Understanding User Behavior in Car Sharing Services Through The Lens of Mobility: Mixing Qualitative and Quantitative Studies

GUANG WANG, Department of Computer Science, Rutgers University, USA
HARSH RAJKUMAR VAISH, Department of Computer Science, Rutgers University, USA
HUIJUN SUN, School of Traffic and Transportation, Beijing Jiaotong University, China
JIANJUN WU, State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, China
SHUAI WANG, School of Computer Science and Engineering, Southeast University, China
DESHENG ZHANG, Department of Computer Science, Rutgers University, USA

Qualitative and quantitative user studies can reveal valuable insights into user behavior, which in turn can assist system designers in providing better user experiences. Car sharing (e.g., Zipcar and car2go), as an emerging App-based online shared mobility mode, has been increasing dramatically worldwide in recent years. However, to date, comprehensive user behavior in car sharing systems has not been investigated, which is essential for understanding their characteristics and promotion roadblocks. With the goal of understanding various facets of user behavior in online car sharing systems, in this paper, we performed a qualitative and quantitative user study by adopting a *mixed-methods approach*. We first designed an attitude-aware online survey with a set of qualitative questions to perceive people's subjective attitudes to online car sharing, where a total of 185 participants (68 females) completed the survey. Next, we quantitatively analyzed a one-year real-world car sharing operation dataset collected from the Chinese city Beijing, which involves over 68,000 unique users and over 587,850 usage records. We dissected this attitude-free dataset to understand the objective car sharing user behavior from different dimensions, e.g., spatial, temporal, and demographic. Furthermore, we conducted a comparative study by utilizing one-year data from other two representative Chinese city Fuzhou and Lanzhou to show if the obtained findings from Beijing data may be generalizable to other cities having different urban features, e.g., different city size, population density, wealth, and climate conditions. We also do a case study by designing a user behavior-aware usage prediction model (i.e., BeXGBoost) based on findings from our user study (e.g., unbalanced spatiotemporal usage patterns, weekly regularity, demographic-related usage difference, and low-frequency revisitation), which is the basis for car sharing service station deployment and vehicle rebalancing. Finally, we summarize a set of findings obtained from our study about the unique user behavior in online car sharing systems, combined with some detailed discussions about implications for design.

CCS Concepts: • **Human-centered computing** → *User studies*; *Ubiquitous and mobile computing*; *Empirical studies in HCI*.

Additional Key Words and Phrases: User behavior; shared mobility; car sharing; qualitative; quantitative

Authors' addresses: Guang Wang, Department of Computer Science, Rutgers University, 110 Frelinghuysen Road, Piscataway, NJ, 08854-8019, USA, guang.wang@rutgers.edu; Harsh Rajkumar Vaish, Department of Computer Science, Rutgers University, Piscataway, USA, harsh.vaish@rutgers.edu; Huijun Sun, School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China, hjsun1@bjtu.edu.cn; Jianjun Wu, State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China, jjwu1@bjtu.edu.cn; Shuai Wang, School of Computer Science and Engineering, Southeast University, Nanjing, Jiangsu, China, shuaiwang@seu.edu.cn; Desheng Zhang, Department of Computer Science, Rutgers University, NJ, USA, desheng.zhang@cs.rutgers.edu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2020 Association for Computing Machinery.
2474-9567/2020/12-ART156 $15.00
https://doi.org/10.1145/3432200

ACM Reference Format:
Guang Wang, Harsh Rajkumar Vaish, Huijun Sun, Jianjun Wu, Shuai Wang, and Desheng Zhang. 2020. Understanding User Behavior in Car Sharing Services Through The Lens of Mobility: Mixing Qualitative and Quantitative Studies. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 4, Article 156 (December 2020), 30 pages. https://doi.org/10.1145/3432200

1 INTRODUCTION

Mobile phones and applications have advanced our mobility modes in recent years, e.g., ride-hailing and car sharing [41]. As an innovative mobility mode [54], online car sharing has been experiencing rapid growth due to its environmentally-friendly nature and flexibility for use [35]. For example, the car sharing service provider car2go reached three million users within sixteen months, becoming the largest operator in the flexible car sharing sector [42]. Another world-leading car sharing provider Zipcar provides over one million users access to more than 12,000 vehicles in over 500 cities around the world [70]. It is estimated that over 2,095 cities worldwide had car sharing services, and about 15 million users were registered in 2016, and the number of users is predicted to be over 36 million by 2025 [43, 46]. With so many car sharing users, a comprehensive user behavior study is necessary to enhance the existing car sharing systems and improve user experiences.

An emerging mobility mode must have some unique features for providing complementary mobility services [61] to existing mobility modes. Hence, as a novel mobility mode, the unique operation strategy and user groups of car sharing may also make their users' usage behavior (e.g., weekly usage, pick-up and drop-off locations, and age groups of users) distinct from other mobility modes. Unfortunately, few existing works have been conducted to investigate the comprehensive user behavior in car sharing systems based on real-world data. Lacking data-driven observations makes it challenging for us to understand unique user behavior, practical challenges during the promotion process, and users' attitudes towards the innovative mobility mode. Hence, in this paper, we conduct the first qualitative and quantitative investigation to reveal comprehensive user behavior in car sharing systems with both attitude-free data and attitude-aware data from different spatiotemporal and demographic dimensions. Specifically, we had the following research questions (RQ):

- **RQ1:** How do people **perceive** online car sharing?
- **RQ2: Who** are more likely to use car sharing? **When, where, why, how long, how regular** do users usually use car sharing? How are the **revisitation** patterns?
- **RQ3:** Are there similar car sharing user behavior in **different cities** with distinct urban features?

To answer these questions, in this paper, we adopted a mixed-methods approach by combining small-scale online survey data and large-scale real-world car sharing operation data. The *attitude-aware* online survey data enables us to qualitatively learn people's *subjective* perception of online car sharing systems (**RQ1**). The large-scale *attitude-free* operation data provides us an opportunity to quantify the *objective* user behavior in car sharing systems (**RQ2** and **RQ3**). Since the goal of this paper is to understand comprehensive user behavior in the emerging shared mobility mode and explore some potential directions for the future work, we provide a set of findings obtained from our study, combined with some detailed discussions about design implications. To verify the usefulness of this user study work, we also do a case study by designing a usage prediction model based on findings from our user study (e.g., unbalanced spatiotemporal usage patterns, weekly regularity, demographic-related usage difference, and low-frequency revisitation).

To this end, our data analysis follows a highly exploratory approach, wherein we analyze the collected online survey data and car sharing operation data through lenses of user behavior that are highly relevant to the IMWUT community. The major contributions of this work are the following:

- To our knowledge, this work is the first comprehensive user study for car sharing systems based on a mixed-methods approach. We first qualitatively learned people's subjective perception of the online car sharing systems via an attitude-aware online survey. We then quantified objective user behavior in car sharing systems based on collected large-scale attitude-free operation data.

- We designed an online survey to answer the **RQ1**, which involved 185 participants. The attitude-aware survey questions can explicitly show some perceptions of users to online car sharing systems. The survey results show that convenience is the most important factor (63%) for users to use car sharing instead of prices. It also indicates users usually use car sharing systems for traveling (38%) instead of commuting (18%). Most participants (69.8%) will still use their private cars in a car sharing+autonomous driving era.
- With the attitude-free operation data, we answered the **RQ2** (e.g., "Who", "When", "Where", "Why", "How long", "How regular", "How revisitation" questions) with different metrics quantitatively. In addition, we explain possible causalities based on various correlation analyses. We found the high-income young (18 to 38 years old) males work in IT industry areas are more likely to use car sharing. For daily commuting usage, users usually pick up vehicles around 19:00 and return vehicles around 9:00. For long-time usage, users are more likely to pick up vehicles on Friday around 18:00. And we also found most user behavior (over 80%) is highly predictable. Some findings are also verified by our online survey.
- To answer the **RQ3**, we conducted a comparative study using the car sharing operation data from other two cities, i.e., the Chinese city Fuzhou and Lanzhou, which shows some findings are similar (e.g., predictability) and some findings are nongeneralizable (e.g., temporal patterns). Hence, we point that our study has several limitations about the generalizability of the findings due to limited data collection. To further generalize our findings, there should be comparative studies with other countries and cities.
- Based on findings from **RQ1** to **RQ3** (e.g., unbalanced spatiotemporal usage patterns, weekly regularity, demographic-related usage difference, and low-frequency revisitation), we do a case study by designing a user behavior-aware usage prediction model called BeXGBoost, which shows the significance of our user study work and also lays the foundation for car sharing service station deployment and vehicle rebalancing.

2 RELATED WORK

In this section, we review the related works from four aspects, i.e., user behavior, human mobility, urban computing and car sharing patterns.

2.1 User Behavior

There is a rich history of UbiComp/IMWUT research concentrating on user behavior analysis [63]. In recent years, with the increasing importance of user experiences in system design, we are witnessing an increasing number of research studies in the IMWUT community [2, 24, 26, 27, 33, 37, 40, 45, 47, 63, 64] to understand user behavior. Zhang et al. [63] tried to understand user behavior in group event decisions based on data collected from a mobile application, and they finally provided detailed novel insights in the event scheduling process of social groups. Srinivasan et al. [45] performed a qualitative user study in order to understand user behaviors on using an interactive rule selection tool, and some insights were gained for designing future rule selection systems. Iwamoto et al. [26] surveyed 60 people with an anonymous questionnaire to investigate the user behavior and found features that are effective to classify passersby. Jeong et al. [27] tried to understand smartwatch wearing behaviors of users from different aspects, e.g., temporal wearing behaviors and taking off behaviors, and their findings have the potential to provide diverse design implications for improving wearability. Page et al. [40] explored people's attitudes towards the Internet of Things and their adoption behavior based on 38 interviews. Mathur et al. [37] presented a mixed-methods longitudinal study to provide a holistic view of the smartphone usage behavior of Indian users, and their findings demonstrated the unique characteristics that are shaping the smartphone usage behavior of Indian users.

In summary, almost all research focusing on user behavior analysis tries to provide some new findings from data, which is also in line with our work.

2.2 Human Mobility

With the development of sensing and communication technology, an increase of ubiquitous devices, we have more and more approaches to understand human mobility. Numerous studies have been done to investigate human mobility in different mobility strategies [20–22, 49, 50, 52, 53, 67] based on vehicle GPS data. In addition, smart card data for fare collection in public transport systems are also utilized to analyze people's mobility in bus or subway systems [31]. With the promotion of bike sharing, human mobility in bike sharing systems has also been studied [54]. With smart devices, crowdsensing mobility data was utilized to understand people's mobility patterns in [8]. There are many innovative mobility strategies due to the development of smartphones and applications, and an increasing number of researchers focus on these emerging mobility modes. The for-hire strategy (e.g., Uber and Lyft) provides an effective mobility sharing mechanism to reduce traffic congestion and air pollution, and its mobility patterns have been investigated in [5]. Car rental patterns are also shown in [23, 38]. Distler et al. [19] measured post-immersion acceptance of autonomous mobility on demand by using a mixed-methods approach, and they showed the attitudes of participants. There are also some efforts [61] to investigate mobility patterns by integrating multi-source data, e.g., GPS data of heterogeneous vehicle fleets and smart card data. All of them advanced our understanding of urban mobility.

2.3 Urban Computing

Our work also lies in the emerging topic of urban computing [68], which has attracted great interests in the ubiquitous computing community in recent years. Urban computing aims to tackle different kinds of urban issues by using the data that has been generated in cities, e.g., GPS data, mobile cellular data [66]. Xiao et al. [55] aimed to estimate the similarity between users according to their GPS trajectories. Cao et al. [9] presented the first large-scale analysis of POI revisitation patterns to model the periodic behavior in human mobility. Cao et al. [10] proposed a novel approach to understanding crowd mobility in a metropolitan area by using mobile cellular accessing trace data of users, and some human mobility patterns have been discovered, e.g., commute patterns and spatial correlation rules. Li et al. [32] aimed to geographically mine the similarity between users based on their location histories by considering both the sequence property of people's movement behaviors and the hierarchy property of geographic spaces. Cao et al. [11] proposed a representation learning-based system called habit2vec to represent user trajectory semantics in vector space, which preserves the original user living habit information. The system is evaluated on social network data of over 123,000 users and the results show the effectiveness of clustering users with similar living habits. Vogel et al. [48] analyzed extensive operational data from bike-sharing systems to derive bike activity patterns, and the imbalanced bike distribution phenomenon was uncovered. Corcoran et al. [16] utilized the trip level data to investigate the spatiotemporal dynamics of a large public bicycle sharing system, specifically the effects of weather and calendar events on the geographic and temporal patterning of public bicycle use. Xu et al. [56] investigated a novel problem of detecting popular temporal modes in population-scale unlabelled trajectory data, and their experiments revealed insightful correlations between the popular temporal modes and individuals' social-economic status. Wang et al. [51] tried to optimize the electric bus charging problem with real-world operation data. Zheng et al. [69] tried to mine interesting locations and classical travel sequences in a given geospatial region based on multiple users' GPS trajectories. However, few existing works have been conducted to investigate user behaviors in car sharing systems, e.g., revisitation patterns. In addition, our work adopts a mixed-methods approach to investigate user behavior from both qualitative and quantitative perspectives, which is also rarely explored in existing urban computing works.

2.4 Car Sharing Pattern

There are also many works to investigate car sharing patterns in recent years. However, almost of existing car sharing works only focus on general patterns or coarse-grained patterns, which fail to capture user subjective attitudes and fine-grained user behaviors. There are typically two types of works on car sharing pattern analysis, i.e., survey-based investigation and operation data-based work. Becker et al. [1] conducted a survey to compare user groups and usage patterns of two kinds of car sharing services operating in Switzerland. Kopp et al. [30] designed a survey to obtain mobility data for travel behavior analysis, which measured behavior largely unaffected by the instrument. However, the real-world operation patterns have not been investigated in these papers, from which the comprehensive user behaviors are challenging to be captured. In recent years, due to the availability of real-world operation data, many works have also been done to study car sharing real operation patterns. Sprei et al. [44] studied general spatiotemporal patterns (i.e., travel time and travel distance) of the car sharing vehicles among the early adopters, and they also compare car sharing with other modes of transport. Although with a quantitative study, the paper also indicated that surveys and qualitative interviews should be conducted to identify and understand user attitudes and usage motivations. Our online survey compensates for this blank. Business models and national variations of car sharing in Europe were analyzed in [18], but fine-grained user behavior and usage patterns have not been studied in this work. Boldrini et al. [3] characterized demand and usage patterns in car sharing, but individual user behavior and qualitative investigation have not been studied.

Different from existing car sharing works, in this paper, we utilize a mixed-methods approach to understand user behavior in car sharing based on both attitude-free real-world operation data and attitude-aware survey data, which has the potential to reveal essential fine-grained mobility features of car sharing and users' attitudes and motivations, and help us for better system design and related applications, e.g., usage prediction, service station deployment, and vehicle rebalancing.

3 DATA COLLECTION AND METHODOLOGY

In this section, we describe our mixed-methods approach, which combined an online survey with a long-term real-world car sharing operation process. We first show how we did the online survey, and then we describe how we collected the car sharing operation data, coupled with a detailed data description.

3.1 Online Survey Design

We first conducted an online survey with a set of questions related to our main topics of interest to perceive people's attitudes to online car sharing. Following [65], we leverage a popular survey website in China, i.e., Wenjuan.com, to conduct our survey. Wenjuan is one of the largest survey companies in China, similar to Qualtrics in the US. We first created our survey in English. The native Chinese authors then translated it into Chinese.

In this work, we adopted the following two ways to recruit participants. (i) Our teammates sent the questionnaires to their friends directly via social media or emails, and their friends will also forward the questionnaires to other people. Each participant who finishes the questionnaire can receive a certain 3 RMB cash bonus from our teammates with the screenshot of the questionnaire. (ii) We share the questionnaire link in some Wechat groups (equivalent to groups on Facebook, where there are many known and unknown people in those groups), which makes our participants more diverse. To attract more people participating in our survey, we provide some monetary incentives in a Wechat Red Packet way, and each respondent will receive a random bonus from 0.1 to 15 RMB after they finish the survey. To attract more people to participate in our survey, we try to make our questions as short and as friendly as possible. There are 11 questions in the survey in total (including the attention check question). The first question asks participants if they have used carsharing services. For the next three questions, we ask participants some close-ended questions about their personal information, e.g., age, gender, and the cities they are in. To protect the privacy of participants, we set some age ranges for them to select instead of asking them their exact ages, e.g., 18-23 and 24-29. Each age range is about 5 years to 6 years, and the reason why we set the value starting from 18 is that it is the minimum age that a person can have a driving license in China. Then we set an attention check question to improve the data quality. In the second part of the survey, we ask them four semi close-ended questions about the reasons why they utilize car sharing and the potential reasons that will hinder them from using car sharing. In the last part, we ask two questions about their attitudes to the future autonomous car sharing services.

The survey was active on April 29, 2019, for one week. It took 70.6 seconds to complete on average (Median= 63, SD=36.7). At the end of the survey, we received 198 responses. We removed the respondents who were not attending to the survey (i.e., failed the attention check question, completed too quickly). Our final dataset includes 185 completed survey responses. Our sample was diverse in terms of demographics and geographic location. There are 68 female and 117 male participants ranged in age from 18 to over 50 (15% in 18-23, 69% in 24-29, 11.9% in 30-35, 2.7% in 36-40, 0.5% in 36-40, 0 in 46-50, and 1.1% in 50+).

The detailed results drawn from our online survey will be shown in Section 4.

3.2 Online Car Sharing Systems

After operating the car sharing system with a mobile App as shown in Figure 1 for a long time, our collaborators have collected a large-scale real-world car sharing operation dataset from over 100 Chinese cities, and they provide the data for us to improve their business intelligence. In this paper, we carefully select three representative cities, i.e., Beijing, Fuzhou, and Lanzhou to show our findings. In this subsection, we first introduce the car sharing system, and then we use Beijing as an example to describe the collected dataset.

Fig. 1. Mobile App

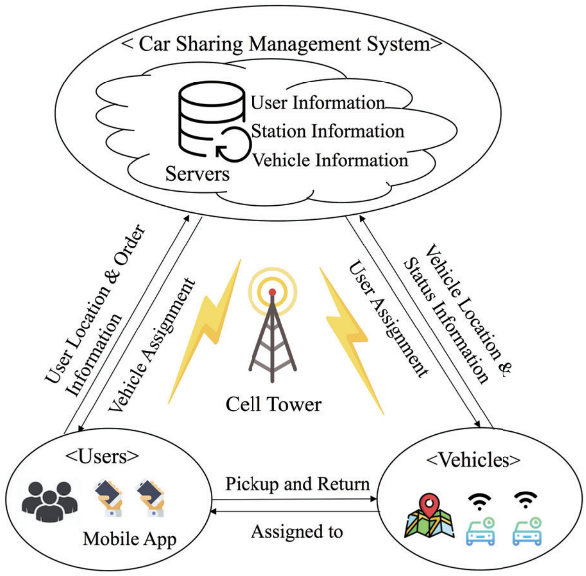

Fig. 2. Car sharing system

3.2.1 Car Sharing Infrastructure. Figure 2 shows our car sharing system. A car sharing vehicle usage cycle includes the following four steps:

- Step 1: Users register for using car sharing systems via our mobile App, and users' ages and genders are required when registering for security purposes.

- Step 2: A user can find a vehicle immediately or reserve a vehicle in advance and then pick it up within the given time.
- Step 3: Users drive the car sharing vehicles to their destinations.
- Step 4: Users park vehicles in dedicated parking spots, and then return the vehicles via the App. However, each parking lot only has limited parking spots for car sharing vehicles.

The real-time location information and order information of users are recorded and uploaded when they use the App. The real-time location and status information of vehicles are also periodically uploaded to our servers via communication devices. The transaction data are recorded when users return car sharing vehicles. All user data and real-time vehicle information are collected and stored in cloud infrastructure for management and analysis.

3.2.2 Car Sharing Usage Pricing. Four usage pricing packages are provided for car sharing services in Beijing. The four pricing mechanisms include the same distance-based mileage fee (i.e., \$0.28/mile) but with different hourly rental rates, which are shown in Table 1. The basic package is hourly, which means users pay based on how long they rent, and the price is \$1.76/hour. There are also three special packages, i.e., daily rate (\$17.6/day), nightly rate (\$4.4/night), and weekend package (\$26.4/weekend).

Table 1. Car sharing pricing packages

Pricing	Hourly Rate	Daily Rate	Nightly Rate	Weekend Package
Time	on hourly	24 hours	17:00-9:00	Friday 17:00 to Monday 9:00

3.2.3 Collected Operation Data. All car sharing vehicles in our dataset have the same vehicle type, which is an SUV model with a maximum driving range of 200 km, so they have similar vehicle properties and performance. There are three main data sources used in our investigation, and their details are shown below:

- **Vehicle Usage Data** include all users' vehicle usage records. Each usage record consists of 26 fields describing vehicle information, user demographic information, and order information, e.g., the order number, the user ID, age, gender, occupation, order time, vehicle pick-up and returning time, pick-up and returning station, and the vehicle ID, etc.
- **Transaction Fare Data** include a detailed cost record of each vehicle usage. Each transaction record contains 24 fields describing vehicle usage and discount information, e.g., the real payment, the travel distance, the travel time, the discounted cost, and the vehicle ID.
- **Car Sharing Service Station Data** describe the service station information, e.g., the station IDs, the station names, the GPS locations, and the number of parking spots in each station. There were 326 stations in Beijing operated by our collaborator in 2019.

4 QUALITATIVE RESULTS

In this section, we conduct a thematic analysis to show our semi-structural online survey results, which is a widely used qualitative research method to identify salient patterns or themes in the data [4]. The survey results potentially indicate some explicit attitudes of people to the innovative online car sharing.

The set of themes and codes resulting from our thematic analysis are presented in Table 2, where (P2) indicates the participant 2 and (All) means aggregate results. In this section, we introduce each code and theme in detail.

Theme 1. Were Participants Willing to Use Car Sharing? To understand people's general attitudes towards car sharing, we ask a question about the perception of car sharing after the basic demographic information questions, as we indicated in Section 3.1. As shown in Figure 3, we found about 43% of participants indicating car sharing is common in their cities, and about 16% of participants said that they could see car sharing in some areas of their cities, so we can infer that car sharing has been promoted in many Chinese cities. We also found that most people (84.3%) are willing to use car sharing services due to the reasons like convenience, as shown in Figure 4. "*I use car sharing because I do not have a private car, and it is convenient for me to use car sharing since there is a service station in the parking lot of my company.*" (P33).

Even though the concept and practice may have differences, the results show that car sharing has the potential to be accepted by most people and has penetrated in many Chinese cities.

Table 2. Results of our thematic analysis show four themes: (1) Willingness, (2) Purposes, (3) Barriers, and (4) Sustainability.

Theme	Code	Example
Were Participants **Willing** to Use Car Sharing?	Willing Due to Convenience	*"It is convenient for me to use car sharing since there is a service station in the parking lot of my company."* (P33)
	Not Willing Due to Some Practice Considerations	*"I do not want to drive by myself. I will use it if the shared vehicles are autonomous driving."* (P17)
What are The **Purposes** for Participants to Use Car Sharing?	Commuting and Business Affairs	28% of participants indicate they use car sharing for commuting or business affairs. (All)
	Leisure Activities	(i) *"Sometimes, I use car sharing vehicles to go to shopping malls since many it is convenient to return vehicles there."* (P12) (ii) *"I have used car sharing for road trips."* (P16) (iii) 38% participants utilize car sharing for traveling. (All)
	Special Circumstances	(i) *"Sometimes I want to drive, but I do not have a private car, so I use car sharing services."* (P33) (ii) *"If I am not in a hurry to get home, I will use car sharing."* (P25) (iii) *"I will compare the cost of car sharing and taxis. If the price of car sharing is lower than taking taxis, I will choose car sharing."* (P23) (iv) 26% of participants use car sharing for an emergency. (All)
What Would Be The **Barriers** that Prevent Participants from Using Car Sharing?	Inefficient Management Mode	(i) *"The procedure to use car sharing services is too complicated. I need to spend a lot of time to learn how to use it."* (P10) (ii) *"I find the current consumer protection system is unsound. It is not easy for us to make complaints."* (P7) (iii) 63% of participants think the vehicle availability will hinder their usage, 54% of participants think unavailable parking spaces is a major barrier. (All) (iv) A quarter think the price of car sharing is too expensive. (All)
	Vehicle Condition	(i) *"I am afraid that the vehicle enegry is not enough for me to reach my destination."* (P7) (ii) *"I do not know if the car sharing vehicle is safe since it may be used by many people for many times."* (P10)
	Personal Reasons	(i) *"I have a private car, so I rarely use car sharing services."* (P2) (ii) *"I do not want to drive by myself. I will use it if the shared vehicles are autonomous driving."* (P17) (iii) *"My driving skill is poor, so I do not drive."* (P5)
How Do Participants Think About The **Sustainability** of Car Sharing?	Optimistic	*"Car sharing is a promising mobility mode and it is a prototype of the future mobility."* (P3)
	Pessimistic	*"The current operation mode of car sharing may make it die soon."* (P7)
	Neutral	*"It is hard to say what will happen in the future, so it still needs time to validate."* (P6)

Theme 2. What are The Purposes for Participants to Use Car Sharing? Next, we designed two semi close-ended questions to understand why people use car sharing systems and what are the reasons that prevent them from using car sharing systems. Figure 5 shows the purposes of participants to use car sharing. We found most participants use car sharing for leisure activities, e.g., 38% of them used car sharing for traveling, which potentially indicates the highest usage may happen on weekends and holidays. *"Sometimes, I use car sharing vehicles to go to shopping malls since many it is convenient to return vehicles there"* (P12). *"I have used car sharing for road trips"*

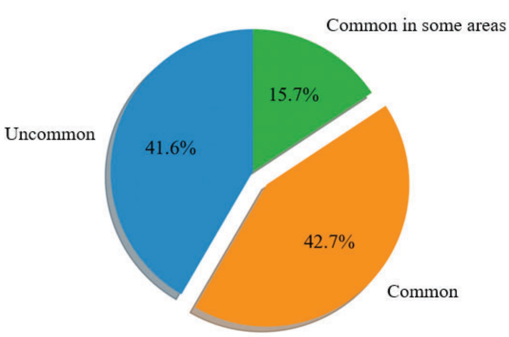

Fig. 3. Is it common to find car sharing in your city

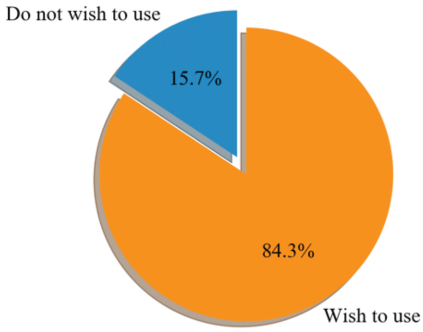

Fig. 4. Are participants willing to use car sharing

(P16). Many participants use car sharing in special circumstances. *"Sometimes I want to drive, but I do not have a private car, so I use car sharing services"* (P33). The price is also a consideration of users. *"I will compare the cost of car sharing and taxis. If the price of car sharing is lower than taking taxis, I will choose car sharing"* (P23). The second-most-popular purpose of using car sharing is for an emergency, which accounts for 26%. An interesting finding is that only 18% of participants use car sharing for commuting, which indicates car sharing may not be able to replace existing commuting ways. In addition, the results also show 10% of participants use car sharing for business affairs.

Based on participants' inputs, we found users will consider the cost and convenience when they use car sharing, so it is important for car sharing providers to provide convenient and cost-competitive services.

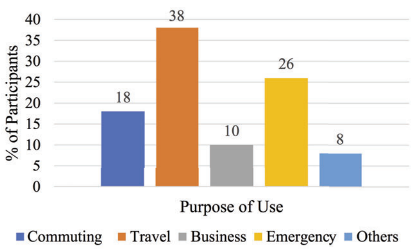

Fig. 5. Purposes to use car sharing

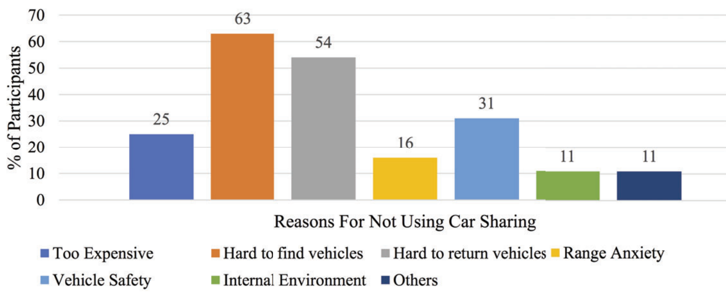

Fig. 6. Reasons that prevent participants from using car sharing

Theme 3. What Would Be The Barriers that Prevent Participants from Using Car Sharing? *Figure 6 shows the barriers that may prevent participants from using car sharing services. For this semi close-ended question,* we provide six pre-defined choices, and users can select multiple answers. Similarly, we also provide a blank space for participants to input their own reasons. We coded three categories of reasons, i.e., Inefficient Management Mode, Vehicle Condition, and Personal Reasons. We found that 63% of participants think that the vehicle availability will hinder their usage, followed by the vehicle returning (54%, this is caused by the limited parking spots in each service station). We found that both the two reasons are related to convenience, so we argue that convenience would be the most important factor to promote car sharing services. *"The procedure to use car sharing services is too complicated. I need to spend a lot of time to learn how to use it"* (P10). In addition, a key concern for participants is the vehicle condition, and we found 31% of participants worry about the vehicle safety. For example, *"I do not know if the car sharing vehicle is safe since it may be used by many people for many times"* (P10). Even though we think the cost would be a key concern, we found only 25% of participants think they will not use car sharing because of the cost, which potentially indicates that the prices of car sharing are acceptable for most people. Some participants will also not use car sharing due to their personal reasons, e.g., *"I have a private car, so I rarely use car sharing services"* (P2). In addition, many participants think they do not want to drive, so they do not use car sharing. *"I do not want to drive by myself. I will use it if the shared vehicles are autonomous driving"* (P17). Hence, we also investigate the attitudes of participants to future car sharing with autonomous vehicles.

Although it is a common belief that the future mobility would be shared autonomous driving [6], which is similar to car sharing with autonomous vehicles, we found less than one-third (30.3%) of participants will abandon their private cars and use car sharing only when car sharing vehicles are available everywhere, as shown in Figure 7. Most participants (46%) will still use their private cars only because of safety concerns. The remaining participants (23.8%) think they will use the car sharing with autonomous vehicles as an auxiliary mode. Hence, we argue that private cars are displaceable even though widely deploying large-scale autonomous car sharing vehicles under the current setting.

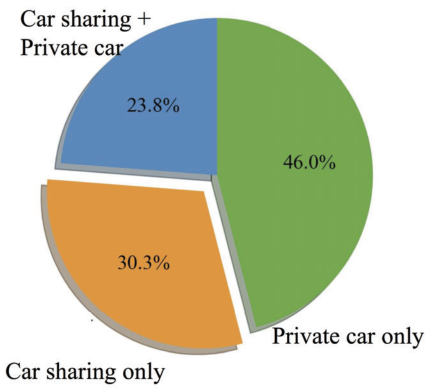

Fig. 7. Will you use car sharing with autonomous vehicles only in the future?

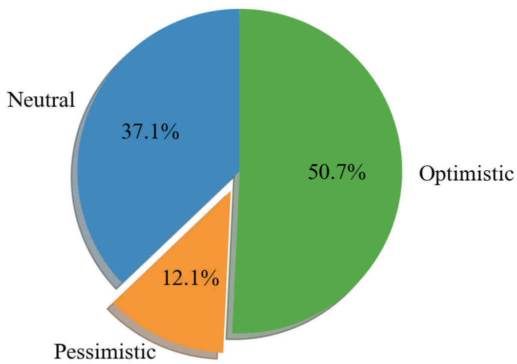

Fig. 8. How do participants think of car sharing

Theme 4. How Do Participants Think About The Sustainability of Car Sharing? *We further investigate people's attitudes toward the sustainability of car sharing.* As shown in Figure 8, we found over half (50.7%) of participants have an optimistic attitude to car sharing, and they think it is a promising mobility mode and the future. *"Car sharing is a promising mobility mode and it is a prototype of the future mobility"* (P3). Only 12.1% of participants have a pessimistic view, and they think car sharing cannot operate in the long run. Besides, over one-third (37.1%) of users have a neutral attitude, and they believe it still needs some time to see the future of car sharing. *"It is hard to say what will happen in the future, so it still needs time to validate"* (P6).

Even though our online survey can show some explicit attitudes of users to online car sharing, it is challenging for us to have a quantitative description of how users use car sharing. Hence, in the next section, we will provide some detailed quantitative results based on real-world operation data.

5 QUANTITATIVE RESULTS

In this section, we show our comprehensive quantitative user behavior investigation results with different metrics (e.g., When, where, why, how long, how regular do users usually use car sharing? How are the revisitation patterns?) In addition, we conduct a comparative study to verify the generalizability of obtained findings.

5.1 When & Why Do Users Use Car Sharing

We first investigate **when** users pick up and return shared vehicles. Later we describe why users might have these preferences. Figure 9 shows the shared vehicle pick-up and returning distributions in one week, and we found other weeks have similar patterns.

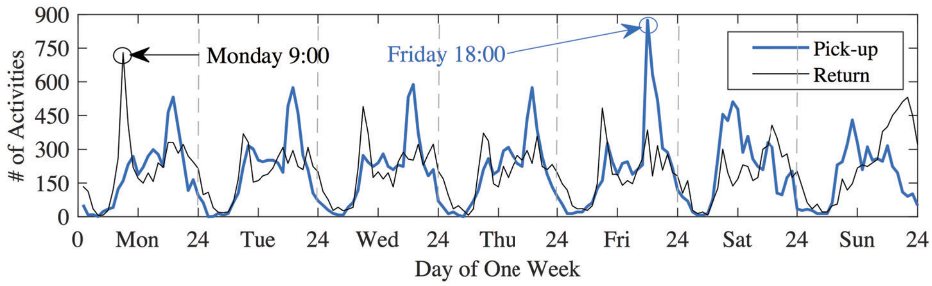

Fig. 9. Pick-up and returning distributions in one week

Weekly patterns. We found there is a distinct vehicle pick-up peak around Friday 18:00 (18:00-19:00), and the returning peak happens on Sunday midnight and Monday 9:00 (9:00-10:00). A possible reason is that many users leverage car sharing services for traveling during weekends. After combining the pricing mechanism of this car sharing service provider that we show in Section 3.2.2, we found this weekly usage pattern is consistent with a special weekend package (from Friday 17:00 to Monday 9:00), during which users can use vehicles with a low fixed cost (i.e., \$26.4) plus the mileage fee (\$0.28/mile). This indicates the cheap weekend package has the potential to incentivize users to use shared vehicles even though the number of users may also increase due to regular demand without such an economic incentive.

Weekday patterns. For Monday to Thursday, we found all their pick-up peaks happen during 19:00-20:00, which is one hour later than on Friday. For the returning peaks, we found all weekdays' returning peaks happen during 9:00-10:00. The reason may be that many users use the car sharing services *for commuting* (i.e., users drive the shared vehicles to home after work and then drive back to workplaces in the next morning), which makes a difference between car sharing and car rental since not many people use car rental for daily commuting.

Weekend patterns. The temporal patterns of usage in weekends are very different from weekdays'. In particular, the pick-up peaks of weekends happen around 11:00, and the returning peaks are around 21:00. This is because most people do not need to work during weekends, so most car sharing usages are for *leisure activities like traveling* instead of commuting on weekdays.

5.2 How Long Do Users Use Car Sharing

From Figure 10, we found that about 50% of car sharing usages are within 5 hours, and 85% of usages are in 15 hours. Even though most per-usage time is shorter than 10 hours, there are several surges in longer durations, i.e., 15 hours (corresponding to the night usage), 24 hours, 48 hours, and 63 hours (corresponding to one night + one weekend usage). We further investigate when these surges happen, and we found that they all happen in the three special pricing packages durations. Hence,

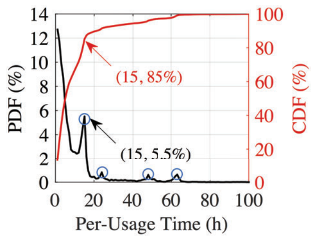

Fig. 10. Per-usage time

the temporal patterns of aggregate car sharing usage may suggest *low special pricing packages have the potential to incentivize users to use car sharing services.* This phenomenon is similar to that of 2017. Hence, this finding may help car sharing providers to customize pricing to enlarge their user groups.

5.3 Where Do Users Use Car Sharing

5.3.1 Connectivity. We first investigate the service station distribution by defining *connectivity* of a service station $\mathcal{S}$ to show the coverage of the service stations, which is represented by the number of neighbor service stations within a certain range of $\mathcal{S}$. From Figure 11, we can see that about 45% of the service stations have no neighbor stations within 1 km, which indicates they have low *connectivity* to other service stations even though most service stations are within the 5^{th} Ring Road of Beijing as shown in Figure 12. Even within 3 km, there are still 13% of service stations have no neighbors. One direct implication is that users need to return vehicles to a remote station if there is no parking space in the dedicated station, which may potentially lead to poor user experience and cause low penetration. Compared to the station distribution in 2017, we found the average *connectivity* of stations in 2019 has increased, e.g., there are 50% of the service stations have no neighbor stations within 1 km and there are 16% of service stations have no neighbors in 2017.

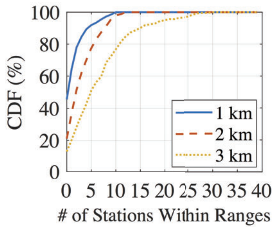

Fig. 11. Station distribution

5.3.2 Aggregate Spatial Behavior: Qualitative. We then investigate the aggregate spatial patterns of car sharing usage, i.e., **where** do users pick up and return shared vehicles? We visualize the total pick-ups and returns during the 10 months, and we found the aggregate spatial patterns of pick-ups and returns are similar since most trips are round-trip, so we only show the pick-up distribution as an example.

As shown in Figure 12, each circle stands for a car sharing service station, and the sizes of the circles mean the number of pick-up activities in stations, i.e., larger circles denote more activities in these stations. We found that there are several stations with very high car sharing usages. We then investigate the geographic locations of these high demand stations, and we found that the most visited stations are all located in **the CBD area and high-tech IT industry bases**, e.g., Zhongguancun and Yongyou Software Base. A possible reason may be that employees in these places usually pick up the shared vehicles after work and then return the vehicles to their workplaces in the next morning. One underlying reason may be that IT people have high incomes and are more open to new technologies. This finding can potentially help car sharing service providers to deploy service stations optimally, e.g., deploying more service stations or more shared vehicles around high-tech companies and city central business areas, and also they need to consider the social fairness when they enlarge the car sharing business.

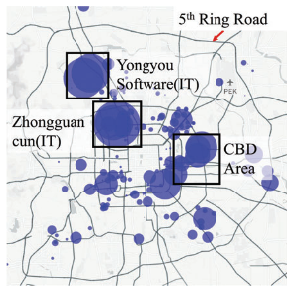

Fig. 12. Aggregate pick-up

5.3.3 Aggregate Spatial Behavior: Quantitative. We show the CDF of pick-ups in all stations to quantitatively measure the usage distribution in the car sharing services. As shown in Figure 13, we found that over half (52%) car sharing usages happen in only 20% of service stations, and 74% of usages are in 40% of service stations, which indicate that there is an unbalanced usage phenomenon among different stations. Compared to the usage distribution in 2017, we found the unbalanced phenomenon has improved, e.g., there are over 57% usages happen in 20% of service stations, but the unbalanced phenomenon is still severe. This unbalanced phenomenon can potentially result in the vehicles in high-demand stations are frequently used while some vehicles in unpopular

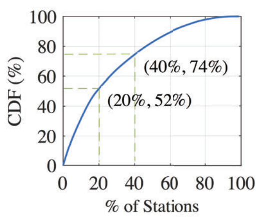

Fig. 13. Pick-up distr.

stations will be rarely used, which may pose huge challenges for vehicle maintenance management. For example, under the current setting, car sharing service providers conduct periodic maintenance to shared vehicles, while the unbalanced usage of vehicles makes some of the vehicles required to maintenance in advance, which causes a challenge for centralized management.

5.3.4 Hourly Spatial Behavior: Qualitative. Even though most usages are in the CBD area and high-tech IT industry bases, we found the top three stations of the highest vehicle net outflows (i.e., # of pick-ups − # of returns) are in residential communities. We then further investigate more fine-grained usage patterns in each station on an hourly basis since users usually use shared vehicles also on an hourly basis, so it may indicate the real-time usage distributions in stations. We found that there are highly unbalanced pick-ups and returns in some stations at different hours. Here we utilize 9:00-10:00 as an example to show this hourly unbalanced usage phenomenon. Figure 14 shows the difference between pick-ups and returns in each service station during 9:00-10:00, where the blue points mean there are more returns than pick-ups in these stations and red points indicate there are more pick-ups than returns in these stations. More specifically, we found there are more returns in IT industry bases and the CBD area, e.g., the top five stations with the most vehicle net inflows (i.e., # of returns − # of pick-ups) are in IT industry bases and the CBD area.

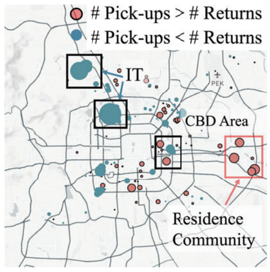

Fig. 14. Unbalance during 9:00-10:00

A possible reason may be that many users live in residential areas but work in IT companies, which cause an intensive returning phenomenon during 9:00-10:00 when they drive shared vehicles to their workplaces. This specific spatial usage pattern differentiate car sharing from other transport modalities (e.g., bike sharing [54] and car rental [23]). It should be noted that the working time of most IT companies in Beijing (e.g., Baidu and Alibaba) is between 10:00-19:00, so employees usually come to their companies between 9:00-10:00 and get off work after 19:00. One possible consequence is that it requires more workers to move and balance the vehicles across the city in real-time to satisfy later demand, which will potentially increase the operating costs of car sharing operators.

5.3.5 Hourly Spatial Behavior: Quantitative. To quantitatively show the unbalanced demand and supply phenomenon is common among different service stations, we show the net flows (i.e., $|\#of\ pickups - \#\ of\ returns|$, an absolute value) of vehicles in each station in one hour, where we also use 9:00-10:00 as an example. As shown in Figure 15, we found that 31% of service stations have a net flow of vehicles over 20. In particular, the station with the largest net flow is 110, which indicates it has a very unbalanced usage in the station. We also verified there is a similar phenomenon in other hours using our real-world data, so we do not show them.

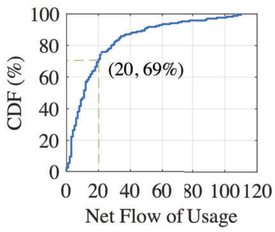

Fig. 15. Net flow of vehicle usages in each station at 9:00

5.3.6 Differences from Bike Sharing. Even though there exist unbalanced usages among different stations in bike sharing services, most users in Beijing usually utilize shared bikes for short trips, e.g., from their homes/workplaces to near subway stations or bus stations, or from the subway/bus stations to homes [54], and it is typically hard for users to ride only bikes from their homes to workplaces in Beijing because of the extremely expensive housing in CBD areas and IT industry bases. Therefore, usage purposes of car sharing services and bike sharing services result in dissimilar temporally unbalanced usage.

In addition, if we try to relocate shared cars to address the unbalanced phenomenon, we need to assign a driver for each car, which is also different from bike relocation, where each worker can relocate many bikes from one station to other stations for each time. In this case, the car sharing rebalancing problem is more challenging, costly, and time-sensitive than bike sharing rebalancing. These two features indicate the car sharing and bike sharing are indeed different.

5.4 How Regular Do Users Use Car Sharing

We introduce the entropy and conditional entropy of users' destinations (i.e., returning stations) and returning times to quantify the **regularity** of car sharing users' user behavior, which are indicators to show if car sharing users have stable user behavior and if their usages are **predictable**. The entropy of a specific user u's destinations H(D_u), the destination conditional entropy given u's original pick-up station H($D_u|O_u$), the destination conditional entropy given u's original pick-up station and pick-up time H($D_u|O_u$, T_u) are shown as below. D_u is the destination of user u; O_u is the original pick-up station of user u; $T_u \in \Theta$ is the pick-up time of user u, and Θ = {[00:00, 01:00), [01:00, 02:00), ... , [167:00, 168:00)} is the set of hours in a week.

$$H(D_u) = \sum_{D_u \in S} p\left(D_u\right) \log \frac{1}{p\left(D_u\right)} \tag{1}$$

$$H(D_u|O_u) = \sum_{D_u, O_u \in S} p\left(D_u, O_u\right) \log \frac{p\left(O_u\right)}{p\left(D_u, O_u\right)} \tag{2}$$

$$H(D_u|O_u, T_u) = \sum_{\substack{D_u, O_u \in S \\ T_u \in \Theta}} p\left(D_u, O_u, T_u\right) \log \frac{p\left(O_u, T_u\right)}{p\left(D_u, O_u, T_u\right)} \tag{3}$$

We replace D_u in Equation 2 and Equation 3 by R_u to calculate the returning time conditional entropy given u's original pick-up station H($R_u|O_u$), and the returning time conditional entropy given u's original pick-up station and pick-up time H($R_u|O_u$, T_u), respectively, which are shown as Equation 4 and Equation 5. Where S is the collection of car sharing service stations; $R_u \in \Theta$ is the returning time of user u, and Θ = {[00:00, 01:00), [01:00, 02:00), ... , [167:00, 168:00)} is the set of hours in a week. The reason why we use a week as time cycle is because there is a strong weekly usage pattern, and the usage pattern in weekends is different from the usage pattern in weekdays as shown in Figure 9.

$$H(R_u|O_u) = \sum_{\substack{O_u \in S \\ R_u \in \Theta}} p\left(R_u, O_u\right) \log \frac{p\left(O_u\right)}{p\left(R_u, O_u\right)} \tag{4}$$

$$H(R_u|O_u, T_u) = \sum_{\substack{O_u \in S \\ T_u, R_u \in \Theta}} p\left(R_u, O_u, T_u\right) \log \frac{p\left(O_u, T_u\right)}{p\left(R_u, O_u, T_u\right)} \tag{5}$$

The destination entropy and returning time entropy are shown in Figure 16 and Figure 17, respectively. From Figure 16, we found that if users' original pick-up station O_u is known, the destination conditional entropy H($D_u|O_u$) is very low, which means their destination stations are more stable and may be predicted. In addition, if their pick-up time T_u is also given, the conditional entropy H($D_u|O_u$, T_u) is near to zero, which indicates most users have stable shared vehicle station usage patterns, e.g., picking up a vehicle after work at their workplaces and returning the vehicle to their workplaces in the next day.

From Figure 17, we found that conditional entropy of users' returning time R_u given their original pick-up stations and pick-up time is small. Specifically, the returning time conditional entropy of about 77% of users is 0, which means their returning time given their current order and historical information is more predictable. Moreover, the returning time conditional entropy for almost all users is less than 0.5, which indicates their returning time is easy to predict. *This finding can potentially help us to design proactive car sharing rebalancing mechanisms based on users' historical and real-time order information.*

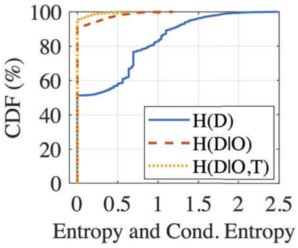

Fig. 16. Destination entropy

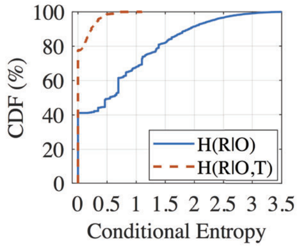

Fig. 17. Return time entropy

5.5 How Are the Revisitation Patterns in Car Sharing

Understanding the revisitation patterns in human mobility is significant and meaningful [9], which not only reveals some latent urban rhythms as we can know the frequency of people revisiting a particular location but also helps us to characterize locations and users into different behavioral clusters. Borrowing the idea from [9, 25, 28], in this paper, we adopt a non-deterministic view to model revisitation as a probability distribution over time, which represents the likelihood of a service station being revisited or a user revisiting a service station after a certain time interval.

5.5.1 User Revisitation and Station Revisitation Patterns. Figure 18 shows the average and median revisitation of all users. We found only 8% of users revisit car sharing daily on average, which indicates that not many people use it every day. There are 26.9% of users revisit car sharing every three days and 64.6% of them revisit car sharing services weekly on average, which potentially indicates most users use car sharing for leisure activities. The median revisitation is smaller than the average revisitation frequency, which potentially indicates that the revisitation patterns of users are diverse. Figure 19 shows the revisitation of service stations, and we show both the pickup and return revisitation patterns. We found about 30% of service stations have pickups every hour on average while only 16% of them have returns every hour on average, this indicates the pickup revisitation pattern of stations is more intensive compared to the return revisitation pattern. Figure 20 shows the number of unique users visited each station. We found that few users visited many service stations, e.g., only 0.6% of users visited more than 10 service stations. About 73% of users visited no more than 2 service stations, which potentially indicates they have stable usage and revisitation behavior.

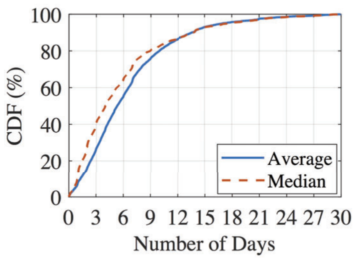

Fig. 18. User revisitation distribution

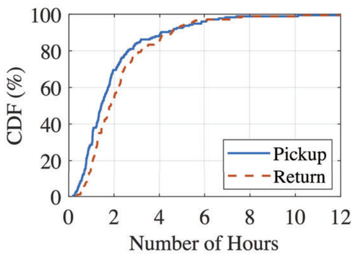

Fig. 19. Service station revisitation

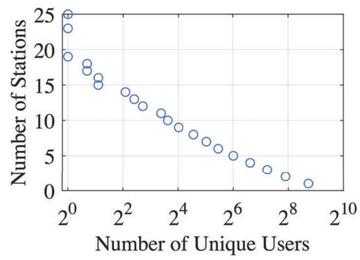

Fig. 20. Unique users per station

5.5.2 Usage Activity Identification. In this paper, we propose to use a spatiotemporal constraints-based method to identify user activities related to car sharing usages. (i) We adopt a widely-used stopping event extract method in [62] to obtain all stopping events of all users when they are using the car sharing services, which is identified by the events that the vehicle stays for a long time (e.g., $\tau_1 > 30$ *minutes*) at the same point (i.e., same longitude and latitude) according to the GPS records. (ii) We then utilize the popular home and workplace inference techniques in [7, 15] to obtain their location revisitation patterns. (iii) Combined the findings from our online survey and data analysis (e.g., Fig. 9), for each user, if the stopping event is in 18:00-7:00 of weekdays and the stopping duration is longer than a threshold (e.g., $\tau_2 = 3\ hours$), then it is labeled as his/her home, and the service station locations are labeled as their workplaces if they return vehicles during 7:00-10:00. If both the stopping events are during weekends and holidays, then they are identified as leisure activities, from where we can obtain their traveling destinations and time. We also utilize some spatiotemporal constraints to refine our results, e.g., the locations of leisure activities should be in certain places like shopping malls, tourist attractions, and restaurants, etc. We utilize Google Maps and Baidu Map to manually check and filter our extracted leisure activities from stopping events. Finally, we extract users' usage activities for different purposes, e.g., home → workplace, workplace → home, home → leisure activities. We eventually identified 22.4% of home → workplace activities, 23.2% of workplace → home activities, and 35.3% of home → leisure activities, and about 19% of usages are for other purposes, e.g., emergency. And we found about 35.7% of users usually use car sharing for commuting.

5.5.3 User Grouping. We then further group all users into different groups based on their usage patterns, including usage activities, usage locations, usage time, etc. The user grouping process is as follows.

Step 1: Constructing a feature vector for each user. We construct the feature vector based on the above data analysis. Each vector includes the following six categories of features: (i) median usage time; (ii) median usage distance; (iii) total usage count; (iv) the total number of visited service stations; (v) median revisitation frequency; (vi) number of different usage activities (e.g., # of home → workplace, # of workplace → home, # of home → leisure activities). Finally, we obtain 9 features for each vector. These features based on our data analysis capture fine-grained usage patterns of each user during a long time period, which enables an effective clustering.

Step 2: Clustering users into different groups. Based on feature vectors, we cluster users into different groups by a Gaussian Mixture Model (GMM) [71] as in Equation 6.

$$p(x) = \sum_{k=1}^{K} \pi_k \mathcal{N}\left(x|\mu_k, \Sigma_k\right) \tag{6}$$

where x is a feature vector in our model, $\mathcal{N}$ is a Gaussian distribution with μ_k as the mean and Σ_k as the covariance matrix. We apply a standard Expect Maximization algorithm to maximize the likelihood iteratively. The output of the clustering gives the centroid μ of each cluster and the corresponding probability of x being in a cluster. We apply a Gaussian-based clustering method since Gaussian distributions are fit into mobility metrics of crowds in many scenarios [14]. Finally, we obtain mainly 4 user groups, and their revisitation frequency and purposes are shown in Table 3.

Table 3. Revisitation patterns and usage purposes of different user groups

Groups	Average Revisitation (day)	Average Usage Duration (hour)	home → workplace : workplace → home : home → leisure
Group1	5.1	4.5	0.39: 0.21: 0.40
Group2	4.3	13.3	0.20: 0.40: 0.40
Group3	17.6	39.9	0.14: 0.20: 0.66
Group4	43.5	6.3	0.26: 0.21: 0.52

We found the highest usage purpose of all the four user groups is for leisure activities, which potentially indicates most people use car sharing for leisure activity purposes. For the Group1, we found that many usages are from home → workplace, and their revisitation frequency is high, which potentially indicates that many users in this groups use car sharing for commuting purposes and usually pick shared cars around their homes and return cars around their workplaces. For the Group2, we found many users in this group use car sharing from workplace → home, and their revisitation frequency is also high, which also indicates these users use car sharing for commuting. We also found their average usage duration is high, i.e., 13.3 hours, which is corresponding to our findings since users may park cars in their homes at nights and then return cars when they come back to their workplaces in the morning of the next day. For the Group3, we found most usages are from home ⟶ leisure, and their average revisitation frequency is low and usage duration is long, so this group may correspond to users who use car sharing for activities. For the Group4, we found that their average revisitation time is very long, so most of them may be irregular users. This group of users may usually use car sharing for other purposes, e.g., emergency or using car sharing when they want to drive. We also compare the results with the k-means combined with the elbow point-based user grouping method [9], and it turns out that the results of our GMM combined with Davies-Bouldin index method are more interpretable and in line with our online survey results.

5.6 Who Are More Likely to Use Car Sharing

Since people of different demographic characteristics may have different shared vehicle usage purposes, their behavior might be different. Hence, we consider car sharing users' demographic information in our dataset (e.g., age and gender) to perform a fine-grained demography-based user behavior study, which answers which groups of people may have a higher probability to use car sharing services.

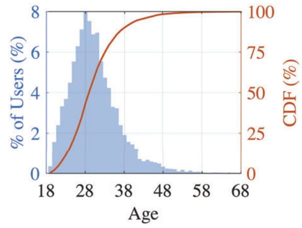

Fig. 21. User age distribution

5.6.1 Age-Based User Behavior. Figure 21 shows the age distribution of car sharing users. We found that about 92% of users are between 18 years old to 38 years old, and the oldest user is 68 years old. Among all users, about 81% of them are male, and only 19% are female. We adopt an equal width-based method to divide users of different ages into different groups, and we set the interval as 3. Finally, we have 17 groups in our dataset (i.e., 18-20, 21-23, ... , 66-68).

Figure 22 shows a heatmap of the total car sharing usage patterns of different age groups. We found that the users between 27-29 have the highest usage demand, and their most frequent usage time durations are 18:00-20:00. One reason could be that there is the largest number of users in this group (i.e., 27-29), and they may use the shared vehicles for commuting. In general, the shared vehicle usage frequency of each group is proportional to the number of users in each age group, and the aggregate temporal pattern is obvious in Figure 22, e.g., concentrating on 18:00-20:00 and 9:00-13:00 from users between 24-32 years old.

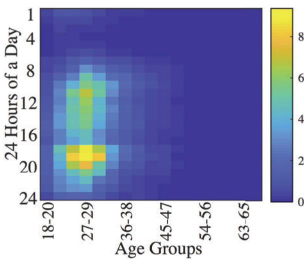

Fig. 22. Usage of age groups

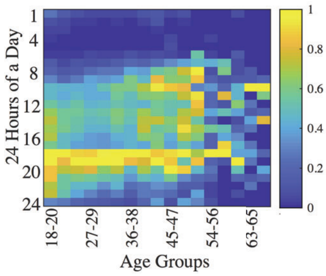

Fig. 23. Weekday usage

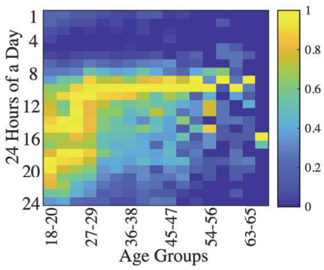

Fig. 24. Weekend usage

We further investigate the temporal usage patterns of different age groups. Due to the different usage patterns between weekdays and weekends, as shown in Figure 9, we also compare the usage distributions of different age groups for weekdays and weekends. As shown in Figure 23 and Figure 24, each column stands for a temporal usage distribution for each age group of users.

From Figure 23, we found that younger groups have different patterns as compared to senior users. More specifically, for users between 18 to 44 years old, they use car sharing services mostly during the evening rush hours (i.e., 18:00-20:00). However, for users older than 45 years old, their temporal usage patterns are not conspicuous, and they may have intensive usage between 8:00-11:00 or 18:00-20:00, which indicates that only a part of them use shared vehicles for commuting and some of them use shared vehicles for other leisure activities during weekdays.

Comparing Figure 24 with Figure 23, we found that their temporal patterns are very different. For example, users between 18 to 26 years old have two distinct usage peaks. One is 14:00-15:00, and another one is the same as weekdays (i.e., 18:00-20:00). The possible reason is that some users use the shared vehicles for traveling, while many young users still need to work during weekends, and they use the shared vehicles for commuting. Working at weekends is a common phenomenon for young people in some IT companies in Beijing, which is consistent with our observations in Figure 12. For users between 27 to 32 years old, their usages are more diverse (i.e., from 9:00 to 19:00), so users of these age groups may have different purposes for using shared vehicles. The weekend usage patterns of users between 33 to 44 years old are very different from weekdays', and their usage peak is around 10:00, which indicates they mostly use the shared vehicles for traveling during weekends. Even though the spatial usage patterns of users older than 45 are still irregular, we found the peaks are also around 10:00.

In summary, users of different ages have distinct shared vehicle usage patterns, and their patterns are also different between weekdays and weekends, which indicate they may have different usage purposes. *This finding can potentially help service providers to customize pricing packages for different user groups.* For example, car sharing service providers can provide weekly commuting pricing packages (including weekends) or coupons for young people who use shared vehicles for commuting, and they can also provide daytime pricing packages/coupons for older people to incentivize more elders to use this innovative mobility mode.

5.6.2 Gender-Based Usage Behavior. We also compare the car sharing user behavior of female users and male users. Figure 25 shows the destination entropy and conditional entropy of female users $H_f(D)$, $H_f(D|O)$ and male users $H_m(D)$, $H_m(D|O)$. We found that the entropy of female users is smaller than male users', which means that female users have more stable destinations than male users, and their usage patterns are more regular. We further investigate the most popular locations of female users and male users. As shown in Figure 26, the three blue circles stand for the top three popular service stations for male users, and we found they are all located in IT bases, which indicates most male users may work in IT companies. However, the top three service stations for female users are in the CBD area and the financial harbor, which potentially indicates that females who work in business companies have higher shared vehicle usage demand.

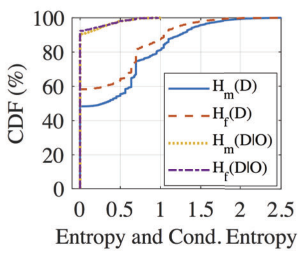

Fig. 25. Entropy of female and male users

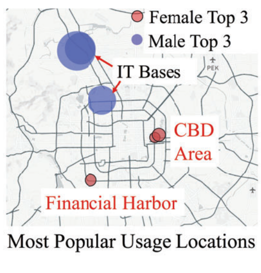

Fig. 26. Top 3 locations of female and male users

This finding has the potential to help car sharing service providers to deploy shared vehicles in different stations optimally. In addition, service providers can adopt different vehicle types or colors for female and male users to provide a user-friendly car sharing service.

5.7 Cross-City Study: Are the Findings Generalizable

To investigate if the results obtained from Beijing data are extrapolated to other cities, we conduct a cross-city study using the car sharing data from the other two Chinese cities, i.e., Fuzhou and Lanzhou. Even though we have access to data of over 100 cities, the reason why we choose the three cities is that the three cities have significantly different urban characteristics, e.g., different geographical distributions, city sizes, population densities, economic levels, and climate conditions, so they have the potential to represent different types of cities. (i) Beijing is the capital city of China and is one of the largest cities in China, which is located in the north of China and it has over 21 million population. Beijing also has the highest economic level in China, and the climate there is a sub-humid warm temperate continental monsoon climate. (ii) Fuzhou is located in the southeast coastal region in China, which has over 7.8 million population. Fuzhou has a subtropical maritime monsoon climate and relatively high economic level in China. (iii) Lanzhou is one of the most important central cities in the west of China, and it has about 3.8 million population. Lanzhou has a temperate continental climate and also has a relatively high economic level in China. In addition, even though we have access to the data of over 30 cities, we found the people in the high economic-level cities are more likely to use car sharing services, so there are enough data for us to analyze. Hence, we believe the three cities we choose are representative of the development of car sharing services. The dataset from Fuzhou and Lanzhou contains one-year usage records from over 42,000 users. The data formats of Fuzhou and Lanzhou are the same as that of Beijing.

Due to the space limitation, we only report part of results based on Fuzhou and Lanzhou data to verify our major findings from Beijing data. Specifically, we show the unbalanced temporal and spatial patterns of car sharing usage in Fuzhou and Lanzhou, and we also show the spatial patterns and demography-based usage.

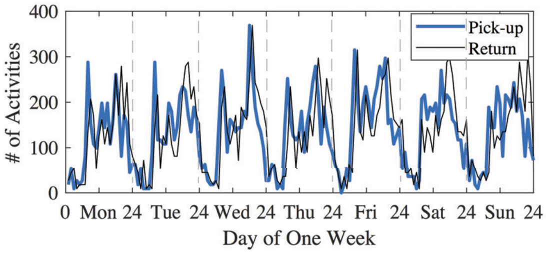

Fig. 27. Pick-up and returning distributions in Fuzhou

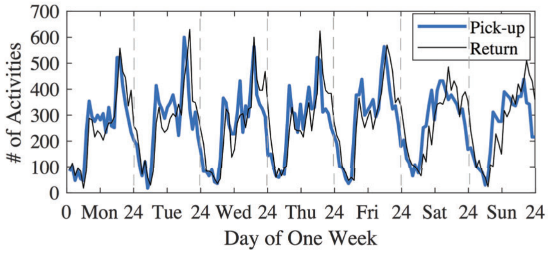

Fig. 28. Pick-up and returning distributions in Lanzhou

5.7.1 Unbalanced Temporal User Behavior. From Figure 27 and Figure 28, we found that the usage patterns of car sharing in Fuzhou and Lanzhou is not regular as that of Beijing, and their pick-up and return patterns are similar. However, the usage during one day is also highly dynamic. Both pick-up and return demands are high during morning rush hour and evening rush hour in Fuzhou, but they are high only in evening rush hour in Lanzhou.

Besides, we also investigate the temporal usage patterns of users in different age groups. As shown in Figure 29, we found that most car sharing usages are happening in daytime in Fuzhou, e.g., 8:00-20:00 from young users between 21-32. The reason may be that most users are in this age range, and car sharing is used for daily short-time trips. However, the temporal usage patterns are different in Lanzhou, e.g., young people usually use car sharing in the afternoon and night, but there are two usage peaks for users between 27-47, i.e., morning rush hour and evening rush hour, which potentially indicates they use car sharing for commuting.

5.7.2 Unbalanced Spatial User Behavior. We also investigate the spatial user behavior of car sharing in Fuzhou and Lanzhou. As we can see from Figure 31, about 89% of pick-ups only happen in 50% of services stations in both Fuzhou and Lanzhou, which indicates that there is an unbalanced usage phenomenon among different stations in the two cities. This unbalanced phenomenon can potentially result in the vehicles in high-demand stations are frequently used while some vehicles in unpopular stations will be rarely used. In addition, from Figure 32, we found that the usage between different stations has a huge gap during some certain time slots in Fuzhou, e.g., more pick-ups in the CBD area while very few returns during 9:00-10:00.

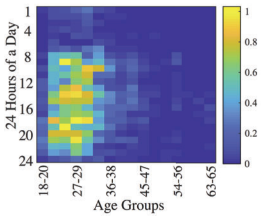

Fig. 29. Usage of age groups in Fuzhou (Compare to Figure 22 of Beijing)

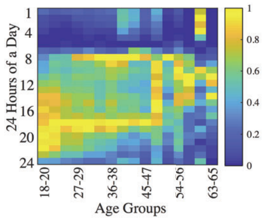

Fig. 30. Usage of age groups in Lanzhou (Compare to Figure 22 of Beijing)

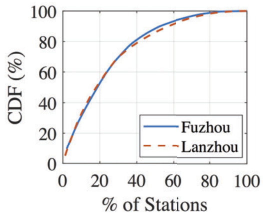

Fig. 31. Pick-up distribution of Fuzhou and Lanzhou (Compare to Figure 13 of Beijing)

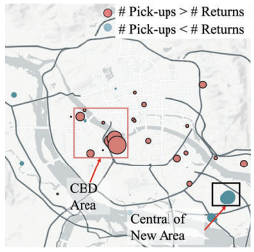

Fig. 32. Spatial unbalance during 9:00-10:00 in Fuzhou

In summary, even though there is a different spatial user behavior between the three cities, the unbalanced usage phenomenon between different stations exists in the three cities.

5.7.3 Entropy-Based Comparison. Similarly, we utilize Equation 1 to Equation 5 to compute the return time entropy and conditional entropy. We also compare the destination entropy of female and male users. The results are shown in Figure 33 to Figure 36. Finally, we have conclusions similar to the ones drawn from Beijing data. (i) The destination and returning time of Fuzhou users are potentially easy to predict than that of Beijing and Lanzhou, which can guide us to design effective balancing mechanisms for car sharing networks. (ii) Female users in Fuzhou have similar destinations with male users. Female users in Lanzhou have more stable destinations compared to male users, so their usage patterns may be more predictable.

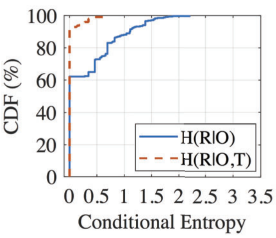

Fig. 33. Return time entropy of Fuzhou (Compare to Figure 17 for Beijing)

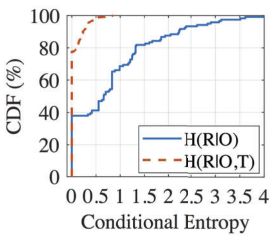

Fig. 34. Return time entropy of Lanzhou (Compare to Figure 17 for Beijing)

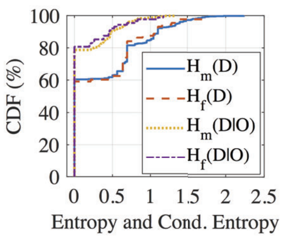

Fig. 35. Entropy of female and male users of Fuzhou (Compare to Figure 25)

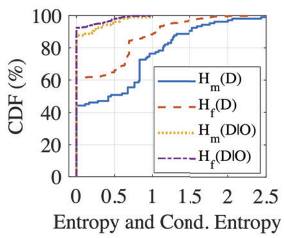

Fig. 36. Entropy of female and male users of Lanzhou (Compare to Figure 25)

To summarize, even though there are some similar usage patterns (e.g., highly predictable), we found many user behaviors in Fuzhou and Lanzhou is different from that of Beijing. **Hence, we still do not have sufficient evidence to claim that our findings are generalizable to other cities or countries, so more comparative studies should be conducted in the future.**

5.8 How Is Car Sharing Different from Other Mobility Strategies

In this part, we compare car sharing with the most relevant mobility strategies (e.g., bike sharing, taxi, and car rental) from both spatial and temporal aspects. We utilize the daily usage distribution and per-usage distance as the comparison metrics. All data used for the comparative study are collected from the same city (i.e., Beijing).

From Figure 37, we found that the three mobility strategies have different temporal patterns. In particularly, bike sharing has two usage peaks (7:00-9:00 and 18:00-19:00, which are the rush hours in Beijing); there is a high taxi demand in Beijing during 8:00-22:00 but no obvious peaks; while the peak of car sharing pick-up is during evening rush hours (i.e., 18:00-19:00), and the peak is very obvious compared to other durations.

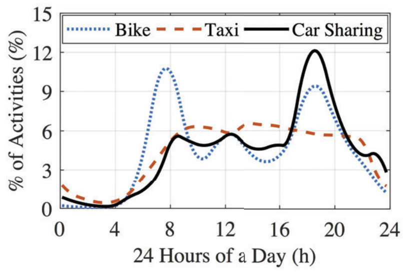

Fig. 37. Temporal comparison

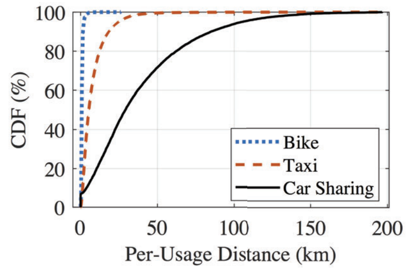

Fig. 38. Spatial comparison

Figure 38 shows the per-usage distance distribution of the three mobility strategies. We found that about 95% of bike per-usage distance is within 2 km; about 92% of taxi trips are shorter than 20 km; while only 11% of the per-usage distance is shorter than 10 km, and 65% of car sharing per-usage distance is longer than 20 km.

Their different spatial and temporal patterns potentially indicate people have different usage purposes for using different mobility services, e.g., car sharing is usually used for leisure activities and commuting purposes as shown in our online survey and data analysis; bike sharing is usually used for short trips like from homes to subway stations; taxis are usually used for immediate travels. Hence, based on their different spatiotemporal patterns, we conclude that car sharing is different from bike sharing and taxis, and they can be seen as complementary strategies for each other.

Even though car rental is also similar to car sharing, it is usually used for daily or weekly usage and long trips based on the current operation mode [18, 23, 38], and we will show more detailed differences in the following parts (e.g., Section 5.1). Another key difference between car sharing and car rental is their operation modes. For example, car rental companies usually operate some dealerships, and users need to go to those dealerships to rent and return cars; for car sharing, operators usually use some parking lots as their service stations, and users can find nearby shared cars via Apps. Hence, there are three key differences between car sharing and car rental: (i) different spatiotemporal usage patterns, (ii) there is no dealership requirement for car sharing, and (iii) car sharing is more convenient for users to use.

6 CASE STUDY: USAGE PREDICTION

As we found from our qualitative and quantitative analysis, the pickup and return behavior of users are highly dynamic in both spatial and temporal dimensions, and one of the most important tasks for online car sharing operation is the usage prediction, which is not only impacts vehicle rebalancing and station siting for operators, but also impacts user experiences, e.g., both starved and jammed stations will damage user experience if users want to pick up or return shared cars. That is to say, there should be shared cars in the station if users want to pick vehicles up and there should be empty parking spaces in the station if users want to return vehicles in this station. Fortunately, from our user behavior investigation, we found some key factors that correlated to users' usage behavior, e.g., time of day, day of week, functional areas, user demographic characteristics, and station revisitation frequency, which provide good opportunities for us to model and predict usages more accurate. Hence, in this part, we utilize **usage prediction as a concrete example of the use case of our user study work** and predict the number of pick-ups and returns in the service station level based on obtained findings from our qualitative and quantitative analysis, from where we extract some fine-grained features.

6.1 User Behavior-Aware Prediction Model

In this paper, we propose to use a user Behavior-aware XGBoost (called BeXGboost) to predict the number of pick-ups and returns in car sharing systems at a fine-grained spatiotemporal level based on our qualitative and quantitative user behavior study. Since the number of pick-ups and returns are highly related to different user behavior, we first extracted a set of user behavior related features that we learned from our online survey and statistical data analysis. We exclude the features that not many participants think important to car sharing usage (e.g., vehicle conditions like the internal environment of vehicles) as we learned from our online survey. Finally, we extract four categories of features that are highly related to car sharing users' usage behavior, i.e., Spatiotemporal Feature $\mathcal{F}_{\mathcal{S}}$, Historical Feature $\mathcal{F}_{\mathcal{H}}$, User Demographic Features $\mathcal{F}_{\mathcal{D}}$, Revisitation Feature $\mathcal{F}_{\mathcal{R}}$. In addition, we also consider a widely adopted feature for better prediction performance, i.e., Contextual Features $\mathcal{F}_C$ including weather conditions and temperature [34, 54]. The detailed descriptions of user behavior related features are shown in Table 4. We then feed these user behavior related features into the decision-tree-based XGBoost framework [12], and the predicted value can be represented as

$$\hat{y}_i = \sum_{k=1}^{K} h_k(\mathbf{x}_i), h_k \in \mathcal{H} \tag{7}$$

where K is the number of trees; $\mathbf{x}_i$ is the ith input, including the 4 categories of user behavior features and contextual features (14 in total); $\hat{y}_i$ is the corresponding predicted output, which is learned by a tree ensemble model with a collection $\mathcal{H}$ of K functions h_k. Each tree is a function h_k that maps the input features to a score. Then the objective function at training round t iteration can be denoted as

$$J^{(t)} = \sum_{i=1}^{n}(l(y_i, \hat{y}_i)) + \sum_{k=1}^{t}\Omega(f_k) \tag{8}$$

where $l(.)$ is the loss function (e.g., Square loss); Ω is the regularization term (e.g., $L2$ norm), which measures the complexity of the model. Since the base model of BeXGBoost is decision tree, it has the potential to show better performance against overfitting.

6.2 Baselines

We compare our user behavior-aware BeXGBoost with the following state-of-the-art prediction methods, which are widely adopted in recent bike sharing or car sharing usage prediction [54, 57, 58, 60].

- **Vector Autoregression (VAR)** is a stochastic process model used to capture the linear interdependencies among multiple time series, which allows more than one evolving variable [36].
- **Random Forest (RF)** is an ensemble learning method that widely used for regression. Yang et al. [57] utilized RF to predict the bike sharing usage and achieved good performance.
- **Gradient Boosting Decision Tree (GBDT)** utilizes an ensemble of decision trees to predict targets and has been widely adopted to predict usage behaviors recently, e.g., Wang et al. [54] utilized GBDT to predict bike sharing usages based on some extracted features.
- **Deep Learning-Based Methods**, we utilize different state-of-the-art deep learning methods as baselines, including Long Short-Term Memory networks (**LSTM**), Convolutional Neural Networks (**CNN**), and Graph

Table 4. A set of user behavior related features that we learned from our qualitative and quantitative studies

Feature Type	Feature Name	Meaning of Each Feature
Spatiotemporal Features $\mathcal{F}_{\mathcal{S}}$	$\mathscr{F}_{tod}$	the fine-grained temporal feature, i.e, time of a day. We discrete one day into four time slots, morning peak hour, evening peak hour, late-night hour, and other time
	$\mathscr{F}_{dow}$	the day of the week, which has five categories, i.e., Monday, Friday, weekend, holiday, and other days
	$\mathscr{F}_{ut}$	the usage time, i.e., average time of usages in each station
	$\mathscr{F}_{area}$	the urban area of each service station. We found users have different purposes for using car sharing systems, which result in different spatial usage patterns. Hence, we divide the city into seven categories of functional areas based on the method in [59] (i.e., residence, entertainment, business, industry, education, scenery spot, and suburb) to capture the spatial characteristics of service stations
	$\mathscr{F}_{ud}$	the usage distance, i.e, the average usage distance of each station
Historical Feature $\mathcal{F}_{\mathcal{H}}$	$\mathscr{F}_{his}$	we utilize our long-term car sharing operation data to capture the historical usage patterns. We extract the number of pick-ups and returns of each service station in the same hour of three previous consecutive weeks (e.g., $\mathscr{F}_{his1}$, $\mathscr{F}_{his2}$, $\mathscr{F}_{his3}$) as the historical usage features.
Revisitation Feature $\mathcal{F}_{\mathcal{R}}$	$\mathscr{F}_{revi}$	the revisitation distribution of each station, and we use the average revisitation and median revisitation frequency of pick-up and return in each station
User Demographic Features $\mathcal{F}_{\mathcal{D}}$	$\mathscr{F}_{age}$	the user age group distribution in different stations, which we utilize the average age $\bar{\mathscr{F}}_{age}$ and median age $\tilde{\mathscr{F}}_{age}$ of all users visited each station
	$\mathscr{F}_{gen}$	the user gender distribution in different stations, which we utilize the ratio of male users to female users in each station
Contextual Features $\mathcal{F}_{C}$	$\mathscr{F}_{wea}$	current weather conditions, which is divided into three categories, i.e., sunny, rainy, and snowy
	$\mathscr{F}_{temp}$	current temperature, which includes three types of values: cold (lower than 15 °C), mild (15 -30 °C), and hot (over 30 °C).

Convolutional Network (**GCN**), which are widely used in bike sharing and car sharing usage prediction recently [34, 58, 60].

6.3 Performance Metrics

We utilize two widely used metrics to quantify the prediction performance of different methods, including Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), shown as below.

$$MAE = \frac{1}{n}\sum\nolimits_{i=1}^{n}\left|y^{(i)} - \hat{y}^{(i)}\right| \tag{9}$$

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

where $y^{(i)}$ and $\hat{y}^{(i)}$ are the true value and predicted value in the ith time slot respectively while n is the total number of predictions.

6.4 Prediction Results

Since car sharing is usually used on an hourly basis, hence, in this part, we perform a fine-grained prediction to predict the usages at hour level. We utilize 70% of data for training and the remaining for testing. All experiments are conducted in the same environment on Windows 10 with GPU Tesla V100-PCIE-16GB.

From Figure 39 to Figure 42, we found some deep learning-based methods (LSTM and CNN) show worse performance compared to the ensemble and boosting learning-based methods (RF, GBDT, XGBoost, and BeXGBoost). Even though GCN shows good prediction performance on the pick-up prediction, its performance is poor on return prediction. However, our BeXGBoost shows the best performance on both pick-up prediction and return prediction in terms of MAE and RMSE. The reason may be that the car sharing usage during a long-term period is evolving, which results in the deep learning-based methods challenging to capture the dynamics without user behavior consideration or overfitting, while XGBoost-based methods normally show excellent performance for the problems with small-to-medium structured/tabular data, which is exactly our extracted user behavior related features. The difference between our XGBoost and BeXGBoost is that XGBoost does not consider the user demographic features and latent revisitation patterns, which are rarely considered by existing bike sharing and car sharing usage prediction works [34, 54, 57, 58]. Another reason may be that the advantage of the user behavior highly-related features we obtained from our qualitative and quantitative user studies, which potentially indicates the significance of our work on user studies.

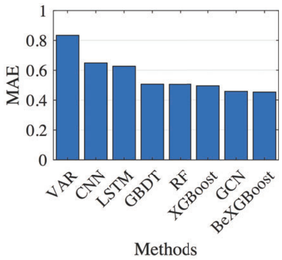

Fig. 39. MAE of pick-up prediction

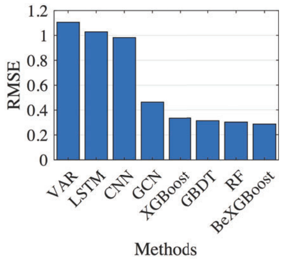

Fig. 40. RMSE of pick-up prediction

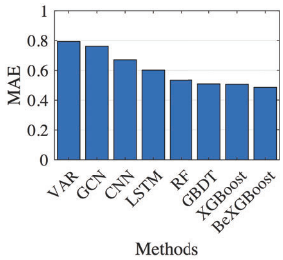

Fig. 41. MAE of return prediction

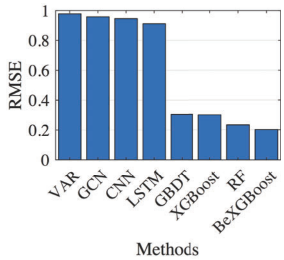

Fig. 42. RMSE of return prediction

7 DISCUSSIONS

In this section, we provide a set of findings obtained from our user behavior study, and then we discuss some potential implications and limitations of our work.

7.1 Summary of Finding

We summarize some key findings of car sharing user behavior from our study in Table 5.

7.2 Implications for Design

We present some potential implications for design, and we also link these implications to the findings in Table 5.

Service Station Deployment: Since the vital concern that prevents users from using car sharing is the convenience (**Q1** in Table 5), suitable station siting is essential to improve user experiences.

We found that most service stations have low connectivity to other service stations (**S3**). One direct implication is that some users need to return vehicles to a far station if there is no parking spot near their destination,

Table 5. Findings of user behavior in car sharing

Comprehensive User Behavior	Qualitative	**(Q1) Willingness.** Most people (84.3%) are willing to use car sharing, as shown in Figure 4, but only 30.3% of people indicated they would rely on future autonomous car sharing (Figure 7). **(Q2) Purpose.** 38% of use car sharing for traveling, but only 18% of participants use it for commuting. **(Q3) Convenience.** The most significant concern that prevents users from using car sharing is the convenience, as 63% of participants think that the vehicle availability will hinder their usage, followed by the vehicle returning (54%, due to limited parking spots in each station). **(Q4) Sustainability.** Over half (50.7%) of participants have an optimistic attitude to car sharing, and they think it is a promising mobility mode and the future.
	Temporal	**(T1) Duration.** Users usually use car sharing services with 10 hours for each time (e.g., 50% of usages are within 5 hours). There are also several surges in longer durations, i.e., 15 hours, 24 hours, 48 hours, and 63 hours as shown in Figure 10. **(T2) Daily.** On weekdays, most users pick up shared vehicles around 19:00, and they return around 9:00. On weekends, most users pick up shared vehicles around 11:00, and they return vehicles around 20:00, as shown in Figure 9. **(T3) Weekly.** There is a periodical weekly usage behavior, e.g., most users pick up shared vehicles on Friday 18:00 (18:00-19:00), and they return vehicles on Monday 9:00 (9:00-10:00), as shown in Figure 9. A possible reason is that users use car sharing to travel during weekends, as indicated by Figure 5. **(T4) Temporal Usage Regularity.** Most users have a high temporal regularity to use car sharing services, e.g., about 80% of usages' returning time is predictable given pickup station and time, as shown in Figure 17.
	Spatial	**(S1) Location.** High-income users work in IT industry bases and the CBD area are more likely to use car sharing services, as shown in Figure 12. About 57% of car sharing services happen in only 20% of service stations and 77% of usages are in 40% of service stations (Figure 13), which indicate that there is an unbalanced usage phenomenon among stations in different regions (also drawn from Figure 14 and 15). **(S2) Connectivity.** Low connectivity of the car sharing stations may be a key reason for poor user experience, e.g., over 16% of stations have no neighbors in 3 km, and about 50% of the stations have no neighbor within 1 km, as shown in Figure 11. **(S3) Spatial Usage Regularity.** Most users have a high spatial regularity to use vehicles in certain stations, e.g., over 95% of usages' destinations are very predictable given the pickup stations and times, as shown in Figure 16.
	Revisitation	**(R1) Frequency.** Only 8% of users revisit car sharing daily, and 64.6% of them revisit car sharing services weekly on average, which potentially indicates most users use car sharing for leisure activities.
	Age Related	**(A1) Distribution.** Most users are young people, e.g., about 90% of users are between 18 to 38 years old, as shown in Figure 21. **(A2) The Most Active Users.** Users between 27-29 have the highest car sharing usage demand, and they usually use shared vehicles in 18:00-20:00 (Figure 22). **(A3) Weekdays vs. Weekends.** Users in different age groups have different car sharing usage behavior, and they are also different between weekdays and weekends. For example, on weekdays, users between 18 to 44 years old pick up shared vehicles mostly during the evening rush hours (i.e.,18:00-20:00, while many elders pick up shared vehicles in the morning 8:00-11:00 (Figure 23). More morning pick-ups (e.g., around 10:00) are from young users in weekends (Figure 24).
	Gender Related	**(G1) Distribution.** Most (80%) car sharing users are male. **(G2) Predictability.** The destinations of female users are more predictable, as shown in the entropy distribution of Figure 25. **(G3) Different Popular Locations.** The most popular stations for male users are in the IT industry bases, while the most popular stations for female users are in the CBD area and financial harbor as shown in Figure 26.

which will severely impair user experience and cause a lower penetration. Hence, it is necessary to make car sharing networks more connected. We found that the CBD areas and the places with centralized IT companies have higher car sharing usage demand. One possible reason is that people that work at these places usually work late at night or even work during weekends, and they use car sharing vehicles for commuting. *Hence, the optimal locations for deploying more service stations could be some places with centralized office buildings, high-tech companies, or financial centers, where there are many young people who are highly likely to utilize car sharing vehicles for commuting, especially when they need to work late. In addition, for female and male users, they have different frequent visiting service stations as shown in Figure 26, so service providers can also tentatively deploy different vehicle models (or vehicles in different colors) in different stations for the better user experience of both female and male users **(G3)**.*

Vehicle Rebalancing: Due to the station-based nature of the car sharing service, users need to pick up vehicles from a set of pre-selected stations, and they also need to return vehicles to these stations. Since the vehicle counts in each station will directly impact the availability of vehicles and parking spots in the station **(Q3)**, an effective vehicle rebalancing mechanism is also important to enhance user experiences. We found there are extremely unbalanced pick-up and returning distributions among stations in some time slots, e.g., 9:00 and 18:00 **(S2)**. As shown in Figure 14, some stations have huge gaps between the pick-ups and returns. One possible result is that it requires more workers to move and balance the vehicles across the city to satisfy demand in later hours, which will potentially increase the operating costs of car sharing operators. However, *we found the destinations and returning times of users are highly predictable, given their pick-up stations and pick-up times, which is shown in **(S4 & T4)**, so we can design a usage rebalancing mechanism based on predicting their returning stations and times. This prediction-based rebalancing mechanism can be used to proactively balance the number of car sharing vehicles in each station and provide returning suggestions for users in real-time.*

Pricing Mechanism: Even though users have an optimistic attitude towards car sharing **(Q4)**, an important factor to promote it is to design a reasonable pricing mechanism. Attractive pricing packages have the potential to incentivize people to use car sharing vehicles, as shown in **(T1) & (T2)**. Current service providers only provide several fixed pricing packages without considering the dynamic usage patterns. Based on Figure 9, we found there are different peak times and off-peak times during weekdays and weekends **(T2) & (T3)**, *so service providers can design a dynamic pricing mechanism to increase the prices during peak durations and lower the prices in off-peak hours in real-time, which is similar to the current for-hire vehicle pricing mechanism. In addition, for users of different ages, they also have different user behavior as shown in **(A1) & (A2)**, so service providers can also provide some dedicated pricing packages for users in different age groups.* For example, they can provide weekly commuting packages (including weekends) **(A3)** or coupons for young users **(A1)**. In addition, we can convert the current mileage and usage duration-based strategy $F(mileage, duration)$ into a more comprehensive pricing strategy $F(mileage, duration, season, age, gender)$ by considering seasons, age and gender of users **(G1) & (G2)**, etc.

7.3 Limitations and Future Work

Our study has several limitations about the generalizability of findings, which are similar to many other IMWUT works [13, 17, 27, 29, 39] due to the specific region or product investigation.

First of all, we only investigated car sharing user behavior in China. The measurement results could be biased due to the particular characteristics of Chinese cities. Even though a comparative study may help us to understand the generalizability of our findings, we still do not have sufficient evidence to claim that our findings are generalizable to other cities or countries due to different economic environment, policies, meteorological conditions, demographics, etc. Overall, to generalize our findings and to understand if there are statistically significant differences in the general population, there should be comparative studies with other cities and countries in the future.

Another limitation is that the survey results might also be biased because the participants were all from China. The country-specific perception may not be the same as participants in other countries. Furthermore, we only collected 185 survey results, which may not be representative of public attitudes. In the near future, we will perform a large-scale online survey to broader users to understand their user behavior.

In general, our study was an initial exploration with the limitations that come from the specific sample used. Because exploratory work is generative, we were able to identify a number of perspectives and trends that are relevant to understanding user behavior towards adopting car sharing services. However, a different sample may reveal additional insight and nuance. Furthermore, while studying user behavior in online car sharing services helped us understand the real-world factors that prevent them from adoption, finding-based applications will be useful for verifying potential benefits. Hence, in the future, we will integrate our findings to later car sharing operation, which may potentially enhance user experiences.

8 CONCLUSION

In this paper, we conducted a mixing qualitative and quantitative study to understand comprehensive user behavior in online car sharing systems. We performed this study by adopting a mixed-methods approach, which includes an online survey and real-world car sharing operation. The attitude-aware online survey provides us some qualitative perceptions of people to online car sharing. We also conducted a detailed quantitative analysis based on a real-world car sharing operation dataset collected from the Chinese city Beijing. We dissected this attitude-free dataset to understand the objective car sharing user behavior from diverse dimensions, e.g., spatiotemporal, demographic, and revisitation. In addition, we conducted a comparative study to show if the obtained findings are generalizable to other cities of different urban features. We also do a case study by designing a user behavior-aware usage prediction model (i.e., BeXGBoost) to show the significance of our user study work. Finally, we provided a set of findings obtained from our study, combined with some detailed discussions about design implications.

ACKNOWLEDGMENTS

The authors would like to thank anonymous reviewers for their valuable comments and suggestions. Huijun Sun is partially supported by NSFC (91846202), State Key Laboratory of Rail Traffic Control and Safety (RCS2020ZI002), Beijing Jiaotong University, and Science and Technology Foundation of Beijing Jiaotong University (2016RC012).

REFERENCES

[1] Henrik Becker, Francesco Ciari, and Kay W Axhausen. 2017. Comparing Car-Sharing Schemes in Switzerland: User Groups and Usage Patterns. *Transportation Research Part A-policy and Practice* 97 (2017), 17–29.

[2] Joan-Isaac Biel, Nathalie Martin, David Labbe, and Daniel Gatica-Perez. 2018. Bites 'n'Bits: Inferring Eating Behavior from Contextual Mobile Data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 1–33.

[3] Chiara Boldrini, Raffaele Bruno, and Marco Conti. 2016. Characterising demand and usage patterns in a large station-based car sharing system. In *IEEE Conference on Computer Communications (INFOCOM)*. 572–577.

[4] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101.

[5] Robin N Brewer and Vaishnav Kameswaran. 2019. Understanding Trust, Transportation, and Accessibility through Ridesharing. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–11.

[6] Lawrence D Burns. 2013. Sustainable mobility: a vision of our transport future. *Nature* 497, 7448 (2013), 181.

[7] Francesco Calabrese, Francisco C Pereira, Giusy Di Lorenzo, Liang Liu, and Carlo Ratti. 2010. The geography of taste: analyzing cell-phone mobility and social events. In *international conference on pervasive computing*. 22–37.

[8] Chu Cao, Zhidan Liu, Mo Li, Wenqiang Wang, and Zheng Qin. 2018. Walkway discovery from large scale crowdsensing. In *2018 17th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. IEEE, 13–24.

[9] Hancheng Cao, Zhilong Chen, Fengli Xu, Yong Li, and Vassilis Kostakos. 2018. Revisitation in urban space vs. online: A comparison across POIs, websites, and smartphone apps. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 4 (2018), 1–24.

[10] Hancheng Cao, Jagan Sankaranarayanan, Jie Feng, Yong Li, and Hanan Samet. 2019. Understanding metropolitan crowd mobility via mobile cellular accessing data. *ACM Transactions on Spatial Algorithms and Systems (TSAS)* 5, 2 (2019), 1–18.

[11] Hancheng Cao, Fengli Xu, Jagan Sankaranarayanan, Yong Li, and Hanan Samet. 2019. Habit2vec: Trajectory semantic embedding for living pattern recognition in population. *IEEE Transactions on Mobile Computing* (2019).

[12] Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 785–794.

[13] Ying-Yu Chen, Ziyue Li, Daniela Rosner, and Alexis Hiniker. 2019. Understanding Parents' Perspectives on Mealtime Technology. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 1 (2019), 1–19.

[14] Eunjoon Cho, Seth A Myers, and Jure Leskovec. 2011. Friendship and mobility: user movement in location-based social networks. (2011), 1082–1090.

[15] Serdar Colak, Lauren P Alexander, Bernardo G Alvim, Shomik Raj Mehndiratta, and Marta C Gonzalez. 2015. Analyzing Cell Phone Location Data for Urban Travel: Current Methods, Limitations, and Opportunities. *Transportation Research Record* 2526, 2526 (2015), 126–135.

[16] Jonathan Corcoran, Tiebei Li, David Rohde, Elin Charles-Edwards, and Derlie Mateo-Babiano. 2014. Spatio-temporal patterns of a Public Bicycle Sharing Program: the effect of weather and calendar events. *Journal of Transport Geography* 41 (2014), 292–305.

[17] Vedant Das Swain, Koustuv Saha, Hemang Rajvanshy, Anusha Sirigiri, Julie M Gregg, Suwen Lin, Gonzalo J Martinez, Stephen M Mattingly, Shayan Mirjafari, Raghu Mulukutla, et al. 2019. A Multisensor Person-Centered Approach to Understand the Role of Daily Activities in Job Performance with Organizational Personas. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 4 (2019), 1–27.

[18] Monitor Deloitte. 2017. Car Sharing in Europe Business Models, National Variations and Upcoming Disruptions. https://www2.deloitte.com/content/dam/Deloitte/de/Documents/consumer-industrial-products/CIP-Automotive-Car-Sharing-in-Europe.pdf.

[19] Verena Distler, Carine Lallemand, and Thierry Bellet. 2018. Acceptability and acceptance of autonomous mobility on demand: The impact of an immersive experience. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–10.

[20] 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), 1–21.

[21] Raghu Ganti, Mudhakar Srivatsa, Anand Ranganathan, and Jiawei Han. 2013. Inferring Human Mobility Patterns from Taxicab Location Traces. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '13)*. ACM, New York, NY, USA, 459–468.

[22] Marta C Gonzalez, Cesar A Hidalgo, and Albert-Laszlo Barabasi. 2008. Understanding individual human mobility patterns. *nature* 453, 7196 (2008), 779.

[23] Prasad Gupte. 2015. What is the difference between car rental and car sharing? https://www.quora.com/What-is-the-difference\-between-car-rental-and-car-sharing.

[24] 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), 1–19.

[25] 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*. 239–252.

[26] Eiichi Iwamoto, Masaki Matsubara, Chihiro Ota, Satoshi Nakamura, Tsutomu Terada, Hiroyuki Kitagawa, and Atsuyuki Morishima. 2018. Passerby Crowdsourcing: Workers' Behavior and Data Quality Management. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 4 (2018), 1–20.

[27] Hayeon Jeong, Heepyung Kim, Rihun Kim, Uichin Lee, and Yong Jeong. 2017. Smartwatch wearing behavior analysis: a longitudinal study. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–31.

[28] Simon L. Jones, Denzil Ferreira, Simo Hosio, Jorge Goncalves, and Vassilis Kostakos. 2015. Revisitation Analysis of Smartphone App Use. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 1197–1208.

[29] Vaishnav Kameswaran, Lindsey Cameron, and Tawanna R Dillahunt. 2018. Support for social and cultural capital development in real-time ridesharing services. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–12.

[30] Johanna Kopp, Regine Gerike, and Kay W Axhausen. 2015. Do sharing people behave differently? An empirical evaluation of the distinctive mobility patterns of free-floating car-sharing members. *Transportation* 42, 3 (2015), 449–469.

[31] Neal Lathia and Licia Capra. 2011. How smart is your smartcard?: measuring travel behaviours, perceptions, and incentives. In *Proceedings of the 13th international conference on Ubiquitous computing (UbiComp)*. ACM, 291–300.

[32] Quannan Li, Yu Zheng, Xing Xie, Yukun Chen, Wenyu Liu, and Wei-Ying Ma. 2008. Mining user similarity based on location history. In *Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems*. 1–10.

[33] Xuan Lu, Zhenpeng Chen, Xuanzhe Liu, Huoran Li, Tao Xie, and Qiaozhu Mei. 2017. Prado: Predicting app adoption by learning the correlation between developer-controllable properties and user behaviors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–30.

[34] Man Luo, Bowen Du, Konstantin Klemmer, Hongming Zhu, Hakan Ferhatosmanoglu, and Hongkai Wen. 2020. D3P: Data-Driven Demand Prediction for Fast Expanding Electric Vehicle Sharing Systems. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 1, Article 21 (March 2020), 21 pages.

[35] Man Luo, Hongkai Wen, Yi Luo, Bowen Du, Konstantin Klemmer, and Hongming Zhu. 2019. Dynamic Demand Prediction for Expanding Electric Vehicle Sharing Systems: A Graph Sequence Learning Approach. *arXiv preprint arXiv:1903.04051* (2019).

[36] Daniel Malinsky and Peter Spirtes. 2019. Learning the Structure of a Nonstationary Vector Autoregression. *Proceedings of machine learning research* 89 (2019), 2986.

[37] Akhil Mathur, Lakshmi Manasa Kalanadhabhatta, Rahul Majethia, and Fahim Kawsar. 2017. Moving Beyond Market Research: Demystifying Smartphone User Behavior in India. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–27.

[38] Neil McMulkin. 2016. Is There a Convergence of Car Rental and Car Sharing Models? https://medium.com/@neil.mcmulkin/is-there-a-convergence-of-car-rental-and-car-sharing-models-cceddd3fc1f7.

[39] Abhinav Mehrotra, Sandrine R Müller, Gabriella M Harari, Samuel D Gosling, Cecilia Mascolo, Mirco Musolesi, and Peter J Rentfrow. 2017. Understanding the role of places and activities on mobile phone interaction and usage patterns. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–22.

[40] Xinru Page, Paritosh Bahirat, Muhammad I Safi, Bart P Knijnenburg, and Pamela Wisniewski. 2018. The Internet of What? Understanding Differences in Perceptions and Adoption for the Internet of Things. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 4 (2018), 1–22.

[41] Sasank Reddy, Min Mun, Jeff Burke, Deborah Estrin, Mark Hansen, and Mani Srivastava. 2010. Using mobile phones to determine transportation modes. *ACM Transactions on Sensor Networks (TOSN)* 6, 2 (2010), 1–27.

[42] Angela Saunders. 2018. WE ARE 3 MILLION! https://blog.car2go.com/2018/02/07/we-are-3-million/.

[43] Susan Shaheen, Adam Cohen, and Mark Jaffee. 2018. Innovative mobility: carsharing outlook. https://escholarship.org/uc/item/49j961wb.

[44] Frances Sprei, Shiva Habibi, Cristofer Englund, Stefan Pettersson, Alexey Voronov, and Johan Wedlin. 2019. Free-floating car-sharing electrification and mode displacement: Travel time and usage patterns from 12 cities in Europe and the United States. *Transportation Research Part D-transport and Environment* 71 (2019), 127–140.

[45] Vijay Srinivasan, Christian Koehler, and Hongxia Jin. 2018. RuleSelector: Selecting conditional action rules from user behavior patterns. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 1–34.

[46] Frost & Sullivan. 2019. Number of car sharing users worldwide from 2006 to 2025 (in millions). https://www.statista.com/statistics/415636/car-sharing-number-of-users-worldwide/.

[47] Zhen Tu, Runtong Li, Yong Li, Gang Wang, Di Wu, Pan Hui, Li Su, and Depeng Jin. 2018. Your apps give you away: distinguishing mobile users by their app usage fingerprints. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 1–23.

[48] Patrick Vogel, Torsten Greiser, and Dirk Christian Mattfeld. 2011. Understanding bike-sharing systems using data mining: Exploring activity patterns. *Procedia-Social and Behavioral Sciences* 20 (2011), 514–523.

[49] Guang Wang, Xiuyuan Chen, Fan Zhang, Yang Wang, and Desheng Zhang. 2019. Experience: Understanding long-term evolving patterns of shared electric vehicle networks. In *The 25th Annual International Conference on Mobile Computing and Networking*. 1–12.

[50] Guang Wang, Wenzhong Li, Jun Zhang, Yingqiang Ge, Zuohui Fu, Fan Zhang, Yang Wang, and Desheng Zhang. 2019. sharedCharging: Data-driven shared charging for large-scale heterogeneous electric vehicle fleets. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–25.

[51] Guang Wang, Xiaoyang Xie, Fan Zhang, Yunhuai Liu, and Desheng Zhang. 2018. bCharge: Data-Driven Real-Time Charging Scheduling for Large-Scale Electric Bus Fleets. In *2018 IEEE Real-Time Systems Symposium (RTSS)*. IEEE, 45–55.

[52] Guang Wang, Fan Zhang, Huijun Sun, Yang Wang, and Desheng Zhang. 2020. Understanding the Long-Term Evolution of Electric Taxi Networks: A Longitudinal Measurement Study on Mobility and Charging Patterns. *ACM Transactions on Intelligent Systems and Technology (TIST)* 11, 4 (2020), 1–27.

[53] Guang Wang, Yongfeng Zhang, Zhihan Fang, Shuai Wang, Fan Zhang, and Desheng Zhang. 2020. FairCharge: A data-driven fairness-aware charging recommendation system for large-scale electric taxi fleets. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 1–25.

[54] Shuai Wang, Tian He, Desheng Zhang, Yuanchao Shu, Yunhuai Liu, Yu Gu, Cong Liu, Haengju Lee, and Sang H Son. 2018. BRAVO: Improving the Rebalancing Operation in Bike Sharing with Rebalancing Range Prediction. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 2, 1 (2018), 44.

[55] Xiangye Xiao, Yu Zheng, Qiong Luo, and Xing Xie. 2010. Finding similar users using category-based location history. In *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*. 442–445.

[56] Fengli Xu, Tong Xia, Hancheng Cao, Yong Li, Funing Sun, and Fanchao Meng. 2018. Detecting popular temporal modes in population-scale unlabelled trajectory data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 1–25.

[57] Zidong Yang, Ji Hu, Yuanchao Shu, Peng Cheng, Jiming Chen, and Thomas Moscibroda. 2016. Mobility modeling and prediction in bike-sharing systems. In *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys)*. ACM, 165–178.

[58] Daben Yu, Zongping Li, Qinglun Zhong, Yi Ai, and Wei Chen. 2020. Demand Management of Station-Based Car Sharing System Based on Deep Learning Forecasting. *Journal of Advanced Transportation* 2020 (2020).

[59] 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.

[60] Cheng Zhang, Linan Zhang, Yangdong Liu, and Xiaoguang Yang. 2018. Short-term Prediction of Bike-sharing Usage Considering Public Transport: A LSTM Approach. In *International conference on intelligent transportation systems*. 1564–1571.

[61] 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 (MobiCom)*. ACM, 201–212.

[62] Fuzheng Zhang, Nicholas Jing Yuan, David Wilkie, Yu Zheng, and Xing Xie. 2015. Sensing the pulse of urban refueling behavior: A perspective from taxi mobility. *ACM Transactions on Intelligent Systems and Technology (TIST)* 6, 3 (2015), 37.

[63] Shuo Zhang, Khaled Alanezi, Mike Gartrell, Richard Han, Qin Lv, and Shivakant Mishra. 2018. Understanding group event scheduling via the outwithfriendz mobile application. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 1–19.

[64] Tengxiang Zhang, Xin Yi, Ruolin Wang, Jiayuan Gao, Yuntao Wang, Chun Yu, Simin Li, and Yuanchun Shi. 2019. Facilitating Temporal Synchronous Target Selection through User Behavior Modeling. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 4 (2019), 1–24.

[65] Liming Zhao, Jianli Yin, and Yao Song. 2016. An exploration of rumor combating behavior on social media in the context of social crises. https://www.sciencedirect.com/science/article/pii/S0747563215302685. *Computers in Human Behavior* 58 (2016), 25–36.

[66] Yu Zheng, Licia Capra, Ouri Wolfson, and Hai Yang. 2014. Urban Computing: Concepts, Methodologies, and Applications. *ACM Trans. Intell. Syst. Technol.* 5, 3 (Sept. 2014), 38:1–38:55.

[67] Yu Zheng, Quannan Li, Yukun Chen, Xing Xie, and Wei-Ying Ma. 2008. Understanding mobility based on GPS data. In *Proceedings of the 10th international conference on Ubiquitous computing (UbiComp)*. ACM, 312–321.

[68] Yu Zheng, Yanchi Liu, Jing Yuan, and Xing Xie. 2011. Urban computing with taxicabs. In *International Conference on Ubiquitous Computing (Ubicomp)*. 89–98.

[69] Yu Zheng, Lizhu Zhang, Xing Xie, and Wei-Ying Ma. 2009. Mining interesting locations and travel sequences from GPS trajectories. In *Proceedings of the 18th international conference on World wide web*. 791–800.

[70] Zipcar. 2019. zipcar overview. https://www.zipcar.com/press/overview.

[71] Zoran Zivkovic. 2004. Improved adaptive Gaussian mixture model for background subtraction. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, Vol. 2. IEEE, 28–31.