

1. はじめに

近年、我が国では豪雪を起因とする交通障害が頻発しており、深刻な社会問題となっている。代表的な事例としては、極端な降雪に伴う路面状況の悪化により交通事故や車両のスタックが発生し、それを契機として後続車両が停止を余儀なくされ、大規模な車両滞留へと発展するケースが挙げられる。近年では、集中的な豪雪の頻発化に対応するため各種の対策が講じられている¹⁾ものの、それにもかかわらず降雪に起因する交通障害の発生は依然として確認されている。

上記を踏まえ本研究では、豪雪期間での交通事故や車両スタックが多数報告される関越自動車道を対象に、交通事故や車両スタックが発生しない「安全」または「交通事故」、「車両スタック」のいずれが発生しやすいかを推定する。しかし、交通事故や車両スタックの発生は稀な事象であることから、サンプル数に偏りのある不均衡データとなり、事象の推定モデル構築において実績を目的変数とするにはデータ不足であった。このような不均衡データに対応するために、本研究では交通事故及び車両スタック発生事象に対する合成データ生成手法を採用しデータ不均衡を補正した。さらに推定精度向上のため、路面状況を表す指標を、観測される車両速度や気象状況から推定し活用した。推定モデルには、離散選択モデル (Multinomial Logit, Nested Logit, Mixed Logit) を使用し、コネクテッドカー²⁾ (以下、CAV と略す) から得られるタイヤ空転制御機能や横滑り制御機能等の運転サポート履歴や車両の走行状況等の情報、交通・気象情報、道路線形、路面状況といった候補変数群から各事象の発生確率の予測モデルを構築し、各事象と各変数との関係性を明らかにするとともに、発生リスク指標の確立と実用性について検討する。

2. 関連する既往研究及び本研究の目的・新規性

前章で述べた研究背景を踏まえて、本研究の目的、目的に関連する既往研究、既往研究を踏まえた本研究の新規性について整理する。

(1) 本研究の目的

本研究の目的は、CAV から取得できるプローブデータを活用して、冬期高速道路における交通事故および車両スタックの発生メカニズムを構造的に明らかにするとともに、低頻度で発生するこれらの事象に対する発生リスクを定量的に評価し、大規模な車両滞留を未然に防止するための道路管理および交通運用方策に資する知見を得るとともに、各事象の発生リスク指標の確立と実用性に

ついて検討することである。

(2) 交通事故及び車両スタックの予測に関する研究

冬期の事故予測に関して、Kim ら³⁾は韓国的高速道路における天候起因交通事故リスクの予測と提供を目的として、事故データ・気象データ・道路線形・地形情報を統合したデータセットを使用し、SMOTE でデータ不均衡を補正したうえでロジスティック回帰、ニューラルネットワーク、XGBoost、ランダムフォレストにより分類と重要度分析をした。各モデルの評価指標として、Accuracy・Kappa・AUC に加え、混同行列から得る F-measure・G-mean で性能を評価した結果、ランダムフォレストが最良であるとした。また交通事故と各説明変数との関係より、気象要因では湿度・気温・風速・気圧が重要であり、道路線形では曲線半径が小さい急曲線区間や方向変化角が大きい屈曲区間ほどリスクが高いことを明らかにした。

冬期のスタック予測に関して、針原ら⁴⁾は冬期高速道路における路面状況を考慮した将来の区間平均車速の予測とスタック発生リスク低減方策の検討を目的とした研究を行った。気象・除雪通過履歴と CCTV 画像に基づく決定木で路面状況を推定し、その低下影響を仮想変数として定量化して交通流率と併せて、時系列的な変化と不確実性を考慮可能な線形ガウス状態空間モデルに組み込んだ。モデルで推定した区間平均車速の低下自体をスタック発生リスクの指標として扱い、実際の車両スタックの発生期間で適用した。その結果、関越自動車道での豪雪時、スタック発生の約 12 時間前から速度低下の進行を捉え、黒シャーベット相当で 30km/h 付近まで低下した局面を高リスクと位置づけるとともに、30 分降雪量が約 1.25cm を超える場合には予防的通行止め等の対策判断に有用な予測指標となることを明らかにした。

(3) Multinomial Logit による事故予測に関する研究

Multinomial Logit を使用した交通事故予測に関して、Chen・Fan ら⁵⁾は、ノースカロライナ州における歩行者と車両衝突の傷害重大度の決定要因の解明を目的として、高速道路の 2005~2012 年データを用い、重大度 5 水準 (物損のみ/可能傷害/明白な傷害/失能傷害/致死) を被説明変数とし、限界効果を算出した。Multinomial Logit を選定した理由は、交通安全分野で広く用いられる離散選択モデルであり、非順序な複数カテゴリに対して各水準の発生確率を直接推定しやすく結果解釈 (限界効果) の平易さに優れるためである。結果として、運転者の身体状態不良、車種、歩行者年齢、週末ダミー、光環境、平面線形、路面状態、道路種別、および速度規制の有無が致死・失能傷害の確率を有意に高める要因とし

て明らかとなった。

(4) Nested Logit による事故予測に関する研究

Nested Logit を使用した交通事故予測に関して、Tavakoli Kashani ら⁹⁾は、イランの高速道路における事故タイプ別の発生要因の同定を目的として、交通警察の事故記録からコロナ禍前の3年間に収集した 10,919 件のデータを用い、単独/多重を上位に、単独事故（固定物衝突・道路逸脱・横転）、多重（対車両・多車両衝突）と定義したネスト構造で分析した。結果の解釈には、各係数の符号と z 統計量・p 値に基づく有意性判定に加えて、説明変数を 1 単位変化させたときに当該タイプの発生確率がどの程度増減するかを示す限界効果を用い、ネスト階層ごとに提示した。具体的には、上位では単独側で速度違反・前方不注意・操作不能が、多重側で急な進路変更・車間距離不保持が発生確率を有意に高め、下位では固定物衝突に対する天候悪化が+4.74%、多重・対車両に対する疲労・眠気が+14.15%、路肩幅の拡大は対車両衝突を-22.6%とする限界効果が示された。モデル全体は尤度比検定により有意であり、推定結果の妥当性が確認された。

(5) Mixed Logit による事故予測に関する研究

Mixed Logit を使用した交通事故予測に関して、Milton ら⁷⁾は、高速道路区間で発生した交通事故データを用い、事故の重症度を被説明変数としてモデルを構築している。説明変数として、交通量、大型車混入率、道路幾何構造、気象・路面条件等を用い、事故重症度に対する影響を分析した。その結果、交通量や大型車混入率、路面状態などの影響は区間ごとに大きく異なり、固定係数を仮定する Multinomial Logit では捉えきれない未観測異質性が存在することを示した。特に、Mixed Logit により、一部の区間では悪天候時に重症事故が顕著に増加する一方、他区間では影響が限定的であることが明らかとなり、事故リスクの空間的・構造的な異質性を考慮する重要性を指摘している。また、Das ら⁸⁾は、有料高速道路で発生した事故データを対象に、Mixed Logit を用いて事故重症度の発生要因を分析している。説明変数には、気象条件、路面状態、交通条件、運転者属性、車両特性などが含まれており、事故重症度を複数カテゴリとして定式化した。分析の結果、降雨や視程低下といった気象条件、夜間時間帯、高速走行条件などが重症事故の発生確率を高めることが示される一方、これらの影響は事故や区間ごとにばらつきが大きいことが確認された。Mixed Logit を用いることで、従来の固定係数モデルでは表現できない未観測要因の影響や感応度の異質性を考慮可能であり、有料道路における事故リスク評価の高度化に有効であることを示している。

(6) 本研究の目的及び新規性

既往研究を踏まえて本研究の目的及び新規性を以下に

示す。交通事故および車両スタックの発生要因や予測に関しては、気象条件、路面状態、交通量、車種構成、道路線形などを対象とした多くの研究が行われてきた。しかし、これらの研究の多くは交通事故または車両スタックのいずれか一方を対象としており、豪雪時に同一の道路環境下で発生しうる交通事故と車両スタックを統合的に扱った研究は限られている。複数の交通状態を同時に扱う手法として、Mixed Logit を用いた事故予測研究が行われてきた。Mixed Logit は複数事象の発生確率を同一の枠組みで推定可能である一方、選択肢間の相関を考慮できず、「安全」から「危険」への移行や、その後の事故類型への分岐といった階層的な発生構造を明示的に表現できないという課題がある。これに対し、Nested Logit は類似事象をグルーピングすることで階層構造を導入できるが、係数を固定とする場合には、区間差や極端気象時の影響のばらつきといった未観測異質性を十分に捉えられないという限界が指摘されている。近年では、Mixed Logit を用いて事故発生特性の異質性を考慮する研究が進展しているものの、事故と車両スタックを同一の枠組みで構造的に同時モデル化した研究はほとんど見られない。以上を踏まえ、本研究では冬期高速道路における交通状態を「安全」「事故」「車両スタック」として定式化し、階層構造および未観測異質性を考慮可能な離散選択モデルを適用することで、豪雪時の交通障害発生メカニズムを構造的に解明する点に新規性を有する。

3. 使用データの概要

(1) 分析の対象範囲

本研究では、関越自動車道上り線において 2020 年 12 月の豪雪により大規模な車両滞留が発生した区間（長岡 IC (244.5KP 地点)～土樽 PA (157.5KP 地点)）を対象として分析した。分析対象範囲は、関越自動車道上り線において 2020 年 12 月の豪雪により大規模な車両滞留が発生した区間である。区間距離は 87km であり、日本有数の豪雪地帯である南魚沼や湯沢を通過する。

(2) CAV から得るデータ

本研究では市販の CAV から得るデータを集計・活用した。CAV とは、車両自体を ICT 端末として機能し、車載の専用通信装置を介して、「走行状況データ」および「車両状態データ」をリアルタイムに取得できる。これにより、車両は単なる移動手段としての役割にとどまらず、交通流・運転挙動・環境条件等に関する高頻度かつ空間的に広がりをもつ情報の収集手段として機能するようになる。本研究においても、これらの走行データを活用することで、交通事故および車両スタックといった事象の発生傾向を把握し、その要因を分析するための基盤データとして用いる。CAV から得るデータ概要を表 1 に示す。また、モデルに用いる説明変数とするために集

計・整形したデータ概要を表2に示す。なお、アクセルペダル量に関しては、欠測を多数確認したため、本研究では除外した。

表1 CAVから取得したデータ概要

項目名	単位	備考
GPS時間	yyyy-mm-dd hh	GPSからの取得時間
GPS緯度	度	GPSから得た緯度
GPS経度		GPSから得た経度
TRC作動履歴	-	データから判断し、作動した場合に1というフラグを立てた
ABS作動履歴		
VSC作動履歴		
外気温	°C	GPS緯度・経度で観測された各種データが格納される
車速	km/h	
前後加速度	m/s ²	
左右加速度		
車輪速(4輪)	km/h	
アクセルペダル量	%	

表2 CAV取得データの集計概要

項目名	単位	備考
TRCダミー	-	作動あり=1, 作動なし=0
ABSダミー		
VSCダミー		
TRC作動率	件	区間内作動件数/対象区間内道路延長/CAV台数
ABS作動率	/km/	
VSC作動率	台	
平均車速	km/h	区間内平均値
車速分散	-	区間内で全観測データの分散
車速差	km/h	CAV速度-ETC2.0速度(区間内1ヶ月平均)
スリップ率	%	区間内最小値
車輪速差	km/h	4輪の最大値-最小値
左右加速度差	m/s ²	個車別に差を求めた値の平均値
前後加速度差		

(3) 使用データセット

本研究では、CAVから取得するデータに加えて、交通事故・車両スタックデータ、気象データ、交通状況データ、道路線形データを使用した。モデルに用いる説明変数とするために集計・整形したデータ概要を表3に示す。これらは東日本高速道路株式会社から提供を受けた。

表3 集計・整形したデータ概要

項目名	単位	備考
気象データ		
降雪量	cm/1h	対象区間から一番近い観測所のデータを使用する。 ※データに欠損がある場合は次に近い観測所のデータを使用する。一番近い観測所は緯度経度からプログラムにより判別する。
	cm/3h	
	cm/6h	
	cm/9h	
	cm/12h	
10分最大降雪量	cm/10min	
最大積雪深	cm	
平均外気温	°C	
最低外気温		
道路線形データ		
曲率	%	対象区間内の最大値
縦断勾配		対象区間内の最大値
交通データ		
全交通量	台/h	対象区間内から一番近い車両感知器のデータを使用する。
大型車交通量		
大型車混入率	-	
その他データ		
交通事故発生ダミー	-	集計時間内で発生=1, 未発生=0
車両スタック発生ダミー		除雪データより判断
除雪車通過ダミー		
月別ダミー		センサスデータより判断
山間部ダミー		
トンネルダミー		
合流/分流ダミー	ストリートビューより判断	
IC区間別ダミー		

4. 研究の全体像

(1) 合成データ生成手法の概要

本研究では、目的変数として事象発生の有無を用いることを検討したが、発生自体が稀な事象であることからサンプル数に偏りのある不均衡データとなり、有意な推定を行うために必要なデータ数を確保することが難しい。そのため、本研究ではクラス不均衡を緩和する手法の導入が不可欠であった。クラス不均衡への代表的な対処法としては、多数クラスのアンダーサンプリング、少数クラスの単純なオーバーサンプリング、コスト感度学習、合成データ生成手法が挙げられる。アンダーサンプリングは多数クラスのデータを削減するため、交通状況や気象条件などの多様な状況を十分に学習できなくなる恐れがある。単純なオーバーサンプリングは少数クラスのデータを重複させる手法であり、過学習を引き起こしやすいという問題がある。また、コスト感度学習は誤分類コストの設定に恣意性が入りやすく、結果の解釈が複雑になるという課題がある。

以上の理由から、データセットの不均衡を解消する合成データ生成手法を採用した。合成データ生成手法の代表的な手法として、SMOTE (Synthetic Minority Over-

sampling Technique) 9)がある。本研究では、通常の SMOTE手法に加え、SMOTEの拡張手法である Borderline-SMOTE¹⁰⁾ (以下、BL-SMOTE と略す) および ADASYN¹¹⁾ (Adaptive Synthetic Sampling) を併せて適用し、モデルに適合する最適な合成データ生成手法について検討したが、既往研究¹²⁾から、SMOTE は判別性能の均衡という観点で有効である一方、誤検知が多く残る傾向が確認された。また、BL SMOTE はモデルの説明力や未発生事象の判別性能に優れるものの、発生事象の検知性能が十分とは言えないことが明らかとなった。これに対し、ADASYN は発生事象の検知性能が相対的に高く、見逃しが重大な影響を及ぼす事象に対して、より安全側の判定が可能であるという特性を有していることが確認された。以上のことから、本研究では、発生事象の早期検知およびリスク把握を重視する立場に基づき、合成データ生成手法として ADASYN を採用する。これらの合成データ生成手法を適用する際には、データ生成における k-近傍点の数および生成されるデータの数を制御するオーバーサンプリング率について、最適な組み合わせをグリッドサーチにより求めた。最適な組み合わせを決定する評価指標には ROC 曲線下面積 (AUC) を用いた。この方法は Blagus らの研究においても採用され、SMOTE の効果を最大化するために有効であることが示されている。

(2) Multinomial Logit モデルの概要

本研究では対象区間・期間に対して、1時間毎・0.1KP 毎に「安全 (safe)」「交通事故 (acc)」「車両スタック (stk)」の3クラスからなる選択肢集合を定義し、その発生確率について推定する。Multinomial Logit モデル構築にあたって、各効用関数の定義を以下に示す。ここで、識別化のため基準選択肢をとする。

$$V_{safe} = 0 \quad (2)$$

$$V_{acc} = \alpha_{acc} + \sum_k \beta_{k,acc} x_k \quad (3)$$

$$V_{stk} = \alpha_{stk} + \sum_k \beta_{k,stk} x_k \quad (4)$$

ここで、

α_{acc} : 交通事故選択時の選択肢固有定数

α_{stk} : 車両スタック選択時の選択肢固有定数

x_k : 観測共変量

$\beta_{k,acc}$: 共変量 x_k の交通事故に対する係数

$\beta_{k,stk}$: 共変量 x_k の車両スタックに対する係数

観測に対する選択確率の算出式は

$$P_i(safe) = \frac{1}{1 + \exp\{V_{acc}\} + \exp\{V_{stk}\}} \quad (5)$$

$$P_i(acc) = \frac{\exp\{V_{acc}\}}{1 + \exp\{V_{acc}\} + \exp\{V_{stk}\}} \quad (6)$$

$$P_i(stk) = \frac{\exp\{V_{stk}\}}{1 + \exp\{V_{acc}\} + \exp\{V_{stk}\}} \quad (7)$$

と表され、パラメータは最尤推定法により推定する。説明変数の選択に関しては、合成データ生成によるデータセットの不均衡補正後に二方向でのステップワイズ法により、各ステップで候補変数を上位ネスト又は下位ネストにひとつずつ仮追加し、AIC 改善、閾値以下の VIF、閾値以上の McFaddenR² の改善を同時に満たす場合に採択した。なお、Multinomial Logit モデルは選択肢の独立 (IIA) を仮定するため、似た選択肢があると適合に限界が生じる。そのため、本仮定はモデルの簡潔さと推定の安定性をもたらす一方、似た選択肢が存在する状況では、視程不良等の未観測要因の共有による誤差相関を十分に表現できない可能性が示唆される。

(3) Nested Logit モデルの概要

本研究で扱う「交通事故」および「車両スタック」は、いずれも冬期の悪天候下において発生しやすい交通障害であり、「危険」といった共通の潜在的状態に起因する可能性が高い。このため、選択肢間の独立性 (Independence of Irrelevant Alternatives : IIA) を前提とする Multinomial Logit モデルでは、交通事故と車両スタックの間に存在する誤差相関を十分に表現できないことが示唆される。

そこで本研究では、IIA 制約を緩和し「危険」状態の内的相関を明示的に考慮可能な二層構造の Nested Logit モデルを採用する。本研究におけるネスト構造は、上位ネストとして「安全 (safe)」と「危険 (danger)」の二択を設定し、下位ネストとして「危険 (danger)」から派生する「交通事故 (acc)」および「車両スタック (stk)」の二択を定義する。この構造により、上位ネストでは交通事故と車両スタックに共通する発生要因の特定が可能となり、下位ネストでは両者を分岐させる要因の特定が可能となる。Nested Logit モデル構築にあたって、各効用関数の定義は Multinomial Logit モデルで使用する式(1)~式(3)と同様に定義する。危険ネストの包括値 (Inclusive Value : IV) を以下に示す。

$$IV_{danger} = \log\{\exp(V_{acc}/\lambda) + \exp(V_{stk}/\lambda)\} \quad (8)$$

ここで、 λ はネスト内のスケール (相関) パラメータである。 $\lambda=1$ が成立する場合、Nested Logit モデルは事実上 Multinomial Logit モデルに一致し、危険ネスト内の IIA が成立する。 $\lambda < 1$ が成立する場合、危険ネスト内の選択肢間に正の相関が許容され、Multinomial Logit モデルと比較してより柔軟な構造を表現できる。

上位 (ネスト) 確率を以下に示す。

$$P(danger) = \frac{\exp\{\lambda \cdot IV_{danger}\}}{\exp(V_{safe}) + \exp\{\lambda \cdot IV_{danger}\}} \quad (9)$$

$$P(\text{safe}) = \frac{\exp(V_{\text{safe}})}{\exp(V_{\text{safe}}) + \exp\{\lambda \cdot IV_{\text{danger}}\}} \quad (10)$$

下位 (ネスト) 確率を以下に示す。

$$P(\text{acc} | \text{danger}) = \frac{\exp(V_{\text{acc}}/\lambda)}{\exp(V_{\text{acc}}/\lambda) + \exp(V_{\text{stk}}/\lambda)} \quad (11)$$

$$P(\text{stk} | \text{danger}) = \frac{\exp(V_{\text{stk}}/\lambda)}{\exp(V_{\text{acc}}/\lambda) + \exp(V_{\text{stk}}/\lambda)} \quad (12)$$

最終確率を以下に示す。

$$P(\text{safe}) = P(\text{safe}) \quad (13)$$

$$P(\text{acc}) = P(\text{danger}) \cdot P(\text{acc} | \text{danger}) \quad (14)$$

$$P(\text{stk}) = P(\text{danger}) \cdot P(\text{stk} | \text{danger}) \quad (15)$$

説明変数の選択に関しては、合成データ生成手法によるデータセットの不均衡補正後に二方向でのステップワイズ法により、各ステップで候補変数を上位ネスト又は下位ネストにひとつずつ仮追加し、AIC 改善、閾値以下の VIF、閾値以上の McFaddenR²の改善を同時に満たす場合に採択した。さらに、下位ネストでは交通事故・車両スタックの等係数仮設 ($\beta_{k,\text{acc}} = \beta_{k,\text{stk}}$) の棄却と、同時ゼロ仮設 ($\beta_{k,\text{acc}} = \beta_{k,\text{stk}} = 0$) の棄却を Wald 検定で確認した。包括値パラメータ λ は $0 < \lambda \leq 1$ の範囲条件を満たす場合のみを採択した。

(4) Mixed Logit モデルの概要

本研究で扱う、Mixed Logit モデルは、Multinomial Logit モデルやNested Logit モデルを拡張し、選択主体間に存在する未観測な異質性を考慮可能とする離散選択モデルである。本研究ではまず、Multinomial Logit モデルおよびNested Logit モデルを構築し、交通事故と車両スタックの発生予測モデルの有効性について検討する。一方で、いずれのモデルにおいても説明変数の係数を固定値として仮定しているため、同一の気象条件や交通条件下であっても、区間特性や時間帯、交通管理状況、個人属性の違いにより生じる交通事故および車両スタックの発生リスクのばらつきを十分に捉えられない可能性がある。そこで本研究では、Multinomial Logit モデルおよびNested Logit モデルの比較により適合度および判別性能の観点から選定されたモデルを基礎として、説明変数の係数を確率分布に従う確率変数として扱う混合ロジットモデルへ拡張し、未観測異質性を考慮することでモデルの適合度の向上を図る。上記の考え方にに基づき、本研究では、Multinomial Logit モデルおよびNested Logit モデルにより選定されたモデル構造を基礎として、説明変数の係数に未観測異質性を導入した Mixed Logit モデル (Mixed Multinomial Logit, Mixed Nested Logit) を構築する。

各効用関数の定義は Multinomial Logit モデルで使用する式(2)~式(4)と同様に定義する。Mixed Logit モデルでは、係数 $\beta_{k,i,j}$ を以下のように分解する。

$$\beta_{k,i,j} = \bar{\beta}_{k,j} + \eta_{k,i,j} \quad (16)$$

ここで、

$\bar{\beta}_{k,j}$: 母集団平均を表す固定成分

$\eta_{k,i,j}$: 平均 0, 分散 σ_k^2 の確率項

$$\eta_{k,i,j} = N(0, \sigma_k^2) \quad (17)$$

このとき、条件付き (β_i が与えられた下での) 選択確率は、通常の Multinomial Logit モデルと同様に、

$$P_{ij}(\beta_i) = \frac{\exp(V_{ij})}{\sum_{m \in \{\text{safe}, \text{acc}, \text{stk}\}} \exp(V_{im})} \quad (18)$$

と表される。

Mixed Logit モデルにおける最終的な選択確率は、ランダム係数分布を考慮した周辺化により、

$$P_{ij} = \int \frac{\exp(V_{ij}(\beta))}{\sum_m \exp(V_{im}(\beta))} f(\beta | \theta) d\beta \quad (19)$$

として与えられる。ここで、 $f(\beta | \theta)$ は係数ベクトル β の確率密度関数、 θ はその分布パラメータである。

上式は解析的に閉形式解を持たないため、本研究ではシミュレーション最尤法を用いて選択確率を近似し、モデルパラメータを推定する。

(5) モデル評価指標

本研究では、モデルの分類性能を評価するために、混同行列 (表 4) に基づく各種評価指標を用いる。

表 4 混同行列の概要

		Actual (実測結果)	
		0	1
Prediction (モデル 予測結果)	0	TN (真陰性)	FN (偽陰性)
	1	FP (偽陽性)	TP (真陽性)

ここで、0 は「事象が発生しない場合」、1 は「事象が発生する場合」を表す。この判定については、モデルから推定する発生確率の高い事象について事象発生と判定した。

混同行列は真陽性 (TP)、偽陽性 (FP)、真陰性 (TN)、偽陰性 (FN) の 4 要素から構成される。これらに基づき、Precision, Recall, Specificity, Balanced Accuracy, および F₁-Score を算出した。

各評価指標の定義を以下に示す。

Precision (適合率) は、陽性と予測した事例のうち実際に陽性である割合を表す。

$$\text{Precision} = \frac{TP}{TP + FP} \quad (20)$$

Recall (再現率) は、実際に陽性である事例のうち正しく陽性と識別できた割合を表す。

$$\text{Recall} = \frac{TP}{TP + FN} \quad (21)$$

Specificity (特異度) は、実際に陰性である事例のうち正しく陰性と識別できた割合であり、真陰性率とも呼ば

れる。偽陽性の抑制が重要な場合に重視される。

$$Specificity = \frac{TN}{TN + FP} \quad (22)$$

Balanced Accuracy は、再現率（真陽性率）と特異度（真陰性率）の平均であり、クラス不均衡下でも両クラスを同等に扱えるため有用である。

$$Balanced Accuracy = \frac{1}{2} \cdot (Recall + Specificity) \quad (23)$$

F_β -Score は、Precision と Recall の加重調和平均である。パラメータ β は Recall と Precision の相対的な重みを決め、 $\beta > 1$ で Recall を、 $\beta < 1$ で Precision を相対的に強調する。

$$F_\beta\text{-Score} = (1 + \beta^2) \cdot \frac{Precision \cdot Recall}{(\beta^2 \cdot Precision) + Recall} \quad (24)$$

本研究では $\beta = 1$ を採用し、Precision と Recall を同等に重視する F1-Score によりモデル性能を評価した。

5. モデル構築結果

(1) Mixed Multinomial Logit モデルの構築

Mixed Multinomial Logit モデルを構築した結果を表 5 に示す。交通事故と車両スタックに共通して有意な影響を示した要因として、車速分散、車速差、大型車交通量、大型車混入率が挙げられる。車速分散および車速差はいずれの事象においても負の係数を示しており、速度が低く、かつ交通流が均一化した状況で両事象の発生確率が高まる傾向が確認された。これは、渋滞や低速走行状態が継続する環境が、交通事故および車両スタックの双方にとってリスクを高める要因となることを示唆している。また、大型車交通量は正の係数、大型車混入率は負の係数を示し、大型車の絶対量が増加する一方で交通流全体が不安定化する状況が、両事象の発生に参与している可能性が示された。

表 5 Mixed Multinomial Logit モデル説明変数

説明変数	交通事故			車両スタック		
	係数	標準誤差	Z 値	係数	標準誤差	Z 値
車速分散	-0.608***	0.0511	-11.9	-1.12***	0.0913	-12.3
車速差	-0.290***	0.0411	-7.06	-2.24***	0.0838	-26.7
左右加速度差	-1.960***	0.0767	-25.5	0.943***	0.0608	15.5
前後加速度差	0.845***	0.0502	16.8	-0.0305	0.0626	-0.487
大型車交通量	0.519***	0.0393	13.2	0.298***	0.0547	5.46
大型車混入率	-0.996***	0.0490	-20.3	-0.852***	0.0599	-14.2
除雪車通過ダミー	0.0687	0.0373	1.84	1.15***	0.0780	14.7
縦断勾配	-0.551***	0.0364	-15.1	2.43***	0.0803	30.3
山間部ダミー	0.771***	0.0443	17.4	0.0114	0.0559	0.204
越後川口 IC 区間ダミー	-0.454***	0.0391	-11.6	-2178	329536	-0.00661
AIC	10,200.8					
McFaddenR ²	0.4656					

*** p-value < 0.001, ** p-value < 0.01, * p-value < 0.05

交通事故に特有の影響を示す要因として、左右加速度差および前後加速度差が挙げられる。左右加速度差は強い負の影響を示し、横方向の挙動変化が大きい区間では事故発生確率が高まる傾向が明確に確認された。また、前後加速度差は正の係数を示し、加減速変動が大きい不安定な走行状態が、事故の突発的な発生と関連していることが示唆される。車両スタックに特有の影響を示す要因として、除雪車通過ダミーおよび縦断勾配が挙げられる。除雪車通過ダミーは正の係数を示しており、除雪作業が行われるほど厳しい降雪・路面条件下において車両スタック発生リスクが顕在化することが示された。また、縦断勾配は事故では負の影響を示す一方で、車両スタックでは強い正の影響を示し、上り坂区間において車両スタックが発生しやすいという現象が明確に捉えられている。これは、積雪・低摩擦環境下での駆動力不足が車両

スタック発生の主要因となることを反映した結果と解釈できる。空間的要因に着目すると、山間部ダミーは交通事故でのみ有意な正の影響を示し、山間部特有の道路線形や視認性の低下が事故発生に参与している可能性が示唆される。一方で、越後川口 IC 区間ダミーは両事象に対して有意な影響を示さず、区間特性が一樣に交通事故および車両スタックに作用するわけではないことが確認された。

(2) Mixed Nested Logit モデルの構築

Nested Logit モデルでは、交通事故および車両スタックを「大規模滞留を引き起こす事象」という上位概念の下に位置付け、上位ネストと下位ネストの二段階構造を仮定して推定を行った。Mixed Nested Logit モデルにおける上位ネストの結果を

表 6 に、下位ネストの結果を

表 7 に示す。上位ネストについて、車速分散は負の有意な係数を示しており、走行速度が均一であるほど、大規模滞留を引き起こす事象全体の発生確率が高まる傾向が確認された。また、前後加速度差は正の有意な係数を示し、加減速変動の大きい不安定な走行状態が、事故・スタックのいずれかが発生する状況と強く関連することが示唆される。さらに、山間部ダミーおよび小出 IC 区間ダミーはいずれも正で有意であり、地形条件や特定区間が大規模滞留発生の基盤的なリスク要因であることが

明らかとなった。下位ネストでは、交通事故と車両スタックでは、発生に寄与する要因の方向性および影響の大きさに明確な差異が確認できる。交通事故については、車速差および左右加速度差が負の有意な係数を示し、平常時と比べて低速な走行や横方向挙動の不安定化が進行する過程で事故発生確率が高まる傾向が確認された。一方、大型車交通量や最大 10 分間降雪量は正の有意な係数を示し、交通密度の上昇や短時間の強い降雪が、事故発生の直接的な誘因となる可能性が示唆される。

表 6 Mixed Nested Logit モデル説明変数（上位ネスト）

説明変数	係数	標準誤差	Z 値
車速分散	-0.609***	0.0517	-11.8
前後加速度差	0.500***	0.0268	18.6
山間部ダミー	0.510***	0.0390	13.1
小出 IC 区間ダミー	0.386***	0.0269	14.4

表 7 Mixed Nested Logit モデル説明変数（下位ネスト）

説明変数	交通事故			車両スタック		
	係数	標準誤差	Z 値	係数	標準誤差	Z 値
車速差	-0.527***	0.0427	-12.3	-1.42***	0.0763	-18.5
左右加速度差	-1.11***	0.0849	-13.1	0.130***	0.0327	3.97
大型車交通量	0.486***	0.0413	11.8	0.348**	0.0425	8.19
大型車混入率	-0.927***	0.0500	-18.5	-0.564***	0.0483	-11.7
最大 10 分間降雪量	0.0984**	0.0377	2.61	0.319***	0.0390	8.17
縦断勾配	-0.0112	0.0370	-0.303	1.36***	0.0763	-18.5
AIC	9,791.5					
McFaddenR ²	0.4868					

*** p-value < 0.001, ** p-value < 0.01, * p-value < 0.05

縦断勾配については有意な影響が認められず、事故発生は勾配条件よりも走行状態や交通環境の変化に強く依存することが示唆される。車両スタックでは、車速差が強い負の影響を示す一方で、左右加速度差、縦断勾配、大型車交通量、最大 10 分間降雪量はいずれも正で有意となっている。特に縦断勾配の係数は大きく、上り坂条件下において駆動力不足や車両性能差が顕在化し、スタック発生に直結しやすい構造が統計的に裏付けられた。また、大型車混入率はいずれの事象においても負の有意な係数を示し、大型車比率が高い時間帯では、交通流が比較的安定している状況が反映されている可能性が示唆される。

(3) モデル比較

表 8、表 9 に示す分類結果、表 10、表 11 に示す評価指標に基づき、Mixed Multinomial Logit モデルと Mixed Nested Logit モデルの性能を比較する。

表 8 Mixed Multinomial Logit モデル

Prediction	Actual		
	安全	交通事故	車両スタック
安全	165,959	10	5
交通事故	17,966	14	1
車両スタック	14,944	1	19

表 9 Mixed Nested Logit モデル

Prediction	Actual		
	安全	交通事故	車両スタック
安全	159,928	9	4
交通事故	19,688	16	0
車両スタック	19,253	0	21

表 8 より、交通事故については真陽性が 14 件、偽陰性が 10 件であり、発生事象を一定程度検知できているものの、見逃しも依然として存在することが確認できる。一方で、安全状態を交通事故と誤判定する偽陽性が 17,966 件と非常に多く、未発生事象を誤って交通事故と

判定するケースが大量に生じている。この結果は、少数クラスである交通事故を過剰に予測する傾向が強いことを示す。車両スタックについては、真陽性が19件、偽陰性が9件であり、交通事故と同様に一定数の発生事象を検知できている。一方で、偽陽性は14,944件と極めて多く、安全状態を誤って車両スタックと判定するケースが多数確認された。このことから、車両スタックの見逃しはある程度抑制されているものの、誤検知が顕著であり警報の過剰発生につながる可能性が高い。表9より、交通事故についてみると、真陽性は14件であり、一定数の発生事象を検知できている。一方で、偽陰性は11件であり、発生した交通事故の約半数近くを見逃していることが確認できる。偽陽性は16,373件と多く、安全状態を交通事故と誤判定するケースが大量に発生している。車両スタックについては、真陽性が20件、偽陰性が5件であり、発生事象の大部分を検知できている。一方で、偽陽性は14,427件と非常に多く、安全状態を車両スタックと誤判定する事例が多数生じている。このことから、Mixed Nested Logit モデルは車両スタックの検知性能自体は比較的高いものの、発生確率を過大に評価する傾向があり、誤検知が大きな課題となる。両モデルの評価指標を示す表10、表11より、全体的なモデル適合度に着目すると、Mixed Nested Logit モデルはAICが9791.5と、Mixed Multinomial Logit モデルの10200.8よりも小さく、McFaddenR²も0.4868と高い値を示す。このことから、事象間の階層構造を考慮したMixed Nested Logit モデルの方が、データへの当てはまりという観点では優れていることが確認できる。交通事故について比較すると、Recallはいずれのモデルでも0.56と同程度であり、発生事象の検知性能に大きな差は確認できない。一方で、SpecificityはNested Logit モデルが0.9135とわずかに高く、未発生事象の誤検知抑制において若干の改善が確認できる。その結果、Balanced AccuracyもMixed Nested Logit モデルが0.7363と、Mixed Multinomial Logit モデルの0.7348を僅かに上回る。ただし、F1はいずれも極めて低く、交通事故クラスの判別が依然として難しい課題である。

表10 Mixed Multinomial Logit モデル評価指標

	安全	交通事故	車両スタック
Recall	0.8345	0.5600	0.7600
Specificity	0.7000	0.9097	0.9249
Balanced Accuracy	0.7673	0.7348	0.8424
F1	0.9098	0.0016	0.0025
Accuracy	0.8345		
AIC	10200.8		
MacFaddenR ²	0.4656		

表11 Mixed Nested Logit モデル評価指標

	安全	交通事故	車両スタック
Recall	0.8292	0.5600	0.8000
Specificity	0.6800	0.9135	0.9158
Balanced Accuracy	0.7546	0.7363	0.8579
F1	0.9066	0.0016	0.0024
Accuracy	0.8292		
AIC	9791.5		
MacFaddenR ²	0.4868		

車両スタックについて比較すると、Mixed Nested Logit モデルではRecallが0.80と、Mixed Multinomial Logit モデルの0.76より高く、発生事象の検知性能が向上することが確認できる。一方で、Specificityは両モデルとも0.92前後と高い水準を維持し、未発生事象の誤判定は一定程度抑制される。その結果、Balanced AccuracyはMixed Nested Logit モデルが0.8579と、Mixed Multinomial Logit モデルの0.8424を上回り、車両スタックに関してはMixed Nested Logit モデルの方が総合的な判別性能に優れる結果となる。

6. リスク指標の検討

前章で良好な性能を示したMixed Nested Logit モデルから算出する交通事故・車両スタック発生の選択確率を事象発生リスクとして位置付け、実際の交通事故および車両スタックの発生を適切に予測できているか、また発生前の段階で検知が可能であるかといった観点から、モデルの実用的有効性を評価する。

(1) リスク指標の時間推移

リスク指標 (Mixed Nested Logit モデルにより算出された選択確率) の時間推移から、交通事故や車両スタックの発生時にリスクが高水準となるか、また発生前の時間帯から高リスクとして事前検知できるかについて検討した。図1に示す、交通事故に関して、発生リスクの平均値は-9h~-6hでは大きな変化が見られない一方、-5h以降に上昇が明確となり、t=0近傍で最大となる。分布も発生時刻に近づくほど高リスク側へ移動することから、交通事故は、発生直前にリスクが顕在化する特徴を有することが示唆される。一方で分布のばらつきが大きく、発生直前であっても低リスクとなり得るため、取りこぼしが生じる可能性がある。図2に示す、車両スタックに関して、発生リスクの平均値は発生の-6h前後から上昇し、-5h~-1hで特に高水準となり、t=0近傍で最大となる。このことは、車両スタックでは交通事故に比べて早期から兆候が確認しやすく、より前段階での検知可能性が高い。ただし全体的に高リスク側へ偏りやすい分布であるため、固定閾値のみの運用では誤警報のリスクが高

まることに留意が必要である。

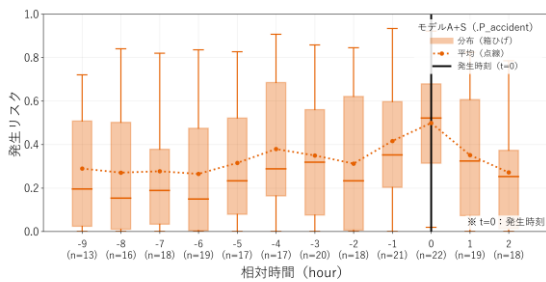


図 1 交通事故発生リスクの時間推移

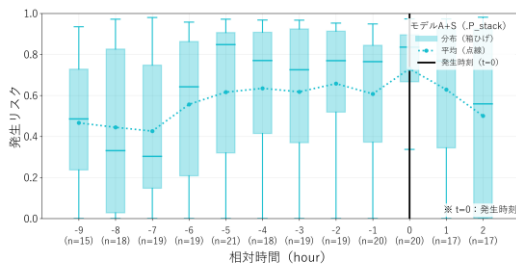


図 2 車両スタック発生リスクの時間推移

以上より、本モデルの出力確率をリスク指標として用いることで、交通事故・車両スタックともに発生時刻に整合したリスク上昇を捉えられることが示された。加えて、交通事故は直前上昇型、車両スタックは早期高止まり型という時間推移の差異が観察され、事象特性に応じた運用設計（短時間先の警戒強化か、早期介入か）に資する可能性がある。一方で分布のばらつきや高水準化の傾向を踏まえると、実運用では確率の絶対値に加えて、上昇量や高リスク継続時間など時系列情報を併用した判定基準の導入が望ましい。

(2) 未発生時と発生前・発生時のリスク指標比較

交通事故の未発生時と発生前・発生時のリスク指標比較について、提示した確率密度分布から、未発生群は低リスク側に集中する一方、発生群は相対的に高リスク側へ分布が移動しており、両者の分布形状の差を通じてモデル出力が「発生しやすさ」の状態差を表現できていることが確認できる。図 3 に示す、交通事故発生前（-1h～-6h）では、発生群は未発生群に比べて高リスク側への裾を一定程度持つが、低～中リスク域での重なりが大きく、分離は限定的である。これは、事故の兆候が発生前段階で現れる場合はある一方、必ずしも一貫して高リスクとして出力されるわけではなく、事前検知には不確実性が残ることを示唆する。また、図 4 に示す、交通事故発生時（t=0）では、未発生群が 0 付近に強く集中することに対し、発生群は 0.5～0.8 程度の中～高リスク域に明確な山を持ち、両者の分布の乖離が拡大する。

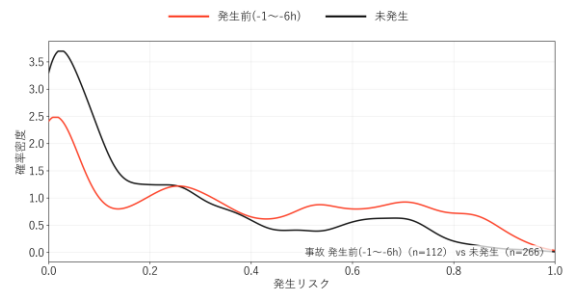


図 3 交通事故発生前と未発生時の分布比較

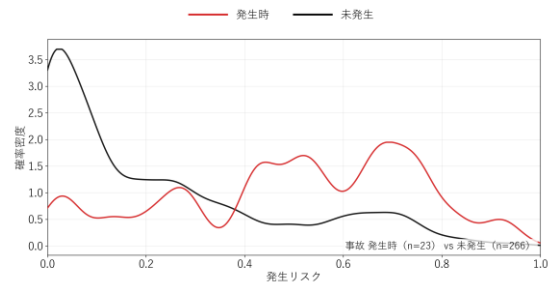


図 4 交通事故発生時と未発生時の分布比較

車両スタックの未発生時と発生前時のリスク指標比較について、車両スタックに関する確率密度分布から、未発生群は低リスク側に大きな山を持つ一方で、高リスク側にも一定の密度が残る。これに対し発生群は、高リスク側（概ね 0.7～1.0）で密度が顕著に高まり、未発生群との分布差が明確化している。すなわち、スタックではリスク指標が高リスク域に強く反応しやすく、発生に関連する状態差を分布として捉えられていることが示唆される。図 5 に示す、車両スタック発生前（-1h～-6h）では、発生群は低リスク域の密度が相対的に低下し、高リスク域（特に 0.8～1.0 付近）に大きな山を形成している。未発生群も低リスク側に集中しつつ高リスク側へ裾を持つが、発生群はそれを上回る密度で高リスク側に偏っており、発生前段階から一定の分離が成立している。これは、スタックでは事故と比較して兆候が早期から顕在化しやすく、発生前検知の可能性が相対的に高いことを示す。また、図 6 に示す、車両スタック発生時（t=0）では、高リスク域（0.8～1.0）における発生群の山がさらに鋭くなり、未発生群との差が一層拡大する。このことは、スタック発生時点ではモデルが発生ケースを高リスクとして強く出力する傾向が明確であり、識別性が発生前よりも向上していることを示唆する。ただし、未発生群においても高リスク側の密度が一定程度存在するため、高リスク値が常に発生を意味するわけではなく、固定閾値のみでは誤警報が残り得る点に留意が必要である。

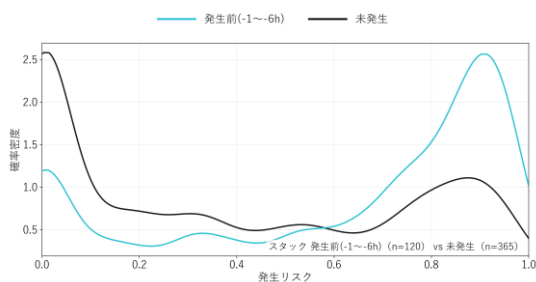


図 5 車両スタック発生前と未発生時の分布比較

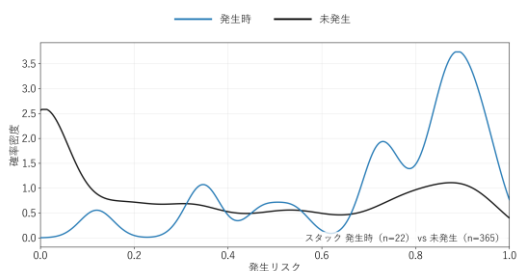


図 6 車両スタック発生時と未発生時の分布比較

7. 結論

本研究は、CAV から取得可能なプローブデータを活用し、冬期高速道路における交通事故および車両スタックの発生メカニズムを構造的に解明するとともに、低頻度で発生するこれら事象の発生リスクを定量的に評価し、①大規模な車両滞留を未然に防止する道路管理、交通運用方策に資する知見を得ること、さらに②各事象の発生リスク指標の確立と実用性を検討することを目的とした。第一に、CAV の運転支援作動履歴や走行挙動、交通、気象、道路線形、推定路面状況といった複数要因を統合した離散選択モデルにより、「安全」から「危険」への移行、および危険状態における「事故/スタック」の分岐という階層的な発生構造を表現し、豪雪時の交通障害を一体として扱う推定枠組みを提示した。また、モデル比較の結果、Mixed Nested Logit は事象間の類似性や階層構造を考慮することで、特に車両スタックの検知性能や適合度の面で改善が確認され、実務上の選択肢として有効であることが示唆された。第二に、低頻度事象に対する定量的リスク評価として、推定確率をリスク指標として解釈し、発生前、発生時における分布の差異から指標の性質を検討した。その結果、車両スタックは発生前段階から高リスク側への偏りが相対的に強く、発生時には高リスク域での分離が一層拡大する特徴を示した。これは、スタックに関しては発生前検知と予防的介入の余地が比較的大きい可能性を示す一方、未発生時にも一定の高リスク値が出現し得るため、単一の固定閾値のみでの運用では誤警報が残る点に留意が必要である。大規模な

車両滞留の未然防止に向けては、上記のリスク指標を、「高リスクの継続時間」、「リスク値上昇幅」、等の時系列条件と併用して判断する設計が有効である。これにより、除雪作業の重点化、注意喚起や速度管理の前倒し、予防的通行止め、入口制御等の実施判断の高度化といった、道路管理・交通運用方策への接続が期待できる。

REFERENCES

- 1) 東日本高速道路株式会社：関越道大雪に伴う大規模車両滞留を踏まえた NEXCO 東日本の雪氷対策の強化・改善策について、URL：https://www.nexco.co.jp/assets/pdf/pressroom/data_room/regular_mtg/r03/0127/01.pdf, (閲覧年月日 2025 年 10 月 4 日)
- 2) 総務省：平成 27 年版情報通信白書 | コネクテッドカー, URL：<https://www.soumu.go.jp/johotsusintokei/whitepaper/jah27/html/nc241210.html>, (閲覧年月日 2024 年 10 月 4 日)
- 3) Kim, D., Jung, S., Yoon, S.: Risk Prediction for Winter Road Accidents on Expressways, *Applied Sciences*, Vol. 11, Article 9534, 2021.
- 4) 針原拳太, 佐野可寸志, 加藤哲平, 原山哲郎：冬期高速道路における路面状況を考慮した走行速度の時系列モデルの構築, *交通工学研究会論文集* 43(0), 567-574, 2023
- 5) Chen, Z., Fan, W. (David): A Multinomial Logit Model of Pedestrian-Vehicle Crash Severity in North Carolina, *International Journal of Transportation Science and Technology*, Vol. 8, pp. 43-52, 2019.
- 6) Tavakoli Kashani, A., Rashidi, A., Amirifar, S.: Identifying Types of Freeway Crashes Using Nested Logit Model, *Amirkabir Journal of Civil Engineering*, 56(1), pp. 23-26, 2024.
- 7) Milton, J. C., Shankar, V. N., Mannering, F. L.: Highway accident severities and the mixed logit model: An exploratory empirical analysis, *Accident Analysis and Prevention*, Vol. 40, No. 1, pp. 260-266, 2008, DOI: 10.1016/j.aap.2007.06.006.
- 8) Das, S., Jafari, M., Hossain, A., Chakraborty, R., Mimi, M. S.: Toll road crash severity using mixed logit model incorporating heterogeneous mean structures, *Transportmetrica A: Transport Science*, 2024
- 9) Chawla, N. V., Bowyer, K. W., Hall, L. O., Kegelmeyer, W. P.: SMOTE: Synthetic Minority Over-sampling Technique, *Journal of Artificial Intelligence Research*, Vol. 16, pp. 321-357, 2002.
- 10) Han, H., Wang, W. Y., Mao, B. H.: Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning, *Advances in Intelligent Computing*, Vol. 3644, pp. 878-887, 2005.
- 11) He, H., Yan, S., Zhang, J.: ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning, *IEEE International Conference on Data Mining (ICDM)*, pp. 272-277, 2008.
- 12) Kataoka, K., Sano, K., Kato, T., Harayama, T.: Development of a model for risk of occurrence of stuck on winter highways, *Journal of the Eastern Asia Society for Transportation Studies*, Vol. 16, PP4172, 2025.