Journal of Intelligent Transportation Systems
Technology, Planning, and Operations

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/gits20

Network-level turning movement counts estimation using traffic controller event-based data

Peipei Xu, Xiaofeng Li, Yao-Jan Wu & Hyunsoo Noh

JOURNAL OF INTELLIGENT TRANSPORTATION SYSTEMS
https://doi.org/10.1080/15472450.2022.2075701

Network-level turning movement counts estimation using traffic controller event-based data

Peipei Xu[a], Xiaofeng Li[a], Yao-Jan Wu[a], and Hyunsoo Noh[b]

[a]Department of Civil and Architectural Engineering and Mechanics, The University of Arizona, Tucson, AZ, USA; [b]Pima Association of Governments, Tucson, AZ, USA

ABSTRACT

Accurate turning movement counts (TMCs) data collected from regional-wide signalized intersections is critical to regional transportation planning and simulation modeling. A variety of existing traffic sensors, configured at intersections for traffic detection and signal control, can generate a large amount of real-time high-resolution event-based data from traffic controllers but few of these sensors are configured to collect TMC. This paper proposes a methodology for estimating network-level TMC using existing traffic controller event-based data without installing additional sensors. First, relevant features that can indicate traffic arrival are extracted from existing event-based data, including detector occupancy time, detector-triggered count, and green time duration. With these features, a multi-output multilayer neural network model is developed to estimate TMC. To further improve network-level TMC estimation accuracy, intersection infrastructure data and point-of-interest (POI) data are included as exogenous variables for the proposed model. Ninety-three signalized intersections are chosen from the Pima County region, Arizona, to calibrate and verify the developed model. The validation results show that the proposed model can accurately estimate TMC, as indicated by a median Root Mean Square Error (RMSE) of 41 veh/15 min, 11 veh/15 min, and 12 veh/15 min for through movement, left-turn movement, and right-turn movement volume estimation, respectively. This research provides a new possibility of utilizing existing data sources to obtain network-level TMC data without additional infrastructure and labor costs for transportation agencies.

ARTICLE HISTORY
Received 4 August 2020
Revised 22 April 2022
Accepted 4 May 2022

KEYWORDS
multi-output multilayer neural network model; signalized intersections; traffic controller event-based data; turning movement counts

Introduction

Turning movement counts (TMCs) at signalized intersections are essential inputs needed by transportation agencies for transportation planning, simulation modeling, and traffic signal timing optimization. Traditional methods for TMC collection, such as manual counting, are both labor-intensive and expensive, especially for region-wide traffic data collection (Li, 2021). As vehicle detection technology has developed, intelligent sensors such as inductive-loop detectors, video-based sensors, and radar sensors have begun to be used for automatic traffic volume data collection. Although these intelligent sensors can automatically collect omnidirectional TMC data 24/7, regionally purchasing and installing these sensors is extremely expensive. For transportation agencies with funding constraints, collecting regional TMC data using smart sensors is infeasible due to the cost. In order to cost-effectively obtain TMC data at all intersections within a region, many practitioners and researchers have worked on developing indirect methods of estimating traffic volume or TMC based on existing data sources to avoid purchasing and installing expensive sensors.

Among previous relevant studies, traffic volume estimation could refer to estimating traffic volume on street segments, or for one movement (such as through movement) at signalized intersections. In contrast with traffic volume estimation, TMC estimation needs to quantify the traffic volume for several movements, i.e., left turn, right turn, and through movement. One type of TMC estimation method is the estimation of turning movement proportions based on available entry and exit traffic counts at intersections from loop detectors (Ghods and Fu, 2014; Mirchandani et al., 2001; Virkler and Kumar, 1998). The concept of this method is a turning movement proportion matrices estimation conducted by

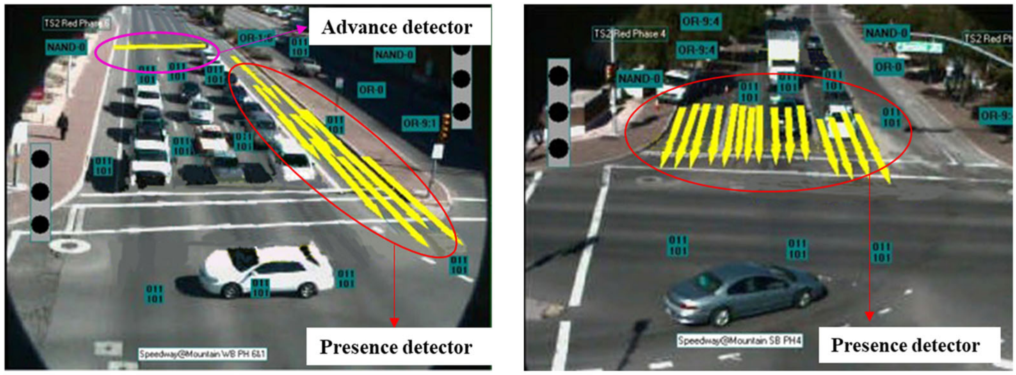
CONTACT Xiaofeng Li xfli@email.arizona.edu Department of Civil and Architectural Engineering and Mechanics, The University of Arizona, Tucson, AZ 85721, USA.

minimizing the error between predicted and observed exit counts or maximizing the likelihood of TMC observations, based on the traffic flow conservation law. However, a region-wide implementation of this method is still challenging because most network roads do not have loop detectors installed for exit count collection at intersections.

Multiple emerging data sources have recently become available in the transportation domain and have been used for TMC estimation. For example, Global Position System (GPS) trajectory data is collected from probe vehicles and connected vehicles. Some studies have proposed methods to estimate TMC at signalized intersections using the probe vehicle or connected vehicle data (Zheng and Liu, 2017). Because connected vehicle technology is still developing, the sample vehicles (e.g., taxi, probe vehicle, and connected vehicle) only account for a small fraction of total traffic at some intersections. This GPS data is insufficient for estimating network-level TMC and may limit the accuracy of the proposed methods. In a further study (Zhao et al., 2019), the penetration rate of vehicles with GPS devices was estimated by using historical stopping positions of vehicles based on the probability theory to mitigate the limitation due to insufficient GPS data. However, the proposed method (Zhao et al., 2019) is still unable to estimate the right-turn volume, or be applied at an intersection with shared lanes. Furthermore, existing GPS data-based methods require a large number of historical trajectory details in a region for TMC estimation. Using this detailed trajectory information is difficult for many transportation agencies due to privacy concerns and costs.

Another emerging large-scale data source is the high-resolution traffic controller event-based data (hereafter "event-based data"). This data can be easily obtained by most local transportation agencies because it is collected by modern signal controllers (Liu and Ma, 2008; Smaglik et al., 2007). Event-based data records a dataset containing signal changes, detector actuation events data, pedestrian-related events data, communication events, etc. Multiple systems have been created by university research teams to collect event-based data from traffic controllers (Balke et al., 2005; Liu and Ma, 2009), and commercial products, such as the ASC/3 controllers developed by Econolite©, have been developed to enable agencies to collect region-wide event-based data. These event-based data were collected and applied to calculate cycle-by-cycle signal performance measures (Day et al., 2008), which were further used to quantify traffic signal performance in later studies (Day et al., 2014, 2010, 2012, 2009). These measures are known as Automated Traffic Signal Performance Measures (ATSPMs) and are used to support traffic signal operations and management. Although ATSPMs can provide TMC data by configuring lane-by-lane stop-bar detectors (Day et al., 2014), obtaining accurate TMC data through ATSPMs is still challenging for agencies that are using an old detector layout with one detector covering multiple lanes. In addition to supporting ATSPM, event-based data has been successfully applied for solving traditional traffic problems, such as travel time estimation (Liu et al., 2008), long queue length estimation (An et al., 2018; Liu et al., 2009), pedestrian volume estimation (Li et al., 2021; Li and Wu, 2021), signal control optimization (Day and Bullock, 2020; Hu and Liu, 2013), and signal timing evaluation (Li, Weber, et al., 2019). Despite these successful applications, the current application of event-based data for TMC estimation is still limited, especially for intersections equipped with old detector layouts. A study conducted by Li et al. has used event-based data to estimate traffic volume based on the proposed dynamic Hidden Markov Model (Li, Wu, et al., 2019). However, this proposed method can only be used for estimating the through movement volume on roads configured with an advance detector.

It was found that previous research on estimating TMC using existing data sources is very limited due to data availability. In comparison with other available data sources, such as GPS data, event-based data has the advantages of region-wide coverage and low cost because most local transportation agencies have configured traffic detectors for actuated signal control. According to the latest National Traffic Signal Report Card (National Transportation Operations Coalition, 2012), 73% of U.S. agencies use actuated signal control. Real-time vehicle detection information and signal status information from event-based data has been proven to have a high correlation with traffic conditions and traffic arrival, for example, event-based data has been successfully used for queue length estimation (An et al., 2018), classification (Liu and Sun, 2014), and through movement volume estimation (Li, Wu, et al., 2019). Because of this, event-based data can be used as an ideal data source for network-level TMC estimation. In this study, we proposed a method focusing on using various information extracted from event-based data to accurately estimate network-level TMC at signalized intersections. This proposed method can be applied to individual signalized intersections in different sizes of road networks, providing



a. WB (Major road) b. SB (Minor road)

Figure 1. Detection system configuration at Speedway Blvd & Mountain Ave in Tucson, Arizona.

essential input for most traffic studies including urban road network planning and modeling.

The rest of the paper is organized as follows: the problem statement section first explores three issues of applying event-based data to estimate TMC. After this, all data sources used in the study are briefly introduced, followed by a description of the TMC estimation methodology. Next, the method performance is evaluated using ground-truth data. The final section summarized the conclusions and directions for future research.

Problem statement

In Pima County, Arizona, approximately 80% of signalized intersections apply an actuated signal control strategy. At a typical signalized intersection using actuated signal control, by default, advance detectors and presence detectors in video-based sensors are commonly configured for each of the approaches at an intersection for signal control. The advance detectors are located upstream from the stop bar for the green time extension, while the presence detectors are located on the stop bar for vehicle presence detection, or both (Wu et al., 2010). These two types of default traffic detectors are usually wired together, and only used for vehicle actuation detection because controllers require vehicle arrival data rather than vehicle counts. Most signalized intersections in the Pima County region have a default configuration with advance detectors and presence detectors without the traffic data collection module, thus, the existing vehicle-actuated detectors in video-based sensors are not capable of collecting TMC data at most intersections due to the two major issues described in the following paragraph.

Figure 1(a) shows the configuration of the detection system on a major road approach (westbound approach) in the Advanced Transportation Management Systems (ATMS). The advance detector (illustrated by a horizontal bar) covers two through lanes and one shared right-turn lane of the westbound direction. The advance detector is wired-together and provides a single call to the controller. The presence detector (illustrated by arrows) is configured on left-turn lanes. Because advance detectors commonly have multiple lanes wired together to provide a single call to the controller, the advance detectors could count multiple side-by-side vehicles as one vehicle and therefore underestimate the actual volume. The count recorded by the advance detectors is also unable to separate the through and right-turn volume when a shared right-turn lane exists.

Figure 1(b) shows the configuration of detectors on a minor road approach (southbound direction), with presence detectors (illustrated by the shape of arrows) on all lanes. The presence detector configured on left-turn lanes is separated from the presence detector configured on both through lanes and right-turn lanes because these two presence detectors provide separate calls to the signal controller. For detectors on the through lane and right-turn lane, the major road is configured with the advance detector, while the minor road is configured with the presence detector. The same time duration recorded by advance detectors and presence detectors could refer to different traffic counts due to the different lengths of these two types of detectors. Advance detectors are used to detect moving vehicles, thus, the time duration is most likely to be the vehicle passing time. The presence detectors are typically used for detecting stopped vehicles, so the time duration is most likely to be the vehicle stopping time.

Based on the detection system configuration presented above, exploiting the event-based data from

Table 1. Description of events used for TMC estimation.

Event categories	Event ID	Name	Description
Detector actuation events	82	Detector On	Triggered when vehicles arrive at the detector
	81	Detector Off	Triggered when vehicles depart the detector
Signal change events	1	Phase Green Beginning	Activated when the green time of a traffic signal has begun
	7	Phase Green Termination	Activated when the green time of a traffic signal is terminated
Controller communication events	500	Total Comm Attempts	Total communication attempts of the controller during an interval
	501	Failed Comm Attempts	Number of failed communication attempts during an interval
	502	Percent Comm Loss	Percent of the failed communication attempts during an interval

advance and presence detectors to estimate TMC has the following challenges:

- The detectors configured at signalized intersections are single-channel detectors covering multiple lanes. The count recorded by these single-channel detectors underestimates the actual volume.
- Detector configuration varies with through and left-turn movements on major and minor roads at signalized intersections.
- The layouts of intersections in the road network are inconsistent. For intersections with a shared left-turn or shared right-turn lane, the detection information cannot separate the vehicles detected by single-channel detectors into through vehicles, left-turning vehicles, or right-turning vehicles.

Data description

The data used in this study is collected from the Pima County region and consists of three categories: event-based data, intersection infrastructure data, and POI data. Each of these three categories will be instrumental in preparing the dataset to train our proposed TMC estimation model.

Event-based data

Traffic controller event-based data consists of a series of events (detector actuation events, signal change events, pedestrian-related events, controller communication events, etc.) generated in real-time. The event-based data used in this study is collected by the MaxView© system, which is a type of Advanced Traffic Management System (ATMS) developed by Intelight Inc. This system can continuously collect event-based data from traffic controllers, including detector actuation events, signal change events, pedestrian button-push events, etc. Historical event-based data before 2016 in Pima County haven't been archived and stored by the local transportation agency, thus, the event-based data during the period of 2016 – 2020 are used in this study.

Three types of event datasets, consisting of detector actuation events, signal change events, and communication events, are used for TMC estimation, as shown in Table 1. Detector actuation event data can provide the start and end times of each detector actuation triggered by vehicles. The time difference between the detector off and detector on represents the detector occupancy time, which can be used to indicate traffic arrival for each movement. Signal change events record each signal phase change. Phase green beginning and phase green termination are chosen for green time interval calculation. The green time interval reflects the maximum number of vehicles that could pass the intersection. Controller communication event data indicates the controller's communication quality. When a controller loses communication, the ATMS server has difficulties in collecting event-based data from the traffic controller, leading to missing event-based data (An et al., 2017). Thus, the communication event is used for event-based data quality control and check before applying event-based data in the TMC estimation.

Figure 2(a) shows sample data for each of these three event types. The DeviceId column is the signal controller ID, referred to as an intersection location. The EventId column correspondingly represents the event categories. The Parameter column in detector actuation events is the detector number associated with the turning movement. The Parameter column in signal change events shows the signal phase sequence number. The Parameter column in communication events shows the value of each event in the last period. As shown in Figure 2(b), an even number is typically associated with through movement, while odd numbers are typically associated with left-turn movements as defined in the Signal Timing Manual (Urbanik et al., 2015).

Infrastructure data

As described in the Problem Statement section, the detectors configured at signalized intersections are single-channel detectors covering multiple lanes, with variation in detector configuration on major and

Detector actuation events

TimeStamp	DeviceId	EventId	Parameter
2019-11-04 00:00:08.900	218	81	8
2019-11-04 00:00:08.900	222	82	2
2019-11-04 00:00:09.000	430	82	2
2019-11-04 00:00:09.000	666	82	6
2019-11-04 00:00:09.000	40	81	2
2019-11-04 00:00:09.000	685	81	6
2019-11-04 00:00:09.000	703	82	1

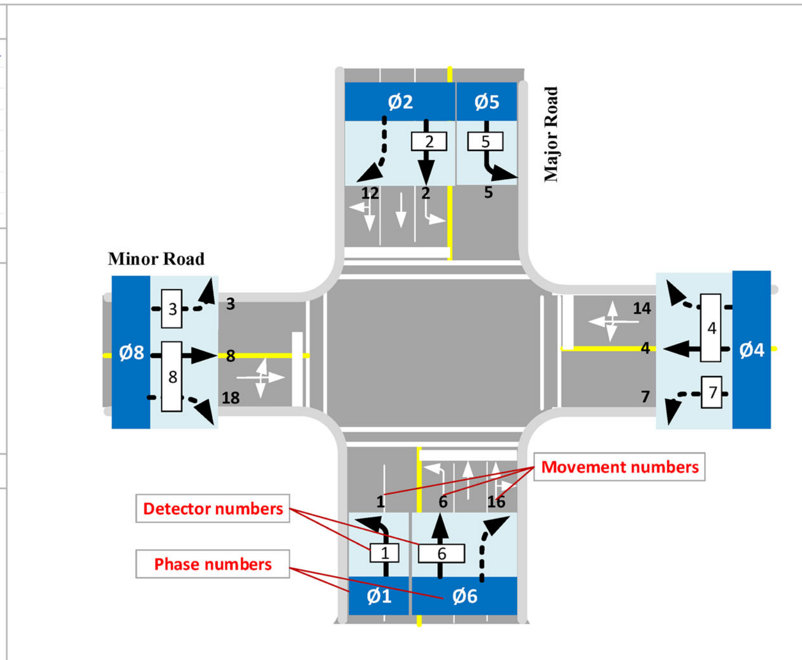
Signal change events

TimeStamp	DeviceId	EventId	Parameter
2019-05-06 01:36:37.200	703	2	8
2019-05-06 01:36:37.200	214	7	2
2019-05-06 01:36:37.200	214	4	2
2019-05-06 01:36:37.200	214	3	2
2019-05-06 01:36:37.300	505	0	2
2019-05-06 01:36:37.300	505	1	2

Communication events

TimeStamp	DeviceId	EventId	Parameter
2019-10-21 00:10:25.987	3	500	19
2019-10-21 00:10:25.987	3	501	15
2019-10-21 00:10:25.987	3	502	78
2019-10-21 00:15:23.860	3	400	-30
2019-10-21 00:15:26.033	3	500	21

(a) Sample data



(b) Detector & signal parameter layout

Figure 2. Sample event-based data collected at intersections.

minor roads. Event-based data from detectors is insufficient for TMC data estimation, therefore, intersection infrastructure data including the type of road (major road or minor road), number of left-turn lanes, number of shared left-turn lanes, number of through lanes, number of right-turn lanes, and number of shared right-turn lanes are introduced in the proposed method. Intersection infrastructure data is manually collected from Google Maps in 2020.

POI data

POI data could indicate urban land use context and economic activities related to traffic attraction and production. The POI data in this study was collected from the Pima Association of Governments (PAG) employment database in which the Google Places API and sample review were utilized to validate the existence of business and control data quality (Noh et al., 2019). The POI data includes 17 categories and 35 subcategories, such as education and business services. The categorization in the POI data follows the North American Industry Classification System (NAICS). In addition to the POI categories, the number of employees and longitude and latitude coordinates are contained in the dataset. In this study, the counts of categories and the number of employees are extracted from raw POI data within a 400 m buffer of an intersection. The category counts can reflect the diversity of POI and land use characteristics surrounding intersections, and the scale of each POI category can be shown by the number of employees per category. These two parameters are expected to be highly related to traveler characteristics and traffic patterns at nearby intersections.

Study intersection selection

The Pima Association of Governments (PAG) conducted a traffic count program to collect TMC data annually from sample intersections within Pima County since 1999. Pima County, with a total area of 9,189 square miles and a population of 1.043 million recorded in 2020 Census, is located in the south-central region of Arizona. The TMC data provided by PAG was an aggregate of 15-min interval TMC data during peak hours (7:00-9:00 AM, 4:00-6:00 PM). These data were usually collected by hired consultant firms using both manual data collection and individual sensors.

Ninety-three intersections in the Pima County area were selected as study locations based on the ground-truth data and event-based data availability, as shown in Figure 3. Ground-truth data and event-based data are selected if both data are identified at the same period and the same location. Then, data from 2016 to 2020 are used in this study to verify and evaluate the proposed method. On average, each intersection

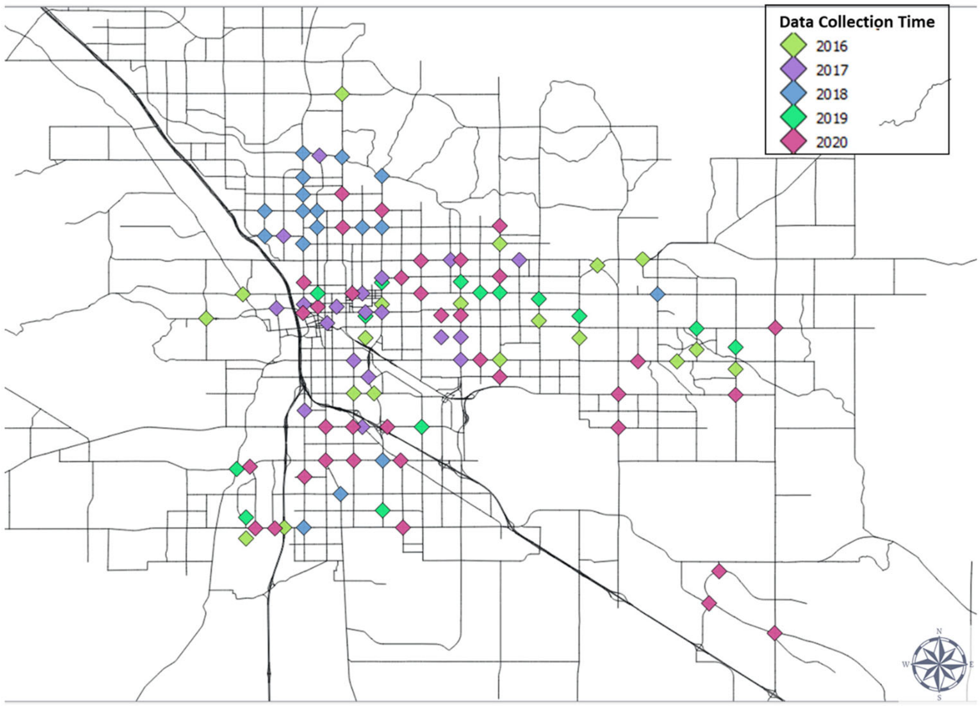


Figure 3. Location of study intersections in Pima County, AZ.

has one to four days of data used for method training and validation. These study intersections are four-leg signalized intersections with 54% of signals using permissive-only left-turn phasing, 36% using protected-permissive left-turn (PPLT) phasing, and 10% using protected-only left-turn phasing.

Methodology

In this section, we first describe how to process event-based data and extract features relevant to TMC estimation. Afterward, a multi-output multi-layer perceptron neural network (MLP) model is developed to estimate TMC based on event-based features and exogenous features such as POI and intersection infrastructure data.

Event-based feature extraction and processing

As mentioned in the data description section, event-based data is much more informative than other data sources because traffic signal changing event data and vehicle-detector actuation event data can record the start time and end time of each detector actuation triggered by vehicles, reporting a real-time signal change status. Based on the information provided by these event-based data, traffic flow-related variables such as detector occupancy time, detector-triggered count, and green time duration can be measured and used to indicate traffic arrival and conditions. In this section, a procedure for these traffic flow-related variables extraction is designed to prepare event-based features for our proposed TMC estimation model.

Event-based feature extraction and processing consist of three steps: communication loss check, event-based feature calculating, and event-based feature combination as shown in Figure 4.

Step 1: Communication loss checking. Event-based data includes a large amount of data but may have some missing data and outliers due to controller communication loss. The communication loss issue of traffic controllers can lead to the loss of event-based data records. For example, the detector-actuation event data loss caused by communication loss could result in an unpaired detector on and off events, or outliers of duration, leading to a biased estimation result. Communication loss can be measured by the event "percentage communication loss (POCL)" in the communication event dataset recorded by controllers. The detector actuation event and signal change event data are removed within the period when the value of POCL is greater than 0.

Step 2: Event-based feature calculating. Four event-based features including detector occupancy time, detector-triggered count, green time duration, and permissive left-turn green time are extracted in this step. These features are calculated and aggregated into 15-min intervals because the ground-truth data were collected in a 15-min interval.

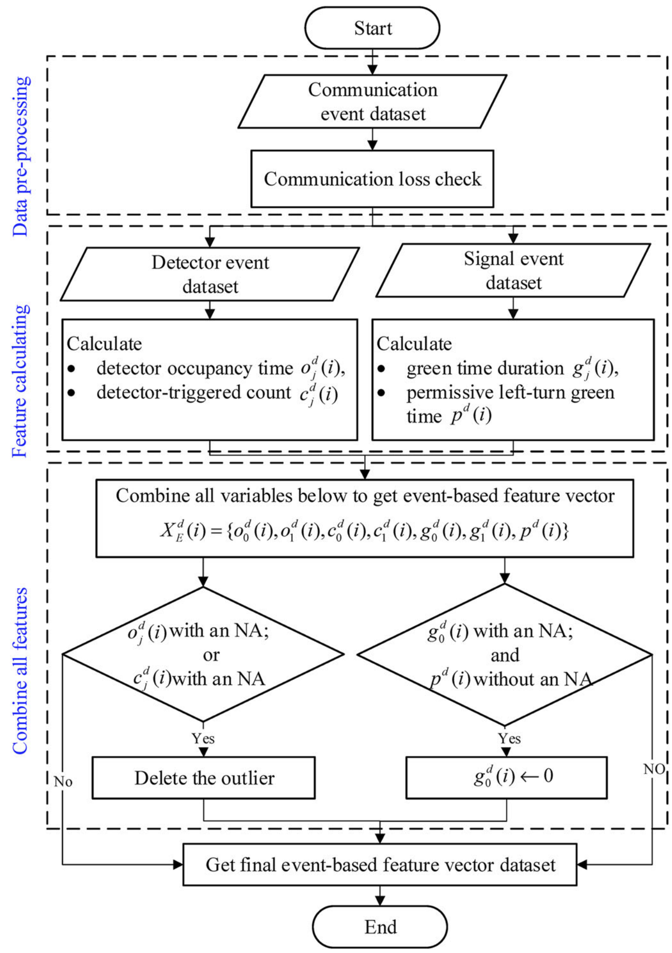


Figure 4. Flowchart of the event-based feature extraction process.

The detector occupancy time (also called "vehicle on-detector time") (Wu and Liu, 2014) can be measured by the time difference between the detector-off timestamp and the corresponding detector-on timestamp. The detector-off event should be followed by the detector-on event. All paired detection events (detector-on and detector-off) are grouped to calculate the detector occupancy time, and all detector occupancy time is summed up in 15-minute intervals. The detector-triggered count is referred to as the total number of times detector-on events occur during a 15-min interval. These two features are formulated as the following:

$$o_j^d(i) = \sum_{n \in N} t_{n,81}^{d,j}(i) - t_{n,82}^{d,j}(i); \;\; j \in \{0,\;1\} \tag{1}$$

$$c_j^d(i) = \sum_{n \in N} [E_n^{d,j}(i) = 81]; \quad j \in \{0,\;1\} \tag{2}$$

where $o_j^d(i)$ is the detector occupancy time for approach i at intersection d; $j = 0$ indicates that a detector number is an even number, which is typically associated with through movements; $j = 1$ is an odd number, which is typically associated with left-turn movements (as shown in Figure 2 (b)) . $t_{n,81}^{d,j}(i)$, $t_{n,82}^{d,j}(i)$ are the timestamp of detector-off and timestamp of detector-on for approach i at intersection d, respectively. N is the number of groups of paired detection events during the 15-min period. $c_j^d(i)$ is the detector-triggered count for approach i at intersection d during the 15-min period. $E_n^{d,j}(i) = 81$ shows that the detector j for approach i at intersection d in group n is detector-on.

Signal change events include the real-time green phase beginning and green phase termination events. After grouping the paired green phase events based on signal cycle ID, green time duration for a whole 15-min interval can be calculated through the sum of the time difference between the phase green beginning event and phase green termination event of each signal phase during a 15-min period, as shown in the following:

$$g_j^d(i) = \sum_{n \in N} t_{n,7}^{d,j}(i) - t_{n,1}^{d,j}(i); j \;\; \in \{0,\;1\} \tag{3}$$

where $g_j^d(i)$ is the green time for phase number j approach i at intersection d during a 15-min interval.

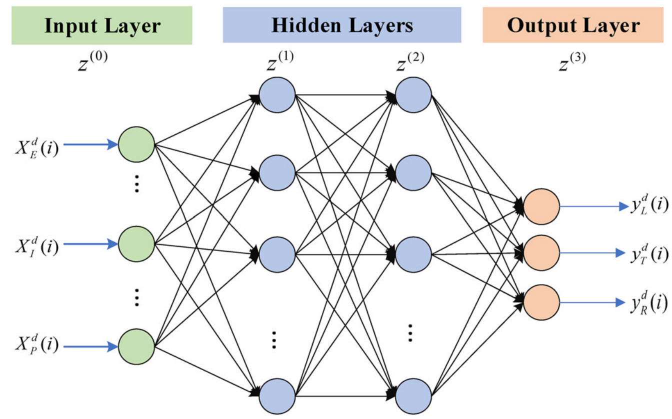


Figure 5. The topology of the MLP model for TMC estimation.

$t_{n,1}^{d,j}(i)$, $t_{n,7}^{d,j}(i)$ are the timestamps of phase green beginning and phase green termination for phase number j approach i at intersection d, respectively.

Through Equation 3, the green time duration for the left-turn and through movement can be calculated. However, only the protected left-turn duration time can be reflected when calculating the left-turn green time. For the permissive-only left-turn phasing, the green time duration is null because no phase parameter is assigned to the left-turn movement (as shown in Figure 2(b) the minor road). For the protected-permissive left-turn phase, only the protected-only green time can be calculated through Equation 3, however, the permissive parts of green time cannot be directly indicated in the event-based data. Thus, permissive left-turn green time is introduced as one input to the model, in addition to protected green time. The permissive left-turn green time calculation is formulated as shown in Equation 4.

$$p^d(i) = \begin{cases} \sum_{c=1}^{C}\left(g_{c,0}^d\left(i'\right) - g_{c,1}^d\left(i'\right)\right); \; S^d\left(i,i'\right) = 1 \\ \sum_{c=1}^{C}\left(g_{c,0}^d\left(i'\right)\right); \; S^d\left(i,i'\right) = 0 \end{cases} \tag{4}$$

where $p^d(i)$ indicates the permissive left-turn green time for approach i; $g_{c,0}^d(i')$ is opposing through movement green time at cycle c; $g_{c,1}^d(i')$ is the opposing left-turn protected green time at cycle c; $S^d(i,i') = 0$ is used to designate that the signal of two opposing approaches is operated concurrently; and $S^d(i,i') = 1$ is used to designate that the signal of two opposing approaches is operated separately.

Step 3: Combining all the event-based features. This step aims to obtain the event-based feature vector by matching these event-based features based on the intersection approach and the time period. On each approach, the event-based feature vector should have seven elements, as shown in Equation 5.

$$X_E^d(i) = \{o_0^d(i), o_1^d(i), c_0^d(i), c_1^d(i), \; g_0^d(i), \; g_1^d(i), p^d(i)\} \tag{5}$$

where $X_E^d(i)$ indicates the event-based feature vector for approach i at intersection d, including seven elements, which can be measured based on step 2.

After extracting and calculating the event-based feature dataset, the dataset requires further cleaning because missing values may exist in this vector due to data quality issues or different types of left-turn phasing.

Multi-layer perceptron neural network model

As described in the Problem Statement section, three major challenges need to be resolved for estimating TMC when using event-based data. The challenges caused by single-channel detectors covering multiple lanes and different types of detector configurations on different approaches result in complex uncertainties and scenarios to estimate TMC. To capture the non-linear and uncertain relationship between event-based data and TMC, a Multi-Layer Perceptron (MLP) Neural Network model is introduced. The MLP model is a class of fully connected feedforward artificial neural networks that can capture underlying relationships expressed in the problem without the need for prior assumptions (Zhang et al., 1998). The MLP model has been successfully used for segment traffic volume and annual average daily volume (AADT) estimation (Gastaldi et al., 2014; Sekuła et al., 2018). The developed MLP model used in this study includes four layers: the input layer, two hidden layers, the output layer, and the model structure, as is shown in Figure 5. Every node in a layer connects to each node in the following layer, making a fully connected network. The left layer is the input layer, which consists of a set of neurons representing

Table 2. Descriptive statistics of detector occupancy time during 15-min intervals.

Detector occupancy time (seconds)	Major road Left-turn	Major road Through	Minor road Left-turn	Minor road Through
Mean	468	445	526	590
SD	241	189	208	196
Minimum	2.9	17	0.1	12
25% Percentile	270	304	389	463
Median	458	431	538	621
75% Percentile	672	580	680	740
Maximum	900	900	900	900

the input variables. The input layer is formulated using Equation 6.

$$X^d(i) = \begin{bmatrix} X_E^d(i) \\ X_I^d(i) \\ X_P^d(i) \end{bmatrix} \tag{6}$$

In Equation 6, $X^d(i)$ is the input layer vectors for approach i at intersection d; $X_E^d(i)$ is the event-based feature vector; $X_I^d(i)$ is the vector consisting of features from intersection infrastructure data; and $X_P^d(i)$ is the vector consisting of features from POI data.

$$X_E^d(i) = \begin{bmatrix} o_0^d(i) \\ o_1^d(i) \\ c_0^d(i) \\ c_1^d(i) \\ g_0^d(i) \\ g_1^d(i) \\ p^d(i) \end{bmatrix}, X_I^d(i) = \begin{bmatrix} m^d(i) \\ n_l^d(i) \\ n_{ls}^d(i) \\ n_{th}^d(i) \\ n_r^d(i) \\ n_{rs}^d(i) \end{bmatrix} \tag{7}$$

In Equation 7, $X_E^d(i)$ consists of $o_0^d(i)$, $o_1^d(i)$, $c_0^d(i)$, $c_1^d(i)$, $g_0^d(i)$, $g_1^d(i)$, $p^d(i)$, and the definition and calculation methods are based on Equation 1 through Equation 5. In the vector $X_I^d(i)$, $m^d(i)$ is the road type at intersection d; $n_l^d(i)$, $n_{ls}^d(i)$, $n_{th}^d(i)$, $n_r^d(i)$, $n_{rs}^d(i)$ represent the number of the dedicated left-turn lane, shared-left turn lane, through lane, dedicated right-turn lane, and shared right-turn lane for approach i at intersection d.

Each neuron in the hidden layer transforms the values from the previous layer with a weighted linear summation followed by a non-linear activation function. The mathematical expression of the MLP model is represented in Equation 8.

$$z_k^{(q+1)} = f(\boldsymbol{w}_k^{(q+1)} \cdot \boldsymbol{z}^{(q)} + b_k^{(q+1)}) \tag{8}$$

where $z_k^{(q+1)}$ represents the output from k th neuron in layer $q+1$. $f()$ is an activation function that is used to capture nonlinear relationships; $\boldsymbol{w}_k^{(q+1)}$ denotes the vector of weights between the k th neuron of q layer and $q+1$ layer; b_k^{q+1} is the bias associated with k th neuron of the $q+1$ layer; $\boldsymbol{z}^{(q)}$ denotes the output vector from the neurons in layer q.

Accordingly, $z^{(0)}$ is the input layer vector; $z^{(1)}$, $z^{(2)}$ are vectors of two hidden layers; $z^{(3)}$ is the vector of the output layer. The output can also be represented as Equation 9.

$$Z^{(3)} = \begin{bmatrix} y_L^d(i) \\ y_T^d(i) \\ y_R^d(i) \end{bmatrix} = \begin{bmatrix} f\left(\sum_{s=1}^{n} w_{s,L}{}^{(2)} z_s{}^{(2)} + b_L^{(3)}\right) \\ f\left(\sum_{s=1}^{n} w_{s,T}{}^{(2)} z_s{}^{(2)} + b_T^{(3)}\right) \\ f\left(\sum_{s=1}^{n} w_{s,R}{}^{(2)} z_s{}^{(2)} + b_R^{(3)}\right) \end{bmatrix} \tag{9}$$

Where $Z^{(3)}$ is the vector of the output layer, consisting of three neurons. $y_L^d\ (i)$, $y_T^d\ (i)$ and $y_R^d\ (i)$ represent the left-turn volume, through movement volume, and right-turn volume on approach i at intersection d, respectively. $z_s{}^{(2)}$ represents the output value of s th neuron in layer 2. $w_{s,L}{}^{(2)}$, $w_{s,T}{}^{(2)}$, $w_{s,R}{}^{(2)}$ represent the weight value between the s^{th} neuron of layer 2 and the left-turn neuron, through movement neuron and right-turn neuron of the output layer, respectively. $b_L^{(3)}$, $b_T^{(3)}$, $b_R^{(3)}$ are the bias for left-turn, through movement, and right-turn of the output layer, respectively.

Implementation and results

Event-based feature analysis

Based on the proposed event-based feature extraction and processing framework, we obtained the event-based features from ninety-three study intersections in the Pima County region. Because the detector configuration on major roads and minor roads at intersections are inconsistent, we examined the distribution of detector occupancy time and detector-triggered count over these ninety-three intersections to further learn about the event-based feature differences between major roads and minor roads. The results are summarized in Tables 2 and 3. The maximum detector occupancy time is consistently 900 seconds, indicating the detector is fully occupied by vehicles during a 15-minute interval. The detector occupancy time on minor roads is higher than that of major roads. The detector-triggered count for the through movement on major roads is significantly greater than others. This outcome can be explained by the fact that the through movement on a major road is equipped with advance detectors that are more sensitive to the short gaps between two vehicle groups than presence detectors. A higher volume of through movement on major roads can also lead to a higher detector-triggered count than other movements.

Hyperparameter determination

The dataset used for the evaluation study is collected from ninety-three signalized intersections (refer to Figure 3 for study intersections). Each data entry corresponds to the number of 15-min interval TMC during peak hours (7:00-9:00 AM, 4:00-6:00 PM) on an approach of an intersection. The dataset is divided into training (80%) and testing (20%) by intersection. The training data has 45% of data from AM peak hours and 55% of data from PM peak hours. The testing data has 42% of data from AM peak hours and 58% of data from PM peak hours. Due to the communication loss of controllers during some periods, more ground-truth data from PM peak hours is used after overlapping with event-based data.

Table 3. Descriptive statistics of the detector-triggered count during 15-min intervals.

Detector-triggered count	Major road		Minor road	
	Left turn	Through	Left turn	Through
Mean	12	87	9	20
SD	12	35	4	9
Minimum	1	3	0	1
25% Percentile	7	65	7	13
Median	10	84	9	18
75% Percentile	13	105	11	24
Maximum	179	469	33	100

Before training the MLP model, the optimal number of neurons for each hidden layer has to be determined. A sensitivity analysis experiment is carried out to select the optimum number of hidden neurons, with the number of neurons in each layer increasing from 10 to 100 in 10 neurons increments. The performance, in terms of Root Mean Square Error (RMSE), is chosen as the error indicator in the number of neurons sensitivity analysis. Figure 6 shows the best performance of the model is achieved when 60 neurons are in hidden layer 1 and 40 neurons are in hidden layer 2. When the number of neurons is greater than 40 neurons, the decrease in the RMSE is not significant. After determining the number of neurons, the activation function and optimizer parameters are then selected based on the model performance. Based on the results, the MLP model using the Rectified Linear Unit (ReLu) for activation function and a quasi-Newton method L-BFGS (Byrd et al., 1995) as an optimizer is found to outperform with faster convergence.

Model performance analysis

The four indicators used to evaluate and quantify the MLP model performance, including Root Mean Square Error (RMSE), Mean Absolute Error (MAE),

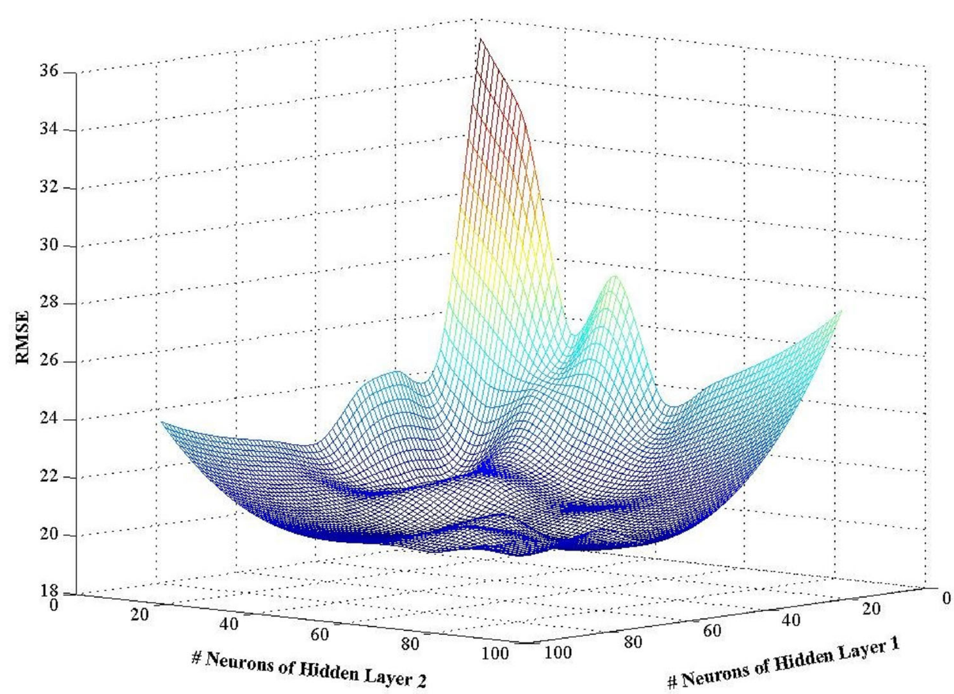


Figure 6. Model performance using different numbers of neurons.

Table 4. Performance measures for MLP-based TMC estimation.

	RMSE (veh/15 min)			MAPE			R^2			MAE (veh/15 min)		
5-Fold	T	L	R	T	L	R	T	L	R	T	L	R
1	47	15	22	38%	38%	63%	0.82	0.58	0.20	35	11	15
2	43	20	18	32%	51%	67%	0.83	0.67	0.34	30	12	13
3	48	14	29	38%	43%	52%	0.85	0.58	0.23	33	10	17
4	47	14	18	30%	40%	53%	0.80	0.56	0.32	33	10	12
5	61	17	19	33%	49%	62%	0.76	0.63	0.25	41	12	13

T: through movement; L: left-turn movement; R: right-turn movement.

statistic R-Squared, and Mean Absolute Percentage Error (MAPE) which is calculated after removing data with ground-truth data values of zero if applicable, are defined as follows:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{\left(Q_i^{Gr} - Q_i^{Es}\right)^2}{n}} \tag{10}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(Q_i^{Gt} - Q_i^{Es}\right)^2}{\sum_{i=1}^{n} \left(Q_i^{Gt} - \overline{Q^{Gt}}\right)^2} \tag{11}$$

$$MAE = \sum_{i=1}^{n} \left|Q_i^{Gr} - Q_i^{Es}\right| \tag{12}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left|\frac{Q_i^{Gr} - Q_i^{Es}}{Q_i^{Gr}}\right| \tag{13}$$

where Q_i^{Gr} is the ground-truth TMC; Q_i^{Es} is the estimated TMC; $\overline{Q^{Gt}}$ is the mean of ground-truth TMC; and n is the number of samples for verification.

To avoid under-fitting and over-fitting during model training with a limited training dataset, 5-fold cross-validation is used to test the effectiveness of the proposed model. This dataset is split into five groups by intersection, each four dataset groups are used for model training, and the fifth dataset group is used for model testing to evaluate the model performance on the dataset not used in the training process. Table 4 shows that the standard deviation of each indicator for different movements is always low, demonstrating that the MLP model has a robust performance using the different unseen datasets. The R-squared for through movement ranges from 0.76 to 0.85, and for left-turn movement range from 0.56 to 0.67. The right-turn movement has a slightly lower R-squared of 0.20 to 0.40. The results show the performance of the proposed model is robust during model training and the proposed model can be used for accurately estimating TMC at signalized intersections.

Figure 7 shows the model performance using different measures by intersection. The median RMSEs of the left-turn, through and right-turn volume estimation is 11 veh/15 min, 41 veh/15 min, and 12 veh/15 min, respectively. The median MAEs of left-turn, through movement, and right-turn volume estimation are 9 veh/15 min, 33 veh/15 min, and 10 veh/15 min, respectively. RMSEs and MAEs of through movement are higher than that of the left-turn and right-turn movements because the average traffic volume of through movement is relatively higher than the other two movements at most intersections. The median R-squared of the left turn, through movement, and right-turn movement is 0.48, 0.73, and 0.16, respectively. The median MAPEs of the left turn, through movement, and right-turn movement are 30%, 23%, and 39%, respectively. According to R-squared and MAPE, the TMC estimation of through movement yields the best performance among these three movements, likely because of the higher volume of through movement. Another reason for this high performance is the configuration of advance detectors on the major corridor through lanes, which is more sensitive to vehicle arrival than presence detectors on left-turn lanes. Also, right-turn vehicles may turn right at any time with a safe gap, regardless of the red light. Therefore, the number of right-turn vehicles is more difficult to accurately indicate using signal events. The left-turn movement count estimation slightly outperformed the right-turn movement, probably because some advance detectors and presence detectors configurations do not extend to right-turn lanes. A lack of right-turn event-based information is a probable cause of the lower R-squared value of the right-turn volume estimation.

To evaluate model performance at different intersections with different left-turn signal phases and intersection layouts, the model accuracy is investigated under different scenarios and summarized in Table 5. Table 5 shows that the model performances under different scenarios are relatively consistent, indicating the proposed model is reliable for accurately estimating TMC at most signalized intersections with varying characteristics. The intersections with permissive-only left-turn signal phases have a higher MAPE of estimated left-turn and right volume, possibly due to the lower volume at these intersections. Similarly, intersections with shared left-turn lanes may have a higher error of estimated through and right-turn volume, because two study intersections (River Rd & Campbell Ave (NB), Swan Rd & Camp Lowell Dr (EB) at Tucson, AZ) have higher error caused by one

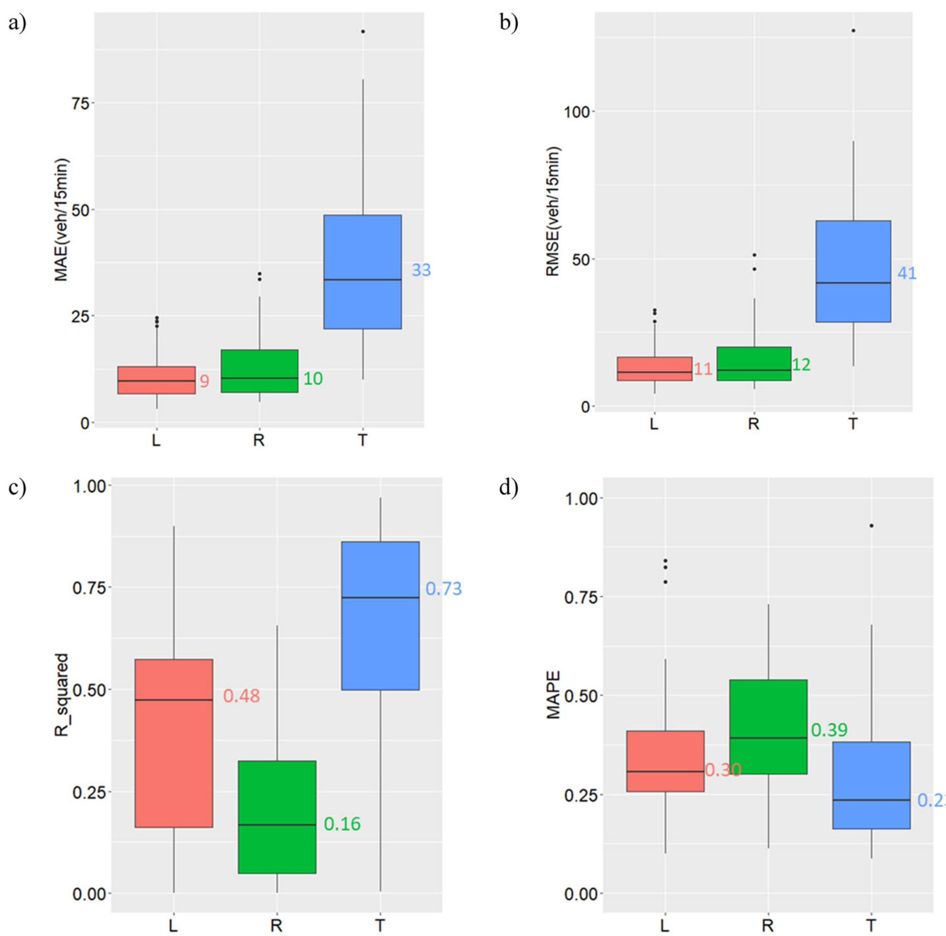


Figure 7. MLP-based model evaluation for three movements using the measures of (a) MAE, (b) RMSE, (c) R-squared, and (d) MAPE.

Table 5. Performance measure comparison under different scenarios.

		MAE (veh/15 min)			MAPE		
	Scenario	LT	TH	RT	LT	TH	RT
LT signal phase	PPLT	10	39	11	35%	30%	40%
	Permissive only LT	9	25	15	40%	32%	50%
	Protected only LT	12	40	17	28%	29%	30%
LT lane type	Exclusive LT	10	36	12	35%	31%	43%
	Shared LT	13	23	20	34%	80%	59%
RT lane type	Exclusive RT	12	38	14	36%	38%	45%
	Shared RT	8	35	12	33%	26%	42%
Number of TH lanes	1	10	11	18	38%	38%	52%
	2	10	31	12	35%	31%	44%
	3	9	61	11	33%	28%	36%
Number of LT lanes	1	9	35	12	34%	31%	43%
	2	18	40	16	41%	39%	45%

TH: through movement; LT: left-turn movement; RT: right-turn movement; PPLT: protected-permissive left-turn.

exclusive left-turn lane and one shared left-turn lane. This lane configuration cannot be captured well by the model, due to limited samples. In addition, the left-turn volume estimation error is less accurate on two left-turn lanes because of the dual left-turn lane configuration for a few study intersections, and this characteristic is not fully captured by the training model.

Sensitivity analysis of input information

The input data for the developed MLP model includes event-based features, infrastructure features, and POI information features. To analyze the effects of the input information on model performance, three types of MLPs with different input information are constructed.

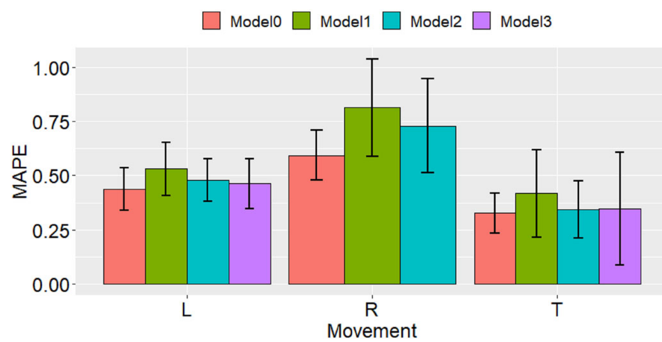


Figure 8. Influence of different input information on the MLP-based model performance using MAPE measures.

Model 0 is the proposed model with all input information and three output layers. Model 1 is an MLP-based model without infrastructure data. Model 2 is an MLP-based model without the POI information. Model 3 consists of two single-output MLP models to estimate through and left-turn volume, respectively. Two MLP models in Model 3, one uses the event-based data associated with through movement (i.e. occupancy time of the through movement detector, triggered count of the through movement detector, the green time duration of the through movement signal), infrastructure information, and POI information to estimate through movement volume. The other model uses the event-based data associated with left-turn movement (i.e. occupancy time of the left-turn movement detector, triggered count of the left-turn movement detector, and green time duration of the left-turn movement signal), as well as infrastructure and POI information to estimate the left-turn movement volume.

These models' performance for estimating 15-min TMC is evaluated and compared by using 10-fold cross-validation. Figure 8 visualizes both the median and variation of the MAPE of different models. The horizontal axis represents the movement at intersections. The MAPE of Model 0 is significantly lower than Model 1 in right-turn, left-turn, and through volume estimation, indicating that infrastructure information has a significant influence on the performance of the proposed model. Model 2 has a higher error than the proposed method in left-turn and right-turn volume estimation, and similar performance of through movement volume estimation. One possible explanation for this result is the impact POI data has on left-turn and right-turn volume, as most turning vehicles are going to nearby stores and thus lessening through volume. The proposed method slightly overperforms Model 3 in estimating left-turn and through volume, however, because Model 3 is unable to estimate right-turn volume or be applied at intersections with a shared left-turn lane. The interactions between left-turn and through-movement associated variables can slightly improve the model performance as well as estimate right-turn volume. The MAPE variation of Model 0 using different datasets is lower than the other three models, showing the reliability and robustness of the proposed method.

Conclusions

TMC data is critical for transportation planning, simulation modeling, and traffic signal timing optimization. Due to most intersections' lack of sensors configured for direct TMC collection, the region-wide TMC data collection process is challenging and time-consuming when using a manual collection method. The existing TMC estimation methods are difficult for region-wide implementation due to data availability and budget limit. This paper proposed a framework of event-based data processing and features extraction and developed an MLP-based approach for estimating 15-min interval TMC at signalized intersections using existing event-based data. Because event-based data is collected from existing traffic controllers, the data is wide-coverage and low-cost for use. The proposed method can benefit transportation agencies by providing a method to cost-effectively obtain region-wide TMC data. Three types of event-based features are extracted for use as crucial variables of MLP model inputs in this study, including vehicle occupancy time, detector-triggered count, and green time duration. To improve the TMC estimation accuracy for signalized intersections with different layouts, infrastructure data and POI data are included as exogenous variables for the MLP model. The developed MLP model is a multi-output neural network that can estimate left-turn, through movement, and right-turn volume for each approach at a signalized intersection. The ground-truth data is collected from ninety-three signalized intersections in the Pima County region, Arizona, and is used for model calibration and verification. The validation results yield a median RMSE of 41 veh/15 min, 11 veh/15 min, and 12 veh/15 min for through movement, left-turn movement, and right-turn movement volume estimation, respectively. These evaluation results indicate that the proposed MLP model can estimate TMC, showing that this research can benefit transportation agencies by enabling them to obtain region-level TMC through existing data sources, avoiding additional infrastructure and labor costs.

Although only the ground-truth data during peak hours is used to calibrate and verify the proposed method, this method can also be applied to estimate TMC during off-peak periods because event-based data can provide real-time information at any time of the day. Due to a lack of detectors configured on right-turn lanes at some intersections, information of the right-turn movement vehicles is difficult to obtain, causing the lower accuracy of right-turn volume estimation. Adding more features to the proposed model could be a potential solution to improving the accuracy of right-turn volume estimation. Evaluation results using MAPE show that left-turn and right-turn volume estimation has a high error because of typical low volume on left-turn and right-turn lanes. However, the TMC data provided by the proposed method has wide coverage and low cost, and the 15-min TMC could be aggregated into daily, monthly, and yearly volume data to yield more meaningful information to support most traffic studies.

Besides, some agencies may not have collected event-based data, one research direction is to use static variables, such as POI and infrastructure data, to estimate the average TMC at intersections. In the future, more ground-truth data will be collected from more sample locations to evaluate the feasibility of only using static variables to estimate the average TMC, such as annual average daily TMC. The proposed method will be further evaluated with data collected by intelligent sensors in the future. Regardless of the above-mentioned limitations, this paper proposed a method to estimate TMC data using existing data sources, saving time and funding.

Acknowledgments

The authors gratefully acknowledge the funding and data support received from the Pima Association of Governments. We would like to thank the City of Tucson for providing the traffic controller event-based data. We would also like to thank James Tokishi, Ryan Hatch, Josh Pope, and Paul Casertano for providing valuable advice and technical support in this project. Also, we want to thank David Klebosky for collecting the infrastructure data. Special thanks to Lilly Cottam and Adrian Cottam for their assistance in English proofreading.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Peipei Xu http://orcid.org/0000-0002-1781-3820
Xiaofeng Li http://orcid.org/0000-0001-5526-9961
Yao-Jan Wu http://orcid.org/0000-0002-0456-7915

References

An, C., Wu, Y.-J., Xia, J., & Huang, W. (2018). Real-time queue length estimation using event-based advance detector data. *Journal of Intelligent Transportation Systems*, *22*(4), 277–290. https://doi.org/10.1080/15472450.2017.1299011

An, C., Wu, Y.-J., Xia, J., & Lu, Z. (2017). Investigating impacts of communication loss on signal performance with use of event-based data. *Transportation Research Record: Journal of the Transportation Research Board*, *2645*(1), 38–49. https://doi.org/10.3141/2645-05

Balke, K. N., Charara, H., & Parker, R. (2005). *Development of a Traffic Signal Performance Measurement System (TSPMS) (No. FHWA/TX-05/0-4422-2)*. Texas Transportation Institute, The Texas A&M University System College Station, Texas.

Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, *16*(5), 1190–1208. https://doi.org/10.1137/0916069

Day, C., Bullock, D., Li, H., Remias, S., Hainen, A., Freije, R., Stevens, A., Sturdevant, J., & Brennan, T. (2014). *Performance measures for traffic signal systems: An outcome-oriented approach*. JTRP Affiliated Reports. Purdue University, West Lafayette, Indiana. https://doi.org/10.5703/1288284315333

Day, C., Haseman, R., Premachandra, H., Brennan, T., Wasson, J., Sturdevant, J., & Bullock, D. (2010). Visualization and assessment of arterial progression quality using high resolution signal event data and measured travel time. Transportation Research Record.

Day, C., Smaglik, E., Bullock, D., & Sturdevant, J. (2008). Real-time arterial traffic signal performance measures (No. FHWA/IN/JTRP-2008/09). Joint Transportation Research Program, Indiana Department of Transportation and Purdue University.

Day, C. M., & Bullock, D. M. (2020). Optimization of traffic signal offsets with high resolution event data. *Journal of Transportation Engineering, Part A: Systems*, *146*(3), 04019076. https://doi.org/10.1061/JTEPBS.0000309

Day, C. M., Bullock, D. M., & Sturdevant, J. R. (2009). Cycle-length performance measures: Revisiting and extending fundamentals. *Transportation Research Record: Journal of the Transportation Research Board*, *2128*(1), 48–57. https://doi.org/10.3141/2128-05

Day, C. M., Ernst, J. M., Brennan, T. M., Chou, C.-S., Hainen, A. M., Remias, S. M., Nichols, A., Griggs, B. D., & Bullock, D. M. (2012). Performance measures for adaptive signal control: Case study of system-in-the-loop simulation. *Transportation Research Record: Journal of the Transportation Research Board*, *2311*(1), 1–15. https://doi.org/10.3141/2311-01

Gastaldi, M., Gecchele, G., & Rossi, R. (2014). Estimation of annual average daily traffic from one-week traffic counts. A combined ANN-Fuzzy approach. *Transportation Research Part C: Emerging Technologies, Special Issue: Towards Efficient and Reliable Transportation Systems*, *47*, 86–99. https://doi.org/10.1016/j.trc.2014.06.002

Ghods, A. H., & Fu, L. (2014). Real-time estimation of turning movement counts at signalized intersections using signal phase information. *Transportation Research Part C: Emerging Technologies, Special Issue: Towards Efficient and Reliable Transportation Systems*, *47*, 128–138. https://doi.org/10.1016/j.trc.2014.06.010

Hu, H., & Liu, H. X. (2013). Arterial offset optimization using archived high-resolution traffic signal data. *Transportation Research Part C: Emerging Technologies*, *37*, 131–144. https://doi.org/10.1016/j.trc.2013.10.001

Li, X. (2021). *High-resolution data-based methods for arterial traffic volume estimation*. PhD dissertation, The University of Arizona.

Li, X., Weber, A., Cottam, A., & Wu, Y.-J. (2019). Impacts of changing from permissive/protected left-turn to protected-only phasing: case study in the city of Tucson, Arizona. *Transportation Research Record: Journal of the Transportation Research Board*, *2673*(4), 616–626. https://doi.org/10.1177/0361198119842108

Li, X., & Wu, Y.-J. (2021). Real-time estimation of pedestrian volume at button-activated midblock crosswalks using traffic controller event-based data. *Transportation Research Part C: Emerging Technologies*, *122*, 102876. https://doi.org/10.1016/j.trc.2020.102876

Li, X., Wu, Y.-J., & Chiu, Y.-C. (2019). Volume estimation using traffic signal event-based data from video-based sensors. *Transportation Research Record: Journal of the Transportation Research Board*, *2673*(6), 22–32. https://doi.org/10.1177/0361198119842120

Li, X., Xu, P., & Wu, Y.-J. (2021). Pedestrian crossing volume estimation at signalized intersections using Bayesian additive regression trees. *Journal of Intelligent Transportation Systems*, *2021*, 1–15. https://doi.org/10.1080/15472450.2021.1933471

Liu, H. X., & Ma, W. (2009). A virtual vehicle probe model for time-dependent travel time estimation on signalized arterials. *Transportation Research Part C: Emerging Technologies*, *17*(1), 11–26. https://doi.org/10.1016/j.trc.2008.05.002

Liu, H. X., & Ma, W. (2008). *Real-time performance measurement system for arterial traffic signals* [Paper presentation]. Presented at the Transportation Research Board 87th Annual MeetingTransportation Research Board.

Liu, H. X., Ma, W., Wu, X., & Hu, H. (2008). Development of a real-time arterial performance monitoring system using traffic data available from existing signal systems (Report). Minnesota Department of Transportation.

Liu, H. X., & Sun, J. (2014). Length-based vehicle classification using event-based loop detector data. *Transportation Research Part C: Emerging Technologies*, *38*, 156–166. https://doi.org/10.1016/j.trc.2013.11.010

Liu, H. X., Wu, X., Ma, W., & Hu, H. (2009). Real-time queue length estimation for congested signalized intersections. *Transportation Research Part C: Emerging Technologies*, *17*(4), 412–427. https://doi.org/10.1016/j.trc.2009.02.003

Mirchandani, P. B., Nobe, S. A., & Wu, W. W. (2001). Online turning proportion estimation in real-time traffic-adaptive signal control. *Transportation Research Record: Journal of the Transportation Research Board*, *1748*(1), 80–86. https://doi.org/10.3141/1748-10

National Transportation Operations Coalition. (2012). National traffic signal report card. ITE, Washington, DC.

Noh, H., Kramer, E., & Sun, A. (2019). Development of strategic regional employment data assessment using google places API. *Transportation Research Record: Journal of the Transportation Research Board*, *2673*(11), 254–263. https://doi.org/10.1177/0361198119852068

Sekuła, P., Marković, N., Vander Laan, Z., & Sadabadi, K. F. (2018). Estimating historical hourly traffic volumes via machine learning and vehicle probe data: A Maryland case study. *Transportation Research Part C: Emerging Technologies*, *97*, 147–158. https://doi.org/10.1016/j.trc.2018.10.012

Smaglik, E. J., Sharma, A., Bullock, D. M., Sturdevant, J. R., & Duncan, G. (2007). Event-based data collection for generating actuated controller performance measures. *Transportation Research Record: Journal of the Transportation Research Board*, *2035*(1), 97–106. https://doi.org/10.3141/2035-11

Urbanik, T., Tamaka, A., & Lozner, B. (2015). *Signal timing manual*. Transportation Research Board.

Virkler, M. R., & Kumar, N. R. (1998). System to identify turning movements at signalized intersections. *Journal of Transportation Engineering*, *124*(6), 607–609. https://doi.org/10.1061/(ASCE)0733-947X(1998)124:6(607)

Wu, X., & Liu, H. X. (2014). Using high-resolution event-based data for traffic modeling and control: An overview. *Transportation Research Part C: Emerging Technologies*, *42*, 28–43. https://doi.org/10.1016/j.trc.2014.02.001

Wu, X., Liu, H. X., & Gettman, D. (2010). Identification of oversaturated intersections using high-resolution traffic signal data. *Transportation Research Part C: Emerging Technologies*, *18*(4), 626–638. https://doi.org/10.1016/j.trc.2010.01.003

Zhang, G., Eddy Patuwo, B., & Y. Hu, M. (1998). Forecasting with artificial neural networks. The state of the art. *International Journal of Forecasting*, *14*(1), 35–62. https://doi.org/10.1016/S0169-2070(97)00044-7

Zhao, Y., Zheng, J., Wong, W., Wang, X., Meng, Y., & Liu, H. X. (2019). Various methods for queue length and traffic volume estimation using probe vehicle trajectories. *Transportation Research Part C: Emerging Technologies*, *107*, 70–91. https://doi.org/10.1016/j.trc.2019.07.008

Zheng, J., & Liu, H. X. (2017). Estimating traffic volumes for signalized intersections using connected vehicle data. *Transportation Research Part C: Emerging Technologies*, *79*, 347–362. https://doi.org/10.1016/j.trc.2017.03.007