

令和 4 年 6 月 21 日現在

機関番号：16301

研究種目：若手研究

研究期間：2020～2021

課題番号：20K14851

研究課題名（和文）ディープラーニングによる交通事故リスク予測モデルの開発

研究課題名（英文）Deep learning model for accident risk prediction

研究代表者

坪田 隆宏 (Tsubota, Takahiro)

愛媛大学・理工学研究科（工学系）・講師

研究者番号：00780066

交付決定額（研究期間全体）：（直接経費） 3,100,000円

研究成果の概要（和文）：本研究は、ディープラーニングモデルを活用した交通事故リスク予測モデルの構築を行い、同モデルの予測精度向上に資する入力データを明らかにすることを旨とするものであり、都市間高速道路を対象とする近未来の交通事故リスク予測モデルを構築した。モデルは畳込みニューラルネットワークを基本とし、入力には車両感知器の速度・交通量・時間占有率（OCC）を使用した。また、入力データの質による影響を検討するために、交通状況の時間変動成分を考慮した分析を実施した。その結果、予測精度向上には、交通量とOCCの時間変動成分が有効であるとの知見を得た。また、構築したモデルの出力は実際の事故発生確率と概ね一致することを確認した。

研究成果の学術的意義や社会的意義

交通事故は物的・人的に多大な損失をもたらす為、事故の起こりやすさ、すなわち事故リスクを予測し、ドライバーへの情報提供や交通対策により事故の発生を未然に防ぐことが重要となる。本研究で構築した事故リスク予測モデルは近未来の事故リスクを高精度に予測する為、事故リスク低減を目指した交通管制に活用可能であると期待される。また、深層学習モデルの精度向上において、交通工学分野の知見が有用であることを示した点においても、同分野におけるAI技術の活用において重要な示唆を与えたといえる。

研究成果の概要（英文）：This study aims to build a traffic accident risk prediction model using a deep learning model and to identify input data that can contribute to improving the prediction accuracy of the model. A near-future traffic accident risk prediction model was built for an intercity expressway. The model is based on a convolutional neural network, and speed, traffic volume, and time occupancy (OCC) of vehicle detectors were used as inputs. In order to examine the impact of the quality of the input data, an analysis was conducted considering the time-varying components of the traffic conditions. The results showed that the time-varying components of traffic volume and OCC were effective in improving forecast accuracy. It was also confirmed that the output of the constructed model was generally consistent with the actual accident probability.

研究分野：交通工学

キーワード：Traffic accident Traffic engineering Deep learning 畳込みニューラルネットワーク

1. 研究開始当初の背景

社会的・経済的に深刻な影響をもたらす交通事故の削減に資する効果的な対策立案に向けて交通事故の発生要因分析が実施されてきており、その中では、道路線形や土地利用状況、天候等を含めた走行環境変数と事故多発地点との関係を示す集計分析や、それらの要因を説明変数に用いた線形回帰モデルによる多変量解析が従来より行われてきた。そして、これらの分析結果に基づく事故対策として、事故多発地点を対象としたハード対策が実施されてきた。このような従来型の事故多発地点を対象とした対策、すなわち交通量が多く交通事故の発生が顕在化した地点/状況への対策による効果は、近年の交通事故件数の減少傾向に表れている。それに加えて、今後の更なる事故削減に向けては、交通量が少なく事故発生件数は少ないが、潜在的に事故が発生する可能性が高まる地点/状況に対する対策を考案し実施していくことが有効であると考えられる。そのためには、事故発生件数が少なく、事故が顕在化していない箇所についても潜在的な事故発生の可能性を評価することが必要となる。そこで、交通事故データ、道路構造データ、交通流観測データならびに気象データを融合し、稀少事象に適應するポアソン分布による一般化線形モデルを用いた多変量解析を活用した、特定の道路線形、交通流状態さらには天候といった交通状況や交通環境に依存して変化する動的な交通事故発生リスク推定モデルが提案されてきた。これらの試みでは、事故発生リスクは、道路線形、時空間上のマクロな交通流状態およびミクロな交通流状態、天候や路面の状態といった様々な要因による影響を受けるとの仮定のもと、基本的には要因の線形和によって事故リスクを表現してきた。あるいは、要因間の交互作用を考慮する場合であっても、その組み合わせが膨大であることから、交互作用項の導入については限定的であった。しかしながら、事故リスクは複数の要因が複雑に作用しあって変動すると考えられるため、線形モデルで事故リスクを回帰する手法には限界がある。そこで、非線形回帰モデルの手法として深層学習モデルによる事故発生リスクの評価への期待が寄せられている。

一般に、多数のパラメータを有するニューラルネットワークは、モデルの空間移転性を確保することができないものの、特定の道路区間を対象として十分な量のデータ収集が可能な場合には、回帰モデルなどの従来の予測モデルとの比較において、高い精度で事故発生を予測することが可能となると期待される。そこで、多様かつ大量のデータを獲得することが可能となりつつ現状を踏まえると、ニューラルネットワークを用いた予測モデルが、従来の予測モデルに代わって実務における交通事故発生予測の役割を担うことが想定される。ここで、高い精度を有するニューラルネットワークモデルを構築するためには、ネットワーク構造やハイパーパラメータの異なる複数のモデルを構築し、各モデルの予測精度比較を行ったうえで、入力変数の質と入力データ量を考慮したモデルの予測精度の向上が求められる。

2. 研究の目的

本研究では、長期間蓄積した多様なビッグデータを活用し、交通事故リスクを動的に評価可能なディープラーニングモデルを構築する。モデルの構築に際しては、画像認識分野で培われた畳み込み演算をはじめとする人工知能技術に交通工学の知見を組み合わせることで交通事故リスクをリアルタイム、かつ高精度に推定可能なシステムの構築を目指し、本問題に適した入力データの選択方法を提案し、最適なディープラーニングモデルの構造を探索的に発見することを目的とする。具体的には、車両感知器データを入力に用いて、近未来における事故発生確率を予測する Convolutional Neural Network (CNN) モデルを構築し、モデルへの入力変数の質がモデルの説明力に与える影響を分析する。

3. 研究の方法

(1) 交通事故発生リスク予測 AI モデルの基本的な考え方

本研究では、予測開始時点直前の 1 時間の交通状況の推移、すなわち時刻 $t - 11$ ~ 時刻 t における検知器地点毎の交通流状態観測値 (速度、交通量、OCC) を入力し、予測開始時点から 2 時間先までにおける事故発生状況、すなわち時刻 $t + 1$ ~ 時刻 $t + 12$ における当該区間での事故発生の有無、および事故発生確率を予測する。

同モデルの構築には時間的・空間的に広がりを持った交通状況を入力する必要がある。また、交通状況は前後の時間・空間と相互に影響しあって推移するため、感知器間、および時刻間の関係を考慮して学習することが望ましい。そこで、本研究では 5 分ごとに得られる感知器データ：速度、交通量、OCC のそれぞれを感知器数 \times 時間帯の行列として入力する。各行は、予測時点直前 1 時間分の 5 分データを時間の順に並べたものであり、各列は、各感知器区間における観測データを上流から下流に向けて感知器が位置する順に並べたものである。このような行列状に配置されたデータの形式は画像データと類似していることから、画像の特徴抽出に広く活用される深層学習手法の一種である CNN モデルを使用する。

入力データの作成においては、本研究では感知器データから得られる速度、交通量、OCC の 5 分間集計値を行列形式に配置し、3 種類の交通データを重ね合わせるにより入力データを作成する。一方の事故データについては、予測開始時点から 2 時間先までの各 5 分間での事故発生

の有無を，事故有り（1），事故無し（0）として時間の順に並べた列ベクトルとして入力する．学習サンプルは予測開始時点の直前 1 時間の交通データに対して，2 時間先までの事故発生の有無をラベルとして付与することで 30 分毎に作成する．なお，何らかのエラーコードを含むデータや，集中工事期間のデータについてはサンプルから除外する．

（2）AI モデルの構築

事故リスク予測モデルには CNN モデルを用い，予測時点の 1 時間前から 5 分前までの車両感知器情報を入力して，予測時点から 120 分先までの事故リスクを [0.0, 1.0] の実数で出力するモデルを構築する．なお，最適化手法には Adam，活性化関数には ReLU 関数を用いるが，全結合層の活性化関数にはシグモイド関数を用いる．

既往の研究により，渋滞領域や臨界領域など，交通状況が時間的に大きく変動するような状況において事故リスクが高まることが指摘されている．そこで，CNN の入力に，交通状況の時間変動成分を明示的に取り込む場合の予測精度への影響を明らかにする．具体的には，入力層について，交通量，OCC の 3 チャンネルからなる標準入力に対して，交通状況の時間変動成分を追加することによる予測精度の変化を確認する．交通状況の時間変動成分は式 (1)-(3) によって計算したうえで感知器数 \times 時間帯の行列形式に整理した上で，標準入力に新たなチャンネルとして追加する．感知器区間 k ，時刻 t における速度，交通量，OCC の時間変動は，それぞれ以下のように表される．

$$dv_{k,t} = v_{k,t} - v_{k,t-1} \quad (1)$$

$$dq_{k,t} = q_{k,t} - q_{k,t-1} \quad (2)$$

$$do_{k,t} = o_{k,t} - o_{k,t-1} \quad (3)$$

ただし，

$v_{k,t}$, $q_{k,t}$, $o_{k,t}$: 感知器区間 k ，時刻 t における速度，交通量，OCC

$dv_{k,t}$, $dq_{k,t}$, $do_{k,t}$: 感知器区間 k ，時刻 t における速度，交通量，OCC の時間変動成分である．

（3）評価指標

本研究で構築するモデルは，「事故有り」「事故無し」のいずれかにて予測結果を出力する方法と，事故が発生する確率を数値で出力する方法を考え，以下の 2 指標を用いて予測結果の評価検証を行う．

事故発生の有無について，直接的に「事故有り」「事故無し」のいずれかにて予測を出力するモデルを構築した際の予測結果評価指標として ROC の AUC 値を用いる．事故の有無で 2 値のいずれかを出力するモデルでは，「事故有り」と予測した際に「実際に事故が起きた」True Positive (TP)，「実際には事故が起きなかった」False Positive (FP) のいずれかの状況が出現する．また，「事故無し」と予測した際も同様に「実際に事故が起きた」False Negative (FN)，「実際には事故が起きなかった」True Negative (TN) のいずれかの状況，計 4 つの状況が出現する．このとき，モデルの予測精度は的中率として算出評価することが可能である．もちろん予測的中の TP と TN の割合が高くなれば予測精度が高いことを示す．ここで，本研究のように稀少事象を対象として取り扱う場合には，例えばすべての出力を「事故無し」としておくことで高い中率が実現するとの状況が出現しうる．一方，本研究によるモデルは，計算によって [0, 1] の出力値 p ($0 \leq p \leq 1$) を出力するモデルとなっており，予め設定された閾値と p との大小関係から「事故有り」「事故無し」の判定を行うモデルであることから，閾値の値を変化させた場合に，上記 4 つの状況の出現割合が異なる結果を得ることができる．そこで，閾値の値を [0, 1] で変化させた場合のそれぞれで，False Positive Rate (以下では“FPR”) と True Positive Rate (以下では“TPR”) を算出する．前者の FPR は，事故が起きなかった場合に「事故無し」と予測していた割合，後者の TPR は，事故が起きた場合に「事故有り」と予測していた割合を示し，閾値の増加に伴って前者の値は単調に増加し，後者の値は単調に減少する．例えば，閾値を 0.0 とし，全ての予測結果を「事故有り」とした場合には，(FPR, TPR) = (0.0, 1.0) となり，逆に閾値を 1.0 とし全ての予測結果を「事故無し」とした場合には (FPR, TPR) = (1.0, 0.0) の結果が得られる．算出結果を FPR - TPR 平面にプロットし，閾値ごとに異なる各点を連続的につなぐ ROC 曲線 (Receiver Operating Characteristic curve) を用いて予測結果を評価する．すなわち，ROC 曲線右下部の面積である AUC 値を以て予測精度の評価指標とする．

「事故有り」「事故無し」，2 値での予測結果出力方法に代えて，事故が発生する確率を数値で出力するモデルを構築した際の予測結果評価手法には，事故発生確率の的中度合いを用いる．具体的には，[0.0, 1.0] の範囲で出力される事故発生確率を 0.1 刻みの階級に分けて予測値出現回数を集計し，予測値が出現した回数の中に占める実際に事故が発生した回数の割合 (以下“事故発生割合”) を算定する．同割合が各階級の階級値に近い値を示せば高い精度で事故発生確率を予測していることとなることから，事故発生割合と階級値との乖離を以て的中度合いの評価指標とする．

4. 研究成果

分析対象路線は、東名高速道路上り方向の御殿場 IC から東京 IC に至る範囲である。道路の総延長は 83.7km で、分析対象路線を 20km 前後の 4 つの分析区間：御殿場 IC 大井松田 IC、大井松田 IC 厚木 IC、厚木 IC 横浜町田 IC、横浜町田 IC 東京 IC に分割し、各区間に対応するモデルの構築を行う。

分析データとしては、各路線に概ね 2km 毎に設置された車両感知器の平均速度、交通量、時間占有率 (OCC) の各 5 分集計値を使用する。さらに、各区間上 5 分単位での事故発生の有無を示した施設接触を除く事故データを用いる。データの収集期間は 2009 年 1 月 1 日から 2018 年 12 月 31 日までの 10 年間分であり、このうち、全期間の 2018 月以外のデータを学習用データ、全期間の 2018 年データを検証用データとする。なお、データ収集期間の 10 年間にわたって、分析対象路線内における車線運用の変更や IC の新設等、道路構造や道路ネットワークの大幅な変更がなかったこと、また 10 年間を通して対象とする事故発件数に大幅な変化が認められないことを確認している。ただし、対象路線外においてはネットワーク新設が行われており、その影響によって対象路線内の交通需要や交通状況に経年的な変化が生じている可能性は否定できないことに留意が必要である。

標準入力、すなわち速度、交通量、OCC から構成される 3 チャネルの入力データを用いた場合の、各分析区間における予測再現性について説明する (表 1)。検証用データを対象に得られた ROC-AUC の値は 4 区間全てにおいて 0.78 を超えており、概ね良好な予測結果が得られたと言える。次に、区間による予測再現性の差異に着目すると、区間によって予測再現性が異なるとの結果が得られた。その一因として、学習サンプルに含まれる陽性サンプル、すなわち事故有りのサンプル数が区間によって異なることが影響していると考えられる。すなわち、学習サンプル数と比較して陽性サンプル数を十分に確保できない場合には、CNN モデルが陽性、すなわち事故有りに至る直前の交通状況を十分に学習できないものと推察される。実際、陽性サンプル数が 4 区間で最も多い区間において最も高い予測再現性を確認できた。

表 1 標準入力を用いた場合の ROC-AUC とサンプル数

	区間1	区間2	区間3	区間4
ROC-AUC(検証)	0.786	0.835	0.791	0.804
学習サンプル	141,003			
陽性サンプル(学習)	3,238	4,740	3,842	2,376
検証サンプル	16,592			
陽性サンプル(検証)	464	529	424	375

続いて、標準入力に式(1)-(3)により算出される交通状況の時間変動成分 dv, dq, do を追加した場合の予測再現性の変化を確認する (図 1)。標準入力を用いた“基準モデル”による ROC-AUC 値に対して、各変動成分を追加した場合の ROC-AUC の変化率を確認した結果、時間変動成分の追加による再現性への影響は、分析区間によって大きく異なる結果となった。再現性の変化傾向が区間によって異なる一因としては、道路線形等の要因を含む詳細な事故発生状況は分析区間によって異なる為であると推察される。

このように、分析区間に共通した特徴の把握は困難ではあったものの、交通量の時間変動成分 dq と、OCC の時間変動成分 do の 2 種類を追加したケース ($dq + do$) においては、4 区間で再現性の向上が認められ、残る 1 つの区間においても基準モデルと概ね等しい再現性が得られる結果となった。すなわち、事故発生に至る直前の交通状況の特徴を抽出する上で、標準入力データに加えて、交通量と OCC の時間変化を明示的に考慮することが有効である可能性を示している。

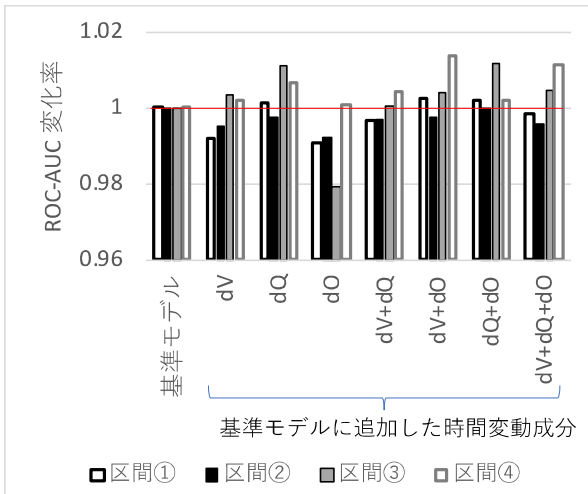


図1 交通状況の時間変動成分追加による
予測再現性の変化

最後に 2 値分類の出力に代えて、事故が発生する確率を数値で出力する運用を想定する際の予測結果を、事故発生割合との比較により評価した。事故発生割合による評価結果からは、いずれの区間においてもモデルが出力値が実際の事故発生割合と概ね一致するとの結果を得ていることから、モデルの出力値は実際の事故発生確率に近い値を示していると言える。

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件 / うち国際共著 0件 / うちオープンアクセス 0件）

1. 著者名 Takahiro TSUBOTA, Toshio YOSHII, Jian XING	4. 巻 1
2. 論文標題 PREDICTION OF TRAFFIC ACCIDENT LIKELIHOOD ON INTERCITY EXPRESSWAY BY CONVOLUTIONAL NEURAL NETWORK	5. 発行年 2020年
3. 雑誌名 AI・データサイエンス論文集	6. 最初と最後の頁 11-17
掲載論文のDOI（デジタルオブジェクト識別子） 10.11532/jsceiii.1.1_11	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計6件（うち招待講演 0件 / うち国際学会 3件）

1. 発表者名 Celso Luis FERNANDO, 吉井稔雄, 坪田隆宏, 白柳洋俊
2. 発表標題 交互作用項の多重共線性が ANN モデルのパフォーマンスに及ぼす影響
3. 学会等名 第18回ITSシンポジウム2020
4. 発表年 2020年

1. 発表者名 田中淳, 吉井稔雄, 坪田隆宏, 田畑大, 川松祐太, Chhatkuli Subas, 城所貴之
2. 発表標題 首都高速道路におけるAIを用いたオンライン事故リスク算定モデルの構築と活用可能性の検証
3. 学会等名 第18回ITSシンポジウム2020
4. 発表年 2020年

1. 発表者名 坪田 隆宏, 吉井 稔雄, XING Jian
2. 発表標題 CNNを用いた都市間高速道路の交通事故リスク予測モデル
3. 学会等名 第40回交通工学研究発表会
4. 発表年 2020年

1. 発表者名 Celso Luis FERNANDO, Toshio YOSHII, Takahiro TSUBOTA, Hirotooshi SHIRAYANAGI
2. 発表標題 Effect of the Multicollinearity of Interaction Terms on the Performance of the ANN Model
3. 学会等名 Proceedings of the Eastern Asia Society for Transportation Studies (国際学会)
4. 発表年 2021年

1. 発表者名 Celso Luis Fernando, Toshio Yoshii, Takahiro Tsubota, Hirotooshi Shirayanagi
2. 発表標題 A Factor Extraction Method using Deep Learning Technique on Traffic Accident Risk
3. 学会等名 2021 International Symposium on Transportation Data & Modelling (ISTDM 2021) (国際学会)
4. 発表年 2021年

1. 発表者名 Takahiro Tsubota, Mamoru Shimmizu, Toshio Yoshii, Hirotooshi Shirayanagi
2. 発表標題 Deep Learning Model for Predicting Traffic Accident Risk on an Expressway
3. 学会等名 2021 International Symposium on Transportation Data & Modelling (ISTDM 2021) (国際学会)
4. 発表年 2021年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------