Reinforcement Learning and Mechanism Design for Routing of Connected and Autonomous Vehicles

Doctoral Consortium

Behrad Koohy
Citizen-Centric Artificial Intelligence Systems, University of Southampton
Southampton, United Kingdom
behrad.koohy@soton.ac.uk

ABSTRACT

The data provided by Connected and Autonomous Vehicles (CAVs) is a powerful tool, providing insight into user incentives and preferences, and combined with existing road data sources, provides a number of new research avenues for intelligent traffic systems. In this paper, we propose the use of Reinforcement Learning (RL) for adaptive pricing of travel systems such as trains, buses and toll-road, in simulations which consider multiple transport providers and traffic management systems, known as the multi-market pricing problem. We also propose two research directions for this problem, the use of incentives when user preferences are included and development of detection and prevention of unintentional collusion between RL pricing agents.

KEYWORDS

Connected and Autonomous Vehicles; Vehicle Routing; Reinforcement Learning; Adaptive Pricing

ACM Reference Format:
Behrad Koohy. 2023. Reinforcement Learning and Mechanism Design for Routing of Connected and Autonomous Vehicles: Doctoral Consortium. In *Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), London, United Kingdom, May 29 – June 2, 2023*, IFAAMAS, 3 pages.

1 INTRODUCTION

The emergence of Connected and Autonomous Vehicles (CAVs) offers significant opportunities for enhancing infrastructure efficiency and planning for new infrastructure. The data provided by CAVs, when combined with other data sources like induction loop sensors, offers significant potential for developing efficient, fair, and intelligent routing systems [3]. This increase in data from CAVs (combined with existing sources of data such as road-side sensors) have made it feasible to apply developments in Reinforcement Learning (RL) to intelligent traffic systems in contexts such as Traffic Signal Control (TSC) [13], electric vehicle charging [11] and autonomous vehicles [1]. One domain in intelligent traffic systems which could benefit from the ability to learn from complex, real-world data that is provided by RL is road and congestion pricing.

Congestion pricing models require a quantitative value for the marginal slowdown caused by the individual, which is often determined through stylized traffic models that assume deterministic conditions and may not reflect real-world scenarios [12]. A commonly proposed method of congestion pricing, known to increase traffic flow is to charge road users a cost proportional to the negative externalities (i.e. increased travel time) caused to other road users, known as a Pigouvian tax [10]. Sharon et al. [12] introduces this pricing strategy as a micro-tolling paradigm. RL has also found use in the context of road pricing, also known as congestion pricing; Mirzaei et al. [9] introduced enhanced delta-tolling, which uses RL to find the optimal parameters within the micro-tolling paradigm introduced by Sharon et al. Micro-tolling calculates the price for each link in a transport network by using two parameters: the known free-flow travel time and the current travel time [12]. This method of pricing improves on macro-tolling which is limited by assumptions of constant demand and capacity for links [12].

Existing approaches in the area of road pricing often view this problem from the perspective of a single travel provider or traffic management tool (i.e. TSC [2, 4, 8, 14] and micro-tolling [9, 12]. Furthermore, proposed adaptive solutions to this problem do not consider user preferences and intention, and are limited to demand-based adaptations. Our research abstracts the micro-tolling road routing problem, referred to as the multi-market pricing problem, and expands it to include multiple travel providers such as buses, trams, and toll roads. This approach provides a more comprehensive view of the road state and can inform optimal policy decisions for stakeholders such as users, local authorities, and government bodies. Additionally, it enables research into cooperation between systems and travel providers that traditionally may not interact.

In section 2, we introduce the Multi-Market Routing Problem (MMRP). In section 3, we discuss our previous work utilising RL for TSC, and how it relates to the MMRP. Section 4 introduces the in progress solution, and sections 5 & 6 are proposed areas of future work.

2 MULTI-MARKET ROUTING PROBLEM

The MMRP is defined by the tuple $(O, D, T, F, \pi(x), V, U)$, where $o \in O$ is the set of origins, $d \in D$ is the set of destinations, $t_x \in T$ where T represents the set of travel providers and t_x is the provider which connects o_x to d_x, F represents the set of free travel options which require a travel cost ϵ to get to from all origins but do not cost the user anything to use, $\pi(x)$ represents the set of pricing policies for the travel providers, $v(x, c) \in V$ is the set of volume delay functions which determine the travel time on travel provider x where c is the current capacity and $u \in U$ is the set of road users. Users u arrive at random to one of the origins o with a randomly chosen destination. We also provide each user $u \in U$ with a maximum budget, where routes above this cost are infeasible,

Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), A. Ricci, W. Yeoh, N. Agmon, B. An (eds.), May 29 – June 2, 2023, London, United Kingdom.

a value that is generated randomly when the users arrive at one of the origins o. Users are also able to see the travel time of all travel options before they make a decision.

We stipulate that the travel time of $\forall f \in F, \forall t \in T, V(f, 0) < V(t, 0)$. This provides a competitive advantage for the competing travel providers over the free travel route. The objective of this problem is to maximise the profits through the use of adaptive pricing policies in place of $\pi(x)$. This problem is a subset of the larger traffic routing problem, consisting of multiple such link-based problems.

3 PROBLEM OF NON-STATIONARITY

Reinforcement learning has recently emerged as a powerful tool for addressing the complex, dynamic nature of traffic signal control. In recent years, there has been a surge of interest in using RL for TSC, as evidenced by the growing number of papers published on the topic [2, 4, 14]. Whilst TSC often views the problem of multiple connected intersections, decisions made at single intersections can have significant impacts on surrounding intersections. Whilst our problem contrasts TSC as we view single links in a traffic network rather than the macroscopic problem, lessons from RL for TSC can be applied to ensure our solution is reliable and effective when utilised in the real world.

In previous work, I reviewed the use of reward functions to mitigate the problem of non-stationarity in multi-agent TSC [8]. I found that the choice of reward function is crucial in dealing with non-stationarity, and highlighted the importance of using real-world data and robust simulations. These lessons will be applied to our research on the multi-market routing problem.

Our proposed future work will consider the Problem of Non-Stationarity, as agents may learn as they are compete, and their future policies will be impacted by all agent's previous policies The reward function could also be a powerful tool in managing unintentional agent collusion.

4 ADAPTIVE PRICING MODELS

An optimal pricing policy is one which provides the maximum profit for the travel provider [15]. The dynamic nature of traffic requires an adaptive pricing policy to avoid over-usage or under-usage of available capacity. Setting the price too high may discourage use, while setting it too low may cause over-subscription and delays, making the option unattractive and losing potential profits.

For an adaptive pricing policy, we propose the use of RL as analytical solutions are infeasible, and RL will learn dynamically from complex real-world data. RL can recognize the relationship between controllable parameters and traffic flow [9], leading to more effective traffic management. RL has been demonstrated to be effective in oligopoly settings, like our research on dynamic pricing under competition in eCommerce, as shown in Kastius et al. [7]. State-of-the-art RL methods such as Twin Delayed Deep Deterministic Policy Gradients have shown potential in similar control problems such as oscillation damping control [5].

To evaluate our RL model, we will use data from CAVs, roadside sensors, and origin-destination matrices for a realistic assessment. We will also compare it against various agents such as deterministic agents (fixed price, limited/unlimited two-bound, and values derived from analytical solutions) and non-deterministic agents (random pricing models, noisy pricing models, including adaptive noisy models). Additionally, we will test the effectiveness of the adaptive pricing policies under uncertainty by incorporating user behavior data and likely future road states into the pricing model.

5 INCENTIVES AND TRAFFIC MANAGEMENT

One area of future work is to introduce a mechanism which allows for the RL agents to offer incentives to users to take certain routes. In [16], an intention-aware routing algorithm is proposed with incentives, and the authors find that operational costs for a fleet of delivery vehicles is reduced by up to 30%. It is important to highlight that one of the methods of accessing road user intention is to incorporate data from CAVs. User preferences will be included when calculating incentives, including a dynamic response in decision-making to pricing policy changes.

An effective, equitable and reliable incentive system would allow for better usage of travel infrastructure, including public transport, where schemes such as "Park and Ride" [6] have been implemented in an attempt to reduce the number of vehicles on the road in city centres.

One further use case of incentives can be to manage environmental considerations in areas by balancing demand in specific areas at specific times (e.g. limiting air pollution around schools). This could be a valuable tool for local authorities, from known scenarios such as the school example, to situations such as accidents and disaster response where sections of the road network are required by emergency services.

6 UNINTENTIONAL COLLUSION AVOIDANCE

Kastius et al. [7] found that RL agents in oligopolies can be forced into collusion without direct communication. To prevent manipulation and protect the interests of road users, local authorities, and governments, it is essential to evaluate strategies that introduce collusion. This can be done by designing adaptive agents which have the objective of forcing collusion, and testing our proposed solution's resilience to these strategies. Additionally, research should explore methods to identify unintentional collusion and adapt pricing strategies accordingly. The effects of these strategies should be evaluated in both early-stage and optimized agents, taking into account the non-stationarity of the problem, which can favor the antagonist agent. To ensure transparency and equity, this research will also draw on explainable RL techniques.

7 CONCLUSIONS

This extended abstract highlights multiple open research questions in the vehicle routing and congestion pricing problem, the feasibility of which is reliant on data from CAVs. The use of RL for this problem would introduce adaptive pricing strategies which can respond to traffic scenarios in a way that analytical solutions are not able to. The research also aims to investigate the use of mechanism design for incentives to manage congestion and the potential for collusion avoidance in RL pricing strategies, which could have wide-reaching implications beyond the transport sector.

ACKNOWLEDGMENTS

This research is supported by an ICASE studentship from EPSRC and Yunex Traffic, the EPSRC AutoTrust platform grant (EP/R029563/1), and an EPSRC Turing AI Acceleration Fellowship on Citizen-Centric AI Systems (EP/V022067/1).

REFERENCES

[1] Szilárd Aradi. 2020. Survey of deep reinforcement learning for motion planning of autonomous vehicles. *IEEE Transactions on Intelligent Transportation Systems* (2020).

[2] James Ault and Guni Sharon. 2021. Reinforcement Learning Benchmarks for Traffic Signal Control. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 1)*.

[3] Saeed Asadi Bagloee, Madjid Tavana, Mohsen Asadi, and Tracey Oliver. 2016. Autonomous vehicles: challenges, opportunities, and future implications for transportation policies. *Journal of modern transportation* 24, 4 (2016), 284–303.

[4] Chacha Chen, Hua Wei, Nan Xu, Guanjie Zheng, Ming Yang, Yuanhao Xiong, Kai Xu, and Zhenhui Li. 2020. Toward a thousand lights: Decentralized deep reinforcement learning for large-scale traffic signal control. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 3414–3421.

[5] Qiushi Cui, Gyoungjae Kim, and Yang Weng. 2021. Twin-Delayed Deep Deterministic Policy Gradient for Low-Frequency Oscillation Damping Control. *Energies* 14, 20 (2021), 6695.

[6] Ardeshir Faghri, Adam Lang, Khaled Hamad, and Heather Henck. 2002. Integrated knowledge-based geographic information system for determining optimal location of park-and-ride facilities. *Journal of urban planning and development* 128, 1 (2002), 18–41.

[7] Alexander Kastius and Rainer Schlosser. 2022. Dynamic pricing under competition using reinforcement learning. *Journal of Revenue and Pricing Management* 21, 1 (2022), 50–63.

[8] Behrad Koohy, Sebastian Stein, Enrico Gerding, and Ghaithaa Manla. 2021. Reward Function Design in Multi-Agent Reinforcement Learning for Traffic Signal Control. (2021).

[9] Hamid Mirzaei, Guni Sharon, Stephen Boyles, Tony Givargis, and Peter Stone. 2018. Enhanced delta-tolling: Traffic optimization via policy gradient reinforcement learning. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 47–52.

[10] Arthur C Pigou. 1920. Some problems of foreign exchange. *The Economic Journal* 30, 120 (1920), 460–472.

[11] Tao Qian, Chengcheng Shao, Xiuli Wang, and Mohammad Shahidehpour. 2019. Deep reinforcement learning for EV charging navigation by coordinating smart grid and intelligent transportation system. *IEEE transactions on smart grid* 11, 2 (2019), 1714–1723.

[12] Guni Sharon, Josiah Hanna, Tarun Rambha, Michael Albert, Peter Stone, and Stephen D Boyles. 2016. Delta-tolling: Adaptive tolling for optimizing traffic throughput. In *ATT@ IJCAI*.

[13] Hua Wei, Guanjie Zheng, Vikash Gayah, and Zhenhui Li. 2021. Recent advances in reinforcement learning for traffic signal control: A survey of models and evaluation. *ACM SIGKDD Explorations Newsletter* 22, 2 (2021), 12–18.

[14] Hua Wei, Guanjie Zheng, Huaxiu Yao, and Zhenhui Li. 2018. IntelliLight: A Reinforcement Learning Approach for Intelligent Traffic Light Control. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (London, United Kingdom) *(KDD '18)*. Association for Computing Machinery, New York, NY, USA, 2496–2505. https://doi.org/10.1145/3219819.3220096

[15] Charles A Wilson. 1988. On the optimal pricing policy of a monopolist. *Journal of Political Economy* 96, 1 (1988), 164–176.

[16] Canqi Yao, Shibo Chen, Mauro Salazar, and Zaiyue Yang. 2022. Incentive-aware Electric Vehicle Routing Problem: a Bi-level Model and a Joint Solution Algorithm. In *2022 American Control Conference (ACC)*. IEEE, 4662–4667.