

時刻間の類似性を考慮した 出発時刻選択モデルの高分解能化

石嶋 悠嗣¹・柳沼 秀樹²・寺部 慎太郎³・海野 遥香⁴

¹ 学生非会員 東京理科大学大学院 理工学研究科土木工学専攻修士課程

(〒278-8510 千葉県野田市山崎 2641) E-mail: 7621502@ed.tus.ac.jp

² 正会員 東京理科大学 理工学部土木工学科 (〒278-8510 千葉県野田市山崎2641)

E-mail: yaginuma@rs.tus.ac.jp

³ 正会員 東京理科大学 理工学部土木工学科 (〒278-8510 千葉県野田市山崎2641)

E-mail: terabe@rs.tus.ac.jp

⁴ 正会員 東京理科大学 理工学部土木工学科 (〒278-8510 千葉県野田市山崎2641)

E-mail: unoharuka@rs.tus.ac.jp

従前の交通需要予測では、新規路線の整備や料金政策など都市圏レベルのマクロな評価を対象に、四段階推定法が適用されてきた。近年では、駅周辺の再開発やテレワーク等の個人レベルのマイクロなライフスタイルの変化を評価することが求められている。そのため、個人の活動レベルに着目した ABM(Activity-Based-Model)に基づく評価の必要性が高まっている。しかしながら、時間分解能が粗いため出発時刻選択に影響する施策の効果を適切に記述できない状況にある。本研究では、出発時刻選択モデルの高時間分解能化を行い、アクティビティモデルの精緻化を試みる。具体的には、時刻選択肢間の類似性を考慮した複数の離散選択モデルを適用して、高時間分解能とモデルの精度向上を図った。これまでの分析では把握できなかったアクティビティの時間軸上での変化を詳細に捉えることが可能となり、施策評価の高度化に寄与することが期待される。

Key Words: Activity Based Model, Discrete Choice Model, Departure Time Choice Model

1. はじめに

交通政策や地域拠点整備を行う際には、それらの効果や影響を評価するために四段階推定法を用いた交通需要予測が行われてきた。近年では、駅周辺の再開発やテレワーク等の個人レベルのマイクロなライフスタイルの変化を評価することが求められている。そのため、従来型のトリップベースモデルの需要予測では、これらの施策を適切に評価することは困難であり、個人の活動レベルに着目したアクティビティベースモデル(ABM)に基づく新たな評価手法の必要性が高まっている。

しかしながら、ABMにおいて、時刻選択は分解能が荒く出発時刻選択に影響する政策の効果を適切に記述できていない状況にある。特に近年では、フレックスタイムや時差出勤などの朝の時間帯の移動だけでなく、在宅勤務後における移動も増加している。そのため、局所的な時間帯だけでなく、1日全体を対象とした広範囲かつ詳細な時刻選択モデルを構築する必要がある。

本研究では上記の施策評価に向けた ABM の開発を念頭に、出発時刻選択モデルの高時間分解能化を行い、アクティビティモデルの緻密化を試みる。具体的には時刻選択肢間の類似性を考慮した離散選択モデルによる推定に加え、機械学習を応用したモデルの構築を行い出発時刻選択モデルの高分解能化と精度向上を図った。

2. 既往研究のレビューと本研究の位置付け

(1) 既往研究のレビュー

Bowman, Ben-Akivaら¹は、ボストンを対象とした ABM の構築を行なった。Nested Logit モデルにより各選択モデルを記述している。しかし、出発時刻選択モデルにおいては、選択肢が朝ピーク、昼、午後ピーク、夜の4つしかなく、分解能の荒さが問題である。

また、近年では我が国を対象とした ABM の開発も行われている。例えば、亀谷、福田ら²は東京都市圏の鉄道

利用者を対象とした ABM の構築を行なった。ABM は時間帯選択、目的地選択、経路選択、アクセスモード選択とし、Nested Logit モデルで記述している。しかし、Bowman, Ben-Akiva と同様に、時間帯選択モデルの選択肢は、朝、昼、夕、夜の 4 つしかなく分解能の荒さが問題となっている。

両者の時刻選択モデルでは時刻の選択肢が 4 つしかないため類似性を考慮していないが、分解能を上げると選択肢間の誤差相関が発生する可能性がある。藤原ら³⁾は、時刻選択を離散選択モデルで表現する際、時刻という連続量を離散値にカテゴリズすることによって類似性が発生することを明らかにしている。MNL モデルと PCL モデルを比較し、選択肢間の類似性を考慮できるモデルの有用性を示した。しかし、PCL モデルは隣接する 2 つの時間帯の類似性しか考慮せず、3 つ以上の選択肢は考慮できない。本研究では、2 つの時間帯だけでなく複数の時刻間の類似性を考慮するため、CNL を用いた出発時刻選択モデルの構築を行う。

(2) 本研究の位置付け

本研究では、時刻間の類似性を考慮した出発時刻選択モデルの高分解能化を行う。5時から21時の出発時刻を範囲とし、1時間、30分の分解能を持つモデルを構築する。多項ロジットモデル(MNL)による推定だけでなく、複数の時刻間の類似性を考慮できるクロスネスティッドロジットモデル(CNL)による推定を行い、どちらの構造が望ましいかどうかを比較する。またモデルの構築にあたっては、移動目的ごとにセグメンテーションし、各移動目的に適した説明変数を選択することで精度向上を目指す。さらに、出発時刻選択モデルの記述性能の向上を念頭に、統計的機械学習理論の適用を試みる。

3. データの概要と基礎集計分析

(1) データの概要

本研究では、H30 東京パーソントリップ調査データ⁴⁾ (以下、東京 PT データ) を用いた。東京 PT データは個人の平日 1 日に行なったトリップデータを収集している。東京を中心とする半径約 80km 圏域を対象とした調査であり、平成 30 年の 9-11 月に行われた。本研究では、この東京 PT データを用いて出発時刻選択モデルを構築する。

また、データ形成として出発時刻のカテゴリズを行なった。東京 PT データ内の出発時刻は回答者が直接時間を記述するため連続的なデータとなっている。そのため離散選択モデルの目的変数として使用する際には、それぞれの分解能に応じて時間帯を区切る必要があった。

表-1 東京 PT の移動目的分類

セグメンテーション	東京PT 移動目的
通勤	勤務先 (出勤・帰社), 通学先 (登校・帰校)
業務	打ち合わせ・会議・商談, 販売・配達・仕入れ・購入先, その他の業務
買い物	日用品の買い物, 日用品以外の買い物
私事	食事・社交, 文化活動, 通院・リハビリ, デイサービス, 他者の用事のつきそい, 他者の送り迎え, 塾・習い事・学習, 散歩・ジョギング・運動, 観光・行楽・レジャー, 地域活動・ボランティア, その他の私用
帰宅	自宅 (帰宅)

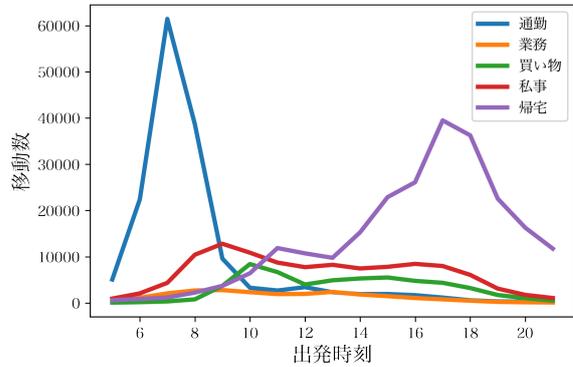


図-1 移動目的別のトリップ数

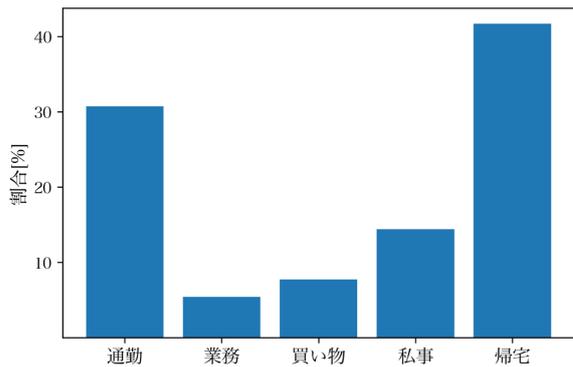


図-2 移動目的割合

例えば、1 時間単位のモデルを構築する際には、7:00~7:59 間に出発したトリップを全て 7 時発のトリップとして考え、他の時刻についても同じようにカテゴリズした。30 分単位のモデルも同様に、7:00~7:29 間に移動したトリップを同じ出発時刻としてカテゴリズした。

(2) 基礎集計分析

モデリングにあたり、基礎集計を行いデータの傾向を把握した。本研究では対象トリップを 20-64 歳または 65 歳以上の就業者 (以下、現役世代) に絞った。H30 東京 PT データにおいて現役世代が占める割合は約 7 割で、約 470000 トリップであった。また本研究では移動目的でセグメンテーションするため、東京 PT データの移動目的を表-1 のように 20 目的から 5 目的に集約した。それぞれの移動目的ごとに説明変数を選択することによって特性を考慮し、詳細なモデルの構築を図った。

全トリップにおけるそれぞれの移動目的割合、1 時間

ごとのトリップ数の推移を図-1、図-2に示す。移動目的の割合に関しては、通勤、帰宅の割合が多く全体の約6割を占めていることがわかる。時間帯別のトリップ数に関しては、通勤目的が朝のピーク時に集中しており、帰宅は14時から18時にかけて増加していることがわかる。業務や買い物、私事に関しては時間帯で大きな変化は見られなかった。

4. DCMに基づく出発時刻選択モデルの構築

(1) 出発時刻選択モデルの定式化

本研究では、離散選択モデル(DCM)のうち MNL と CNL を適用し、両者のパラメータ推定結果の比較を通じて選択肢間の類似性について考察する。

MNLにおいて、選択肢 n を選ぶ確率は式(1)で表される。

$$P(n) = \frac{\exp(V_n)}{\sum_l \exp(V_l)} \quad (1)$$

CNLにおいて、選択肢 n を選ぶ確率は式(2)で表される。

$$P(n) = \sum_m \frac{(\sum_n (\alpha_{mn} e^{V_n})^\lambda)^\lambda (\alpha_{mn} e^{V_n})^\lambda}{\sum_b (\sum_n (\alpha_{bn} e^{V_n})^\lambda)^\lambda \sum_l (\alpha_{ml} e^{V_l})^\lambda} \quad (2)$$

ここで $P(n)$ は時刻 n の選択確率、 V_n は効用関数の確定項、 λ はスケールパラメータ、 α_{mn} はアロケーションパラメータである。

本研究で仮定した CNL のネスト構造を図-3に示す。隣接する3つの時間帯に類似性があるとし、それらと同じネストに入れることで相関関係を表現した。また分解能を細かくしてもネスト構造は変わらず、3つの隣接時間帯を同一ネストに組み込んだ。そして、アロケーションパラメータ α_{mn} を用いることで、隣接時間帯間での類似性を表現することが可能になる。アロケーションパラメータ α_{mn} は各ネスト m のゾーン n に割り当てられ、各出発時間帯においてどのグループに所属するかの割合を示す。本研究においてアロケーションパラメータ α_{mn} は1つネスト内では等間隔に帰属すると仮定するため、全て1/3とした。またスケールパラメータ λ は全てのネストで同じ値をとると仮定した。

(2) 効用関数の特定化

本研究ではMNLとCNLを用いて出発時刻選択モデルのパラメータ推定を行う。効用関数には基礎集計から得られた結果と杉恵ら⁹⁾のモデルを参考に変数を選択した。具体的には移動時間、職業や性別などの個人属性をダミー変数として組み込んだ。移動時間は東京PTから各時

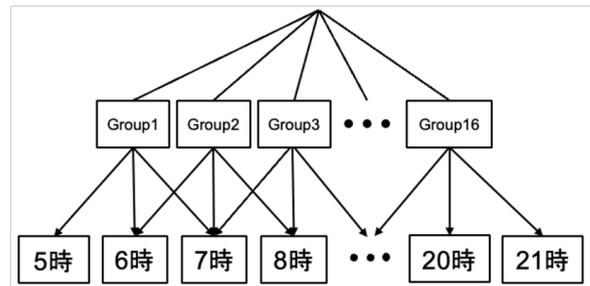


図-3 CNLのネスト構造

間帯の混雑率を計算し、期待移動時間として各時間帯の効用関数に導入した。ダミー変数に関しては基礎集計で得られた結果からトリップ数が多い時間帯の効用関数に導入した。またMNLとCNLでは同じ効用関数を用いているためモデル構造のみが異なっている。

選択肢集合については前後のトリップを加味し、活動時間と移動時間以外を選択肢集合に設定した。具体的には、2回目の行動の場合、1回目の移動時間と活動時間以外の時間帯から選択することとし、2回目以降もトリップがあればその活動時間と移動時間も選択肢集合から除外した。

(3) パラメータ推定結果と考察

推定結果を表-2、表-3に示す。まずMNLの推定結果に関しては、1時間レベルの時刻選択モデルではほとんどの変数が5%有意となった。また、移動時間のパラメータの符号も負であり条件を満たしている。尤度比を見てみる通勤目的で最も高い値となっており、これは通勤目的のトリップが朝の時間帯に集中していることが原因だと考えられる。その次に買い物目的の尤度比が高くなっているが、買い物に関しては前後に私事や帰宅などのトリップをしていることが多く、選択肢集合が狭まった結果高くなったと考えられる。30分単位のモデルを見てみると尤度比、的中率はどの移動目的でも大きく低下してしまい、変数も有意になっていないことがわかる。移動時間は有意で符号条件も満たしているが、ダミー変数は時刻の選択肢の増加に伴い有意性が失われてしまう結果となった。

CNLではどの移動目的に関するモデルでも類似性パラメータ λ が0から1の範囲に収まっており、ネスト構造に問題はないと考えられる。説明変数に関してはMNLと同様に、1時間単位のモデルにおいて有意ではあるが、30分単位の分解能になるとほとんどの変数が有意ではなくなる結果となった。

また、MNLとCNLの尤度比を比較してみると、通勤と帰宅ではCNLの方が高くなっているが、大きな改善はなく的中率もほぼ変化しない結果となった。このことから、類似性はあるものの精度を大きく改善するための

表-2 1時間単位モデルの推定結果

	通勤		業務		買い物		私事		帰宅	
	MNL	CNL	MNL	CNL	MNL	CNL	MNL	CNL	MNL	CNL
期待所要時間	0.470 **	-0.264 **	-0.340 **	-0.793 **	-0.131 **	-0.186 **	-0.298 **	-0.718 **	-0.175 **	-0.0752 **
勤務時間固定	0.546 **	0.678 **								
鉄道ダミー			0.454 **	1.07 **	0.779 **	2.65 **	0.347 **	2.43 **	0.424 **	0.298
女性ダミー	1.11 **	0.889 **	0.576 **	1.37 **			0.170	0.336		
正社員ダミー			0.455 **	2.45 **	0.802 **	3.11 **	0.466 **	1.12 **	0.757 **	1.95 **
主婦ダミー					0.215	0.848	0.338	0.779	0.265	0.804
パートダミー			0.464 **	0.991 **						
類似性パラメータ		0.510 **		0.137 **		0.280 **		0.298 **		0.201 **
初期尤度	-1932	-1848	-4825	-4810	-3437	-3433	-3355	-3323	-3867	-3804
最終尤度	-416	-383	-3366	-3117	-1223	-1187	-1536	-1491	-1857	-1762
修正済み尤度比	0.774	0.782	0.339	0.377	0.638	0.648	0.536	0.549	0.510	0.529
的中率	0.875	0.859	0.465	0.424	0.744	0.744	0.658	0.658	0.483	0.487

「**」:5%有意

表-3 30分単位モデルの推定結果

	通勤		業務		買い物		私事		帰宅	
	MNL	CNL	MNL	CNL	MNL	CNL	MNL	CNL	MNL	CNL
期待所要時間	-0.525 **	-0.427 **	-0.362 **	-0.592 **	-0.455 **	-1.66 **	-0.383 **	-0.589 **	-0.283 **	-0.345
勤務時間固定	0.031	0.150								
鉄道ダミー			0.692	0.249	0.563	2.59	0.605 **	0.595	0.506	0.488
都心発ダミー					0.395	2.16 **	0.209	0.498		
女性ダミー	0.0509	0.190	0.0223	0.135			0.224	0.443		
正社員ダミー			0.159	0.186	0.635 **	1.21	0.283	0.608	0.219	0.955
主婦ダミー					0.117	0.727	0.220	0.547		
類似性パラメータ		0.162 **		0.159 **		0.271 **		0.367 **		0.208 **
初期尤度	-1972	-1972	-2020	-2047	-2165	-2171	-2199	-2199	-2440	-2416
最終尤度	-1329	-1466	-1459	-1753	-871	-853	-1148	-1117	-1344	-1271
修正済み尤度比	0.307	0.237	0.269	0.144	0.584	0.588	0.459	0.472	0.434	0.458
的中率	0.309	0.320	0.274	0.270	0.590	0.596	0.478	0.476	0.340	0.358

「**」:5%有意

モデルとしては適正ではないことが示唆された。

5. 機械学習に基づく出発時間選択モデルの構築

MNLとCNLの推定結果を踏まえると、離散選択モデルによる時刻選択は困難であると考えられる。1時間単位のモデルでは的中率は高いが、30分単位になると説明変数も有意ではなくなり精度も下がることが明らかになった。もし仮に構築したモデルの説明変数が全て有意になったとしても、連続量が移動時間しかなくほぼダミー変数で表現されるモデルであるため解釈が困難である。

そこで本研究では出発時刻を予測する新たな取り組みとして機械学習を応用したモデルの検討を行う。具体的にはランダムフォレストとニューラルネットワークを用いて予測を行い、出発時刻選択モデルとして適用できるか否かの検討を行う。離散選択モデルと同様に1時間、30分単位の時刻選択モデルを構築し精度向上を目指す。

また特徴量の重要度を判断するために Permutation Importance を算出し、出発時刻を決定する要因を明らかにする。最後に離散選択モデルとの結果を比較し、再現性や精度についても検討を行う。

(1) ランダムフォレスト(RF)の概要

RFはアンサンブル学習の一つであり、Leo Breimanによって提案されたモデルである⁹⁾。ここでアンサンブル学習とは複数の機械学習モデルを組み合わせることでより強力なモデルを構築するやり方で、RFは複数の決定木を組み合わせることによって回帰やクラスタリングを行う機械学習手法である。決定木単体では機械学習しやすいため、この問題に対応するために提案されたモデルである。

RFは、パラメータとして決定木の数を設定する必要がある。本研究では決定木の数を1000本に設定し、それらの決定木の多数決により出発時刻の分類を行う。

(2) ニューラルネットワーク(NN)の概要

NN は脳の神経細胞（ニューロン）を模倣したものであり、図-4のようにモデル化される。1つのニューロンには、ニューロン同士の繋がり（強さ）を表す重みと、受け取った値を次層にどのように出力するかを決める活性化関数で構成され、これらを多数結合させたものが NN である。大きく分けると外部からの情報を受け取る入力層、上位の中間層または出力層に情報を送る中間層、外部に情報を送る出力層の3つで構成される。

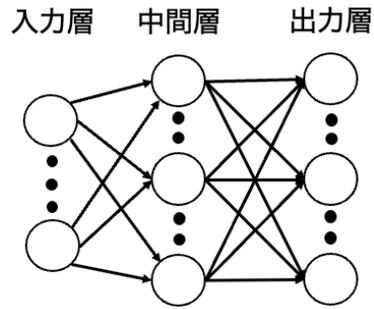


図-4 ニューラルネットワークのモデル化

本研究では、入力層のニューロン数は 13 個に設定した。入力層でのニューロン数は分析に用いる特徴量の数と一致しており、東京 PT データから選択した 13 個の特徴量に対応している。中間層に関しては、1 時間単位のモデルと 30 分単位のモデルで層の数とニューロンの数を変更した。1 時間単位では層の数を 3 つにし、それぞれ 50, 40, 30 個の順にニューロンを設定し、30 分単位においては層の数を 5 つとし 100, 100, 70, 60, 50 の順にニューロンを設定した。出力層でのニューロン数は出発時刻の選択枝数と一致するため、1 時間単位では 17 個、30 分単位では 34 個に設定した。また、活性化関数に関しては中間層で ReLU 関数、出力層には Softmax 関数を使用した。

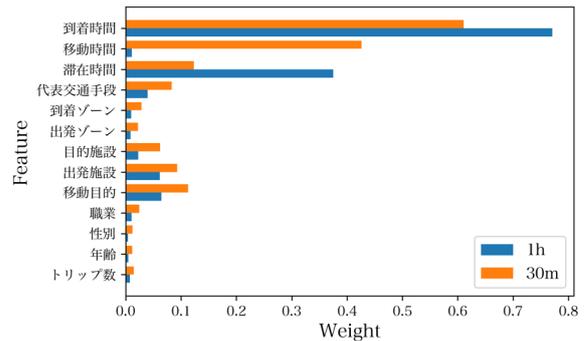


図-5 RFにおける特徴量の重要度

(3) 交差検証

本研究では K-分割交差検証を用いた評価を行う。交差検証は、過学習を防ぎ汎化性能を改善させるために使用される手法である。今回使用する K-分割交差検証はデータを K 個に分割してそのうち 1 つをテストデータにし、残りの K-1 個を学習用データとして正解率の評価を行う。本研究では分割数を 10 にして交差検証を行い、評価の検討を行なった。

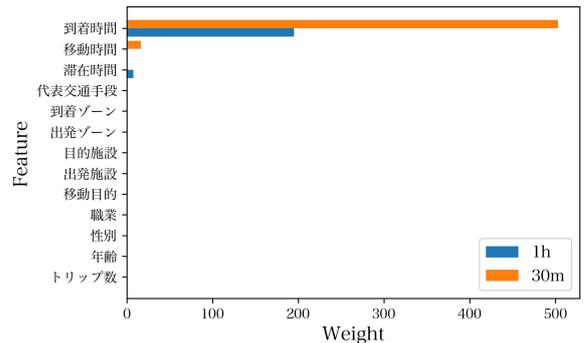


図-6 NNにおける特徴量の重要度

(4) モデル構築

予測をするにあたっては RF と NN で同じ特徴量を使用した。離散選択モデルによる出発時刻選択と同様に年齢や性別などの個人属性に加え、出発地や目的地の情報を特徴量として導入した。また特徴量に到着時刻や移動時間を入れているが、離散選択モデルでも到着時刻や移動時間を元に選択枝集合を決めているため、機械学習による予測においても特徴量として導入した。離散選択モデルでは、移動目的でセグメンテーションを行い各移動目的に対して説明変数を選択したが、機械学習においては移動目的も特徴量の 1 つとして分析を行なった。

本研究では、特徴量の重要性を解釈する指標として Permutation Importance を使用する。従来までの機械学習による予測では、学習過程がブラックボックスであるためどの特徴量がどれだけ学習に寄与したのか判断することが困難であった。仮に本研究で用いているランダムフォ

表-4 RF と NN の正解率

	ランダムフォレスト	ニューラルネットワーク
1hモデル	0.872	0.923
30mモデル	0.666	0.911

レストのようなアルゴリズムであれば、Gini 係数を用いて重要度を計測することは可能であるが、NN のように複雑なアルゴリズムには適応することができない。しかし、本研究で用いる Permutation Importance はアルゴリズムに依存せず特徴量の重要度を計算することができる。Gini 係数などの指標と比較して全体への影響は小さいものの、局所的な影響が強い特徴量への感度は低いという特徴がある。本研究ではこの Permutation Importance を用いて RF と NN の特徴量の重要度を算出し解釈を試みた。

(5) モデル学習結果と考察

RF と NN における的中率を表-4 に示す。1 時間、30 分のどちらのモデルでも NN の方が精度が高くなっていることがわかる。NN は 30 分レベルにおいても的中率は 9 割以上あるため、時間分解能を細かくして予測するには有効な手法だと考えられる。

Permutation Importance によって算出した特徴量の重要度を図-5、図-6 に示す。RF では、1 時間単位、30 分単位のどちらのモデルでも到着時刻の影響が一番大きく、その次に移動時間や滞在時間が影響していることがわかる。NN に関しても同様に、到着時刻の重要度が最も高くなる結果となった。このことから時刻選択においては、特徴量選択の重要性が示唆される。特に到着時刻や移動時間などの前後の行動を決める特徴量の影響は大きく、個人属性などの特徴量は出発時刻に寄与しないということが確認できる。

5. おわりに

本研究では、離散選択モデルを用いた出発時刻選択モデルの高分解能化と機械学習による精度向上を試みた。離散選択モデルを用いた出発時刻選択モデルでは、MNL に加え CNL を用いることによって時刻間の類似性を加味した。1 時間単位のモデルでは尤度比、的中率共に良好な結果を得ることができた。特に通勤目的に関してのトリップは、朝の時間帯に集中するため精度が高くなった。しかし、30 分レベルのモデルになると説明変数が有意とはならず、尤度比の低下も確認された。このことから 30 分以上に分解能を細かくしても、的中率や尤度比、説明変数の有意性に関しては良好な結果を得ることができないと考えられ、離散選択モデルでの出発時刻予測は限界があるということが示唆された。

よって本研究では新たな時刻選択モデルの提案として機械学習を用いた分析、特に RF と NN による出発時間帯の予測を行なった。結果として機械学習を用いたこと

により離散選択モデルによる予測よりも精度が向上し、特に NN での予測は時間分解能が細かくなっても効果的であることが確認された。到着時刻や移動時間などの特徴量に依存するモデルではあるが、出発時刻選択の精度向上が可能になった。

今後の課題として、時間分解能を細かくすること、到着時刻や移動時間に依存しないモデルの構築が挙げられる。分解能に関しては 15 分、10 分、5 分と細かくし、場合によっては NN に留まらず、中間層を増やしたディープラーニングを用いた時刻選択の分析を行なっていく必要もあるだろう。特徴量に関しては、到着時刻や移動時間以外で予測可能な NN を検討していく必要がある。

参考文献

- 1) 金森亮, 森川高行, 山本俊行, 三輪富生: 総合交通戦略の策定に向けた統合型交通需要予測モデルの開発. 土木学会論文集 D, Vol.65, No.4, pp.503-518, 2009.
- 2) 亀谷淳平, 福田大輔: 鉄道利用者を対象とした Activity-based 交通行動モデルに関する研究. 土木計画学研究・講演集, Vol.53, pp.380-389, 2016.
- 3) 藤原章正, 神田佑亮, 杉恵頼寧, 岡村敏之: 時刻選択モデルの選択肢区分が選択肢間の類似性及び不均一性に及ぼす影響. 土木計画学研究・論文集, No. 17, pp.599-604, 2000.
- 4) 東京都市圏パーソントリップ調査: 東京都市圏交通計画協議会 https://www.tokyo-pt.jp/data/01_01
- 5) 杉恵頼寧, 張 峻屹, 岡村 敏之, 藤原章正, 周藤 浩司: フレックスタイム制度の導入が出勤・退社時刻選択行動に及ぼす影響. 土木計画学研究・論文集, Vol. 19, pp.383-390, 2002.
- 6) Leo Breiman: Random Forest. Machine Learning, 45, pp.5-32, 2001
- 7) 島崎 敏一, 安田誠一: ニューラルネットワークによる交通手段選択モデル, 土木学会論文集, No.494, pp.79-86, 1994

(Received ?)
(Accepted ?)

HIGH TEMPORAL RESOLUTION OF DEPARTURE TIME SELECTION MODEL CONSIDERING SIMILARITY BETWEEN TIMES

Yushi ISHIJIMA, Hideki YAGINUMA, Shintaro TERABE and Haruka UNO

The evaluation of the impact of infrastructure development and transportation policies has shifted from a macro perspective to a micro perspective. In recent years, ABM has been used as a new method of analysis. However, the departure time selection model of ABM has some problems such as limited time study and coarse time resolution. In this study, we will refine the activity model of the Tokyo metropolitan area by increasing the resolution of the departure time selection structure.