

IC カードデータに基づく継続性と利用頻度に着目した利用日数予測モデルの構築に関する研究

高知工科大学 竹内 淑貴
指導教員 西内 裕晶

1. 背景と目的

現在、日本では公共交通機関の利用者数は減少傾向にある。特に、地方部ではモータリゼーション、少子高齢化、大都市への人口流動に伴う人口減少によりその影響が顕著である。また、利用者数の減少による公共交通機関の利益の低下や、運転手の担い手不足の影響による運賃の値上げや便数の減少といったサービスレベル、利便性の低下により、さらなる利用者数の減少が懸念される。こうした負のスパイラルと呼ばれる循環が地方部での公共交通機関の課題である。公共交通の負のスパイラルから脱却するためには、持続的な公共交通機関が不可欠であり、利用者の行動特性を考慮し、利用促進やサービス改善するための提案が必要である。そこで、本研究では約 14 年間の移動履歴が蓄積された IC カードデータを活用し、継続性や利用頻度に着目した利用日数の予測モデルの構築や継続性に影響している特徴を把握することで、継続性や利用頻度の高い利用者がどのような交通行動をしているかを明らかにすることを目的とする。

2. 研究の概要

本研究では、高知県で発行されている IC カード「ですか」の約 14 年間のデータを使用する。表 1 に使用する IC カードデータの概要を示す。今回使用するデータは 2009 年に発行された 47,316 件の ID とし、その ID を各月における 5 段階の頻度（利用なし 0・低頻度 1・中頻度 2・高頻度 3・最高頻度 4）に分類する。分類手法は月利用日数、月利用回数を説明変数としたクラスター分析を用いて算出した。長期的な継続性を追うために、各利用者について 13 年間の各月のクラスター番号の平均値を算出した。これを平均クラスター値として利用頻度段階の指標とした。表 2 に分類した頻度段階とサンプル数を示す。

表 1 使用する IC カードデータの概要

提供者	とさでん交通株式会社
期間	2009 年 1 月 25 日～2023 年 3 月 31 日
カード種別	大人記名、大人無記名、大人身障、ナイスエイジ、身障介護、小児、小児身障
情報	利用日、カード ID、利用交通機関、乗車時間、降車時間、乗車停留所、降車停留所、利用区分、乗車距離

表 2 13 年間の平均クラスター値に基づく利用分類

平均クラスター値	13 年後までの利用頻度	サンプル数(件) N=47,316
0～1	長期的な低頻度以下	32,718
1～2	長期的な低頻度	3,378
2～3	長期的な中頻度	651
3～4	長期的な高頻度	92

3. 研究手法

本研究では、IC カードデータを用いて、各 ID の未来の月利用日数を予測した上で、頻度段階に応じた影響のある特徴を把握する。利用日数の予測には、決定木と勾配ブースティングを組み込んだ機械学習の一種である Light GBM (Light Gradient Boosting Machine) という手法を用いた。算出した予測値の精度は、平均絶対誤差 MAE (Mean Absolute Error) を用いて評価する。学習データの期間は、どの ID でも最低 1 年間は学習できるように 2010 年 1 月～2010 年 12 月までを学習期間とし、2011 年 1 月の利用日数を予測した。また、予測が可能な期間を明らかにするために、予測した結果も学習に加える逐次学習を行い、2023 年 3 月までの月の利用日数を予測し、全ての期間において MAE を算出して評価をした。さらに、頻度段階ごとの利用時間および交通手段選択（バス・電車）に関する利用特性を把握するために、PDP (Partial Dependence Plot) を用いた。PDP とは特徴量の値が変化したときに、定めたラベルごとにモデル出力の平均的な変化を折れ線グラフで可視化したものであり、PDP のグラフは学習データとテストデータを変化させたときに、グラフの形が一定でなければならないという性質をもつ。そこで、学習データとテストデータをランダムに分割し、曲線の形・曲線の大きさ・変化の割合を示す値が収束するまで複数回グラフを作成し、それらの一致率による評価をした。

キーワード IC カードデータ、頻度、継続性、クラスター分析、機械学習、Light GBM、PDP、

4. 結果と考察

(1) 予測モデルの評価

予測モデルの説明変数を各 ID の 1 年間の月利用日数とその変化率の推移とし、逐次学習によって各頻度の予測した値を MAE で評価した。1 ヶ月ごとの結果の推移を図 1 に示す。図 1 より、1 年の学習期間で全ての頻度において 17 ヶ月目までは MAE の値 4 以下で予測することが可能である。17 ヶ月目以降には低頻度・中頻度の順に MAE の値の増加の傾向が変化している。また、コロナ期以前は MAE が低かった高頻度の ID もコロナ期以降は MAE が増加している。これより、約 10 年間高頻度で利用していても、パンデミックのような社会現象が発生すると利用を離脱することを明らかにした。MAE の値が増加するのは、利用開始からの 1 年間の学習をしているため、数年先の減少傾向を学習できないからだといえる。また、図 2 に示した平均月利用日数の推移と図 1 を比較すると、利用日数の減少と同時に MAE の値が上昇していることから減少傾向を学習できていないことがいえる。

(2) PDP での特徴量ごとのグラフ

この分析で扱う頻度段階の分類は表 3 に示すように（高利用者・低利用者）2 つのラベルとする。この分類は、表 2 で示した 13 年後の平均クラスター値の 2.5 未満を低利用者とし、2.5 以上を高利用者とした。また、特徴量の算出に使用する説明変数を表 4 に示す。PDP グラフの結果を図 3 に示す。図 3 の縦軸は高利用者に属する確率、横軸は変数の数値の変化である。時間帯に関しては、平日 5-9 時の時間帯利用割合が高い人、休日 18-23 時の時間帯利用割合が低くなる人ほど高利用者になることを示唆している。これは、朝の通学・通勤利用の影響や休日夜間に街に出かけない人が多いことが考えられる。移動距離に関しては、バス・電車の中程度の距離の利用者や家からの距離が長い人ほど高利用者になりやすいことを示唆している。これは、利用回数が多い人、乗り換えが多い人や普段から車を利用しない人が多いと考えられる。公共交通の選択に関しては、バスの利用割合が高い人ほど高利用者になることを示唆している。これは、目的地までの移動距離が長くて、車の利用が困難な人が多いことが考えられる。

5. まとめと今後の課題

本研究では、IC カードデータを用いて、継続性及び利用頻度に着目した月利用日数の予測モデルの構築と PDP を用いて特徴量を算出した。その結果、1 年間の学習をすることで、その後の約 2 年間の月利用日数を低い MAE の値で予測できることを明らかにした。以上より、約 2 年間の利用状況の把握が可能となり、利用が減少するタイミングに応じた利用促進対策の提案が可能であることが示唆された。また、交通手段選択の予測や個人行動の要因の特定は今回の分析手法およびデータの枠組みでは不可能である。今後は、交通行動の選択に関する予測モデルの構築や個人の利用行動に着目した特徴量の把握を行う必要がある。

参考文献

- 1) 西内裕晶：IC カードデータを活用したトリップ強度指標による公共交通利用継続特性の把握，第 66 回土木計画学研究・講演集，45-04，2022。

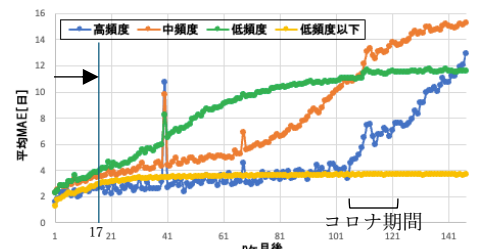


図 1 各頻度の平均 MAE の推移

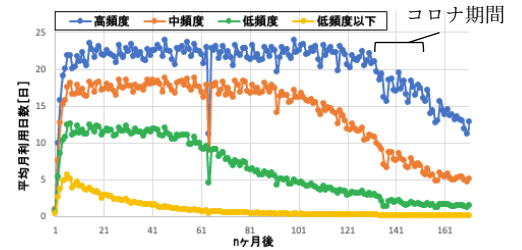


図 2 各頻度の平均月利用日数の推移

表 3 PDP の分析で使用するラベルの概要

平均クラスター値	ラベル	頻度	サンプル数(件) N=47,316
0~2.5	0	低利用	46,910
2.5~4	1	高利用	406

表 4 PDP の分析で使用する説明変数

説明変数
平日・休日時間帯別利用割合
バス・電車の利用割合
バス・電車の平均移動距離(km)
1日の最初に乗車する平均移動距離(km)
定期・定期外の利用割合

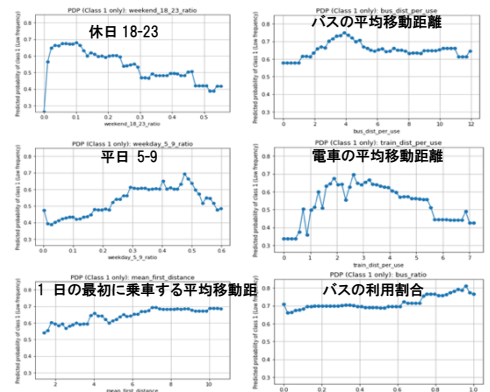


図 3 ラベル 1 の PDP グラフ