Data Analytics for Location-Based Services: Enabling User-Based Relocation of Carsharing Vehicles

CONFERENCE PAPER · DECEMBER 2015
DOI: 10.13140/RG.2.1.1595.6966

4 AUTHORS, INCLUDING:

Sebastian Wagner
University of Freiburg
6 PUBLICATIONS 4 CITATIONS

Christoph Willing
University of Freiburg
1 PUBLICATION 0 CITATIONS

Tobias Brandt
University of Freiburg
15 PUBLICATIONS 11 CITATIONS
Data Analytics for Location-Based Services: Enabling User-Based Relocation of Carsharing Vehicles

Completed Research Paper

Sebastian Wagner  
University of Freiburg  
Platz der Alten Synagoge  
79085 Freiburg, Germany  
sebastian.wagner@is.uni-freiburg.de

Christoph Willing  
University of Freiburg  
Platz der Alten Synagoge  
79085 Freiburg, Germany  
christoph.willing@is.uni-freiburg.de

Tobias Brandt  
University of Freiburg  
Platz der Alten Synagoge  
79085 Freiburg, Germany  
tobias.brandt@is.uni-freiburg.de

Dirk Neumann  
University of Freiburg  
Platz der Alten Synagoge  
79085 Freiburg, Germany  
dirk.neumann@is.uni-freiburg.de

Abstract

In this paper we demonstrate the potential of data analytics methods for location-based services. We develop a support system that enables user-based relocation of vehicles in free-floating carsharing models. In these businesses, customers can rent and leave cars anywhere within a predefined operational area. However, due to this flexibility, free-floating carsharing is prone to supply and demand imbalance. The support system detects imbalances by analyzing patterns in vehicle idle times. Alternative rental destinations are proposed to customers in exchange for a discount. Using data on 250,000 rentals in the city of Vancouver, we evaluate the relocation system through a simulation. The results show that our approach decreases the average vehicle idle time by up to 16 percent, suggesting a more balanced state of supply and demand. Employing the system results in a higher degree of vehicle utilization and leads to a substantial increase of profits for providers.

Keywords: Carsharing, Data Analytics, Location-Based Services, Spatial Analysis

Introduction

Over the past decade, carsharing as a new mode of transportation has received great attention, both in public discussions and academia. This is reflected in carsharing membership numbers, which have increased from 16,000 to more than a million between 2002 and 2013 in North America alone (Shaheen and Cohen 2013). The recent success of carsharing is caused by several factors. On the one hand, carsharing can contribute to solving today’s urban transportation problems—congested streets, a shortage of parking spaces, and harmful emissions (Baptista et al. 2014; Firnkorn and Müller 2011). On the other hand, carsharing benefits from—and is, simultaneously, representative for—the rise of the sharing economy and collaborative consumption (Belk 2014). People increasingly value access over ownership and owning a car is becoming less of a status symbol than, for instance, owning the newest-generation smartphone.

Carsharing has also captured the interest of information systems (IS) scholars because it merges several current IS research streams. First and foremost, it is a location-based service (Dhar and Varshney 2011; Gerpott and Berg 2011; Junglas and Watson 2008). While it is linked to a specific physical asset, reservations, unlocking, and payment are handled using mobile devices and electronic commerce.
applications. Particularly free-floating carsharing (FFCS) poses tremendous location-based challenges to the provider. In such a business model, customers can end their rentals anywhere within the operation area of the carsharing provider and are not restricted to stations. As a result, supply and demand for vehicles in particular areas of a city vary throughout the day and may not always be in balance. This problem is more pronounced than for station-based systems since the comparatively low number of stations focuses demand at these points. Addressing this imbalance between supply and demand for FFCS systems relates to work within the IS discipline on geographical information systems (GIS) and spatial analysis (Bendler et al. 2014; Pick 2004; Walsham and Sahay 1999). Finally, carsharing models generate huge amounts of data on trips, vehicles, customers, and transactions, resulting in enormous potential for big data analytics applications (Hsinchun et al. 2012). IS scholars have already begun to apply novel analytics methods to these data sources to provide strategic and operational decision support to carsharing providers (Rickenberg et al. 2013; Wagner et al. 2014).

Building upon this foundation, the following paper demonstrates how data analytics can improve operation and efficiency of carsharing and mobility services in general. We address the issue of demand and supply imbalances by analyzing information on the idle times of vehicles for particular areas in the city of Vancouver. These imbalances constitute a major challenge to the financial viability of FFCS business models. Since customers do not need to return cars to fixed stations, they may end their rentals in areas of low demand. For instance, customers returning from the high-demand city center to residential areas on the outskirts of the city leave the cars in these low-demand areas. As a result, there is an excess of idle vehicles in areas where they are not needed and, simultaneously, a lack of vehicles in the city center. The provider needs to send out technicians who relocate the vehicles to areas of high demand, which is a substantial cost driver of FFCS models. We develop a bi-modular support system for FFCS providers, illustrated in Figure 1, which uses data analytics methods and incentive pricing to enable user-based relocation of carsharing vehicles. Using data on more than 250,000 rentals, the analytics module employs density estimation to analyze spatial and temporal variations of idle times across the provider’s business area in Vancouver. Based on this big data analysis, the relocation module suggests parking zones with a low expected idle time to the customer to end the rental in and provides price incentives, such as free rental minutes. We determine the impact of user-based relocation by conducting a simulation based on real carsharing usage patterns. Customers accept the proposal with a certain probability, based on the distance between the suggested parking zone and the original destination. Hence, this paper aims to provide answers to the following research questions.

1. How does the idle time of vehicles vary with respect to the spatial and temporal dimensions?
2. How can the share of rentals ended in areas of low idle time be increased?
3. What is the impact of the user-based relocation system on the FFCS business model?

Our solution aims at balancing and thereby improving the system for both customers and providers. Vehicles will be more frequently available when and where they are truly needed. As a result, cars are expected to be better utilized, generate more rentals, and create additional revenue for carsharing providers.

The remainder of this paper is structured as follows. In the next section, we summarize research related to the relocation problem for carsharing vehicles and put our data-driven approach in this context. In Section 3, a detailed description of our dataset is given and the analytics module is introduced. Section 4 explains the user-based relocation methodology enabled by the relocation module. We discuss the results and give managerial implications in Section 5 before the paper concludes in Section 6.

---

**Figure 1. FFCS Support System for User-Based Relocation**
Related Work

The success of free-floating carsharing as a novel business model is determined by the usage behavior of the customers. Throughout the entire day, cars are used for different purposes, such as running errands, driving to work, or dining out (Millard-Ball 2005). The change in demand for vehicles over time and how demand is distributed across the city are a result of changes in the relevance of these destinations during the day. Celsor and Millard-Ball (2007) point out that demographic factors may influence demand, as well, since one-person households and households that do not own a vehicle are more inclined to use carsharing. Wagner et al. (2014) show that both demographic indicators and information about the structural environment can be used to estimate demand variations in an urban area.

Herrmann et al. (2014) emphasize that the variations of vehicle demand and supply pose a crucial challenge to FFCS models and necessitate the implementation of relocation strategies. However, most research on carsharing relocation has been conducted for station-based systems (see Cepolina and Farina (2012) and Jorge and Correia (2013) for an overview of studies). Relocation strategies are generally divided into two groups: user-based and operator-based relocation.

For one-way station-based carsharing systems, Jorge et al. (2014) show that operator-based relocation leads to a more balanced system. Staff of the carsharing provider carries out the relocation by means of towing or ride-sharing (Barth and Todd 1999; Bruglieri et al. 2014; Cepolina and Farina 2012; Kek et al. 2009). (Barth and Todd 1999; Bruglieri et al. 2014; Cepolina and Farina 2012; Kek et al. 2009). An approach similar to towing, relocation in bikesharing systems is generally achieved by using trucks to pick up and redistribute bicycles (Benchimol et al. 2011; Raviv et al. 2013). Naturally, due to the limited capacity of trucks, this approach is not as suitable for carsharing. Kek et al. (2009) develop decision criteria to balance supply between overstocked and understocked stations. Clemente et al. (2013) outline that operator-based relocation is more costly than user-based relocation and, thus, user-based mechanisms are preferable. For user-based strategies, ride-sharing or ride-splitting (Barth et al. 2004), suggesting alternative destination to users (Di Febraro et al. 2012), and other price incentives (Clemente et al. 2013) have been proposed as possible strategies.

While FFCS can be imagined as a station-based system with an infinite number of stations, it is hard to apply the aforementioned relocation strategies due to fundamental differences. For instance, a key factor in station-based relocation strategies is the limited number of parking spaces at stations (Kek et al. 2009). For FFCS, parking spaces are a different issue, since only the distance between the parking space and the intended destination is relevant. Up to now, there has only been little research examining the relocation problem in an FFCS environment. Weikl and Bogenberger (2013) develop conceptual models, which help to determine optimal vehicle distributions and derive appropriate relocation strategies. They introduce an offline demand clustering model to predict recurring demand patterns based on historical data and an online module to determine the difference between the current vehicle distribution and the optimum distribution. For actual relocation, they consider both operator-based and user-based approaches and provide general guidance on which relocation strategy to use under what circumstances. The operator-based strategies include relocation by staff or using buffer depots to balance demand. For user-based mechanisms, price discounts and trip sharing or splitting are proposed. However, in contrast to our research, they do not evaluate the models with real-world data. Herrmann et al. (2014) build upon Weikl and Bogenberger’s (2013) work and test the acceptance and practicality of user-based relocation in an FFCS context. Using a survey of current and potential carsharing users, they found that a majority of respondents are willing to “book a more distant car for a 10 cent/mile price reduction” (85%) and are willing to “indicate their destination at the beginning of the trip” (89%) (Herrmann et al. 2014, p. 156). Due to the price sensitivity of carsharing users, the authors derive four user-based relocation methods. Providing a discount to book a more distant car (1); providing a discount to leave the vehicle at a more distant destination (2); “paying” users, e.g. with free minutes or kilometers, for specific relocation routes (3); and offering ride and cost-sharing in high demand areas with low availability of cars (4).

The research presented in this paper extends the state-of-the-art in several ways. By employing spatial data analytics methods, we develop a user-based relocation system that incentivizes users through free minutes. This approach combines geographical information systems and novel big data analytics techniques. We specifically design this system for free-floating carsharing models, since they exhibit the highest degree of flexibility, and thereby complementing public transportation for a more environmentally friendly hybrid...
system of urban transportation. The relocation system uses data on idle times of vehicles in Vancouver, which are the reflection of demand and supply imbalances in the city. Thereby, we are able to evaluate a theoretical model with real-world data to improve on current relocation models.

**Analytics Module: Visualizing Carsharing Usage in Vancouver**

In Vancouver, carsharing is well established and a common business model with several active providers, such as zipcar, modo, or car2go. Therefore, we choose this city to analyze carsharing usage patterns during two months in early 2015. In this section, we provide descriptive statistics of the dataset and outline how the analytics module (cf. Figure 1) of the FFCS support system produces the insights necessary for user-based relocation procedures.

**Operation Area and Spatial Distribution of Rentals**

In FFCS systems, the provider has to determine a specific region where people are allowed to end their car rentals. For our reference city, this business area covers approximately 120 square kilometers and consists of four polygons illustrated in Figure 2a. We have one home area in the center of Vancouver and three smaller regions, one in the north (North Vancouver), one in the south (Richmond), and a last one in the west (Wesbrook).

In total, the analytics module investigates more than 250,000 rentals of 730 cars over two months. A rental can be regarded as a single one-way trip with an origin and a destination. Each trip contains information on the time and date the car is rented and when it is returned, which allows calculating the duration of each rental. Due to privacy reasons, it is not possible to obtain detailed information on the exact route customers have taken, as well as on all stops in between. However, the minimum distance a car is driven has to be from the start to the end point of the rental. We use this information to calculate the minimum trip distance and duration. The Bing-API is used to consider the actual course of Vancouver’s roads given a situation without any heavy traffic. This enables the calculation of lower bounds for distance and time. After a rental ends, the car is idle until the next customer rents it again. In summary, a rental $r_i$ is defined as the following 6-tuple

$$r_i = (start_{lat,lon}, end_{lat,lon}, \tau_{start}, \tau_{end}, \tau_{drive}, \tau_{idle})$$

(Eq. 1)
We define the granularity of the business area and affect the computational complexity. The values vary even stronger. It averages at about two hours, but the standard deviation is more than twice as high. Since idle vehicles are a major cost driver of FFCS systems, the purpose of the analytics module is to provide insights on how idle time varies across time and space.

The set of all rentals \( r_i \) in our dataset is \( R \). Table 1 provides descriptive statistics of the dataset on rentals and vehicles. Evidently, the dataset is quite heterogeneous. Trip distances according to our Bing-estimates vary between only a few hundred meters (rounded to zero kilometers) and 75 kilometers, with a mean and a standard deviation of four kilometers, respectively. The estimated duration of the trip according to the Bing-API varies similarly. Interestingly, the idle time, i.e. the time between successive rentals of a vehicle, varies even stronger. It averages at about two hours, but the standard deviation is more than twice as high.

The idle time is not a characteristic of a specific rental or vehicle, but rather of the location where a rental ends. It reflects how supply and demand for carsharing services match in the vicinity. To map the information on idle times of individual vehicles to the business area, we first divide the latter into a grid of thousands of individual tiles. Each tile \( t_i \) is defined by GPS-coordinates for its center and a predefined edge length. Hence, the business area is given by a set \( T \) of tiles \( t_i \) as follows:

\[
T(\Delta \lambda, \Delta \phi, n) = \{t_1, t_2, ..., t_n\}
\]

\( n \in \mathbb{N}^+ \) positive number of tiles

\( t_i \) Tile \( t_i \) with center at GPS-latitude, \( \lambda_i \), GPS-longitude, \( \phi_i \)

\( \Delta \lambda, \Delta \phi \) Changes in latitude, longitude to define edge length of tiles

The values \( \Delta \lambda \) and \( \Delta \phi \) determine the edge length of the tiles and, thus, their size. Therefore, the values also define the granularity of the business area and affect the computational complexity of the subsequent relocation approach. The smaller the edge length the greater the number of tiles considered by the FFCS support system. However, as soon as the edge length drops below a certain threshold, neighboring tiles barely differ. We tested different thresholds for the edge length (50m, 100m, 200m, 500m) and decided to use a latitude delta of \( \Delta \phi = 0.0018 \) and a longitude delta of \( \Delta \lambda = 0.002764 \). These values result in an edge length of 200 meters and likewise ensure both high granularity and a manageable complexity. Applying these values to QGIS – an open-source desktop geographical information system application –, the resulting business area for Vancouver consists of \( n = 3,088 \) tiles.

As a next step, we need to estimate the expected idle time for each tile \( t_i \in T \). For this purpose, we use a kernel density estimation (KDE) to map each rental \( r_k \) to a number of tiles based on the rental's end time and destination. For a given tile, the kernel considers all rentals within a radius of 500 meter (bandwidth \( \sigma = 0.5 \)). We assume a linearly decreasing relationship based on the distance between rental and tile. Moreover, we analyze variations in the idle time values in the temporal dimension in addition to the spatial dimension. For this purpose, we apply a moving average that extends an hour into the future and into the

Table 1. Descriptive statistics of FFCS in Vancouver

<table>
<thead>
<tr>
<th></th>
<th>Minimum</th>
<th>Average</th>
<th>Maximum</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trip distance (Bing)</td>
<td>0 km</td>
<td>4 km</td>
<td>75 km</td>
<td>4 km</td>
</tr>
<tr>
<td>Trip duration (Bing)</td>
<td>0 min</td>
<td>9 min</td>
<td>62 min</td>
<td>6 min</td>
</tr>
<tr>
<td>Idle time</td>
<td>2 min</td>
<td>116 min</td>
<td>6675 min</td>
<td>236 min</td>
</tr>
<tr>
<td>#rentals per day</td>
<td>2,742</td>
<td>5,480</td>
<td>7,222</td>
<td>823</td>
</tr>
<tr>
<td>#rentals per car</td>
<td>0</td>
<td>8</td>
<td>22</td>
<td>3</td>
</tr>
</tbody>
</table>

Thirty Sixth International Conference on Information Systems, Fort Worth 2015
past for each hour of the day. For instance, the average idle time for the time slot between 2 p.m. and 3 p.m. is calculated by including all rentals between 1 p.m. and 4 p.m. that end within a 500 meter radius of the tile. The chosen geographic bandwidth relies on different theories on human behavior. For example according to Tobler’s first law of geography, “everything is related to everything else, but near things are more related than distant things” (Tobler 1970). This is of particular interest in the case of relocation, since customers always try to get as close as possible to the desired destination, in order to minimize the walking distance. If relocation is offered and accepted by the user, the walking distance increases with respect to the proposed location.

Figure 2b shows the daily rental frequency based on the kernel for the carsharing business in Vancouver. The plot shows a heat map based on all 250,000 rentals without considering temporal differences. While cars in North Vancouver and Richmond are rented quite rarely, the downtown area of Vancouver is highly attractive for carsharing. This provides a first indication of where cars might run idle due to a lack of demand.

Figure 2b also illustrates FFCS characteristics that lead to the exceptionally high standard deviations in Table 1. Since people are allowed to perform any desired trip as long as the location where the rental ends is within the business area (cf. Figure 2a), one resulting phenomenon is that cars are often parked close to the borders. This happens mainly in the following two cases. First, customers drive far beyond the enclosed areas in Figure 2, such as another city in case of business or amusement trips. In this case, usually the shortest path back to the business area is selected to reduce the costs for returning the car. In the second scenario, the final destination is not located inside the operation area. However, customers use carsharing to get as close as possible. The intention is to minimize either walking distance or to select a place from which it is easy to continue the trip by alternative means such as public transport. In both cases, the destination of the rental is close to the borders of the business area. As a result, cars may run idle afterwards because the borders are usually less attractive for carsharing compared to the city center or other highly frequented locations.

The phenomenon described above appears only in free-floating models, since the final destinations of station-based carsharing are predefined. Another main difference is that parking spaces at carsharing stations are given, while customers using free-floating models are allowed to park at any legal parking spot. Consequently, the imbalance between supply and demand of vehicles at a certain location is limited in station-based systems, while it is almost unlimited for FFCS. Furthermore, supply and demand characteristics do not have to be constant during the day. For instance, while it may be necessary to have a large supply of cars downtown, since the average rental frequency is high, in the morning hours vehicles may also be needed in residential areas outside the city center to convey people to work.

To account for these features, the analytics module calculates the expected idle time $z_{ij}$ for each hour of the day $j$ and for each tile $t_i \in T$ based on the set of rentals $R$. First, a subset of rentals $R_{ij}$ is defined for each tile that contains all rentals that end within a certain distance $\sigma$ of that tile during the time slot $j$.

$$R_{ij} \subseteq R \text{ with } R_{ij} = \{ r_k \in R \mid \text{dist}(t_i, r_k) \leq \sigma \land \text{hour}(\tau_{k,\text{end}}) = j \}$$

(Eq. 3)

$$\text{dist}(t_i, r_k) \quad \text{distance between tile } t_i \text{ and rental } r_k$$

$$\sigma \quad \text{distance threshold}$$

$$\text{hour}(\tau_{k,\text{end}}) \quad \text{hour of the time the rental ends}$$

The expected idle time is subsequently calculated as

$$z_{ij} = \frac{\sum_{k=1}^{\left| R_{ij} \right|} \tau_{k,\text{idle}} \left( 1 - \frac{\text{dist}(t_i, r_k)}{\sigma} \right)}{\frac{\sum_{k=1}^{\left| R_{ij} \right|}}{\left( 1 - \frac{\text{dist}(t_i, r_k)}{\sigma} \right)}} \quad \text{with}$$

(Eq. 4)

$$\tau_{k,\text{idle}} \quad \text{as the idle time at hour } j \text{ of tile } t \in T \text{ at position } i.$$

Hence, the time a car is idle after a rental ends adds to the expected idle times of all tiles within its vicinity bounded by $\sigma$ at the hour during which the rental ended, but is weighted by the distance between tile and rental. As a result, we receive values for the idle time for each hour of the day and all tiles in $T$. 

---

*Data Analytics for Location-Based Services*

*Thirty Sixth International Conference on Information Systems, Fort Worth 2015*
Visualizing Idle Time

Figure 3 shows the geographically distributed idle times at two different times of the day, thereby providing insights on the first research question. It illustrates the change in supply and demand dynamics between different times of day. As already indicated by Figure 2b, the northern parts of Vancouver’s home area provide low idle times at any given point in time. By comparing Figure 3a with Figure 3b we can further observe that carsharing is used more frequently during the day. Especially at night, North Vancouver and Richmond show idle times of several hours. Since every location has its own characteristics, such as financial districts or shopping streets, they are appealing to different customers and at different times. With some establishments being closed over the weekend, these locations might be interesting only during weekdays. Hence, a relocation mechanism needs to reflect these spatial and temporal variations and incentivize users to end rentals in areas with low idle times that are close to their original destination. The resulting questions of where, when, and how to relocate carsharing vehicles to the right spots will be investigated in the subsequent section.

Relocation Module: Incentivizing Customers

The illustrations in the previous sections visualize how carsharing supply and demand dynamics shift during the course of the day. These shifts also reflect the changes in the purposes of trips at different daytimes. For instance, cars may be used to commute to work in the morning hours, while amusement activities become more common in the evening and at night. As areas within the city cater to different interests, carsharing usage patterns change. Following Di Febbraro et al. (2012) and Herrmann et al. (2014), we consider a relocation system where customers get a discount for parking at a proposed location. In practice, such a system could be realized by offering the customer a certain amount of free minutes for the next ride if they accept to park somewhere else. The amount of free minutes would have to be sufficiently large to motivate customers to relocate. There are also upper limits to giving away premiums, the most obvious being that the premium must not exceed the expected reduction in idle time. Furthermore, the premium approach should not be too complicated for customers to grasp. For illustrative purposes, we, therefore, assume a static incentive of five free minutes. Figure 4 provides a high-level representation of the user-based relocation procedure that is executed by the relocation module (cf. Figure 1).

The relocation process has three main phases. In the initialization phase, the customer enters the desired destination into the car’s navigation system. If the destination of the rental is not given initially, relocation

![Figure 3. Business area idle times during different hours of the day in Vancouver](image-url)
cannot be offered and the possibility to earn free minutes is missed. Furthermore, the provider has to make sure that it is not possible to change the final destination afterwards and that relocation is offered only once. If the customer still decides to park somewhere else, a normal rental is conducted and it is not possible to earn the free minutes for the offered relocation.

Determining the relocation zone and providing an offer to the customer are the main tasks of the second phase. The analytics module calculates the expected idle time for the entered destination at the projected time of arrival and determines the corresponding idle time of tiles in the near vicinity. Without loss of generality, we assume that this vicinity parameter is equal to \( \sigma \) and use a radius of 500 meters for both. The relocation module then calculates the difference in idle times between the original destination and these tiles. If the difference exceeds a threshold value \( \omega \), which should be larger than the amount of free minutes credited to the customer, the on-board computer suggests to park in this area—the so-called relocation zone—instead of the original destination. Specifically, the relocation zone \( Y_{ij} \subset T \) for tile \( i \) (the original destination) during time slot \( j \) is defined as

\[
Y_{ij} = \{ t_h \in T \mid \text{dist}(t_i, t_h) \leq \sigma \land z_{hj} < z_{ij} - \omega \}. \tag{Eq. 5}
\]

This is further illustrated in Figure 5. The relocation module attempts to identify an area, in which the calculated idle times are lower than at the original destination. Consider the case that the customer’s final destination is at the top of the mountain in the three-dimensional illustration of Figure 5a. Here, the system tries to relocate the car into the surrounding valleys. The greater the difference in altitude, the lower is the expected idle time and the better is the result from vehicle relocation. Hence, the relocation zone only contains locations, in which the altitude (expected idle time) is lower by \( \omega \) or more, compared to the current location.

In Figure 5b, we can see a two-dimensional representation of Figure 5a as a contour plot. The lines represent the different expected idle times, while the relocation module again attempts to route cars from low demand (large number) to high demand (small number) zones. It might even be beneficial to include tiles into the relocation zone, which are within the same contour as the original destination. In this case, we have an “altitude” difference of 50 minutes between each contour. Thus, relocating a vehicle from a location with an expected idle time of 140 minutes to a location with 110 minutes still makes sense, although both locations are within the same contour. As we can see, the illustrated region only extends across one square kilometer but shows large variations in idle time from more than 400 minutes to less than 100 minutes. This implies that even if cars are relocated only by a short distance, the effect on the carsharing business model can be enormous. However, in some cases, it might happen that the original end location is already the best place to return the car for the given time of day, e.g., in the city center during lunchtime. If such a situation occurs, the module will not offer relocation.

In the last phase, the customer decides, based on the proposed offer, to either take the free minutes and park within the relocation zone, or to decline the offer and park at the original destination. Depending on the customer’s decision, the car is parked at a different location implying different expected idle times for future rentals.
Evaluating User-Based Relocation

As shown in Equation 5, the relocation area consists of a set of tiles, for which the difference in idle times compared to the original destination exceeds $\omega$. This threshold is one of the key values within the relocation process, since it determines the number of tiles that qualify for the relocation zone and the minimal expected reduction in idle time. On the one hand, if we choose an exorbitantly large $\omega$, the number of tiles in $Y_{th}$ converges to zero and relocation will never be offered to the customer. On the other hand, if the $\omega$ value is too low, relocation will be proposed at almost any location, even ones directly adjacent to the tile of the original destination. Hence, the $\omega$ value is indirectly linked to the distance a customer has to drive to receive the free minutes bonus. Furthermore, low $\omega$ values may require the carsharing provider to give out more free minutes than are compensated by reduced idle times. Therefore, we test several $\omega$ values to identify the optimal configuration for the relocation system.

To evaluate the FFCS support system, the probability that customers accept the relocation offers has to be modeled. Since relocation is not mandatory and only a special offer to the customer to earn extra credits for the next rental, there is a certain probability that the offer is refused. One of the main reasons people decide to park at a specific location is to minimize the necessary walking distance afterwards. Hence, if the proposed relocation spot increases the walking distance to the final destination of the customer, the probability of accepting the relocation offer decreases. Another factor that may influence the decision is the additional carsharing time and associated costs to the customer to get to the relocation zone. However, in contrast to the walking distance which always increases since the customer enters the exact desired destination, the driving distance to the relocation zone does not necessarily have to be longer than to the original destination. The system might suggest an area closer to the point of origin and, thus, decrease time and distance of the rental, while also offering free minutes to the customer. In the opposite case, the relocation zone is further away from the origin. As a result, the free minutes are less beneficial to the customer because she partly uses them to get to the relocation zone. Weighting both cases, we assume that the influence of additional driving is negligible on average and model the acceptance decision only dependent on the increased walking distance.

Similar to Hermann et al. (2014), we assume that the probability $\rho$ of accepting to relocate from a given location $t_i$ to a proposed location $t_{ih} \in Y_{ij}$ is based on the distance in between. We define this probability $\rho_{ih}$ as

![Figure 5. Exemplary idle times for an area of one square kilometer in Vancouver](image)
\[
\rho_{ih} = 1 - \frac{\text{dist}(t_i, t_h)}{\sigma} \quad .
\]

(Eq. 6)

Since the relocation area \(Y_{ih}\) usually consists of a set of different locations, we assume that the probability of accepting relocation in general is determined by

\[
\rho_{T} = \max_{\rho_{ih}} Y_{ih} \quad ,
\]

i.e. the maximum probability of any tile in \(Y_{ih}\). If relocation is accepted, the new destination tile is determined by weighting all probabilities of tiles in \(Y_{ih}\).

After a car is relocated, future trips are expected to start earlier, which creates a completely new dynamic and is challenging to simulate for evaluation. Since the vehicle is now parked at a different location than in the original dataset and is expected to have a decreased idle time, the next rental is likely to diverge from the original dataset, as well. The simulation addresses this issue by selecting a random rental that starts in the \(\sigma\)-vicinity of the new location and during the new time slot from the set of all rentals. If multiple rentals from within that vicinity have the same destination, each has a separate chance of being drawn. Thereby, the appeal of particular destinations is implicitly taken into account.

However, before the impact of relocation is simulated, a data cleaning process is applied to the set of historical rentals. We remove all rentals that show an abnormal behavior or which are affected by transmission errors and sensor failures. For instance, the cleaning process deletes rentals with a negative or exorbitantly high duration. As a next step, we split the original dataset into an independent training set and a smaller test set, in order to ensure robustness of the evaluation results. Hence, more than 250,000 rentals within the training set are used to estimate the expected idle time at a specific time and place and determine relocation suggestions. This data is used to train the model (calculate expected idle time) in order to select optimal relocation spots for future rentals. The rentals from the test set are subsequently used to validate the approach by determining actual idle times for the simulated vehicle. Each iteration of the simulation takes a random car on a random day from the test set and calculates the impact of user-based relocation. Algorithm 1 describes this process in pseudo code.

As a first step, an initialization takes place (line 1) during which a random car is selected from the test set on a random day. The first trip of this vehicle on this day \(r_0\) is taken as a starting point, along with its end time. For instance, if the first rental of a car ends at 8:30 a.m., the time value is initialized with 8 hours 30 minutes \((j = 8)\) and the destination of the rental is the first tile to check for relocation. In lines 5 and 6, the relocation area \(Y_{ij}\) and the probability of accepting the offered relocation service are calculated for the current time slot. It is important to note that the suggested relocation area is based on data from the training set. In most cases, the system is able to offer a relocation zone but the service has to be accepted as well (if-condition in line 7).

Once the customer accepts, the actual relocation takes place in lines 9 to 11. The new location of the rental is determined by selecting a tile \(t_h \subseteq Y_{ij}\) based on the customer acceptance probabilities \(\rho_{ih}\). The time value is updated using the average idle time of the relocation tile based on the test set. At this point—after the first relocation event has occurred—cars follow new rental sequences different from the original dataset. For the selection of future destinations, we select a random rental from the test set that has started within a radius of \(\sigma\) of the relocation tile during the updated time slot \(j\). Thus, the simulation does not just pick a random rental out of the dataset but rather selects a trip that was actually conducted in the vicinity at this time of day.

If relocation is rejected, the idle time of the original destination \(t_i\) is added to the time value and a subsequent rental is selected (lines 13 to 18). As mentioned above, the future rental is either the historical next rental of the respective car or a common rental based on the current time and location if relocation had previously occurred during the simulated day. Since the time value is updated after each rental regardless of whether relocation has taken place, the while-condition in line 3 is false after a certain number of rentals and the day for the current car has ended. The performance of the simulated sequence of rentals is subsequently compared to the performance of the car in the original dataset.
For illustration, Figure 6 shows a number of relocation events for the North Vancouver area between midnight and 1 a.m. The beginning of each arrow represents the original destination, while the end shows the offered relocation spot. The arrow length implies the distance between original and proposed destination. As we can see, the FFCS support system employs user-based relocation to get out of the red (high idle time) and into the yellow or even green (low idle time) zones. Figure 6 also confirms that the system works as intended since there are no arrows in the opposite direction pointing from yellow or green to red zones.

Table 2 summarizes the simulation results for different idle time thresholds $\omega$ and provides insights on research question 3. As a baseline, we set the threshold to infinity ($\omega = \infty$). As a result, the number of tiles in the relocation zone $|Y_{ij}|$ is always 0, according to Equation 5. This leads to a simulation run where no relocation is offered to the customer. The average idle time is slightly less than documented in Table 1 as a result of the sample split but reasonably close given the high standard deviation.

Table 2 shows that the idle time is reduced by five percent for each rental on average, leading to a simulation run where no relocation is offered to the customer. The average idle time is slightly less than documented in Table 1 as a result of the sample split but reasonably close given the high standard deviation.

The first value we test is $\omega = 10$. This is effectively the minimal realistic threshold since it more than compensates the offered premium to the customer (5 minutes) as the idle time is expected to be reduced by at least ten minutes. Table 2 shows that the idle time is reduced by five percent for each rental on average, and as a result, eight percent more trips will be performed per car and day. While this may appear to be a substantial improvement for the carsharing provider, we need to consider the costs of offering free minutes to the customer. Relocation is offered for 58 percent of all rentals and the acceptance rate is very high (78 percent) due to a large number of tiles within the relocation zone that are close to the original destination. As a result, the number of relocation events is very high, as well, and the provider has to credit five free minutes after almost every other rental.

**Algorithm 1: Pseudo code of the relocation algorithm using a test set**

```
1 Initialize car, time, and first rental $r_0$ with time slot $j$,
2 WHILE time < 24 hours:
3    Get destination tile $t_i$ for rental $r_x$
4    Calculate $Y_{ij}$ (from training set)
5    Calculate probability $\rho_Y$ for $Y_{ij}$
6    IF $|Y_{ij}| > 0$ AND relocation is accepted based on $\rho_Y$:
7        Get relocation tile $t_h \in Y_{ij}$
8        time = time + $z_{hj}$ (from test set)
9        new time slot $j = hour(time)$
10       Select next rental $r_{x+1}$ out of rental list in $t_h$ at $j$
11    ELSE:
12        time = time + $z_{hi}$
13        new time slot $j = hour(time)$
14        IF relocation has previously occurred:
15            Select next rental $r_{x+1}$ out of rental list in $t_i$ at $j$
16        ELSE:
17            Continue with next rental $r_{x+1}$ out of rental list in car
18            time = time + $t_{end,x+1} - t_{start,x+1}$
19        new time slot $j = hour(time)$
20  new time slot $j = hour(time)$
```
With an increasing threshold $\omega$, the acceptance rate drops and the share of rentals for which relocation is offered decreases. Since $\omega$ correlates with the average distance customers have to drive to get to the proposed relocation destination, this distance increases slightly for larger thresholds. Interestingly, the average distance to the center of Vancouver remains the same considering both the start and end locations of the rentals. The results imply that relocation does not necessarily route cars to the center of the city but rather to nearby locations with low expected idle times.

Table 2 further shows that a threshold of $\omega = 60$ leads to the highest reduction of idle time at 16 percent, while increasing vehicle utilization by 15 percent. Additionally, relocation is offered for only 29 percent of all rentals and accepted by the customers with a probability of 72 percent. By comparing the configurations of $\omega = 30$ and $\omega = 60$ we can see that the values for idle time and number of rentals per car are almost the same. However, ten percent more relocation events are conducted by using a threshold of $\omega = 30$. As a result, the provider’s investment in terms of free minutes is substantially higher for the same outcome.

**Managerial Implications**

The results presented in the last section show that the FFCS support system introduced in this paper decreases the average idle time per vehicle and increases vehicle utilization by offering user-based relocation.
relocation incentives to customers. Depending on the structure of the provider’s variable costs, the approach also increases profits as long as the threshold \( \omega \) is appropriately chosen. Taking the results for \( \omega = 60 \) from Table 2 into account, without relocation 8.3 rentals per day and car are each followed by 106 minutes of idle time. Hence, vehicles are rented on average for 560 minutes per day. With relocation, vehicles are rented 9.5 times per day with each rental followed by an idle time of 89 minutes, resulting in 595 rental minutes per day, a gain of 35 minutes. The rental price per minute in our study is CAD 0.41, however, this includes gas costs and depreciation. Assuming that the provider has a marginal profit of CAD 0.3 per minute results in CAD 7,665 per day over 730 vehicles. If 20 percent of all rentals are associated with a relocation as the simulations results suggest, the provider grants free minutes 1.9 times per car and day. Taking the example of five free minutes into account results in a revenue loss of CAD 2,843 per day. This results in a substantial net profit increase of CAD 4,822 per day or CAD 1,760,000 per year just for the City of Vancouver (approximately USD 1,450,000). In addition, cars are used more efficiently and the free minute bonus can serve as a customer relationship measure, binding them more closely to the company. Such a static price incentive is easily implemented and communicated to customers. For instance, FFCS providers generally rely on a smartphone app for vehicle location and booking. Hence, this app can be enhanced to offer relocation to the customers or the onboard navigation system can be used to suggest alternative destinations. When booking a car via the respective app, customers could already be asked for their desired destination to offer relocation even before the car is moved. The chosen destination could then be submitted to the car’s navigation system automatically, making the carsharing experience even more user-friendly.

To increase user acceptance of the relocation proposal and in order to make use of price discrimination, providers could also consider a dynamic zone-based incentive system. A possible application could be a system where users see the real-time demand dynamics in their vicinity while they are driving. From a technical point of view, the onboard navigation system or a mobile app would display different zones on the screen. They represent variations in the expected idle times, illustrated by Figure 7. For an easy identification, zones are emphasized by color; the distance and estimated driving time to the closest “good” zone can be calculated and displayed based on the current location. Users receive premiums according to the zone in which they end their rental. For instance, the incentive system could offer a low premium of one free minutes for parking in the yellow zone, a medium premium of two minutes for the green zone, up to a high premium of three free minutes for the blue zone. Depending on the current location users might have to drive more than 500 meters to switch the zone (position 1 in Figure 7). On the other hand, customers who originally intended to park in a high-demand area can receive the premium without adapting their behavior (position 2 in Figure 7) and are rewarded for parking in areas with low idle times. Such a system has many benefits for both customer and provider. Carsharing users who do not know their exact destination in advance, such as tourists, are able to collect credits without major drawbacks. Customers do

![Figure 7. Dynamic user-based relocation using zones](image)
not have to give any information on their final destination beforehand and, thus, destinations can change in the course of the rental. However, the implementation of a dynamic incentive scheme is slightly more complex, compared to the static incentive, as it requires the constantly updated visualization of real-time data. It also raises several questions for future research, such as, how to design the user interface in a manner that is easy to use. After all, drivers should not get lost in looking at the screen but still drive safely. Furthermore, it needs to be investigated whether customers accept the additional complexity and whether the system is prone to exploitation by customers who aim to maximize their premiums.

Conclusion

In this paper, we analyze and address the relocation problem in free-floating carsharing models for the city of Vancouver as showcase for the application of data analytics techniques to location-based services. We introduce a system supporting user-based relocation in free-floating carsharing, which offers free minutes for the next ride to the customer in return for parking at a specific location. The approach is trained and evaluated on a real-world data set, in order to obtain accurate and realistic results. Thereby, we are able to provide general decision support for any kind of free-floating carsharing model independent of the respective location, business area, or price system. Furthermore, our proposed algorithm has machine-learning character and is self-adapting to customer behavior. The spatially and temporally dependent idle times become more accurate as new data is collected by the system, thereby, improving precision. If usage patterns change, such as through the opening of a new shopping mall, these changes are subsequently reflected by the relocation process. In such a case, the idle time may drop and the respective area will be recommended as part of the relocation zone for arriving customers.

We have shown that spatial and temporal differences are highly relevant to develop effective and profitable relocation algorithms for carsharing businesses. In order to offer the right relocation spot at the right time, methods for spatial data analysis are applied. Our results show that a user-based relocation approach is able to reduce the idle time of carsharing cars substantially. Thereby, the utilization of vehicles is increased. Furthermore, the introduced methodology does not route cars to the city center per se but rather to urban areas where they are more likely to be picked up. This provides a general contribution in the area of carsharing research since most previous approaches (mainly operator-based) aim to relocate cars closer to the center without analyzing actual rental data.

Additionally, the developed algorithm is easy to implement within existing smartphone apps. Therefore, providers are able to test the approach and validate our findings by offering relocation during a restricted period, such as a Christmas special. In this case, investment costs remain assessable for the provider. Hence, we are able to create a win-win situation for both customers by satisfying mobility needs and providers by increasing vehicle utilization tremendously. Eventually, the insights of this paper take us one step further to reach the overall goal of building a self-regulating system where cars are located where and when they are truly needed. We expect that different kinds of location-based services, such as station-based carsharing models, bike sharing, or taxi and ride-sharing services, will also benefit from these insights.

For future research, we will apply this approach to datasets from 30 different cities all over the world including the United States, Canada, Germany, Austria, Italy, Denmark, Sweden, and the Netherlands. We will test different incentive systems, including variations of the thresholds for distance and free minutes, to further refine the overall business model. Acknowledging that the respondents in Herrmann et al. (2014) were only from the city of Hamburg, Germany, it could be valuable to study relocation acceptance in further detail for different cities where FFCS is offered, either via additional surveys or via real world testing by providers. Recently there have been efforts to estimate parking space availability in urban areas (Richter et al. 2014). Incorporating this information into the relocation proposal could prove highly beneficial to customers and studying its effect on the relocation acceptance rate would be an interesting extension. Additionally, we will include an operator-based relocation strategy to offer a possibility to the provider to move a car if relocation was denied. Comparisons between operator-based and user-based relocation with regard to costs and effectiveness will provide a comprehensive picture on how to best balance the fleet. Furthermore, we will analyze side effects and endogeneity of idle times in more detail. Assuming perfect relocation implies more cars in highly frequented areas. Therefore, a vehicle threshold needs to be calculated at which a certain area is saturated for a given point in time. If this threshold is reached or even exceeded, the relocation approach needs to exclude the respective area from the recommended zones. For our analysis, we believe that this effect is marginal, since only a gradual relocation to the near vicinity is
proposed. Furthermore, relocation strategies can be enhanced by considering urban Points of Interests (POIs). Research has shown that POIs are able to estimate demand for areas with limited information. Thus, their inclusion into the analysis could circumvent the problem of hidden demand in areas without sufficient historical carsharing data.

References


