

類似時期データの活用による季節変動を考慮した交通事故予測精度の向上

Improving Traffic Accident Prediction Accuracy by Utilizing Temporally Similar Data under Seasonal Variation

井土 雅耀[†]

Masateru Izuchi

李 嘉誠[†]

Jiacheng Li

能登 正人[†]

Masato Noto

1. はじめに

交通事故の予測とその未然防止は、安全で持続可能な社会を実現するうえで極めて重要な課題である。近年では、電動キックボードやシェアサイクルの急速な普及により、新たな形態の交通事故が増加傾向にある。こうした背景を受け、従来の事後対応型の安全施策に加え、データ分析や機械学習を活用した予測的アプローチへの転換が注目されている。従来の交通事故予測研究では、過去の事故統計や交通量、道路構造、天候などの変数を用いた回帰モデルが多く用いられてきた。これらの研究は一定の成果を挙げているものの、「地域依存性」「季節変動性」「外部要因の不確実性」などが予測精度向上の妨げとなっている。特に、12月や1月といった降雪期や日照時間が短くなる時期には、事故件数の急増や傾向の乱れが見られ、従来モデルでは予測精度が大きく低下する傾向がある。

本研究では、これまでの交通事故予測研究で見落とされがちであった「地域間における事故傾向の類似性」および「季節ごとのパターン変動」に着目し、これらの要素を予測モデルに組み込むことで、より汎化性能が高く、実用性に優れた二輪車事故件数予測システムの構築を目指した。

2. 先行研究

ビッグデータを取り扱うことで、高い予測精度を発揮することから近年では交通事故予測に、機械学習や深層学習などのAIモデルを活用する例が増えている。先行研究では、Changらが、不特定多数の人の位置情報（GPS）をもとに、交通事故件数を予測する深層学習モデルを開発し61%の精度を達成した。リアルタイムなGPS位置情報を用いていることから、よりリアルタイムに近い予測結果となっている[1]。Yuanらは、降水量や交通量データを用い、過去7日間の交通事故件数を元に、次の7日間の交通事故件数を予測する深層学習モデルを開発した。特徴として、時間解像度が1時間と高い点が挙げられる[2]。鳥居らは、予測に8種類のデータを用いて、多様な交通事故要因をカバーしようとする特色がある。イベントデータを交通事故要因として、初めてAIモデルに取り入れている[3]。

3. 本研究のアプローチ

本研究では、検討対象とする機械学習モデルとして、回帰型モデルを採用する。モデルの構築にあたっては、まず学習の効率性と精度の向上を目的として、ハイパーパラメータのチューニングを行う。さらに、モデルの学習を適切に導くための最適化関数の選定と、予測精度を客観的に評価するための損失関数についても説明する。その上で、最も高い精度を示した予測モデルを用い、最終的には特定地域（例：横浜駅周辺）における二輪車事故リスクの可視化や、事故防止施策の立案支援に資するシステム構築の手法を示す。

4. 実験

本研究では、神奈川県横浜市西区・横浜駅を中心とする半径約10kmの一般道区間をケーススタディとし、二輪車事故件数の予測精度向上を目的に、地域間事故傾向の類似性と季節的パターン変動を取り入れた機械学習モデルを構築した。教師データには、警察庁公開の交通人身事故データ（2019–2023年）を用い、信号や交差点が存在せず事故の性質が大きく異なる高速道路上の事故は除外している。本節では、使用データの詳細、モデル構築の過程、および評価指標について述べる。

4.1 使用データと前処理

実験には、警察庁が公開する過去5年間の交通事故統計データを使用した。事故発生時刻、事故類型、道路環境、気象条件、交通規制情報などの多様な変数を取り込み、1kmメッシュ単位で集計した。また、予測対象月と類似する気候傾向を持つ他地域のデータを「類似時期データ」として抽出し、学習データに追加した。本研究では、予測対象となる月の事故件数に焦点を当て、その前後にあたる月のデータのみを学習・予測に用いている。これは、冬季に特有の気象条件（降雪や路面凍結、日照時間の減少など）が事故発生傾向に強く影響を及ぼすことから、この時期の事故件数予測は特に困難であり、従来のモデルでは精度が大きく低下する傾向があるためである。

データ前処理では、欠損値の補完、カテゴリ変数のエンコーディング、スケーリングなどを行い、特徴量として約20項目を最終的に選定した。

[†] 神奈川大学, Kanagawa University

4.2 機械学習モデルの構築

使用モデルの概要を表1に示す。これらの機械学習モデルを使用した理由としては、特徴量が多い本研究のデータに対し、分類ではなく事故件数という連続値を高精度で予測する必要があるためである。

使用モデル	機械学習の種類
XGBoost	回帰型
LightGBM	回帰型
CatBoost	回帰型
MLP	ニューラルネットワーク型

4.3 評価方法

モデルの性能評価にあたっては、予測誤差の大きさを定量的に測定するための指標として、平均絶対誤差(MAE)および二乗平均平方根誤差(RMSE)を用いた。これらは、いずれも実測値と予測値との差を評価する指標であるが、RMSEは誤差が大きい予測に対してより敏感であるため、外れ値の影響も考慮した比較を可能にする。

5. 結果・考察

本研究では、横浜駅を中心とする半径10km圏内の2019–2022年11・12・1月データを学習用、2023年12月データをテスト用として、4手法による二輪車事故件数予測を行った。評価指標としてMAEとRMSEを用いた結果を表2に示す。

使用モデル	MAE	RMSE
XGBoost	2.338	3.134
LightGBM	2.378	3.156
CatBoost	2.245	2.988
MLP	2.668	3.581

CatBoostはカテゴリ変数の順序統計エンコーディングを内部で自動処理できるため、高基数のメッシュコードを前処理なしで利用できた。この結果、4手法中でも低い誤差を示した。学習データが比較的少量でも過学習を抑制しつつ空間的局所パターンを捉えられた点が優位性の要因と考えられる。

表3は、季節パターン適合効果を検証するために学習期間を変更して得た結果である。11・12・1月に限定した冬季モデルは、通年モデルと比べてCatBoostで5.7%、XGBoostで4.8%のMAE低減を達成した。サンプル数は通年の約25%に減少したにもかかわらず、MAEは約5%低下していることがわかる。季節パターンを取り入れることが冬季事故予測の精度向上に大きく寄与することが示された。

表3: 季節フィルタリングの効果検証

使用モデル	学習データ	MAE
CatBoost	11・12・1月	2.245
CatBoost	1–12月(通年)	2.380
XGBoost	11・12・1月	2.338
XGBoost	1–12月(通年)	2.456

表4は、横浜駅10km圏内のCatBoostを用いた学習(ベースライン)に対して無作為に選んだ外部メッシュを30%追加した場合と類似地点データを30%追加した場合を比較した結果である。類似地点データを追加した場合には、MAEが2.245となりベースラインと比較すると−0.069(約3%低減)を達成した。以上より、予測精度向上の主因は、事故傾向が統計的に似通った地域データを追加したことにあると判断できる。

表4: 類似地点データの効果検証

比較条件	学習データ	MAE
ベースライン	横浜駅10km圏内	2.312
無作為追加	+無作為メッシュ	2.304
類似地点追加	+類似地点メッシュ	2.245

6. おわりに

本研究では、横浜駅を中心とする半径10km圏内を対象に、1kmメッシュ単位で冬季の交通事故件数を予測し、季節パターンの適合効果と地域間における事故傾向の類似性の有効性を検証した。季節フィルタリングでは、通年モデルに対して、CatBoostで5.7%、XGBoostで4.8%のMAEの低減を確認した。さらに類似地点メッシュを追加した場合は、MAEを約3%改善し、傾向の近さが予測精度向上の主要因であることを明らかにした。今後は、多変量特徴量を拡充することでモデルの説明力と汎化性能を高める。

参考文献

- [1] Chang, L.-Y.: Analysis of Freeway Accident Frequencies: Negative Binomial Regression Versus Artificial Neural Network, *Safety Science*, Vol. 43, No. 8, pp. 541–557 (2005).
- [2] Yuan, Z., Zhou, X. and Yang, T.: Hetero-ConvLSTM: A Deep Learning Approach to Traffic Accident Prediction on Heterogeneous Spatio-Temporal Data, *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 984–992 (2018).
- [3] 鳥居和史, 水野由裕, 遠山一則, 清水茂樹, 向後颯太: 構造化データと衛星画像のマルチモーダルAIモデルによる交通事故発生確率の予測手法の開発, 自動車技術会論文集, Vol. 53, No. 2, pp. 404–409 (2022).