Review of “Autonomous Bus Fleet Control Using Multiagent Reinforcement Learning”

Original Article published in Hindawi Journal of Advanced Transportation by Dr. Sung-Hung Wang and Dr. S. K. Jason Chang

Carroll, Kenneth J.

2021

With the rise of autonomous vehicles, every industry is investigating their potential applications given that they have the potential to remove human error thus reducing cost and increasing efficiency. While many approaches focus on the vehicles themselves, in a fleet capacity (such as a taxi or bus service) there will be another autonomous agent present in the form of coordination of the group of autonomous vehicles. The training and evaluation of these coordination models presents a unique area of investigation and given that vehicle fleets are typically used as public or enterprise services, highly effective models for their coordination will be highly present in the public if adopted. Being in a highly visible, well-funded space, the investigation of proper autonomous fleet coordination could hold three benefits: monetary, notoriety, and altruistic benefits. Both infrastructure and enterprise utilities, vehicle fleets typically have moderate to strong budgets and are applicable in areas ranging from college campuses to intercity travel, and even multilocation travel for campus size businesses like Google. In any event, technology adopted by these groups will have a monetary reward. With how synonymous Uber, Lyft, Lime, and Bird are in the consumer vehicle space, the ever-present nature of intercity transportation in the form of taxi and city buses, an autonomous approach to these fleets will hold a degree of notoriety. When a technology or service is widely adopted, the name essentially enters a collective unconscious i.e., regardless of your personal use of it, you have heard of it to enough of a degree to recognize the adoption and use of the technology or service. The researcher or business who brings forth the next wide adoption will have a degree of notoriety behind them for doing so. Lastly is the altruistic benefit, which is a primary benefit of removing the human aspect from a widespread industry. Reduction in accidents, routing errors, and wait times could mean a large improvement of the quality of public, commercial, and enterprise travel solutions. One could argue the monetary savings from this reduction could also lead to the need for less funding, which could in part route more funding to other community services in the usage for public transit, or towards employee benefits in the commercial and enterprise spaces. Out of these possible motivations, the altruistic one is the one directly stated in the paper but the first and second motivations mentioned here are also inevitabilities for any successful research in the current technology driven climate.

The contribution of this research is to establish the usage of reinforcement learning in the training of agents to provide fleet coordination in a decentralized fashion to solve the last mile problem in transportation. The primary argument is that most machine learning approaches to the fleet management problem focus on a single entity, a central determining agent, which makes the action decisions necessary to adapt the fleet to varying conditions. In this approach, Dr. Wang and Dr. Chang argue that the nature of autonomous vehicles is decentralized, where the real time nature of their action predictions does not apply well to a centralized control model. Instead, they focus on the development of multiagent models through reinforcement learning that allows for the training and integration of multiple vehicles into a fleet to accommodate multiple area scenarios in volume, stops, and interruptions without the usage of a centralized control. Alongside a multi agent reinforcement learning model, there is also a temporal difference learning algorithm in place, given that bus scheduling, travel time, and passenger load/unload rate are time driven elements, and the waiting time of passengers is the metric under score for their model. The focus of their paper is generating a realistic simulation environment where the agents are trained, based off a set loop of bus stops with a scalable number of autonomous vehicles, and working to minimize the wait time of passengers to be loading into a vehicle. They compare their utilization of multiagent deep deterministic policy gradient (MADDPG) vs multi agent discrete Q network (MADQN) as a means of judging scalability in respect to passenger wait time. They do draw comparisons to vehicle distance traveled, which can also be used as a metric for designing a similar system but acknowledge that it is not the main point of optimization for their model which leads them to have a tradeoff of higher distance traveled for vehicle in exchange for reducing passengers’ time. This distance metric is often used to associate the cost efficiency of running the vehicle, so comparisons do demonstrate that focus on passenger wait time does directly increase the cost of operation.

The methodology used in this research is simulating multiple environments and multiple vehicle demand conditions to train the multiagent model using SimPy for the environment with processes living in an OpenAI Gym structure where the only “intelligent” agent is the autonomous vehicle. The environment consisted of a fixed loop route for the bus stops and a number *m* buses which would operate in the environment under the conditions that each bus can occupy the same space without collision and under the ability to make a U-Turn between stops to adjust for demand. The number of stops could be varied on the environment but were not dynamic in the sense that new stops would not be added mid simulation or after establishment of the given environment under test. Distance between stops is defined on environment initialization but can vary from 200 to 450 meters. Passengers are simulated with the option to enter a vehicle at a given stop or be able to disembark at a given stop where load time can be varied from 0.6-6.0 seconds with a mean of 1.75 seconds per passenger. The measurement of time for doors opening and closing along with the time for passenger flow is defined as the dwell time. The passenger capacity of each vehicle is 12 seats and 6 standing slots. The speed of the vehicles is defined at a maximum of 20 km/h with a stop deceleration or an acceleration to get up to speed being 12 seconds. Using these parameters in each environment, the agents (vehicles) define a state space that holds the passenger, vehicle, and environmental factors under action determination; using this state space the agent chooses one of ten actions defined in the action space ranging from moving nonstop, idling with the intent to idle, and transitory actions between those states and the expected next state (i.e., idling intending to move in direction of some station). With reinforcement learning there is the need for objective rewards based off a metric. The metric used here is passenger wait time but is discounted based on Bellman Error.

With the parameters for the methodology defined, there are two scenarios under training: a single agent in a five stop/station environment with a 15-minute headway where travel patterns can consist of two passenger request rates of 90 and 180 passenger requests per hour and a 5 agent 15 stop/station environment with 5 minutes of headway and passenger request rates of 810 and 1620 passengers per hour. The first scenario was used to validate that both single agents using MADDGP and DQN can be trained in this scenario, with a simulation of 1 hour and 300 instances of simulation for a total of 300 simulated hours. Once those results were validated, the five-vehicle fleet scenario was run with MADDGP and MADQN for the models for a simulation of 1 hours and 300 instances of simulation. The given performance of the two agents led to the selection of MADDGP being used for the following scenarios under comparison with traditional scheduled loops that do not respond dynamically to request load: a single vehicle simulation for a university campus with free flow, a dual vehicle simulation for an industrial zone with stable unconstrained flow, and a five-vehicle fleet simulation a downtown street with stable flow under interference. The resultant passenger wait times and the vehicle’s average speeds for both systems was charted in each scenario against the passenger service volume per hour for the final comparison and conclusion of their research.

In the comparison of the MADDGP fleet system versus a standard scheduled bus fleet in each of the previously defined environments under test, the MADDPG fleet outperforms the standard schedule fleet in scenario 1 (university) and scenario 2 (industrial). On the metric of passenger wait time, the MADDGP fleet was consistently lower in weight times by nearly 50s in both scenarios but only achieved higher vehicle speeds in scenario 2. In the 3rd scenario (downtown), MADDGP starts with significantly lower passenger wait times but as volume approaches 1620 passenger requests per hour, the wait time spikes and underperforms the scheduled fleet through 2200 passenger requests per hour before again yielding slightly better wait times than a scheduled fleet for the remainder of request volumes. The number of vehicles required to handle larger volumes was also decreased significantly with one example referenced being the passenger volume of two buses being 405 passengers per hour using MADDGP agents vs 180 in the scheduled bus fleet. This argues that while passenger wait times are decreasing, the effective volume the system can handle without scaling to additional vehicles is also increasing. The data under simulation implies that a MADDGP approach to fleet control could be a valid option under coordination but poses the question on efficiency as the dynamic nature of the buss behavior led to significant increases in vehicle distance traveled for the same passenger volume. Depending on how the system is designed, vehicle distance traveled can be a significant expense which drives many decisions – additional data would be needed to compare if the increased capacity per bus and decreased wait times justified the additional fuel cost and vehicle degradation requiring additional maintenance over time. The paper argues that this implementation provides a proof of concept that can be adopted to increase quality of service in existing fleets by reducing waiting times or by allowing fleets to reduce size to save capitol while maintaining a similar level of performance.

The first critique of this paper is that it is comparing two systems designed around completely different metrics to justify their model’s performance. If we look at a standard schedule bus route, it operates on fixed time fixed routes with minor variations to account for traffic conditions like construction or accidents. These fixed-schedule routes are not designed with passenger wait time as the primary metric – they are designed around cost factors that restrict their budgetary options. Limited government funding and grant funding are the primary sources for public transport with the bulk of costs being operating costs. These operating costs include the fuel required to operate the bus and the maintenance required at fixed distance intervals. Both gas and maintenance have a distance component to them, so a reduction in vehicle distance is a point of optimization for these systems. Along with that optimization, public transit schedules off rider density to balance availability to passengers with that reduction in operating costs. In the paper, one point identified by the authors is that the dynamic nature of the MADDGP approach is that the autonomy of the buses allowed them to adapt to demands. This meant that the buses could perform U-Turn actions which moderately changed their loop order to optimize for low wait times. This behavior led to an increase in vehicle distance traveled which is a cost factor that is meant to be limited in scheduled bus loop. While the cost argument could be made that the increase in capacity offsets the increase gas and maintenance costs associated with the increased vehicle distance, it still leaves the impression that the main comparison is between a system bounded by one attribute versus one that is not bound which may impact the accuracy of the results. If the MADDGP were also bounded in a similar way to a scheduled bus loop where they had to yield results under similar vehicle distance measurements, it could be a more accurate measurement of the MADDGP performance in routing decisions over a fixed loop.

The second critique for this paper is that the constructed simulation environment and fleet model does not demonstrate the model’s operation under subpar conditions. The claim that “the dispatching model presented in this paper can be employed by autonomous bus fleet operators that plan to offer an on-demand service” is only validated in the sense of the optimal. The conditions are such that outside of metrics mentioned in the first critique, there is no demonstration of the fleet’s ability to optimally route in the event of a direct failure of one of the units. If we look at scenario 3, and at a large volume of passengers on the order of 1920 per hour, the current model does not demonstrate how it would scale should one of the five vehicles undergo a failure. The scalability did demonstrate that going from 4 to 5 vehicles does improve the load, but the adaptability of the multiagent model to account for larger volumes than optimized for in the event of a failure is not demonstrated. Figure 8a demonstrates a model operating under saturated passenger demands but the case of a redistribution across already saturated vehicles is not present. Building off of this situation where each bus is saturated in passenger demand, the decision metric for which stop the bus should go to given that each stop is saturated evenly is also not stated. If stops A, B and C have the same passenger demand, and the buses are loaded evenly, how will the model respond in what appears to be a tied scenario? Should the claim be that the system is at a point ready for adoption into an autonomous fleet, then metrics for failure response should be present and measured within the three scenarios used for evaluation of performance. Additionally, if this model is focusing more on a proof of concept that this type of system could be developed, it should not state that the model could be employed, only that the model could be further developed into a deployable model with additional research and validation.

The final critique for the paper is that using the average wait time for passengers as the performance metric only covers half of the passenger experience. The multiagent model is capable of making U-Turns and can adapt to passenger demand – this has been well established in the model. The metric that is not stated in the research is passenger time in transit, and by extension, a measure of the amount of time spent in transit above the expected for a schedule route-based transit system. For changes in routing, there is the possibility that the adjusted route may be more efficient for passenger demand to get on the bus but that the adjusted route increases the transit time for passengers already on the bus. For example, let us assume stops A, B, C, and D. A passenger boards at stop A with the intention to get off the bus at stop D. During this time, the bus is never more than half full which allows a volume of passengers to be constantly embarking and disembarking which is performing properly to the average wait time metric. Demand to be picked up at stop D remains quite low, while demand at stops A, B, and C are significantly higher. The bus may make it to stop D, but it may also reroute to stops A, B, and C multiple times. The metric for the time passengers spend waiting will appear optimal in this setting; however, the measurement of passenger travel time, or the difference in passenger travel time compared to the standard time on a route based bus (that the system poses itself as a replacement for) may be drastically higher. There is a degree of consistency with the time schedule bus, and there is not an available metric visible in this research for the equivalent of that time on the multiagent model’s fleet. Using the scenario previously mentioned occur, it could be a difficult problem to sway passengers to use a system that poses the potential of significantly delaying their schedule at the improvement of a couple minutes of waiting at a bus stop. The counter argument may be that a bus stuck on route cannot adapt to a massively increased load, but there is the argument of if the dynamic addition is a higher benefit of the model than the actual decision processes at work. Essentially, the question of if you dynamically allocated buses to a time based route to meet demand, would it perform comparably to the system while yielding better travel times, becomes the focus from there.

This research demonstrates that deep reinforcement learning can be used to train a multiagent system under this simulation environment. To build off where the research finished, an application based in consumer vehicle fleets vs public and private transit options would be a strong area of exploration. Rather than modeling the performance across a series of stops, a company with large amounts of travel data like Uber, Lyft, Waze, or any other company operating in the consumer transit sector could use common areas of transit and a limited range of GPS positions to model similar multiagent systems that operate under larger areas and driving conditions. The further research section of the paper specified looking into autonomous consumer vehicles, but the current usages of autonomous vehicles are slim due to safety concerns and regulation. This makes autonomous vehicles a currently niche area in the consumer space and coordinating a fleet of autonomous vehicles is a niche of a niche topic at this point in time. Significant widespread adoption of autonomous vehicles would be needed for coordination to become a more everyday problem. Instead of focusing on a potential future optimization, using this type of model to coordinate driver operation areas for ride sharing could be a point of expansion since consumer wait time would be a point of optimization and the dynamic nature of gig work would require the ability to scale to demand and handle random losses of a given vehicle. Since each transit only satisfies a single transaction, the measurement of time to destination over the average disappears as well given that the only two “stops” in a consumer ride share is the pickup and drop-off locations

The current operation of consumer ride sharing uses a mix of a queue-based system to reduce rider wait times with distance optimizations to make drivers’ efforts worth the pay of the ride. While this system does work well to a certain degree, the ability to have that delegation of rider pickup times and locations to better match the distribution of riders rather than balancing first come first serve with cost effectiveness could be a larger cost saving measure in the long run. The system would have to behave differently, in the sense that a vehicle would have to disappear from availability from the moment of a passenger being picked up to a passenger being dropped off, and from there be considered available again. The system would need to be able to handle multiple areas and optimize driver distributions off an everchanging distribution of vehicle availability. In comparison with the bus stops application, where each possible destination is known, there would need to be an overarching distance calculation in place or a heatmap that would assist in matching drivers to optimal riders without negatively affecting rider experience. The system would also need to be significantly more tolerant to traffic conditions, since the degree of traffic could vary wildly where in short route scenarios as described in scenarios 1, 2, and 3 the overall traffic behavior and history should be well known due to the degree of planning needed for proposing, funding, and creating public bus routes. A similar, but larger in scope simulation for adapting this research to consumer ride sharing could model suburban and downtown travel of short, intermediate, and long distances using existing ride data if the assumption is made that the dataset from a large ridesharing company is available or a similar dataset is available to the public. Even if those databases are not available, traffic patterns are public information and readily available so information could be scraped from public sites to pull the expected behavior for given areas at given times for the purposes of training or modeling similar scenarios in a simulation environment.