

深層学習による高速道路の交通状態ナウキャストに関する研究

岐阜大学 学生会員 ○杜 然
岐阜大学 正会員 倉内 文孝

1. はじめに

円滑な道路交通を実現するためには、交通需要をその容量以下に抑制することが重要であり、そのための一方策として渋滞情報の提供が実施されている。しかしながら、提供されている現在時間情報は、特に遠方に行けば行くほど情報が古くなり、その精度が低下することが知られている。また、現在時間情報に基づく交通情報提供の改善を求める声が多いのも事実であり、提供情報の改善が求められるといえる。本研究では、近年様々な分野での適用が進む深層学習を活用し、高速道路の近未来交通状態予測（ナウキャスト）を試みる。なお、合流部や追い越し車線での車線利用の偏りが渋滞原因の一つとして知られているため、車線レベルで交通状況を知ることは渋滞発生予測のためには重要である。これに対し、現在車線レベルで交通状況を観測できるのは車両検知器のみであるため、本研究では主に車両検知器データを活用することを考えている。車両検知器は容易にデータを収集できるが、定点での観測のため、観測地点周辺以外の交通状況把握が困難である。このような問題に対して車両検知器データとプローブデータを活用することで、時系列予測を空間的に広げる可能性があるのではないかと考え、双方のデータを活用した推定手法を検討することとする。

本研究では、まず説明力が高い車両検知器データ（特徴量）の選択方法を検討する。次に、高速道路の短期交通状態をナウキャストするモデルを構築し、予測結果の有用性を評価する。また、車両検知器データ以外に、プローブデータなどを活用し、時空間的な交通状態予測を行うことを検討する。

2. 使用データ

本研究では中国自動車道に位置する宝塚西トンネル入口付近のボトルネックの渋滞を短期予測する深層学習モデルの構築に取り組む。新名神高速道路が開通するまでは、宝塚西トンネルは関西地区における屈指の渋滞発生箇所として知られていた。渋滞が頻発したため、渋滞予測が適切にできているかどうかを判断しやすいと考え、予測対象地点とした。図-1に示すように、対象地点の下流側は中国道吹田ジャンクション(JCT)まで、上流側は山陽道木見支線神



図-1 データ取得区間と予測対象地点

戸西、山陽道福石パーキングエリア(PA)、中国道佐用、舞鶴若狭道小浜までの車両検知器データ（上り方向）を用いる。車両検知器データは車線別に1分単位で記録されている。データ期間については、2017年12月に新名神高速道路の開通により宝塚西トンネル付近の渋滞がかなり緩和されたため、その影響を避けるために新名神開通前約一年半のものを用了。

3. データ処理とモデル構築

3.1. 欠損値処理

本研究では、連続1時間以内の欠損値は線形補間し、1時間以上の欠損値は「欠損ラベル」を付けることで推定に用いないこととした。学習データに振り分けるときにラベル付きのデータセットを削除する。

3.2. 特徴量選択

深層学習において時系列データの予測に適しているモデルのひとつとして、Long short-term memory (LSTM)¹⁾がある。LSTMは過去の情報を適切なタイミングまで次の層へ伝播する特徴がある。なお保持する過去の情報量と特徴量の数は自由に設定することができる。一方で、各特徴量の欠損部分が異なるため、特徴量の数を増やすほど、学習データの中の「欠損ラベル」付きのデータが増え、欠損値削除により学習に用いることができるデータ数が減る。すなわち、特徴量の数と学習データ数がトレードオフの関係である。本研究では、まず可能な限り学習データ数（渋滞回数）を保有しながら、保持する過去のデータ量を1時間、2時間、3時間、4時間の4種類を用意し、予測精度が最良となる設定を調査する。そこで決定された設

定を用いて、ステップワイズ法による特徴量選択を行う。選択手法はまた検討中であるが、現段階ではベクトル自己回帰モデル (VAR) を参考に、対象地点の追越車線の占有率と他の特徴量の全期間の時間を通じた関係を捉えた自己相関行列で、相関が高い特徴量を選択している。次には渋滞時と非渋滞時を分けて検討を進める。選択された特徴量の位置情報と種類を把握するため、図-2 に示すように、実際の道路距離を考慮した可視化を行った。

3.3. モデル構築

本研究では、宝塚西トンネルの上流に位置する 23.12kp を渋滞把握のための地点とし、占有率を用い、占有率 25%以上かつ連続 15 分以上の時間帯を「渋滞：1」、他「非渋滞：0」と対象地点の交通状態を予め「渋滞ダミー」で定義する。そして先述で選択された車線別占有率、交通量、大型車交通量、速度、車線利用率などの特徴量に加え、時間帯ダミーと曜日ダミーを入力データとし、予測対象地点の渋滞ダミーを出力データとする。全期間の前8割を訓練データと検証データとし、残り 2 割をテストデータとする。

4. 結果分析

モデルの学習曲線の 1 つを図-3 に示す。モデルの誤差推移と精度推移から見ると、学習が収束し、過学習はしていないことがわかった。精度は 97%と非常に高いが、渋滞時間に対して非渋滞時間が圧倒的に長いためともいえる。そのため、渋滞発生時点と解消時点の予測精度に着目した。予測値と観測値の比較の一部を図-4 に示す。予測渋滞開始時刻と実の渋滞開始時刻の乖離を集計するため、渋滞判定の閾値を設定する必要がある。また、閾値はモデルの利用目的によって変えることができる。確実に渋滞中を検知したい時を高め、わずかでも渋滞の兆候を検知したい時を低めに設定すればよい。

実際発生した渋滞に対して、予測の発生時点を実際の発生時点の前後 10 分以内であれば、渋滞発生予測ができたと判断される。渋滞発生あるいは解消時点の評価指標は混同行列を参考に予測渋滞発生時に実際に渋滞が発生した割合 (精度) と実際に渋滞発生した時に予測できている割合 (再現率) の調和平均 (F 値) を用いる。異なる閾値により予測結果がどのように変わるかを調査するため、閾値を 0.1 から 0.9 まで 9 つの値を用意し、予測精度が最良となる閾値を調査する。一例として、閾値を 0.5 とする場合の混同行列と乖離の誤差分布を表-1 と図-5 に示す。渋滞発生より渋滞解消の方は精度が高いが、全体的に渋滞発生予測の時間誤差は小さくないことが読み取れる。

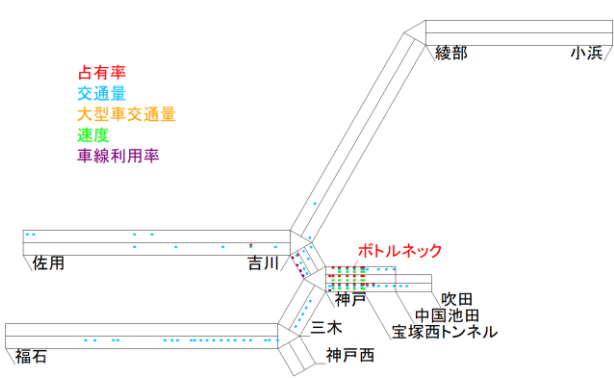


図-2 特徴量の空間的分布

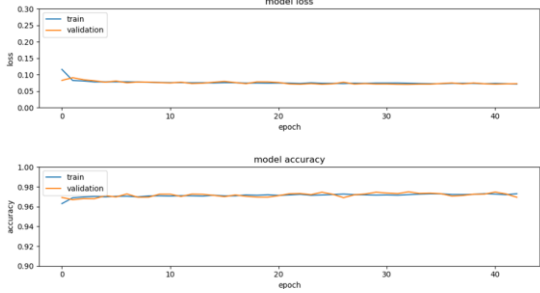


図-3 学習曲線

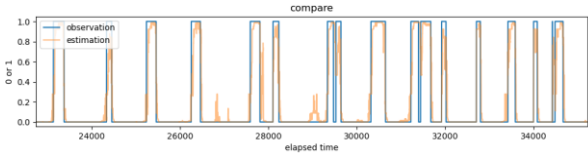
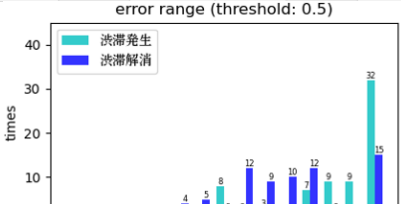


図-4 予測値と観測値 (一部)

表-1 混同行列 (閾値 0.5)

発生混同行列	予測発生	予測非発生	F値	0.2	外れ回数
実際発生	14	58	精度	0.21	-5
実際非発生	53		再現率	0.19	
解消混同行列	予測解消	予測非解消	F値	0.46	
実際解消	32	40	精度	0.48	
実際非解消	35		再現率	0.44	閾値: 0.5



5. おわりに

本研究では深層学習ベースの宝塚西トンネル付近の短期渋滞予測モデルを構築した。現状では精度が望ましくないため、さらなる改善が必要である。今後渋滞状況を分割した特徴量選択と対象区間の事象データや ETC2.0 データの活用を検討を進める。

謝辞：本研究は西日本高速道路株式会社に協力いただいたものである。記してここに謝意を示す。

参考文献

1) Hochreiter, K. and Schmidhuber, J. : Long Short-Term Memory, Neural Computation, 9(8)pp1735-1780, 1997.