車聯網與人工智慧下之動態號誌控制初探

The Study of Adopting Connected Vechicle and Artificial Intelligence in Urban Traffic Signal Control

運輸資訊組 周家慶

研究期間：民國110年2月至110年12月

摘　要

近年來車聯網(V2X)與人工智慧(AI)等科技的發展，提供號誌控制不同思考，例如：相較於傳統車輛偵測技術，車聯網(V2X)可提供精確車輛位置與每秒 10 筆的微觀車流動態，在車聯網裝機率達 20%下即可執行號誌控制工作。而人工智慧強化學習則提供號誌控制演算於獨立路口、幹道系統以及路網系統等的模擬模式。

本所自 110 年起執行「5G 智慧交通數位神經中樞」計畫，將利用 5G 通訊環境，發展符合我國主要交通環境下之人工智慧號誌導入，與持續精進人工智慧在自動車流偵測應用，並藉由與「構建 5G 智慧交通數位神經中樞」實驗場域結合，來支援其號誌控制與管理需求，取得加乘效果，以提升車流運作效率及交通安全；且自 111 年起，本所進行「我國人工智慧車聯網之號誌控制模式探討」計畫。爰本研究期透過先期研究蒐集在應用車聯網與人工智慧等技術與模式，在獨立路口與幹道系統之號誌控制理論發展與實作案例，以為我國後續年度研究計畫之執行參考應用。

關鍵詞：

車聯網、人工智慧、強化學習、號誌控制。

車聯網與人工智慧下之動態號誌控制初探

一、計畫背景與目的

號誌化路口依其交通特性與分析範圍，可分類為獨立路口、幹道系統以及路網系統，採用控制方式為固定時制的定時號誌控制及動態查表或適應性號誌控制的動態號誌控制等方式，而我國都會區之號誌控制設計則多採用定時號誌控制。定時號誌控制為利用路口鄰近路段上歷史車流流量資料作分析，將週內日與週休日分為包含尖離峰數個時段，透過時制產生最佳化軟體(例如：Synchro)與現場微調方式，產生多套最佳化時制計畫儲存於該路口號誌控制器，號誌控制器則依預定排程執行不同時制計畫。動態號誌控制則需搭配車流偵測技術，即時掌握車流動態以對應調整時制計畫內容，動態號誌控制技術包括：觸動式、動態查表、適應性等號誌控制方式，而國內亦發展將定時號誌控制配合車流尖離峰變化的動態時段型態號誌控制。另號誌控制也針對特定車輛提供優先通行的優先號誌控制設計，例如：緊急車輛(救護車、消防車及救災車)優先、公車優先及輕軌優先等。

近年來車聯網(V2X)與人工智慧(AI)等科技的發展，提供號誌控制不同思考，例如：相較於傳統車輛偵測技術，車聯網(V2X)可提供精確車輛位置與每秒 10 筆的微觀車流動態，根據美國密西根大學運輸研究中心 Yiheng Feng 博士的研究，在車聯網(V2X)裝機率達 20%下即可執行號誌控制工作，又如美國華盛頓大學 Xuegang (Jeff) Ban 教授利用車聯網資料進行號誌控制最佳化研發。而人工智慧則提供號誌控制演算不同思考，例如：美國卡內基美隆大學 Stephen Smith 教授結合人工智慧與車流理論，採用分散式處理架構，進行多方向競爭車流量變化大之都市交通號誌控制最佳化演算，又如美國賓州大學 Hua Wei 博士等發展應用人工智慧強化學習於獨立路口、幹道系統以及路網系統等的模擬模式，而本所與桃園市政府亦相關研究與試作。

本所自 110 年起執行「5G 智慧交通數位神經中樞」計畫，利用 5G 通訊環境，發展符合我國主要交通環境下之人工智慧號誌導入，與持續精進人工智慧在自動車流偵測應用，並藉由與「構建 5G 智慧交通數位神經中樞」實驗場域結合，來支援其號誌控制與管理需求，取得加乘效果，以提升車流運作效率及交通安全。爰本研究期透過先期研究蒐集在應用車聯網與人工智慧等技術與模式，在獨立路口與幹

道系統之號誌控制理論發展與實作案例，以為我國後續年度研究計畫之執行參考。

二、車聯網在號誌控制應用

隨著無線通訊技術的進步，車載隨意網路(Vehicular ad hoc networks, VANETs)通過行動隨意網路(Mobile ad hoc networks, MANETs)原理，將之應用於車輛領域，並成為智慧型運輸系統的重要組成。在 VANETs 中，車輛能夠通過專用短距通訊(DSRC)或蜂巢通訊(C-V2X)相互通訊的車間通訊(V2V)與車路通訊(V2I)技術，並被稱為車聯網(Connected vehicle, CV)。來自車聯網的數據提供包括位置、速度、加速度和其他車輛動態數據的車輛狀態。相較於傳統車輛偵測器的點資料，車聯網提供的交通資訊更為穩定和持久，因為單一車輛的故障只會略微降低滲透率(Penetration rate)，車聯網系統仍然可以提供相對準確的車流資訊。此外，此架構可節省車輛偵測器的安裝和維護成本。相關研究曾利用車聯網所得車流資訊，搭配 Webster 號誌周期計算公式來發展適應性號誌控制，透過即時的時相順序或時比調整，來因應車流需求與車流抵達型態的動態變化。相關研究顯示，車聯網所得車流資訊，具備在號誌控制最佳化的應用潛力。

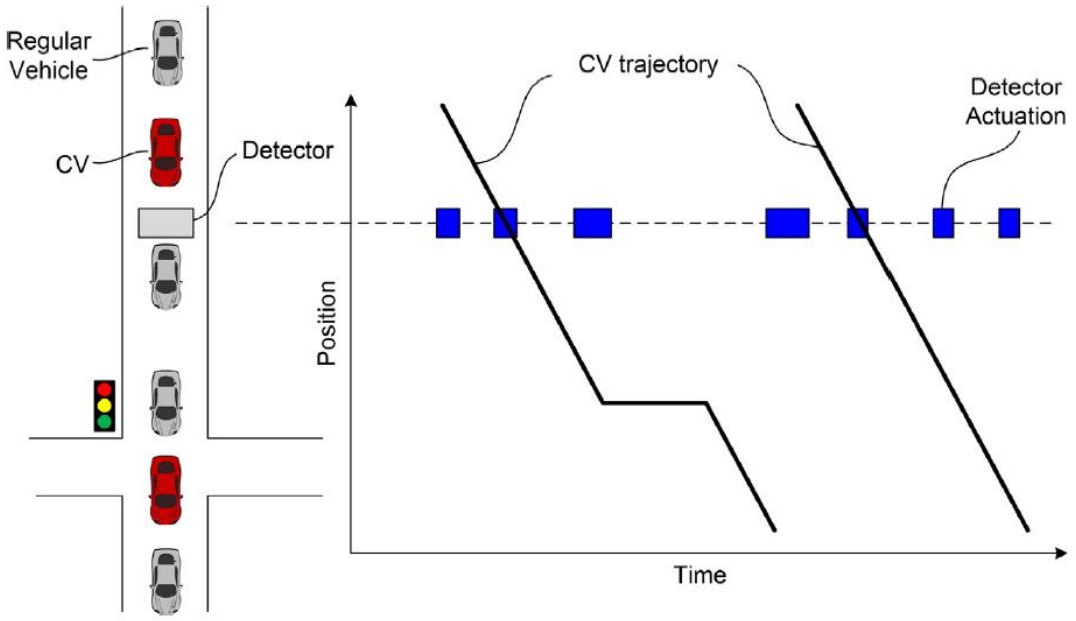
美國國家公路交通安全管理局(NHTSA)將車輛自動化定義為五個級別，從沒有任何自動控制功能車輛的第 0 級到全自動車輛的第 4 級。在未來很長一段時間內，不同程度的自動控制功能車輛將會併存於交通系統中；因此，如何考量第 0 級到第 4 級不同自動駕駛程度車輛併存於既有道路環境交通管理，將是重要課題。另一方面，通過聯網車輛提供的資訊，可以容易區別車輛類型，此將有助於對於緊急車輛與公共運輸車輛提供號誌優先通行(Traffic Signal Priority, TSP)的服務。

目前車聯網(V2X)通訊標準主要分成兩大類，分別為歐美日所支持的專用短距通訊(Dedicated Short Range Communications , DSRC)與以電信技術為基礎的 LTE C-V2X 與 5G。DSRC 專用短距通訊技術實體層採用 IEEE 802.11p 通訊標準開發，國際上主導 DSRC 標準制定的主要有美國、歐盟和日本，雖然歐美日的 DSRC 標準有所不同，但均以 IEEE 802.11p 標準為基礎，經由架設 DSRC 路側設備(或基地台)進行通訊，DSRC 不需透過電信商網絡，由車間或車輛與路側設備進

行短距通信。以電信技術為基礎 V2X 標準主導國際組織為 3GPP 標準組織，2018 年起進行 R.16 第五代蜂巢式車聯網(Rel.16 5G C-V2X)的標準化研究，以符合 5G 蜂巢式車聯網通訊需求。根據 3GPP 規劃，在訊息層將沿用美規的基本安全訊息(Basic Safety Message, BSM)及歐規 CAM (Cooperative Awareness Message)與 DENM (Decentralized Environmental Notification Message)訊息，其他各層通訊則使用 3GPP 相關設計。

不論是專用短距通訊(DSRC)或是電信車聯網(C-V2X)，車輛的車載單元(OBU)均是利用其全球衛星定位(GNSS)晶片所取得即時車輛位置、速度、行進方位角，以美規 BSM 訊息或歐規 CAM 訊息每秒 10 筆的頻率來進行廣播。相較於 1 分鐘或 5 分鐘的傳統車輛偵測器點速率、eTag 讀取器區段旅行時間或速率，以及 15 秒或 30 秒的公車動資訊，車聯網提供高頻率車輛即時動態軌跡，根據美國密西根大學運輸研究中心 Yiheng Feng 博士研究，在車聯網(V2X)裝機率達 20%下即利用此軌跡資訊來執行號誌控制。

車流模式可分為巨觀模式與微觀模式，微觀模式車流參數為個別車輛車聯網所得之動態資訊，例如：即時車輛位置、速度、航向、加減速度、車輛型態等，而巨觀模式車流參數為路段車流之交通量、等候線長度等。相較於傳統車輛偵測器所得的點車流資訊，車聯網則可取得車輛的即時軌跡資訊(如圖 1 所示)。號誌控制過程同時也會納入天候因素與事件的影響。號誌控制最佳化係透過包括時相計畫、號誌周期長度、時比與時差等在內之最佳化，來獲得路口平均停等延滯(等候時間)、等候車隊長度與旅行時間的減少。號誌控制最佳化亦考量納入環境因素於目標函數，以減少油耗與廢氣汙染。另目標函數以可考量複合指標，例如：號誌最佳化模式中可將延遲、停止次數和減速度組合的最小化，來進行號誌時制最佳化設計。



資料來源：Zheng 等, 2017, Estimating traffic volumes for signalized intersections using connected vehicle data

圖 1 車聯網與傳統車輛偵測器所得車流資訊示意圖

Priemer 與 Friedrich(2009)發展利用車路整合(V2I 資料的分散式適應性號誌控制演算法，該演算法每 5 秒運算 1 次，來預測下個 20 秒交通狀況，目標函數為等候線長度，最佳化演匱法為動態規劃(Dynamic programming)。Berg 等(2010)考量在 DSRC 車機幾乎為 0 情形下，發展或評估車聯網號誌控制演算法非常受限，因此他在加州 DENSO 公司附近構建可控制環境的紅燈所短與綠燈延長演算法，該演算法根據駛抵停止線的車輛數、速度與距離，給予路口 4 個方向不同權重，但此演算法過於單純，因此無法於真實環境下使用。Datesh 等(2011)發展在車聯網環境下，以車隊(Platoon)為基礎的號誌控制演算法，該演算法將號誌時相分為 2 個階段，第 1 階段視為等候佇列(Standing queue)，第 2 階段為車輛接近路口。根據車輛所廣播的 BSM 可取得其位置與速度，以及抵達路口的旅行時間，這些接近路口車輛利用 K-means 分為 2 群(Cluster)，並分配足夠綠燈時間給在綠燈群(Green cluster)的車輛通過路口，同時讓位於紅燈群(Red cluster)的車輛停止行進。根據實驗數據，DSRC 車機的裝機率須達 50%方能確保運作效率。Venkatanarayana 等(2011)考量在過飽和車流狀況下發展車聯網號誌控制演算法，該演算法同樣利用車輛 BSM 之位置與速度進行。演算法即時監控下游路口等候線長度，並動態調整號誌時差與時比來避免車流的溢流，不過此演算法似乎只能處理 2 個單行道的單純

路網。

He 等(2012)發展利用車路整合(V2I)資料的分散式適應性號誌控制演算法，以車隊車間距辨識(Headway-based platoon recognition)的PAMSCOD (Platoon-based Arterial Multi-modal Signal Control with Online Data)模式演算法來發展虛擬車隊(Pseudo-platoons)。最佳化部分則採用 MILP 方法，來計算不同車流狀況與號誌步階下的最佳化號誌時制。利用 VISSIM 模擬結果顯示，PAMSCOD 可以應用於處理非飽和與飽和車流狀況下之公共運輸車輛與小客車之雙運具環境，同時該模式需要 40%的 DSRC 裝機率方能產生效益，另在實務交通問題的決策變數增加時，該演算法需要強大運算能力的支持。2014 年 He 等進一步將其模式納入緊急車輛、商用車輛與行人，同時考量號誌連鎖與號誌觸動控制運作模式。

He 等(2012)同時利用 PAMSCOD 模式來評估不同滲透率情境下之車聯網號誌控制效能，根據模擬分析結果，車聯網號誌控制在約40%滲透率下，表現優於現有利用車輛偵測器的號誌控制方式。Lee 與 Park (2012)提出不需號誌資訊的協同式多代理人路口號誌控制演算法(Cooperative vehicle intersection control, CVIC)，CVIC 演算法去除路口所有衝突方向車輛的潛在重疊軌跡，為接近路口的每輛車輛尋求安全的通過的機制。此外，Lee 與 Par 還考量因為路口不可避免的車輛軌跡重疊和不可行解決方案所導致的系統故障情況，設計配合演算法進行處理。模擬結果顯示在不同擁擠程度十字路口案例，相較於傳統觸動號誌控制，CVIC 演算法分別獲得 99%與 33%的停等延滯和總旅行時間的減少。He 等(2012)亦利用 PAMSCOD 模式，在模式可以在車輛與號誌控制器兼具備通訊能力下，同時針對不同運具(公車與小客車)來進行路網號誌控最佳化。根據微觀模擬結果顯示，該模式可以同時考量公車與小客車對於後至控制的需求，並能顯著減少其停等延滯。

Kari (2014)開發以代理人為基礎的線上適應性號誌控制(Adaptive signal control, ASC)策略。相較於以 QEM (Quick Estimation Method)與 HCM 的策略，在穩定的車流需求環境下，該系統在旅行時間方面節省 5-14%，在燃油經濟性方面節省 0-5%。Guler 等(2014)發現車聯網滲透率從 0%增加到 60%，將可以顯著降低所提演算法的平均延滯。不過短期內車聯網高滲透率情境在實務上並不可行，因此如何在車聯

網低滲透率環境下，利用真實環境車聯網資料來改善號誌控制運作是一個緊迫的問題。Feng 等(2015)利用車聯網資料發展適應性動態時相演算法，針對低滲透率的車聯網課題，該演算法透過 EVLS (Estimation of Vehicle Location and Speed)模式來根據所蒐集車聯網資料，推估未配備車聯網車機車輛之動態資訊，以構建完整車流抵達資訊，根據模擬分析結果，車聯網號誌控制在高滲透率情境下，在總延滯減少上優於觸動號誌控制達 16.33%，而在低滲透率情境下，其表現則與觸動號誌控制類似。

Hu 等(2015)提出在車聯網連續多個路口連鎖環境下，以人延遲為基礎的智慧型公車優先號誌邏輯(Coordinated TSP Logic Based on Connected Vehicle Technology, TSPCV)最佳化模式，該模式除透過模擬軟體測試外，並於現場進行實作與評估。優先號誌控制除單純地考慮行駛時間和可靠性等因素外，尚可納入車輛重量、道路坡度和貨車引擎型態等因素，以減少貨運路廊貨車運行之能源使用與廢氣排放。該概念可以推廣到不同交通模式具有不同優先級的系統。

Zhu 與 Ukkusuri (2015)利用車聯網環境的車流資訊來發展路口自主號誌控制(Linear programming formulation for autonomous intersection control, LPAIC)的線性規劃公式。首先，導入以車道為基礎的雙層優化模式來進行車流的推進，系統最佳化路網模式中考量動態出發時間、動態路徑選擇與路口自主號誌控制。Feng 等 (2015)利用車聯網資料來進行適應性號誌控制，利用雙層式最佳化模式來進行時相順序與長度來取得總車流延滯與等候線長度最小之最佳化演算法提出了一種使用連接車輛數據的實時自適應信號相位分配算法。所提出的算法通過解決兩級優化問題來優化相序和持續時間，以最小化總車輛延遲和隊列長度，並利用 VISSIM 軟體離進行模擬與驗證。結果顯示在車聯網高滲透率(20%~30%)下，相較於定時號誌控制，所提出演算法可降低總延遲達 16.33%。Li 等 (2016)考量個別車輛之燃料消耗和行駛時間車輛特性，來進行號誌時制最佳化的混合整數非線性規劃(Mixed-integer nonlinear program)模式發展，該模式應用智慧駕駛模式(Intelligent driving model, IDM))來預測車聯網環境下之車輛軌跡。

Zheng and Liu (2017)將車輛到達號誌化路口的車輛到達方式以時間相關的卜松程序((Poisson Process)來表示，並利用期望值最大化(Expectation maximization, EM)程序來處理參數推估問題。期望值最大

化(EM)演算法是一個迭代過程，用於存在未觀察到或部分觀察到變數時來找到最適合 MLE。EM 演算法包括兩個主要步驟「E-step」與「M-step」。「E-step」根據初始參數來計算未觀察到或部分觀察到變數的條件期望值，以及可能性的條件期望值；然後「M-step」從最大概似估計(Maximizing the likelihood)中來蒐尋最佳參數的更新。EM 演算法重複這兩個步驟，直到更新收斂為止。

Liang 等(2020)針對無左轉車流十字路口，利用車聯網資料開發及時號誌控制演算法，根據所得車聯網車輛資訊，來掌握停等於路口非車聯網車輛的存在與否，然後識別出車流中的車隊(Platoon)型態。接下來演算法選擇最佳時相順序來紓解車流，以取得所有車輛平均延滯之最小化。而連鎖控制目標是透過所有號誌控制器間協調，來取得幹道或路網中的整體最佳化。Wang 等(2020)利用所有車聯網車輛資料進行幹道號誌連鎖的整合控制模式，來取得車輛速度的最佳化；該演算法將車聯網車輛形成若干車隊(Platoon)，而號誌設計方式讓車隊可以不用停等地通過路口或為最小停等狀況下通過路口，如此將可以降低因為號誌產生的延滯與增加通過量(Throughput)。根據真實路網的模擬顯示，該整合控制模式可以分別減少停等時間和停等次數達 53.69%和 41.15%，每輛車的路口延滯減少 13.19%。

Ishlam (2020)提出在都市交通路網的所有車輛與各路口均連結的假設下，發展分散式連鎖(Distributed-Coordinated)的號誌控制最佳化方法，此方式新穎之處在於分散式處理因此實務上可以進行即時控制且具備擴充性。但分散式系統的缺點為將導致區域間流量均衡的實現延後。因此，潛在探討方向是導入分層式結構，以結合集中式控制和分散式協同控制的優點。Rafter 等(2020)提出多模式適應性號誌控制(Multimode Adaptive Traffic Signals, MATS)創新號誌控制演算法，MATS 演算法整合聯網車輛的位置資訊與傳統環路線圈資料，以及號誌時制計畫內容來進行都會區的分散式號誌控制，所提 MATS 演算法可以在聯網車輛數量較少(滲透率較低)的情境下運作。

三、人工智慧強化學習在號誌控制應用

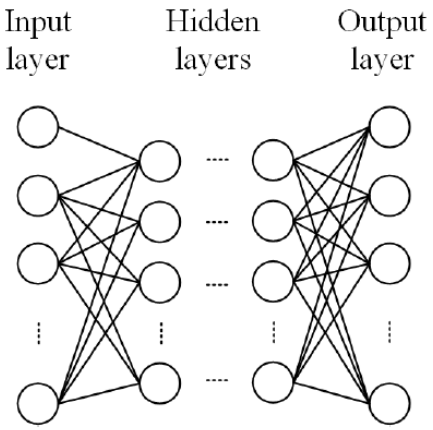
人工智慧(AI)被定義為對智慧代理人的研究，人工智慧感知外在環境並採取適當動作以最大限度地提高所追求目標的成功機會。因為人工智慧可以在系統不假設任何模式或其動態參數等先驗知識下，從

環境中進行學習，因此有望隨著時間的推進來改進動作，以及適應環境的變化；而無須如現有模式解方式須進行複雜的模式校估與優化過程，此特性適合即時取得微觀車流資訊車聯網環境特性下的號誌控制應用。相關研究案例包括多代理人系統、模糊理論(Fuzzy theory)、類神經網路(Artificial neural network)和強化學習(Reinforcement learning)。在強化學習中，車輛和號誌控制器都可以被視為「代理人(Agent)」，從這個角度來看，整個號誌控制系統就變成所謂的多代理人系統。多代理人系統具有無模式(Model free)、連鎖(Coordinated)，以及共同學習(Co-learning)等特性，且適合平行處理。

深度學習(Deep Learning, DL)是一種先進人工智慧方法，由深度神經網路(deep neural network, DNN)組成，例如：全連接層網路(Fully-connected layer network, FCLN)。所謂「深(deep)」的意義在於神經網路是由大量隱藏層(Hidden layer)組成(例如：多達 150 層)，它們之間可能是全連接(Fully-connected, FC)，而傳統的神經網路通常由數量少得多的隱藏層組成(例如：兩層或三層)。圖 2 顯示一個全連接層網路架構示意圖，它由三種主要類型的層組成，即輸入層(Input layer)、隱藏層(Hidden layer)和輸出層(Output layer)，它能夠學習非結構化和複雜數據的神經元(即處理元素)的互連組件。在訓練過程中，數據從輸入層流向輸出層。隱藏層神經元 k 的輸出 y_k 輸出層如下：

$$y_k = \varphi\left(\sum_{j=0} w_{kj} - x_j\right)$$

其中w_{kj}表示權重(或網路參數)，它是根據輸入x_j與其他輸入相比的相對重要性分配的，ϕ(.)表示神經元(Neuron) k 的激活函數(Activation function)。有多種深度學習架構應用於號誌控制(Traffic signal control, TSC)，包括傳統全連接層網路(FCLN)、卷積神經網路(Convolutional neural network , CNN)、堆疊自動編碼器(Stacked auto encoder, SAE)、決鬥網絡(Dueling network)和長短期記憶(Long short-term memory, LSTM)。



資料來源：Rasheed 等, 2020, Deep Reinforcement Learning for Traffic Signal Control: A Review

圖 2 全連接層網路示意圖

強化學習(Reinforcement learning, RL)是人工智慧的第三類型，它不同於監督學習(supervised learning)和非監督學習(unsupervised learning)方法。它使代理人(Agent)利用不同的狀態(State)與(Action)動作配對(pair)來進行探索，以便隨著時間推進(t=1, 2, 3, 4…..)，來實現系統性能增強的最佳可能正獎勵(Positive reward)(或負成本(Negative cost))。傳統強化學習演算法在某時刻 t∈T，代理人在動態(Dynamic)與隨機(Stochastic)操作環境中觀察到其馬可夫(Markovian)(或無記憶(Memoryless))決策(Decision-making)因素(Factors) (或狀態(State), st∈S)，並選擇與執行某個動作 a_t，而 $a_t \in A$。隨後，代理人觀察到下一個狀態 s_{t+1} 並接收立即獎勵(r_{t+1})(或成本(s_{t+1}))，它取決於狀態-動作配對(s_t, a_t)的下一個狀態 s_{t+1}。然後，在某個狀態-動作配對(s_t, a_t)，它更新代表知識的 Q 值($Q_t(S_t, a_t)$)。Q 值$Q_t(s_t, a_t)$表示在狀態(st)下採取行動(at)的適當性(Appropriateness)，Q 函數更新方式為$Q_{t+1}(s_t, a_t) \leftarrow Q_t(S_t, a_t) + \alpha\delta_t(s_t, a_t)$，其中α為學習率(Learning rate)，0≤α≤1，而$\delta_t(s_t, a_t)$是時序差分(Temporal difference, TD)，TD 是基於 Bellman 方程式，代表 2 個連續推估(Estimations)的立即(Immediate)獎勵與折扣(Discounted)獎勵差異。

$$\delta_t(s_t, a_t) = r_{t+1}(s_{t+1}) + \gamma \max_{a \in A} Q_t(s_{t+1}, a) - Q_t(s_t, a_t)$$

其中$\gamma \max_{a \in A} Q_t(s_{t+1}, a)$代表折扣獎勵，為在 t+1 時間的最大 Q 值；γ為折扣率(discount factor)，代表對於折扣獎勵的偏好，0≤γ≤1。換句話說，立即獎勵$r_{t+1}(s_{t+1})$代表短期獎勵，而$\gamma \max_{a \in A} Q_t(s_{t+1}, a)$代表長期獎勵。隨著時間的推進，代理人對於所有狀態與動作配對(s_t, a_t)進行探勘(explore)、更新與儲存$Q_t(s_{t+1}, a)$(Q 值)，因而產生 2 維的 Q 值

表。在動作選擇過程，代理人透過 2 種方式進行，分別為探勘(Exploration) 或開發 (Exploitation) 。探勘 (Exploration) 以小機率(Probability)ε選擇一個隨機動作來更新其 Q 值，以便在隨著時間推進，在動態和隨機的操作環境中識別出更好的動作。而開發(Exploitation)則是選擇機率為$1-\varepsilon$的最知名(貪婪(Greedy))的動作，利用如下價值函數(Value function)來取得狀態值的最大化。

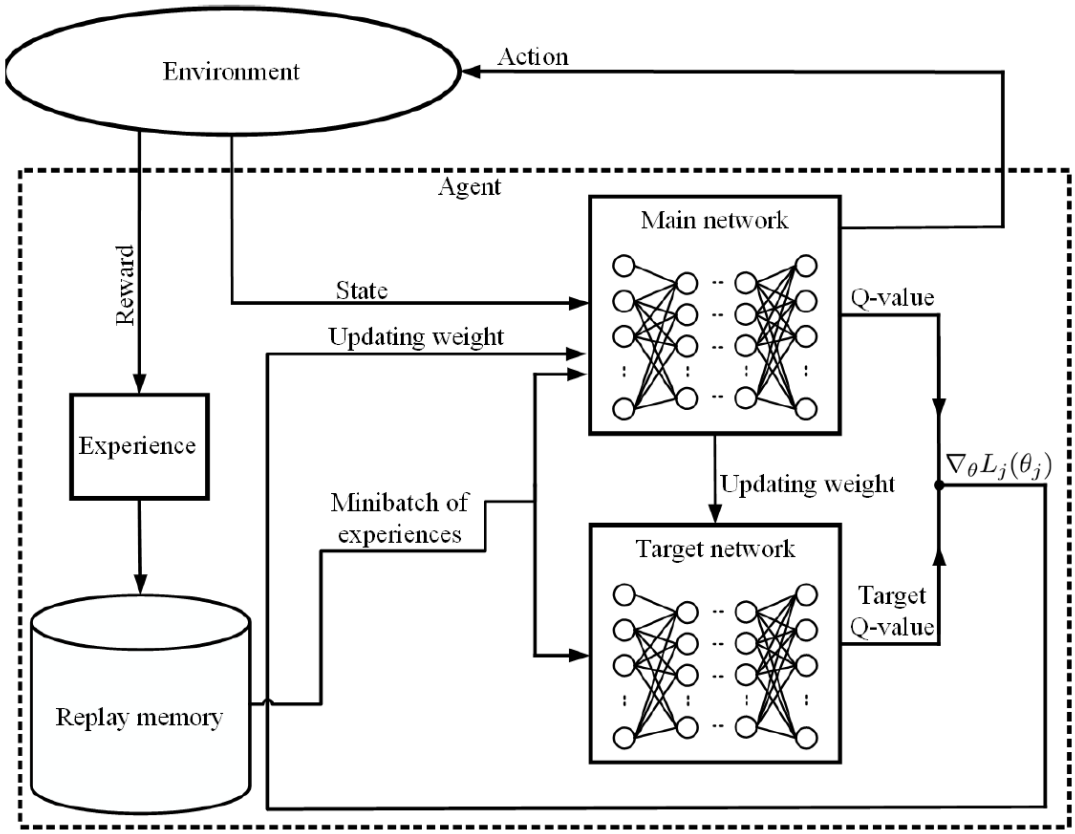
$$v_t^{\pi}(s_t) - \max_{a \in A} Q_t(s_t, a)$$

其中π為政策(policy)，係代理人用來根據目前狀態(s_t)來決定下一個動作a_{t+1}，π的定義如下：

$$\pi(s_t) = arg \max_{a \in A} Q_t(s_t, a)$$

因此代理人根據最大 Q 值來進行動作的選擇。

深度強化學習(DRL)是兩種人工智慧方法(即深度學習和強化學習)的結合。Deep Q-network (DQN)是 DeepMind 提出的第一個深度強化學習方法，廣泛應用於人工智慧號誌控制研究。DQN 有兩個主要特點，即經驗回放(Experience replay)和目標網路(Target network)。使用經驗回放，代理人將經驗儲存於重放記憶(Replay memory)，然後使用從重放記憶中，隨機選擇經驗來進行自我訓練。而使用目標網路，代理人利用主網路(Main network)的副本，並使用其權重來計算目標 Q 值，隨後用來使用梯度下降來求得損失函數(Loss function)的最小化。目標網路的權重值是固定的(或在一定次數的迭代後進行更新)，以提高訓練穩定性。在訓練過程中，目標 Q 值用於計算所選動作的損失以穩定(Stabilize)訓練，並且於一定次數迭代(Iteration)進行更新。主網路使代理人能夠在環境中在觀察其狀態後選擇一個動作，並隨後更新其主要 Q 值。DQN 擁有一種不同類型的深度學習架構，例如：全連接層網路、卷積神經網路、堆疊自動編碼器、3DQN 和長短期記憶。全連接層網路已廣泛用於 DQN，圖 3 為 DQN 架構圖。



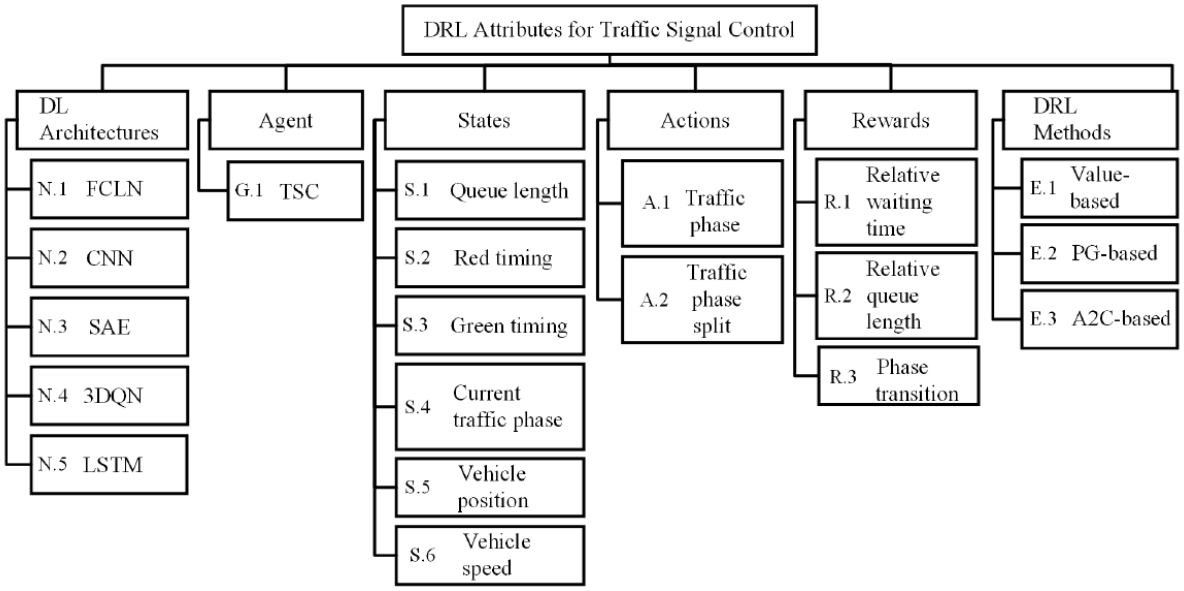
資料來源：Rasheed 等, 2020, Deep Reinforcement Learning for Traffic Signal Control: A Review

圖 3 DQN 架構圖

代理人具有三個主要組件，分別為重放記憶體(replay memory)、主網路(main network)和目標網路(target network)。重放記憶體指的是代理人經驗的資料集 $D_t=D(e_1, e_2,\ldots., e_t, \ldots..)$，資料集係隨著時間推進(t=1,2,3,…)，由代理人與環境互動過程蒐集產生。接下來，經驗資料集 Dt 使用於訓練過程中。主網路由一個全連接層網路組成，全連接層網路的權重θ_k用於在第 k 個迭代時逼近(approximate)其 Q 值，該 Q 值為$Q(s,a;\theta_k)$。主網路為從環境中觀察到的特定狀態 st 選擇一個動作，以便在下一個時段(t +1)時獲得可能的最佳獎勵$r_{t+1}(s_{t+1})$與下一個狀態s_{t+1}。

目標網路是主網路的副本，全連接層網路的權重值θ^-用來於第 k 次迭代後逼近其 Q 值$Q(s,a;\theta^-)$。主網路和目標網路的 Q 值有兩個主要區別，首先，主網路用於動作選擇和訓練，而目標網路僅用於訓練。目標網路提高訓練的穩定性，沒有目標網路，策略(policy)可能會在單個網路中的主要 Q 值和目標 Q 值之間振盪。其次，主網絡的 Q 值$(Q(s,a;\theta_k))$在每次迭代 k 中更新，而目標網錄的權重值θ^-通過在每 C 步時，複製主網路的權重值θ_j來更新，此相當於第 k 次迭代。

Rasheed 等(2020)回顧應用深度強化學習於單一路口、多路口、真實世界(Real world)和網格(Grid)等不同路網型態號誌控制研究文獻，並從深度學習架構、深度學習方法、模擬平台、代理人(Agent)、狀態(State)參數、動作(Action)、獎勵(Reward)等面向進行探討，圖 4 為深度強化學習於號誌控制研究不同面向之屬性彙整。

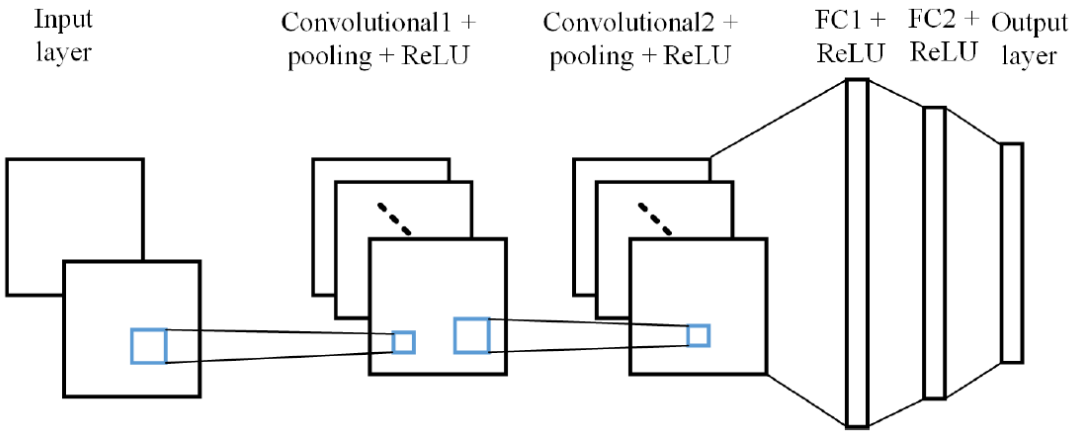


資料來源：Rasheed 等, 2020, Deep Reinforcement Learning for Traffic Signal Control: A Review

圖 4 深度強化學習於號誌控制不同面向屬性

一、深度學習架構

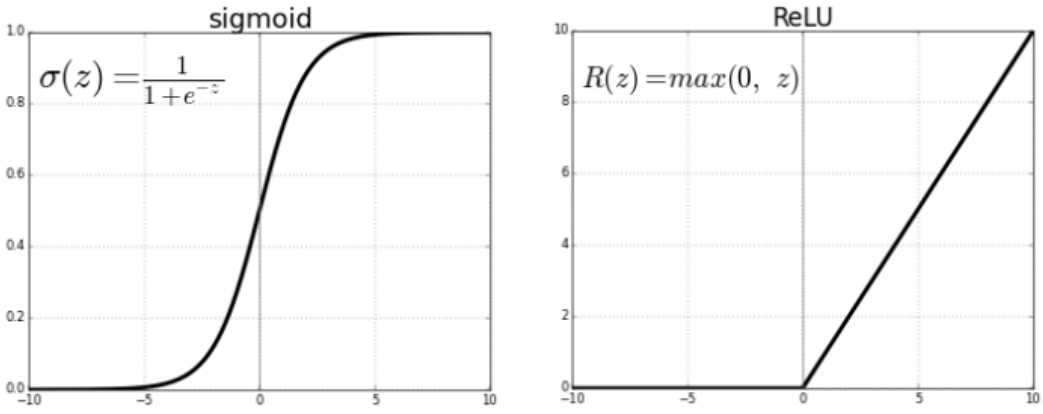
深度強化學習在號誌控制研究的深度學習架構有全連接層網路(FCLN)、卷積神經網路(CNN)、堆疊自動編碼器(Stacked auto encoder, SAE)、雙決鬥深度 Q 網路(Double dueling deep Q-network, 3DQN)、長短期記憶(Long short-term memory, LSTM)神經網路。全連接層網路與卷積神經網路(CNN)均應用來逼近(approximate)號誌控制的 Q 值。卷積神經網路有兩種主要類型的層 - 卷積層(Convolutional layer)和傳統的 FC 層，以圖 5 為例，CNN 有一個輸入層、兩個卷積層、兩個 FC 層和一個輸出層。每個卷積層由三部分組成，即卷積(Convolution)、池化(Pooling)和激活(Activation)，資料是從輸入層流向輸出層。圖 5 為 CNN 架構圖範例。



資料來源：Rasheed 等, 2020, Deep Reinforcement Learning for Traffic Signal Control: A Review

圖 5 CNN 架構範例

輸入層代表狀態(s_t^i)，例如：車輛位置與速度，以及假設在一個 60X60 網格(grid)的路口，而 1 輛車有 2 個狀態(位置與速度)，則該輸入層的矩陣大小為 60X60X2。2 個卷積層包括 k 個濾波器(或核心(Kernels))，每個濾波器包含一組權重值。每個權重將前一層的局部補丁(The local patches)(例如：影像的像素)進行聚合(Aggregates)，並根據每次步幅定義的固定步數(Steps)來移動前所聚合的局部補丁。透過池化將局部補丁的顯著值(Salient value)置換為整個補丁(The whole patch)，以刪除較不重要的資訊並降低輸入狀態(s_t^i)的維度。接下來，激活函數(例如：ReLU (Rectified Linear Unit))來激活補丁單元(The units of patches)。常用的 Sigmoid 與 ReLU 激活函數如圖 6 所示

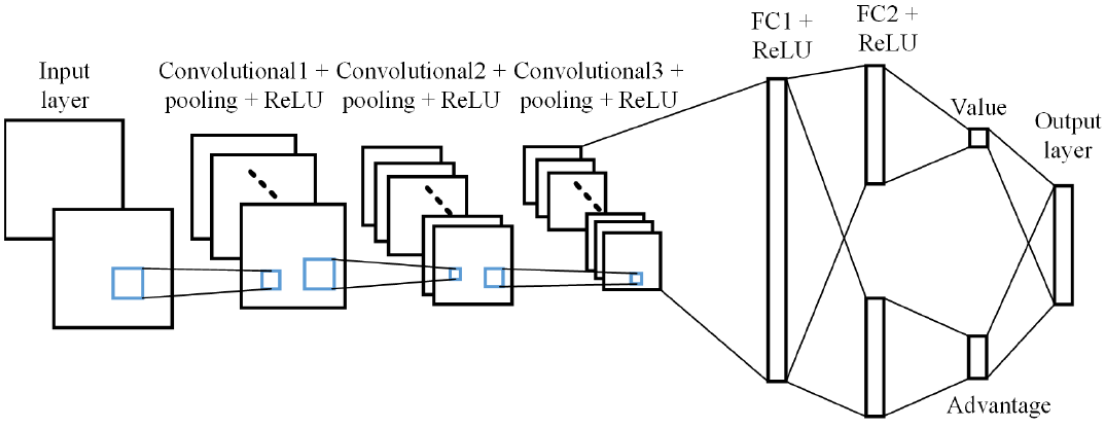


資料來源：https://towardsdatascience.com/ activation-functions-neural-networks-1cbd9f8d91d6

圖 6 Sigmoid 與 ReLU 激活函數

堆疊自動編碼器(Stacked auto encoder, SAE)神經網路執行編碼和解碼功能，用於逼近(Approximate)號誌控制的 Q 值，編碼器(encoder)將輸入資料映射(Map)到隱藏表示(Hidden representations)(意即進行

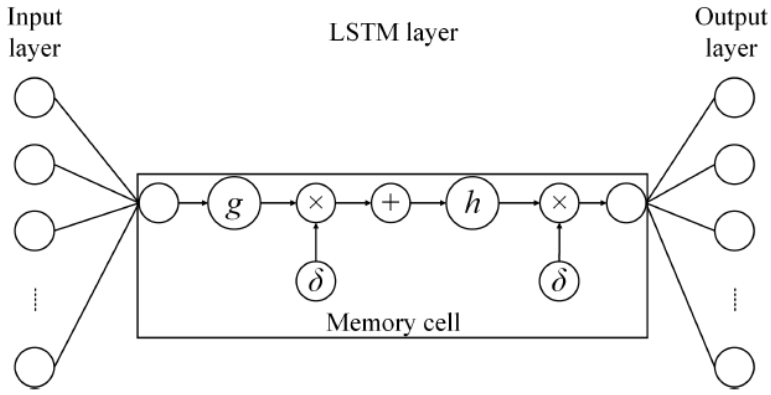
特徵萃取(Feature extraction))，解碼器(Decoder)從隱藏表示來重建輸入資料。雙決鬥深度 Q 網路(Double dueling deep Q-network, 3DQN)，其全連結(FC)層分為兩個獨立的流(Stream)來逼近(Approximate)號誌控制的 Q 值。3DQN 架構由雙 Q 學習(Q-learning)和決鬥網路(dueling network)組成，如圖 7 所示。在雙 Q 學習中，最大運算子(Max)在動作評估過程中，拆解出所選擇的動作，而在傳統的 Q 學習，最大運算子(Max)使用相同值來進行動作的選擇與評估。決鬥網路有兩個獨立的流(Stream)來分別估計狀態值和每個動作的優勢。資料由輸入層流向輸出層。



資料來源：Rasheed 等, 2020, Deep Reinforcement Learning for Traffic Signal Control: A Review

圖 7 3DQN 架構圖

基於循環神經網路(Recurrent neural network, RNN)的長短期記憶(Long short-term memory, LSTM)神經網路係由一個記憶單元組成；並且它已被採用於逼近(approximate)號誌控制的 Q 值(如圖 8 所示)，資料由輸入層流向輸出層。LSTM 層由一個保持狀態時間窗口的記憶單元(memory cell)組成，圖 8 中的記憶單元有一個輸入閘(g)和一個輸出閘(h)。此外還有兩個節點(node)用於乘法，一個節點用於加總(summation)「+」，以及兩個閘的激發函數(δ)將數據轉換為 0 到 1 間的數值。



資料來源：Rasheed 等, 2020, Deep Reinforcement Learning for Traffic Signal Control: A Review

圖 8 LSTM 架構案例

二、深度學習方法

有三種深度強化學習(DRL)方法可應用於號誌控制的深度學習，說明如下。

(一)價值基礎方法(Value-based method)：價值基礎方法是傳統 DQN 方法，而價值基礎 DQN 利用價值函數將每個狀態-動作(s_t, a_t)映射狀態值$V_t(s_t)$，以便識別每個狀態下的最佳可能動作。

(二)策略梯度基礎(PG)方法(Policy-gradient-based method)：策略梯度基礎方法的 DQN 係在給定狀態$(s_t \in S)$之概率分佈$\pi(a_t|s_t;\theta)$來選擇動作$a_t \in A$，其中概率分佈是通過對策略參數執行梯度下降來學習的(意即 DQN 的權重量θ)。因此原公式可改寫為

$$\nabla_\theta L_j(\theta_j) = E_{s_j,a_j\sim p(.)}\left[\sum \nabla_\theta \, log\pi_\theta(a_t|s_t)R_t\right]$$

其中$R_t = \sum_t \gamma r_t$為獎勵函數

(三)優勢演員評論(Advantage actor critic, A2C)基礎方法：A2C 基礎方法的 DQN 是混合方法，為價值基礎方法與策略基礎方法的結合。每個代理人都有一個控制其行為方式的演員(Actor)，而評論(Critic)用來衡量所選擇動作的適用性(suitability)。

$$\nabla_\theta L_j(\theta_j) = E_{s_j,a_j\sim p(.)}\left[\sum \nabla_\theta \, log\pi_\theta(a_t|s_t)A_t(s_t,a_t)\right]$$

其中$A_t(s_t, a_t) = Q_t(s_t, a_t) - V_t(s_t)$

三、模擬平台

車流模擬器是用來評估、比較和優化以深度強化學習為基礎的號誌控制模擬平台。一般來說，車流模擬器提供圖形化使用者介面與必要功能，以微觀或巨觀方式進行號誌控制、車輛與道路的模擬，以蒐集路口或路網的車輛等候線長度。微觀模擬會蒐集個別車輛的特性，例如：車輛速度與方向，而巨觀模擬則會蒐集車流的密度、車道速限等。大多數深度強化學習在號誌控制研究多利用微觀模擬進行，使用 SUMO、Paramics、VISSIM 和 Aimsun Nex 等軟體進行。

四、代理人(Agent)：指的是做出動作的執行者，可以為交控中心的主控電腦或智慧化路口的工業級電腦。

五、狀態(State)：指的是目前的車流與號誌的及時狀態，可以包括等候線長度、紅燈時相秒數、綠燈時相秒數、目前時相、車輛位置與速度等。

1. 紅燈時相秒數：表示紅燈時相時的秒數，當紅燈時，$s_t^{i,l}$=1，當綠燈或黃燈亮起時$s_t^{i,l}$=0。狀態s_t^{i,l^i}可以代表路口 i 的車道方向l^i的車道$d^{t^i} \in D^{l^i}$的紅燈秒數$t_{r,t}^{i,l^i}$。

2. 綠燈時相秒數：表示綠燈時相時的秒數，當綠燈時，$s_t^{i,l}$=1，當紅燈或黃燈亮起時$s_t^{i,l}$=0。狀態s_t^{i,l^i}可以代表路口 i 的車道方向l^i的車道$d^{t^i} \in D^{l^i}$的紅燈秒數$t_{r,t}^{i,l^i}$。

3. 目前號誌時相：代表經由決策過程後啟動之號誌時相，在 i 路口 t 時間狀態s_t^i，其子號誌時相可用$s_t^i = \left(s_t^{i,1}, s_t^{i,2}, s_t^{i,3}, s_t^{i,4}, \ldots, s_t^{i,8}\right) = (1,0,0,0,0,0,0,0)$來表示在時間 t 只有號誌時相 1 有作用。

4. 車輛位置：代表在路口 i 車行方向l^i車道d^{l^i}等候車輛實際位置，假設在一個車道自路口 i 開始往上游切成需多小片段，每個小片段放置一輛車，此時$s_t^i = (s_t^{i,1}, s_t^{i,2}, s_t^{i,3}, s_t^{i,4}, \ldots, s_t^{i,j})$表示每個小片段$(s_t^{i,1})$的位置，而$s_t^{i,j}$就是最大等候線長度。

5. 車輛速度：代表在路口 i 車行方向l^i車道d^{l^i}移動車輛的速度，假設在一個車道自路口 i 開始往上游切成需多小片段，每個

小片段放置一輛車，此時$s_t^i = (s_t^{i,1}, s_t^{i,2}, s_t^{i,3}, s_t^{i,4}, \ldots., s_t^{i,j})$表示每個小片段$(s_t^{i,1})$的速度，而$s_t^{i,*} = \{1,0.1,0.2, \ldots\ldots,0.9,1\}$，$s_1^i$=1代表這個車行方向的速限，而$s_1^i$=0 代表這個車行方向的最低速度為 0 kph。

六、動作(Action)：指的是代理人(Agent)所執行的動作內容，可以為號誌時相型態變換或時比調整。

1. 號誌時相型態(Traffic phase type)：表示在路口非衝突車流同步分配的綠燈號誌組合選擇，而在時間 t，路口 i 的動作集合a_t^i為$\{a_1^i, a_2^i, a_3^i, .., a_n^i\}$，候選動作的數量等於號誌時相數量。

2. 號誌時相時比調整(Traffic phase split)：表示在路口 i 選擇號誌時相的時間間隔，動作$a_t^i = \{a_1^i, a_2^i\}$表示不論代理人是維持現在號誌時相(a_1^i)或在一段時間未收到最佳獎勵時轉壞到另一動作(a_2^i)。

七、獎勵(Reward)：指的是在代理人(Agent)執行動作之後的車流變化，可以為相對等候時間與等候線長度。

1. 相對停等時間(Relative waiting time)：在這種表示方式中，代理人收到的獎勵(或成本)將會隨著路口車輛的平均等待時間而變化。由路口壅塞與號誌控制運作，車輛在路口的平均等待時間可能會增加。$r_{t+1}^i(s_{t+1}^i)$為相對值，例如：$r_{t+1}^i(s_{t+1}^i) = W_t^{i+1} - W_{t+1}^i$代表所有車輛在路口 i 的時間 t 與時間 t+1(或號誌時相間)的平均停等時間差。

2. 相對等候線長度(Relative queue length)：在此種表示方式中，代理人接收獎勵(或成本)，獎勵(或成本)將隨著路口車輛隊等候線長度增加或減少而變化。$r_{t+1}^i(s_{t+1}^i)$為相對值，例如：$r_{t+1}^i(s_{t+1}^i) = n_{c,t+1}^i - n_{q,t+1}^i$代表$n_{c,t+1}^i$車輛在路口(i)時的$n_{q,t+1}^i$等候線長度，此代表綠燈時間的時間是否夠。

3. 時相變換(Phase transition)：時相變換表示一個交通時相轉變的成本，例如：號誌時相轉變過程中產生的時間延遲。

Rasheed 等同時從導入深度強化學習在號誌控所面臨的挑戰、車流到達率、號誌控制架構、績效指標等個面向進行探討。Rasheed 等認為深度強化學習以針對不合適的號誌時相順序，以及不合適的時比

所造成的壅塞進行處理。在車流到達率部分，首先為 Poisson 為基礎的車流到達率，使用 Poisson 處理，有三個主要屬性來呈現真實交通模型的屬性，分別為(a)到達間隔時間呈現指數分佈(exponentially distributed)；(b)到達間隔時間是無記憶(memoryless)；(c)不同車道的抵達車輛數彼此相互獨立。其次為考量真實環境的車流到達率係根據車輛的穿越屬性(例如：車道切換、超車、行駛方向、行駛速度和目的地的物理特性)，來決定在一段時間內抵達路口車輛數的機率，真實環境車流模式包括跟車模式與 Nagel-Schereckenberg 模式。

在號誌控制系統架構方面，Rasheed 等認為集中式架構下的代理人(Agent)可以蒐集所有或鄰近代理人之車流統計資料(例如：等候線長度)，並選擇適當的動作(Action)(例如：時相型態變換)來取得整體系統效能的最佳化。接下來該集中式代理人可以執行該動作或將將該動作或知識(Knowledge)傳送至所有或周邊代理人，此周邊代理人可執行集中式代理人所傳送之動作，或利用集中式代理人所傳送的知識來產生新的動作。集中式架構有效率(Efficiency)、擴充性(Scalability)與穩健性(Robustness)等 3 項課題，首先集中式架構存在單點故障(Single point of failure)課題，集中式代理人的故障會影響整個交通路網的車流狀況。其次，集中式代理人會面臨系統擴充時的挑戰。最後，集中式代理人會因大量的資訊交換因而產生通訊上的負擔。

而分散式架構使多個分散式代理人能夠各自蒐集當地車流統計資訊，並選擇各自的動作進行操作。因此，一個複雜的問題被分成由分散式代理人解決的子問題，有助於提高效率和穩健性。在深度強化學習背景下，分散式架構使代理人能夠優化交通路網中的全域(Global) Q 值，以實現交通路網的整體目標。 為提供操作環境的整體視野，分散式代理人觀察它們各自的本地操作環境並了解其相鄰代理人的資訊(例如：獎勵和 Q 值)。代理人選擇各自對應的動作，隨著時間的推移，全域(global) Q 值收斂到最優平衡狀態，以達到聯合動作的最優化。

表 1 為 Rasheed 等彙整各種深度強化學習模式於號誌控制應用，每個深度強化學習模式都有其優勢，例如；3DQN 廣泛用於提高學習速度，CNN 廣泛用於影像處理與分析。

表 1 深度強化學習號誌控制之模式描述及其優勢

深度強化學習模式	模式描述	優勢
傳統價值基礎方法的 FCLN	代理人對應每個狀態與動作配對到價值，以產生某狀態下之最可能動作	儲存較有效率
價值基礎方法的卷積神經網絡(CNN)	卷積代理人對應每個狀態與動作配對到價值，以產生某狀態下之最可能動作	影響處理與分析較有效率
策略梯度基礎的卷積神經網絡(CNN)	卷積代理人根據透過梯度下降參數學習到的機率分布為基礎的策略，來選擇動作	影響處理與分析較有效率
價值基礎方法的堆疊自動編碼器(SAE)	使用加解碼函數於代理人在對應每個狀態與動作配對到價值，以產生某狀態下之最可能動作過程	壓縮資料時較有效率
價值基礎方法的雙決鬥深度 Q 網路(3DQN)	在代理人對應每個狀態與動作配對到價值，以產生某狀態下之最可能動作時，代理人會將全連結(FC)層分為 2 個個別流程(stream)	增加學習速度
勢演員評論(A2C)基礎方法的長短期記憶(LSTM)	代理人使用演員(actor)來控制它的行為，並藉由評論(critic)來評估所選擇動作之合適性(suitability)	在提供記憶體於記憶前次輸入上較有效率

Rasheed 等認為在導入深度強化學習於號誌控制時須考慮兩個主要面向。首先，需要充分理解所欲解決的交通問題，包括目標、問題敘述以及該問題的研究課題。其次，該問題的研究課題提出解決構想。例如：應用 3DQN 架構和基於價值方法的深度強化學習模式號誌控制，來減少車輛的平均旅行時間的解決問題目標。接下來，進行狀態的定義，動作和獎勵的表示方式的設計，並討論深度強化學習方法的選擇。最後，定義深度學習的架構。通常狀態選擇等候車隊長度、紅燈時相時間、綠燈時相時間、目前號誌時相、車輛位置和車速等。狀態、動作和獎勵表示等是在深度強化學習的方法和網路架構前須先定義的。

代理人(agent)在實務環境中觀察到的決策參數應該明確加以定義，獎勵最大化目標是減少車輛在路口的平均等待時間，因此代理人可使用車輛位置與速度來表示狀態。通過狀態的觀察，代理人可以依據狀態來決定其下一個動作(Action)。深度強化學習的輸入層係由輸

入神經元組成，輸入層可以 60 X 60 大小的網格來表示表示，而其中有 2 個 60 X 60 輸入狀態來表示車輛位置與速度。代理人透過適當的動作來追求獎勵的最大化，例如：對於減少車輛平均等待時間的目標，代理人必須選擇一個合適的時相的適當間隔，此時的動作可選擇維持現有時相，或在下一個時段於預設時相順序中切換到下一個時相，以解決不合適時比所帶來的挑戰，例如：輸出層可由 9 個神經元組成，而每個神經元代表一個動作。獎勵反應出代理人在狀態下執行某個動作後實現的目標，例如：獎勵可以是路口車輛平均等待時間的增加或減少， 因此代理人使用相對等待時間來代表獎勵。透過獎勵的增加，代理人可以在達到目標的同時，亦可以增加系統效能。

在深度強化學習於號誌控制的方法選擇方面，與模式相關方法的目標(例如：動態調整的折扣因子，或將多個種機制整合於一個架構)應該有良好的定義，例如：為提高學習率，而將雙 Q 學習、決鬥(dueling)網路和優先經驗回放等 3 種方法整合於一個架構中，以增加學習率。高學習率可以減少學習時間，有助於縮短在探索所有狀態-動作組合找出最佳動作所需的時間。例如透過時相選擇最佳化的號誌控制來實現更順暢的車流。基於價值的方法將每個狀態-動作組合映射(Map)到一個值 $Vt\ (s_t)$以便識別每個狀態下的最佳可能動作。

為應對號誌控制的挑戰，應為深度強化學習定義出合適的深度學習架構，例如：3DQN 架構由 3 個卷積層和 2 個 FC 層組成，用於獲得車輛位置與速度，以應對不當時比所帶來的挑戰。由於車輛位置和速度是透過圖形(或影像)方式取得，因此選擇具有捲積層的 3DQN 架構。同時選擇合適的網數層數目是一個重要課題，較少層數的類神經網路可能難以擬合(Fit)訓練數據，而較少層數的類神經網路可能會由於訓練數據的記憶效應特性導致過度擬合(Overfit)，因而對性能產生負面影響。同樣，在每一層中確定合適數量的神經元(Neuron)是另一個重要課題，一般來說，輸入層的神經元個數等於特徵個數，例如：輸入層中有 80 個神經元來表示路口的 80 個單元，從而能代表車輛的等候線長度與位置；然而，隱藏層的神經元數量設計並不單純，根據經驗儘管更多的神經元可以提高系統效能，但是將增加模式的複雜度。

四、人工智慧車聯網在號誌控制應用

Surtrac (Scalable URban TRAffic Control)系統是由內基美隆大學 Stephen Smith 博士智慧協調與物流實驗室(Intelligent Coordination and Logistics Laboratory, ICLL)團隊所發展之適應性號誌控制系統，曾在匹茲堡市進行實測，並獲得 25%的旅行時間縮短與 40%的路口停等時間減少等量化效益。Surtrac 結合人工智慧與車流理論，採用分散式處理架構進行多方向競爭車流量變化大之都市交通號誌控制最佳化演算，每個路口透過各自的影像式或雷達式車輛偵測器進行車流偵測與產生最佳化號誌控制內容，各路口所偵測交通量與預測流出量也會透過通訊傳送至周邊路口，以利取得幹道或路網之最佳化號誌控制。同時由與採用分散式控制架構，所以並無實際交控中心，且系統擴充彈性佳，目前匹茲堡市部分路口由 Surtrac 進行號誌控制。Surtrac 納入行人自行車與公車，並結合車聯網與適應性號誌控制，在 Baum Boulevard 與 Centre Avenue 走廊 24 個路口裝有 DSRC 車聯網路側設備，進行整合式公共運輸優先號誌控制與及時車輛路徑導引。Surtrac 於 2014 年 6 月 18 日申請美國專利，並於 2015 年 10 月 13 日取得美國專利該專利證號為 US 9,159,229 B2。

Surtrac 近來利用公車 DSRC 車機所廣播即時資訊，整合於核心路口排程模組(core intersection scheduling procedure)來進行公共運輸優先號誌運作。該模式運作需要知道公車站位置、乘客上下車時間、車輛乘客數、班表、車門開啟狀態等以更精確掌握公共運輸車輛抵達路口時間，以及優先號誌運作邏輯。根據文件說明，其演算法考量並無特別獨特處，我國優先號誌邏輯與時也發展多時，最近較大規模實作是在臺南市，多條幹線公車在所行駛幹道先進行號誌時制重整後，實施優先號誌。相較於 Surtrac 所說明之號誌優先處理，我國尚未納入考量的為乘客上下車時間、車輛乘客數與車門開啟狀態，同較無考量對橫向車流之衝擊；一般而言，我國的交通環境與車流量均較 Surtrac 目前在匹茲堡市為複雜與大。因此其演算法在我國交通環境下之適用性，尚有待進一步進行驗證。

Liu (2017)提出基於強化親和推播(Enhanced affinity propagation)特性，結合車聯網 V2X 路網動態群聚演算法(V2X networks' dynamic clustering algorithm)；經由整合群聚演算法，提出協同式強化學習控制方案來平衡車流負載。為了處理強化學習的高維度(Dimensionality)

課題，該研究導入分散式路口間合作機制，並提出快速梯度下降函數逼近(Fast gradient-descent function approximation)方法來提高即時控制的實用性。根據模擬顯示，該演算法獲得最短的路口停等時間，以及最小的等候車隊長度。Cheng (2017)應用強化學習來微調路網的控制參數，使其適應各種車流狀況，模擬顯示在高動態車流環境下，該演算法表現良好。Wang (2021)將號誌與車聯網控制問題以強化學習(RL)問題來表述，其動作(Action)和狀態(State)空間的設計是特別考慮車聯網車輛特性，在獎勵設計部分，考慮改道對於路網流量效率的衝擊，數值模擬顯示該模式的有效性和效率。

五、我國人工智慧車聯網在號誌控制發展

我國目前在車聯網相關研究所應用多著重在 V2X 交通安全課題，以及提供自動駕駛車輛或緊急車輛於通過號誌化路口之號誌資訊或優先號誌服務，尚未就利用車聯網所得車流資訊，來進行動態號誌控制進行相關研究。

許添本與程楷祐(2020)以深度強化學習方式建構混合車流之 AI 最佳化號誌時制計畫，該研究以臺北市松山路與永吉路路口為對象，使用 VISSIM 軟體建立車流模擬環境，以 106 年交通流量調查作為研究車流輸入參考依據。深度強化學習之狀態包括車輛位置、車速以及當前時相。在決策時間點上所能做的動作即是否需切換至下一時相，因此號誌系統運作的方式為四時相之輪替，時相間具有順序性，代理人每 5 秒會進行狀態偵測並判斷是否切換至下一時相，但必須每個時相須滿足最小綠燈時間 15 秒，以及在時相轉換過程的黃燈時間 3 秒以及全紅時間 2 秒。

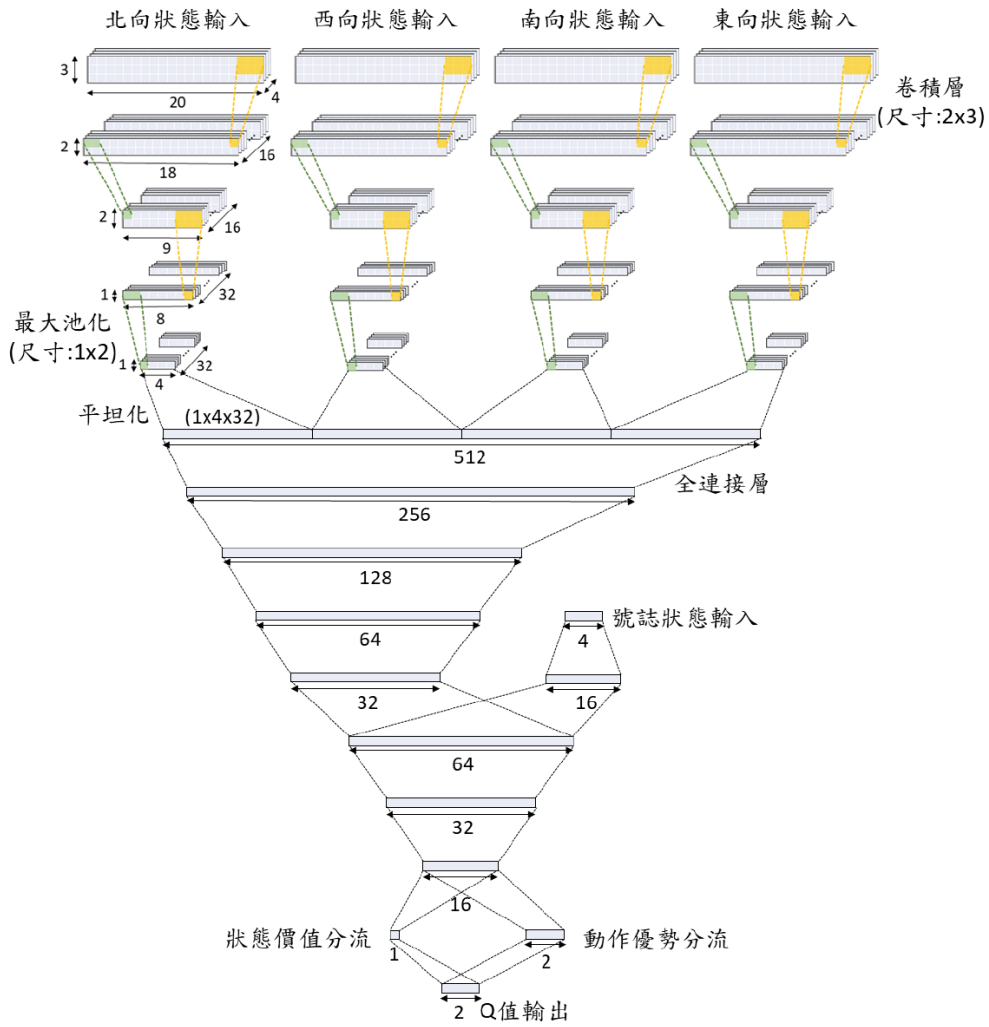
該研究採用加權指標作為獎勵，包括路口之車輛累積停等時間、路口通過量以及切換時相之懲罰。代理人在訓練初期時，因仍在探索階段使得所做出之決策通常並非最佳之動作，當連續錯誤之決策執行導致路口容易發生過飽和狀態，因此加入路口通過量以增加號誌系統訓練效率。由於各績效指標單位不同，因此在獎勵公式中給予各績效指標權重，以避免尺度不同導致部分績效指標過小而被忽略之情形發生。此設計方式會使代理人不論是在訓練環境、高流量、低流量、對稱流量、非對稱流量等 4 種情境下，代理人會依據當前時相所通行之車流紓解狀況來決定是否切換時相，而偏好繼續紓解該流向之尚未紓

解的車流。

該研究採用的人工智慧強化學習演算法為深度 Q 學習(Deep Q-learning, DQN)於被結合雙 Q 學習(Double Q-learning)、競爭網路架構(Dueling Network Architecture)及優先經驗回放(Experience replay)技術。在狀態輸入部分包含各方向之格位狀態以及號誌時相狀態，首先透過卷積網路萃取各方向之特徵，在各方向之卷積網路中包含了兩層卷積層及兩層最大池化層向下傳遞特徵，最終將各方向特徵平坦化後進入全連接層，在後續類神經網路中會另外加入號誌時相狀態，網路架構如圖 9 所示。由於號誌時相狀態的特徵數量較少，為避免號誌時相狀態少量特徵被忽略，因此在特徵數量較少的隱藏層才另外輸入號誌時相特徵。最終 Q 值輸出部分，該研究採用競爭網路架構，因此 Q 值在輸出之前，會先有動作優勢和狀態價值之分流，兩者相加過後才等於是 Q 值輸出。Q 值表示在此輸入狀態下動作之價值，因該研究動作為切換時相或不切換時相，故輸出之兩個 Q 值即分別表示切換時相及不切換時相之價值，代理人即以此價值做為決策判斷。

許添本與蔡沐軒(2021)延續前期研究，進一步考量深度強化學習存在學習效率課題，進行示範式深度強化學習在號誌時制最佳化應用，以現有號誌控制的最佳化定時控制時制計畫及全觸動時制計畫為示範對象，來建構示範式深度強化學習模型。結果顯示經由示範進行之深度強化學習能有效縮短所需訓練回合數，且路網績效亦較示範系統佳，於相近訓練回合數，該研究模型較單純深度強化學習模型更能適應車流的變化。最後透過不同交通量、轉向比與汽機車混合比情境模擬，驗證其對於路網變化具有一定適應能力。

許添本等(2020 與 2021)研究之後續建議有導入可辨識車種與車輛位置及速度之影像式車輛偵測器、考量時相跳躍(Skip)的動作、每次週期時相秒數不同之用路人心理因素、導入考量「綠燈續進帶」/「時差」/「車輛停等次數」等獎勵之幹道多路口多代理人等。



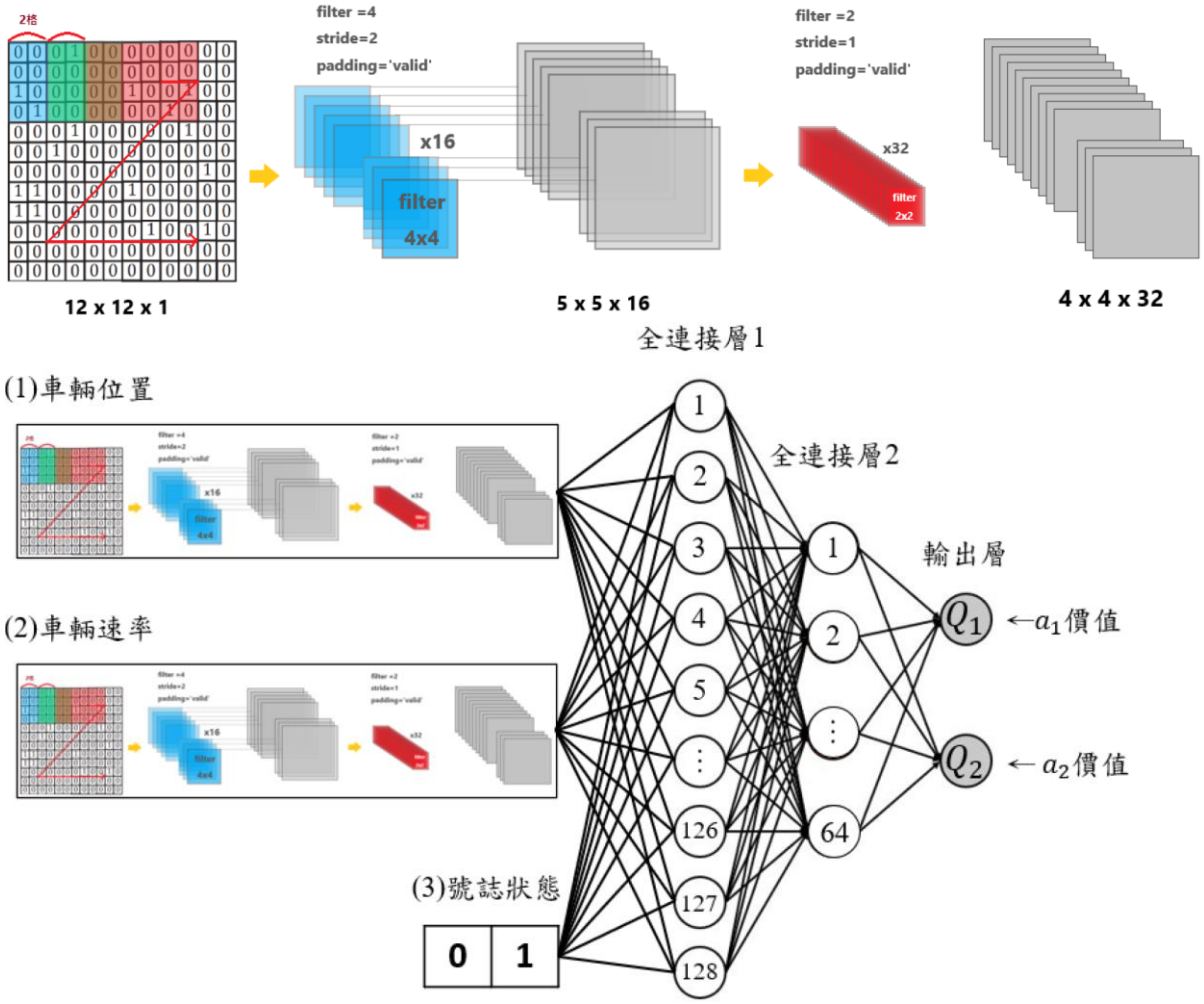
資料來源：許添本與程楷祐，2020 年以深度強化學習方式建構混合車流之 AI 最佳化號誌時制計畫

圖 9 Q 網路(卷積神經網路)架構圖

胡大瀛與李卓育(2021)就深度強化學習下號誌控制設計獎勵機制進行探討，比較與探討不同獎勵機制對於號誌代理的訓練過程與結果。該研究以臺南市東區林森路與東寧路路口為對象，利用 SUMO (Simulation of urban mobility)車流模擬軟體作為深度強化學習訓練環境，代理人能夠觀察路口時相狀態並偵測各個車輛的位置與速率特徵，經由深度強化學習模式處理與輸出動作價值，提供代理人作為判斷依據，而代理人做出動作決策後，環境將會回饋獎勵值做為代理人動作的貢獻程度。該研究利用卷積神經網路(CNN)的 4x4 濾波器(filter)與 2x2 濾波器權重改變，搭配 ReLU 激活函數，最後與號誌狀態一同輸入三層全連接層，分別為 128 個與 64 個神經元，最後輸出動作價值，供代理人選擇動作策略，CNN 與 DL 網路架構如圖如圖 10 所示。該研究利用 DDQN 方式來進行學習，故建置兩個相同的深度神經網路架構，分別為 Main Q-Network 與 Target Q-Network。

該研究於 SUMO 利用 TraCI (Traffic control interface)取得模擬過程所需車流資訊，例如：車輛所位車道、瞬時速率、位置等相關資訊；考量的狀態包括：某時刻小客車於路口的空間分布位置、速率與當前號誌的時相狀態；設置「南北向直右綠燈」和「東西向直右綠燈」等兩個動作；所探討之獎勵包括「車隊等候長度」、「路口通行量」、「車輛停等時間」等。使用三種不同獎勵會有不同的獎勵與代理人學習過程或走向，只要所回饋獎勵合乎邏輯且能夠分辨代理人動作優劣，代理人便能進行學習，逐漸改善神經網路的權重與產生輸出相對應的動作價值。

在後續研究建議上，胡大瀛與李卓育(2021)建議可以結合不同的交通參數來設計獎勵，代理人能夠更全面的往正確方向學習；此外，建議訓練一個代理人來進行兩個路口的號誌控制，以考量相鄰路口見的連鎖運作效率。

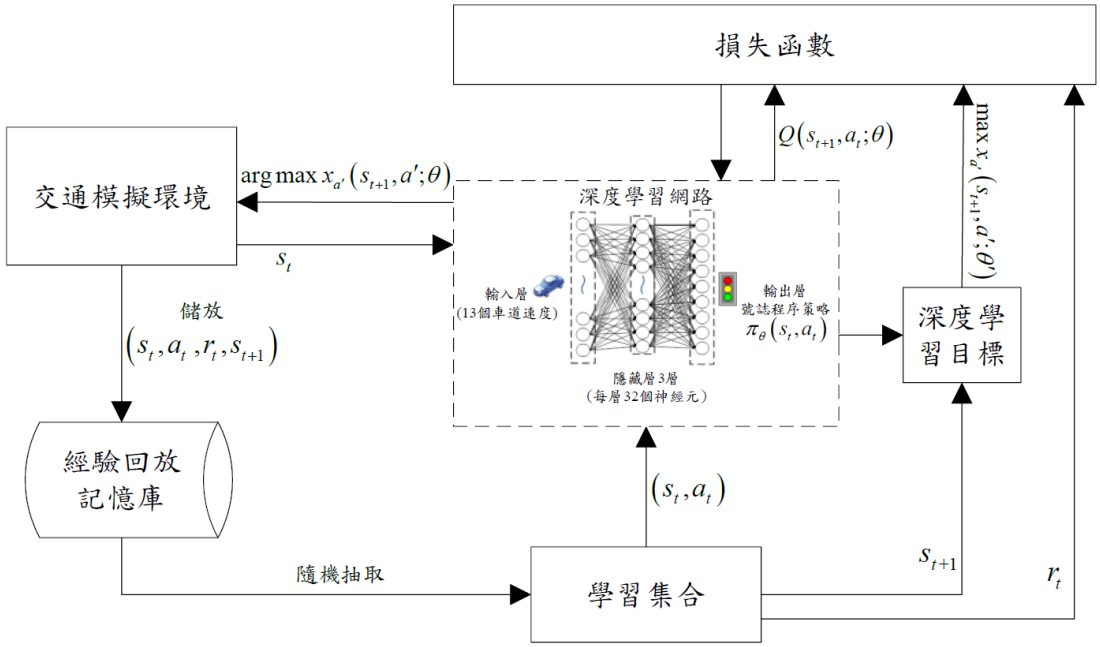


資料來源：胡大瀛、李卓育，2021 年，深度強化學習下號誌控制設計獎勵機制之探討

圖 10 CNN 與 DL 網路架構

趙世祺、邱俊偉、嚴國基(2021)應用深度強化學習應用於國道都會區(國 1 北上 27 公里(三重北上入口匝道)至北上 23 公里(圓山交流道濱江街出口)間路段為研究環境，進行匝道儀控之預定號誌程序深度強化學習模型(Deep Reinforcement Learning with Predefined Signal Program, DRLPSP)研發。該研究利用 SUMO 車流模擬軟體作為深度強化學習訓練環境，透過 Python 程式擷取模擬過程匯出車流數據，經過邏輯判斷或運算，動態調整模擬軟體中的交通參數，例如：號誌秒數或車道變換機制等。DRLPSP 模式的狀態選用 ETC 門架所擷取個別車道速度與入口匝道 5 分鐘平均速度，動作為採用預定號誌程序的 4 個匝道儀控秒數(入口匝道綠燈及平面道路(圓山交流道濱江街出口)紅燈秒數)；獎勵為國道主線各車道的平均速度(狀態)之權重相乘後的綜合指標。

DRLPSP 在深度學習訓練過程中，匝道號誌利用模擬環境中的偵測器所得狀態(各車道速度，S_t)，以產生對應的號誌程序(動作)與獎勵，並可以由 t 時階的動作，得到 t+1 時階的狀態(S_{t+1})，將其存放於經驗回放記憶庫。經驗回放記憶庫中隨機抽取一組(S_t ,a_t ,r_t ,S_{t+1})進入學習集合。藉由學習集合的輸出，產生深度學習與最大化學習目標進而得到損失函數誤差，模式以試誤法來選擇模擬環境當時狀態下的正確動作，圖 11 為該研究之預定號誌程序之深度強化學習流程圖。DRLPSP 模型所採用神經網路有 5 層，輸入層負責將所有狀態導入神經網路中，計 13 個輸入值；隱藏層有 3 層，各有 32 個神經元選擇 ReLU 函數為激活函數；輸出層計 9 個輸出值，代表為 9 個方案，再選取 9 個方案中機率最大者，做為代理人所要採取動作。



資料來源：趙世祺、邱俊偉、嚴國基，2021 年，深度強化學習模型應用於國道都會區重現性壅塞問題

圖 11 預定號誌程序深度強化學習

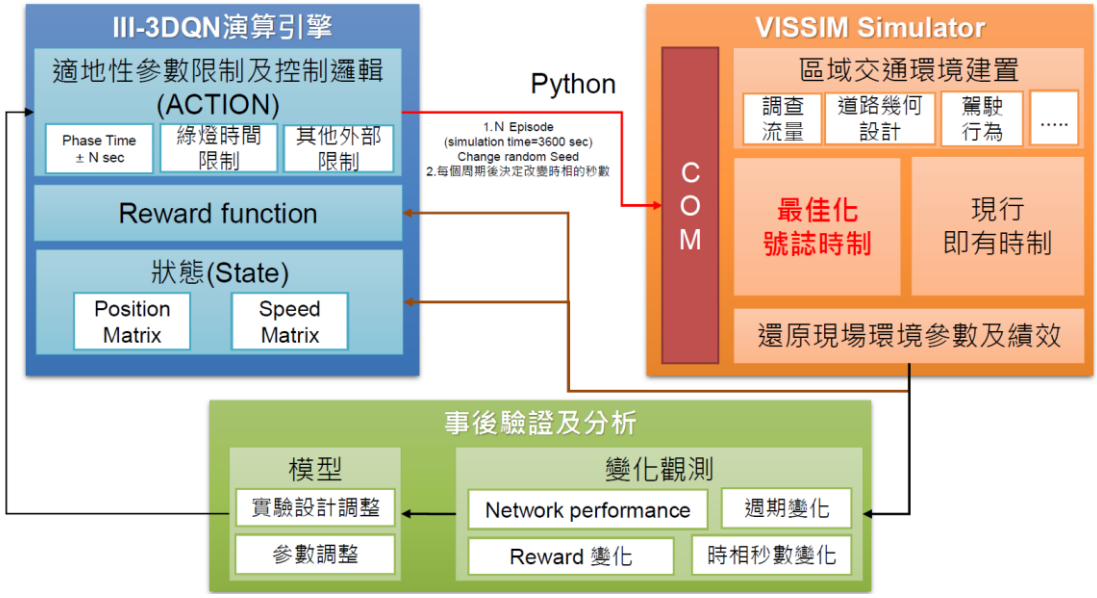
趙世祺、邱俊偉、嚴國基(2021)建議之後續研究方向，包括：(1)考量突發性狀況(事件)，然須此將面對事件不確定性及歷史數據不足等問題，對於深度強化學習模式學習訓練的挑戰；(2)納入封閉匝道、路肩開放或可變速限等策略等更多元的動作；(3)納入與平面道路號誌控制之整合運作課題等。

睿星科技(2021)研發分散式多代理人的深度強化學習之號誌控制方法。深度強化學習方法通常以交通量、等候線長度、號誌時相、綠燈時相秒數等作為狀態(State)參數，並嘗試根據當前狀態做出改善交通的動作(Action)。在多代理人強化學習中，每個代理人按順序執行動作以提高整體系統性能。訓練後，每個代理人將可學習如何彼此間相互協調，以實現整個系統的最大獎勵。

狀態包括車流狀態和號誌狀態；車流狀況包括經佈設停止線前影像偵測設備所取得之每個車道上的大型車輛、小客車與機車即時數量；號誌狀態包括每個路口的號誌時相與剩餘綠燈秒數。獎勵為研究路段的總停車延滯；每個代理人都經過訓練來取得獎勵的最小化，每個代理人的動作為下個時相的綠燈秒數。在訓練階段，我們使用微觀模擬軟體 VISSIM 來最為與代理人互動的環境。每個代理人都根據模擬器所產生的狀態-動作-獎勵樣本來進行訓練，一旦訓練過程收斂後，每

個經過學習的代理人可以協調方式執行動作，來取得整個路段的總停車延滯最小化。每個路口布設經過訓練的邊緣運算代理人，在取得影像式車輛偵測器的車流狀態後，各路口的邊緣運算代理人彼此間即時交換號誌資訊，以及進行協調式的號誌控制。該演算法已於桃園市大園交流道及大竹交流道等 2 處場域，以及高雄港洲際貨櫃中心聯外道路南星路及台 17 線沿海路段進行佈署與運作，根據實測數據顯示在停等延滯與旅行時間上均較原定時號誌控制有 13%~20%改善。

江立仁(2020)以臺北市松山路與永吉路、新生南路/和平東路、堤頂大道/樂群二路等 3 個路口構建實驗場域，發展人工智慧強化學習號誌控制，使用強化學習架構為 Double Dueling DQN (3DQN)演算法，來發展 III-3DQN 演算引擎，而該演算引擎透過 COM (Component Object Model)與 VISSIM 進行連動。VISSIM 模擬環境還原實驗場域現場車流狀態，並根據 VISSIM 即時輸出結果進行 III-3DQN 號誌最佳化的運算。III-3DQN 演算引擎根據車輛位置與速度矩陣之狀態產生改變時相時機、時相秒數增或減之號誌控制動作，利用所開發 Python 程式將每個週期決定改變後之時相秒數，經由 COM 輸入 VISSIM 模擬環境，產生模擬環境績效值(延滯時間、旅行時間、停等長度等)，再由 VISSIM 將績效值(獎勵)回饋 III-3DQN 演算引擎，再由 III-3DQN 演算引擎產生新的動作來持續學習，圖 12 為 III-3DQN 號誌控制最佳化運算邏輯。



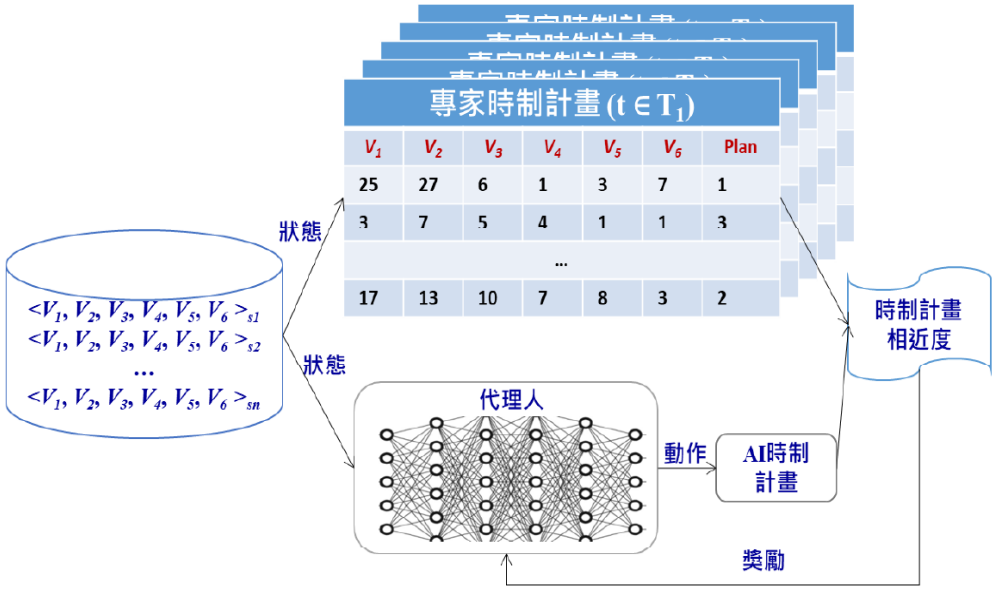
資料來源：網路資料，2020，江立仁

圖 12 III-3DQN 號誌控制最佳化運算邏輯

根據 3 個實驗場域的模擬測試結果，III-3DQN 演算引擎所產生績效均優於定時號誌控制。江立仁(2020)在後續研究發展建議上，除持續進行演算法優化與升級、及時與長期資料的導入與訓練外，在單一路口實驗場域實驗場域部分，該研究建議嘗試導入不同交通參數與進行不同類型路口之學習與測試，在區域路網方面，建議嘗試選擇壅塞路廊重要道路交織等區域進行實驗與測試。

本所(2020)發展示範型強化學習之號誌控制系統，以人工智慧進行適應性之號誌控制，該模式首先以巨觀時制最佳化分析軟體 PaSO 負責產生時制計畫，再由強化學習模式學習 PaSO 之號誌控制經驗。為評估方式之有效性，該研究以臺中市「樂業-十甲東」、「樂業-東英路」兩路口作為實驗場域，透過微觀交通模擬軟體 SUMO 進行人工智慧號誌控制之績效評估。根據模擬分析與實測結果顯示，可有效改善實驗場域之總延滯與總旅行時間。

該研究實驗設計共分兩階段進行，第一階段為單純模仿專家於時制計畫上之決策，第二階段為擴充模仿專家於尖離峰及平假日採以不同時制計畫之評斷。最終產生之模型，代理人可於實際號誌路口，依時間與交通環境，採用趨近專家時制計畫。第一階段中，隨機取樣環境中之狀態，即為六方向之車流量，查詢此狀態對應於專家應用 PaSO 所研擬之時制計畫，並同時輸入代理人之深度 Q 網路做為狀態，進而比對兩者選訂之時制計畫差異作為網路更新獎勵(與專家時制計畫之相似度)，整體流程如圖 13 所示。針對網路更新部分，在網路初始化後，代理人不斷接收輸入狀態及輸出動作，並進行與專家動作相似度比較計算獎勵，待收集到預設的狀態-動作-獎勵數量。基於預設學習率以時間分差方式更新網路，直到代理人達到指定獎勵為止。第二階段網路設計，代理人網路包含雙層結構，第一層由輸入之時間決定採用之時制計畫總週期長，第二層結合車流量與第一層之輸出以挑選在指定時間上與專家相近之時制計畫。



資料來源：吳沛儒等，109 年，示範型強化學習之人工智慧號誌控制

圖 13 AI 號誌控制示範型強化學習

六、結論與建議

相較於傳統車輛偵測器資料，來自車聯網的高頻率與低延遲車輛軌跡數據(車輛位置、速度、加速度和其他車輛動態數據的車輛狀態)，車聯網提供更為穩定和持久微觀交通資訊，相關研究顯示，車聯網所得車流資訊，具備在號誌控制最佳化的應用潛力。然而使用車聯網資料於號誌控制的最大挑戰來自於滲透率(Penetration rate)的不足，相關研究顯示在 40%滲透率情形下，車聯網號誌控制將優於現有號誌控制方式。儘管車聯網技術迅速普及，但滲透率不足的現象將持續很長時間，因此短期內車聯網高滲透率情境在實務上並不可行，雖然已有研究利用機率概念處理該問題，但實務應用可行性仍不高，因此探討在車聯網低滲透率下的號誌控制將是關鍵問題；相較之下，將人工智慧強化學習導入於號誌控制，應該是更快與容易實現的作法與趨勢。

但人工智慧號誌控制仍面臨深度強化學習模式學習效率、代理人學習過程安全性等課題，因此雖然目前深度強化學習號誌控制相關研究，其模擬結果顯示均優於現有的定時號誌控制，但多為模擬環境下的成果。目前我國產業界嘗試發展分散式多代理人的深度強化學習號誌控制模式作法，有如自動駕駛車輛的人工智慧學習方式，先在實驗室透過適當的實驗設計，利用模擬環境來訓練深度強化學習號誌控制模式，以避免發生代理人於學習過程產生不合理動作所產生的交通安

全，俟訓練學習成果達到可接受(或收斂)程度，再部署於實務環境來運作，並無法利用即時車流資訊來進行即時線上(online)學習。因此未來可能如同自動駕駛車輛方式，一段時間當車流型態改變後，仍須回到實驗室進行「再訓練」工作。

另目前號誌控制對於因事件所產生車流干擾現象的交通控制並無實務可行之模式，人工智慧的「無模式」學習方式，或可提供新的思考方向，例如：LSTM 具備存儲歷史訊息(或數據)的記憶單元，此特性可以隨著時間的推進進行探索，減少事件干擾所產生的影響，進而提高預測的準確性。

集中式或分散式的號誌控制架構課題同時存在於車聯網與人工智慧導入過程的思考課題。集中式號誌控制架構可以掌握整體路網車流狀況，來取得整體路網的最佳解，但在大型路網中，龐大的運算需求將是一項重大挑戰；此時分散式的邊緣運算，可以扮演使號誌控制架構更具彈性設計與擴展性的選項。然而分散式號誌控制架構所新增的路側設備建置與維運成本，是各縣市政府非常關心的課題，因此雖然分散式的邊緣運算存在號誌控制反映即時性高的優點，但在國內實務上的推動可能受限。

根據國內外在人工智慧車聯網相關研究案例回顧後，本研究提出下列幾點後續發展建議。

1. 在人工智慧深度強化學習號誌控制之代理人(Agent)、狀態(State)、獎勵(Reward)、動作(Action)等設計上，建議可再透過模擬環境來探討更多元組合與可能選項，例如：在特定情境下導入時相跳躍(Skip)的動作、導入多路口多代理人的分散式運作機制、嘗試幹道多路口之時差與車輛停等次數等獎勵等。

2. 目前桃園市與高雄市雖已導入人工智慧深度強化學習號誌控制於特定場域，但所經訓練的深度強化學習代理人是否具備複製應用於其他車流性質類似場域，甚或車流性質有差異場域？該課題及其代理人之「再訓練」的作法與可行性等課題應可進一步加以探討。

3. 目前人工智慧車聯網號誌控制相關研究與實作之績效評估，多以定時號誌控制為基線(Baseline)來進行比較，建議後續應可嘗試選擇已實施或即將實施傳統動態或適應性號誌控制實驗場

域，同步導入人工智慧(或含車聯網)深度強化學習號誌控制，並進行同屬動態號誌控制之績效比較。

4. 雖然因為路側設備建置與維運成本考量，目前各縣市政府多傾向採用集中式號誌控制架構，然分散式邊緣運算在號誌控制仍具有高即時性的優勢，因此建議後續可探討導入分層式架構，來結合集中式控制和分散式協同控制優點的可行作法。

5. 我國目前在車聯網相關研究並未利用車聯網所得車流資訊，來進行動態號誌控制相關研究。目前臺南市與高雄市已建置緊急車輛優先號誌通行系統，利用 DSRC 或 C-V2X 通訊技術取得緊急車輛動態資訊，搭配號誌控制之特勤控制或觸動控制方式來執行優先號誌路口通行。建議後續可進一步將緊急車輛運行之車聯網資料構建成資料庫，分析其在我國交通環境下與傳統車輛偵測器之特性差異，並透過模擬環境來發展優先號誌控制模式，甚或可結合人工智慧深度強化學習來發展具有車聯網資料特性之人工智慧優先號誌控制模式。

6. 目前國內外在人工智慧車聯網相關研究多著重於都市單一路口或幹道，然高速公路交流道區域往往是我國都市交通的瓶頸，因此後續可參考整合式路廊交通管理概念，將都會區的高速公路上匝道儀控與平面道路號誌控制整合運作納為人工智慧深度強化學習的研究課題。

7. 突發性狀況(或事件)具有不確定性及稀少性的歷史數據不足等問題，對於深度強化學習模式學習訓練的造成挑戰，但 LSTM 具備存儲歷史訊息(或數據)的記憶單元特性，提供可能的處理方式，建議後續亦可將此課題進行探討。

參考文獻

1. 趙世祺、邱俊偉、嚴國基，深度強化學習模型應用於國道都會區重現性壅塞問題，2021 年中華民國運輸學會學術論文研討會
2. 胡大瀛、李卓育，深度強化學習下號誌控制設計獎勵機制之探討，2021 年中華民國運輸學會學術論文研討會
3. 許添本、蔡沐軒，示範式深度強化學習應用於號誌時制最佳化之研究，2021 年中華民國運輸學會學術論文研討

會
4. 桃園市政府交通局，110 年，AI 智慧號誌控制系統，2021 年中華智慧運輸協會「智慧運輸應用獎」申請書
5. 睿星科技股份有限公司，110 年，AI 智慧號誌控制方法，2021 年中華智慧運輸協會「智慧運輸產業創新獎」申請書
6. 許添本、程楷祐，以深度強化學習方式建構混合車流之 AI 最佳化號誌時制計畫，2020 年中華民國運輸學會學術論文研討會
7. 吳沛儒等，109 年，示範型強化學習之人工智慧號誌控制，2020 年中華民國運輸學會學術論文研討會
8. 高雄市政府交通局，高雄市運用 AI 智慧號誌紓解交通瓶頸計畫獲選 IDC 2022 年亞太智慧城市獎(SCAPA)，https://www.tbkc.gov.tw/Message/Bulletin/News?ID=ba8a0620-02b4-4d22-a6c6-254d161f49bf
9. Hua Wei 等，2020，A Survey on Traffic Signal Control Methods
10. Martin Greguri 等，2020，Application of Deep Reinforcement Learning in Traffic Signal Control: An Overview and Impact of Open Traffic Data，Applied Science
11. Wan Li and Xuegang Ban，2020，Connected Vehicle-Based Traffic Signal Coordination, Engineering, Volume 6, Issue 12, December 2020, Pages 1463-1472
12. Faizan Rasheed, etc, 2020, Deep Reinforcement Learning for Traffic Signal Control: A Review, IEEE Access
13. Qiangqiang Guoa, etc., 2019, Urban traffic Signal Control with Connected and Automated Vehicles: A Survey, Transportation Research Part C, Volume 101, April 2019
14. Guanjie Zheng 等，2019，Learning Phase Competition for Traffic Signal Control，CIKM ’19, November 3–7, 2019, Beijing, China
15. Xuegang Ban and Wan Li，2018，Connected Vehicle Based Traffic Signal Optimization - C2Smart
16. Kok-Lim Alvin Yau 等，2017，A Survey on Reinforcement

Learning Models and Algorithms for Traffic Signal Control，ACM Computing Surveys, Vol. 50, No. 3, Article 34
17. Zheng 等, 2017, Estimating traffic volumes for signalized intersections using connected vehicle data, Transportation Research Part C
18. Rasheed 等, 2020, Deep Reinforcement Learning for Traffic Signal Control: A Review, IEEE Access