

交通速度の予測における Graph Neural Network

学習の効率化に関する研究

Graph neural network in traffic speed prediction: a study on efficiency of training

緒方 陸^{*1}
Riku Ogata宮崎 利行^{*1}
Toshiyuki Miyazaki菊池 恵和^{*1}
Yoshikazu Kikuchi村野 祐太郎^{*1}
Yutaro Murano菅原 宏明^{*1}
Hiroaki Sugawara^{*1} 八千代エンジニアリング株式会社
Yachiyo Engineering Co., Ltd. #1

Real-time traffic speed prediction and dynamic traffic control are needed to reduce traffic congestion. In recent years, many examples using one of the deep learning methods “Graph Neural Network (GNN)” have been reported. However, the problem is the long training time and large memory usage when the data size is large. Therefore, the important issue is to train models efficiently while aiming for high accuracy. In this paper, we attempt to reduce the training time using the open data, METR-LA dataset and road traffic data in England, in traffic speed prediction. A sensitivity analysis of the training times and accuracies was conducted when the initial values of the adjacency matrix are manipulatively changed, and it was found that the optimal initial values differ depending on the data. For data in England, the method proposed in this paper reduced training time without sacrificing accuracy compared to previous methods.

1. はじめに

道路施策検討や現場管理等への活用等を目的に、国土交通省はオープンなデータプラットフォームとして xROAD の構築を進めている。将来的には xROAD 上で、ETC 2.0 プローブ情報や交通量を含む道路データが API としてリアルタイムで公開される予定である[国土交通省 2022]。交通量等のデータ公開が進み、交通予測技術と組み合わせることで、渋滞発生回避や動的な交通制御が可能となる。

近年では機械学習を用いた交通予測に関する研究も盛んに行われており、国内では LightGBM [宮崎 2022] や LSTM [小川 2021] を用いた例がある。また最近では海外をはじめ、Graph Neural Network (GNN) を用いた例も多い ([Yu 2018], [Wu 2019] など)。[Wu 2019] が用いた Graph WaveNet は隣接行列自体を学習パラメータとしており、昨今ではこの方法が主流となってきている。

GNN による交通予測では、対象とする地点の入力を一度に読み込むため、大規模データを用いると大きいメモリ容量が必要となることや学習に長時間を要することが課題となる。ここで、GNN の計算において隣接行列はノード間の関係を表すものであるが、隠れ層などのパラメータと同様、その初期値により損失の収束の速さが異なると考えた。しかしながら、著者らが確認した限りでは、時系列予測における GNN において隣接行列の初期値について研究された例は無い。

本稿では、学習時間の削減を目的とし、隣接行列の初期値と収束の速さや精度の関係から感度分析を行い、学習効率化の評価を実施した。

2. 既往研究

カリフォルニア州の交通データである PeMSD7 や METR-LA など各種ベンチマークデータセットを用いた検討において、

GNN は高い精度を発揮している ([Yu 2018], [Wu 2019] など)。[Yu 2018] は従来手法である LSTM などのリカレントネットワークでは地点間の空間的な特徴を得られないことが課題と述べており、これに対し、効率的に空間的な特徴量の取得することを目的として位置情報を用いた式(1)から隣接行列を生成している。なお、 w_{ij} は行番号 i 、列番号 j のパラメータ、 d_{ij} は地点 i, j のユークリッド距離、 σ^2 と ϵ は隣接行列を制御するための閾値であり、[Yu 2018] はそれぞれ 10 と 0.5 が用いている。

$$w_{ij} = \begin{cases} \exp\left(-\frac{d_{ij}^2}{\sigma^2}\right), & i \neq j \text{ and } \exp\left(-\frac{d_{ij}^2}{\sigma^2}\right) \geq \epsilon \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

しかし実現現象としては地理空間情報のみでは地点間の関係性を全て表すことはできない。この問題に対し、[Wu 2019] は “self-adaptive adjacency matrix” という手法をネットワークに組み込み、隣接行列自体を学習パラメータとすることで地点間の関係を捉えることで問題解決に貢献した。交通予測 GNN においてこの方法は主流となっており、[Wu 2019] はノード埋め込みベクトルの初期値に、正規分布に従う乱数を与えている。他に、時系列データの異常検知タスクにおいて、コサイン類似度を用いて隣接行列の初期化を行ったものもある ([Deng 2021])。また、隣接行列を学習パラメータとして扱わない時系列予測においては、シンプルな GCN による株価予測タスクで、事例としては多くないものの、相関係数や Dynamic Time Warping (DTW) を隣接行列として採用した例もある ([Loe 2021], [Peng 2021])。

ニューラルネットワークのパラメータ初期値の検討としては [Glorot 2010] や [He 2015] があるが、先述の通り時系列予測 GNN において隣接行列初期値について研究された例は無い。相関係数や DTW により生成した初期値でデータ間関係を捉えられれば、交通予測に限らず時系列予測において同様の方法で GNN の学習を効率化できる可能性がある。本稿では [Wu 2019] の Graph WaveNet を用い、隣接行列初期値の違いが学習時間や精度に与える影響を明らかにすることを目的として、隣接行列初期値と学習時間および精度の感度分析を実施した。

連絡先: 緒方陸, 八千代エンジニアリング株式会社, 東京都台東区浅草橋 5-20-8 CS タワー 技術創発研究所, TEL: 03-5822-6395, Mail: rk-ogata@yachiyo-eng.co.jp

3. 実験

3.1 データセット

本稿では[Wu 2019]の Graph WaveNetを用い、METR-LAと Englandの2種類のデータセットで検討を実施した。

METR-LAは[Li 2018]が公開したデータセットであり、ロサンゼルス的高速道路に設置された207の速度センサーにより計測されたデータである(図1)。本稿で用いる Graph WaveNetの論文([Wu 2019])においてもベンチマークデータセットとして用いられている。データは5分単位で整備されており、2012/3/1から2012/6/27までの約4ヶ月間のデータである。

また、[宮崎 2022]が使用した「WebTRIS Traffic Flow API」よりダウンロードした Englandのオープンデータも用いて検討を実施した([Highways England 2022])。収集した全15,743地点のデータのうち、アクティブ状況や欠損状況等を踏まえ、Birmingham Town Hallから半径30km以内の範囲の計170地点を検討対象地点とした(図2)。なお、COVID-19が交通状況に影響を与えた可能性を考慮し、2016年から2017年を学習データ、2018年を検証データ、2019年をテストデータとして用いた。データは15分単位で整備されており、平均速度(Avg mph)と合計台数(Total Volume)を用いた。なお、入力データには現況から2時間45分前まで3時間分(12時刻)の平均速度と合計台数、および時刻(15分単位時刻を正規化、0~1)と曜日のダミー変数(0~6)を用いた。出力は1時刻から12時刻先(15分先から180分先)の全12時刻分の平均速度とした。

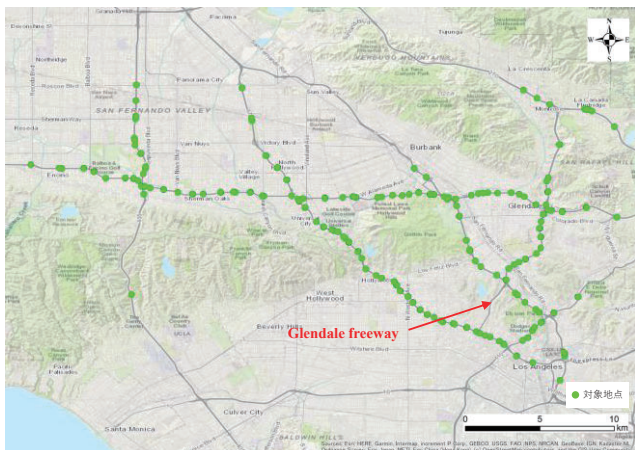


図1 METR-LA 対象地点

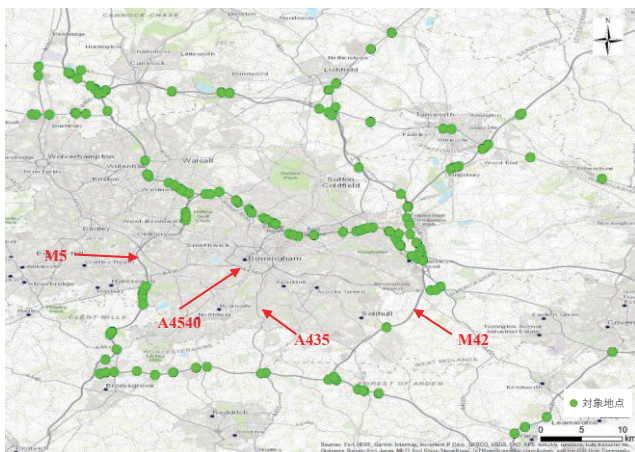


図2 England 対象地点

3.2 モデル

交通予測において、GNNのうち Graph Convolutional Networks (GCN)が良く用いられる。GCNの隣接行列と入力・出力の関係は、下の式で表される([Wu 2019])。

$$Z = \tilde{A}XW \quad (2)$$

ここで、 $Z \in R^{N \times M}$ は出力、 $\tilde{A} \in R^{N \times N}$ は隣接行列、 $X \in R^{N \times D}$ は入力、 $W \in R^{N \times M}$ はモデルのパラメータを表す。Graph WaveNetでは隣接行列として、先述の”self-adaptive adjacency matrix (\tilde{A}_{adp})”という方法を取り入れ、データ間の関係を取得することで精度向上を達成した([Wu 2019])。

$$\tilde{A}_{adp} = \text{SoftMax}(\text{ReLU}(E_1, E_2^T)) \quad (3)$$

ここで $E_1, E_2 \in R^{N \times c}$ はソース/ターゲットノードに対し埋め込みをしたベクトルを表す。[Wu 2019]はこれらを正規分布に従う乱数により初期化し、学習を行った。学習時にはReLU関数により弱い関係は除かれ、各地点間の関係を表す数値はSoftMax関数により正規化される。以上により空間に依存しない地点間関係を取得する構造としたことで高い精度を発揮した。

本稿では以上で紹介した Graph WaveNetを用い検討を実施した。詳細については[Wu 2019]の論文を参考にされたい。

3.3 実験方法

本稿では Graph WaveNetを用い、隣接行列の初期値を変化させた際の学習時間や精度への影響を評価した。具体的には、既往研究で用いられている以下のうち(1)Norm および(4)Euclidに加え、(2)Uni, (3) Corr, (5)DTWの方法で初期値を設定し、学習時間や精度への影響の感度分析を実施した。

- (1) Norm: 式(3)の E_1, E_2 を正規分布に従う乱数で初期化
- (2) Uni: 式(3)の E_1, E_2 を一様分布に従う乱数(0-1)で初期化
- (3) Corr: ピアソンの積率相関係数で初期化
- (4) Euclid: 位置情報から式(1)により隣接行列を初期化
- (5) DTW: Dynamic Time Warping (DTW)により距離を算出し、式(4)により算出した値で隣接行列を初期化(後述)

ここで、GNNの隣接行列はノード間の関係を表しているため、予め各データ間の関係を初期値として与えることで、学習を効率化できると考えた。

DTWは時系列データ間の類似度をデータ間距離により算出する手法であり、時系列データ間の関係を捉える方法である。DTW適用例のイメージを図3に示す。DTWの計算においては、まずデータ間(data1, data2)の各time stepの距離を総当たりで求め、最短となる距離(図中赤点線)を算出する。さらに総和をとることで類似度を計算する。これにより、時系列同士の長さや周期が異なる場合でも類似度を算出できる。本稿ではDTWを高速化したPythonライブラリ”fastdtw”([Salvador 2004])を用い、各データ間の類似度を算出し、式(4)により隣接行列の初期値を設定した。

$$w_{ij} = 1/(dtw_{ij} + \alpha) \quad (4)$$

w_{ij} は行番号 i 、列番号 j のパラメータ、 dtw_{ij} は地点 i, j の類似度、 α は制御パラメータを表し、ここではMETR-LAは 10^5 、Englandは 10^6 とした。 dtw_{ij} を逆数として与えた理由としては、DTWでは類似度の高いデータ間の数値ほど小さい値となる一方、隣接行列においては関係性の高いデータ間の数値ほど高い数値となるためである。なお、DTWの性質上、対角成分は0となるため、対角成分には非対角成分の平均を与えた。なお、隣接行列初期値以外のハイパーパラメータはデフォルトとし、エポック数は150とした。

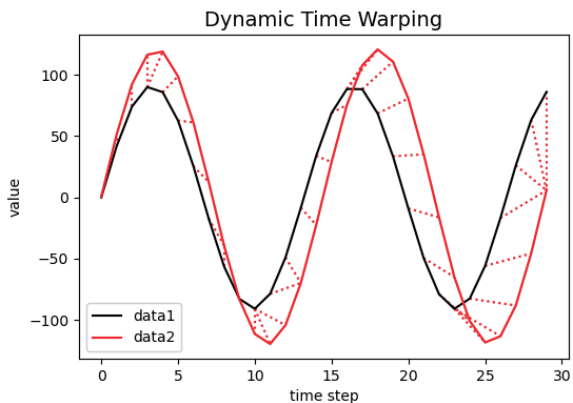


図3 Dynamic Time Warping 計算例

4. 結果・考察

4.1 実験結果

結果を図4および図5に示す。なお、学習曲線は大きく振動するため、図4はloss最小値が更新した場合のみを示した。また精度評価には平均絶対パーセント誤差(MAPE)と決定係数(R^2)を用いた。表1には validation loss の収束にかかった時間と全予測時刻平均精度(MAPE/ R^2)を示す。

図4(左図)より、METR-LAでは、(5)DTWが最も学習時間が長い一方、(4)Euclidは最も学習時間が短い結果となった。また図5(左図)および表1からは(5)DTWで精度が低く、(3)Corrと(4)Euclidの精度が高いものの、(2)~(4)で精度は大差無い結果となった。表1に示す本稿の学習環境において学習時間が最長・最短の(4)Euclidと(5)DTWの validation loss が最小値に達するまでに要した時間を比較すると、72 minの差があった。

Englandのデータでは、図4(右図)より(2)Uniで初期値を与えた場合の収束が最も遅く、(5)DTWで初期値を与えた場合が最も収束が速かった。図5(右図)および表1より精度は(3)Corrと(5)DTWが他に比較してややMAPEが低い値を示した。表1より学習時間が最長・最短の(2)Uniと(5)DTWの validation loss が最小値に達するまでに要した時間を比較すると、83 minの差があった。

また、表1の(3)Corrに着目すると、両データセットにて学習時間は平均的な性能を、精度についてはやや高い値をとっている。

4.2 結果の考察

二つのデータセットの結果を比較すると、METR-LAでは(4)Euclidが効果を発揮している一方、(5)DTWは学習時間、精度ともに5ケースのうち最低の結果となった。対してEnglandのデータではMETR-LAとは逆に(5)DTWの収束が速い。ここで図1を見ると、METR-LAの観測地点がGlendale freewayを除く対象地域内の幹線道路を網羅しており、観測機器も密に設置されている。そのため、位置情報から算出したユークリッド距離は実空間上での近接関係を表し、値が小さい(隣接行列の数値が高い)地点間においては、交通速度予測において有効な特徴を有すると考えた。これを初期値として与えることで、METR-LAの予測においては(4)Euclidが効果を発揮したと推察した。図5(左図)MAPEの指標において、他と比較して(4)Euclidが長時間先になるほど精度が落ちるのもこれに起因すると推察した。

一方で図2を見るとEnglandのデータは幹線道路のA435やA4540にはデータが無く、M5やM42の道路はデータを有する地点間隔が疎になっている。また前述の通り、Englandのデータにおいては(4)Euclidが有効に働いていない。以上よりデータセットの地点密度が疎である場合、必ずしもユークリッド距離が近い地点が予測に寄与する特徴を有しないため、(4)Euclidで隣接行列初期値を与える方法は有効でなかったと推察した。対して(5)DTWはデータ間の距離(類似度)を測るため、時系列データの波形が類似していると値は小さくなる。この算出方法から、地点が離れていても地点間の関係性を特徴として捉えることができると考えた。以上から地点密度が疎なEnglandのデータにおいては、実空間上のユークリッド距離よりもDTWで算出した距離が近い方が有効な特徴を有し、これを基に生成した初期値を与えたことが学習時間削減、精度向上に繋がったと推察した。

また、(3)Corrのケースは、各データセットの最も効果を発揮したケース(METR-LA; (4)Euclid / England; (5)DTW)には及ばないものの、それらに次ぐ効果を発揮している。相関係数はその算出方法から、標準的な平均値との距離を示している。このため、(4)Euclidや(5)DTWのように地点間/データ間の直接的な関係は得られないものの、平均的なデータ間関係を取得できた

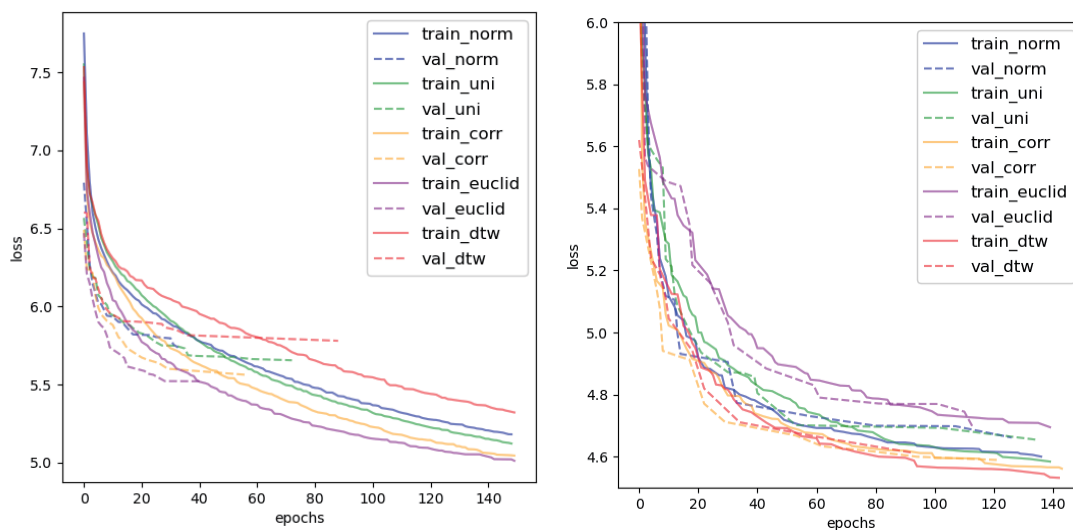
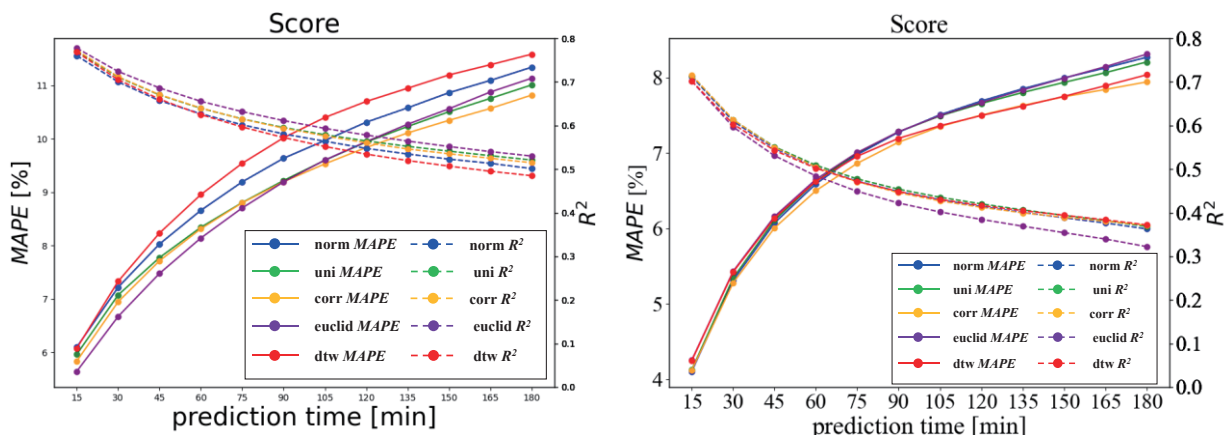


図4 学習曲線(左図:METR-LA, 右図:England)(実線:train, 点線:validation)

図5 精度(左図:METR-LA, 右図:England) (実線:MAPE, 点線:R²)

と考えた。相関係数を隣接行列初期値として与えることで大きな貢献は無いものの、様々なデータセットに対して平均的に学習時間削減や精度向上に貢献できるものと推察した。

以上より、データにより最も有効な初期値設定方法は異なるものの、(3)Corrの方法により平均的に学習時間削減・精度向上へ貢献できる可能性を示した。

表1 Validation loss 収束にかかる時間と精度および学習環境

	METR-LA		England	
	Convergence time	MAPE/R ²	Convergence time	MAPE/R ²
(1)Norm	80 min	9.41/0.62	383 min	6.98/0.47
(2)Uni	69 min	9.10/0.62	410 min	6.97/0.47
(3)Corr	89 min	9.00/0.61	368 min	6.82/0.47
(4)Euclid	65 min	9.02/0.62	401 min	7.02/0.45
(5)DTW	137 min	9.70/0.58	327 min	6.89/0.47
Environment	vCPU: 4/ Memory: 16 GiB / NVIDIA T4 Tensor Core GPU		vCPU: 16/ Memory: 64 GiB / NVIDIA T4 Tensor Core GPU	

5. おわりに

本稿では、交通予測における GNN の学習効率化を目的とし、隣接行列初期値を変化させた場合の学習時間や精度への影響の感度分析を実施した。結果から、本稿で用いた METR-LA のデータにおいては位置情報からユークリッド距離を用いて生成した初期値で、England のデータにおいては DTW により生成した初期値で学習させることで学習時間短縮や精度向上に繋がることを示した。また、相関係数を初期値として与えることでも学習時間短縮や精度向上に貢献する可能性を示した。

今後の課題として、本稿で用いたネットワークは Graph WaveNet のみであり、データによっても性能が異なる可能性がある。今後は、異なるネットワークやデータを用いた場合に初期値が与える学習効率や精度への影響について検討する。また、今回実空間上のユークリッド距離や DTW により算出した距離が近い地点の特徴の交通速度予測に対する貢献度を定量的に評価できていない。これについても今後の課題とする。

参考文献

[Deng 2021] Ailin Deng, Bryan Hooi: Graph Neural Network-Based Anomaly Detection in Multivariate Time Series, arXiv preprint arXiv:2106.06947, 2021.
 [Glorot 2010] Xavier Glorot and Yoshua Bengio: Understanding the difficulty of training deep feedforward neural networks,

International Conference on Artificial Intelligence and Statistics, 2010.

[He 2015] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. ICCV, 2015.

[Highways England 2022] Highways England: WebTRIS Traffic Flow API, <https://webtris.highwaysengland.co.uk/api/swagger/ui/index>. (2023/02/24 アクセス)

[Li 2018] Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu: Diffusion convolutional recurrent neural network: Data-driven traffic forecasting, ICLR, 2018.

[Loe 2021] Dylan Loe, Sung-lin Chang, and Jason Chau: Stock Market Movement Prediction Using Graph Convolutional Networks, UCSD Data Science Capstone Projects: 2020-2021, 2021.

[Peng 2021] Shuyi Peng: Stock Forecasting using Neural Network with Graphs, University of York Computer Science, 2021.

[Salvador 2004] Stan Salvador and Philip Chan: FastDTW: Toward Accurate Dynamic Time Warping in Linear Time and Space, Intelligent Data Analysis, 2007.

[Wu 2019] Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, Chengqi Zhang: Graph WaveNet for Deep Spatial-Temporal Graph Modeling, arXiv preprint arXiv:1906.00121v1, 2019.

[Yu 2018] Bing Yu, Haoteng Yin, Zhanxing Zhu: Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting, arXiv preprint arXiv:1709.04875v4, 2018.

[小川 2021] 小川晃平・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, 土木学会論文集 D3 (土木計画学), 土木学会 土木計画学研究委員会, 2021.

[国土交通省 2022] 国土交通省 道路局: 道路システムの DX～xROAD の推進～, 第 31 回国と地方のシステムワーキング・グループ 会議資料, 2022.

[宮崎 2022] 宮崎利行・大沢明瑠・菊池恵和・菅原宏明:イングランドの交通オープンデータと AutoML による短時間渋滞予測, AI・データサイエンス論文集, 土木学会 構造工学委員会, 2022.