

# スマートカードの利用履歴データを用いた 公共交通利用者の選好特性分析

宮崎 一貴<sup>1</sup>・溝上 章志<sup>2</sup>・八戸 龍馬<sup>3</sup>

<sup>1</sup> 学生会員 熊本大学 大学院自然科学教育部 (〒860-8555 熊本県熊本市中央区黒髪2丁目39番1号)

E-mail: 203d2218@st.kumamoto-u.ac.jp

<sup>2</sup> 正会員 熊本大学 大学院先端科学研究部 (〒860-8555 熊本県熊本市中央区黒髪2丁目39番1号)

E-mail: smizo@gpo.kumamoto-u.ac.jp

<sup>3</sup> 学生会員 熊本大学 大学院自然科学教育部 (〒860-8555 熊本県熊本市中央区黒髪2丁目39番1号)

E-mail: 194d8361@st.kumamoto-u.ac.jp

近年、熊本市内の公共交通利用者数が減少傾向にある。スマートカードが熊本市内の公共交通機関に導入されている。スマートカードの利用履歴データには、乗車日時や停留所名などの多く情報が記録されている。スマートカードの利用履歴データを用いた研究は多く行われてはいるが、個人を対象にした長期的な分析は行われていない。そこで本研究では、状態空間モデルを用いて熊本市電の利用者の時間的な選好特性について分析を行なった。得られた状態推定値を用いてクラスター分析を行い、いくつかのグループに分かれることが分かった。最後に、各グループの乗車回数を増やす方法について検討した。

**Key Words :** *smart card data, Kumamoto city tram, general state space model, particle filter, cluster analysis*

## 1. はじめに

### (1) 公共交通機関の現状とスマートカード

熊本市内において、公共交通利用者数は減少傾向にあり、その中でも、バスの利用者の減少が著しい<sup>1)</sup>。路面電車（以下、熊本市電）は直近10年間において微増傾向にあるが具体的な根拠となる要因については分からない<sup>2)</sup>。利用者数を増加させるには便利で信頼性の高いサービスを行うことが重要である。

近年、スマートカードの導入が全国的に進められている。代表的なものとしてJR東日本が発行しているSuicaがあり、その発行枚数は7,587万枚で年々増加している<sup>3)</sup>。熊本市内の公共交通においても、Suicaやnimocaなどの全国で利用可能なICカードが10種類（以下、全国相互利用交通系ICカード）と「くまもんのIC CARD」と親しまれている熊本地域振興カードが利用可能となっている<sup>4)</sup>。

このスマートカードの利用履歴データを用いた研究が国内外問わず行われている。西内らは、土佐電気鐵道を対象にスマートカードですかの利用履歴データ4ヶ月分を用いて、減便によるサービスの悪化が利用者及び影響について状態空間モデルや生存時間モデルを用いて分析を行った。減便については影響力が小さいこと、電

停OD間の立地条件が影響していることを示した<sup>5)</sup>。西内らは、公共交通マーケティング手法を検討するために、スマートカードですかの利用履歴データ1ヶ月分を用いて、カードの種類や定期の有無などが、時間的トリップ依存度や空間的トリップ依存度に影響があるか調べた。低頻度でトリップパターンを持たない利用者が多くを占めていることが示された。また、非定期利用者に対する利用促進政策の重要性について示した<sup>6)</sup>。熊本市電においての既往研究として、森田らは、全国相互利用交通系ICカードの利用履歴データ1年分を用いて利用回数の分析や、nimocaやSuicaなどのカードの販売場所に違いがあるか調べた。利用回数に違いがあるものの、カードごとに分けるのは難しい。また、定期券購入者の利用回数は増加していることや、クラスター分析を用いて、クラスターごとに異なることを示した。ICカードの信頼性についても示した<sup>7)</sup>。吉塚は、熊本市電の2014年度から2017年度の利用履歴データより、利用者数を観測モデルとした集計型の線形ガウス型状態空間モデルを構築し、カルマンフィルターを用いてトレンドやサービスについての影響人数について状態推定とパラメータ推定を行った。また、1年間の利用日数が100日を超える利用者3人を対象に、非線形ガウス型状態空間モデルを構築し、粒子フ

フィルタを用いて状態推定を行った。状態変数の値が異なることや、利用者の異質性について示した<sup>8)</sup>。

## (2) 研究目的と論文の構成

これらの研究はいずれも需要者のような集計値を利用することや、個人を対象にしたとしても任意の短期間の分析を行っているだけで、1年以上を超える長期的な分析は行われていない。また、CRMやOne to Oneマーケティングと呼ばれる一人一人や小規模なグループを対象とし、それぞれの行動や嗜好などの情報から最適なアプローチを行うということが注目されている<sup>9)</sup>。

これらを踏まえ、本研究では公共交通機関全体の利用促進を目的と掲げ、熊本市電利用者を対象に個人レベルの非集計的かつ長期的な分析を行う。分析の結果から利用者ごとに特性の時間的変化について明らかにし、今後行われる交通政策や交通マーケティングに繋げるための方法について探っていく。

本論文は6章から構成されている。2章で分析に用いたデータの概要について説明をする。3章では、分析に用いるモデルの検討や導入したモデルの概要について説明する。4章では、状態空間モデルの推定結果についてまとめる。5章では、利用促進のためのマイクロマーケティング分析として、状態空間モデルで求めた状態変数を用いてクラスター分析をする。最後に6章にて本論の結論と今後取り組むべき課題について述べる。

## 2. 分析対象データの概要

本研究で使用するデータは熊本市交通局から提供された全国相互利用交通系ICカードの利用履歴データである。サンプルは、2014年4月1日から2018年9月30日までの1,644日（54か月）分、観測IDは1,187,464分である。本研究では、観測IDと利用日時、定期の種別のデータを用いて分析を行った。利用者が複数持っている可能性もあるが、本研究においては、1つのIDについて1人の利用者が利用しているものとする。対象期間における観測IDの時間的変化については図1に示す。1回の乗車・降車について1カウントとし、時間的推移を示している。往復で利用した際にはそれぞれカウントしている。また、同時期の熊本市電の沿革について表1に示す。導入当初から利用されており、IC定期券運用開始から急速に利用が伸びていることが分かる。熊本地震後や正月、9月の連休あたりに利用者が減少する傾向が見られる。

本来であれば、全データを用いて分析を行うべきであるが、分析の対象となるデータは1,187,464人×1,644日と多くあるため、全国相互利用交通系ICカードの種類の割合に従って抽出する比例割当法を用いて標本を作成した。

各ICカードの抽出数について表2に示す。

熊本市電では、ICカードを利用した定期が2014年10月1日から発行されており、定期は主に通勤、通学を対象に導入されている。利用履歴データには、1ヶ月、3ヶ月、通学の3種類が記録される。乗車日の定期の種別については判別することができるが、乗車しなかった日についてはその人が定期利用者なのか非定期利用者なのかの判別ができない。そのため今回は以下の定期利用を判定する方法を用いて適宜、補間した。

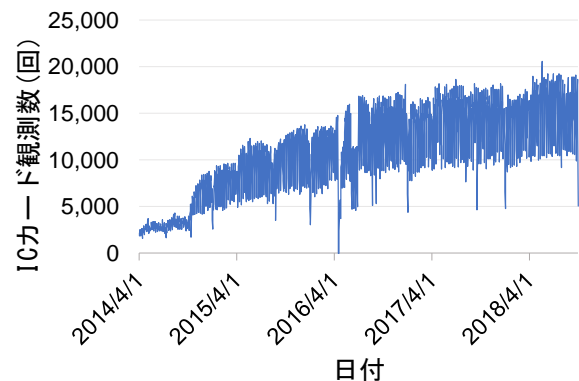


図1 対象期間における観測IDの推移

表1 熊本市交通局の沿革

時期	内容
2014/3/28	でんでんnimoca(全国相互利用交通系ICカード)
2014/10/1	IC定期券運用開始
2015/8/7	熊本地域振興ICカードの運用開始
2016/2/1	150円から170円に値上げ
(2016/3/13)	バスで全国相互利用交通系ICカードが運用開始
2016/4/14	熊本地震（前震）
2016/4/16-19	熊本地震の影響より運行取りやめ（一部を含め）

表2 利用者の概要

カード	全データ数	抽出ID	18日以上
ICOCA	126,962	1,069	27
Kitaca	3,948	33	0
manaca	24,090	202	3
nimoca	166,310	1,401	271
PASMO	206,269	1,738	49
PiTaPa	26,422	222	3
SUGOCA	160,416	1,351	200
Suica	428,386	3,609	133
TOICA	6,615	55	0
はやかけん	38,046	320	11
定期利用者		145	130

- Step-0: 定期券が導入される以前は、定期利用者はいないこと、定期は平日利用すると仮定する。
- Step-1: 定期の種別が記録されていない日に移動する。
- Step-2: 前日の情報を取得する。
- Step-3: 次の日の値とその日が平日かどうか記録する。
- Step-4: Step-3で値が観測された場合はStep-6へ、観測されなかった場合は更に次の日の値と平日かどうかを記録し、平日が5日観測されるまで繰り返す。
- Step-5: 平日が5日観測されたとしても値が観測されなかった場合は非定期としStep-6へ。
- Step-6: 前日と観測された値が同じか否かを判定する。
- Step-7: 同じである場合はその値を、異なる場合はいずれかをランダムに採用する。
- Step-8: 上記を最終日まで繰り返す。

その結果定期利用者数は145人となった。今回の研究では、公共交通の利用促進を目的としていることから、利用頻度が低い利用者にターゲットを置くのは適切とは言えない。そこで、3ヶ月に1回以上（合計18日以上）利用している人を対象に分析を行うこととする。その結果、1万人から697人へ減少した。また、定期利用者も145人から130人になった。内訳の詳細については表2に示す。熊本県内で購入可能な株式会社ニモカのnimocaやJR九州のSUGOCAの利用する頻度が高い人が多い。JR北海道のKitacaやJR東海のTOICAについては、18日以上利用の条件に該当する人がいなかった。

### 3. 選好特性の分析モデル

#### (1) 適応モデルの検討

時間的变化について着目した分析として時系列分析がある。時系列分析は目的変数の変化に対する分析であり要因について議論するのが難しい<sup>10)</sup>。乗車の有無に着目するモデルとして、一般化混合線形モデルなどが考えられるが、時間的变化について組み込むことが難しい<sup>11)</sup>。どちらも特性の時間的变化と乗車の有無の因果関係を同時に考慮することが出来ない。吉塚は個人レベルの分析を行うために状態空間モデルを採用し、状態推定を行っている<sup>8)</sup>。状態空間モデルを採用し、状態推定を行い、推定された特性についての時間的变化について分析する。

#### (2) 一般化状態空間モデル

##### a) モデルの概要

萩原らを参考にモデル概要について説明する。状態モデルと観測モデルを組み合わせたモデルである。状態ベクトル $x$ は観測されない値から成り立つ。観測ベクトル $y$ は観測された値から成り立つ。状態モデルは一次のマルコフ過程が成立するとして式(1)のように定式化され

る。観測モデルは状態モデルから得られる状態ベクトルからしか影響を受けないと仮定されたモデルで式(2)のように定式化される。時間変化については、 $t$ と表され式(3)のように定義する。

$$p(x_t | x_{0:t-1}, y_{1:t-1}) = p(x_t | x_{t-1}) \quad (1)$$

$$p(y_t | x_{0:t-1}, y_{1:t-1}) = p(y_t | x_t) \quad (2)$$

$$t = (1, 2, 3, \dots, T) \quad (3)$$

初期値は状態ベクトル $x_0$ に導入し、分析者が設定する任意の確率変数に従うものと仮定し、式(4)に与える<sup>12)</sup>。

$$x_0 = p(x_0) \quad (4)$$

#### b) 推定方法

萩原ら、佐藤ら、北川を参考に状態空間モデルの推定方法について説明する。模式図を図4に示す。

Step-0 (初期分布) :  $t$ 時点が初日 ( $t = 0$ ) である場合、初期値となる任意の初期分布 $x_0$ を与える。

Step-1 (一期先予測分布) : 状態ベクトル $x_{t-1}$ と状態モデルを用いて $t$ 地点の予測分布 $p(x_t | y_{1:t-1})$ を求める。

Step-2 (一期先予測尤度分布) :  $t$ 時点の状態ベクトルを観測モデルの尤度関数や対数尤度関数に当てはめ、尤度または対数尤度を求める。

Step-3 (リサンプリング) : 尤度などから、 $t$ 地点において最適な状態ベクトル $x_t$ に更新し $t$ 時点の分布形とする。得られた分布形はフィルター分布 $p(x_t | y_{1:t})$ と呼ばれる。

Step-4 (平滑化) : 未来の値が観測された時未来の値を用いて $t$ 地点の値 $x_t$ を更新する。得られた分布は平滑化分布 $p(x_t | y_{1:T})$ と呼ばれる<sup>9),12),13)</sup>。

#### (3) 採用した状態空間モデル

##### a) 状態モデルと観測モデル

状態モデルは一次のマルコフ過程である線形モデルと

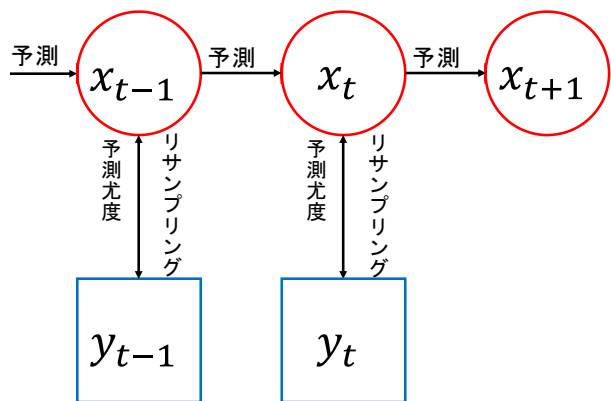


図2 モデル構造模式図

して式(5)のように表現する. 状態ベクトル $x$ は後述する個人 $i$ が熊本市電を利用する確率を予測する変数であり, 選択モデルの効用を構成する変数にかかる未知パラメータベクトルである.

$$x_{i,t} = Fx_{i,t-1} + Gv \quad (5)$$

$i$ は推定を行う利用者個人を示しており,  $F$ ,  $G$ は時変回帰係数,  $v$ はホワイトノイズを示している.

観測モデルは式(6), (7)の非集計二項選択ロジットモデルで表現する. 状態モデル式(5)で定義された状態ベクトル $x_t$ は式(8)の効用関数 $u$ の未知パラメータになる.

$$y_{i,t} = \begin{cases} 0, & p_{i,t} < \frac{1}{2} \\ 1, & p_{i,t} \geq \frac{1}{2} \end{cases} \quad (6)$$

$$p_{i,t} = \left( \frac{e^{u_{i,t}}}{1 + e^{u_{i,t}}} \right)^{y_{i,t}} \left( \frac{1}{1 + e^{u_{i,t}}} \right)^{(1-y_{i,t})} \quad (7)$$

$$u_{i,t} = H_{i,t}x_{i,t} + \omega \quad (8)$$

$H$ はデザイン行列,  $\omega$ はホワイトノイズである. ロジットモデルの性質上, ホワイトノイズ $\omega$ はガンベル分布に従うため式(9)の極値分布の位置パラメータ $a$ , 尺度パラメータ $b$ で表現する.

$$\omega = \frac{1}{b} e^{\left( \frac{a-x}{b} - e^{\left( \frac{a-x}{b} \right)} \right)} \quad (9)$$

リサンプリングの際は, 観測モデル用いて尤度関数を構築し, 尤度 $\theta$ が高いものを多く採用する.

## b) 状態ベクトル

状態ベクトル $x$ には以下の変数を導入した.

トレンドについては, 未知の時間変化を表す変数である.  $t-1$ 時点と $t$ 地点において, ならだかに変動すると仮定し,  $t-1$ 地点と $t$ 地点の差は0に近似することを期待している. 乖離部分については, 平均0, 分散 $\tau^2$ の正規分布に従うものとする. 式(10)に示す.

$$(\text{trend})_{i,t} = (\text{trend})_{i,t-1} + v(\text{trend}) \quad (10)$$

曜日トレンドについては識別性の問題が発生する可能性があるため, 曜日ごとではなく, 今日, 1日前, 2日前, ..., 6日前のように表すことで曜日を表現する. ダミー変数として考慮するため, 6つのダミー変数を用いる. 成分については, 1週間のトレンド成分の和は任意の値でほとんど一定であると仮定する. 任意の値は0とした.

乖離部分はトレンド成分と同様に平均0, 分散 $\tau^2$ とし式(11)に示す. また,  $t-1$ 地点の1日前の値は $t$ 地点における2日前の値である. これらの関係を式(12)~(17)に示す. なお, 右辺に関しては $t-1$ 時点の変数である.

$$\begin{aligned} (\text{week})_{i,t} = & -(\text{week})_{i,t} - (\text{week})_{i,t-1} - (\text{week})_{i,t-2} \\ & - (\text{week})_{i,t-3} - (\text{week})_{i,t-4} \\ & - (\text{week})_{i,t-5} + v(\text{week}) \end{aligned} \quad (11)$$

$$(\text{week})_{i,t-1} = (\text{week})_{i,t-2} \quad (12)$$

$$(\text{week})_{i,t-2} = (\text{week})_{i,t-3} \quad (13)$$

$$(\text{week})_{i,t-3} = (\text{week})_{i,t-4} \quad (14)$$

$$(\text{week})_{i,t-4} = (\text{week})_{i,t-5} \quad (15)$$

$$(\text{week})_{i,t-5} = (\text{week})_{i,t-6} \quad (16)$$

$$(\text{week})_{i,t-6} = (\text{week})_{i,t-7} \quad (17)$$

雨など降雨, 1日平均気温, 祝日・盆・正月, 運賃, 定期などについては, 影響度は未知であるため, これらの影響に関するパラメータを状態変数として状態ベクトルに導入する. ティレンドと同様乖離部分は平均0, 分散 $\tau^2$ の正規分布に従うとし, 式(18)~(22)に示す.

$$(\text{rain})_{i,t} = (\text{rain})_{i,t-1} + v(\text{rain}) \quad (18)$$

$$(\text{temp})_{i,t} = (\text{temp})_{i,t-1} + v(\text{temp}) \quad (19)$$

$$(\text{holiday})_{i,t} = (\text{holiday})_{i,t-1} + v(\text{holiday}) \quad (20)$$

$$(\text{fare})_{i,t} = (\text{fare})_{i,t-1} + v(\text{fare}) \quad (21)$$

$$(\text{regular})_{i,t} = (\text{regular})_{i,t-1} + v(\text{regular}) \quad (22)$$

以上の状態変数の定義をベクトル表示にした状態ベクトル $x$ を式(23)に示す.

$$x_{i,t} = \begin{pmatrix} (\text{trend})_{i,t} \\ (\text{week})_{i,t} \\ (\text{week})_{i,t-1} \\ (\text{week})_{i,t-2} \\ (\text{week})_{i,t-3} \\ (\text{week})_{i,t-4} \\ (\text{week})_{i,t-5} \\ (\text{rain})_{i,t} \\ (\text{temp})_{i,t} \\ (\text{holiday})_{i,t} \\ (\text{fare})_{i,t} \\ (\text{regular})_{i,t} \end{pmatrix} \quad (23)$$

## c) デザイン行列

観測モデルの線形効用関数を構成するデザイン行列について説明する. ティレンド, 曜日トレンドに関して対応する観測値は存在しないため係数を1とした. 曜日トレンドの前日などの前の日に関しては観測モデルに直接影響を与えないものとした. 雨など降雨データについては,



気象庁のホームページから1時間雨量について取得した<sup>17)</sup>。6時から18時の朝から夕方までの12時間の間に1時間雨量が $3\text{mm}$ 以上観測された日を1とするダミー変数を導入する。対象期間内に観測された日数は214日であった。一日平均気温については、気象庁のホームページから対象期間の1日の平均気温を取得した。小数点1桁までを数量データとして観測値に導入する<sup>14)</sup>。祝日・盆・正月については祝日などのダミー変数を導入した。祝日は、国民の祝日を、盆は8月13日から16日の4日間とした。正月は、12月29日から1月3日の6日間とした。運賃については、熊本市電は原則、均一運賃である。観測値として150円、170円の数値データを導入する。定期については、定期期間内を1、それ以外を0としたダミー変数で表現した。デザイン行列を式(24)に示す。

$$H_t = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & rain & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & temp & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & fare \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & regular \end{pmatrix} \quad (24)$$

なお、rain, temp, fare, regularには観測値が入る。

#### d) 観測ベクトル

熊本市電は6時頃から翌日1時頃まで運行している。乗車の判定は利用履歴データを用いて、IDが観測された日は1を、IDが観測されなかった日は0のダミー変数を観測ベクトル $y_t$ に導入した。降車時の日付より判定した。

### (3) 初期分布の設定と状態の推定方法

初期分布については、任意の分布に従うと仮定する。中西らも示しているように、初期値によって推定結果が異なることがある<sup>19)</sup>。一般的に状態ベクトルの分布が既知である場合はその分布形を、未知であれば正規分布または、無情報事前分布として $-\infty$ から $\infty$ の一様分布、分散 $\infty$ の正規分布を仮定する。今回は、分散100の正規分布を仮定したもので推定を行う。

非線形状態空間モデルを求める手法とし逐次的にベイズ推定を行う粒子フィルターがある。粒子フィルターを用いて状態推定を行うため、ホワイトノイズの分布を事前に与えなければならない。正規分布は平均0, 分散1の標準正規分布と仮定した。効用関数の誤差項である極値分布については、位置パラメータ0, 尺度パラメータ1の

標準正規分布に近似するガンベル分布とした。中西らはこれらの仮定が状態推定や予測精度に大きな影響があると示しているが、正確な値を求めるには膨大な時間がかかる<sup>15)</sup>。正確な値よりも、状態変数の個人ごと傾向や変容に着目することが重要と考えたからだ。

#### a) 粒子フィルター

多くの状態変数を発生させて最適な状態ベクトルの分布形を求める。この値を粒子と見立て、粒子フィルターと呼ばれる。今回のモデルでは、12次元空間に10,000個の粒子を発生させた。発生させた粒子から、それぞれの尤度を求める。計算の途中で、効用の値が極端な値となり、計算不能となることがあるため、上限と下限(-50,50)を設定した。その間で尤度の高いものを多く採用するように、系統リサンプリングにより分布と重み付き平均の値を推定した。重み平均の値を状態変数の代表値として扱う。系統リサンプリングは、正規化を行った尤度の累積和を用いて、等間隔に取り出す手法である。

#### b) 平滑化

未来の値を用いて過去の時点を補正する手法として平滑化がある。粒子フィルターを用いた際の平滑化手法としては固定ラグ平滑化という手法がある。この手法は $t$ 時点から $n$ 日先の未来の観測値が得られた際に $t+1$ 日から $t+n$ 日間の状態ベクトルのフィルター分布を用いて、 $t$ 時点の状態ベクトルに補正を行う。未来の値 $n$ は全期間ではなく予め日数を決めておく。例えば、3日先のデータを使って補正すると決めていた場合は、4日先のデータが取得された際は補正を行わずほぼ同じであると仮定する。一般的に固定ラグ平滑化に用いる日数は長すぎない方がいいとされている<sup>9)</sup>。今回は1ヶ月先のデータまで取得すると考え、30日とした。計算手法は、北川アルゴリズムを採用した<sup>9)(12)(13)</sup>。平滑化の一連の流れを模式図とし図3に示す。緑の点が、リサンプリングで得られたフィルター分布であり、青線は、粒子の親子の関係を示している。黄色の分布が平滑化分布を示す。

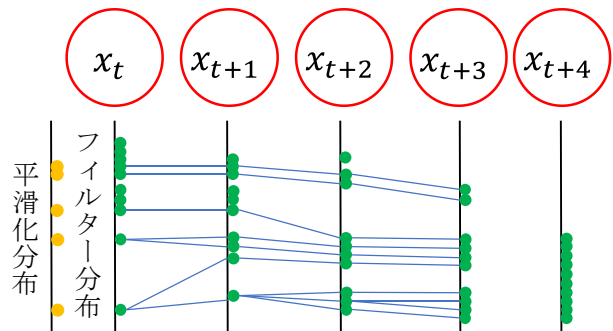


図3 平滑化模式図

#### 4. 状態空間モデルの推定結果

状態空間モデルから得られた平滑化分布の状態変数の値で比較を行った。本来であれば全利用者の推定結果について比較的検討を行うべきであるが、ここでは無作為に抽出した10人の利用者の推定結果について示す。

トレンドの状態変数の時間的推移について図4に示す。初期値の影響で値は0付近から広がっていく分布である。そのため、分散を大きくし、粒子数を増やすなどの計算手法の改良の余地があることが分かる。日付が進むごとに大きく効用が下がっている利用者や、横ばいの利用者などいくつかのグループに分けることが可能であると考えられる。

図5に曜日トレンドの分布を示す。曜日トレンドについては、周期性を持った変化であるため、状態変数は小刻みに変化している。利用者ごとに傾向があるように見えるが分かりにくいいため、曜日ごとに分けて比較する。日曜日について図6に示す。曜日ごとに分けるとそれぞれなだらかな変化している。

雨など降雨の変化は、推定開始から雨が観測されるまでの期間は雨のダミー変数は0であるためランダムに推

移している。雨など降雨の状態変数の推移を図7に示す。10人はそれぞれで雨の影響について異なった値を示しており、一人は効用が高くなっていることから、雨が観測されると乗車が大きく見込まれる利用者である可能性があると言える。また、雨が降ると利用しなくなる利用者についても観測出来た。

一日平均気温については、気温1度当たりの状態変数の変化について図8に示す。利用者それぞれの値を示している。利用者は気温の影響については、大きな変動を受けながら横ばいに推移している。しかし、値の変化については、理解が難しい分布になった。

祝日・盆・正月の影響については、この変数も観測されるまでは、初期分布からランダムに推移している。図9に状態変数の推移について示す。推定開始からほぼ横ばいを推移しているが、利用者それぞれある地点から多く増加または、減少を示す。最終的には、祝日・盆・正月に利用が見込まれるグループなど3つのグループに分かれそうな値となった。

運賃の1円当たりの状態変数についてみると、運賃が150または170と大きな値であるため、状態変数の値は小さく出ていることが分かる。推移については図10に示

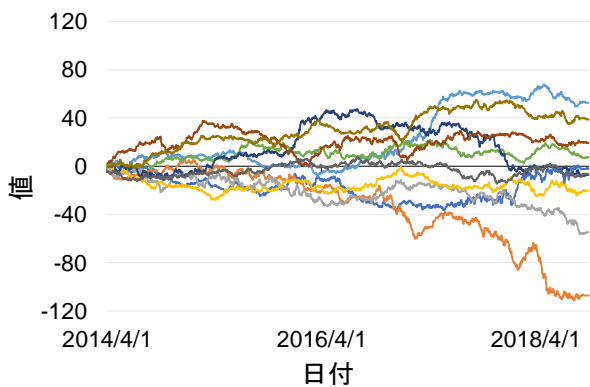


図4 トレンドの状態変数の推移

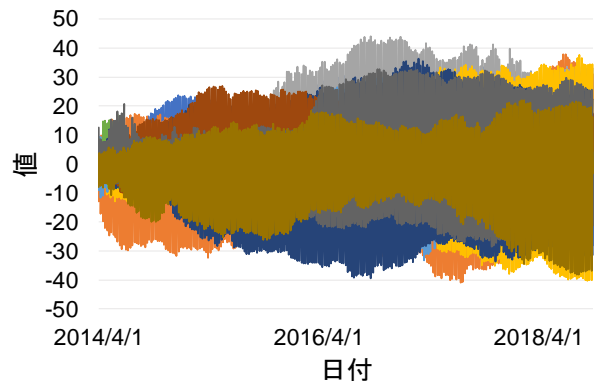


図5 曜日トレンドの状態変数の推移

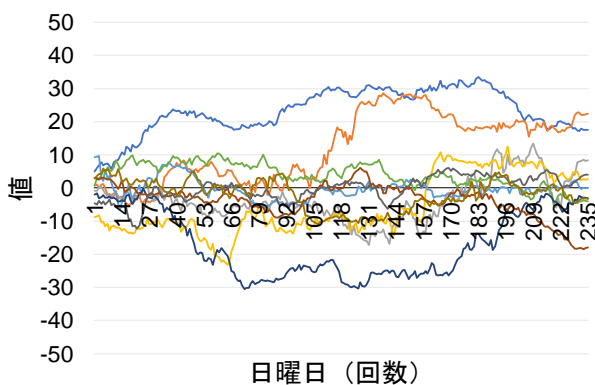


図6 日曜日の状態変数の変化

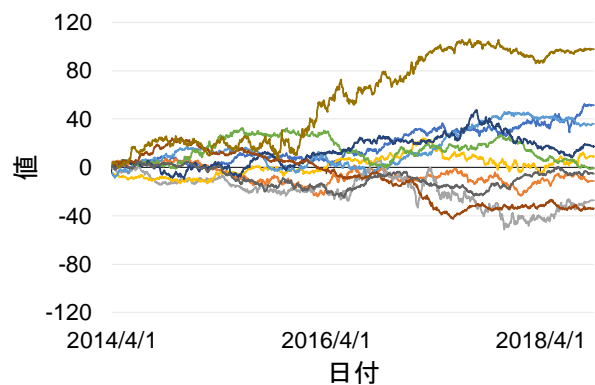


図7 雨など降雨の状態変数の推移

す。運賃が高いことにより乗車しなくなるという負の値を示すことを期待していたが、推定してみると正の値を示す利用者がいることが分かる。推定上、運賃を値上げした際も利用が見込まれることから、価格弾力性小さい利用者と思われる。また、状態変数は1月頃になると極端に値が小さくなることが分かった。運賃の影響については冬よりも夏の方が高いことが分かる。

最後に、定期的状態変数についてみる。状態変数の推移について図11に示す。定期と観測されるまでは、初期値からランダムに進み、効用に与える影響は0である。利用するまで未知数であるが、ランダムに推定された分布が潜在的な状態変数である。

以上より、状態変数の影響について分解できたことや、利用者がそれぞれ異なった値を示す特異性について表現できたと言える。

## 5. 熊本市電利用促進のためのマイクロマーケティング分析

状態空間モデルから得られた状態変数の特徴によって利用者ごとの特性を分析できれば、個人ごとにどのよう

な交通サービスを提供すれば良いか明らかになり、交通政策上有用である。得られた状態変数によって利用者をいくつかの類似するグループに分けて、グループごとにマーケティングを行うことによって利用回数の向上を図る。

### (1) データ

得られたデータは、697人分の時間的に変化する状態変数、効用関数を構成する説明変数にかかるパラメータの推定値である。推定値されたパラメータを時間的に変化する波形と捉えて、動的時間伸縮法（以下、DTW:Dynamic Time Warping）を用いて階層型クラスタ分析を行った。定期的状態変数は1度でも定期利用が確認された130人を対象にした。

### (2) 時間的クラスタ分析

時系列データから傾向や規則性などのパターンを抽出するため、2つの時系列の類似度を比較する手法としてDTWやその改良モデルが多く提案されている。DTWは音声識別や株価など情報分野で利用されている方法であり、2つの波形の値の類似度 $D$ を計算する<sup>16)17)</sup>。

697人の組み合わせを総当たりで行い、類似度 $D$ を算出する。算出された類似度 $D$ を距離行列と捉え、階層型

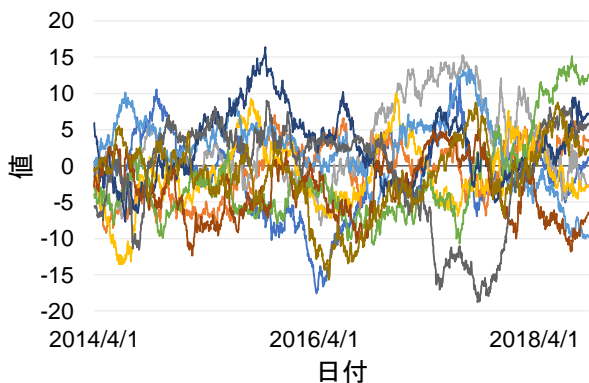


図 8 一日平均気温の状態変数の推移

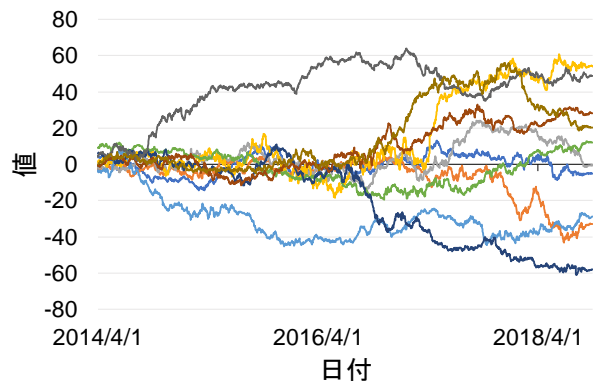


図 9 祝日・盆・正月の状態変数の推移

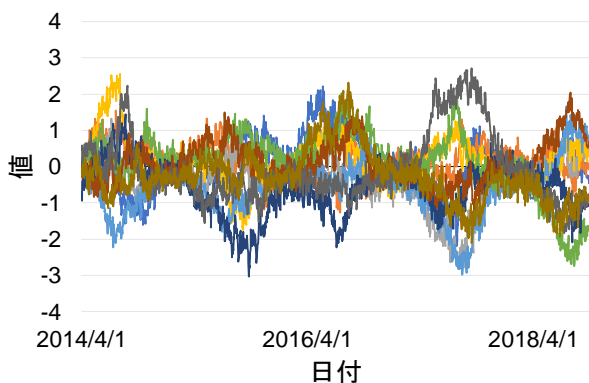


図 10 運賃の状態変数の推移

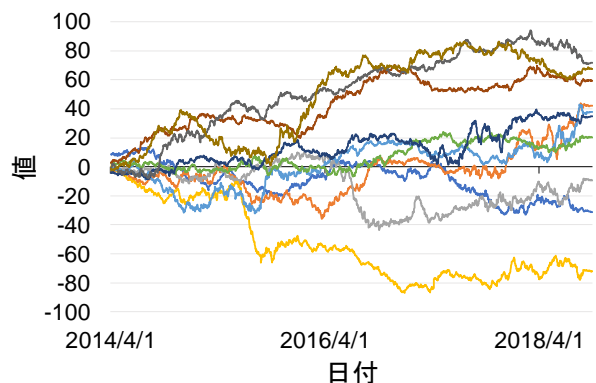


図 11 定期的状態変数の推移

クラスター分析を用いて、クラスタリングを行う。階層型クラスター分析はそれぞれの値から距離に近いもの同士を結合しクラスタリングを行う手法である。今回はウォード法を用いて分析を行った。

得られたデンドログラムから2つ、または3つのタイプに分けた。状態変数が増加するタイプ、横ばいに推移するタイプ、減少するタイプに分類分けができた。各変数について考察する。人数については表3に示す。

時間変化に関する変数については、トレンドの結果は図12、曜日トレンドの結果を図13に示す。95人は増加し、261人は減少することが分かる。曜日トレンドについては、大半の利用者ほぼ一定である。これらの結果から、継続的な利用が見込まれる利用者を明らかにすることが可能である。利用が見込めないタイプの利用者や、ほぼ一定に推移するタイプの利用者何らかの働きかけを行うことで乗車回数を増やすことが可能である。

表3 タイプごとの人数

	トレンド	曜日 トレンド	雨など 降雨	一日平均 気温	祝日・ 盆・正月	運賃	定期
増加傾向	95	103	278	184	372	206	54
一定	341	487	246	370		267	47
減少傾向	261	107	173	143	325	224	29

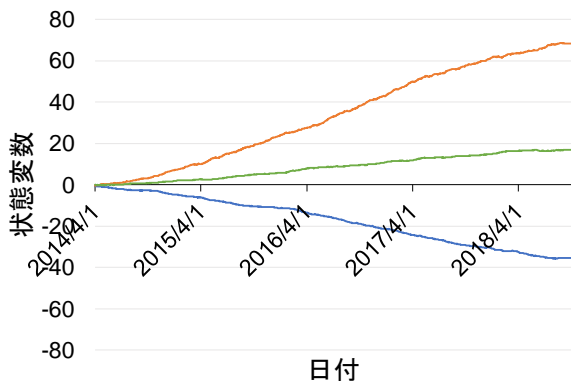


図12 トレンドのタイプ

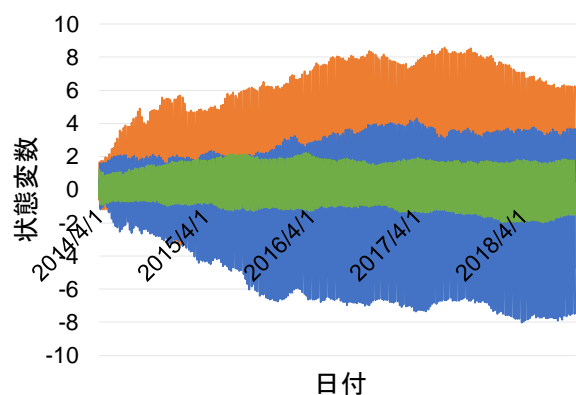


図13 曜日トレンドのタイプ

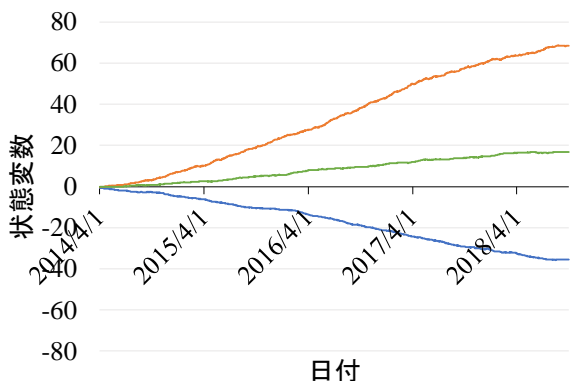


図14 雨など降雨のタイプ

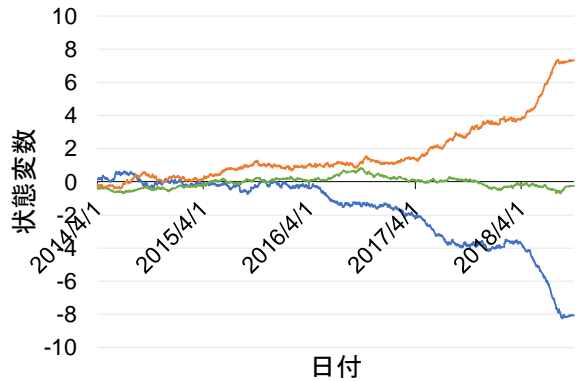


図15 一日平均気温のタイプ

気候に関する変数については、雨など降雨の影響のタイプを図14、一日平均気温のタイプを図15に示す。雨など降雨の影響は、観測された際に状態変数が高くなる利用者は、増加278人、一定246人であるため、正に働いている利用者が多いことが分かる。減少173人は雨の際に利用を避ける可能性がある。しかし、これらの要因については不明である。例えば、雨で利用しなくなる利用者について雨に濡れるから利用しないといった知見が得られた際には、屋根を設置すると利用を促せる可能性がある。一日平均気温に関しては、全体を見るとほぼ一定に推移している。しかし、2016年4月ごろからタイプ間に大きな差が表れた。それ以前は全体的に影響が小さいことが分かる。

祝日・盆・正月であることの利用に関するタイプについて図16に示す。状態変数が高くなるタイプと低くなるタイプに約半数ずつに分かれた。これらのタイプはトレードオフの関係であるため、政策の検討は難しい。

熊本市電のサービスについて、運賃についてのタイプを図17に示す。状態変数と同様に周期性がみられ、1月頃はタイプ間の差が少なく、夏頃になると差が大きくなる傾向がある。増加するタイプと減少するタイプともに、2018年2月頃から状態変数が大きく変化している。これ



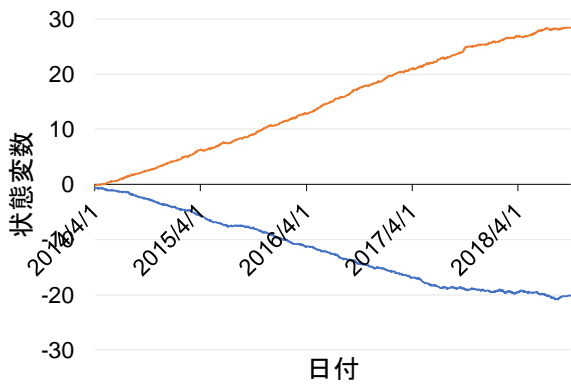


図 16 祝日・盆・正月のタイプ

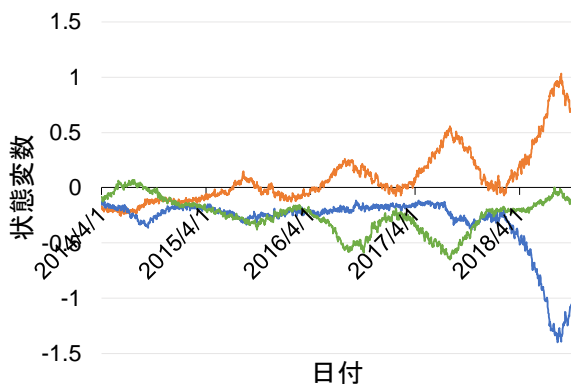


図 17 運賃のタイプ

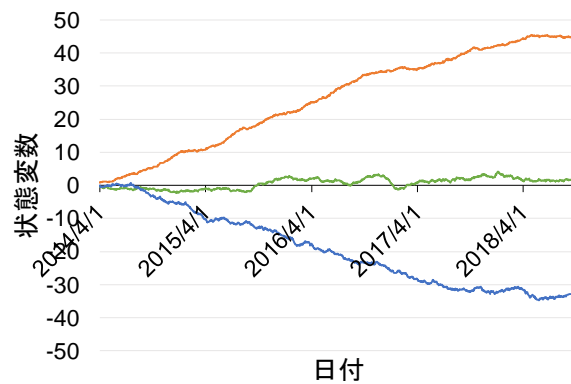


図 18 定期のタイプ

らの具体的な要因については不明である。運賃が高くて利用する利用者が206人、利用が減少する利用者が224人いることが分かる。減少するタイプの利用者に利用を促すためには、運賃の値下げと考えられる。しかし、増加している利用者が減少する恐れがあるため、適切な運賃設定を行う必要がある。

定期利用について図18に示す。多くの利用者は高いまたは、ほぼ一定のタイプに分類されるが、状態変数が減少する利用者もいることが確認できる。そのため、定期形態について見直す利用を促せる可能性がある。

## 6. おわりに

本研究では、スマートカードから得られる利用履歴データを用いて、観測モデルに利用の有無と観測値を導入した二項選択ロジットモデルを、効用関数に導入した変数のパラメータを状態モデルの変数として状態空間モデル構築し、状態推定を行った。さらに、得られた状態変数を用いて利用促進のためのマーケティングを行うため、サンプルのグループ分けを行った。

本研究で得られた成果と課題についてまとめる。

- 1) 全国相互利用交通系ICカードの種類ごとに利用が異なることが分かった。
- 2) 今回は、トレンド、曜日トレンド、雨など降雨、一日平均気温、祝日・盆・正月、運賃、定期について、状態推定を行った。
- 3) 状態変数は、項目間で異なる値を示していることや、利用者ごとに異なる値を示し、利用者の異質性について示すことが出来た。
- 4) 状態変数は0付近から推定を始めたが、次第に大きな値を示す項目もある。
- 5) 状態ベクトルの状態変数を用いて時系列クラスター分析を行った結果、2つまたは3つのグループに分けることが出来た。
- 6) 得られたグループから政策について考察を行った。得られたクラスター分析の結果を基に対応した具体的な公共交通政策などのマイクロマイクロマーケティングについて検討したい。
- 7) 状態空間モデルについて、情報量規準の算出や、パラメータ推定などの精度向上や、利用者の共通性を考慮などモデルの改善を行う必要がある。
- 8) 状態空間モデルを用いた推定には時間がかかるため、最適な標本抽出手法について検討を行う必要がある。
- 9) 利用履歴データを用いて状態空間モデルを適用できた。非集計データであっても、変数の影響を見ることが出来るため、状態空間モデルを適用した分析が幅広く行えるものであると考えられる。

**謝辞：**貴重なデータを提供していただいた熊本市交通局の皆さまには深く感謝申し上げます。

### 参考文献

- 1) 熊本市ホームページ バス網再編：  
[http://www.city.kumamoto.jp/common/Upload-FileDsp.aspx?c\\_id=5&id=2432&sub\\_id=23&flid=178767](http://www.city.kumamoto.jp/common/Upload-FileDsp.aspx?c_id=5&id=2432&sub_id=23&flid=178767)，2020.02.09 最終閲覧
- 2) 熊本市交通局 平成 30 年度事業決済について：  
[http://www.kotsu-kumamoto.jp/kihon/pub/detail.aspx?c\\_id=56&id=1042&pg=1](http://www.kotsu-kumamoto.jp/kihon/pub/detail.aspx?c_id=56&id=1042&pg=1)，2020.02.09 最終閲覧
- 3) 東日本旅客鉄道株式会社 IT・Suica 事業：

- [https://www.jreast.co.jp/youran/pdf/2019-2020/jre\\_youran\\_group\\_p69-74.pdf](https://www.jreast.co.jp/youran/pdf/2019-2020/jre_youran_group_p69-74.pdf), 2020.02.09 最終閲覧
- 4) 熊本市交通局 営業・乗車案内：[http://www.kotsu-kumamoto.jp/one\\_html3/pub/default.aspx?c\\_id=7](http://www.kotsu-kumamoto.jp/one_html3/pub/default.aspx?c_id=7), 2020.02.09 最終閲覧
  - 5) 西内 裕晶, 轟 朝幸, 川崎 智也：生存時間分析を用いた路面電車の利用者数の変化に関する研究-土佐電気鐵道を対象として-, 交通学研究, 2015, 58 巻, pp. 113-120
  - 6) 西内 裕晶, 轟 朝幸：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, Vol.68, No.2, pp.II\_8-II\_17,2012.
  - 7) 森田琢雅 溝上章志 中村 嘉明：IC カードデータによる熊本市電利用者の行動特性分析とダイヤ編成への活用, 土木学会論文集 D3, Vol.73, No.5, pp.I\_993-I\_1001,2017.
  - 8) 吉塚卓史：スマートカードから得られる交通系ビッグデータを用いた熊本市電利用需要の特性分析, 熊本大学大学院自然科学研究科社会環境工学専攻, 平成 30 年度修士論文, 2019.
  - 9) 佐藤忠彦, 樋口知之：ビッグデータ時代のマーケティング ベイジアンモデリングの活用, 講談社, pp.1-131, pp.147-186, 2013.
  - 10) 馬場真哉：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.35-112, pp.161-177, 2018.
  - 11) 久保拓弥：データ解析のための統計モデリング入門—一般化線形モデル・階層ベイズモデル・MCMC (確率と情報の科学), 岩波書店, 2012.
  - 12) 萩原淳一郎, 瓜生真也, 牧山幸史：基礎からわかる時系列分析 R で実践するカルマンフィルタ・MCMC・粒子フィルタ, 石田基広, pp.66-92, pp.218-251, pp.356-358, 2018
  - 13) 北川源四郎：モンテカルロ・フィルタおよび平滑化について, 統計数理, 第 44 巻, 第 1 号, pp.31-48, 1996.
  - 14) 国土交通省気象庁 過去の気象データ検索：<http://www.data.jma.go.jp/gmd/risk/obsdl/index.php>, 2020.02.09 最終閲覧
  - 15) 中西航, 高橋真美, 布施孝志：歩行者挙動モデルのパラメータ推定への一般化状態空間モデルの適用, 土木学会論文集 D3, Vol.71, No.5, pp.I\_559-I\_566,2015
  - 16) 中川慧, 今村光良, 吉田健一：株価の変動パターン類似性を用いた株価予測, 人工知能学会全国大会論文集, JSAI2017 巻, 第 31 回全国大会, セッション ID 2D1-1, p. 2D11,, 2017.
  - 17) 白浜公章：経済データに対する値と形状に基づく時系列類似度の比較, 国民経済誌, 204(5):71-79, 2011.

(20???.?? 受付)

## ANALYSIS ON CHARACTERISTICS OF KUMAMOTO CITY TRAM USERS' PREFERENCE USING SMART CARD USAGE HISTORY DATA

Kazuki MIYAZAKI, Shoshi MIZOKAMI and Ryoma YAE

The number of public transportation users have been decreasing in Kumamoto city recently . The smart cards have been introduced to public transportation in Kumamoto City. In the smart usage history data includes many kinds of information such as the getting-on date and time and the name of stops. By using these data, many analyses have been done, however, no studies have analyzed individual long-term preference changes. In this study, we analyze the time-varying characteristics of users' preference for city trams by using a state-space model. As a result, it is founded that all users are divided into several groups by the state estimates. Finally, we examine ways to increase the number of rides for each group.