MIMU: Mobile WiFi Usage Inference by Mining Diverse User Behaviors

ZHOU QIN, Rutgers University, USA
YIKUN XIAN, Rutgers University, USA
FAN ZHANG, Shenzhen Institutes of Advanced Technology, China
DESHENG ZHANG, Rutgers University, USA

Mobile WiFi is a newly emerging service in recent years, which provides convenience for users to access online resources and increases revenues for operators via services such as advertisements and application promotions. However, in practice, the prohibitively high system implementation and operational costs, especially the costs of perpetual data traffic, hinder the further deployment of mobile WiFi services. In this paper, we present *MIMU*, a usage inference system for data traffic saving suitable for ubiquitous mobile WiFi services. We demonstrate the performance of the system via an example from the real-world nationwide edge computing mobile WiFi infrastructure. To address the impact of diverse user behaviors, we investigate the WiFi network usage from the perspective of users and devices, focusing on two unique features of mobile WiFi: user mobility regularity and access irregularity. In particular, we first design a deep learning-based two-dimension usage predictor to infer the future mobile WiFi usage with 1) a user dimension model with temporal attention addressing dominant users with heavy bus WiFi usage, and 2) a device dimension model with spatial attention addressing diverse WiFi usage and connection. Based on the results of the predictor, an application of content caching is implemented in an iterative fashion to save the data traffic. We evaluate *MIMU* by real-world bus WiFi system data sets of three major cities with 6,643 bus WiFi devices and 150k daily active users in total. Our results show that *MIMU* outperforms state-of-the-art methods in terms of usage inference. Moreover, we summarize the lessons learned from our large-scale bus WiFi system investigation.

CCS Concepts: • **Information systems → Mobile information processing systems**.

Additional Key Words and Phrases: Bus WiFi, Behavior, Mobile Networks, Deep Learning

ACM Reference Format:
Zhou Qin, Yikun Xian, Fan Zhang, and Desheng Zhang. 2020. MIMU: Mobile WiFi Usage Inference by Mining Diverse User Behaviors. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 4, Article 149 (December 2020), 22 pages. https://doi.org/10.1145/3432226

1 INTRODUCTION

In recent years, we are witnessing a rapid growth of mobile WiFi deployment and usage on public transportation systems (e.g., subways, buses, and even airplanes) due to its convenience for passengers and potential commercial values for service providers [22, 40]. For example, there were more than 100,000 WiFi devices deployed on urban transit buses in China by 2015 [18], and it is expected more public transportation systems will offer WiFi within a few years [1]. Besides, it is reported that there are more than 3.39 million unique users using WiFi on public

Authors' addresses: Zhou Qin, Rutgers University, 96 Frelinghuysen Rd, Piscataway, NJ, 08854, USA, zq58@cs.rutgers.edu; Yikun Xian, Rutgers University, USA, siriusxyk@gmail.com; Fan Zhang, Shenzhen Institutes of Advanced Technology, China, zhangfan@siat.ac.cn; Desheng Zhang, Rutgers University, USA, desheng.zhang@cs.rutgers.edu.

2474-9567/2020/12-ART149 $15.00
https://doi.org/10.1145/3432226

transport in 9 European cities under the UK Government's SuperConnected cities initiative [17]. As a promising technology, mobile WiFi is beneficial for passengers, vehicular fleet companies, and mobile WiFi service providers. Taking the bus system as an example: 1) for the bus passengers, bus WiFi enables them to access the Internet more conveniently with a low-cost and high-quality service compared to cellular data plans [35]; 2) for the bus company, it has the potential to improve bus passenger experience; 3) for the bus WiFi service providers, personalized advertisements and app promotion can be cast based on user profiling, which may potentially increase their revenues. Currently, there are many real-world bus WiFi systems, e.g., Google bus WiFi [2], New York bus WiFi [6], and Shenzhen Huashi bus WiFi [15], and most of them are free for bus passengers to use.

Despite the foreseeable benefits of bus WiFi for all passengers, company, and service providers, bus WiFi services have not yet been available extensively due to various real-world reasons, among which the high operational cost (mostly for monthly cellular data fees) is the key concern that hinders the bus WiFi from being further promoted [26, 38]. Hence, comprehending bus WiFi user behavior patterns in terms of mobility and resources accessing is essential for reducing operational costs via developing practical systems such as content caching.

Research efforts have been taken to study several aspects regarding mobile WiFi operation and related applications [13, 30, 41], including making use of deployed WiFi Access Point (AP) to boost applications such as predicting bus arrival time in urban regions [30], and analyzing log data from the system to understand WiFi networks status and associated users [7, 22]. Some works are dedicated to building edge-computing platforms [22, 35]. Techniques used to investigate cellular networks can also be used to investigate mobile WiFi due to similar settings, i.e., stationary cellular towers [42].

However, the above works focused on stationary WiFi or cellular networks, which did not involve the key features of mobile WiFi users, i.e., *mobility regularity* and *access irregularity* in the setting of public transportation. In particular, the bus passengers typically have pretty regular mobility and access patterns, e.g., taking the same bus line during workdays and watching videos during commuting. However, most of the bus passengers have connected to different on-board WiFi devices because they will mostly get on different buses despite taking the same bus line every day. This unique combination of mobility regularity and access irregularity makes the *bus WiFi usage highly irregular* at the device dimension and thus significantly differs from the above previous works on stationary WiFi and cellular networks. For those two networks, the mobility regularity of users will lead to the access regularity (due to stationary APs), which makes *network usage regular* at the device dimension.

In this paper, we collaborate with Shenzhen Huashi WiFi company, which is one of the biggest bus WiFi service providers in China. We utilize Huashi Bus WiFi infrastructure to design a content cache system based on the comprehensive investigation towards spatiotemporal patterns of bus WiFi from the user dimension and device dimension in the same city. Through the observations and insights obtained from the in-depth WiFi usage pattern investigation, the caching system is developed with a sophisticated content predictor and implemented in an iterative updating fashion. The implementation of this content caching application on bus WiFi devices can reduce traffic data and improves the system performance of the mobile WiFi networks. The major contributions of this paper can be summarized:

- To our best knowledge, we conduct the first work to investigate usage patterns of large-scale mobile bus WiFi networks in the real world from the perspective of mobility regularity and access irregularity. More specifically, we conduct an in-depth case study by leveraging bus WiFi log data from an average of 110k active users per day, which are collected from 4,384 bus WiFi devices deployed in Shenzhen. **Our sample data will be released for reproducibility and following work in this direction.**
- Based on the comprehensive investigation on the bus WiFi infrastructure, especially on the usage patterns from two different perspectives, i.e., the user perspective and aggregate device perspective, we design a dedicated two-dimension (user dimension and device dimension) caching system called *MIMU* to infer the future bus WiFi usage based on mobility regularity and access irregularity. Our deep learning-based

predictor considers the patterns from log data investigation, such as user-device interaction, the similarity of usage from close timing, and the similarity of usage from the close locations, etc.

- Our *MIMU* is evaluated by two-month long bus WiFi log data from three cities in China, i.e., Shenzhen, Wuxi, and Nanjing, including 4,384/1,478/782 bus WiFi devices, with 114k/32k/14k daily active users, respectively. It is shown that *MIMU* achieves better performance than state-of-the-art methods. Evaluations on our content caching application for bus WiFi also show that data traffic can be saved for WiFi service providers, which encourages future system deployment at a larger scale.
- We provide some in-depth discussions for the lessons learned, including usage patterns and potential commercial values. Our data-driven insights and modeling have the potential to help bus WiFi operators and similar mobile WiFi systems on public transportation such as subways to understand usage patterns, and better develop profitable operation models considering the mobility regularity and access irregularity.

2 MOTIVATION

2.1 WiFi System Operation Cost

The cost of implementing and maintaining a city-scale bus WiFi system is undoubtedly high; for example, a WiFi service company called 16WiFi invested around $90 million in the bus WiFi system in the Chinese city Beijing but still faces capital shortage challenges [44]. Commercial service providers whose primary purpose is to gain profits may approach from two perspectives: 1) one is reducing the system operation cost, mainly the cost of continuous data usage considering the fact that others such as the costs of WiFi devices are unavoidable and one-time payments; 2) the other is expanding business such as advertising based on the system and corresponding sources, e.g., advertising on videos, audios, online games, and online purchasing platforms.

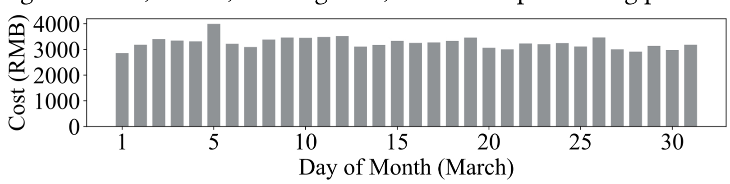

Fig. 1. Daily Data Traffic Cost Distribution in Shenzhen

High Data Usage Cost Unlike one-time costs such as payments to bus companies and governmental transportation department, or expense on WiFi devices purchases, perpetual data usage cost is one of the significant practical challenges for the commercial bus WiFi systems, leading to the shutdown of many bus WiFi projects despite that the cost of cellular data keeps decreasing in China [44, 45]. The current bus WiFi system in the Chinese city Shenzhen run by Huashi Company also faces the same challenge. To offset the loss, they store multimedia contents on the hard disk of router devices and offer free access to bus passengers to save some data traffic, but it does not work well since the contents are fixed. The data usage demand is high due to long-running hours of bus systems (average 7.55 hours per bus per day in Shenzhen), a large number of WiFi users (an average of 114 thousand users per day), and the large amount of data consumed, as shown in Fig. 1, where the daily data traffic cost across Shenzhen for one month in 2017 is presented. We find that average daily data traffic is as high as 108.56 GB, equating to around 3,256 RMB according to the data plan rate, which is three thousand times expensive as the bus fare (1 RMB) [9]. Such a high demand for data traffic and high data costs pose a vital threat to the continuous operation of the bus WiFi system. National deployment of such systems thus poses a considerable capital challenge for the company.

2.2 Challenges

To tackle the issue of high cost, we aim to design a content caching system via data usage predictor to better predict future usage. To achieve this, we have to predict visiting contents (e.g., URLs of *HTTP*) accurately given a limited-space hard disk. By checking the data sets, we find out that there are several real-world challenges related to data quality and usage patterns. Below we summarize the challenges related to building an accurate predictor.

Unbalanced Usage Distribution: Considering the large number of users involved in the bus WiFi system, one may expect rich historical data for each user since we can gather users' usage from different devices. However, our observation is against this assumption. On the one hand, users may visit the bus WiFi for limited times, as shown by Fig. 2, where we find that users have used WiFi for fewer times per day than we expect. Specifically, 55% of users have only one *HTTP* visit per day (as demonstrated in Fig. 2). Meanwhile, the heavy users can have hundreds of *HTTP* visits records in a single day. We further find that data usage is severely unbalanced among users where the top 10% heavy users account for more than 90% of the data usage. For users without enough visiting data, it is challenging for us to predict future usage based on users. On the other hand, for users with rich single-day visiting data, the number of days with WiFi connection may be limited. Explicitly, in our two-month-long data sets, users may only use bus WiFi for a specified number of days, as illustrated in Fig. 3, where we can see that 96% of users have visited *HTTP* via bus WiFi for fewer than 7 days. It means that most users have only a few days of historical data available for future prediction investigation. In short, the historical data are limited for the majority of users on a daily basis and across different days.

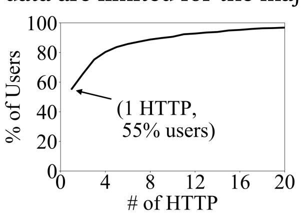

Fig. 2. WiFi Visits Distribution

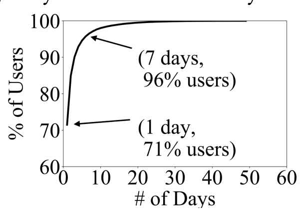

Fig. 3. Visiting Distribution

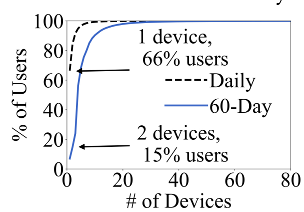

Fig. 4. Devices Connection

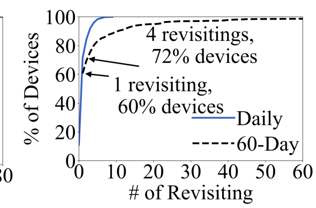

Fig. 5. Users Connection

Diverse Usage Distribution: Even the historical data are abundant for heavy users, we still face the challenge brought by user behavior irregularity. The bus WiFi setting is different from the traditional WiFi setting regarding the user-machine interaction: 1) the user with regular commuting patterns may still connect to different WiFi devices, e.g., different buses within the same bus line; 2) the WiFi device is mobile and gets connected to diverse groups of users. These fundamental differences lead to several challenges.

Firstly, a user may connect to different devices per day, as shown in Fig. 4. Around 34% of users connect to more than 1 bus devices on a daily basis, 85% of users connect to more than 2 bus devices in our two-month-long data set, suggesting high connection dynamics. Even the usage at the user level is regular, e.g., a user takes the same bus line every day and has a specific online surfing pattern, the usage may be distributed to different devices, disrupting the regularity from individual users. From both a daily basis and multiple days based perspectives, users tend to connect to numerous bus devices, determined by the characteristics of the bus WiFi systems.

Secondly, from the device level, the connection is similarly dynamic, i.e., a specific device is connected by different users within a day or across multiple days, suggesting a user seldom revisits the same device s/he connected before. We show the revisiting times (i.e., visiting more than once) from the same user for all the devices in Fig. 5, with the x-axis denoting the number of revisiting from the same user. We observe that 60% of the devices only have 1 revisiting from each user on average per day, meaning 40% of the devices never have users revisited. For the two-month-long period, 72% of the devices have fewer than 4 times of revisiting in accumulation. Directly dealing with such diverse user participation on the device level would be hard since each user contributes to different usage.

2.3 Opportunity: User Behavior Regularity

User WiFi Usage Behavior: As we have shown above, the bus WiFi usage on devices is dynamic and diverse, making gathering behaviors of different users difficult. However, we argue that the users themselves can be a potential opportunity. The bus WiFi system log data include several fields regarding user behaviors, such as the connection status and preferences of the users in terms of Internet surfing, which can provide us with rich spatiotemporal and contextual information to assist understanding and predicting data usage. Details of related data fields will be introduced in the next section. By leveraging data fields associated with users, we can analyze the usage patterns of the users by historical connection and browsing records of users, for instance, when did the user connect to which bus WiFi device and what kind of content s/he visited (we discuss the possible privacy issues in the discussion section). According to the relative regular pattern of user behaviors [11, 46], we then can predict when will which user visit what contents. Also, based on the historical connected devices and bus routes, we can have the connection status of users given a specific time, i.e., which device the user is likely to connect to. Similar connection patterns and usage behaviors extracted from different groups of users can be beneficial to the data usage prediction.

3 MIMU FRAMEWORK

3.1 Bus WiFi System

In this section, we use the bus WiFi system as an example to introduce the infrastructure and data sets, together with the design overview of the caching system.

3.1.1 Infrastructure. The bus WiFi system is a commercial mobile WiFi system aiming to provide free WiFi access to bus passengers as part of the smart city initiative. The system contains several major components, including WiFi devices (we will use "device" for brevity and there is only one device on each bus) deployed on the buses, external cellular networks in the city which are already functioning, and an operation center connected to the Internet where all buses devices can be under centralized management. The overview of the physical infrastructure is shown in Fig. 6.

The device here is the critical component in the bus WiFi system since it has multiple functions, such as uploading data through cellular networks via SIM cards and providing WiFi access to passengers. Besides, the device also contains a 32GB hard disk that can store multimedia contents and can be updated when the bus is back to the garage. The device can be working as an Access Point (AP), providing WiFi services to on-board users via hotspot sharing by the SIM card. Users can also directly visit local content on the hard disk of the device at high accessing speed. When the buses are back in the parking lot, the devices connect to the Internet via the wired connection and can then be under unified management. Under this case, the devices can update local multimedia contents, if any.

Fig. 6. Overview of Bus WiFi System Infrastructure

3.1.2 Data. The bus WiFi log data is two-month-long (from 2017-02-01 to 2017-04-01), covering 22 major cities with 34,377 bus devices in China. Our investigation here focuses on Shenzhen as an example for the most time, but we also address the generalizability of *MIMU* in the evaluation through implementation in multiple cities. The log data are collected via the device when there is communication between device and user terminals as offline data and uploaded when buses are back in the parking lot. The data can be generally classified into several categories, with rich information regarding users' bus WiFi connection and usage. Partial data fields and samples related to our investigation are shown in Table 1. For brevity and convenience, we reformat the data into four comprehensive data sets, which are vital for further investigation.

- **Traffic Flow Data:** The traffic flow data set contains all the data traffic consumption records, i.e., how much data are uploaded and downloaded via which device by which user at what time.
- **Content Data:** Two types of visiting records are stored. One is *HTTP* visiting records, i.e., URL addresses; the other is *portal* contents visiting records, i.e., local contents stored in the hard disk, we use "*portal*" for short in the rest of the paper.
- **Connection Data:** The connection data contain the records about connection and disconnection time, together with the associated user ID (encoded string based on user WiFi account) and unique device ID, which can be used to determine whether the user is on-board and the duration of bus WiFi connection session.
- **Location Data:** The device also stores the GPS locations of the running bus but at a low updating frequency (at minutes level). Besides, arrival time at different bus stops is also stored.

Table 1. Partial Bus WiFi Data Example

Field	Value	Field	Value
Timestamp	20170407T16:03:25.000Z	Upload (byte)	12624
Device ID	9303LL201507080004	Download (byte)	11088
User ID	X0MkcDjsujzP4ML+I74/tQ	Portal ID	Tank Game
HTTP	http://m.sogou.com/	RSSI (dB)	-60
Longitude	118827420	Latitude	31982690
Bus Line	E16	Bus Plate	BBA106
City	Shenzhen	Storage	31.64 GB
Total number of WiFi Devices in Shenzhen: 4,384			
Total number of WiFi Users in Shenzhen: 114k			

3.2 Design Overview

3.2.1 Problem Formulation. We start by introducing the formulation of WiFi usage prediction problem. Formally, let $\mathcal{D}$ and C respectively be the sets of devices and contents. Let $\mathcal{U}$ be the set of dominant users that contribute frequent visitation and heavy data traffic to the bus WiFi system. The *Bus WiFi Usage Prediction (BPred)* problem can be formulated as follows.

Definition 3.1. (*BPred* Problem) Given device $d \in \mathcal{D}$ and content $c \in C$, the goal is to predict the WiFi usage of content c on device d, denoted by $y_{d,c} \in \mathbb{R}$.

The bus WiFi usage $y_{d,c}$ here can be decomposed along two dimensions: user dimension and device dimension. User dimension refers to that large proportion of WiFi usage which is contributed by a small group of dominant users from $\mathcal{U}$; while device dimension accounts for the aggregation of the rest of users whose historical records are sparse and in turn their usage is modeled at the device level for compensation. Therefore, the overall usage $y_{d,c}$ can be estimated as follows.

$$y_{d,c} = \sum_{u \in \mathcal{U}} f_\theta(d,c,u) + \alpha g_\phi(d,c), \tag{1}$$

where $f_\theta(d,c,u)$ denotes the user dimension function that estimates the WiFi usage of content c on device d given the dominant user u, and $g_\phi(d,c)$ denotes the device dimension function for aggregated usage prediction. α is a hyperparameter that controls the tradeoff between the WiFi usage results from two dimensions. Accordingly, our target is equivalent to learning two functions $f_\theta(d,c,u)$ (user dimension) and $g_\phi(d,c)$ (device dimension) from historical WiFi usage data so that they can be applied for future usage prediction.

3.2.2 Design Objectives. The objectives of our design are two-fold: i) we want to extract patterns existing in users' WiFi usage, e.g., the regularity of visiting certain contents; ii) based on these patterns, we want to predict future usage in the bus WiFi scenario accurately for content caching.

Better Historical Usage Understanding: The first goal of our design requires a comprehensive investigation of possible patterns in users' WiFi usage. The understanding can be approached from three aspects: i) the connection patterns of users, e.g., when will the user get on the bus or start using the bus WiFi service; ii) what kinds of contents is the user likely to visit, according to the historical visiting records; iii) is there any correlation existing in the bus WiFi scenario, e.g., similar usage across different users, usage dynamics difference and similarity for different contents, different bus lines, etc.

Better Future Usage Prediction: The ability to predict future bus WiFi usage can be beneficial for the bus WiFi company, e.g., they can better manage the future cellular data package plan, schedule future advertisement casting, and regulate the bus WiFi hardware, etc. Based on the understanding of usage patterns, a more accurate prediction may be achieved compared to vanilla temporal sequence prediction schemes. In this paper, we design *MIMU* for caching via predicting future bus WiFi usage from two dimensions, i.e., user dimension and device dimension. The reason that we propose such a two-dimension prediction scheme is based on two vital data-driven factors, i.e., usage imbalance among users and the similarity of buses in the same bus line. For users with heavy bus WiFi usage, we can make use of their behavior patterns to model and predict their usage patterns. For users without many historical usage records, we can aggregate them at the bus device level to accumulate enough historical data for prediction.

To achieve the design objectives, we briefly introduce the design overview of a predictor in *MIMU*, which is built from two dimensions: user dimension and device dimension, as shown in Fig. 7.

- **Historical Data Investigation:** First, we investigate the patterns from historical data to guide further modeling, e.g., usage dynamics of users and buses, similarities among them, and spatiotemporal correlations.
- **Future Usage Prediction:** Based on the findings, we build a two-dimension model, where the user modeling part deals with users with rich historical data, e.g., the top 10% users with a large amount of WiFi usage data. It enables usage prediction with three kinds of contextual information, i.e., visiting timing, the device connected and visiting contents. The device modeling part is built at the device level, where we predict the future usage of the device by aggregated historical usage.

The advantages of this two-dimension design lie in several aspects: 1) user patterns such as behavior regularity are considered; 2) similarities among different users, devices, and locations are included; 3) modeling can work even there are few user data. Design details are elaborated in the following subsection.

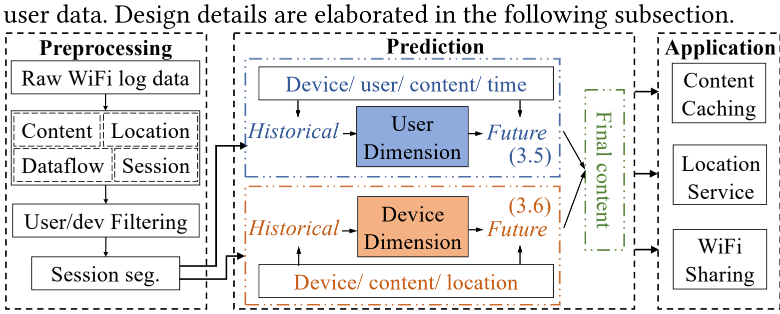

Fig. 7. MIMU Overall Framework

3.3 Pre-processing

For most raw data collected from the infrastructures, they suffer from several issues that make direct investigation ineffective, e.g., missing values, inconsistent data format, repetitive data fields in separate tables. Further processing is required to better understand the data, such as trip segmentation, outliers filtering, etc.

Data Formatting: In the raw bus WiFi log data, the records are separately stored in 17 different tables with different data fields format. To begin with, we first reformat the raw data into four types by joining different tables in the same fields. For example, the data traffic flow of users is kept in one data set, and the online content (*HTTP* content or *portal* content) visiting records are kept in another data set. Analyzing the data consumption during a user visiting a certain URL requires calculation involving both data sets. In the end, we fuse these data sets into one data set with a consistent format, including information important to usage understanding:

`[userID, deviceID, timing, location, URL, portal, uploaded data, downloaded data]`

Specifically, the *userID* and *deviceID* are encrypted unique IDs for users and devices. For the timing, we partition 24 hours of a day into 288 5-minute-long time slots for further processing (fine-grained enough compared to the average online session length, which is 5.4 minutes). For the location, we partition the Shenzhen city into 901×471 $100m \times 100m$ grids for better investigation, and hash the indexes of a grid into values for later use [36].

Targets Filtering: As mentioned before, there are more than 114k users recorded and around 4.4k bus WiFi devices deployed in the Shenzhen bus WiFi system. However, not all the users use the bus WiFi every day, or all the devices operate every day. Here we filter the active users and devices with at least 7 days of data in our two-month-long data set as the data for further processing to achieve better performance [16]. As a result, 20,168 users and 3,889 buses are involved in the user dimension modeling and device dimension modeling, respectively.

3.4 Data Investigation For Overall Model Design

To accurately predict the bus WiFi usage, an intuitive idea is to build a temporal sequence prediction model for all the visited contents. However, this method misses contextual information brought by the users. After all, all the usage is contributed by users; a specific user may prefer to visit particular contents, use the bus WiFi at a certain time on some specific bus lines. We analyze the entropy [5] of visiting content across multiple days including *HTTP* and *portal* and find out that, in general, the user-level entropy is clearly smaller than device-level entropy, denoting a better predictability (as shown in Fig. 8 and Fig. 9).

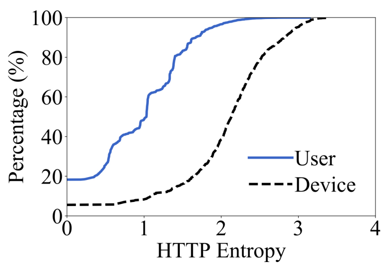

Fig. 8. *HTTP* Entropy

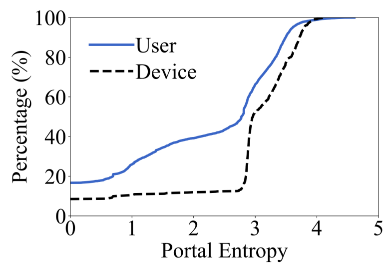

Fig. 9. *Portal* Entropy

However, we cannot simplify the model focusing solely on the user level, since not all the users use the bus WiFi frequently. According to Fig. 2 and Fig. 3, we know that more than half of the users have *HTTP*/*portal* request for only once a day. One step further, we find that on average, 58.59% of users have not uploaded any data, and 68.32% of users have not downloaded any data during a day. At the same time, the top 10% of users consume more than 90% of the data. Considering the data availability and user behavior regularity, we design our modeling process as two dimensions in *MIMU* to tackle the bus WiFi usage prediction. For user dimension modeling, we focus on the dominant users who are selected by three metrics: average daily WiFi usage, average daily connection times, number of days with WiFi usage records. For the device dimension, we focus on the aggregated WiFi usage from the rest of the non-dominant users.

3.5 MIMU: User Dimension

3.5.1 Data Investigation for User Dimension Modeling. The ultimate goal of the predictor is to infer future usage (*HTTP* and *portal*) on the device, despite that usage comes from both the users and the devices. That being said, we have to distribute users' usage to different bus devices. To achieve so, we also have to consider the future user-device connection. Luckily, we find the device connection entropy is low, as shown in Fig. 10, where 46.02% of users only connect to one bus WiFi device per day and 35.09% of users connect to two bus WiFi devices per day. Fewer than 20% of users connect to multiple (> 2) devices, making device connection easier to be considered. We also investigate the usage similarity regarding usage interval, i.e., the Cosine similarity of two vectors where each vector represents the contents visited by all users at a given time. We find out that usage from close time tends to have higher similarity, suggesting the high temporal dependence for the bus WiFi usage, as shown in Fig. 11. The darker color denotes a higher occurrence of corresponding similarity and interval.

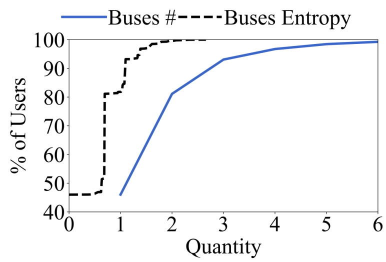

Fig. 10. Bus Entropy

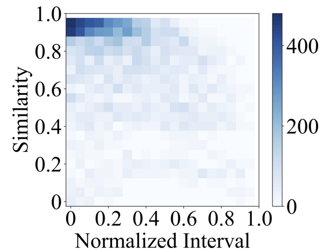

Fig. 11. Timing Correlation

3.5.2 Design Overview of User Dimension. Modeling in the user dimension handles users with abundant historical and massive WiFi data usage, considering the efficiency and necessity of doing so. For efficiency, patterns can be considered, such as the regularity of users' bus WiFi usage from temporal perspectives, e.g., a user may tend to use bus WiFi at the regular time and connect to regular bus devices (several devices belonging to the same bus line). For necessity, considering the timing and connection of WiFi usage is essential to obtain the device level usage, i.e., the device level usage can be aggregated from user-level usage. Thus, we format the problem into a prediction problem with temporal attention, i.e., using the previous $N-1$ days multi-dimension data to predict the N_{th} day multi-dimension data.

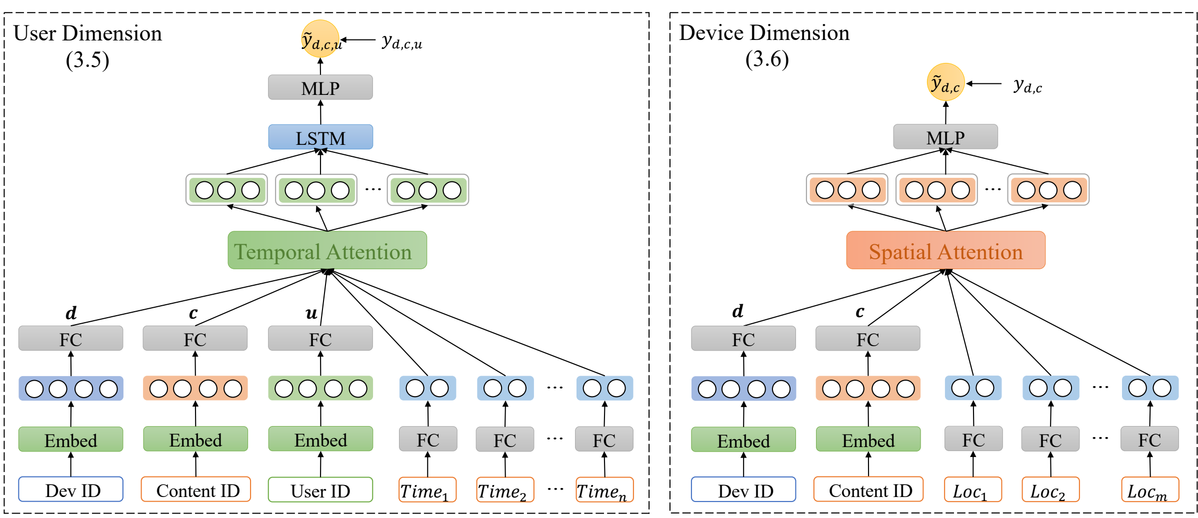

Fig. 12. Structure of the Two-dimension Predictor

3.5.3 Predictor with Temporal Attention. In the user dimension usage prediction, only data of the dominant users will be considered in the modeling, since these users contribute most of the bus WiFi usage and they also have rich historical data for individual level modeling. The details regarding the modeling are illustrated as follows.

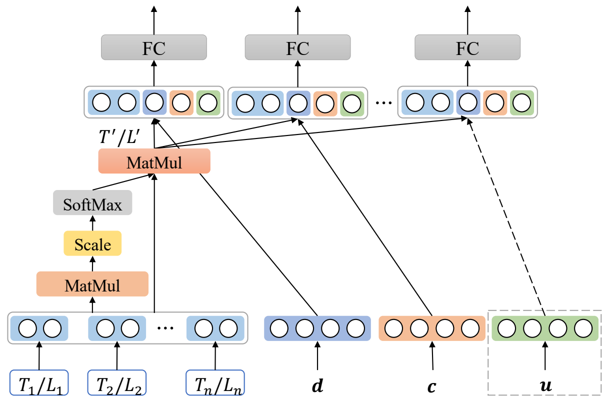

Fig. 13. Details of Temporal and Spatial Attention

We first denote the set of all the bus WiFi devices by $\mathcal{D}$, the content set by C and the set of dominant users by $\mathcal{U}$. Given a bus WiFi device $d \in \mathcal{D}$, a content $c \in C$ and a user $u \in \mathcal{U}$, the user-dimensional function $f_\theta(d, c, u)$ parameterized by θ aims to output estimated WiFi usage $y_{d,c,u} \in \mathbb{R}$ of content c on device d contributed by the user u. Basically, the inputs d, c, u are all identifiers and can be represented as one-hot vectors, i.e., $d \in \{0,1\}^{|\mathcal{D}|}, c \in \{0,1\}^{|C|}, u \in \{0,1\}^{|\mathcal{U}|}$, where only one dimension of each vector is set to one with rest being zeros.

For better representation power and generalizability, we approximate the function $f_\theta(d, c, u)$ by a neural network whose architecture is shown as the user dimension component in Fig. 12, also defined as follows. First, device d, content c, and user u are encoded into latent vectors denoted by $\boldsymbol{d}, \boldsymbol{c}, \boldsymbol{u}$ via embedding layers with nonlinear projection. That is

$$\boldsymbol{d} = V_d \sigma(W_d d), \quad \boldsymbol{c} = V_c \sigma(W_c c), \quad \boldsymbol{u} = V_u \sigma(W_u u), \tag{2}$$

where $W_d \in \mathbb{R}^{l_d \times |\mathcal{D}|}, W_c \in \mathbb{R}^{l_c \times |C|}, W_u \in \mathbb{R}^{l_u \times |\mathcal{U}|}$ are embedding layers, $V_d \in \mathbb{R}^{l \times l_d}, V_c \in \mathbb{R}^{l \times l_c}, V_u \in \mathbb{R}^{l \times l_u}$ are linear layers and $\sigma(\cdot)$ denotes nonlinear activation function (we adopt ReLU [12] here).

We also take temporal factors into consideration for two reasons: 1) The temporal information reveals resemblance among different contents, as we have demonstrated in Fig. 11, where WiFi usage presents a higher similarity for close timing; 2) User level behaviors are closely related to time, e.g., taking buses at regular time for commuting, etc. Specifically, let $\{t_1, \ldots, t_n\}$ be a sequence of timestamps, which can also be embedded into learnable latent vectors in the same low-dimensional space. Considering the ordering relationship among n timestamps, we impose positional encoding on each of the timestamp latent representations and denote the derived vectors by $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n \in \mathbb{R}^{l_t}$. In order to automatically recover the proportion of contribution from different timing, we propose to leverage Self-Attention Block (SAB) [43] over $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n$, as illustrated in Fig. 13. (Notice that we illustrate both temporal and spatial attention in the same figure, more details regarding spatial attention of device dimension are in the next section). Therefore, the output attentive timestamp matrix $\boldsymbol{T}' \in \mathbb{R}^{n \times l_t}$ can be acquired by $\boldsymbol{T}' = \mathrm{SAB}(\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n)$, where the SAB block is generally defined as:

$$\mathrm{SAB}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \mathrm{softmax}\left(\frac{\boldsymbol{X}\boldsymbol{X}^\top}{\sqrt{d}}\right)\boldsymbol{X}, \text{ where } \boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n]^\top, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathbb{R}^d, \tag{3}$$

where $\boldsymbol{X}$ can be temporal matrix $\boldsymbol{T}$ or spatial matrix $\boldsymbol{L}$, and d is the scaling factor. Now, given the encoded vectors of device $\boldsymbol{d}$, content $\boldsymbol{c}$, user $\boldsymbol{u}$, and the temporal matrix $\boldsymbol{T}'$ output by SAB, we aggregate all four kinds of information to make prediction on WiFi usage. Specifically, let $\boldsymbol{t}'_i$ be the i-th row of matrix $\boldsymbol{T}'$. The aggregated vector $\boldsymbol{z}_i$ at timestamp t_i is defined as

$$\boldsymbol{z}_i = W_z[\boldsymbol{d}; \boldsymbol{c}; \boldsymbol{u}; \boldsymbol{t}'_i], \ i = 1, \ldots, n, \tag{4}$$

where $[;]$ denotes concatenation operation and $W_z \in \mathbb{R}^{l_x \times (3l+l_t)}$ is the learnable parameter. To capture sequential information at different timestamps, we define the final prediction block as follows:

$$f_\theta(d, c, u) = \mathrm{MLP}(\mathrm{LSTM}(\{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n\})), \tag{5}$$

where $\mathrm{LSTM}(\cdot)$ stands for one layer of Long Short Term Memory networks [24] to process the sequential input, and $\mathrm{MLP}(\cdot)$ stands for Multilayer Perceptron for output reshaping, i.e., two fully connected layers compressing $(3l + l_t)$ into 1 dimension. Finally, by aggregating all tuples of devices, contents and users, we adopt L2 distance as the objective function to learn the model parameters for user dimension WiFi usage prediction:

$$\ell_u(\theta) = \sum_{d \in \mathcal{D}} \sum_{c \in C} \sum_{u \in \mathcal{U}} \| f_\theta(d, c, u) - y_{d,c,u} \|_2^2. \tag{6}$$

3.6 MIMU: Device Dimension

3.6.1 Data Investigation for Device Dimension Modeling. User-level modeling alone is not enough, due to the sparse historical records for the majority of users. When the number of records is insufficient for user dimension modeling, we turn to model at the device level for compensation since aggregation leads to more historical records. We find that devices at the same bus line show higher similarity than devices at different bus lines, suggested by Fig. 14, which justifies the model training at the device level. Moreover, the device level usage shows spatial dependence in Fig. 15, i.e., the bus WiFi usage suggests high similarity at close locations (higher frequency indicated by darker color), which can be helpful in the modeling.

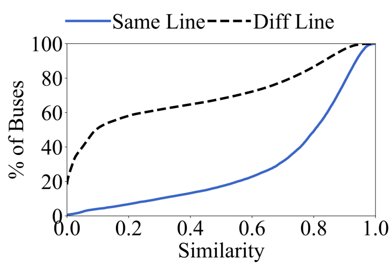

Fig. 14. Bus Similarity

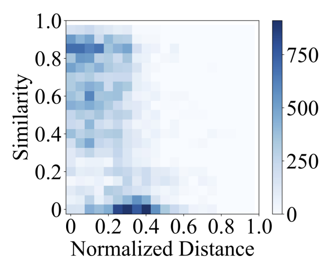

Fig. 15. Location Correlation

3.6.2 Design Overview of Device Dimension. Taking the existing similarity at the device level, we format the problem into a prediction problem with spatial attention, where each content is treated as an item, and the corresponding bus WiFi device is regarded as a consumer, similar to the setting of recommendation. system. The number of times content visiting on the device can be treated as ratings [34]. The goal is to infer future ratings, i.e., predict future bus WiFi usage. In this way, we can treat all the devices as consistent "consumers", and their bus line level similarity and spatial dependency can be involved in the deep neural networks with spatial attention. The high-level goal is still to predict bus WiFi usage on N_{th} day by the previous $N - 1$ days.

3.6.3 Predictor with Spatial Attention. In the device dimension usage prediction, we target the modeling on those non-dominant users since their historical WiFi usage data are too sparse to be considered for individual modeling. Thus, we aggregate all the usage from these users on the device level for further processing, i.e., dumping the individual level identification. The details regarding the modeling are illustrated as follows.

Similar to the user dimension, given a bus WiFi device $d \in \mathcal{D}$ and a content $c \in C$, the device-dimensional function $g_\phi(d, c)$ parameterized by ϕ seeks to estimate WiFi usage $y_{d,c}$ of content c on device d. We also approximate the function $g_\phi(d, c)$ by a neural network with architecture shown as the device dimension in Fig. 12. First, device d and content c are encoded into latent vectors via the shared embedding layers in Eq. 2 but different nonlinear projections:

$$\boldsymbol{d} = V_d' \sigma(W_d d), \quad \boldsymbol{c} = V_c' \sigma(W_c c), \tag{7}$$

where $V_d' \in \mathbb{R}^{l \times l_d}, V_c' \in \mathbb{R}^{l \times l_c}$ are linear layers.

For the device dimension, we focus more on the spatial factors in the modeling because of two unique characteristics: 1) As presented in Fig. 15, we observe that WiFi usage show higher similarity for close locations (indicated by GPS location data), which can be manipulated to assist our usage prediction; 2) Bus line information can be considered, i.e., the devices belonging to the same bus line shows higher usage similarity, as shown in Fig. 14. In details, let $\{l_1, \ldots, l_m\}$ be a set of m locations, which can also be embedded into latent vectors in the same low-dimensional space, denoted by $\boldsymbol{l}_1, \ldots, \boldsymbol{l}_m \in \mathbb{R}^{l_l}$. Notice that the locations are limited for a given device d, because the bus lines are fixed and the location grids are thus limited. In this way, the mapping relations of bus lines to devices (buses) are considered in the device dimension. Similarly, to automatically calculate the contribution proportion of different locations, we also leverage the SAB block defined in Eq. 3 over $\boldsymbol{l}_1, \ldots, \boldsymbol{l}_m$, as shown in Fig. 13. In the device dimension, there are no more identifiers of users, so we mark the user input by dotted frame. The output attentive location vectors $\boldsymbol{L}' \in \mathbb{R}^{m \times l_l}$ can be computed as $\boldsymbol{L}' = \mathrm{SAB}(\boldsymbol{l}_1, \ldots, \boldsymbol{l}_m)$.

We then aggregate the encoded vectors of device $\boldsymbol{d}$, content $\boldsymbol{c}$, and the spatial matrix $\boldsymbol{L}'$ to predict device level WiFi usage. Let $\boldsymbol{l}_i'$ be the i-th row of matrix $\boldsymbol{L}'$. Then the aggregated vector $\boldsymbol{z}_i'$ at location l_i can be expressed as

$$\boldsymbol{z}_i' = W_z' [\boldsymbol{d}; \boldsymbol{c}; \boldsymbol{l}_i'], \ i = 1, \ldots, m, \tag{8}$$

where $[;]$ denotes concatenation and $W_z' \in \mathbb{R}^{l_x \times (2l + l_l)}$ is the learnable parameter. Then the final prediction block can be presented as:

$$g_\phi(d, c) = \mathrm{MLP}(\{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_m\}), \tag{9}$$

where $\mathrm{MLP}(\cdot)$ stands for another Multilayer Perceptron with two fully connected layers for output reshaping, i.e., compressing $(2l + l_l)$ into 1 dimension. L2 distance is used as the objective function here to learn the model parameters for device dimension:

$$\ell_l(\theta) = \sum_{d \in \mathcal{D}} \sum_{c \in C} \|g_\phi(d, c) - y_{d,c}\|_2^2. \tag{10}$$

3.7 Fusion of Two Dimensions

The predicted contents from the user dimension and device dimension are then combined, which is used as the final results of *MIMU* on a certain day, i.e., all usage from dominant users in the user dimension output and all usage form non-dominant users in the device dimension output. In detail, define the usage of content c visited by user u via device d from the user dimension as $f_\theta(d, c, u)$, and the usage of content c which is visited by all other non-dominant users via device d as $g_\phi(d, c)$. Then the overall usage $y_{d,c}$ of content c on device d can be expressed as:

$$y_{d,c} = \sum_{u \in \mathcal{U}_d} f_\theta(d, c, u) + \alpha g_\phi(d, c), \tag{11}$$

where $\mathcal{U}_d$ represents all the dominant users.

4 EVALUATION

4.1 Evaluation Methodology

4.1.1 Evaluation Data Management. In this project, we work with Shenzhen transportation committee and Huashi WiFi company, who operate all buses and all bus WiFi services in Shenzhen, respectively. They provide us with access to the log data to enhance the commercial values of bus WiFi service. Due to the large size of our bus WiFi data, the comprehensive data cleaning, pre-processing, and modeling are implemented on a high-performance cluster with Spark. The detailed configurations are given as follows: (i) 2 HP machines with 2 Tesla K80c each; (ii) 2 Dell machines with 4 Tesla K80c each.

4.1.2 Data Sets. The evaluation is implemented upon the two-month-long bus WiFi data from Shenzhen, Wuxi, and Nanjing. For the majority of the evaluation, we use Shenzhen as an example to show the effectiveness of *MIMU*; multiple cities are considered in the multi-city evaluation. The heat map of bus WiFi usage across four cities in China is shown in Fig. 16, where the red color denoting heavy bus WiFi usage. Our evaluation is based on three of them, i.e., Shenzhen, Wuxi, and Nanjing. Regarding the user dimension implementation, for each user, we extract content visiting data, including timing, connected device, and visiting contents (*HTTP* and *Portal*). For device dimension implementation, we extract device level content visiting records, including the visiting timing, device ID, and contents. All the data are further formatted into inputs clarified in Section 3.

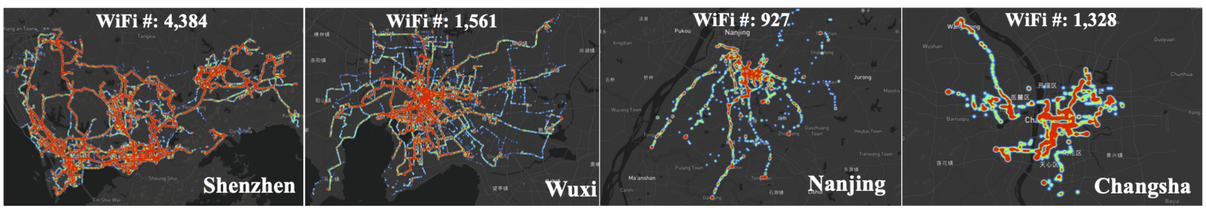

Fig. 16. Distribution of WiFi usage across four cities in China, the red color denotes locations with heavier bus WiFi data usage compared to the green/blue color.

4.1.3 Evaluation Setting. 1) Cross-validation and label selection. The evaluation is based on the cross-validation scheme. For both user dimension and device dimension, the previous $N-1$ days data are considered as training data for the model, and the last (N^{th}) day data are considered as the test data. For the user dimension, *userID*, *deviceID*, *contentID*, and *timing* are used as inputs, and the corresponding content usage is used as the label. For the device dimension, the way we implement cross-validation and label selection is the same as the user dimension except there is no *userID* in the inputs. The contents prediction results from the two dimensions are joined as the final prediction result. 2) Parameter setting. We implement *MIMU* with *PyTorch*. In both dimensions, we embed the identifiers into $\mathbb{R}^{128}$, the time features into $\mathbb{R}^{9}$, and location features into $\mathbb{R}^{64}$. For simplicity, we use the same dimension for the output of the attention module where we grid search the combinations. The model performs the best with dimension 64. The trade-off parameter α in Eq. 1 is empirically set in range (0.7, 1.2). The default learning rate is set as 0.001, and the batch size is set as 32.

4.1.4 Evaluation Metrics. For *MIMU* and all other baselines, we use one metric to measure the accuracy of our model, i.e., *accuracy* based on Mean Absolute Percentage Error (MAPE), defined as:

$$MAPE_{pred,d_i} = \frac{1}{N}\sum_{j=1}^{N}\frac{|usage_{d_i,c_j,pred}-usage_{d_i,c_j,real}|}{usage_{d_i,c_j,real}} \tag{12}$$

$$Accuracy_{pred,d_i} = 1 - MAPE_{pred,d_i} \tag{13}$$

Here the $usage_{d_i,c_j,pred}$ means the total usage of content c_j on device d_i, e.g., the predicted value $usage_{d_i,c_j,pred}$ is 2 while the $device_i$ may only access the content in the real world once ($usage_{d_i,c_j,real}$=1). $MAPE_{pred,d_i}$ means the average MAPE of usage prediction on device d_i. As aforementioned, we aggregate the usage from the user dimension usage and the bus device dimension to have the final predicted results at the device dimension for evaluation. All the final results are evaluated on a daily basis, e.g., we compare the predicted contents associated with the device on a certain day to the usage of real contents on this day.

4.2 Baselines

We also compare our results to multiple state-of-the-art methods, all these methods are implemented at the aggregate device level without considering users' characteristics, i.e., all inputs are formatted as visiting records focusing only on devices without any user information, such as $[device, content, usage]$. The training and test data are formatted and split in the same way as for all the methods.

- **Naive**: For the naive model, we use the most frequently visited content as the predicted content, and use the average visiting times as the predicted visiting frequency. This baseline is to testify whether a naive statistical model can achieve a satisfying performance.
- **LSTM** [24]: We implement a predictor by LSTM on the device dimension without any contextual information as one baseline. Specifically, we only use $N-1$ days' content sequences as inputs and N_{th} day's contents as labels.
- **NMF** [23]: NMF (Neural Collaborative Filtering) is a method considering neural networks into traditional collaborative filtering. Non-linearities are embedded into the framework to represent the relationship between users and items [23]. We treat devices as "users" and contents as "items" to implement the model.
- **DeepST** [47]: Considering that our data can be formulated into spatiotemporal data, e.g., records with time and location information. We also implement the DeepST prediction method proposed in [47].

4.3 Evaluation Results

4.3.1 Overall Results. Consider all the predicted results are based on devices, i.e., for each device and a certain day, we have a predicted result. We show the accuracy of our *MIMU* and baselines in Fig. 17. We can see that the naive method does not have a good performance, but *MIMU* achieves the most satisfying performance in terms of average accuracy and accuracy distribution. The average accuracy of all devices is 72.15%, and the standard deviation (STD) of the accuracy is 27.57%. DeepST also shows good performance with an average accuracy of 60.88% and STD of 28.93%, while the other two baselines perform worse regarding average accuracy.

The main reason our framework performs better than the rest baselines is that we have separate consideration towards user dimension and device dimension. Specifically, for DeepST implementation, content usage data are formatted for all devices as a spatial-temporal 2D matrix, consistent with the setting in DeepST [47]. For each day, a 2D matrix is calculated where one dimension represents all the devices and the other dimension represents all the contents in consideration; then each element $m_{i,j}$ in the matrix shows the usage of $content_j$ on $device_i$. The difference between such a structure and our *MIMU* can then be concluded: 1) In the bus WiFi setting, one of the advantages of DeepST, the spatial correlation or closeness of geographic partition, is lost since there is no such direct correlation among different bus WiFi devices; 2) The spatial similarity and difference of usage across different bus lines (as shown in Fig. 14) is missing since there is no consideration in relationships across different geographical locations in DeepST, e.g., whether the two regions have the same function; 3) The user related information such as behavior regularity is completely missing in DeepST. Different from the human mobility tasks in DeepST, the majority of the bus WiFi usage comes from a small group of dominant users. Modeling the dominant users helps the overall usage inference. Another thing that can be observed in the results is that the accuracy varies across different devices, we believe it is brought by the diverse usage distribution over all devices. We also show that the current accuracy could still be beneficial to our content caching application later.

Moreover, we show the performance of *MIMU* with only device dimension modeling and user dimension modeling in Fig. 18. MIMU-D means modeling only on the device dimension and MIMU-U means modeling only on the user dimension. We can observe an accuracy decrease from *MIMU* since we miss the combination of both dimensions. Besides, the device level modeling alone performs worse than the user dimension modeling, which may be caused by the fact that users visit different devices across different days, making usage patterns inconsistent on device dimension.

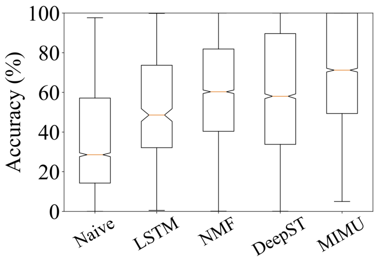

Fig. 17. Overall Performance

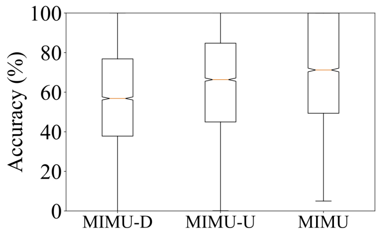

Fig. 18. Dimension Ablation

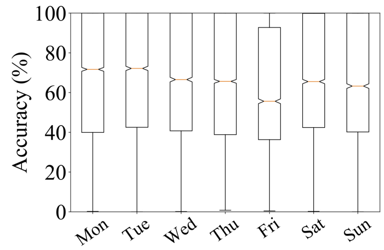

Fig. 19. Day of Week

4.3.2 Contextual Information. We are also interested in the effects brought by some contextual information given our two-month-long data set, i.e., the day of the week. By training and testing *MIMU* by data from different days, e.g., data all from Mondays, we show the performance of *MIMU* across the different days of the week in Fig. 19. We can see that there is no clear temporal pattern as studied by general data analysis, e.g., higher dynamics on weekends [37]. However, we can see that the accuracy of Fridays is lower than on other days. By analyzing our data set, we find that the bus WiFi usage dynamics (entropy) of Fridays is a little higher than the rest days of the week.

Moreover, we also test the performance of *MIMU* by different data proportions, i.e., using the different percentages of training data. The results are shown in Fig. 20. We can observe the general trend of increasing accuracy as the data proportion increases, suggesting that more data can achieve a better prediction performance.

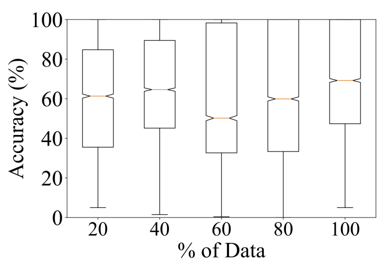

Fig. 20. Data Proportion Impacts

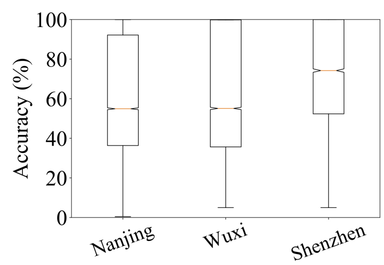

Fig. 21. Multiple Cities Evaluation

4.3.3 Multi-City Evaluation. Considering our major investigation and prediction are based on data from one city, one may be concerned with the generalizability of *MIMU*. Thus we analyzed data set from five other cities, i.e., Beijing, Shanghai, Changsha, Nanjing, and Wuxi. However, there are only 45, 13, and 109 active users in Beijing, Shanghai, and Changsha from Huashi company, so we focus on the other two cities. Due to the page limit of the paper, we directly present the final performance of *MIMU* across these cities. The details are in Fig. 21. We can see that *MIMU* achieves poorer performance in the other two cities, i.e., Nanjing and Wuxi, compared to Shenzhen. We argue the main reason behind this is the size of the data in different cities. Specifically, there are only 14,818 and 32,189 users in Nanjing and Wuxi, equivalent to around 10% and 20% of users number in Shenzhen.

4.4 Content Caching Application

Once accurate usage prediction is achieved, multiple applications can be developed based on it, such as content caching, commercial advertisement casting, and bus WiFi resource rebalancing, etc. Here we address the caching application with two approaches, i.e., daily based approach and iterative online approach.

4.4.1 Approach 1: Daily Content Caching. The hard disk in the WiFi device provides us with a hardware basis for content caching. The caching based on usage prediction can save data if the operator downloads all the predicted contents on the device in advance via cabled Internet access. Data transmission speed is also an important reason for considering a caching scheme based on bus WiFi usage prediction since the data transmission speed varies under different cases in the bus WiFi scenario. The speed of downloading Intranet contents is much faster than downloading directly from WiFi provided via a 4G SIM card. Reports about Huashi WiFi show that downloading directly from WiFi can only achieve a speed of 100-500 KB/s while downloading from stored contents can obtain a speed of 3-5 MB/s, which is much faster than accessing the Internet directly [25]. Thus, caching contents on the hard disk can be beneficial for both sides, saving data traffic for service providers and improving user experience due to faster transmission speed.

We also evaluate the performance on the content storage application by the metric Saving Rate (Saving) at the device level, defined as the ratio of accurately predicted data consumption out of all the data consumption of a device in a day.

$$Saving\ (\%) = \frac{DataTraffic_{predicted}}{DataTraffic_{all}} \tag{14}$$

The results is shown in Fig. 22. We can see that 80% of devices can achieve a saving rate lower than 31%, also meaning there are 20% of devices which can achieve a saving rate higher than 31%. Besides, we show the hit rate and miss rate of *MIMU*, i.e., for each device, the percentage of correct content prediction out of all contents as hit rate and the incorrect prediction as miss rate. The results are shown in Fig. 23. For 60% of devices, the hit rate is around 32%, while the miss rate is around 44%. Moreover, we also show the real data traffic saved based on the prediction results in Fig. 24. We can observe a similar distribution as the usage distribution among users, i.e., a small group of devices achieve a high saving amount. Specifically, 20% of devices can save more than 51 MB for a day and 20% of bus lines can save more than 870 MB for a day. We argue that the number may seem trivial in the beginning, while the cost can accumulate to a huge number for long running time scenarios. For example, a single bus line with 870 MB data saving per day can save up to 9,303 $(365 * 30 * 870/1024)$ RMB for a year [10].

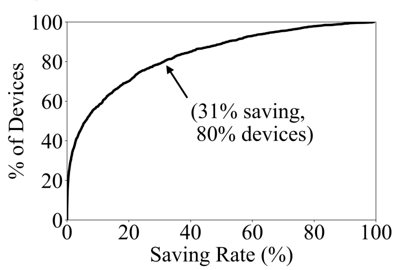

Fig. 22. Saving (%) CDF

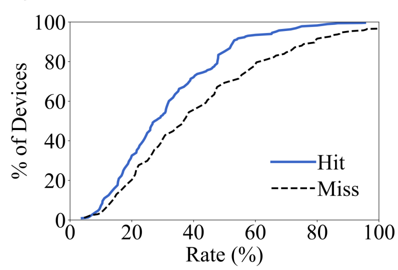

Fig. 23. Hit & Miss

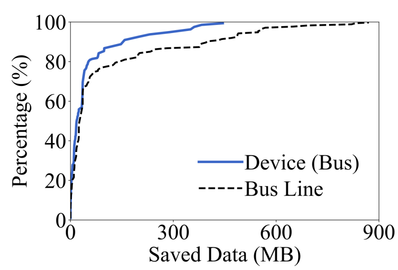

Fig. 24. Traffic Saving

To demonstrate the effectiveness of caching scheme based on our prediction, we compare our caching results to two classic caching methods, i.e., Least Recently Used (LRU) and Least Frequently Used (LFU). These two baselines are implemented by setting the device storage disk as the cache and updated on a daily basis. The results are shown in Fig. 25, where the distributions of saved data traffic amount across different devices are compared. We can observe a better performance of *MIMU* than LRU and LFU. Also, LFU performs slightly better than LRU since LFU achieves a longer tail in the distribution, meaning there are several devices saving a large amount of data.

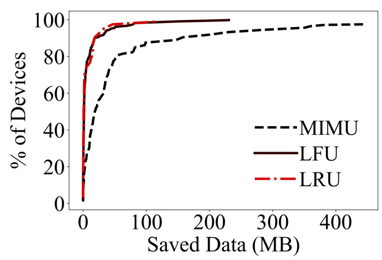

Fig. 25. Caching Comparison

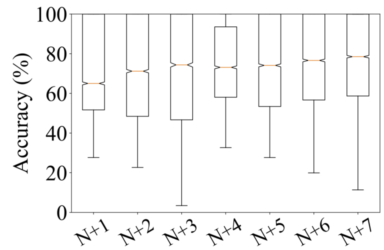

Fig. 26. Saving (%) CDF

4.4.2 Approach 2: Iterative Online Content Caching. In practice, data can be continuously collected and consumed from user activity on an iterative daily basis, which requires the system to handle model training and prediction in an online paradigm. However, it is both space inefficient to save all the accumulated data in the system and computationally expensive to retrain the model from scratch every day. To the end, we consider an online caching system that only stores a small portion of distilled history data and incrementally updates model parameters using the mix of history and newly arrived data.

Specifically, the online caching process of our system is illustrated in Fig. 27. On the n-th day (t_n), the system collects the data from each user as usual and store them in the storage together with the history data (D_n). The recently collected data are exploited to incrementally train the new model f_n based on the previous version f_{n-1}. Since this new training set is smaller than the whole accumulated dataset, the training process is efficient and can be regarded as a fine-tuning step to capture user new patterns without forgetting the previous ones. Once the training procedure is completed, the model f_n can be used to make a prediction, and simultaneously to distill a small set of data from D_n. The small distilled subset consists of useful historical information and will be used to update the model in the future t_{n+1}.

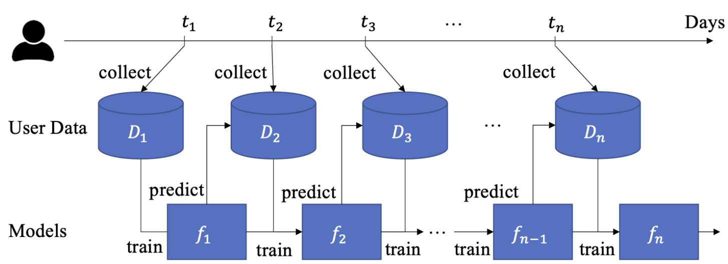

Fig. 27. Illustration of Online Caching Process

To testify the effectiveness of online caching, we use the first 40 ($N = 40$) days data for the model offline training and fine-tune the model during the next day ($(N + 1) - th$ day). Then for $(N + 2) - th$ day, we fine-tune the model by the newly coming data from the $(N + 1) - th$ day. By this incremental training, we emulate the process of online content caching. The evaluation results based on data in Shenzhen are shown in Fig. 26. We can see that the average accuracy generally keeps increasing during the continuous online content caching process, potentially leading to a higher hitting rate and data saving rate.

5 DISCUSSIONS

In this section, we provide a few discussions about our work, including the limitations, lessons learned, data management, data privacy issue, and more potential applications.

5.1 Limitation & Future Work

Limitation: 1) In this work, we only leverage bus WiFi data from one WiFi service provider to study the operational patterns and user behaviors. In practice, different providers may have different operation modes, e.g., users may need to install an app before connecting bus WiFi, and there are many advertisements, which may have an impact on users' usage patterns. 2) For the prediction modeling, we did not consider derived information conceived by the content data, e.g., what kind of topic the content is about, is there any correlation in terms of semantic meaning unveiled by the visited contents, which can be further utilized for a more personalized prediction. With increasing services and applications like live streaming, predicting a correct visiting URL may not guarantee a satisfying performance on data traffic saving, e.g., a live streaming website may be broadcasting different contents all the time, making caching less useful. It can be potentially addressed by a more sophisticated data collection approach at the service provider level, while in this paper we focus on the network provider level. 3) We filter the partial users and devices for the modeling implementation, while it remains an open question for processing data with long-tailed distribution [28], e.g., users with few historical records in our case, which is out of the scope of this work. 4) Another limitation is the newly coming URL to the system may deteriorate the performance. Meanwhile, we argue that this work can be extended to incorporating new URLs in either the training stage or inference stage: i) For the training stage, in addition to the labeled URLs (seen in $N-1$ days), one can also merge a large number of unlabeled URLs into the training set. This problem is similar to the cold-start problem in recommendation [33] where newly-arrived items have very little interactions with users. One solution is to leverage the features extracted from the URLs and content-based filtering algorithms [34] to make predictions. Recent advanced models for cold-start recommendation using graph neural networks [29] can also be adapted to this URL prediction problem. ii) For the inference stage, the problem can be cast to the inductive setting by semi-supervised learning (SSL) [49]. That is during the inference step, unseen URLs will also be utilized as the candidates in making predictions. In this case, one may adopt some off-the-shelf SSL techniques to handle the issue such as a graph-based approach [20] or a perturbation-based approach [39].
Future Work: 1) In the next step, we will try to do a comprehensive multi-city investigation, i.e., looking into the bus WiFi operational and usage pattern across multiple cities. Built upon the pattern investigation, we plan to study the transfer learning problem regarding different inputs distribution across different cities, which can provide more insights for fellow researchers and commercial public transport WiFi operators. 2) An in-depth investigation into the visiting content on bus WiFi devices may enhance the performance of predictors and provide more information about users' interests, i.e., topics revealed by semantics may be more general to understand and easier to predict. Applications based on user interests may benefit the operators more.

5.2 Lessons Learned and Insights

We also summarize the insights found in this work:

- **Usage Spatiotemporal Similarity:** We find that usage closer in temporal dimension and spatial dimension shows higher similarity, as supported by Fig. 11 and Fig. 15, i.e., bus WiFi visiting contents tend to be similar if visiting time is close or visiting location is close.
- **Diverse Usage Dynamics:** We find that individual bus WiFi usage dynamic is lower than device level bus WiFi usage (Fig. 8 and Fig. 9), denoting the different predicability.
- **Usage Participation:** Only around 25% of users of all bus WiFi users visit bus WiFi contents repeatedly, around 7% of users access the bus WiFi contents for more than 10 days during the two-month period.
- **Unbalanced Usage:** As shown in Fig. 2 and Fig. 3, the bus WiFi usage across all users is extremely unbalanced, 10% of users contribute more than 90% of WiFi data traffic. It is similar for different bus lines, e.g., one bus line can account for 12% of all the data consumed in the city.

5.3 Ethic and Privacy Protection

All the WiFi data used in this project are legally collected by the service providers and city governments. For example, GPS data from bus networks are collected by the transportation committee who owns and manages all buses in Shenzhen. The WiFi data collected by WiFi service providers and all WiFi users were informed about data collection and agreed (by selecting a WiFi usage agreement before connection) to provide their usage data to improve service performance, which is the crucial goal of this paper. The key challenge in bus WiFi data processing is to protect the privacy of users and ensure the utility of the models at the same time. To help us understand the user behavior anonymously without privacy concerns, the company staff at the bus WiFi company has removed or replaced all private personal information from the WiFi log data, e.g., the information of users' cellphones are removed, and user IDs are replaced by random strings, the authors do not have access to the raw data and will never reveal the identity of each user. As a result, our work is approved by the institutional IRB.

5.4 More Applications

- **Commercial App Promotion:** We find online game web pages are one of the most popular visiting web pages by the bus WiFi users. The reason may be that most young passengers will use bus WiFi to download and play games when they are on buses, so there is potential commercial value for bus WiFi operators to add some game websites on the WiFi login interface to attract more bus passengers and increase their revenue via collaboration with game companies.
- **WiFi Sharing:** Another possible application is to share bus WiFi to nearby commercial vehicles, such as a taxi. Considering the dense distribution of buses and taxis, we consider sharing bus WiFi under the circumstances, including when vehicles are stuck in the traffic jam, waiting for the traffic light, etc. Under such cases, passengers on both buses and taxis could make full use of bus WiFi when vehicles are close to each other. While this requires further operations such as simplification of the connecting process, etc.

5.5 Public Data Access

Large-scale commercial bus WiFi log data are hard to obtain due to various real-world issues. As the initial step, we commit to release our data sets across the time period: from Feb 1, 2017 to Feb 7, 2017. Four different kinds of log data will be released, including device location data, user connection data, and visiting topic data. All data sets will be processed into the dictionary data structure with details as follows. The data link is: https://www.cs.rutgers.edu/~dz220/data.html.

- Device location data: key: (*deviceID, date, ToD*), value: *gridID*. The *deviceID* and *gridID* are hashed into integer values. *ToD* (time of the day) is given at the granularity of 5 minutes. *gridID* is given at the granularity of $300m \times 300m$.
- User connection data: key: (*userID, date, ToD*), value: *deviceID*. *userID* is hashed by integer encoding. This data set reveals which user connects to which device at a given time.
- Visiting topic data: key: (*userID, date, ToD*), value: *topicID*. *topicID* is processed by integer encoding, corresponding topic names will also be provided for potential semantic investigation.

6 RELATED WORK

Lots of researches have been done regarding WiFi networks, generally, we can classify them into small-scale and large-scale according to numbers of devices (AP) involved, here we use 100 as the threshold to distinguish two categories; it can also be classified into stationary and mobile by the mobility of devices, i.e., whether they are set as fixed or mobile.

For the small-scale and stationary category, studies utilize small-scale implementation of networks to address related issues and applications. In [19], a family of handoff schemes based driven by clients are designed and

Table 2. Summary of WiFi Networks Related Research

	Stationary	Mobile
Small-scale(≤100)	[19] [21] [27] [30] [3]	[22] [31] [13] [4] [8] [48] [35]
Large-scale(>100)	[32] [14] [7] [41]	Our paper

testified by 4,000 users networks. WiFi working as an offloading method is briefly introduced in [21]. A more detailed study of 3G network data offloading via WiFi is given in [27], where results show that WiFi offloads 65% of mobile data and saves battery power by 55%, even larger gain is achieved by longer data transfer delay. WiLocator [30], as a tool for bus tracking and bus arrival time prediction, uses WiFi devices and its associated Signal Voronoi Diagram (SVD) to provide accurate bus arrival time prediction. Moreover, in [3], researchers focus on the negative impacts on users on buses brought by WiFi devices in bus stops, quantitative results and reasons behind them are also revealed.

For the small-scale and mobile category, researches are usually based on a few devices installed on vehicles. Hare *et al.* [22] deploy WiRover system on 5 public buses, data analyses based on it presents user and data usage characteristics, they also propose several schemes for network optimization. WiFi connections between base stations and vehicles are studied in [31] by measuring a testbed named VanLan. It specifically focuses on the connection performance when vehicles coming in and out of the range of base stations; and claims that the poor performance under this case is caused by urban radio environment rather than the motion of vehicles. In [13], researchers develop new data delivery and handoff schemes by predicting users' mobility and WiFi using patterns. Experiments show that new strategies can significantly improve download performance. Measurements from two testbeds are used to design an opportunistic protocol named ViFi, which doubles both TCP transfers and VoIP sessions [4]. A lot of detailed measurement results are given in [8] to suggest that WiFi networks providing intermittent connectivity can be useful in lots of applications. Also, in [48], traces from WiFi nodes in buses are used to analyze the inter-contact of bus pairs and their impacts on DTN (Delay Tolerant Network) routing performance. However, these small WiFi networks cannot fully reveal the complexity and characteristics of city-scale bus WiFi networks.

For the large-scale and static category, researchers make use of large-scale fixed devices to carry studies. Birk *et al.* [7] present a measurement study of urban commercial mesh WiFi network with 250 devices and 1.7 million log entries, which reveals useful observations such as mesh devices on utility poles lead to the performance degradation. Cabernet, as a system using devices around the road to transfer data among vehicles, is introduced in [14]; two new components are added for data delivery improvement, i.e., reducing connection setup time and increasing throughput. Mota *et al.* [32] provide a measurement of 21,649 devices during bus routes in Paris, revealing some findings such as deployed WiFi could offload 30% data during 30 seconds in a set scenario. Shen *et al.* [41] propose BaG by using the number of bursts collected via WiFi probes to analyze group detection.

To our best knowledge, our work is the first comprehensive investigation to analyze usage patterns and infer future usage of a large-scale mobile bus WiFi system.

7 CONCLUSION

In this paper, we perform, to the best of our knowledge, the first comprehensive investigation on the operational patterns and usage patterns of the bus WiFi systems. Specifically, we conduct an in-depth case study by leveraging bus WiFi usage log data of 110k users, which were collected from WiFi devices deployed on 4,384 buses in the Chinese city Shenzhen. Based on the investigation, we propose a sophisticated two-dimension bus WiFi usage prediction scheme called *MIMU* to predict and cache future bus WiFi usage, which shows superior performance compared to state-of-the-art methods. Application based on the prediction results, i.e., content caching, is also evaluated. Lessons learned and future work are also discussed. Investigation from such a large-scale bus WiFi network unveils the unique patterns of bus WiFi service and user behavior, which can benefit fellow researchers and commercial mobile WiFi operators.

REFERENCES

[1] AALSMEER. 2017. Staying connected: why WiFi in public transportation is the future. https://www.mondicon.com/news/Why-Wifi-in-transportation-is-the-future.html.

[2] Mikhail Afanasyev, Tsuwei Chen, Geoffrey M Voelker, and Alex C Snoeren. 2010. Usage patterns in an urban WiFi network. *IEEE/ACM Transactions on Networking (TON)* 18, 5 (2010), 1359–1372.

[3] Sehyun Bae, Daehyun Ban, Dahyeon Han, Jiyoung Kim, Kyu-haeng Lee, Sangsoon Lim, Woojin Park, and Chong-kwon Kim. 2015. Streetsense: Effect of bus wi-fi aps on pedestrian smartphone. In *Proceedings of the 2015 Internet Measurement Conference*. ACM, 347–353.

[4] Aruna Balasubramanian, Ratul Mahajan, Arun Venkataramani, Brian Neil Levine, and John Zahorjan. 2008. Interactive wifi connectivity for moving vehicles. *ACM SIGCOMM Computer Communication Review* 38, 4 (2008), 427–438.

[5] Hugo Barbosa, Marc Barthelemy, Gourab Ghoshal, Charlotte R James, Maxime Lenormand, Thomas Louail, Ronaldo Menezes, José J Ramasco, Filippo Simini, and Marcello Tomasini. 2018. Human mobility: Models and applications. *Physics Reports* 734 (2018), 1–74.

[6] David P Blinn, Tristan Henderson, and David Kotz. 2005. Analysis of a Wi-Fi hotspot network. In *Papers presented at the 2005 workshop on Wireless traffic measurements and modeling*. USENIX Association, 1–6.

[7] Vladimir Brik, Shravan Rayanchu, Sharad Saha, Sayandeep Sen, Vivek Shrivastava, and Suman Banerjee. 2008. A measurement study of a commercial-grade urban wifi mesh. In *Proceedings of the 8th ACM SIGCOMM conference on Internet measurement*. ACM, 111–124.

[8] Vladimir Bychkovsky, Bret Hull, Allen Miu, Hari Balakrishnan, and Samuel Madden. 2006. A measurement study of vehicular internet access using in situ Wi-Fi networks. In *Proceedings of the 12th annual international conference on Mobile computing and networking*. ACM, 50–61.

[9] China Telecom. 2017. China Telecom 4G Data Plan Rate. https://www.189.cn/bj/support/tariff/ydtc/71929.html.

[10] China Telecom. 2017. China Telecom 4G Data Plan Rate. https://www.189.cn/ln/support/tariff/package/52346.html.

[11] E Clayirci and Ian F. Akyildiz. 2002. User mobility pattern scheme for location update and paging in wireless systems. *IEEE Transactions on Mobile Computing* 99, 3 (2002), 236–247.

[12] George E Dahl, Tara N Sainath, and Geoffrey E Hinton. 2013. Improving deep neural networks for LVCSR using rectified linear units and dropout. In *2013 IEEE international conference on acoustics, speech and signal processing*. IEEE, 8609–8613.

[13] Pralhad Deshpande, Anand Kashyap, Chul Sung, and Samir R Das. 2009. Predictive methods for improved vehicular WiFi access. In *Proceedings of the 7th international conference on Mobile systems, applications, and services*. ACM, 263–276.

[14] Jakob Eriksson, Hari Balakrishnan, and Samuel Madden. 2008. Cabernet: vehicular content delivery using WiFi. In *Proceedings of the 14th ACM international conference on Mobile computing and networking*. ACM, 199–210.

[15] Zhihan Fang, Boyang Fu, Zhou Qin, Fan Zhang, and Desheng Zhang. 2020. PrivateBus: Privacy Identification and Protection in Large-Scale Bus WiFi Systems. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 1–23.

[16] 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), 42.

[17] Monica Ferrari. 2017. Benefits of free on-board Wi-Fi for public transport. https://www.tanaza.com/blog/benefits-of-free-on-board-wi-fi-for-public-transport/.

[18] Sina Finance. 2016. Do you have WiFi access in your city? http://finance.sina.com.cn/roll/2016-11-17/doc-ifxxwmws3013474.shtml.

[19] Anastasios Giannoulis, Marco Fiore, and Edward W Knightly. 2008. Supporting vehicular mobility in urban multi-hop wireless networks. In *Proceedings of the 6th international conference on Mobile systems, applications, and services*. ACM, 54–66.

[20] Will Hamilton, Zhitao Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. In *Advances in neural information processing systems*. 1024–1034.

[21] Bo Han, Pan Hui, and Aravind Srinivasan. 2011. Mobile data offloading in metropolitan area networks. *ACM SIGMOBILE Mobile Computing and Communications Review* 14, 4 (2011), 28–30.

[22] Joshua Hare, Lance Hartung, and Suman Banerjee. 2012. Beyond deployments and testbeds: experiences with public usage on vehicular WiFi hotspots. In *Proceedings of the 10th international conference on Mobile systems, applications, and services*. ACM, 393–406.

[23] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural collaborative filtering. In *Proceedings of the 26th international conference on world wide web*. International World Wide Web Conferences Steering Committee, 173–182.

[24] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

[25] Huashi. 2017. Huashi ViFi Introduction. http://www.visionchina.tv/news_nei.asp?pid=1438.

[26] Coby Joseph. 2014. How providing Wi-Fi can increase mass transit ridership. http://thecityfix.com/blog/wifi-increase-mass-transit-transport-ridership-bus-metro-coby-joseph/.

[27] Kyunghan Lee, Joohyun Lee, Yung Yi, Injong Rhee, and Song Chong. 2013. Mobile data offloading: How much can WiFi deliver? *IEEE/ACM Transactions on Networking (ToN)* 21, 2 (2013), 536–550.

[28] Jialun Liu, Yifan Sun, Chuchu Han, Zhaopeng Dou, and Wenhui Li. 2020. Deep Representation Learning on Long-tailed Data: A Learnable Embedding Augmentation Perspective. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2970–2979.

[29] Siwei Liu, Iadh Ounis, Craig Macdonald, and Zaiqiao Meng. 2020. A Heterogeneous Graph Neural Model for Cold-Start Recommendation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2029–2032.

[30] Wenping Liu, Jiangchuan Liu, Hongbo Jiang, Bicheng Xu, Hongzhi Lin, Guoyin Jiang, and Jing Xing. 2016. WiLocator: WiFi-sensing based real-time bus tracking and arrival time prediction in urban environments. In *Distributed Computing Systems (ICDCS), 2016 IEEE 36th International Conference on*. IEEE, 529–538.

[31] Ratul Mahajan, John Zahorjan, and Brian Zill. 2007. Understanding WiFi-based connectivity from moving vehicles. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*. ACM, 321–326.

[32] Vinicius FS Mota, Daniel F Macedo, Yacine Ghamri-Doudane, and José Marcos S Nogueira. 2013. On the feasibility of WiFi offloading in urban areas: The Paris case study. In *Wireless Days (WD), 2013 IFIP*. IEEE, 1–6.

[33] Seung-Taek Park and Wei Chu. 2009. Pairwise preference regression for cold-start recommendation. In *Proceedings of the third ACM conference on Recommender systems*. 21–28.

[34] Michael J Pazzani and Daniel Billsus. 2007. Content-based recommendation systems. In *The adaptive web*. Springer, 325–341.

[35] Bozhao Qi, Lei Kang, and Suman Banerjee. 2017. A vehicle-based edge computing platform for transit and human mobility analytics. In *Proceedings of the Second ACM/IEEE Symposium on Edge Computing*. ACM, 1.

[36] Zhou Qin, Fang Cao, Yu Yang, Shuai Wang, Yunhuai Liu, Chang Tan, and Desheng Zhang. 2020. CellPred: A Behavior-aware Scheme for Cellular Data Usage Prediction. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 1–24.

[37] Zhou Qin, Zhihan Fang, Yunhuai Liu, Chang Tan, Wei Chang, and Desheng Zhang. 2018. EXIMIUS: A measurement framework for explicit and implicit urban traffic sensing. In *Proceedings of the 16th ACM Conference on Embedded Networked Sensor Systems*. 1–14.

[38] Zhou Qin, Yikun Xian, and Desheng Zhang. 2019. A neural networks based caching scheme for mobile edge networks. In *Proceedings of the 17th Conference on Embedded Networked Sensor Systems*. 408–409.

[39] Antti Rasmus, Mathias Berglund, Mikko Honkala, Harri Valpola, and Tapani Raiko. 2015. Semi-supervised learning with ladder networks. In *Advances in neural information processing systems*. 3546–3554.

[40] John P Rula, James Newman, Fabián E Bustamante, Arash Molavi Kakhki, and David Choffnes. 2018. Mile high wifi: A first look at in-flight internet connectivity. In *Proceedings of the 2018 World Wide Web Conference*. International World Wide Web Conferences Steering Committee, 1449–1458.

[41] Jiaxing Shen, Jiannong Cao, and Xuefeng Liu. 2019. BaG: Behavior-aware Group Detection in Crowded Urban Spaces using WiFi Probes. In *The World Wide Web Conference*. ACM, 1669–1678.

[42] Rajkarn Singh, Marco Fiore, Mahesh Marina, Alberto Tarable, and Alessandro Nordio. 2019. Urban Vibes and Rural Charms: Analysis of Geographic Diversity in Mobile Service Usage at National Scale. In *The World Wide Web Conference (WWW '19)*. Association for Computing Machinery, New York, NY, USA, 1724–1734. https://doi.org/10.1145/3308558.3313628

[43] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*. 5998–6008.

[44] XinhuaNet. 2017. Beijing: The Free Bus WiFi is Gone. http://www.xinhuanet.com/2017-08/19/c_1121508023.htm.

[45] XinhuaNet. 2019. Reports of Cellular Data Rate from Major Cellular Operators. http://www.xinhuanet.com/fortune/2019-07/30/c_1124813473.htm.

[46] 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.

[47] Junbo Zhang, Yu Zheng, and Dekang Qi. 2017. Deep spatio-temporal residual networks for citywide crowd flows prediction. In *Thirty-First AAAI Conference on Artificial Intelligence*.

[48] Xiaolan Zhang, Jim Kurose, Brian Neil Levine, Don Towsley, and Honggang Zhang. 2007. Study of a bus-based disruption-tolerant network: mobility modeling and impact on routing. In *Proceedings of the 13th annual ACM international conference on Mobile computing and networking*. ACM, 195–206.

[49] Xiaojin Zhu and Andrew B Goldberg. 2009. Introduction to semi-supervised learning. *Synthesis lectures on artificial intelligence and machine learning* 3, 1 (2009), 1–130.