Article

A Predictive Vehicle Ride Sharing Recommendation System for Smart Cities Commuting

Theodoros Anagnostopoulos [1,2] (ID)

1   DigiT.DSS.Lab, Department of Business Administration, University of West Attica, 250 Thivon & P. Ralli Str, Egaleo, 12241 Athens, Greece; Theodoros.Anagnostopoulos@uniwa.gr
2   Department of Infocommunication Technologies, ITMO University, Kronverksiy Prospect, 49, 197101 St. Petersburg, Russia

Abstract: Smart Cities (or Cities 2.0) are an evolution in citizen habitation. In such cities, transport commuting is changing rapidly with the proliferation of contemporary vehicular technology. New models of vehicle ride sharing systems are changing the way citizens commute in their daily movement schedule. The use of a private vehicle per single passenger transportation is no longer viable in sustainable Smart Cities (SC) because of the vehicles' resource allocation and urban pollution. The current research on car ride sharing systems is widely expanding in a range of contemporary technologies, however, without covering a multidisciplinary approach. In this paper, the focus is on performing a multidisciplinary research on car riding systems taking into consideration personalized user mobility behavior by providing next destination prediction as well as a recommender system based on riders' personalized information. Specifically, it proposes a predictive vehicle ride sharing system for commuting, which has impact on the SC green ecosystem. The adopted system also provides a recommendation to citizens to select the persons they would like to commute with. An Artificial Intelligence (AI)-enabled weighted pattern matching model is used to assess user movement behavior in SC and provide the best predicted recommendation list of commuting users. Citizens are then able to engage a current trip to next destination with the more suitable user provided by the list. An experimented is conducted with real data from the municipality of New Philadelphia, in SC of Athens, Greece, to implement the proposed system and observe certain user movement behavior. The results are promising for the incorporation of the adopted system to other SCs.

Keywords: smart cities; vehicle ride sharing; user commuting; prediction; recommendation; artificial intelligence

1. Introduction

Smart Cities (or Cities 2.0) are the future of citizen habitation. The Internet of Things (IoT) evangelize a new era of transport commuting, which is changing rapidly with the proliferation of contemporary vehicular technology. Vehicle ride sharing systems are changing the way citizens commute in their daily movement schedule by adopting new models of transportation scheduling. Innovation is mainly focused on the principle that a private vehicle can be used to commute more than one passenger in sustainable Smart Cities (SC). Such utility can overcome the vehicles' resource allocation and urban pollution by car ride sharing.

SC is a sustainable environment, which should be efficiently faced. Specifically, cyber resilience and incident response are emerging issues in SC. IoT can address cyber resilience and support digital forensic incident response aspects towards a green ecosystem [1]. Efficient and integrated SC is discussed in [2], where the authors present a research network for sustainable SC to develop a methodology for strategic planning. Such planning enables cities of certain region to achieve sustainable solutions for their citizens combining strengths from participating research groups and collaborating institutions. Digital systems as

a service are proposed in [3], where authors treat SC infrastructure issue as a viable solution for citizens. Digitalization is used in every aspect of daily experience in SC to achieve the independence of the physical infrastructure. Specifically, priority is given to well-designed digital solutions towards a sustainable ecosystem.

IoT is the backbone of SC emergence in new era evangelizing the use of sensors and actuators to everyday life. Sustainable SC are based on advances in IoT technology where certain challenges and approaches are emerged. Addressing the exponential growth of contemporary urbanization and population IoT is key for outlining innovations applied to SC infrastructure [4]. IoT also enables the incorporation of disruptive technologies in SC where current trends and challenges can be examined in more depth [5]. The benefits of urbanization create emerged challenges related to limited resources and infrastructure. Such problems are faced by contemporary technologies, which generate changes in key SC sectors to be further handled by disruptive technologies. However, IoT potentiality to SC ecosystem requires citizens' consent to provide advanced services. In such cases, ethics and law requirements should be carefully examined to ensure citizens' privacy and address data security concerns [6].

Vehicular technology is an advanced solution to leverage intelligent transportation systems towards smart vehicles, which exploits crowdsourcing potentiality. Specifically, vehicular social networks can use current SC infrastructure to support mobile, spatial, and passive sensing crowdsourcing techniques. Such technology can improve connectivity and higher throughput where smart vehicles can exploit with embedded sensors and actuators to provide innovative distributed processing techniques for efficient transportation in SC [7]. Software-Defined Networking (SDN) provides the necessary technology to achieve a vehicle routing protocol for improving rush hour delay in SC environments. SDN technology can offer an overview of the examined network to control vehicle transport application specifications. It enables vehicles to be part of a challenging SC infrastructure, which handles delays in everyday transport [8]. The monitoring of vehicle speed in SC is important for the city ecosystem since it proves a context, where drivers are aware of their behavior as well to the impact it has to other vehicle drivers' and pedestrian's safety [9]. Such monitoring is feasible due to contemporary crowdsensing-based technologies, which allow citizens to use their smartphones for large-scale data collection and further processing to provide a safe SC green ecosystem.

Electric Refuse Collection Vehicles (ERCV) are also incorporated in sustainable SC infrastructure to achieve zero-emission municipality services. Such ERCV use a customized battery capacity to enable energy efficient refuse collection, thus providing a green ecosystem [10]. The online and dynamic prediction of actual passengers' transport with rail transit systems is significant for sustainable SC infrastructure. Specifically, to achieve such an effort, the authors in [11] propose the use of deep learning methods to predict efficiently the urban rail transit passenger flow. The system can predict the passenger flow of rail stations during peaks in the week as well as during the limited usage of rail in weekends. Ride hailing is a transport behavior emerging in SC, which affects the mobility choices of public transit as well as ownership of private vehicles. Such behavior leads to the adoption of public transit trips, which changes daily life and transport practices in contemporary SC [12].

A public transport commute system, which is based on mobility recommendations is presented in [13]. Specifically, the authors introduce daily life in certain sustainable SC by focusing on the optimization of transport due to public transport systems. Such a system incorporates every SC public transport like electric bus, public bicycles and electric scooters. The key feature of the system is its ability to recommend a public transport based on citizens' mobility behavior. Interconnected public spaces as areas where citizens use contemporary smart bus commute technology can provide transport services to elderly and impaired users. Such services are able to increase social inclusiveness and provide mobility assistance for senior citizens' well-being [14]. Social commute can be expanded to other forms of vehicular transportation such as taxi commuting. In such a system, the mining of

citizens' mobile patterns is significant to provide a personalized transport service. Dealing with taxi routing trips incorporates certain citizen mobility distribution rules of pick-up and drop-off SC locations, which provide accurate information of the origin and the destination of the commute [15].

An activity-based ride matching (ABRM) system exploits user behavior to provide optimal matching ride requests with ride offers, [16]. The possible reaching destination is achieved with the incorporation of alternative destinations in cases where intended activity can be obtained. A sustainable shared mobility car riding system is proposed for commute, aiming to reduce traffic at rush hours in SCs [17]. Multimodality is examined in [18] where the authors propose a car riding system modelling and solving to assist daily commute of many modes of transport in cities. Context aware technology supports an Autonomous Vehicle (AV) infrastructure for providing congestion-aware ride sharing in SCs [19]. Such a system aims to catalyze car ride sharing ubiquity by incorporating an efficient disruptive technology, which computes optimal travel plans based on traffic load generated by multimodal means of transportation.

Dynamic scheduling is analyzed in conjunction with topology dependence for achieving the online demands of a car riding system during daily commute in rural and urban areas [20]. The authors study the effects of several underlying traffic conditions on the system's performance to provide an optimal solution. The authors in [21] propose an eco-mobility on demand car ride sharing system applied on a fleet of vehicles to eliminate fuel consumption. Their research focuses on Connected Autonomous Vehicles (CAV) technology exploiting mobility on demand for a data driven model to provide fleet control. Operational policies are examined in [22], where the authors focus on occasional ride sharing with regards to mobility on demand services in SCs. In [23], the authors study the effects of privacy regulations on dynamic car riding sharing systems, where they experiment with spatiotemporal data optimization to provide an optimal privacy-based approach for commute in sustainable cities. Pricing options are studied in [24], where the authors present a joint pricing and matching architecture to enable cost efficient ride sharing systems in green ecosystems.

The Uber marketplace is presented in [25], where the authors propose a matching policy for riders and drivers as well as batch matching policy to serve increased ride sharing demand in the SC. Both policies are optimized based on historical data to reduce riders waiting time. Uber's shared mobility system is also examined in [26], where the authors focus on ride sharing services available for AV technology. Ride hailing is also considered as a commute option, which focuses mainly on vehicle distance traveled in rural and urban areas. Uber-Pool is considered in [27], where the research exploits the data analyses of car ride sharing systems with regards to obtained service time and detour guarantees provided by taxi policies. Operational policies in dynamic environments are studied in [28], where the authors focus on analyzing Uber's mobility services. The research focuses mainly on trip pattern matching and searching efficiency to provide detailed knowledge on emerged Uber's operational policies.

Dynamic routing strategies are researched in [29], where the focus is on designing a scalable matching system for exploiting online routing to face car ride sharing demands. Such a system is evaluated by assessing the prediction accuracy of the proposed routes, which are available to riders. An electric vehicles (EV) car ride sharing system is evaluated in [30], where the authors focus on providing efficient data driven on demand prediction accuracy. A car ride sharing system for commuting in SCs is proposed, in [31], where human behavior characteristics are evaluated to propose a recommender system. Such a system also exploits the prediction accuracy of routing trips selected by riders. A recommender system is examined in [32], which uses an expansive search-based model to provide advanced ride sharing services in SCs' daily commute. A multipath planning system for car ride sharing services is proposed in [33], where the authors provide a recommendation system, which is evaluated with dynamic optimization processes. Shortest path clustering is used in [34], to enhance the spatial data mining of a high-capacity car

ride sharing system, which incorporates a recommender system for efficient commute in sustainable cities.

Current research on car ride sharing systems is widely expanding in a range of contemporary technologies. These approaches focus mainly on certain research areas of SC commuting rather than covering a multidisciplinary approach. In this paper, the focus is on performing a multidisciplinary research on car riding systems taking into consideration personalized user mobility behavior, which is further optimized to serve a population of riders. Location prediction is also part of the research effort, which is evaluated with prediction accuracy on selected destinations in SCs. The efficiency of the provided commute choices is also evaluated by the incorporation of a recommender system based on personalized riders' information. Specifically, this paper presents research on smart vehicle commuting systems and proposes a predictive vehicle ride sharing system, which has a positive impact on SCs' green and sustainable environment. The adopted system also provides a recommendation capability to citizens to select the persons they would like to commute with. An Artificial Intelligence (AI)-enabled weighted pattern matching model is used to assess user movement behavior in SC and provide the best predicted recommendation list of commuting users. Citizens are then able to engage a current trip to next destination with the more suitable user provided by the list. An experiment is conducted with real data from the municipality of New Philadelphia, in SC of Athens, Greece, to implement the proposed system and observe certain user movement behavior. The results are promising for the incorporation of the adopted system to other SCs.

2. Materials and Methods

2.1. System Architecture Overview

The proposed system is based on a certain architecture. The system collects users' daily trips in the SC and preforms profiling according to the citizens' movement preferences. For example, a user on certain day starts her daily commute from home early in the morning then reach school to drop-off her children, subsequently go to work and stay there until afternoon. After work, she picks-up kids from the school and is directed to the super market, while at the end of the working day she returns back to home. During this daily trajectory user follows a predefined movement which is based on certain profiling preferences.

Subsequently, the system is able to track each movement of the trajectory and capture valuable information, which can be used to analyze the user's profile. Specifically, on a daily basis, it captures the location of the user's movement in the SC. Such process encodes implicitly time as a sequence of consecutive historic places where the user has visited in the near past during her daily movement trajectory. However, the system stores each user's trajectory and annotates it with certain day temporal information. Concretely, this process is repeated for every user commute in the city, forming a detailed knowledge base of the users' movements in the SC.

Concretely, there is an app which uses information of the previous two tiers to provide a recommendation system based on user predicted future location within her daily schedule. The system app is able to locate the user in the SC coverage area and according to stochastic information to infer, this predicts the location the user is going to visit next according to a certain trajectory. In addition, to assess a commute the system traverses the knowledge base to find other users who have similar movement trajectories and the same predicted destination compared to the pivot user. Then, the system recommends a list of other users who can share the same trip with the pivot user, to take on a shared commute to the same predicted destination with the same vehicle in the SC. The architectural overview of the proposed system is presented in Figure 1.
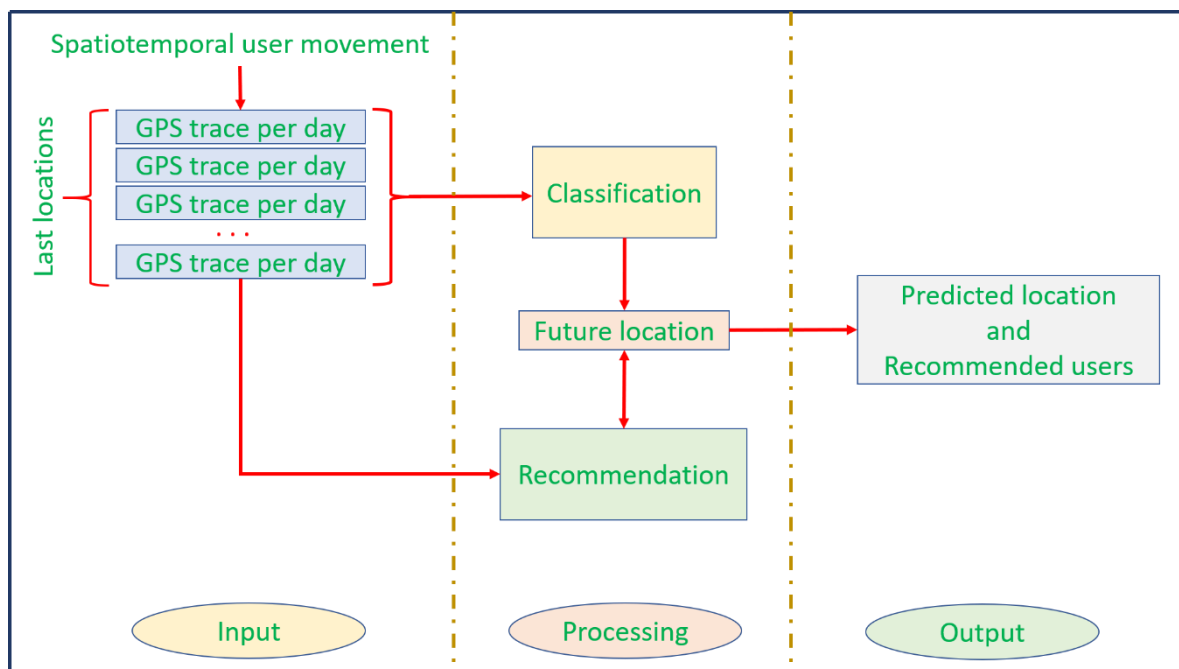
Figure 1. Proposed system's architecture: (1) System contains user's daily trips and provide movement profiling; (2) Subsequently, system implicitly annotates time per day of week and handles the system's knowledge base; (3) Concretely, system provides prediction and recommendation of similar users' trajectories to support a shared vehicle commute.

2.2. AI-Enabled Weighted Pattern Matching Model

The system's back-end intelligence is based on an AI-enabled weighted pattern matching model. Such a model can take as input the current location of a pivot user invoke the stochastic historic places (i.e., historic window size *m*) she has visited in the near past and predict the future location of the user (i.e., prediction window size *l*). Time is encoded implicitly with regards to the sequentially modeling of the represented ordered historic places the user has visited, thus a user cannot get to work early in the morning if she does not leave first her home.

However, note that each user's movement is assigned to a certain day temporal information of movement activity. This consecutive representation of visited places forms a user's movement trajectory. Vehicle ride sharing is feasible since the model searches the knowledge base for other users located nearly to pivot user's location with similar historic visited behavior per certain day of week. In addition, per certain returned user the model is able to check the future location after the current location.

If the future location of a user (i.e., predicted next location) is similar to the pivot user's next location then the user is a candidate to be added to the recommendation list. It holds that similarity between two separate users' trajectories is inferred taking into consideration the spatial historic similarity threshold, μ, which assesses the similarity of the historic locations except of the current location. In addition, the prediction similarity threshold, ϑ, assesses the similarity of the current and the future prediction locations. Historic similarity threshold, μ, is more relaxed compared with prediction similarity threshold, ϑ, since the history information of the trajectory may vary a little between two trajectories. However, this is not the case with the prediction similarity threshold, ϑ, which is less relaxed since it is required that the current and future prediction commute locations should be similar enough in order the model to be more accurate. So, generally, it holds that:

$$\vartheta < \mu \tag{1}$$

Both similarity thresholds are set experimentally to assure that the compared trajectories are similar enough to considered as a match for the adopted model. When both similarity thresholds comparison holds for the whole examined trajectory of a user and the pivot user's trajectory, then that user's trajectory is finally added to the recommendation list. The proposed model returns a list of recommended users (i.e., number of returned recommendations in defined as k), where the pivot user can choose to have a ride sharing in the SC. The characteristic of the shared commute users is that they travel towards to the same direction of the predicted location. The algorithm of the proposed model is presented in Table 1.

Table 1. Prediction and recommendation algorithm of the proposed ride sharing system.

#	Prediction and Recommendation Algorithm
1	**Input:** KB //knowledge base
2	i //examined instance
3	d //day of the week
4	m //historic window size
5	l //prediction window size
6	μ //spatial historic similarity threshold
7	ϑ //spatial prediction similarity threshold
8	k //recommendation list size
9	**Output:** N //returned recommendation list
10	**Begin**
11	$N \leftarrow Null$ //returned recommendation list is empty
12	$k \leftarrow read()$ //initialize recommendation list size
13	$i \leftarrow read()$ //read the examined instance from user mobile app
14	$j \leftarrow read(KB)$ //read the first instance of the KB
15	**While** $(j \neq KB(EoF))$ **Do**//traverse KB
16	**If** $((d(i) = d(j))\ AND\ (i(m) - j(m) \leq \vartheta)\ AND\ (i(l) - (j(l) \leq \vartheta))$ **Then**
17	//if current and predicted locations of i, j are similar w.r.t. ϑ similarity for certain day
18	**For** $(n \in [0,\ m-1])$ **Do**//traverse from first to last historic location of the trajectory
19	**If** $(i(n) - j(n) \leq \mu)$ **Then**//step by step historic comparison
20	$\varphi \leftarrow \varphi + 1$ //historic similarity flag increases
21	**End If**
22	**End For**
23	**If** $(\varphi = n)$ **Then**
24	//if historic similarity condition w.r.t. μ holds proceed to recommendation list step
25	**If** $(size(N) \leq k)$ **Then**//if size of N is less than or equal to k
26	$N \leftarrow N + j$ //recommendation list is expanded
27	**Else**
28	$sort(N)$ //sort recommendations in ascending order of similarity
29	$return(N)$ //return recommendation list and exit
30	**End If**
31	**End If**
32	**End If**
33	**End While**
34	**End**

2.3. Evaluation Method and Metrics

2.3.1. 10-Fold Cross Validation Evaluation Method

The model is evaluated with 10-fold cross validation method, which is a commonly used evaluation method for assessing machine learning classifiers' efficiency. In such a method, the aim is that all instances of the examined dataset will be used in both states for training set and testing set. The training set and testing set are formed by the initial dataset by dividing it to 10 equal sized sets. In a loop of 10 iterations in each iteration 9 of the equal sized sets are used for forming the training set and the one remaining set is used to form the testing set. To avoid overfitting each instance cannot belong at the same time to both the training and testing set. Instead, during this iterative process at each separate

loop an instance can be part of either the training set, or the testing set. Note that since all the instances are used during the evaluation this is not a random process.

2.3.2. Prediction Accuracy Evaluation Metric

Prediction accuracy is an evaluation metric, which is used for assessing the prediction effectiveness of the proposed model with regards to the prediction dimension. Specifically, prediction accuracy, a, is defined as follows:

$$a = \frac{t_p + t_n}{t_p + t_n + f_p + f_n} \tag{2}$$

where t_p are the correctly classified positive instances, t_n are the correctly classified negative instances, f_p are the falsely classified positive instances, and f_n are the falsely classified negative instances. It holds that $a \in [0, 1]$.

2.3.3. Recommendation MAP@N Evaluation Metric

Precision@k is an evaluation metric that is used to assess the precision efficiency of the proposed model with regards to the recommendation dimension. Specifically, precision@k, $p(k)$, is defined as follows:

$$p(k) = \frac{t_p@k}{t_p@k + f_p@k} \tag{3}$$

where $t_p@k$ are the correctly classified positive instances, and $f_p@k$ are the falsely classified positive instances at k recommendations, respectively. In addition, the term, k, denotes the returned number of predicted recommendations per examined instance, such as $k = t_p@k + f_p@k$. As such, Equation (3) is formed as:

$$p(k) = \frac{t_p}{k} \tag{4}$$

It holds that $p(k) \in [0, 1]$. However, $p(k)$ is not that accurate when generalize to a N number of recommended instances for evaluation. In this case, AveragePrecision@N, $ap@N$, is used, which is defined as follows:

$$ap@N = \frac{1}{|KB|} \sum_{k=1}^{N} p(k) \cdot rel(k) \tag{5}$$

where $rel(k)$ is an indicator, which expresses that k^{th} instance was relevant with regards to recommendation, i.e., $rel(k) = 1$ or not, i.e., $rel(k) = 0$. In addition, $|KB|$, is the total number of instances and N is the number of recommended instances. It holds that $ap@N \in [0, 1]$.

However, $ap@N$ is referred to a single user's instance. In case there is a need to expand the range of $ap@N$ to the total number of users, the MeanAveragePrecision@N, $map@N$, evaluation metric should be defined as follows:

$$map@N = \frac{1}{|U|} \sum_{u=1}^{|U|} (map@)_u = \frac{1}{|U|} \sum_{u=1}^{|U|} \frac{1}{|KB|} \sum_{k=1}^{N} p(k) \cdot rel(k) \tag{6}$$

where u is a certain user and $|U|$ is the total number of users. It holds that $map@N \in [0, 1]$.

3. Results

3.1. Experimental Parameters

To assess the efficiency of the research effort, the proposed system is experimented on with real data provided by citizens' movement trajectories observed in the coverage area of New Philadelphia, which is a municipality of the SC of Athens, Greece. The dataset has a size of $|KB| = 2958$ instances containing GPS location information [35]. The experimental dataset is produced by the concurrent movement trajectories of $|U| = 100$ users for $d = 7$ days per

week, in a total period of $w \in [1, 8]$ weeks, within the coverage area of the municipality of New Philadelphia. The adopted dataset is visualized in Figure 2. $|U|$ is the number of the users created the experimental dataset, while $|KB|$ is the size of the dataset in GPS instances that was created by the users. Thus, $|U|$ and $|KB|$ are totally different quantities. w is the number of the weeks, which means that the adopted dataset covers a total period of two months. Specifically, the collected data contain information from 9 November to 9 January 2009. This means that there are not much seasonal data in the examined dataset to exploit stochastic user behavior at a seasonal level. However, there are daily, weekly, and monthly user movement patterns, which are exploited in the performed experiments in Section 3.2. Raw data are collected from the Open Street Maps open-source website. This website provides free available datasets for research and industrial purposes. It also provides a user-friendly Application Programming Interface (API) for prosperous users who would like to access the provided datasets. However, note that all the provide datasets are created from users who have upload them to support the academic and industrial community for further use. Such datasets have certain structure. The experimental dataset it is used in this paper has the following Extensible Markup Language (XML) structure as depicted in Figure 3.
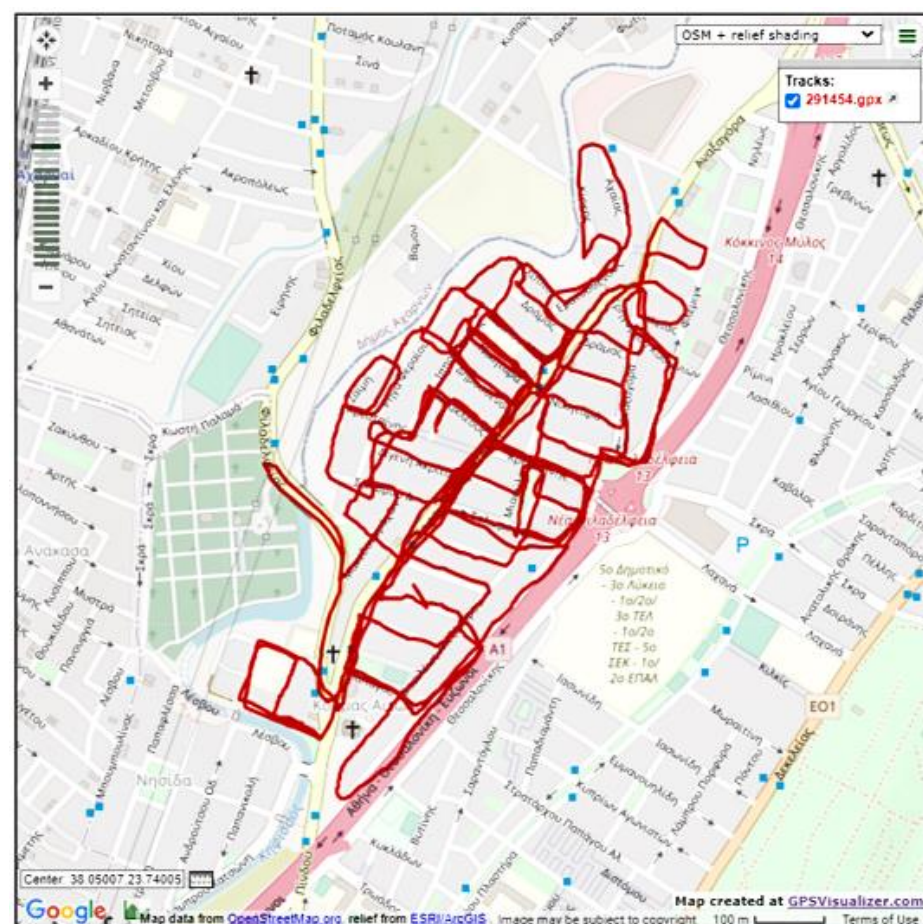

Figure 2. Adopted dataset visualized from the municipality of New Philadelphia in the SC of Athens, Greece.

```
<?xml version="1.0" encoding="UTF-8"?>
<gpx version="1.1" creator="OSMTrack"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns="http://www.topografix.com/GPX/1/1"
xsi:schemaLocation="http://www.topografix.com/GPX/1/1
http://www.topografix.com/GPX/1/1/gpx.xsd">
     <trk>
          <trkseg>
               <trkpt lat="38.048604716667"
lon="23.739073266667">
                    <ele>200.0</ele>
                    <time>2009-01-09T19:44:34Z</time>
               </trkpt>
               <trkpt lat="38.048368683333"
lon="23.739266383333">
                    <ele>160.0</ele>
                    <time>2009-01-09T19:44:48Z</time>
               </trkpt>
               <trkpt lat="38.048368683333"
lon="23.739395133333">
                    <ele>158.0</ele>
                    <time>2009-01-09T19:44:59Z</time>
               </trkpt>
               <trkpt lat="38.048357950000"
lon="23.739416583333">
                    <ele>158.0</ele>
                    <time>2009-01-09T19:44:59Z</time>
               </trkpt>
               <trkpt lat="38.048347233333"
lon="23.739480966667">
                    <ele>157.0</ele>
                    <time>2009-01-09T19:45:00Z</time>
               </trkpt>
               <trkpt lat="38.048336500000"
lon="23.739502416667">
                    <ele>157.0</ele>
                    <time>2009-01-09T19:45:01Z</time>
               </trkpt>
```

Figure 3. Open Street Maps experimental dataset XML structure.

GPS traces have length of 8 decimal digits which is equal of sensitivity of 10 meters. The minimum latitude of the dataset is 38.04582595 and the minimum longitude is 23.73619793. In addition, the maximum latitude is 38.05432318 while the maximum longitude is 23.74390125. Such GPS coordinates form a coverage area of 0.64 square kilometers. The coverage area of New Philadelphia is 2.85 square kilometers. The adopted dataset covers almost 1/4.5 of the coverage area of the municipality. The experimental parameters of the adopted dataset are presented in Table 2.

Table 2. Adopted dataset experimental parameters.

Parameter	Value
GPS traces length	8 decimal digits
Sensitivity	10 meters
Minimum latitude	38.04582595
Minimum longitude	23.73619793
Maximum latitude	38.05432318
Maximum longitude	23.74390125
Coverage area	0.64 square kilometers

Regarding the model, the historic places visited by the user have length of m places (e.g., where historic window size m should be defined experimentally in Section 3.2.1),

while one more place is used to depict the future place (i.e., prediction window size $l = 1$). It holds that the prediction similarity threshold, which measures the similarity of two examined trajectories w.r.t current and future prediction GPS locations is set to $\vartheta = 0.00000001$, (i.e., prediction similarity threshold has sensitivity of 10 meters, which means that it exploits locations with the same city block) [36]. In addition, the historic similarity threshold, which measures the similarity of two examined trajectories w.r.t. historic GPS location is set to $\mu = 0.000001$, (i.e., historic similarity threshold has sensitivity of 100 meters, which means it exploits locations between neighboring city blocks). The user population of the available users required by the system to be operational is defined to be $|U| = 100$ users (i.e., total amount of users). The total number of instances is equal to the adopted dataset size, which is $|KB| = 2958$ instances. When a prediction is achieved the number of the recommended users', which have similar trajectories with the pivot user, is denoted with N (e.g., returned recommendation list should be defined experimentally in Section 3.2.2). Experimental parameters of the proposed model are presented in Table 3.

Table 3. Proposed model experimental parameters.

Parameter	Value
l	1 GPS predicted location
ϑ	0.00000001 (10 m)
μ	0.000001 (100 m)
$|U|$	100 users totally
$|KB|$	2958 instances

3.2. Experiments

3.2.1. Prediction Accuracy

Prediction accuracy, *a*, is calculated based on a 10-fold cross validation evaluation method for 1000 iterations based on a user population of $U = 100$ users for $d = 7$ days per week, in a total period of $w \in [1, 8]$ weeks, and dataset size of 2958 instances containing spatial GPS coordinates. Prediction accuracy, *a*, depends on the value of historic window size *m*. To define which value of *m* is optimal for the current experiment it is tried several values of $m \in [1, 10]$. It is found that for value $m = 6$ proposed system achieves more efficient accuracy than the other values of *m*. Prediction accuracy results for the optimal value of historic window size in presented in Figure 4.
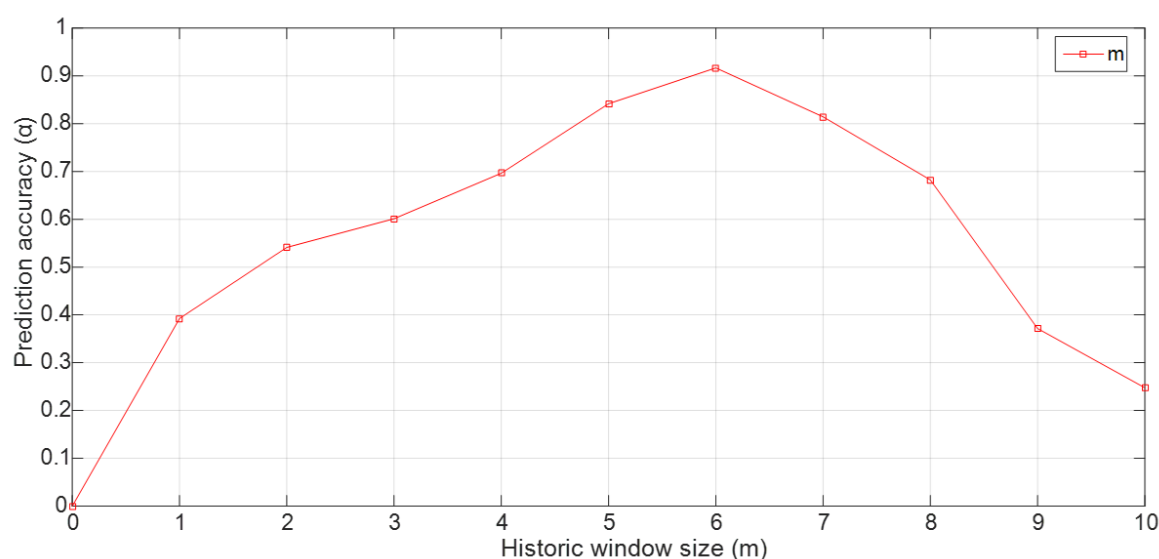

Figure 4. Prediction accuracy results for optimal value of historic window size.

To define in which the value prediction accuracy *a* reaches its optimal value, research effort is experimented with 10-fold cross validation evaluation method for 1000 iterations based on a user population of $U = 100$ users for $d = 7$ days per week, in a total period of $w \in [1,8]$ weeks, and dataset size of 2958 instances containing spatial GPS coordinates. It is observed that *a* is incremental and converges to value $a = 0.9168$ after $w = 4$ week, which means that the adopted system has achieved its higher level of efficiency after the fourth week of experimental evaluation. The prediction accuracy results for optimal system convergence on certain week are presented in Figure 5.
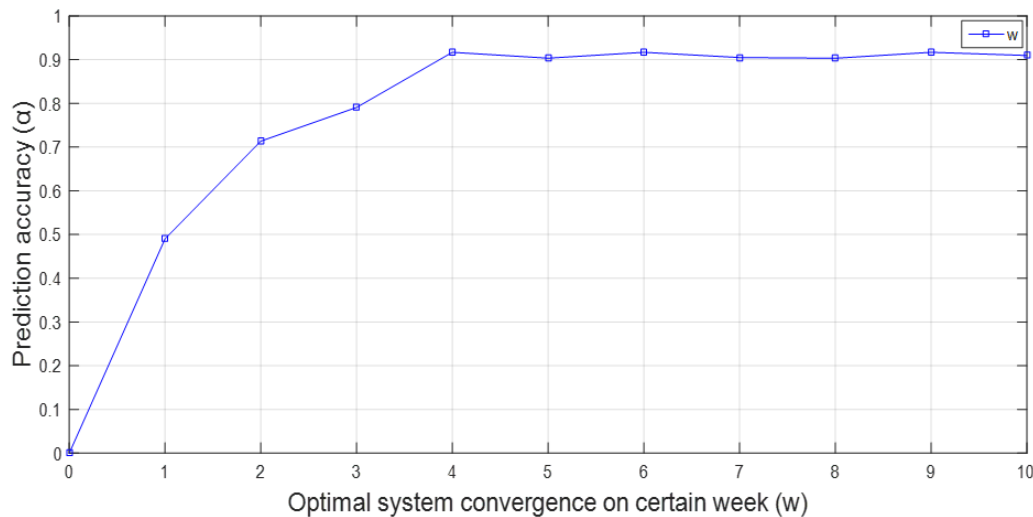

Figure 5. Prediction accuracy results for optimal system convergence on certain week.

3.2.2. Recommendation MAP@N

MAP@N, *map@N*, where $N \in [1, 10]$ are the returned user's recommendations, which are also calculated based on 10-fold cross validation for 1000 iterations run on a user population of $U = 100$ users for $d = 7$ days per week, in a total period of $w \in [1,8]$ weeks, and a dataset size of 2958 instances containing spatial GPS coordinates. Specifically, MAP@N, *map@N*, is incremental and converges to value $map@N = 0.8234$ for value of $N = 5$ recommended users which means that after this value of $N = 5$ recommended users there is not much effect on the system recommendation performance. MAP@N result for optimal value of recommended users are presented in Figure 6.
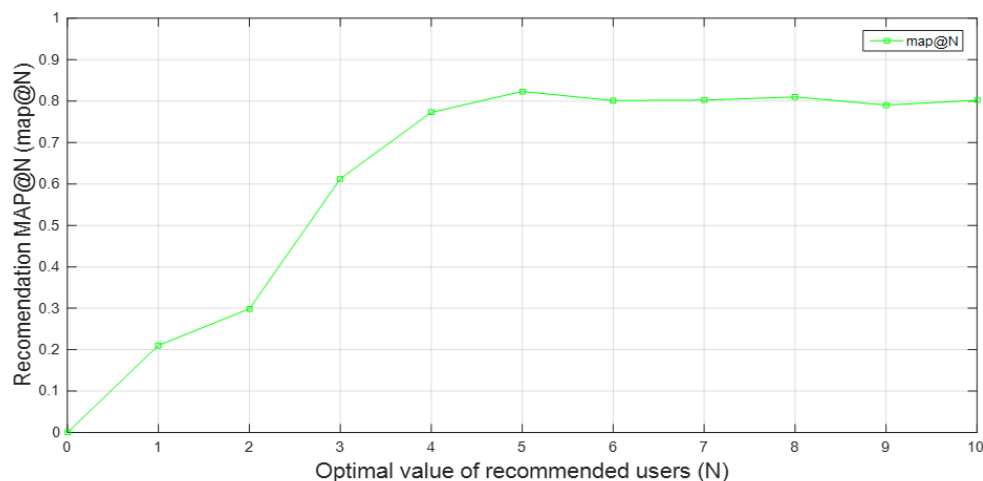

Figure 6. MAP@N result for optimal value of recommended users.

4. Discussion

4.1. Discussion on the Results

The results are based on certain decision of adopted experimental parameters. Such a decision is rational according to common sense for the proposed model and experimental setup. Specifically, the experiments assess the efficiency of prediction accuracy, a, and $map@N$, based on calculated based on 10-fold cross validation for 1000 iterations run on a user population of $U = 100$ users for $d = 7$ days per week, in a total period of $w \in [1,8]$ weeks, and dataset size of 2958 instances containing spatial GPS coordinates. It is experimentally found that the optimal value for historic window size is $m = 6$. This value means that the proposed model is able to converge to an efficient value of prediction accuracy, a, assessing the last $m = 6$ GPS locations of the users' movement. A lower value of m means that the system cannot reach a mature decision since the number of historic locations are few, thus, it is not able to infer a certain user movement pattern. Instead, in case m has a higher value the system also cannot achieve an efficient value since it is not able to capture certain user movement behavior from different GPS places but rather merges them inelegantly.

Having settled to a certain value of historic window size $m = 6$ it is observed how prediction accuracy, a, increments through the adopted experimental setup and converges to a certain value of $a = 0.9168$ after $w = 4$ week of the system evaluation. Actually, the prediction accuracy starts from values near to zero since in the beginning of the evaluation phase there is not enough knowledge to assess the system's effectiveness. However, as time passes more data are gathered and forming a much richer knowledge base. This knowledge base follows an incremental way of maturity till the value of $w = 4$ week. After this value there is no need to train the system since it is already intelligent to assess certain user movement. By forcing the system to be trained more than $w = 4$ week might result in the unfortunate state of overfitting which would deteriorate the efficiency of the system.

Previous experiments evaluated the prediction accuracy, a, of the system with regards to the efficiency that it adopts new instances in order to perform better predictions of the next future location, where the user is going to visit during her daily trip. However, prediction evaluation metrics are not able to assess the effectiveness of the recommendation results, which is the precision of the recommended users proposed by the adopted system. To evaluate the precision of the system, it incorporated the MAP@N, $map@N$, recommendation evaluation metric. This metric defines the optimal number of recommended users the system returns to the pivot user who invokes the proposed system. MAP@N achieves optimal value $map@N = 0.8234$ for $N = 5$ recommended users, which means that for less recommended users the pivot user cannot find a match to ride share a certain commute, while for more recommended users the pivot user does not variate her behavior. This stability of the pivot user behavior after $N = 5$ recommended users explains why the system converges to a certain $map@N = 0.8234$ value.

4.2. Comparison with Other Research Efforts

The proposed system focuses on multidisciplinary research in the field of car ride-sharing systems. The next destination prediction is evaluated through prediction accuracy, a, of the adopted classification model as well as, $map@N$, recommendation evaluation metric. System is based on optimal personalized information of users commute in SCs. Comparing observed results with other research efforts in next destination prediction, such as [29–31], the adopted research effort reaches better values for, a, prediction accuracy evaluation metric. Specifically, the proposed system achieves $a = 0.9168$, while effort in [29] achieves $a = 0.9028$. In addition, the system in [30] achieves $a = 0.9097$, while effort in [31] achieves $a = 0.9125$. Concretely, comparing results with research in other recommender systems, such as [32–34], the proposed system reaches optimal values for, $map@N$, recommendation evaluation metric. Subsequently, the adopted system achieves

$map@N = 0.8234$, while effort in [32] achieves $map@N = 0.8113$. In addition, system in [33] achieves $map@N = 0.8204$, while effort in [34] achieves $map@N = 0.8079$.

Strengths of the proposed approach are the analyses of user mobility behavior towards a sustainable way of transport in rural and urban environments. Such personalized user behavior enables a context-aware car riding system, which faces commute in contemporary green ecosystems. The exact timing for online riders picks up and drop off in areas of interest in the SC is achieved by dynamic scheduling of the proposed system. Concretely its ability to provide online dynamic routing destinations towards next location ride sharing prediction is an advantage of the adopted system. The proposed approach provides its services with free of charge policy, which means that riders are not required to pay for the commute.

The weaknesses of the proposed system are that the adopted approach does not studies multimodal means of transport like other research approaches. Since during the commuting process there is sensitive information that is shared between drivers and riders, compliance with privacy regulations is crucial. However, the proposed system does not cover such a utility in the current stage of research. In addition, current research effort is not tuned to support its efficiency in cases of EV as well as AV and CAV commuting environments. Operational policies are not taken into consideration during this research since it is a separate research area which is not covered by the data provided in this effort.

Compared with Uber-Pool, in [25–28], proposed research effort does not serve multiuser demands. Specifically, Uber batch matching is not provided by the adopted system. However, it supports single users commute services like Uber while it exploits personalized riders' information. The next step of the proposed effort is to incorporate social context as well as social mobility context to provide a more social car ride sharing environment in SCs commute. In addition, the aim is to provide a mobile application where riders can collect desired commutes with drivers that they have common interests with. Such a mobile application will not require pricing options, like Uber does, but it will be a free of charge service for every citizen who wishes to commute in SCs environments.

5. Conclusions

The current research effort focuses on smart vehicle commuting systems. It proposed a predictive vehicle ride sharing system, which has impact to SC green and sustainable environment. The adopted system also provides a recommendation capability to citizens to select the persons they would like to commute with. It analyzes each user's mobility behavior as part of a sustainable commuting transport in SCs. The proposed system focuses on providing personalized services by incorporating user mobility context as part of the online scheduling and dynamic routing processes. A free of charge policy is adopted in contrast with Uber pricing, which enables riders to use the research effort in their daily commute in SCs without paying a trip fee. Subsequently, the research effort outperforms other approaches with regards to a and $map@N$ system evaluation performance metrics.

Future research directions are to focus on expanding the user mobility context by incorporating social and social mobility context to provide riders and drivers a trip where they can share common interests thus make every day commuting in the SC a self-satisfaction experience. Concretely, it is aims to research such social satisfaction in cases of joint car riding trips where more than a single passenger will share the same vehicle. In such cases, research should be done towards examining group integrated satisfaction with the incorporation of clustering techniques applied to pervasive green environments.

Author Contributions: Conceptualization, T.A.; methodology, T.A.; software, T.A.; validation, T.A.; formal analysis, T.A.; investigation, T.A.; resources, T.A.; data curation, T.A.; writing—original draft preparation, T.A.; writing—review and editing, T.A.; visualization, T.A.; supervision, T.A.; project administration, T.A. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Data exploited in this paper where provided by OpenStreetMap, available online: https://www.openstreetmap.org/user/PierrosPapadeas/traces/291454 (accessed on 14 January 2021).

Conflicts of Interest: The author declares no conflict of interest.

References

1. Ahmadi-Assalemi, G.; Al-Khateeb, H.; Epiphaniou, G.; Maple, C. Cyber Resilience and Incident Response in Smart Cities: A Systematic Literature Review. *Smart Cities* **2020**, *3*, 894–927. [CrossRef]
2. Nesmachnow, S.; Hernandez-Callejo, L. CITIES: Ibero-American Research Network for Sustainable, Efficient, and Integrated Smart Cities. *Smart Cities* **2020**, *3*, 758–766. [CrossRef]
3. Serrano, W. Digital Systems in Smart City and Infrastructure: Digital as a Service. *Smart Cities* **2018**, *1*, 134–154. [CrossRef]
4. Belli, L.; Cilfone, A.; Davoli, L.; Ferrari, G.; Adorni, P.; Nocera, D.F.; Olio, A.D.; Pellegrini, C.; Mordacci, M.; Bertolotti, E. IoT-Enabled Smart Sustainable Cities: Challenges and Approaches. *Smart Cities* **2020**, *3*, 1039–1071. [CrossRef]
5. Radu, L.D. Disruptive Technologies in Smart Cities: A Survey on Current Trends and Challenges. *Smart Cities* **2020**, *3*, 1022–1038. [CrossRef]
6. Tzafestas, S.G. Ethics and Law in the Internet of Things World. *Smart Cities* **2018**, *1*, 98–120. [CrossRef]
7. Lucic, M.C.; Wan, X.; Ghazzai, H.; Massoud, Y. Leveraging Intelligent Transportation Systems and Smart Vehicles Using Crowdsourcing: An Overview. *Smart Cities* **2020**, *3*, 341–361. [CrossRef]
8. El-Garoui, L.; Pierre, S.; Chamberland, S. A New SDN-Based Routing Protocol for Improving Delay in Smart City Environments. *Smart Cities* **2020**, *3*, 1004–1021. [CrossRef]
9. Costa, D.G.; Damasceno, A.; Silva, I. CitySpeed: A Crowdsensing-Based Integrated Platform for General-Purpose Monitoring of Vehicular Speeds on Smart Cities. *Smart Cities* **2019**, *2*, 46–65. [CrossRef]
10. Zhao, R.; Stincescu, T.; Ballantyne, E.E.F.; Stone, D.A. Sustainable City: Energy Usage Prediction Method for Electrified Refuse Collection Vehicles. *Smart Cities* **2020**, *3*, 1100–1116. [CrossRef]
11. Xiong, Z.; Zheng, J.; Song, D.; Zhong, S.; Huang, Q. Passenger Flow Prediction of Urban Rail Transit Based on Deep Learning Methods. *Smart Cities* **2019**, *2*, 371–387. [CrossRef]
12. Das, V. Does Adoption of Ridehailing Results in More Frequent Sustainable Mobility Choices? An Investigation Based on National Household Travel Survey (NHTS) 2017 Data. *Smart Cities* **2020**, *3*, 385–400. [CrossRef]
13. Hipogrosso, S.; Nesmachnow, S. Analysis of Sustainable Public Transportation and Mobility Recommendations for Montevideo and Parque Rodo Neighborhood. *Smart Cities* **2020**, *3*, 479–510. [CrossRef]
14. Napoles, V.M.P.; Paez, D.G.; Panelas, J.L.E.; Perez, O.G.; Santacruz, M.J.G.; Pablos, F.M.D. Smart Bus Stops as Interconnected Public Spaces for Increasing Social Inclusiveness and Quality of Life of Elder Users. *Smart Cities* **2020**, *3*, 430–443. [CrossRef]
15. Jiang, Y.; Cao, J.; Liu, Y.; Fan, J. West Lake Tourist: A Visual Analysis System Based on Taxi Data. *Smart Cities* **2019**, *2*, 345–358. [CrossRef]
16. Monteiro, D.L.V.; Perego, R.; Rinzivillo, S.; Times, V.C. Boosting Ride Sharing with Alternative Destinations. *IEEE Trans. Intell. Transp. Syst.* **2018**, *19*, 2290–2300. [CrossRef]
17. Tirachini, A.; Chaniotakis, E.; Abouelela, M.; Antoniou, C. The sustainability of shared mobility: Can a platform for shared rides reduce motorized traffic in cities? *Transp. Res. Part. C* **2020**, *117*, 1–15. [CrossRef]
18. Enzi, M.; Parragh, N.S.; Pisinger, D.; Prandtstetter, M. Modeling and solving the multimodal car- and ride-sharing problem. *Eur. J. Oper. Res.* **2020**, 1–14. [CrossRef]
19. Correa, O.; Mustafizur Rahman Khan, A.K.M.; Tanin, E.; Kulik, L.; Ramamohanarao, K. Congestion-Aware Ride-Sharing. *ACM Trans. Spat. Algorithms Syst.* **2019**, *5*, 1–33.
20. Manik, D.; Molkenthin, N. Topology dependence of on-demand ride-sharing. *Appl. Netw. Sci.* **2020**, *5*, 1–16.
21. Xianan, H.; Li, B.; Peng, H.; Auld, J.A.; Sokolov, V.O. Eco-Mobility-on-Demand Fleet Control with Ride-Sharing. *IEEE Trans. Intell. Transp. Syst.* **2020**, 1–11. [CrossRef]
22. Ruch, C.; Lu, C.Q.; Sieber, L.; Frazzoli, E. Quantifying the Efficiency of Ride Sharing. *IEEE Trans. Intell. Transp. Syst.* **2020**, 1–6. [CrossRef]
23. Goel, P.; Kulik, L.; Ramamohanarao, K. Privacy-Aware Dynamic Ride Sharing. *ACM Trans. Spat. Algorithms Syst.* **2016**, *2*, 1–41. [CrossRef]
24. Ozkan, E. Joint pricing and matching in ride-sharing systems. *Eur. J. Operat. Res.* **2020**, *287*, 1149–1160. [CrossRef]
25. UberMarketplace. How Does Uber Match Riders with Drivers? Available online: https://marketplace.uber.com/matching (accessed on 14 January 2021).
26. Schaller, B. Can Sharing a Ride Make for Less Traffic? Evidence from Uber and Lyft and Implications for Cities. *Transp. Policy* **2020**, 1–25. [CrossRef]
27. Daganzo, C.F.; Ouyang, Y.; Yang, H. Analysis of ride-sharing with service time and detour guarantees. *Transp. Res. Part. B* **2020**, *140*, 130–150. [CrossRef]
28. Qian, X.; Kumar, D.; Zhang, W.; Ukkusuri, S.V. Understanding the Operational Dynamics of Mobility Service Providers: A Case of Uber. *ACM Trans. Spat. Algorithms Syst.* **2020**, *6*, 1–20. [CrossRef]

29. Cao, B.; Hou, C.; Zhao, L.; Alarabi, L.; Fan, J.; Mokbel, M.F.; Basalamah, A. SHAREK*: A Scalable Matching Method for Dynamic Ride Sharing. *Geoinformatica* **2020**, *24*, 881–913. [CrossRef]

30. Luo, M.; Du, B.; Klemmer, K.; Zhu, H.; Ferhatosmanoglu, H.; Wen, H. D3P: Data-driven Demand Prediction for Fast Expanding Electric Vehicle Sharing Systems. *ACM Interact. Mob. Wearable Ubiquitous Technol.* **2020**. [CrossRef]

31. Yatnalkar, G.; Narman, H.S.; Malik, H. An Enhanced Ride Sharing Model Based on Human Characteristics and Machine Learning Recommender System. In Proceedings of the 3rd International Conference on Emerging Data and Industry 4.0 (EDI40), Warsaw, Poland, 6–9 April 2020.

32. Escalona, J.A.; Manalo, B.; Limjoco, W.J.R.; Dizon, C.C. A Ride Sharing System based on An Expansive Search-Based Algorithm. In Proceedings of the IEEE Region 10 Conference (TENCON), Osaka, Japan, 16–19 November 2020.

33. Yousaf, G.; Li, J.; Chen, L.; Tang, I.; Dai, X. Generalized multipath planning model for ride-sharing systems. *Front. Comput. Sci.* **2014**, *8*, 100–118. [CrossRef]

34. Zuo, H.; Zhao, Y.; Shen, B.; Zheng, W.; Huang, Y. High-capacity ride-sharing via shortest path clustering on large road networks. *J. Supercomput.* **2020**, 1–26. [CrossRef]

35. OpenStreetMap. Available online: https://www.openstreetmap.org/user/PierrosPapadeas/traces/291454 (accessed on 14 January 2021).

36. Global Positioning System (GPS). Available online: https://www.gps.gov/systems/gps/performance/accuracy/ (accessed on 14 January 2021).