

ICカードデータを用いた機械学習による 停留所別降車人数推計手法の研究

渡邊 悠介¹・中村 俊之²・宇野 伸宏³・Jan-Dirk Schmöcker⁴・山崎 浩気⁵

¹学生会員 京都大学大学院工学研究科 (〒615-8540 京都府京都市京都大学桂C1-2-437)

E-mail:watanabe@trans.kuciv.kyoto-u.ac.jp

²正会員 京都大学大学院工学研究科 (〒615-8540 京都府京都市京都大学桂C1-2-434)

E-mail:nakamura@trans.kuciv.kyoto-u.ac.jp

³正会員 京都大学経営管理大学院 (〒615-8530 京都府京都市京都大学桂C1-2-436)

E-mail:uno@trans.kuciv.kyoto-u.ac.jp

⁴正会員 京都大学大学院工学研究科 (〒615-8540 京都府京都市京都大学桂C1-2-436)

E-mail:schmoecker@trans.kuciv.kyoto-u.ac.jp

⁵正会員 京都大学大学院工学研究科 (〒615-8530 京都府京都市京都大学桂C1-2-434)

E-mail:yamazaki@trans.kuciv.kyoto-u.ac.jp

交通系ICカードデータは料金支払いのシステムとして導入され、全国で普及が進んでいる。その一方で、システムから収集されるデータ(ICカードデータ)は、利用時間や乗降実態等、人々の行動把握が可能な貴重なデータである。一方で、交通事業者にとって、収集されたデータが十分に活用されているとは言い難い。具体的な事例の1つとして、バス交通において、乗車時もしくは降車時のみICカードを機器にタッチして料金を支払う場合には、乗車停留所、降車停留所のいずれかのデータしか取得できず、利用者がどの停留所で降車(乗車)したのか、さらには利用者の起終点情報については把握できない。

本研究では、その問題点に対して、乗車時にしかデータが収集されていないとの仮定のもとに、停留所別降車人数の推計を試みた。推計には、近年大量かつ継続的なデータに対して適用されることが多い機械学習による手法を用いた。乗車規模別に推定を行った結果、現状で乗降停留所が一致していると仮定した場合と比較して、誤差率に改善がみられた。

Key Words : smart card data, uniform rate system, machine learning, bayesian network, neural network

1. はじめに

鉄道・バスの交通事業者にとって、ダイヤ改正や路線計画等の交通施策を策定及び実施するにあたり、乗降者数や利用者の乗車・降車駅や停留所、起終点駅や停留所、利用時間等を把握することは必要不可欠である。これまで、改札通過による移動実態の把握や都市交通センサ調査、乗り込み調査等で把握を試みてきた。そうした中で、交通系 IC カードデータを活用することでデータ収集可能な環境が整ってきている。IC カードの発行枚数は年々増加しており、2013年3月時点での発行枚数合計は8700万枚にまで増加している¹⁾。

交通事業者によって収集される IC カードデータはどの程度活用されているのかに関して、今井ら²⁾ (2012) が全国 58 事業者に行ったアンケート調査がある。その調査結果では、半数以上の事業者が乗降車数等の流動量

把握に利用と回答している。しかし、データ活用に至っているのは鉄道事業者が大半を占め、大手を除き、バス事業者では活用に至っていないのが実態である。その理由として、収集されるデータ量が飛躍的に大きく、そのデータを分析可能な者が社内にはいないこと、さらには、そもそもデータ活用方法がわからないこと、折角収集しているデータについても、データが活用できる形になっていないとの回答が多くを占めていた。

IC カードによるデータにより、収集すれば直ぐに、乗降者数や利用者の乗車停留所、降車停留所、利用時間等のデータが出来上がるのかと言えば、必ずしもそうではない。バス交通のうち、東京都バスや京都市バスで採用しているような均一料金制度の場合、乗車時もしくは降車時のみ IC カードを機器にタッチするシステムを多くの事業者が採用している。この結果、乗車停留所、降車停留所のいずれかのデータしか取得できず、どの利用

者がどの停留所で降車（乗車）したのか、さらに起終点までは把握できないのが現状である。

上述の問題に対して、本研究ではバス交通において、乗車時のみカードをタッチすると仮定し、乗車停留所のデータが IC カードより収集されたときの降車停留所を推計する手法を構築することを目的とする。

なお、本研究は、停留所別の利用人数を集計的に捉え、停留所別降車人数を推計することとする。もちろん、IC カードデータには、カード固有の ID 情報を有していることが多く、個人単位で乗車停留所に対して、降車停留所を推計し、個人を積み上げて停留所単位とする方法もあり得る。しかしその方法では、利用頻度の低いカードユーザーの影響により推計精度が低くなる研究³⁾が報告されており、研究結果を事業者が利用する場合には、停留所単位で乗車人数、降車人数の集計値が求められているとの見解からである。

本研究の遂行にあたっては、近年大量かつ継続的なデータに対して適用されることが多い機械学習による手法を用いることとする。利用特性の分析を行った研究を参考に、停留所ごとに推計に有効であると考えられる変数を抽出し、その変数をもとに機械学習によって停留所別降車人数の推計する。

2. 既往研究と本研究での着眼点

IC カードデータによる利用実態を捉えた研究は近年数多く存在している。そのような中で、利用頻度に着目したものとして、岡村ら⁴⁾(2012)の研究がある。この研究では、都市圏の対距離制を採用しているバス路線を対象に、利用特性の分析を行った。結果として、高頻度利用者は朝夕に、低頻度利用者は日中に利用が多いことを明らかにしている。また、IC カードデータのカード固有 ID を用いて降車停留所を推計した研究として、Trepanier ら³⁾(2007)の研究が存在する。オタワ市バスを対象に、乗車日時、乗車停留所、カード ID のデータを用いて降車停留所を推定するモデルを構築した。全データ中、モデルが適用できたのは 66%であり、時間帯別で見ると、6~8 時台には 80%に至った。モデル適用が可能なユーザーは 66%に留まっており、この結果が本研究で、降車人数を集計的推計として試みるようにさせた所以である。

次に、土木分野において、収集されたデータに対して、機械学習を用いた事例を紹介する。奥嶋ら⁵⁾(2002)は、高速道路における流入交通量予測のリアルタイム運用を考慮する上で、ニューラルネットワークを用いている。予測精度は問題ない範囲に収まったが、モデル構築に平日データのみを使用する等、制約条件があるため、より

一般的なモデル構築には改善が必要である。モイヌルら⁶⁾(2012)は、高速道路におけるリアルタイム交通事故予測モデルの開発のために、ベイジアンネットワークを用いた。しかし、実際はオンランプ近傍での予測可能性を示すまでに留まっている。

上記の既往研究を踏まえた上で、これまでバス交通においても、IC カードデータを用いることで、大量かつ連続的に収集されていることから、機械学習の適用を試みる。機械学習のうち、ニューラルネットワークやベイジアンネットワークの 2つの手法を用いることで、どの程度の精度での推計が可能なのか、推計を行う際に用いられる変数の違い等を考察する。

3. 分析対象地域及びデータ概要

(1) 分析対象地域

本研究は静岡県静岡市を分析対象地域とする。この地域では、しずてつジャストライン株式会社により、路線バスが運行されている。複数の事業者が存在していないことから、停留所別利用者はしずてつジャストラインのバスにより収集されるデータで包括できる。

なお、対象地域内には、複数のバス路線が通る利用者数が多い停留所から、単一路線で運行本数や利用者数も少ない停留所まで、様々な特徴を持つ停留所が存在している。そのため、将来的に他地域での展開を考慮すると、学習用データとしても適当であると考えられる。

(2) 利用 IC カードデータ

本研究ではしずてつグループが導入している IC カード「LuLuCa」により収集されるデータを利用する。表-1にカード種別、データ期間、データ数、データ項目を示す。1年間の分析期間中の登場カード ID 数は 137,936 である。収集データ項目としては、利用金額データも存在しているが、他都市で収集されているデータ項目を踏まえて、本研究では利用金額データは使用しないこととした。

表-1 利用データ概要

項目	内容
カード種別	LuLuCa パサール/プラス ※公共交通利用が可能なICカード
データ期間	2013年9月1日~2014年8月31日 (1年間)
取得データ数	16,181,546
データ項目	カードID, カード利用日, 利用金額, 乗車時刻, 乗車停留所, 降車停留所

4. 停留所別降車人数推計手法の構築

本章では、実際に停留所別降車人数推定手法の内容と手順の概要について述べた上で、実際に推計の前段階の変数の抽出、基礎分析の結果について説明する。

(1) 停留所別降車人数推計手法の概要

停留所別降車人数の推計手法を図-1の手順に沿って説明する。

最初のステップは、停留所別の降車人数に影響を与えると考えられる変数を抽出する段階である。推定を行う停留所別降車人数に影響を与えると想定される乗車人数や頻度別利用実態、駅や学校に隣接した停留所か否か等の停留所が持つ位置に起因する情報を想定する。

次のステップは、抽出した変数に関する基礎分析の段階である。推計にあたって従属変数となる降車人数に対して、抽出した変数がどのような影響を有しているのかについて把握することを目的としている。具体的には、各独立変数とのクロス集計により特徴を捉える。

最後のステップは、抽出した変数を用いて停留所別降車人数の推計を行う段階である。前述のように推計手法としては、ニューラルネットワークとベイジアンネットワークを用いる。

ここで各手法の特徴の1つとして、ニューラルネットワークは連続変数を扱うことができるが、ベイジアンネットワークは離散変数しか扱うことができないことが挙げられる。ニューラルネットワークでは具体的な推計値が得られるのに対して、ベイジアンネットワークでは幅をもった推計値が得られる。

ベイジアンネットワークを用いて、精度の高い推計値を得るためには、従属変数を細かくカテゴライズする必要があることが知られている。しかし、本研究において、従属変数として設定する降車人数は、停留所別に大きな差異が生じており、カテゴリ数が多くなりすぎてしまう。そこでベイジアンネットワークによる推計では、カテゴリ数を少なくするために、従属変数として乗車人数と降車人数の比（降車人数を乗車人数で除したもの、以下乗降比）を設定する。ニューラルネットワークによる推計にあたっては、上述のような問題は生じないことから、降車人数を直接従属変数とする。

(2) 推計に用いる独立変数の抽出

独立変数としては IC カードデータより集計される動的な変数と停留所が個別に有する静的な変数を定義する。

動的な変数は、利用するデータ期間やエリアにより、その値が異なる変数である。IC カードデータから抽出

される動的な変数として、乗車人数、低頻度利用者割合、高頻度利用者割合、朝時間帯利用割合、夕時間帯利用割合、平日利用割合、片道利用者割合の7変数を抽出した。これらの変数の定義を表-2に示す。

静的な変数は、交通事業者が路線や時刻表等のサービスレベルの更新作業を行わないければ不変の変数であり、停留所に停車する路線数やバス停留所の位置に関する情報を含むものである。静的な変数として、路線重複数、一方向停車ダミー、始バス7時ダミー、終バス20時ダミー、最近隣停留所間距離、そして位置ダミーとして、駅、商業施設、学校、病院、住宅地、レジャー施設、役所の12変数を抽出した。これらの変数の定義を表-3に示す。

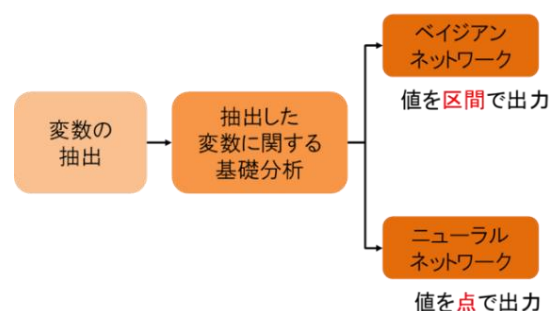


図-1 停留所別降車人数推計の手順

表-2 動的な変数の定義

変数名称	変数の定義
乗車人数	データ期間中の乗車人数の計
低頻度利用者割合	各停留所を利用したIDのうち、利用回数が月5回以下のID割合
高頻度利用者割合	各停留所を利用したIDのうち、利用回数が月20回以上のID割合
朝時間帯利用者割合	乗車人数のうち、5-8時台での乗車人数が占める割合
夕時間帯利用者割合	乗車人数のうち、17-23時台での乗車人数が占める割合
平日利用割合	乗車人数のうち、平日（祝日除く）の乗車人数が占める割合
片道利用者割合	各停留所を利用したIDのうち、往復利用せずに同一日に片道（1回）の利用であるIDの割合

表-3 静的な変数の定義

変数名称	変数の定義
路線重複数	各停留所を走行するバス路線の数
一方向停車ダミー	各停留所に対して、往路・復路で停車している停留所であるか
最近隣停留所間距離	各停留所に対して、直線距離で最も近くに存在する停留所との距離
位置ダミー	停留所の設置位置に基づくダミー変数

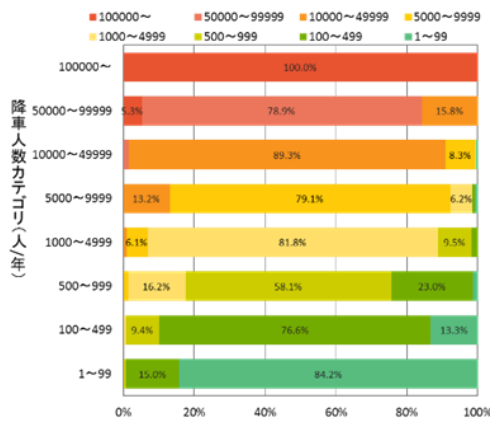


図-2 降車人数カテゴリ別の乗車人数カテゴリ構成比

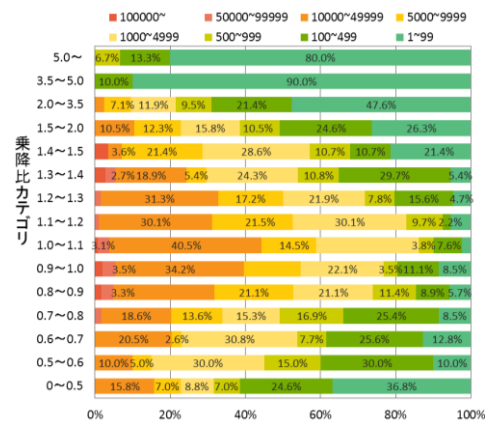


図-3 乗降比カテゴリ別の乗車人数カテゴリ構成比

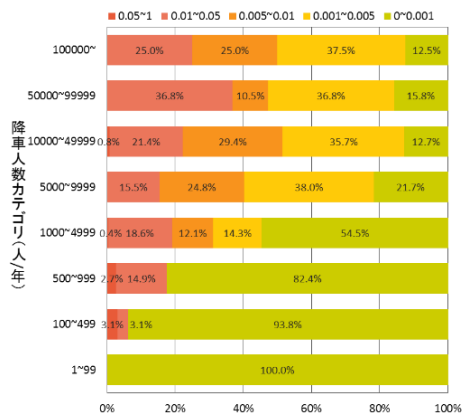


図-4 降車人数カテゴリ別の高頻度利用者割合カテゴリ構成比

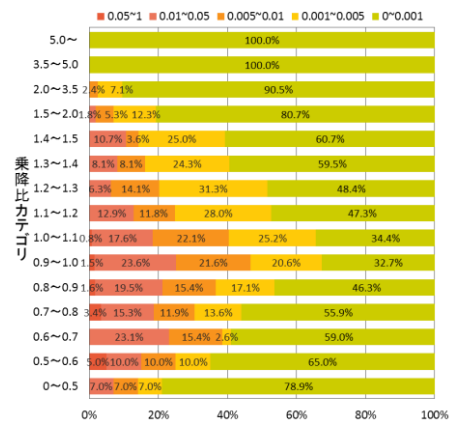


図-5 乗降比カテゴリ別の高頻度利用者割合カテゴリ構成比

(3) 変数に関する基礎分析

図-2は降車人数カテゴリ別の乗車人数カテゴリ構成比であり、この結果からは、往路と復路で同一の停留所を利用している利用者の存在により、停留所の乗車人数と降車人数は、近い値になる傾向が確認できる。また、図-3は乗降比カテゴリ別の乗車人数カテゴリ構成比であり、比率をとった場合、乗車人数が多い停留所ほど、計算上規模の影響を受け、乗降比は1に近い値をとる傾向を示している。

図-4は降車人数カテゴリ別の高頻度利用者割合カテゴリ構成比であり、高頻度利用者割合が小さい停留所では降車人数が少なくなる傾向、すなわち日常的に利用している利用が多く存在する停留所ほど、降車人数が多くなることを示している。また図-5は、乗降比カテゴリ別の高頻度利用者割合カテゴリ構成比であり、高頻度利用者割合が小さい停留所ほど乗降比は1から離れた値をとる傾向、これは図-4で示している結果と同様である。

なお、本論文内では、一部のクロス集計の結果のみを示しているが、機械学習における推計にあたって、利用する全ての独立変数と従属変数間でクロス集計は実施し、独立変数ごとに従属変数に対して、特定の影響を確認した。

5. 停留所別降車人数推計手法の適用結果

前章にて示した手順のうち、機械学習手法により、実際の推計を行う。まず、推計手法の概説と前提条件を述べた上で、推定結果を示す。

(1) 推計手法の概説

ニューラルネットワークは、人間の脳神経細胞の情報処理過程をモデル化したものである⁷⁾。変数間の因果関係は把握できないが、複雑な非線形関係を表現することが可能である。そして連続変数と離散変数の両方を扱うことができ、変数の数に制限がないことが一般的な特徴である。

ベイジアンネットワークは、複数の変数間の定性的な因果関係をグラフ構造によって表し、個々の変数間の定量的な関係を条件付確率で表すモデルである⁸⁾。有向グラフ構造から、変数間の依存関係が表現可能である。さらに、経験や知識などから主観的にグラフ構造を決定することで、より柔軟にモデルを構築することができる。また、変数の数に制限はないが、連続変数を扱うことができないことが一般的な特徴として挙げられる。

本研究では、独立変数と従属変数間の相関分析によっ

表-4 推計結果のまとめ

乗車人数規模	比較対象	平均差	平均誤差率	RMSE
1~99 人	乗降車数一致	138.9	78.2%	1121.1
	ニューラルネットワーク	51.3	239.1%	107.9
	ベイジアンネットワーク	127.3	80.3%	1085.8
100~999 人	乗降車数一致	190.0	58.8%	456.5
	ニューラルネットワーク	157.5	55.0%	339.7
	ベイジアンネットワーク	171.3	37.7%	363.3
1,000~9,999 人	乗降車数一致	977.6	28.6%	1810.7
	ニューラルネットワーク	913.4	27.1%	1467.6
	ベイジアンネットワーク	713.1	17.9%	1633.3
10,000~59,999 人	乗降車数一致	3230.3	18.5%	4911.6
	ニューラルネットワーク	3173.3	18.6%	4623.4
	ベイジアンネットワーク	2627.7	15.2%	4798.9

$$\text{注) 平均差} = \frac{\sum | \text{推計降車人数} - \text{実測降車人数} |}{\text{停留所数}}$$

$$\text{平均誤差率} = \frac{\sum \frac{| \text{推計降車人数} - \text{実測降車人数} |}{\text{実測降車人数}}}{\text{停留所数}} \times 100$$

て有意な相関が見られた変数は必ず従属変数の親になるという条件を設定し、ベイジアンネットワークモデルを構築する。

(2) 推計時の前提条件

既に第4章にて前述した通り、ニューラルネットワークでは降車人数を、ベイジアンネットワークでは乗降比を従属変数として設定する。この際、乗降比は図-3、図-5においても示した15のカテゴリに分け、各カテゴリの中間値を乗車人数に掛けることで降車人数を推計を行う。

本研究における推計値の比較対象は、全利用が同一停留所間での往復利用を仮定し、降車人数を推計したモデル（以下乗降車数一致モデル）を設定する。全利用者の乗降車が往復利用を仮定したモデルでもデータが大量かつ長期間蓄積することで、利用者の特徴の把握がある程度は可能である。機械学習を用いることで、その状況よりも精度良く推計を行うことができるのかを見極める必要がある。

また、乗車人数の規模によって停留所の特徴が異なると考え、本研究では停留所を乗車人数規模によって4つのカテゴリに分類し、推計を行った。

(3) 推計手法の適用結果

4つの乗車人数規模に対してニューラルネットワークとベイジアンネットワークを適用した結果を表-4に示す。この結果から、乗車人数規模 1~99 人での平均誤差率と乗車人数規模 10,000~59,999 人のニューラルネットワークにおける平均誤差率を除き、乗降車数一致モデルと比較して推計精度は改善されていることが確認できる。

乗車人数規模 1~99 人での平均誤差率のみ乗降車数一致モデルよりも悪い結果となった理由は、外れ値に関する学習の影響が考えられる。乗車人数 95 人に対して降

車人数 12,484 人という停留所が存在しており、学習データとしてこのデータを採用することで、推計結果が大きくなる方向に影響を及ぼしたと考えられる。また、ベイジアンネットワークよりも、ニューラルネットワークによる平均誤差率の方が大きく外れた結果となった原因は、乗降比の上限値が 5 に設定してあることが影響していると考えられる。

次に推計値と実測値の散布図によって考察として、乗車人数規模 100~999 人（図-6）と 10,000~59,999 人（図-7）での結果についての考察を行う。

乗車人数規模 100~999 人モデルでの推計結果分布を示したものが図-6 である。この散布図からは、乗降車数一致モデルと比較して、ニューラルネットワークモデルとベイジアンネットワークモデルでの推計値はより 45 度線に近い値が多く存在していることが確認できる。また、ニューラルネットワークモデルでの推計値は、ベイジアンネットワークモデルのものと比較して、実測値よりも大きな推計値が多く存在していることも見て取れる。

乗車人数規模 10,000~59,999 人モデルでの推計結果分布を示したものが図-7 である。この散布図からは、乗降車数一致モデル、ニューラルネットワークモデルでの推計値と比べて、ベイジアンネットワークモデルでの推計値はより 45 度線に近いものが多い結果となる一方で、45 度線から大きく外れた推計値も多くことが確認できる。

以上の結果より、各モデルの特徴を整理する。乗降車数一致モデルは、停留別にデータの集計のみを行えば良く、データ処理も簡単であり、簡潔で分かりやすいが今回の機械学習の結果と比較して、精度の面で劣ることがわかった。

今回適用した機械学習を用いた手法について、まず、ニューラルネットワークモデルは、低精度な推計値が少なく、RMSE では良い結果が得られたものの、高精度な

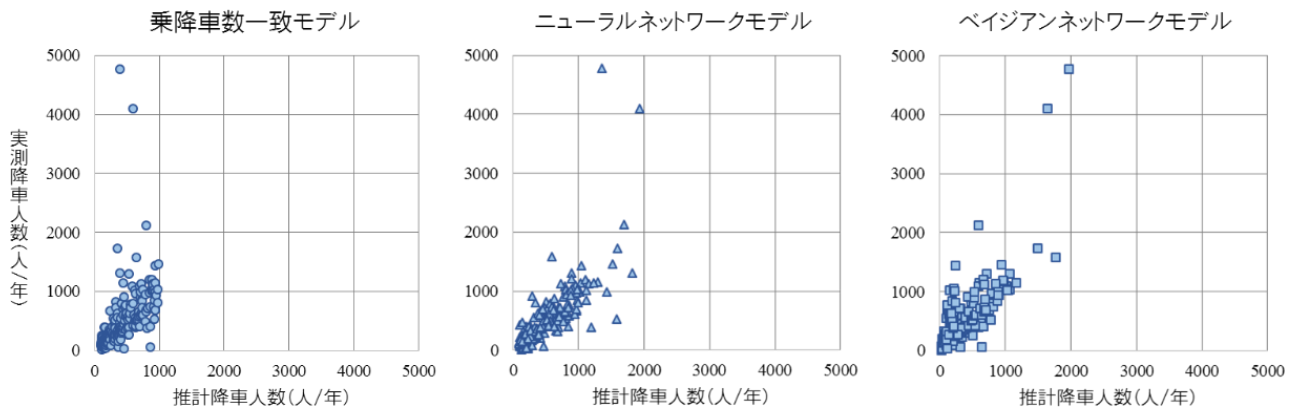


図-6 乗車人数規模 100~999 人モデルの推計結果分布図

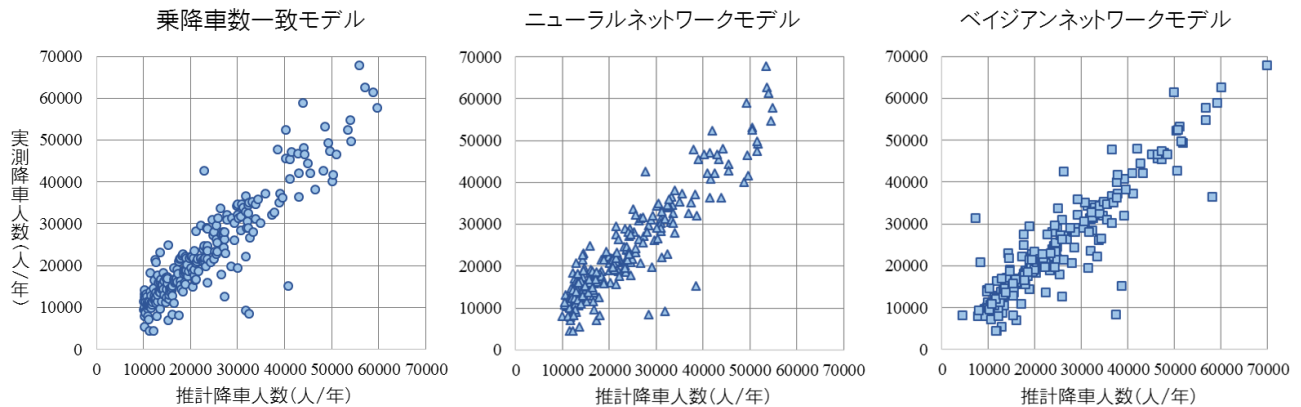


図-7 乗車人数規模 10,000~59,999 人モデルの推計結果分布図

推計値も少ない。ベイジアンネットワークモデルは、高精度な推計値が多く、平均差と平均誤差率では良い結果が得られたものの、低精度な推計値となることも多い結果となった。このように、各モデルには一長一短が存在することから、推計時に用意できる変数や乗車人数規模に応じて、現状ではモデルを使い分けることが求められる。一方でベイジアンネットワーク適用時には利用変数間の関係が把握できる結果として把握できるのに対して、ニューラルネットワーク適用時には変数間の関係は分析者でも把握ができない。すなわち得られた結果を用いたその影響を説明するという場合にはベイジアンネットワークを適用せざるを得ないことも付与しておく。

6. 結論

本研究では、IC カードより得られる大量かつ長期間のデータを利用して、バス乗車時のみデータが収集されている、すなわち乗車人数のみ把握可能な状況を仮定して、降車人数を機械学習による手法を用いて推計を行った。

本研究より得られた成果を以下にまとめる。推計手法としてニューラルネットワークとベイジアンネットワー

クを用いたが、乗降車数一致モデルと比較して両手法による推計精度は良い結果が得られた。そのため、乗車人数のみ把握可能な状況下で、機械学習を用いて適用することで、一定の精度下で降車人数を推計することが可能となった。

ニューラルネットワークによる推計では、乗車人数と降車人数に大きな差があるような停留所でも、サンプルの学習によって比較的精度の良い推計が行えることが明らかになった。また、降車人数が多い停留所では、低精度な推計値が少ないことが示唆された。ベイジアンネットワークによる推計では、降車人数が少ない停留所で比較的高精度な推計値が得られる結果となった。

本研究の遂行にあたって、分析の都合上、乗車人数規模によって 4 カテゴリーに分類したが、今回の研究では分析者が恣意的な数値を用いて分類を行った。しかし、クラスター分析を適用し、分類するなど、今後は理論的な根拠に基づいた分類方法を採用し、結果への影響を捉えることが必要となる。

また、モデル推計時に利用した変数についても、本研究で抽出した変数以外にも変数候補は存在するため、変数と推計精度の関係についても更なる研究として探求すべきである。

今回モデル構築時とモデル検証時のサンプルとしてと

もに静岡市の停留所データを使用しているため、精度が高めに出ていることが懸念される。今後料金均一制のバスが運行しており、乗車時もしくは降車時にしかデータ収集されていない東京都バスや京都市バスのエリア等の他地域へ本手法を適用することを考慮すると、諸地域の様々な特徴をもったサンプルを学習させることが望まれる。

謝辞

本研究の実施に際して、データを提供いただきました静岡鉄道株式会社に謝意を表します。また、個人情報保護の観点で配慮したデータを提供いただきましたことを付記致します。

参考文献

- 1) 国土交通省：鉄道における交通系 IC カードの導入状況について (2015.7.7 時点)
URL : http://www.mlit.go.jp/tetudo/tetudo_ik6_000015.html
- 2) 今井龍一, 井星雄貴, 中村俊之, 森尾淳, 牧村和彦, 濱田俊一 : 交通系 IC カードから取得できる動線データの活用に向けた考察～全国の交通系 IC カード取扱事業者への実態調査

- から得た知見～, 土木計画学研究・講演集, CD-ROM, Vol.45, 2012
- 3) Trepanier, M., Chapleau, R., Tranchant, N. : Individual trip destination estimation in a transit smart card automated fare collection system, Journal of Intelligent Transportation System, 2007
 - 4) 岡村敏之, 中村文彦, 小幡慎二, 王鋭 : IC カード記録に基づく都市内路線バスの利用特性分析, 題 45 回土木計画学研究発表会, 2012
 - 5) 奥嶋政嗣, 大窪剛文, 大藤武彦, 土田貴義, 秋山孝正 : ニューラルネットワークを用いた都市内高速道路における時系列流入交通量予測の適用性の検討, 土木計画学研究・講演集, CD-ROM, Vol.45, 2002
 - 6) モイヌルフセイン, 室町泰徳 : ベイジアンネットワークによるリアルタイム交通事故予測モデルの開発, 交通工学, Vol.47, No.2, pp.39-44, 2012
 - 7) 大内東, 山本雅人共著, 川村秀憲 : マルチエージェントシステムの基礎と応用-複雑系工学の計算パラダイム, (株) コロナ社, 2002
 - 8) 木村陽一, 岩崎弘利著 : ベイジアンネットワーク技術, 東京電気大学出版局, 2006

(2015. 7. 31 受付)

ESTIMATION OF PASSENGERS ALIGHTING AT A BUS STOP BASED ON SMART CARD DATA USING MACHINE LEARNING

Yusuke WATANABE, Toshiyuki NAKAMURA, Nobuhiro UNO,
Jan-Dirk SCHMOECKER and Hiroki YAMAZAKI