Online Incremental Machine Learning Platform for Big Data-Driven Smart Traffic Management

Dinithi Nallaperuma, Rashmika Nawaratne, Tharindu Bandaragoda, Achini Adikari, Su Nguyen, *Member, IEEE*, Thimal Kempitiya, Daswin De Silva, *Member, IEEE*, Damminda Alahakoon, *Member, IEEE*, and Dakshan Pothuhera

***Abstract*—The technological landscape of intelligent transport systems (ITS) has been radically transformed by the emergence of the big data streams generated by the Internet of Things (IoT), smart sensors, surveillance feeds, social media, as well as growing infrastructure needs. It is timely and pertinent that ITS harness the potential of an artificial intelligence (AI) to develop the big data-driven smart traffic management solutions for effective decision-making. The existing AI techniques that function in isolation exhibit clear limitations in developing a comprehensive platform due to the dynamicity of big data streams, high-frequency unlabeled data generation from the heterogeneous data sources, and volatility of traffic conditions. In this paper, we propose an expansive smart traffic management platform (STMP) based on the unsupervised online incremental machine learning, deep learning, and deep reinforcement learning to address these limitations. The STMP integrates the heterogeneous big data streams, such as the IoT, smart sensors, and social media, to detect concept drifts, distinguish between the recurrent and non-recurrent traffic events, and impact propagation, traffic flow forecasting, commuter sentiment analysis, and optimized traffic control decisions. The platform is successfully demonstrated on 190 million records of smart sensor network traffic data generated by 545,851 commuters and corresponding social media data on the arterial road network of Victoria, Australia.**

***Index Terms*—Smart traffic management, concept drift, unsupervised incremental learning, deep learning, deep reinforcement learning, impact propagation, traffic optimization, traffic forecasting, traffic control, social media analytics.**

I. Introduction

ROAD traffic conditions and flow management continue to be an important area of research with many practical implications. During the last decade, the technological landscape of transportation has gradually integrated disruptive technology paradigms into current transportation management systems, leading to Intelligent Transportation Systems (ITS) [1], [2]. The emergence of Internet of Things (IoT), sensor networks and social media has surpassed traditional means of collecting data, by creating voluminous and continuous streams of real-time data. Leveraging such big data environments is a formidable issue, due to the intense volume and velocity at which data is generated by transportation and mobility systems [1]. Furthermore, the dynamic nature of these environments makes the data generation volatile, which impedes the effectiveness of decision-making in ITS.

Manuscript received July 15, 2018; revised November 7, 2018 and April 8, 2019; accepted May 31, 2019. Date of publication July 11, 2019; date of current version December 23, 2019. This work was supported in part by the Australian Government Research Training Program Scholarship and in part by the Data to Decisions Cooperative Research Centre (D2D CRC) through the Analytics and Decision Support Program. The Associate Editor for this paper was J. Sanchez-Medina. *(Corresponding author: Tharindu Bandaragoda.)*

D. Nallaperuma, R. Nawaratne, T. Bandaragoda, A. Adikari, S. Nguyen, T. Kempitiya, D. De Silva, and D. Alahakoon are with the Research Centre for Data Analytics and Cognition, La Trobe Business School, Bundoora, VIC 3083, Australia (e-mail: t.bandaragoda@latrobe.edu.au).

D. Pothuhera is with the Department of Information Management and Technology, VicRoads, Kew, VIC 3101, Australia.

Digital Object Identifier 10.1109/TITS.2019.2924883

The dynamicity of data generated by transportation systems consists of continuously changing patterns and concept drifts. In a traffic context, concept drifts are the changes to the distributions of data in a traffic data stream over time [3]. Based on the nature of fluctuations in data streams, these changes are further classified as recurrent and non-recurrent concept drifts. For example, traffic congestion changes due to peak/off-peak traffic is a recurrent concept drift whereas an accident or breakdown is a non-recurrent concept drift. Special importance should be placed into identifying non-recurrent concept drifts as it could affect the entire road network. Existing literature reports a number of supervised machine learning algorithms that detect drifts and adapt to new concepts [3]–[6]. Although real-time concept drift detection is crucial for effective decision making in transportation, feedback on the type of traffic incident is only received following an unknown delay. This severely limits the applicability of the supervised learning nature of these algorithms. Therefore, we postulate that concept drift detection in road traffic requires unsupervised online incremental machine learning to address the challenges of real-time, unlabeled, volatile data streams.

In this work, we distinguish online learning and incremental learning. Online learning updates the model using each incoming data point that arrives during the operation, without storing [7]. As such, online learning is utilized to handle large volumes of streaming data arriving at high velocity. Incremental learning is learning from batches of data at distinct time intervals, and has the capability to stabilize the historical knowledge of the learning model over novel learnings [8]. Hence, the model is updated to any new data point that is received while keeping its existing knowledge intact. Further, it is essential that non-recurrent concept drifts are identified and utilized for updated traffic propagation and traffic flow prediction models in a real-time manner.

To this end, we further address several key concerns which are underexplored in current ITS, to support the development of a holistic traffic management platform. Following an extensive review of current literature in ITS [1], [2], [9] we identified the following four current challenges which have not been sufficiently addressed. a) the development of real-time machine learning algorithms and prediction schemes for non-recurrent traffic incidents that impact an entire road network. b) a majority of existing approaches focus on freeways and highways, with very limited attention to arterial networks due to the technical challenges of integrating multiple streams of traffic data, thus fails when determining traffic propagation in the entire network. c) current approaches do not account for network level spatiotemporal variables that are expressed as big data streams. d) the human element of road traffic, commuter sentiment and emotions expressed regarding traffic on social media channels, which are largely overlooked in current ITS research. While social media is increasingly being used in emergency events [10]–[12], integrating such data with other traffic-related data would provide a holistic view of the situation from both road dynamics and commuter perspective.

Addressing these limitations, we designed and developed the 'Smart Traffic Management Platform' (STMP), an expansive intelligent traffic data integration and analysis platform, that integrates heterogeneous traffic data sources. With the proposed STMP platform we present the following research contributions.

- a novel online incremental machine learning algorithm to detect real-time concept drifts from big data streams
- a deep learning approach for real-time network level traffic flow prediction and impact propagation estimation in arterial road networks
- a deep reinforcement learning approach to determine optimal traffic control actions based on real-time measurements
- a social media data integration model to capture social behaviors during a non-recurrent traffic event, to determine commuter sentiment and emotion
- demonstrated the STMP platform on 190 million records of smart sensor network traffic data generated by 545,851 commuters and corresponding social media data on the arterial road network of the State of Victoria in Australia.

Rest of the paper is organized as follows. The next section delineates the proposed STMP platform, followed by subsections for each research contribution. STMP is demonstrated in Section III, and Section IV concludes the paper following a discussion on implications and potential for future research.

II. Proposed Platform

The proposed STMP platform is illustrated in Fig. 1. It consists of three layers of functionality, L1-L3. It is expansive in terms of the heterogeneous data sources that can be transformed and integrated into the data transformation layer L1 (Section II-A). The middle layer L2 (Section II-B) consists of the online incremental machine learning algorithm which detects concept drifts and classifies into recurrent/non-recurrent traffic events. These are passed on as input to smart traffic management modules in layer L3 for impact

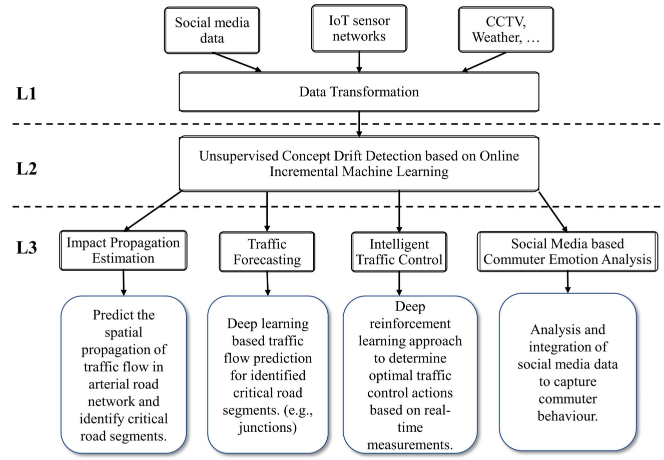

Fig. 1. Smart traffic management platform architecture.

propagation estimation (Section II-C), traffic flow forecasting (Section II-D), optimization for intelligent traffic control (Section II-E) and commuter emotion analysis (Section II-F). L3 is also expansive as further modules can be linked to L2 for other traffic management activities.

A. Data Transformation

The data transformation layer receives heterogeneous sources of big data streams related to road traffic, such as IoT, sensor network data, social media data, video surveillance feeds, weather data, planned public events, and road construction activities. These data sources are pre-processed, transformed and integrated into a computational format that can be effectively ingested by the online incremental machine learning algorithm in L2. Due to space limitations, we scope this study to trajectory data from Bluetooth Traffic Monitoring System (BTMS) and social media data.

1) Traffic Flow Modeling: The flow of traffic in a road network fluctuates due to recurrent events (e.g., peak, off-peak traffic), non-recurrent events (e.g., accidents, road work), and the characteristics of the location (e.g., schools, shopping centers). Therefore, the traffic flow needs to be modeled as a spatiotemporal function.

The traffic flow $T(t, s)$ at location s and time t are a directional measure determined separately for each direction of traffic at s. This is based on the number of vehicles that pass through location s towards a particular direction at time interval t to $t + \Delta t$ [13]. In legacy approaches, traffic flow is measured from several reference points and then extrapolated to determine the traffic flow of the road network based on the density flow theories of hydrodynamics [13]–[16]. However, as the arrival of more comprehensive traffic data collections systems, fully data-driven methods have been recently proposed to estimate the traffic flow [17]–[19].

In addition to point-based traffic flow, the availability of vehicle trajectory data enables the traffic flow to be estimated for road segments (between any two sensor locations). For example, the traffic flow of road segment AB can be determined by $T(t, A \rightarrow B)$ and $T(t, B \rightarrow A)$, which denotes directional traffic flow at time t from A to B and B to A respectively.

2) Social Media Data Modeling: Social media data are user-generated content streams with faster information dissemination. A social media stream can be regarded as a dynamic content stream which is a continuous stream of messages over time $(m_1, m_2, \ldots m_n)$ where each message consists of a short text and a time-stamp. As the first step, the social media stream is sampled using a fixed time interval (Δt) which produces batches of social media messages $(B_1, B_2, \ldots B_i)$ where each batch is a collection of social media messages collected for time Δt. Thus, for a given batch, the message collection for Δt could be defined as,

$$B^i = \{m\}_{t=i\times\Delta t}^{t=(i+1)\times\Delta t} \tag{1}$$

The granularity of the time period Δt (e.g., hourly, daily, weekly) can be customized to suit the velocity of the social media stream as well as to match other application specific requirements. Each batch is then taken for pre-processing. During the pre-processing tasks, duplicated values and stop words are removed. Natural Language Processing (NLP) and regular expressions are used to remove URLs, symbols, and emojis. In addition, geolocation-based filtering is carried out to filter messages relevant to the context being studied. Subsequent to the aforesaid transformation, the data are utilized for the analysis.

B. Unsupervised Concept Drift Detection Based on Online Incremental Machine Learning

Following the data transformation process, feature vectors that are mostly unstructured with unlabeled target variables are passed on to L2. In L2, a novel online incremental machine learning algorithm is proposed for real-time concept drift detection and adaptation (Fig. 2). The underlying base algorithm has been applied to examine context awareness in the Aarhus city of Denmark motor traffic dataset [20] and to study how the driver behavior change can affect the coordination between autonomous and human-driven vehicles [21], and the algorithm has been extended in this work for recurrent and non-recurrent concept drift capture. The algorithm consists of two forms of learning, online and offline. Online learning addresses volume and velocity constraint. Data-driven triggers are used to define the processing time window and level of abstraction. The offline algorithm consists of two learning features:

- Incremental learning to learn from evolving new concepts, which effectively addresses both time and space constraints. As the learning assumes the new incoming data are similar to previously learnt concepts, incremental learning can be used to detect drift between the new concept and old concept.
- Decremental learning to forget the concepts that are not further relevant which allows the algorithm to adapt to the new concept.

1) Online Learning: Online learning is handled by an online adaptive clustering algorithm [22] where the model is updated as the new data are presented. The online adaptive clustering algorithm periodically transfers an aggregated knowledge to the offline learning algorithm. Transfer time window and initial centroids for clustering are defined by the processing triggers. The transfer time window is the time taken by the offline algorithm to learn. Learning time is shorter when the algorithm learns a previously learnt concept whereas it takes longer time to learn a new concept. Weight vectors of the initial centroids for clustering are assigned from the offline learning and this allows the algorithm to quickly adapt to known concepts.

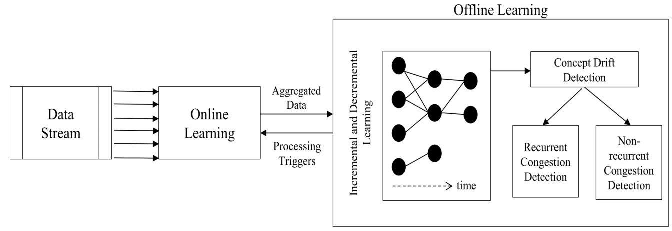

Fig. 2. Proposed data driven unsupervised concept drift detection algorithm.

2) Incremental Learning: Incremental learning is based on Incremental Knowledge Acquisition and Self Learning (IKASL) algorithm [23]. The algorithm continues to learn from new data based on learning and generalization of the past learning outcomes. The functionality of the learning layer is based on Growing Self Organizing Maps (GSOM) [24] that generates a dynamic feature map based on the growing self-organizing process. For each output from the online processing, Euclidean measurement is calculated to find the closest node in the map. The winning node's weight error rate is increased and the weight vector is updated using Eq. 2, where $w_j(t), w_j(t+1)$ are the weight vectors of node j before and after adaptation, and N_{n*} is the neighborhood of the winning node.

$$w_j(t+1) = \begin{cases} w_j(t), & \\ \qquad j \notin N_{n*}(t+1) & \\ w_j(t) + LR(t)\left(x_k - w_j(t)\right), & \\ \qquad j \in N_{n*}(t+1) & \end{cases} \tag{2}$$

Node growth occurs when this error value exceeds a predefined growth threshold, GT. New nodes grow out of the node with the highest accumulated error (N_e) and are initialized to reflect the neighborhood of N_e. Upon learning for a predefined number of epochs, a calibrating phase smooths out irregularities in recent weight adaptations.

In the dynamic feature map, weight vectors of the nodes and its neighborhood represent the current knowledge of the concepts. Hence the proximity matrix, S is calculated where s_{km} (m-th node of the k-th neighborhood) represent the proximity of $n_k(H_i)_m$ corresponding to hit node, H_i. Use of proximity matrix ensure the furthest neighborhood node will have the strongest impact on the aggregate weight vector. All such aggregated nodes form the x-th generalized layer G_x. Each node in the generalization layer has the potential to grow into a feature map. Each subsequent learning phase, L_{x+1}, is started with a winning node from the generalization layer.

This further learning enables incremental learning allowing the algorithm to preserve knowledge.

3) Decremental Learning: In the implementation of decremental learning, generalization nodes that do not become a winner for any of the inputs are identified. These nodes represent knowledge that is no longer relevant, hence will be forgotten from the map. This allows the algorithm to adapt to the new concepts more efficiently.

4) Concept Drift Detection: In concept drift detection, the distance (d) between the generalized nodes in subsequent learning layers (GN^x) is calculated as,

$$\mathrm{d(x)} = \sqrt{\sum_{i=1}^{D}\left(GN_i^x - GN_i^{x-1}\right)^2} \tag{3}$$

where D is the number of dimensions of the input and x is the learning phase (layer). The distance $d(x)$ represents the new knowledge acquired by GN^x. Hence, a concept drift (CD) in layer x can be identified as a concept drift when d(x) is a local maximum.

Among CDs, non-recurrent CDs result in relatively higher knowledge acquisition as they were not learned before, in contrast, recurrent CDs results in lower knowledge acquisition as they were being learned before. Hence, in a non-recurrent CD, $d(x)$ should be significantly higher. To detect a non-recurrent concept drift, the prominence of the distance change is calculated. A non-recurrent concept drift can be identified when there is a significant difference in distance.

Above mentions characteristics of CDs are being captured by f_{CD} in Eq. 4 to identify whether a layer x represent a non-recurrent CD, recurrent CD or neither.

$$f_{CD}(x) = \begin{cases} \text{non-recurrent CD} & \text{if } d(x) > d(x-1) \text{ and,} \\ & d(x) > d(x+1) \text{ and,} \\ & |d(x) - d(x+1)| > \dfrac{\sum_{j=1}^{x} d(j)}{x} \\ \text{recurrent CD} & \text{if } d(x) > d(x-1) \text{ and,} \\ & d(x) > d(x+1) \\ \text{no CD} & \text{otherwise} \end{cases} \tag{4}$$

C. Impact Propagation Estimation

The previously described layer L2 captures concept drifts due to both recurrent and non-recurrent traffic incidents. Such incidents not only impact the location of occurrence but propagates through the road network. While the impact propagation of recurrent incidents is known and accounted in traffic planning, the impact prediction on non-recurrent incidents in near-real times is crucial to remediate its impact. Therefore, it is important to predict the impacted road segment and the proportion of impact propagated.

Due to the lack of historical traffic data, such impact predictions are often carried-out using mathematical model based simulations [25], which are often suboptimal as real-world traffic conditions deviate from the underlying assumptions. In contrast, there have been recent data-driven supervised learning approaches which learn impact propagation models from road occupancy data from induction loop networks and incident data from historic incident reports [26]. However, such an approach can only be accurately applied to locations with a substantial amount of historic incident data.

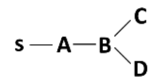

Fig. 3. Sample road network.

This work proposes an unsupervised data-driven approach to predict incident impact propagation across the arterial road network based on historical data on vehicle trajectories. It is based on the hypothesis that when the time and location of the incident is s and t, the propagated impact of that incident to a location S_i is proportionate to the flow of traffic from S_i to s at time t i.e., $T(t, S_i \to s)$. It assumes that when there is an incident at s, it is going to block/slow down the traffic that flows $S_i \to s$, thus impacts the traffic on S_i.

Note that, since the availability of the historical vehicle trajectory data, traffic flow between two points that are even separated by several other junctions can be accurately estimated.

The *relative incident impact* to the location S_i is determined based on the traffic flow $S_i \to s$ relative to the aggregated traffic flow goes through S at time t.

Let the n road segments connecting to S be $\{s_1 \to s, \ldots, s_n \to s\}$, the *relative incident impact* $P_I(t, S_i \to s)$ to the location S_i from an incident in s at time t is defined as,

$$P_I\,(t, S_i \to s) = \frac{T\,(t, S_i \to s)}{\sum_{j=1}^{n} T\left(t, S_j \to s\right)} \tag{5}$$

The *relative incident impact* can be determined across the road network for the reference points (e.g., junctions) that are within a designated distance from the incident location. For example, let A, B, C, and D be the junctions of the road-network that connects to the incident location s as illustrated in Fig. 3, then the *relative incident impact*on each junction will be $P_I(t, A \to s)$, $P_I(t, B \to s)$, $P_I(t, C \to s)$, and $P_I(t, D \to s)$.

The *relative incident impact* decreases as the junctions are further from the incident location. Junctions with a significant *relative incident impact* as well as the road network that connects those junctions can be identified as the impacted critical road segments of the network. Note that, an important advantage of this approach is that, with the availability of historical data, it takes into account the temporality of the traffic flow when determining the *relative incident impact*, as the impacted road segments may vary depending on the usual traffic conditions at the time t of the incident.

D. Traffic Forecasting

Critical road segments pose highly unpredictable traffic conditions and they can be determined using impact propagation estimation. Providing suitable traffic control mechanisms for such critical road segments is an important consideration in ITS to enable a smooth traffic flow over the arterial road network. In order to develop a suitable control mechanism, it is essential to be able to forecast the traffic flow of the critical road segments. In this section, we propose an approach for traffic prediction based on the current traffic flow condition of the surrounding using a Deep Neural Networks (DNN) model.

Previous efforts in road traffic forecasting range from traffic flow modeling to recent advancements in data-driven models. With the recent advancement in deep learning paradigm, the ITS research community has made efforts to extend the ITS capabilities with deep learned modeling [17], [27], [28]. In the proposed STMP platform, we introduce a new DNN based on LSTM [29] and enrich its prediction capabilities by incorporating the surrounding traffic flow of influential road segments.

Standard DNN architectures based on LSTM attempt to predict time series data singularly considering a single time scale, where time series data such as road traffic flow often have a temporal hierarchy with information spread out over multiple time scales. With the introduction of cascade-connected layers of LSTM networks, the abstraction of input observations over multiple time scales can be incrementally extracted to incorporate in modeling the input pattern [30], [31]. Therefore, due to the complex and stochastic nature of the road traffic, it is advantageous to model complex dependencies between the time series input from the traffic flow. This has been recently attempted in [32], where correlations of the traffic flow of targeted road segment and that of its upstream and downstream road segments are considered for the prediction. However, it is imperative to capture the spatial correlation of the arterial road network including the connected road segments that affect the traffic flow of the targeted road segment.

We designed a DNN based on hierarchically stacked three layered LSTM network architecture that consists of 256, 128 and 64 LSTM cells respectively, accommodating deeper abstraction of temporal hierarchy to be trained by the prediction model. The proposed DNN architecture captures the correlation patterns of the targeted road segment and the surrounding road segments that impact the traffic flow of the targeted road segment, which was identified through the impact propagation analysis. This proposed DNN architecture has enabled the prediction model to incorporate longer duration of current and previous traffic condition in order to forecast the future traffic condition.

E. Intelligent Traffic Control

Real-time concept drift detection and traffic forecasting are useful inputs for intelligent traffic control to optimize network performance. Conventional control approaches to intelligent traffic control such as static feedback control (SFC) and optimal control and model predictive control (MPC) are developed based on many assumptions and idealistic models [33]. As a result, these approaches have trouble coping with the dynamics of the traffic networks. Some traditional AI techniques such as case-based reasoning and rule-based systems are used to determine control actions based on recorded similar situations (from historical data or off-line simulation) or based on simple “if-then” rules. However, these techniques do not have a learning mechanism to automatically update their model. Also, they do not have a strong inference power to deal with unseen situations.

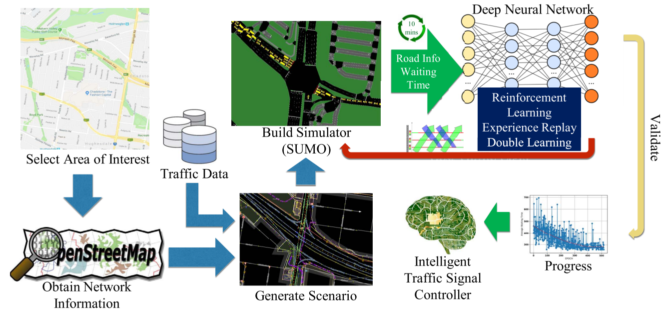

Fig. 4. Deep reinforcement learning for adaptive traffic control.

DNN combined with reinforcement learning usually referred to as Deep Reinforcement Learning (DRL) [34], is a generic and flexible way to develop intelligent and adaptive traffic control systems. Fig. 4 shows a DRL method for intelligent traffic control in the proposed STMP. The goal of DRL is to select the most suitable control program which decides the duration for each time phase for each traffic light in the network. In this method, the area of interest is first selected and the corresponding road infrastructure of this area is obtained using the data from OpenStreetMap (OSM) [35]. Because it is very costly to test the algorithm on real environments, a virtual environment is developed (via simulation) as a mean to validate the effectiveness of the evolved controller. In our implementation, we feed the road information and the traffic data into TraCI-SUMO [36] to generate simulation scenarios. Each simulation run is for two hours and the information (s, a, r, s') of the current state (s), action (a), reward (r), and new state (s') are recorded every 10 minutes in which $s \xrightarrow{a} r, s'$.

The inputs for the DRL algorithm are the numbers of vehicles traveling through all lanes in the network (each road segment can have many lanes). Based on the inputs, DRL will calculate the expected values for each action (e.g., a control program) and the action with the highest expected values will be selected and applied to the traffic network. The proposed DRL method has two fully connected hidden layers. The two hidden layers use Rectified Linear Unit (ReLU) as activators. The dimension of the output layer is the number of control programs M (i.e the action $a \in \{C_1, \ldots, C_M)$ that we want to select and linear activators are employed. At the beginning of each time interval, the most suitable control program is selected by,

$$a_t = \begin{cases} C_r & rand < \epsilon \\ argmax_a Q(s', a) & otherwise \end{cases} \tag{6}$$

where $rand$ is a random number from 0 to 1, ϵ is exploration factor which is decayed over time, C_r is the random control program. In the proposed DRL algorithm, we apply Q-learning in which DNN is used to represent the Q function [39]. To improve the stability of Q function in the training process,

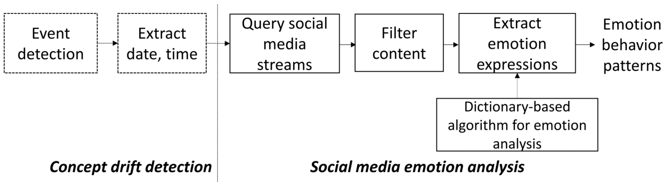

Fig. 5. High-level process of the proposed emotion analysis during an event.

we use the double learning and experience replay:

$$Q(s,a) = r + \gamma\, \tilde{Q}\left(s', argmax_a Q\left(s', a\right)\right) \tag{7}$$

where $Q(s, a)$ is the primary Q network, $\tilde{Q}(s, a)$ is the target network, and γ is the discount factor [50]. This trick has been shown to significantly improve the effectiveness of Deep Q-learning.

The reward obtained for each time interval (e.g., 10 minutes) is calculated as,

$$r_t = \begin{cases} W_t - W_{t-1} & if\ (W_t - W_{t-1}) < 0 \\ -P * (W_t - W_{t-1}) & otherwise \end{cases} \tag{8}$$

where W_t is the average waiting of vehicles traveling on the traffic network in the time interval t, and P is the penalty for slowing down the traffic. Based on the recorded states, actions, and rewards, DRL will update its parameters (i.e., connection weights) to optimize its decisions by using back-propagation. In this method, the traffic controller can be triggered via time-based (e.g., every five minutes) or event-based mechanisms (e.g., by anomaly detection).

F. Social Media Based Commuter Emotion Analysis

In the proposed architecture, once an event is detected using the concept drift module (Section II-B), the emotion analysis is used to capture the emotional behaviors of commuters from the live social media stream. The importance and the applicability of social media in transport sector have been stated in [38] where it mentions that social media data from commuters potentially enrich other data sources by providing a different view of the commuters and their behaviors which are not captured from other traffic data sources. This analysis is conducted as a supporting module to investigate commuter behaviors during detected events, as social media platforms such as Twitter are widely used communication platforms to notify about significant events such as accidents, traffic congestions and roadblocks [10]–[12] and in several studies, sentiment analysis using Twitter has been carried out as a measure of commuter satisfaction and to observe commuter behavior and opinion patterns in order to improve the transportation services [39], [40]. This work extends the traditional sentiment analysis by extracting deeper emotion from twitter, thus enabling a more granular analysis of user opinion and feeling. Fig. 5 is a high-level illustration of the process.

Once an event is identified by the concept drift detection, the Twitter data stream is analyzed to collect tweets that are relevant to the event. Such tweets are defined as originating within a radius r_e of the event for a time interval t_e since the event. Twitter API is queried with r_e and t_e to collect the relevant *georeferenced* tweets of the event. Note that, there are advanced methods of localizing the non-georeferenced tweets [41], however, such approaches are beyond the scope of this work. The radius r_e is set based on the impact propagation analysis of the accident by drawing a bounding circle covering the road segments with significant relative incident impact.

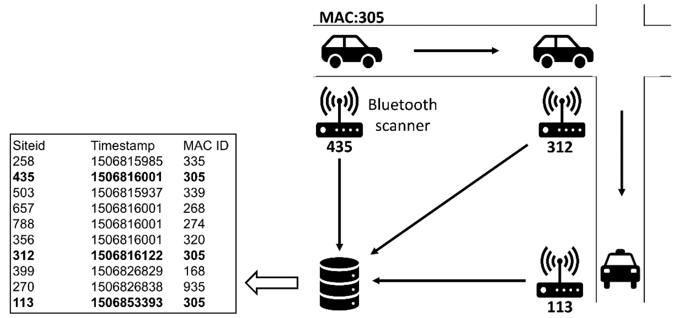

Fig. 6. Bluetooth traffic monitoring system (BTMS).

The emotion extraction is based on Ekman's model of emotions where it presents six basic emotions ("Joy", "Anger", "Surprise", "Disgust", "Fear", "Sadness") which are known as universal emotions [42]. An emotion vocabulary was formed by creating a dictionary of emotional expression per each emotion category. Emotion expressions were obtained from the LIWC dictionary [43] and other published research studies [44]. Text mining and NLP techniques were used to extract the emotion expressions from social media contents. Having identified the emotions of a particular tweet, the weight of each emotion (w_e) was calculated by taking the average of emotional expressions present in each tweet in order to determine the strength of the emotions expressed. Afterwards, the accumulated emotion intensity level was calculated and used to determine if the emotional behavior was significant. Emotion intensity (I_t) can be defined based on the aggregation of emotions over a specified time period, commencing at time t, where n denotes the number of tweets and m denotes the number of emotions in the category, as expressed below.

$$I_{t,t+\Delta t} = \sum_{i=0}^{n}\sum_{j=0}^{m} w_{ij} \tag{9}$$

III. Experiments

This section demonstrates the proposed STMP platform using real traffic data from the arterial road network of the State of Victoria, Australia.

The information on traffic has been acquired from the Bluetooth Traffic Monitoring System (BTMS) that is used to monitor the road traffic of arterial roads in Victoria. BTMS is a type of automatic vehicle detectors that is used to estimate travel times in a road network [45], [46]. For a comprehensive understanding of the Bluetooth sites refer [47].

As shown in Fig. 6, BTMS consists of a network of Bluetooth traffic scanners that are placed in the junctions of arterial roads. These Bluetooth scanners capture the Bluetooth devices that transit the scanning zone, which are either Bluetooth enabled vehicle stereo systems or the mobile devices

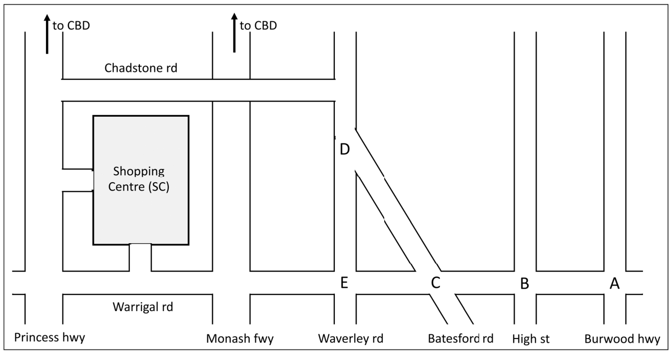

Fig. 7. Schematic of the road network in the area of interest selected for traffic analysis.

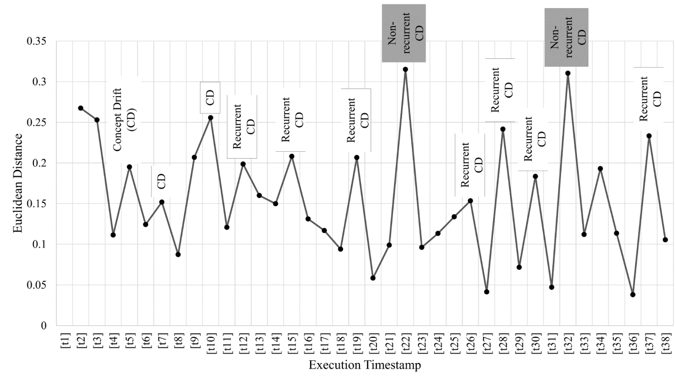

Fig. 8. Recurrent and non-recurrent concept drift detection.

of the occupants. The scanners capture the unique electronic identifier (MAC address) of the Bluetooth devices that transit the scanning area and the transit timestamp [46]. Each Bluetooth scanner periodically transmits those records to a central database. Since the electronic identifier is unique to each Bluetooth devices, its travel path can be traced across the network of Bluetooth scanners placed in the road network.

For this study, the dataset was obtained from Victoria road authority (VicRoads) and comprises all vehicle records for October 2017. This dataset consists of approximately 190 million vehicle records obtained from 1,408 Bluetooth scanners placed at the junctions of arterial roads. It contains records from 545,851 unique MAC-IDs, which is assumed to be unique vehicles.

The capability of the proposed platform is demonstrated by analyzing the traffic behavior around a key Shopping Centre (SC), which is often a volatile traffic region in Victoria. It is one of the largest stand-alone shopping centers in Australia with over 20 million annual visitor turnarounds. Thus, it accounts for a large traffic footprint in surrounding arterial roads. In addition, it is sandwiched between two large freeways (Princess Hwy and Monash Fwy) which are the key freeways that connect Melbourne Metropolitan Area to Southeast Victorian suburbs. The combination of the above-mentioned factors yields unique and highly volatile traffic patterns in the selected region. Moreover, the surrounding arterial roads are often operated in near saturation, thus, any non-recurrent incident can result in significant congestion and its impact often propagated significantly across the arterial road network. Fig. 7 presents a schematic of the selected SC and the surrounding arterial roads.

A. Traffic Data Transformation

Each record in the dataset $D = \{(\upsilon, s, t)\}$ can be denoted by (υ, s, t) where υ is the vehicle denoted by the MAC-ID, s is the location denoted by the site id and t is the time denoted by the timestamp. The traffic flow of road segments was derived from these traffic records.

Based on the definition of traffic flow in Section II-A, the traffic flow $T(t, A \rightarrow B)$ of road segment AB can be defined as the number of vehicles that are first detected in A at time t and subsequently detected in B, which can be denoted as,

$$T(t, A \rightarrow B) = \int_{t}^{t+\Delta t} \sum_{\forall \upsilon} \mathrm{I}(\upsilon, A \rightarrow B, t)dt \tag{10}$$

where Δt is the sampling interval for the traffic flow which can be adjusted to obtain the required granularity of the traffic flow. $I(\upsilon, A \rightarrow B, t)$ is an indicator function which is active if the vehicle υ is first detected at A at time t and subsequently detected at B within a time threshold τ. It can be defined as,

$$\begin{aligned} &\mathrm{I}(\upsilon, A \rightarrow B, t) \\ &= \begin{cases} 1 & \text{if } \exists\ (\upsilon, A, t) \in D \text{ and } \exists\ (\upsilon, B, t') \in D, \\ & \text{where } t < t' \leq t + \tau \\ \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{11}$$

where τ is the trip threshold which is set enough for a single trip in the road segment. The idea of setting this threshold is to filter out noisy trips [48] such as pedestrians as well as vehicles that make a stop inside the segment.

B. Unsupervised Concept Drift Detection

Transformed data were emulated as a real data stream to represent VicRoads live feed using an external application. This data stream is presented to the unsupervised concept drift detection algorithm explained in Section II-B. Fig. 8 illustrates the concept drifts detected in the selected traffic area. The x-axis denotes execution timestamps of incremental learning and distance measure calculations from Eq. 3 of the algorithm are denoted on the y-axis. Fig. 8 demonstrated the recurrent and non-recurrent concept drifts identified by Eq. 4 of the algorithm.

The algorithm identified three recurrent concept changes throughout the data stream. These recurrent concept changes relate to the traffic flow changes due to longer shopping hours, weekdays to weekend traffic flow and vice versa. A summary of the recurrent traffic flow changes is denoted in Table I.

Moreover, the algorithm identified two non-recurrent concept drifts at execution timestamps [t22] and [t32], of which the concept drift at [t32] is employed to demonstrate the functionality of L3 of the proposed STMP platform.

TABLE I
SUMMARY OF RECURRENT CONCEPT DRIFT DETECTION

Traffic flow change	*EXECUTION TIMESTAMP*	*Explanation*
Wednesday -> Thursday	[t5]: 4-5/10/2017 [t12]: 11-12/10/2017 [t19]: 18-19/10/2017 [t30]: 25-26/10/2017	Traffic is affected by the longer shopping hours on Thursday.
Friday -> Saturday	[t7], [t26], [t34]	The area is a central suburb where most of the traffic would cross while travelling from outer suburbs to city. Traffic will reduce on Saturday compared to Friday, as Saturday is a holiday.
Sunday -> Monday	[t10], [t15], [t28], [t37]	The area is a central suburb where most of the traffic would cross while travelling from outer suburbs to city. Traffic will increase on Monday compared to Sunday as most people work in the city.

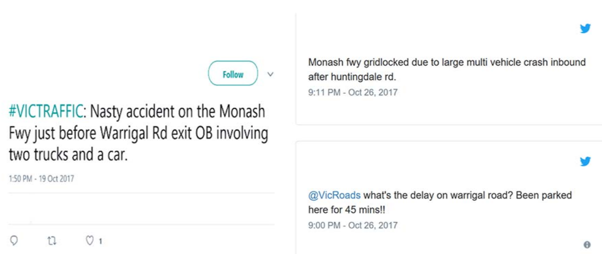

Fig. 9. Evidence for non-recurrent concept drift detection, left: usual traffic congestion is reduced in Warrigal Rd due to an accident in Monash Fwy, right: accident on Warrigal Rd which has created traffic congestion on Warrigal Rd and Monash Fwy.

This selected incident has occurred at the Warrigal Rd-Waverly Rd intersection (see Fig. 7).

The tweets relevant to this incident were collected using the technique delineated in Section II-F, in which the two parameters are set as follows. The radius of the event (r_e) is set to 3.7km which is the distance to the furthest location (Warrigal Rd - Burwood Hwy intersection) with a significant *relative incident impact* identified by the impact propagation analysis in the next section (see Table II). The event duration (t_e) is set experimentally to 1 hour to be suitable for both major and minor incidents. From the collected tweets, it was found that these non-recurrent traffic flow changes had occurred due to an accident (Fig. 9). Tweets in Fig. 9 (right) shows that there was a communication gap of more than 45 minutes to gather information on the situation. Due to the data-driven nature of the algorithm, concept drifts on traffic flow can be detected almost real time. Being able to provide a real-time notification on non-recurrent traffic flow changes will allow better communications and effective optimizations.

The proposed algorithm has the capability of detecting concept drifts with different granularities such as daily or hourly. This can be achieved by changing the growth threshold and distance measure. With the demonstrated granularity, out of detected concept drifts, 92.3% was in-line with the ground truth validated by domain experts.

TABLE II
INCIDENT IMPACT PROPAGATION FOR AN INCIDENT AT WARRIGAL RD-WAVERLEY RD INTERSECTION

Direction	Road segment	Relative incident impact	
		Weekday 8AM	Weekday 8PM
Warrigal Rd south bound	Batesford Rd to Waverley Rd	28.8	24.9
	High St Rd to Batesford Rd	19.0	15.4
	Highbury Rd to High St Rd	12.8	9.6
	Burwood Hwy to Highbury Rd	10.1	8.1
Warrigal Rd north bound	Monash Fwy to Waverley Rd	27.8	38.0
	Middle Rd to Monash Fwy	14.0	19.2
	Princess Hwy to Middle Rd	8.8	9.4
	Atherton Rd to Princess Hwy	5.2	8.1
Waverley Rd east bound	Batesford Rd to Warrigal Rd	17.3	16.0
	Chadstone Rd to Batesford Rd	8.4	7.5
Waverley Rd west bound	Huntingdale Rd to Warrigal Rd	11.4	6.1
	Stephensons Rd to Huntingdale Rd	7.6	2.8
	Forster Rd to Stephensons Rd	5.5	1.8
	Blackburn Rd to Forster Rd	3.4	1.0

C. Impact Propagation Analysis

In Section II-D the impact propagation across the road network due to a non-recurrent incident was derived based on the traffic flow towards that location from other road segments. This capability is demonstrated based on the Warrigal Rd - Waverley Rd intersection (see Fig. 7) where the traffic incident was identified in the previous section (non-recurrent concept drift at timestamps [t32] in Fig. 8). Based on the Eq. 5 (Sec. II-C) the *relative incident impact* was derived for the road segments that are close to the incident location.

Table II presents the road segments with a significant impact from the incident with its respective relative incident impact determined for the weekday 8 am traffic and weekday 8 pm traffic. Note that two time-of-the-day values were selected to compare and contrast differences in incident impact propagation due to different traffic behaviors.

As shown in Table II, the incident impact is mainly propagated along the Warrigal Rd in both directions (south and northbound), compared to relatively less impact on Waverley Rd (east and westbound). This is because Warrigal Rd carries a high traffic flow as it has entrance to the freeway. At 8 am, the highest impact on Warrigal Rd is for Batesford Rd to Waverley Rd traffic flow, as that traffic is bound to cross the Warrigal Rd-Waverley Rd intersection and enter the Monash Fwy towards the City. In contrast, impact at 8 pm is highest on Monash Fwy to Waverley Rd, which consists of the traffic coming from the city and exiting Monash Fwy to Warrigal Rd.

TABLE III
EVALUATION OF ACCURACY BASED ON THE TIME-LAG SELECTED FOR PREDICTION

Time lag	*MSE*	*MAE*
1	*0.0102*	*0.0757*
2	*0.0100*	*0.0746*
3	***0.0097***	***0.0727***
4	*0.0099*	*0.0747*
5	*0.0099*	*0.0737*
6	*0.0101*	*0.0750*
7	*0.0135*	*0.0858*

The incident impact on Waverley Rd is relatively less, which mostly impacts the eastbound traffic while the westbound traffic is only impacted at 8 am.

D. Traffic Forecasting

The primary objective of the traffic forecasting experiment is to evaluate the proposed DNN's fitness in short-term traffic flow forecasting. Based on the impact propagation analysis conducted in the previous step, it was identified that the traffic flow of the road segment on Warrigal Rd (annotated as B to C in Fig. 7), in which heavy traffic congestion befall in frequent time periods, is critically impacted by the road segments; A to B, C to D and C to E. Therefore, in this experiment we have evaluated the effect of the identified road segments with respect to the traffic flow prediction of road segment B to C. In the experiment, the data horizon is taken as 15 minutes, where the data extension was 24 days (1st to 24th October 2017).

Based on the Eq. 10, traffic flow is determined for the selected road segments using 15-minute sampling intervals and utilized as the input sequences of the DNN. The data was divided into training and testing subsets, in which the first 24 days' data were utilized for training of the model and the remaining 7 days' data was utilized for testing the model performance. To validate the effectiveness of the algorithm, the model performance was measured by means of Mean Squared Error (MSE) and Mean Absolute Error (MAE).

Table III demonstrates the prediction performance with respect to the measures selected. We evaluate the prediction performance based on the multiple time-lags, which have presented to the DNN as input, ranging from 15 minutes to 2 hours. The time lag with the highest performance is marked in bold. The findings reveal that with the time lag increases the model performance. However, after the time lag of 3, the model performance begins to decrease, and it starts to drastically reduce after the time lag of 5. Based on the internals of the algorithms and the findings, it is evident that the DNN with three LSTM layers can successfully model dependencies between time series input (i.e., traffic flow data as input) for a time lag of 3. However, the current model is not capable enough to model complex dependencies over 3-time lags, hence, the decline of the performance of the prediction model.

Fig. 10 presents the traffic flow prediction performance comparison on the time period Oct. 25 to Oct. 31, 2017, in order to examine the prediction performance visually.

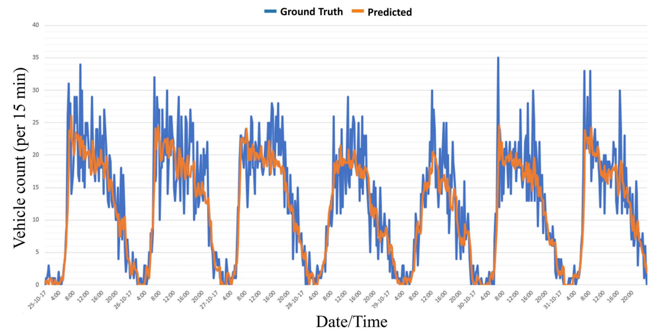

Fig. 10. Traffic flow prediction performance comparison on the time period Oct. 25 to Oct. 31, 2017.

The DNN trained upon input of 3 time-lags (i.e., 45 minutes prior traffic flow data) was selected for the examination in Fig. 10 to compare with the ground truth traffic flow data. Based on the graphical representation of the prediction performance, it is clear that the DNN model can successfully model traffic flow with normal fluctuation, however, there are limitations in predicting traffic flow that has severe fluctuations which occur at a higher frequency. This limitation is expected to be addressed when additional traffic data (e.g. roadworks, incidents, events) is used for training the DNN model.

E. Intelligent Traffic Control

To illustrate the effectiveness of DRL for adaptive traffic control, we apply it to the traffic network around the selected shopping center (Fig. 7) to control the traffic light phases. The selected area has 1805 lanes so the input dimension (i.e., the state S) of the proposed DNN model is 1805. In our experiments, we perform 500 simulation runs (or epochs) to learn the DNN model. Five master control programs are considered, i.e., $M = 5$. The first program is the default program determined by the SUMO's generator [36]. The second program set all phase durations to 60 seconds. The third program doubles the phase durations for phases with durations greater than 40 seconds. The fourth program double the phase durations for all phases. The last program set the phase durations for all phases to 30 seconds. The parameters of the proposed DRL algorithm are: discount factor $\gamma = 0.95$, initial exploration factor $\epsilon = 1$, and decay factor $= 0.99$. The network weights are updated using Adam optimizer with the learning rate of 0.001 at the end of each simulation run using the experience replayed from the agent's memory. For the reward function in Eq. 8, we use the penalty $\mathrm{P} = 5$.

The results of DRL is shown in Fig. 11. It is easy to see that the average waiting time is roughly reduced by half after the 500 epochs. It demonstrates the ability of the proposed DRL to learn effective control decisions even for a very large traffic network.

The learning mechanism of the proposed DRL can be adapted to different networks and other control problems in intelligent traffic systems. Instead of periodically reselect the control program, we can modify the algorithm to adjust the control program based on the signal from concept drift or abnormally detection. Compared to conventional

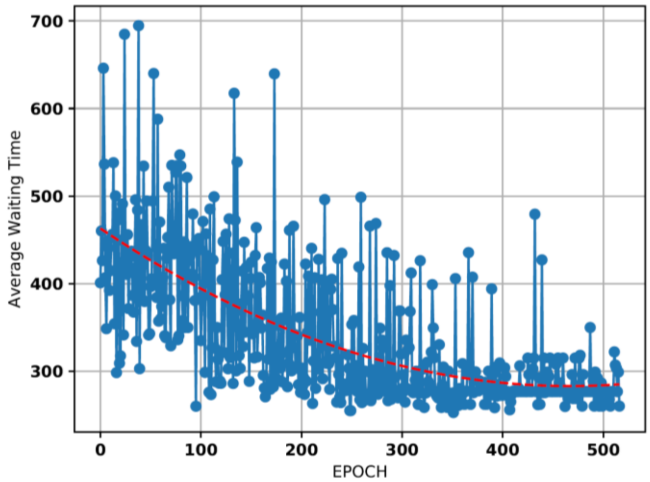

Fig. 11. Training performance of the proposed DRL.

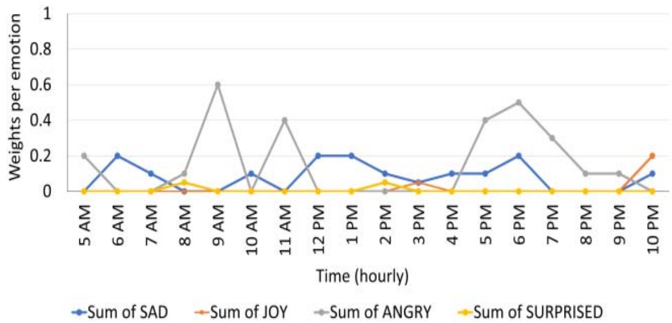

Fig. 12. Emotion fluctuations over time (hourly).

control methods, DRL is more suitable for high-dimensional streaming data because high-level features can be automatically extracted and DRL is data-driven and does not rely on any explicit model.

F. Control Social Media Based Commuter Emotion Analysis

Based on the proposed model in Section II-F, an experiment was carried out to detect significant emotional behaviors of users who post about road conditions on social media. To conduct the experiment, Twitter data was collected on a day where a service disruption had occurred, which has consequently led to an increase in road traffic. Widely used hashtags (#VicTraffic, #VicRoads, #MelbTraffic, #GettrafficVIC) related to transportation in the selected region were used to filter the relevant social media content. Following Fig. 12 illustrates the emotion expression variations overtime on an hourly scale on this day.

Emotion intensity at each time frame was calculated and was compared with a 3-point moving average to determine if the change in the behavior was significant or not. This will enable to identify abrupt peaks with high emotion intensities. Based on the average, Table IV shows the significant times which recorded a higher emotion intensity on Twitter. It can be seen that traffic peak hours demonstrated a higher negative emotion intensity when compared to other times of the day. We further explored the relationship between emotional intensity and congestion using a Granger Causality test [49].

TABLE IV

HIGHER EMOTION INTENSITY TIMELINE

Significant negative emotions	*8:00:00 AM*
	9:00:00 AM
	11:00:00 AM
	5:00:00 PM
	6:00:00 PM
	7:00:00 PM
Significant positive emotions	*10:00:00 PM*

We found out that the association between emotional intensity and congestion is statistically significant ($p < 0.05$).

Depending on the required sensitivity to detect emotion events, the threshold value could be changed, and human monitoring could be used to validate the significant events captured by the algorithm. Incorporating real-time monitoring of crowd-sourced data on social media would enable authorities to take proactive decisions and gain a better understanding of the user's perception.

This section demonstrated each module of the proposed multi-layered STMP platform using a dataset of 190 million records of smart sensor network traffic data generated by 545,851 commuters and corresponding social media data on the arterial road network of Victoria, Australia. Data received from heterogeneous sources is integrated and transformed in L1, in L2, concept drifts, recurrent and non-recurrent events are identified by the online incremental machine learning algorithm. Identified events are fed into the smart traffic management modules in L3.

IV. CONCLUSION

This paper proposed a new smart traffic management platform to capture dynamic patterns from traffic data streams and to integrate AI modules for real-time traffic analysis and adaptive traffic control. The main benefit of the proposed platform is that its AI modules are designed to efficiently cope with the key challenges of future transportation systems where IoT devices are widely adopted, analysis and control technologies must be more responsive and self-evolved, and social behaviors need to be taken into consideration. Moreover, the platform also overcomes the limitations of current algorithms and technologies which rely heavily on limited labeled data and strict assumptions about data and traffic behaviors.

To evaluate the feasibility and effectiveness of the proposed platform, we have conducted a series of experiments based on real-time Bluetooth sensor network data and social media data from the arterial road network in Victoria, Australia. The experimental results show that the platform can successfully and in a timely manner detect recurrent and non-recurrent events, and those results are further validated using the insights automatically captured from social media. The experiments also show that impact propagation and traffic flow prediction modules can efficiently predict short-term impacts of the events. Finally, the simulation of a large-scale traffic network shows that the proposed deep reinforcement learning can learn to improve traffic signal control decision based on many real-time data streams.

We acknowledge a number of potential areas for improvement. Further research directions involve fusing data from heterogeneous data sources such as security cameras, weather information, and other transportation-related data sources. Also, the interpretability of AI modules, especially ones based on complicated techniques such as deep neural networks, are worth investigating in the future to gain the acceptance of the platforms.

Acknowledgment

Authors would like to thank VicRoads–Victorian Road Authority, Australia for their consultations in domain knowledge and providing us with real road traffic data. Authors would also like to acknowledge the in-kind support from the Data to Decisions Cooperative Research Centre (D2D CRC) as part of their analytics and decision support Program.

References

[1] I. Lana, J. D. Ser, M. Velez, and E. I. Vlahogianni, "Road traffic forecasting: Recent advances and new challenges," *IEEE Intell. Transp. Syst. Mag.*, vol. 10, no. 2, pp. 93–109, Apr. 2018.

[2] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Short-term traffic forecasting: Where we are and where we're going," *Transp. Res. C, Emerg. Technol.*, vol. 43, pp. 3–19, Jun. 2014.

[3] J. Gama, I. Žliobaite, A. Bifet, M. Pechenizkiy, and A. Bouchachia, "A survey on concept drift adaptation," *ACM Comput. Surv.*, vol. 46, no. 4, p. 44, Apr. 2014.

[4] J. L. Lobo, J. D. Ser, M. N. Bilbao, C. Perfecto, and S. Salcedo-Sanz, "DRED: An evolutionary diversity generation method for concept drift adaptation in online learning environments," *Appl. Soft Comput.*, vol. 68, pp. 693–709, Jul. 2018.

[5] M. M. Masud *et al.*, "Addressing concept-evolution in concept-drifting data streams," in *Proc. IEEE Int. Conf. Data Mining*, Dec. 2010, pp. 929–934.

[6] A. Bifet, G. Holmes, B. Pfahringer, R. Kirkby, and R. Gavaldà, "New ensemble methods for evolving data streams," in *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jun. 2009, pp. 139–148.

[7] M. Sayed-Mouchaweh, "Learning in dynamic environments," in *Learning from Data Streams in Dynamic Environments*, Springer, New York, NY, USA: 2016.

[8] M. A. Maloof and R. S. Michalski, "Incremental learning with partial instance memory," *Artif. Intell.*, vol. 154, nos. 1–2, pp. 95–126, Apr. 2004.

[9] K. N. Qureshi and A. H. Abdullah, "A survey on intelligent transportation systems," *Middle-East J. Sci. Res.*, vol. 15, no. 5, pp. 629–642, 2013.

[10] C. Khatri, "Real-time road traffic information detection through social media," Jan. 2018, *arXiv:1801.05088*. [Online]. Available: https://arxiv.org/abs/1801.05088

[11] D. Wang, A. Al-Rubaie, J. Davies, and S. S. Clarke, "Real time road traffic monitoring alert based on incremental learning from tweets," in *Proc. IEEE Symp. Evolving Auton. Learn. Syst. (EALS)*, Dec. 2014, pp. 50–57.

[12] P.-T. Chen, F. Chen, and Z. Qian, "Road traffic congestion monitoring in social media with hinge-loss Markov random fields," in *Proc. IEEE Int. Conf. Data Mining*, Dec. 2014, pp. 80–89.

[13] M. J. Lighthill and G. B. Whitham, "On kinematic waves. II. A theory of traffic flow on long crowded roads," *Proc. Roy. Soc. London, Ser. A, Math. Phys. Sci.*, vol. 229, pp. 317–345, May 1955.

[14] C. F. Daganzo, "The cell transmission model. Part I: A simple dynamic representation of highway traffic," *Transp. Res. Part B Methodol.*, vol. 28, no. 4, pp. 269–287, 1994.

[15] P. I. Richards, "Shock waves on the highway," *Oper. Res.*, vol. 4, no. 1, pp. 42–51, 1956.

[16] A. Aw, A. Klar, M. Rascle, and T. Materne, "Derivation of continuum traffic flow models from microscopic follow-the-leader models," *SIAM J. Appl. Math.*, vol. 63, no. 1, pp. 259–278, 2002.

[17] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic flow prediction with big data: A deep learning approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 865–873, Apr. 2015.

[18] R. Yu, Y. Li, C. Shahabi, U. Demiryurek, and Y. Liu, "Deep learning: A generic approach for extreme condition traffic forecasting," in *Proc. SIAM Int. Conf. Data Mining*, Jun. 2017, pp. 777–785.

[19] G. Michau, A. Nantes, A. Bhaskar, E. Chung, P. Abry, and P. Borgnat, "Bluetooth data in an urban context: Retrieving vehicle trajectories," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 9, pp. 2377–2386, Sep. 2017.

[20] D. Nallaperuma, D. D. Silva, D. Alahakoon, and X. Yu, "A cognitive data stream mining technique for context-aware IoT systems," in *Proc. 43rd Annu. Conf. IEEE Ind. Electron. Soc.*, Nov. 2017, pp. 4777–4782.

[21] D. Nallaperuma, D. D. Silva, D. Alahakoon, and X. Yu, "Intelligent detection of driver behavior changes for effective coordination between autonomous and human driven vehicles," in *Proc. 43rd Annu. Conf. IEEE Ind. Electron. Soc.*, Oct. 2018, pp. 3120–3125.

[22] A. Câmpan and G. Şerban, "Adaptive clustering algorithms," in *Proc. Adv. Artif. Intell.*, Oct. 2006, pp. 407–418.

[23] D. D. Silva and D. Alahakoon, "Incremental knowledge acquisition and self learning from text," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2010, pp. 1–8.

[24] D. Alahakoon, S. K. Halgamuge, and B. Srinivasan, "Dynamic self-organizing maps with controlled growth for knowledge discovery," *IEEE Trans. Neural Netw.*, vol. 11, no. 3, pp. 601–614, May 2000.

[25] C. S. Wirasinghe, "Determination of traffic delays from shock-wave analysis," *Transp. Res.*, vol. 12, no. 5, pp. 343–348, Oct. 1978.

[26] B. Pan, U. Demiryurek, C. Shahabi, and C. Gupta, "Forecasting spatiotemporal impact of traffic incidents on road networks," in *Proc. IEEE 13th Int. Conf. Data Mining*, Dec. 2013, pp. 587–596.

[27] W. Huang, G. Song, H. Hong, and K. Xie, "Deep architecture for traffic flow prediction: Deep belief networks with multitask learning," *IEEE Trans Intell. Transp. Syst.*, vol. 15, no. 5, pp. 2191–2201, Oct. 2014.

[28] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transp. Res. Part C, Emerg. Technol.*, vol. 54, pp. 187–197, May 2015.

[29] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1–32, 1997.

[30] M. Hermans and B. Schrauwen, "Training and analysing deep recurrent neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, Jun. 2013, pp. 190–198.

[31] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 6645–6649.

[32] Z. Zhao *et al.*, "LSTM network: A deep learning approach for short-term traffic forecast," *IET Intell. Transp. Syst.*, vol. 11, no. 2, pp. 68–75, 2017.

[33] L. Baskar, B. De Schutter, J. Hellendoorn, and Z. Papp, "Traffic control and intelligent vehicle highway systems: A survey," *IET Intell. Transp. Syst.*, vol. 5, no. 1, pp. 38–52, Mar. 2011.

[34] V. Mnih *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, pp. 529–533, Feb. 2015.

[35] M. Haklay and P. Weber, "OpenStreetMap: User-generated street maps," *IEEE Pervasive Comput.*, vol. 7, no. 4, pp. 12–18, Oct. 2008.

[36] A. Wegener, M. Piórkowski, M. Raya, H. Hellbrück, S. Fischer, and J.-P. Hubaux, "TraCI: An interface for coupling road traffic and network simulators," in *Proc. 11th Commun. Netw. Simulation Symp.*, Apr. 2008, pp. 155–163.

[37] H. V. Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double Q-Learning," Sep. 2015, *arXiv:1509.06461*. [Online]. Available: https://arxiv.org/abs/1509.06461

[38] S. M. Grant-Muller, A. Gal-Tzur, E. Minkov, S. Nocera, T. Kuflik, and I. Shoor, "Enhancing transport data collection through social media sources: Methods, challenges and opportunities for textual data," *IET Intell. Transp. Syst.*, vol. 9, no. 4, pp. 407–417, Nov. 2014.

[39] F. Ali, D. Kwak, P. Khan, S. M. R. Islam, K. H. Kim, and K. S. Kwak, "Fuzzy ontology-based sentiment analysis of transportation and city feature reviews for safe traveling," *Transp. Res. C, Emerg. Technol.*, vol. 77, pp. 33–48, Apr. 2017.

[40] F. C. Pereira, A. L. C. Bazzan, and M. Ben-Akiva, "The role of context in transport prediction," *IEEE Intell. Syst.*, vol. 29, no. 1, pp. 76–80, Jan. 2014.

[41] H. Abdelhaq, C. Sengstock, and M. Gertz, "EvenTweet: Online localized event detection from twitter," *Proc. VLDB Endowment*, vol. 6, no. 12, pp. 1326–1329, Aug. 2013.

[42] P. Ekman, "An argument for basic emotions," *Cognit. Emotion*, vol. 6, nos. 3–4, pp. 169–200, 1992.

[43] J. H. Kahn, R. M. Tobin, A. E. Massey, and J. A. Anderson, "Measuring emotional expression with the linguistic inquiry and word count," *Amer. J. Psychol.*, vol. 120, no. 2, pp. 263–286, Jul. 2007.

[44] R. Valitutti, "WordNet-affect: An affective extension of WordNet," in *Proc. 4th Int. Conf. Lang. Resour. Eval.*, May 2004, pp. 1083–1086.

[45] U. Mori, A. Mendiburu, M. Álvarez, and J. A. Lozano, "A review of travel time estimation and forecasting for advanced traveller information systems," *Transportmetrica A, Transp. Sci.*, vol. 11, no. 2, pp. 119–157, 2015.

[46] A. Bhaskar and E. Chung, "Fundamental understanding on the use of Bluetooth scanner as a complementary transport data," *Transp. Res. C, Emerg. Technol.*, vol. 37, pp. 42–72, Dec. 2013.

[47] VicRoads. *Bluetooth Sites–VicRoads Open Data*. Accessed: Feb. 26, 2019. [Online]. Available: https://vicroadsopendata-vicroadsmaps.opendata.arcgis.com/datasets/48fd4d7e1127453ea5f9bdc757ab00e7_0

[48] A. Nantes, M. P. Miska, A. Bhaskar, and E. Chung, "Noisy Bluetooth traffic data?" *Road Transp. Res., A J. Austral. New Zealand Res. Pract.*, vol. 23, no. 1, pp. 33–43, Mar. 2014.

[49] C. Hiemstra and J. D. Jones, "Testing for linear and nonlinear Granger causality in the stock price-volume relation, " *J. Finance*, vol. 49, no. 5, pp. 1639–1664, Dec. 1994.

Dinithi Nallaperuma received the master's degree in advanced software engineering from the University of Westminster, U.K., and the Ph.D. degree from the Research Center for Data Analytics and Cognition, La Trobe University, Australia. Prior to commencing her Ph.D. degree, she was a Research Engineer with the National University of Singapore. Her research interests include the Internet of Things, data stream mining, unsupervised learning, and cognitive computing.

Rashmika Nawaratne received the bachelor's degree (Hons.) from the Department of Computer Science and Engineering, University of Moratuwa, Sri Lanka, in 2014. He is currently pursuing the Ph.D. degree with the Research Centre for Data Analytics and Cognition, La Trobe University, Australia. Prior to commencing his Ph.D. degree, he was a Technical Lead with the Software Product Development Organization. His research interests include incremental learning, video analytics, deep learning, the Internet of Things, and data stream mining.

Tharindu Bandaragoda received bachelor's degree (Hons.) in engineering from the University of Moratuwa, Sri Lanka, and the M.Phil. degree in IT from Monash University, Australia. He is currently pursuing the Ph.D. degree with the Research Centre for Data Analytics and Cognition, La Trobe University, Australia. He is currently working on the D2D CRC Beat the News Project. His research interests are in cognitive-inspired incremental learning, data mining, social media analytics, large-scale data mining for anomaly detection, and mining insights from social media.

Achini Adikari received bachelor's degree (Hons.) in information technology from the University of Moratuwa, Sri Lanka. She is currently pursuing the Ph.D. degree with the Research Center for Data Analytics and Cognition, La Trobe University, Australia. Prior to commencing her Ph.D. degree, she was an Associate Tech Lead with the Software Product Development Organization. She is currently works on several research projects related to text analysis in public health forums and social media data, with a keen interest in human emotions analysis using self-learning artificial intelligence (AI).

Su Nguyen received the Ph.D. degree in artificial intelligence and data analytics from the Victoria University of Wellington, New Zealand, in 2013. He is currently a Senior Research Fellow with the Centre for Research in Data Analytics and Cognition, La Trobe University, Australia. His primary research interests include computational intelligence, optimization, data analytics, large-scale simulation, and their applications in operations management and social media.

Thimal Kempitiya received the bachelor's degree (Hons.) from the Department of Computer Science and Engineering, University of Moratuwa, Sri Lanka, in 2015. He is currently a senior software engineer in machine learning and digital human simulation fields. His research interests include data mining, business analytics, deep learning, incremental learning, and human simulation.

Daswin De Silva received the Ph.D. degree in machine learning and artificial intelligence from Monash University. He is currently a Senior Lecturer with La Trobe University, Australia. Besides his academic and professional expertise in analytics, he is also actively involved in industry research and engagement. His research interests include cognitive computing, autonomous learning algorithms, incremental knowledge acquisition, stream mining, social media, and text mining.

Damminda Alahakoon received the Ph.D. degree from Monash University, Australia, in 2002. He is currently a Professor of business analytics with the La Trobe University Business School, Melbourne, Australia, and the Director of the Research Centre for Data Analytics and Cognition. He has published over 100 research articles in data mining, clustering, neural networks, machine learning, and cognitive systems. He received the Monash Artificial Intelligence Prize from Monash University.

Dakshan Pothuhera received the M.B.A. and master's degrees in management information systems from Swinburne University, Australia. He is currently a Senior Project Manager with VicRoads, where he is responsible for delivering business and IT projects for VicRoads on behalf of the Victorian government. He has over 20 years of experience in the IT industry, including over three years in USA and 11 years in Australia.