Extending Coverage of Stationary Sensing Systems with Mobile Sensing Systems for Human Mobility Modeling

YU YANG, Rutgers University
ZHIHAN FANG, Rutgers University
XIAOYANG XIE, Rutgers University
FAN ZHANG, Shenzhen Institute of Advanced Technology
YUNHUAI LIU, Peking University
DESHENG ZHANG, Rutgers University

Human mobility modeling has many applications in location-based services, mobile networking, city management, and epidemiology. Previous sensing approaches for human mobility are mainly categorized into two types: stationary sensing systems (e.g., surveillance cameras and toll booths) and mobile sensing systems (e.g., smartphone apps and vehicle tracking devices). However, stationary sensing systems only provide mobility information of human in limited coverage (e.g., camera-equipped roads) and mobile sensing systems only capture a limited number of people (e.g., people using a particular smartphone app). In this work, we design a novel system Mohen to model human mobility with a heterogeneous sensing system. The key novelty of Mohen is to fundamentally extend the sensing coverage of a large-scale stationary sensing system with a small-scale sensing system. Based on the evaluation on data from real-world urban sensing systems, our system outperforms them by 35% and achieves a competitive result to an Oracle method.

CCS Concepts: • **Networks** → **Sensing Systems**; • **Information systems** → *Location based services.*

Additional Key Words and Phrases: Sensor networks, Human mobility, Heterogeneous sensing systems

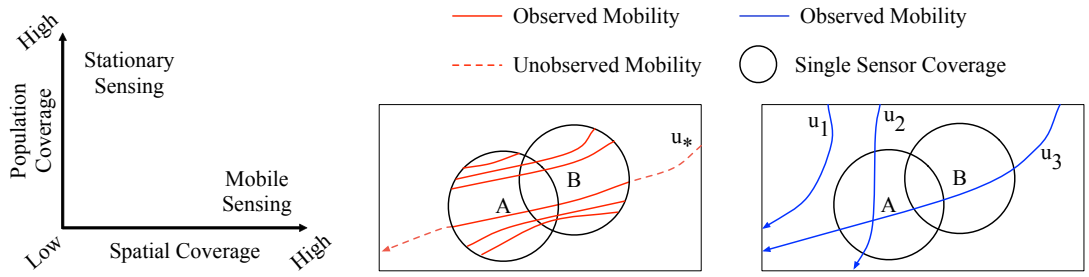
ACM Reference Format:
Yu Yang, Zhihan Fang, Xiaoyang Xie, Fan Zhang, Yunhuai Liu, and Desheng Zhang. 2020. Extending Coverage of Stationary Sensing Systems with Mobile Sensing Systems for Human Mobility Modeling. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 3, Article 100 (September 2020), 21 pages. https://doi.org/10.1145/3411827

1 INTRODUCTION

Human mobility modeling such as characterizing human travel distances and predicting future locations is of great importance for a wide variety of urban applications, ranging from location-based services and urban planning, to public health and safety monitoring and personalized recommendation systems. For example, location-based business companies such as Yelp and Foursquare can advertise potential restaurants or interests in the next locations of users [45]; With the real-world benefits from these applications of human mobility, it has been an important topic to utilize sensing systems to obtain human mobility data.

Authors' addresses: Yu Yang, Rutgers University, yu.yang@rutgers.edu; Zhihan Fang, Rutgers University, zhihan.fang@cs.rutgers.edu; Xiaoyang Xie, Rutgers University, xx88@cs.rutgers.edu; Fan Zhang, Shenzhen Institute of Advanced Technology, zhangfan@siat.ac.cn; Yunhuai Liu, Peking University, yunhuai.liu@pku.edu.cn; Desheng Zhang, Rutgers University, desheng.zhang@cs.rutgers.edu.

2474-9567/2020/9-ART100 $15.00
https://doi.org/10.1145/3411827



(a) Strength and Weakness (b) Stationary Sensing System (c) Mobile Sensing System

Fig. 1. Sensing System Comparison

Based on different types of sensing infrastructures, prior works can be classified into two categories including (i) stationary sensing systems such as cell towers [23] [14] [12], surveillance cameras [42], Automatic Fare Collection (AFC) devices deployed for urban transit (e.g., subways [13] [15], buses [58], and highways [49]), and (ii) mobile sensing systems such as GPS information from smartphone apps [56] and vehicle on-board diagnostics(OBD) devices [52]. We plot the strength and weakness of two types of sensing systems in Fig. 1(a) based on two criteria: population coverage and spatial coverage since temporal coverage is fixed given population and spatial coverage. In general, stationary sensing systems have better population coverage. For example, Fig. 1(b) shows a simple sensing system consisting of two surveillance cameras (i.e., A and B); Fig. 1(c) shows the mobility of 3 residents driving a car captured by an insurance company's tracking system as the mobile sensing system in the same area, i.e., GPS uploading. Based on these two examples, we found (1) the two cameras as an example of stationary sensing can sense the mobility of almost all the nearby residents, but they only have limited spatial sensing range as a combined coverage of two cameras; (2) the insurance tracking system as an example of mobile sensing can sense the complete mobility traces of participating customers, but the participating customers are limited. Considering these complementary sensing features for human mobility, we focus on stationary sensing and explore mobility sensing as a complementary modality, since the stationary sensing is more ubiquitous and pervasive given the recent development of urban infrastructure and urban Internet-of-Things. In the following sections, we refer to human mobility modeling as predicting next locations.

Many works have been done to model human mobility at the sensor-location level such as predicting the next location at the cell tower level [11] or station level [39]. However, they are limited by sensors' spatial coverage that cannot model finer-level mobility such as locations between sensors due to resolution limitation. To address the weakness, some works were conducted to improve the spatial coverage in an interpolation manner. For example, *Yuan et al.* [51] reconstructed missing records from smart card transactions. Some works implemented map-matching algorithms to complete the missing trajectories from low-sampling-rate GPS [30] [32]. With these approaches, the sensing coverage of single sensors is extended to a certain area. The area could be on a small scale such as a campus or a city district with surveillance cameras, or on a large scale such a whole city with cell towers. However, it is not well studied about modeling human mobility after people leave these covered areas. For example, shown as in Fig. 1(a), most of the works model human mobility (i.e., sold lines) in the coverage of sensor A and B (e.g., such as cameras), but are not able model it outside (i.e., such as the dashed lines).

In this work, we aim to extend the sensing coverage of stationary sensing systems with complementary characteristics from mobile sensing systems. Specifically, we aim to predict people's future locations when these locations are not detected by stationary sensing systems. Our key idea is that human moves in their own patterns and are generally independent to the types of sensing systems they are exposed [34]. It indicates that human mobility at the individual level, as a behavior, is inherently consistent in two systems [13]. Based on this intuition, we study the inherent similarity of individual mobility in different systems to extend the sensing coverage in

stationary sensing systems. For example, a trace u_* is observed in a stationary sensing system in Fig. 1(b) and three traces u_1, u_2, u_3 are observed in a mobile sensing system in Fig. 1(c). By studying the similarity between them, we found u_* is most similar to u_3 since they are both sensed by A and B, which is obtained by stationary sensing alone. In this sense, u_* may have similar traces or locations as u_3 after leaving the joint sensing coverage of A and B.

To verify our idea, we conduct a field study based on two concrete sensing systems: a smart toll payment systems (i.e., stationary sensing system) and a vehicle tracking system for insurance (i.e., mobile sensing system). We design a system *Mohen* that extends the sensing coverage of the stationary sensing system with small population (i.e., compared to people captured in stationary sensing systems) sensed by the mobile sensing system. Specifically, we utilize those population with detailed traces from the vehicle tracking system to predict others' final location without traces when they are out of the coverage of the smart toll payment systems. We first formulate the mobility as a spatiotemporal sequence where each node is the encoded representation of a location with a timestamp. Then we learn the representation of the mobility sequence based on sequence embedding, which is formulated based on the relationship between two kinds of sensing systems. Based on the relationship and sequence embedding, we finally predict the future locations of all the population after leaving the coverage of the stationary sensing system. Compared to prior works, the unique advantage of *Mohen* is it can continuously model human mobility even when they are out of the coverage of a sensing system. In addition, we do not require deployed mobile sensing systems in most people, which makes our system more usable in a more general setting. Our key contributions are as follows.

- *Investigating the problem of extending stationary sensing systems for human mobility modeling:* We investigate a novel problem of extending sensing coverage of stationary sensing systems, and utilize human mobility prediction as an concrete example for system design, implementation, and evaluation.
- *Designing a novel sensing system:* Technically, we design a system called *Mohen* based on mobility sequence learning. Sequence embedding is applied to mobility sequences to reserve the relationship between different users' mobility behaviors. Based on the sequences from similar vehicles, we train a customized LSTM network to predict human mobility within and outside of system sensing coverage.
- *Evaluating the system with real-world datasets:* We evaluate *Mohen* based on data from two real world sensing systems: (i) a stationary sensing system: a smart toll payment system with 2 million daily users; (ii) a mobile sensing system: a vehicle tracking system with 114 thousand vehicles. Compared with two state-of-the-art methods, our system outperforms them by 35% and achieves a competitive result to an Oracle method.
- *Demonstrating the utility of mobility prediction in a real-world application:* To show the benefits, we design a measurement approach to optimize urban density control strategies in a controlled urban environment. With careful optimization, we could improve the mobility flow by 36% on average in terms of speed.

2 SYSTEM OVERVIEW

We introduce the overview of this work in Fig. 2 by three components including two physical sensing systems, a cloud component, and a application component.
(i) Our physical sensing systems consist of a smart toll payment system as a stationary sensing system and a vehicle tracking system as a mobile sensing system. Two systems upload their sensing data into the cloud and historical data are stored in the cloud data depository. The details are introduced in Section 3.
(ii) Based on the historical data, we design our model with two main module: an embedding and clustering module which study the similarity of mobility between two systems; and a mobility prediction module that extends the mobility sensing coverage of the stationary sensing system. The details are introduced in Section 4.

(ii) In real time, when the stationary sensing system uploads new data, we apply designed model to predict the mobility, which are useful for various applications such as locations based services and urban planning. In Section 6, we design a novel system for urban density control.

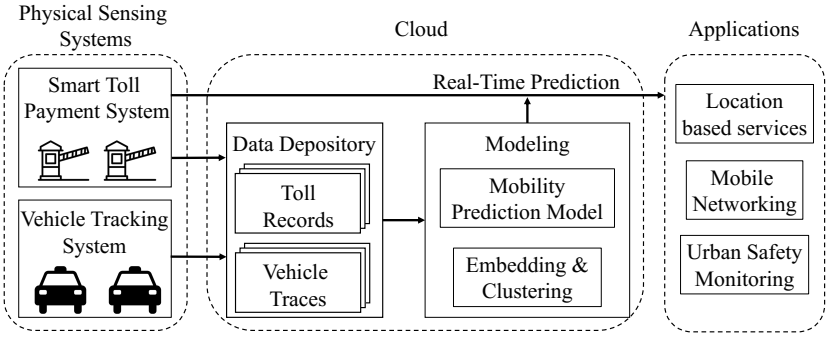


Fig. 2. System Overview

Problem Definition. *Given a set of vehicles V, $N \subseteq V$ are captured by a stationary sensing system (i.e., the smart toll payment system) and $M \subseteq V$ are captured by a mobile sensing system (i.e., the vehicle tracking system) ($|N| >> |M|$), our objective is to infer the final locations of these vehicles in $N - N \cap M$ (i.e., in N but not M) after they leave the coverage of the stationary sensing system.*

3 PHYSICAL SENSING SYSTEMS AND OBSERVATIONS

3.1 Smart Toll Payment System

In our work, we utilize the smart toll payment system (ETC) in a Chinese city (shown in Fig. 3). The red dots are 71 toll booths spreading in the city. Performed as a stationary sensing system, all the vehicles passing booths would be detected with information collected shown in Table 1 as an example. The data available to us consists of approximately 30 million made by vehicles exiting to the city during January 2016. The dataset is pseudonymized and contains no explicit personally identifiable information. We plot the volume from each toll booth in Fig. 4 sorted by the exiting volumes. We observed an unbalanced entering and exiting volumes in the same toll booths. Fig. 5 plots the total volumes of toll booths in a day. Two volume peaks are observed from 8 am to 10 am and 3 pm to 6 pm.

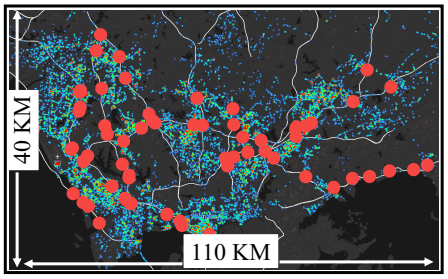


Fig. 3. Toll Payment System and Vehicle Tracking System

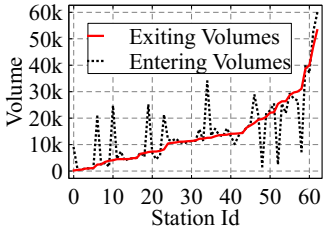


Fig. 4. Volumes by stations

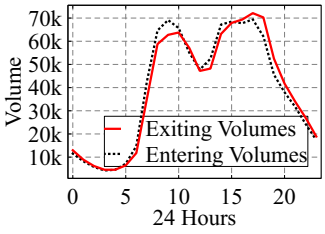


Fig. 5. Volumes by time

Table 1. Toll Transaction Record Format

Name	Value
Entering/Exit Toll Station	Luohu
Entering/Exit DateTime	20160102 13:10:00
Vehicle Id	F14SD1A
Vehicle Type	Car/Bus/Truck
Weights	1300 kg
# Daily Records: 1.1 million	
# Daily Vehicles: 768 thousand	

Table 2. Fleet Data Format

Name	Value
Vehicle Id	460040053409
DateTime	20160601 12:10:00
GPS Coordinates	115.7591, 23.3944
Speed	40km/h
# Daily Records: 200 million	
# Vehicles: 114 thousand	
# Vehicles using highway: 6834	

3.2 Vehicle Tracking System

In our work, we utilize a historical GPS dataset collected by a vehicle tracking system for insurance. Specifically, we utilize OBD devices from an insurance company that are originally deployed for Usage-based Insurance. Each of these vehicles uploads its GPS locations with a frequency of about 10 seconds to a cloud server through an OBD (On-Board Diagnostics) device along with a smartphone app. In total, there are 114 thousand vehicles generating 200 million GPS points on average every day. The detailed format is listed in Table 2. The dataset is also pseudonymized and contains no explicit personally identifiable information. It is clear that not all vehicles travel on highways. We conducted a map matching algorithm [32] on the GPS traces and only keep vehicles with highway traces. The final dataset contains 6834 vehicles in total, whose destinations after highways are plotted as a heatmap in Fig. 3. We observe that those vehicles cover major areas of the city, which shows they are representative in a city-scale study.

3.3 Observations

Opportunity: We show our opportunities about how to establish the connection between two systems for mobility prediction. Our intuition is that two vehicles with similar patterns (i.e., entering/exit locations and time) in the stationary sensing system may also have similar mobility (i.e., future locations such as destinations) afterward. The reason of the intuition is that humans move in their own patterns and are generally independent to the types of sensing systems they are exposed. It indicates that human mobility at the individual level is inherently consistent in two systems [13]. To verify the intuition, we align the vehicles in two systems in three steps: (i) We first apply a map matching algorithm [32] to map the vehicular GPS collected from the vehicle tracking system to the road network, especially the highway road network. (ii) Secondly we extract the time a vehicle passing the toll station on the corresponding road. (iii) Finally, combining both the entering station and exit station, we align the vehicle mobility in the vehicle tracking system to the toll payment system. Given the alignment, we extract the vehicles that are observed in both systems, which is around 10% of all the vehicles in the vehicle tracking system.

Suppose we have two vehicles ve_i and ve_j existing in both data sources. We defined the similarity of the highway mobility $HSim$ between two vehicles as the Jaccard Index:

$$HSim(ve_i, ve_j) = \frac{|S_i \cap S_j|}{|S_i \cup S_j|} \quad DSim(ve_i, ve_j) = \frac{|D_i \cap D_j|}{|D_i \cup D_j|} \tag{1}$$

where S_i is the set of historical passing toll station of the i_{th} vehicle. Similarly, we define the destination similarity as $DSim$ where D_i is the set of historical destinations of the i_{th} vehicle. Fig. 6 plots $HSim$ and $DSim$ of randomly sampled vehicles. Each point represents a comparison between two vehicles and the black line is the fitting function. We found two similarity measurements are positively correlated, which implies the similarity on the highway mobility could potentially imply similar destinations.

Even the similarity relationship in Fig. 6 is not strictly linear, it does not affect the purpose of this observation as our motivation. The key purpose of Fig. 6 is to show there is a positive relationship between the mobility in two systems instead of using it for computation purposes. To utilize this observation in our design, our goal is to extract and represent this relationship more precisely with more sophisticated measurements and techniques.

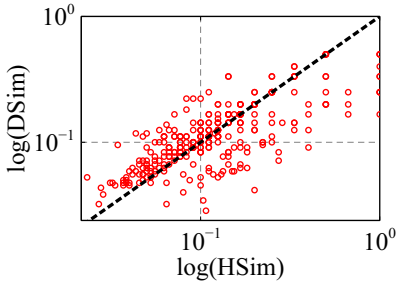


Fig. 6. Opportunities

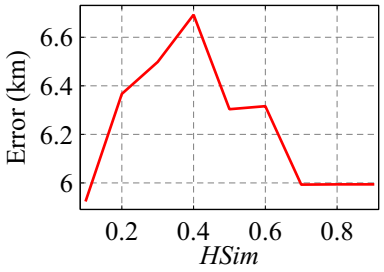


Fig. 7. Challenges

Challenges: Given this opportunity, we implement a base model based on similar vehicles only. Specifically, we select the vehicles with *HSim* greater than a certain threshold to estimate the destination probability based on these vehicles' destinations. We evaluate how different *HSim* impacts the performance in Fig. 7. We found the low and high *HSim* have better performance than the middle *HSim* but all have large errors. Because the simple similarity measurement in Eq. 1 is not enough to quantify the relationship, it motivates us to design more sophisticated measurement methods such as encoding based approaches to extract the hidden relationship more precisely.

4 MODEL DESIGN

4.1 Preliminary

More generally, we abstract mobility into a two-level sequence in Fig. 8. At the in-coverage mobility level, each node represents a toll station that a vehicle enters or leaves. A pair of consecutive nodes (e.g., r_1 and r_2) represents a complete trip on highways. At out-coverage mobility level, each node is a destination of a vehicle, which is the modeling objective. In our study, it is represented as grids of 500-meter width. Combining two levels, we have a complete mobility sequence since entering a highway from a toll station to arriving in a destination (e.g., r_1-r_2-g_1).

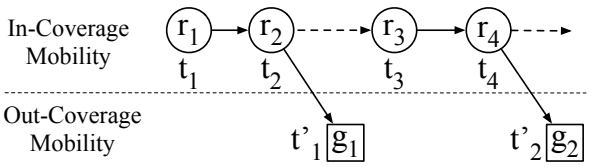


Fig. 8. Mobility Sequence Abstract

4.2 Design Intuition

The key design components in our work are (i) the correlation between in-coverage mobility and out-coverage mobility; (ii) the prediction model. We motivate our design with the following two intuitions.

(i) Correlation between in-coverage mobility and out-coverage mobility: There are a lot of works been done to predict the destinations of vehicles. A practical reason for promising performance is that they generally have plenty of labeled historical data to train the destination prediction model or learn the destination probability. In our scenario, there are only a few vehicles equipped with traceable GPS devices. For more general vehicles without GPS devices, there is no explicit way to obtain their destinations. This motivates us to extract more information by studying the correlation between two levels of mobility to enrich the missing labeled data. We introduce a detailed design in Section 4.3.

(ii) Prediction Model: Further, existing approaches can be categorized as different types of Markov Models. A theoretical assumption behind these approaches is Markov Property that future states depend only on the current state, not on the events that occurred before it. However, human mobility can be periodical that the future destinations could be related to those in a long time ago [29], which cannot be captured under the Markov Property assumption. The recent development of sequence models and deep learning gives a new opportunity to model mobility. Specifically, Long Short-Term Memory (LSTM) networks have been proved as sophisticated models to learn sequential data in many fields, such as natural language processing [20] and video sequential learning [18]. A key advantage of LSTM networks is that they are capable of learning both long-term dependencies and short-term dependencies. In our problem, these two dependencies can be utilized to explain the factor of long-term periodical mobility and short-term none-periodical mobility, which motivates us to model the problem as a sequence learning problem using LSTM networks. We introduce a detailed design in Section 4.4.

4.3 Embedding Learning and Clustering

To establish the connection between vehicles with and without labeled destinations, we utilize the commonly available data from the ETC systems to measure the similarity. Specifically, we aim to learn the similarity of vehicles' destinations only based on in-coverage mobility sequences. To achieve this objective, we design an embedding method to learn the representation of in-coverage mobility sequences that: (i) the similarity can be directly measured; (ii) it is approximated to the destination similarity.

Background: Embedding based methods have been applied in many machine learning tasks such as representation learning. The basic idea is to learn an encoder that maps the data from the original space into the embedding space and reserve the similarity of data in both spaces. There are generally three steps: (i) Define an encoder; (ii) Define the similarity function in the original space; (iii) Define the loss function and optimize the parameters. We illustrate how we adopt the embedding method in our work step by step as follows.

- **Encoder:** The purpose of an encoder is to map the original data into a vector representation. In our problem, the original data is the in-coverage mobility sequence, which can be fed into a LSTM based model for encoding. Fig. 9 illustrates the basic procedure, where r is each station node in the sequence. Due to the limitation of spaces, we omit the mathematical formulas in each LSTM unit, which could be found in [41]. We define this procedure as
$$V_i = Encoder(R_i) \tag{2}$$
where R_i is the input of i_{th} in-coverage mobility sequence and V_i is the encoding vector.

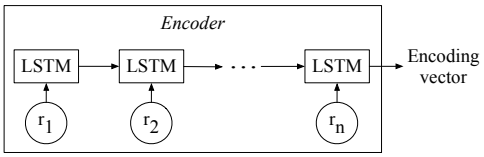


Fig. 9. Encoding Procedure

- **Similarity function:** The similarity function is generally chosen depending on the concrete problem. In our problem, we assume two in-coverage mobility sequences should have similar encoding vectors if they have the same destination. In an in-coverage mobility sequence of length n, there are generally $n/2$ destinations that each of them follows a pair of station nodes. We utilize the Jaccard index of the destinations as the similarity metric. Formally, it is defined as
$$Sim(R_i, R_j) = \frac{|Des_i \cap Des_j|}{|Des_i \cup Des_j|} \tag{3}$$

where Des_i is the set of destinations in the i_{th} sequence; $|\cdot|$ stands for the total number of the entries.

- **Loss function:** Suppose there are n in-coverage mobility sequences, we define our loss function as

$$\mathcal{L}_{encoding} = \sum_{i,j \in n} ||Sim(R_i, R_j) - V_i^T V_j||^2 \tag{4}$$

where $V_i^T V_j$ is the dot product of two encoding vectors. We minimize $\mathcal{L}_{encoding}$ to learn the encoder utilizing the stochastic gradient descent algorithm [2].

Clustering: After we learn the encoder and the encoding vectors, we apply the K-Means clustering algorithm on the encoding vectors of all the sequences. Benefited from the objective function of embedding, we can directly measure the similarity utilizing the dot product on the encoding vectors even we do not know the detailed destinations of the in-coverage mobility sequence. The result of each cluster represents a specific group of sequences with similar destinations.

Why clustering based on embedding instead of the similarity function directly: In our setting, our target is all the vehicles detected by the stationary system, most of which are without GPS trajectories. In this sense, we are not able to directly measure similarity or do clustering on them based on the similarity function, which is based on the destinations from GPS trajectories. In our method, we apply representation learning on the data from the stationary sensing system and make the learned representation consistent with the destination similarity In this way, we could compute the similarity on all the targeted vehicles since the representation of all can be obtained by encoding the data from the stationary sensing system.

In summary, we learn an encoder to measure the similarity between sequences. Similar sequences are clustered together as inputs to train the mobility prediction model.

4.4 Mobility Prediction

Given clusters of the in-coverage mobility sequences, we first learn the representation of each station node. Then we design a context-based LSTM network in each cluster to predict destinations.

4.4.1 SpatioTemporal Representation. To fit in-coverage mobility sequences into any learning network, the first task is to learn the representation of each node. A naive approach is to apply one-hot encoding. For example, given N possible nodes, the one-hot encoding of a node is a vector of length N where only one entry is one while other entries are zeros. Different nodes have ones in different entries. However, these are two issues of one-hot encoding: (i) It does not reserve any relationship between nodes. (ii) It does not provide any temporal information.

To address these issues, we design a spatiotemporal embedding method motivated by the co-occurrence model [35] in the field of natural language processing. Basically, it utilizes the co-occurrence of words to learn the vector representation of each word. We illustrate how we adopt the method in our scenario as follows.

- **Corpus vocabulary:** In the in-coverage mobility sequence, each node is correlated with spatial and temporal information. We represent each node as a tuple whose first and second entry is the location id and temporal id, respectively. For example, the first node in Fig. 8 is represented as (r_1, t_1). Those tuples are considered as vocabularies in the language model.
- **Co-occurrence generating:** The next step is to define co-occurrence. In the original design, similar words tend to occur together and will have a similar context. In our scenario, since our objective is to predict destinations, we consider the previously passing node as similar nodes if their destinations are the same. For example, suppose there is a sequence $r_1 - r_3 - g_1$, $r_2 - r_3 - g_1$ (we omit the temporal id to make it clear). We generate three co-occurrences (r_1, r_3), (r_2, r_3) and (r_1, r_2) since they all leads to the same destination. Given all the co-occurrences of N distinct nodes, we have a matrix O with size $N \times N$.

- **Learning:** To learn the representation of each node, we apply a non-negative matrix factorization (NMF) algorithm. In particular, given the co-occurrence matrix O, we aim to learn two matrices W and H, where matrix W denotes the latent representation of nodes, and matrix H denotes the interaction patterns. We learn these two matrices by minimizing the following loss function:

$$\mathcal{L}_{ST} = ||O - WH^T||_F^2 + \lambda(||W||_F^2 + ||H||_F^2) \tag{5}$$

where $||\cdot||_F$ denotes the Frobenius norm, $\lambda(||W||_F^2 + ||H||_F^2)$ is the item to avoid overfitting. We can solve the equation 5 with a multiplicative update algorithm in [27]. The output W is our objective spatiotemporal representation. We simplify the representation learning procedure as

$$w_i = Embed(r_i, t_i). \tag{6}$$

In summary, we learn the spatiotemporal representation of each node by remaining the similarity between nodes, which is defined as the co-occurrence relationship. Further, since the co-occurrence relationship is extracted according to the destination dependency, it is also helpful to support destination prediction in the prediction model.

4.4.2 cLSTM Prediction Model. Given the spatiotemporal representation of each node in the mobility sequence, we train a LSTM network with contexts (cLSTM) to predict the destination. We show the structure of the model in Fig. 10, which is a many-to-many LSTM network. Basically, there are two components: *Mobility Unit* and *Context Information* as follows.

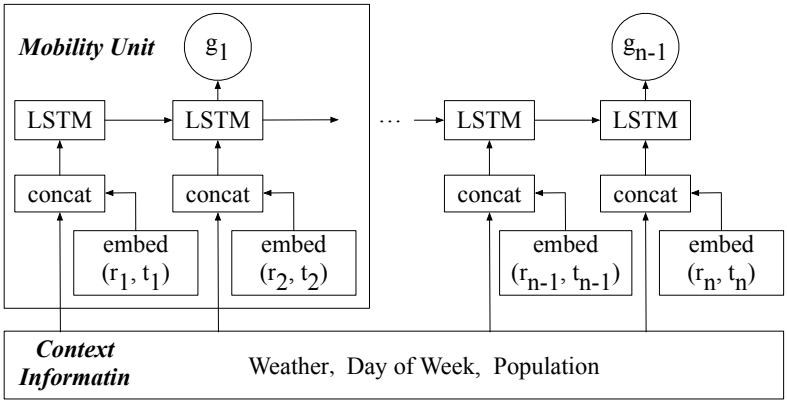


Fig. 10. cLSTM Network Structure

- **Mobility Unit**: A mobility unit is the minimal unit of mobility. In our scenario, it represents the mobility of a vehicle from entering the highway to arriving on the destination. It contains two inputs, each of them is the spatiotemporal representation of (r_i, t_i). The output in the mobility unit is the destination g_i. Our mobility unit is also flexible to other scenarios if there are a various number of inputs. For example, in the setting of surveillance cameras, a vehicle may be captured by the cameras more than twice. We can easily extend the mobility unit with more than two inputs.
- **Context Information**: Since people's mobility is generally affected by many context information such as weather and day of a week, we introduce extra information to improve the design of our LSTM network. The input context information is concatenated with the spatiotemporal representation as inputs of LSTM units. More information can also be feasible to be appended into our model.

Learning: Given the network structure in Fig. 10, we introduce our learning process. Instead of learning a single model for all the vehicles, we learn a separate cLSTM model for each cluster learned in Section 4.3 to model similar mobility. We aim to minimize the error of the destination prediction, which can be formulated as

$$\mathcal{L}_{prediction} = \frac{1}{m}\sum_{i=1}^{m} I(g_i, g_i^{'}) \tag{7}$$

where g_i is the ground truth destination and $g_i^{'}$ is the predicted destination, m is the total number of sequences, I is an indicator function such that $I(g_i, g_i^{'}) = 1$ if $g_i = g_i^{'}$ otherwise $I(g_i, g_i^{'}) = 0$.

4.4.3 Online Prediction. We implement online prediction based on the following three steps.

- Sequence construction: we start the prediction when there is a vehicle detected by a toll station. According to the id of this vehicle, we retrieve its historical ETC records from the cloud. Concatenated with new observations of current locations, we convert them into an in-coverage mobility sequence.
- Clustering Lookup: Given the in-coverage mobility sequence, we utilize the *Encoder* learned in Section 4.3 and calculate the distance of the encoding to each cluster's center. The closest one is our corresponding cluster.
- Prediction: We input the in-coverage mobility sequence into the trained cLSTM to predict the destination.

5 EVALUATION

In this section, we first introduce the implementation and computing complexity analysis of our system. Then we show the metrics and baselines for the evaluation followed by a discussion of the performance under different contexts. Finally, we perform a case study of measuring local traffic pressure in the city.

5.1 Implementation

Our evaluation is based on two sets of real-world data in Chinese province Guangdong, of which the detailed format is described in Section 3.

(1) In our experiment setting, we randomly split the vehicles into two parts: the training part (80%) and the test part (20%). In the training part, we use both the toll payment records and trajectories of 80% vehicles.
(2) We first learn the encoder with all the training data. Then we cluster the training data into different clusters. In the destination prediction, the prediction model is trained by the sequences in each cluster. Then we apply the encoder on the testing data and find the corresponding cluster. Finally, we utilize the corresponding model to predict the destination.
(3) In the test period, we only use the toll payment records to predict the destinations, which are tested against the ground truth extracted from their trajectories. In particular, the model is not aware of trajectories from the tested vehicles and the only input of the model is the toll payment records.

The key advantage of the setting is in its realizability that it is generally very difficult for a system to obtain the trajectories of all the vehicles. However, it is more realistic and reasonable for a system (i.e., highway system) to obtain the toll payment records of all the vehicles and trajectories of a small portion of vehicles (i.e., open data or taxis). Given this setting, we aim to predict the destinations of more general vehicles without trajectories in the system.

5.2 Computing Efficiency

We implement *Mohen* on a server with Intel Xeon E5-1660 3.00GHz CPU, 32GB RAM, and Nvidia Tesla K40c. After the offline processing and training, the speed of real-time prediction is 120 times every second on average,

which means 432 thousand per-hour, and higher than the max demand of 300 thousand transactions at peak hours of the ETC system.

5.3 Evaluation Setting

- **Ground Truth:** The ground truth is the destinations of mobility sequences in our testing data, each of which is represented as an equal size grid (i.e., 500 meters × 500 meters).
- **Performance Metrics:** We utilize accuracy as the metric to represent the prediction performance, which is defined as

$$\text{accuracy} = \frac{\#\text{ of prediction}_{\text{correct}}}{\#\text{ of prediction}_{\text{all}}} \times 100\%. \tag{8}$$

- **Comparisons:** The key advantages of our system lie in the heterogeneous data and similarity learning. We compare our model with three baselines including two approaches based on single data sources and a naive fusion of heterogeneous data.
  - **Mobile Sensing Network (MSN)** [22]: Almost all the previous work of the destination prediction can be categorized as different types of Markov models. We choose the state-of-the-art method [22] to predict the vehicles' final destinations. The basic idea is to predict the destinations by two measurements: one is based on the given origin and the other is based on the trajectory. For the first measurement, we adopt the approach to predict the destination based on the vehicles' origin. For the second measurement, we adopt the approach by dividing the trajectories as highway sub-trajectories and local-trajectories. Compare to our work that is not aware of trajectories of the tested vehicle and the only input of the model is the toll payment records, MSN model is the oracle model since the historical trajectories of the tested vehicles are given, which is the same as the conventional destination prediction task. In this sense, we could observe the historical destinations of these vehicles, which provide labeled data to improve the prediction performance.
  - **Stationary Sensing Network (SSN)** [14]: In general, a stationary sensing network alone only cannot be utilized to predict final destination since it lacks the information after vehicles leave the coverage of sensors. In our context, we estimate the probability of the different destinations by the population density in the nearby area of the exit toll station (i.e., the Voronoi centered by the toll station shown as Fig. 3) according to the finding in [14].
  - **Naive Fusion of Heterogeneous Data(Naive-HSN)**: We set this baseline based on a naive fusion of heterogeneous data to show the advantage of our technique design. Specifically, We implement this baseline as a variant of our model by replacing the clustering module designed in Section 4.3 with a straightforward strategy. Specifically, we design the similarity according to *SSIM* defined in Section 3.3. The more common entering and exit stations they have, the more similar they are. Based on this similarity measurement, we cluster similar sequences and following the same procedure in our model to predict the destinations.
  - **Modeling with a Different Sequence Representation Component (DPLink*)** [17]: Another baseline we consider is to implement our system with a different technique component. Specifically, we consider the work introduced in [17], which is also a work focusing on modeling heterogeneous mobility data. Because of a different scenario in [17](i.e., link identity in different data sources), we adapt it by using the location and trajectory encoder to replace our sequence representation while remaining the rest components. Considering we do not include POI information in our method, to make it a fair comparison, we only encode time and location in the baseline implementation.
- **Impacts of factors:** Several important factors are discussed, including location and time factors, vehicle types, and modeling parameters.

- *Location and time factors:* We evaluate how *Mohen* performs in different locations (i.e., downtown and suburb) and time (i.e., weekday and weekend).
- *Vehicle types:* As shown in Table 1, there are three types of vehicles including cars, buses, and trucks. We study the performance of *Mohen* on different types.
- *Modeling parameters:* There are several sensitive parameters in our model design. We evaluate how different modeling parameters impact the performance including the historical data size of stationary sequences, the number of clustering, and the size of destination grids.
- *Highway length:* We further evaluate how the highway length (i.e., travel distance on highways) affects the performance.

5.4 Evaluation Results

Comparison to MSN and SSN: Fig. 11 plots the result of the destination prediction of *Mohen* and the baselines MSN and SSN. The X-axis is the time of a day, and Y-axis is the accuracy of every hour. Based on the result, *Mohen* achieves the accuracy of 72% on average, which is much higher than 37% of the baseline *SSN*. More importantly, we found the performance of *Mohen* is close to the Oracle baseline *MSN* with an accuracy of 82% on average. This implies even without the destinations of the testing vehicles, we can still achieve a certain level of accuracy. This finding testifies a great opportunity for mobility modeling that largely relaxes the assumption that vehicles have to be equipped with GPS devices. We further introduce a frequently used metric, i.e., Acc@topN [44], which

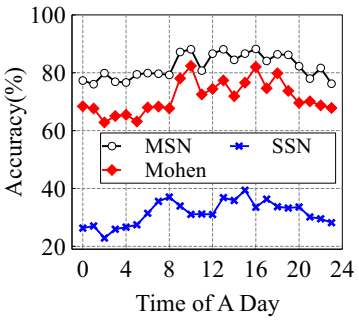


Fig. 11. Baseline Comparison

Table 3. Acc@topN Comparison

Baseline	Acc@top1	Acc@top5
MSN	35%	55%
SSN	82%	90%
Mohen	72%	88%

indicates the accuracy based on candidates with the top N probability. If the predicted destination is in the top N candidates, we consider it as a correct prediction. Then we calculate the accuracy based on the same accuracy equation (i.e., Eq. 8). The result is presented in Table 3. We found Acc@top5 is significantly improved when we relax the top-candidate selection to be top 5. For example, our model *Mohen* improves by 16% and is much closer to our oracle model SSN. The main reason is that people generally only move in a limited time and distance after leaving highways or toll stations, which constrains the possible destination candidates. In this sense, increasing the number of candidates in Acc@topN would greatly improve the accuracy.

We also conduct a statistical hypothesis test to validate the significance of the performance gains. Specifically, we select paired t-test to check if the difference in the mean accuracy between the two models is statistically significant. In our setting, according to the empirical practice [6], we conduct 5 times 2-fold test under the significance level of 0.05. The calculated critical value of the MSN-Mohen pair is 2.093 and that of the SSN-Mohen pair is 2.132. Both values are greater than the standard value 2.0150, which validates the significance of the performance gain.

Comparison to Naive-HSN: As we mentioned, a key component in *Mohen* is to find similar sequences to cluster. We show the advantage of our design by comparing it with a naive fusion of heterogeneous data. Fig. 12

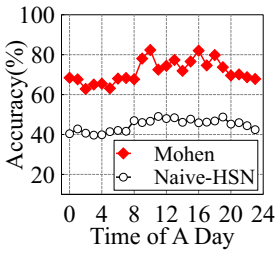


Fig. 12. Naive Fusion

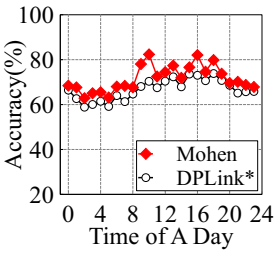


Fig. 13. Different Component

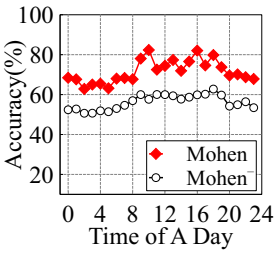


Fig. 14. Variant without Integration

illustrates the performance of *Mohen* compared to Naive-HSN. It shows *Mohen* has much better performance than Naive-HSN (only 45% on average). Even the direct measurement of sequence similarity can qualitatively show the similarity (such as in Fig. 6), it is unable to quantitatively find the correct cluster for a specific objective such as the destination prediction. It is necessary to include objective information (i.e., destination similarity) to learn a better similarity measurement.

Comparison to DPLink*: We compare our method with the baseline DPLink* (shown as in Fig. 13). In general, compared to other baselines, DPLink* has a closer performance as our method. Specifically, the average accuracy of DPLink* is 65% compared to the 72% in *Mohen*. The main reason is that both methods utilize mobility sequences (or trajectories) encoding to extract the relationship between sequences that provides positive effect on the performance. The difference happens in the morning peak (i.e., 9 am and 10 am) and the afternoon peak (i.e., 4 pm and 6 pm) that *Mohen* performs better than DPLink*, which shows the strength of our design when the amount of data is large.

Comparison to a variant without Integration (*Mohen*$^-$): A key advantage of *Mohen* is in its integration of heterogeneous systems. To show its advantage, we further design a variant without integration. In our model, two data sources have their unique functionality that trajectories from the mobile sensing system provide information about destination and toll records from the toll payment system provides rich historical information to measure the similarity of vehicles. In our design, instead of extracting in-coverage mobility sequences from the toll payment system, we utilize the trajectories alone to represent it. Specifically, we map the trajectories on the highway network and extract the entering and exit stations. Then this information is fed into our model in the same manner as our toll records. In this way, the variant model only utilizes the mobile sensing system alone without integration. We show the result in Fig. 14. On average, *Mohen* improves the accuracy by 16% compared to the accuracy of 56% in *Mohen*$^-$. The main reason is that trajectories alone from the mobile sensing system are not representative enough as system integration, which mainly reduces the functionality of embedding learning and clustering.

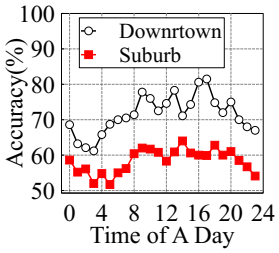


Fig. 15. Loca. Impact

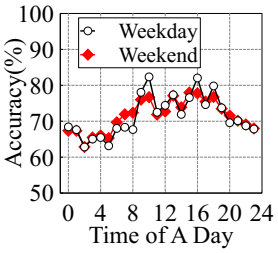


Fig. 16. Time Impact

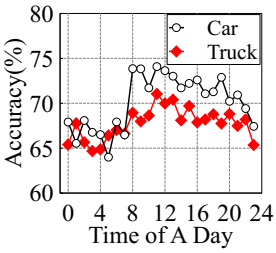


Fig. 17. Vehicle Type

Location and time: Fig. 15 plots the accuracy of *Mohen* in different spatial areas. It shows the accuracy in downtown is better than that in the suburb. We interpret the result as the effect of the clustering component.

Because there are generally more vehicles in the downtown than in the suburb, it gives the opportunity to find more similar vehicles to form the cluster, which improves the performance. Fig. 15 further demonstrates other factors hidden in the downtown and suburban areas. In general, the downtown area has a more complex highway network than the suburban areas because of its denser exits to different areas in the city. We found even the highway network is complex, the performance in downtown areas is still much better than in suburban areas. The main reason is that the destination prediction is constrained by the location of the exit toll station. People generally exit from the station that is nearest to the destination, which constrains the number of destination candidates. Fig. 16 plots the performance of *Mohen* on weekday and weekend, which does not show much difference. This is because we include the context information in the prediction model that helps to capture mobility on both weekdays and weekends.

Vehicle types: There are three types of vehicles including car, truck, and bus. We did not consider the bus in the evaluation because they have defined mobility patterns that can be directly obtained from scheduling centers. Fig. 17 shows the performance of *Mohen* on different types of vehicles. We found in the early morning, the performance on two types is similar. In the day time, the performance of cars is better than trucks. This is because cars are generally personal vehicles that have routine patterns. Trucks can change their destinations depending on different business purposes.

Historical data size: Fig. 18 reveals how the size of the historical data influences the performance. The x-axis is the number of days, which is the unit of the size. As the number of days increases, the accuracy also increases. In particular, the increasing speed is very low when the number of days is less than one week. The performance becomes stable when the days are more than 14 days. It implies that two-week data is potentially representative to reflect the regular mobility of vehicles.

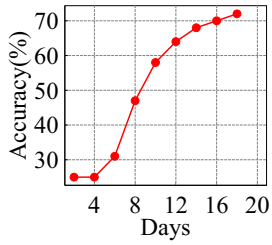


Fig. 18. Historical Data

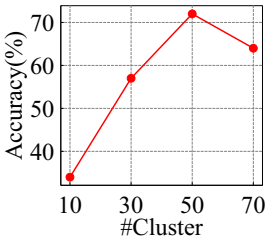


Fig. 19. #Clusters

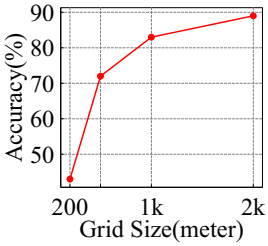


Fig. 20. Grid Size

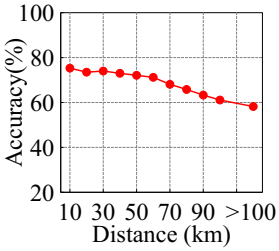


Fig. 21. Distance

Number of clusters: Our prediction models work independently on different clusters. We show the performance changes with the different sizes of clusters. Fig. 19 shows the accuracy increases as the number of clusters increases until it is equal to 50. Then the accuracy decreases when it is equal to 70. We interpret it as the balance between bias and variance of data. A larger number of clusters generally help to differentiate the dissimilar sequences, which reduce the bias of data. As the number is larger than a certain number, there is only a small amount of data in the cluster, which introduces more variance.

Size of destination grids: The performance of *Mohen* is strongly related to the size of the destination grids. Fig. 20 plots the sensitivity. In our experiment, the destination area (i.e., grid size) is manually defined according to different application requirements. Shown as in Fig. 20, in general, a large grid size means a smaller total number of destination candidates, which in turn benefits the prediction accuracy. We found even the grid size is two thousand meters, the prediction is still under 90%, which is lower than the human mobility prediction limit (shown as 93% in [40]). This demonstrates a certain number of vehicles with very irregular mobility patterns, which requires a more sophisticated model to capture their mobility.

Highway length: We further evaluate how the highway length affects the performance, which in our case implies the different travel distances on highways. Shown as in Fig. 21, in general, we found the longer travel

distance makes the prediction more challenging that leads to the performance decrease in a longer distance. Specifically, the performance is relatively stable when the distance is shorter than 70 kilometers but decreases more significantly when it is much longer. The main reason is that most people generally travel within a short distance, which makes long-distance travels less representative of all the mobility. To the best of our knowledge, it is still an open yet interesting problem to model long-distance low-frequency mobility.

6 APPLICATION

Setting: To show the benefits of our model, we conduct a field study in a Chinese city to measure the impact of closing exit ramps. We choose two exit ramps (shown as Fig. 22) near to the downtown areas as our measurement objects, where generally suffer from congested traffic. Basically, there are two peak hours in one day. The morning peak hour is 8:30 am (region A) or 9 am (region B) and the evening peak hour is 6 pm. At the peak hour (i.e., half an hour), there is more than one thousand vehicle exiting highways (i.e., vehicles passing the toll station every two seconds on average).

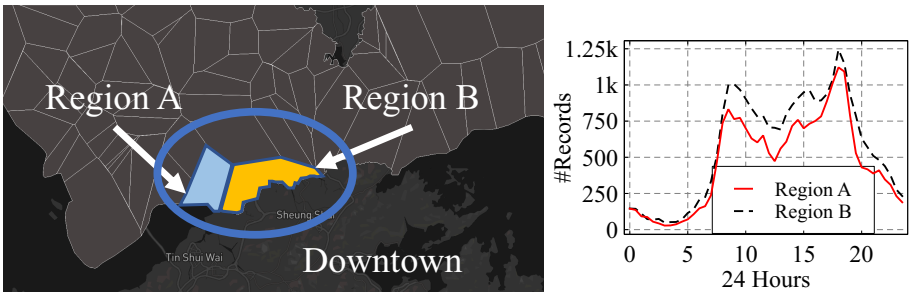
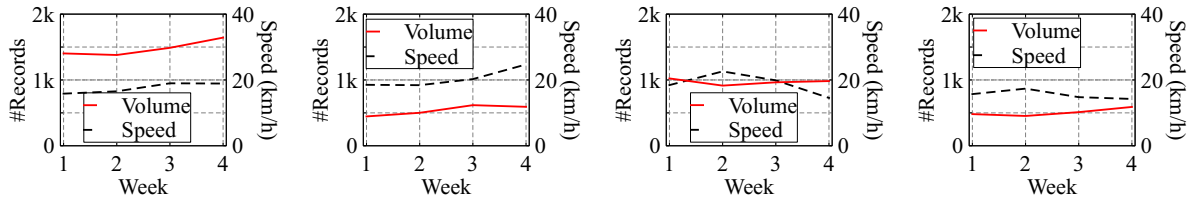


Fig. 22. Setting: we partition the whole city into regions by a Voronoi diagram, where each region has one and only one toll station. Especially, region A and B are located in the downtown area. The plot shows the traffic volume in a day.

Measurement: Formally, we can present our measurement as the equation $Y = f(X, t)$ where X is the number vehicles from highways located in a grid at the time interval t, and Y is the traffic condition (i.e., mean driving speed) in the grid. To learn this function, we first specify how we obtain the input X at time t and the output Y. For each vehicle leaving highways in our system, we predict its destination. Based on the work [25], we infer their passing grids and possible arriving time. By aggregating the result of all the vehicles in a time interval (i.e., 8:00 am to 8:10 am), we know the number of vehicles (those from highways) located in each grid, which is our input X and t. The output Y is obtained from the taxi dataset in the city. Given their driving speed, we can easily obtain the average driving speed in each grid. Based on the input and output, we apply a machine learning method to learn the function f.

However, the traffic condition in each grid is also impacted by vehicles not from highways, which are not observable from our system. To simplify our setting, we assume the vehicles not from highways are stable at different times of a day that we can model the impact by time without explicitly counting the number of none-highway vehicles. To further support our assumption, we study the local traffic condition at different times of the day. Our intuition is that if both the X and Y are stable, then the number of local vehicles is also stable. We show our observations in Fig. 23 considering four timestamps (i.e., 8 am, 12 pm, 4 pm, 8 pm) of the same weekday within a month. Highway and speed represents the number of vehicles leaving highway and speed in the grid, respectively. We found both X and Y are stable that X and Y have a standard deviation less than 2% and 6% of the mean value, which supports our assumption. Finally, we utilize a linear regression model to learn the function, which leads to an accuracy of 88%.



(a) 8 am (b) 12 pm (c) 4 pm (d) 8 pm

Fig. 23. Relationship between X (i.e., vehicles leaving highways) and Y (i.e., average speed in grids).

Strategy: Based on the measurement approach, we design the exit closing strategy. There are three elements to be adjusted in the strategy including: what exit to close; when to close; how long to close. We present a strategy candidate as $S(e, w, l)$, where e, w, l are the adjustable elements, respectively. Specially, w and l are discretized values (i.e., every 15 minutes). Given our measurement approach, we could potentially test the impacts by each possible strategy and select the one resulting in the desired effect. Considering different local traffics may require different effects, we only show the effects of two different strategies as an example including (i) close the exit in region A at 6 pm for 1 hour (ii) close the exit in region A at 6 pm for 2 hours. Fig. 24 shows the estimated average traffic speed in each grid after applying two strategies compared to no strategy is applied. We found the traffic congestion is not evenly distributed in the region if no strategy is applied. Some grids have heavy traffic while others have lighter traffic. Both strategies help reduce the traffic in the heavy traffic area (i.e., increased box low bounds). Strategy II makes the traffic even better since it closes the exit ramp longer and distributes more traffic to other regions. It is worthwhile to mention that the transportation agency could design different strategies to meet the requirements under different conditions. Our system performs as a tool to measure the possible impacts of the strategies.

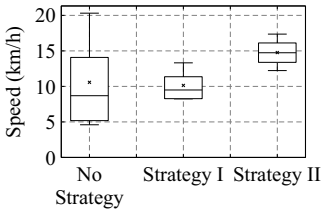


Fig. 24. Performance

7 LESSONS LEARNED AND DISCUSSIONS

Lessons Learned: In our work, we measure different characteristics of vehicular mobility and address the post-highway vehicular mobility problem. We summarize the most important lessons learned in this work.

- For modeling on heterogeneous data, the key is to find the overlapping on different data. In our study, the overlapping is that all the vehicles pass the toll stations no matter they have GPS traces or not. Based on the overlapping, we can establish the connection between them, which is the base of utilizing complementary characteristics. Shown in the evaluation, the single data source can only achieve 37% of accuracy, which testifies the advantage of heterogeneous data.
- When the direct measurement of objective similarity is not available (i.e., destination similarity between vehicles with and without GPS traces), we can take indirect measurements to approximate the objective similarity. In our study, we achieve this by learning an encoder to connect the on-highway sequence similarity and destination similarity. Compared with the naive fusion of heterogeneous data, our designed similarity learning component helps improve the accuracy by 27%.
- Instead of training a single model to predict destinations of all the vehicles (shown as the result of naive prediction in Fig. 12), a more feasible approach is to train separate models for different groups of vehicles.
- As shown in Fig. 11, our model has closed performance with the Oracle baseline, which proves sequential models are good at mobility modeling if the node representation is well defined or learned.

Comparison to single data sources: In our study, the heterogeneous data is a better solution since we need to take information from ETC systems as an extra input. As shown in the evaluation, our model performs well in majority cases. However, it also shows limitations in some cases such as in the suburb and some wrong choice

of the model parameters. In these case, single data sources may have better performance if there are enough spatiotemporal coverage with existing stations or enough vehicles equipped with GPS devices. For irregular or very unique personal mobility, GPS traces could be a better choice since they can observe these behaviors.

Pervasive Heterogeneous Sensing Systems: We emphasize heterogeneous sensing systems are pervasive in the real world. For example, in urban areas, people are captured by surveillance cameras (i.e., stationary sensing system) and some of their smartphones (i.e., mobile sensing system) send locations to map service providers [42]. On highways, all drivers pay at toll booths (i.e., stationary sensing system) while some vehicles (i.e., sensing sensing system) of them also upload their locations through OBD devices [49]. Even in the futuristic scenarios of autonomous vehicles, they could also communicate with stationary urban infrastructures such as traffic signals (i.e., stationary sensing system) while also uploading its real-time status to the cloud through cellular networks (i.e., mobile sensing system) [28] [47]. We envision our work is tied to these scenarios to extend the sensing coverage of those stationary sensing systems.

Generalization to other human mobility data: In our setting, we use two concrete systems (i.e., a toll payment system and a vehicle tracking system) as the sources of the heterogeneous data. In a more general setting, such as GPS from smartphone localization data, it still can be input into our system since our mobile sensing system is general enough to represent GPS locations from different data sources. The only extra effort is to map collected GPS into road segments to find the corresponding passing toll stations. After extracting the passing toll stations, the rest of the process is exactly the same as our vehicle tracking data.

Limitations: In this work, we only evaluate Mohen based on the ETC systems and GPS traces in one city. The performance in other cities with significantly different features may be different. However, we believe the generic design for mobility modeling can be applied to a wide range of cities with similar mobility features. Further, in our evaluation, we only use the 114 thousand vehicles we have data to study their destinations, which are not at the same scale as the real number of vehicles leaving the highways. However, it is almost impossible to obtain detailed information for all vehicles since not every vehicle has GPS equipped. To partially address this issue, we aim to explore cellphone data in the future work to infer mobility patterns of more vehicles.

Privacy and Data for Good: In Mohen, all data are collected legally with drivers' consent and they reserve the right to withdraw personal data at any time, e.g., ETC records or GPS. Current statistics show that more than 99% drivers agree to contribute their data to the studies of urban traveling patterns. Besides the incentives provided by ETC and insurance companies, the trending awareness of 'Data for Good' has influenced more data owners to hold a more collaborative attitude as long as the personal information is well preserved.

Granted with these large scale real-world data, it is non-trivial to protect privacy of drivers especially in the context of fine-grained spatial information. All data used to train Mohen are anonymized, and we have filtered out the most unique drivers with strong patterns to prevent potential re-identification. In future work, rigorous studies about privacy bounds with respect to the data type would be conducted. We are working on sharing sample data to the academia in a responsible, privacy preserving way to improve the highway and local traffic experience which reciprocally benefit the drivers. To start with, urban-scale synthetic data sharing the same utility features will be released through our platform.

8 RELATED WORK

Human Mobility in Mobile Sensing Systems: Most of the existing work models human mobility based on mobile sensing systems, e.g., vehicular OBD devices [6] [9] [53] [46] and smartphone apps [55]. ElSherief et al. [7] propose querying algorithms to exploit limited probing sensing resources for higher spatial coverage detection. Model *ParkMaster* [19] leverages the ubiquitous smartphone as mobile sensing source to help drivers find parking spaces in the urban environment and model *ParkNet* creates a spot-accurate map of parking with GPS devices. Model *WalkCompass* [36] manages to tell the direction of a user by exploiting smartphone sensors and model

GruMon [38] accurately monitors group for dense and complex urban spaces by commodity smartphone sensors. Bejan et al. [1] provide a statistical model to recover speed information from the sparse probe movement data. Some recently works aim to design more delicate models to increase prediction accuracy [22] [8] [10] [16] [24] [37]. Given the full trajectories captured by these sensors, it is promising to model vehicular mobility. However, these approaches are generally constrained by the low penetration rate of vehicles, which limits their potential applications. For more general vehicles without equipped GPS devices, they cannot model the vehicular mobility such as predicting destinations.

Human Mobility in Stationary Sensing Systems: Modeling vehicular mobility through stationary sensing becomes popular as more and more Internet of Thing (IoT) devices are deployed in cities such as cameras and toll stations. For example, Wang et al. [43] analyze and infer traffic patterns with video surveillance information. Yang et al. [49] model the fine-grained travel time on highways with toll stations and infer the locations of each vehicle on highways [48]. Zhou et al. [59] build estimated conflict graphs by collecting limited signal measurements with limited number of stationary sensors. Some other works are based on cell towers [14] [13]. These approaches utilize the full penetration rate of the stationary sensors but cannot track vehicles after leaving the coverage of the sensors.

Human Mobility in Heterogeneous Sensing Systems: A few studies work on mobility modeling based on heterogeneous sensing systems. Meng et al. [31] infer the traffic volume based on the loop detector data and taxi trajectories. Wang et al. [42] track the hit-and-run vehicles with deployed cameras and taxis. Chowdhery et al. [5] propose a layered architecture bridging the local fine-time scale of sensors to the coarser topological and temporal operations of the cloud which facilitates information search and improve its availability. Lane et al. [26] utilize heterogeneous sense data to personalize classifiers with data contributed from other similar users and infer human activities. They did not fully utilize the complementary characteristics of heterogeneous sensing systems. Our work utilizes the strength of heterogeneous sensing systems compared to single sensing systems and deeply investigate the relationship of them. We study the destination prediction problem that cannot be completed by a single sensing system or naive combination of heterogeneous sensing systems.

Other Works on Heterogeneous Data Fusion: There are also many works related to fusing heterogeneous data, which are similar to our work that involves heterogeneous sensing systems. Jie et al. [17] and Cao et al. [3] designed algorithms to link user identities in heterogeneous mobility data. Wang et al. [44] performed location prediction using heterogeneous mobility data. Hsieh et al. [21] fused data from multiple data sources in a city to suggest the best station locations for air quality monitoring. Many other works were also based on heterogeneous data fusion to model the different phenomenon, such as urban anomalies [33] [54], city noise [57], region functions [50], traffic accidents inference [4], etc. However, our work is fundamentally different from these works in two aspects: (i) we do not require id linkage from different data sources, which is different from [17] [3]; (ii) our work aims to predict the mobility of people without data from mobile sensing systems by only fusing heterogeneous data from a small portion of vehicles, which differs from others that are implemented with full heterogeneous data such as [44].

9 CONCLUSION

In this work, we focus on the problem of extending sensing coverage of stationary sensing systems and design a novel system called *Mohen* to model vehicular mobility based on heterogeneous sensing systems. Our key idea is based on that human mobility is inherently consistent in different systems. We implement *Mohen* based on large-scale real-world data including ETC data of 2 million vehicles per day and GPS data of 114 thousand vehicles. The experiment shows its accuracy on the destination prediction and its effects on improving urban flow. We anticipate that this work will motivate public and private agencies to work together to enable a new class of live smart city services. We will release samples of our data to promote further research in this direction.

ACKNOWLEDGMENTS

This work is partially supported by NSF 1849238, 1932223, 1952096, and 2003874. We thank the associate editor and reviewers for the detailed and insightful feedback for this work.

REFERENCES

[1] Andrei Iu Bejan, Richard J Gibbens, David Evans, Alastair R Beresford, Jean Bacon, and Adrian Friday. 2010. Statistical modelling and analysis of sparse bus probe data in urban areas. In *13th International IEEE Conference on Intelligent Transportation Systems*. IEEE, 1256–1263.

[2] Léon Bottou. 2010. Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010*. Springer, 177–186.

[3] Wei Cao, Zhengwei Wu, Dong Wang, Jian Li, and Haishan Wu. 2016. Automatic user identification method across heterogeneous mobility data sources. In *2016 IEEE 32nd International Conference on Data Engineering (ICDE)*. IEEE, 978–989.

[4] Quanjun Chen, Xuan Song, Harutoshi Yamada, and Ryosuke Shibasaki. 2016. Learning deep representation from big and heterogeneous data for traffic accident inference. In *Thirtieth AAAI Conference on Artificial Intelligence*.

[5] Aakanksha Chowdhery, Marco Levorato, Igor Burago, and Sabur Baidya. 2018. Urban iot edge analytics. In *Fog computing in the internet of things*. Springer, 101–120.

[6] Zheng Dong, Cong Liu, YanHua Li, Jie Bao, Yu Gu, and Tian He. 2017. REC: Predictable Charging Scheduling for Electric Taxi Fleets. In *2017 IEEE Real-Time Systems Symposium (RTSS)*. 287–296.

[7] Mai ElSherief, Morgan Vigil-Hayes, Ramya Raghavendra, and Elizabeth Belding. 2017. Whom to Query?: Spatially-Blind Participatory Crowdsensing under Budget Constraints. In *Proceedings of the First ACM Workshop on Mobile Crowdsensing Systems and Applications*. ACM, 31–37.

[8] Zipei Fan, Xuan Song, Renhe Jiang, Quanjun Chen, and Ryosuke Shibasaki. 2019. Decentralized Attention-based Personalized Human Mobility Prediction. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 4 (2019), 1–26.

[9] Zipei Fan, Xuan Song, Ryosuke Shibasaki, and Ryutaro Adachi. 2015. CityMomentum: an online approach for crowd behavior prediction at a citywide level. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 559–569.

[10] Zipei Fan, Xuan Song, Tianqi Xia, Renhe Jiang, Ryosuke Shibasaki, and Ritsu Sakuramachi. 2018. Online Deep Ensemble Learning for Predicting Citywide Human Mobility. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 105.

[11] Zhihan Fang, Guang Wang, and Desheng Zhang. 2020. Modeling Fine-Grained Human Mobility on Cellular Networks. In *The World Wide Web Conference*. 1–2.

[12] Zhihan Fang, Shuai Wang, Guang Wang, Chaoji Zuo, Fan Zhang, and Desheng Zhang. 2020. CellRep: Usage Representativeness Modeling and Correction Based on Multiple City-Scale Cellular Networks. In *The World Wide Web Conference*. 1–11.

[13] Zhihan Fang, Yu Yang, Shuai Wang, Boyang Fu, Zixing Song, Fan Zhang, and Desheng Zhang. 2019. MAC: Measuring the Impacts of Anomalies on Travel Time of Multiple Transportation Systems. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 2 (2019), 1–24.

[14] Zhihan Fang, Fan Zhang, Ling Yin, and Desheng Zhang. 2018. MultiCell: Urban population modeling based on multiple cellphone networks. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 106.

[15] Zhihan Fang, Fan Zhang, and Desheng Zhang. 2019. Fine-grained travel time sensing in heterogeneous mobile networks. In *Proceedings of the 17th Conference on Embedded Networked Sensor Systems*. 420–421.

[16] Jie Feng, Yong Li, Chao Zhang, Funing Sun, Fanchao Meng, Ang Guo, and Depeng Jin. 2018. Deepmove: Predicting human mobility with attentional recurrent networks. In *Proceedings of the 2018 World Wide Web Conference*. International World Wide Web Conferences Steering Committee, 1459–1468.

[17] Jie Feng, Mingyang Zhang, Huandong Wang, Zeyu Yang, Chao Zhang, Yong Li, and Depeng Jin. 2019. DPLink: User Identity Linkage via Deep Neural Network From Heterogeneous Mobility Data. In *The World Wide Web Conference*. 459–469.

[18] Lianli Gao, Zhao Guo, Hanwang Zhang, Xing Xu, and Heng Tao Shen. 2017. Video captioning with attention-based LSTM and semantic consistency. *IEEE Transactions on Multimedia* 19, 9 (2017), 2045–2055.

[19] Giulio Grassi, Kyle Jamieson, Paramvir Bahl, and Giovanni Pau. 2017. Parkmaster: An in-vehicle, edge-based video analytics service for detecting open parking spaces in urban environments. In *Proceedings of the Second ACM/IEEE Symposium on Edge Computing*. ACM, 16.

[20] Klaus Greff, Rupesh K Srivastava, Jan Koutník, Bas R Steunebrink, and Jürgen Schmidhuber. 2017. LSTM: A search space odyssey. *IEEE transactions on neural networks and learning systems* 28, 10 (2017), 2222–2232.

[21] Hsun-Ping Hsieh, Shou-De Lin, and Yu Zheng. 2015. Inferring air quality for station location recommendation based on urban big data. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 437–446.

[22] Ryo Imai, Kota Tsubouchi, Tatsuya Konishi, and Masamichi Shimosaka. 2018. Early Destination Prediction with Spatio-temporal User Behavior Patterns. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 142.

[23] Sibren Isaacman, Richard Becker, Ramón Cáceres, Margaret Martonosi, James Rowland, Alexander Varshavsky, and Walter Willinger. 2012. Human mobility modeling at metropolitan scales. In *Proceedings of the 10th international conference on Mobile systems, applications, and services*. Acm, 239–252.

[24] Renhe Jiang, Xuan Song, Zipei Fan, Tianqi Xia, Quanjun Chen, Qi Chen, and Ryosuke Shibasaki. 2018. Deep roi-based modeling for urban human mobility prediction. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 14.

[25] John Krumm, Robert Gruen, and Daniel Delling. 2013. From destination prediction to route prediction. *Journal of Location Based Services* 7, 2 (2013), 98–120.

[26] Nicholas D Lane, Ye Xu, Hong Lu, Shaohan Hu, Tanzeem Choudhury, Andrew T Campbell, and Feng Zhao. 2011. Enabling large-scale human activity inference on smartphones using community similarity networks (csn). In *Proceedings of the 13th international conference on Ubiquitous computing*. ACM, 355–364.

[27] Daniel D Lee and H Sebastian Seung. 2001. Algorithms for non-negative matrix factorization. In *Advances in neural information processing systems*. 556–562.

[28] Shengbo Eben Li, Shaobing Xu, Xiaoyu Huang, Bo Cheng, and Huei Peng. 2015. Eco-departure of connected vehicles with V2X communication at signalized intersections. *IEEE Transactions on Vehicular Technology* 64, 12 (2015), 5439–5449.

[29] Qiang Liu, Shu Wu, Liang Wang, and Tieniu Tan. 2016. Predicting the Next Location: A Recurrent Model with Spatial and Temporal Contexts.. In *AAAI*. 194–200.

[30] Yin Lou, Chengyang Zhang, Yu Zheng, Xing Xie, Wei Wang, and Yan Huang. 2009. Map-matching for low-sampling-rate GPS trajectories. In *Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems*. 352–361.

[31] Chuishi Meng, Xiuwen Yi, Lu Su, Jing Gao, and Yu Zheng. 2017. City-wide Traffic Volume Inference with Loop Detector Data and Taxi Trajectories. (2017).

[32] Paul Newson and John Krumm. 2009. Hidden Markov map matching through noise and sparseness. In *Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems*. 336–343.

[33] Bei Pan, Yu Zheng, David Wilkie, and Cyrus Shahabi. 2013. Crowd sensing of traffic anomalies based on human mobility and social media. In *Proceedings of the 21st ACM SIGSPATIAL international conference on advances in geographic information systems*. 344–353.

[34] Luca Pappalardo, Filippo Simini, Salvatore Rinzivillo, Dino Pedreschi, Fosca Giannotti, and Albert-László Barabási. 2015. Returners and explorers dichotomy in human mobility. *Nature communications* 6 (2015), 8166.

[35] Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. 1532–1543.

[36] Nirupam Roy, He Wang, and Romit Roy Choudhury. 2014. I am a smartphone and i can tell my user's walking direction. In *Proceedings of the 12th annual international conference on Mobile systems, applications, and services*. ACM, 329–342.

[37] Amin Sadri, Flora D Salim, Yongli Ren, Wei Shao, John C Krumm, and Cecilia Mascolo. 2018. What will you do for the rest of the day?: An approach to continuous trajectory prediction. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 4 (2018), 186.

[38] Rijurekha Sen, Youngki Lee, Kasthuri Jayarajah, Archan Misra, and Rajesh Krishna Balan. 2014. Grumon: Fast and accurate group monitoring for heterogeneous urban spaces. In *Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems*. ACM, 46–60.

[39] Shuo Shang, Danhuai Guo, Jiajun Liu, and Kuien Liu. 2014. Human mobility prediction and unobstructed route planning in public transport networks. In *2014 IEEE 15th international conference on mobile data management*, Vol. 2. IEEE, 43–48.

[40] Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-Laszlo Barabasi. 2010. Limits of Predictability in Human Mobility *(Science)*.

[41] Martin Sundermeyer, Ralf Schlüter, and Hermann Ney. 2012. LSTM neural networks for language modeling. In *Thirteenth annual conference of the international speech communication association*.

[42] Yang Wang, Wuji Chen, Wei Zheng, He Huang, Wen Zhang, and Hengchang Liu. 2017. Tracking Hit-and-Run Vehicle with Sparse Video Surveillance Cameras and Mobile Taxicabs. In *2017 IEEE International Conference on Data Mining (ICDM)*. IEEE, 495–504.

[43] Yang Wang, Yiwei Xiao, Xike Xie, Ruoyu Chen, and Hengchang Liu. 2018. Real-time Traffic Pattern Analysis and Inference with Sparse Video Surveillance Information.. In *IJCAI*. 3571–3577.

[44] Yingzi Wang, Nicholas Jing Yuan, Defu Lian, Linli Xu, Xing Xie, Enhong Chen, and Yong Rui. 2015. Regularity and conformity: Location prediction using heterogeneous mobility data. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1275–1284.

[45] Rowan Wilken and Peter Bayliss. 2014. Locating foursquare: The political economics of mobile social software. In *Locative media*. Routledge, 193–208.

[46] Xiaoyang Xie, Yu Yang, Zhihan Fang, Guang Wang, Fan Zhang, Fan Zhang, Yunhuai Liu, and Desheng Zhang. 2018. coSense: Collaborative Urban-Scale Vehicle Sensing Based on Heterogeneous Fleets. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 4 (2018), 196.

[47] Xiao-Feng Xie and Zun-Jing Wang. 2018. SIV-DSS: Smart in-vehicle decision support system for driving at signalized intersections with V2I communication. *Transportation Research Part C: Emerging Technologies* 90 (2018), 181–197.

[48] Yu Yang, Xiaoyang Xie, Zhihan Fang, Fan Zhang, Yang Wang, and Desheng Zhang. 2019. VeMo: Enabling Transparent Vehicular Mobility Modeling at Individual Levels with Full Penetration. In *The 25th Annual International Conference on Mobile Computing and Networking*. 1–16.

[49] Yu Yang, Fan Zhang, and Desheng Zhang. 2018. SharedEdge: GPS-free fine-grained travel time estimation in state-level highway systems. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 48.

[50] Jing Yuan, Yu Zheng, and Xing Xie. 2012. Discovering regions of different functions in a city using human mobility and POIs. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. 186–194.

[51] Nicholas Jing Yuan, Yingzi Wang, Fuzheng Zhang, Xing Xie, and Guangzhong Sun. 2013. Reconstructing individual mobility from smart card transactions: A space alignment approach. In *Data Mining (ICDM), 2013 IEEE 13th International Conference on*. IEEE, 877–886.

[52] Jorge Zaldivar, Carlos T Calafate, Juan Carlos Cano, and Pietro Manzoni. 2011. Providing accident detection in vehicular networks through OBD-II devices and Android-based smartphones. In *2011 IEEE 36th Conference on Local Computer Networks*. IEEE, 813–819.

[53] Desheng Zhang, Jun Huang, Ye Li, Fan Zhang, Chengzhong Xu, and Tian He. 2014. Exploring human mobility with multi-source data at extremely large metropolitan scales. In *Proceedings of the 20th annual international conference on Mobile computing and networking*. ACM, 201–212.

[54] Mingyang Zhang, Tong Li, Hongzhi Shi, Yong Li, and Pan Hui. 2019. A decomposition approach for urban anomaly detection across spatiotemporal data. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*. AAAI Press, 6043–6049.

[55] Mingmin Zhao, Tao Ye, Ruipeng Gao, Fan Ye, Yizhou Wang, and Guojie Luo. 2015. VeTrack: Real Time Vehicle Tracking in Uninstrumented Indoor Environments. In *SenSys*.

[56] Yi Zhao, Xu Wang, Jianbo Li, Desheng Zhang, and Zheng Yang. [n. d.]. CellTrans: Private Car or Public Transportation? Infer Users' Main Transportation Modes at Urban Scale with Cellular Data. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 2019* ([n. d.]).

[57] Yu Zheng, Tong Liu, Yilun Wang, Yanmin Zhu, Yanchi Liu, and Eric Chang. 2014. Diagnosing New York city's noises with ubiquitous data. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 715–725.

[58] Pengfei Zhou, Yuanqing Zheng, and Mo Li. 2012. How Long to Wait? Predicting Bus Arrival Time With Mobile Phone Based Participatory Sensing. *IEEE Trans. Mob. Comput.* 13 (2012), 1228–1241.

[59] Xia Zhou, Zengbin Zhang, Gang Wang, Xiaoxiao Yu, Ben Y Zhao, and Haitao Zheng. 2013. Practical conflict graphs for dynamic spectrum distribution. *ACM SIGMETRICS Performance Evaluation Review* 41, 1 (2013), 5–16.