

國道事故處理時間影響因素及建模預測

Influencing Factors and Prediction of Traffic Accident Duration on National Highways

卓明君¹、賈毓虎²、賴建宇³、陳勝文⁴

摘要

當國道發生事故時，若干車道旋即被占用，上游車輛須減速或變換車道因應，致使路段發生壅塞。倘若能蒐集歷史資料，運用數據分析、機器學習等技術建立模型，對即時的交通事故進行處理時間預測，並提供用路人更為準確的事故處理時間資訊，則有利於用路人做出因應，降低事故造成的壅塞影響，盡快恢復交通順暢。現階段國道發生事故時，事故班抵達後會記錄事故的狀況，包含涉及的肇事車輛、發生位置等資訊，待事故排除完成後將發生時間點、排除時間點及前述事故相關資料以簡訊方式回報給交控中心，本研究採用該事故簡訊資料作為後續分析。

承上，本研究使用 108-110 年的國道事故簡訊資料分析，觀察資料中各因素的敘述性統計，並以推論性統計之方法，如 Kolmogorov-Smirnov、變異數分析等進行事故處理時間的分布配適及影響因素的確認，其中發生時間(小時)、事故等級、國道、車種、車道以及事故型態為顯著。後續將影響因素透過獨熱編碼轉換成變數欄位，再以迴歸分析、K 近鄰、決策樹集成等不同模型進行訓練及交叉驗證，後用 Optuna 演算法校估超參數後得出最佳模型為決策樹集成方法中的 CatBoost，以 111 年 1-2 月份資料進行預測後得到平均誤差(MAE)約為 11.23 分鐘，中位數則約為 8.59 分鐘

關鍵字：事故處理時間分布、機器學習、事故影響因素、事故處理時間預測

一、緣 起

當國道發生事故時，若干車道旋即被占用，上游車輛須減速或變換車道因應，致使路段發生壅塞。倘若能蒐集歷史資料，運用數據分析、機器學習等技術建立模型，對即時的交通事故進行處理時間預測，並提供用路人更為準確的事故處理時間資訊，則有利於用路人做出因應，降低事故造成的壅塞影響，盡快恢復交通順暢。

交通部高速公路局設有北、中、南區及坪林四個交控中心，分別管轄不同的國

¹交通部高速公路局交通管理組組長。

²交通部高速公路局交通管理組副組長。

³交通部高速公路局交通管理組交通分析科科長。

⁴交通部高速公路局交通管理組交通分析科工程師，新北市泰山區黎明里半山雅 70 號，02-2909-6141#2347，shinefirem@freeway.gov.tw。

道範圍，本文使用的資料為交控中心回傳之事故簡訊資料，記錄了車道、車種、事故型態、發生時間與地點等各項事故資訊，得供管理單位進行研究。

綜上所述，若能根據回報狀況預測出事故處理時間，則能顯示在 1968 App、網站及資訊可變標誌等供用路人參考因應，本文從事故處理時間分布配適切入，以相應的推論統計方法進行檢定，從中找出影響事故處理時間的重要因素，據以建模預測。

二、文獻回顧

本文之目的為透過事故處理時間的分布配適搭配統計檢定找出對事故處理時間影響較大的因素，並據此建立模型及進行驗證比較，最後統計出真實資料集的預測績效。過往研究曾就事故處理時間定義、分布配適、重要因素及建模預測等各面向進行詳盡的文獻回顧(Li, Francisco, Pereira and Moshe, 2018)，文中亦整理了許多學者對於本文主題進行之研究得供參考。

2.1 事故處理時間分布

過往針對事故處理時間分布，部分研究透過加速失效模型相關分析認定為 Weibull 分布(Kaabi, Dissanayake and Bird, 2012;Hojati, Ferreira, Charles and Kabit, 2012)、Log-logistic 分布(Chung, Walubita and Choi, 2010)及 Log-normal 分布(Chung and Yoon, 2012)，此外部分研究則將事故處理時間進一步細分為回報時間、反應時間及清除時間，各自的分布則為 Weibull、Weibull 和 Log-logistic(Nam and Mannering, 2000)。

除加速失效模型外，透過其他方式分析檢定亦得到相似的結果，包含以統計檢定方式得出該分布為 Log-normal 分布(Giuliano, 1989)，以變異數分析檢定出事故分布為 Log-logistic 及 Log-normal 分布(Hojati, Ferreira, Charles and Kabit, 2012)，及運用迴歸分析推論出為 Log-logistic 及 Log-normal 分布(Zhang and Khattak, 2010)，綜上，多數研究認為事故處理時間呈現 Log-Logistic、Weibull 或 Log-normal 分布。

2.2 事故處理時間預測

過往業已有許多模型被應用在事故處理時間預測，包含線性迴歸(Peeta, Ramos and Gedela, 2000;Khattak, Wang and Zhang, 2012)、模糊系統(Dimitriou and Vlahogianni, 2015)、貝氏網路(Ozby and Noyan, 2006;Shen and Huang, 2011)以及生存分析(Qi and Teng, 2008;Chung, 2010;Kang and Fang, 2011)。

近年來由於人工智慧技術的演進，許多模型也被應用在此議題，包含決策樹(Knibbe, Alkim, Otten, and Aidoo, 2006;Chang and Chang, 2013)、深度學習(Wei and Lee, 2005)、支持向量機(Zong, Zhang, Xu, Zhu and Wang, 2013;Wu, Chen and Zheng, 2011)、K 近鄰(Kuang, Yan, Zhu, Tu and Fan, 2019;Wen, Chen, Xiong, Han and Chen, 2012)。

在事故處理時間預測的課題中，最常見的使用績效為 MAPE(Weng, Qiao, Qu and Yan, 2015;Khattak, Wang and Zhang, 2012)、R-squared(Garib, Radwan and Al Deek, 1997;Peeta, Ramos and Gedela, 2000)、RSME(Khattak, Liu, Wali, Li and Ng, 2016)及絕對誤差(Ji, Jiang, Qu and Chung, 2014)，本文參考過往文獻使用之模型與績效，搭配事故簡訊資料進行驗證比較。

三、事故處理時間影響因素

本文所指之事故處理時間係指事故發生至排除事故為止，單位皆為分鐘，為探討對於事故處理時間影響較大的因素，本文使用各交控與行控中心回報之事故簡訊與通報狀況資料，從 108 年至 110 年共 56,235 筆。

3.1 各因素處理時間統計

3.1.1 各年度處理時間

將事故處理時間依照年度及處理時間長度分組統計後如圖 1，各年度的處理時間皆為右偏分配，108 年處理時間在 10 分鐘以內的事務為 25%，110 年度則為 32%，處理時間逐漸降低，但整體而言無法直接觀察出不同的年度對於事故處理時間具有明顯的影響。

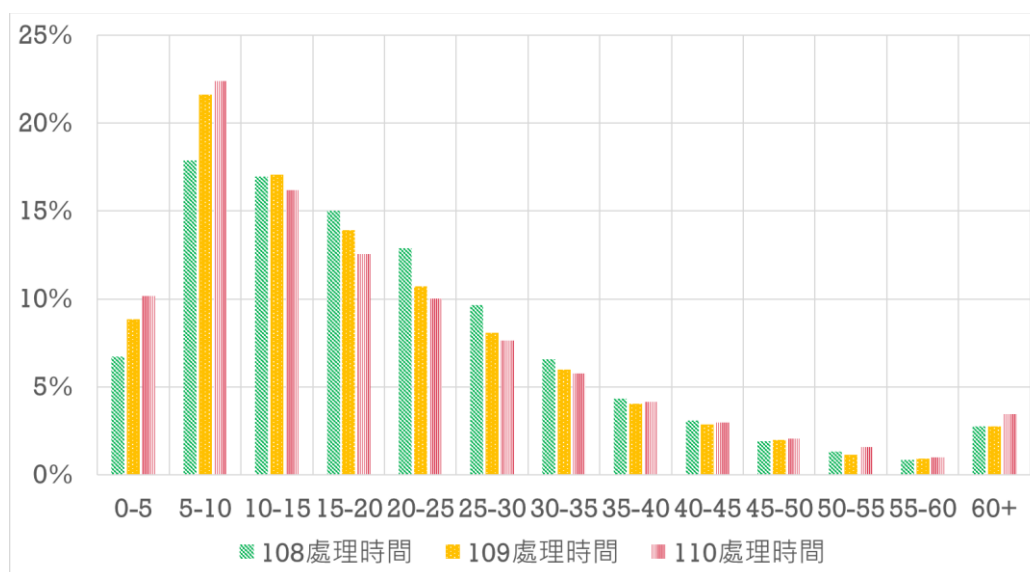


圖 1 各年度處理時間分組比例

3.1.2 分時事故統計

依照事故發生小時統計如圖 2、圖 3，圖中 0 至 6 時事故件數低但 A1 和 A2 的事故比例較高，約為 20-30%，使得統計後處理時間的中位數較長。

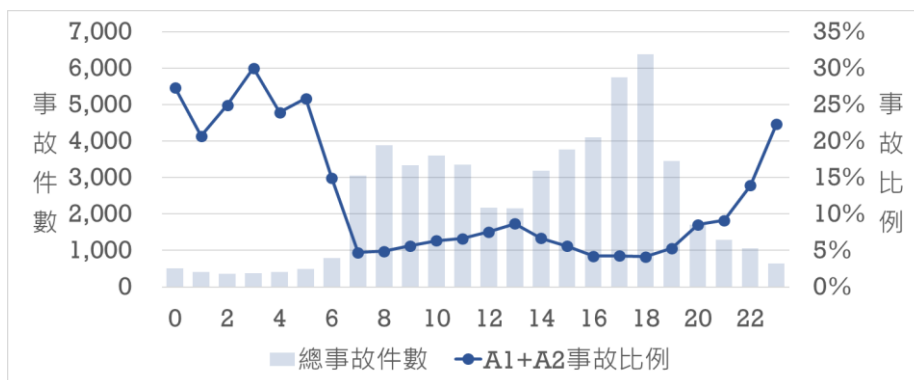


圖 2 分時事故統計

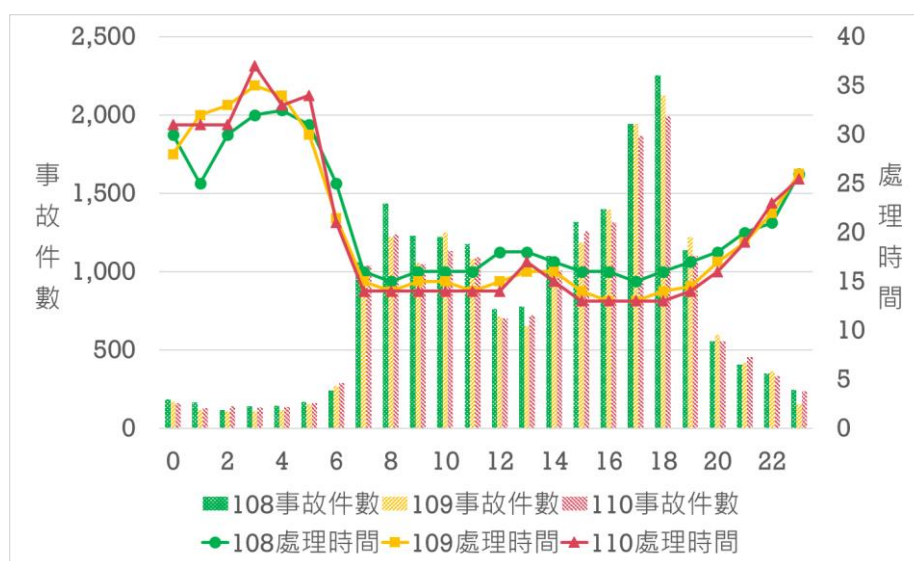


圖 3 分時處理時間統計

3.1.3 事故等級

針對事故等級進行統計後結果如表 1，A1 和 A2 事故的處理時間中位數分別為 A3 事故的 6.5 倍及 2.6 倍。

表 1 事故等級統計

事故等級	事故件數	比例	處理時間
A1	167	0.3%	98
A2	3,765	6.7%	39
A3	52,303	93.0%	15

資料來源：本研究整理

3.1.4 分區國道統計

依交控中心及國道分組統計處理時間中位數及件數如圖 4，除特定分組外處理時間差異小，而部分處理時間較長的分組樣本數低，可能僅為個案狀況。

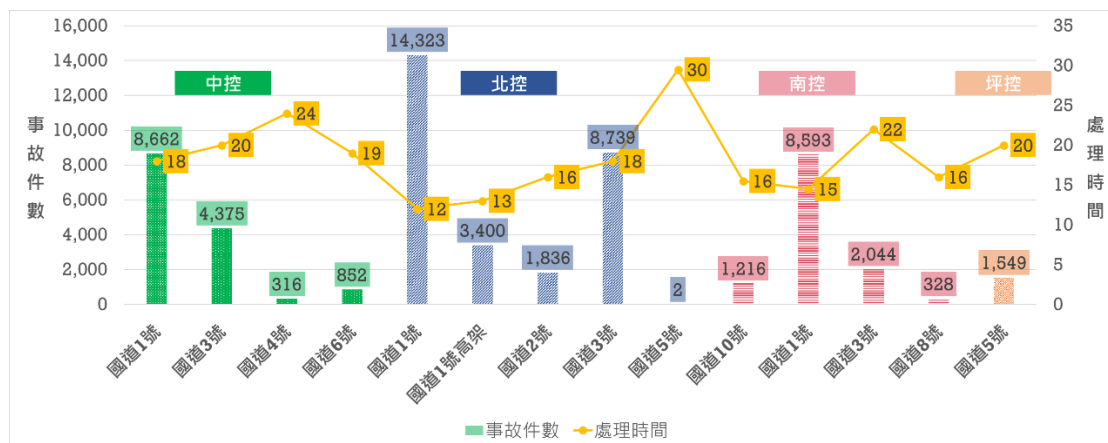


圖 4 分區國道統計

3.1.5 日期型態

除了依發生的小時統計外，另外一個面向則是發生當時的日期型態(如週一、週五、連假等)，各日期型態的統計如表 2，連假與週五的平均件數較多，但處理時間中位數皆約為 15-17 分鐘，日期型態對於處理時間的影響較小。

表 2 日期型態統計

	連假	週一	週二	週三	週四	週五	週六	週日
平均件數	63	48	45	48	47	67	52	51
處理時間	15	17	16	16	16	15	16	16

資料來源：本研究整理

3.1.6 車種

針對車種統計如表 3，由於同一事故可能涉及多車種，因此事故的肇事車輛中包含該車種則納入統計，表中顯示聯結車的事故處理時間中位數約為其它車種的 2 至 3 倍。

表 3 車種統計

	小客車	小貨車	大客車	大貨車	聯結車	其它
件數	48,972	5,460	945	4,074	272	9,160
比例	87.1%	9.7%	1.7%	7.2%	0.5%	16.3%
處理時間	16	20	20	21	40	15

資料來源：本研究整理

3.1.7 車道

事故發生時占用的車道也可能會影響處理時間，統計如表 4，同一事故可能占用多個車道，只要事故占用該車道則納入統計，表中顯示內路肩和中內車道事故處理時間中位數略高，但比例僅佔總事故的 0.5%及 2.3%。

表 4 車道統計

	內路肩	內車道	中內車道	中車道	中外車道	外車道	外路肩
件數	270	21,518	1,292	4,951	1,121	17,819	4,380
比例	0.5%	38.3%	2.3%	8.8%	2.0%	31.7%	7.8%
處理時間	23	16	20	16	17	16	19

資料來源：本研究整理

3.1.8 事故型態

事故型態係指事故概略的發生狀況，統計如表 5，表中顯示衝出邊坡、火燒車和翻覆的處理時間中位數較高。

表 5 事故型態統計

	追撞	擦撞	自撞	衝出邊坡	翻覆	撞護欄	火燒車	打橫
件數	49,407	149	1,482	86	1,270	3,833	480	65
比例	87.9%	0.3%	2.6%	0.2%	2.3%	6.8%	0.9%	0.1%
處理時間	15	17	25	59	41	22	56	30

資料來源：本研究整理

3.1.9 車種、型態及等級交叉統計

上述各變數中，車種、事故型態與事故等級各自影響處理時間甚鉅，為確認變數的特定組合是否有更為明顯的處理時間差異，本文交叉統計各分組的處理時間中位數如表 6，空格代表尚無樣本，持續蒐集後便能補齊表格並供管理人員參考。

表 6 交叉統計

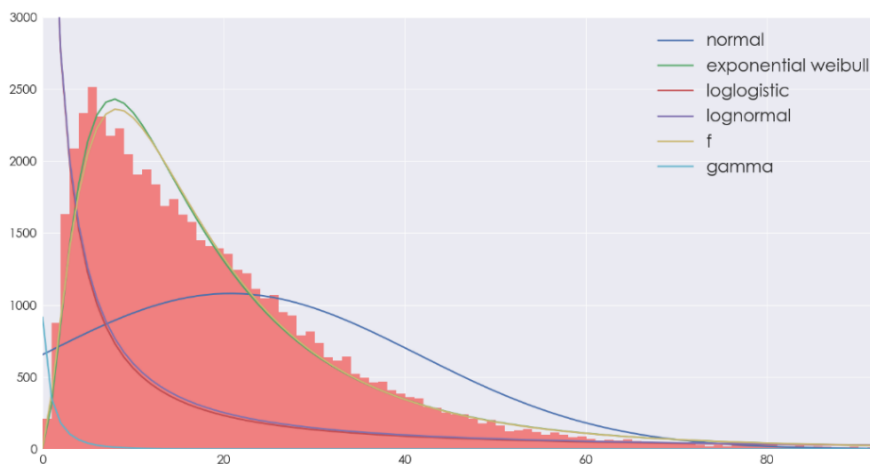
車種	等級	打橫	撞護欄	擦撞	火燒車	翻覆	自撞	衝出邊坡	追撞
小客車	A1		74	105	167	119	117	413	86
	A2	36	32	31	63	43	34	63	38
	A3	15	20	15	52	32	23	31	14
小貨車	A1	84	90		84	96	97		84
	A2	38	39	30	70	48	45	76	44
	A3	25	20	28	51	40	27	65	16
大客車	A1		124					144	71
	A2	67	51				107	678	43
	A3	27	16	14	76	38	63	30	17
大貨車	A1					181		413	113

車種	等級	打橫	撞護欄	擦撞	火燒車	翻覆	自撞	衝出邊坡	追撞
	A2		71	62	94	102	145	166	47
	A3	34	41	10	80	79	37	192	16
聯結車	A1		387			387			208
	A2		139		658	232	161	249	67
	A3	72	61	71	188	183	58	185	22
其它	A1	84	117	105	170	224	224		88
	A2	53	37	29	200	114	40	191	44
	A3	42	27	19	75	52	25	149	13

資料來源：本研究整理

3.2 處理時間分布配適

前一節觀察不同因素進行事件數與處理時間的敘述性統計，但要確認是否影響處理時間，則須透過假設檢定變數確認各組間是否有顯著差異。2.1 節中回顧以往有關事故處理時間之分布，結果皆為右偏之分布，可歸納成以下幾種：1. Weibull 分布、2. Log-logistic 分布、3. Log-normal 分布、4. f 分布及 5. gamma 分布，將上述各分布以及常態分布對事故處理時間進行配適後如圖 5。



資料來源：本研究整理

圖 5 處理時間分布配適

為確認處理時間之分布，以 Kolmogorov-Smirnov 檢定進行統計後如表 7，該檢定係以與理論的累積分布最大垂直距離為依據計算統計值，因此統計值愈低則代表與該理論分布愈接近，與處理時間分布差異最小的為 Exponential Weibull 分布，以此進行後續檢定。

表 7 分布 KS 檢定表

分布	統計值
Normal	0.1970
Exponential Weibull	0.0363
Loglogistic	0.5878
Lognormal	0.5864
F	0.0364
Gamma	0.9851

資料來源：本研究整理

3.3 變異數分析

以變異數分析(ANOVA)得以檢定每項因素各組是否具有差異，然而變異數分析具有三個假設：1.母群體為常態分布 2.變異數同質性 3.樣本為獨立樣本，以下分別進行說明。

3.3.1 變異數分析假設

1.母群體為常態分布：

根據上節的檢定得知處理時間呈現 Exponential Weibull 分布，而根據研究(Black, Ard, Smith and Schibik, 2010)對於實驗結果中型一錯誤的觀察，ANOVA 在資料呈現 Weibull 分布之資料仍具有穩健性，ANOVA 仍可適用於 Weibull 分布的資料。

2.變異數同質性

ANOVA 要求變異數同質性的假設，即各組樣本要來自變異數相等的分布，但由於事故簡訊資料各組間變異數差異過大(以車種、事故型態為例，如表 8)，違反變異數同質性假設，須改以 Welch's ANOVA 進行分析。

表 8 各分組變異數統計表

因素	分組	變異數
車種	小客車	197
	小貨車	558
	大客車	1,328
	大貨車	5,207
	聯結車	18,624
事故型態	追撞	259
	擦撞	341
	自撞	487
	衝出邊坡	7,060
	翻覆	4,553

因素	分組	變異數
	撞護欄	393
	火燒車	2,076

資料來源：本研究整理

3.樣本為獨立樣本

ANOVA 第 3 個假設是樣本為獨立樣本，本文採用的變數中除車種、車道、事故型態三個變數外，其他變數的分組樣本獨立，不互相影響，但上述三個變數因為同一事故可能涉及 2 個以上的分組(如該事故肇事車輛為 1 小客車及 1 小貨車)，因此進行後續分析時，只包含該事故僅涉單一分組的資料，若該筆資料涉及多分組則不予採用。

3.3.2 變異數分析結果

為符合 ANOVA 之假設，本文採用 Welch's ANOVA 並且使用獨立之樣本資料。結果如表 9，但由於樣本數大，各組間微小的變異也會放大，所有變數的 p 值都近似 0 且拒絕虛無假設，無益於本文所提及找出事故處理時間重要影響因素之目的。

表 9 變異數分析(全部樣本)

變數	統計值	p 值
年度	17.67	2.12e-08
小時	51.39	1.73e-219
事故等級	986.19	2.72e-160
日期型態	24.19	4.59e-33
分區國道	64.10	1.16e-31
車種	206.03	3.45e-71
車道	55.20	2.71e-58
事故型態	197.61	1.27e-123

資料來源：本研究整理

承上，為避免 ANOVA 檢定統計各因素皆拒絕虛無假設的問題，本文以隨機取樣的方式代替，各組以 30、40...100 的樣本數隨機取樣後進行分析，由於其隨機性，每個樣本數皆執行多次檢定，結果如表 10，表中的比例代表所有檢定內拒絕虛無假設的比例，本文以在所有檢定中超過半數拒絕虛無假設的因數作為後續事故處理時間預測的輸入因素，包含小時、事故等級、分區國道、車種、車道以及事故型態。

表 10 變異數分析統計表

變數	樣本數							
	30	40	50	60	70	80	90	100
年度	13%	0%	0%	3%	13%	7%	7%	3%
小時	100%	100%	100%	100%	100%	100%	100%	100%

變數	樣本數							
	30	40	50	60	70	80	90	100
事故等級	100%	100%	100%	100%	100%	100%	100%	100%
日期型態	17%	10%	13%	3%	3%	20%	13%	20%
分區國道	30%	43%	70%	70%	90%	93%	87%	93%
車種	100%	100%	100%	100%	100%	100%	100%	100%
車道	27%	50%	60%	73%	93%	83%	90%	97%
事故型態	100%	100%	100%	100%	100%	100%	100%	100%

資料來源：本研究整理

四、事故處理時間預測

4.1 實驗設計

本章以 108-110 年事故簡訊資料進行建模及交叉驗證，有關實驗中使用之變數、預測方法、驗證方式及績效計算方式分述如下：

4.1.1 相關變數

從前章節變異數分析中可知，對事故處理時間有影響的因素包含小時、事故等級、分區國道、車種、車道以及事故型態，故以此為輸入因素進行模型校估。

本文之輸入因素皆為類別型態，須將因素編碼成數值才能進行後續分析，常見的方式包含標籤編碼及獨熱編碼，說明如下：

1. 標籤編碼：將該因素編碼成單一變數，各組別以 1 個數值作為代表，如內路肩為 1、內車道為 2、中內車道為 3。
2. 獨熱編碼：將該因素各組別編碼成不同的變數，每個變數代表是否為該組別的二元變數，如該事故涉及內路肩為 1，否則為 0。

使用標籤編碼時，該變數會隱含順序關係，以本文情形則不適用，故輸入變數以獨熱編碼進行輸入資料建模，而本文之目的為事故處理時間預測，輸出變數為連續變數，整理如表 11：

表 11 相關變數彙整

輸入變數		
因素	組別說明	變數個數
小時	0、1、2...23	24
事故等級	A1、A2、A3	3
分區國道	北控國道 1 號、北控國道 3 號...	14
車種	小客車、小貨車...	5
車道	內路肩、內車道、中內車道...	7

輸入變數		
事故型態	追撞、擦撞、自撞、...	8
輸出變數		
事故處理時間	連續變數，以數值代表處理時間	1

資料來源：本研究整理

4.1.2 預測方法

本文為透過統計及機器學習等方法預測處理時間，統整現今的各模型預測績效以作為後續相關議題或應用之參考，本文採用資料競賽較常見且在競賽中績效卓越的方法進行比較，如下：

1. 線性迴歸(LR)：以最小平方法建立線性模型。
2. K近鄰(KNN)：從訓練資料中根據變數之距離選擇K筆資料進行預測。
3. 多層感知器(MLP)：以深度學習多層感知器搭配激活函數建立非線性模型。
4. 決策樹集成方法：資料科學競賽中常見之方法，許多文獻提出多種變形，本文採用梯度提升迴歸樹(GBRT)、直方梯度提升迴歸樹(HistGBRT)、極限梯度提升(XGBoost)、類別提升(CatBoost)及輕梯度提升機(LightGBM)。

4.1.3 績效計算

迴歸問題常見的績效計算方式包含平均絕對百分誤差(MAPE)、平均絕對誤差(MAE)和平均平方誤差(MSE)等，計算方式統整如表 12：

表 12 績效計算公式

績效	公式
平均絕對百分誤差(1)	$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left \frac{y_i - \hat{y}_i}{y_i} \right $
平均絕對誤差(2)	$MAE = \frac{1}{n} \sum_{i=1}^n y_i - \hat{y}_i $
平均平方誤差(3)	$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$

資料來源：Pramudya, 2018

n ：資料筆數

i ：第 i 筆資料， $i = 1, 2, 3 \dots n$

y_i ：第 i 筆資料的真實處理時間

\hat{y}_i ：第 i 筆資料的預測處理時間

上述三種方式中，平均絕對誤差接代表了預測與真實的事故處理時間相差了幾分鐘，而平均絕對百分誤差代表了誤差佔真實處理時間的比例，若真實處理時間

過低(如僅為 1 分鐘)，即使誤差僅為 1 分鐘，計算成 MAPE 則為 100%，預測績效差但實際誤差並不嚴重，故不適宜用為本研究之參考績效，綜上所述，本文以 MAE 作為後續模型驗證之績效。

4.1.4 交叉驗證

若希望提升模型之泛用程度，避免僅因單一様本資料作為選擇模型之依據，本文以常見之 KFold 方式進行交叉驗證，亦即將資料切分成 K 等份，將其中一份資料作為驗證資料，其它 K-1 份資料則為訓練資料，產出 K 個驗證績效，並以績效平均數值進行模型的調整參數及後續彙整依據。

4.2 模型校估

4.2.1 超參數調校

本文採用線性迴歸、K 近鄰等 8 種建模方式，各自具有許多超參數，需要調校以取得較佳的預測績效，所謂超參數即為透過給定而非訓練建立的參數，通常模型愈複雜則超參數愈多，其組合亦難以窮舉羅列，本文以 Optuna 演算法(Akiba, Sano, Yanase, Ohta and Koyoma, 2019)進行超參數的最佳化。

4.2.2 模型驗證績效

各模型透過 optuna 並設定 20 次試驗(trial，每次 trial 進行 KFold(5)計算平均績效並調整超參數)後超參數最佳化後如表 14，其中最佳績效的模型為 CatBoost，後續以此為模型進行建模預測。

表 14 模型驗證績效表

模型	最佳超參數	最佳績效
LR		10.897
KNN	{'n_neighbors':10,'weights':'uniform'}	10.897
MLP	{'activation':'relu','solver':'sgd','learning_rate':'invscaling','tol':0.00042438459590664394.243e-4}	10.409
GBRT	{'learning_rate':0.13154389978916736,'n_estimators':292,'subsample':0.7704169535624915,'min_samples_split':10,'min_samples_leaf':14,'max_depth':6,'max_features':'log2','tol':0.00010008255229965678}	10.405
HisGBRT	{'learning_rate':0.3162579075179961,'max_depth':8,'min_samples_leaf':5,'tol':0.0004693309174918695}	10.393
XGBoost	{'n_estimators':267,'max_depth':5,'learning_rate':0.16580391670699351,'gamma':18.194113216486908,'min_child_weight':84.54225667361568,'subsample':0.6153552300960966,'colsample_bytree':0.9789022702269768}	10.374
CatBoost	{'learning_rate':0.19007712047431674,'n_estimators':200}	10.338

模型	最佳超參數	最佳績效
	s':225,'min_child_samples':78,'subsample':0.4428237143810469,'max_depth':5}	
LightGBM	{'num_leaves':31,'learning_rate':0.07440135987071475,'n_estimators':257,'min_child_weight':69.31856217363796,'min_child_samples':58}	10.397

資料來源：本研究整理

4.3 預測結果

前一節驗證最佳績效的模型為 CatBoost，以此為模型並將 108-110 年的資料匯入進行訓練，再針對 111 年 1-2 月的資料進行預測，每筆資料都能透過真實值與預測值的差異計算出該筆資料的預測績效，平均而言預測績效為 11.23 分鐘，其它預測績效如表 15，50%的資料能夠預測到 8.59 分鐘內，75%的資料能預測到 12.79 分鐘內：

表 15 預測績效統計彙整

項目	數值
筆數	3,498
平均(分)	11.23
標準差	33.27
25 百分位數(分)	4.54
中位數(分)	8.59
75 百分位數(分)	12.79

資料來源：本研究整理

五、結論與建議

1. 各因素對事故處理時間中位數之影響

- (1) 凌晨時段(0-6 時)由於 A1 及 A2 事故比例較高，處理時間中位數也較高。
- (2) 事故等級對事故處理時間影響甚鉅，A2 約為 A3 的 2.6 倍，A1 約為 A3 的 6.5 倍。
- (3) 所有車種中，涉及聯結車的事故的處理時間約為其他車種的 2-3 倍。
- (4) 當事故發生衝出邊坡、火燒車及翻覆時，事故處理時間大幅上升，中位數約為 59、56、41 分鐘。

2. 為進行事故處理時間的預測，需確認對其有影響的因素並轉換成後續預測的輸入變數，本文以變異數分析進行檢定後，影響因素為小時、事故等級、分區國道、車種、車道以及事故型態。
3. 以線性迴歸、K 近鄰、多層感知器以及決策樹集成等方法建模，並搭配交叉驗證及超參數最佳化後，最佳的模型為 CatBoost，最佳驗證績效為 10.338，即平均而言驗證資料的預測處理時間與真實處理時間相差 10.338 分鐘。
4. 以 4.2 節驗證得到的最佳模型及超參數建模，並將 108-110 年所有資料輸入進行訓練，再以 111 年 1-2 月的資料進行預測，統計後為 11.23，即平均而言預測事故處理時間與真實處理時間相差 11.23 分鐘，中位數則為 8.59，代表 50%的預測與真實處理時間相差在 8.59 分鐘內。
5. 後續透過持續累積歷史資料、加入影響因素、建立其他模型及超參數調校最佳化等方式皆可能提升預測績效，本文建立之流程與模型亦得供後續應用在實際之生活場景及其他研究之參考。

參考文獻

- Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019), "Optuna: A Next-generation Hyperparameter Optimization Framework," DOI: 10.48550/ARXIV.1907.10902.
- Al Kaabi, A., Dissanayake, D., and Bird, R. (2012), "Response Time of Highway Traffic Accidents in Abu Dhabi," *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2278, No. 1, pp. 95-103.
- Black, G., Ard, D., Smith, J., and Schibik, T. (2010), "The impact of the Weibull distribution on the performance of the single-factor ANOVA model," *International Journal of Industrial Engineering Computations*, Vol. 1, No. 2, pp. 185-198.
- Chang, H.-L., and Chang, T.-P. (2013), "Prediction of Freeway Incident Duration based on Classification Tree Analysis," *Journal of the Eastern Asia Society for Transportation Studies*, Vol. 10, pp. 1964-1977.
- Chung, Y. (2010), "Development of an accident duration prediction model on the Korean Freeway Systems," *Accident Analysis & Prevention*, Vol. 42, No. 1, pp. 282-289.
- Chung, Y., Walubita, L.F., and Choi, K. (2010), "Modeling Accident Duration and Its Mitigation Strategies on South Korean Freeway Systems," *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2178, No. 1, pp. 49-57.
- Chung, Y., and Yoon, B.-J. (2012), "Analytical method to estimate accident duration using archived speed profile and its statistical analysis," *KSCE Journal of Civil Engineering*, Vol. 16, No. 6, pp. 1064-1070.
- Dimitriou, L., and Vlahogianni, E.I. (2015), "Fuzzy modeling of freeway accident duration with rainfall and traffic flow interactions," *Analytic Methods in Accident Research*, Vol. 5-6, pp. 59-71.

- Garib, A., Radwan, A.E., and Al-Deek, H. (1997), "Estimating Magnitude and Duration of Incident Delays," *Journal of Transportation Engineering*, Vol. 123, No. 6, pp. 459-466.
- Giuliano, G. (1989), "Incident characteristics, frequency, and duration on a high volume urban freeway," *Transportation Research Part A: General*, Vol. 23, No. 5, pp. 387-396.
- He, Q., Kamarianakis, Y., Jintanakul, K., and Wynter, L. (2013), "Incident Duration Prediction with Hybrid Tree-based Quantile Regression," *Complex Networks and Dynamic Systems*, pp. 287-305.
- Hojati, A. T., Ferreira, L., Charles, P., & bin Kabit, M. R. (2012) , "Analysing freeway traffic-incident duration using an Australian data set," *Road & Transport Research*, Vol. 21, No. 2, pp. 19-31.
- Ji, Y. beibei, Jiang, R., Qu, M., and Chung, E. (2014), "Traffic Incident Clearance Time and Arrival Time Prediction Based on Hazard Models," *Mathematical Problems in Engineering*, Vol. 2014, pp. 1-11.
- Khattak, A., Wang, X., and Zhang, H. (2012), "Incident management integration tool: dynamically predicting incident durations, secondary incident occurrence and incident delays," *IET Intelligent Transport Systems*, Vol. 6, No. 2, p. 204.
- Khattak, A., Wang, X., and Zhang, H. (2012), "Incident management integration tool: dynamically predicting incident durations, secondary incident occurrence and incident delays," *IET Intelligent Transport Systems*, Vol. 6, No. 2, p. 204.
- Khattak, A.J., Liu, J., Wali, B., Li, X., and Ng, M. (2016), "Modeling Traffic Incident Duration Using Quantile Regression: Transportation Research Record," *Journal of the Transportation Research Board*, Vol. 2554, No. 1, pp. 139-148.
- Knibbe, W.J.J., Alkim, T.P., Otten, J.F.W., and Aidoo, M.Y. (2006), "Automated estimation of incident duration on dutch highway," *2006 IEEE Intelligent Transportation Systems Conference*,.
- Kuang, L., Yan, H., Zhu, Y., Tu, S., and Fan, X. (2019), "Predicting duration of traffic accidents based on cost-sensitive Bayesian network and weighted K-nearest neighbor," *Journal of Intelligent Transportation Systems*, Vol. 23, No. 2, pp. 161-174.
- Li, R., Pereira, F.C., and Ben-Akiva, M.E. (2018), "Overview of traffic incident duration analysis and prediction," *European Transport Research Review*, Vol. 10, No. 2.
- Nam, D., and Mannering, F. (2000), "An exploratory hazard-based analysis of highway incident duration," *Transportation Research Part A: Policy and Practice*, Vol. 34, No. 2, pp. 85-102.
- Ozbay, K., and Noyan, N. (2006), "Estimation of incident clearance times using Bayesian Networks approach," *Accident Analysis & Prevention*, Vol. 38, No. 3, pp. 542-555.
- Peeta, S., Ramos, J., and Gedela, S. (2000), *Providing Real-Time Traffic Advisory and Route Guidance to Manage Borman Incidents On-Line Using the Hoosier Helper Program*, Purdue University.

- Pramudya, A.(2018) What is the difference MAE, MAPE, MSE, and RMSE, Retrieved July 1, 2022, website: <https://www.kaggle.com/discussions/getting-started/52081>
- Qi, Y., and Teng, H. (2008), “An Information-Based Time Sequential Approach to Online Incident Duration Prediction,” *Journal of Intelligent Transportation Systems*, Vol. 12, No. 1, pp. 1-12.
- Shen, L., and Huang, M. (2011), “Data Mining Method for Incident duration Prediction,” *Communications in Computer and Information Science*, pp. 484-492.
- Wang, W., Chen, H., and Bell, M. (2002), “A Study of the Characteristics of Traffic Incident Duration on Motorways,” *Traffic And Transportation Studies* (2002),.
- Wei, C.-H., and Lee, Y. (2005), “Applying Data Fusion Techniques to Traveler Information Services in Highway Network,” *Journal of the Eastern Asia Society for Transportation Studies*, Vol. 6, pp. 2457-2472.
- Wen, Y., Chen, S.Y., Xiong, Q.Y., Han, R.B., and Chen, S.Y. (2012), “Traffic Incident Duration Prediction Based on K-Nearest Neighbor,” *Applied Mechanics and Materials*, Vol. 253-255, pp. 1675-1681.
- Weng, J., Qiao, W., Qu, X., and Yan, X. (2015), “Cluster-based lognormal distribution model for accident duration,” *Transportmetrica A: Transport Science*, Vol. 11, No. 4, pp. 345-363.
- Wu, W., Chen, S., and Zheng, C. (2011), “Traffic Incident Duration Prediction Based on Support Vector Regression,” *ICCTP 2011*,.
- Ying Ji (2011), “Prediction of freeway incident duration based on the multi-model fusion algorithm,” *2011 International Conference on Remote Sensing, Environment and Transportation Engineering*,.
- Zhang, H., and Khattak, A.J. (2010), “Analysis of Cascading Incident Event Durations on Urban Freeways,” *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2178, No. 1, pp. 30-39.
- Zong, F., Zhang, H., Xu, H., Zhu, X., and Wang, L. (2013), “Predicting Severity and Duration of Road Traffic Accident,” *Mathematical Problems in Engineering*, Vol. 2013, pp. 1-9.