

グラフ畳み込みニューラルネットを用いた 観光地における旅行速度予測

Huseyin TIRTOM¹・森田 格²・辻 大樹³・村戸 伸行⁴・力石 真⁵

¹ 正会員 日本工営株式会社 先端研究開発センター (〒300-1259 茨城県つくば市稲荷原2304)

E-mail: huseyin-tr@n-koei.jp

² 正会員 日本工営株式会社 先端研究開発センター (〒300-1259 茨城県つくば市稲荷原2304)

E-mail: morita-it@n-koei.jp

³ 非会員 日本工営株式会社 福岡支店 (〒812-0007 福岡県福岡市博多区東比恵1-2-12)

E-mail: a7476@n-koei.jp

⁴ 非会員 日本工営株式会社 大阪支店 (〒530-0047 大阪市北区西天満 1-2-5)

E-mail: a5954@n-koei.jp

⁵ 正会員 広島大学准教授 大学院先進理工系科学研究科 (〒739-8529 広島県東広島市鏡山1-5-1)

E-mail: chikaraishim@hiroshima-u.ac.jp

近年、ディープラーニングを用いた交通流予測モデルは、より高い精度に到達し、実用的なアプリケーションが可能になりつつある。具体的には、高い精度で数日先の渋滞状況を予測できれば、訪れる日・時間を変えて混雑を回避する行動を促進でき、例えば慢性的に混雑している観光地の交通状態の改善に役立つと考えられる。本研究では、GCN (Graph Convolutional Network) と LSTM (Long short-term memory) を組み合わせた渋滞予測 (旅行速度予測) モデルを構築し、実証分析を行った結果を報告する。過去のETC2.0データを用いてモデルを学習し、将来7日の7~20時の交通流を予測した。予測精度を確認したところ、2.3~5.6km/hの平均誤差で旅行速度の予測が可能であることを確認した。

Key Words : *deep learning, traffic speed prediction, graph convolution, ETC2.0*

1. はじめに

近年、ICTやAI技術の向上に伴い、タクシー需要予測、電車の混雑予測、観光地の混雑予測など、交通分野でもディープラーニングアプリケーションが急速に開発・実装されている。例えば、長期的な (本研究では、既存の数時間先を予測する短期予測と区別するために、数日先の予測を長期予測と呼称する) 交通流を正確に予測することで、旅行者に混雑する日・時間を回避する旅行計画を促すことができる。特に、紅葉シーズンなどの観光ピーク期に国内観光による混雑が発生している観光地では、交通渋滞を緩和し、旅行者の利便性を向上させる上でも役立つ事が期待される。混雑した道路での待ち時間が減少することで、大気汚染の削減にも繋がり、長期的な交通流予測の研究は、重要な実用的意義も持っている。

ディープラーニングを用いた予測方法は、実装が容易であり、様々なソースから取得するビッグデータから学習できるため最近広く利用されている。初期のディープラーニングの研究では、複数の隠れ層のみを持つ単純な

FNN (フィードフォワードニューラルネットワーク) が使用されていたが、その後の研究では、LSTM (長期短期記憶) やCNN (畳み込みニューラルネットワーク) など、より複雑なネットワークを使用するアルゴリズムが出てきた。LSTMは、時系列形式の入力データを受け入れ、過去の出力結果も学習に利用できるようにしたモデルである。メモリユニットを使用して、過去のデータから重要な情報を保持し、時間的特徴を抽出する機能がある。一方、CNNには画像のように正方行列形式の入力を受け入れ、空間的特徴を抽出する機能がある。CNNは画像処理アプリケーションでしばしば活用されているが、多変量時系列データを画像形式に変換しCNNに入力することで時間的および空間的特徴を一度に抽出することも行われている。しかし、このような行列形成の問題の1つは、無関係なデータが空間的に接続されているかのように畳み込まれることである。最近では、この問題を克服するためにGCN (グラフ畳み込みネットワーク) が開発された。GCNを使用すると、隣接行列を用いて定義された近傍のデータのみが畳み込ま

れ、多くの場合、CNNよりも適切に空間的特徴が抽出できるようになる。また、LSTM層とCNN層あるいはGCN層を組み合わせることで、1つモデルで空間的特徴（CNN層あるいはGCN層による）と時間的特徴（LSTM層による）の両方を抽出することが可能となる。

本研究では、GCNとLSTMを組み合わせた長期交通流速度予測モデルを構築し、過去のETC2.0データを用いてモデルを学習し、10道路区間の将来7日の7～20時の交通流（速度）の予測を試みる。次の章では既存の研究をレビューし、第3章ではモデル構造について説明する。第4章では、過去のETC2.0データを活用するケーススタディを実行し、提案されたモデルの精度を確認する。最後に、第5章では、結果と今後の課題について記載する。

2. 関連研究

交通流予測はこれまでも多くの検討がなされてきたテーマであり、過去データの単純な平均から複雑な機械学習フレームワークまで、幅広い予測方法が検討されている。Hinsbergenら⁷⁾は、これらの方法をナイーブ、パラメトリック、ノンパラメトリックに分類している。ナイーブ手法は、過去データの平均などモデル形成のない単純な方法である。パラメトリック手法には、モデルの構造が事前に決定され、データからパラメータのみを推定する必要があるシミュレーションモデルや時系列モデル等があり、膨大なデータを必要としないが、交通流のダイナミクスに関する包括的な知識を必要とする。一方、k近傍法、決定木、ニューラルネットワークなどのノンパラメトリックモデルは、膨大な量のデータを必要とするが、渋滞メカニズムに関する深い知識を必要とせず、モデル構造とパラメータの両方がデータによって決定される。最近ではビッグデータ活用が広まり、膨大な交通関連データが入手可能になって、ノンパラメトリックモデルの適用が多くなっている。

予測方法以外には、データ種類、データ期間、予測対象（エリア、時間、速度または移動時間など）および評価方法も様々である。金澤ら²⁾は13カ月のVICS (Vehicle Information and Communication System) 交通情報を用いて決定木モデルでVICSネットワーク全体の10分～240分先の予測をした。Kano³⁾はVICS交通情報を用いてファジークラスタリングとセルオートマトンモデルでネットワーク全体の1時間先の予測をした。Shiraishi⁴⁾はVICS交通情報を用いてリアルタイムシミュレーションモデルでネットワーク全体の1時間先の予測をした。森田⁵⁾はVICS装置がないエリアを考慮して1年間のタクシープローブデータを用いてエージェントベースモデルで1時間先の予測をした。Liang⁶⁾は信号の影響を考慮してシミュレーションソフトで作成した交通情報を基に数理的モデルで市掛地の5分先の予測をした。芦田⁷⁾は高速道路の3か月分の走行データと雨量データを用いて異種混合学習技術モデルで時間帯別渋滞予測をした。大川⁸⁾は一週間のNAVITIMEプローブデータを用いて潜在因子モデルで首都高速道路の48時間先の渋滞を予測した。高橋⁹⁾はシミュレーションモデルから生成された交通データを用いて密度球GCNモデルで走行時間の予測をした。柴田¹⁰⁾は4年間の車両感知器から得られた断面交通量及び平均速度データに加えてカレンダー、気象、イベントデータを用いて、4層FNNモデルで24時間先の予測をした。吉田¹¹⁾は4か月の交通量データを用いてGCNモデルで道路ネットワーク全体の5分先の交通量を予測した。小川¹²⁾は2年間の交通量データに加えて降水量及びイベントデータを用いて、多変量LSTMモデルで複数地点の60分先の通過交通量を予測した。

以上から、ほとんどの研究は、5分先から数時間先までの短期予測に着目している。もっと先の長期的予測は交通流理論から離れパターン認識問題になり、通常、パターンマッチングや自己回帰モデルフィッティングなどの手法が使われている。多くの場合、渋滞状況は時間帯と曜日大きく依存し、渋滞パターンは季節を通じて繰り返されるため、ディープラーニングによりそのパターンを学習することにより、数日先の交通流を良い精度で予測できると考えられる。本研究では、短期予測に適用されてきたディープラーニング手法と交通データを長期予測に活用することを試みる。

以上から、ほとんどの研究は、5分先から数時間先までの短期予測に着目している。もっと先の長期的予測は交通流理論から離れパターン認識問題になり、通常、パターンマッチングや自己回帰モデルフィッティングなどの手法が使われている。多くの場合、渋滞状況は時間帯と曜日に大きく依存し、渋滞パターンは季節を通じて繰り返されるため、ディープラーニングによりそのパターンを学習することにより、数日先の交通流を良い精度で予測できると考えられる。本研究では、短期予測に適用されてきたディープラーニング手法と交通データを長期予測に活用することを試みる。

3. モデル構築

(1) 使用データ

データ取得の即時性を考慮し、使用するデータはETC2.0速度（時間毎の平均）、ETC2.0サンプル数（時間毎の合計）、降水量、曜日とする。ある道路区間の将来の交通流は、上流側と下流側の影響を受けると考えられるため、予測対象道路区間の交通情報（ETC2.0速度とETC2.0サンプル数）に加えて上下流道路区間の情報も混雑予測に使用する。その情報を取得する道路ネットワークを図-1に示す。上流6区間（1～6）・対象区間（7）・下流6区間（8～13）の形で設定する。入力データを固定ネットワーク形状で準備することで、異なる対象道路区間でも同じモデルを使用できる。

予測日の天気を「晴れ・雨・大雨」として設定できるように、日降水量を3グループ（晴れ：<4mm, 雨：≥4mm, <40mm, 大雨：≥40mm）に分けて準備する。曜日データは

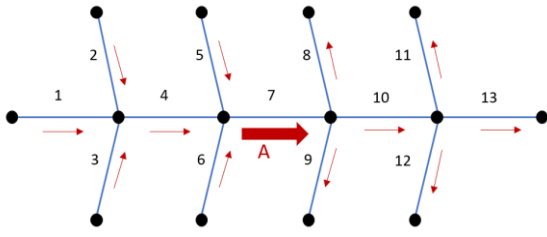


図-1 モデルにおける道路ネットワークの構造
平日・休み・連休という3つのフラグで準備する。

(2) モデル構造

過去2週間のデータを用いて将来1日の予測をするモデルは図-2で示す通り構成されている。入力層は正規化し、GCN層で空間的特徴、LSTM層で時間的特徴を抽出、FNNで出力結果にあわせた予測を行う。GCN層の入力I (図-2の入力行列1) は $N \times M$ 行列で、 N は入力件数、 M は計測件数を表す。GCN層の入力A (図-2の隣接行列) は道路区間間の接続性を示す $N \times N$ 行列である。GCN層の出力は $N \times M \times Z$ 行列 (Z はフィルター数) で、式(1a)によって計算される。

$$y_{ijz} = \sigma(\sum_k (I_{kj} * A_{ik} * F_k) + b_z), k \in \{1, 2, \dots, N\} \quad (1a)$$

ここで σ は活性化関数、フィルタ F_k はランダムに生成された $1 \times N$ ベクトル、 b_z はバイアス項である。GCN層の出力に降水量と曜日を加えてLSTM層の入力とする。さらに、LSTM層の出力に予測日の降水量と曜日を加えてFNN層に入力し、13道路区間の将来1日分の時間帯別(7~20時)の速度とサンプル数を出力する。Dropout層は出力の一部を学習時にランダムで0にする層で、一部のデータが欠損していても正しく予測ができるようにし、過学習も回避することができる。

モデルの入力行列を図-3の通り3つに分けて準備する。速度とサンプル数データから入力行列1を、降水量と曜日データから入力行列2を作成する。予測日のデータも入力行列3として作成する。

損失関数として式(1b)の平均二乗誤差 (MSE) を使用する。ここで、 y_p が計測速度 y'_p が予測速度で p は出力数である。

$$MSE = \frac{1}{p} \sum_{i=1}^p (y_p - y'_p)^2 \quad (1b)$$

(3) 予測モデルの使用方

予測モデルは将来1日分のみを予測するがこのモデルを繰り返し使用することで今日を含めた将来8日分の速度を予測する。まずは、過去1~14日のデータを使用し、15日目の値を予測する。次は、過去2~14日入の力データに加えて15日目の予測値を加え、16日目の値を予測する。これを8回繰り返し、過去14日分のデータで将来8日

分の速度を予測する。

4. ケーススタディ

混雑が発生しているある道路ネットワークを対象にH28.4~R1.12の実際の交通とアメダスの降水量データを用いてモデルを学習し、精度確認を行った。ETC2.0データは図-5に示す33道路区間を対象に整理し、このうち黄色の5道路区間(双方向で10方向分)を学習対象区間として教師データを作成した。

教師データを表-1の通り学習と検証に分け、モデル学習を行った。学習時のパラメータは表-2の通り設定した。

学習したモデルの精度を検証するため、検証データを用いてR1.11.1~27の予測を行った。精度比較のため、道路区間別・曜日別自己回帰和分移動平均モデル (ARIMAモデル) による予測も行った。ARIMAモデルは、 p 次の自己回帰モデルと q 次の移動平均モデルを組み合わせたモデルである自己回帰移動平均モデルに対して、 d 次差分をとることによってトレンド除去を考慮したモデルで、式(1c)によって計算される。

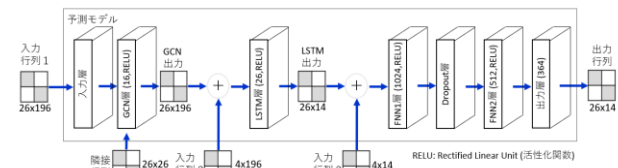


図-2 モデル構造

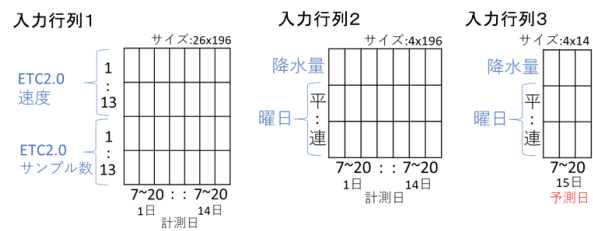


図-3 入力行列

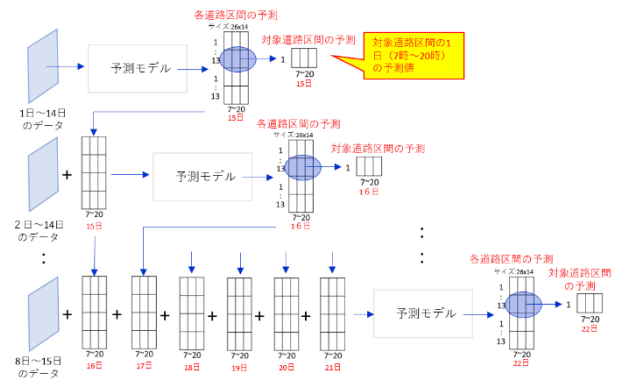


図-4 モデルの予測方法

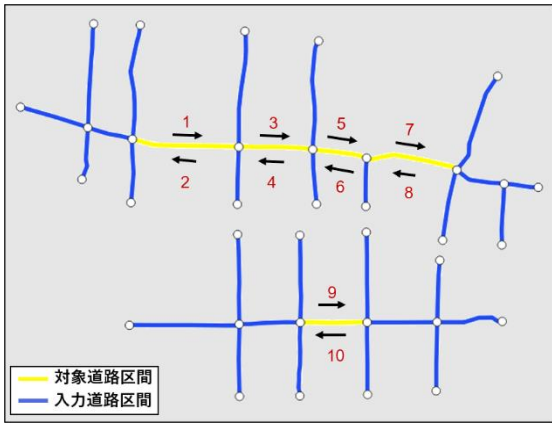


図-5 データ収集道路区間

表-1 学習と検証期間

学習	H28.4~R1.10, 12950 サンプル (10 対象道路区間 X1,295 セット)
検証	R1.11.1~27, 210 サンプル (10 対象道路区間 X21 セット)

表-2 学習パラメータの設定

パラメータ	値
エポック数	1000
学習終了判定	20
誤差関数	平均二乗誤差
学習・テストの割合	0.2
オブティマイザー	Adam

表-3 予測モデルとARIMAモデルの速度誤差 (RMSE) の比較 (時間帯別)

RMSE (km/h)	予測時間帯													
	7時	8時	9時	10時	11時	12時	13時	14時	15時	16時	17時	18時	19時	20時
予測モデル	3.1	2.8	2.6	2.3	2.7	3.2	3.2	3.2	3.3	3.4	4.5	3.7	3.2	3.5
ARIMA	5	3	2.9	3.5	3.7	3.8	3.9	4.1	4.5	5.1	6	3.8	3.8	4

表-4 予測モデルとARIMAモデルの速度誤差 (RMSE) の比較 (対象道路区間別)

RMSE (km/h)	対象道路区間									
	1	2	3	4	5	6	7	8	9	10
予測モデル	3.1	2.4	2.8	2.8	3.9	3.6	3.3	3.4	3.6	2.2
ARIMA	2.8	2.6	4	3.3	5.8	4.6	4.6	5.1	4.1	3.4

表-5 予測モデルとARIMAモデルの速度誤差 (RMSE) の比較 (曜日別)

RMSE (km/h)	予測日							
	月	火	水	木	金	土	日	祝
予測モデル	2.9	3.1	2.6	2.9	2.8	3.2	3.7	5.6
ARIMA	3.6	3.9	3.5	3.5	3.5	3.9	5	6

$$y_t - y_{t-d} = c + \varepsilon_t + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} \quad (1c)$$

ここで c, ϕ, θ は推定するパラメータ、 ε はランダム誤差である。各ケース (道路区間・曜日) の p, d, q パラメータをグリッド検索で指定した。 p (0~10), d (0~3), q (0~4) の値を組み合わせた一連のARIMAモデ

ルをH28.4~R1.10のデータを用いて適合させ、AIC (赤池情報量基準) 値が最も低いモデルを選択した。

この2つ手法の時間帯別・対象道路区間別・曜日別の二乗平均平方根誤差 (RMSE, km/h) を表-3~5に示す。予測モデルの時間帯別RMSEは2.3~4.5と算出されており、各時間帯でも予測モデルの精度が高いが速度は多く変動している17時に少し低下することが示された。対象道路区間別RMSEは2.2~3.9となり、予測モデルがさまざまな道路区間にわたって汎用化されていると思われる。曜日別RMSEは2.6~5.6と算出されており、予測精度は平日の場合高いが土・日は少し低下し、祝日の場合特に落ちていることが示された。予測精度が下がる日とイベント有無の関係は明らかではないがイベントデータ・季節変数・スマートフォン利用者の位置情報・路線バスの乗降客数など祝日の混雑具合に関連すると思われるデータを追加することで、祝日の精度は向上できると考えられる。

また、図-5、図-6 では予測モデルとARIMAモデルそれぞれの散布図 (縦軸: 予測値, 横軸: 実測値) とヒストグラムの比較を示しており、各ケースで予測モデルの精度が高いことが示された。

5. 結論と今後の課題

本研究では、空間的特徴を抽出するGCN層と時間的特徴を抽出するLSTM層を組み合わせた、長期的交通流予測モデルを構築した。そして、ETC2.0, 降雨量, 曜日など、取得可能なデータを用いて予測モデルを学習し、予測精度の検証を行った。構築したモデルでは比較的良い精度で長期予測が可能であることを確認した。今後は、より多くの教師データの取得 (取得期間の延長), 入力変数の検討 (イベント・通行止め・モバイルGPSベースの混雑データ) 等により予測精度の向上を試みる。また、転移学習等による他の地域へ適用について検討を行う予定である。

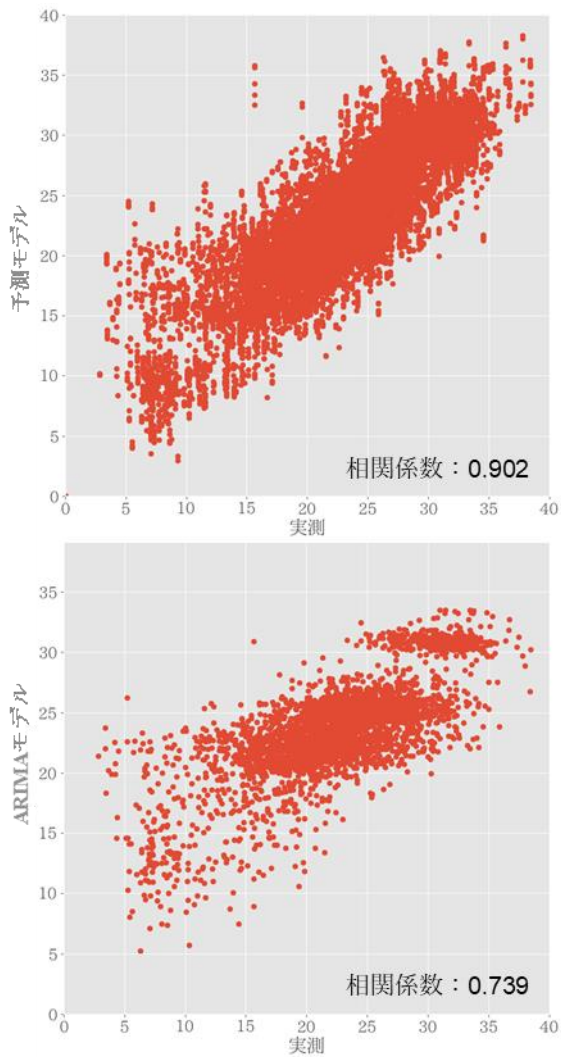


図-5 実測と予測の散布図
(上：予測モデル，下：ARIMAモデル)

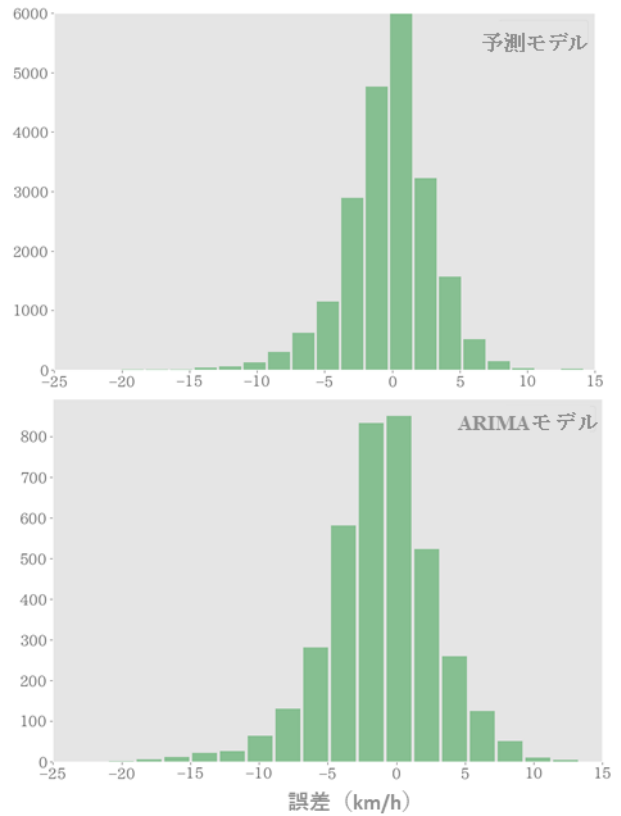


図-6 予測エラーのヒストグラム
(上：予測モデル，下：ARIMAモデル)

参考文献

- 1) Van Hinsbergen, C. P., J. W. Van Lint, and F. M. Sanders : Short term traffic prediction models, *Proc. of 14th ITS World Congress, 2007.*
- 2) 金澤明浩, 杵渕哲也, 毛利仁士, 小川智章, 市河研一, 荒川賢一 : 決定木を利用した交通渋滞予測手法に関する考察, *情処学 ITS 研報, vol.16, no.21, pp.141-148, 2004.*
- 3) Kanoh, H., Furukawa, T., Tsukahara, S., Hara, K., Nishi, H. and Kurokawa, H. : Short-Term Traffic Prediction Using a Fuzzy C-Means and Cellular Automata in a Wide-Area Road Network, *Proc. of 8th International IEEE Conference on Intelligent Transportation Systems, 2005.*
- 4) Shiraishi, T., Horiguchi, R., Kuwahara, M. : A Development of Traffic Prediction System Based on Real-time Simulation, *Conference of Infrastructure Planning Proceedings, Vol. 30, 2005.*
- 5) 森田哲郎 : データマイニングによる交通情報予測システムの開発, *SEI テクニカルレビュー, (178), 45-52.*
- 6) Liang, Z., and Wakahara Y. : City traffic prediction based on real-time traffic information for intelligent transport systems, *13th International Conference on ITS Telecommunications (ITST), 2013.*
- 7) 芦田優太, 西岡到 : 路側データから生成した交通流モデルによる高速道路の交通状況予測, *研究報告モバイルコンピューティングとパーベイシブシステム, 2015.16, 2015.*
- 8) 大川真耶, 金秀明, 戸田浩之 : 畳み込み双線形ポアソン回帰を用いたオンライン交通量予測, *電子情報通信学会論文誌, D 101.4, 2018.*

- 9) 高橋慧, 坂本克好, 山口浩一, 沼尻匠, 曾我部完, 曾我部東馬: 密度球を用いた GraphCNN 深層学習手法による渋滞予測, *人工知能学会全国大会論文集*, 第 33 回全国大会, 2019.
- 10) 柴田文和, 清水英聖, 渡辺柚佳子, 長谷川秀明: ニューラルネットワークを用いた巨視的渋滞予測モデルの高度化と汎用性の検証, *交通工学研究発表会論文集*, 2019.
- 11) 吉田幸司, 井上亮: 道路リンクの接続関係を考慮した
- 畳み込みニューラルネットワークによる自動車交通量の短期予測の検討, *土木学会論文集 D3*, Vol.75, No.5, 2019.
- 12) 小川晃, 福田大輔: 多変量 LSTM による短期交通量予測: 鎌倉市の観光交通を対象として, *第 61 回土木計画学研究発表会・講演集*, 2020.

(2021.6.25 受付)

TRAFFIC SPEED PREDICTION AROUND TOURIST SPOTS USING GRAPH CONVOLUTIONAL NEURAL NETWORK

Huseyin TIRTOM, Itaru MORITA, Hiroki TSUJI, Nobuyuki MURATO, Makoto CHIKARAISHI

Accurate and long term (several days in advance) traffic flow prediction can promote using alternative routes to avoid congestion. Thus, it may help to improve travelers comfort especially around chronically congested touristic hotspots. However, traffic flow prediction is a challenging task due to the complicated spatial dependencies on road networks and time-varying traffic patterns. In this research, we attempt to build a traffic flow (speed) prediction model by combining a graph convolutional network (GCN) and an LSTM (long short term memoy) network. The model was trained using real traffic (ETC2.0) data to predict hourly speed of 10 road sections for the next 7 days from 7:00 to 20:00. Results confirm that the model is able to predict traffic speed with 2.3 to 5.6 km/h average error.