Realtime 事故評估系統

Highway

摘要

目錄

1. 研究主軸 ……………………………………………………01
2. 現況與問題盤點評估
   1. 國道現況……………………………………………… 01
   2. 問題盤點評估………………………………………… 01
3. 資料來源及其處理過程
   1. 資料來源……………………………………………… 02
   2. 處理過程……………………………………………… 02
4. 資料分析模式及方法……………………………………… 03
5. 研究流程、推論邏輯及理論依據
   1. 研究流程……………………………………………… 04
   2. 推論邏輯……………………………………………… 05
   3. 理論依據……………………………………………… 06
6. 數據之因果關係詮釋……………………………………… 07
7. 資料未來應用與推廣……………………………………… 08
8. 其它有助於文獻、案例
   1. 國外文獻……………………………………………… 09
   2. 關於高速公路風險評估……………………………… 11

壹、研究主軸

歷年交通事故頻傳，是交通壅塞原因之一，造成駕駛的生命、財產的損失，對於高速公路高速行駛，駕駛須對環境、路段保持專注，否則在高速公路交通事故的損害將較一般道路嚴重，本團隊著重在提供駕駛在行駛之餘的服務，參考國道的交通事故以及違規情形，進行資料探勘統計分析，從易肇事路段建立風險評估模型，協助駕駛提高警覺，避免交通事故之嚴重，期能減少國道交通事故發生，以增加用路人行車安全為主要目的。

貳、現況與問題盤點評估

1. 國道現況

根據交通部高速公路局資料統計，目前路線網路為國道1號、國道2號、國道3號、國道4號、國道5號、國道6號、國道8號、國道10號，總長度達1,049.7公里，其中國道1號和國道3號是西部南北重要交通運輸幹道，國道5號是臺灣東、西部重要聯絡公路，國道2、4、6、8及10號則為臺灣整體國道路線橫向重要幹道。

國道計費門架共計有335處，其中163處計費門架位於國道一號(15處於國道一號高架)、158處於國道三號(4處於國道三號甲)、14處於國道五號，電子計程收費總長度達到926公里，平均間距約為5.54公里，ETC電子收費平均使用率為92.75%(統計為107年度)。

以國道103年第一季

1. 問題盤點評估

內政部警政署國道公路警察局統計在國道事故資料，分為A1、A2、A3三大類，A1為交通事故係指人員當場或24小時內死亡之交通事故，A2類交通事故則為造成人員受傷或24小時後死亡之案件，A3類指車輛碰撞僅造成財損，據統計結果，106年共發生24,473件國道事故，其中A1類70件、A2類1,709件、A3類22,694件，相較105年A1類交通事故死亡人數增加50人(56.2%)、受傷人數增加15人(16.9%)，A2類交通事故增加25件(1.7%)、受傷人數增加142人(6%)，A3類交通事故增加1009件(5.4%)。

在違規取締總件數部分，106年件數為588,741件，較105年減少42,129件(6.7%)。

參、資料來源及其處理過程

一、資料來源

交通部高速公路局(以下簡稱：高公局)：

* 1. 103-107年交通資料集
  2. 103-107年違規資料集(僅取國道)
  3. 103-107年事故資料集(僅取國道)
  4. 國道計費門架座標及里程牌價表(政府開放平台)

二、處理流程

1. 問題動機發想-整合性探討

2. 資料前處理

（1.）資料搜集：資源來源

（2.）資料匯入--小樣本

（3.）資料特徵工程

（4.）資料欄位統計

（5.）資料清理

3. 探索式資料分析（EDA）

（1.）單變量分析

（2.）多變量分析

（3.）資料視覺化

4. 資料模型建立

（1.）參照關係

（2.）資料關聯性

5. 驗證式資料分析（CDA）

（1.）分類或分群

（2.）線性回歸

（3.）演算法理論實作

6. 資料報告建立

肆、資料分析模式及方法

本團隊在初步探勘方式以Weka開源軟體作為工具，並透過Python及R程式語言輔助實作資料分析，檢視各欄位值的分佈，刪除空值或資料值錯誤的欄位，並簡易繪圖形式理解該欄位相關係數及分佈情形。在選擇可用之欄位後，以Hadoop Ecosystem開源數據平台協作導入數據，分散式系統架構減少單一電腦運算效能及儲存空間，提升大量資料建立模型的效率，應用所建立之模型，依不同資料產生風險評估值。

以下圖4-1為導入Hadoop Ecosystem開源數據平台協作示意圖

一張含有 螢幕擷取畫面 的圖片

自動產生的描述圖4-1 Hadoop Ecosystem開源數據平台

於參點之處理流程所用到探索式資料分析以及驗證式資料分析，前者為在未熟悉資料前運用統計方法以及視覺化的呈現，整體性檢視資料屬性，並從中摘取部分可用之欄位，以利本團隊後續假設問題。後者為將前者的假設性問題透過大量資料進行驗證性判斷，決定後續是否採納前者的假設性問題，築起初步的應用方案概念。

伍、研究流程、推論邏輯及理論依據

一、研究流程

對於前述國道盤點問題進行探勘，由於提供資料為2014-2017年度，初步整體檢視資料後，篩選不必要欄位以利後續分析使用。以國道違規資料集、事故資料集為主，結合時段進行時間分析，依序由車種分析、道路型態分析（依國道分類）、車輛違規原因探勘、肇事事件探勘（可細分為：肇事原因、肇事者型態、肇事情況），以上步驟後，本團隊進行綜合性應用發想，考量「可行性方案」和「效益性評估」，最後提出最後應用方案。

以下圖5-1為資料研究流程示意圖，本團隊透過多層資料探勘，協助日後對創意應用發想實踐。

一張含有 螢幕擷取畫面 的圖片

自動產生的描述圖5-1 資料研究流程

二、推論邏輯

本圖隊初步整體檢視資料分布情形後，篩選不必要欄位以利後續分析使用。以國道違規資料集、事故資料集為主，理解各種類(A1、A2)件數趨勢情形，結合時段進行時間分析。接著本團隊思考有哪些因素會造成事故的發生，深入推論在何種車種方面肇事率較高，以及路段交叉分析，發現在不同型態的道路屬性有些許差異，由於用路人影響駕駛因素眾多，本團隊挑選幾項較為明顯的欄位，根據當時情境的天氣狀況，和車流量交叉分析，在降雨時因視線不良，行車速度較慢，肇事件數也相對較少，僅少數在壅塞路段事故發生，是否能提早透過推播形式告知用路人前方資訊將可作參考。

經過來回資料探勘結果顯示，車輛違規資料，其一項影響違規主要為駕駛人特性，可見多數為用路人可避免的，其一項少數為機車違規和逆向違規因素造成壅塞或需要派遣人員支援驅離動作。最後以探勘肇事事件進行，本團隊成員長期藉由媒體報導得知酒駕於高速公路的事件頻傳，藉此將酒精濃度列為優先因素，作是否影響國道其他車輛事故的假設，近一步結合時段、道路種別、天氣等等欄位。彙整國道事故資料有多半事件為人為造成，顯示該事故應可事前避免，本團隊由此發想出提供給用路人的「風險評估模型」，根據歷史資料的訓練，以及當前用路人行駛情形，推導出風險評估值和當前路段狀況，提高用路人行車警覺和降低事故發生，而歷史資料亦可去識別化方式供民眾查詢，盡量避免相關因素（例：時段、車流量）行駛，保障自身安全。

三、理論依據

(1.) 隨機森林模型(Random Forest Model)：

隨機森林是一個包含多個決策樹的分類工具，並且其輸出的類別是由個別樹輸出的類別的眾數而定。Leo Breiman和Adele Cutler發展出推論出隨機森林的演算法。

優點：對於很多種資料，此模型產生高準確度的分類器，它可以處理大量的輸入變數，並在決定類別時，評估變數的重要性。而對內部一般化後的誤差產生不偏差的估計。如果有很大一部分的資料遺失，仍可以維持準確度。於計算各例中的親近度，對於[資料探勘](https://zh.wikipedia.org/wiki/%E6%95%B0%E6%8D%AE%E6%8C%96%E6%8E%98" \o "資料探勘)、偵測[離群點](https://zh.wikipedia.org/w/index.php?title=%E9%9B%A2%E7%BE%A4%E9%BB%9E&action=edit&redlink=1" \o "離群點（頁面不存在）)和將資料視覺化非常有用。綜合上述。它可被延伸應用在未標記的資料上，這類資料通常是使用非監督式資料集。也可偵測偏離者和觀看資料。

(2.) 長短期記憶理論(LSTM)：

達到最小化訓練誤差，類似於梯度下降法（Gradient descent），可用來依據錯誤修改每次的權重。誤差梯度隨著事件間的時間長度成指數般的消失。當設置LSTM 區塊時，誤差也隨著倒回計算，從output影響回input階段的每一個gate，直到這個數值被過濾掉。因此正常的倒傳遞類神經是一個有效訓練LSTM區塊記住長時間數值的方法。

(3.) 支持向量機(SVM)：

由 Vapnik等根據統計學習理論提出的一種新的機器學習方法。在解決小樣本、非線性及高維模式識別問題中表現出許多特有的優勢。性能優於已有的學習方法，表現出良好的學習能力。從有限訓練樣本得到的決策規則對獨立的測試集仍能夠得到較小的誤差

陸、數據之因果關係詮釋

時間分析：7-10時和16-19時因車輛較多、車速較慢，A1類事故較少，卻容易忽略行車安全距離之重要，因此A2類件事較多，相對深夜清晨時段，A1事故件數較其他時段多，可能因車流量少，用路人容易不注意而有開快之情形，肇事件數也相對增加。

柒、資料未來應用與推廣

本團隊經過近ㄧ個月的資料初探後，認為有下列三點資料趨勢可作為未來應用方針。其一、道路型態分析，其二、違規車輛行為，其三、國道事故肇因分析，提供用路人更多即時資訊，降低肇事或連帶事故等問題，亦提升駕駛體驗感受，打造行車道路安全環境，

在未來應用層面，本團隊考量如取得三項資料集串檔資料，將能直接關聯三個資料集，因此使用以上三點之特徵欄位，參考業界風險評估標準，運用競賽資料和開放API，建立屬於國道的「風險評估模型」。下列描述為此評估模型應用方式：

針對歷史肇事和違規資料集之欄位，例如：道路特性、天氣、光線、違規條款、肇事原因等相關欄位，利用非監督式模型或其他可參考迴歸模型運算，將所建立的風險評估模型顯示至駕駛APP應用中，藉由用路人行車前輸入相關內容，例如：駕駛資格、車輛種類、目的地等等，配合手持裝置定位功能，判斷該車輛行駛於何處國道、路段，以及天氣、行駛時間、道路情況等，並由歷史資料所評估出該路段風險值，以提示功能通知駕駛該路段時間的風險值，警示駕駛人應保持適當距離，注意兩側來車，和減少行駛速度避免超速或肇事等情況發生。

對於管理者層面結合監視管理平台將即時數據回傳至Spark Streaming架構，能解析數據並以儀表板形式呈現於網頁平台，一旦有車輛事故或壅塞臨時狀況發生，立即顯示於網頁螢幕警示，指派相關人員於現場處理，大幅縮短應變時間，有效紓解尖峰時刻車流，除了達到「掌握性」、「預判力」，間接對事件後各單位探討，延伸出「可調查性」，以便建立事故處理派遣機制。

在此創意應用之推廣管道，本團隊希望延續「1968APP」進行功能銜接。截至2017年底，已累積210萬次下載量和1.15億次瀏覽量的創舉，搭配改善的使用者介面，相信在有效的推播，同時多元化的服務，將會吸引更可觀的用路人下載。亦可於國定假日車流量較大時刻，推動三方宣導計畫，其一、於平面文宣告知避免疲勞駕駛、請勿刻意超車等危險標語，其二、於影音廣告提醒用路人應於駕駛前使用風險評估功能，其三、製作社群網路懶人包以利告知民眾注意事項、離峰優惠等。

綜觀資料集的應用與推廣，本團隊從各角度進行分析，不論是特徵欄位的選擇，抑或應用方式的實踐，甚至協助高公局管理上的建議，本團隊針對目前國道事故與違規數量逐年遞增的問題著手，評估單位未來可執行的方案，提供使用者欲改善的回饋作為參考，共同建立通暢且安全的高速公路。

捌、其它有助於文獻、案例

一、國外文獻：

1. 哈佛大學對醫療做風險預測模型：http://ncook.bwh.harvard.edu/

2. 哥倫比亞大學對公衛領域所作風險預測模型：

https://www.mailman.columbia.edu/research/population-health-  
methods/risk-prediction

3. 模型論文：

(1.) Xia, E., Yu, Y., Xu, E., Mei, J., & Sun, W. (2019). From Risk Prediction Models to Risk Assessment Service: A Formulation of Development Paradigm. arXiv preprint arXiv:1903.07551.

(2.) Shigemizu, D., Akiyama, S., Asanomi, Y., Boroevich, K. A., Sharma, A., Tsunoda, T., ... & Sakurai, T. (2019). Risk prediction models for dementia constructed by supervised principal component analysis using miRNA expression data. Communications biology, 2(1), 77.

(3.) Steyerberg, E. W., van der Ploeg, T., & Van Calster, B. (2014). Risk prediction with machine learning and regression methods. Biometrical Journal, 56(4), 601-606.

二、關於高速公路風險評估：

1. 論文：

(1.) Hannan, A., Ahmed, A., Ashraf, T., & Bai, Q. (2016). Estimation of Highway Project Cost using Probabilistic Technique. DEStech Transactions on Engineering and Technology Research, (ictim).

2. 案例：

(1.) 美國交通部聯邦高速公路進行成本風險評估：

華盛頓州DOT公路建築成本指數、成本風險估算