Dear Mr. Benjamin Riviere,

The review of your paper, "Urban Ridesharing with Hybrid Distributed Reinforcement Learning," T-ITS-19-09-1038, has been completed. The reviewers’ comments and those of the Associate Editor are copied below. Based on these comments and recommendation of the Associate Editor your paper is not ready for publication in its present form, A properly revised version that takes care of the concerns and drawbacks pointed out by the reviewers and Associate Editor is potentially publishable.

Therefore, I suggest that you revise your paper along the lines described by the reviewers and resubmit the paper. Please include a description on how you took into account the reviewers' comments in preparing your revision. We would then hope to determine a publication decision soon thereafter. Please note that if your revision is not submitted within the next 3 months (90 days), your paper will be treated as a new paper.

You may want to resubmit your paper as a REGULAR PAPER (suggested length: 10 Transactions pages, authors' biographies included). The Associate Editor will determine whether your paper is best suited to be considered as a Regular Paper or shortened to a Short Paper.

Thank you for submitting your manuscript to the Transactions on Intelligent Transportation Systems.

Sincerely,

Azim Eskandarian

Editor-in-Chief

Transactions on Intelligent Transportation Systems

Editor's comments:

Editor

Comments to the Author:

Dear authors. Four reviewers have provided interesting comments for this paper. Three reviewers recommend major revision, one reviewer minor revision. Among the aspects that require improvements: 1) literature review and discussion of the methodological contributions of the paper in view of the literature, 2) inclusion of a fair comparison with state-of-the-art methods, 3) evaluation of the scalability of the method, for instance by including simulation results with different fleet size (large-scale performance), 4) evaluation of objective functions/performance indicators that are of interest from the transportation perspective: travel times, waiting times, delays (extra waiting and travel time due to the dynamic operation of the system), operating costs, etc., 5) various clarifications of the proposed method (see comments of reviewers). Finally, reviewers indicate that the overall readability of the paper should improve, including as well the quality of the figures and overall structure of the paper. After my own assessment I concur with the reviewers and recommend revise and resubmit as regular paper.

Reviewer: 1

Recommendation: Prepare A Major Revision For A New Review

Comments:

Regarding the prior, assignment for serving ridesharing requests is related to the mobility profiling of individuals. Mobility profiling with road network (e.g. POIs) is even richer: it predicts forthcoming demand by origin AND destination. Both are valuable in the optimization process of the automatic ride-matching process. However, no mobility (except OD), road network and traffic have been integrated into the MDP formulation.

Regarding the latter, in my eyes, coordination would be much more interesting for this optimization of the allocation to vehicle capacity, rather than for the sharing of vehicles to reach common or nearby destinations. How to coordinate the schedules between participants, issue a ride-matching process to determine vehicle routes and the assignment of passengers to vehicles are critical ridesharing problems. Usually, we consider the objectives of maximizing the number of serviced passengers, minimizing the operating cost, and minimizing passenger inconvenience. I don't see many correlations to ridesharing here.

My questions are supported by the presented summary of the contributions of the paper on Page 2. Contribution 1 claims “MDP formulation for optimal routing” but fails to say how to pick up new passengers en-route or make a detour. It is unclear to me. The authors could provide a case study.

This work seems to be an extension of the existing TD Q-learning algorithm to the transportation area. The authors may consider identifying the challenging issues in using reinforcement learning to solve the problem when people sharing a trip. Besides, for a fair comparison with others, using the existing reinforcement learning methods is a little bit insufficient. Some of the original studies in ridesharing should be also highlighted in the comparison.

In detail:

The first paragraph of the introduction is irrelevant to the ridesharing problem at hand.

Is a "multi-rider ridesharing"? have already multiple riders compromised on their schedules? Nothing is talking about multiple riders.

What’s the “unknown environment”? this statement needs references.

Why is ||(xti-ytj)-(xtj-ytj)||2<=Rc ?

The application of the proposed algorithm in online learning scenarios:

-What assumptions are made in the Chicago taxis dataset?

-How might the proposed algorithm be applied/tested with real-time data?

-What might be the timeliness of assignment (i.e., horizons) - this would be critical to the performance of any future mobility system (refer to recent papers from Savelsbergh) – on-demand, dial-a-ride, etc.

-What challenges/benefits might exist here?

-Can parts of your algorithm be adapted to other ridesharing forms?

I think the task assignment, is not sufficiently described. Some papers are cited in this area, but these are limited and there is little alignment with the manuscript about how these might be integrated, particularly in regard to trip purpose and utility. I'm confused about how this relates to traditional transport theory? Why aren’t other aspects of utility mentioned?

-Choice factors

-Choice hierarchies/temporal variability

-Concepts of Spatio-temporal flexibility

-Trust

-Other activities-based intentions

If a system is suggesting ridesharing assignment based on dynamic demands, how might alternative actions be suggested? My choice factors/hierarchy may change with a range of outside effects (e.g., weather), and a broader choice set may be desirable for increased utility (refer recommender systems research).

The parameters of algorithms were not discussed in detail.

Additional Questions:

Summary of Evaluation: Good

If the paper is rejected for publication, the authors should: Prepare a major revision and resubmit to The ITS Transactions as a new paper

textbox:

Organization: 2

Clarity: 3

Length: 3

References: 4

Correctness: 5

Significance: 4

Originality: 5

Attachments: 3

If Survey Coverage: 4

Contribution: 4

What are the contributions of this paper?: The paper starts from the concept of task assignment for ridesharing, defined by MDP-based oricess for coordination, and uses a Q-learning to maxmize the profits.

What are some ways in which the paper could be improved? Please supply any additional important references that you feel the author omitted which should be noted in the paper.: see below

Reviewer: 2

Recommendation: Prepare A Major Revision For A New Review

Comments:

(There are no comments. Please check to see if comments were included as a file attachment with this e-mail or as an attachment in your Author Center.)

Additional Questions:

Summary of Evaluation: Good

If the paper is rejected for publication, the authors should: Prepare a major revision and resubmit to The ITS Transactions as a new paper

textbox:

Organization: 2

Clarity: 2

Length: 3

References: 3

Correctness: 3

Significance: 3

Originality: 3

Attachments: 3

If Survey Coverage: 3

Contribution: 3

What are the contributions of this paper?: The paper considers the problem of assigning self-driving taxis to taxi users to maximize the profit. The method in this work combines methods in different fields which includes machine learning, operations research, etc. The problem is formulated as MDP, and solved initially at a central node. Then, the problem is passed as the initial policy to the taxis, and the taxis will use local information to capture the changes in the dynamic environment. The model keeps monitoring the solution quality and will send the data to the central node again to solve the problem if the solution is drifted too much. In this manner, the authors claim that the model is highly scalable and the model is tested using Chicago dataset.

What are some ways in which the paper could be improved? Please supply any additional important references that you feel the author omitted which should be noted in the paper.: Overall the problem and the model in the paper are timely and interesting, but the paper can be improved by discussing some of the following concerns:

1. The paper seems to lack some discussion among the common models solving taxi-customer assignment problem in the experiments. For example, I think it will be beneficial to discuss what are the pros and cons of this model and some recent popular models like the model in [1], and even better if we can have experimental results comparison. In this paper, the results are comparing to the closest algorithm which is naive since its basically to pickup the closest customer. The learning-based model may be better to maximize long-term profit, but it would be better to validate the gap using experiments if authors could compare the performance with some batched assignment models not using ml.

2. The paper is considering a fleet of 100, which is very small. In reality, the current taxi fleet can be thousands or more than ten thousand. I am wondering how will the computation increase as the scale increases. Also, we are divided the map into grids. What will be the tradeoff if we increase the number of grids.

3. The paper's clarity can still be improved.

[1] Alonso-Mora, Javier, et al. "On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment." Proceedings of the National Academy of Sciences 114.3 (2017): 462-467.

Reviewer: 3

Recommendation: Prepare A Major Revision For A New Review

Comments:

Besides the suggestions for improvement. Here are some comments that I wish to help improving the paper.

-- The motivation for using a "Hybrid" algorithm is not well justified? Why not using only RL for both sub-problems you are solving?

-- You are claiming that your algorithm is distributed. However, you still coordinate between vehicles using a game-theoretic approach, which makes it centralized in some sort or at the very least semi-distributed. In practice, drivers work greedy to maximize their profit. How such coordination could be implemented in practice?

--Since vehicles use only local information to update its policy, the algorithm performance degrades as it lacks some external environment information. Any idea?

-- The transition probability is binary and deterministic, which simplifies the problem a lot! These probabilities should be either learned over time or estimated from past data. In this case, prob. would not be binary.

-- Fig2. is very unclear. Remove the color in the background to make it better.

-- Actions are taking by vehicles, right? However, the reward is written for customers.

Shouldn't the reward reflect the revenue or the gain of the driver/vehicle/agent?. However, the reward is a function of the customer's payment, time to arrive ...etc.

--ttc^k should be fixed if the shortest path is chosen to drop off the passenger from the pick-up location to the destination. Why it is part of the reward, then! Further, tts^k also is calculated based on norm2 as in Table 1. Therefore, vehicles will likely choose the nearest customer to serve in a gready manner (so that better reward is obtained based on eqn 1), which makes the matching problem not optimal. So, where is the optimization part for matching customers to vehicles in your algorithm?

-- It is always better to use the present simple, instead of using the past simple like "chose" in page 3 second column...etc.

-- Why C1 is fixed? Are fair and tts/ttc on the same scale so no one will dominate?

-- It is preferred to have all equations with numbers to make it easy for the reviewers to refer to an equation if need be. The equation after equation (1) is not clear enough to me. How do you choose \lambda?

-- how do u get both tts and ttc? not clear.

-- functions in Alg1 are too abstracted. Please expand on everyone. expand the algorithm.

-- The evaluation is limited. Very few cars with very number of requests. One day may not be enough since the demand changes from one day to another, from early morning to late at night. From morning on Sunday to morning on a weekday like Monday, for example. Consider more days to test your algorithm.

-- Clarity of figures is poor. Use markers to distinguish between policies.

-- Your evaluation does not show some important QoE measures like wait time per request, the utilization of vehicles, ...etc.

-- With respect to literature, some recent work on fleet management and vehicle dispatching is not cited. The authors should position their work with the recent work on this domain. How is your algorithm is different than this work? Such work includes

Oda, Takuma, and Carlee Joe-Wong. "Movi: A model-free approach to dynamic fleet management." IEEE INFOCOM 2018-IEEE Conference on Computer Communications. IEEE, 2018.

Alabbasi, A, A Ghosh, and V Aggarwal. "Deeppool: Distributed model-free algorithm for ride-sharing using deep reinforcement learning." arXiv preprint arXiv:1903.03882 (2019).

Ke, Jintao, et al. "Optimizing Online Matching for Ride-Sourcing Services with Multi-Agent Deep Reinforcement Learning." arXiv preprint arXiv:1902.06228 (2019).

-- first sentence in the abstract,-- service ----> serve

-- proofreading is definitely needed.

Additional Questions:

Summary of Evaluation: Fair

If the paper is rejected for publication, the authors should: Prepare a major revision and resubmit to The ITS Transactions as a new paper

textbox:

Organization: 3

Clarity: 3

Length: 4

References: 2

Correctness: 4

Significance: 3

Originality: 3

Attachments: 3

If Survey Coverage:

Contribution: 3

What are the contributions of this paper?: The paper proposes a hybrid distributed RL for fleet management. The solution is decomposed into a cell-based MDP with coordination individual agents/vehicles through a game-theoretic approach Distributed Kalman filter is used to give an optimal learning rate. As part of their evaluation, Chicago taxi dataset is used.

What are some ways in which the paper could be improved? Please supply any additional important references that you feel the author omitted which should be noted in the paper.: -- Evaluation is limited with only a very number of vehicles (100 vehicles) and few ride requests.

-- Reward does not take into consideration the drivers/vehicle preferences. Thus, the proposed algorithm is not very appealing to the vehicles' drivers.

-- Scalability is still an issue here since Q-learning (and not deep Q-learning) is used. With larger number of states, your algorithm may fail to produce results in a timely manner.

-- Figures in all the papers are of low quality. Use markers to distinguish between policies.

-- Improve the presentation is not very well and lacks the coherency.

Reviewer: 4

Recommendation: Accept With Minor Changes

Comments:

Authors propose a learning-based algorithm to assign taxis to customer requests, by maximizing profits in an unknown and dynamic environment. The proposed ideas are multidisciplinary, interesting and have being validated experimentally. Some comments for improving the final version of the manuscript are listed as follows.

1) The paper is difficult to follow because of the use of terms that have not being defined previously, such as hybrid learning architecture, dynamic consensus algorithm, incremental TD, TNC architecture and distributed framework. What are these terms? Also, the literature about frameworks has not been included in the article and there are many types of frameworks (i.e., conceptual frameworks, application frameworks). Thus, it is not known why the term framework is used throughout the article. From the technical point of view, all terms used in the article must be introduced earlier and chosen more carefully.

2) In Section I, authors claim that a hybrid learning architecture has been developed and is proposed. However, no formalism has been used to detail such system architecture. An overview of the proposed system architecture is missing and the test configurations should be detailed as well, in order to better illustrate the centralized solution and the distributed scenario that are compared in the article.

3) The article is not well structured since the related work section is missing. Section I is not well structured and is too long. It does not introduce the work done and represents a mix of a related work section with a high level description of the proposed solution. The main contributions are listed too late in the article and are not indicated clearly because many terms from different areas (machine learning, distributed estimation, prescriptive game theory) are used without previous definition.

4) Revise the text with respect to English writing improvements. Some typos and weird sentences (e.g. " our MDP is a framework is agnostic..") were found.

5) Reinforcement learning algorithms are mentioned in Section C. However, no algorithms are listed in this section. With respect to Algorithms, Section IV contains the Algorithm 1 that has not been explained in the text. For the sake of readability, authors should adopt the use of main constructors, such as: algorithms, procedures, etc Algorithm 1 is only a high level description of the distributed routing algorithm since it does not provide the input/ output data nor provides the implementation details of the functions/procedures used. Authors should provide all algorithms used in the experimental section, including those that correspond to the proposed hybrid method, the centralized approach and to the distributed solutions as well. This diminishes the impact of the work and its usage by others.

6) It is hard to follow the paper, since the authors have not employed a running example or some examples to lead the reader to grasp the intuition of the work or of the several theorems that are proved.

7) Figures should be better explained in the text. Also, figure 1 should show the inputs and outputs of each box for the sake of readability and understanding.

8) With respect to test settings, it is not clear how the value of the constant parameter C1 was identified? Preliminary computations? Is considered that such value is uniform or the same for all customer requests? Why? Explain better.

9) The conclusion section is a mere summary of what has been presented in the paper. What is missing though is a discussion, carefully comparing strengths and weaknesses of the proposed approach in comparison with related work.

Additional Questions:

Summary of Evaluation: Excellent

If the paper is rejected for publication, the authors should:

textbox:

Organization: 2

Clarity: 2

Length: 4

References: 4

Correctness: 4

Significance: 5

Originality: 5

Attachments: 4

If Survey Coverage: 4

Contribution: 5

What are the contributions of this paper?: The article presents a cell-based Markov Decision Process for enabling an optimal routing process and proposes a means of coordinating individual agents with a game theoretic task assignment. Also, by using main concepts from distributed TD Q-learning, Bellman iteration and distributed estimation analysis, an optimal adaptive learning rate is built to tune the trade-off between computational costs and estimation accuracy. Finally, the proposed ideas are validated using both a simulation dataset and a real taxi dataset collected for the city of Chicago.

What are some ways in which the paper could be improved? Please supply any additional important references that you feel the author omitted which should be noted in the paper.: Some suggestions for improvement are listed as follows:

1) The current manuscript contains many terms (or combination of terms) that are not common sense but are used throughout the article.

2) The article is not well structured and section 1 does not provide a clear introduction.

3) A mix of definitions and corresponding examples should be given.

4) The most negative aspect of the article is the lack of detailed algorithms, specially those that were used in the performance evaluation.

5) Both figures and assumptions made should be better explained.

Please note that some reviewers may have attached more comments; please log into <https://mc.manuscriptcentral.com/t-its> to download them.