

自然言語処理モデルを活用した降雪時における交通障害の発生予測

南 紘雅*¹ 猪井 博登*²

富山大学大学院 理工学研究科*¹

富山大学 学術研究部 都市デザイン学系*²

大雪時における自動車の立ち往生などの交通障害への対策が求められるなか、交通障害の発生自体を予測することが期待されている。本研究では、プローブデータや気象データなどをもとに自然言語処理モデル GPT を活用した機械学習を行い、その有用性と予測精度の向上方法について検証した。その結果、一定の信頼性が得られたほか、学習データに用いた用語の説明を加えて学習させた場合に予測精度が向上するなどの傾向が確認できた。加えて、地域によって予測精度が異なり、交通量などの地域特性による影響が示唆された。今後、リアルタイムデータを活用した予測方策の検討を行う際に本研究の枠組みが活用できると期待される。

Prediction of Traffic during Snowfall Using NLP Models

Hiromasa Minami*¹ Hiroto INOI*²

Graduate School of Science and Engineering, University of Toyama*¹

Faculty of Sustainable Design, University of Toyama*²

As countermeasures against traffic disruptions such as stranded vehicles during heavy snowfall are required, it is expected that the occurrence of traffic disruptions themselves can be predicted. In this study, machine learning using the GPT natural language processing model was conducted based on probe data and weather data, and its usefulness and methods for improving prediction accuracy were verified. As a result, a certain degree of reliability was obtained, and a trend such as an improvement in forecast accuracy was confirmed when the model was trained with explanations of terms used in the training data. In addition, the prediction accuracy varied by region, suggesting the influence of regional characteristics such as traffic volume. It is expected that the framework of this study can be utilized in future studies of forecasting measures that utilize real-time data.

Keyword: Heavy snow, ETC2.0 Probe data, GPT (Generative Pre-trained Transformer)

1. はじめに

近年、都市部における大雪時の交通障害が大きく

報じられるなか、地方部では毎年のように各地で立ち往生などの交通障害が発生している。2020年12月

から 2021 年 2 月にかけての冬期期間、日本海側を中心に記録的な大雪が発生し、市民生活に大きな影響を与えた。特に、2021 年 1 月に発生した大雪では、北陸地方を中心に各地で立ち往生などの交通障害が発生し、大雪時における自動車交通への対策が大きな課題として取り上げられた。

しかし、近年、短時間において集中的な降雪により大雪が発生するケースが増加しており、それらに十分対応できる除雪体制の構築は、現状においては困難である¹⁾。よって、現時点では大雪時における交通障害の発生を抑制するには、第一に交通量の削減を図る必要がある。そのため、道路情報をドライバーへ迅速かつ正確に発信することが求められる。現在、リアルタイムでの渋滞情報や規制情報はカーナビゲーションシステムなどで確認することができ、大雪時においてもこれらを活用できる。しかし、これらは速度低下や交通規制が発生した後に提供されることから、大雪時における交通障害の発生を抑制する根本的な解決には至らない。

そこで、本研究では機械学習を活用し、交通障害の発生予測が可能かを検証する。機械学習を活用した交通需要予測に関する研究として、小林ら²⁾はスマートフォン上で取得した位置情報データをもとに自然言語処理モデル GPT-2 を適用した人流のモデルを構築し、メッシュ単位での交通需要を予測する手法を提案した。また、Jin³⁾らは BERT(Bidirectional Encoder Representations from Transformers)を用いて交通ビッグデータを学習させた様々な道路に適したモデルを提案し、交通流予測の有用性を検証した。このような平常時における将来予測に関する研究は散見されるものの、記録的な大雪を含む、積雪時における交通障害の発生予測に関する研究は見られない。また、近年幅広い分野で活用が期待されている自然言語処理モデル GPT がある。GPT は 2018 年に OpenAI によってリリースされて以降、様々な分野において研究が進められているが、依然黎明期である。特に、交通工学分野においては GPT を含む言語処理を活用した研究自体極めて少ない。

これらのことから、本研究では自然言語処理モデル GPT を活用した機械学習を行い、大雪時を含む降雪時における交通障害の発生を予測し、その有用性と予測精度の向上方法について検証する。なお、分析対象期間のうち大雪による影響を受けたとされる期間とその他の降雪期間の両方を予測対象とすることから、大雪時に限らず、降雪時における交通障害の発生予測とする。

2. 対象とする地域・期間及び学習データの概要

2-1 対象とする地域の概要

本研究では、富山県富山市を分析対象地域とした。富山市は全域が豪雪地帯又は特別豪雪地帯に指定されており、年度により変動はあるものの、例年数十センチから数百センチの累計降雪量が観測されている。また、富山地方气象台によると、積雪深が 250cm に達した昭和 38 年の 38 豪雪や死傷者が千人を超えた昭和 56 年の 56 豪雪をはじめ、過去には記録的な大雪災害が発生している。近年においては、平成 30 年に累計積雪量が 430cm に達する大雪も観測されている⁴⁾。

2-2 対象とする期間の概要

本研究では、2021 年 1 月 1 日から 31 日までの 1 か月間を分析対象として選定した。このうち、14 日間は降雪が記録されている⁵⁾。また、1 月 7 日から 11 日にかけて、断続的な降雪により 1986 年以来 35 年ぶりに積雪が 100cm を超えるなど、記録的な大雪災害となった⁶⁾。市内の在来線全区間が運休し、北陸新幹線が開業以降初めて大雪により運休するなど、交通機関への影響も大きかった。道路交通においては、生活道路をはじめ、高速道路や幹線道路など、各地で立ち往生が多発した。特に、降雪が強まった 1 月 7 日の夕方以降が平日の帰宅時間帯であったことから、スタックの発生やそれに伴う渋滞が顕著となった。富山県によると、除雪作業や通行規制など、大雪による影響は降り始めの 10 日後の 1 月 17 日まで続いたと報告されている⁷⁾。

2-3 学習データの概要

(1) ETC2.0 プローブデータ

ETC2.0 プローブデータの走行履歴と挙動履歴ではデータを記録するタイミングが異なる。時刻や位置情報(緯度経度)を含む走行履歴は、200m 毎または 45 度以上方向転換した際に記録される⁸⁾。一方、ヨー角速度、左右加速度、前後加速度を含む挙動履歴は、それぞれの閾値を超えた時の最大値が記録される。なお、それぞれの閾値を超える挙動には、車線変更や定速走行時の挙動が含まれる可能性は非常に低いことが検証されている⁹⁾。よって、異常走行の検出における閾値設定としては妥当だとされている。

本研究では、ETC2.0 プローブデータ出力様式 1-4 挙動履歴情報及び出力様式 2-3 道路プローブ DRM 区間単位時間帯別平均旅行時間旅行速度を用いて分

析を行う。挙動履歴においてはヨー角速度、左右加速度、前後加速度のいずれかが閾値を超えた際に記録されるため、すべてが閾値を超えるデータと1つまたは2つのみが閾値を超えるデータが混在しているため、データ同士の比較を行う際には注意する必要がある。また、平均旅行速度はデジタルロードマップ（以下、「DRM」という）のリンクごとの平均速度が1時間単位で集計されるが、時間帯によっては件数が0件のため、データが欠損している場合があることについても留意する必要がある。

(2) メッシュ気象データ

降雪時における交通障害の発生は、気象条件による影響も大きいと考えられる。そこで、本研究では解析雨量、気温、相対湿度を用いた。降雪による影響を評価する指標として、解析雨量、解析降雪量、解析積雪深を用いることが考えられる。しかし、解析降雪量及び解析積雪深は5kmメッシュで推定されたものであり、富山市内を対象とした本研究では精度が確保できない。また、気象庁も、メッシュの平均値であり、局地的な降雪の多寡を表現できないことや量的には過小に予想する傾向があることから、値を直接活用することを推奨していない¹⁰⁾。これらのことから、本研究では1kmメッシュで推定され、比較的精度が担保できる解析雨量を用いる。なお、解析降雪量や解析積雪深は解析雨量などを用いて算出されることから、降雪時の気象状況の把握には解析雨量をも活用できると考えられる。また、松尾¹¹⁾によると、雪、みぞれ、雨の区別は気温と相対湿度で説明される。さらに、道路交通に関する研究においても、舟田¹²⁾は路面温度と路面湿度をもとに路面凍結予測モデルを構築している。これらのことから、気温及び相対湿度が降雪を説明できる指標であると言える。

(3) 除雪車両走行履歴データ

降雪時の路面状況を説明する指標として気象条件の他にも、除雪作業の有無が交通障害の発生に大きく影響すると考えられる。除雪車両の走行履歴に関するヒアリング調査を実施した結果、市道においては、除雪車両が稼働した全時間帯の走行履歴データが全ての車両台数分記録されていた。そこで、本研究では位置情報の精度が比較的高く、且つ広範囲の除雪状況を把握できる市道における除雪車両の走行履歴データを用いることとした。

富山市へのヒアリング調査の結果、2021年1月7

日から1月14日までの期間における除雪車両の走行履歴データの提供を受けた。データは業者番号及び機械番号、日時、緯度経度で構成されている。除雪車両は合計で約800台にのぼり、それぞれの車両の位置情報が5秒間隔で記録されている。

(4) 断面交通量データ

降雪時、特に大雪時においては交通量が交通障害の発生に大きく影響すると考えられる。一般に、交通量が増加するにつれて通過速度は低下し、立ち往生のリスクが高くなる。2021年1月に発生した大雪においても、交通量の多い幹線道路を中心に立ち往生などの交通障害が多発したと報告されている⁷⁾。そこで、本研究では日本道路交通情報センター（JARTIC）から提供を受けたオープンデータに加え、公益財団法人日本交通管理技術協会が提供する計測地点の位置情報をもとに分析を行った。富山市内に計測地点は計505か所あり、データは5分毎に記録される。ただし、計測地点がなく、断面交通量を把握できない地域があることに留意する必要がある。

3. 自然言語処理モデルによる予測

3-1 学習データの整理

ETC2.0プローブデータ、メッシュ気象データ、除雪車両走行履歴データ、断面交通量データのいずれも3次メッシュ単位の1時間毎のデータに加工し、分析を行うこととした。これは、メッシュ気象データが1kmメッシュで推定され、学習データの粒度の統一を図るためである。また、路線単位などで分析した場合には、欠損値や外れ値を含む可能性が高く、予測精度に影響を与える可能性が懸念される。

ETC2.0プローブデータにおいては、挙動履歴と走行履歴でデータの扱いが異なる。挙動履歴においては、閾値を超えた場合の発生地点をGPSによって測位された緯度経度をもとにマップマッチングを行う。南ら¹⁾の研究では、挙動履歴情報の記録件数が交通障害の発生を説明する指標として有用であると示したが、山間部における急カーブなどで閾値を超えるデータが記録され、大雪の影響によらないデータが含まれるという課題が残った。そこで、本研究では菊池ら¹³⁾の研究をもとに、閾値設定を以下のように設定した（表1）。これらの閾値をすべて超えるデータの記録件数をメッシュごとに算出した。一方、本研究で用いる走行履歴の旅行平均速度は区間の平均値であるため、単一メッシュ内のリンク旅行平均速度の平均値を算出した。

表 1 挙動履歴情報の閾値設定

ヨー角速度	±8.5deg/s
前後加速度	-0.3G
左右加速度	±0.3G

その他、メッシュ気象データや除雪車両走行履歴データの記録件数も同様に3次メッシュ単位の1時間毎のデータに整理した。また、断面交通量はメッシュ内の1計測地点あたりの平均断面交通量とした。

3-2 GPT モデルのファインチューニングと予測

降雪時における交通障害の発生予測において、自然言語処理モデルを活用するメリットとして以下のようなことがあげられる。

1. 局所的な降雪や短時間の降雪への対応が期待できる。降雪時においてはデータ特性が急変する可能性が高く、空間的・時間的自己相関性が期待できない。また、画像認識等による従来の手法では予測が困難である。2. 粒度が異なるデータ、文章データなどの学習が可能。これにより、対話型 AI サービスへの援用が期待できる。3. 自然言語処理モデルの交通工学分野への適用可能性の検証。

自然言語処理モデル GPT-1 及び GPT-2 は OpenAI によってオープンソースとして公開されているが、GPT-3 以降は有償の API 経由でのみ提供している。また、トレーニングされた既存の大規模言語モデルを特定のタスクやドメインに適応させるファインチューニング (Fine-Tuning) はトークン数に応じて利用料金が定められている。本研究では、オープンソースの GPT-2 モデルを用いてファインチューニングを行い、交通障害の発生を予測する。なお、GPT-2 の実装は Hugging Face transformers を用いた。予測手法はおおよそ以下の通りである。

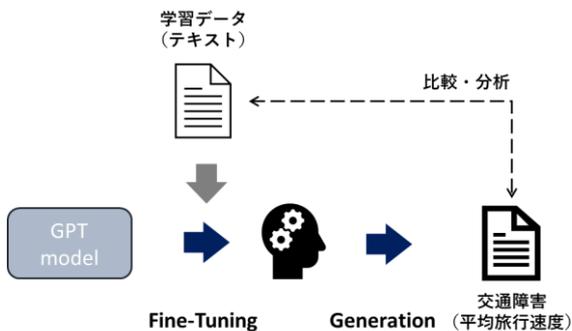


図 1 GPT モデルを活用した予測手法

4. 予測結果と考察

4-1 有用性の検証

交通障害の発生を予測するにあたり、交通障害を定義し、予測する指標を明確にする必要がある。南からは、2021年1月に発生した大雪において、自動車の立ち往生などにより平均旅行速度が著しく低下し、交通障害の発生に影響していることを明らかにした。そこで、本研究では平均旅行速度を交通障害として予測を行う。

ファインチューニングされたモデルによって出力される平均旅行速度の予測結果を10に設定し、それらの中央値または平均値を正解データと比較し、相関係数、決定係数及び二乗平均平方根誤差 (RMSE) を用いて評価する。なお、2021年1月7日から11日にかけての5日間を大雪時、1日、2日、6日、29日、30日を降雪時とし、両期間における交通障害の発生を予測した (表 2)。

表 2 大雪時と降雪時の定義

日付	降雪 (合計)	最深積雪 (値)	日付	降雪 (合計)	最深積雪 (値)
1	降雪時 ↑ 6	23	17	4 ↓	53
2	18	33	18	9	59
3		32	19	6	64
4		26	20		61
5		21	21		57
6	降雪時 ↑ 1	18	22		55
7	26 ↑	40	23		50
8	61	100	24		47
9	大雪時 ↑ 31	127	25		44
10	6 ↓	128	26		42
11	10 ↓	125	27		37
12		109	28		32
13	大雪による影響を受けた	96	29	降雪時 ↑ 17	47
14	と報告される期間	84	30	降雪時 ↓ 6	53
15		74	31		42
16	1	66			

ここでは、富山市中心部を含む3次メッシュコード 55370126、55370127、55370136、55370137 の4メッシュの範囲内を対象として予測した結果について述べる。なお、対象時刻は2021年1月に発生した大雪において最も交通障害の発生が多かったとされる午前9時台とした。

予測の結果、大雪時を対象とした場合において、実測値と予測値間の相関係数は0.5958、決定係数は0.3550、RMSEは7.0622であった。一方、降雪時を対象とした場合においては相関係数は0.4892、決定係数は0.2393、RMSEは13.8935であった。これらのことから、自然言語処理モデルを活用した交通障

害の発生予測においては一定の精度は認められると言える。また、降雪時と比較して大雪時を対象とした予測の場合に、精度が高くなることが確認できた。これは、学習データが大雪時における交通障害の発生を説明する指標で構成されており、記録的な大雪時を対象とした予測の方が適していることが要因だと推測される。

表3 予測精度の比較

	大雪時	降雪時
R(Avg.)	0.5958	0.4892
R(Med.)	0.4379	0.3059
R ² (Avg.)	0.3550	0.2393
R ² (Med.)	0.1918	0.0936
RMSE(Avg.)	7.0622	13.8935
RMSE(Med.)	7.1645	12.3926

4-2 予測精度の向上に関する検証

入力する単語の組み合わせや質問を工夫して、意図するコンテンツを出力させる手法をプロンプトエンジニアリングという。本研究においても、学習データをどのような形式で入力すると予測精度の向上につながるのかを検証した。

本研究における学習データはテキスト形式で入力している。そこで、学習データ内で用いられている用語の説明を学習データ自体に追加し、予測精度の向上に影響があるかを検証した。その結果、RMSEは10.1543から7.9136になり、予測精度が向上することが確認できた。

表4 予測精度の向上

	追加前	追加後
R(Avg.)	0.3372	0.3844
R(Med.)	0.5383	0.6138
R ² (Avg.)	0.1137	0.1478
R ² (Med.)	0.2897	0.3768
RMSE(Avg.)	10.1543	7.9136
RMSE(Med.)	10.1492	9.2654

その他、外れ値の出力を除外するため、平均旅行速度の値が取り得るおおよその出力範囲を指定した。また、地域メッシュ単位（1次メッシュと2次メッシュ、2次メッシュと3次メッシュ）に区別を付け、他の数値データとの区別を行った。しかし、これらを反映した入力方法では、有意な外れ値の減少は見られなかった。

シユ、2次メッシュと3次メッシュ）に区別を付け、他の数値データとの区別を行った。しかし、これらを反映した入力方法では、有意な外れ値の減少は見られなかった。

4-3 予測結果と地域特性

学習データのうち、プローブデータや除雪車両走行履歴データ、断面交通量データは、データの記録がないことから、欠損となっているメッシュが多数存在する。これらの欠損値の影響によりメッシュごとに予測精度の差が生じる可能性が考えられる。そこで、市内中心地域と郊外地域において予測精度を比較した。

中心地域においては、予測値と実測値の乖離が小さく、その誤差の平均値は4.43km/hであった。一方、郊外地域においては、中心地域と比較して乖離が大きく、誤差の平均値は9.68km/hであり、最大14.85km/hの誤差が生じた。これらのことから、中心地域と比較して、郊外地域における交通障害の発生予測は精度が低くなることが確認できた。これは、前述の通り、郊外地域においては学習データの欠損箇所が多いことに起因すると考えられる。

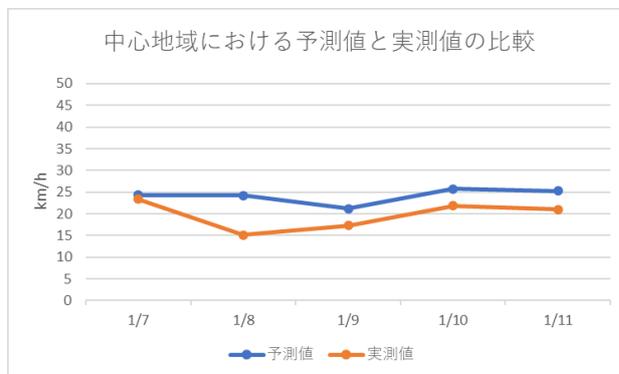


図2 中心地域における予測値と実測値

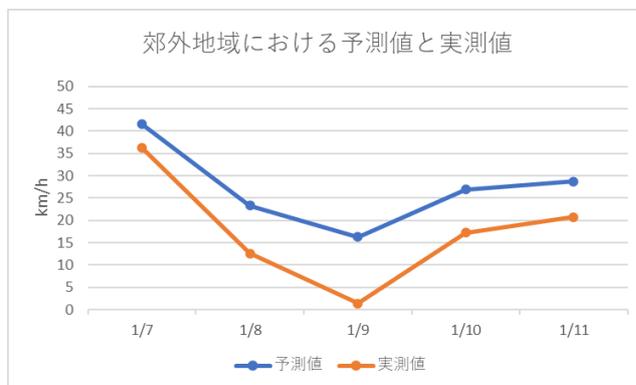


図3 郊外地域における予測値と実測値

5. おわりに

本研究では、自然言語処理モデル GPT を活用した交通障害の発生予測を行い、その有用性と予測精度の向上方法について検証した。その結果、交通障害の発生予測が一定程度可能であることを示した。特に、通常の降雪時と比較して、記録的な大雪時においてその予測精度が向上することが確認できた。また、学習データに関する用語の説明を追加した場合に予測精度が向上することが明らかになった。さらに、市内中心地域と比較して、学習データの欠損箇所が多い郊外地域において予測精度が低下することが確認できた。これらのことから、自然言語処理モデルを交通障害の発生予測に活用できること、さらに、その特性について明らかになった。

一方で、本研究においては、出力される予測結果を 10 に設定し、その平均値または中央値で評価していることから、出力設定を変更した場合に結果が異なる可能性がある。地域や時間帯を変更した場合も同様に、予測精度が異なることがある。また、本研究では自然言語処理モデル GPT-2 を活用したが、OpenAI によって GPT-3 や GPT-4 をはじめとする大規模言語モデルが公開されている。パラメータ数の多いこれらのモデルを活用することで、予測精度の向上が期待される。さらに、本研究では ETC2.0 プローブデータ等の記録蓄積データを学習データとして活用していることから、リアルタイムでの交通障害の発生予測は困難である。今後、リアルタイムデータを活用した予測方策の検討を行う際に本研究の枠組みが活用できると期待される。

謝辞

本研究の実施にあたり、三井住友海上福祉財団より研究助成を受けました。また、国土交通省北陸地方整備局富山河川国道事務所、富山市建設部道路河川管理課、日本道路交通情報センターよりデータをご提供頂きました。ここに記して、厚く御礼申し上げます。

参考文献

- 1) 南紘雅, 猪井博登:大雪時における自動車の急ハンドル及び急加減速の時系列変化に関する研究—交通障害の予測方策の確立に向けて—, 第 42 回交通工学研究発表会論文集, 2022
- 2) 小林亮博, 上坂大輔, 武田直人, 南川敦宣, 森本章倫: GPT-2 を用いた位置情報ビッグデータに基づく交通需要推定技術, 第 65 回土木計画学研

究発表会・講演集, 2022

- 3) KyoHoon Jin, JeongA Wi, EunJu Lee, ShinJin Kang, SooKyun Kim, YoungBin Kim: TrafficBERT: Pre-trained model with large-scale data for long-range traffic flow forecasting, Expert Systems with Applications Volume 186, 2021
- 4) 富山県: 降積雪の経年変化, https://www.pref.toyama.jp/1711/kurashi/seikatsu/seikatsueisei/yuki/know/know_keinen/index.html, (最終訪問日: 2022.01.27), 2021
- 5) 気象庁: 富山(富山県)2021年1月(日ごとの値)主な要素, https://www.data.jma.go.jp/obd/stats/etrn/view/daily_sl.php?prec_no=55&block_no=47607&year=2021&month=1&day=&view=, (最終訪問日: 2023.10.30), 2021
- 6) 朝日新聞デジタル: 富山で記録的大雪、帰宅困難者もコンビニ品薄で休業へ, <https://www.asahi.com/articles/ASP196HW7P19PISC00T.html> (最終訪問日: 2022.02.03), 2021.01.09
- 7) 富山県: 令和3年1月7日~11日大雪に係る交通対策の検証結果について, 検証会議第1回配布資料, p.9, <https://www.pref.toyama.jp/document/s/18574/r30121haihusiryoku.pdf>, (最終訪問日: 2021.06.28), 2021
- 8) 田伏雅也, 遠藤徹, 佐々木啓司: スリップは見えるのか?—道路維持管理における ETC2.0 プローブデータ活用検討—, 第 62 回(2018 年度)北海道開発技術研究発表会発表論文集, 2019
- 9) 秀島哲雄: 5.8GHz 帯 DSRC ビーコンによる車両挙動情報の活用~プローブ情報の活用と以上走行の検証~, <https://www.hido.or.jp/itsapq/jsp/auth/H21/kenkyu/H21-04.pdf>, (最終訪問日: 2022.02.04)
- 10) 気象庁: 解析積雪深・解析降雪量、降雪短時間予報, <https://www.jma.go.jp/jma/kishou/know/kurashi/snow.html> (最終訪問日: 2022.01.27)
- 11) 松尾敬世: 雪と雨をわけもの, https://www.metsoc.jp/tenki/pdf/2001/2001_01_0033.pdf, (最終訪問日: 2023.10.31), 2001
- 12) 舟田久之: 道路凍結と降雪の予測, 地学雑誌 Journal of Geography 101(6) 491-505, 1992
- 13) 菊地春海, 岡田朝男, 水野裕彰, 絹田裕一, 中村俊之, 萩原剛, 牧村和彦: 道路交通安全対策事業における急減速挙動データの活用可能性に関する研究, 土木計画学論文集 D3, Vol.68, No.5, 1_1193-1_1204, 2012