

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

芦田 優太^{†1} 西岡 到^{†1}

概要：交通渋滞の予測情報は利用者、道路事業者双方にとって有用であり、交通状況予測の研究はこれまで盛んに行われてきた。しかし、交通流は天候や路面の状態などの環境状態に加え、利用者の運転特性などによって複雑に変化するため、単一のモデルで交通状態を予測することは困難である。この課題に対するアプローチとして、交通状態を変化させる条件に応じたデータの分割や、道路区間毎のモデルパラメータ調整などが考えられるが、これらは高度な事前分析が必要であり、道路全体の渋滞予測には高いコストが必要になる。この課題に対して、本稿では路側データから学習した道路交通流モデルを用いて高速道路における交通状況予測を行った結果を報告する。またその結果から、交通流の特性が大きく変化する環境条件と、各条件下の交通流モデルをデータから機械的に学習することにより、従来の手法では予測困難であった広範囲にわたる交通渋滞が、時間や気候の変化に関わらず高精度に予測出来る事を示す。

キーワード：交通流、渋滞予測、データ分析、機械学習

Highway Traffic State Prediction by Using Traffic Flow Model Derived from Road-Side Sensor Data

YUTA ASHIDA^{†1} ITARU NISHIOKA^{†1}

Abstract: Traffic state prediction is useful for both road operators and users. If future state information is provided across the road, road operators can improve road management operation drastically and provide more efficient transport service. Although many techniques are proposed to predict traffic state in these few decades, traffic state prediction with high accuracy is still difficult problem because dynamics of traffic flow might change in accordance with weather condition, road surface state and driver's characteristic. For this problem, we studied model developing method based on Heterogeneous Mixture Learning (HML) that is predictive analysis technology developed by NEC. In this paper, we report traffic state prediction using traffic model learned by HML. HML can derive not only state prediction model but environmental factor which changes traffic flow dynamics. Furthermore, we show our method can predict traffic congestion anywhere on the road based on prediction results and traffic models derived from actual data.

Keywords: Traffic flow, Congestion prediction, Data analysis, Machine learning

1. はじめに

交通状況の予測情報は、移動時間の見積りや出発時間の決定、事故など異常発生への備えなど、利用者、道路事業者双方にとって有用な情報である。特に、広域にわたる道路区間で渋滞予測が可能になれば、交通状態の時間発展を俯瞰的に見える化でき、利用者の利便性の向上や道路管制業務の効率化を実現することができると考えられる。交通状況の予測手法に関しては、これまで多数の研究が行われてきており、そのアプローチは交通流の挙動を記述した物理モデルを用いて交通状態の時間変化を算出する物理モデルアプローチと、物理モデルを用いず交通状況変化の規則性を実際のデータから学習して予測モデルを構築する機械学習アプローチに大別できる。

物理モデルアプローチでは、交通流を波の伝播としてモデル化し、その伝播を算出する kinematic wave モデル[1]や、道路を細かいセルに分割し、各車両のセル移動によって交

通流をモデル化する Cell Automaton モデル[2]など複数のモデルが提案されており、実際のオペレーションにも活用されている。しかしながら、これら物理モデリングによるアプローチでは、道路地点ごとに精密なパラメータの調整が必要である上に、道路構造が複雑になるとさらに多数のパラメータを必要とする点がデメリットとして存在する。

一方、機械学習によるアプローチは、モデルパラメータの大部分をデータから学習できることから、十分な学習用データサンプルが存在すれば物理モデルと比較してモデル構築が容易であり、近年盛んに研究されている[3]。しかしながら、交通流は天候や曜日、時間帯等に応じて異なった規則性を持つため、そのモデルを高精度に学習することは難しい。従来このようなデータの分析に際しては規則性毎のデータ分割を事前に行い、その上で適切な予測モデルを構築することで対応されてきた。しかしこのような作業は専門知識と多くの工数が必要であり、多数の地点で高精度な予測モデルを構築することは困難である。また、新たな道路の建設等に応じた修正など予測モデルの維持管理コス

^{†1} NEC 情報・ナレッジ研究所
NEC Knowledge Discovery Research laboratories

とも問題となる。

この課題に対して、我々は交通流の規則性が変化する要因を自動で抽出し、各条件下における予測モデルをデータから学習する交通流モデル生成方法を用いて高速道路全体の交通状況予測を行った。本手法では、交通流の規則性が変化する条件の抽出と、条件毎の予測モデル学習の双方を自動的に行うことが出来るため、任意の道路地点における予測モデルを低コストで構築できる。本稿では上記の方法によって気候や時間の変化に対しても高精度を維持する予測モデルが低コストで構築できるとともに、広域道路全点において高精度な交通状況予測が可能であることを示す。

2. 関連研究

上述のように、交通流の変化は条件に応じて異なる規則性に従う。このような複雑な規則性を持つデータに対する予測モデル学習のアプローチとして非線形モデルによる予測方式が考えられるが、そのような複雑なモデルを選択した場合、高い予測精度が得られたとしても学習の結果得られたモデルを人が解釈することが困難である。特に、交通工学の領域では、データから学習されたモデルによって、ある時間の交通状況は過去のどの地点の交通状態に影響を受けるか、という交通分析としての役割も重視されることから、予測精度に加え、モデルの解釈性も重要な要因となる[3]。また、実用面でも複雑なモデルは利用者が理解し辛く、予測精度が悪化した際に、入力データのエラーなどシステム上の問題であるのか、学習済みのモデルと現実との乖離が問題であるのか切り分けることが難しい。そのため、予測に用いるモデルは人が解釈可能なシンプルなモデル、例えば線形回帰モデルや重回帰モデルなどが望ましい。

このような要求を踏まえ、MinらはVARMA(ベクトル自己回帰移動平均)モデルに状況毎に異なる空間相関行列を組み込むモデルを提案している[4]。空間相関行列はある地点の交通状態と、他の地点の過去の交通状態が相関を持つか否かを示す行列であり、VARMAモデルに組み込むことで目的変数と時空間上で相関を持つ説明変数のみを使用したモデルを構築することが出来る。Minらはこの空間相関行列を渋滞時や天候毎に用意し、状況毎の予測モデルを構築することで高精度な多変数の交通状態予測を可能としている。しかしながら空間相関行列の構築は事前に別途行わなければならないため、気候条件や曜日等交通状況が変化する要因の組み合わせに応じた事前分析が必要となり、多様な環境条件で実用可能な予測モデルを学習するには高いコストが必要となる。

3. 路側データからの交通流予測モデル生成

3.1 異種混合学習技術

交通流の規則性は天候や道路構造など複数の要因が絡み合って変化する。つまり、交通状況データは複数の規則に

従うデータが混合された形で収集・分析されることになる。我々はこのような異なる種類の規則性に従うデータ(異種混合データ)を分析するため、異種混合学習技術を開発し実用化している[5]。異種混合学習技術は、異種混合データを、人にとって解釈しやすいモデルによって高精度に予測するため、入力データを決定木形式のルールによって場合分けし、各場合で異なる説明変数を組合せた線形モデルを学習する(図1)。このモデルを異種混合予測モデルと呼んでおり、モデル選択を行う決定木から規則性変化の要因となっている変数を、また選択された線形予測モデルからその状況において強い相関を持っている変数をそれぞれ読み取ることが出来る。異種混合学習技術はこの異種混合予測モデルをデータから自動的に学習する技術であり、データの分割と分割されたデータの予測モデル学習を機械的に行う[6][7]。

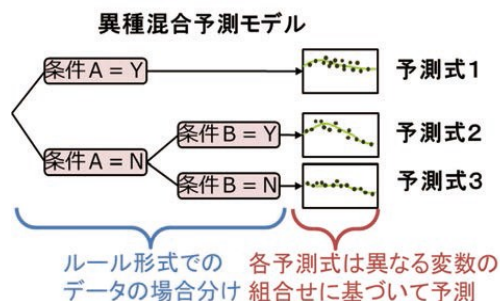


図1 異種混合予測モデル

交通流ではデータの規則性は路面勾配や道路線形、料金所や入口・出口からの距離など、地点ごとの特性によって異なることが考えられる。このように複数の規則性を持つとともに、その規則性変化も異なるデータのモデルを構築する場合は物理モデルに基づく方法や従来の機械学習技術では高いコストが必要となる。異種混合学習技術を用いることにより、道路全地点における予測モデルを機械的に学習することが出来、複雑な事前分析を行うことなく道路全体に渡って交通状況の予測が可能となる。

3.2 異種混合学習技術による交通流モデル生成

異種混合学習技術によって、規則性毎の予測モデルが学習できるが、交通状況予測においては規則性変化の要因となりうるデータ属性と、予測式の説明変数となりうるデータ属性はそれぞれ異なる性質を持つ。交通管制や、道路利用者への情報提供という観点では、予測された情報を基に交通流の制御や、出発時間、経路の選定という意味決定を行う時間的猶予が必要であるため、数時間後の交通状況予測が効果的である。二時間後の交通状況予測の場合、後方の交通流が対象地点に到着することによる影響や、前方の渋滞状況が対象地点まで延伸するなど、他の地点との時空間上の相関が強い影響を及ぼすと考えられる。そのため、ある予測式における説明変数は各地点の交通状況が採用さ

れ、その回帰係数が該当する条件下での相関の強さを説明するモデルとなっているのが適切であると考えられる。また、交通流の規則性変化は天候による路面状態の一時的変化や、曜日、時間帯、休日などに応じて変化する移動目的や大型車両割合などトラフィック特性の変化が重要な要因である。そのため、これらの交通状況ではない環境情報はモデルの決定木における説明変数として使用されることが望ましい。

以上の事から、モデル決定木の説明変数の候補には曜日、時間帯、休日などの環境情報を、予測式の説明変数の候補には各地点の交通状況をそれぞれ用い、異種混合学習技術によって道路各地点の予測モデルを学習する。各地点の予測モデルは他の地点の交通状況が予測対象地点の将来交通状況に与える影響を表現する、交通流の時空間相関モデルと捉えることが出来、道路上の全地点における予測モデルを構築することで路側データから道路全体の交通流モデルを得ることが出来る。

4. 生成した交通流モデルによる予測結果

4.1 分析内容

本章では前記の方法で路側データから構築した交通流モデルを用い、実際の高速道路における交通状況を予測した結果、及び構築されたモデルについて述べる。予測対象は各時刻から2時間後の全車線平均速度とした。予測及び学習に用いたデータは、アジア圏内に位置する高速道路約200kmの道路区間で収集された3か月間の路側データであり、支線や接続している他路線のデータは含まない。このデータセットの内2か月分のデータを用いて予測モデルを学習し、残る1か月分を予測した。比較対象として、時間帯などの環境情報と交通情報を含む重回帰分析、交通情報を含まない重回帰分析の2手法で同一の実験を行った。各分析方法における説明変数は表1のとおりである。

表1 各分析手法の説明変数

説明変数	異種混合学習 (2時間後)	重回帰分析 (2時間後)	重回帰分析
曜日	✓	✓	✓
時間帯	✓	✓	✓
雨量	✓	✓	✓
視界(良, 不良)	✓	✓	✓
地点交通量	✓	✓	
走行速度	✓	✓	

4.2 全地点の予測結果

まず、予測モデルの精度について考察するため、対象区間全地点の予測結果を図2に示す。図2はある一日の予測結果を、縦軸を時刻、横軸を終点からの距離、走行速度を

色としたヒートマップで可視化した結果である。可視化にあたっては渋滞状態が予測できているか否かを重視するため、時速40km以下の値を強調して示している。(c)の交通状態情報を含まない重回帰モデルによる予測では渋滞形状、すなわち走行速度が急激に減少している領域を予測できていないが、(b)の異種混合学習による速度予測では特にピークの渋滞長および解消部分が高精度に予測できている。(d)の異種混合学習と同じ説明変数を用いた重回帰分析では(c)と比較して渋滞形状が判別できる程度に予測精度が向上している。しかしながら(b)と比較すると渋滞/非渋滞領域の境界が不明瞭であり、特にピーク時の渋滞長や、渋滞解消に関しては明確に判別することが難しい。

これをより明確に示すため、実績値と各手法による予測値との差分を同じくヒートマップとして可視化した図3を示す。図3上の青色の領域は実績値-予測値が正の領域、すなわち渋滞していないにもかかわらず渋滞と判定してしまったことを示している。また、赤色の領域は実績値-予測値が負の領域、すなわち渋滞しているにもかかわらず渋滞でないと判定してしまったことを示している。図3を見ると異種混合学習による予測では解消に転じてからの渋滞推移で誤差が少ないことがわかる。特に(g)では渋滞が解消に転じている領域で正・負両方の誤差が生じているため、渋滞解消部分の誤差が大きいことがわかる。

図4は各地点の平均二乗平方根誤差(RMSE)を示している。青実線が異種混合学習におけるRMSE、黒破線が同じ説明変数を用いた重回帰分析、黒実線が交通情報を含まない重回帰分析におけるRMSEを示す。図4を見ると終点から50kmの範囲ではそれぞれの誤差にほとんど差がない。これは、実験で対象としている道路区間の前方に存在する路線の情報がデータセットに存在しないため各地点から見て前方の交通状態情報が利用できないためであると考えられる。そのため、今回の対象区間より広域の道路情報が利用できれば予測精度が向上する可能性がある。一方、終点より50km以上先、特に図2において渋滞が発生している区間である50kmから120kmの区間においては交通状態情報を用いることにより、異種混合学習、重回帰分析双方ともに約5km/h低減されており、進行方向先の交通状態が利用できることにより予測精度が向上していることがわかる。

また、予測速度の精度だけでなく、渋滞状態の予測精度での比較を行うため、各時刻の予測速度が渋滞(40km/h以下)/非渋滞(40km/h超)のどちらに含まれるかを判定し、そのF値を算出した。図5に各手法による渋滞状態予測のF値を示す。渋滞が実際に発生した地点でのF値は最大1であるが、渋滞が未発生地点については0の値を採用している。F値による比較では異種混合学習が最も高い評価値を示している。RMSEでは顕著な差が見られなかった重回帰分析とは最大で0.2ポイント差異があり、速度値

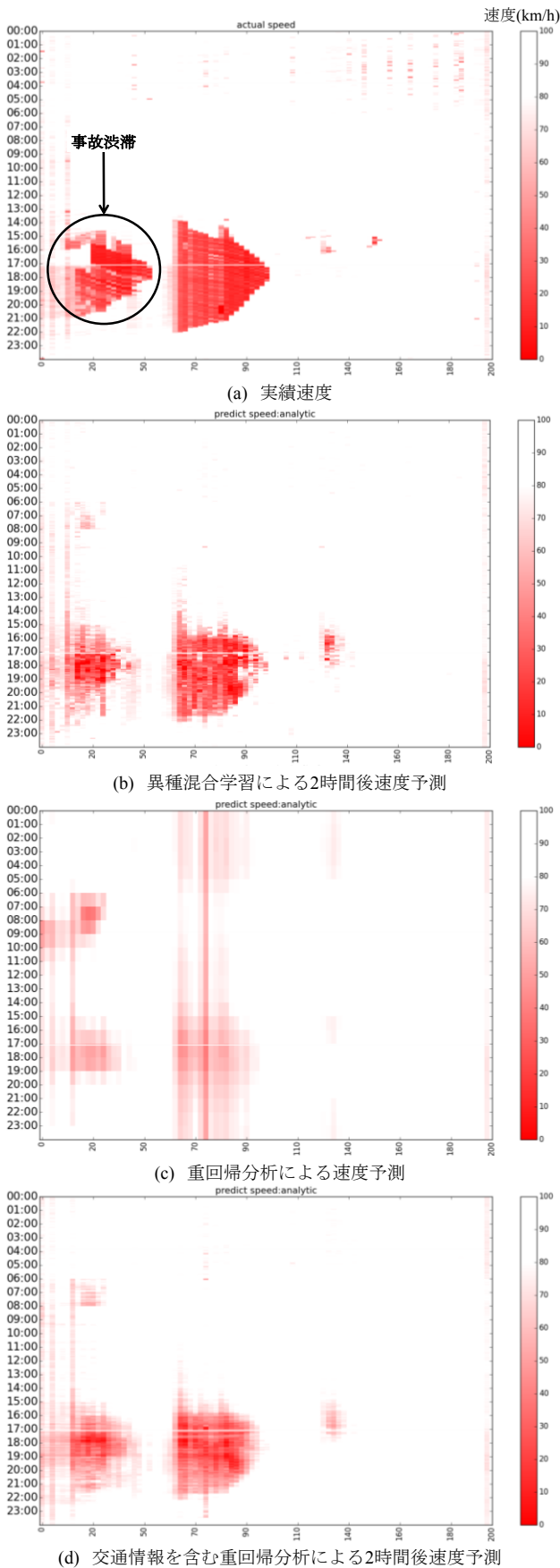


図 2 実際の各地点走行速度と走行速度予測結果

そのものの予測精度では大きな差が出ないが、渋滞/非渋

滞の状態予測では異種混合学習が高い精度となっている。

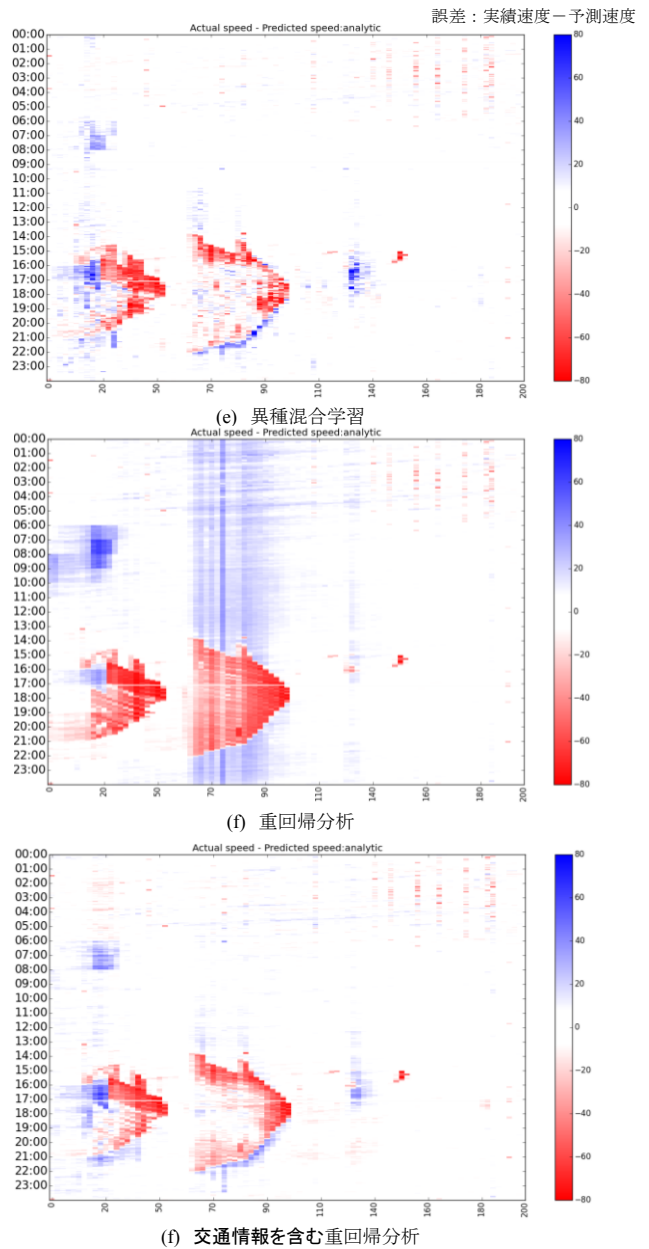


図 3 走行速度予測値と実績値の誤差

4.3 渋滞発生地点の速度予測結果

次に、走行速度の時間変化を詳細に考察するために、同日における渋滞発生区間内の一地点における速度予測結果を図 6 に示す。図 6 の赤線が実績値、青実線が異種混合学習による予測値、黒破線が交通情報を含む重回帰分析、黒実線が交通情報を含まない重回帰分析による予測値である。交通情報を含まない重回帰分析では実際に渋滞が発生している 14 時 30 分から 22 時の間でやや低い速度が予測されているが、70km/h 以下にはなっておらず、渋滞が発生しているとみることが困難である。また、交通情報を用いることによって重回帰分析でも 40km/h 以下の渋滞状態を予測できているが、その発生開始は 16 時と 1 時間 30 分の遅

延が存在している。一方、異種混合学習による予測値では渋滞発生開始時刻はおよそ 15 時と予測されており、30 分程度遅れがあるものの、それ以降継続して 40km/h 以下の速度が予測されており、かつ渋滞終了に関しては実績とほぼ同じ時刻が予測できているため、渋滞／非渋滞の遷移がより高精度に予測できている。

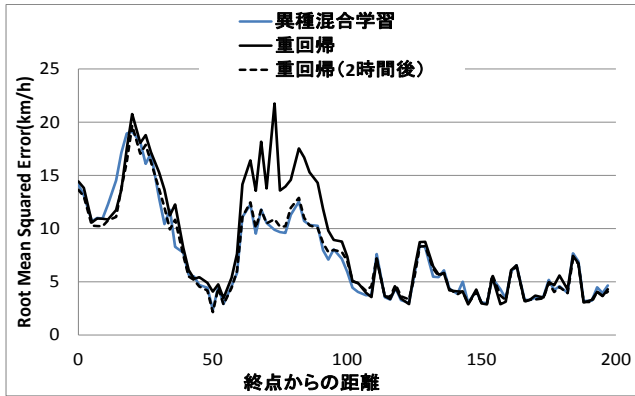


図 4 各分析方式の RMSE

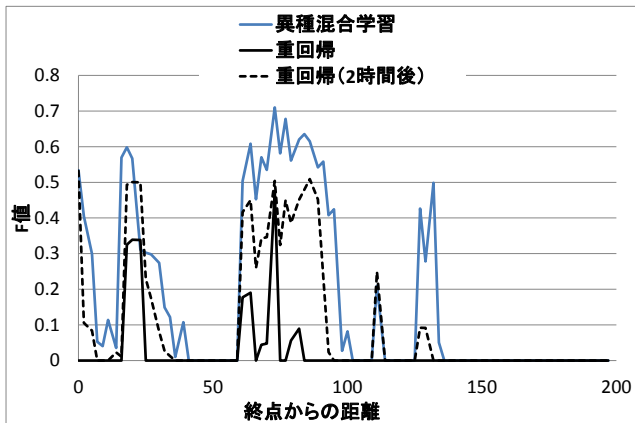


図 5 各分析方式による渋滞予測の F 値

図 6 上の黒点は、異種混合学習において各時刻の速度予測値を算出するのに用いられた予測式番号を示している。

4.4 予測モデルの考察

図 7 にこの地点で学習された予測式選択木を示す。この決定木では主に夕方の時間帯、および週末であるかどうかでモデルが決定されており、時間帯と曜日に応じて交通状態変化の規則性を切り替えて予測していることがわかる。図 8、図 9 には上記のモデルのうち、同日中で最も高頻度で使用された予測式 7 と、渋滞中から渋滞解消への遷移時に使用された予測式 2 において高い係数をもつ説明変数をそれぞれ示している。予測式 7 では予測対象地点から見て前方後方両方の交通状態が重視されており、後方の交通量が 2 時間後に予測対象地点の走行速度に影響を与える一方で、前方の交通状態が渋滞状態である限り、対象地点の将来の走行速度は渋滞状態に準じるものになるということを示している。予測式 2 では後方の交通状態が大部分を占めており、現時刻において後方の交通状態が減少している

場合、前方の渋滞がいずれ解消に遷移していく現象をとらえていると考えられる。これらの予測式の特徴は交通流の経験的な知見と合致しており、状況によって変化する交通流変化の規則性を複数の予測式を切り替えることで再現できている。

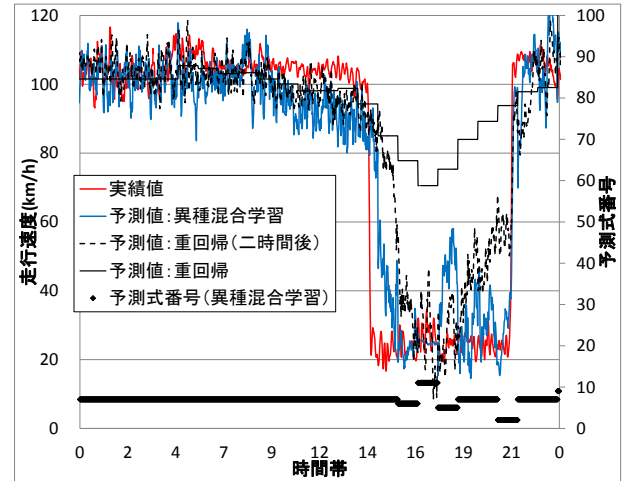


図 6 渋滞発生地点の速度予測結果

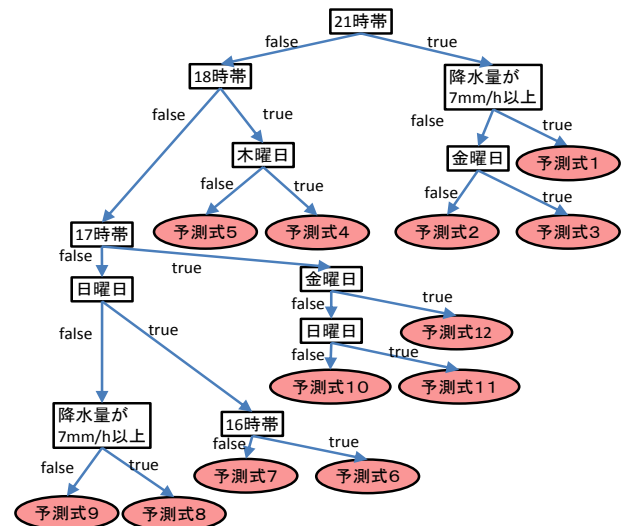


図 7 異種混合学習で得られたモデル決定木

また、図 10 は雨天（降水量が 7mm/h 以上）時と晴天時にそれぞれ使用される予測式における、各道路区間の交通状況の相関度を示している。縦軸は各 20km 区間の交通状況の相関係数の絶対値を取ったものである。晴天時の予測式と雨天時の予測式で最も異なる点は、晴天時は雨天時と比べて広範囲の交通状況を考慮している点である。雨天時に考慮されているのは予測対象地点から約 100km 上流である 140km から 160km 区間まで、下流側はほぼ考慮されていない。一方、晴天時には予測対象地点から約 140km 上流である対象区間始点から約 60km 下流である対象区間終点までの交通状況を考慮している。この理由として、降雨による平均速度の低下が考えられる。実際に、実績データ

を

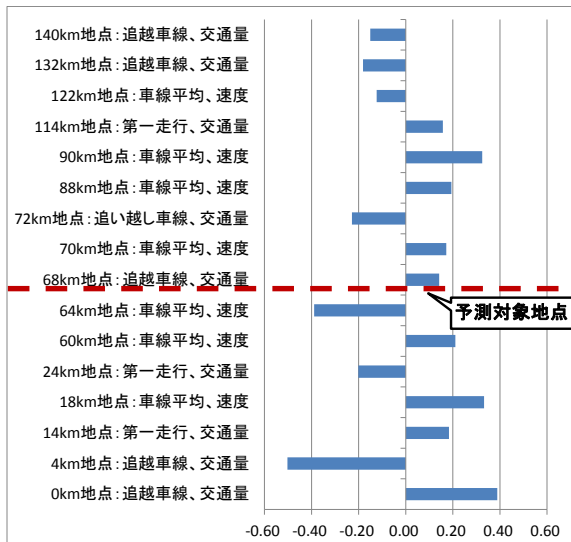


図 8 予測式 7 で使用される主要な説明変数

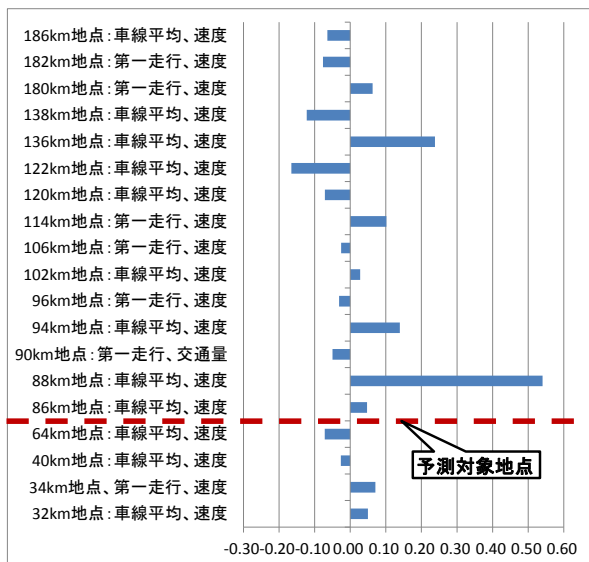


図 9 予測式 2 で使用される主要な説明変数

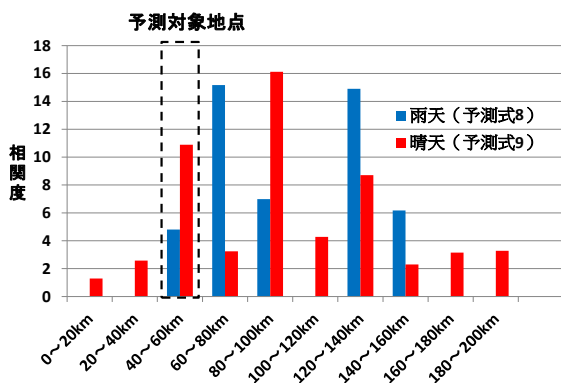


図 10 雨天時と晴天時における各区間交通状況の相関度

観察すると、降雨時は平均速度が 10~15km/h 低下しており、ある区間の交通状況に関わりのある地理範囲が天候によ

て変化していることが分かる。このような規則性の変化を図 7 のモデル決定木は表現しており、モデルで決定されるそれぞれの予測式は状況に応じた交通状況の時空間相関を表現できていると言える。

5. おわりに

本稿では路側データから学習した交通条件毎の予測モデルによる交通状況予測について報告した。交通流の規則性変化の要因と、それぞれの規則性を表現する予測式を過去の交通データから機械的に学習することにより、複雑な事前分析を行うことなく高精度な予測モデルを構築することが可能である。実際の交通データに本手法を適用することで、200km という広範囲な道路全点において高精度な交通流予測モデルを構築することが可能である事が実験結果から示された。本手法は路側データから機械的に交通流モデルを学習することが出来るため、データ収集が可能であれば任意の地点、区間において時間や気候の変化に関わらず高精度なモデルを構築することが出来、従来の手法では困難であった広範囲にわたる交通状況予測が可能である。

今後、より広範な環境への適用を考え、センサが十分でない高速道路や、一般道路への適用に向けて、車両プローブデータの分析技術との組み合わせによる高速・一般道路合わせた交通渋滞予測へ発展させていく予定である。

参考文献

- 1) Lighthill, Michael J., and Gerald Beresford Whitham. "On kinematic waves. II. A theory of traffic flow on long crowded roads." Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences. Vol. 229. No. 1178. The Royal Society, 1955.
- 2) Nagel, Kai, and Michael Schreckenberg. "A cellular automaton model for freeway traffic." Journal de physique I 2.12 (1992): 2221-2229.
- 3) Vlahogianni, Eleni I., Matthew G. Karlaftis, and John C. Golias. "Short-term traffic forecasting: Where we are and where we're going." Transportation Research Part C: Emerging Technologies 43 (2014): 3-19.
- 4) Min, Wanli, and Laura Wynter. "Real-time road traffic prediction with spatio-temporal correlations." Transportation Research Part C: Emerging Technologies 19.4 (2011): 606-616.
- 5) 藤巻遼平, and 森永聡. "ビッグデータ時代の最先端データマイニング." NEC 技報 65.2 (2012): 81.
- 6) Fujimaki, Ryohei, and Satoshi Morinaga. "Factorized asymptotic bayesian inference for mixture modeling." International Conference on Artificial Intelligence and Statistics. 2012.
- 7) Eto, Riki, et al. "Fully-Automatic Bayesian Piecewise Sparse Linear Models." Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics. 2014.