Journal of Computers Vol. 31 No. 3, 2020, pp. 206-215
doi:10.3966/199115992020063103016

Ride to Work Together: Commuter Ridesharing Meeting Points Design

Guangyao Li*

Software Engineering, College of Software, Taiyuan University of Technology, Taiyuan, China
davey_lee6533@qq.com

Received 28 May 2019; Revised 2 August 2019; Accepted 6 January 2020

Abstract. Commuter ridesharing is an effective way of reducing traffic congestion during peak hours, building friendship with colleagues and neighbors and saving commuting costs. However, to ride together efficiently, the riders need to find a meeting point. In this paper, the commuter ridesharing meeting points design problem is defined, from both set partitioning and centroid-based clustering perspectives, and an adapted k-medians clustering algorithm is proposed as solution. Validation of the effectiveness of the algorithm with a real-world dataset and study of the effects of different parameters including the cluster number and the vehicle capacity constraint are carried out. Simulation results show that the algorithm could save up to 67% mileage with a vehicle capacity of 10, 60% mileage with a vehicle capacity of 5 and 30% mileage with a vehicle capacity of 2. Simulation carried out on uniform platform based on real-world travel data is used as the research method, conclusion is derived according to the interpretation of simulation results. Contributions in this paper are mani-fold. Firstly, a systematic and thorough study of the commuter ridesharing meeting points design problem and the process of formulating it as an optimization problem and the prove of the equivalence between two formats, the set partitioning and the centroid-based clustering are elaborated. Then, an adapted k-medians clustering algorithm that can efficiently solve the ridesharing meeting points design problem theoretically is proposed. Finally, the utilization of a real-world travel dataset to further confirm the effectiveness and running time of the adapted k-medians clustering algorithm is introduced.

Keywords: commuter ridesharing, k-medians clustering, transportation

1 Introduction

With the expansion of the big cities and the growth of population, a lot of challenges are posed in the city life, among which the traffic congestion is one of the most severe problems. With more people driving their own vehicles, the road soon becomes too crowded and everybody becomes stuck in congested traffic, especially during morning and evening peak hours on weekdays.

While public transportation including buses and subways has a larger capacity, the fixed route and stops could be troublesome for riders. Also taking the public transportation usually means a longer travel time, especially when there are too many stations. In addition, given the gradually serious traffic congestion of the present days, which can be manifested everywhere in China, especially in major cities like Beijing and Shanghai. The overall growing tendency of total population and amount of automobiles is obvious, so public transportation pressure is bound to be strengthened. Therefore, carrying out ride sharing is of vital significance.

Commuter ridesharing [1] has been proposed as an effective solution, which encourages people to share a vehicle when commuting from home to work and from work to home. With less vehicles on the road, the benefits are not only economical, e.g., saving people's traveling time, but also environmental, e.g., alleviating air pollution. By sharing the same vehicle with only a small group of individuals, the route can be negotiated and designed efficiently.

---

* Corresponding Author

While commuter ridesharing seems promising, to incentivize sharing among strangers, both privacy and efficiency need to be taken into consideration. Using meeting points, strangers can find a place to gather first, and then take the same vehicle to work, while keep the privacy of their home addresses. However, the proper choice of the meeting points would highly affect the efficiency of ridesharing as well as the user experience, thus requiring a systematic study.

In this paper, the commuter ridesharing meeting points design problem is been formalized from two perspectives, set partitioning and centroid-based clustering. Then an adapted k-medians clustering algorithm that can solve this problem efficiently is proposed. Lastly, validation concerning the effectiveness and running time of the algorithm with a real-world travel dataset provided by Didi Chuxing [2] is completed.

Main contributions in this paper is three-fold:

(1) Formulating a systematic study of the commuter ridesharing meeting points design problem as an optimization problem, proving the equivalence between two formats, setting partitioning and centroid-based clustering.

(2) An adapted k-medians clustering algorithm that can solve the ridesharing meeting points design problem theoretically is proposed.

(3) Utilization of a real-world travel dataset to further confirm the effectiveness and running time of the adapted k-medians clustering algorithm.

The rest of the paper is organized as follows: in Section 2, a short review of the related work is given; in Section 3, formulation of the commuter ridesharing meeting points design problem in mathematical formats; in Section 4, explanation of simulation settings and results; conclusion is drawn in Section 5.

2 Related Work

In this section, a short review of related work is provided.

2.1 Ridesharing

Ridesharing or carpooling [3-4] has drawn a lot of attention from the academia in the past few decades, even before the appearance of the Internet and smartphones, as it has been proved to bring traveling convenience, both economic and environmental benefits [3], and a complement for public transit during holidays [5-6].

Commuter ridesharing [7-12] is also discussed in the literature. In [1], the authors illustrate a ride matching method for commuting trips based on a trajectory clustering. In [7], the authors evaluate the benefits of meeting points, which allow more matches and improve mileage savings. In [8], the authors focus on developing a privacy-preserving service to compute meeting points in ridesharing, such that each user remains in control of his location data. In [9], the authors use an advanced k-means method for ride matching, with a case study of Gazela Bridge in Belgrade, Serbia.

For now, who to share and where to meet remain as two major concerns for commuter ridesharing. In this paper, these two concerns will be solved simultaneously with proposed adapted k-medians clustering algorithms, when the clusters could be seen as the ride share groups and the clustering centroids could be seen as the meeting points.

2.2 Human Mobility Pattern Analysis

The wide availability of trajectory data and the methods that transform trajectories into other data formats are introduced in [14], such as graphs, matrices, and tensors, to which more data mining and machine learning techniques can be applied. It provided an overview on how to deal with trajectories.

To describe and study human mobility patterns, GPS trajectories collected from several time periods has been utilized in enormous studies. (1) Analyzing taxi drivers' behaviors. In study [15-16], researchers extract both passenger-delivery and passenger-searching trip information from GPS trajectories and evaluate taxi drivers' working conditions, and find that drivers with top efficiencies tend to search locally, drive faster for passenger-delivery. These differences may help the taxi drivers to make improvements [15-16]. (2) Studying the effects of Didi Chuxing on Beijing's taxi industry by conducting a multiperiod analysis of the taxi drivers' behaviors. [17] (3) Evaluating the double-apping phenomenon in the e-hailing service. Researchers of this literature gave the first systematic study of the phenomenon of double-apping and clarify its definition in a smartphone-based e-hailing market model and derived four cruising modes by conducting extensive simulations based on real taxi GPS trajectories [18].

In this paper, the GPS dataset provided by Didi Chuxing is utilized and validation of the effectiveness of proposed algorithm is carried out.

2.3 Customized-bus Sharing

Customized bus [19-24] is another solution for commuting. Several previous researchers have carried out studied concerning CB and provided a systematic analysis of the current state of Customized bus service. Customized bus is one of a innovative public-transport modes. It serves as a good alternative and effectively improve traffic safety and alleviate energy consumption problems.

Since there are diversified passenger-travel demand, various CB systems have been developed, such as: Customized commuter bus, Customized school bus, Customized business bus, and Customized community bus. CB is a totally demand-driven and user-oriented transportation system. [19]. Also, researchers in [19] summed up several advantages of CB including its flexibility, personalization, reliability, rapidness, smoothness, etc.

After comprehensive in-depth analysis of CB conducted by former researchers, some important achievements have been made: (1) formulation of a holistic traveler mobility optimization approach to determine bus stops, detailed bus route and schedule; (2) how to integrate and solve capacitated trip-to-bus assignment and bus timetabling problems for large-scale networks [20].

To sum up, different from the traditional buses with fixed route and stations, customized bus serves a small group of people with the same origins and destinations or live nearby. Different from consideration for commuter ridesharing, customized bus is more suitable for the case when no rider owns a vehicle and the bus is equipped with a driver. For commuter ridesharing, some rider may act as the driver.

3 Problem Formulation

3.1 Problem Definition

A simplified setting of sharing rides when commuting to work in the morning peak hours and commuting to home in the evening peak hours is considered, which is referred as commuter ridesharing in this paper. The users are divided into different ride sharing groups firstly, with the objective of minimizing the summation of the costs from each share group, i.e., the total traveling distance, subjecting to the vehicle capacity constraint. The users in one ride sharing group agree to meet up in a pickup meeting point. And they share a single vehicle with a capacity constraint to a certain drop-off meeting point. These users travel individually from home to the pickup meeting point as well as from the drop-off meeting point to work. In practice, one user may drive his/her own car and the other users may use other forms of transport, e.g., station-based or free-floating shared bikes, which is suitable for a short-distance travel. Choosing the pickup & drop-off meeting locations is needed, with the objective of minimizing the traveling distance for a particular ride sharing group. The illustration of ridesharing with meeting points is shown in Fig. 1.

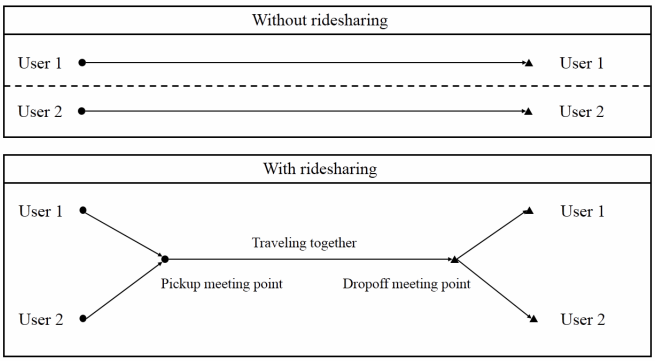


Fig. 1. The situations without and with ridesharing

For simplicity, the pickup and drop-off duration as well as users' time cost by waiting for other fellow travellers for arriving early will not be taken into consideration. Also, the problem of fare splitting will be left for future consideration.

3.2 Mathematical Formulation

3.2.1 Set Partitioning Perspective

Assuming there are N users $\mathcal{U}$. For a user $u \in \mathcal{U}$, the origin is denoted as O_u and the destination as D_u , where $O_u = (x_{O_u}, y_{O_u}), D_u = (x_{D_u}, y_{D_u}) \in \mathcal{R}^2$. The origin and destination correspond to the home and work locations in the morning peak hours, and vice versa in the evening peak hours. The Manhattan distance [25] is used to measure the user's traveling distance $d(O_u, D_u) = |x_{O_u} - x_{D_u}| + |y_{O_u} - y_{D_u}|$, which is suitable and reasonable for the city scenario considered in this study.

A ride share group $S \in \mathcal{S}$ is defined as a subset of users, where $|S| \le c$, c is the vehicle capacity constraint, $\mathcal{S}$ is the set of all subsets. For a ride share group S , defining the pickup & drop-off meeting points $\mu_S = (O_{\mu_S}, D_{\mu_S})$, with the objective of minimizing the traveling distance of the ride share group, which can be represented as

$$\mu_S = \arg\min_\mu d(S,\mu) = \arg\min_\mu \sum_{u \in S} [d(O_u, O_\mu) + d(D_\mu, D_u)] + d(O_\mu, D_\mu) \quad \textbf{(3.1)}$$

where $d(O_u, O_\mu)$ represents the distance of traveling from the origin to the pickup meeting point, $d(D_\mu, D_u)$ represents the distance of traveling from the drop-off meeting point to the destination, and $d(O_\mu, D_\mu)$ represents the distance of the shared trip. Notice that the meeting point μ may not be necessarily the origin & destination of a user.

Now the meeting points design problem is ready to be formulated as a set partitioning problem with the vehicle capacity constraint. Denoting the solution as a vector $v = (v_S)_{S \in \mathcal{S}}$, where $v_S = 1$ indicates that S is a selected ride share group in the solution. Then the set partitioning problem is formulated as follows:

$$\begin{aligned} &\min_v \sum_{S \in \mathcal{S}} v_S d(S, \mu_S) \\ \text{subject to} \quad &\sum_{S: u \in S} v_S = 1, \forall u \in \mathcal{U} \\ &|S| \le c, \forall S \in \mathcal{S} \\ &v \in \{0,1\}^N \end{aligned} \quad \textbf{(3.2)}$$

If a k-set constraint is included, k-set partitioning problem is as follows, where the number of selected subset is exactly k :

$$\begin{aligned} &\min_v \sum_{S \in \mathcal{S}} v_S d(S, \mu_S) \\ \text{subject to} \quad &\sum_{S: u \in S} v_S = 1, \forall u \in \mathcal{U} \\ &|S| \le c, \forall S \in \mathcal{S} \\ &v \in \{0,1\}^N \\ &\sum_{S \in \mathcal{S}} v_S = k \end{aligned} \quad \textbf{(3.3)}$$

3.2.2 Centroid-based Clustering Perspective

By adding the k-set constraint, the problem can be converted into a centroid-based clustering problem with the cluster size constraint as follows:

$$\begin{aligned} & \min_{(T_1, T_2, \ldots, T_k)} \sum_{i=1}^{k} d(T_i, \mu_{T_i}) \\ \text{subject to} \quad & \cup_{i \in \{1,2,\ldots,k\}} T_i = \mathcal{U} \\ & T_i \cap T_j = \emptyset, \forall i \neq j \\ & |T_i| \leq c, \forall i \in \{1,2,\ldots,k\} \end{aligned} \tag{3.4}$$

where

$$\mu_{T_i} = \arg\min_{\mu} d(T_i, \mu) = \arg\min_{\mu} \sum_{u \in T_i} [d(O_u, O_\mu) + d(D_\mu, D_u)] + d(O_\mu, D_\mu) \tag{3.5}$$

when each cluster is regarded as a feasible ride sharing group.

3.2.3 Adapted k-medians Clustering Algorithm

To solve the meeting points Design problem efficiently, the standard k-medians clustering algorithm [26] with a new update rule of centroids is adopted. The algorithm is stated in Algorithm 1, which is still a variant of the generalized expectation-maximization algorithm as the standard version.

Table 1. Adapted k-medians clustering algorithm

Algorithm 1. The adapted k-medians clustering algorithm

1. Initialization step: select k users as initial centroids $\mu_1^{(1)}, \ldots, \mu_k^{(1)}$.
2. **repeat**
3. Assignment step: assign each user to the cluster with the closest centroid:

$$T_i^{(t)} = \{u : d(O_u, O_{\mu_i^{(t)}}) + d(D_{\mu_i^{(t)}}, D_u) \leq d(O_u, O_{\mu_j^{(t)}}) + d(D_{\mu_j^{(t)}}, D_u), \forall j, 1 \leq j \leq k\} \tag{3.6}$$

. where a user is assigned to exactly one cluster $T_i^{(t)}$ randomly, if it could be assigned

. to two or more of them.

4. Update step: calculate the new centroids to be the "medians" of the new clusters.

$$O_{\mu_i^{(t+1)}} = \text{median}(\underbrace{O_{u_1}, \ldots, O_{u_{|T_i^{(t)}|}}}_{\text{users in cluster } T_i^{(t)}}, \underbrace{D_{\mu_i^{(t)}}, \ldots, D_{\mu_i^{(t)}}}_{\left\lceil \frac{|T_i^{(t)}|}{c} \right\rceil \text{ duplicates}}) \tag{3.7}$$

$$D_{\mu_i^{(t+1)}} = \text{median}(\underbrace{D_{u_1}, \ldots, D_{u_{|T_i^{(t)}|}}}_{\text{users in cluster } T_i^{(t)}}, \underbrace{O_{\mu_i^{(t)}}, \ldots, O_{\mu_i^{(t)}}}_{\left\lceil \frac{|T_i^{(t)}|}{c} \right\rceil \text{ duplicates}}) \tag{3.8}$$

5. **until** Centroids do not change.

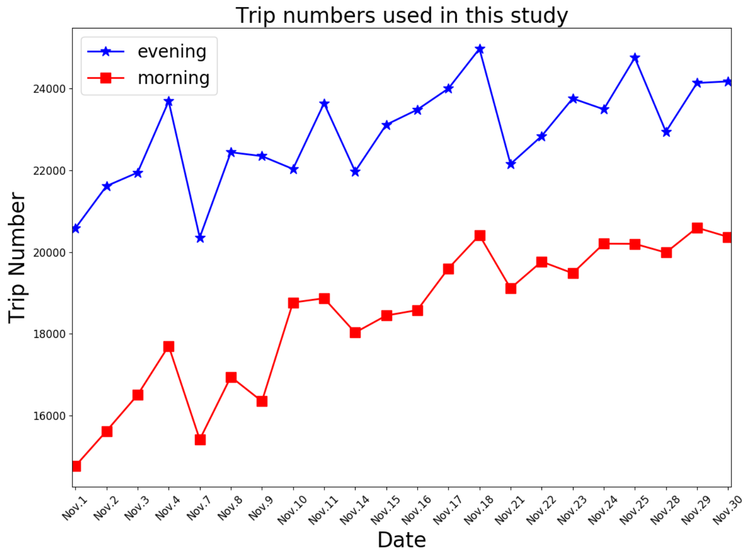
4 Simulation

4.1 Dataset Description

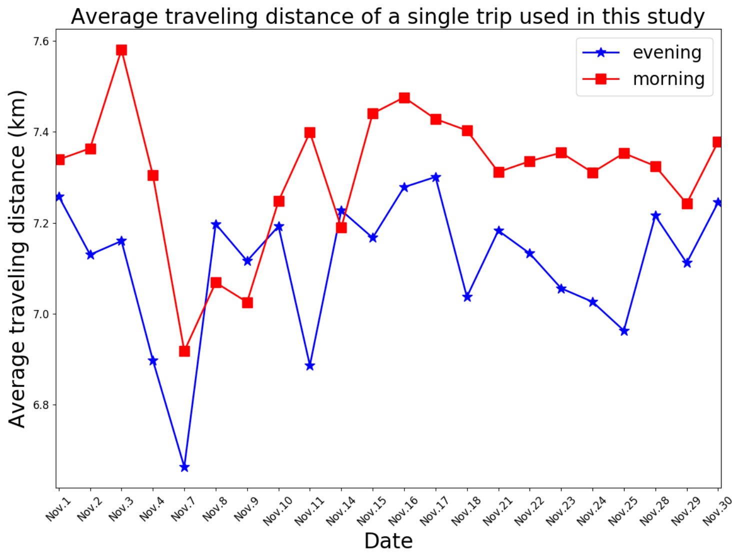
In this study, an Internet based ride-hailing trip dataset provided by Didi Chuxing is utilized, the largest transportation network company in China. The trip dataset covers the order and GPS trace information of Didi's ride-hailing service in Chengdu, the capital of China's Sichuan province, from November 1 to November 30, 2016, which contains 22 weekdays and 8 weekends. Each trip record contains trip id, begin time, end time, origin longitude & latitude, and destination longitude & latitude. More details can be found on Didi Chuxing's website for the dataset.

As the goal is to study the commuter ridesharing phenomenon, only the trips on weekdays and within the peak hours (morning: 7am to 9am, evening: 6pm to 8 pm) are used. The spatial range is bounded to the central part of Chengdu city (longitude: 103.9 to 104.21, latitude: 30.53 to 30.8). those orders which are not within the spatial range and the abnormal orders with a duration less than 1 minute are filtered out.

The number of trips and the average traveling distance of a single trip are summarized in Fig. 2.



(a) The number of trips in November, 2016



(b) The average traveling distance in November, 2016

Fig. 2.

4.2 Parameter Setting

To evaluate the influence of different parameters, I choose the number of clusters k as 500, 1000, 2000, 4000, and 8000, and the vehicle capacity constraint c as 2, 5 and 10. The constraint of 2 corresponds to the case when two passengers share a vehicle, when the coordination cost between passengers is the least. The constraint of 5 corresponds to the case when a small car's capacity is fully utilized. The constraint of 10 simulates the case of a minibus. These parameters are considered for realistic travelling requirements and pre-literature Settings.

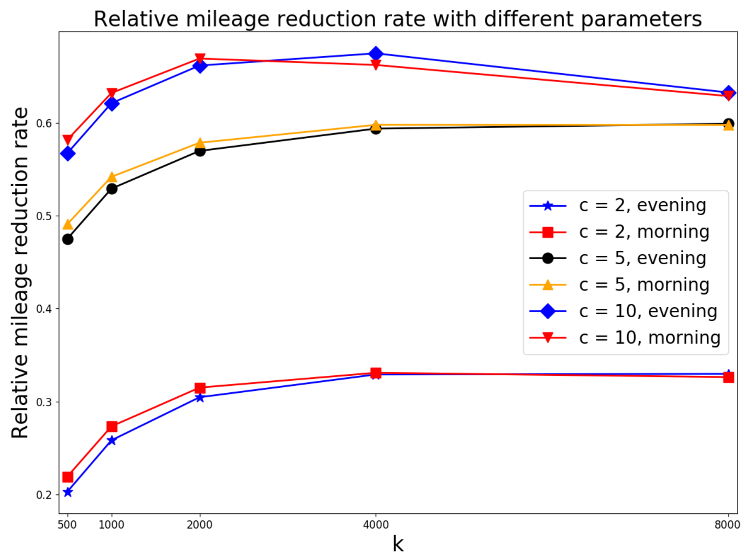
4.3 Evaluation Metric

In this study, the relative mileage reduction rate is used as evaluation metric, which evaluates the benefits of ridesharing, as compared to the original case without ridesharing. Let $M0$ denote the total mileage before ridesharing and $M1$ denote the total mileage after ridesharing, the relative mileage reduction rate is defined as $(M0 - M1) / M0$.

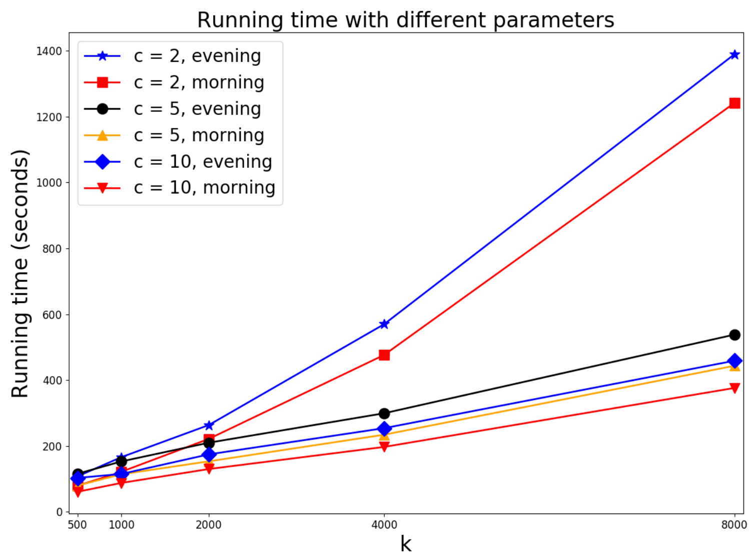
Since there is no guarantee that the optimum is found immediately by the adapted k-medians clustering algorithm, the clustering algorithm for each day by 5 times and then take the maximum relative mileage reduction rate is taken. The final result is averaged among 22 weekdays used in this study. I also evaluated the running time of the algorithm with different parameters, and the result is the average of the 5 runs and the 22 weekdays.

4.4 Results

The results are shown in Fig. 3. The simulations are conducted on a Windows OS machine, with a CPU operating frequency of 2.4 GHz.



(a) The relative mileage reduction rate



(b) The running time

Fig. 3.

As it can be seen from Fig. 3, a larger vehicle capacity constraint c can achieve a higher relative mileage reduction rate with a smaller running time, which are both desirable properties. However, the challenge for using a larger vehicle capacity in practice is the coordination cost between passengers, when everyone agrees to meet in the same location.

With a larger clustering number k, a larger relative mileage reduction rate would be expected, with the cost of a larger running time. However, as k reaches 4000 in our case, there is no need to increase k, as the relative mileage reduction rate remains stable, while the running time keeps increasing dramatically.

Since there are less trips in the morning peak hours, the running time of morning trips are smaller. However, more evening trips make it even harder to save more mileage, as shown in Fig. 3(a). The possible reason could be that with more diverse trips, it is harder to find a ride share group that has close origins and destinations for them all.

5 Conclusion

In this study, the commuter ridesharing meeting points design problem is defined and an adapted k-medians clustering algorithm to solve this problem is proposed. Validation regarding the effectiveness of the algorithm with a real-world dataset and study of the effects of different parameters are carried out. Results show that the algorithm could save up to 67 mileage with a vehicle capacity of 10, 60% mileage with a vehicle capacity of 5 and 30% mileage with a vehicle capacity of 2.

Some exciting achievements and multiple contributions are successfully made through this study. A systematic study, formulating as well as the proving process with respect to the commuter ridesharing problems are elaborated. An adapted k-medians clustering algorithm that can efficiently solve the ridesharing meeting points design problem theoretically is proposed. In addition, the utilization of a real-world travel dataset to further confirm the effectiveness and running time of the adapted k-medians clustering algorithm is introduced.

While the algorithm shows relatively promising results, this study is based on many preconceived simplifications, without considering the variety and complexity of time constraints and pricing schemes which are usually commonplace in the real world.

To put work into practice, a system to implement the algorithm need to be built, which requires more resource and effort. For future work, other than the system implementation, I would also improve the model and algorithm in a theoretical aspect, by taking into consideration more constraints. In addition, explanations of the detailed theoretical analysis of the feasibility, effectiveness, etc. regarding to specific methods towards system deployment and implementation would also be necessary. Further more, it would be more suitable to carry out a survey regarding to users' willingness to use the system, through which we get statistics of their acceptance rate of the new system.

Acknowledgements

Data retrieved from Didi Chuxing.

References

[1] Z. Hong, Y. Chen, H.S. Mahmassani, S. Xu, Commuter ride-sharing using topology-based vehicle trajectory clustering: Methodology, application and impact evaluation, Transportation Research Part C: Emerging Technologies 85(2017) 573-590.

[2] Didi Chuxing. <https://www.didiglobal.com/> 2014 (accessed 19.06.14).

[3] M. Furuhata, M. Dessouky, F. Ordóñez, M.-E. Brunet, X. Wang, S. Koenig, Ridesharing: The state-of-the-art and future directions, Transportation Research Part B: Methodological 57(C)(2013) 28-46.

[4] H. Huang, D. Bucher, J. Kissling, R. Weibel, M. Raubal, Multimodal route planning with public transport and carpooling, IEEE Transactions on Intelligent Transportation Systems 20(9)(2019) 3513-3525..

[5] W. Jiang, C. Jiang, P. Zhang, L. Zhang, Poster: CountryRoads: Large-Scale Nationwide Ridesharing System, in: Proc. the 13th ACM Conference on Embedded Networked Sensor Systems, 2015.

[6] W. Jiang, C.R. Dominguez, P. Zhang, M. Shen, L. Zhang, Large-scale nationwide ridesharing system: a case study of Chunyun, International Journal of Transportation Science and Technology 7(1)(2018) 45-59.

[7] M. Stiglic, N. Agatz, M. Savelsbergh, M. Gradisar, The benefits of meeting points in ride-sharing systems, Transportation Research Part B: Methodological 82(C)(2015) 36-53.

[8] U.M. Aïvodji, S. Gambs, M.J. Huguet, M.-O. Killijian, Meeting points in ridesharing: a privacy-preserving approach, Transportation Research Part C: Emerging Technologies 72(2016) 239-253.

[9] M. Šelmić, D. Macura, D. Teodorović, Ride matching using K-means method: case study of Gazela Bridge in Belgrade, Serbia, Journal of Transportation Engineering 138(1)(2011) 132-140.

[10] P. Czioska, R. Kutadinata, A. Trifunović, et al, Real-world Meeting Points for Shared Demand-Responsive Transportation Systems, Public Transport, (2017). DOI:10.1007/s12469-019-00207-y.

[11] P. Czioska, M. Sester, Determination of efficient meeting points in ride sharing scenarios, in: Proc. 18th AGILE Conference on Geographic Information Science, 2015.

[12] X. Li, S. Hu, W. Fan, K. Deng, Modeling an enhanced ridesharing system with meet points and time windows, PloS one 13(5)(2018) e0195927.

[13] P. Czioska, A. Trifunović, S. Dennisen, M. Sester, Location-and time-dependent meeting point recommendations for shared interurban rides, Journal of Location Based Services 11(3-4)(2017) 181-203.

[14] Y. Zheng, Trajectory data mining: an overview, ACM Transactions on Intelligent Systems and Technology 6(3)(2015) 29.

[15] W. Jiang, J. Lian, M. Shen, L. Zhang, A multi-period analysis of taxi drivers' behaviors based on GPS trajectories, in: Proc. 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), 2017.

[16] Y. Dong, S. Wang, L. Li, Z. Zhang, An empirical study on travel patterns of internet based ride-sharing, Transportation Research Part C: Emerging Technologies 86(2018) 1-22. DOI:10.1016/j.trc.2017.10.022.

[17] W. Jiang, L. Zhang, The impact of the transportation network companies on the taxi industry: evidence from Beijing's GPS taxi trajectory data, IEEE Access 6(2018) 12438-12450.

[18] W. Jiang, L. Zhang, Evaluating the effects of double-apping on the smartphone-based e-hailing service: a simulation-based study, IEEE Access 6(2018) 6654-6667.

[19] T. Liu, A.A. Ceder, Analysis of a new public-transport-service concept: customized bus in China, Transport Policy 39(2015) 63-76.

[20] L.C. Tong, L. Zhou, J. Liu, X. Zhou, Customized bus service design for jointly optimizing passenger-to-vehicle assignment and vehicle routing, Transportation Research Part C: Emerging Technologies 85(2017) 451-475.

[21] Y. Lyu, C.Y. Chow, V.C.S. Lee, J.K.Y. Ng, Y. Li, J. Zeng, CB-Planner: A bus line planning framework for customized bus systems, Transportation Research Part C: Emerging Technologies 101(2019) 233-253.

[22] R. Guo, W. Guan, W. Zhang, Route design problem of customized buses: mixed integer programming model and case study, Journal of Transportation Engineering, Part A: Systems 144(11)(2018) 04018069.

[23] T. Perera, A. Prakash, T. Srikanthan, A hybrid methodology for optimal fleet management in an electric vehicle based flexible bus service, in: Proc. 2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV), 2018.

[24] Y. He, J. Ni, B. Niu, F. Li, X. Shen, Privacy-preserving ride clustering for customized-bus sharing: a fog-assisted approach, in: Proc. 2018 16th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), 2018.

[25] P.E. Black, Manhattan distance, Dictionary of Algorithms and Data Structures 18(2006) 2012.

[26] A.K. Jain, R.C. Dubes, Algorithms for Clustering Data, Prentice Hall, Englewood Cliff, NJ, 1988.