

公共交通ビッグデータを用いた 潜在的な優良顧客の発見手法

谷本 圭志¹・廣畑 健也²・伊藤 昌毅³・宮崎 耕輔⁴

¹正会員 鳥取大学 教授 工学研究科社会基盤工学専攻 (〒680-8552 鳥取市湖山町南4丁目101)
E-mail: tanimoto@tottori-u.ac.jp

²非会員 鳥取大学工学部社会開発システム工学科 (〒680-8552 鳥取市湖山町南4丁目101)
E-mail: B14T7042X@edu.tottori-u.ac.jp

³正会員 東京大学生産技術研究所 助教 (〒153-8505 東京都目黒区駒場4-6-1)
E-mail: mito@iis.u-tokyo.ac.jp

⁴正会員 香川高等専門学校 准教授 (〒761-8058 香川県高松市勅使町355)
E-mail: miyazaki@t.kagawa-nct.ac.jp

地方では、少子高齢化や人口減少の影響に伴って公共交通の利用者数が減少している。近年では交通ICカードの普及が進み、そこで得られるビッグデータを活用したマーケティングが可能になっており、それを踏まえた対策が期待されている。そこで本研究では、香川県の高松琴平電気鉄道を対象に、長期の利用履歴に基づいて顧客をクラスタリングすることで優良顧客に相当するグループを特定し、その上で、各地域で予測される顧客数と実際の利用者数を比較することで、どの地域でどれほどの優良顧客の発見が可能かを見出す手法を開発する。

Key Words: public transportation service, big data, lucrative customers, clustering

1. はじめに

公共交通は通勤や通学、病院、買い物など、人々の日常的な移動には欠かせないサービスである。しかし、地方では自家用車の世帯当たりの保有台数が多いことに加え、少子高齢化、人口減少が進んでいるため、公共交通の利用者数が従来に比べて少なくなっている。このため、経営状況は今後ますます厳しくなることが予想されており、サービスの維持が大きな課題となっている。

サービスを維持するための一つの方策として、利用の促進が考えられる。この取り組みを企画するためには、誰に対して働きかけることが有効かといったマーケティング的な視点に基づく検討が不可欠である。そのためには利用に関する十分なデータが必要となるが、以前はその蓄積がなされていないか、データがあってもその活用が十分になされていない場合が少なくなかった。

非接触式 IC カードの Suica が登場して以来、多くの公

共交通機関に交通 IC カードが導入されるようになった。交通 IC カードの登場により、利用者の利用履歴を中心としたデータが容易に蓄積できるようになった。これにより、従来実施できなかった交通ビッグデータによる分析やマーケティングが可能となっている。

そこで本研究では、交通ビッグデータを活用して潜在的な優良顧客を把握する手法を検討する。具体的には、利用履歴のデータを用いて顧客を利用回数に基づいたいくつかのグループに分類し、優良顧客に相当するグループを特定するとともに、各年齢層にどのグループの顧客がどれほど期待できるのかを導出する。その結果を踏まえ、各地域におけるグループごとの顧客数を予測するとともに、それらを実際の顧客数と比較することで、さらなる利用が見込める地域を見出す。以上の検討に際しては、高松琴平電気鉄道の沿線地域を対象として実証的にアプローチする。

2. 本研究の基本的な考え方

2.1 既往の研究

公共交通の利用促進に関する研究は数多く存在する。木内ら¹⁾は、アンケート調査に基づいて意識・交通行動を把握し、モビリティ・マネジメントによる鉄道の利用促進効果を確認している。また、大野ら²⁾は地方部の通学生を対象にモビリティ・マネジメントを効果的に実施する方法を明らかにすることを目的とし、送迎による通学に関する非個人的・社会的なデメリットを対象者に伝えることが送迎を抑制し、公共交通での通学を促す上で効果があることを明らかにした。高木ら³⁾は、潜在需要を分析するにあたって、地方都市において住民の意識調査を行い、バスの利用需要を年齢別に推計している。本研究では優良顧客を発見することを目的としており、潜在需要に着目したこの研究とは問題意識を一にする。

鉄道によるサービスの改善に関する研究も少なからずある。三寺ら⁴⁾は地方鉄道の再生のために効果的な駅周辺地区の整備方策を導き出すことを目的とした検討を行っている。藤居ら⁵⁾は、地方都市の鉄道利用の減少に対して、沿線の地域住民にアンケート調査を行い、鉄道利用の施策に対する限界評価額を試算し、アンケートから得られた施策として運賃の値下げや運行間隔の短縮の評価が高いことを明らかにした。

これら既往の研究では、分析に際してアンケート調査の結果を用いており、豊富な実績データを扱ったアプローチではない。近年、交通 IC カードを用いた研究が盛んに行なわれるようになっており、長期の需要変動を分析した嶋本ら⁶⁾、利用者の行動特性に着目した西内・轟⁷⁾や森田ら⁸⁾、運賃制度と利用者の行動を分析した近藤・嶋本⁹⁾などがある。しかし、これらはいずれも、優良顧客という顧客の特性に着目し、その発見の可能性に着目するという視点は持ち合わせていない。

2.2 潜在的な優良顧客の発見のアプローチ

本研究では、利用履歴に基づいて顧客をクラスタリングし、優良顧客を見出すとともに、どの年齢にどのクラスターがどれほど存在するのかを明らかにする。一方で、各地域における年齢別の居住者数は国勢調査で把握することができる。よって、ビッグデータ解析によって導出される年齢別のクラスターの分布と国勢調査に基づく年齢別の人口を乗じることで、クラスターごとの期待利用者数を地域ごとに関して算出できる。この値と実際の利用者数を比較し、前者が後者よりも上回っている地域であることが判明すれば、その地域は当該のクラスターに関して利用促進の余地があると考えられる。これによって、どの地域にも一律的に利用促進を働きかけるのではなく、特定の地域に対して重点的に働きかけることが可

能となり、効果的な利用促進が可能になる。

以上が研究のプロセスであるが、技術的な面では以下のようなアプローチに基づく。まずは、各利用者の利用履歴に基づいて一ヶ月の利用回数を集計し、その時系列データを用いてクラスター分析を行うことで、利用の多さに応じて利用者を分類する。これにより、どのようなクラスターが存在し、どのクラスターが優良顧客に相当するのかを特定するとともに、各利用者が属するクラスターを明らかにする。すると、各利用者の年齢が分かれば、「年齢」と「所属するクラスター」の組み合わせのデータが得られる。

そこで、次にはこれらの組み合わせデータを用いて、年齢ごとに各クラスターに所属する確率を求める。その際、カーネル密度推定により各クラスターにおける年齢の分布を推計した上で、ベイズの公式によって年齢別の各クラスターの所属確率を求める。最後に、この所属確率と、国勢調査による各地域の年齢別人口と乗じることによって、地域ごとの利用者の予測を行う。こうして予測した利用者数と実際の利用者数を比較し、利用者予測が上回った地域を、さらなる利用が見込める地域であるとする。

2.3 対象地域

香川県の高松琴平電気鉄道（ことでん）を対象として検討する。ことでんには、高松市とその沿線を結んでいる志度線、長尾線、琴平線の 3 つの路線があり、52 の駅が存在する。ことでんの交通 IC カードには IruCa があり、これによりことでんの電車、バスが利用できるとともに、関連する施設での買い物も可能になっている。IruCa には IC カード乗車券と IC カード定期券がある。

2.4 利用者のクラスタリング

(1) k-平均法の概要

利用者の中にどのようなグループが存在し、どのグループが優良顧客に相当するのか特定するために、利用者の利用履歴を用いてクラスター分析を行う。クラスター分析には様々な手法があるが、本研究では非常に大きなサンプルを対象とすることから、その場合によく用いられる k-平均法を用いる。k-平均法は式(1)を最小化することを目的変数とし、各クラスターと各サンプルの距離が最も短くなるように計算が行われる。ただし、 c は任意のクラスター、 K はクラスターの数、 x_i はサンプル i のデータ、 X_c はクラスター c に属するサンプルのデータの集合、 μ_c はクラスター c の重心である。

$$\arg \min_{\mu_1, \dots, \mu_K} \sum_{c=1}^K \sum_{i \in X_c} \|x_i - \mu_c\|^2 \quad (1)$$

(2) クラスタ数決定

クラスタ数 K の決定には様々な手法が提案されているが、その中でもよく知られた簡易な手法がエルボー法である。エルボー法はクラスタリングの良し悪しのうち凝集性に着目した方法であり、どれだけクラスタ内にデータが集まっているかを表す。エルボー法では、それぞれのクラスタ数のもとでの各クラスタの距離の総和(within-cluster sum of squares, WCSS)を目的関数としてグラフに表し、グラフの傾きが大きく変わっている点を最適なクラスタ数として決定する。WCSS は式(2)で表すことができる。

$$WCSS = \sum_{c=1}^K \sum_{i \in X_c} \|x_i - \mu_c\|^2 \quad (2)$$

3. 分析の流れ

3.1 所属クラスタの特定

過去の利用者の利用履歴を一ヶ月単位で集計し、個々の利用者ごとに調査対象期間における最初の月から最終の月までの利用回数を要素としたベクトルデータを作成する。これを各サンプルのデータとしてクラスタ分析を行う。これにより、任意のサンプル i がどのクラスタに所属するのかが明らかになる。これを次式で示す ϕ_c で表す。

$$\phi_c = \begin{cases} 1 & (i \text{ がクラスタ } c \text{ に属する場合}) \\ 0 & (i \text{ がクラスタ } c \text{ に属さない場合}) \end{cases} \quad (3)$$

すると、任意のクラスタ c の所属確率 $P(c)$ は次式で表される。ただし、 n はサンプル数である。

$$P(c) = \frac{\sum_{i=1}^n \phi_{ic}}{n} \quad (4)$$

3.2 年齢別のクラスタの所属確率

前節の結果として各利用者が所属するクラスタが判明するが、各利用者の年齢が把握できれば、年齢と所属クラスタの組み合わせデータが得られる。このデータを用いて、クラスタ別の年齢分布を求めることができる。その方法として、クラスタ別にカーネル密度推定を適用し、平滑化カーネル分布を特定することで、クラスタ別の年齢分布を求める。こうして得られた任意の

クラスタ c における年齢 a の(累積)分布関数を $F_c(a)$ で表す。すると、クラスタ c における年齢 a の確率 $P(a|c)$ は次式で求められる。

$$P(a|c) = F_c(a+1) - F_c(a) \quad (5)$$

なお、上式は年齢が一歳刻みの場合を表しているが、その他の刻み幅についても同様に定式化できる。式(4)、(5)より、ベイズの公式を用いることで、年齢別におけるクラスタの所属確率は次式で算出することができる。

$$P(c|a) = \frac{P(a|c)P(c)}{\sum_{c'=1}^K P(a|c')P(c')} \quad (6)$$

3.3 各地域におけるクラスタ別の利用者の予測

式(6)で得られた年齢別のクラスタ所属確率を、各地域の年齢別の人口に乘じることにより、各クラスタの利用者数を地域ごとに予測する。すなわち、任意の地域 j における年齢 a の人口を $N_j(a)$ で表すと、 $P(c|a)N_j(a)$ によってクラスタ c の利用者数が推計できる。

しかし、クラスタはそもそも利用者を分類したものであることから、確率 $P(c|a)$ は利用者を対象としたクラスタの分布に過ぎない。したがって、 $P(c|a)N_j(a)$ はそもそも利用しないという可能性を反映しておらず、この推計値は過大な値である。

そこで、 $P(c|a)N_j(a)$ に利用しない可能性を乘じることで、利用しないという可能性も反映した推計値に修正する。具体的には、地域 j において現在ことを利用している人の割合 λ_j をデータから求め、これを $P(c|a)N_j(a)$ に乘じる。なお、以下では便宜的に λ_j を補正係数と呼ぶ。以上より、任意の地域 j におけるクラスタ c 、年齢 a の利用者数の予測は、次式で求められる。

$$P(c|a)N_j(a)\lambda_j \quad (7)$$

すると、地域 j におけるクラスタ c の利用者数の予測値 $u_j(c)$ は次式で求められる。

$$u_j(c) = \sum_a P(c|a)N_j(a)\lambda_j \quad (8)$$

3.4 優良顧客が見込める地域の特定

優良顧客が見込める地域を特定するために、各地域で利用者を予測するとともに、現在の利用者数の実績を集計し、これらを比較する。具体的には、地域 j における

クラスター c の利用者数の予測を式(8)で求め、実績は次に示す $v_j(c)$ により求める。ただし、 y_{ij} はサンプル i が地域 j に居住する確率であり、データから居住地が分かるサンプルについては確定的に y_{ij} が与えられる。一方、データから居住地が分からないサンプルについては、サンプルが最もよく利用している駅をデータから特定するとともに、その駅の利用者の居住地の分布を求め、その上で、この分布が当該のサンプルがどの地域に居住しているのかの確率分布であるとして、居住地を確率的に与える。

潜在的な優良顧客の発見の可能性は、予測と実績の差と比の二つの観点での評価が考えられる。そこで、 $u_j(c) - v_j(c)$ 、ならびに、 $u_j(c)/v_j(c)$ の二つの指標で可能性を評価する。

$$v_j(c) = \sum_{i=1}^n \phi_{ic} y_{ij} \quad (9)$$

4. 分析結果

4.1 利用者のクラスタリング

本研究ではクラスタリングするデータとして、2013年4月～2015年2月の23ヶ月における15歳以上のことでの利用者 111,594 人の利用履歴データを用いた。また、地域としては、ことでの沿線市町である高松市、さぬき市、丸亀市、綾川町、三木町、琴平町、まんのう町を対象に、大字で区分される364の地域とした。

クラスター数を決定するにあたって、クラスター数を3～12と設定してWCSSを算出した。その結果を図-1に示す。この図より、クラスター数を9のときにグラフの傾きが緩やかになる変化が認められたため、エルボー法により、クラスター数を9とした。

クラスター数が9の場合のクラスタリングの結果を図-2、また各クラスターの特徴を表-1に示す。これらより、利用回数の高低だけでなく、途中から利用が増加したクラスター、逆に途中から利用が減少したクラスターがあることが分かる。本研究では、優良顧客を発見することが目的であるため、以下では利用が多い、もしくは今後利用が増えていく可能性があるクラスター1,3,5,8,9に着目し、以下に示す図表ではこれらにアンダーラインを付け、他と区別する。

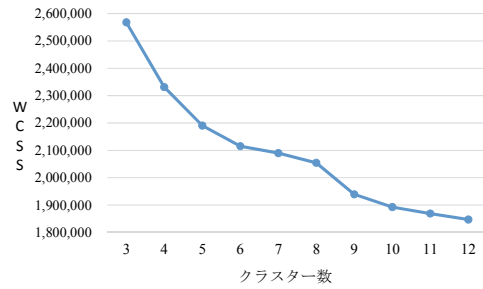


図-1 クラスター数とWCSSの結果

表-1 各クラスターの特徴

クラスター	特徴
<u>1</u>	週に2回程度使っている利用者。
2	当初は週に5回程度の高い頻度で利用するが、2014年に利用が減少傾向となり6月以降の利用はほとんどなくなった利用者。
<u>3</u>	ほぼ毎日利用しており全クラスターの中で利用頻度が最も高い利用者。
4	当初は定期的に利用するが、クラスター2と同様2014年以降の利用はほとんど見られない利用者。
<u>5</u>	2014年4月以降新たに使い始めた利用者であり、2014年10月以降は週に2回の頻度で利用している利用者。
6	年間を通じて利用の頻度が低い利用者。
7	月に2,3回程度使う利用者。
<u>8</u>	週に5回程度利用しており、クラスター3に次いで利用頻度の高い利用者。
<u>9</u>	2014年4月以降新たに使い始めた利用者であり、頻度は週に5回程度使う利用者。

※下線を付しているクラスターが優良顧客

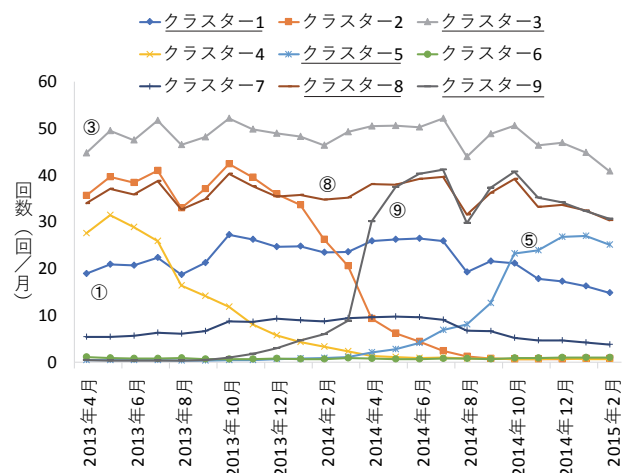


図-2 クラスタリングの結果

(図中の○印内の数値はクラスターの番号)

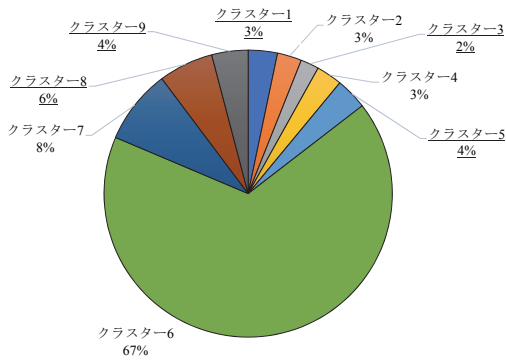


図-3 各クラスタの所属確率

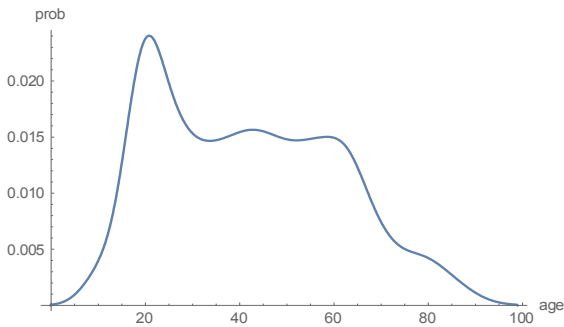


図-4 クラスタ-1における年齢分布

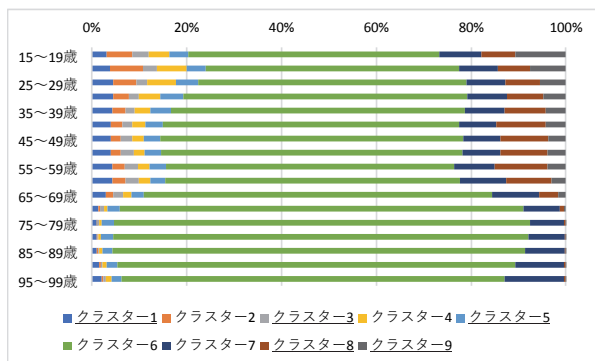


図-5 年齢別のクラスタの所属確率

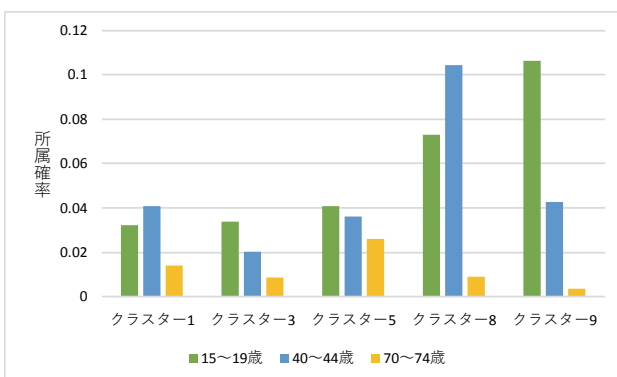


図-6 代表的な年齢におけるクラスタの所属確率

4.2 クラスタの所属確率

式(4)に基づき、各クラスタの所属確率 $P(c)$ を算出した。その結果を図-3 に示す。この結果から、利用回数

が最も少ないクラスタ-6 の割合が最も高いことが分かる。一方、優良顧客であることが判明したクラスタ-1, 3, 5, 8, 9 は全体の 2 割程度である。

4.3 年齢別のクラスタの所属確率

各利用者の年齢とクラスタ分析の結果から得られた所属クラスタを用いて、カーネル密度推定により各クラスタの年齢分布を算出した。その結果の一例を図-4 に示す。次いで、式(6)を用いて、年齢別の各クラスタの所属確率 $P(c|a)$ を算出した。その結果を図-5 に示す。

図-5, 6 より、最も利用頻度の高いクラスタ-3 および 2014 年 4 月以降に利用が増えるクラスタ-5, 9 は年齢が高いほど所属確率が小さくなる傾向にある一方、比較的使用頻度の高いクラスタ-1, 8 は、中年層で所属確率が高い傾向にある。また、高齢者層では優良顧客は少なく、ほとんど使わないというクラスタ-6 が相対的に多くを占める。

4.4 予測値の妥当性の検証

式(8)に基づいて、地域におけるクラスタ別の利用者数の予測値を算出した。なお、 $N(a)$ については、平成 27 年度の国勢調査の人口を用いた。一方、データから式(9)によって表される実績を算出した。

優良顧客が見込める地域の特定に先立って、利用者数の予測値が実績と著しく異なっていれば、その予測値には妥当性がないと考えられる。そこで、その検証のために、クラスタ別に予測値と実績の散布図を求めた。その結果を図-7 に示す。回帰式より、予測値と実績が大きく離れている結果ではないことが分かる。

4.5 潜在的な優良顧客が見込める地域の特定

各地域における予測と実績の差ならびに比を求めた。その際、データから居住地が分からないサンプルは 36,796 人であったため、これらについては先述の方法で確率的に居住地を求めた。その上で、少なくとも一つ以上のクラスタに関して差が 10 人以上でかつ比が 2 以上である地域を特に利用が見込める地域とし、その結果を図-8 に示す。なお、ある地域では複数のクラスタに関して利用が見込めることから、利用が見込めるクラスタを色で区別した。

特に利用が見込まれるのは、どのクラスタでも目的地が集中する地域や JR 線の沿線地域が多い。ただし、この結果には注意が必要である。本研究では、年齢に応じてクラスタの分布が決まるとしているが、そもそも上記の地域は他と比べて優良顧客の割合が少ないと考えられる。すなわち、目的地が近ければ徒歩や自転車など他の交通手段も利用できることから、他の地域に比べて外出の頻度が同様であってもことごとくに乗車する頻度は

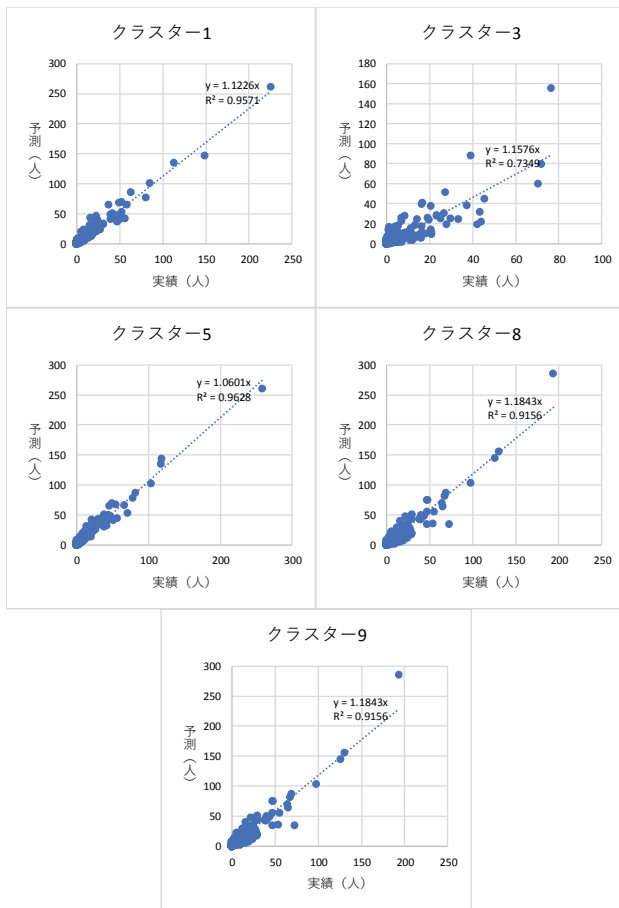


図-7 クラスタ別の利用者数の散布図

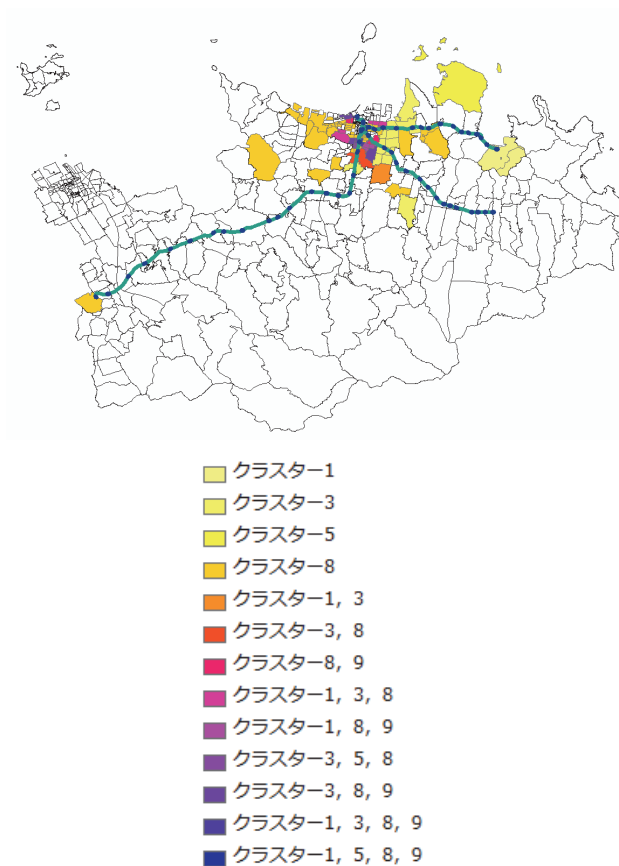


図-8 特に利用が見込める地域

少ないと考えられる。JR の沿線地域についても同様である。よって、これらの地域については利用の見込みの高さは割り引かなくてはならない。

一方、そうではない地域については、利用の見込みが高いと考えられる。具体的には、二つの路線にはさまれた地域、中心部から離れた地域である。特に、琴平線と長尾線（それぞれ、図の左側と中央の路線）にはさまれた地域では、複数のクラスタで利用が見込まれるため、これらの地域について重点的に利用を促すのは効果的であるとされる。

また、必ずしも複数のクラスタではないものの、志度線（図の上側の路線）の北に位置している地域についても利用が見込める可能性がある。この地域は駅から遠く離れた場所に位置しており、北部だと自動車でも十数分かかる。よって、パークアンドライドといった駅の駐車場の整備や公共交通の充実によって、利用者の確保の可能性もあることも示唆される。このように、どの地域に見込みがあるのかだけでなく、地域の条件と照らし合わせることにより、どのような対策が考えられるのかについて検討する材料を与えることができる。

5. おわりに

本研究では、IC カードに蓄積された長期の利用履歴を用いて、顧客をクラスタリングすることで優良顧客を明らかにした。その上で、各地域において予測される顧客数と実際の利用者数を比較することで、どの地域で優良顧客の発見が可能かを見出す手法の開発を試みた。

ことで沿線地域でその手法を実証し、予測値についても良好な値を示すとともに、その値を用いて優良顧客が見込める具体的な地域をクラスタ別に見出し、その中でも特に利用促進にふさわしい地域を特定することができた。

一方で、この手法によって特定された地域が実際にふさわしい地域なのかの検証が必要である。このため、過去にことでんが実施した何らかの利用促進策の反応とこの手法での結果を照合するなどの付加的な検討が今後必要となる。

謝辞：本研究は JSPS 科研費 16K06548 ならびに鳥取大学持続的過疎社会形成研究プロジェクトの助成を受けた。また、全般について高松琴平電気鉄道株式会社に多くの助力を得た。付して謝辞と致します。

参考文献

- 1) 木内徹, 土井勉, 藤井聡: 鉄道の利用促進を目的としたモビリティ・マネジメントの取組と課題, 土木学会論文集 D, Vol.64, No.1, pp.111-121, 2008.

- 2) 大野悠貴, 北原啓司: 通学利用の確保に資する効果的なモビリティ・マネジメントの実施に関する研究—高校生を対象とした弘南鉄道大鰐線における利用促進方策を事例として—, 都市計画論文集, Vol.52, No.3, pp.825-832, 2017.
- 3) 高木昌也, 奥嶋政嗣, 近藤光男: 地方都市のバス利用促進のための行動意図法による潜在需要分析, 都市計画論文集, No.45-3, 2010.
- 4) 三寺潤, 本多義明: 地方鉄道の再生のための駅周辺地区の評価と整備方策に関する研究, 都市計画論文集, No.39-3, 2004.
- 5) 藤居良夫, 佐竹わか菜: 地方都市における地方鉄道利用の社会経済的評価に関する研究, 都市計画論文集, No.41-3, 2006.
- 6) 嶋本寛, 北脇徹, 宇野伸宏, 中村俊之: 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), 70 巻, 5 号, pp.1_605-I_610, 2014.
- 7) 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 (土木情報学), 68 巻, 2 号, pp.11_8-II_17, 2012.
- 8) 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 (土木計画学), 73 巻, 5 号, pp.1_993-I_1001, 2017.
- 9) 近藤篤史, 嶋本寛: IC カードデータを用いた公共交通の運賃制度と乗客行動の関係性分析, 土木学会論文集 D3 (土木計画学), 72 巻, 5 号, pp.1_653-I_660, 2016.

(2018. 4. 27 受付)

FINDING POTENTIAL HIGH-VALUE CUSTOMERS OF PUBLIC TRANSPORTATION SERVICE

Keishi TANIMOTO, Kenya HIROHATA, Masaki ITO and Kosuke MIYAZAKI

In recent years, the marketing activity is possible for the traffic carriers by using big data obtained from the traffic IC cards record data. Currently the promotion using the data is paid attention. In this study we clarify the characteristics of lucrative customers by clustering the customers using long term history big data of Takamatsu-Kotohira Electric Railway in Kagawa prefecture, and then estimate the number of the lucrative customers expected in each region. In addition, we develop a method to find out the regions in which potential lucrative customers live by comparing expected number of customers to the actual number of customers.