

交通系 IC カードログによる駅ごとの乗降者数格差の検証

加藤 弘則¹・伊藤 昌毅²・瀬崎 薫³

¹ 東京大学大学院総合文化研究科広域科学専攻基礎科学系博士課程

E-mail: hironori_kato11s@gmail.com

² 東京大学生産技術研究所

³ 東京大学生産技術研究所, 東京大学空間情報科学研究センター

Suica をはじめとする交通系 IC カードからは、乗降客の乗降履歴が取得可能である。本研究では、高松琴平電気鉄道（香川県）が発行する IC カード IruCa の乗降履歴データを用いて、一人あたりの乗車回数観点からみた駅間のジニ係数の算出と、駅間のジニ係数をもとに駅のクラスタリングを試みた。

交通センサスなどの調査結果では取得不可能な、データユニークなカード ID に紐づく、乗車駅、降車駅、乗車開始時間などの情報を集計することにより、ある乗降客一人あたりの乗車回数の分布の確認が可能となる。また、一人あたりの乗車回数の分布を駅に紐付けて、駅ごとにジニ係数それぞれ算出することで、駅間の乗降客に関する不平等度合いを確認すると共に、「乗り換え駅」「繁華街」「ビジネス街」などの駅のクラスタリングを試みた。

Key Words: IC カード, ジニ係数, 乗車回数分布

1. はじめに

Suica をはじめとする日本における交通系 IC カードは、全国 78 事業者により、平成 25 年 3 月約 8,700 万枚が発行されている¹⁾。ユーザが交通インフラを利用するごとに交通系 IC カードが生み出すログでは、乗降駅名や乗降が行われた時間などの移動に関する情報が、カードに付随するユニークな ID に紐づく。その結果、ほぼ日本全国の鉄道・バスをカバーしている²⁾と言っても過言ではない膨大なカードによって、それらのログは莫大となり、リアルな人々の移動が分析可能となる。

交通系 IC カードのログ分析を巡っては、観光・スマート化の文脈で実施が進む³⁾。特に、高齢化・持続可能性の観点からは、エネルギー効率の観点からダイヤの適正化を図り、環境効率が高い交通環境の整備が求められる。ログの欠損が少なく、ユニークなユーザ ID に紐づく交通系 IC カードのログを用いることで、路線の需要予測や混雑予測が可能となる。

既存の交通分析に関しては、大都市交通センサスが代表的に用いられる。大都市交通センサスとは、首都圏、中京圏、近畿圏の鉄道・バス等の大量公共交通機関の利用実態を明らかにする事を目的に国土交通省が実施する

調査である。調査項目は、鉄道 OD 調査、鉄道定期券・普通券等利用者調査、鉄道輸送サービス実態調査、バス・路面電車定期券・普通券等利用者調査、アンケート票回収数、バス・路面電車 OD 調査、バス・路面電車輸送サービス実態調査、乗換え施設実態調査で構成される⁴⁾。

しかし交通センサスは 5 年ごとに実施され、特定の 3 日間に配布した質問紙の回答の集計結果を元に調査結果を得るため、緻密な交通インフラの利用実態の把握が難しい他、平成 27 年実施分からは質問紙のオンライン回答が可能となったもののサンプリング調査である事には変わらない。また、定期券ユーザの動向については、輸送人員の実態と調査結果で誤差が生じる可能性がある⁵⁾。

また、地方交通の利用実態については、鉄道に関しては各地方における運輸局が年度ごとに取りまとめる。調査項目は営業キロ、輸送人員、輸送人キロ、一人平均乗車キロ、1 キロ 1 日平均輸送人員、旅客運賃収入である⁶⁾⁷⁾。調査は各鉄道事業者に対する質問紙調査によって行われる。

本稿では、交通系 IC カードの鉄道利用に関するログを用いて、既存の調査手法で実施される調査項目の一部について実施の可能性を探るとともに、IC カードのロ

グならではの鉄道の利用実態調査の方針について検討を行う。

2. データセットについて

香川県の高松琴平電気鉄道における、IC カードを通して得られた乗降ログ 14,535,764 件（2013 年 3 月 1 日～2015 年 2 月 28 日の全乗降履歴）を分析対象として用いた。データセットの乗降履歴に関するカラムは表-1 より構成される。

3. 高松琴平電気鉄道について

高松琴平電気鉄道（正式名称、高松琴平電気鉄道株式会社）は香川県の鉄道・バス会社である。琴平線、長尾線、志度線の 3 路線で構成され、総営業キロは 60.0 キロにのぼる（図-1）。

駅ごとの距離は 2～3 キロと、付近に位置する JR と比較して短い。

なお、1 キロ 1 日平均輸送人員は 5,355 輸送人キロは 117,277 である。

高松琴平電気鉄道の IC カードの利用状況については、2005 年には非接触式の IC カードシステム「IruCa（イルカ）」が、四国の鉄道事業者としては初めて導入された。2018 年より、全国相互利用サービスの対応が開始された。

それぞれの券種の利用状況については、ユーザ数ベースでは非定期券ユーザと定期券ユーザの比率は約 2:1 なのに対し、乗車回数の比率は約 3:7 程度であり、乗車回数のお大半が定期券ユーザによって構成されている（表-2、表-3）。

また、定期券と非定期券の時間帯別乗車状況をみると、定期券ユーザの朝・夕の時間帯に乗車が集中しており、平日の通勤・通学の需要によって乗車回数が構成されていると考えられる。

表-1 IC カードのログフォーマット

型	意味
INTEGER	乗車駅が属する路線コード
INTEGER	降車駅が属する路線コード
TIMESTAMP	乗車駅改札通過時間
TIMESTAMP	降車駅改札通過時間

INTEGER	乗車時間 (秒)
TIME	乗車時間 (mm:ss)
INTEGER	乗車駅コード
STRING	乗車駅名
STRING	乗車駅 ID
INTEGER	降車駅コード
STRING	降車駅 ID
STRING	降車駅 ID
FLOAT	距離 (キロ)
INTEGER	駅数
INTEGER	券種コード
STRING	券種



図-1 高松琴平電気鉄道の駅の分布状況

表-2 高松琴平電気鉄道の外観について（2017年度）⁹⁾

営業キロ		60
輸送人員 (千人)	定期	8,396
	定期外	5,746
	計	14,142
輸送人キロ (千人)	定期	78,641
	定期外	38,636
	計	117,277
一人平均 乗車キロ	定期	9.4
	定期外	6.7
1 キロ 1 日平均輸送人員		5,355

表-3 高松琴平電気鉄道の券種ごとの利用状況
(2014 年 10 月)

	非定期券	定期券
乗車回数計	195,796	462,198
ユニークユーザ数	28,029	13,391

4. 券種ごとの乗降状況について

交通センサスでは、定期券の利用実態に関して、アンケートによる利用実態をもとに駅ごとの定期券販売枚数による補正を加えて結果を得る。定期券販売枚数による補正を加えるのは、駅ごとに得られる利用者サンプルが異なるためである。加えて、交通センサスによって得られるのは、ある日における定期券を用いて移動した地点間の回数であり、月あたり、週あたりなどの定期券を用いて鉄道を利用した回数は得られない。

また、交通センサスにおいて定期券による通勤・通学は、ある駅において定期券を保有している場合、定期保有者の 100%が通勤・通学をしているという仮定が設けられている。したがって、実際の定期利用者と比較して、1, 2 割程度利用者数が多く算出される傾向がある⁵⁾。

高松琴平電気鉄道において券種別に月あたりの利用回数の分布をユーザ ID 単位で確認したところ、定期券ユーザの乗降回数はこぶ型が一つの正規分布に近い分布であることが認められた(図-3)。こぶのピークは 44 回であり、これは 22 回程度、つまり平日に同じ地点を往復したことが要因を考えられる。定期券ユーザの乗降回数に関しては、一人あたりの乗車回数の分布を駅ごとに考慮した上で利用実態を検討する必要性が示唆される。なお、データのユーザ抽出条件は、2014 年 10 月における全定期ユーザで、定期券の期限が過ぎた後に乗車回数が低下することと、定期券の購入時点から乗車が始まることを考慮して、月内の初乗車が 1~5 日、月内の最終乗車日が 25 日以降のユーザを抽出した。

また、定期外の一般券種ユーザに関しても同様に、月内の利用回数分布を取得したところ、指数分布に近い分布を示した(図-4)。これはユーザが鉄道利用を伴うイベントの発生確率がランダムであり、ポアソン過程に従うことから理解可能である。鉄道利用を伴うイベントが定期的に複数回発生し、かつその費用が定期券購入額を上回る場合、定期券ユーザへ転換されると考えられる。なお、データのユーザ抽出条件は、2014 年 10 月における全普通券ユーザである。

カード ID ごとの乗降回数の分布は、券種によって変化することが認められた。しかし、乗降回数の分布は、駅の周囲の人口や、オフィス街や歓楽街といった駅の周囲の環境によって左右されると考えられる。そこで、駅ごとに、その駅を起点に乗車したユーザと一定期間内の乗降回数の分布を取得した(以下、駅ごとユーザあたり乗車回数分布、と表す)。なお、分析対象期間は 2014 年 10 月である。

その結果、駅ごとユーザあたり乗車回数分布は、駅ごとに分布の形状が異なることが認められた(図-5)。

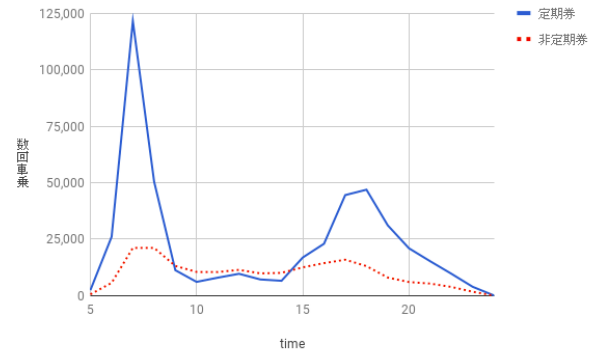


図-2 定期券・非定期券ユーザの時間ごと乗車回数

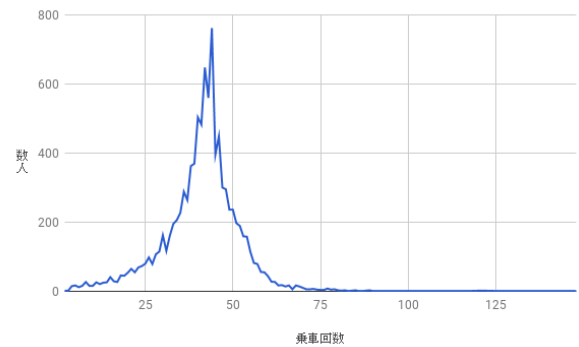


図-3 定期券ユーザの月間の乗車回数分布

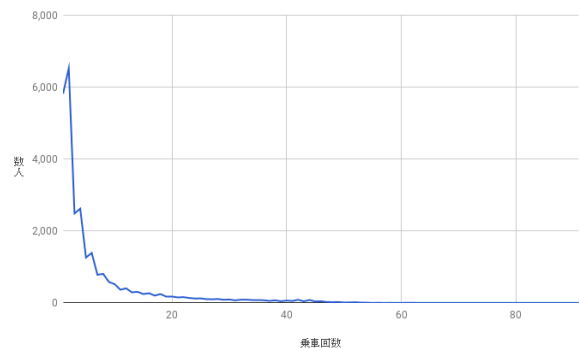


図-4 定期券以外ユーザの月間の乗車回数分布

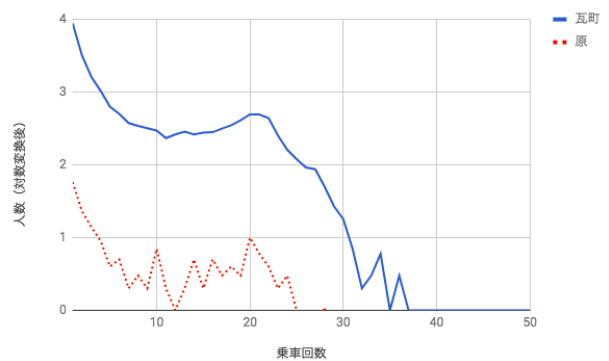


図-5 瓦町駅と原駅の 1 人あたり乗車回数の分布比較
(一ヶ月間の乗車回数)

5. 駅ごととユーザあたり乗車回数分布のジニ係数

駅ごととユーザあたり乗車回数分布における分布の形状の差異は、ある期間における乗降回数の合計が同じであっても、それを構成する乗降客数の違いによって変化が生じることを示唆する。したがって、駅ごととユーザあたり乗車回数分布の形状の差異を駅ごとに比較することで、その駅の乗車回数が相対的に偏った乗車客数によってどの程度支えられているか、を求められる。また、その度合いは、例えば繁華街における様々な乗客によって乗車客数の総量が構成されている状況や、住宅地における固定された乗客によって乗車客数の総量が構成されている状況などによって変化すると考えられる。

そこで、駅ごととユーザあたり乗車回数分布から、分布の不平等度合いを数値として表すジニ係数を駅ごとに算出して比較した。

(1) ジニ係数とは

ジニ係数とは、所得分配の不平等さなど、統計上の分布の状況を計るための指標である。2 変数のそれぞれの累積相対度数を二軸に並べ、原点と各点を結んだ曲線をローレンツ曲線と言ひ、ローレンツ曲線と 45 度線で囲まれる面積を 2 倍した値をジニ係数と呼ぶ。なお、ジニ係数が 1 に近づくほど不平等度が高いといえる。

n 個に分割された各階層をもとに、下から i 番目の階層値の平均 y_i とする ($y_1 < y_2 < \dots < y_n$)。その場合のジニ係数の一般式は

$$g = \frac{1}{2n^2} \sum_{i=1}^n \sum_{j=1}^n |y_i - y_j|$$

となる。

(2) 駅ごとのジニ係数

駅ごととユーザあたり乗車回数分布から得たジニ係数は、高松築港駅を筆頭に、琴電琴平駅、瓦町駅とジニ係数が大きな駅が続く (表-4)。反対に、ジニ係数が小さい駅は、塩屋駅、房前駅、西前田駅と、ジニ係数が小さい順に続く。

駅ごとに比較してジニ係数が大きい (不平等) ということは、ある人に乗車回数が集中していることを意味する。ジニ係数が大きな駅は、定期券利用客が多い、あるいは通勤・通学ルート上の乗り換え駅であることが示唆される。ジニ係数が高い駅をみると、高松築港駅は本州への高速フェリーに乗船できる港の駅であるし、琴電琴平駅や瓦町駅は繁華街や複数路線が乗り入れるターミナル駅である。また、綾川駅は琴平線の途中駅に位置する

が、駅の近隣に大型ショッピングモールがあり、就業者の定期券利用客が多い、もしくは頻繁にショッピングモールを利用するユーザの存在が示唆される。

反対に、ジニ係数が小さな駅は、ある人に乗車回数が集中していないことを示唆するため、定期券利用客が少ない場所に位置した駅であるといえる。ジニ係数が小さな塩屋駅、房前駅、西前田駅などは、商業地域ではなく、宅地、あるいは農地が広がる。

このように、駅ごとのジニ係数をみることで、その駅が抱えるユーザの属性を分析可能となる。また、ユーザ属性から、駅の地理的環境が推測可能といえる。しかし、ジニ係数は構成される集団内における不平等度 (格差) をみる数値であるため、鉄道駅の周囲の環境を推し量るために考慮すべき規模 (乗客回数) については考慮がなされない。

そこで次節では、ジニ係数と乗車回数を駅ごとに考慮した駅分類を実施する。

表-4 駅ごとのジニ係数一覧

駅名	ジニ係数
高松築港	0.958
琴電琴平	0.945
瓦町	0.944
綾川	0.934
松島二丁目	0.932
片原町	0.926
中略	
井戸	0.799
公文明	0.797
羽床	0.796
西前田	0.773
房前	0.771
塩屋	0.755

6. ジニ係数を用いた駅の種類手法

前項で得た、駅ごとユーザーあたり乗車回数分布のジニ係数と、駅ごと乗車回数、平均乗車距離の3点を用いて散布図を得た。既存の駅分類の研究では、交通センサスで得られたデータをもとに駅のラベリングが実施されている。しかし、前述の通り、交通センサスに関してはデータの特異性を考慮すべきであり、本節では簡単ではあるが、IC カードのログを用いた駅のクラスタリングを試行する。

なお、住宅地と他鉄道会社への乗り換え駅の距離や、商業地域との距離が駅の周囲の環境を推し量るための指標になるという仮定の元、平均距離を変数に加えた(図-6)。横軸にジニ係数、縦軸に平均乗車距離、各点の大きさに乗車回数を示させた場合、点が右に位置し、点が大きい場合は、ユーザーに紐付いた乗車回数の格差が大きく、その駅が抱える乗客数も大きいことを示す。したがって、大規模商業地域である可能性が示唆される。また、ジニ係数が大きく、乗客数が小さい場合は小型の商業地域であるか、大型のショッピングモールなどが駅の周囲にある可能性が示唆される。

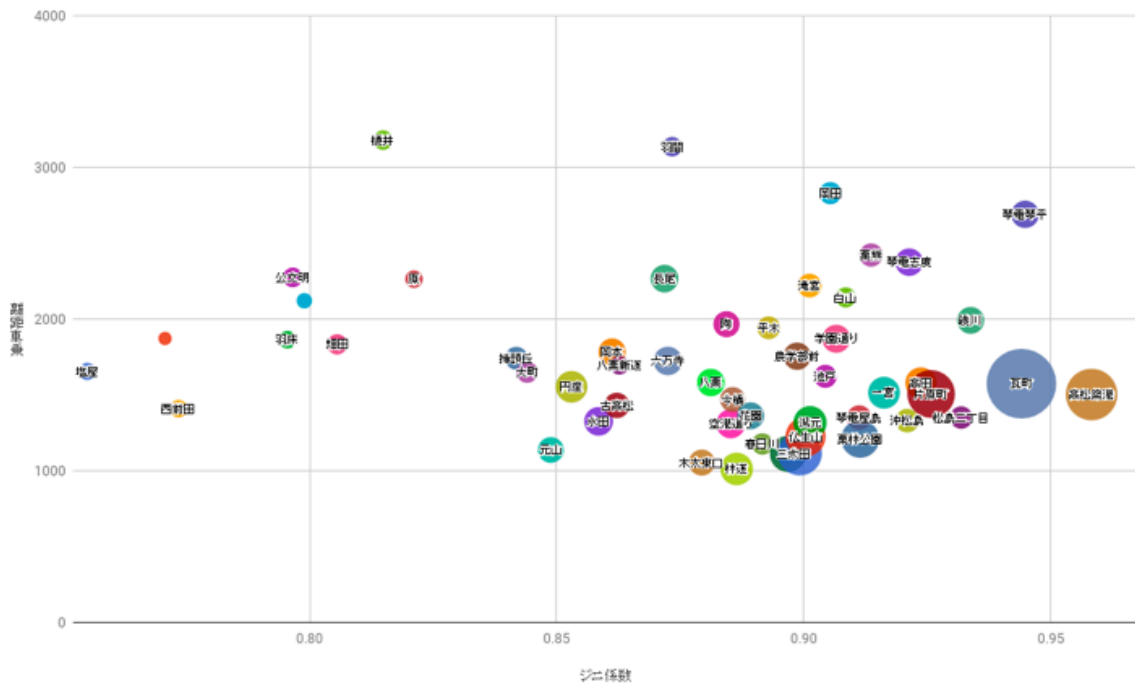


図-6 駅ごとのジニ係数,乗車回数,平均乗車距離の散布図
(縦軸:平均乗車距離,横軸:ジニ係数,点の大きさ:乗車回数)

また、駅ごとのジニ係数と乗車回数、平均乗車時間の 3 点を用いて、各要素に対してその要素を含んだクラスタの重心までの距離の二乗の総和を最小にするウォード法によるクラスタ分析を実施した。デンドログラムでは、各要素（駅名）同士が結合する線の距離に近いほど似た要素と言える（図-7）。

クラスタ分析の結果をもとに、片原町、瓦町、高松築港をクラスタ 1、円座、林道、三条、太田仏生山、湯元、栗林公園をクラスタ 2、その他駅をクラスタ 3 とする。

実際の地図にクラスタ情報を重ねると、クラスタ 1 は本州へ向けたフェリー発着港へのアクセスやビジネス街、

乗り換えターミナル駅を抱える。クラスタ 2 では、観光スポット、繁華街、住宅密集地内に駅が位置する。そこでクラスタ 1 は繁華街・乗り換え駅クラスタ、クラスタ 2 は準繁華街クラスタ、クラスタ 3 は住宅地クラスタと表現できる（図-8）。

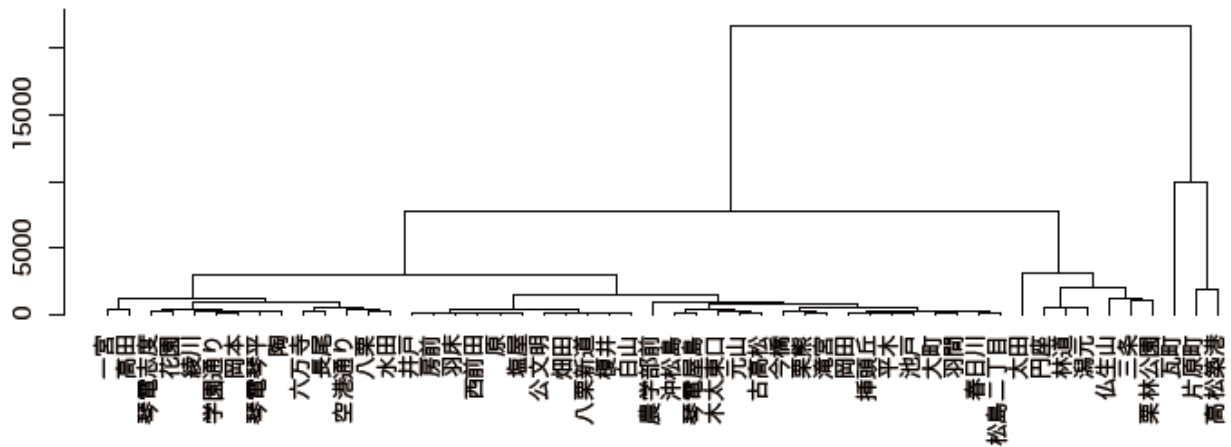


図-7 駅ごとのジニ係数, 乗車回数, 平均乗車距離を元に行ったクラスタ分析の結果 (デンドログラム)



図-8 デンドログラムによるクラスタ分類を地図上にマッピングしたもの

(■:クラスタ 1, ◆:クラスタ 2, ▲:クラスタ 3)

7. 分析を通して得られた示唆と現実世界との差異

駅間ごとの乗降者数によるジニ係数を用いて得られた駅のクラスタを、実際の街並と比較したところ、クラスタ 1 に分類された瓦町駅では、幹線道路やデパート、オフィスビル、飲食店が立ち並ぶ市街地であることは明白であった。片原町駅は商店街に隣接し、大型スーパーマーケット、ミュージアムが並ぶ。高松築港駅は高松港の旅客ターミナルに隣接し、高層建築や大型のホテルも見られる。その他、高松城や裁判所、中央郵便局など街機能の中心的役割を果たす機能が集中している印象である。クラスタ 1 を通して感じられたのは、街の中心地かつ、賑わっている印象である。

クラスタ 2 では仏生山駅に関しては多目的な運動施設の仏生山公園が位置する。栗林公園駅では、付近に歴史的な庭園を擁する栗林公園が隣接している。その他の駅を通して見た感想として、住宅街かつ、休日に訪れる可能性がある施設がある駅、という印象であった。

クラスタ 3 に関しては、駅前にコンビニエンスストアや商店、飲食店が見受けられない駅が多く、典型的な住宅街である印象であった。その一方で、乗り換え駅である琴電琴平駅や琴電志度駅もクラスタ 3 に含まれており、利用シーンが明らかに特徴的と言えるであろう両駅が住宅街の駅と同じクラスタとして分類された結果には違和感が残る。

このように、本手法によるクラスタ分類は、市街地と準市街地、住宅地を大まかに分類する点においては概ね実用的な精度を得られたと考えられる。

8. まとめ

IC カードから得た、一意のカード ID に紐づく乗降客ログから、1 人あたり乗車回数とジニ係数を得た。駅ごとのジニ係数の差異は、駅が抱えるユーザの属性と、駅周囲の環境を示唆することが認められた。また、駅ごとのジニ係数と乗車回数、乗車時間を考慮し、駅の分類の可能性が示唆された。

本論を通して得た手法、示唆は、IC カードのログが持つ「カード ID に紐づくデータ」であることによって実現された。駅ごとのジニ係数は、駅の利用状況を分解する足がかりとなり、特定の属性のユーザを増加させる意図で行われた街づくりの効果検証や、乗車回数を高めることで運賃収益を向上させるための施策立案に寄与できると考えられる。また、ジニ係数とローレンツ曲線による解析は所得格差や資源の分布など、経済学の分野でこれまで広く用いられてきたため、運賃収入によって主に運営される鉄道分野における経営分析や沿線の商業施

設の分析、交通広告の分析を経済学における既存研究の応用によって実施できる可能性が示唆される。

しかし、ジニ係数によるユーザ属性の推定がなされた一方で、ジニ係数と平均乗車時間、乗車回数から実施した駅の分類手法に関しては、鉄道事業者 1 社のデータから行った都合上、大都市圏の鉄道駅でも同様の結果が得られるか疑問が残る。また、ウォード法を用いたクラスタリング手法に関しては、ターミナル駅か住宅街の駅か程度の判別は可能であるものの、どのような施設が周囲に存在するかなど周囲の環境に関して精緻に分類できる精度には至っていない。今後はその他の鉄道事業者が保有する IC カードのログを用いて駅のクラスタリングが行えるか実証を目指す一方で、IC カードのログに付帯する商品の購買履歴などの付帯情報を総合的に分析することを検討したい。

謝辞：香川高等専門学校、宮崎耕輔准教授、および鳥取大学谷本圭志教授には、データの取得、及び整備において多大なるご協力を賜りました。また、本研究に際し、ログの提供を頂きました高松琴平電気鉄道に心より感謝いたします。

参考文献

- 1) 株式会社矢野経済研究所：交通系 IC カードに関する調査を実施（2017 年）
<https://www.yanoict.com/summary/show/id/490>, 2017.
- 2) 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 とりまとめ概要, pp. 110-119, 1996.
- 3) 総務省：平成 29 年度 IoT おもてなしクラウド事業について, 2017
- 4) 国土交通省：第 1 2 回 大都市交通センサス調査（概要版）, 2017.
- 5) 加藤浩徳：大都市交通センサスの調査手法上の課題と改善の方向性：特に鉄道輸送データの観点から（特集 交通統計調査のいま）, 運輸と経済 72(6), pp. 31-38, 運輸調査局, 2012.
- 6) 国土交通省：鉄道統計年報 [平成 24 年度], http://www.mlit.go.jp/tetudo/tetudo_tk6_000036.html
- 7) 国土交通省：鉄道統計年報 [平成 25 年度], http://www.mlit.go.jp/tetudo/tetudo_tk2_000027.html
- 8) 国土交通省：鉄道統計年報 [平成 26 年度], http://www.mlit.go.jp/tetudo/tetudo_tk2_000038.html
- 9) 国土交通省：鉄道統計年報 [平成 27 年度], http://www.mlit.go.jp/tetudo/tetudo_tk2_000040.html

RESEARCH FOR INEQUALITY OF NUMBER OF PASSENGERS CARRIED BETWEEN STATIONS WITH TRANSPORT IC CARD DATA

Hironori KATO, Masaki ITO and Kaoru SEZAKI

Smart card systems gather big quantities of traveller informations. In this study, we analyzed train passenger data acquired from transportation smart card “IruCa”, which is managed by TAKAMATSU-KOTOHIRA ELECTRIC RAILROAD Co. , Ltd. (in Kagawa prefecture) . With “Gini coefficient”, which is a measure of statistical dispersion, for measuring the inequality among boarding counts by stations, this paper classify stations.

Road traffic census, which is surveyed by Ministry of Land, Infrastructure, Transport and Tourism in Japan, collects data that has no unique ID. This paper makes frequency distribution table for boarding counts with passenger data associated with unique ID. By matching the Gini coefficient about stations with the average boarding time and the total passenger number, it makes possible to classify each station as “transit station”, “downtown” and “business town”, etc.