{Sun} : 日曜日,

D_{PH} : 祝日

とする。

(III) 誤差成分 v_t

観測誤差 v_t は、時系列変動の水準を表すレベル成分と曜日や祝日の特徴を表す回帰成分では説明できない要因を表現することから、突発的需要を表現する成分と定義する。本研究では、経路検索数と交通系 IC カード利用件数の誤差成分をそれぞれ $v_{S,t}$ 、 $v_{I,t}$ と表すこととする。この誤差成分 v_t が突発的な経路検索数と突発的な交通系 IC カードの利用数を表すものとする。

5.6 2 変量状態空間モデルによる成分分解結果

5.6.1 経路検索数と交通系 IC カード利用数の各成分の特徴

日別経路検索数と日別交通系 IC カード利用数の変動にレベル成分、回帰成分（曜日成分と祝日成分）、誤差成分を仮定した 2 変量状態空間のモデルのパラメータ推定結果を表 5.2 に示す（図 5.10 の step2）。さらに、図 5.11 に日別経路検索数と日別交通系 IC カード利用数についてのレベル成分、回帰成分、誤差成分の推移をそれぞれ示す。そして、経路検索数と交通系 IC カード利用数のそれぞれで各成分の値の特徴を把握する。

表 5.2 2 変量状態空間モデルのパラメータ推定結果

パラメータ	推定値	パラメータ	推定値
$R_{S,S}$	8.21×10^3	$Q_{S,S}$	9.49×10^2
$R_{I,S}$	3.17×10^4	$Q_{I,S}$	-2.93×10^4
$R_{I,I}$	1.09×10^6	$Q_{I,I}$	1.46×10^6
$D_{S,Mon}$	-1.19×10^1	$D_{I,Mon}$	4.50×10^3
$D_{S,Tue}$	-3.82×10^1	$D_{I,Tue}$	5.29×10^3
$D_{S,Wed}$	-5.80×10^1	$D_{I,Wed}$	5.54×10^3
$D_{S,Thu}$	-5.69×10^1	$D_{I,Thu}$	5.23×10^3
$D_{S,Fri}$	4.28×10^1	$D_{I,Fri}$	5.32×10^3
$D_{S,Sat}$	1.22×10^2	$D_{I,Sat}$	-8.50×10^3
$D_{S,Sun}$	1.52×10^1	$D_{I,Sun}$	-1.17×10^4
$D_{S,PH}$	4.67×10^{-21}	$D_{I,PH}$	-1.58×10^4



図 5.11 日別経路検索数（左）と交通系 IC カード利用数（右）の各成分の推移

(I) 経路検索数の各成分の特徴

図 5.11 に基づき、経路検索数に対応する各成分の値を図 5.11 a) ~ d) に再掲しながら、各成分について考察する。

(1) レベル成分 $x_{S,t}$

図 5.11 a) のレベル成分 $x_{S,t}$ に着目すると、経路検索数の原系列（黒線）に存在する周期的な日ごとの変動や突発的な変動は、変動の水準を表すレベル成分 $x_{S,t}$ では取り除かれていることが確認できる。経

経路検索数の全期間におけるレベル成分の平均値は 533 件で、値が年末年始や 3 月中旬、8 月中旬にかけて増加している。これは冬休みや春休み、お盆休みといった長期休暇期間にかけて経路検索が多く行われる傾向があることを示している。

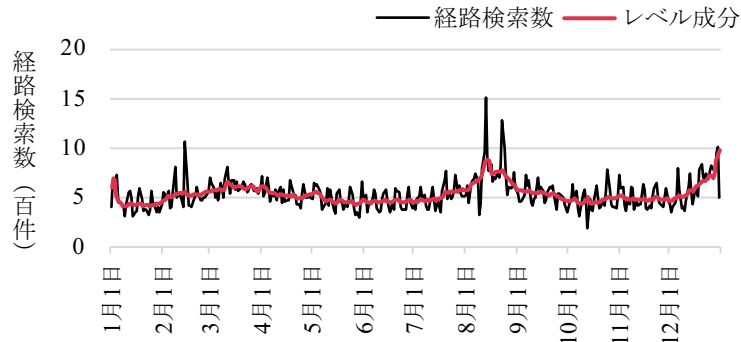


図 5.11 a) 経路検索数のレベル成分

(2) 曜日成分 $D_5 d_t$

図 5.11 b) の曜日成分からは曜日による周期性が確認できる。曜日成分の曜日ごとの成分値の推定結果を表 5.4 に示す。周期成分は確定的要素としてパラメータ推定を行なっているため、年間を通じて曜日ごとに一定の値を持つ。経路検索数の曜日成分の推定値は月曜日から木曜日にかけては負である一方、金曜日から日曜日にかけては正である。また、土曜日ほどの曜日よりも大きい値である。この傾向は図 5.3 と同様の傾向であることから、パラメータ推定が適切に行われていることを裏付け、推定値の大小関係から、水曜日と木曜日は特に経路検索数が少なく、週末は多いことがわかる。

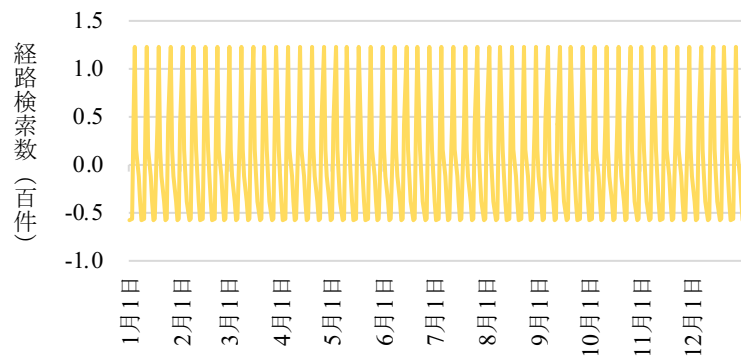


図 5.11 b) 経路検索数の曜日成分

表 5.3 経路検索数の曜日成分のパラメータ推定結果 (表 5.2 より抜粋)

月曜日	火曜日	水曜日	木曜日	金曜日	土曜日	日曜日
-1.19×10	-3.82×10	-5.80×10	-5.69×10	4.28×10	1.22×10^2	1.52×10

(3) 祝日成分 $D_S d_t$

図 5.11 c)の祝日成分からは祝日による周期性が確認できる。図 5.11 c)より、経路検索数の祝日成分の推定値は正であり、祝日はそれ以外の日よりも経路検索が多く行われていることが示された。ただし、表 5.3 に示す金曜日、土曜日、日曜日に対応する推定値よりも祝日の推定値は低いことから、月曜日から木曜日において祝日である場合でも経路検索数は通常の週末よりも少ないと推察できる。

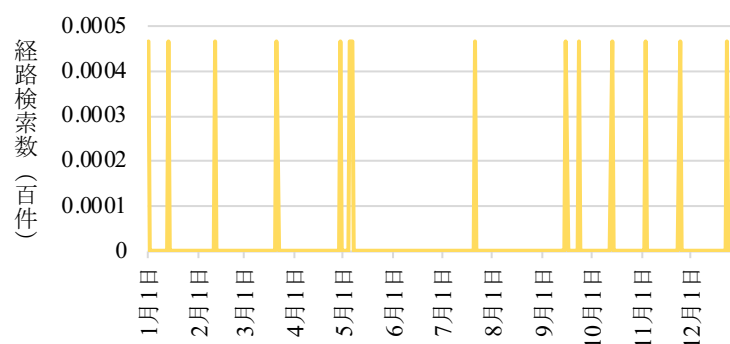


図 5.11 c) 経路検索数の祝日成分

(4) 誤差成分 $v_{S,t}$

図 5.11 d)の誤差成分からは突発的な経路検索数を確認する。本研究では時系列の変動をレベル成分、周期成分、誤差成分に分けたときに、レベル成分と周期成分の和で表現されるのが定常的需要、誤差成分で表現されるのが突発的需要と考え、経路検索数の誤差成分をそれぞれ抽出した(図 5.10 の step3)。

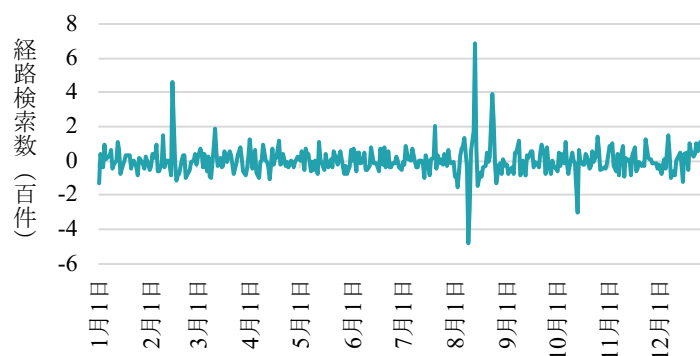


図 5.11 d) 経路検索数の誤差成分

(II) 交通系 IC カード利用数の各成分の特徴

図 5.11 に基づき、交通系 IC カード利用数に対応する各成分の値を図 5.11 e)~h)に再掲しながら、各成分を考察する。

(1) レベル成分 $x_{l,t}$

図 5.11 e)より、交通系 IC カード利用数の原系列（黒線）に着目すると、経路検索数と同様に周期的な変動や突発的な変動があるとわかる。変動の水準を表すレベル成分 $x_{l,t}$ を見ると、それらの周期的な変動や突発的な変動が取り除かれている。一方、交通系カード利用数の全期間における平均値は 19,679 件であり、値が年末年始と 5 月上旬、8 月中旬、10 月中旬において減少している。年末年始と 5 月上旬、8 月中旬の減少はそれぞれ冬休みとゴールデンウィーク、お盆休みの長期休暇期間であることから、通勤・通学客数が減少したことに起因すると考えられる。

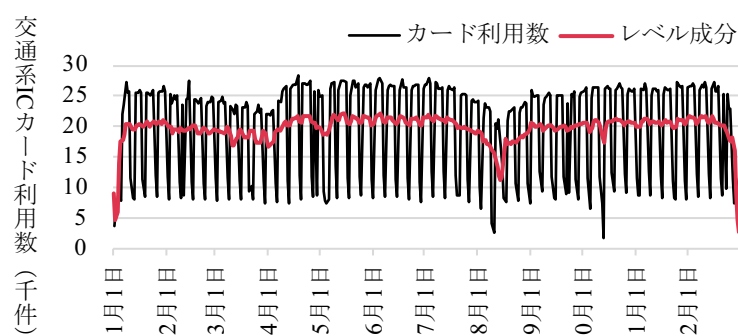


図 5.11 e) 交通系 IC カード利用数のレベル成分

(2) 曜日成分 $D_1 d_t$

図 5.11 f)の交通系 IC カード利用数の曜日成分からは曜日による周期性が確認できる。曜日成分の曜日ごとの成分値の推定結果を表 5.5 に示す。交通系 IC カード利用数の曜日成分の推定値は月曜日から金曜日では大きな変化はないが、土曜日と日曜日では負となり、日曜日に最小値をとる。この傾向は図 5.6 と同様のものであり、平日は通勤や通学などによる利用が定常的に発生するのに対して、土曜日と日曜日は会社や学校が休みにより定常的な利用が減少することを表しているといえる。

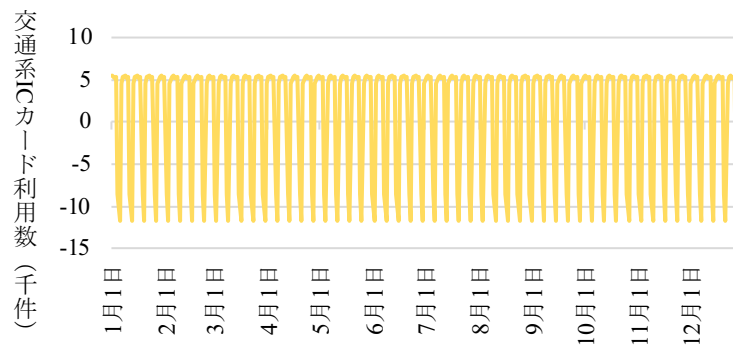


図 5.11 f) 交通系 IC カード利用数の曜日成分

表 5.4 交通系カード利用数の曜日成分のパラメータ推定結果 (表 5.2 より抜粋)

月曜日	火曜日	水曜日	木曜日	金曜日	土曜日	日曜日
4.50×10^3	5.29×10^3	5.54×10^3	5.23×10^3	5.32×10^3	-8.50×10^3	-1.17×10^4

(3) 祝日成分 $D_1 d_t$

図 5.11 g) から祝日の交通系 IC カード利用数はそれ以外の日よりも少ないことが示された。また、表 5.4 に示す月曜日から金曜日の推定値と祝日の推定値との和はいずれも -2.00×10^4 程度であり、日曜日の推定値 (-1.17×10^4) と差がないことから、平日が祝日の場合の交通系 IC カード利用数は日曜日と同程度に少ないといえる。これは祝日が日曜日と同様に会社や学校が休みであるためと考えられる。

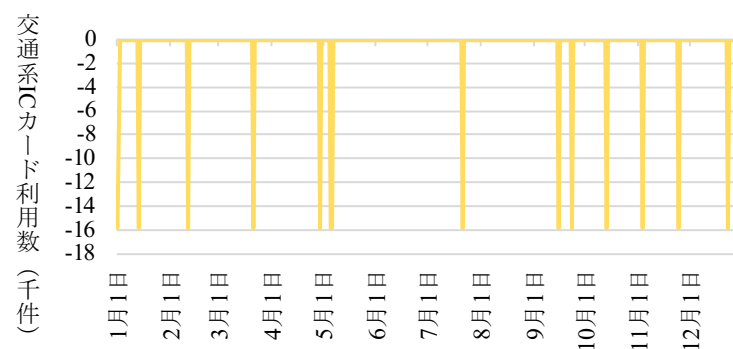


図 5.11 g) 交通系 IC カード利用数の祝日成分

(4) 誤差成分 $v_{1,t}$

図 5.11 h) より、カード利用数についてもレベル成分と周期成分からは説明できない誤差成分を抽出できた (図 5.10 の step3)。

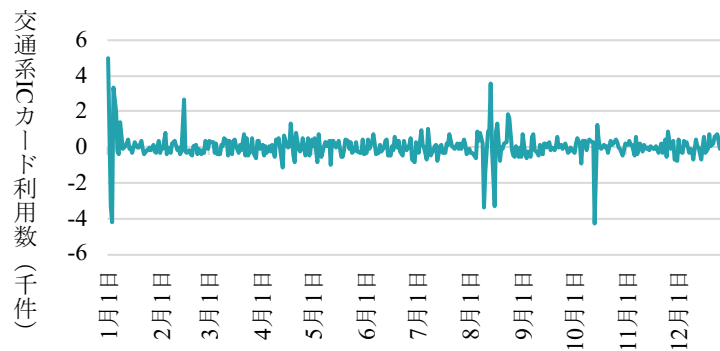


図 5.11 h) 交通系 IC カード利用数の誤差成分

5.6.2 経路検索数と交通系 IC カード利用数の誤差成分と突発的需要の同等性

ここで、誤差成分が突発的な経路検索数と突発的な交通系 IC カード利用数を表しているのかの妥当性を検討する。そのために、誤差成分の変動が特に大きい日（特異日）を抽出し、突発的需要が発生した日のイベント開催や天候変化との関連を考察する。

本研究では経路検索数と交通系 IC カード利用数の誤差成分それぞれの標準化した値が、自由度 364 (=365-1) を持つ t 分布の両側 95% 点に基づき、1.967 を上回る、あるいは -1.967 を下回る誤差成分を抽出する。その結果、経路検索数については該当日が 11 日、交通系 IC カード利用数については該当日が 15 日存在した。経路検索数と交通系 IC カード利用数の誤差成分が大きく変動した日とその要因として考えられる事象をそれぞれ表 5.5、表 5.6 に示す。なお、誤差成分の増減に関係があると考えられる情報は Web サイトで香川県を対象に該当日の出来事を検索した。

表 5.5 の経路検索数の誤差成分に基づく特異日について誤差成分の値が正である日に着目すると、2 月 14 日は大雪によるダイヤの乱れを考慮した経路検索数の増加が関連すると考えられる。7 月 20 日や 8 月 12 日、8 月 13 日は連休中の外出の増加に伴い経路検索数も増加したと考えられる。また、8 月 13 日は高松市内で開催された花火大会に向かう人々の移動需要によるものと推察される。8 月 23 日、8 月 24 日は国営讃岐まんのう公園で開催された中四国最大規模の野外音楽フェスに伴い経路検索数が増加したと考えられる。一方で、誤差成分の値が負である 8 月 9 日、8 月 10 日および 10 月 13 日は台風のため外出を見合わせ、普段よりも経路検索数が少なかったといえる。12 月 31 日は年末であったことが関係していると考えられる。なお、3 月 11 日については要因となる事象を見つけられなかった。

表 5.5 経路検索数の誤差成分に基づく特異日とその関連事象

該当日	誤差成分の値	関連性の高い事象
2月14日	4.63	大雪
3月11日	1.90	？
7月20日	2.06	三連休
8月9日	-4.83	台風
8月10日	-2.79	台風
8月12日	1.71	盆休み
8月13日	6.85	盆休み・花火大会
8月23日	3.88	音楽フェス
8月24日	2.31	音楽フェス
10月13日	-3.00	台風
12月31日	-4.18	年末

一方、表 5.6 の交通系 IC カード利用数の誤差成分に基づく特異日について誤差成分の値が正である日に着目すると、2月14日、8月13日、8月23日、8月24日は経路検索数の抽出結果と共通している。負の値を示す特異日については8月9日、10月13日、12月31日が経路検索数の抽出結果と一致した。経路検索数の誤差成分に基づく特異日にはなかった日としては、1月1日から5日が挙げられる。この期間の誤差成分の値は正負が統一していないが、年始の連休に起因する交通系 IC カードの利用と関連していると考えられる。また、12月27日、12月28日は高松市内のいくつかの施設で開催された音楽コンサートに行くためにことடன்利用者数が増えたと推察される。

表 5.6 交通系 IC カード利用数の誤差成分に基づく特異日とその関連事象

該当日	誤差成分の値	関連性の高い事象
1月1日	5.02	年始
1月2日	-3.33	年始
1月3日	-4.22	年始
1月4日	3.36	年始
1月5日	2.00	年始
2月14日	2.65	大雪
8月9日	-3.35	台風
8月13日	3.60	盆休み・花火大会
8月15日	-3.30	盆休み
8月23日	1.87	音楽フェス
8月24日	1.63	音楽フェス
10月13日	-4.28	台風
12月27日	1.56	コンサート
12月28日	3.21	コンサート
12月31日	-3.79	年末

以上を整理すると、経路検索数の誤差成分および交通系 IC カード利用数の誤差成分ともに、天候やイベント、連休に関連して変動していると推察され、普段の行動とは異なる突発的な変動として捉えることは妥当といえる。また、経路検索数の誤差成分に基づく特異日と交通系 IC カード利用数の誤差成分に基づく特異日の多くが一致することから、突発的な経路検索数は突発的な交通系 IC カード利用数を反映している可能性が高い。

5.6.3 分散共分散行列の考察

2 変量状態空間モデルのパラメータ推定結果から、経路検索数と交通系 IC カード利用数の関係を把握する。観測誤差 \mathbf{v}_t の相関と過程誤差 \mathbf{w}_t の相関を表すパラメータの推定結果を表 5.7 に示す。

観測誤差の共分散 $R_{I,S}$ の値が正であることから、経路検索数と交通系 IC カード利用数の観測誤差 \mathbf{v}_t 間には正の相関があり、過程誤差の共分散 $Q_{I,S}$ の値が負であることから、経路検索数のレベル成分と交通系 IC カード利用数のレベル成分の過程誤差 \mathbf{w}_t 間には負の相関があることが示された。

表 5.7 観測誤差と過程誤差の分散共分散 (表 5.2 より誤差分散のみ抜粋)

$R_{S,S}$	8.21×10^3	$Q_{S,S}$	9.49×10^2
$R_{I,S}$	3.17×10^4	$Q_{I,S}$	-2.93×10^4
$R_{I,I}$	1.09×10^6	$Q_{I,I}$	1.46×10^6

5.6.4 経路検索数と交通系 IC カード利用数の関係性のまとめ

誤差成分とレベル成分の相関を算出すると、誤差成分については 0.55、レベル成分については-0.76 になった。誤差成分は正の相関になることから、突発的な経路検索数が増加するときに、突発的な交通系 IC カード利用数も増加する傾向にあるといえる。一方、レベル成分は負の相関となった。レベル成分 \mathbf{x}_t が時系列変動の水準を表現していることを踏まえると、レベル成分 \mathbf{x}_t は定常的な経路検索数と定常的な交通系 IC カード利用数を表している。よって、レベル成分の相関は、「普段通りの行動が発生するような日には交通系 IC カード利用数は多いが、このとき経路検索数は少ない」、あるいは、「普段とは違う行動が発生するような日は経路検索数は多いが、このとき交通系 IC カード利用数は少ない。」

という関係性を意味しており、平常時における経路検索数と交通系 IC カード利用数の関係性が適切に表現できたといえる。

5.7 経路検索数と交通系 IC カード利用数の誤差成分の関係分析

突発的需要を表す誤差成分同士の関係、すなわち経路検索数の誤差成分と交通系 IC カード利用数の誤差成分の関係を重み付き回帰モデルを用いて分析する（図 5.10 の step4）。本研究では、目的変数を交通系 IC カード利用数の誤差成分、説明変数を経路検索数の誤差成分として、式(5.23)のような回帰モデルを設定する。

$$v_{I,t} = \alpha v_{S,t} + \beta + \varepsilon \quad (5.23)$$

ただし、

$v_{S,t}$: 経路検索数の誤差成分、

$v_{I,t}$: 交通系 IC カード利用数の誤差成分、

α : 回帰係数、

β : 定数項、

ε : 誤差項

である。

誤差成分の変動が大きい日の交通系 IC カード利用数の推定誤差を小さくするために、式(5.24)のような重み付き最小二乗問題を解くことで、パラメータを推定する。

$$\arg \min_{\alpha, \beta} \sum_{t=1}^{365} w_t \left(v_{I,t} - (\alpha v_{S,t} + \beta) \right)^2 \quad (5.24)$$

ここで、重み w_t は、本研究では式(5.25)によって与えることとする。

$$w_t = \frac{|v_{s,t}|}{\sum_{t=1}^{365} |v_{s,t}| / 365} \quad (5.25)$$

重み付き回帰分析の推定結果を表 5.8 に、各誤差成分の散布図を図 5.12 に示す。重み付き回帰分析の結果、回帰係数の推定値は 5.87 となり、これは経路検索数の誤差成分が 1 件多くなる、すなわち定常時よりも経路検索数が 1 件増減すると、そのとき交通系 IC カード利用数は 5.87 件増減することを意味する。前節では、経路検索数と交通系 IC カード利用数の誤差成分の絶対値が特に大きい日に着目し、どちらも天候やイベント、連休などに伴う行動による変動と推察したが、重み付き回帰分析の結果から絶対値が大きい日の誤差成分の関係を定量的に明らかにできた。

表 5.8 重み付き回帰分析の推定結果

変数	推定値
定数項	-80.79
経路検索数の誤差成分	5.87 *
サンプル数	365
決定係数	0.68
自由度調整済み決定係数	0.68

* : 有意水準 1%

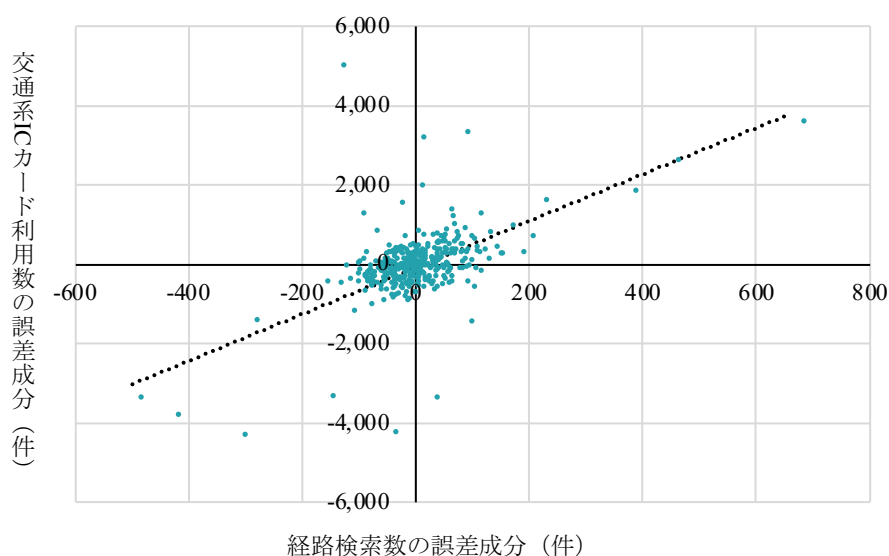


図 5.12 誤差成分の散布図

5.8 経路検索数に基づく交通系 IC カード利用数の予測モデルの構築

仮説 2「経路検索数と交通系 IC カード利用数に包含される突発的事象に起因する需要部分を適切にモデル化できれば、経路検索数は交通系 IC カード利用数の予測精度向上に寄与する」ことを仮定し、前節で明らかにした経路検索数の突発的需要を表す経路検索数の誤差成分 $v_{S,t}$ と交通系 IC カード利用数の突発的需要を表す交通系 IC カード利用数の誤差成分 $v_{I,t}$ の関係性を用いて、経路検索数の予測値に基づく交通系 IC カード利用数の予測モデルを構築する（図 5.10 の step5）。

図 5.13 に提案モデルの予測フローを示す。予測に用いる 2 変量状態空間モデルの観測方程式と状態方程式をそれぞれ式(5.26)と式(5.27)に表す。

$$\begin{bmatrix} y_S \\ y_I \end{bmatrix}_t = \begin{bmatrix} x_S \\ x_I \end{bmatrix}_t + \begin{bmatrix} \hat{D}_{S,Mon} & \cdots & \hat{D}_{S,Sun} & \hat{D}_{S,PH} \\ \hat{D}_{I,Mon} & \cdots & \hat{D}_{I,Sun} & \hat{D}_{I,PH} \end{bmatrix} \begin{bmatrix} d_{Mon} \\ \vdots \\ d_{Sun} \\ d_{PH} \end{bmatrix}_t + \begin{bmatrix} v_S \\ v_I \end{bmatrix}_t, \quad (5.26)$$

$$\begin{bmatrix} v_S \\ v_I \end{bmatrix}_t \sim \text{MVN} \left(\mathbf{0}, \begin{bmatrix} R_{S,S} & R_{I,S} \\ R_{I,S} & R_{I,I} \end{bmatrix} \right)$$

$$\begin{bmatrix} x_S \\ x_I \end{bmatrix}_t = \begin{bmatrix} x_S \\ x_I \end{bmatrix}_{t-1} + \begin{bmatrix} w_S \\ w_I \end{bmatrix}_t, \quad \begin{bmatrix} w_S \\ w_I \end{bmatrix}_t \sim \text{MVN} \left(\mathbf{0}, \begin{bmatrix} Q_{S,S} & Q_{I,S} \\ Q_{I,S} & Q_{I,I} \end{bmatrix} \right) \quad (5.27)$$

ここで、

$y_{S,t}$: 経路検索数の観測値

$y_{I,t}$: 交通系 IC カード利用数の観測値

$x_{S,t}$: 経路検索数のレベル成分（状態値）

$x_{I,t}$: 交通系 IC カード利用件数のレベル成分（状態値）

$d_{Mon,t}, \dots, d_{Sun,t}, d_{PH,t}$: 曜日・祝日ダミー変数

$\hat{D}_{S,Mon}, \dots, \hat{D}_{S,Sun}, \hat{D}_{S,PH}$: 経路検索数の曜日・祝日ダミー変数のパラメータ推定値

$\hat{D}_{I,Mon}, \dots, \hat{D}_{I,Sun}, \hat{D}_{I,PH}$: 交通系 IC カード利用数の曜日・祝日ダミー変数のパラメータ推定値

$v_{S,t}$: 経路検索数の誤差成分（観測誤差）

$v_{I,t}$: 交通系 IC カード利用数の誤差成分（観測誤差）

$w_{S,t}$: 経路検索数の過程誤差

$w_{I,t}$: 交通系 IC カード利用件数の過程誤差

である。

1段階目：2変量状態空間モデルによる成分分解
 2段階目：重み付き回帰モデルによる関係分析

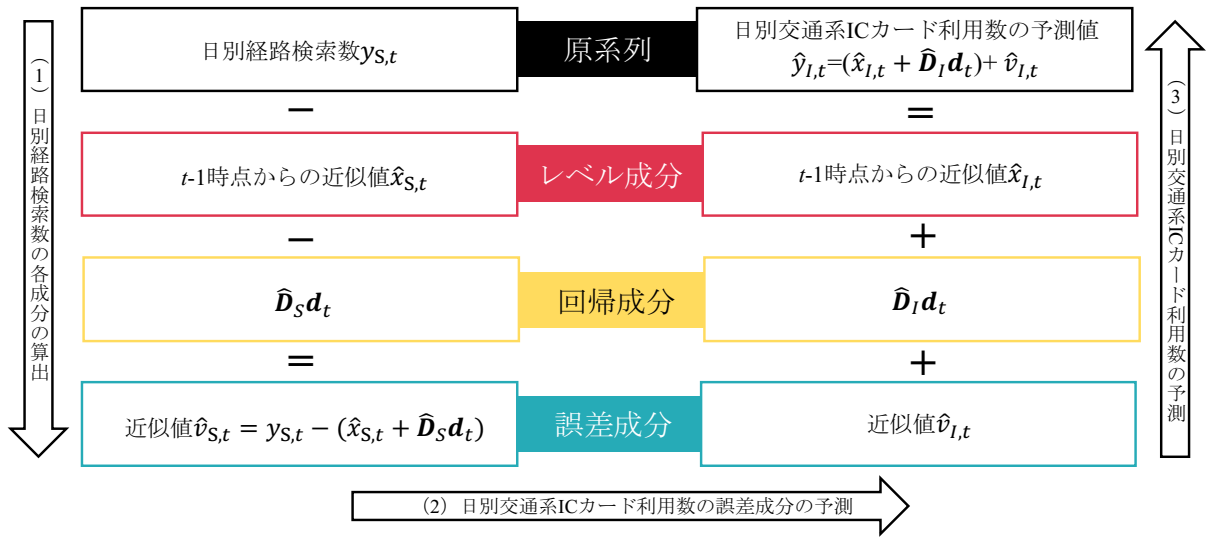


図 5.13 提案モデルの予測フロー

予測にあたっては、まず、式(5.27)に表す状態方程式の時点 t の状態値 $x_{S,t}$, $x_{I,t}$ を推定する必要がある。式(5.27)は、1 時点前の状態値 $x_{S,t-1}$, $x_{I,t-1}$ と時点 t の過程誤差 $w_{S,t}$, $w_{I,t}$ で構成される。ここで、 $x_{S,t-1}$, $x_{I,t-1}$ は状態空間モデルの適用により推定された時点 $t-1$ の状態の推定値であるため予測可能である。一方、過程誤差 $w_{S,t}$, $w_{I,t}$ は時点 t においてどのような値を取るかを事前に予測できない。そこで、予測にあたっては、 $w_{S,t}$, $w_{I,t}$ を0とし、過程誤差の相関は捨象する。すなわち、状態方程式の時点 t の状態値 $x_{S,t}$, $x_{I,t}$ は近似値として式(5.28)と式(5.29)から算出する。

$$\hat{x}_{S,t} = x_{S,t-1} + w_{S,t} = x_{S,t-1} + 0 = x_{S,t-1} \quad (5.28)$$

$$\hat{x}_{I,t} = x_{I,t-1} + w_{I,t} = x_{I,t-1} + 0 = x_{I,t-1} \quad (5.29)$$

このとき、2 変量状態空間モデルによる経路検索数の予測値 $\hat{y}_{S,t}$ は、式(5.26)の変形によって式(5.30)で与えられる。

$$\hat{y}_{S,t} = \hat{x}_{S,t} + \hat{D}_S d_t + \hat{v}_{S,t} \quad (5.30)$$

式(5.30)において、 $\hat{x}_{S,t}$ は式(5.28)から求まる。また、 $\hat{D}_S d_t$ は状態空間モデルの推定結果から得られる。そのため、式(5.30)において、観測誤差 $\hat{v}_{S,t}$ のみ未知である。

時点 t における経路検索数を事前に得ているものとする。このとき、経路検索数 $y_{s,t}$ に基づく、経路検索数の誤差成分の近似値 $\hat{v}_{s,t}$ を式(5.31)のように求める。なお、ここでも観測誤差の相関は捨象する。

$$\hat{v}_{s,t} = y_{s,t} - \hat{y}_{s,t} = y_{s,t} - (\hat{x}_{s,t} + \hat{\mathbf{D}}_s \mathbf{d}_t) \quad (5.31)$$

式(5.31)によって経路検索数の観測誤差の近似値 $\hat{v}_{s,t}$ が求めれば、式(5.32)によって、 $\hat{v}_{s,t}$ から交通系 IC カード利用数の観測誤差の近似値 $\hat{v}_{l,t}$ が算出される。

$$\hat{v}_{l,t} = 5.87 \times \hat{v}_{s,t} - 80.79 \quad (5.32)$$

$\hat{v}_{l,t}$ を式(5.26)に代入し、式(5.33)から交通系 IC カード利用数の予測値 $\hat{y}_{l,t}$ を求める。

$$\hat{y}_{l,t} = (\hat{x}_{l,t} + \hat{\mathbf{D}}_l \mathbf{d}_t) + \hat{v}_{l,t} \quad (5.33)$$

5.9 現況再現性の確認

本節と次節では、仮説 2「経路検索数と交通系 IC カード利用数に包含される突発的事象に起因する需要部分を適切にモデル化できれば、経路検索数は交通系 IC カード利用数の予測精度向上に寄与する」ことを検証する。まず、提案モデルの現況再現性を確認する（図 5.10 の step6）。経路検索数 $y_{s,t}$ に基づく経路検索数の観測誤差 $\hat{v}_{s,t}$ を算出し、予測モデルから交通系 IC カード利用数の予測値 $\hat{y}_{l,t}$ を求める。分析の結果得られた、経路検索数に基づく交通系 IC カード利用数の予測値と観測値の推移を図 5.14 に示す。現況再現性を示す RMSE（平均平方二乗誤差：Root Mean Square Error）は 1,328 件となり、平均交通系 IC カード利用数 19,904 件/日に対して約 7%の誤差で再現できた。日々の予測値と観測値の誤差に着目すると、予測序盤の 1 月 2 日～5 日と予測終盤の 12 月 28 日～31 日の当てはまりが著しく低い。これらの日においては、提案手法では予測できないような突発的需要が発生している可能性も考えられるが、状態空間モデルの性質上、初期値依存が大きいことや予測終盤が年末という特異日となったことで状態値（レベル成分）を補正できなかったために、予測期間の序盤と終盤の誤差が大きくなったと考えられる。そこで、前後 1 週間を削除して現況再現すると、RMSE は 904 件で、1 日あたりの平均交通系 IC カード利用数に対して約 5%の誤差とさらに高い精度で再現できていることが確認できた。

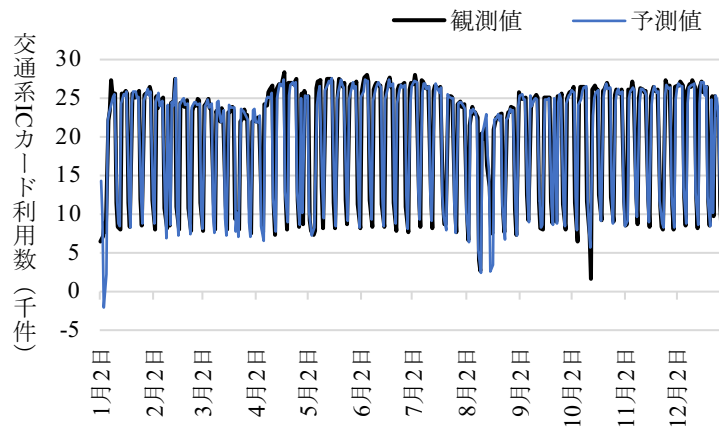


図 5.14 経路検索数の予測値に基づく交通系 IC カード利用数の予測値と観測値

本提案モデルは、経路検索数の観測誤差 $\hat{\vartheta}_{s,t}$ を算出した後、式(5.32)によって交通系 IC カード利用数の観測誤差 $\hat{\vartheta}_{l,t}$ を計算し、式(5.33)に示すように交通系 IC カード利用数のレベル成分と回帰成分に誤差成分 $\hat{\vartheta}_{l,t}$ を加えて、交通系 IC カード利用数を算出する。

$$\hat{y}_{l,t} = (\hat{x}_{l,t} + \hat{D}_l d_t) + \hat{\vartheta}_{l,t} \quad \text{再掲} \quad (5.33)$$

交通系 IC カード利用数の観測誤差 $\hat{\vartheta}_{l,t}$ の予測値の平均は 597 件/日であり、これは交通系 IC カード利用数の予測値 $\hat{y}_{l,t}$ (平均 19,892 件/日) の約 3%を占める。すなわち、経路検索数が説明できる交通系 IC カード利用数は日平均では 3%程度であることを意味し、この値は必ずしも大きいとはいえない。

ここで、5.2.2 項で示した、経路検索数の誤差成分に基づいて判定した特異日と交通系 IC カード利用数の誤差成分に基づいて判定した特異日で共通する 2 月 14 日、8 月 9 日、8 月 13 日、8 月 23 日、8 月 24 日、10 月 13 日、12 月 31 日の 7 日間に注目する。これら特異日 7 日間の交通系 IC カード利用数の観測誤差 $\hat{\vartheta}_{l,t}$ の予測値の平均は 2,866 件/日となった。これは特異日 7 日間の交通系 IC カード利用数の予測値の平均 11,023 件/日に対して約 26%を占めることになる。

本研究では、経路検索データと交通系 IC カードデータは突発的需要部分に関連するという仮説に基づき、2 変量状態空間モデルと重み付き回帰モデルによる 2 段階モデルを構築した。突発的需要は日々頻繁に発生するものではないため、平均的には経路検索数の増減が交通系 IC カード利用数の増減に与える影響はわずかである。しかし、突発的事象が発生したといえる日に着目すると、経路検索数の増減が交通系 IC カード利用数の増減に与える影響は大きく、両者の関係を適切に表現できたといえる。

5.10 他モデルと提案モデルの現況再現性の比較

本節では、以下、4つのモデルと提案手法の RMSE を比較し、提案モデルの有効性を検証する（図 5.10 の step7）。

モデル1) 交通系 IC カード利用数と経路検索数の相関関係を表現する単回帰モデル

モデル2) モデル 1 では表現できない周期性を考慮するために曜日ダミーと祝日ダミーを組み込んだ重回帰モデル

モデル3) 交通系 IC カード利用数のみを用いた単変量状態空間モデル

モデル4) 経路検索数を回帰成分に組み込んだ単変量状態空間モデル

モデル 1) は、経路検索数で交通系 IC カード利用数を直接予測するものであり、最も単純な考えに基づくモデル化といえる。

モデル 2) は、周期性が存在することは既知のため、その要因を考慮したモデルである。

モデル 3) は、経路検索数を考慮せずに単純な時系列データとして交通系 IC カード利用数を捉え、ランダムウォークする誤差成分を 0 と考えた期待値による予測モデルである。

モデル 4) は、モデル 3) を改良し、交通系 IC カード利用数を予測するために経路検索数を説明変数として用いるモデルである。

モデル 1~4) は以下の通り定式化される。

モデル 1) 単回帰モデル

$$y_{I,t} = \alpha y_{S,t} + \beta + \varepsilon \quad (5.33)$$

モデル 2) 重回帰モデル

$$y_{I,t} = \alpha y_{S,t} + \mathbf{D}_I \mathbf{d}_t + \beta + \varepsilon \quad (5.34)$$

モデル 3) 単変量状態空間モデル（交通系 IC カード利用のみ）

$$y_{I,t} = x_{I,t} + \mathbf{D}_I \mathbf{d}_t + v_t, \quad v_t \sim \text{MVN}(0, R) \quad (5.35)$$

$$x_t = x_{t-1} + w_t, \quad w_t \sim \text{MVN}(0, Q) \quad (5.36)$$

モデル 4) 単変量状態空間モデル（回帰成分に経路検索数を組み込む）

$$y_{I,t} = x_{I,t} + \mathbf{D}_I \mathbf{d}_t + \alpha y_{S,t} + v_t, \quad v_t \sim \text{MVN}(0, R) \quad (5.37)$$

$$x_t = x_{t-1} + w_t, \quad w_t \sim \text{MVN}(0, Q) \quad (5.38)$$

ただし,

$y_{S,t}$: 経路検索数の観測値

$y_{I,t}$: 交通系 IC カード利用数の観測値

$x_{S,t}$: 経路検索数のレベル成分 (状態値)

$x_{I,t}$: 交通系 IC カード利用数のレベル成分 (状態値)

$\mathbf{d}_t = d_{\text{Mon},t}, \dots, d_{\text{Sun},t}, d_{\text{PH},t}$: 曜日祝日ダミー変数

$\mathbf{D}_I = D_{I,\text{Mon}}, \dots, D_{I,\text{Sun}}, D_{I,\text{PH}}$: 交通系 IC カード利用数の曜日祝日ダミー変数のパラメータ

v_t : 交通系 IC カード利用数の誤差成分 (観測誤差)

w_t : 交通系 IC カード利用数の過程誤差

である.

各モデルにより推定される交通系 IC カード利用数の予測値と観測値を図 5.15 から図 5.19 に示す. 表 5.9 には各モデルの RMSE を示す. 表 5.10 から, 提案手法の RMSE が最小になることがわかった. これより,

仮説 2: 「経路検索数と交通系 IC カード利用数に包含される突発的事象に起因する需要部分を適切

にモデル化できれば, 経路検索数は交通系 IC カード利用数の予測精度向上に寄与する」

ことが立証された.

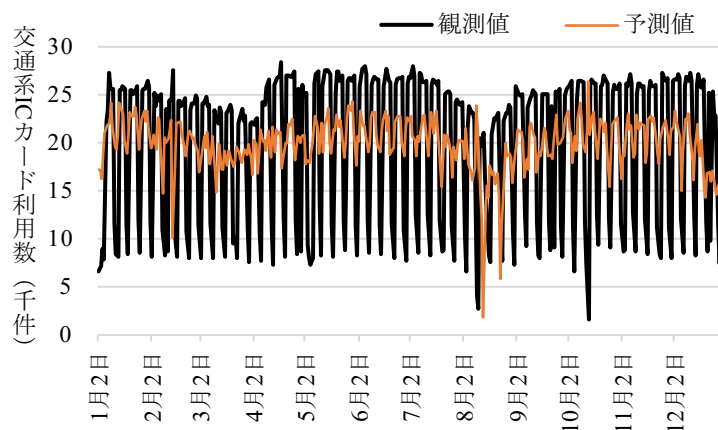


図 5.15 交通系 IC カード利用数の予測値と観測値 (モデル 1)

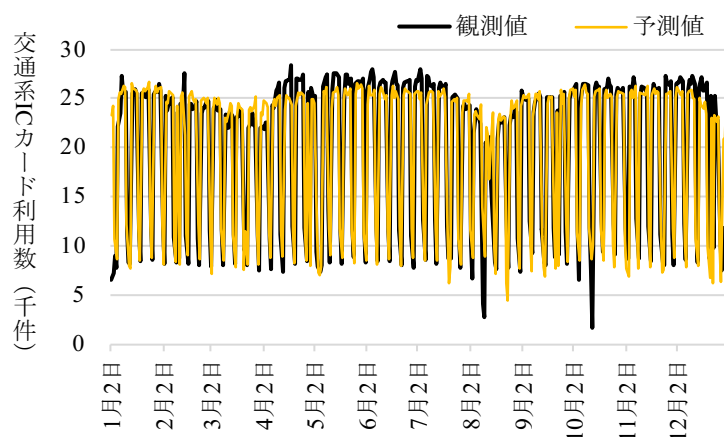


図 5.16 交通系 IC カード利用数の予測値と観測値 (モデル 2)

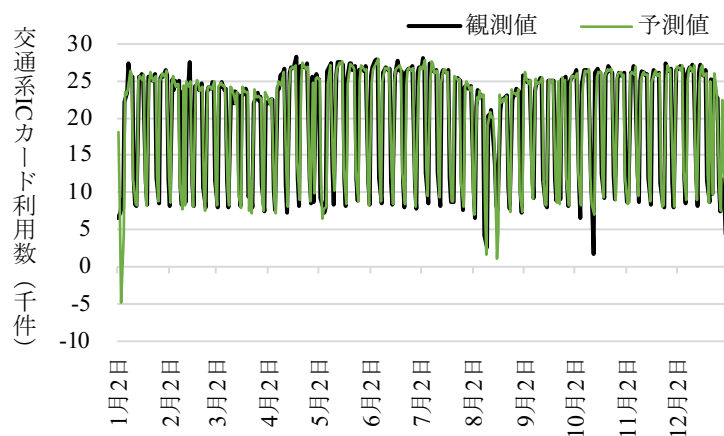


図 5.17 交通系 IC カード利用数の予測値と観測値 (モデル 3)

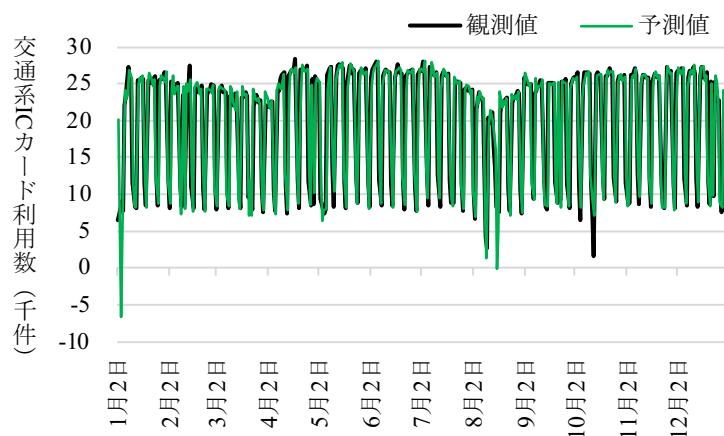


図 5.18 交通系 IC カード利用数の予測値と観測値 (モデル 4)

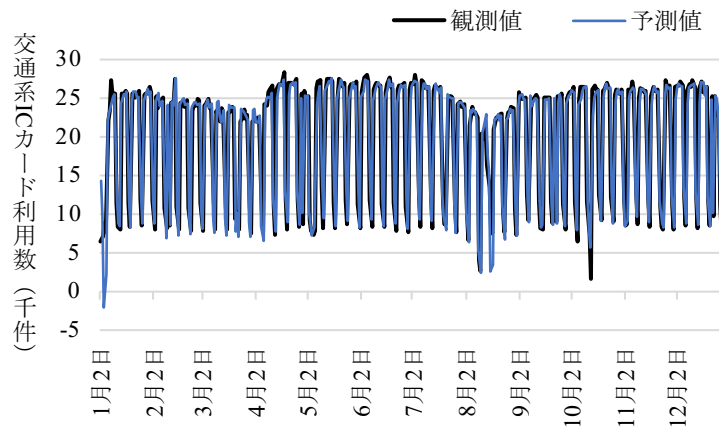


図 5.19 交通系 IC カード利用数の予測値と観測値 (提案手法)

表 5.9 各モデルの RMSE

	RMSE (件)
モデル 1: 単回帰モデル	7,219
モデル 2: 重回帰モデル	2,393
モデル 3: 単変量状態空間モデル	1,588
モデル 4: 回帰成分付き単変量状態空間モデル	1,765
提案手法	1,432

次に、経路検索数の考慮により交通系 IC カード利用数の予測精度が向上したことを検証するために、提案手法の RMSE の次に低い RMSE となったモデル 3 と予測値の変動区間の比較を行う。提案手法においては、交通系 IC カード利用数の過程誤差 $w_{l,t}$ を 0、すなわち交通系 IC カード利用数の状態値 $x_{l,t}$ に平均値を用いて利用件数 $\hat{y}_{l,t}$ の期待値を予測する。

ここで、交通系 IC カード利用数の予測値は過程誤差 $w_{l,t}$ に基づきランダムウォークしている。さらに、突発的な交通系 IC カード利用数の予測値 $\hat{y}_{l,t}$ は誤差分散による変動幅を持つ。一方、経路検索数を考慮しないモデル (モデル 3) の場合は、提案モデルと同様に交通系 IC カード利用件数の予測値は過程誤差 $w_{l,t}$ に基づきランダムウォークする。加えて、観測誤差 $v_{l,t}$ の分散によっても変動する。

過程誤差と観測誤差のそれぞれの標準偏差の大きさに着目し、2 変量状態空間モデルによる経路検索数に基づく交通系 IC カード利用数 $\hat{y}_{l,t}$ の期待値の変動 (提案モデル) と、単変量状態空間モデルによる交通系 IC カード利用数のみを用いたモデル (モデル 3) で予測したときの期待値の変動を比較する。このとき、提案モデルに比べモデル 3 の分散のほうが大きければ、期待値が取りうる値の範囲が広がり、提案モデルよりもモデル 3 の方が突発的な交通系 IC カード利用数を含む交通系 IC カード利用数の予測可能性が高いことになる。つまり、これは突発的な交通系 IC カード利用数の予測に経路検索数が寄与しないことを意味し、経路検索数を考慮したことで交通系 IC カード利用数の予測精度が向上したと

はいえない。

そこで、式(5.39), (5.40)のように交通系 IC カード利用数の過程誤差 $w_{I,t}$ に過程誤差の標準偏差 $\sqrt{Q_{I,I}}$ を, 重み付き回帰モデルにおける誤差に標準偏差 $\sqrt{\sigma_\varepsilon}$ を, 観測誤差 $v_{I,t}$ に観測誤差の標準偏差 $\sqrt{R_{I,I}}$ を仮定し, 提案手法とモデル 3 の交通系 IC カード利用数 $\hat{y}_{I,t}$ の期待値の信頼区間をそれぞれ求める。

$$\text{提案手法の}\hat{y}_{I,t}\text{の変動区間} = (x_{I,t-1} \pm \sqrt{Q_{I,I}}) + \hat{D}_I d_t \pm \sqrt{\sigma_\varepsilon} \quad (5.39)$$

$$\text{モデル 3 の}\hat{y}_{I,t}\text{の変動区間} = (x_{I,t-1} \pm \sqrt{Q_{I,I}}) + \hat{D}_I d_t \pm \sqrt{R_{I,I}} \quad (5.40)$$

図 5.20 に経路検索数に基づく交通系 IC カード利用数の予測値の推移を, 図 5.21 に交通系 IC カード利用数のみによる予測値の推移を示す。表 5.5 と表 5.6 に示した経路検索数とカード利用数の誤差成分の分散がともに大きい日を特異日とし, 特異日における提案モデルとモデル 3 の予測値の信頼区間に観測値が収まっているかを把握する。なお, 対象とする特異日は 2 月 14 日, 8 月 9 日, 8 月 13 日, 8 月 23 日, 8 月 24 日, 10 月 13 日, 12 月 31 日の 7 日である。

図 5.20 において, 経路検索数に基づく予測値の信頼区間に観測値が含まれない日は 7 日間のうち 2 日存在する。一方, 図 5.21 において交通系 IC カード利用数のみの予測値の信頼区間に観測値が含まれない日は 4 日であった。これより, 交通系 IC カード利用数できないような区間を超える変動, すなわち, 突発的に生じる交通系 IC カード利用数の変動を, 経路検索数を用いることによって予測可能であることが示され, 経路検索数を用いた提案モデルの有効性が示された。

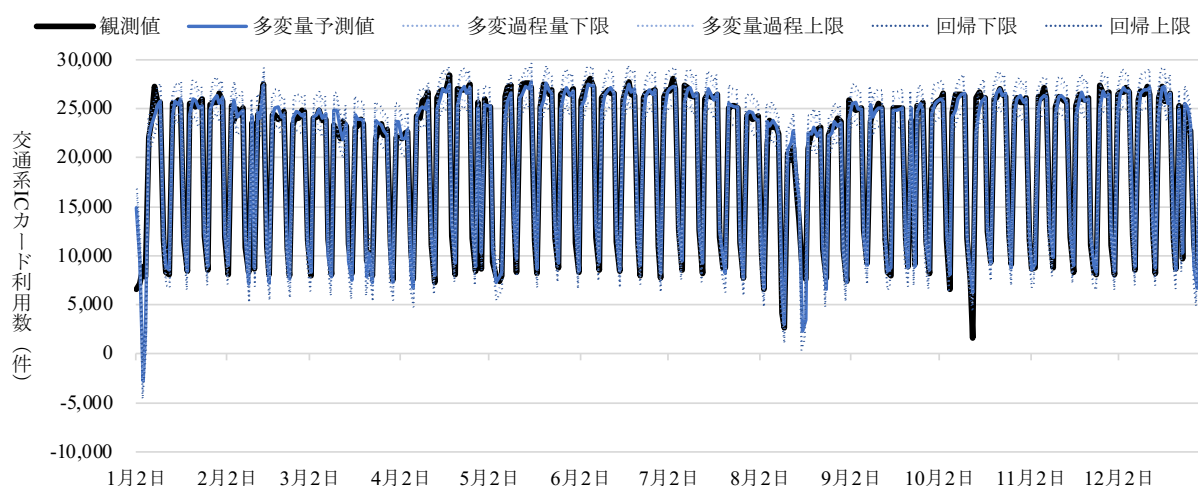


図 5.20 経路検索数に基づく交通系 IC カード利用数の予測値の推移

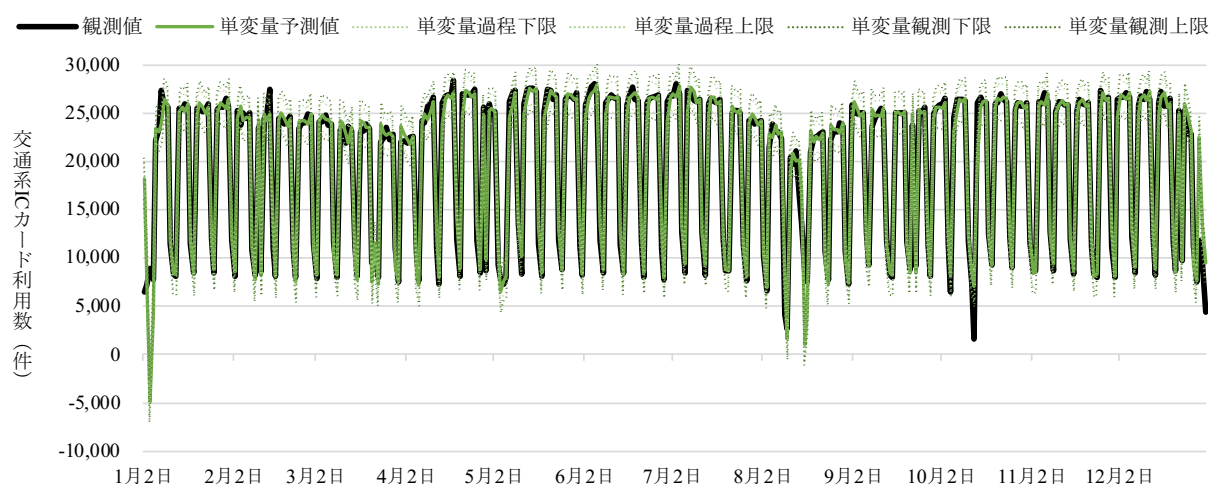


図 5.21 交通系 IC カード利用数のみによる予測値の推移

5.11 まとめ

5.11.1 5章の研究成果と次章への展開

本章では、Web 上に表明される移動意向と実空間上に顕在化する交通需要を関連づけるための方法論を提案し、経路検索データを用いた交通需要予測モデルを構築した。そして、提案モデルの有用性を検証することによって、従前より不明確であった「経路検索データは公共交通の需要予測に役立つのか」という問いに答えた。

直接的には結びつけられない経路検索データと交通系 IC カードデータを関連づけるために、両データに内在する定常的需要と突発的需要を分解し、突発的経路検索数と突発的交通系 IC カード利用数の関係を明らかにした。

第 1 段階では、2 変量状態空間モデルにより時系列データの変動を分解する成分に、時系列変動の水準を表すレベル成分、周期的な変動を表す曜日成分、祝日による利用者数の変化を表す祝日成分を仮定し、残る時系列変動を誤差成分として抽出した。これらの成分の抽出に際し、2 変量間の相関関係も同時に考慮できるのが 2 変量状態空間モデルの特徴である。観測誤差と過程誤差に関する推定結果から、観測誤差の相関が正值であることから、「突発的な経路検索数が増減するときに、突発的な交通系 IC カード利用数も増減する」という関係性が、過程誤差の相関が負値であることから、「普段通りの行

動が発生するような日には交通系 IC カード利用数が多いが、このとき経路検索数は少ない」あるいは「普段通りの行動が発生しない日には経路検索数は多いが、このときは交通系 IC カード利用数少ない」という関係性がわかった。2 変量以上の多変量状態空間モデルは主に生態学分野に適用事例があるものの (Hinrichsen, 2009; Stowe et al., 2020; Elghafghuf et al., 2018), 交通行動分析への適用例は国内外でも極めて少ないため (Stathopoulos, 2003), 交通分野への適用例を提示できたことは本研究の成果の 1 つであるといえよう。

第 2 段階では、抽出された誤差成分に重み付き回帰分析を適用し、突発的な経路検索数と突発的な交通系 IC カード利用数の関係性を明らかにした。

最後に、これら 2 段階で構成される分析手法により、経路検索数を用いて交通系 IC カードの利用者数を現況再現した結果、RMSE は 1,432 件と、1 日あたりの平均交通系 IC カード利用数に対して 7% の誤差で再現できることを示した。また、提案手法と 4 つの予測モデルの RMSE を比較したところ、提案手法の RMSE が最小となり、経路検索データを用いて突発的な交通系 IC カード利用数を予測することによって、日々の交通系 IC カード利用数の予測精度を高められることが示された。特に、突発的需要の変動が大きい日においては、交通系 IC カードデータ単体で予測するよりも、経路検索データを用いた本提案手法の方が突発的需要を含む日々の交通需要を予測できることが明らかとなった。

しかし、本章での現況再現はその日の経路検索数に基づくその日の交通系 IC カード利用数の推計に留まっており、あくまで再現は可能とするものの、事前予測はできない。そこで、次章で事前予測を可能とするモデルを開発する。

5.11.2 5 章における今後の課題

本章ではパラメータ推定に際して、カルマンフィルタを用いた。カルマンフィルタは最終的に得られる式が簡明であるために、推定方法として一般的に用いられる方法である。しかし、アルゴリズム内の初期値の設定が明確ではなく、初期値依存も大きいという欠点があり、与える初期値によっては時点初めの推定が上手くいかず、誤差が大きくなってしまふ。そのため、今後は、カルマンフィルタを拡張した手法やその他のフィルタリング手法の利点と欠点を整理し、推定方法の検討が必要と考える。

また、本提案手法は 1 日あたりの平均交通系 IC カード利用数に対して 7% と高いレベルで予測できたが、予測期間の前後 1 週間を除外すると、予測精度は 5% とさらに向上する。図 5.11 e) より、12 月 29 日から 31 日の 3 日間の交通系 IC カード利用数のレベル成分は平均的推移になっているとはいえ

ず、それまでのレベル成分の推移から極端に落ちている。このことから、3日間のレベル成分が適切に推計できていないことが全体の予測精度に大きく影響したと考えられる。12月の最後3日間は当たらないのは、データの期間の最後に年末という特異日が来てしまったことが原因と考えられ、何も無い平日や休日であれば予測精度が向上したことが期待される。あるいは、データ期間を拡張し、2015年1月1日以降のデータを含めた推計ができれば、年末という特異な動きの一部をレベル成分として吸収可能と考えられる。

第5章 参考文献

- Elghafghuf, A., Canderstichel, R., St-Hilaire, S., Stryhn H. (2018), Using state-space models to predict the abundance of juvenile and adult sea lice on Atlantic salmon, *Epidemics* 24, pp.76-87.
- Holmes, E. E., Ward, E. J., Wills, K. (2012), MARSS: Multivariate Autoregressive State-space Models for Analyzing Time-series Data, *The R Journal*, Vol.4, 19 pages.
- Hinrichsen, R., Holmes, E. E. (2009), Using multivariate state-space models to study spatial structure and dynamics, *Spatial Ecology*, 22 pages.
- Holmes, E. E., Ward, E. J., Scheuerell M. D. (2020), Analysis of multivariate time-series using the MARSS package version 3.10.12, Northwest Fisheries Science Center NOAA Seattle, WA, USA.
- Stathopoulos, A., Karlaftis, M. G. (2003), A multivariate state space approach for urban traffic flow modeling and prediction, *Transportation Research Part C*, Vol.11, pp.121-135.
- Stowe, E. S., Wenger, S.J., Freeman, M.C., Freeman, B.J. (2020), Incorporating spatial synchrony in the status assessment of a threatened species with multivariate analysis, *Biological Conservation* 248, 108612.
- 石田基広, 萩原淳一郎, 瓜生真也, 牧山幸史 (2018), 基礎からわかる時系列分析, 技術評論社, pp.37-38, pp.72-151.
- 沖本竜義 (2010), 経済・ファイナンスデータの計量時系列分析, 朝倉書店, pp.4-5.
- 北川源四郎 (2005), 時系列解析入門, 岩波書店, pp.128-131.
- J.J.F.コマンダー, S.J.クープマン (2008) 状態空間時系列分析入門, シーエーピー出版, pp.9-127.
- 野村俊一 (2016), カルマンフィルタ -Rを使った時系列予測と状態空間モデル-, 共立出版, pp.71-112.
- 馬場真哉 (2018), 時系列分析と状態空間モデルの基礎 R と Stan で学ぶ理論と実装, プレアデス出版, pp.39-40.

経路検索数に基づく交通系 IC カード利用数予測モデルの開発

〈第 6 章の概要〉

前章で提案したモデルは、(1) 2 変量状態空間モデルによって経路検索数と交通系 IC カード利用数の突発的需要を抽出し、(2) 重み付き回帰モデルによって突発的需要間の関係性を定量化するという 2 段階分析方法であった。

ここで、日々の交通系 IC カード利用数を予測するために肝要となるのが、提案モデルの 2 段階目において推計される回帰係数である。データの収集期間や分析対象地域などによってモデル推計に用いるデータが変わることで、突発的経路検索数と突発的交通系 IC カード利用数の関係性を表現する回帰係数が大きく変化する場合、分析システムの信頼性、妥当性があるとはいえない。本章では、回帰係数の時間的、空間的変動性の有無を検証し、分析対象期間および地域を変更した場合の提案モデルの適用可能性を検討する。

その上で、事前予測のために、予測日のどのくらい前に収集した経路検索数を用いれば、どの程度の予測精度をもって、突発的交通系 IC カード利用数を含む日々の交通系 IC カード利用数が予測可能かを検証することによって、本研究の最終目的である経路検索数に基づいて突発的需要を含む日々の交通需要を予測するための方法論を提案する。

6.1 はじめに

本章では、提案する交通需要予測モデルの性質の整理と、導入する経路検索データを工夫し、その有用性を検証する。

本章の分析フローを図 6.1 に示す。第 5 章で構築した提案手法は、第 1 段階として 2 変量状態空間モデルによって経路検索数と交通系 IC カード利用数の突発的変動成分を抽出し、第 2 段階として重み付き回帰モデルによって突発的経路検索数 $v_{s,t}$ の突発的交通系 IC カード利用数 $v_{l,t}$ の反映度合いを表す回帰係数 (5.6 節の式(5.23)中の回帰係数 α) を算出するものであった。すなわち、経路検索数によって交通系 IC カード利用数を予測するための鍵となるのが、回帰係数 α である。

$$v_{l,t} = \alpha v_{s,t} + \beta + \varepsilon$$

再掲
(5.23)

提案手法の有用性を示すために、この回帰係数が時間的に変動するか否かの確認し、モデルの性質上の問題、あるいは導入するデータの問題で予測できない可能性があるのかを明確にする。モデルの移転性を考えたときに、推定されるパラメータは可変性がない方が望ましい (高瀬ら, 2000)。しかし、例えば、2014 年の 1 年間に収集されたデータと 2016 年の 1 年間に収集されたデータで、突発的経路検索数の突発的交通系 IC カード利用数への影響が異なる、つまり回帰係数が異なる場合には、データの問題で 2014 年モデルは 2016 年のデータには当てはまらないことになる。そのため、日々データを更新しながらモデル推計し、回帰係数も更新していく必要がある。

さらに、空間的安定性についても議論しなければならない。第 4 章で議論した通り、経路検索数の多寡は駅周辺環境によって異なる。経路検索数が異なれば、当然、経路検索数と交通系 IC カード利用数の関係性も異なるため、回帰係数の値も変化すると予想される。回帰係数の空間的安定性を検証するためには、他地域において提案手法を適用し、結果の比較を行うことが望ましい。しかし、データ取得の問題から本研究では他地域での適用可能性を検証することができない。そこで、第 5 章ではことடன்全体での予測モデルを構築したのに対し、本章では路線ごとに分析を行うことによって、回帰係数の空間的安定性の検証を行う。具体的には、ことடன்を構成する琴平線、長尾線、志度線の 3 路線に対して回帰係数を算出し、その特性について考察する。

提案モデルは交通需要モデルとして役立つのか？

6.4節	回帰係数の時間的・空間的安定性の検証 問題意識：モデルの性質上の問題か？導入するデータの問題か？
6.5節	事前収集時間別予測精度の検証 問題意識：経路検索データは交通需要の先行指標となるのか？

図 6.1 第 6 章の分析フロー

以上、経路検索数から交通系 IC カード利用数を予測するために用いる突発的需要の回帰係数の時間的、空間的安定性の有無を検証することによって、

- (1) 分析対象期間を変更しても提案手法は適用可能か（時間的安定性）
- (2) 路線（地域）を変更しても提案手法は適用可能か（空間的安定性）

を検討する。

さらに、突発的需要を含む日々の交通需要の予測に際して、予測日終日までに収集された経路検索数に基づくモデルでは、当然、事前予測はできない。したがって、事前に収集された経路検索数に基づくモデルの開発が必要である。そのために、前章で構築した、2 変量状態空間モデルと重み付き回帰モデルを組み合わせた 2 段階モデルに導入するデータを変更し、事前に収集された経路検索数を用いた交通需要予測モデルを開発する。具体的には、第 5 章で示した予測日終日までに収集された経路検索数に基づくモデルに対して、予測日当日朝までの経路検索数のみを用いた場合、前日夕方までの経路検索数で予測した場合などの複数のケースにおいて精度比較を行う。これにより、

- (3) どのくらい前に収集された経路検索数であれば予測可能か（事前予測可能性）

を検討する。

以上、(1) 時間的安定性、(2) 空間的安定性、(3) 事前予測可能性の 3 つを検討することによって、提案手法の適用限界と有用性を明らかにする。

(1) 時間的安定性と (2) 空間的安定性の検証に際しては、2014 年の 1 年間で収集された、交通系 IC カード「IruCa」の乗降履歴データと経路検索システム「NAVITIME」を用いてことのでんを対象に検索された検索履歴データに加え、2016 年 10 月 1 日から 2019 年 9 月 30 日に収集された交通系 IC カードデータと経路検索データを用いる。ただし、2014 年の経路検索データが検索者がシステムにアクセ

スした時間、つまり検索時間を記録しているのに対し、2016年10月からの3年間の経路検索データには検索時間は記録されていない。これはNAVITIME社の個人情報の保護に関する基本方針が変更になったためである。そのため、(3) 事前予測可能性については2014年データを用いる。

6.2 使用データ

6.2.1 経路検索システム「NAVITIME」の検索履歴データ

本章では、経路検索システム「NAVITIME」を用いて、起終点にことでの駅を指定し、利用指定日を2014年1月1日からの12ヶ月間内とした経路検索データに加え、2016年10月1日から2019年9月30日を指定した経路検索データを用いる。2014年データの詳細については5.2節を参照されたい。

ここでは、2016年10月1日からの3年間に収集された経路検索データ2,381,689件についての基礎集計を行う。図6.2に3年間における日別経路検索数の推移を示す。1日あたりの平均経路検索数は約2,175件であり、2014年(535件/日)に対して4倍増えたことがわかった。これは、経路検索システムが2014年に比べて、さらに普及し、利用者が増えたことが影響したと考えられる。最も経路検索数が多かったのは2019年8月24日の5,929件、最も経路検索数が少なかったのは2019年8月15日の742件であり、同年同月でも日別経路検索数に大きく差があることが明らかとなった。

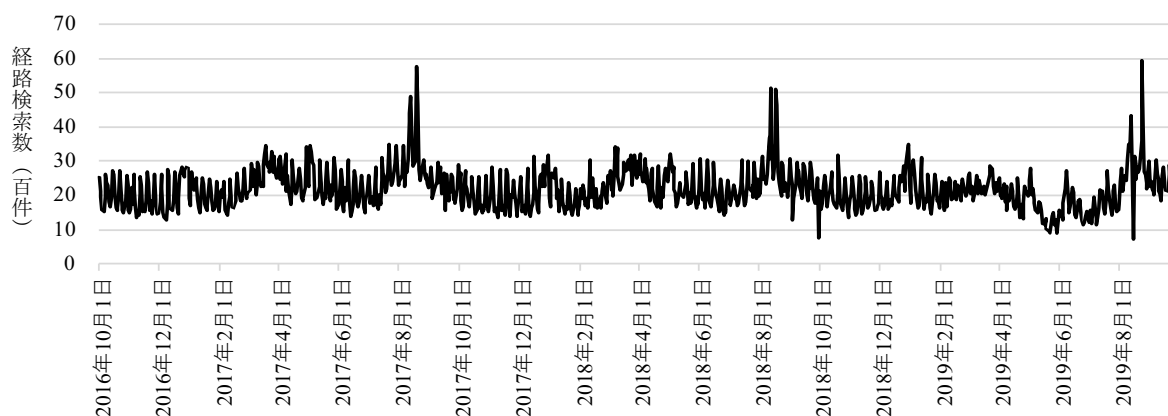


図 6.2 日別経路検索数の推移

図 6.3 には曜日別の平均経路検索数を示す。曜日別の平均経路検索数は月曜日から木曜日にかけて減少し、火曜日に最も少ない。そして、金曜日と土曜日に増加し、土曜日に最も多いが、日曜日に減少

する。また、週末の金曜日から日曜日の平均経路検索数が他の平日よりも多いことがわかる。この傾向は2014年と同様である。さらに、図6.4の経路検索数のコレログラムからは、7日の周期で強い自己相関があるとわかり、経路検索数には曜日による周期性の存在が3年間に収集された経路検索数についても示唆された。

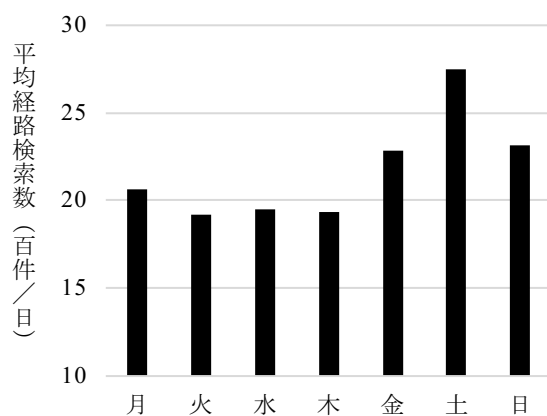


図 6.3 曜日別平均経路検索数

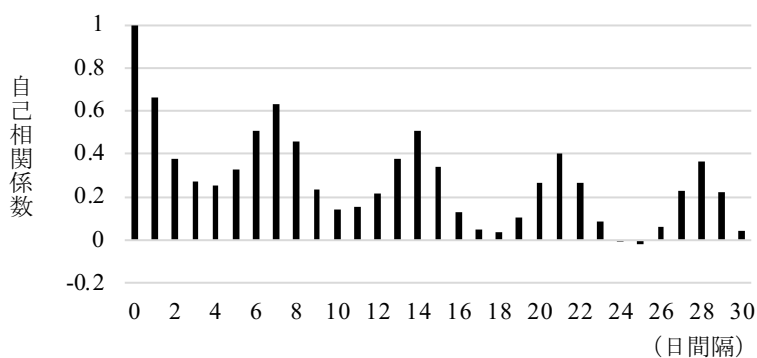


図 6.4 経路検索数のコレログラム

6.2.2 交通系 IC カード「IruCa」の乗降履歴データ

本章では、前章までと同様に実行動を表すデータとして、ことのでんに導入されている交通系 IC カード「IruCa」の乗降履歴データを用いる。データの収集期間は経路検索データと同様に2014年1月1日からの12ヶ月間に加えて、2016年10月1日から2019年9月30日までの3年間とする。2014年データの詳細については5.2節を参照されたい。

ここでは、2016年10月1日からの3年間に収集された交通系ICカードデータ 25,000,387件についての基礎集計を行う。図 6.5 に3年間における日別の交通系ICカード利用数を示す。1日あたりの平均交通系ICカード利用数は約22,831件であり、2014年の平均交通系ICカード利用数（19,904件/日）に比べて1.15倍になっている。このことから、経路検索数と比較して、IruCaの利用数は経年的に大きく増加しているとはいえない。最も交通系ICカード利用数が多かったのは2019年7月19日の33,408件、最も交通系ICカード利用数が少なかったのは2019年8月15日の0件であった。交通系ICカード利用数が最も少なかった日は経路検索数の最少日と一致する。この日に交通系ICカード利用数が0件となった理由には、2019年の台風10号の影響により、ことடன்全線が運休となったことが挙げられる。

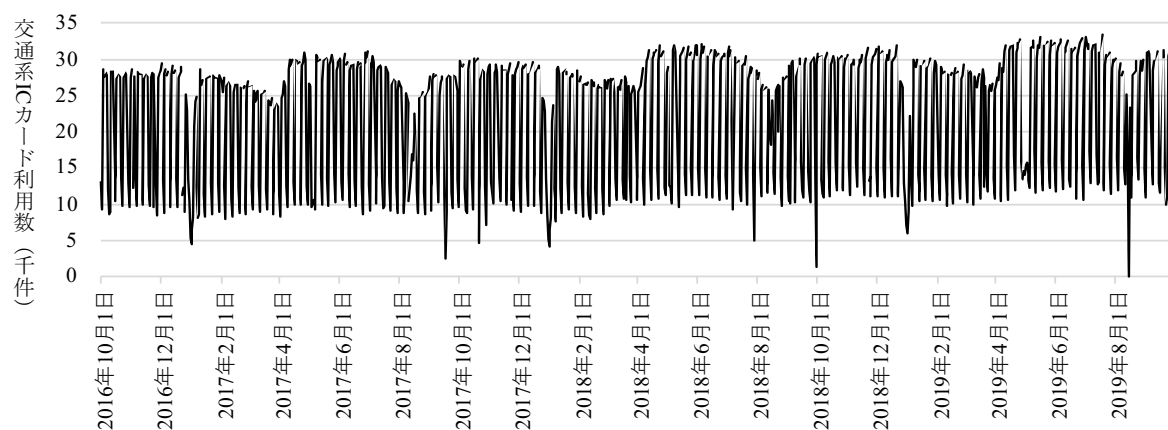


図 6.5 日別交通系ICカード利用数の推移

図 6.6 には曜日別の平均交通系ICカード利用数を示す。曜日別の平均交通系ICカード利用数は月曜日から金曜日にかけて増加傾向にあり、土曜日から日曜日にかけては減少している。また、平日の平均交通系ICカード利用数27,214件に対して休日の平均交通系ICカード利用数は11,938件であることから、平日と休日とでは交通系ICカード利用数に大きな差があることがわかった。この傾向は2014年と同様である。さらに、図 6.7 の交通系ICカード利用数のコレログラムからは、7日の周期で強い自己相関があるとわかり、経路検索数と同様に交通系ICカード利用数にも曜日による周期性があると考えられる。

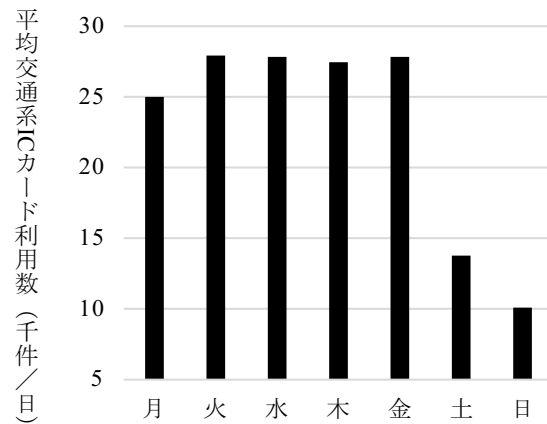


図 6.6 曜日別平均カード利用数

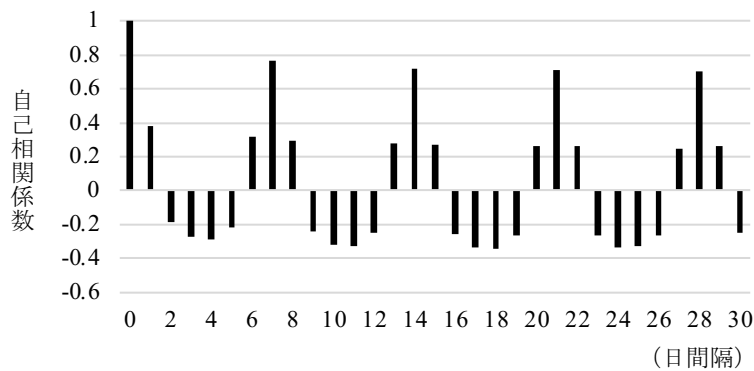


図 6.7 カード利用数のコレログラム

6.2.3 経路検索数と交通系 IC カード利用数の特徴把握

図 6.8 に平日と土日祝日における経路検索数と交通系 IC カード利用数の散布図を示す。図 6.8 より、経路検索数と交通系 IC カード利用数の相関係数を算出したところ、 -0.45 と負の相関があることがわかった。平日のみに着目すると、経路検索数と交通系 IC カード利用数の相関係数は -0.40 と算出され、両データ間には負の相関がある。一方、土日祝日のみに着目すると、経路検索数と交通系 IC カード利用数の相関係数は 0.32 と算出され、両データ間には正の相関がある。このように、平日と土日祝日のそれぞれで経路検索数と交通系 IC カード利用数の相関関係もまた、2014 年データ（5.3 節の図 5.8：相関係数は全体 -0.35 、平日 -0.52 、休日 0.50 ）と同様の傾向となることがわかった。

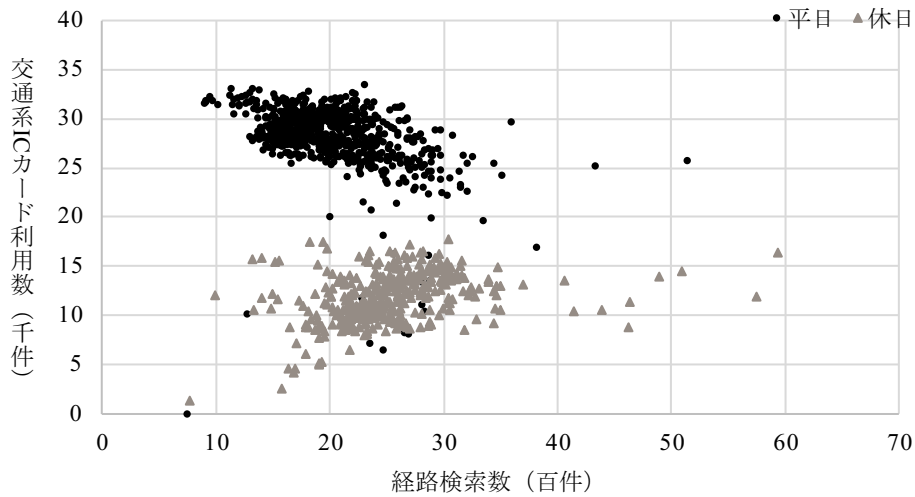


図 6.8 平日と休日における経路検索数と交通系 IC カード利用数の散布図

6.3 経路検索数の予測値に基づく交通系 IC カード利用数の予測モデルの再構築

成分分解に用いる状態空間モデルの観測方程式と状態方程式をそれぞれ式(6.1)と式(6.2)に表す.

$$\begin{bmatrix} y_S^l \\ y_I^l \end{bmatrix}_t = \begin{bmatrix} x_S^l \\ x_I^l \end{bmatrix}_t + \begin{bmatrix} \widehat{D}_{S,Mon}^l & \cdots & \widehat{D}_{S,Sun}^l & \widehat{D}_{S,PH}^l \\ \widehat{D}_{I,Mon}^l & \cdots & \widehat{D}_{I,Sun}^l & \widehat{D}_{I,PH}^l \end{bmatrix} \begin{bmatrix} d_{Mon}^l \\ \vdots \\ d_{Sun}^l \\ d_{PH}^l \end{bmatrix}_t + \begin{bmatrix} v_S^l \\ v_I^l \end{bmatrix}_t, \quad (6.1)$$

$$\begin{bmatrix} v_S^l \\ v_I^l \end{bmatrix}_t \sim \text{MVN} \left(\mathbf{0}, \begin{bmatrix} R_{S,S}^l & R_{I,S}^l \\ R_{I,S}^l & R_{I,I}^l \end{bmatrix} \right)$$

$$\begin{bmatrix} x_S^l \\ x_I^l \end{bmatrix}_t = \begin{bmatrix} x_S^l \\ x_I^l \end{bmatrix}_{t-1} + \begin{bmatrix} w_S^l \\ w_I^l \end{bmatrix}_t, \quad \begin{bmatrix} w_S^l \\ w_I^l \end{bmatrix}_t \sim \text{MVN} \left(\mathbf{0}, \begin{bmatrix} Q_{S,S}^l & Q_{I,S}^l \\ Q_{I,S}^l & Q_{I,I}^l \end{bmatrix} \right) \quad (6.2)$$

ここで,

$y_{S,t}^l$: 路線 l の経路検索数の観測値 (l = 琴平線, 長尾線, 志度線)

$y_{I,t}^l$: 路線 l の交通系 IC カード利用数の観測値

$x_{S,t}^l$: 路線 l の経路検索数のレベル成分 (状態値)

$x_{I,t}^l$: 路線 l の交通系 IC カード利用数のレベル成分 (状態値)

$d_{\text{Mon},t}^l, \dots, d_{\text{Sun},t}^l, d_{\text{PH},t}^l$: 路線 l の曜日祝日ダミー変数

$\widehat{D}_{S,\text{Mon}}^l, \dots, \widehat{D}_{S,\text{Sun}}^l, \widehat{D}_{S,\text{PH}}^l$: 路線 l の経路検索数の曜日祝日ダミー変数のパラメータ推定値

$\widehat{D}_{I,\text{Mon}}^l, \dots, \widehat{D}_{I,\text{Sun}}^l, \widehat{D}_{I,\text{PH}}^l$: 路線 l の交通系 IC カード利用数の曜日祝日ダミー変数のパラメータ推定値

$v_{S,t}^l$: 路線 l の経路検索数の観測誤差 (誤差成分)

$v_{I,t}^l$: 路線 l の交通系 IC カード利用数の観測誤差 (誤差成分)

$w_{S,t}^l$: 路線 l の経路検索数の過程誤差

$w_{I,t}^l$: 路線 l の交通系 IC カード利用数の過程誤差

である。

予測にあたっては、路線ごとにパラメータ推定を行う。そのために、まず、式(6.1)に表す状態方程式の時点 t の状態値 $x_{S,t}^l, x_{I,t}^l$ を推定する必要がある。式(6.1)は、1 時点前の状態値 $x_{S,t-1}^l, x_{I,t-1}^l$ と時点 t の過程誤差 $w_{S,t}^l, w_{I,t}^l$ で構成される。ここで、 $x_{S,t-1}^l, x_{I,t-1}^l$ は状態空間モデルの適用により推定された時点 $t-1$ の状態の推定値であるため予測可能である。一方、過程誤差 $w_{S,t}^l, w_{I,t}^l$ は時点 t においてどのような値を取るかを事前に予測できない。そこで、予測にあたっては、 $w_{S,t}^l, w_{I,t}^l$ を 0 とする。これは 5.7 節のモデルと同様である。したがって、状態方程式の時点 t の状態値 $\hat{x}_{S,t}^l, \hat{x}_{I,t}^l$ は式(6.3)と式(6.4)から計算する。

$$\hat{x}_{S,t}^l = x_{S,t-1}^l + w_{S,t}^l = x_{S,t-1}^l + 0 = x_{S,t-1}^l \quad (6.3)$$

$$\hat{x}_{I,t}^l = x_{I,t-1}^l + w_{I,t}^l = x_{I,t-1}^l + 0 = x_{I,t-1}^l \quad (6.4)$$

このとき、2 変量状態空間モデルによる路線 l の経路検索数の予測値 $\hat{y}_{S,t}^l$ は、式(6.1)の変形によって式(6.5)で与えられる。

$$\hat{y}_{S,t}^l = \hat{x}_{S,t}^l + \widehat{D}_S^l d_S^l + \hat{v}_{S,t}^l \quad (6.5)$$

式(6.5)において、 $\hat{x}_{S,t}^l$ は式(6.3)から求まる。また、 $\widehat{D}_S^l d_S^l$ は 2 変量状態空間モデルの推定結果から得られる。そのため、式(6.5)において、観測誤差 $\hat{v}_{S,t}^l$ のみ未知である。

時点 t における路線 l の経路検索数を事前に収集する (6.5 節で述べる)。このとき、路線 l の経路検索数 $y_{S,t}^l$ に基づく、路線 l の経路検索数の観測誤差の推定値 $\hat{v}_{S,t}^l$ を式(6.6)のように求める。

$$\hat{v}_{S,t}^l = y_{S,t}^l - \hat{y}_{S,t}^l = y_{S,t}^l - (\hat{x}_{S,t}^l + \hat{D}_S^l \mathbf{d}_S^l) \quad (6.6)$$

式(6.6)によって経路検索数の観測誤差の推定値 $\hat{v}_{S,t}^l$ が求めれば、式(6.7)によって、 $\hat{v}_{S,t}^l$ から路線 l の交通系 IC カード利用数の観測誤差の予測値 $\hat{v}_{I,t}^l$ が算出される。

$$\hat{v}_{I,t}^l = \alpha^l \times \hat{v}_{S,t}^l + \beta^l \quad (6.7)$$

ここで、

α^l : 回帰係数

β^l : 定数項

である。

そして、 $\hat{v}_{I,t}^l$ を式(6.1)に代入し、式(6.8)から路線 l の交通系 IC カード利用数の予測値 $\hat{y}_{I,t}^l$ を求める。

$$\hat{y}_{I,t}^l = (\hat{x}_{I,t}^l + \hat{D}_I^l \mathbf{d}_I^l) + \hat{v}_{I,t}^l \quad (6.8)$$

6.4 回帰係数の時間的・空間的安定性の検証

6.4.1 回帰係数の時間的安定性の検証

(I) 回帰係数の検証方法

突発的経路検索数と突発的交通系 IC カード利用数の関係、つまり両者を関連づけるための回帰係数((6.7) 式の回帰係数 α^l) は時間的に安定しているか否かを確認する。そのために、2014 年データを用いて推計した回帰係数と、2016 年 10 月 1 日から 2019 年 9 月 30 日までの 3 年間（以降、2016～2019 年データ）から抽出した 1 年間分（時点数 365 日）のサンプルを用いて推計した回帰係数を比較する。ここで、データの収集期間の差によるバイアスを除くため、2016～2019 年データからのサンプルの抽出期間、すなわち時点数は 2014 年の 1 年間に合わせて 365 日としている。つまり、2016 年 10 月 1 日を開始時点としてサンプル数 365 日分を抽出してモデル推計を行い ($t_{start} = 1$)、次に時点を 1 日送り 2016 年 10 月 2 日からの 365 日サンプルでモデル推定を行う ($t_{start} = 2$)。これを、2018 年 10 月 1 日 ($t_{start} = 731$) まで繰り返し行う。具体的な 2016～2019 年データ ($t = 1 \sim 1095$) を用いた回帰係数の

推計手順は以下の通りである。

- step1 : 2016~2019 年データからのサンプルの抽出期間の開始時点を 2016 年 10 月 1 日 ($t_{start} = 1$)、終了時点を 2017 年 9 月 30 日 ($t_{finish} = 365$) とし、1 年間分の日別経路検索数と日別交通系 IC カード利用数を抽出する。
- step2 : step1 で抽出された日別経路検索数と日別交通系 IC カード利用数に対して、2 変量状態空間モデルを適用し、成分分解を行う。モデル設計は 6.3 節に示した通りである。ただし、ここでは時間的安定性のみを検証するために、路線周辺環境の違いによる推定結果への影響を避け、路線別にモデルは構築しない。したがって、3 路線を合算したことで全体の日別経路検索数と日別交通系 IC カード利用数の成分分解を行う。
- step3 : step2 で抽出した誤差成分に対して、重み付き回帰モデルを適用し、回帰係数 $\alpha_{t_{start}}$ を推計する。
- step4 : 2016~2019 年データからのサンプルの抽出期間の開始時点と終了時点を 1 日ずらし、開始時点は 2016 年 10 月 2 日 ($t_{start}=2$)、終了時点は 2017 年 10 月 1 日 ($t_{finish} = 366$) とし、新たに 1 年間分のデータを抽出し、step2 と step3 を実施する。

分析対象期間は 2016 年 10 月 1 日からの 3 年間であるため、step1 から step4 は $1095 - (t_{start} - t_{finish}) = 1095 - 364 = 731$ 回繰り返すことになる。

そして、2014 年データでの回帰係数 ($\alpha_{2014 \text{年}} = 5.87$) と 2016~2019 年データから 1 年間分のサンプルを抽出して推計した 730 個の回帰係数 $\alpha_{t_{start}}$ を比較する。本節では、2014 年データでの回帰係数と 2016~2019 年データから 1 年間分のサンプルを抽出して推計した回帰係数の平均値の比較を行い、2014 年から 2016~2019 年にかけて回帰係数が時間的に安定しているのかを検証する。さらに、2016~2019 年において、サンプルを抽出する期間によって、回帰係数が変動するか否かを検証する。

(II) 回帰係数の時間的観点からの検証

図 6.9 に 2016~2019 年データを用いて推計された回帰係数の時系列の推移を示す。横軸に示す年月日はサンプルの抽出期間の開始時点であり、それに対応する縦軸は開始時点を t_{start} としたときに推計された回帰係数 $\alpha_{t_{start}}$ である。例えば横軸左端の 2016 年 10 月 1 日は、2016 年 10 月 1 日~2017 年 9 月 30 日までのサンプルを用いてモデル推計した場合の回帰係数 $\alpha_{t_{start}=1}$ の値を示す。

2016~2019 年データから、サンプル抽出期間の開始時点 t_{start} と終了時点 t_{finish} を 1 日ずつずらし、1 年間分のサンプルを抽出して推計した 731 個の回帰係数の平均値 $\bar{\alpha}_{t_{start}}$ は 4.25 であった。この値は 2014 年データで推計した回帰係数 $\alpha_{2014 \text{年}} = 5.87$ に比べて小さい。6.2.1 項で述べたように、2014 年は 1 日あ

たりの平均経路検索数が 535 件に対して、2016 年 10 月 1 日からの 3 年間の 1 日あたりの平均経路検索数は 2,175 件と増加している。一方、交通系 IC カード利用数は 2014 年は平均 19,904 件に対して 3 年間の平均値は 22,831 件と微増である。このことから、 $\alpha_{2014\text{年}}$ に対して $\alpha_{t_{\text{start}}}$ の平均値が小さくなったのは、2014 年から 2016 年 10 月 1 日からの 3 年間にかける、経路検索数の大幅な変化量に対して、交通系 IC カード利用数の変化量が小さいかったことが影響したと考えられる。

2014 年データと 2016~2019 年データとでは、2016~2019 年データから 1 年間分のサンプルを抽出して推計した場合の方が、モデルに導入した経路検索数のオーダーが大きくなった一方で、交通系 IC カード利用数は微増だったため、突発的経路検索数が突発的交通系 IC カード利用数に与える影響を表現する回帰係数が小さくなったと解釈できる。

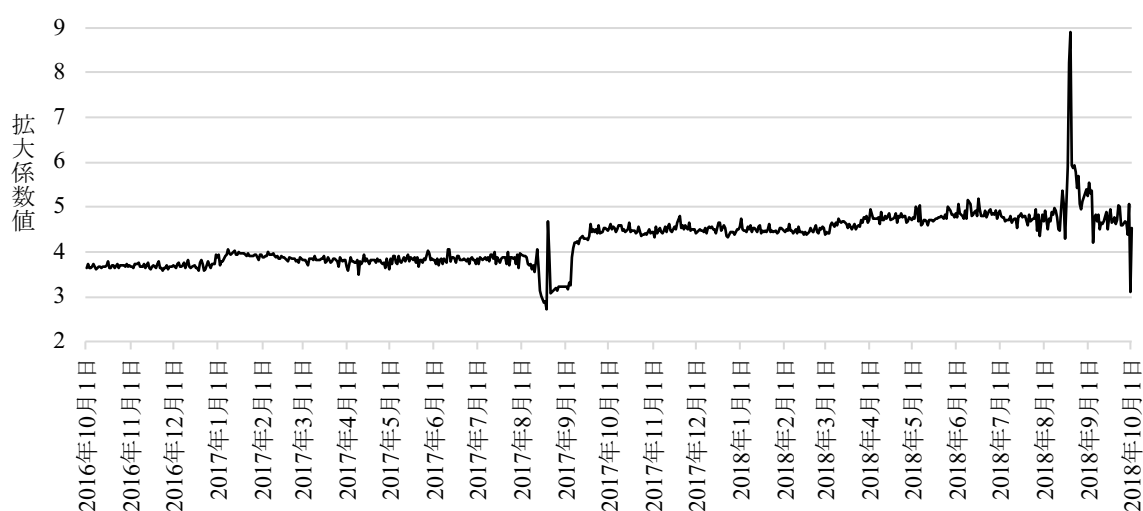


図 6.9 回帰係数の推移

次に、2016 年 10 月 1 日からの 3 年間で 1 年間分のサンプルを抽出する期間によって、回帰係数が変動する、すなわち突発的経路検索数と突発的交通系 IC カード利用数の関係が変化するのかを検証する。そのために、1 年間分のサンプル抽出期間の開始時点として 2016 年 10 月 1 日 ($t_{\text{start}}=1$) としたときの回帰係数 3.65 とその標準誤差 0.17 に基づいて、開始時点として 2016 年 10 月 2 日 ($t_{\text{start}}=2$)、2016 年 10 月 3 日 ($t_{\text{start}}=3$)、2016 年 10 月 4 日 ($t_{\text{start}}=4$)、... と 1 日ずつずらして、推計された回帰係数が、平均値 3.65、標準偏差 0.17 の正規分布に従うかを有意水準 5% でコルモゴロフ・スミルノフ検定する。ここで、1 年間分のサンプル抽出期間の開始時点によって異なる標本間の回帰係数を同時に比較するためにダネット法による多重比較検定を行う。そして、抽出期間の開始時点をいつまでとしたときの 1 年間分のサンプルを用いて推計された回帰係数が、平均値 3.65、標準偏差 0.17 の正規分布と同一の分布

とみなせるかを検定することで、回帰係数がデータ収集期間によって安定しているのかを検討する。

分析の結果、サンプル抽出期間の開始時点を 2016 年 10 月 2 日～2017 年 8 月 15 日で 1 日ずつずらし、推計される全 318 個の回帰係数は、サンプル抽出期間の開始時点を 2016 年 10 月 1 日としたときの回帰係数 $\alpha_{t_{start}=1}$ と同一の正規分布に従うことがわかった。図 6.9 に示すように、サンプル抽出期間の開始時点を 2016 年 10 月 1 日～2017 年 8 月 15 日としたときの回帰係数は開始時点によって微量に変化しているように見えるものの、それらは有意水準 5%で平均値 3.65、標準偏差 0.17 の正規分布に従うことから、サンプル抽出期間の開始時点が 2016 年 10 月 1～2017 年 8 月 15 日の間の回帰係数は時間的安定性を持つといえる。

しかし、サンプル抽出期間の開始時点を 2017 年 8 月 16 日以降としたときに推計された回帰係数は、平均値 3.65、標準偏差 0.17 の正規分布に従わなかった。図 6.9 に示すように、サンプル抽出期間の開始時点が 2017 年 8 月 16 日～2017 年 9 月 10 日の間では、推計された回帰係数は乱高下している。このような回帰係数の大きな変化の原因については明らかではなく、今後さらに検証する必要がある。しかし、この期間で回帰係数が安定していない有力な原因として、図 6.2 に示すように、2017 年 8 月 16 日から 2017 年 9 月 10 日の日別経路検索数が大きく変化している。2 変量状態空間モデルによる成分分解の初期時点がこの特異日と重なり、経路検索数のレベル成分が初期時点の状態値推計の影響を受けて適切に推計されず、それに伴って経路検索数の誤差成分が変化し、結果として回帰係数である突発的経路検索数と突発的交通系 IC カード利用数の重み付き回帰分析のパラメータが不安定になったと考えられる。

サンプル抽出期間の開始時点を 2017 年 9 月 11 日以降としたときに推計された回帰係数においても、平均値 3.65、標準偏差 0.17 の正規分布に従わなかった。そこで、2017 年 9 月 11 日以降で経路検索数と交通系 IC カード利用数の間に構造変化が起きたと考え、新たな正規分布を仮定し、多重比較検定を行った。具体的には、開始時点を 2017 年 9 月 11 日 ($t_{start}=346$) としたときの回帰係数 $\alpha_{t_{start}=346}=4.30$ とその標準誤差 0.18 に基づいて、開始時点を 2017 年 9 月 12 日 ($t_{start}=347$)、2017 年 9 月 13 日 ($t_{start}=348$)、2017 年 9 月 14 日 ($t_{start}=349$)、...と 1 日ずつずらして、推計された回帰係数が、平均値 4.30、標準偏差 0.18 の正規分布に従うかを検定した。その結果、開始時点が 2017 年 9 月 12 日～2018 年 5 月 2 日の間の回帰係数は、2017 年 9 月 11 日の回帰係数と同一の正規分布に従うことがわかった。したがって、サンプル抽出期間の開始時点が 2017 年 9 月 11 日～2018 年 5 月 2 日の間の回帰係数は時間的安定性を持つといえる。

同様に、開始時点を変更しながらダネット法による多重比較検定を行い、回帰係数が同一の正規分布に従う期間を分析した。表 6.1 に回帰係数の時間的安定性に関する結果をまとめる。表 6.1 より、

2016～2019年データにおいて、回帰係数はサンプル抽出期間の開始時点によって部分的に安定していることがわかった。

表 6.1 回帰係数の時間的安定性

サンプル抽出期間の開始時点	回帰係数の値	標準偏差	判定
2016年10月01日～2017年08月15日	3.65	0.17	安定
2017年08月16日～2017年09月10日			不安定
2017年09月11日～2018年05月02日	4.30	0.18	安定
2018年05月03日～2018年08月05日	5.02	0.20	安定
2018年08月06日～2018年08月13日	4.88	0.17	安定
2018年08月14日～2018年08月19日			不安定
2018年08月20日～2018年09月03日	5.94	0.12	安定
2018年09月04日			不安定
2018年09月05日～2018年09月27日	4.77	0.11	安定
2018年09月28日～2018年10月01日			不安定

ここで、2016～2019年データについて、サンプル抽出期間の開始時点の違いによる回帰係数の値の違いが、交通系 IC カード利用数の予測に大きく影響するならば、回帰係数は2016年10月1～2017年8月15日、2017年9月11～2018年8月2日...というように、部分的な期間で安定するといえる。一方で、回帰係数の値の違いが、交通系 IC カード利用数の予測に大きく影響しないならば、2016年10月1日からの3年間を通して、相対的に回帰係数は安定するといえる。そこで、回帰係数の値の違いが交通系 IC カード利用数の予測にどの程度影響を及ぼすのかを検討するために、2017年1月3日から2018年1月2日の回帰係数 $\alpha_{t_{start}=95} = 3.80$ と、2018年7月2日から2019年7月1日の回帰係数 $\alpha_{t_{start}=640} = 4.80$ を用いて、それぞれ推計される突発的交通系 IC カード利用数の予測値 $\hat{v}_{i,t}$ が交通系 IC カード利用数の予測値 $\hat{y}_{i,t}$ に占める割合の平均値を比較する。すなわち、式(6.8)で示した $\hat{y}_{i,t} = (\hat{x}_{i,t} + \hat{D}_i \mathbf{d}_i) + \hat{v}_{i,t}$ における $\hat{v}_{i,t}/\hat{y}_{i,t}$ の平均値を求める。モデルで推計される全体の交通系 IC カード利用数 $\hat{y}_{i,t}$ の平均値と突発的交通系 IC カード利用数 $\hat{v}_{i,t}$ の平均値、およびその比である $\hat{v}_{i,t}/\hat{y}_{i,t}$ の平均値を表 6.2 に示す。

表 6.2 より、突発的交通系 IC カード利用数が交通系 IC カード利用数に占める1日あたりの平均的な割合は、両サンプル抽出期間において約5%と変わりなかった。つまり、2017年1月3日～2018年1月2日の回帰係数と、2018年7月2日～2019年7月1日の回帰係数にはおよそ1程度の違いがあるものの、両期間ともに平均的には概ね5%程度の交通系 IC カード利用数の変動を経路検索数で説

明しているといえる。

サンプル抽出期間の開始時点によって一部、回帰係数の推移が大きく変動しているものの、相対して回帰係数は安定して推移していることから、提案手法の頑健性が示された。しかし一方で、部分的な回帰係数の安定がどの程度継続するかは明らかではないため、予測の立場からは回帰係数は長期的には変動していると考えらるべきである。つまり、突発的経路検索数の突発的交通系 IC カード利用数の反映度合いは変化するものと考えられることから、例えば、2014 年に収集された経路検索数と交通系 IC カード利用数を用いて推計されたモデルを、2016 年の予測モデルとして適用することは適切とはいえない。部分的には安定していることから短期的にモデルを更新する必要はないが、定期的に提案手法による回帰係数の推計、確認を行う必要があるといえる。

表 6.2 回帰係数の値の差による交通系 IC カード利用数予測への影響

サンプル抽出期間	平均交通系 IC カード利用数 ($\hat{y}_{i,t}$)	平均突発的 交通系 IC カード利用数 ($\hat{v}_{i,t}$)	平均比 ($\hat{v}_{i,t}/\hat{y}_{i,t}$)
2017 年 1 月 3 日から 2018 年 1 月 2 日まで	21,722 件/日	1,086 件/日	0.05
2018 年 7 月 2 日から 2019 年 7 月 1 日まで	23,219 件/日	1,160 件/日	0.05

6.4.2 回帰係数の空間的安定性の検証

(I) 回帰係数の検証方法

回帰係数が空間的に安定しているのか、あるいは変化するのかを確認する。そのために、2014 年データをを用いて推計した路線別に回帰係数と、2016～2019 年データから 1 年間分のサンプルを抽出し、路線別に推計した回帰係数を比較する。路線ごとの回帰係数の推計手順は時間的安定性の検証 (step1～4) のときと同様である。

(II) 回帰係数の空間的観点からの検証

表 6.3 に 2014 年データを用いて日別経路検索数と日別交通系 IC カード利用数の成分分解を行い、抽出した誤差成分を用いて推計した路線別 (琴平線, 長尾線, 志度線) の重み付き回帰モデルの結果と

系全体の推計結果を示す。表 6.3 より、琴平線の回帰係数は 7.13、長尾線の回帰係数は 2.79、志度線の回帰係数は 3.01 と系全体での回帰係数 5.87 と大きく異なることがわかった。これは、路線ごとに駅周辺環境が異なるため、突発的経路検索数の影響度合いが変わることが原因と考えられる。最も回帰係数の大きい琴平線周辺は他の 2 つの路線に対して大型商業施設や観光施設が路線周辺に多く分布しており、路線上に位置する駅数も多い。そのため、イベント開催によることでん利用や天候変化によることでん利用が他路線と比べて多くなると考えられる。実際、2014 年における 1 日あたりの平均交通系 IC カード利用数が琴平線は 11,321 件、長尾線は 6,059 件、志度線は 4,130 件に対して、1 日あたりの平均経路検索線数が琴平線は 264 件、長尾線は 200 件、志度線は 160 件であり、琴平線は平均経路検索数に対して平均交通系 IC カード利用数のオーダーが大きい。したがって、琴平線上の駅を指定した突発的経路検索数のオーダーに対して、突発的交通系 IC カード利用数のオーダーが大きいことから、回帰係数も大きくなったと考えられる。一方、長尾線と志度線については、経路検索数と交通系 IC カード利用数のオーダーに琴平線ほどの差はないために、回帰係数が小さいといえる。

表 6.3 路線別重み付き回帰モデルの推定結果と系全体の重み付き回帰モデルの推定結果

	琴平線	長尾線	志度線	系全体
変数	推定値	推定値	推定値	推定値
定数項	-58.65	-7.66	-16.21	-80.76
回帰係数	7.13*	2.79*	3.01*	5.87*
サンプル数	365	365	365	365
決定係数	0.85	0.53	0.86	0.68
自由度調整済み決定係数	0.85	0.52	0.86	0.65

*：有意水準 1%

次に、2014 年データの路線別回帰係数と 2016～2019 年データから抽出した 1 年間分のサンプルを用いて推計した路線別回帰係数を比較する。図 6.10 から 6.12 に路線ごとの回帰係数の推定値の推移を示す。

琴平線における回帰係数について、2014 年データでの推計（回帰係数 7.13）に対して、2016～2019 年データから抽出した 1 年間分のサンプルを用いて推計した場合の回帰係数の平均値は 3.26 と小さくなった。この要因は時間的観点での検証で述べたように、2014 年データに対して 2016～2019 年データの、経路検索数のオーダーが大きくなった一方で、交通系 IC カード利用数は微増だったためと考えられる。

長尾線における回帰係数について、2014 年データでの回帰係数 2.79 と比較すると、2016～2019 年データから抽出した 1 年間分のサンプルを用いて推計した場合の回帰係数の平均値は 3.72 となり、大き

くなることがわかった。

志度線における回帰係数については、志度線の 2014 年データを用いた推計では回帰係数は 3.01 であった。対して、2016～2019 年データから抽出した 1 年間分のサンプルを用いて推計した場合の回帰係数の平均値は 3.30 であり、志度線については平均値で見ると回帰係数はあまり変化しないことがわかった。

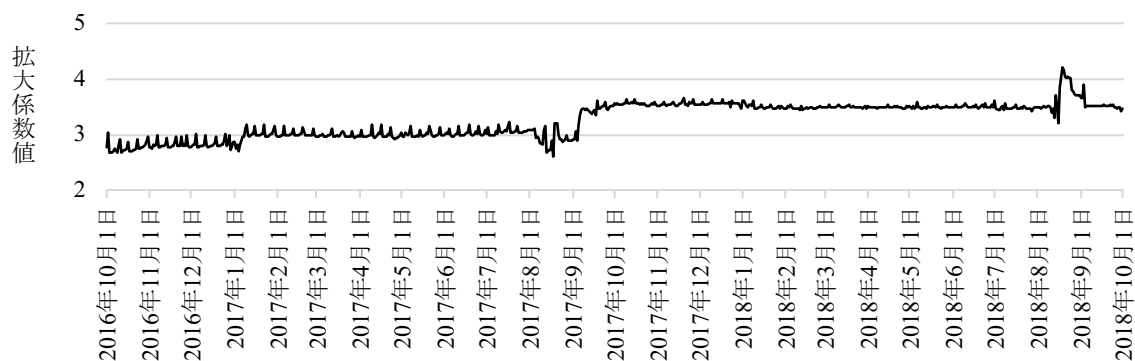


図 6.10 琴平線の回帰係数の推移

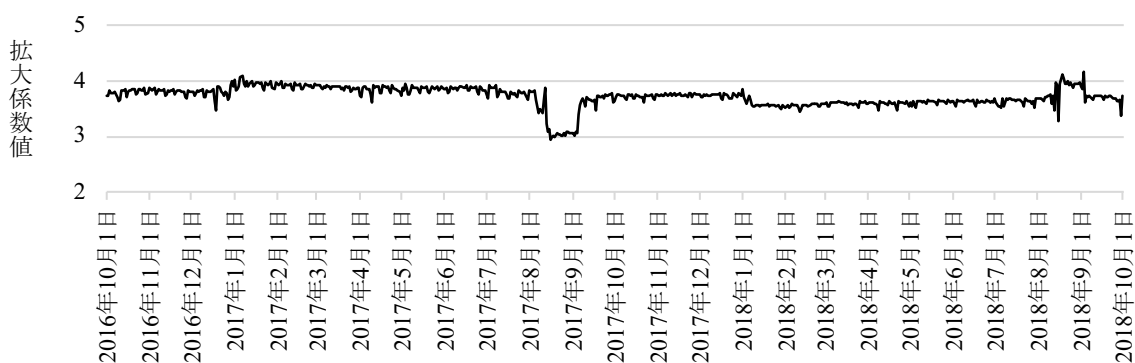


図 6.11 長尾線の回帰係数の推移

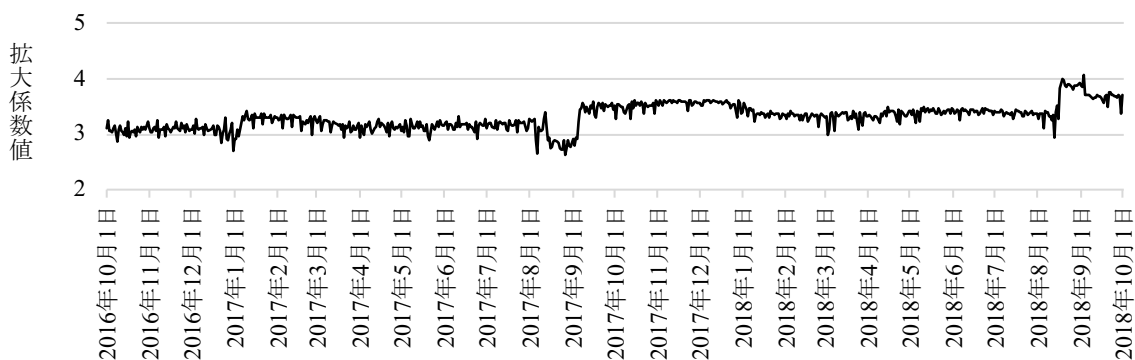


図 6.12 志度線の回帰係数の推移

次に、路線別に、2016年10月1日からの3年間で1年間分のサンプルを抽出する期間によって回帰係数が変動するのかを、時間的観点での検証で述べた同様の手順で多重比較検定によって検証した。各路線の回帰係数の時間的安定性を持つといえる開始時点を図6.13にまとめる。

図6.13より、琴平線においては、開始時点が

- ・2016年10月1日～2017年9月5日（黄色①）
- ・2017年9月6日～2018年8月16日（黄色②）
- ・2018年8月18日～2018年8月24日（黄色③）
- ・2018年8月25日～2018年9月3日（黄色④）
- ・2018年9月4日～2018年9月30日（黄色⑤）

で、異なる回帰係数を持つが、各期間内では回帰係数は同一の正規分布に従うと判定され、時間的に安定しているといえる。

同様に、長尾線については、

- ・2016年10月1日～2017年8月13日（緑色①）
- ・2017年8月14日～2017年9月5日（緑色②）
- ・2017年9月6日～2018年8月17日（緑色③）
- ・2018年8月18日～2018年9月3日（緑色④）
- ・2018年9月4日～2018年9月30日（緑色⑤）

で、志度線については、

- ・2016年10月1日～2018年8月6日（桃色①）
- ・2018年8月17日～2018年9月19日（桃色②）
- ・2018年9月21日～2018年9月29日（桃色③）

で、それぞれの期間で回帰係数は時間的安定性を持つことがわかった。

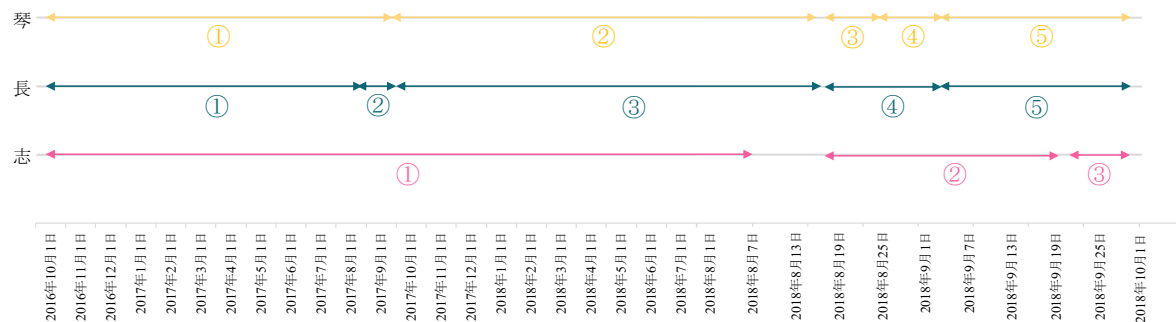


図 6.13 各路線の回帰係数の時間的安定性

図 6.13 より、琴平線と長尾線は回帰係数の値が変化する開始時点に類似性があり、変化の回数も多い。志度線は 2 路線とは異なり、開始時点が 2016 年 10 月 1 日から 2018 年 8 月 6 日と長い期間において、回帰係数が安定している。3 路線で共通するのは、2018 年 8 月 17 日頃を境に回帰係数の値が変化することがわかった。この原因については明らかではないため今後検証する必要がある。

以上の検証から、

- (1) 分析対象期間を変更しても提案手法は適用可能か（時間的安全性）
- (2) 路線（地域）を変更しても提案手法は適用可能か（空間的安全性）

に対して、

- (1) 短期的に回帰係数は安定しているものの、回帰係数が変化する時期の特定は難しい。
- (2) 路線別に回帰係数は異なるものの、(1) 時間的安全性の特性は路線間で共通である。

が明らかとなった。

6.5 事前収集時間別予測精度の検証

6.5.1 事前収集時間別の予測精度の確認方法

前節で構築した予測モデルをもとに、どのくらい前に収集された経路検索数であれば突発的需要を含む交通需要が予測可能かを検討する。本項では、経路検索数を事前に集める時間帯について、

- (1) 始発直前（予測日当日 4 時台）まで、
- (2) 始発 3 時間前（予測日当日 2 時台）まで、
- (3) 始発 6 時間前（予測日前日 23 時台）まで、
- (4) 始発 12 時間前（予測日前日 17 時台）まで

を設定し、予測精度を比較する。以後、便宜上、(1) ～ (4) を事前収集時間と呼ぶ。

図 6.14 に事前収集時間別経路検索数の概念図を示す。図 6.14 に示すように、事前収集時間までに累積された経路検索数（色付部分の面積）でどの程度交通需要が予測可能かを検証する。

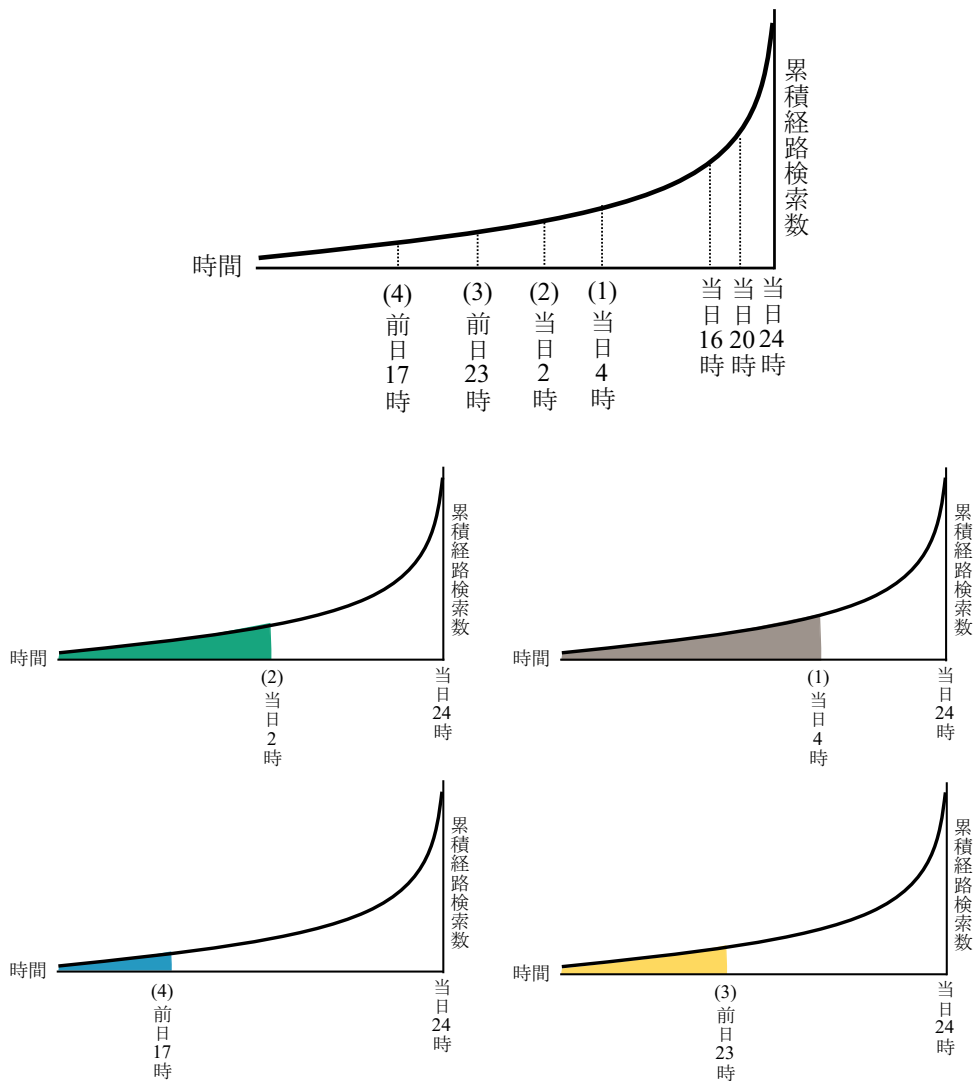


図 6.14 事前収集時間別経路検索数の概念図

図 6.15 に予測フローを再掲する (5.7 節の図 5.13 再掲). 事前収集時間を (1) とし, 路線 l の 2014 年 12 月 1 日の交通系 IC カード利用数を予測する場合, データ収集期間は 2014 年 1 月 1 日から 11 月 30 日の始発直前までである. このとき, 経路検索数について, 各日においてその日を指定した検索の検索時間 (検索者が検索を行なった時間) がその日の 4 時台までの経路検索数, つまり 2014 年 1 月 1 日を指定した検索で検索時間が 1 月 1 日 4 時台までの経路検索数, 1 月 2 日を指定した検索で検索時間が 1 月 2 日 4 時台までの経路検索数, ..., 11 月 30 日を指定した検索で検索時間が 11 月 30 日 4 時台までの経路検索数を日別経路検索数として用いる. そして, 各日の 4 時台までに収集された経路検索数と, 各日の交通系 IC カード利用数を用いてパラメータ推定を行い, 12 月 1 日 4 時台までに収集された経路検索数に基づいて 12 月 1 日の交通系 IC カード利用数を予測する.

次に、12月2日の交通系ICカード利用数を予測する場合、データ収集期間を1日増やし、2014年1月1日から12月1日とする。このとき、12月1日の交通系ICカード利用数を予測したときと同様に、経路検索数については各日を指定した検索の検索時間がその日の4時台までの経路検索データである。そして、各日の4時台までに収集された経路検索数と、各日の交通系ICカード利用数を用いてパラメータ推定を行い、12月2日4時台までに収集された経路検索数に基づいて12月2日の交通系ICカード利用数を予測する。以後、次の時点の交通系ICカード利用数を予測する場合、データの収集期間を1日ずつ増やししながら、パラメータ推計を行い、予測日4時台までに収集された経路検索数に基づいて交通系ICカード利用数を予測する。なお、モデル推計は路線ごとに行う。

1段階目：2変量状態空間モデルによる成分分解
 2段階目：重み付き回帰モデルによる関係分析

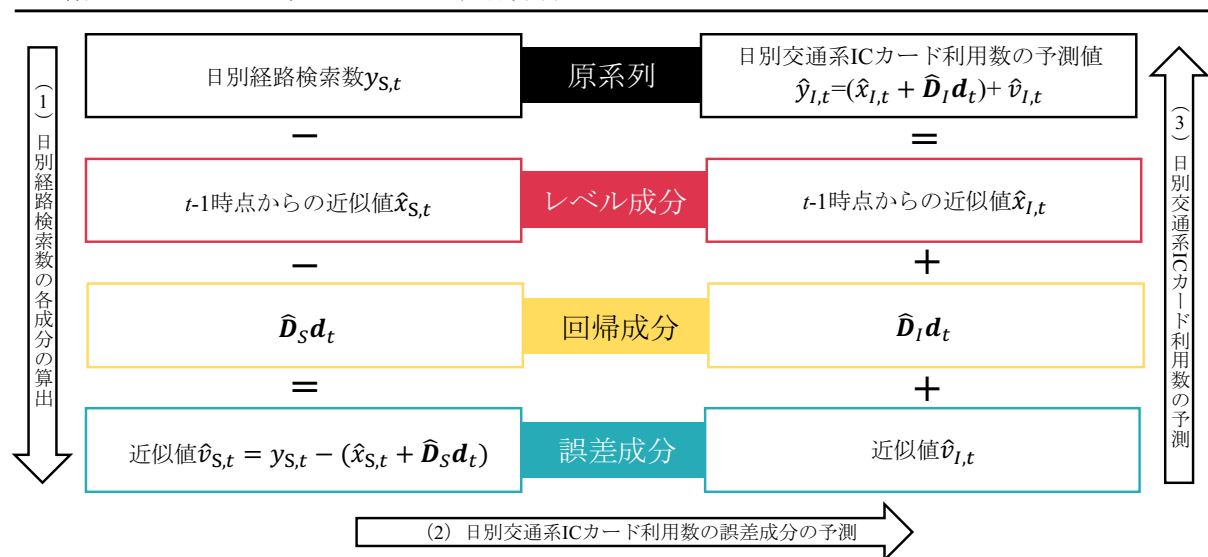


図 6.15 提案モデルの予測フロー (再掲)

6.5.2 事前収集時間別経路検索数の推移

路線ごとに事前収集時間別経路検索数を比較する。図 6.16 から図 6.18 に琴平線、長尾線、志度線の事前収集時間別経路検索数の推移をそれぞれ示す。また、表 6.4 に事前収集時間を考慮しない場合の日別経路検索数と事前収集時間を考慮した場合の日別経路検索数の相関係数を示す。図 6.16 より、琴平線について、事前収集時間を考慮すると、事前収集時間が長いほど徐々に経路検索数は少なくなっている。しかし、表 6.4 より、事前収集時間を考慮しない場合の日別経路検索数と事前収集時間を考慮し

た場合の日別経路検索数の相関関係をみると、経路検索数の時系列の変動は事前収集時間に関わらず、概ね同様の傾向を示すことがわかった。図 6.17, 6.18 および表 6.4 より、長尾線と志度線についても琴平線と同様である。

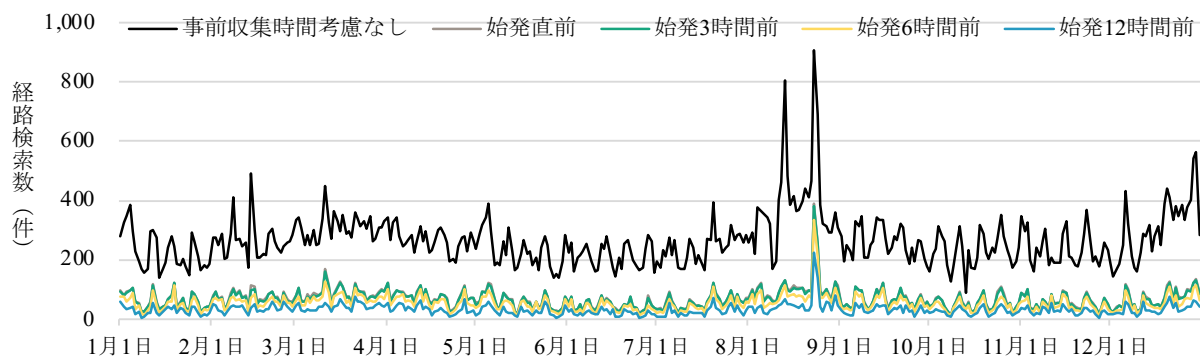


図 6.16 琴平線の事前収集時間別経路検索数の推移

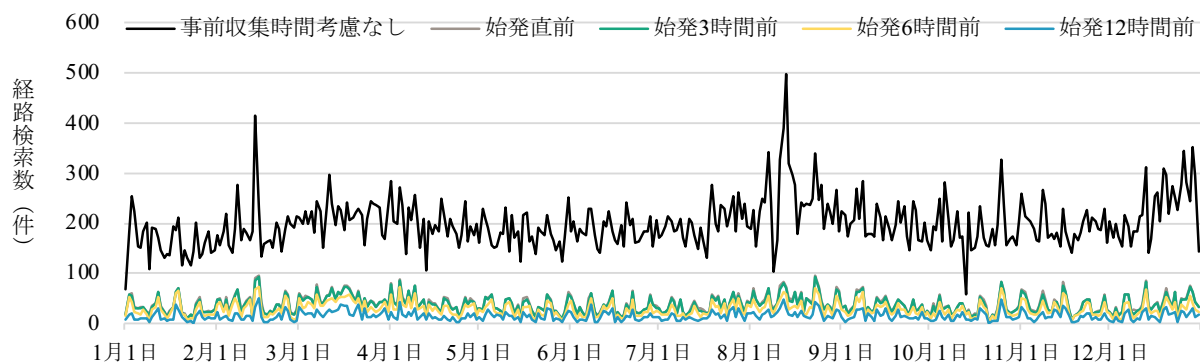


図 6.17 長尾線の事前収集時間別経路検索数の推移

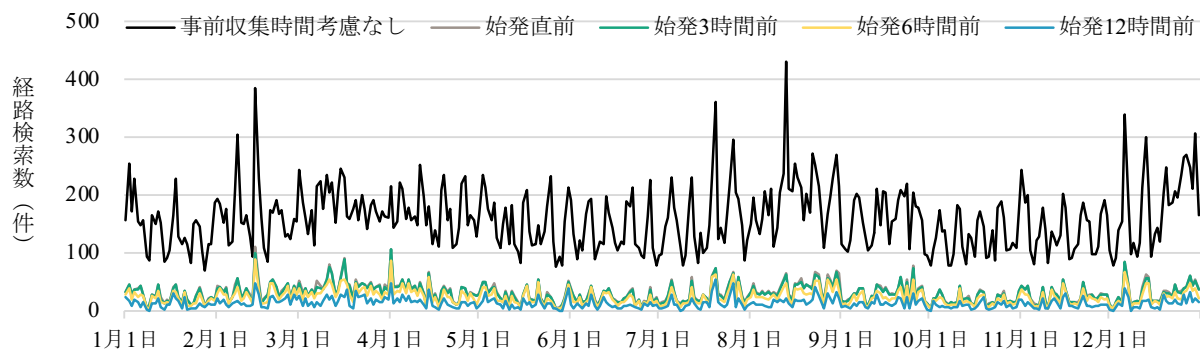


図 6.18 志度線の事前収集時間別経路検索数の推移

表 6.4 各路線の事前収集時間別相関係数

路線	始発直前	始発 3 時間前	始発 6 時間前	始発 12 時間前
琴平線	0.85	0.84	0.82	0.76
長尾線	0.64	0.63	0.56	0.54
志度線	0.78	0.77	0.74	0.65

6.5.3 路線ごとの事前収集時間別予測精度の検証

前項で集計した事前収集時間別の経路検索データと、交通系 IC カードデータを用いて、路線ごとに 12 月の各日の交通需要の事前予測を行ったときの予測精度を事前収集時間別に比較する。予測精度の指標には RMSE を用いる。

3 路線ごとの事前収集時間別 RMSE を表 6.5 に示す。表 6.5 中の丸括弧内は 12 月の平均交通系 IC カード利用数（件/日）に対する誤差を示す。表 6.5 より、予測精度を表す RMSE は事前収集時間別に路線ごとで大差はない。事前収集時間を考慮しない経路検索数に基づくモデルの路線ごとの RMSE は、琴平線 1,460 件（誤差 13%）、長尾線 706 件（誤差 11%）、志度線 531 件（誤差 13%）であり、どの路線もいずれの事前収集時間の経路検索数を用いた場合でも、良好な予測精度を得られているといえる。ただし、どの路線においても 12 月 29 日以降の予測値は観測値と大きくずれていた。これは予測の終盤が年末という特異日になったことが影響していると考えられる。そこで、28 日までを予測対象として RMSE を算出したところ、各路線の 12 月の平均交通系 IC カード利用数（件/日）に対して 6~9%の誤差で予測できることが明らかとなった。

表 6.5 路線ごとの事前収集時間別 RMSE

事前収集時間	RMSE（件）：全日			RMSE（件）：28 日まで		
	琴平線	長尾線	志度線	琴平線	長尾線	志度線
直前	1,653（15%）	843（14%）	687（17%）	953（9%）	411（7%）	237（6%）
3 時間前	1,648（15%）	844（14%）	689（17%）	939（8%）	410（7%）	237（6%）
6 時間前	1,634（15%）	842（14%）	677（17%）	902（8%）	413（7%）	237（6%）
12 時間前	1,571（14%）	855（14%）	673（17%）	841（8%）	417（7%）	233（6%）

6.6 まとめ

6.6.1 6章の研究成果

本章では、第1段階として2変量状態空間モデルによって経路検索数と交通系ICカード利用数の突発的変動成分を抽出し、第2段階として重み付き回帰モデルによって突発的経路検索数と突発的交通系ICカード利用数の関係を分析する提案モデルの有用性を検討した。そのために、(1)分析対象期間の変更による提案手法の適用可能性と、(2)分析対象地域(路線)の変更による提案手法の適用可能性を検証した。その結果、2016年10月1日から2019年9月30日の3年間に於いて、2017年9月頃や2018年9月頃に回帰係数が大きく変化するものの、回帰係数は時間的に安定していることが明らかとなった。また、路線別の回帰係数の推定値とその推移から、路線ごとに回帰係数は異なるものの、路線別には回帰係数が時間的安定性を持つことがわかった。

そして、時間的、空間的観点からの回帰係数の変動性の検証の上で、事前に収集された経路検索数を用いた交通需要予測モデルを開発し、(3)どのくらい前に収集された経路検索数であれば、高精度に予測可能かを検討した。その結果、事前収集時間を考慮することによって、経路検索数のサンプル数が少なくなるものの、その時系列変動は事前収集時間を考慮しない場合と概ね同じ傾向を示し、予測日前日夕方までに収集した経路検索データであっても、比較的サンプル数の多い予測日直前に収集した経路検索データと同程度の精度で予測できることがわかった。

本提案手法は、経路検索システムの検索履歴データと交通系ICカードの乗降履歴データの両方を用いて、将来の交通需要を予測するものである。本章の結果から、経路検索データを用いることによって、突発的交通系ICカード利用数を含む日々の交通系ICカード利用数が事前に予測できることが示された。これによって、データ収集が困難なイベント情報や天候情報を用いずとも、日々変動する交通需要を予測できることを実証的に明らかにした。

6.6.2 6章における今後の課題

本章では時間的、空間的観点からの回帰係数の安定性の検証する際、サンプルを抽出する期間を1年間(時点数365)とした。これは時系列モデルのパラメータ推計において、既往研究の多くが1年以上に渡って収集したデータを用いており、適切なパラメータ推計には最低1年間分のサンプルがあれば

十分と考えたためである。一方、サンプル数が多いほどパラメータ推計は安定するといえる。しかし、サンプル数が多いほど計算に時間がかかるだけでなく、ノイズが含まれる可能性が高まる。そこで、本提案モデルにおいて、どの程度の期間のサンプル数を用いることが高精度に予測できるかを検証する必要がある。

本事例においては、地方都市の交通ビッグデータが持つサンプル数の小ささを踏まえて、系全体と路線ごとの分析を行なった。しかし、タクシーの配置問題や、デマンドバスの運行区間、共助交通の実際の導入を考えると、乗車駅・降車駅別に分析単位を落としたより詳細な交通需要の予測が必要と考えられる。その際、収集される乗車駅・降車駅別経路検索数と、乗車駅・降車駅別交通系 IC カード利用数には、ノイズとなるような発生量しか持たない乗車駅・降車駅が存在すると考えられる。そのため、すべての乗車駅・降車駅について交通需要が予測できるとはいえない。そこで、本研究成果により明らかとなった、相対的に経路検索数が多いとされる観光施設や商業施設が立地するような場所に分析範囲を絞り、交通需要の予測を行う。そして、乗車駅・降車駅周辺の特徴と予測精度との関係が明らかとなれば、提案モデルのサンプル量に対する頑健性の議論が可能といえる。また、本提案モデルに基づく交通需要予測に必要な最低サンプル量が明らかとなれば、経路検索データが十分に収集できない場合であっても、第 4 章で推計されたパラメータを用いて人口分布と立地施設に基づく潜在的需要が表現可能になると考えられる。

第6章 参考文献

高瀬達夫, 森川高行, 阿野貴史, 長澤貴憲 (2000), モデルの移転可能性に焦点をあてた空港アクセス交通の日集計分析, 土木計画学研究・論文集, Vol.17, pp.855-862.

結論

〈第 7 章の概要〉

「経路検索データは実行動者数の予測に貢献するのか？」

経路検索データは新たな行動データとして活用が期待されているものの、日常的な外出時に経路検索システムは利用しないことや、経路検索を行うだけでシステム上で掲示された経路通りの行動を実行するのかが不明であることなどから、この問いに対して、誰も答えることができなかった。

本研究では、「経路検索数と実行動者数には直接的な関係はなく、両者に包含される突発的事象の発生に起因する需要のみが関係する」という仮説に基づき、2 変量状態空間モデルと重み付き回帰モデルによる経路検索データを用いた交通需要予測を提案し、実証分析を通じて提案手法の有効性を示した。

これにより、Web 検索してから実行動を行う現代において、これからの交通サービスの根幹になると考えられる経路検索システムの新たな活用方法を示すことができた。

本章では、得られた結果を整理するとともに、本研究の成果を新たな移動サービスの導入検討に活用するための道筋を示す。

7.1 本研究のまとめ

本研究では、公共交通の需要変動の先行指標として経路検索データに着目し、経路検索数から交通系 IC カード利用数を予測するための分析手法を構築した。その際、

仮説 1 : 「経路検索システムと交通系 IC カードは利用されやすい状況が異なる」

ことを検証し、

仮説 2 : 「両者に包含される突発的事象に起因する需要部分を適切にモデル化できれば、経路検索数は交通系 IC カード利用数の予測精度向上に寄与する」

ことに基づき、Web 上に表明される移動意向と実空間上に顕在化する交通需要を関連づける方法論を提案した。以下に、各章の分析から得られた知見をまとめる。

第 1 章と第 2 章では、従前の交通需要予測モデルの需要応答型交通や MaaS への非対応性、突発的需要を含む日々の交通需要予測の必要性、SP データとしての経路検索データの活用可能性を整理し、本研究の目的を示した。その上で、RP 交通ビッグデータに基づく交通需要予測モデルに関する研究と経路検索データに関する研究を整理し、本研究の位置づけを示した。

第 3 章と第 4 章では、「経路検索システムと交通系 IC カードは利用されやすい状況が異なる」という仮説 1 の妥当性を検討するために、経路検索行動の特性把握を行い、経路検索データに内在する需要を明らかにした。

第 3 章では、「いつ」「どこから」「どこへ」の移動が検索されやすいのかを概観することを目的に、曜日、乗車駅、降車駅の 3 次テンソルデータからデータが持つ特徴を抽出した。本章では、高次数・高次元データを分析する際、次数（要素軸）の組合せの順序によって発生する誤差の生じ方の違いを回避するために、非負値テンソル因子分解を用いたパターン抽出方法を提案した。3 次数を同時に考慮した分析によって抽出された検索パターンから、本事例においては週の後半曜日（水曜日、木曜日、金曜日）に高松市中心部に位置する駅から、中心部の特定の駅へ移動するための検索が最も多いことがわかり、高松市中心部内の移動に対してニーズがあることが明らかとなった。非負値テンソル因子分解は平均的パターンの抽出に長けているという特性を踏まえると、抽出されたパターンでの検索は安定的に行われるものであり、集計レベルでは定常的に発生すると考えられる。

第 4 章では、検索者が指定する駅が検索されやすい要因を明らかにすることを目的に、経路検索数の多寡と駅周辺環境の関連性を推定した。その際、本事例で扱う経路検索データが地方都市の交通ビ

ッグデータであることに起因する、(1) 乗車駅・降車駅別経路検索数に 0 件データが多量に存在する問題、(2) 乗車駅・降車駅別経路検索数の多少に大きく差がありデータ分布に偏在性が存在する問題と、モデル構築の観点からは、(3) モデルに導入する説明変数の候補が大量にあるという変数選択問題を回避するために、Bayesian Adaptive Lasso トービット分位点回帰モデルを構築し、パラメータ推定した。その結果、経路検索数の多さには駅周辺の医療施設や商業施設、観光施設の立地が影響することが明らかとなり、検索者は非日常時の通院や購買行動、観光行動に伴って経路検索システムを利用すると推察された。

以上、第 3 章と第 4 章の分析に基づいて整理された経路検索データが表現する行動と需要に加え、交通系 IC カードデータが表現する行動と需要を表 7.1 に示す (1.3 節の表 1.1 再掲)。表 7.1 より、経路検索データと交通系 IC カードデータでは日々一定数発生する行動が異なる。このことは、「**経路検索システムと交通系 IC カードは利用されやすい状況が異なる**」ことを示唆し、仮説 1 の考え方は妥当といえる。さらに、経路検索データと交通系 IC カードデータでは反映される個人レベルの行動に違いがあることから、それら行動を集計した定常的需要の性質は異なる。また、集計レベルの需要を構成する定常的需要と突発的需要の比率も両データでは相違がある。したがって、経路検索数と交通系 IC カード利用数は、経路検索数が多いと交通系 IC カード利用数も多いという単純な関係では表現できないと考えられる。

表 7.1 経路検索データと交通系 IC カードデータが表現可能な「行動」と「需要」(再掲)

経路検索データ		交通系 IC カードデータ	
個人レベル	集計レベル	個人レベル	集計レベル
<ul style="list-style-type: none"> ・ 非定期的行動 (例) 非定期的な買い物や通院, 観光や出張 	<ul style="list-style-type: none"> ・ 定常的需要 	<ul style="list-style-type: none"> ・ 定期的行動 (例) 通勤・通学, 定期的な買い物や通院 ・ 非定期的行動 (例) 非定期的な買い物や通院, 観光や出張 	<ul style="list-style-type: none"> ・ 定常的需要
<ul style="list-style-type: none"> ・ 非定期的行動 (例) イベント参加や悪天候回避 	<ul style="list-style-type: none"> ・ 突発的需要 	<ul style="list-style-type: none"> ・ 非定期的行動 (例) イベント参加や悪天候回避 	<ul style="list-style-type: none"> ・ 突発的需要

そこで、続く、第 5 章と第 6 章では、仮説 1 を検証した上で、経路検索数と交通系 IC カード利用数に関連づけるための方法論を提案し、経路検索数に基づく交通需要予測モデルを構築した。そして、

「両者に包含される突発的事象に起因する需要部分を適切にモデル化できれば、経路検索数は交通系 IC カード利用数の予測精度向上に寄与する」という仮説 2 を検証するとともに、提案モデルの汎用性と適用限界、および公共交通の需要変動の事前予測可能性を検証した。

第 5 章では、経路検索数と交通系 IC カード利用数の関係性を明らかにすることを目的に、両者の関係を適切に表現するための方法論を提案した。その上で、経路検索数を用いた公共交通の需要予測モデルを構築した。

第 3 章および第 4 章では、仮説 1 の妥当性は検討できたものの、検証はできていない。そこで、2014 年の日別経路検索数と日別交通系 IC カード利用数の全日数での相関係数と、平日と休日で分けた相関係数を算出した。その結果、全体での相関と全体を分割した層間の相関は異なり、シンプソンのパラドックスの発生が確認された。これより、「経路検索システムと交通系 IC カードは利用されやすい状況が異なる」という仮説 1 が立証され、2 つのデータは直接的に結び付けられないことが示された。

そこで、第 5 章では、2 変量状態空間モデルを用いて、経路検索データと交通系 IC カードデータの両データに内在する需要を定常的需要と突発的需要に分解し、2 種類の需要の相関関係から、経路検索数と交通系 IC カード利用数の関係性の解明を試みた。両データの定常時の相関を表すレベル成分の過程誤差の相関が負値となったことから、通勤・通学のような定期的行動が多く行われる日は交通系 IC カード利用数が増加し、経路検索数は減少する、あるいは出張や観光のような非定期的行動が多く行われる日は経路検索数が増加し、交通系 IC カード利用数は減少するという関係性が明らかとなった。一方、イベント開催や天候変化のような突発的事象発生時の相関を表す誤差成分の観測誤差の相関が正値となったことから、何らかの要因によって突発的に経路検索数が増減すると交通系 IC カード利用数も増減するという関係性が明らかとなった。これより、経路検索数と交通系 IC カードは、突発的事象の発生を共通要因とする需要の部分で関連づけられることが示された。そして、この突発的需要の関係、すなわち突発的経路検索数と突発的交通系 IC カード利用数の関係を重み付き回帰モデルによってモデル化した。

その上で、経路検索数に基づく交通系 IC カード利用数の予測手法を構築した。提案モデルの需要予測のフローを図 7.1 に示す（5.7 節の図 5.13 再掲）。提案手法により推計した日別交通系 IC カード利用数の予測値は日平均交通系 IC カード利用数に対して 7% の誤差で現況再現できることが明らかとなった。さらに、提案手法の予測精度は従前モデルよりも向上することが示された。これによって、「両者に包含される突発的事象に起因する需要部分を適切にモデル化できれば、経路検索数は交通系 IC カード利用数の予測精度向上に寄与する」ことを立証できた。

1段階目：2変量状態空間モデルによる成分分解
 2段階目：重み付き回帰モデルによる関係分析

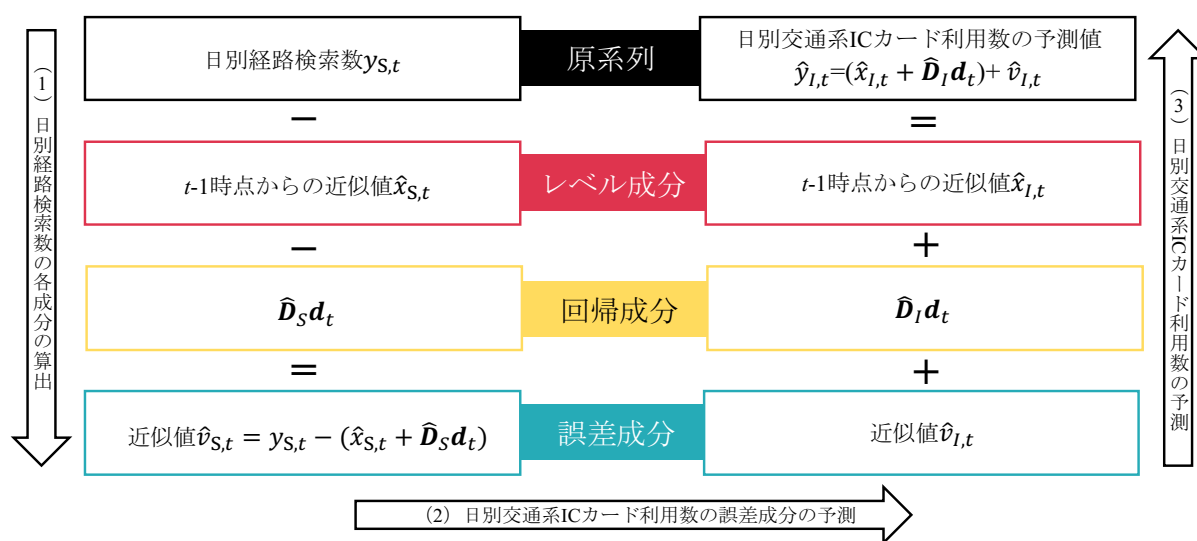


図 7.1 提案モデルの予測フロー（再掲）

第 6 章では、提案手法の適用限界と有用性を検証することを目的に、(1) 分析対象期間の変更による提案手法の適用可能性、(2) 分析対象地域（路線）の変更による提案手法の適用可能性、(3) 需要変動の事前予測可能性を明らかにした。(1) と (2) については、突発的経路検索数の突発的交通系 IC カード利用数への反映度合いを表す回帰係数の時間的・空間的安定性の観点から検証し、回帰係数は短期的に安定していることと、路線別では変化するが路線間で共通して短期的な時間的安定性を持つことを示した。(3) については、予測日始発以前に収集可能な経路検索数を用いて、提案モデルの予測精度を検証し、経路検索データは交通需要の変動の先行指標になることを示した。

以上、第 3 章から第 6 章までの分析より、本研究の目的である「経路検索数に基づいて日々の交通需要を事前に予測可能とする方法論の開発とその有用性を示すこと」は達成できたといえる。

7.2 本研究の成果

7.2.1 交通計画における経路検索データの有用性の実証

経路検索データは公共交通の潜在需要を反映した交通ビッグデータとして注目されていたものの、交通計画において有用なデータであるのかはこれまで明示されてこなかった。さらに、SP 交通ビッグ

データとしての活用可能性についても議論が十分ではなかった。これは、

- 検索者の個人属性や移動目的には偏りが存在する可能性があること
- 定期的な外出では経路検索システムは利用しないこと
- 検索者が必ずしも公共交通を利用するとは限らないこと

などに起因し、(1) 経路検索データが表現する潜在需要の特性や、(2) 実際に顕在化する公共交通の需要との関連性が不明確であったためである。

それゆえ、

「経路検索データを用いて、実際の公共交通利用者数を予測できるのか？」

という問いに対して、誰も言及できなかった。

そこで、本研究では、Web 上の潜在需要の特性に関する仮説 1 と Web 上の潜在需要と実空間上の顕在需要を関連づける方法に関する仮説 2 を仮定し、第 3 章と第 4 章の分析から (1) 経路検索データが表現する潜在需要の特性を明らかにした上で、第 5 章と第 6 章の分析で、(2) 実際に顕在化する公共交通の需要との関連性をモデル化し、経路検索データの交通需要予測への有用性を示すことができた。

経路検索システムは、都市部に限らず、地方部、あるいは過疎地域の公共交通の情報まで提供されるようになり、我が国においてはかなりの広域エリアをカバーしている。また、GTFS (General Transit Feed Specification) によって公共交通に関するデータの標準化が進んでおり、経路検索システムの利用範囲は今後一層高まるものと考えられる。経路検索システムを提供する主体は、Yahoo, Google, NAVITIME などの大手企業、各地の交通事業者や地方自治体など多数ある。これら主体から提供されるシステムにおいて、本研究で提示したような経路検索データは日々蓄積されている。経路検索データは、ETC2.0 や交通系 IC カードデータ、モバイル空間統計データなどに比べ、情報開示が進んでいるとはいえず、誰でもアクセスできるようなデータではない。しかし、ETC2.0 や交通系 IC カードデータも以前はデータにアクセスできる人は限られ、簡単に入手できるものではなかったが、サービスの利用者が増えるとともに情報開示されるようになった。さらに、ビッグデータ時代、DX 時代と呼ばれるようにデータ分析の価値が共有されるようになり、急速にその開示が進んでいる。したがって、経路検索データの開示も今後積極的に行われるであろう。それによって、これまでデータの普及とともにそれに関わる研究蓄積が増えていったように、経路検索数から実行動者数を予測するような研究も増えていくと考えられる。研究蓄積を増やすことで経路検索数と実行動者数の関係性がより明確になれば、実行動に関するデータがなくても経路検索データだけで実行動者数を予測できるようになる可能性もあり、実行動のデータ収集が困難な地方都市やその郊外部の交通需要を予測や交通計画立案における経路検索データの重要性は一層増すと考えられる。

7.2.2 経路検索データを用いた公共交通の需要変動の予見

本研究では、経路検索データで交通需要が予測できることを示したうえで、さらに、経路検索データが突発的な交通需要の変動の先行指標になることも示した。

働き方や暮らし方の変化を通して人々の行動が多様化している中で、これからの行動分析において求められることは、突発的変動を捉えることだと考える。突発的事象が起きたときには、日常とは異なる行動が発生するが、非日常の行動は稀にしか計測できないため、そのモニタリングは難しいとされてきた。一方、人々は非日常時の状態に置かれたときこそ、一層情報を求める。そのようなときに経路検索システムは使われ、公共交通における非日常時の行動をモニタリングできることが本研究成果より示された。今後、経路検索システムが単なる経路情報だけでなく、検索者の状況にあった付加情報が提供されるようになれば、経路検索データにはこれまで以上に多様な突発的事象に基づく変動が反映されるといえる。そこで、経路検索データから突発的交差需要の発生を予見することで、人々の急な行動変容に対応するための策を講じることや、非日常時に役立つ情報を提供することが可能になると考えられる。

7.2.3 データ間の潜在的関係性の抽出

外出という活動 1 つを取っても、その活動目的は多様化しており、外出するタイミングも活動主体者の心的要因以外に外出前に得る情報によって変化する。本研究では、公共交通における移動に焦点を絞り、外出前あるいは外出中に利用する経路検索システムに蓄積される経路検索データが交通需要の変動の先行指標になる可能性に着目し、交通需要予測モデルを構築した。経路検索データはあくまで Web 上に表明される潜在需要であり、直接的には顕在需要と結びつけられるものではなかった。しかし、Web 上と実空間上で異なる空間で発生する需要であっても、データの背後に潜む関係性を抽出することで、両者を関連づけられる。本研究の分析における核心部は、経路検索データと交通系 IC カードデータそれぞれについて、時系列データが表す交通需要の種類を整理し、変動のタイプによってデータに内在する需要を分解した上で、突発的需要だけを関連づける点にある。データを成分分解して、関連のある部分を抽出するという考え方は、ビッグデータの活用可能性を広げることに繋がるといえる。観光の計画段階から実行までを例に考える。1.1.5 項で整理した公共交通の交通需要予測上の AISAS に基づくと、

- (1) Attention : 観光施設を認知する
- (2) Interest : 観光施設に興味を持ち、行動意図が形成される
- (3) Search : 観光施設の情報や、観光施設までの経路を検索する
- (4) Action : 選択した経路および手段で移動する
- (5) Share : 観光施設での体験を SNS で共有する (ただし、必要不可欠な段階ではない)

となり、観光行動であれば、移動前に訪れたい観光施設の情報や観光施設までの経路を事前に検索する。そして、人々の観光施設を訪れたいという欲求は観光施設の Web サイトのアクセスログ数や経路検索データに反映されると考えられる。そこで、観光施設の Web サイトのアクセスログに反映される興味・関心度合いと経路検索データに反映される潜在需要との関係を分析し、さらに、それらと実空間上の顕在需要を結びつけることによって、観光を移動目的とした公共交通利用の突発的需要を事前に予測できる可能性がある。Web サイトのアクセスログデータはサイト自体の改善点の発見や利益の向上に向けた示唆を得るために商業分野で活用されることが多いが、アクセス数を興味・関心度合いと捉え、需要変動の先行指標として活用できれば、交通分野においても有用なデータになると考えられる。今後、実証分析は必要であるが、複数のビッグデータ間の関係を記述する場合でも、本研究の考え方を応用し、2 変量の分析を多変量へと拡張することで、あらゆる分野のビッグデータの活用可能性を広げ、交通需要予測に適用できるといえる。

7.2.4 地方部のビッグデータの活用可能性

本研究では、日平均検索件数が 535 件程度の必ずしもサンプル数が多いとはいえない地方都市のビッグデータを用いて、データ間の関係性を解明することに成功した。サンプル数が少ないために、ビッグデータ解析という俎上に乗り切れていない地方都市のビッグデータを用いてその活用可能性を示すことができたことは本研究の学術面における成果だと考える。

都市部に比べ、データの種類やサンプル数は十分でないとはいえ、地方都市においてもいくつかビッグデータは存在する。しかし、先に述べたサンプル数が少ないことや、サンプルの地理的分布に偏在性があるために、母集団を捉えることができず、十分に活用できない場合がある。そこで、本提案モデルのように、データの変動性によってデータが表す需要を成分分解し、共通要因を持つ部分に着目して関係性を分析すれば、サンプル数の小さい地方都市のビッグデータでも現象解析や予測に耐えうる可能性がある。

一方で、データの変動を分析する場合には、サンプル数が少ないと、微量な誤差としての変動か、事象の変化を表す重要な変動かが判断しにくい。本研究では、どの程度のサンプル数であれば、現象解析や予測に耐えうるのかについては明らかにできておらず、今後議論を深める必要がある。本研究では路線ごとのモデルを構築し分析した。しかし、例えば、乗降車駅の組合せごとで予測した場合に、中心部の駅間にはでは交通需要を予測できるが、郊外部の駅間は検索がほとんどなされないため、予測できないということがわかれば、予測に耐えうるサンプル数とともに予測できる場所の特徴を明らかにできる可能性がある。提案手法を、乗降駅単位、あるいは乗降車駅の組合せ単位と、分析単位を小さくしていくことで、予測に耐えうるサンプル数が明らかになるはずである。

7.3 今後の展望

本研究では、Web上に表明される移動意向と実空間上に顕在化する交通需要の関係性に着目し、公共交通における需要予測モデルを提案した。従前の交通需要予測では、実空間上の需要を分析してきたが、今後は本研究のようにデジタル空間上に表明される需要と実空間上の顕在需要の関係性を明らかにすることの重要性が増すと考えられる。

ICTの発展によって、自宅にいてもスマートフォンさえあれば買い物（ネットショッピング）や、自宅から離れた人ともビデオ通話やテレビ会議ができるようになり、「情報」を通して、外出せずともモノや人と繋がれるようになった。また、食べ物の配送サービスが充実し、外食のために出かける必要もない。VR技術による仮想空間で現地にいかなくともライブや芸術鑑賞、アウトドアレジャーといったさまざまな体験を楽しむこともできる。このように、実空間上からデジタル空間上へも行動の選択肢が広がったことで、人々は自身の欲求を満たすために、実空間とデジタル空間を使い分けできるようになったといえる。

オンライン会議、宅配、仮想空間上での体験などが今後さらに社会に浸透すると考えると、従前の実空間上だけの移動に着目した交通需要予測手法では、将来の需要を予測することはできない。そこで、個人が自身の欲求を満たすためにデジタル空間と実空間を使い分けしていることを踏まえて、全体のニーズを把握した上で、デジタルと実空間でニーズを配分するような考え方が必要となるといえる。

現在、公共交通利用者数、自動車交通量などの交通データだけでなく、施設利用者数や商品売上など施設利用に関するデータ、宅配やデリバリーの注文件数や商品売上に関するデータ、Web注文数に関するデータなど実空間とデジタル上でのあらゆるデータはビッグデータとして蓄積されている。今後

は、これらデータを組み合わせ、各データによって観測される需要を統合し、交通需要だけでなく、人々の一連の行動に関わるさまざまな需要を把握していきたい。ここで、交通手段、宅配業者、オンラインショッピングサイトなどサービス提供者は無数に存在する。そのため、すべてのデータを収集することはできず、分析者が収集できる極一部のデータだけで、実空間とデジタル空間の関係性を解明しなければならない。このようなサービスの利用者像が不明確で母集団代表性があるとはいえないようなデータの組合せにおいても、本研究で提案した非負値テンソル因子分解や Adaptive Tobit 分位点回帰モデルによるデータの特性分析を行った上で、2 変量状態空間モデルと重み付き回帰モデルの 2 段階分析で関係性を分析するという考え方は適用できると考えられる。また、2 変量分析を多変量分析へと発展させることで、デジタル空間と実行動、複数の行動間の関係性の解明に、本研究を展開していきたいと考える。しかし、個人レベルの行動を分析する場合には、集計レベルの需要に着目した本提案モデルは適用できない可能性がある。これは本研究でも確認されたように、シンプソンのパラドックスの発生や生態学的誤謬に起因するといえる。個人レベルでは経路検索データでは予見できないような要因が存在する可能性は高い。一方で、分析単位を個人レベルにしても、突発的行動は発生すると考えられる。そこで、突発的行動をデータ間および行動と需要を関連づけるキーワードとして、シンプソンのパラドックスや生態学的誤謬を回避できる方法論を構築し、個人レベルの行動と集計レベルの需要を同時に予測できるモデルへと提案モデルを展開する必要がある。そして、都市全体の人々の行動を包括的に予測できるような方法論を開発し、ニーズに即した都市計画、持続可能な地域づくり、地方都市の活性化に寄与する施策を検討していきたい。

本研究では、外出前に検索を行う点に注目し、経路検索数を公共交通利用者数の変動の先行指標と考えモデル化を行ったが、公共交通利用に限らず、自動車による外出、あるいは買い物、外食、通院、レジャーなどの移動を発生させる要因についても、その予兆となる Web 上での事前検索が発生していると考えられる。2020 年以降、COVID-19 によって、日常的な行動が失われてしまったが、COVID-19 収束後の生活に期待する人も多く、現在制限されている行動が完全に解禁されれば、それまで耐え忍んでいた欲求を満たすために、広範囲で多様な人々が活動目的を達成するために移動するといえる。また、COVID-19 の感染拡大以前から、情報収集 1 つをとってもその手段は多様化しており、その中から人々は自分に適した手段を取捨選択し、個々の欲求を満たすために動いている。その個々の欲求を満たすために、社会は大域的なサービス提供よりも、多種多様な限定的なサービス提供にシフトしつつある。したがって、COVID-19 による制限期間が明け、特定の人に向けた限定的なサービスが解禁されれば、多様な人々の行動が変動し、突発的需要はこれまで以上に激しいものとして発生するに違い

ない。また、突発的需要が発生する要因とタイミングの違いから、突発的需要にも予測日の直前に収集した経路検索数で予測できるものと半日前に収集した経路検索数で予測できるものなどいくつか種類が存在するはずである。さらに、発生要因によっては経路検索数に織り込まれていない需要が存在し、経路検索データですべての突発的需要を予測できるとは限らない。そのため、今後は、さまざまな行動やニーズに対して先行指標となりうる SP データを検討し、突発的需要についても要因とタイミングに基づく成分分解が必要と考える。

謝辞

本研究の遂行に際して、多くの方々からご指導とご協力を賜りました。以下に感謝の意を表します。

鳥取大学大学院工学研究科 桑野将司 教授には、本研究のすべての研究過程においてご指導賜りました。思い起こせば、2016年、今から6年前の春、都市計画研究室に配属され、桑野先生に卒業論文のご指導を頂く機会を得たところから、私の人生は大きく変わりました。学部4年生当時は研究というものを理解していませんでしたが、次第に研究の楽しさを知り、研究成果が得られたとき大きな達成感に包まれました。この経験ができたのは、私が越えられる課題を見極め、常に温かく見守り、立ち止まったときには的確なアドバイスをしてくださった桑野先生のおかげです。修士課程、博士課程へと進んでからは、日々の研究や論文執筆でのご指導はもちろんのこと、国内外の学会発表、研究会やセミナーなど数多くの経験の機会を与えてくださいました。殊に、FIT2017での受賞経験は桑野先生のご指導なくして有り得ませんでした。自身が取り組んだ研究が客観的評価を得たときの喜びは今でも忘れません。この受賞をきっかけに、博士課程への進学を強くすすめてくださり、大学で研究する機会を与えてくださいました。その後は、研究遂行のための思考力やテクニックのほか、学生指導に至るまですべてをたたき込んでいただきました。発表練習や論文執筆では、毎度余裕がない状態で教を乞う私ですが、決して見放さず、時に苛立ちながらも迅速かつ丁寧に指導してくださいました。学部生時代より、一貫して直接ご指導頂き、桑野先生のもとで学ばせていただけたことすべてが大きな財産です。言葉では言い表せられないほど、深く、深く感謝しております。この6年間、私への評価はさほど口にしない桑野先生でしたが、「細江を博士課程に呼んでよかった」と一度だけ言ってくださったときがあります。その理由はわからずじまいですが、平凡な自分に落ち込み、不安だった心が救われました。今後は、桑野先生からの数々の教を胸に、日々の研究、教育に邁進し、いつか「一人前の研究者になったね」とお褒めの言葉をいただけるよう、まだ見ぬ春を心待ちに、精進して参ります。

鳥取大学大学院工学研究科 森山卓 助教には、ゼミの時間に限らず、幾度も本研究について議論し、ご助言を頂きました。時には、夜遅くまでお付き合いいただき、モデルの推定結果の数学的解釈についてご教授いただくこともありました。森山先生から頂いたテクニック面でのアドバイスのおかげで、困難を切り開く糸口を掴め、本研究を遂行させることができ、感謝の念に堪えません。また、精神面でも気にかけてくださり、落ち込んでいる時には励ましのお言葉をかけてくださいました。森山先生は年齢が近く、最も身近な研究者であり、目標とする研究者の1人です。森山先生の研究者、そして教育者としての在り方を間近で感じながら、今後も研鑽を積んでいきたいと思っております。

鳥取大学大学院工学研究科 福山敬 教授、谷本圭志 教授、さらに広島大学大学院先進理工系科学研究科 塚井誠人 准教授には、本論文の審査員をお引き受けいただき、研究内容から論文内容まで核心的なご意見と丁寧なご指導を賜りました。福山先生には、学部4年生時の研究室配属からお世話になり、情報システム研究室に配属が変わってからも、合同ゼミや研究会でアドバイスを頂きました。昨

年、都市計画研究室の助教として着任してからは、博士論文に集中できるようにサポートいただき、ご負担を増やしてしまうことがあったと思いますが、いつも温かいお言葉をかけてくださいました。ここに厚くお礼を申し上げます。谷本先生には、6年間、幾度のタイミングでの確なご助言と鋭いご意見をいただき、大変お世話になりました。深く感謝しております。塚井先生には、公聴会後も本研究について直接ご指導いただく機会を頂戴し、大変有益な示唆をいただき、深謝いたします。

香川高等専門学校 建設環境工学科 宮崎耕輔 教授、東京大学大学院情報理工学系研究科 伊藤昌毅 准教授には、本研究の遂行にあたり、ことごと NAVITIME のデータを入力、ご提供いただき、心より感謝申し上げます。

岐阜大学工学部社会基盤工学科 倉内文孝 教授、豊橋技術科学大学建築・都市システム学系 杉木直 准教授、名城大学理工学部社会基盤デザイン工学科 鈴木温 教授には、研究会にて本研究の遂行のためのご助言とご意見を賜りました。自身では気づけない視点から、研究を発展させるヒントとなる情報を与えていただき、厚くお礼を申し上げます。

鳥取大学大学院工学研究科 菅原一孔 教授には、BUSNET のデータをご提供いただき、本研究に関わる重要な研究を共著論文として執筆する機会を幾度も頂きました。ここに深く感謝の意を表します。

鳥取大学大学院工学研究科 石井晃 教授には、学部4年生のときから大変お世話になりました。構内でお会いした時は、本研究の進捗具合や私の進路について気にかけてくださり、アドバイスをいただいたこと、心から感謝申し上げます。昨年、突然の訃報に接し、ただただ悲しい気持ちでいっぱいです。在りし日のお姿を偲び、謹んでご冥福を申し上げます。

鳥取大学大学院工学研究科 南野友香 准教授、長曾我部まどか 助教、同大学技術部 岩田千加良さんには、本研究を遂行するうえで、精神的に支えていただきました。論文の提出、予備審査、公聴会などの際には励ましのお言葉をいただき、とても嬉しかったです。ここに厚く感謝の意を表します。

桑野研究室のOB・OGのみなさん、在籍学生のみなさんには、大変お世話になりました。みなさんのおかげで、楽しい研究室生活を送ることができ、大切な思い出となっています。また、本論文に関わる研究の遂行に際しては、吉永隆太さん、古川ゆりさん、古江克成さん、前田真作さん、中井健太郎さんに、分析のご協力を得ました。特に、前田さんと中井さんには、未熟な指導をお詫びすると共に、辛抱強く研究にお付き合いいただいたこと、心から感謝しております。

最後に、これまでの大学生活の間、私を温かく見守り、精神的に強く長く支え続けてくれた母と友人に、心からの感謝を捧げて結びとします。

2022年1月 細江 美欧