Traffic Black Spots Identification Based on Machine Learning

-- Taking Camden in London as An Example

Abstract

Road traffic accidents have long-term been the main safety threat to urban residents.

The study outlines an attempt to train random forest classification models to identify the potential traffic black spots based on the junctions built environment features, including street view information, POI data and street network configurations.

Around the accident risk of junctions, the trained models present a relatively high overall accuracy of 72%. However, due to the uneven distribution of accident data, the accuracy of models in classifying high-risk junctions can be insufficient.

Keywords: Traffic Black Spot, Road Junction, Random Forest Classifier

# Introduction

Road traffic accidents have long-term been the main safety threat to urban residents. According to the statistics of the World Health Organization (WHO), an estimate of 1.35 million people dies in road traffic accidents every year and in high-income countries the average death rate is 8.3 per 100,000 population(World Health Organization, 2018). Due to the huge hazard of traffic accidents, the prediction and prevention of accident risks are of great significance.

The study outlines an attempt to train random forest classification models to identify the potential traffic black spots based on the junctions built environment features, including street view information, POI data and street network configurations.

Around the accident risk of junctions, the trained models present a relatively high overall accuracy of 72%. However, due to the uneven distribution of accident data, the accuracy of models in classifying high-risk junctions can be insufficient.

# Literature Review

## Definition and Identification of Traffic Black Spot

Generally, the traffic black spots refer to concentration areas of traffic accidents on the road (Dereli & Erdogan, 2017). However, the specific criteria and identification methods of traffic black spots can vary greatly, according to the differences in urban form and social characteristics. Objects considered as traffic black spots can be road intersections, short road sections or mid-mid-blocks (Kowtanapanich, Tanaboriboon, & Chadbunchachai, 2006). Common defining elements of traffic black spots may include the frequency of accidents in a specific spatial and temporal range, the severity of accidents, and the proportion of accident numbers to the traffic volume (Aziz & Ram, 2022). For example, China defines urban road black spots as: road sections within 500 meters or intersections within 150 meters that have at least three casualty accidents every year (Yuan, Zeng & Shi, 2020); In Thailand, the black spot is the place where at least 3 serious accidents or 5 injury accidents have been recorded within 100m within three years (Tanprasert, T. et al., 2020). Considering the complexity of urban traffic environment and the potential requirements to applying multiple defining elements, some studies do not adopt fixed threshold standards, but identify the clustering structure of collision accidents as traffic black spots. Modified clustering algorithm of K-Means (Zhang, Shu & Yan, 2019), Firefly (Yuan, Zeng & Shi, 2020) and DBSCAN(Ye et al., 2010) have been proved to improve the accuracy of the identification of traffic black spots. Besides, some emerging urban data, such as social media data, have also been found effective in identifying the traffic black spots (Sinnott & Yin, 2015).

## Built Environment and Traffic Black Spots

Numerous studies have explored and examined the built-environment triggers for traffic accidents. Most studies focus on road crashes dominated by vehicle traffic, and the types of accidents include vehicle-to-vehicle (Asadi et al., 2022), vehicle-to-pedestrian (Ukkusuri et al., 2012), and vehicle-to-bicycle (Osama & Sayed, 2017), and small geographic areas such as census tracts or travel analysis areas are common study units. The corresponding built environment variables may include population and employment density, land use and diversity, street network configuration (connectivity, road grade, intersection density), characteristics of transportation and public services, and the socioeconomic status of communities (Fischer et al., Sternfeld & Melnick, 2013; Chen & Zhou, 2016; Osama & Sayed, 2017; Yang & Loo, 2016; Marshall & Garrick, 2011; Choi & Ewing, 2021).

Considering that the impact of the same built environment element on collision accidents may be ambiguous or contradictory in some studies, due to regional differences or incomplete selection of variables, some scholars have attempted to summarise the roles of different environmental characteristics in the accident process. Hakkert and Braimaster (2002) divided the occurrence of collision accidents into three parts: exposure, risk and injury probability. On this basis, Guerra and Dumbaugh (2020) argued that users' exposure and collision risks as primary mediums are affected by the built environment factors, thereby increasing or reducing the probability of traffic accidents. The variables associated with exposure may include traffic volume and the number of potential traffic participants (population and employment density); Variables associated with risk include vehicle speed and number of intersections. Land use, street configuration and facilities may have mixed or unconfirmed effects.

In the face of smaller-scale research scenarios, such as drivers, or safety from the perspective of cycling and walking, Street View imagery is gradually playing a more important role in identifying elements of the built environment.

## Street View

Tanprasert and his colleagues have explored the method to recognize and classify traffic black spots in Thailand based on street view images(Tanprasert et al., 2020). A distance-aware image segmentation model was developed to extract information of objects surrounding the road and a fully-connected neural networks was trained to identify the black points. The model has shown an accuracy of 69.91% in classifying black and safe spots in Thailand where recall of black points is 75.86%. The combination of street view information and the classification task of traffic black spots is proved to have sufficient potential.

# Research Question

The research focuses on the application of street view images in improving the traffic safety around urban road junctions, and the question explored can be illustrated as:

Q1. To what extent, the visual info extracted from street view images can reflect or indicate the built environment features and social context info around urban road junctions?

Q2. Can street view images independently be applied to identify whether a road junction is a potential traffic black spot?

# Methodology

## Research Path

The research path of the study can be presented as (Figure 1).

* Task1 is heavily relied on the the collision accident dataset, where accident records within 20m distance from the road junctions are selected as research objects. With severity class of accidents as dependent variable, and a series of potentially related environmental records as independent variables, a random forest model classifier is trained to predict the possible severity class of accidents. The independent variables include: ‘Year', 'Month', 'Day', 'Hour', 'Road Type', 'Speed Limit’, ‘Junction Detail', 'Junction Control', 'Pedestrian Crossing', 'Light Condition Band', 'Weather', 'Road Surface', 'Special Conditions', 'Carriage Way Hazard'
* Task2 has shift focus from accident points to the road junctions. For junctions having accidents records in 20m radius, environmental info is extracted from street view images corresponding to the junction locations, and the POI points within 50m radius from the junctions are collected to represent the activity condition around the junction. Similarly, a random forest classifier is trained based on the street view and POI info, to detect whether a junction could be a potential traffic black spot.
* The Side Task is an attempt to train and apply the image segmentation model with FastAI. Considering the workload for training and tuning the model,the analysis is listed as an independent task.

|  |
| --- |
|  |
| Figure 2 |

## Data Source

The research is mainly based on the ‘Road Collision Attendants In Camden’ dataset provided by Transport for London(TfL). The dataset contains 4555 road collisions in Camden during 2015-2019, recording the time, location, road type, speed limit, junction detail and other external environment details about the accidents and the severity class(fatal,serious,slight) and casualty number of accidents. Table 1 provide the definition of severity class by Department of Transportation in the UK. Besides, the research also apply the road central line data of from Ordnance Survey (OS), the Point of Interest(POI) data from OpenStreetMap(OSM) and the administrative boundary data of Camden town. An image segmentation model is trained based on the Cambridge-driving Labelled Video Database(CamVid)，and street view images of road junctions are collected from Google Street View Static API.

|  |  |
| --- | --- |
| Table 1 The Definition of Accident Severity | |
| Severity | Definition |
| Fatal Accident | An accident in which at least one person is killed; other casualties (if any) may have serious or slightly injuries. |
| Serious Accident | One in which at least one person is seriously injured but no person is killed. |
| Slight Accident | One in which at least one person is slightly injured but no person is killed or seriously injured. |

## Identification of Traffic Black Spots

### Method A - Frequency

An attempt has been made to extract the road junctions from the Camden road network. The analysis extract the coordinates of the first and the last nodes from the geometry of each road，and drop duplicated nodes. There are 2303 junctions nodes are extracted. Junctions with accidents in 20m radius are selected and the count of accidents on each junctions are summarized. There are 789 junctions with accidents in 20m distance.

### Method B - Clustering

## Street View Image Processing

POI data are relabelled in different categories and joined to the nearest junction(including all junctions and the junctions with accidents in 20m). Count of POI in each category and around each junction are summarized.

For each road junction, the corresponding street view images are split into 4 pieces and downloaded respectively.

## Random Forest Model

# Analysis and Results

## Exploratory Analysis

In this section, preliminary exploratory analysis is conducted to investigate the accident, junction, POI info and street view info datasets.

### Accidents

According to the accident records, for each year during 2015-2019 about 70% accidents happened near a road junciton. The significantly high proportion indicates a potential high risk of road junction as traffic black points. The section also presents the distribution of severity class of collision accidents. It is found that for each year most of the accidents belong to the 'slight' type and 'serious' type account to about 1/6 of the slight type. Few accidents can be labeled as 'fatal'.

The section then plots all the road junctions and highlight their difference in accident count. The accident count follows an exponential distribution, where most of the junctions have no accident. Besides, it is found that most of the accidents happen on the junctions near local commercial centres or high streets, which indicates a possible relation between commercial activities and accidents.

|  |  |
| --- | --- |
| 图表, 条形图  描述已自动生成 | 图表, 条形图  描述已自动生成 |
| Figure 2 | Figure 3 |

|  |
| --- |
| 散点图  描述已自动生成 |
| Figure 4 |

### Built Environment Features

Then plots the proportion of 4 types of street view information(sky, tree, building, tree) on the road junctions and count of 6 types of POI points(pub\_bar, food\_drink, shop\_mall, hotel, school, culture) near junctions, are plotted on the sub maps.

According to the street view plots, from southeast to northwest, there is an increase trend in the proportion of sky and tree elements and a decrease trend in the proportion of building elements. The change of road elements is not obvious.

|  |
| --- |
| 图形用户界面, 图表, 地图, 散点图  描述已自动生成 |
|  |

For POI points, POI points with labels of pub\_bar, food\_drink, shop\_mall and culture see a relatively even spatial distribution. While points with hotel and school labels mainly gathered in the southeast of Camden, which is also the centre of inner London.

|  |
| --- |
| 图表, 散点图  描述已自动生成 |
|  |

|  |
| --- |
|  |
|  |

## Traffic Black Spots Classification

In this section, two attempts have been made to train random forest classifiers to recognize whether a road junction could be a potential traffic black spots, based on the street view and POI information around the junctions.

### Model Training Based on Junctions with Accidents Around (To be replaced)

In the first attempt, the classifier is trained based on the accident count, street view and POI info of only juncitons with accidents around. The juncitons are reclassified as three types:'0' -- with only 1 accident, '1' -- with 2 accidents, '2'-- with 3 or more accidents. On this basis, 8 street view feature proportions, 6 POI counts and the longitude and latitude info are selected as the independent variables.

Based on the model training and tunning, the classifier gets a best accuracy score of 0.51. Though the accuracy score of the traffict black spots model is much lower that that in the severity prediction model, it shows a more balanced performance on the precision for classifying all the three types of different junctions. Besides, the model has a relatively high recall scores for juntions with 1 accident(recall = 0.77) and for junctions with 3 or more accidents( recall = 0.4).

In term of the feature importance, the road feature stablely ranks first as the most important feature. This indicate that the proportion of road feature in street view(related to the road width and junciton size) can contribute more to the correct recogniztation of potential traffic black spots. Besides, street view features such as buildings, sky and sidewalk, and POI points such as pubs and bars also play a relatively important role.

### Model Training Based on All Junctions

To increase the accuracy of model, another random forest classifier is trained based on all the road juncitons in Camden. The difference of the model lies on that juncitons are reclassified as :'0' -- with 0 accident, '1' -- with 1 or 2 accidents, '2'-- with 3 or more accidents. Junctions with no accidents are joined as a new class. The seleciton of independent variables keeps the same.

The new model has seen an signifcant increase of best accuracy score from 0.51 to 0.68, indicating that an increase of sample number in trainning dataset may help improve the model performance. Even though, due to the larger number of junctions with no accident, the model also show an unbalanced prediction performance like severity prediction model. The model has a relatively high precision, recall and f1-score on 0 accident junctions, while for the other junctions the accuracy of prediction and classification is low.

precision recall f1-score support

0 0.80 0.88 0.84 380

1 0.39 0.28 0.32 137

2 0.46 0.46 0.46 59

accuracy 0.70 576

macro avg 0.55 0.54 0.54 576

weighted avg 0.67 0.70 0.68 576

表格

中度可信度描述已自动生成

# Discussion and Conclusion

## Summary of Main Findings

According to the analysis, there are a series of findings to report:

The accident severity prediction model gets a relatively high accuracy of 0.87, but its accuracy is limited in classifying and predicting serious accidents. Possibly, there are two reasons to explain the situation: on the one hand, the proportion of slight accidents and serious and fatal accidents in the dataset is seriously unbalanced. The total number of fatal and serious accidents is far less than that of silight accidents. The features learned and acquired by the model from serious accidents can be insufficient; On the other hand, the original data set is mainly composed of categorical variables, which contain a considerable number of options representing special or rare options. These options are not representative in the overall accident records. However, for the integrity of the analysis, these options are still retained and entered into the model, which may lead to a low efficiency of the model in extracting accident characteristics.

There are two traffic black spots recognization models trained in this research, one on the road junctions with accidents(789) and the ohter on all the juncitons within Camden boundary(2302). The first model shows a more balanced performance in classifying all the different junciton types while the second model has higher accuracy but get poor performence in recognizing jucnitons with 2 or accidents. Road, building and sky features in street view and some POI types have been observed to play a positive role in correctly classifying the road juncitons.

## Performance of Envrionmental variables in Models

Considering that the two models trained in the study mainly take static environment and spatial factors as independent variables, the performance differences of models in different tasks is worth further discussion.

The exploration of accident severity prediction model shows that the prediction of serious accidents which have strong uncertainty and randomness, may rely on more continuous variables and larger sample number. Analysis in the notebook only includes part of the environment, road and intersection characteristics and time factors of accidents. While specific information about drivers, injured persons and vehicles involved in the accidents are not included. The missing factors may limit the further improvement of model accuracy.

In contrast, models for traffic black spots recognization reflect that environmental factors may play a more significant and positive role in predicting cumulative impact and change in urban space. In this study, the cumulative number of accidents at the road junctions is applied to reflect the possibility of junctions as a traffic black spot. And the road and building feature in street view -- the environmental perception characteristics that keep stable in a long time, togerther with the POI distribution -- a result of evolution of activities can also be regarded as the cumulative output of urban system. The similarity between the two may explain the more balanced performance of traffic black spots recognization model.

## Limitations and Possible Improvement

One of the main limitations in the research is that, significant data distribution difference could be ignored in variables included in two random forest classifiers. According to works by Alkheder, Taamneh and Taamneh (2017) and Iranitalab and Khattak(2017) k-means clustering methods could be a possible solution to reduce the variable difference and increase the model accuracy. However due to the workload limitation, the relavant analysis is not included in the notebook. The method can be tested and applied in the future exploration.

# Acknowledgements

# References

Dereli, M.A. and Erdogan, S. (2017) ‘A new model for determining the traffic accident black spots using GIS-aided spatial statistical methods’, *Transportation Research Part A: Policy and Practice*, 103, pp. 106–117. doi:10.1016/j.tra.2017.05.031.

World Health Organization. (2018). Global status report on road safety 2018: Summary. [online] p.2. Available at: <https://www.who.int/publications/i/item/WHO-NMH-NVI-18.20> [Accessed 25 April 2022].

Tanprasert, T. et al. (2020) ‘Recognizing Traffic Black Spots From Street View Images Using Environment-Aware Image Processing and Neural Network’, *IEEE Access*, 8, pp. 121469–121478. doi:10.1109/ACCESS.2020.3006493.