Urban Traffic Prediction through the Second Use of Inexpensive Big Data from Buildings

Zimu Zheng
Dept. Computing,
Hong Kong Polytechnic Univ.

Dan Wang
Dept. Computing,
Hong Kong Polytechnic Univ.

Jian Pei
School of Computing Science,
Simon Fraser University

Yi Yuan
Dept. MIG Operation,
Tencent Ltd.

Cheng Fan
Dept. Building Services Eng.,
Hong Kong Polytechnic Univ.

Fu Xiao
Dept. Building Services Eng.,
Hong Kong Polytechnic Univ.

ABSTRACT

Traffic prediction, particularly in urban regions, is an important application of tremendous practical value. In this paper, we report a novel and interesting case study of urban traffic prediction in Central, Hong Kong, one of the densest urban areas in the world. The novelty of our study is that we make good second use of inexpensive big data collected from the Hong Kong International Commerce Centre (ICC), a 118-story building in Hong Kong where more than 10,000 people work. As building environment data are much cheaper to obtain than traffic data, we demonstrate that it is highly effective to estimate building occupancy information using building environment data, and then to further use the information on occupancy to provide traffic predictions in the proximate area. Scientifically, we investigate how and to what extent building data can complement traffic data in predicting traffic. In general, this study sheds new light on the development of accurate data mining applications through the second use of inexpensive big data.

Keywords

Traffic Prediction; Building Occupancy;

1. INTRODUCTION

Traffic prediction, particularly in urban regions, is well recognized as an important application with tremendous practical value. The ability to predict traffic levels is a starting point for dealing with traffic congestion, which can be a costly problem that continues to grow in magnitude. "If, in 2013 traffic congestion cost Americans \$124 billion in direct and indirect losses, this number will rise to \$186 billion in 2030." [7]

The objective of *traffic prediction* is to predict the traffic status of a road or region. It has long been a focus of research in the field of transport engineering, mathematics, and computer science. Many existing traffic prediction methods heavily rely on accurate and rich current or past traffic data. Such traffic data are often obtained from *traffic sensing*, that is, from a traffic monitoring system that is permanently deployed on a road to monitor its current traffic status. However, such traffic monitoring systems are often expensive. For example, high-accuracy traffic monitoring systems use closed-circuit camera detectors [3]. A set of such devices easily costs USD\$2500, plus a 10% top-up service fee for maintenance and installation [35]. In the tunnel crossing Victoria Harbor, Hong Kong, which we focus on in this paper, there are over 30 such devices and 20 other supplementary detectors, with a total cost of USD\$71,500 for this one road. In Hong Kong, the traffic monitoring system costs over one hundred million Hong Kong dollars, but covers only a quarter of the roads in Hong Kong [10].

In this era of big data, as many different kinds of data have been collected, it is natural to ask whether we can take advantage of some other data to facilitate effective traffic prediction. This is the motivation behind our project on using big data collected from buildings to make traffic predictions. The intuition is that changes in *occupancy* of buildings may contribute to changes in traffic in the surrounding areas.

In this paper, through a case study, we try to answer two important questions: 1) whether we can improve the accuracy of traffic predictions using building occupancy information and 2) whether we can use such building occupancy to *replace* traffic data in traffic predictions: this endeavor has significant practical value, making it possible to reduce the cost of setting up a thorough traffic sensing system.

In this paper, we show that the answers to both questions are positive. We overcome a set of technical challenges.

First, building occupancy information is not directly available. We have to estimate occupancy information using *building environment data*, such as electricity usage, CO_2 concentration, elevator status, and so on. We make good second use of such data to estimate occupancy information.

Second, to make occupancy data useful for traffic predictions, and in particular, to replace traffic data in traffic predictions, we need to conduct domain transformation, so that traffic predictions can be conducted by occupancy data only. We develop an occupancy-traffic (OccTra) model for the relationship between occupancy data and traffic data. As a result, we can deploy a temporary traffic monitoring

CIKM'16 , October 24 - 28, 2016, Indianapolis, IN, USA

ISBN 978-1-4503-4073-1/16/10...\$15.00

DOI: http://dx.doi.org/10.1145/2983323.2983357

(a) (b)

Figure 1: The ICC and its neighborhood.

system, which is far less costly than setting up a permanent one, and train a model that uses both traffic data and occupancy data. After that, the traffic monitoring system can be removed and traffic predictions can be conducted using only occupancy data. From a practical point of view, this technique has the potential to save millions of dollars on the costs of the permanent traffic monitoring systems.

Third, we need a unified traffic prediction framework that can naturally adapt to the situations with or without traffic data when predicting. We carefully design the training and prediction phases. Our traffic prediction framework consists of a set of algorithms to extract features from space and time, through Lasso [30], Recursive Feature Elimination (RFE) [8], and Locally Weighted Regression (LWR) [25].

We report a comprehensive evaluation of our traffic prediction framework using a real world case. With support from the Hong Kong Transport Department and the Hong Kong International Commerce Centre (ICC), we collected four months of building data on the ICC (Fig. 1 (a)), and traffic data on the neighboring roads of the ICC (Fig. 1 (b)); such neighboring roads include the West Tunnel, one of the busiest roads in Hong Kong, and Lin Cheung Road. ICC is a 118-story building where more than 10,000 people work, hosting companies such as Morgan Stanley, Credit Suisse, and Deutsche Bank, to name just a few. The West Tunnel connects the emerging business area in Kowloon, where the ICC is located, to the business center in Hong Kong Island across the Victoria Harbor (Fig. 1 (b) shows the locations of the International Finance Center, the Bank of China Tower, and the headquarter of HSBC). The building data and traffic data were recorded every two minutes and every six minutes, respectively. The ICC is divided into individual zones and our building data consist of 124 zones. We show that, given occupancy data on individual zones, we can further improve the traffic prediction results. The intuition is that people in some zones may be more likely to drive and people in some other zones may be more likely to take public transportation. People in different zones may have different effects on traffic status when leaving the building.

The total data is more than 1TB. Cleaning such a massive amount of data is not a small task. We configure a private cloud for our experiments. We show that our traffic prediction approach outperforms the state-of-the-art traffic prediction algorithms in transport engineering by up to 10 times during off-duty times, which are when the predictions are needed most and which are the most challenging periods for predicting traffic. We also compare our approach with the traffic prediction service from Google. Google's approach takes data from Google Map users only, a much sparser set of data; this results in a generally lower prediction accuracy. In one example, Google's traffic prediction can have an error rate of more than 13 times to that of our approach. We further show that we can take one month of occupancy data and traffic data as inputs in the training phase, and conduct traffic predictions for the next three months using occupancy data only. The implication of the cost saving can be remarkable. For example, the traffic sensing system on a road can be reused on three other roads during this period [10]. Using West Tunnel as an example, this has a potential to save USD$201,500 to avoid setting up three additional sets of traffic sensing systems.

The rest of the paper is organized as follows. In Section 2, we review related work. We present feasibility analysis, problem definitions and our traffic prediction framework in Section 3. In Section 4, we present our model of using the CO_2 data to predict the occupancy dynamics. Sections 5 is devoted to the detail models of our framework. We conduct a comprehensive set of evaluations in Section 6. We discuss the business value of our approach in Section 7 and we conclude our paper in Section 8.

2. RELATED WORK

Traffic Prediction: Traffic prediction (or traffic forecasting) has long been a topic of research in the fields of transport engineering, mathematics, and computer science (see a research tree in Fig. 2). Traffic prediction has two broad directions: the long-term and the short-term traffic prediction. In long term traffic predictions, attempts are made to model the physical process that governs the evolution of traffic [11][12]. Such traffic predictions can be used for city blueprints, road system planning, and so on. These studies start from baseline mathematical models, and their success relies heavily on the effort to calibrate the parameters from city to city, which is often a labor-intensive task.

Our work falls into the category of short-term traffic prediction, the aim of which is to predict the day-to-day, hour-to-hour status of traffic. Two directions are common in short-term traffic prediction: predicting the traffic at a location where the computation involves multiple locations, i.e., considering the road network as a graph and jointly considering the traffic at multi-locations; and predicting the traffic at a location by using the (historical) data for this location only. The modeling complexity of multi-location graph-based research [26] is usually high. In this paper, we take the direction of considering only single location data.

Predicting traffic using single location data is the most important direction in traffic prediction, and one of the most intensively studied areas of research on the subject. There are two further directions of research under this area: one in which only traffic data are considered (e.g., history traffic sensing data) and the other in which additional information is taken into consideration. For traffic prediction using traffic sensing data, a great many learning and inference algorithms have been tried, including linear regression [2], univariate and multivariate state-space methods (ARIMA) [33], neural networks [18], k-nearest neighbors [22], locally-weighted regression (LWR) [25], Kalman filtering [17], artificial neural networks and knowledge-based methods [34], and others. Of all these, LWR has been shown to have the best results [4].

For traffic prediction assisted by additional information, two commonly considered classes of information are weather

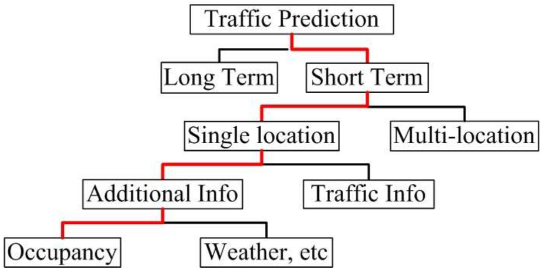


Figure 2: The research line of traffic prediction.

information [28], and weekday-weekend/holiday information [32]. In [39], online information, such as on the weather, sporting events, and holidays, is used to assist in making traffic predictions. Another recent work reports using multi-source data related to traffic (such as taxicabs, buses, trucks, subway and cellular data) to predict traffic status [37]. Such additional information can be seen as assistance information, the objective of which is to assist traffic prediction or avoid major errors of prediction; in other words, if there are changes to such additional information (e.g., changes in weather, a weekday changing into a holiday), the traffic prediction results will change. Building occupancy information has more direct correlation with the traffic status of nearby roads. It not only directly improves the accuracy of prediction, but more importantly, the occupancy data can replace traffic data. This kind of substitution is difficult or even not possible using the aforementioned data.

Traffic Sensing: In traffic sensing, the traffic speed at the *current* time is estimated through traffic monitoring systems. There are many studies on traffic sensing; see a good comparison of different methods in [19]. The key trade-off is between the cost and accuracy.

Various types of sensors have been investigated such as infrared sensors, acoustic sensors, probe sensors, cell phones and participatory sensing in academia [19][20][21][36]. These approaches try to study low cost sensors. They usually have low accuracy, data sparsity or cannot scale in practice [14]. Waze [6] is a traffic sensing application currently in use. It takes advantage of the users' participatory uploads of the traffic status of their nearby roads. Such approach is low cost, yet has an unstable coverage and accuracy.

High-accuracy traffic sensing systems in practice include loop and closed-circuit camera detectors [3]. However, such solutions for traffic sensing suffer from high costs. A set of loop detectors costs USD$9000. A set of camera detectors, with associated RFID sensors, cost USD$2500 and a 10% top up service fee for installation and maintenance [35].

Building Occupancy Sensing: Building systems have recently attracted interest from sensor networking and system researchers. One direction of research is occupancy sensing, i.e., detecting the presence and the number of occupants, so as to turn off unnecessary equipment, adjust HVAC intensity and so on. There are studies on the use of different types of sensors, such as passive infrared (PIR) sensors, reed switches, and motion sensors [1]. There are also studies using camera sensors [5] or electricity consumption [13] to estimate building occupancy. These studies either suffer from scalability problems, or require additional sensors to be deployed that may not be widely adopted in buildings.

In this paper, we choose to use CO_2 concentration for occupancy sensing. CO_2 sensors are widely available in building management systems. CO_2 concentration is the most

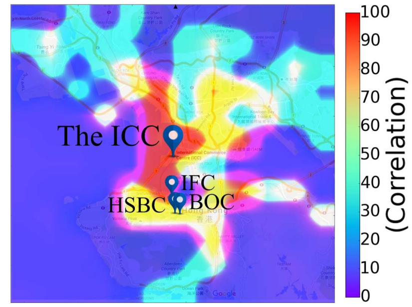


Figure 3: The correlations between the ICC occupancy data and the traffic data for different roads.

readily available information, the main purpose of which is to assist ventilation functions of a building. Intrinsically, higher occupancy levels result in higher concentration of CO_2. The CO_2 approach is scalable and it is more accurate when the number of occupants is large [27][9].

One challenge for using CO_2 is that CO_2 has a delay factor, e.g., the CO_2 level may only reflect occupancy status 10 - 15 minutes ago. In this paper, we first developed a general model that extends the conventional steady-state model in [27] and conduct a real world validation of its effectiveness. We then carefully develop our occupancy-traffic model and prove that the delay factor will not affect the prediction accuracy of our model.

3. FRAMEWORK OVERVIEW

3.1 Analysis on the Feasibility of Using Occupancy Data to assist Traffic Predictions

We first conduct a study of the correlation between the ICC occupancy data and the traffic data for the roads of different distances from the ICC. This can be seen as a proof of concept to show the feasibility of our study.

Note that the further a location is from the ICC, the more time is needed for the impact of the ICC to reach this location. As such, we shift the time series of the traffic data accordingly, using the average traffic speeds [31].

We compute the correlation and show the results in Fig. 3. We see that the closer the roads are to the ICC, the higher the correlation. This is strong evidence of the overall feasibility of using occupancy data to help predict traffic. [1]

3.2 The Problems

We now formally present our traffic prediction problems. Let $\mathcal{S}$, $\mathcal{T}$ be the data series on occupancy and traffic speed, in which $S_m \in \mathbb{R}$, $T_n \in \mathbb{R}$ denote the number of occupants at time m and traffic speed at time n.[2] Let the current time be c. Let the time to be predicted be $c+h$, where h denotes the prediction length.

Problem Traffic Prediction with Occupancy (TPO): Given a training set of $\mathcal{S}$ and $\mathcal{T}$ with time interval $[0, c]$, develop a traffic prediction scheme $\mathcal{G}(\cdot)$ which outputs T_{c+h}, i.e., $T_{c+h} = \mathcal{G}(\mathcal{S}, \mathcal{T})$.

We study two versions of **TPO** in our paper. First, we study a scheme where we always have traffic data. We call

[1] We also find that the average correlation drops to below 50% over four km away, which means the occupancy dynamics of ICC has little effect on the traffic status of roads over four km away.

[2] We may abuse the notations and use S_{ij}, T_{ij} to denote the occupancy and traffic speed at time j on day i.

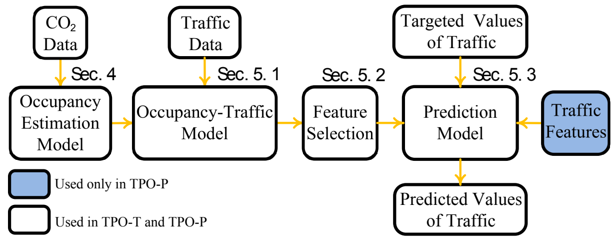


Figure 4: The Traffic Prediction Framework.

this problem **Traffic Prediction with Occupancy under Permanent Traffic Sensing (TPO-P)**. Second, we study a scheme where we use occupancy data to replace traffic data in the prediction phase; thus we only need temporary traffic sensing for the training phase. This problem is called **Traffic Prediction with Occupancy under Temporary Traffic Sensing (TPO-T)**. We assume that occupancy data are always available since the building system is deployed for building functions; and we make a second use of such data.[3]

We divide **TPO** into a training phase and a prediction phase. Let $\mathcal{S}_t$, $\mathcal{T}_t$ be the set of occupancy data and traffic data in the training phase. Let $F(\cdot)$ be the function developed for the training phase; it takes $\mathcal{S}_t$, $\mathcal{T}_t$ as inputs. For simplicity, we denote the trained model outputted by $F(\cdot)$ as $R_F = F(\mathcal{S}_t, \mathcal{T}_t)$. Let $\mathcal{S}_p$, $\mathcal{T}_p$ be the set of occupancy data and traffic data in the prediction phase. Let $G(\cdot)$ be the function developed for the prediction phase, which outputs T_{c+h} with the trained model R_F and $\mathcal{S}_p$, $\mathcal{T}_p$ as inputs.

Problem TPO-P: Given a training set of $\mathcal{S}$ and $\mathcal{T}$ with time interval $[0, c]$, develop a function $F(\cdot)$ for the training phase with output $R_F = F(\mathcal{S}_t, \mathcal{T}_t)$, and a function $G(\cdot)$ for the prediction phase such that $T_{c+h} = G(\mathcal{S}_p, \mathcal{T}_p | R_F)$.

Problem TPO-T: Given a training set of $\mathcal{S}$ and $\mathcal{T}$ with time interval $[0, c]$, develop a function $F(\cdot)$ for the training phase with output $R_F = F(\mathcal{S}_t, \mathcal{T}_t)$, and a function $G(\cdot)$ for the prediction phase such that $T_{c+h} = G(\mathcal{S}_p | R_F)$.

3.3 The Traffic Prediction Framework

In this paper, we develop a unified framework that solves both **TPO-P** and **TPO-T**, i.e., we can naturally adapt to prediction with or without traffic data in the prediction phase. Especially in **TPO-T**, we only have occupancy data in the prediction phase. The core challenge is that the relationship between occupancy and traffic status is dynamically changing from time to time. Thus, the occupancy data cannot become effectively features alone, conventionally with a fixed synthetic model between occupancy and traffic. Consequently, instead of using a fixed model, we train a dynamic *occupancy-traffic (OccTra) model* for the relationship between the occupancy and traffic data when extracting features. The output of such models is *the time-warped offset*, i.e., a non-uniform time difference between the occupancy data and traffic data. The time-warped offset is used to extract raw occupancy data into useful features in the prediction phase. In addition, we do not directly have raw occupancy data. We develop an *occupancy estimation model (OEM)* to transform the CO_2 data into raw occupancy data. Besides the OEM and the OccTra model, we also develop a *prediction model*, i.e., a weighting matrix on features, to predict the targeted traffic status with our features.

We show our framework in Fig. 4. The framework is divided into a training phase and a prediction phase. In the training phase, our OEM model takes CO_2 data as inputs and outputs raw occupancy data. The OccTra model takes the occupancy data and traffic data as inputs. With it, we develop time-warped offsets and extract useful features from the occupancy data. We then have feature selection on both temporal and spacial dimensions. The temporal selection is based on different time slots and the spacial selection is based on different zones of the building. Intuitively, different time slots and different zones of a building have different impacts on the prediction. Our prediction model is a learning algorithm that outputs a matrix of weighting coefficient. In the prediction phase, the occupancy estimation model transforms CO_2 data into raw occupancy data. By combining time-warped offsets developed in the training phase, the raw occupancy data become useful features. After feature selection, traffic prediction is conducted through the matrix of weighting coefficient developed in the training phase.

If we have permanent traffic data, for both the training and prediction phase, feature selection will be performed additionally on the traffic features for the temporal dimension. Thus, we have a unified traffic prediction framework.

4. OCCUPANCY ESTIMATION MODEL

We now describe our occupancy estimation model (OEM) for our traffic prediction framework. OEM is a joint cyber-physical model based on ASHRAE 62-1989R. [4]

ASHRAE 62-1989R proposed a steady-state model to estimate the occupancy of a building as follows.

$$U^{(tot)}E + R^{(s)}(C^{(out)} - C^{(in)}) = 0 \qquad (1)$$

Where $U^{(tot)}$ is the number of occupants in a room (or a zone); E is the amount of CO_2 a person generates per second; $R^{(s)}$ is the volume of air flow per second; and $C^{(in)}$ and $C^{(out)}$ are the CO_2 concentration in the room and outdoor air, respectively. Thus, the first part, $U^{(tot)}E$, is the total amount of CO_2 generated by all occupants and the second part, $R^{(s)}(C^{(out)} - C^{(in)})$, is total amount of CO_2 evacuated from the room by air flow.

The above model is used for steady-state situations. Yet we focus on *dynamic scenarios* with finer granularity. Thus, we discretize the time series and develop our OEM model:

$$(U_i^{(tot)}E + R_i^{(s)}(C_i^{(out)} - C_i^{(in)}))(t_{i+1} - t_i) = V(C_{i+1}^{(in)} - C_i^{(in)}) \qquad (2)$$

Here, $(t_{i+1} - t_i)$ is the time interval between two records. In the dynamic scenario, one key difference is that in a time interval, there may be CO_2 left over from one time interval to the next time interval. This is captured by $(C_{i+1}^{(in)} - C_i^{(in)})$, which is not available in the steady-state equation, i.e., Eq. 1. We also find that the model needs to be calibrated for different rooms with different sizes. This is reflected by V, the volume of the room. In other words, since our Equation 2 is not in the steady state, the left hand side of it computes the CO_2 left over in this time interval (not zero), which is equal to the right hand side of the equation.

[3] We clearly admit that there are business, privacy, and other issues regarding whether and how the building data should be shared. In this paper, we do not consider such issues, but instead focus on technical problems.

[4] American Society of Heating, Refrigerating and Air Conditioning Engineers (ASHRAE) 62-1989R can be seen as a standard; e.g., RFC 791/793 (TCP/IP) in the context of computer science.

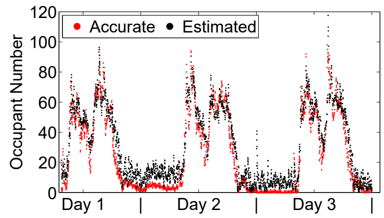


Figure 5: Validation of the OEM model in the ICC.

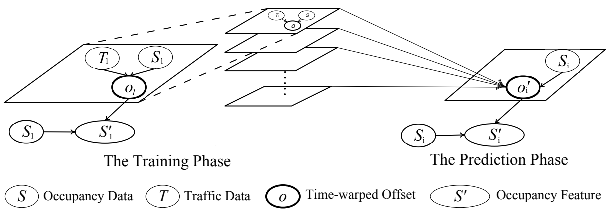


Figure 6: An example of the OccTra Model.

A comprehensive evaluation of our OEM model is available in our previous work [27]. To strengthen the confidence of our OEM model, we further conducted three-day experiments of the OEM model in a zone of the ICC. The maximum occupancy in this zone is 98. The zone has an area of 570.562m^2, with height of 2.85m. Thus, $V = 1626.1\text{m}^3$. We recorded the CO_2 data of the model from BAS. To obtain the ground truth of the occupancy data, we sent one student on-site to count the occupancy data manually in this zone for three days. We show the result in Fig. 5. We can see that our model is fairly accurate. We also found that error is normally distributed, with the mean of the error is -5.7 and the root mean square error is 9.3.

Our validation confirms that our OEM model is applicable in general. In practice, building management are zone based. We thus can estimate building occupancy on a per zone basis and add together the occupancy for all zones.

There is also a delay factor because time is needed to distribute the CO_2 concentration. In Section 5.1.2, Theorem 1, we will show that in the OccTra model, such a delay factor will be absorbed and will not affect the prediction accuracy of our traffic prediction framework.

5. THE TRAFFIC PREDICTION MODELS

We now describe the details of the OccTra model, the feature selection and the prediction model in our framework.

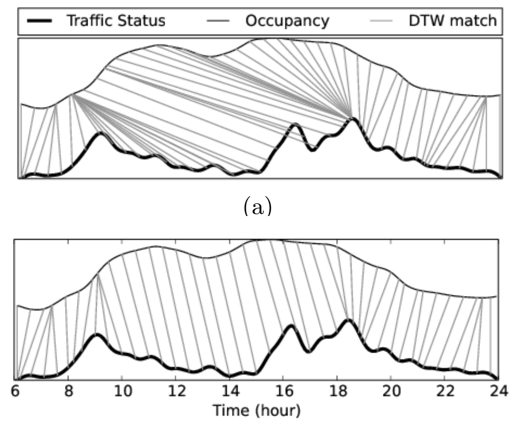
5.1 The Occupancy-Traffic (OccTra) Model

5.1.1 Preliminaries

As discussed, we need a domain transformation, so that we are able to extract useful occupancy features for the traffic prediction. We thus develop a model for the relationship between the occupancy data and traffic data.

Intuitively, the changing of occupancy at one moment can result in the changing of traffic status at another moment. In other words, there is a time shift between the dynamics of occupancy and traffic status, which may come from occupants taking elevators, powering up their cars, and so on. The core challenge here is to clarify such a time shift between occupancy and traffic, and find out a useful feature space of occupancy to predict traffic.

Our idea is to introduce *time-warped offsets* between the two data series. For example, if the occupancy status at 4:15pm matches the traffic status at time 5:15pm, then the value of the offset is one hour. Note that such offsets are *non-uniform*, i.e., they differ from time to time. Formally,



(b)

Figure 7: ICC occupancy and traffic speed of West Tunnel (the data come from a randomly selected day in the four months): (a) the default distance function; (b) the new distance function.

DEFINITION 1. *(Time-warped offset) Assuming m is the time index of a targeted value in data series $\mathcal{T}$, and n is the time index of a feature in data series $\mathcal{S}$ by using a matching function between $\mathcal{S}$ and $\mathcal{T}$. The time-warped offset o_m is the difference between m and n, i.e., o_m=n-m.*

We show an overview of our model in Fig. 6. In the training phase, we develop an algorithm based on the well-known Dynamic Time Warping (DTW) [16], the Occupancy-Traffic-DTW Algorithm (DTW-OT), to compute the time-warped offsets. DTW-OT (Section 5.1.2) takes raw occupancy data and traffic data as inputs. It computes the matching between $\mathcal{S}$ and $\mathcal{T}$ and outputs the time-warped offsets. We prove that the DTW-OT algorithm naturally absorbs the delay factor discussed in Section 4 (Section 5.1.3). We develop an offset-feature shift function to transform occupancy data into occupancy features with these offsets (Section 5.1.4). The occupancy features are later used in feature selection.

In the prediction phase, we develop a Kernel Average algorithm (Section 5.1.5) to extract occupancy features. Without current traffic data in this phase, we explore the past days that have similar occupancy patterns with the current day. The time-warped offsets of these past days are used. The offset-feature shift function is again used to transform occupancy data into occupancy features with these offsets. The occupancy features are later used in feature selection.

5.1.2 Computing Time-warped Offsets

Our objective is to find good point-to-point matching between $\mathcal{S}$ and $\mathcal{T}$, and calculate the time-warped offsets. Note that we care about matching the shape of $\mathcal{S}$ and $\mathcal{T}$. We develop an algorithm based on Dynamic Time Warping (DTW) [16]. In DTW, the shape of two time series is important. DTW outputs point-to-point matching of the two data series (see Fig. 7 (a)), where such matching minimizes the total distance between the two data series.

DTW is originally developed to connect data series of the same type, e.g., two audio records. It cannot be directly applied to connecting two different types of data series, such as occupancy data and traffic data. We make three important modifications and develop our DTW-OT algorithm.

First, DTW does not limit the length of the connection between two nodes in its distance function. For example, using default matching, we found that DTW may match occupancy at 10:00am to traffic speed at 7:30pm. This is called *over matching*, and it is clearly unreasonable in our context. We thus modify the distance function as follows:

$$D(S_m, T_n) = \begin{cases} |S_m - T_n| & |m-n| \leq \Delta \\ \frac{|S_m - T_n|}{\alpha^{|m-n|}} & |m-n| > \Delta \end{cases} \tag{3}$$

where α is a constant satisfying $0 < \alpha < 1$ and Δ is a threshold. We set $\alpha = 0.2$, the same to the UTCS system in Washington, D.C. [17]. In this distance function, after the difference exceeds a threshold Δ, the distance increases quickly. In other words, the two data points become less likely to match with each other.

Second, two different types of data commonly have different value ranges, e.g., the value of occupancy is always greater than 1000 and that of traffic speed is around 60. As such, all points in the occupancy data series will match with the point of maximum value in the traffic data series. In our DTW-based algorithm, both $\mathcal{S}$ and $\mathcal{T}$ are first normalized.

Third, two different types of data series can have different types of correlations, e.g., positive or negative; yet the assumption in DTW is that two time series should be positively correlated. In our scenario, $\mathcal{S}$ and $\mathcal{T}$ are negatively correlated. We thus define *Traffic Status*: $T^{-1} = \frac{1}{\text{Traffic Speed}}$, and use T^{-1} as the $\mathcal{T}$ data series.

Our DTW-OT is DTW considering all of the three problems. We particularly compare the results of the matching, using the default and the new distance function in Fig. 7, both with normalization and positive correlation. We see that in Fig. 7(b), over matching is avoided.

5.1.3 The CO_2 Delay Effect under DTW-OT

We now show that DTW-OT rectifies the CO_2 delay effect discussed in Section 4. Recall that using CO_2 concentration, we cannot estimate the current occupancy, but the occupancy of some time ago. Let δ_m be the *estimation delay* of S_m. In other words, S_m intrinsically is $S_{m'}$, where time point $m' = m + \delta_m$. In practice, δ_m is around 10 - 15 minutes.

THEOREM 1. *Assume that without estimation delay, a point S_m is connected to T_n by DTW-OT. S_m intrinsically is $S_{m'}$ where $m' = m + \delta_m$. Then, under DTW-OT, $S_{m'}$ is still connected to T_n, if the computed offset $|m' - n| \leq \Delta$.*

PROOF. Due to page limitations, our formal proof is in [38]. Intrinsically, both DTW and DTW-OT match the shape. The CO_2 delay effect does not change the shape of $\mathcal{S}$. It only moves $\mathcal{S}$ forward. For example, it moves S_m forward for δ_m time slots. If the delay δ_m will be consumed by DTW-OT, if it is small or we set a reasonable Δ so that $|m' - n| < \Delta$. We show extensive illustrations in [38]. $\square$

5.1.4 Extracting Occupancy Features for Training

After we have the time-warped offsets o_m, we develop an offset-feature shift function to extract occupancy features. Intrinsically, the origin data series of occupancy is replaced by another with the index shifted by o_m. Then, we take the $\Delta_o = o_m - o_{m-1}$ as the occupancy feature, because changes of traffic status are correlated with changes in occupancy rather than absolute value of occupancy.

One nuance is that after transfer, occupancy data may not exist on the shifted time index. For such cases, we use the average of the occupancy of two neighboring time indices.

5.1.5 Extracting Occupancy Features for Prediction

However, in the prediction phase, traffic data is no longer available. Then, the above DTW-OT, which takes both occupancy and traffic data series as input, can not be applied here. Thus, we need another approach for feature extraction in this phase. Our approach, as mentioned at the beginning of this section in Fig. 6, is to combine the current occupancy data in prediction phase and the time-warped offsets developed in the training phase.

With the knowledge learnt in the training phase, we can still acquire the time-warped offset in the prediction phase, even without the current traffic data on the very day. To find such offsets for the current occupancy, we first find the days similar to the current day, in terms of the occupancy data pattern. Intuitively, the more similar two days are, the more likely their offsets are to be the same, which is verified with experiments in our technical report [38].

With the similarity of days measured, we apply the Kernel Average (KA) algorithm to obtain the offset, so that the occupancy feature can be extracted for our predictor using the offset-feature shift function. KA is based on the weighted average, obtained by imposing different weights on each day. The more similar the two days are, the higher the weight given to the time-warped offset of this history day, by developing the kernel function in KA. Our experience shows that KA is better than other algorithms such as k-NN, since k-NN suffers from errors due to its locality.

5.2 Feature Selection

We now describe our feature selection, on both temporal and spacial dimensions. The objectives are:

1. To select those time slots that are more important for prediction. As an example, to predict the traffic at 21:10pm, training the model with features at 21:00pm, 21:06pm may be much more effective than at 10:00am, 10:06am (both the traffic and the occupancy). The less-important features may also introduce *over-fitting*, which brings even more prediction errors.
2. To remove those building zones that do not contribute to the prediction. This follows our insight that some building zones contribute less to the traffic status than others since they host more people who take public transportation, while the other zones host more people who can afford to drive and park.

5.2.1 Selecting Time Slots

For both occupancy and traffic features, some time slots are consistently much more important than others. If we keep these features, we can avoid over-fitting and improve prediction accuracy and computation efficiency.

For time slot features, we use *Lasso* [30], a well-known feature selection method. Lasso outputs a subset of features, by formalizing and optimizing the objective function of feature selection. One important improvement for Lasso is to consider *Lag Effects*. In general, the time points that are closer to the prediction time k should have higher weights. We chose the adaptive Lasso in [40], where a lag-weighted Lasso is used to reflect the lag effects.

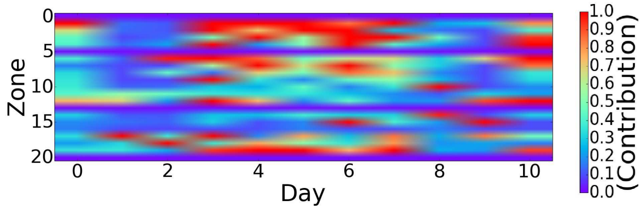


Figure 8: The contribution of the zone feature is measured as its prediction accuracy. Such contributions are computed on different days.

5.2.2 *Selecting Zones*

Zone features differs greatly from time slot features since we found that there are no zones that are consistently important. Important zones keep changing every day. Nevertheless, we found that there are zones that are consistently irrelevant and only introduce noise. Intuitively, these zones are likely to host people who only take public transportation. We show this using a heat map in Fig. 8.

Thus, instead of selecting important zones, we choose to remove irrelevant zones. We use *Recursive Feature Elimination* (RFE) [8] to remove zones. RFE selects features by recursively reducing more irrelevant features, given an external predictor that assigns weights to features. As for the zone features that are important on different days, we will assign different weights in our final learning model. In each iteration, we remove the zone inducting greatest error and obtain the final size of feature space for occupancy.

5.3 The Prediction Model

We adopt the state-of-the-art Locally Weighted linear Regression (LWR) [24] as our learning and prediction algorithm to train the weighting coefficient matrix. It has been shown to outperform many existing traffic prediction algorithms such as k-NN and Neural Network methods [4].

Note that for **TPO-T**, the weighting coefficient matrix is a set of weights on the occupancy features and historical traffic features. For **TPO-P**, there are occupancy and traffic features on the predicting day. Therefore our algorithm also take the traffic features as input. When the contributions of occupancy features are low, e.g., with a low correlation between the two data series, occupancy features are assigned lower weights and traffic features play a more important role.

Last, but not least, our model can be extended to multiple buildings since our model is zone-based. Each zone contributes a dimension in the weighting coefficient matrix. We emulate multiple buildings and show that our model is general in Fig. 20 of Section 6.

5.4 Computational Complexity of the Models

Now we analyse the computational complexity. Our model consists of the training phase and the prediction phase. Below shows the computation complexity of the training phase:

THEOREM 2. *The computational complexity of our model in the training phase is $O(ZQ(P^2))$, where Z is the number of building zones, Q is the number of days, and P is the number of time slots in each day.*

Due to page limitations, our formal proof is in [38]. Note that there is also a non-trivial time for data loading. The computational complexity is of $O(Q)$. This is done once.

We emphasize more on the computational complexity of the prediction phase, because it is necessary for real time traffic prediction. We evaluate both theoretically, and in evaluation. The prediction phase consists of multiplying the coefficient matrix of the regression, which is linear to the number of training examples ZP:

THEOREM 3. *The computational complexity of our model in the prediction phase is $O(ZP)$, where Z is the number of zones, and P is the number of time slots in each day.*

Due to page limitations, our formal proof and evaluation are in [38]. In evaluation, the running time for prediction is 0.37s on average, which is suitable for real time usage.

Data Sets		Traffic	Occupancy
♯ attributes		53,308,800	64,281,600
♯ features		13,327,200	8,035,200
Permanent Sensing	training set	9,995,400	6,026,400
	testing set	3,331,800	2,008,800
Temporary Sensing	training set	3,331,800	2,008,800
	testing set	9,995,400	6,026,400

Table 1: Summary of the data sets.

6. EVALUATION

6.1 Evaluation Setup

Data sets: We use real data sets collected from the Hong Kong Transport Department and the ICC for four months (May to August). The data sets consist of two parts, which are shown in Table 1 and elaborated below.

1. **Traffic sensing data**: The data were collected every six minutes for 617 roads installed with traffic monitoring systems, which consist more than one fourth of the roads in Hong Kong. Each traffic sensing data record contains four important attributes: the road ID, traffic speed, date and insert time.

2. **Occupancy sensing data**: The data were collected from the Building Management System of the ICC. The data were collected every two minutes in 124 building zones of the ICC, hosting over 70 companies on 118 stories. Each BMS record contains five important attributes for our occupancy estimation: the zone ID, time stamp, CO_2 concentration, the operation and parameter of air flow controls.

The full data set is more than 1TB in storage size in an uncompressed CSV format. After cleaning unnecessary roads and attributes, the total size of the data we used in this evaluation section is 132.47G.

Execution environments: To process the big data, we establish a private cloud to run our experiments. The cloud has 12 cores, each with 2.6GHz, and a total memory of 64G.

Comparing scenarios: We evaluate **TPO-P** and **TPO-T**. To compare **TPO-P**, we use a state-of-the-art traffic prediction algorithm [24]. We denote this algorithm as **TP-P**, **T**raffic **P**rediction under **P**ermanent traffic sensing. We also compare our algorithm **TPO-P** with the traffic service provided by Google [29]. We crawl data directly from Google Map traffic service. As discussed, Google traffic prediction takes data from Google Map users, and is sparse. We found that it performs worse than the state-of-the-art traffic prediction algorithm and our algorithm (see Fig. 9). Thus, our evaluation is mainly between **TPO-P** and **TP-P**.

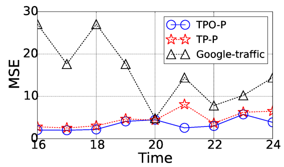


Figure 9: MSE as a function of time, between 16:00 and 24:00.

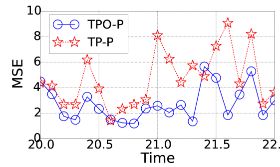


Figure 10: MSE as a function of time, between 20:00 and 22:00.

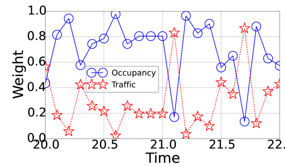


Figure 11: Weight of Occupancy against Traffic as a function of time.

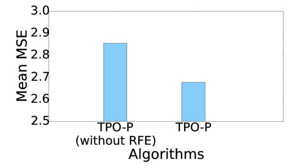


Figure 12: MSE of TPO-P without and with the algorithm of RFE.

To the best of our knowledge, there is no specialized algorithm for traffic prediction with temporary traffic sensing. For comparison, we use *historical average* [23] since it does not require features of the days to be predicted. We denote this algorithm as **TP-T**. We also compare **TPO-T** with **TP-P** and **TPO-P** to show that our traffic prediction scheme can replace permanent traffic sensing with a very moderate sacrifice, if any, of prediction accuracy.

Evaluation criteria: We adopt one most commonly used mean squared error (MSE) [15] to compare the prediction performance. MSE is defined as $\frac{1}{P}\sum_{j}^{P}(t_j - t'_j)^2$, where t'_j is the predicted traffic at time $j \leq P$ and t_j is the real traffic.

Default evaluation parameters: We split our four-month data into a training set and a prediction set. For all *permanent* scenarios, the default training set is three months and the prediction set is one month. For all *temporary* scenarios, the default training set is one month and the prediction set is three months. The default period for evaluation is 16:00 to 24:00. The default prediction length h is one hour, e.g., for a current time of 16:00, we predict 17:00. The default road for prediction is the West Tunnel.

6.2 Evaluation Results

6.2.1 Permanent Traffic Sensing

Prediction accuracy: We first compare the accuracy of **TPO-P** and **TP-P**, i.e., traffic prediction algorithms with and without occupancy data under permanent traffic sensing; as well as **Google-traffic**. Fig. 9 shows the MSE as a function of time. Since the prediction is for a whole month (August), the MSE is an average of all days in this month. For example, the MSE of **TPO-P** in 21:00 is 2.57, which shows the average error of all days at 21:00 in August.

In Fig. 9, we can see that with occupancy, **TPO-P** almost always outperforms **TP-P**. In particular, we see that the prediction error of **TP-P** is much higher in the period of 20:00-22:00. For example, the MSE of **TP-P** in 21:00 are 14.44 and 8.10 respectively, much higher as compared to other time periods. As a comparison, our algorithm **TPO-P** remains stable in terms of prediction error; e.g., in 21:00, the MSE is only 2.57. That is to say, **TP-P** has an error that is 3.14 times to that of **TPO-P**. This clearly shows the effectiveness using occupancy data and our algorithm.

Google-traffic performs worse than both **TPO-P** and **TP-P**. For example, at 18:00, **Google-traffic** has an error rate of 13.07 times to that of **TPO-P**. On average, **Google-traffic** has an error rate of 5.38 times to that of **TPO-P** and 4.43 times to that of **TP-P**. Thus, in the remaining part of the paper, we only compare our algorithm with **TP-P**.

We show a more fine-grained comparison between 20:00 and 22:00 in Fig. 10; with comparison made every six minutes, which the minimum period in our data. The results further confirm that **TPO-P** outperforms **TP-P**.

We next show the weights that are automatically assigned to the occupancy features and traffic features during the execution of the training phase of our algorithms. Figure 11 shows that the occupancy features have higher weights. This provides a strong justification from the micro view for the effectiveness of using occupancy data in traffic predictions.

Figure 12 shows that RFE (selecting zones) can provide an additional improvement of 7.59%. Note that RFE is a greedy heuristic in removing zones. We believe that improvement is possible and designing a more optimized zone selection algorithm would make an interesting future study.

Impact of training length: In Fig. 13, we show the impact of training length on the prediction accuracy in **TPO-P**. We use different training length of 0.5 month, one month, and up to three months. We can see that the longer we train, the better the prediction accuracy. We see that when the training length is greater than one month, in particular, after 1.5 months, the prediction results start to become stable.

Impact of prediction length: We study the impact of prediction length h on the prediction accuracy. In Fig. 14, we take an average of the prediction results of 20:00 - 22:00. The x-axis shows the prediction length, e.g., predicts the future 12 minutes, 24 minutes, 36 minutes, and so on. In general, our algorithm **TPO-P** remains stable even when the prediction length increases. The prediction error of **TP-P** gradually increases from 4.7 to 5.3 as the prediction length increases. This demonstrates that the use of occupancy data can result in *earlier* forecasts, whereas with traffic data only, the longer the prediction length, the higher the error.

6.2.2 Temporary Traffic Sensing

We next compare the accuracy of **TPO-T** and **TP-T**, i.e., traffic prediction with and without occupancy data with temporary traffic sensing. Fig. 15 shows the MSE as a function of time. We see that with occupancy information, **TPO-T** almost always outperforms **TP-T**.

On many regular days, the traffic can be very stable, and it is much easier to achieve a good prediction result. The focus of a traffic prediction algorithm should be more on predicting situations where there are changes. To evaluate this aspect, we rank our evaluation cases according to the amount of change that occurs in traffic. We extract the top 25% of cases of change in traffic. We plot the ratio between **TPO-T** MSE and **TP-T** MSE in Fig. 16. We see that **TPO-T** can outperform **TP-T** by as much as 72.85 times. On average, **TPO-T** outperforms **TP-T** by 12.80 times in the top 25% of cases of change in traffic.

We next compare **TPO-T** with **TP-P**, i.e., traffic prediction with occupancy data under temporary traffic sensing and traffic prediction under permanent traffic sensing. This can show the extent to which occupancy data can replace the expensive permanent traffic sensing. For **TPO-T**, we use May for training and predict the traffic in August. For

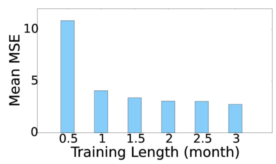


Figure 13: Mean MSE as a function of training length of months.

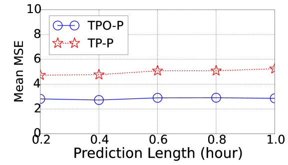


Figure 14: Mean MSE as a function of prediction length, 20:00-22:00.

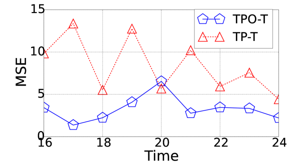


Figure 15: MSE as a function of time, between 16:00 and 24:00.

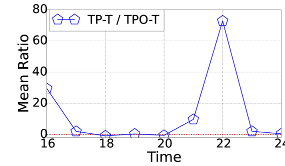


Figure 16: Mean Ratio as a function of time, the most volatile 25% cases.

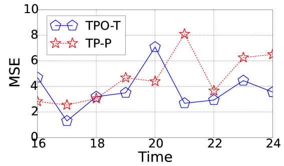


Figure 17: MSE as a function of time, between 16:00 and 24:00.

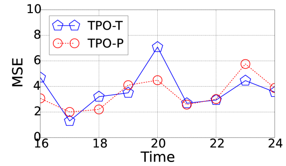


Figure 18: MSE as a function of time, between 16:00 and 24:00.

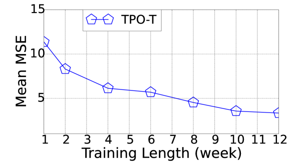


Figure 19: Mean MSE as a function of training length, in one month.

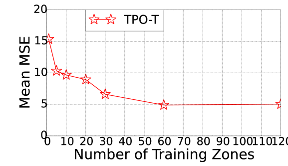


Figure 20: Mean MSE as a function of the number of zones, in one month.

TP-P, we use May, June, and July for training and predict the traffic in August. The results are in Fig. 17. Our algorithm has a performance that is comparable to **TP-P**. In particular, we are better in the period of 21:00 - 24:00.

We next compare **TPO-T** with **TPO-P** in Fig. 18. On average, **TPO-P** has better performance; yet the performance of our algorithm **TPO-T** is close to that of **TPO-P**. Again, this illustrates the feasibility of replacing permanent traffic sensing with temporary traffic sensing.

In Fig. 19 we show the impact of training length on prediction accuracy. We use different training length of one week, two weeks, four weeks (the default value), and up to eight weeks. We can see that the longer we train, the better. We see that when the training length is greater than four weeks, the performance gain becomes smaller.

We next emulate multiple buildings. This is done by dividing the ICC into multiple buildings based on different numbers of zones (from one zone to 120 zones). Each of them can be considered as different buildings. For example, the first building contains the first zone; the second building contains the first five zones; the third building contains the first ten zones, and so on. We show the results in Fig. 20. We can see that, the more building zones one has, the better.

We also evaluate the performance of our traffic prediction scheme on Lin Cheung Road, the impact of Lasso, the impact of RFE on our traffic prediction framework, and the running time of our algorithms. Due to page limitations, these results can be found in [38].

7. DISCUSSIONS

Traffic prediction has been investigated heavily in various background contexts; and traffic prediction systems exist in practice. Our approach is suitable for densely populated urban areas with a large number of buildings, where traffic prediction is needed most, such as Hong Kong, New York, Singapore, Tokyo and many cities in Europe and China.

The traffic status may be the ensemble of occupancy dynamics from multiple buildings. We have evaluated the cases with various numbers of building zones. We can see that, the more building zones one has, the better. Thus, one may consider taking data from more building zones, if he does not have data from a population-dense center as ICC but still want to achieve high prediction accuracy.

We have demonstrated that we can outperform the state-of-the-art traffic prediction algorithms. Nevertheless, our traffic prediction scheme is not only an algorithm focusing on improving the traffic prediction accuracy over past algorithms. The implication of the saving by our traffic prediction approach on the traffic sensing systems is remarkable. For example, assume there are four roads. Rather than deploying four sets of permanent traffic sensing systems, we can deploy one set of traffic sensing system as a temporary traffic sensing system, reuse it on each road for one month. As a matter of fact, we are confident that the interval of temporary traffic sensing on a road can be extended based on our evaluation results, making it possible to further reduce the number of traffic sensing systems.

Take Hong Kong as one example. Our traffic sensing systems use camera and RFID detectors. The cost of a set of the camera detector we used is USD$1500 ($1000 for a camera, $200 for a controller and $300 for a modem), and the cost of a set of RFID detector we used is USD$1000 ($500 for a reader, $200 for a controller and $300 for a modem). There is another 10% top up installation and service fee. On West Tunnel, there are 30 sets of cameras and 20 sets of RFID detectors. The total cost of four sets of permanent traffic sensing systems to cover four roads can be USD$286,000. If we use one set of such system as a temporary sensing system, i.e., we reuse these devices, and assume the installation and service fees increase for three times, the total cost will be USD$84,500, a more than 70% saving. Given that there are thousands of roads in Hong Kong, and three quarters of the roads are not covered primarily due to high costs, the potential saving can be millions of dollars [10].

To sum up, our study can promote the research and practice not only of traffic predictions, but also of traffic sensing systems. Portable traffic-sensing systems can become popular, as they become serious choices with our techniques.

8. CONCLUSION

In this paper, we studied the second use of building data for predicting traffic on nearby roads. While traffic prediction algorithms have traditionally relied on only traffic sensing data, we showed that building data could replace traffic sensing data and improve traffic prediction accuracy.

We developed a traffic prediction framework with a set of novel models for the training and prediction phases. We reported a comprehensive evaluation of our traffic prediction framework using four months of fine-grained data from the ICC and neighboring roads in Hong Kong.

We showed that our traffic prediction framework outperforms state-of-the-art traffic prediction algorithms. By using our traffic prediction framework, we can also use building occupancy data to replace traffic data from the permanent traffic sensing systems. From a practical point of view, the implication of the saving by our framework can be over 70% on traffic sensing systems. This can be translated into millions of dollars in Hong Kong alone, since the traffic sensing system costs over one hundred million Hong Kong dollars, but covers only a quarter of the roads in Hong Kong.

9. ACKNOWLEDGMENTS

The authors would like to thank Peiyuan Zhou, Shuhang Gu, Sijia Cai, Prof. Lei Zhang, Prof. Wenjie Li of The Hong Kong Polytechnic University and Junjie Hu of Carnegie Mellon University for their valuable discussion and feedback.

Parts of this work were supported by an NSERC Discovery grant, the Canada Research Chair program, the National Natural Science Foundation of China under grant No. 61272464, RGC/GRF PolyU 5264/13E, RGC/GRF 152181/14E, the MIG Operation Department of Tencent Co. Ltd., and a sponsored research grant from Huawei Technologies Co. Ltd. All opinions, findings, conclusions and recommendations in this paper are those of the authors and do not necessarily reflect the views of the funding agencies.

10. REFERENCES

[1] Y. Agarwal, B. Balaji, R. Gupta, J. Lyles, M. Wei, and T. Weng. Occupancy-driven energy management for smart building automation. In *Proc. ACM BuildSys'10*, 2010.

[2] E. Chung and N. Rosalion. Short term traffic flow prediction. In *Proc. ATRF'01*, pages 1897–1901, Hobart, Tasmania, Australia, Jan. 2001.

[3] B. Coifman and M. Cassidy. Vehicle reidentification and travel time measurement on congested freeways. *Transportation Research Part A: Policy and Practice*, 36(10):899–917, 2002.

[4] C.P.I. Van Hinsbergen, J.W.C. Van Lint and F.M. Sanders. Short term traffic prediction models. In *Proc. the 14th ITS World Congress, Beijing, China*, 2007.

[5] V. L. Erickson, Y. Lin, A. Kamthe, R. Brahme, A. Surana, A. E. Cerpa, M. D. Sohn, and S. Narayanan. Energy efficient building environment control strategies using real-time occupancy measurements. In *Proc. ACM BuildSys'09*, 2009.

[6] V. Goel. Maps that live and breathe with data. *New York Times*, 2013.

[7] F. Guerrini. Traffic Congestion Costs Americans $124 Billion A Year, Report Says. *Forbes*, October 2014.

[8] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik. Gene selection for cancer classification using support vector machines. *Machine learning*, 46(1–3):389–422, 2002.

[9] H. Han, K.-J. Janga, C. Han, and J. Lee. Occupancy estimation based on CO_2 concentration using dynamic neural network model. In *Proc. AIVC'13*, 2013.

[10] B. Huang. *Personal communication, Transport Department of Hong Kong.* 2015.

[11] INRIX home page. http://www.inrix.com/default.asp, 2016.

[12] F. Jin and S. Sun. Neural network multitask learning for traffic flow forecasting. In *Proc. IEEE IJCNN'08*, pages 1897–1901, Nagoya, Japan, Jan. 2008.

[13] W. Kleiminger, C. Beckel, T. Staake, and S. Santini. Occupancy detection from electricity consumption data. In *Proc. ACM BuildSys'13*, 2013.

[14] H. K. Le, J. Pasternack, H. Ahmadi, M. Gupta, Y. Sun, T. Abdelzaher, J. Han, D. Roth, B. Szymanski, and S. Adali. Apollo: towards factfinding in participatory sensing. In *Proc. ACM/IEEE IPSN'11*, pages 129–130, 2011.

[15] A. M. Mood. Mean-squared error. In *Introduction to the Theory of Statistics*, chapter 7, pages 291–294. 1950.

[16] M. Muller. Dynamic time warping. *Information retrieval for music and motion*, pages 69–84, 2007.

[17] I. Okutani and Y. J. Stephanedes. Dynamic prediction of traffic volume through Kalman filtering theory. *Transportation Research Part B: Methodological*, 18(1):1–11, 1984.

[18] D. Park and L. Rilett. Forecasting multiple-period freeway link travel times using modular neural networks. *Transportation Research Record: Journal of the Transportation Research Board*, 1617:163–170, 1998.

[19] R. Sen, A. Maurya, B. Raman, R. Mehta, R. Kalyanaraman, N. Vankadhara, S. Roy, and P. Sharma. Kyun Queue: a sensor network system to monitor road traffic queues. In *Proc. ACM SenSys'12*, 2012.

[20] R. Sen, B. Raman, and P. Sharma. Horn-ok-please. In *Proc. ACM MobiSys'10*, pages 137–150, 2010.

[21] A. Skordylis and N. Trigoni. Delay-bounded routing in vehicular ad-hoc networks. In *Proc. ACM MobiHoc'08*.

[22] B. Smith and M. Demetsky. Multiple-interval freeway traffic flow forecasting. *Transportation Research Record: Journal of the Transportation Research Board*, (1554):136–141, 1996.

[23] B. L. Smith and M. J. Demetsky. Traffic flow forecasting: comparison of modeling approaches. *Journal of transportation engineering*, 1997.

[24] F. Stulp and O. Sigaud. Many regression algorithms, one unified model: a review. *Neural Networks*, 69:60–79, 2015.

[25] H. Sun, H. X. Liu, H. Xiao, R. R. He, and B. Ran. Short term traffic forecasting using the local linear regression model. In *the 82nd Annual Meeting of the Transportation Research Board, Washington, DC*, 2003.

[26] S. Sun, C. Zhang, and Y. Zhang. Traffic flow forecasting using a spatio-temporal bayesian network predictor. In *Artificial Neural Networks: Formal Models and Their Applications – ICANN'05*, pages 273–278. Springer, 2005.

[27] Z. Sun, S. Wang, and Z. Ma. In-situ implementation and validation of a CO_2-based adaptive demand-controlled ventilation strategy in a multi-zone office building. *Building and Environment*, 46(1):124–133, 2011.

[28] Y. H. Tang and B. Xi. Dynamic forecasting of traffic volume based on quantificational dynamics: a nearness perspective. *Scientific Research and Essays*, 5(4):389–394, 2010.

[29] The bright side of sitting in traffic: Crowdsourcing road congestion data. https://goo.gl/UfZ45j, 2016.

[30] R. Tibshirani. Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[31] Transport Advisory Committee. Report on Study of Road Traffic Congestion in Hong Kong. http://www.thb.gov.hk/eng/boards/transport/land/Full_Eng_C_cover.pdf, 2014.

[32] M. Van Der Voort, M. Dougherty, and S. Watson. Combining Kohonen maps with ARIMA time series models to forecast traffic flow. *Transportation Research Part C: Emerging Technologies*, 4(5):307–318, 1996.

[33] B. M. Williams and L. Hoel. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: theoretical basis and empirical results. *Journal of Transportation Engineering*, 129(6):664–672, 2003.

[34] H. Yin, S. Wong, J. Xu, and C. Wong. Urban traffic flow prediction using a fuzzy-neural approach. *Transportation Research Part C: Emerging Technologies*, 10(2):85–98, 2002.

[35] J. Yoon, B. Noble, and M. Liu. Surface street traffic estimation. In *Proc. ACM MobiSys'07*, pages 220–232, 2007.

[36] J. Yuan, Y. Zheng, X. Xie, and G. Sun. T-drive: enhancing driving directions with taxi drivers' intelligence. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 25(1):220–232, 2013.

[37] D. Zhang, J. Zhao, F. Zhang, and T. He. Urbancps: a cyber-physical system based on multi-source big infrastructure data for heterogeneous model integration. In *Proc. the ACM/IEEE 6th International Conference on Cyber-Physical Systems (ICCPS'15)*, pages 238–247, 2015.

[38] 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 - technical report. https://goo.gl/c71A7W, 2016.

[39] T. Zhou, L. Gao, and D. Ni. Road traffic prediction by incorporating online information. In *Proc. Springer WWW'14*, pages 1235–1240, 2014.

[40] H. Zou. The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476), 2006.