BuildSenSys: Reusing Building Sensing Data for Traffic Prediction with Cross-domain Learning

Xiaochen Fan[*], Chaocan Xiang[*], Chao Chen, Panlong Yang
Liangyi Gong, Xudong Song, Priyadarsi Nanda, Xiangjiang He

Abstract—With the rapid development of smart cities, smart buildings are generating a massive amount of building sensing data by the equipped sensors. Indeed, building sensing data provides a promising way to enrich a series of data-demanding and cost-expensive urban mobile applications. In this paper, as a preliminary exploration, we study how to reuse building sensing data to predict traffic volume on nearby roads. Compared with existing studies, reusing building sensing data has considerable merits of cost-efficiency and high-reliability. Nevertheless, it is non-trivial to achieve accurate prediction on such cross-domain data with two major challenges. First, relationships between building sensing data and traffic data are not unknown as prior, and the spatio-temporal complexities impose more difficulties to uncover the underlying reasons behind the above relationships. Second, it is even more daunting to accurately predict traffic volume with dynamic building-traffic correlations, which are cross-domain, non-linear, and time-varying. To address the above challenges, we design and implement *BuildSenSys*, a first-of-its-kind system for nearby traffic volume prediction by reusing building sensing data. Our work consists of two parts, *i.e.*, *Correlation Analysis* and *Cross-domain Learning*. First, we conduct a comprehensive building-traffic analysis based on multi-source datasets, disclosing *how* and *why* building sensing data is correlated with nearby traffic volume. Second, we propose a novel recurrent neural network for traffic volume prediction based on cross-domain learning with two attention mechanisms. Specifically, a cross-domain attention mechanism captures the building-traffic correlations and adaptively extracts the most relevant building sensing data at each predicting step. Then, a temporal attention mechanism is employed to model the temporal dependencies of data across historical time intervals. The extensive experimental studies demonstrate that *BuildSenSys* outperforms all baseline methods with up to 65.3% accuracy improvement (*e.g.*, 2.2% MAPE) in predicting nearby traffic volume. We believe that this work can open a new gate of reusing building sensing data for urban traffic sensing, thus establishing connections between smart buildings and intelligent transportation.

Index Terms—Internet of Things, Building Sensing Data, Machine Learning, Traffic Prediction, Cross-domain Learning

✦

1 INTRODUCTION

SMART buildings equipped with an increasing number of IoT sensors are rising rapidly, thus producing large amount of building sensing data (also called *building data*[1] [1]–[3]). According to Statista's reports in [4], the explosive volume of sensing data collected by global smart buildings was nearly 7.8 ZB (about 7.8×2^{40}G) in 2015, and it is expected to be growing up to nearly 37.2 ZB by 2020. For instance, as shown in Fig. 1(a), a CBD building in Sydney is installed with more than 2,000 sensors, generating over 100 million sensor readings in monitoring the status of
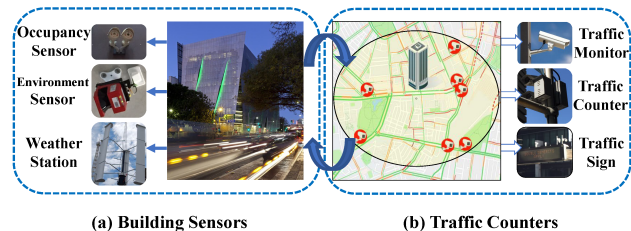

(a) Building Sensors **(b) Traffic Counters**

Fig. 1: Illustrations for reusing building sensing data to predict nearby traffic volume with cross-domain learning.

- *Xiaochen Fan, Xudong Song, Priyadarsi Nanda and Xiangjian He are with the School of Electrical and Data Engineering, Faculty of Engineering and Information Technology, University of Technology Sydney, Australia, NSW 2007; E-mails: {xiaochen.fan, xudong.song, priyadarsi.nanda, xiangjian.he}@uts.edu.au.*
- *Chaocan Xiang and Chao Chen are with the College of Computer Science, Chongqing University, Chongqing, China, 400044; E-mails: {xiang.chaocan, ivanchao.chen}@gmail.com.*
- *Panlong Yang is with the School of Computer Science and Technology, University of Science and Technology of China, Hefei, Anhui, China, 230026. E-mail: panlongyang@gmail.com.*
- *Liangyi Gong is with the School of Software and BNRist, Tsinghua University, Beijing 100084, China, E-mail: gongliangyi@gmail.com.*
- *Xiaochen Fan and Chaocan Xiang contributed equally to this work and share the first authorship. The corresponding authors are Chaocan Xiang and Xiangjian He.*

1. In the remaining of this paper, we will use the terms building sensing data and building data interchangeably unless otherwise stated.

the building, including building occupancy, indoor/outdoor environment, *etc.*

Reusing existing building sensing data is with great significance for ubiquitous sensing of smart cities, as building data is cost-efficiency and sustainable [5]. Assuming building sensing data is available for the long-term, it has recently enabled a series of new applications in urban sensing, such as developing urban 3D mobility models [6] and identifying abnormal appliances [7], [8]. In this paper, as a preliminary exploration (illustrated in Fig. 1), we are motivated to reuse building sensing data for nearby traffic sensing for the following reasons. Intuitively, most urban buildings are connected by roads, while residents move among different buildings, mainly via commuting on these

roads [9]. According to the report in [10], the American citizens averagely spend over 93% of their daily time in enclosed buildings and vehicles (accounting for 87% and 6%, respectively). Thus, there could exist some underlying relationships between sensing data of buildings and traffic data of buildings' surrounding roads, named as *building-traffic correlations* [8]. Such relationships reveal a promising direction of reusing building sensing data, *i.e.*, for predicting traffic of nearby roads. As a proof-of-concept, this work mainly focuses on predicting *traffic volume, which is defined as the number of vehicles traversing on a road segment per hour* [11]. Accurate predictions on traffic volume are fundamentally crucial for Intelligent Transportation Systems (ITSs), such as traffic light control, road navigation, and estimation of vehicle emission [8], [12], [13].

Compared with the existing methods [11], [14]–[18] of predicting nearby traffic volume, reusing building data has considerable merits of both *low cost* and *high reliability*, which are crucial to traffic sensing [13]. Most of the conventional traffic prediction methods heavily rely on the fixed road-based traffic sensing systems, such as loop detectors [14] and traffic surveillance cameras [15]. Although these methods can provide accurate traffic volume information, they incur extremely high costs on installation and maintenance, impeding their extensions to large-scale cities [8]. On the contrary, re-using off-the-shelf building sensing data will significantly reduce the cost, as with no extra deployment and maintenance [8], [11]. In terms of opportunistic sensing data of floating vehicles, GPS trajectories [11] and cellular records of passengers [18] are recently exploited to infer traffic volume to address the high-cost issue. However, opportunistic sensing suffers from the uncontrollable property of users and data deficiency of particular roads, resulting in unreliable performance in traffic volume prediction [18]. Buildings and their surrounding roads, as two basic constructions in urban infrastructure, are both stationary with long-term spatial relations. Therefore, in comparison to the opportunistic sensing [11], buildings can generate more reliable and sustainable sensing data, providing opportunities for accurately predicting nearby traffic volume for long term [17], [19].

Nevertheless, it is necessary to formally address two principal challenges in achieving reusing building sensing data for predicting nearby traffic volume.

- **Challenge 1**: *Investigating unknown building-traffic relationships*. While a few works [6], [8] have shown certain evidence of building-traffic correlations, they failed to investigate what exactly these relationships are. To precisely reveal building-traffic correlations is difficult with great complexity. Furthermore, it is extremely challenging to uncover the underlying reasons behind these correlations.
- **Challenge 2**: *Accurate prediction with cross-domain building-traffic correlations*. The building-traffic relationships are cross-domain and non-linear, which further renders accurate prediction of traffic volume as non-trivial. Even more challenging, these correlations would vary dynamically along with time, imposing more difficulties on accurate traffic prediction [20].

To address these two challenges, we design and implement *BuildSenSys*, a first-of-its-kind *Build*ing *Sen*sing data-based *Sys*tem for nearby traffic volume prediction. First, for Challenge 1, we conduct extensive experiments based on multi-source real-world datasets (in Section 4). We delve into the relationships between building data and nearby traffic volume, and the experimental results indicate that building-traffic correlations are *non-linear*, *time-varying*, and *cross-domain*. Second, for Challenge 2, we propose a novel recurrent neural network based on cross-domain learning with two attention mechanisms for traffic volume prediction (in Section 5). Specifically, a cross-domain attention mechanism captures the building-traffic correlations and adaptively extracts the most relevant building data at each time interval. Then, a temporal attention mechanism is employed to model the temporal dependencies across historical time intervals. Finally, we implement a prototype system of *BuildSenSys* and conduct comprehensive evaluations with one-year real-world datasets.

In summary, we make three key contributions in this paper as follows:

- To the best of our knowledge, we are the first to conduct comprehensive building-traffic analysis with multi-source real-world datasets. By applying multi-source cross-verification, this work not only discloses *how* but also sheds light on *why* the building data is correlated with nearby traffic volume. Essentially, the changes in building occupancy induced by the commuting activities of occupants are highly related to the dynamics of traffic volume on nearby roads. In addition, the higher probability building occupants pass through a road, the stronger building-traffic correlation there exists.
- We propose a novel recurrent neural network for traffic volume prediction based on cross-domain learning. It leverages a cross-domain attention-based encoder and a temporal attention-based decoder to extract the *non-linear*, *time-varying*, *cross-domain* building-traffic correlations accurately and further achieves accurate traffic volume prediction.
- The extensive experimental studies demonstrate that *BuildSenSys* outperforms all baseline methods with up to 65.3% accuracy improvement (*e.g.*, 2.2% mean absolute percentage error) in predicting nearby traffic volume. We believe this work can open a new gate of reusing building sensing data for traffic sensing, and further establish connections between smart buildings and intelligent transportation.

The rest of this paper is organized as follows. Section 2 reviews related literature and Section 3 introduces an overview of the *BuildSenSys* system. Then, Section 4 presents the correlation analysis of building data and nearby traffic data. Section 5 formulates the prediction problem and presents the cross-domain learning-based recurrent neural network for traffic prediction. Section 6 evaluates *BuildSenSys* through extensive experimental studies with real-world datasets. Finally, Section 7 discusses some critical issues of reusing building data, and Section 8 concludes the paper.

2 RELATED WORK

Reusing building data: Building sensing data is originally dedicated to management and control purposes [21]; therefore, reusing building sensing data for traffic prediction will not cost any extra cost. Beyond building management, reusing building sensing data has attracted considerable research interests from smart city applications, including urban transportation [8], crowd flow patterns [6], and data integration [22]. According to [23], a single set of traffic monitoring systems with camera detectors could easily cost $2500 USD, and over 100 million dollars of such devices can only cover a quarter of roads in a typical metropolitan city [8]. In comparison with conventional methods [11], [15]–[17] that fully rely on sensing data from traffic monitoring systems, it is considerably low-cost while highly reliable to make a second use of building sensing data to predict nearby traffic. For example, Zheng *et al.* [6] studied the impacts of buildings on human movements and further developed a new urban mobility model for urban planning. Hu *et al.* [24] proposed a communication sharing architecture for smart buildings to organize in-building IoT devices with heterogeneous data communication. Moreover, Zheng *et al.* [8] proposed to use indoor CO_2 data to estimate building occupancy and further developed an occupancy-traffic model for traffic speed prediction. Inspired by the above existing works, it is of great significance to unify traffic monitoring systems with external sensing infrastructures to enhance traffic prediction accuracy and reduce marginal cost. In this study, we propose innovative reuse of multi-dimension building sensing data for traffic prediction. As a preliminary exploration, we extensively investigate the building-traffic correlations and apply cross-domain learning to achieve accurate traffic prediction.

Traffic volume prediction: Most existing works of predicting traffic volume use data from fixed road-based traffic sensors, such as loop detectors, microwave radars, and video cameras [13]. The main advantage of road-based sensors is that they can provide reliable data by capturing all vehicles passing by the corresponding roads [25]. Recently, opportunistic sensing and crowdsensing techniques have been utilized to collect GPS data [11], [26], [27], and cellular record data [28]–[30] from floating vehicles and mobile passengers. These trajectories provide detailed mobility traces for network-wide traffic sensing and prediction. Many existing works have integrated both traffic sensor data and opportunistic sensing data for traffic prediction. For example, Meng *et al.* [31] further combined loop detector data and taxi trajectories with a spatio-temporal semi-supervised learning model to infer traffic volume. However, with the inherent biases and random data deficiency, most trajectory data cannot cover the entire traffic dynamics. For instance, GPS data of a 6,000-taxi network can only cover 28% of the overall road segments in a large city with distinctive operating time [18]. With such data sparsity, trajectory sensing data cannot guarantee the sustainability and reliability in traffic prediction, especially on explicitly targeted road segments. In this work, the source data for traffic prediction is building sensing data, which is quite different from most existing works. Both the buildings and their surrounding roads are stationary, and inherently they have permanent spatial
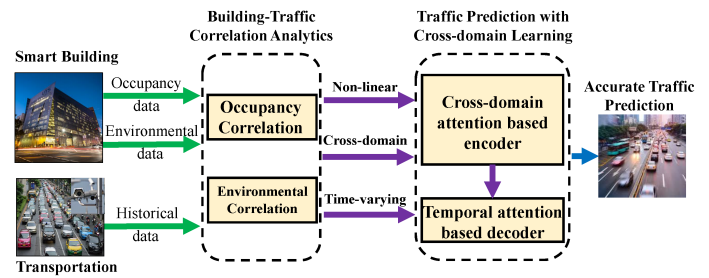

Fig. 2: An overview of the BuildSenSys system for reusing building data for nearby traffic prediction.

relations. Most importantly, as building-traffic correlations are sustainable, reusing building data is cost-efficiency and highly reliable for long-term traffic prediction.

Traffic prediction models: Conventional short-term traffic prediction approaches mainly apply parameter-based prediction models, for example, Autoregressive Integrated Moving Average (ARIMA) [32], Vector Autoregression (VAR) [33] and Locally Weighted Linear Regression (LWR) [8]. With the rising of deep learning, deep neural networks [34] have been adopted for traffic prediction as they can capture complex temporal-spatial dependencies through feature learning [35]. Moreover, Recurrent Neural Networks (RNN) have also been adopted by [36] to perform sequence learning on historical traffic data. However, RNNs are not capable of preserving long-term dependencies on historical traffic data, as their performance would deteriorate with longer input. As a result, Long Short-Term Memory (LSTM) networks are further adopted by many research studies [37], [38] to perform long-term prediction tasks. Inspired by the human's ability to capture a focus in certain visions, attention mechanisms have been integrated into the neural networks for sequence-to-sequence learning [39]. For instance, [40] proposed a spatial-temporal dynamic network with a periodically shifted attention mechanism to capture periodic temporal similarity in traffic predictions. In this paper, we devise a cross-domain learning-based recurrent neural network with a cross-domain attention mechanism and a temporal attention mechanism. Our model can effectively extract cross-domain, non-linear, and time-varying building-traffic correlations for accurate traffic prediction.

3 SYSTEM OVERVIEW

In this section, we briefly present the system overview of reusing building data for traffic volume prediction. As shown in Fig. 2, the data sources of *BuildSenSys* include a smart building (generating occupancy data and environmental data) and nearby traffic monitors (providing traffic volume data). More importantly, *BuildSenSys* consists of two main components as follows.

1) **Building-traffic correlation analysis with multi-source real-world datasets** (in Section 4). Extensive experimental analysis is conducted based on real-world smart building data, traffic data of fixed-road sensors, and Google traffic data. We investigate not only whether building data is correlated with the traffic data but also what the underlying reasons are. The results show that the answer is positive,
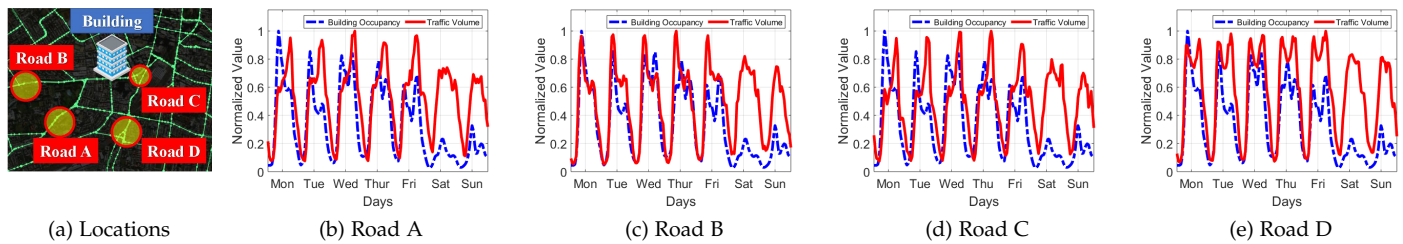
Fig. 3: Comparisons of normalized building occupancy and normalized traffic volume on different roads.

and there are two types of cross-domain correlations, including i) **occupancy correlation** (in Section 4.1) between the building occupancy and traffic data, and ii) **environmental correlation** (in Section 4.2) between the building environmental data and traffic data. Besides, the building-traffic correlations are dynamic and time-varying, generating the temporal correlations along with the time.

2) **Cross-Domain Learning for Accurate Traffic Prediction** (in Section 5). In correspondence to cross-domain correlations and temporal correlations, we propose cross-domain learning to predict traffic volume accurately. Specifically, we propose a cross-domain attention mechanism for the encoder to learn features from non-linear, cross-domain correlations between building data and traffic data (in Section 5.2). By correlating these features with traffic volume data, the most relevant cross-domain correlations are extracted through the training process of the *BuildSenSys* model. Furthermore, we present a temporal attention mechanism for the decoder, which aims to learn the temporal dependencies for the predicted traffic data (in Section 5.2.2). After that, the output context factors are leveraged to predict traffic volume.

4 BUILDING-TRAFFIC CORRELATION ANALYSIS WITH MULTI-SOURCE DATASETS

In this section, we conduct comprehensive experiments to explore spatial and temporal building-traffic correlations with multi-source real-world datasets as follows. To begin with, we briefly introduce the basics of building sensing data and traffic volume data as follows.

First, we collect building sensing data from the Faculty of Engineering and Information Technology's Building (FEIT Building) at the University of Technology Sydney, New South Wales, Australia. The FEIT building is a 16-level campus building with a total usable floor area of 23,500 m^2 and it has the capacity to accommodate a maximum population of over 5000. As illustrated in Fig. 1(a), the FEIT building is described as a 'living laboratory,' with around 2,500 internal environment sensors installed across all floors and public spaces. These sensors constantly monitor the internal and external environments of the building, including indoor environment (by environmental sensors), outdoor environment (by a roof-top weather station), and building occupancy (by smart cameras). We access all building sensing data through an online database server via MySQL workbench, which contains 33 types of building sensing data with a total volume of over 10 GB [41]. In this paper,

we have leveraged 10 most relevant building sensing data from FEIT building to achieve cross-domain traffic sensing and prediction.

Second, as shown in Fig. 1(b), the traffic data are the traffic volumes collected by permanently deployed traffic counters on several road segments that are proximate to the FEIT building. Each traffic counter is integrated to a traffic monitoring system for calculating hourly traffic volume data. We access historical traffic volume data from the official website of the Department of Roads and Maritime Services, New South Wales State [42]. In the following, we present a detailed cross-domain correlation analysis with traffic volume data and different building sensing data.

4.1 Correlation analysis with building occupancy data

In this section, we investigate cross-domain correlations between building occupancy data and traffic volume. The building occupancy data is collected by a number of cameras installed at building entrances, stairways, and walkways to monitor the people's movement inside the building. Based on the sensing data of these cameras, we use the PLCount algorithm [43] to compute the total building occupancy accurately. As the calculation of building occupancy is not our focus, we omit its details that can be referred from [43].

4.1.1 Correlation quantification with metrics

With the preprocessed building occupancy, we closely study the correlations between building occupancy and nearby traffic volume. First, we compare building occupancy with the traffic volume of nearby roads in one week. Fig. 3(a) shows the comparison of normalized building occupancy (dashed blue line) and traffic volume (solid red line) of roads A-D. Note that the above four roads are different types of roads, where road A is a minor highway, road B is a major highway, road C is a main street, and road D is a primary street. We observe that the building occupancy and traffic volume exhibit similar hourly patterns on weekdays and different patterns on weekends, respectively. This is because the high volume of building occupancy is driven by the working days, and there are just a few overtime workers inside the building on weekends. In particular, the building occupancy closely follows the rise of traffic volume in both morning rush hours and evening rush hours.

However, the comparison between normalized building occupancy and traffic volume only shows the general dynamics and correlations. To further quantify the correlations between building occupancy and traffic
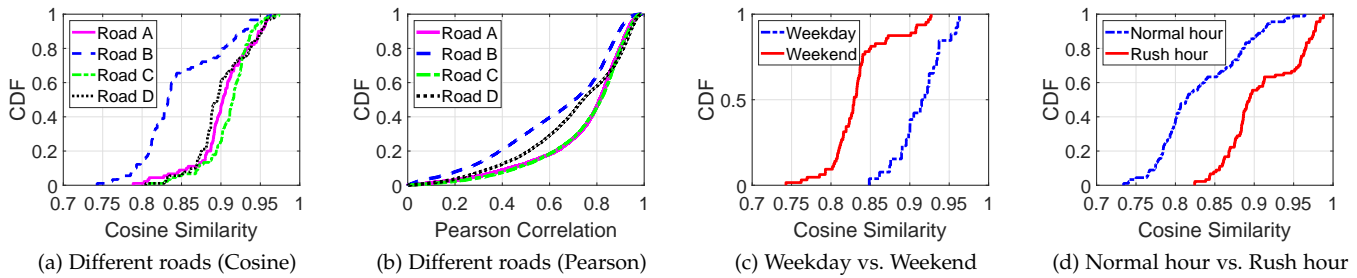
Fig. 4: Quantification of correlations between building occupancy data and traffic volume data with two metrics.

volume, we employ two correlation metrics, *i.e.*, Cosine Similarity and Pearson Correlation Coefficient, to evaluate the above correlations. The *Cosine Similarity* quantifies the similarity between two non-zero vectors on an inner product space. We use s_i to represent the cosine similarity between building occupancy and traffic volume on the day i. It is computed as $s_i = \cos(\mathbf{b}_i, \mathbf{t}_i)$, where $\mathbf{b_i} = [b_{i,1}, b_{i,2}, ..., b_{i,24}]$ and $\mathbf{t_i} = [t_{i,1}, t_{i,2}, ..., t_{i,24}]$ denote the vectors of building occupancy and traffic volume over 24 hours on the day i, respectively. In this way, the overall cosine similarity between building occupancy and traffic volume for n days is denoted by $\mathbf{s} = [s_1, ...s_i, ...s_n]$. Moreover, the *Pearson Correlation* is computed as $p = \sum_{j=1}^{k} (b_{i,j} - \bar{b}_i)(t_{i,j} - \bar{t}_i) / \sqrt{\sum_{j=1}^{k} (b_{i,j} - \bar{b}_i)^2 \sum_{j=1}^{k} (t_{i,j} - \bar{t}_i)^2}$, where k denotes the time interval for each Pearson Coefficient, $\bar{b}_i$ and $\bar{t}_i$ are the average building occupancy and traffic volume at the day i, respectively. Note that the absolute value of Pearson Correlation indicates the strength of the correlation, ranging from 0 (weak correlation) to 1 (strong correlation).

We quantify correlations between building occupancy and traffic volume on four different road segments, using both Cosine Similarity and Pearson Correlation. As illustrated in Fig. 4a, the cosine similarities on all four roads are greater than 0.8, showing that building occupancy is strongly correlated with traffic volume. Meanwhile, Fig. 4b shows that over 70% of traffic volume data has strong and positive (≥ 0.5) Pearson Correlation with building occupancy data. In addition, we evaluate the cosine similarity of building occupancy and traffic volume by dividing the data into different groups, *i.e.*, weekdays/weekends and rush hours/normal hours, respectively. From Fig. 4c and Fig. 4d, it is clear that building occupancy and traffic volume have higher correlations on the weekdays (≥ 0.85) and rush hours (≥ 0.82) than the weekends (≥ 0.75) and normal hours (≥ 0.75), respectively.

4.1.2 Correlation verification with Google Maps

The results in Section 4.1.1 indicate that the building occupancy data is highly correlated with the traffic volume on nearby road segments. Through an in-depth analysis, the main reason is that most of the building occupants entering or leaving a building would pass the surrounding roads via transportation. Intuitively, the building-traffic correlations are dependent on the probability of building occupants passing by each road segment. That is, the building-traffic

(a) Bird's-eye view of navigation map

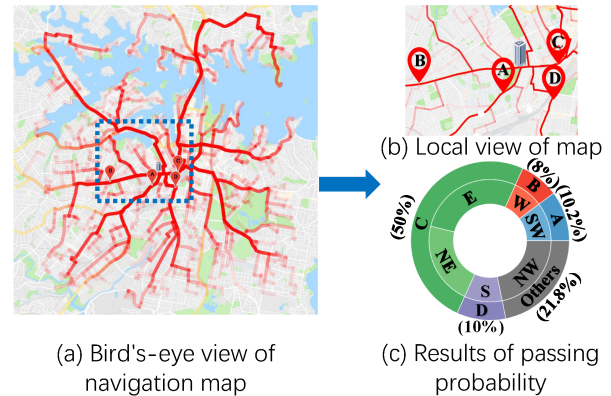(b) Local view of map

(c) Results of passing probability

Fig. 5: Cross-verification for building-traffic correlations via Google Maps Navigation.

correlation would be stronger when a road is with higher passing probability by the building's occupants.

To verify this assumption, we further investigate the relationship between building occupancy and traffic volume by conducting cross-verification experiments with Google Maps. First, as shown in Fig. 5(a), we set this building's location as the starting points or the endpoints, while randomly selecting 500 locations within 10 km distance of this building as the end/starting points. Each randomly selected point together with the building's location are packaged as a navigation request, which is sent to Google Maps for navigation. Then, the best-fit road geometries are returned with a series of GPS locations. We visualize 500 navigation results in Fig. 5, where the color of each route indicates its passing probability and a route with deeper color has higher passing probability.

As illustrated in Fig. 5(a), some long-distance routes would pass through the main roads, for example, major highways and primary streets. Moreover, Fig. 5(b) shows a zoom-in view of roads A-D. Specifically, the passing probabilities of roads A-D are shown in Fig. 5(c), where the outer circle represents the passing probability of four road segments, and the inner circle represents the main directions of navigation routes. It can be observed that the passing probabilities of roads A, B, C, D are 10.2%, 8%, 50% and 10%, respectively, while the others only have about 20% passing ratio. By using the cross-verification method, we compare these results based on Google Maps with building-traffic correlations in Section 4.1.1. As indicated in Table 1,

Road Name	Passing Ratio	Cosine	Pearson	Distance
Road C	50%	1st	1st	0.3 km
Road A	10.2%	2nd	2nd	1.6 km
Road D	10%	3rd	3rd	1.9 km
Road B	8%	4th	4th	3.0 km

TABLE 1: Cross-verification: comparison between roads A, B, C and D in navigation passing probability, Cosine similarity, Pearson correlation and distance to the building.
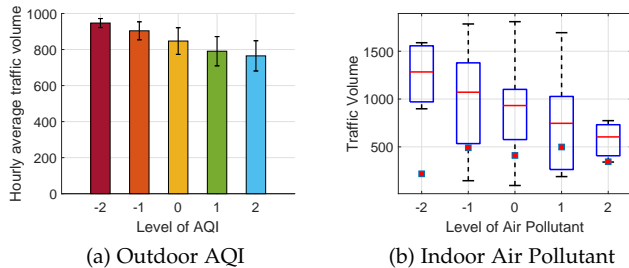

|(a) Outdoor AQI | (b) Indoor Air Pollutant|

Fig. 6: Correlation analysis between the building environmental data and traffic data (-2 stands for the worst level, 0 for the moderate level and 2 for the best level).

the results from multi-source datasets are consistent with each other. That is, a road segment with higher passing probability and shorter distance to the building would have stronger building-traffic correlations with this building. As a result, the above results of cross-verification further verify the reason behind building-traffic correlation as follows. *The building occupants entering or leaving a building would pass the surrounding roads via transportation. Hence, the change of building occupancy data (induced by the commuting activity of the building's occupants) is highly related to the traffic volume of nearby road segments.*

4.2 Correlation analysis with environmental data

Besides building occupancy, environmental data have been proved to affect traffic on the nearby roads evidently, and they can also contribute to the improvement of traffic prediction accuracy [16], [44]. In this study, for cross-domain traffic prediction, we typically select 5 types of indoor environmental data (CO_2 concentration, building humidity, O_2 concentration, building temperature and building air pollution) and 4 types of outdoor environmental data (including outdoor temperature, rainfall, wind speed and air quality index) from the FEIT building. Through conducting a series of correlation studies with environmental data and traffic volume data, we find that different types of environmental data have different levels of correlations with traffic volume data. While some of the environmental data have stronger correlations with outdoor traffic volume, others show weak correlations. To this end, we report two representative cases for correlations between environmental data and traffic data as follows.

Fig. 6a presents correlations between the outdoor Air Quality Index (AQI) and traffic volume by averaging values of traffic data under categorized air quality conditions. We categorize air quality values into five levels from the worst to the best, represented by -2 to 2. It can be observed that

AQI stays at a lower level when the traffic volume is in higher values. The main reason is that when the daily traffic volume is at a high level, more emissions from vehicles can directly pollute the air quality, thereby affecting outdoor AQI.

Moreover, we evaluate the correlations between traffic volume and indoor air pollutant levels. As shown in Fig. 6b, the box plot illustrates averaged traffic volume with categorized air pollutant levels. It can be observed that the indoor air pollutant becomes worse with higher traffic volume. Although indoor environmental factors are not directly related to outdoor traffic volume, the indoor environmental data can be used as the indirect sensing data for building occupants [8]. An essential objective of using indoor environmental data is to assist traffic prediction when the value of building occupancy is at abnormal values. For example, in case of an emergency evacuation, there will be a rapid removal of people inside the building, which will result in higher but false occupancy data generated by smart cameras. If we combine indoor environmental data in the prediction model, the prediction errors in the above cases can be eliminated effectively.

To conclude, the environmental data is correlated with the traffic volume; meanwhile, their correlations are not very strong. Nevertheless, the environmental data and the occupancy data can be complementary to each other in cross-domain traffic prediction with building data.

5 ACCURATE TRAFFIC PREDICTION WITH CROSS-DOMAIN LEARNING OF BUILDING DATA

In this section, based on the correlation analysis in Section 4, we propose a cross-domain learning method for traffic volume prediction by reusing building sensing data. First, we formulate the traffic volume prediction problem of using cross-domain building sensing data. Second, to address this problem, we propose a cross-domain learning-based recurrent neural network to learn the non-linear, time-varying, and cross-domain building-traffic correlations for accurate traffic volume prediction.

5.1 Problem Formulation

In this work, our goal is to predict traffic volume by leveraging building sensing data and historical traffic volume data. Assume that there are N types of building sensing data for T time intervals, we further introduce the following notations.

- **Building data types**: a building's IoT sensors generate N types of sensing data that are used for traffic prediction, including N_o types of occupancy sensing data (*e.g.*, covering different public zones) and N_e types of environmental sensing data (*e.g.*, indoor environmental and outdoor environmental data). Intuitively, $N = N_o + N_e$.
- **Building sensing data**: let $x_t(i)$ denote i-th building sensing data at time t and $\mathbf{x}_t = \{x_t(i) | 1 \leq i \leq N\}$ denote a vector of all building sensing data at time t. Accordingly, $\mathbf{X}_{[1:T]} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_T]_{T \times n}$ denotes a measurement matrix of all building data across T time intervals.

- **Traffic volume data**: let y_t represent traffic volume of a target road segment at time t, $1 \le t \le T-1$. Accordingly, $\mathbf{y}_{[1:T-1]}$ denotes a vector of historical traffic volume of the target road segment across $T-1$ time intervals, where $\mathbf{y} = \{y(j)|1 \le j \le T-1\}$.
- **Future Traffic volume**: the predicted traffic volume of a target road segment is denoted as $\widehat{\mathbf{Y}}_{[T:T+\tau]} = \{y(j)|T \le j \le T+\tau\}$ while using $\mathbf{Y}_{[T:T+\tau]}$ to represent its ground truth, where τ is the time intervals for prediction.

Problem Definition: formally, given n types of building sensing data over T time intervals and historical data of a target road segment over $T-1$ time intervals, the traffic volume prediction problem for τ future time intervals is to optimize the prediction errors and defined as:

$$\text{Minimize} \quad ||\widehat{\mathbf{Y}}_{[T:T+\tau]} - \mathbf{Y}_{[T:T+\tau]}||_F^2, \qquad (1)$$

$$\text{where} \quad \widehat{\mathbf{Y}}_{[T:T+\tau]} = F(\mathbf{y}_{[1:T-1]}, \mathbf{X}_{[1:T]}), \qquad (2)$$

where $F(\cdot)$ denotes the nonlinear mapping function from building sensing data to traffic data that a predicting model needs to learn.

5.2 Attention mechanisms-based encoder-decoder Recurrent Neural Network

To solve the cross-domain learning-based traffic volume prediction problem, we propose *BuildSenSys*, an LSTM based encoder-decoder architecture with dual-attentions mechanisms. First, we employ cross-domain attention on input data to capture building-traffic correlations and adaptively select the most relevant building sensing data at each step of prediction. Second, we apply temporal attention to capture temporal features from historical dependencies. Then, *BuildSenSys* adaptively select the most relevant encoder hidden states across all time intervals. By integrating the above attention mechanisms with LSTM-based recurrent neural networks, we can further jointly train *BuildSenSys* model with standard back propagation. As a result, *BuildSenSys* is capable of selecting the most relevant building sensing data for traffic prediction and capture long-term temporal features of traffic volume. The overall framework of *BuildSenSys* is presented in Fig. 7.

5.2.1 Encoder with cross-domain attention

The encoder in the *BuildSenSys* framework is an LSTM-based recurrent neural network, and it encodes the input sequence into a feature vector. For cross-domain traffic volume prediction, given N types of building sensing data, we denote the input sequence as $\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_t, ..., \mathbf{x}_T)$, where $\mathbf{x}_t \in \mathbb{R}^N$. The hidden state of the encoder at time interval t is computed by

$$\mathbf{h}_t = f_e(\mathbf{h}_{t-1}, \mathbf{x}_t), \qquad (3)$$

where $\mathbf{h}_{t-1} \in \mathbb{R}^p$ is the previous hidden state of the encoder at time interval $t-1$; p is the size of the hidden state in the encoder; and f_e is an LSTM based recurrent neural network. Since LSTMs are capable of learning long-term dependencies, we employ the classic LSTM unit that is with one memory cell and three sigmoid gates. At time interval t, the cell state of memory is $\mathbf{s}_t$, the forget gate is $\mathbf{f}_t$, and the
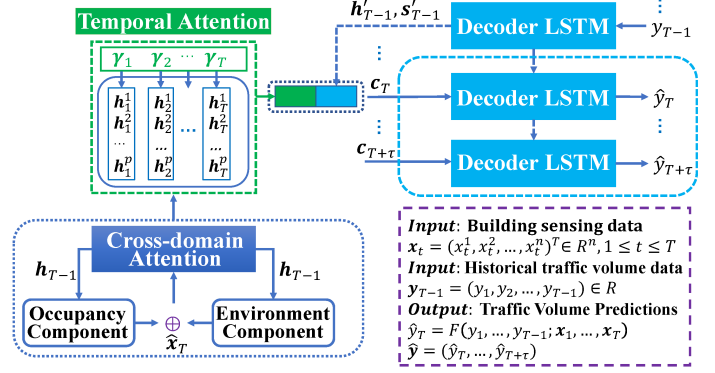

Fig. 7: The graphical architecture of cross-domain attention-based recurrent neural networks for cross-domain traffic prediction.

input gate is $\mathbf{i}_t$. The encoder LSTM updates its hidden state by

$$\mathbf{f}_t = \sigma(\mathbf{W}_f[\mathbf{h}_{t-1}; \mathbf{x}_t] + \mathbf{b}_f), \qquad (4)$$

$$\mathbf{i}_t = \sigma(\mathbf{W}_i[\mathbf{h}_{t-1}; \mathbf{x}_t] + \mathbf{b}_i), \qquad (5)$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_o[\mathbf{h}_{t-1}; \mathbf{x}_t] + \mathbf{b}_o), \qquad (6)$$

$$\mathbf{s}_t = \mathbf{f}_t \odot \mathbf{s}_{t-1} + \mathbf{i}_t \odot tanh(\mathbf{W}_s[\mathbf{h}_{t-1}; \mathbf{x}_t] + \mathbf{b}_s), \qquad (7)$$

$$\mathbf{h}_t = \mathbf{o}_t \odot tanh(\mathbf{s}_t), \qquad (8)$$

where $[\cdot;\cdot]$ is a concatenation operation; σ is a logistic sigmoid function; $\odot$ is a pointwise multiplication; $\mathbf{W}_f$, $\mathbf{W}_i$, $\mathbf{W}_o$, $\mathbf{W}_s$ and $\mathbf{b}_f$, $\mathbf{b}_i$, $\mathbf{b}_o$, $\mathbf{b}_s$ are the learnable parameters.

Inspired by that visual attention allows human to focus on a certain region of images or sentences for creating the perception of information, attention mechanism has become an integral part of the compelling sequence modeling and transduction models [39]. In general, attention mechanisms are proposed to make a soft selection over historical data by calculating and assigning different weights to them. In this work, we aim to predict the traffic volume with different types of building sensing data and historical traffic volume data. To achieve this goal, we propose cross-domain attention to capture complex building-traffic correlations and enhance feature representation of all input data. In specific, based on the results of cross-domain correlation analysis, we propose to learn different correlations with the occupancy component and environmental component, respectively.

Occupancy Component: we envision a multi-zone scenario for traffic volume prediction, with j_{th} zone's occupancy as $\mathbf{x}^j = (x_1^j, x_2^j, ..., x_t^j)^\top \in \mathbb{R}^T$ and $1 \le j \le N_o$. The cross-domain attention on occupancy components is calculated by referring the previous hidden state $\mathbf{h}_{t-1}$ and previous cell state $\mathbf{s}_{t-1}$ of encoder LSTM by

$$o_t^j = \mathbf{v}_o^\top \tanh(\mathbf{W}_o[\mathbf{h}_{t-1}; \mathbf{s}_{t-1}] + \mathbf{U}_o\mathbf{x}^j + \mathbf{b}_o), \qquad (9)$$

$$\beta_t^j = \frac{\exp(o_t^j)}{\sum_{i=1}^{N_o} \exp(o_t^i)}, \qquad (10)$$

where $\mathbf{v}_o, \mathbf{b}_o \in \mathbb{R}^T$, $\mathbf{W}_o \in \mathbb{R}^{T \times 2p}$ and $\mathbf{U}_o \in \mathbb{R}^{T \times T}$ are learning parameters. By using another softmax function, we ensure that all attention weights with occupancy component are normalized, and the total weight for all occupancy input

data is 1. With cross-domain attention weights for all public zones, we acquire the output vector from the occupancy component at time interval t as:

$$\hat{\mathbf{x}}_t^{occ} = (\beta_t^1 x_t^1, ... \beta_t^j x_t^j, ... \beta_t^{N_o} x_t^{N_o})^\top. \tag{11}$$

Environmental Component: given k_{th} type of environmental data input as $\mathbf{x}^k$, where $\mathbf{x}^k = (x_1^k, x_2^k, ..., x_t^k)^\top \in \mathbb{R}^T$ and $1 \le k \le N_e$. We employ the cross-domain attention for the environmental component to adaptively capture the dynamic correlation between traffic volume and k_{th} type of environmental data by

$$e_t^k = \mathbf{v}_e^\top \tanh(\mathbf{W}_e[\mathbf{h}_{t-1}; \mathbf{s}_{t-1}] + \mathbf{U}_e \mathbf{x}^k + \mathbf{b}_e), \tag{12}$$

$$\alpha_t^k = \frac{\exp(e_t^k)}{\sum_{i=1}^{N_e} \exp(e_t^i)}, \tag{13}$$

where $[\cdot ; \cdot]$ is a concatenation operation. Here, $\mathbf{v}_e, \mathbf{b}_e \in \mathbb{R}^T$, $\mathbf{W}_e \in \mathbb{R}^{T \times 2p}$ and $\mathbf{U}_e \in \mathbb{R}^{T \times T}$ are learnable parameters. By applying a softmax function to e_t^k, we obtain the normalized attention weight α_t^k for k_{th} environmental data at the time interval t. With cross-domain attention weights for all types of environmental data input, the output vector of the environmental component at time interval t can be adaptively acquired by

$$\hat{\mathbf{x}}_t^{env} = (\alpha_t^1 x_t^1, ... \alpha_t^k x_t^k, ... \alpha_t^{N_e} x_t^{N_e})^\top. \tag{14}$$

Finally, for encoder LSTM, we adaptively concatenate above output vectors from different components and extract the final output vector of cross-domain attention mechanism as:

$$\hat{\mathbf{x}}_t = [\hat{\mathbf{x}}_t^{env}; \hat{\mathbf{x}}_t^{occ}], \tag{15}$$

where $\hat{x}_t \in \mathbb{R}^N$. We feed the final output vector $\hat{\mathbf{x}}_t$ into the encoder LSTM as its new input at time interval t. Consequently, the hidden state of encoder LSTM in Eq. 3 is updated by

$$\mathbf{h}_t = f_e(\mathbf{h}_{t-1}, \hat{\mathbf{x}}_t), \tag{16}$$

where f_e is the encoder LSTM network described in Eq. 4 to Eq. 8.

5.2.2 Decoder with temporal attention

The basic idea of attention mechanism is to distinguish task-related importance of input data in historical time intervals. Intuitively, the decoder with a temporal attention mechanism can be trained to capture temporal dependencies between traffic volume and building data by calculating attention weights for all encoder hidden states. With the temporal attention weights assigned to all encoder hidden states, decoder LSTM can focus on the most relevant input data when decoding [39]. To this end, we employ a temporal attention mechanism that enables the decoder to select relevant encoder hidden states across all time intervals adaptively. In specific, when computing the attention vector for encoder hidden state at time interval t, the temporal attention mechanism refers to the previous hidden state $\mathbf{h}'_{t-1}$ and previous cell state $\mathbf{s}'_{t-1}$ of decoder LSTM by

$$\mu_t^i = \mathbf{v}_d^\top \tanh(\mathbf{W}_d[\mathbf{h}'_{t-1}; \mathbf{s}'_{t-1}] + \mathbf{U}_d \mathbf{h}_i + \mathbf{b}_d), 1 \le i \le T, \tag{17}$$

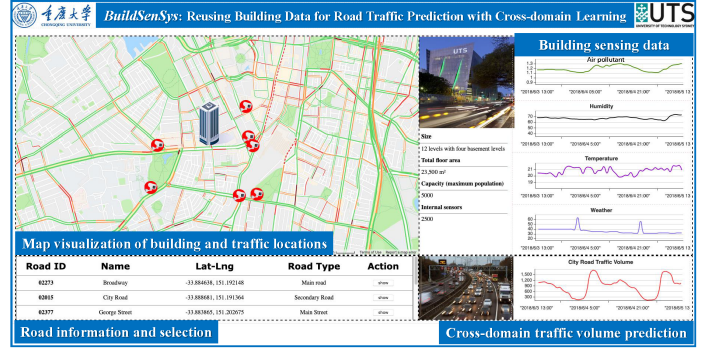$$\gamma_t^i = \frac{\exp(\mu_t^i)}{\sum_{l=1}^{T} \exp(\mu_t^l)}, \tag{18}$$

Fig. 8: The prototype system of *BuildSenSys* for data visualization and traffic prediction.

where $[\cdot]$ is a concatenation operation; $\mathbf{v}_d, \mathbf{b}_d \in \mathbb{R}^p$; $\mathbf{W}_d \in \mathbb{R}^{p \times 2q}$; $\mathbf{U}_d \in \mathbb{R}^{p \times p}$ are learning parameters. Through a softmax layer, the temporal attention weight γ_t^i is calculated for i_{th} encoder hidden state at time interval t. As the component of the input sequence is temporally mapped to each encoder, we calculate the context vector $\mathbf{c}_t$ as a weighted sum of all encoder hidden states by

$$\mathbf{c}_t = \sum_{i=1}^{T} \gamma_t^i \mathbf{h}_i. \tag{19}$$

To capture the dynamic temporal correlation in traffic volume data, we further combine the context vector with $\mathbf{y} = (y_1, ..., y_t, ..., y_{T-1})$ as follows:

$$\tilde{y}_{t-1} = \tilde{w}^\top [y_{t-1}; \mathbf{c}_{t-1}] + \tilde{b}, \tag{20}$$

$$\mathbf{h}'_t = f_d(\mathbf{h}'_{t-1}, \tilde{y}_{t-1}), \tag{21}$$

where f_d is an LSTM-based recurrent neural network as decoder; $\tilde{w} \in \mathbb{R}^{p+1}$ and $\tilde{b} \in \mathbb{R}$ are parameters to map the concatenation result to the size of the decoder input. As the structure of LSTM unit in the decoder is exactly same as the encoder (referred to Eq. 4 to 8), we omit the update process of f_d. Finally, attention mechanisms based recurrent neural network concatenates the context vector $\mathbf{c}_T$ with decoder hidden state $\mathbf{h}'_T$, predicting the traffic volume at time interval T as

$$\hat{y}_T = \mathbf{v}_y^\top (\mathbf{W}_y[\mathbf{c}_T; \mathbf{h}'_T] + \mathbf{b}_y) + b, \tag{22}$$

where $[\mathbf{c}_T; \mathbf{h}'_T] \in \mathbb{R}^{p+q}$ is a concatenation operation; parameters $\mathbf{W}_y \in \mathbb{R}^{q \times (p+q)}$ and $\mathbf{b}_y \in \mathbb{R}^q$ together map the concatenation to the size of decoder hidden states. The final output is generated by a mapping function with weights $\mathbf{v}_y \in \mathbb{R}^q$ and $\mathbf{b}_y \in \mathbb{R}$.

6 PERFORMANCE EVALUATION

To evaluate the performance of *BuildSenSys*, we develop a prototype system, as shown in Fig. 8. We first train the cross-domain learning-based RNN model with a training set of building sensing data and traffic data. With the pre-trained model, *BuildSenSys* can output predicted traffic volume on nearby roads only with building sensing data as the input. In the following, we first introduce the experimental settings, including dataset, baseline methods, evaluation

metrics, and model parameters. Then, we conduct extensive experimental studies on *BuildSenSys* and evaluate its performance in terms of baseline comparison, parameter study, ablation study, attention weight, and other extensive comparison.

6.1 Experimental Methodology and Settings

6.1.1 Dataset Description

Based on the results of the observation study, we select traffic data from nearby roads and relevant building sensing data as the training input data for the *BuildSenSys* model. For traffic data, we collect traffic volume count data from the official website of the Department of Roads and Maritime Services, New South Wales State [42]. The traffic volume count data is generated by permanent and temporary roadside collection devices, monitoring the number of passing vehicles on each road with the one-hour interval. We collect 12-month traffic volume data (from 2018/1/1 to 2018/12/31) on four nearby roads of the building. For building sensing data, we collect data from three categories, *i.e.*, building occupancy, building environmental data, outdoor environmental data. First, building occupancy data is generated by camera sensors distributed at Point-of-Interest Zones inside the building. We process and aggregate these data with the PLCount algorithm [43] for overall building occupancy. Second, the building environmental data include CO_2 concentration, building humidity, O_2 concentration, building temperature, and building air pollution. Third, the outdoor environmental data is collected from rooftop weather station of our building, including outdoor temperature, rainfall, and wind speed. Besides, as vehicle emissions have a direct influence on outdoor air quality, we further adopt the hourly AQI data from the Bureau of Meteorology's official website and integrate the AQI data into outdoor environmental data for traffic prediction. At last, we synchronize all building sensing data and traffic volume data with the one-hour interval for training and testing purposes.

6.1.2 Baseline Methods

To comprehensively evaluate the performance of *BuildSenSys*, we compare it with seven baseline methods as follows.

- **HA** [45]: The historical average (HA) model, which predicts the traffic volume by averaging the historical value of all corresponding time intervals.
- **ARIMA** [46]: The autoregressive integrated moving average (ARIMA) model, which is a classic model to predict future time series.
- **VAR** [47]: The Vector Auto-regressive (VAR) model is an extension of the univariate autoregressive model, which has been widely used for multivariate time series forecasting.
- **LWR** [8]: The Locally Weighted Linear Regression (LWR) model is a non-parametric model, which performs regressions around points of interest.
- **LSTM** [37]: The Long Short-Term Memory (LSTM) network is a variation of recurrent neural networks designed for avoiding the vanishing gradient problem.

- **Seq2Seq** [48]: The Sequence to Sequence model based on an encoder-decoder architecture and recurrent neural networks, consisting of three parts, *i.e.*, encoder, context vector, and decoder.
- **Seq2Seqw/attn** [48]: The Sequence to Sequence model with a temporal attention mechanism.

Note that all baseline methods take the time of a day as an essential feature of input data, and they use this feature for traffic prediction from different perspectives. For example, the HA leverages time of the day with statistical regression, while ARIMA and VAR process this feature with autoregression. Moreover, the neural networks of the LSTM and Seq2Seq models learn the time of the day as a feature with their hidden states. In the proposed *BuildSenSys* model, we further address the time of the day by employing a temporal attention mechanism to learn temporal features of traffic data.

6.1.3 Evaluation Metrics and Parameter Settings

To evaluate the prediction accuracy, we employ three widely used evaluation metrics, *i.e.*, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) [49]. Both MAE and RMSE are scale-dependent metrics, and MAPE is a scale-independent metric. Specifically, MAE measures the average magnitude of errors in prediction results as Eq. 23. RMSE measures square root of the average squared differences between prediction results and ground truth as Eq. 24. MAPE measures the size of errors in percentage terms to quantify the prediction accuracy as Eq. 25.

$$\mathbf{MAE} = \frac{1}{\tau} \sum_{[T:T+\tau] \in \mathbb{R}^T} \left| \mathbf{Y}_{[T:T+\tau]} - \hat{\mathbf{Y}}_{[T:T+\tau]} \right|, \quad (23)$$

$$\mathbf{RMSE} = \sqrt{\frac{1}{\tau} \sum_{[T:T+\tau] \in \mathbb{R}^T} \left(\mathbf{Y}_{[T:T+\tau]} - \hat{\mathbf{Y}}_{[T:T+\tau]} \right)^2}, \quad (24)$$

$$\mathbf{MAPE} = \frac{1}{\tau} \sum_{[T:T+\tau] \in \mathbb{R}^T} \left| \frac{\mathbf{Y}_{[T:T+\tau]} - \hat{\mathbf{Y}}_{[T:T+\tau]}}{\mathbf{Y}_{[T:T+\tau]}} \right|, \quad (25)$$

where $\hat{\mathbf{Y}}_{[T:T+\tau]}$ and $\mathbf{Y}_{[T:T+\tau]}$ are the prediction results of traffic volume from time interval T to $T+\tau$, respectively.

For model training, the tunable hyperparameters for *BuildSenSys* include the time window L (*i.e.*, the length of input data in hours), the length of predicting window τ and the size of hidden states in encoder/decoder (denoted by h_a and h_b, respectively). Specifically, for LSTM, Seq2Seq, Seq2Seqw/attn, and *BuildSenSys* model, h_a and h_b are tuned with the same values in 32, 64, 126, 256, 512, and L varies from 4, 6, 12, 18, 24 to 48 (hours), respectively. The predicting window τ denotes the number of days for future traffic prediction. In the training session, the batch size, learning rate, and dropout rate are set to 256, 0.001, and 0.2, respectively. Also, we use the Adam as the optimizer of *BuildSenSys* model. For all building sensing data and traffic volume data, we split the dataset into the training set (70%), validation set (10%) and test set (20%) in chronological order.
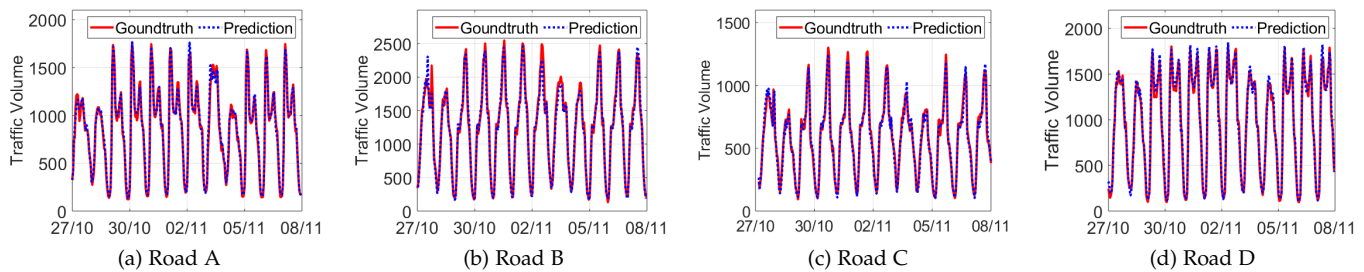
Fig. 9: Comparison between predicted traffic volume of *BuildSenSys* and the ground truth on four different roads.

TABLE 2: Performance comparison with baseline methods on different roads.

Models	Road A			Road B			Road C			Road D		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
LWR	244.51	155.69	22.78%	257.86	176.94	25.91%	206.52	176.35	22.61%	239.48	190.79	22.13%
ARIMA	152.74	141.22	14.51%	192.73	142.66	15.58%	149.11	138.09	14.70	173.32	124.31	14.22%
VAR	117.35	110.83	12.99%	120.37	116.9	13.06%	98.65	94.15	11.05	124.61	110.68	12.46%
HA	108.01	89.72	11.37%	125.88	93.57	12.78%	90.71	65.20	10.26%	117.03	89.45	11.12%
LSTM	74.37	58.30	9.83%	91.71	72.47	11.04%	70.71	58.9	9.31%	90.18	67.37	10.39%
Seq2Seq	72.86	51.06	7.40%	89.25	64.78	8.29%	66.19	53.32	7.12%	87.22	63.20	8.14%
Seq2Seqw/attn (128)	59.58	45.18	6.89%	73.5	59.12	7.75%	54.07	46.98	6.73%	81.15	56.95	7.49%
Seq2Seqw/attn (512)	50.18	39.74	5.71%	68.71	50.71	6.97%	45.51	37.33	5.49%	65.3	46.72	6.4%
BuildSenSys (128)	46.89	33.35	4.85%	60.18	41.89	4.36%	36.42	27.14	5.25%	50.45	38.56	5.2%
***BuildSenSys* (512)**	**33.08**	**22.78**	**2.05%**	**58.33**	**39.71**	**3.64%**	**30.49**	**20.29**	**2.51%**	**35.10**	**25.75**	**2.22%**

6.2 Experimental Evaluations

6.2.1 Evaluations on overall prediction results

First, we evaluate the overall prediction accuracy of *Build-SenSys* by comparing prediction results with the ground truth on four different types of roads, *i.e.*, roads A (minor highway), B (major highway), C (main street), and D (primary street). Fig. 9 illustrates the predicted traffic volume and the ground truth over 12 days. As shown in Fig. 9a to Fig. 9d, the prediction results are highly close to the ground truth. Thus, *BuildSenSys* can successfully capture the cross-domain correlations and temporal dependencies in reusing building data for traffic volume prediction. Moreover, the above results indicate that the distance between a road segment and the building can significantly affect the prediction accuracy. For example, as illustrated in Fig. 9b, prediction errors on road B are much larger than those of roads A, C and D. As shown in Table 1, road B is the furthest (3.0 km) to the building among all four roads.

6.2.2 Comparisons with Baselines

We further quantify the performance of *BuildSenSys* by comparing it with seven baseline methods by using three evaluation metrics. The evaluation experiments are conducted on four road segments. The basic parameter settings of baseline methods are adopted from [50]. For *BuildSenSys*, we set the length of input time window $L = 24$ and the size of hidden states in both encoder LSTM and decoder LSTM as 128 and 512, respectively. Table 2 presents the performance comparison between *BuildSenSys* and all baseline methods.

Overall, *BuildSenSys* achieves the best prediction accuracy with the lowest RMSE at 30.49, lowest MAE at 20.29, and lowest MAPE at 2.05%, respectively. It outperforms the best prediction of Seq2Seqw/attn by up to 46.2% in RMSE on road D, 45.6% in MAE on road C and 65.3% in

MAPE on road D. In contrast, the LWR model, which also uses building data to predict traffic volume [8], shows the worst performance with RMSE at 239.48, MAE at 190.79 and MAPE at 22.13% on road D. The key reason here is that the LWR model [8] is based on the assumption of linear building-traffic correlation. Meanwhile, the actual building-traffic correlations are proved to be non-linear, time-varying, and far more complex than the linear model. For other baseline methods that are based on historical traffic data, we present the detailed comparison results as follows.

First, ARIMA and VAR methods show the worst prediction performance, as they are not capable of making accurate predictions for the long term, especially on the 'turning points' (*e.g.*, rush hours). Interestingly, the Historical Average (HA) method, as a most naivest scheme, performs better than ARIMA and VAR for the following reasons. Traffic volumes have daily patterns and weekly patterns that are relatively stable in our dataset. Nevertheless, the HA method can produce unsatisfied prediction results with large errors in holidays, extreme weather, and social events, because it can not predict traffic volume that does not follow regular patterns.

Second, the Recurrent Neural Network-based methods, *i.e.*, LSTM, Seq2Seq, and Seq2Seqw/attn, show superior performance with MAEs lower than 70 and RMSEs lower than 90. In specific, the performance of LSTM and Seq2Seq are competitive, while Seq2Seq outperforms the LSTM by 2.35% and 2.43% in MAPE metric on Road D and Road A, respectively. As for Seq2Seq model based on the encoder-decoder architecture, it encodes the input traffic volume data into a feature vector, from which the decoder generates the fixed-length prediction iteratively. Nevertheless, the performance of encoder-decoder networks will deteriorate rapidly when the input sequence data becomes longer. Both
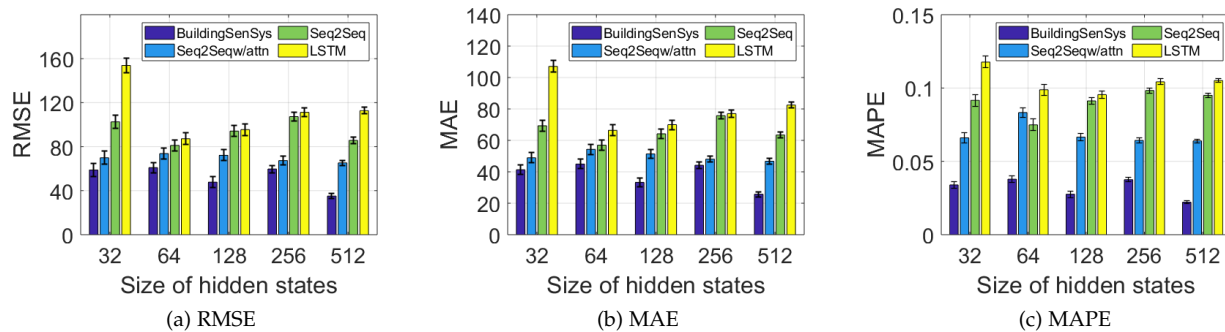
Fig. 10: Impact of the hidden states on the performance of *BuildSenSys* and three RNN-based baseline methods.
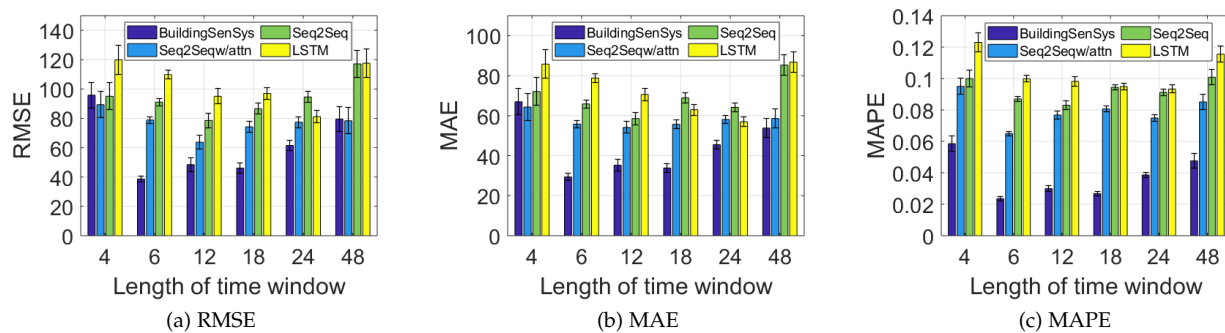

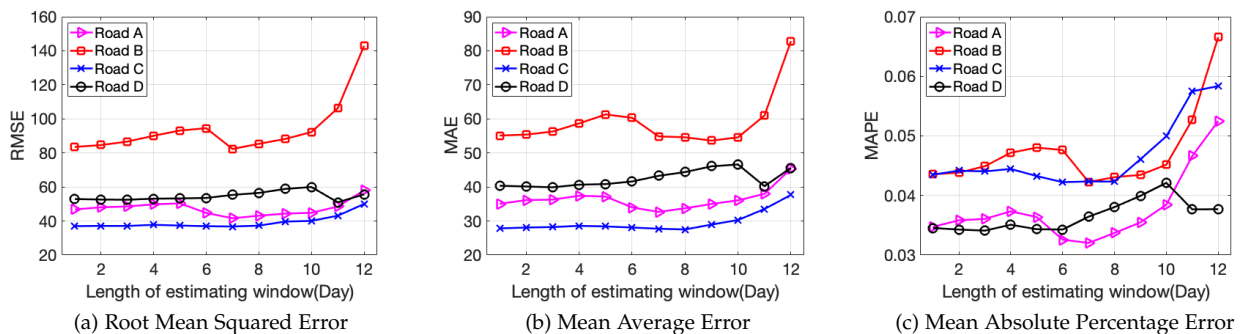Fig. 11: Impact of the input time window on the performance of *BuildSenSys* and three RNN-based baseline methods.

Fig. 12: Prediction accuracy of *BuildSenSys* on four roads by varying different lengths of predicting window.

BuildSenSys and Seq2Seqw/attn employ temporal attention mechanisms. Hence, the decoders of the above models can select the most relevant encoder hidden states adaptively to improve prediction accuracy. The effectiveness of attention mechanism is validated by the more accurate prediction results by Seq2Seqw/attn and *BuildSenSys*. With 512 hidden states, both Seq2Seqw/attn and *BuildSenSys* greatly improve the prediction accuracy in comparison with models that have 128 hidden states. More importantly, the prediction accuracy of *BuildSenSys* is further enhanced by the cross-domain attention that learns building-traffic correlations. In comparison with Seq2Seqw/attn, *BuildSenSys* shows 32.3% and 37.8% improvements in MAE and RMSE, respectively. *In summary, compared with RNN-based baseline methods, BuildSenSys jointly leverages cross-domain attention and temporal*

attention to adaptively learn cross-domain, time-varying, and non-linear building-traffic correlations. As a result, BuildSenSys outperforms all baseline methods with up to 65.3% accuracy improvement (e.g., 2.2% MAPE) in predicting nearby traffic volume.

6.2.3 Evaluation of Parameters

We further evaluate the impact of parameters in RNN networks on *BuildSenSys*'s performance, since *BuildSenSys* is based on RNN. The parameters include the size of the hidden states in the encoder (h_a) and decoder (h_b), the lengths of the input time window (L) and the prediction window (τ). Moreover, three RNN-based baseline methods, *i.e.*, LSTM, Seq2Seq, and Seq2Seqw/attn, are used as the comprised benchmarks.
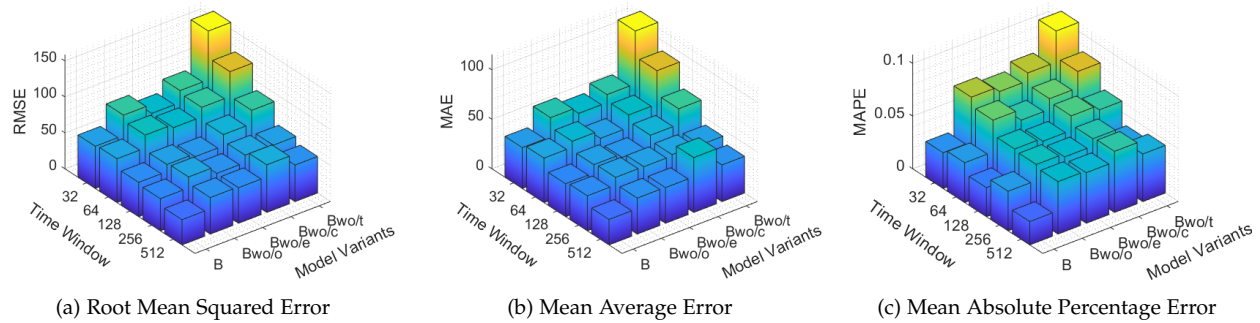
Fig. 13: Performance comparison among different variants of *BuildSenSys* with varying size of hidden states.
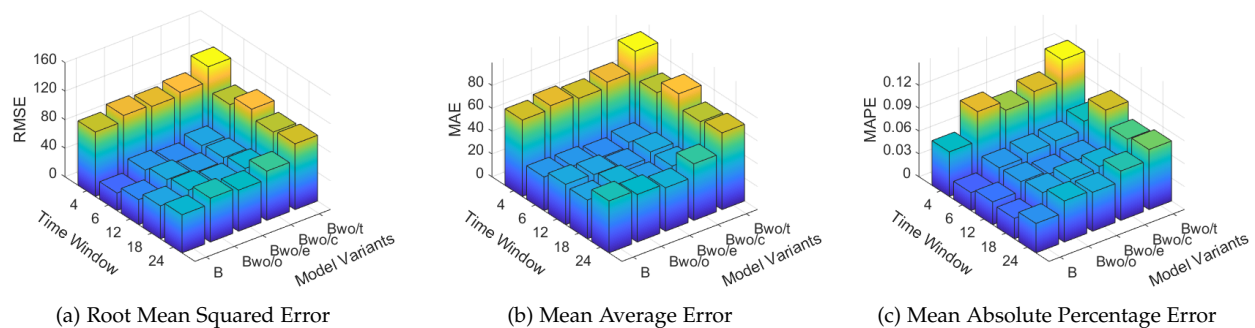

Fig. 14: Performance comparison among different variants of *BuildSenSys* with varying length of the time window (in hours).

1) **Size of hidden states**: Following [51], we set the size of hidden states in both encoder and decoder as the same, *i.e.*, $h_a = h_b$. We change the value of h_a (h_b) from 32 to 64, 128, 256, and 512 with a grid search. As shown in Figs. 10a, 10b and 10c, *BuildSenSys* outperforms LSTM, Seq2Seq and Seq2Seqw/attn in all settings of hidden states in terms of RMSE, MAE and MAPE. Moreover, the experimental results show that the prediction accuracy of *BuildSenSys* roughly improves with the larger size of hidden states. Overall, *BuildSenSys* achieves the best prediction results (*i.e.*, RMSE ≤ 33, MAE ≤ 25 and MAPE ≤ 0.22) when $h_a = h_b = 512$.

2) **Length of input time window**: The time window L denotes how many hours of historical data are fed into the traffic prediction models (1 hour as the basic unit). In order to evaluate the impact of L, we change its value from 4 to 6, 12, 18, 24 and 48 hours. Thereby, the historical data of the last 4, 6, 12, 18, 24, and 48 hours will be used to predict traffic volume in the next hour. As illustrated in Figs. 11a, 11b, and 11c, the prediction accuracy of all RNN-based methods increases with the greater length of the time window. Meanwhile, the performance of RNN-based methods also decreases when the time window is too large, *e.g.*, $L = 48$. Accordingly, *BuildSenSys* achieves the best performance when $L = 6$, *i.e.*, RMSE= 31.7, MAE= 24, and MAPE= 0.019.

3) **Length of prediction window**: We evaluate the prediction accuracy of *BuildSenSys* in terms of the length of predicting window τ. We change τ from 1 day to 12 days. From Figs. 12a, 12b, and 12c, we observe that the RMSE,

MAE, and MAPE increase with the greater length of predicting windows, *i.e.*, the prediction accuracy decreases with the longer predicting windows. However, except for Road B, the decreasing speed of prediction accuracy is very slow. Therefore, *BuildSenSys* can achieve a stable prediction with high accuracy by reusing building data, *e.g.*, RMSE and MAE increase from 36 to 52 and 28 to 30, respectively. Meanwhile, as Road B is the furthest road from the building, its building-traffic correlation is not as strong as other roads. Therefore, the prediction errors on Road B increase significantly with the increasing value of τ.

6.2.4 Evaluation of Cross-domain Learning

BuildSenSys incorporates a cross-domain attention mechanism and a temporal attention mechanism to jointly extract spatio-temporal features from building-traffic correlations. To evaluate the impact of each component on the overall performance, we conduct an ablation study on *BuildSenSys* with its variants as follows.

- **B**, *i.e.*, *BuildSenSys*, leverages cross-domain attention and temporal attention to jointly learn building-traffic correlations and temporal correlations with both occupancy data and environmental data.
- **Bwo/o**, a variant of **B** without **o**ccupancy component.
- **Bwo/e**, a variant of **B** without **e**nvironmental component.
- **Bwo/c**, a variant of **B** without **c**ross-domain attention, which only employs temporal attention to learn

temporal dependencies of building sensing data and traffic volume data.

- **Bwo/t**, a variant of **B** without **t**emporal attention, which only employs cross-domain attention to learn cross-domain building-traffic correlations.

First, we evaluate the performance of different components of *BuildSenSys* by changing the size of hidden states. As shown in Fig. 13, *BuildSenSys* demonstrates the best performance, while Bwo/t shows the worst performance as it lacks temporal correlations for traffic prediction. Meanwhile, the prediction accuracy of both Bwo/o and Bwo/e is better than Bwo/c. It indicates that cross-domain learning with building data can effectively improve overall performance. Moreover, as illustrated in Figs. 13a, 13b. and 13c, the prediction accuracy of Bwo/e is higher than that of Bwo/o. This result shows that the building occupancy data has more contribution to cross-domain traffic prediction than building environmental data.

Second, we evaluate the performance of different components of *BuildSenSys* by changing the length of the input time window. As shown in Fig. 14, the performance of Bwo/t deteriorates significantly as it has no temporal attention to capture the temporal correlations in building data and traffic data. As a result, the prediction accuracy of Bwo/t is the worst, *e.g.*, the RMSE over 115, MAE over 80, and MAPE over 10%, respectively. Meanwhile, *BuildSenSys* shows the best prediction accuracy when the input time window is 6. However, when the length of the input time window increases to 12, 18, and 24, the prediction accuracy of all variants decreases continuously. This can be explained as the longer inputs could pose greater difficulty in capturing building-traffic correlations and temporal correlations.

To sum up, the above results indicate that each component in cross-domain learning has its contribution to enhancing the prediction accuracy of *BuildSenSys*. Specifically, the occupancy component and the environmental component are complementary to each other in reusing building data to predict nearby traffic volume. Moreover, it is highly significant to jointly leverage cross-domain attention and temporal attention in cross-domain learning of building data. Each attention mechanism has its unique contribution to the improvement of prediction accuracy, *e.g.*, up to 45.5% improvement by temporal attention and 30.9% improvement by cross-domain attention.

6.2.5 Evaluation of Attention Weight

1) **Cross-domain attention**: We analyze the attention weights of the proposed traffic prediction model in two folds, *i.e.*, cross-domain attention weight and temporal attention weight. First, as the correlations between traffic volume data and different building sensing data are non-linear, *BuildSenSys* employs a cross-domain attention mechanism to achieve the non-linear mapping from building sensing data to traffic data in predicting traffic volume. To demonstrate the different attention weights from different types of building sensing data, we conduct extensive experiments to analyze the attention weight of each type of building sensing data. As shown in Fig. 15a, the 10 groups of attention weights correspond to 10 types of sensing data, including building occupancy (B_occ), CO_2 concentration (B_carb), building
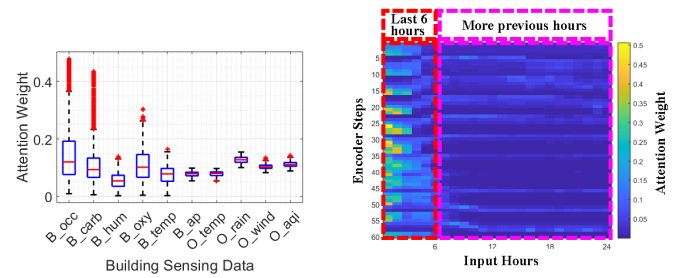
(a) Cross-domain Attention Weight   (b) Temporal Attention Weight

Fig. 15: Visualization of Cross-domain Attention and Temporal Attention.

humidity (B_hum), O_2 concentration (B_oxy), building temperature (B_temp), building air pollution (B_ap), as well as outdoor temperature (O_temp), rainfall (O_rain) and air quality index (O_aqi). The box plot in Fig. 15a compares the distributions of attention weights by each type of building sensing data.

As demonstrated in Fig. 15a, each type of building sensing data has its own contribution to traffic prediction while building occupancy data has the highest attention weights across all predicting steps. Specifically, the whiskers (*i.e.*, lines extending above and below each box) of B_occ have the most extensive range (from 0.01 to 0.36). B_occ covers the furthest adjacent values of attention weight, showing the significant importance of building occupancy data in cross-domain traffic prediction. Moreover, the B_occ and B_carb have the largest number of outliers with the highest values (*i.e.*, up to 0.48 and 0.39, respectively). It indicates that they have stronger correlations with traffic data and higher attention weights in traffic prediction. Finally, the median values (*i.e.*, the central marks) of attention weights boxes are roughly close, illustrating that all types of building data make non-negligible contributions in traffic predictions.

In summary, the above experimental results indicate that, each building data has its own contribution to enhancing the prediction accuracy of *BuildSenSys*. Meanwhile, the building occupancy data and the environmental data are complementary to each other in reusing building data for traffic sensing and prediction.

2) **Temporal attention**: We further investigate how the temporal attention mechanism captures the correlations between predicted traffic values and historical traffic data. More specifically, we conduct the exploring experiment to visualize how *BuildSenSys* assigns different attention weights to the historical traffic data in the time window 'L'. As explained in the experimental setups, the input time window 'L' represents the length of input data measured in hours, and we set 'L' as 24 (hours) in the following experiment.

As a result, the heat map in Fig. 15b visualize the temporal attention weights over 24 hours and 60 predicting steps. We divide the input time window into two categories, *i.e.*, the latest 6 hours and more previous (last 6-24) hours. The above two categories have great disparities in assigned attention weights. For the input data of the latest 6 hours, *BuildSenSys* assigns more attention weight to them, as each
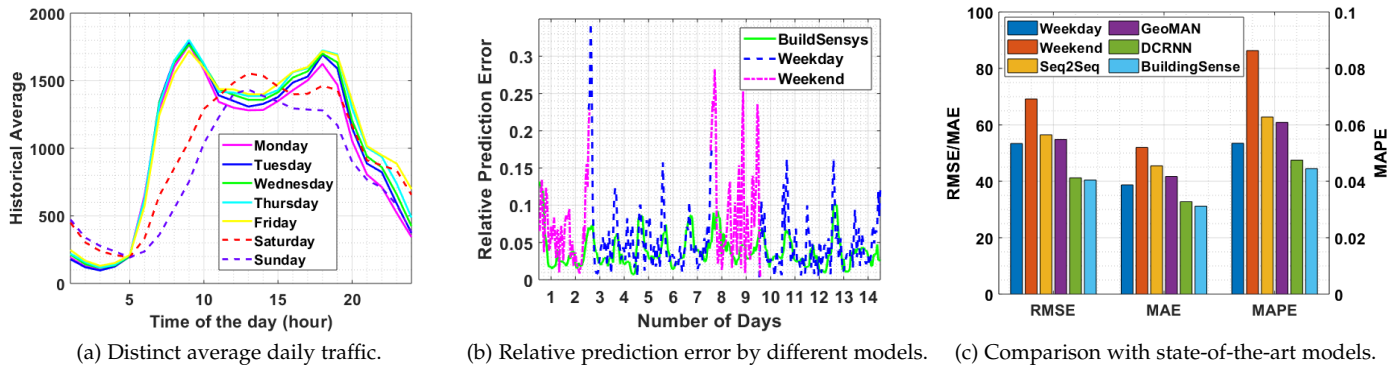
(a) Distinct average daily traffic.   (b) Relative prediction error by different models.   (c) Comparison with state-of-the-art models.

Fig. 16: Performance comparison with Weekday model and Weekend model.

hour's data has up to 0.5 attention weight. In contrast, most input data in last 6-24 hours have particularly small values of attention weights (under 0.05). The above result shows that the temporal attention mechanism in *BuildSenSys* would focus more on the latest 6 hours of the input data. Meanwhile, as all of the temporal attention weights sum to 1, with the longer input time windows (*e.g.*, 12, 18, 24 and 48 in our ablation studies), the latest 6 hours will share less attention weights. In such cases, *BuildSenSys* will not be able to specifically focus on the most recent data, but to process both of short-term and long-term temporal correlations for traffic volume prediction. Consequently, the prediction accuracy of *BuildSenSys* will be influenced, as demonstrated and verified by the evaluation results in previous experimental studies.

6.2.6   Extensive evaluations of comparison

Baseline methods with the weekday's data and the weekend's data: To further explore the impact of data patterns on the traffic prediction results, we conduct an extensive experiment by applying two different models to predict traffic volumes on weekdays and weekends, respectively. First, we visualize a distinct daily average of traffic data on a target road in Fig.16a, where different colors plot the average values of one-year traffic data on each day of the week. We can observe that the traffic volumes on weekdays and weekends follow two different patterns, respectively. For weekdays, the peak hours range from 8 a.m. to 10 a.m. and 4 p.m. to 7 p.m. Meanwhile, for weekends, there is a longer range for peak hours, i.e., 12 a.m. to 6 p.m. We aim to study whether using two different models of *BuildSenSys* would impact the results of overall results of traffic volume prediction.

The basic settings for this experiment are the same as experiments in 7. The only difference is that we divide one-year traffic data into two groups of data (weekdays and weekends), and correspondingly we train two models (weekday model and weekend model) of BuildSenSys for each group. At the same time, we train a BuildSenSys model based on the whole dataset, and the total epochs for all models are set as 2,500. The prediction results of one-model prediction and two-model prediction are shown in Fig. 16b and Fig. 16c. Weekday represents a special model of *BuildSenSys* that is trained only with weekday data to predict the data on weekdays. Similarly, Weekend stands for another special model of *BuildSenSys* that is trained only with weekend data predicts the data on weekends. As illustrated in Fig. 16b and Fig. 16c, we have compared the prediction accuracy of the Weekday model and the Weekend model, Seq2Seq model, and *BuildSenSys* model. The results show that in comparison with the Seq2Seq model, both the Weekday model and the Weekend model can achieve some improvements in prediction accuracy. However, the combined prediction results are still not satisfactory with even lower accuracy than the Weekday model. Meanwhile, by using a temporal attention mechanism, the proposed *BuildSenSys* can dynamically learn the impact of historical data on the predicting target, thereby it outperforms both the Weekday model and the Weekend model. The experimental results also validate that, breaking the continuity of the dataset in the temporal dimension will undermine the performance of temporal attention mechanisms.

Comparison with state-of-the-art methods based on different data sources: To address how different data measurements would impact the traffic prediction results, we further employ three baseline methods for comparison. Among these baseline methods, the Sequence to Sequence model (Seq2Seq) [48] uses traffic volume data from traffic sensing system on one road for prediction; the Multi-level Attention Networks for Geospatial Sensors (GeoMAN) [51] exploits the data of one road along with the weather data; and Diffusion Convolutional Recurrent Neural Network (D-CRNN) [52] leverages the sensing data from multiple roads with the road graph. In this work, BuildSenSys leverages building sensing data to predict the traffic volume on nearby roads. The experimental results in Fig. 16c show that Build-SenSys achieves the best performance, even compared with DCRNN, which uses traffic data from 10 road segments to predict traffic volume on the target road. The GeoMAN out-performs Seq2Seq, as it considers both spatial and temporal attention with weather data and historical road data for traffic prediction. Meanwhile, DCRNN further improves the prediction accuracy by using both diffusion convolution and sequence to sequence learning framework on traffic volume data of multiple roads with sensor topology information. The above results also assess the practicability and useful-ness of cross-domain traffic prediction by re-using building sensing data.

7 DISCUSSION

In this work, we explore the possibility of reusing building sensing data to predict the traffic volume of nearby road segments. Although we have made some advances toward this new direction, there are still several issues that need to be further investigated as follows.

Applicable conditions of *BuildSenSys*: It is of great importance to discuss: 1) which kind of buildings are suitable for accurate traffic sensing and prediction, and 2) what kind of real-world applications could benefit from cross-domain traffic prediction by reusing building sensing data. Based on our initial studies, we roughly summarize three critical factors of a building for cross-domain traffic prediction, including location, capacity and building types. First, human movement patterns (including traffic) among urban sites are affected by buildings, which can 'temporarily hold' human mobility [6]. Therefore, buildings located close to road networks are preferable for traffic sensing, as they would have more chances to affect human mobility on the road [6]. Second, the capacity of a building, *i.e.*, the the volume of occupants it can hold, is essential for accurate cross-domain traffic sensing. For instance, Zheng*et al.* [8] showed a commercial building with a capacity of 10000 occupants can effectually impact the traffic on nearby roads with substantial evidence. Third, the types of buildings can affect the applicable conditions of *BuildSenSys*. Different types of buildings have different patterns of occupancy dynamics, which will directly affect building-traffic correlations. To this end, we recommend commercial building (*e.g.*, office buildings and retail buildings) are the first choice when performing cross-domain traffic sensing with *BuildSenSys*. In practice, *BuildSenSys* can be used in a variety of real-world applications. For example, it can be exploited to provide real-time traffic volume data to support controls of intelligent traffic light [53], [54]. Besides, *BuildSenSys* can also be applied in traffic management systems, which requires accurate traffic predictions, especially in rush hours [55].

Sensing coverage of *BuildSenSys*: In this paper, *BuildSenSys* is a proof-of-concept in reusing building data for traffic sensing. As shown in the experimental evaluations of Section 6, the prediction accuracy of *BuildSenSys* is in accordance with the building-traffic correlation of a road. Thus, we further conduct experiments to investigate the coverage of cross-domain traffic prediction by reusing building sensing data. In specific, as illustrated in Fig. 17, we calculate the correlations between building occupancy and traffic data on nearby roads using the Pearson Correlation Coefficient, and further use the building-traffic correlations as the index of sensing coverage of *BuildSenSys*. Fig. 17 demonstrates that BuildSenSys generally has a prediction coverage within 5 km to the building.

Extension to large-scale scenarios: According to the investigation of sensing coverage, the *BuildSenSys* system can predict traffic volume for road segments within 5 km of the building. To extend this coverage of traffic predictions to a larger scale, it requires to incorporate more sensing data that are correlated with traffic volume. Intuitively, traffic data of road segments in the same district would have spatial and temporal correlations [23], [56]. For example, Liu *et al.* [27], [57] probed traffic conditions on different road
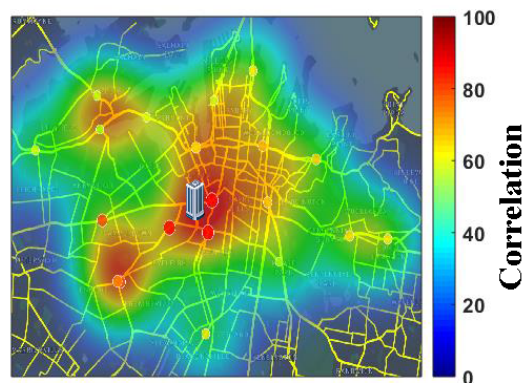

Fig. 17: Illustrations of prediction coverage by *BuildSenSys*.

segments with GPS data from bus riders. In our future work, we will further exploit the above multi-road and multi-region correlations to extend the coverage of *BuildSenSys* in traffic sensing and prediction. Besides, the pervasive street cameras are the cost-efficient data source for aggregating urban crowd flow, which may have direct and indirect correlations with traffic volume on nearby roads. Existing works in computer vision have proposed highly effective approaches to generate density maps [58] and achieve accurate crowd counting [59], which have the potential to contribute to cross-domain traffic sensing and prediction.

Privacy and security of building data: With the massive amount of IoT data generated by smart buildings, it is essential to reuse building data in a privacy-preserving manner. In our work, the building datasets, including occupancy data and environmental data, are anonymized by the data provider. As a consequence, they cannot be used to trace back to any personal information. Notably, the people counters (cameras) do not have facial recognition capability and there would be no personal privacy leakage from the building occupancy data. Moreover, if needed, we can adapt existing privacy protection methods [60], [61] to prevent privacy leakage when utilizing building sensing data.

8 CONCLUSION

In this paper, we have proposed *BuildSenSys*, a first-of-its-kind building sensing data-based traffic volume prediction system with cross-domain learning. Firstly, we have conducted extensive experimental analysis on building-traffic correlations based on multi-source real-world datasets. It discloses that building data has strong correlations with traffic data. Then, we have proposed a cross-domain learning-based RNN with cross-domain and temporal attention mechanisms to jointly extract building-traffic correlations for accurate traffic prediction. Moreover, we have implemented a prototype system of *BuildSenSys* and conducted extensive experiments. The experimental results demonstrated that *BuildSenSys* outperforms seven baseline methods with up to 65.3% accuracy improvement in predicting nearby traffic volume. We believe this work can open a new gate to reuse building sensing data for traffic sensing and prediction, hence indicating an interconnection between smart buildings and intelligent transportation.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2020.2976936, IEEE Transactions on Mobile Computing

16

REFERENCES

[1] A. P. Plageras, K. E. Psannis, C. Stergiou, H. Wang, and B. B. Gupta, "Efficient iot-based sensor big data collection–processing and analysis in smart buildings," *Future Generation Computer Systems*, vol. 82, pp. 349–357, 2018.

[2] O. Bates and A. Friday, "Beyond data in the smart city: repurposing existing campus iot," *IEEE Pervasive Computing*, vol. 16, no. 2, pp. 54–60, 2017.

[3] N. Nesa and I. Banerjee, "IoT-based sensor data fusion for occupancy sensing using dempster–shafer evidence theory for smart buildings," *IEEE Internet of Things Journal*, vol. 4, no. 5, pp. 1563–1570, 2017.

[4] T. C. Team, "A to Z of smart buildings," https://www.comfyapp.com/blog/d-data/, accessed August, 2019.

[5] D. Snoonian, "Smart buildings," *IEEE Spectrum*, vol. 40, pp. 18–23, 2003.

[6] Z. Zheng, F. Wang, D. Wang, and L. Zhang, "Buildings affect mobile patterns: developing a new urban mobility model," in *Proceedings of the 5th Conference on Systems for Built Environments*. ACM, 2018, pp. 83–92.

[7] H. Rashid, N. Batra, and P. Singh, "Rimor: towards identifying anomalous appliances in buildings," in *Proceedings of the 5th Conference on Systems for Built Environments*. ACM, 2018, pp. 33–42.

[8] Z. Zheng, D. Wang, J. Pei, Y. Yuan, C. Fan, and F. Xiao, "Urban traffic prediction through the second use of inexpensive big data from buildings," in *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. ACM, 2016, pp. 1363–1372.

[9] D. Jo, B. Yu, H. Jeon, and K. Sohn, "Image-to-image learning to predict traffic speeds by considering area-wide spatio-temporal dependencies," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 2, pp. 1188–1197, 2018.

[10] N. E. Klepeis, W. C. Nelson, W. R. Ott, J. P. Robinson, A. M. Tsang, P. Switzer, J. V. Behar, S. C. Hern, and W. H. Engelmann, "The national human activity pattern survey (nhaps): a resource for assessing exposure to environmental pollutants," *Journal of Exposure Science and Environmental Epidemiology*, vol. 11, no. 3, p. 231, 2001.

[11] X. Zhan, Y. Zheng, X. Yi, and S. V. Ukkusuri, "Citywide traffic volume estimation using trajectory data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 2, pp. 272–285, 2016.

[12] D. Deng, C. Shahabi, U. Demiryurek, L. Zhu, R. Yu, and Y. Liu, "Latent Space Model for Road Networks to Predict Time-Varying Traffic," in *ACM KDD*. New York, New York, USA: ACM Press, 2016, pp. 1525–1534.

[13] A. M. Nagy and V. Simon, "Survey on traffic prediction in smart cities," *Pervasive and Mobile Computing*, 2018.

[14] J. Tang, G. Zhang, Y. Wang, H. Wang, and F. Liu, "A hybrid approach to integrate fuzzy c-means based imputation method with genetic algorithm for missing traffic volume data estimation," *Transportation Research Part C: Emerging Technologies*, vol. 51, pp. 29–40, 2015.

[15] T. Idé, T. Katsuki, T. Morimura, and R. Morris, "City-wide traffic flow estimation from a limited number of low-quality cameras," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 4, pp. 950–959, 2016.

[16] L. Lin, J. Li, F. Chen, J. Ye, and J. Huai, "Road traffic speed prediction: a probabilistic model fusing multi-source data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 7, pp. 1310–1323, 2018.

[17] J. Wan, J. Liu, Z. Shao, A. V. Vasilakos, M. Imran, and K. Zhou, "Mobile crowd sensing for traffic prediction in internet of vehicles," *Sensors*, vol. 16, no. 1, p. 88, 2016.

[18] Z. Qin, Z. Fang, Y. Liu, C. Tan, W. Chang, and D. Zhang, "Eximius: A measurement framework for explicit and implicit urban traffic sensing," in *Proceedings of the 16th ACM Conference on Embedded Networked Sensor Systems*. ACM, 2018, pp. 1–14.

[19] B. Balaji, A. Bhattacharya, G. Fierro, J. Gao, J. Gluck, D. Hong, A. Johansen, J. Koh, J. Ploennigs, Y. Agarwal *et al.*, "Brick: Towards a unified metadata schema for buildings," in *Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments*. ACM, 2016, pp. 41–50.

[20] Z. Pan, Y. Liang, W. Wang, Y. Yu, Y. Zheng, and J. Zhang, "Urban traffic prediction from spatio-temporal data using deep meta learning," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 2019, pp. 1720–1730.

[21] S. P. Mohanty, U. Choppali, and E. Kougianos, "Everything you wanted to know about smart cities: The internet of things is the backbone," *IEEE Consumer Electronics Magazine*, vol. 5, no. 3, pp. 60–70, 2016.

[22] D. Minoli, K. Sohraby, and B. Occhiogrosso, "Iot considerations, requirements, and architectures for smart buildings±energy optimization and next-generation building management systems," *IEEE Internet of Things Journal*, vol. 4, no. 1, pp. 269–283, 2017.

[23] Z. Liu, Z. Li, M. Li, W. Xing, and D. Lu, "Mining road network correlation for traffic estimation via compressive sensing," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 7, pp. 1880–1893, 2016.

[24] C. Hu, W. Bao, D. Wang, Y. Qian, M. Zheng, and S. Wang, "stube+: an iot communication sharing architecture for smart after-sales maintenance in buildings," *ACM Transactions on Sensor Networks (TOSN)*, vol. 14, no. 3-4, p. 29, 2018.

[25] K. Nellore and G. P. Hancke, "A survey on urban traffic management system using wireless sensor networks," *Sensors*, vol. 16, no. 2, p. 157, 2016.

[26] X. Kong, X. Song, F. Xia, H. Guo, J. Wang, and A. Tolba, "Lotad: long-term traffic anomaly detection based on crowdsourced bus trajectory data," *World Wide Web*, vol. 21, no. 3, pp. 825–847, 2018.

[27] Z. Liu, P. Zhou, Z. Li, and M. Li, "Think like a graph: Real-time traffic estimation at city-scale," *IEEE Transactions on Mobile Computing*, 2018.

[28] Y. Cui, B. Jin, F. Zhang, B. Han, and D. Zhang, "Mining spatial-temporal correlation of sensory data for estimating traffic volumes on highways," in *Proceedings of the 14th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*. ACM, 2017, pp. 343–352.

[29] A. Janecek, D. Valerio, K. A. Hummel, F. Ricciato, and H. Hlavacs, "The cellular network as a sensor: From mobile phone data to real-time road traffic monitoring," *IEEE transactions on intelligent transportation systems*, vol. 16, no. 5, pp. 2551–2572, 2015.

[30] C. Costa, G. Chatzimilioudis, D. Zeinalipour-Yazti, and M. F. Mokbel, "Towards real-time road traffic analytics using telco big data," in *Proceedings of the International Workshop on Real-Time Business Intelligence and Analytics*. ACM, 2017, p. 5.

[31] C. Meng, X. Yi, L. Su, J. Gao, and Y. Zheng, "City-wide traffic volume inference with loop detector data and taxi trajectories," in *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 2017, p. 1.

[32] A. Sarker, H. Shen, and J. A. Stankovic, "Morp: data-driven multi-objective route planning and optimization for electric vehicles," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 4, p. 162, 2018.

[33] D. Pavlyuk, "Short-term traffic forecasting using multivariate autoregressive models," *Procedia Engineering*, vol. 178, pp. 57–66, 2017.

[34] Y. Lv, Y. Duan, W. Kang, Z. Li, F.-Y. Wang *et al.*, "Traffic flow prediction with big data: A deep learning approach." *IEEE Trans. Intelligent Transportation Systems*, vol. 16, no. 2, pp. 865–873, 2015.

[35] Z. Liu, Z. Li, K. Wu, and M. Li, "Urban traffic prediction from mobility data using deep learning," *IEEE Network*, vol. 32, no. 4, pp. 40–46, 2018.

[36] J. Xu, R. Rahmatizadeh, L. Bölöni, and D. Turgut, "Real-time prediction of taxi demand using recurrent neural networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 8, pp. 2572–2581, 2018.

[37] Z. Zhao, W. Chen, X. Wu, P. C. Chen, and J. Liu, "Lstm network: a deep learning approach for short-term traffic forecast," *IET Intelligent Transport Systems*, vol. 11, no. 2, pp. 68–75, 2017.

[38] Y. Tian, K. Zhang, J. Li, X. Lin, and B. Yang, "Lstm-based traffic flow prediction with missing data," *Neurocomputing*, vol. 318, pp. 297–305, 2018.

[39] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, 2017, pp. 5998–6008.

[40] H. Yao, X. Tang, H. Wei, G. Zheng, and Z. Li, "Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction," in *AAAI Conference on Artificial Intelligence*, 2019.

[41] "EIF research data interface," https://eif-research.feit.uts.edu.au/, accessed August, 2019.

[42] Roads and N. Maritime Services, "Roads and maritime services collects traffic volume information from roadside traffic collection devices across the nsw road network,"

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2020.2976936, IEEE Transactions on Mobile Computing

17

https://www.rms.nsw.gov.au/about/corporate-publications/statistics/traffic-volumes/index.html, accessed August, 2019.

[43] F. C. Sangoboye and M. B. Kjærgaard, "Plcount: A probabilistic fusion algorithm for accurately estimating occupancy from 3d camera counts," in *Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments*. ACM, 2016, pp. 147–156.

[44] A. Koesdwiady, R. Soua, and F. Karray, "Improving traffic flow prediction with weather information in connected cars: a deep learning approach," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 12, pp. 9508–9517, 2016.

[45] J. Zhang, Y. Zheng, and D. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction." in *AAAI*, 2017, pp. 1655–1661.

[46] W. Min and L. Wynter, "Real-time road traffic prediction with spatio-temporal correlations," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 4, pp. 606–616, 2011.

[47] S. R. Chandra and H. Al-Deek, "Predictions of freeway traffic speeds and volumes using vector autoregressive models," *Journal of Intelligent Transportation Systems*, vol. 13, no. 2, pp. 53–72, 2009.

[48] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Advances in neural information processing systems*, 2014, pp. 3104–3112.

[49] Y. Qin, D. Song, H. Chen, W. Cheng, G. Jiang, and G. Cottrell, "A dual-stage attention-based recurrent neural network for time series prediction," *arXiv preprint arXiv:1704.02971*, 2017.

[50] G. Lai, W.-C. Chang, Y. Yang, and H. Liu, "Modeling long-and short-term temporal patterns with deep neural networks," in *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. ACM, 2018, pp. 95–104.

[51] Y. Liang, S. Ke, J. Zhang, X. Yi, and Y. Zheng, "Geoman: Multi-level attention networks for geo-sensory time series prediction." in *IJCAI*, 2018, pp. 3428–3434.

[52] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," *https-s://openreview.net/forum?id=SJiHXGWAZ*, 2018.

[53] H. Zhang, S. Feng, C. Liu, Y. Ding, Y. Zhu, Z. Zhou, W. Zhang, Y. Yu, H. Jin, and Z. Li, "Cityflow: A multi-agent reinforcement learning environment for large scale city traffic scenario," in *The World Wide Web Conference*. ACM, 2019, pp. 3620–3624.

[54] H. Wei, G. Zheng, H. Yao, and Z. Li, "Intellilight: A reinforcement learning approach for intelligent traffic light control," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 2018, pp. 2496–2505.

[55] L. Zhu, F. R. Yu, Y. Wang, B. Ning, and T. Tang, "Big data analytics in intelligent transportation systems: a survey," *IEEE Transactions on Intelligent Transportation Systems*, no. 99, pp. 1–16, 2018.

[56] H. Zhu, Y. Zhu, M. Li, and L. M. Ni, "Seer: metropolitan-scale traffic perception based on lossy sensory data," in *IEEE INFOCOM 2009*. IEEE, 2009, pp. 217–225.

[57] Z. Liu, S. Jiang, P. Zhou, and M. Li, "A participatory urban traffic monitoring system: the power of bus riders," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 10, pp. 2851–2864, 2017.

[58] C. Xu, K. Qiu, J. Fu, S. Bai, Y. Xu, and X. Bai, "Learn to scale: Generating multipolar normalized density maps for crowd counting," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 8382–8390.

[59] X. Cao, Z. Wang, Y. Zhao, and F. Su, "Scale aggregation network for accurate and efficient crowd counting," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 734–750.

[60] R. Jia, F. C. Sangogboye, T. Hong, C. Spanos, and M. B. Kjærgaard, "Privacy-preserving building-related data publication using pad," in *Proceedings of the 4th ACM International Conference on Systems for Energy-Efficient Built Environments*. ACM, 2017, p. 32.

[61] K. Zhang, J. Ni, K. Yang, X. Liang, J. Ren, and X. S. Shen, "Security and privacy in smart city applications: Challenges and solutions," *IEEE Communications Magazine*, vol. 55, no. 1, pp. 122–129, 2017.

Xiaochen Fan received his B.S. degree in computer science from the Beijing Institute of Technology, China, in 2013. He is currently a Ph.D. candidate in the School of Electrical and Data Engineering at the University of Technology Sydney (UTS), Australia. His research interests include mobile/pervasive computing, deep learning, and IoT.

Chaocan Xiang is an associate professor at the College of Computer Science, Chongqing University, Chongqing, China. He received the BS and Ph.D. degrees in computer science and engineering from the Nanjing Institute of Communication Engineering, China, in 2009 and 2014, respectively. He studied at the University of Michigan-Ann Arbor in 2017. His current research interests include wireless sensor networks, crowd-sensing networks, and IoT.
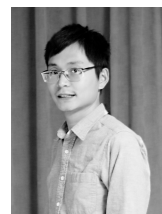
Chao Chen is a full professor at the College of Computer Science, Chongqing University, Chongqing, China. He obtained his Ph.D. degree from Pierre and Marie Curie University and Institut Mines-Télécom/Télécom SudParis, France in 2014. He received the B.Sc. and M.Sc. degrees in control science and control engineering from Northwestern Polytechnical University, Xi'an, China, in 2007 and 2010, respectively. Dr. Chen has published over 80 papers, including 20 ACM/IEEE Transactions. His research interests include pervasive computing, mobile computing, urban logistics, data mining from large-scale GPS trajectory data, and big data analysis for smart cities.
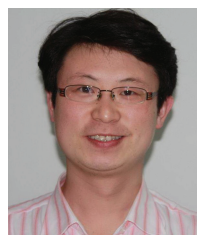
Panlong Yang is now a professor in the School of Computer Science and Technology, University of Science and Technology of China. He received the BS, MS, and Ph.D. degrees in communication and information system from the Nanjing Institute of Communication Engineering, China, in 1999, 2002, and 2005, respectively. His research interests include Mobile/Pervasive Computing, IoT.

LIANGYI GONG is a postdoc researcher at the School of Software and BNRist, Tsinghua University, and an assistant professor at the School of Computer Science and Engineering, Tianjin University of Technology, China. He received the B.E. degree in 2010 and Ph.D. degree in 2016 from Harbin Engineering University, China. His major research interests cover Mobile/Pervasive Computing and Network Security.
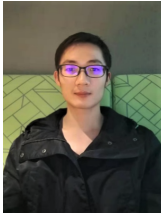
Xudong Song received his B.S. degree in electrical engineering and automation from the Qingdao University of Science and Technology, Qingdao, China, in 2016, and he received the M.S. degree in control science and engineering from the Dalian University, Dalian, China, in 2019. He is currently a visiting scholar at the University of Technology, Sydney. His research interests include IoT, Indoor Localization, Deep Learning, and Edge Computing.

Priyadarsi Nanda earned his Ph.D. degree from the University of Technology Sydney, Australia. He is a senior lecturer in the School of Computing and Communications at University of Technology Sydney, Australia. He has more than 26 years of experience in cybersecurity, Internet of Things security, network quality of service, assisted health care using sensor networks, and wireless sensor networks. He has published more than 70 refereed research articles.

Xiangjian He (M'99 SM'05) received the Ph.D. degree in computing sciences from the University of Technology Sydney, Australia, in 1999. Since 1999, he has been with the University of Technology Sydney. He is currently a Full Professor and the Director of the Computer Vision and Pattern Recognition Laboratory.