

機械学習によるプローブ旅行速度の欠測値補間モデルの改良と検証

河合 芳之¹・河野 九三夫²・竹平 誠治³

¹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

²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

³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: takehira@oriconsul.com

信号制御は、各利用者の遅れを最小化する制御設計が基本である。現状の信号制御は、非飽和を前提としたものが多く、遅れ時間のかなりの部分を発生させている過飽和の遅れ時間の定量的な把握は重要である。信号交差点流入部毎の一定時間単位ごとの正確な交通量と旅行速度が得られれば、交差点の遅れ時間を定量的に把握することが可能となる。しかし、現状では、旅行速度データは、欠測があり、十分に入手できている状況にはない。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た機械学習によりプローブ旅行速度の欠測値を補間するモデルを改良し、検証をおこなった。旅行速度の計測範囲を150m、300m及び500m、計測時間を15分、30分及び60分として、計測範囲が150m、計測時間15分の最小単位で有効な精度が得られる最適な機械学習アルゴリズムを構築した。動的に変動する交通状況を測定し、遅れ時間を最小化するように信号制御パラメータを一定量ずつ更新することが可能となる。

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

1. はじめに

国土交通省は、ICTを活用して、2020年度からETC2.0からODデータを作成する技術開発や必要なシステム改修を実施し、ODデータの常時観測化を進めている。交通量は、AIカメラ（画像認識型交通量観測装置）の全国配備を行って、常時観測エリアを拡大するとともに、ビデオ観測などの機械観測を推進し、従来の人による観測を廃止する。旅行速度調査は、ETC2.0データの活用により、直轄管理、地方管理道路とともに、常時観測体制へ移行を進めている。また、国土交通省では、ETC2.0データの利活用促進を目的として、国が収集したETC2.0データを配信するサービスについて、プライバシー保護やセキュリティ確保を考慮しつつ、配信サービスの具体的な内容の検討等を行っている。

一方、2017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2018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）

表-1 断面交通量及び交差点制御情報の公表箇所数

| 項目 | 公表箇所数 ² |
|--------------------|--------------------|
| 断面交通量 ¹ | 37,936断面 |
| 交差点制御情報 | 10,822交差点 |

1: 上り下り別 2: 2019年7月データのホームページで行っている（表1）。

ETC2.0プローブデータや様々な民間プローブデータは、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。

本研究では、これまでの研究成果¹⁾を改良し、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、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、主要な信号交差点では、上流側の交通量情報や交差点制御情報として、サイクル長や青信号スプリットが公表されている。

機械学習の目的は、「まだ手に入れていないデータを予測すること」である。その目的をどれだけ達成できているのかを評価するために、汎化誤差が使われる。汎化誤差とは「まだ手に入れていないデータを予測した時の誤差」のことである。機械学習の目的は、言い換えると「汎化誤差の小さな（汎化能力が高い）予測を出すこと」と言い換えることができる。

既存研究¹⁾では、青信号スプリット、流入部交通量、対向交通量比を説明変数とした機械学習による非線形モデルを構築した。本研究は、信号交差点流入部における機械学習によるプローブ旅行速度の欠測値補間モデルの改良を行うことを目的とし、実施したものである。

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。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り、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見る必要がある。

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、流入部交通量、対向交通量比、交差道路の交通量比の各々について、(15,30,60分×150,300,500m)の6つの組み合わせとなる9種類のデータベースを作成する。次に、作成したデータセットを用いて、8つの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（汎化誤差の小さい）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分単位の信号機の青信号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。

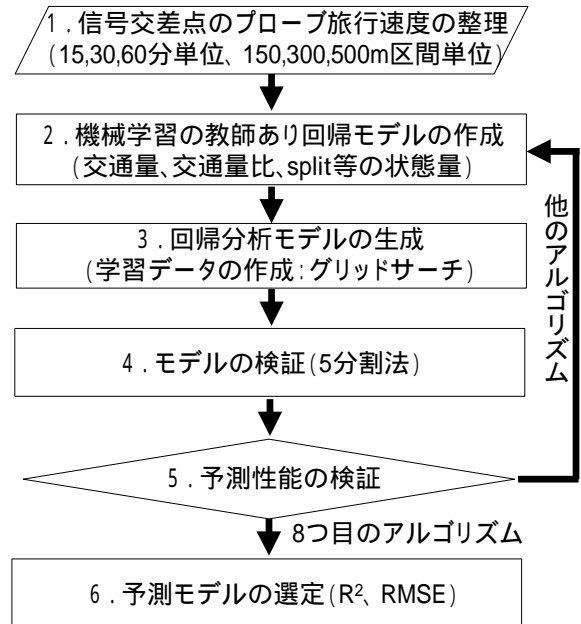


図-1 本研究における実施フロー

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2）し、平成30年2月について整理した。既存研究¹⁾では、旅行速度に影響を及ぼす要因として、流入部交通量、青信号スプリット及び対向交通量比が得られたが、本研究では、新たに、交差道路の交通量比を追加している。

旅行速度の分析対象区間は、MODERATO制御の車両感知器配置基準である「重要交差点流入部の幹線側に停止線より150mおよび300m上流に設置」に準拠して、150m、300m及び500mとした。分析対象時間は、15分、30分及び60分とし、これらの組み合わせとなる9種類のデータベースを成した。対向交通量比は、時間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、交差道路の交通量比は、交差道路の上下交通量の比率とした（図3）。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

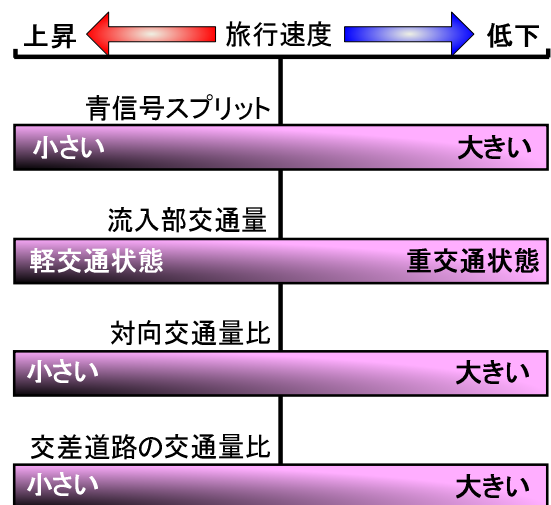


図-2 旅行速度影響要因

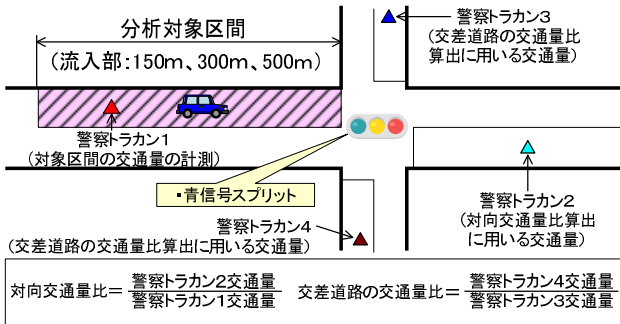


図-3 対象箇所の分析区間及び交通量比の算定方法

表-2 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 | 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 | |
| 青信号スプリット(%) | 33 ~ 48 | 29 ~ 38 | 37 ~ 52 | |
| 交通量(台/15分) | 24 ~ 324 | 3 ~ 948 | 12 ~ 193 | |
| 対向交通量比 | 0.5 ~ 4.7 | 0.02 ~ 27.3 | 0.3 ~ 1.8 | |
| 交差道路の交通量比 | 0.3 ~ 2.3 | 0.1 ~ 10.9 | 0.4 ~ 4.8 | |
| 評価延長(m) | 150m | 173 | 119 | |
| | 300m | 337 | 298 | |
| | 500m | 503 | 437 | |
| サンプル数(取得率) | 15分 | 1,464 (54%) | 1,247 (46%) | 953 (35%) |
| | 30分 | 971 (72%) | 842 (63%) | 727 (54%) |
| | 1時間 | 577 (86%) | 510 (76%) | 493 (73%) |

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2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、既存研究¹⁾で選定した、Random(Decision) Forest (RF)、Support vector (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種類に加えて、デジジョンツリーRegressor (DTR)、XGBoost (XGB) の8種類とした。

追加したデジジョンツリーRegressor (DTR) は、他のアルゴリズムと比較すると、入力データから特徴を学習し、決定木と呼ばれる樹木状の構造で学習結果を視覚化でき、ルールをシンプルに表現できる特徴があり、決定木はホワイトボックスのモデルだといえ、論理的に解釈することが容易である。

XGBoost (XGB) は、勾配ブースティング木を改良したブースティングシステム、分類器とその構築アルゴリズム、ライブラリの総称である(複数の決定木や回帰木を弱学習器として、勾配ブースティングというアンサンブル学習手法を用いて作られた分類器をブースティング木

と呼ぶ。)

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(15,30,60分×150,300,500m)の旅行速度、流入部交通量、青信号スプリット、対向交通量比及び交差道路の交通量比のデータを用いて、8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。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。次に残りの3割のデータを用いて、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、8つのアルゴリズムの中から最もR²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-learn²⁾を用いて実施する。

3.3 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8つのアルゴリズムで回帰モデルを作成した。機械学習の結果、得られた回帰モデルは、全てのケースでXGBoostとなった(図4)。

XGBoost は従来のブースティング木に比べて並列化が容易であり、学習データがスパースであるとき高速に動き、大規模データをメモリ上で効率よく扱えるといった多くの利点がある。

ある回帰木がベクトル x を入力したときに $f(x)$ を出力するとする。XGBoost が K 個の回帰木により構成されていたとき、入力 x に対する XGBoost の出力は、

$\phi(x) = \sum_{k=1}^K f_k(x)$ で表される。XGBoost の学習に用いる学習用データを $((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n))$ とすると、XGBoost の目的関数は以下の式で表される。

$$L(\phi) = \sum_{i=1}^n l(y_i, \phi(x_i)) + \sum_{k=1}^K \Omega(f_k)$$

第1項目が損失項であり、 l は損失関数を表す。損失関数には対数尤度や平均二乗誤差などが使われる。第2項目が正則化項であり、 Ω は正則化関数を表す。正則化には L1 ノルムや L2 ノルムなどが用いられる。XGBoost は目的関数 L を最小化するように回帰木を追加することで学習を行う。XGboostは、基本的に内部で行われていることは決定木を複数個作り、学習結果の修正を行って

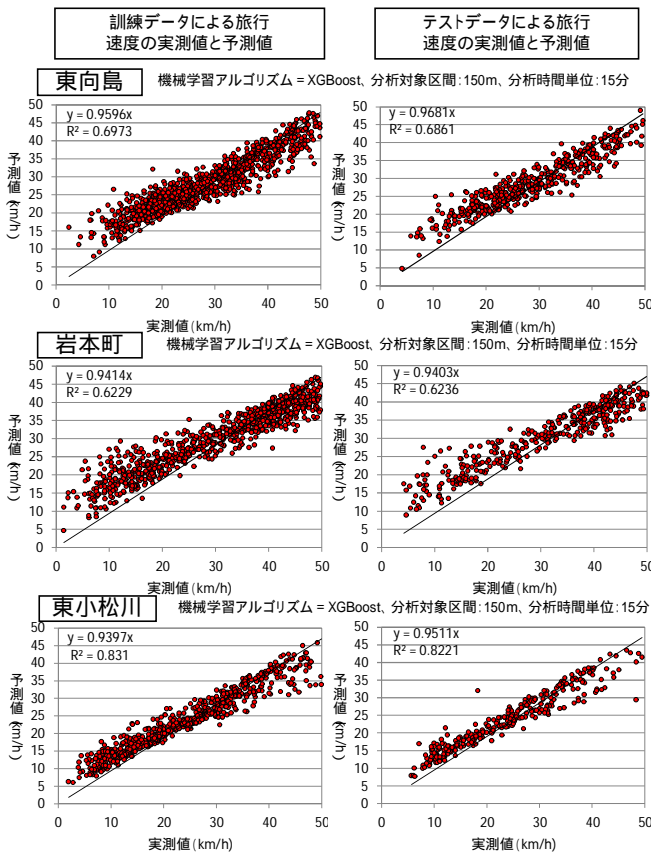


図-4 東向島・岩本町・東小松川の旅行速度の予測結果 (15分単位, 150mのケース)

いる。既存研究¹⁾で用いたランダムフォレストとは異なり、学習結果の修正が入っている。正しく分類されたインスタンスはますます重みが小さくなり、正しく分類されないインスタンスはますます重みが大きくなり続ける。最終的に、信頼度の重み付け平均の結果を用いて、最終的なモデルを構築する。

3.4 旅行速度予測モデルの特徴量(説明変数)の重要度

決定木をベースにしたアルゴリズムは、特徴量の重要度という指標が存在する。特徴量の重要度は、説明変数を用いてXGBoostで作成した非線形回帰モデルで算出した、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

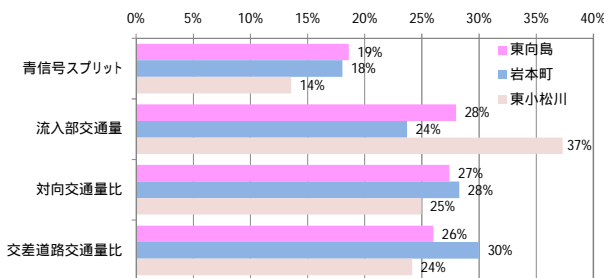


図-5 東向島の旅行速度に及ぼす説明変数の重要度

表-3 評価結果 (R², RMSE, 的中率)

| 評価項目 | 分析単位 | 東向島 | | | 岩本町 | | | 東小松川 | | | |
|--------------------|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|------|
| | | 150m | 300m | 500m | 150m | 300m | 500m | 150m | 300m | 500m | |
| 決定係数R ² | 15分 | 0.697 | 0.821 | 0.727 | 0.623 | 0.699 | 0.825 | 0.697 | 0.821 | 0.727 | |
| | 30分 | 0.773 | 0.730 | 0.873 | 0.891 | 0.703 | 0.784 | 0.773 | 0.730 | 0.873 | |
| | 1時間 | 0.886 | 0.820 | 0.843 | 0.869 | 0.887 | 0.869 | 0.886 | 0.820 | 0.843 | |
| 平均二乗誤差 (RMSE) | 15分 | 4.20 | 3.11 | 3.60 | 5.50 | 5.45 | 4.00 | 3.66 | 4.15 | 3.75 | |
| | 30分 | 3.49 | 3.09 | 2.58 | 2.93 | 4.73 | 4.18 | 3.53 | 2.71 | 3.17 | |
| | 1時間 | 2.98 | 2.73 | 2.51 | 3.27 | 3.28 | 3.61 | 3.25 | 3.74 | 2.41 | |
| 的中率 | ± 1 以内 | 15分 | 98% | 100% | 99% | 98% | 99% | 99% | 98% | 100% | 99% |
| | | 30分 | 99% | 99% | 99% | 100% | 99% | 98% | 99% | 99% | 99% |
| | | 1時間 | 99% | 98% | 99% | 99% | 99% | 99% | 99% | 98% | 99% |
| | ± 2 以内 | 15分 | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% |
| | | 30分 | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% |
| | | 1時間 | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% |

る。

各説明変数の重要度は、流入部交通量、対向交通量比、交差道路の交通量比の値が高く、青信号スプリットの値は低い結果となった(図5)。特徴量の重要度を確認することで、流入部交通量、対向交通量比、交差道路の交通量比及び青信号スプリットのいずれの指標も交差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。

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R²、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以内にはまらないデータは、外れ値と言われている。このことから、本研究では、標準化残差が±1及び±2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3)。

予測精度でみると、全ての箇所分析単位時間が60分と長いケースの方がR²及びRMSEが高い傾向がみられるものの、分析単位時間が15分かつ分析単位区間が150mと短いケースでも、旅行速度の欠測値を補間する十分な精度が得られた。

4. 非線形回帰モデルを用いた遅れ時間の評価

信号交差点の円滑性の評価指標は、遅れ時間、信号待ち回数と停止回数、渋滞長と渋滞長時間及びサービス水準などがある。これらの指標のうち、遅れ時間は、道路利用者が交通に伴う便益を表す直接的指標として最も普遍的なものである。これまで、交差点が無かったと仮定した場合の対象とする区間の旅行時間の推定は、容易ではなく、遅れ時間の評価は、実際の旅行時間の差分で評価することが一般的であった。また、我が国で一般的に用いられている指標は、観測のしやすさから渋滞長と渋滞時間が一般的である。

遅れ時間を推定する方法として、従来から種々のモデ

ルが提案されてきたが、その代表的なものとして、

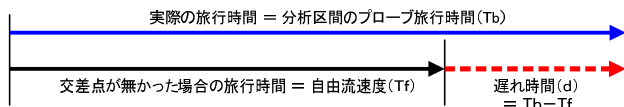


図-6 旅行時間と遅れ時間の関係

Webster³⁾やHighway Capacity Manual等による遅れ時間が考案されている。前者は、交通流が飽和状態に近づくにつれて遅れ時間は無限大に発散するため、適用にあたり、交通状態が非飽和であることが条件となる。一方、後者は非飽和状態から過飽和状態まで連続的に扱うことが特徴である。我が国では、Websterの遅れ式の簡易式である下式を用いている。

$$d = \frac{(1-g)^2}{2(1-\lambda)} C + \frac{x^2}{2q(1-x)}$$

本研究では、推定した旅行時間を用いて、信号交差点の遅れ時間を算定することで、実際の遅れ時間とモデル式及び遅れの式との比較を行い、当該モデルの有用性について検討をおこなった(図6)。

4.1 信号交差点遅れ時間の評価方法

信号交差点における遅れ時間は、交差点が無かったと仮定した場合の分析区間の旅行時間と実際の旅行時間との差である(図7)。

本研究では、実際の旅行時間(Tb)をプローブ速度から得られた値とし、交差点の無かった場合の旅行速度(自由流速度：Tf)として、分析区間の全旅行時間デ

ータから得られた90パーセンタイル速度値とし、TbとTfの差分を遅れ時間(d)と定義した。

本研究の遅れ時間は、1台当たりの平均遅れ時間(秒/台)とし、プローブ旅行時間から得られる遅れ時間を「実測」とし、本研究で得られた機械学習によるXGBoostの非線形回帰モデルによる遅れ時間を「モデル式」とし、無作為に半数の分析時間単位(15, 30, 60分)の実測旅行時間データを削除しモデル式で旅行時間を補間した遅れ時間を「補間(50%)」とし、Websterの簡易式で得られる遅れ時間を「遅れ式」と定義した。分析は、実測の遅れ時間に対する各遅れ時間の比率を整理することで、モデル式及び補間(50%)の遅れ時間の精度の確認をおこなった。

信号交差点の評価に用いたデータは、評価対象時間の2時間全ての分析時間単位でプローブ旅行時間データが得られた日を抽出した。東向島は、2018年2月6日(火)の14~16時で、ピーク時2時間前の交通状況であり、岩本町は、2018年2月26日(月)の14~16時でピーク時の2時間帯であり、東小松川は、2018年2月7日(水)の10~12時で、オフピークの時間帯である。

4.2 信号交差点遅れ時間の評価結果

信号交差点の遅れ時間及び実測遅れ時間に対する比率等は、表4に示すとおりである。

モデル式は、各交差点ともに分析時間単位が15分の時は、実測値の遅れ時間に比べて過小評価する傾向があり、

表-4 東向島、岩本町及び東小松川における実測、モデル式、補間(50%)及び遅れ式による平均遅れ時間

| 分析時間 | 分析区間 | 実測速度(km/h) | 交通量(台/2h) | 自由流速度(km/h) | 平均遅れ時間(秒/台) | | | | 実測遅れ時間に対する比率(%) | | |
|------|------|------------|-----------|-------------|-------------|-------|---------|-------|-----------------|---------|------|
| | | | | | 実測 | モデル式 | 補間(50%) | 遅れ式 | モデル式 | 補間(50%) | 遅れ式 |
| 東向島 | 15分 | 150m | 22.2 | 1,971 | 45.1 | 14.2 | 10.8 | 12.6 | 76% | 89% | 266% |
| | | 300m | 27.2 | | 46.6 | 18.5 | 17.2 | 18.4 | 93% | 99% | 182% |
| | | 500m | 30.3 | | 47.0 | 21.3 | 21.0 | 22.1 | 98% | 104% | 151% |
| | 30分 | 150m | 23.1 | | 45.1 | 13.1 | 11.5 | 10.6 | 88% | 81% | 316% |
| | | 300m | 27.7 | | 46.6 | 17.8 | 17.4 | 16.0 | 98% | 90% | 210% |
| | | 500m | 30.5 | | 47.0 | 20.8 | 21.0 | 19.5 | 101% | 94% | 172% |
| | 60分 | 150m | 25.1 | | 45.1 | 11.0 | 9.5 | 8.9 | 86% | 81% | 377% |
| | | 300m | 28.9 | | 46.6 | 16.0 | 15.2 | 14.1 | 95% | 88% | 238% |
| | | 500m | 31.2 | | 47.0 | 19.4 | 21.0 | 17.5 | 108% | 90% | 191% |
| 岩本町 | 15分 | 150m | 20.4 | 3,449 | 50.0 | 12.4 | 9.7 | 12.4 | 78% | 100% | 424% |
| | | 300m | 17.1 | | 49.5 | 41.0 | 26.9 | 37.4 | 66% | 91% | 140% |
| | | 500m | 11.5 | | 44.7 | 101.2 | 82.7 | 95.2 | 82% | 94% | 55% |
| | 30分 | 150m | 22.9 | | 50.0 | 10.1 | 10.0 | 10.0 | 99% | 99% | 522% |
| | | 300m | 20.7 | | 49.5 | 30.1 | 28.2 | 28.4 | 94% | 94% | 185% |
| | | 500m | 13.8 | | 44.7 | 78.6 | 81.9 | 74.5 | 104% | 95% | 70% |
| | 60分 | 150m | 24.9 | | 50.0 | 8.7 | 8.9 | 9.6 | 103% | 110% | 549% |
| | | 300m | 22.1 | | 49.5 | 26.9 | 30.6 | 28.6 | 114% | 106% | 183% |
| | | 500m | 14.4 | | 44.7 | 73.9 | 72.8 | 70.2 | 99% | 95% | 75% |
| 東小松川 | 15分 | 150m | 19.5 | 998 | 41.5 | 19.6 | 20.0 | 19.9 | 102% | 102% | 140% |
| | | 300m | 16.8 | | 47.7 | 63.6 | 60.4 | 65.6 | 95% | 103% | 43% |
| | | 500m | 14.2 | | 53.0 | 104.0 | 99.3 | 105.0 | 95% | 101% | 27% |
| | 30分 | 150m | 21.0 | | 41.5 | 16.9 | 15.9 | 17.7 | 94% | 105% | 158% |
| | | 300m | 17.2 | | 47.7 | 61.2 | 55.9 | 59.1 | 91% | 97% | 47% |
| | | 500m | 15.0 | | 53.0 | 96.6 | 96.2 | 93.9 | 100% | 97% | 30% |
| | 60分 | 150m | 21.6 | | 41.5 | 16.0 | 19.3 | 16.2 | 120% | 101% | 172% |
| | | 300m | 17.7 | | 47.7 | 58.2 | 48.9 | 52.1 | 84% | 90% | 54% |
| | | 500m | 15.5 | | 53.0 | 92.0 | 91.0 | 93.8 | 99% | 102% | 30% |

分析区間単位が長くなりに従って、適合性が向上する傾向がある。30分及び60分の分析時間単位の時は、実測値と概ね合致する傾向がある。

補間（50%）では、各交差点ともに、全ての分析時間単位および分析区間単位で、実測値と概ね合致している。

遅れ式は、多くのケースで過大評価となっており、分析区間単位が長くなる箇所では、過小評価しているケースもみられる。また、全てのケースで、実測値との乖離が非常に大きくなっている。

これらの結果を踏まえると、本研究の目的であるプローブ旅行時間の欠測時間帯の補間は、50%程度の欠測時間帯が生じている場合であっても、本モデル式を適用することで、実測値に近似した旅行時間の補間が行えることが確認された。

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を行い、説明変数に流入部交通量、青信号スプリット、対向交通量比及び交差道路の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、説明変数の重要度をXGBoost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、XGBoostによる非線形回

帰モデルの汎化性能が高い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。

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。また、単路部における旅行速度の推定への活用も視野に入れ、地方管理道路のETC2.0カバー率の低い路線等での旅行速度の推定を行うモデルの構築を行うことも、今後の課題である。

参考文献

- 1) 河合芳之，河野九三夫，田中淳：機械学習による信号交差点流入部の旅行速度予測，第 58 回土木計画学研究発表会・講演集，2018.11
- 2) Jason Brownlee：Gradient Boosted Tree With XGBoost and scikit-learn，2018
- 3) F.V.Webster，B.M.Cobbe：Traffic Signals,Road Research Technical paper No.56，1966

(2019.10.3 受付)

IMPROVEMENT AND VERIFICATION OF MISSING VALUE INTERPOLATION MODEL FOR PROBE TRAVEL SPEED USING MACHINE LEARNING

Yoshiyuki KAWAI，Kumio KOHNO，Seiji TAKEHIRA