新潟県における冬期の高速道路での交通事故発生要因に関する分析

渡邊　和志[1]・佐野　可寸志[2]・加藤　哲平[3]
原山　哲郎[4]・松田　曜子[5]

[1]非会員　長岡技術科学大学大学院 環境社会基盤工学専攻（〒940-2188　新潟県長岡市上富岡町 1603-1）
E-mail: s191081@stn.nagaokaut.ac.jp

[2]正会員　長岡技術科学大学 環境社会基盤系 教授（〒940-2188　新潟県長岡市上富岡町 1603-1）
E-mail: sano@vos.nagaokaut.ac.jp

[3]正会員　長岡技術科学大学 環境社会基盤系 講師（〒940-2188　新潟県長岡市上富岡町 1603-1）
E-mail: tkato@vos.nagaokaut.ac.jp

[4]正会員　長岡技術科学大学大学院 工学研究科（〒940-2188　新潟県長岡市上富岡町 1603-1）
E-mail: s237007@stn.nagaokaut.ac.jp

[5]正会員　長岡技術科学大学大学院 准教授（〒940-2188　新潟県長岡市上富岡町 1603-1）
E-mail: ymatsuda@vos.nagaokaut.ac.jp

本研究では，冬期の高速道路での交通事故発生頻度が年間で特に多いことから，高速道路における降雪条件下での交通事故発生要因を分析し，冬期の高速道路での交通事故発生確率を予測するモデルを作成し，降雪時の交通事故リスクを事前に予見し，事故抑制に役立てることを目的とする．本研究では，新潟県全域をカバーする５つの主要道路で発生した交通事故を対象としている．

先行研究では，負の二項回帰モデルを作成し，降雪時の高速道路における１時間の交通事故発生確率を予測していた．しかし，このモデルは積雪深さや路面状況などの要因をモデルに考慮できていない点や，希少な事象である事故件数を目的変数としているので，高精度のモデルを作成する上で，問題が生じている．モデルの精度向上のために，これらの問題の解決を行う．

Key Words: *heavy snowfall, accident, Rare data, Probability Prediction, Negative binomial regression*

1. はじめに

冬期に発生する交通事故は，大雪や吹雪，強風，霧などにより，ドライバーの視程不良や路面状態の悪化などを引き起こす．また本研究で対象とする，新潟県における冬期の高速道路での事故発生件数は，他の季節に比べ多くなっていることが（事故データから）分かった．

冬期の交通事故を減らすために，交通事故発生に影響を及ぼす要因を分析し，対策を講じることが求められる．

このため本研究では，新潟県を対象とし，高速道路における積雪条件下での交通事故発生要因を分析し，24時間当たりの交通事故発生傾向を予測する確率モデルの開発を行うことで，事故の発生確率の高い時間帯や区間の計画的通行止め，事故の発生しやすい区間の整備・改善策の検討に活用することを目的とする．

2. 先行研究および本研究の新規性

本研究を行うにあたり，先行研究および新規性について述べる．

冬期の高速道路における降雪条件下での事故に関する研究では，佐野ら [1]は，事故の重大性や事故リスクに影響を及ぼす要因の分析が中心であった．また，Maら[2]は，高速道路で発生する交通事故の発生頻度に影響を及ぼす要因を分析し，交通事故発生確率を予測するモデルを開発した研究もあるが，降雪・積雪は考慮されていなかった．Manhら[3]は，負の二項回帰モデルを作成し，1時間当たりの交通事故発生確率の予測を行っていたが，交通事故という希少な事象を対象としているため，モデルの精度は低くなっていた．

そこで，本研究では，先行研究と同様に負の二項回帰モデルを用いて，交通事故発生確率の予測を行う期間を1時間から24時間に拡大することで，データセットに含

まれる多くの事故件数が 0 件のサンプルを減らし，モデルの予測精度の向上を試みた．

本研究の新規性は，負の二項回帰モデルを用いて，短期間での交通事故発生確率に影響を及ぼす要因を分析し，降雪状況などを考慮した短期間の高速道路での交通事故発生確率を予測することにある．

3. 研究の全体像

(1) 研究の流れ

本研究では，1 時間ごとの事故データセットを用いて，24 時間ごとの事故データセットを作成する．データセット作成後，モデルに使用する説明変数を選択し，選択した説明変数に対して，多重共線性を避けるため，相関分析を用いて変数間の相関を明らかにする．その後，負の二項回帰モデルを用いて，パラメータの推定を行い，交通事故発生に影響を及ぼす要因を分析し，交通事故発生確率の予測を行い，分析結果と，先行研究で得られた 1 時間当たりの交通事故発生確率の予測結果との比較を行う．

(2) 使用したデータセット

本研究で使用した 1 時間ごとの事故データセットには，東日本高速道路株式会社（NEXCO）から提供していただいた，新潟県を通る 5 本の高速道路内で起きた交通事故のデータ，交通状況データ，道路線形データ，天候データが含まれる．それぞれのデータを，データの集計時間と区間に基づいて結合し，24 時間当たりの事故データセットを作成した．

図-1 は，事故件数の分布である．**図-1** より，事故データセットの総サンプル数は，1 時間当たりの事故データセットが 20,276,880 件，24 時間当たりの事故データセットが 844,870 件となった．これにより，19,431,725 個の 0 件データを解消することができた．また，総事故数は，1時間当たりの事故データセットが6,306件，24時間当たりの事故データセットが 6,021 件となった．

事故件数(件)	個数(個)	
	1h	24h
0	20,270,574	838,849
1	6,020	5,514
2	286	457
3	0	38
4	0	10
5	0	2

図-1 事故件数の分布

(3) 使用するモデル

本研究では，負の二項回帰モデルを用いて分析を行う．負の二項モデルは，他の統計モデルと比較して，交通事故予測のモデリングに広く適用されている．負の二項回帰モデルは，従属変数 Y が説明変数に依存する平均値 μ の負の二項回帰に従うと仮定すると，確率密度関数を持つ負の二項回帰モデルは以下の式(1)が与えられる．

$$P_r(Y = y_i|\mu_i, \alpha) = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(\alpha^{-1})\Gamma(y_i + 1)}\left(\frac{1}{1 + \alpha\mu_i}\right)^{\alpha^{-1}}\left(\frac{\alpha\mu_i}{1 + \alpha\mu_i}\right)^{y_i} \quad (1)$$

ここで，$P_r(Y = y_i|\mu_i, \alpha)$ は，j 時間の間に区間 i^{th} に発生する y_{ij} の確率，y_i は，j 時間の間に区間 i^{th} に発生した事故の数，α は，分散パラメータ，μ_i は，区間 i^{th} における事故発生件数の期待値で，本研究では，以下の関数(2)に従う

$$\mu_i = e^{\beta_0} L_i \times TV_i^{\beta_1} \times e^{(\beta_2 X_{2i} + \cdots\cdots + \beta_k X_{ki} + \varepsilon_i)} \quad (2)$$

L_i は，区間 i^{th} の長さ，TV_i は，区間 i^{th} の j 時間当たりの交通量，X_{ki} は，降雪量，気温，鉛直勾配など，j 時間目の第 i 区間における説明変数 k 番目，ε_i は，平均 1，分散α2 のガンマ分布の誤差項，β_k は，平均 1，分散α2 のガンマ分布の誤差項である．係数 β_k と過分散パラメータαは，対数尤度関数を最大化することで推定される．

4. 結果と考察

(1) 選択した説明変数

図-2 に，選択した説明変数の基本統計量を示す．モデルを作成するために選択した説明変数は，区間長(km)，24 時間当たりの交通量(台/日)，トラックの割合(%)，24 時間当たりの車両の平均速度(km/h)，24 時間当たりの降雪量(cm/日)，気温(℃)，勾配(%)，中央分離帯の有無である．

説明変数	最小	平均	最大	標準偏差
事故件数(件)	0	0.007802	5	0.09635
区間長(km)	0.2060	0.9484	3.905	0.5639
交通量(台/日)	82	10,234	44,777	6,463
トラックの割合(%)	0	21.94	42.12	8.727
平均速度(km/h)	6.639	84.54	126.3	8.515
降雪量(cm/日)	0	0.6045	19.33	1.347
気温(℃)	-6.688	4.529	19.60	3.323
勾配(%)	-4.500	-0.02003	4.500	1.503

説明変数		区間の個数(個)
中央分離帯の有無	有り	168,780
	無し	676,090

図-2 説明変数の基本統計量

(2) 相関分析結果

図-3 に，説明変数間の相関分析の結果のヒートマップを示す．モデル内の変数の相関係数は，-0.538～0.431 となった．これは，モデル内の説明変数間に多重共線性が

ないことを示している．

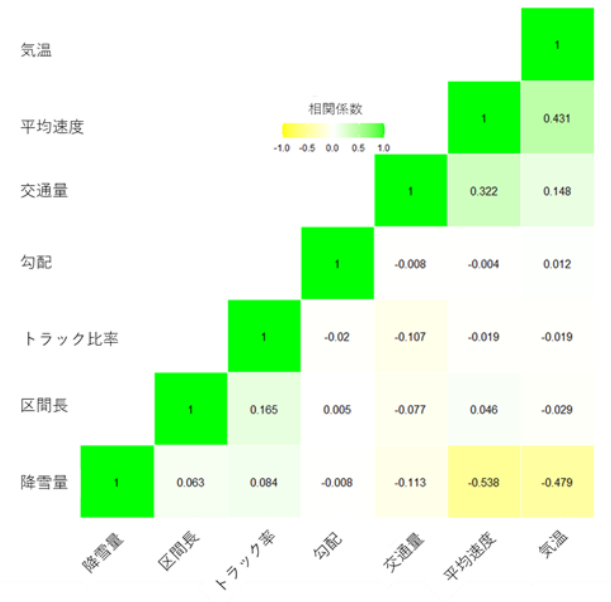


図-3　相関係数のヒートマップ

(3)　負の二項回帰モデルのパラメータ推定結果

作成した事故データセットに対して，事故データセットの 80%をトレーニングデータ，残りの 20%をテストデータとし，トレーニングデータを用いて予測を行う．その後，予測結果とテストデータを比較し，モデルの評価を行う．

図-4 に，負の二項回帰モデルを用いて，24 時間当たりの事故データセットからパラメータを推定した結果を示し，**図-5** に，負の二項回帰モデルを用いて，1 時間当たりの事故データセットからパラメータを推定した結果を示す．

図-4 より，有意水準 1%において，説明変数は統計的に有意であることが分かった．これは，冬期の高速道路で発生する 24 時間当たりの交通事故発生確率に対して，すべての要因が統計的に有意であるということを示している．交通量と平均降雪量は，交通事故発生確率に正の影響を与えることが分かった．また，大型車混入率，車両の平均速度，気温，勾配は，交通事故発生確率に負の影響を与えることが分かった．

図-5 より，負の二項回帰モデルによる，1 時間当たりの事故データセットからパラメータを推定した結果と，24 時間当たりの事故データセットからパラメータを推定した結果を比較すると，どちらも説明変数が統計的に有意であり，それぞれの説明変数の推定値の符号も同じであることが分かった．また，24 時間当たりの事故データセットから推定されたパラメータは，勾配と気温が事故発生確率に与える負の影響が大きくなり，他の変数は正の影響，負の影響が小さくなった．赤池情報量基準の値は，24 時間の事故データセットから推定されたパラメータの方が小さくなっていたが，デビアンス残差は大きくなった．

説明変数	推定値	標準誤差	z 値	Pr(>|z|)
切片	-3.015	0.1371	-21.99	< 2e-16
交通量(台/日)	5.583.E-05	2.486.E-06	22.46	< 2e-16
トラックの割合(%)	-1.203	0.1612	-7.465	8.3E-14
降雪量(cm/日)	0.09217	0.008537	10.80	< 2e-16
気温(℃)	-0.09774	0.005951	-16.42	< 2e-16
平均速度(km/h)	-0.04856	0.001605	-30.26	< 2e-16
中央分離帯の有無	1.019	0.06964	14.63	< 2e-16
勾配(%)	-0.1108	0.009422	-11.76	< 2e-16
標準偏差	0.0047			
Null deviance	33045 on 675895			
Residual deviance	29117 on 675888			
赤池情報量基準	56046			

図-4　負の二項回帰モデルのパラメータ推定結果（24 時間）

説明変数	推定値	標準誤差	z 値	Pr(>|z|)
切片	-3.830	0.2148	-17.83	< 2e-16
交通量(台/日)	1.002E-03	7.756E-05	12.92	< 2e-16
トラックの割合(%)	-2.757	0.2297	-12.00	< 2e-16
降雪量(cm/日)	1.484	0.2515	5.902	3.58E-09
気温(℃)	-0.06046	0.007666	-7.886	3.12E-15
平均速度(km/h)	-0.06004	0.002553	-23.52	< 2e-16
中央分離帯の有無	1.096	0.09538	11.49	< 2e-16
勾配(%)	-0.08392	0.01842	-4.555	5.23E-06
標準偏差	1.30E-06			
Null deviance	9198.6 on 16221503			
Residual deviance	7610.4 on 16221496			
赤池情報量基準	97366			

図-5　負の二項回帰モデルのパラメータ推定結果（1 時間）

評価指標	1h	24h
絶対平均誤差	0.0006474	0.01559

図-6　負の二項回帰モデルの予測性能の評価結果

(4)　負の二項回帰モデルの事故発生確率の予測結果

負の二項回帰モデルにトレーニングデータを用いて，事故発生確率の予測を行った．その後，テストデータを用いてモデルの予測性能の評価を行った．

図-6 に，モデルの予測性能の評価結果を示す．予測性能の評価には平均絶対誤差を用いた．絶対平均誤差の値は，24 時間の事故データセットから推定されたパラメータの方が大きくなった．

(5)　考察

負の二項回帰モデルを用いたパラメータの推定結果と予測性能の評価結果より，モデルを評価する．

赤池情報量基準の値より，24 時間当たりの事故データセットを用いたモデルの方が，良いモデルだといえる．

しかし，デビアンス残差や平均絶対誤差の値が，1 時間当たりの事故データセットを用いたモデルより大きくなった理由として，データセットに含まれる事故件数の範囲が広くなったことが挙げられる．

1 時間当たりの事故データセットには，事故件数が 0 から 2件までの範囲が含まれており，24時間当たりの事故データセットには，事故件数が 0 件から 5 件までの範囲が含まれていたので，事故件数の範囲が広がったことによる影響が考えられる．

また，事故件数が 5 件のサンプルは 2 個しかないため，データセットをトレーニングデータとテストデータに分ける際に，トレーニングデータの方に 5 件のサンプルが二つとも入ってしまったことも考えられる．

これらのことから，予測する範囲が広がったこと，トレーニングデータとテストデータに偏りがあったことから，予測することが難しくなったと考える．

5. 結論

本研究では冬期の高速道路での 24 時間当たりの交通事故発生確率を予測するモデルを開発し，交通事故の発生に影響を与える要因についての分析を行った．その後，予測期間が 1 時間のモデルと，24 時間のモデルの予測結果の比較を行った．

24 時間当たりの事故データセットに変更したことで，赤池情報量基準の値は小さくなり，1 時間当たりの事故データセットで作成したモデルより，データに適合できており，良いモデルになったといえる．一方で，事故発生確率の予測では，予測する事故の件数の範囲が広がったため，複雑になったと考えられる．

今後の展望としては，予測期間の他に，予測を行う区間を，IC 間に変更し，さらに事故件数が 0 件のサンプル数を減らす，交通事故発生の有無をダミー変数と設定し，モデルの複雑さを解消していく，路面状態を考慮したモデルを作成する．

謝辞

本研究の遂行にあたり，東日本高速道路株式会社にデータの提供をはじめ，多大なるご協力をいただいた．ここに深く謝意を表する．

REFERENCES

1) Kazushi SANO，Touru INAGAKI，Jouji NAKANO，Nguyen Cao Y：An Analysis on Traffic Accidents on Undivided Expressway in Cold and Snow Area，Eastern Asia Society for Transportation Students，vol.8，2010.
2) Zhuanglin Ma，Honglu Zhang，Steven I-Jy Chien，Jin Wang，Chunjiao Dong： Predicting expressway crash frequency using a random effect negative binomial model: A case study in China，Accident Analysis and Prevention，vol.98，2017.
3) MANH Bui Tuen, PERERA KPD Frank, SANO Kazushi, TKAKURA Takumi, KATO teppei：高速道路で積雪条件下の交通事故傾向を予測する確率モデルの開発，土木計画学研究・講演集，vol.66，2022.

Traffic Accidents on Highways in Niigata Prefecture in Winter

Kazushi WATANABE，Kazushi SANO，Teppei KATO，Tetsuro HARAYAMA，Yoko MATSUDA

The purpose of this study is to analyze the factors that cause traffic accidents on expressways under snowfall conditions, and to develop a model to predict the probability of traffic accidents on expressways in winter, in order to predict the risk of traffic accidents during snowfall in advance and to help reduce the frequency of accidents. The model was developed to predict the probability of traffic accidents on expressways during winter. In this study, traffic accidents that occurred on five major roads covering the entire Niigata Prefecture were used.
In a previous study, a negative binomial regression model was developed to predict the one-hour probability of traffic accidents on expressways during snowfall.
However, the model does not take into account factors such as snow depth and road surface conditions, and the number of accidents, which is a rare event, is used as the objective variable.
To improve the accuracy of the model, we will solve these problems.