

# 機械学習による信号交差点流入部の旅行速度予測

河合 芳之<sup>1</sup>・河野 九三夫<sup>2</sup>・田中 淳<sup>3</sup>

<sup>1</sup>非会員 株式会社オリエンタルコンサルタンツ (〒151-0071 東京都渋谷区本町3-12-1)

E-mail: kawai-ys@oriconsul.com

<sup>2</sup>非会員 株式会社オリエンタルコンサルタンツ (〒151-0071 東京都渋谷区本町3-12-1)

E-mail: kohno-km@oriconsul.com

<sup>3</sup>正会員 株式会社オリエンタルコンサルタンツ (〒151-0071 東京都渋谷区本町3-12-1)

E-mail: tanaka-at@oriconsul.com

現状では、信号交差点流入部の旅行速度は、ETC2.0や民間プローブデータを用いてもサンプル数の制約から日別15分単位の旅行速度として活用できる状況にはない。本研究では、警察庁が保有している断面交通量情報と交差点制御情報を活用して、機械学習によりプローブ旅行速度の欠測値を補間する予測モデルを構築した。その結果、信号交差点流入部の旅行速度予測は、実測値と予測値の標準化残差が±2以内となる割合が95～98%の的中率となった。本研究では、機械学習の課題であるホワイトボックス性を図るために、ランダムフォレストの説明変数の重要度を適用した。No Free Lunchの定理についても、箇所により最適な機械学習アルゴリズムが異なる結果となり、複数の機械学習アルゴリズムの適用の重要性を示すことができた。

**Key Words** : Travel speed prediction, Signalized intersections, machine learning

## 1. はじめに

渋滞対策については、これまでの取組により緩和が図られてきたものの、全国の渋滞対策協議会において特定した主要渋滞箇所は、約9,000箇所と依然として多数存在する。特に、一般道路における主な渋滞箇所は、信号交差点を起点に発生している。渋滞の原因を把握するためには、日々刻々と変化する旅行速度を把握することが重要である。

近年、ETC2.0プローブデータや様々な民間プローブデータを大量に収集・活用できる環境が整いつつあり、収集されたプローブデータから得られる様々な情報が、有益に活用できる可能性が明らかになってきている。しかしながら、現状では、旅行速度の分析精度を高めるために、日別、15分毎など分析単位を細分化すると、データに欠測値が生じる状況である(図1)。

一方、平成28年12月14日に公布・施行された「官民データ活用推進基本法」を踏まえて、平成29年3月31日より、警察庁では断面交通量情報の提供、加えて平成30年2月1日より、交差点制御情報の提供を(公財)日本道路交通情報センター(JARTIC)のホームページで行っている。

時間帯	信号交差点										進行方向	
	交通調査基本区間											
14:00 ~ 14:15	7179	7176	7176	7171	7171	7168	7161	7161	7159	7153	7153	7152
14:15 ~ 14:30	5.8	6.9	6.4	18.7	24.0	20.2	10.1	19.6	22.4	6.4	20.1	10.2
14:30 ~ 14:45	-	-	-	-	-	-	-	-	-	-	-	-
14:45 ~ 15:00	-	-	3.2	31.6	46.6	45.1	41.2	48.3	-	-	-	-
15:00 ~ 15:15	31.8	-	-	-	-	-	-	-	-	-	-	-
15:15 ~ 15:30	35.0	-	3.3	30.3	32.6	31.2	39.0	39.4	-	-	-	-
15:30 ~ 15:45	-	-	-	-	-	-	-	-	-	-	-	-

\* は、プローブデータの欠測

図-1 岩本町の旅行速度データ取得状況(H30.2.14)の例

欠測の無い旅行速度データとするためには、プローブデータの特性を踏まえた補間が必要となる。補間手法には、過去のデータ履歴の同時刻平均値を補間データとする手法、類似した旅行速度の変動パターンのデータを適用して補間<sup>1)</sup>する手法、Kinematic Wave理論の交通流の変分理論(VT)による推定する手法<sup>2)</sup>等がある。

本研究では、欠測率が高いプローブデータの旅行速度に対して、高度な補間が可能な新たな機械学習による手法を提案する。

## 2. 旅行速度予測に必要なデータの作成

### 2.1 基本的な考え方

交通流理論における交通量と旅行速度の関係は、交通量が交通容量に達しない領域では、交通量の単調増加に

対して、旅行速度は、単調減少であるが、交通量が交通容量を超える渋滞領域では、交通量と旅行速度は共に減少するといった、1つの交通量に対して2つの旅行速度を持つ二価関数となることが知られている。したがって、交通量と旅行速度の関係は、非線形関数を推定する必要がある。

また、都市部における主要な信号交差点では、MODERATO制御が行われている。これは、信号交差点上流部の車両感知器（以下、警察トラカン）から交通需要を推計し、交差点流入部毎に負荷率と呼ばれる指標を算出し、この指標値からサイクル長やスプリットを導出するものである。

回帰問題の目的は、連続値、離散値を問わない説明変数を用いて、連続値である目的変数を表す数理モデルを求める問題である。しかしながら、統計学的な重回帰モデルは線形関係であり、統計学的な非線形モデルは、パラメトリックであると見なされ、非線形回帰として記述される。一般的に、ノンパラメトリック非線形回帰には機械学習方法が使用される。

本研究は、流入部交通量、サイクル長、スプリット、時間帯、平休の別などが信号交差点流入部の旅行速度を決定するのではないかと推察し、これらの関係を機械学習によりモデル化することに着目した。すなわち、本研究は、信号交差点流入部における旅行速度を目的変数とし、説明変数に交通量、交通量比、サイクル長、スプリット、時間帯、平休の別とした機械学習を用いたノンパラメトリック非線形回帰モデルで旅行速度の欠測値を補間するものである。

本研究における機械学習アルゴリズムの選定に当たっては、ホワイトボックス（可読）性と、No Free Lunch定理に留意して実施する。ホワイトボックス性は、説明変数と目的変数間の関係が陽に解釈できるかどうか依存する。入出力の関係性がブラックボックスなモデルの場合は精度が高くとも説明変数の何がどれほど予測結果に寄与しているか言及できない。

No Free Lunch定理は、どのような問題やデータに対しても最高の精度を出せる万能な機械学習手法は存在しないという定理である。機械学習で最高の精度のモデルを構築するためには、様々なアルゴリズムを検証してみる必要がある。

## 2.2 研究のフロー

本研究は、図2に示すフローに従って実施する。初めに、入力データとなる旅行速度、交通量、交通量比、サイクル長、スプリット、時間帯及び平休の別を15分単位のレコードとなるデータベースを作成する。次に、作成

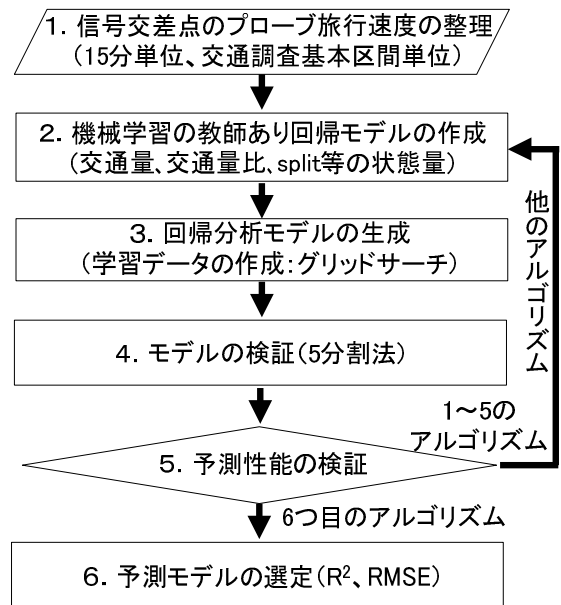


図-2 本研究における実施フロー



※ 平休別は、箇所により異なる。

図-3 旅行速度影響要因

したデータセットを用いて、アルゴリズム毎に機械学習の教師あり回帰モデルの作成し、グリッドサーチと交差検証を実施して、予測性能の最も高い予測モデルを選定する。

## 2.3 入力データの作成

プローブ旅行速度データは、(株)富士通交通・道路データサービスが保有・管理する15分・DRM区間単位の商用車プローブデータを活用する。交差点制御情報は、5分単位の信号機のサイクル長及びスプリット情報を活用する。交通量は、警察トラカンの断面交通量情報を活用する。上記データを用いて、旅行速度に影響を及ぼすと考えられる要因を抽出(図3)し、平成30年2月につい

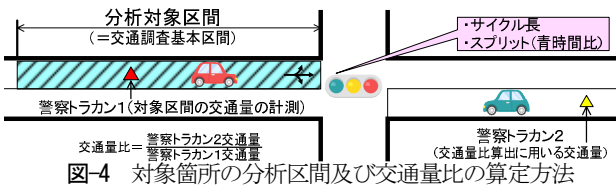


図-4 対象箇所の分析区間及び交通量比の算定方法

表-1 選定箇所及びデータ整理結果の概要

交差点名	東向島	岩本町	東小松川
流入部名	水戸街道下り	靖国通り東行き	船堀街道南行き
車線数(単路)	2	3	1
サイクル長(秒)	110~140	132~150	100~192
スプリット(%)	33~48	29~38	37~52
交通量(台/15分)	24~324	3~948	12~193
交通量(台/日)	11,028~16,977	18,694~32,718	7,983~9,768
交通量比	0.5~4.7	0.02~27.3	0.3~1.8
交通調査基本区間(km)	1.6	0.4	0.6
サンプル数	1,674/2,688	1,301/2,688	1,045/2,688
データ取得率	62%	48%	39%

て整理した。旅行速度は、主要渋滞箇所の算定方法に準拠して信号交差点に接続する交通調査基本区間について、15分単位で整理した。交通量、サイクル長及びスプリットは、15分単位に整理した。時間帯は、朝[7-9]、昼間[9-17]、夕方[17-19]、夜[19-22]、深夜[22-7]に区分した。平休別は、土曜日、日曜日及び祝日を休日とし、それ以外を平日に区分した。交通量比は、15分単位に当該流入部の交通量と対向流入部の交通量の比率とした(図4)。交通量比は、サイクル長やスプリットが大きい場合に、対象流入部の交通需要が大きいケースと対向流入部の交通需要が大きいケースが存在するため、この状況を表す指標として組み入れたものである。

分析地点は、旅行速度データが1ヶ月と短いため、プローブ取得率の高い区間で、かつ車線数、駐停車需要(駐停車が少ない)等を考慮して、3箇所を選定した。表1に、選定箇所及びデータ整理結果の概要を示す。

### 3. 旅行速度予測モデルの構築

#### 3.1 アルゴリズムの選定

本研究で用いる機械学習のアルゴリズムは、No Free Lunch定理を考慮して、比較的データ数が少ないケースでも適用可能なアルゴリズムを複数選定する。選定したアルゴリズムは、ランダムフォレスト(RF)、サポートベクター回帰(SVR)、k近傍法(KNN)、アンサンブルバギング回帰(EBR)、勾配ブースティング(GBRT)、Extremely Randomized Tree(ERT)の6つとする。

#### 3.2 旅行速度予測の手順

分析区間となる各箇所流入部を対象に、整理した交通調査基本区間の15分単位の旅行速度、交通量、交通量比、スプリット、サイクル長、時刻、平休別のデータを用い

て、6つの機械学習の教師あり非線形回帰アルゴリズムを適用する。その際、時刻、平休別は、ダミー変数を作成して量的変数とした。初めに、全サンプルデータのうちランダムに選出した7割を訓練データとし、各説明変数は、機械学習への寄与が入力の大さに影響するため、平均0、標準偏差1に標準化する。次にグリッドサーチを行って予測モデルの生成を行う。その後、各アルゴリズムに対して5分割法による交差検証を行い、得られた回帰モデルを用いて、訓練データの平均値と標準偏差を使って、テストデータの説明変数を標準化して予測値を算出する。最後に、6つのアルゴリズムの中から最もR<sup>2</sup>が大きく、RMSEが小さい予測性能の高い予測モデルを選定する(図1)。分析は、Pythonのオープンソース機械学習ライブラリであるscikit-learn<sup>3)</sup>を用いて実施する。

#### 3.3 全ての説明変数を用いた旅行速度予測モデル

全ての説明変数を適用して、訓練データを用いて6つのアルゴリズムで回帰モデルを作成した。作成した回帰モデル式とテストデータの説明変数を用いて、予測値である旅行速度を算出したところ、全てのケースで過学習が発生して、適切な旅行速度の予測値を得られなかった。過学習は、機械学習において、訓練データに対して学習されているが、未知データに対しては適合できていない、汎化できていない状態を指す。汎化能力の不足に起因する。過学習に対して最も効果がある対処法は訓練データを増やすことである。データを増やすことが出来ない場合には、モデルへの適合の説明変数が多すぎる等、訓練データの個数に比べて、モデルが複雑で自由度が高すぎることもある。

#### 3.4 説明変数を減じた旅行速度予測モデル

全ての説明変数を用いた旅行速度予測モデルでは、過学習が発生したため、説明変数を減じることとする。減じる説明変数の選定は、全ての説明変数を用いてランダムフォレスト(RF)で作成した回帰モデルで算出した説明変数の重要度を使用して変数選択を行う。RFの説明変数の重要度は、OOB(out-of-bag)データにおける予測精度から算出される。OOBデータとは、決定木を作成する際のブートストラップ標本に含まれなかったサンプルを指す。OOBデータを使うことによって、モデルの作成時に使われていないデータへの予測精度を評価することができる。

各説明変数の重要度は、スプリット、交通量、交通量比の値が高く、平休及びサイクル長の値は低い結果となった(図5)。このため、旅行速度予測に用いる説明変数は、スプリット、交通量、交通量比の3変数とする。

説明変数をスプリット、交通量及び交通量比の3変数、

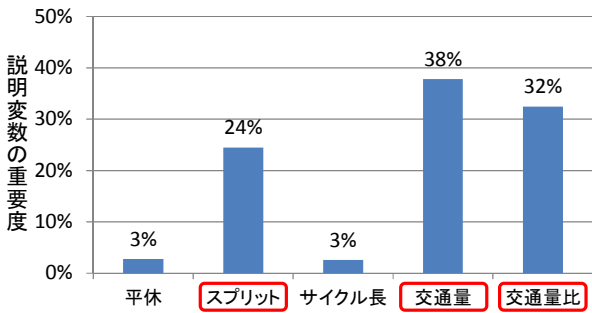


図-5 FRによる説明変数の重要度

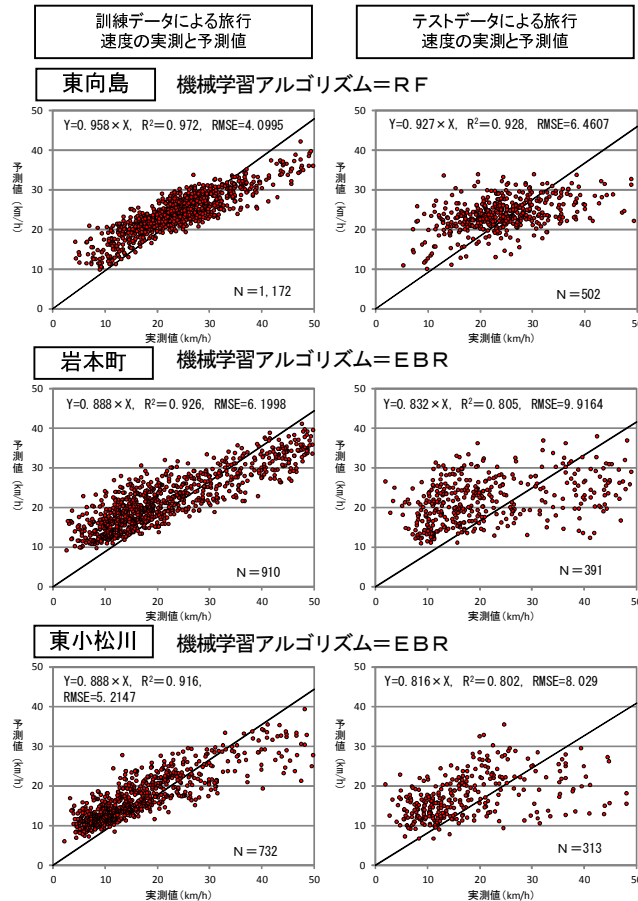


図-6 東向島・岩本町・東小松川の旅行速度の予測結果

目的変数を旅行速度とした6つのアルゴリズムを適用して訓練データを用いて、旅行速度予測の回帰モデルを作成する。機械学習の結果、得られた回帰モデルは、東向島では、ランダムフォレスト (RF) であり、岩本町と東小松川では、アンサンブルバギング回帰 (EBR) となった (図6)。

RFは、訓練データのブートストラップで描画された決定木で構築される。ブートストラップは、バイアス・高分散の推定のセットの分散を低減させ、過学習を回避するために使用される。一方、EBRは、元の訓練データのランダムなサブセットに決定木を構築し、個々の予測を集約して最終的な予測を構築するアルゴリズムである。この手法は、ランダム化をその構築手順に導入し、アン

サンプルを作成し、その予測を平均化によって決定木の分散を低減させ、過学習を避けるメタ推定手法である。ランダムフォレストで用いられているブースティングは、通常、弱いモデル (浅い決定木) が効果的であるが、バギングは、強く複雑なモデルで最も効果的である。

### 3.5 旅行速度予測モデルの評価

本研究では、得られた旅行速度予測モデルの評価を行う。評価は、決定係数R<sup>2</sup>、平均二乗誤差 (RMSE) 及び的中率の3指標を用いて行う。本研究で用いた的中率は、テストデータにおける各予測値と実測値の標準化残差を用いて定義した。標準化残差は、残差を標準偏差で規格化した値であり、一般的に標準化残差が±2もしくは±3以内に当てはまらないデータは、外れ値とされている。このことから、本研究では、標準化残差が±2及び±3以内のデータの割合を的中率とし、各箇所の予測データについて整理した (表2)。

予測精度でみると、岩本町が最も高く、次いで東向島、東小松川の順となる。これらの結果から、信号交差点流入部の旅行速度は、スプリット、交通量、交通量比を説明変数とした機械学習を用いることで、予測精度を著しく低下させることなく予測が可能であることが示された。しかしながら、いずれの箇所においても実測値の低速度を高め、高速度を低めに予測する傾向がみられる。これは、訓練データによる予測モデル作成時点においてもその傾向がみられることから、1ヶ月の旅行速度データのサンプル数によることが原因と推察される。

東向島の予測結果を用いて、説明変数の重要度をみると、旅行速度に影響を及ぼす要因のうち、交通量の重要度が39%と最も高く、次いで、交通量比35%、スプリット27%の順となっていることが示された (図7)。交通量、交通量比及びスプリットのいずれの指標も交差

表-2 箇所別の旅行速度予測モデル評価結果

交差点名	東向島	岩本町	東小松川
決定係数R <sup>2</sup>	0.928	0.926	0.802
平均二乗誤差 (RMSE)	6.46	6.20	8.03
的中率	(±3以内)	100%	100%
	(±2以内)	95%	98%

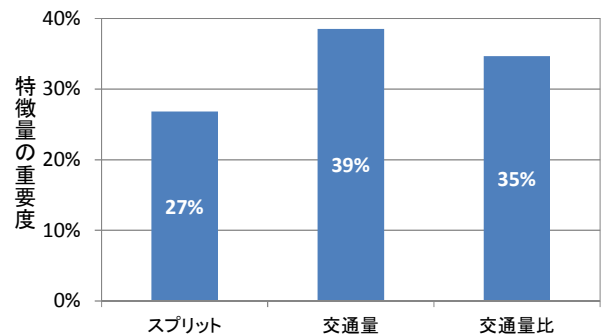


図-7 東向島の旅行速度に及ぼす説明変数の重要度

点流入部速度に影響を及ぼす要因であることが定量的に示された。

#### 4. おわりに

本研究では、機械学習アルゴリズムを用いた信号交差点流入部の旅行速度予測行い、説明変数にスプリット、交通量、交通量比を用いることで旅行速度の予測が行えることを示した。

本研究では、機械学習アルゴリズムの課題である「ホワイトボックス（可読）性」について、ニューラルネットワークなどではブラックボックスとなる説明変数の重要度をランダムフォレストを利用することで、分析可能であることを示された。また、「No Free Lunchの定理」においても、最適な機械学習アルゴリズムは、東向島ではRF、岩本町と東小松川では、EBRとなり、箇所によって異なる結果となり、同じ入力条件であっても箇所によって汎化性能が高くなるアルゴリズムは異なることを明らかにしたことで、予測モデル検討に際しては、複数の機械学習アルゴリズムを適用する重要性が示された。

本研究では、使用したサンプル数が1ヶ月間と少ないため、今後は、データ量を増やして機械学習の精度を高める必要がある。その際には、No Free Lunchの定理に従

って、サンプル数の制約から今回の研究では適用できなかったニューラルネットワークやガウス過程回帰モデルを含む複数の機械学習アルゴリズムを適用することと、ホワイトボックス性を確保するためにランダムフォレストを用いて、説明変数の重要度を算出することで目的変数と説明変数の因果関係を可視化した信号交差点流入部旅行速度予測モデルを構築する必要がある。

さらに、精度の高い信号交差点流入部の旅行速度予測モデルを構築後には、プローブデータと本手法を組み合わせ、欠測値のない網羅的な交通情報の提供を行う検討を進めていくことが、今後の課題である。

#### 参考文献

- 1) 舟橋賢二ほか：VICS 蓄積データを用いた旅行時間短期予測手法に関する研究，第 27 回土木計画学研究発表会講演集，CD-ROM(2003).
- 2) 桑原雅夫ほか：プローブデータと車両感知器データを融合活用した都市間高速道路における交通状態の推定，第 47 回土木計画学研究発表会，(2013)
- 3) Julian Avila etc. : scikit-learn Cookbook Second Edition, Packt Publishing, (2017)

(2018.7.31 受付)

## TRAVEL SPEED PREDICTION OF THE APPROACH AT SIGNALIZED INTERSECTIONS USING MACHINE LEARNING

Yoshiyuki KAWAI, Kumio KOHNO, Atsushi TANAKA