More Knowledge, More Efficiency: Using Non-Local Information on Multiple Traffic Attributes

Ana L. C. Bazzan, Henrique Uhlmann Gobbi, and Guilherme D. dos Santos

Universidade Federal do Rio Grande do Sul, Brazil
{bazzan,hugobbi,gdsantos}@inf.ufrgs.br

Abstract. New technologies have the potential to transform urban mobility. Among the contributions, providing timely information to drivers via, e.g., apps, is proving valuable. However, providing the same information to nearly everyone is counterproductive. In this paper we extend previous works in which vehicles and the road infrastructure exchange information to allow drivers to make better informed decisions when using reinforcement learning. Here, we use non-local information to augment the knowledge elements of the infrastructure have. Moreover, we connect these elements when they have similar patterns related to multiple attributes, including emission of gases. Our results show that using augmented information leads to more efficiency.

CCS Concepts: • **Computing methodologies → Multi-agent reinforcement learning.**

Keywords: Multi-Agent Reinforcement Learning, Route Choice, Non-Local Information

1. INTRODUCTION

Congestion in urban traffic networks poses challenges to many economies, as well as to the environment. Reducing emissions and lost hours in traffic are among the top priorities of our societies. Conventional traffic management solutions may have reached their limits, as the available methods and tools are not flexible enough, or were not developed having in mind traffic patterns that are arising in mega-cities, or due to new transportation modes, mobility-on-demand, etc. Moreover, those solutions do not necessarily exploit new technologies, such as communication systems, including vehicle to infrastructure (V2I) communication. In fact, although Intelligent Transportation Systems (ITS) have received a lot of attention in the past decades, only recently has this area focused on the avenues opened by fast communication and vehicular networks. By effectively using communication-based approaches, congestion and, consequently, emissions and lost hours can be reduced.

New technologies can be employed in several ways. In this paper we concentrate on V2I as a tool to improve drivers' decisions on how to travel from A to B (pointers on other research directions appear in Section 2.2). Under this particular perspective, while the current pattern is that each individual driver selects a route based on his/her own experience, this is changing with the increasing penetration of new technologies that allow information exchange. Examples of these technologies are not only based on broadcast (e.g., GPS or cellphone information) but also a two-way communication channel, where drivers provide and receive traffic information.

Key here is that these technologies change the paradigm. While currently many traffic management systems are based on a central authority in charge of assigning routes for drivers, or at least providing information (e.g., Waze, Google apps, etc.) for them to decide, communication among vehicles, between vehicles and the road infrastructure, or even among elements of the infrastructure are transformative. In fact, roads are already undergoing the same changes that are seen in the economy, as well as in the society, namely, a decentralization of the decision-making process and, not least, the prominence of several players, such as IT, big tech, and, most importantly, the citizen her/himself.

Right now we are experiencing a situation in which these technologies and platforms are trying to establish themselves, and are still focusing very much on an agenda that is decades old, namely saving travel time. However, more and more, other aspects are being considered when formulating public policies related to urban mobility. One of these aspects concerns the environment, since stop-and-go traffic may cause more emissions. Therefore, while in the past traffic engineering has focused mostly on reducing travel time, emissions are often being taken into account as well.

As discussed in the next section, there are many ways to help improve how the demand (persons, trips, goods) can efficiently use the existing supply (road infrastructure). Most of them rely on centralized approaches though. One way to mitigate this is by letting drivers experience and decide in a decentralized way by means of reinforcement learning (RL), where, given the collective nature of this process, we in fact should use multi-agent reinforcement learning (MARL), aiming at investigating how drivers (or agents) choose their preferable route based on their own learning experiences.

In previous works [Santos and Bazzan 2021; Santos et al. 2021], we have connected MARL to V2I communication, in order to investigate how it could augment the information drivers use in their route choices. Later, we have considered multiobjective RL [Santos and Bazzan 2022], where drivers have two objectives: reduce not only travel time, but also emission of carbon monoxide.

In the present work, we connect the research on multiple attributes, while also considering V2I. Moreover, we add a third element to this, namely, information about non-local interactions. Section 3 details the methodology. Here, it suffices to say that we propose the use of a relationship graph where sections of the traffic network (links) that have similar values for those attributes are connected in such graph, and exchange information about travel time and emissions so that the infrastructure has augmented information, which is then passed to vehicles to allow them to make more informed decisions. To the best of our knowledge, employing this kind of graph is novel in this context. Section 4 reports experiments that show the efficiency of our approach, where informed drivers are able to make decisions that, despite aiming at reducing their travel times, also reduce emissions.

2. BACKGROUND AND RELATED WORK

In this section, we cover some key concepts that underlie our approach. Due to lack of space, for more details on conventional traffic assignment, we refer the reader to Chapter 10 in [Ortúzar and Willumsen 2011]. For our purposes it suffices to mention that conventional approaches are centralized. Instead, this section focuses on MARL-based approaches that allow a decentralized decision-making. Next, we give a brief introduction to RL and MARL. Section 2.2 then discusses the related literature.

2.1 Reinforcement Learning

Reinforcement learning (RL) is a machine learning method, in which agents learn how to map a given state to a given action, by means of a value function. RL can be modeled as a Markov decision process (MDP), where there is a set of states S, a set of actions A, a reward function $R : S \times A \rightarrow \mathbb{R}$, and a probabilistic state transition function $T(s, a, s') \rightarrow [0, 1]$, where $s \in S$ is a state the agent is currently in, $a \in A$ is the action the agent takes, and $s' \in S$ is a state the agent might end up, taking action a in state s. The tuple (s, a, s', r) represents that an agent was in state s, then took action a, ended up in state s' and received a reward r. The key idea of RL is to find an optimal policy π^*, which maps states to actions in a way that maximizes future rewards.

In model-free formulations of RL – such as Q-Learning (QL) –, the agents learn R and T by interacting with an environment. In QL, the agent keeps a table of Q-values that estimate how good it is for it to take an action a in state s; thus a Q-value $Q(s, a)$ holds the maximum discounted value of going from state s, taking an action a and keep going through an optimal policy. In each learning episode, the agents update their Q-values as in Equation 1, where α and γ are the learning rate and the discounting factor for future values, respectively.

$$Q(s,a) = Q(s,a) + \alpha(r + \gamma max_a[Q(s',a') - Q(s,a)]) \tag{1}$$

In RL tasks, it is also important to define how the agent selects actions, while also exploring the environment. A common action selection strategy is the ϵ-greedy, in which the agent chooses to follow the optimal values with a probability $1-\epsilon$, and takes a random action with a probability ϵ.

2.2 Related Work

Traffic assignment problem is not a new problem; there have been several works that aim at solving it. Besides conventional methods, which mostly deal with planning (long term) tasks, and are centralized, RL is turning popular. In this front, methods usually fall into two categories: a traditional (state-based) RL method, and a stateless one. In the latter, each agent d actually is in only one state (its origin location), where it selects a route to travel. A route is defined as a sequence of links that take d from its origin to its destination, thus no en-route decision is necessary. Works in this category are [Ramos and Grunitzki 2015; Ramos et al. 2017; Zhou et al. 2020]. Multiobjective, stateless RL was employed in [Huanca-Anquise 2021], where agents optimize travel time and a second objective, toll.

In the state-based front, each agent d makes decisions at each junction, regarding which link to select next, so that it will eventually reach its destination. In [Bazzan and Grunitzki 2016] this is used to allow agents to learn how to build routes. However, they use a macroscopic perspective by means of cost functions that compute the abstract travel time. In the present paper, the actual travel time is computed by means of a microscopic simulator (see Section 4).

As aforementioned, our approach includes V2I communication, as this kind of new technologies may lead agents to benefit from sharing their experiences, thus reducing exploration. The use of communication in transportation systems, as proposed in the present paper, has also been studied previously ([Grunitzki and Bazzan 2016], [Auld et al. 2019]). In a different perspective, works like [Yu et al. 2020] evaluate the impact of incomplete information sharing.

Also, in order to exploit the potential of V2I communication, in previous works [Santos and Bazzan 2020; 2021; Santos et al. 2021], we have connected it to MARL, in order to investigate how V2I communication could benefit drivers use in their route choices. In these works, the infrastructure is able to communicate with the vehicles, both collecting information about their most recent travel times (on given links), as well as providing them with information that was collected from other vehicles. However, links in the infrastructure only exchange information if they are connected by a junction, i.e., only local information is considered.

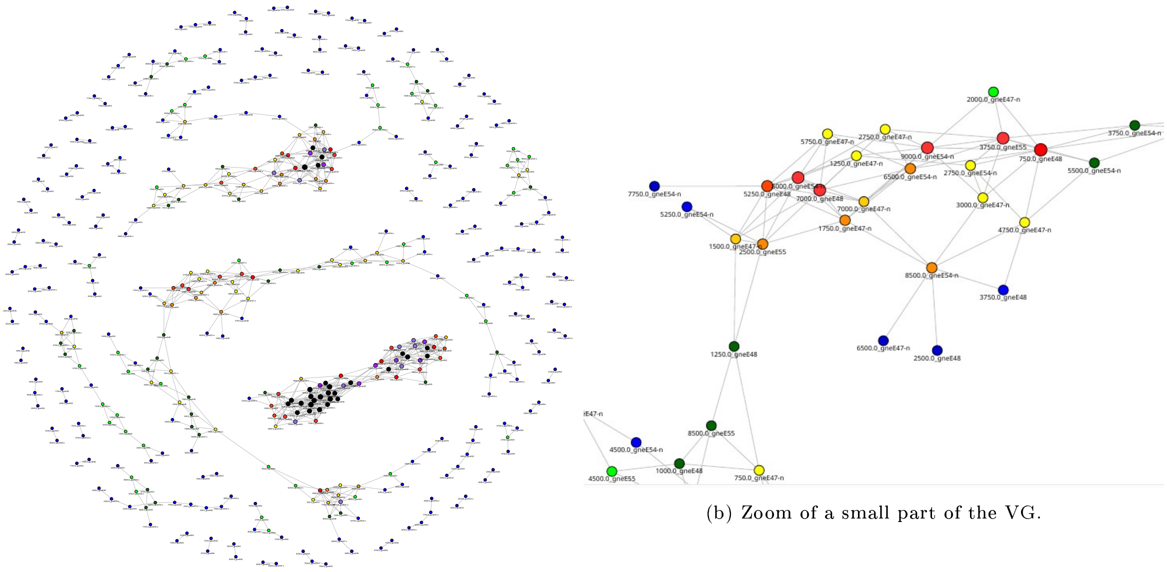
The value of V2I communication has started to receive attention also in the traffic engineering community. The reader is referred to [Mahmassani 2016] (focusing on how autonomous vehicles and connected vehicles are expected to increase the throughput of highway facilities, as well as improve the stability of the traffic stream), and [Maimaris and Papageorgiou 2016] (applications).

Finally, the method we proposed here grounds on graph-based methods. Due to lack of space and the fact that few of them do tackle communication, we refer the reader to a survey: [Cui et al. 2022].

3. METHODOLOGY

3.1 Terminology: Road Network and Virtual Graph

We deal with two sorts of graphs. First, a road network is a (planar) graph $G = (J, L)$, where J is the set of junctions (intersections), and L is the set of links. We use the term link, since it is more commonly used in traffic engineering (and then reserve the term edge for the second graph, as described next). For example, in Fig. 2 links `gneE54` and `gneE55` are both connected to the junction



(b) Zoom of a small part of the VG.

(a) Full VG.

Fig. 1: Instance of a virtual graph VG.

that appears in the center of the figure. As for the second graph, once our approach relies on non-local information, such graph is the one that connects two links $l_1 \in L$ and $l_2 \in L$ which are *not necessarily physically close* (as, e.g., `gneE49` and `gneE45` in Fig. 2), but that have similar patterns. We call this a virtual graph denoted by $VG = (L, E)$, where L is the set of links (note that now they act as vertices in VG), and E is the set of edges that connect two links that have similar patterns, as described next.

In order to define when two links are to be connected in VG, historical information is collected for a network G. This information refers to several attributes: travel time, fuel consumption and several kinds of gas emissions[1], per link, per time interval. We aggregate such information using a time window w_h and normalize the values of all attributes between zero and one. Then, the values of such attributes for each two pairs of links are compared. If two links l_1 and l_2 have the same values for all attributes (given a tolerance value, i.e. $\pm\delta_a$), then an edge connecting l_1 and l_2 is inserted in VG. Fig. 1a shows an instance of such a virtual graph, whereas Fig. 1b depicts a zoom of that graph, where some relationships among similar links can be better seen. The labels of the vertices are formed by the link ID plus the time interval in which their values were found to be similar.

3.2 How Communication Works

Next we briefly explain how the communication is performed by the elements of the road network G. We assume that every junction $j \in J$ and every link $l \in L$ is equipped with a communication device (henceforth, CommDev) that is able to send and receive messages to and from nearby vehicles, as well as among themselves. For instance, in Fig. 2, the red vehicle informs the CommDev of the corresponding link about its rewards in terms of travel time and other attributes, once it has travelled that particular link. Similarly, the CommDev informs the green vehicle the expected travel time in the links ahead, so that the green vehicle is able to decide which link to take, once it reaches the next junction (i.e., the next decision state).

A junction CommDev collects information from its incoming links, defined in the physical road network G. Additionally, given that these incoming links may have virtual neighbors in the virtual

[1] We collect CO, CO2, HC (hydrocarbon), PMx (particulate matter), and NOx.

graph VG, information about their virtual neighbors are also passed to the links CommDev's and, from these to the junction CommDev. Once a CommDev at a junction j has collected such information, it updates a table in which the last 30 entries are kept in a FIFO way, for each attribute. This value was used in [Santos and Bazzan 2020] for the same scenario. Moreover, in the present paper, a CommDev stores information also about travel time and CO emission.

CommDev's then communicate to each nearby driver agent an aggregation of those values kept in the tables[2], i.e., potential rewards that the agent may obtain if selecting each action in that particular state. The agent then perceives this information as expected rewards for the actions available to it.

3.3 MDP Formulation

As mentioned in Section 2, a RL learning task is formulated by an MDP. In our case, given a network G, the set of states is defined by J. Two particular states are the origin and the destination of an agent. They define the so-called origin-destination (OD) matrix or set of OD pairs, which basically shows how many trips start and end in each location of the network. A_d^j denotes the set of actions available to agent d at j, which are the links that leave j. Since we deal with a maximization task, each reward is given by the negative of the travel time experienced by d at link l. This value is provided by the microscopic simulator we use (see next section).

Note that in the standard QL algorithm, the agents update their Q-values based on the feedback from the action they have just taken. However, in our case agents also update their Q-values based on the expected rewards received by the CommDev's. This means that every time they reach an intersection, they also update their Q-values with the information provided by the CommDevs.

4. EXPERIMENTS, RESULTS, AND ANALYSIS

4.1 Scenario: Network and Demand

Simulations were performed using a microscopic simulator called SUMO [Lopez et al. 2018], whose API was used to allow vehicle agents to interact with the simulator during simulation time. The network used is shown in Fig. 2, where the main links are two-way. Trips originate in each of the four most external links (gneE63, gneE64, gneE65, and gneE66), and have the other three of these links as destination (as, e.g., gneE63 to gneE64, gneE65, and gneE66), thus defining 12 OD pairs, with 400 trips each. This demand was set to maintain the network populated at around 30% of its maximum capacity, (given that a vehicle occupies $5m$), which is considered a high occupation.

4.2 Model Parameters, Performance Metrics, and Results

For the various parameters of the model, we have used the same values as in [Santos et al. 2021], in which two different networks used these values. This way, we have set learning rate $\alpha = 0.5$, the discount factor $\gamma = 0.9$, and $\epsilon = 0.05$. These values guarantee that the future rewards have a considerable amount of influence in the agent's current choice, since γ has a high value. Other parameters take these values: $w_d = 250$ and $\delta_a = 0.005$.

To measure the performance, we collect travel time and CO emission, over all links $l \in L$ of the network. Given the probabilistic nature of the process, 30 runs were performed. Plots ahead, in which shadows account for the deviations over the runs, show a comparison between three methods: without learning, and QL with and without the information of the virtual graph VG. Note that it

[2] The information about CO is not used by the agent, given that QL only optimizes for one objective – in this case travel time – but, as discussed in Section 5, this will be addressed in a future work, in a similar way as in [Santos and Bazzan 2022].

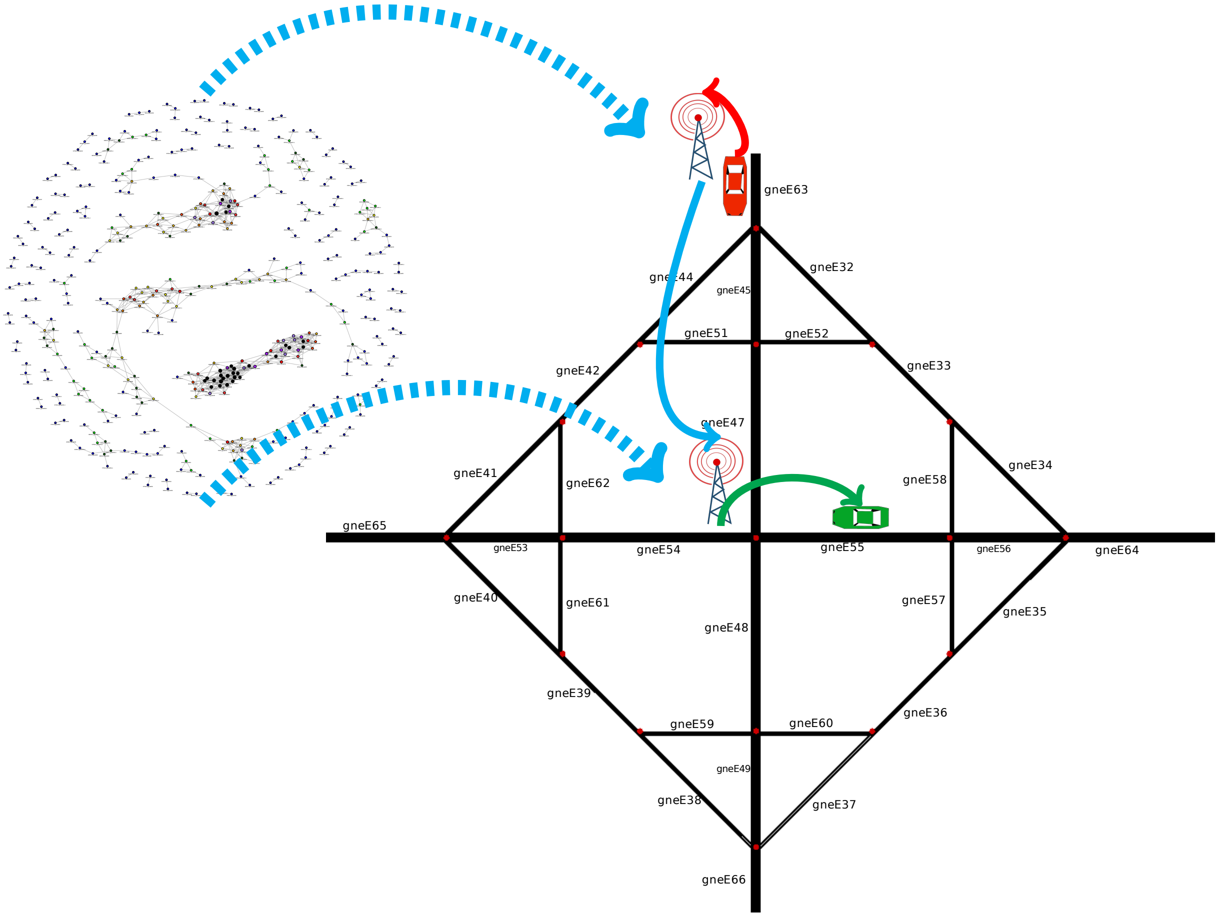
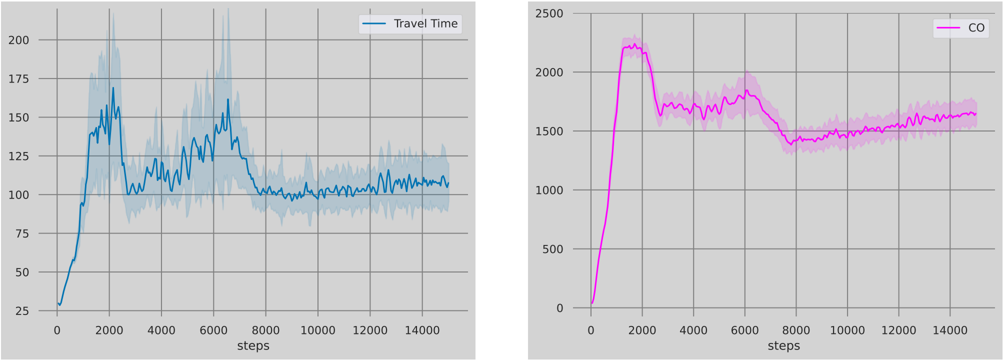


Fig. 2: Road network used in the experiments (labels for some links that run in an opposite direction are omitted; they are similarly labelled: `gneE66-n` for example.). This figure also depicts a scheme of the communication among various elements of the road infrastructure, as well as reproduces the VG shown in Fig. 1a.

takes some time for all vehicles to be loaded, hence the initial oscillations in all plots. We start by discussing the plots that refer to how travel time changes along time: Fig. 3a, Fig. 4a, and Fig. 5a. In the former, agents do not learn. The travel time starts to stabilize around step 8,000, roughly at 100 steps or seconds (per link). When agents use QL, they have to experiment in the beginning, until they converge to decisions (about choice of links), that lead to lesser travel times. Thus, after step 10,000, the travel time shown in Fig. 4a is lower than that in Fig. 3a.

When the virtual graph is used, the convergence to a lower travel time happens earlier, due to the fact that links with similar patterns exchange information that helps CommDev's better inform drivers about which links to select. Also, note that there are less deviations (blue shadow).

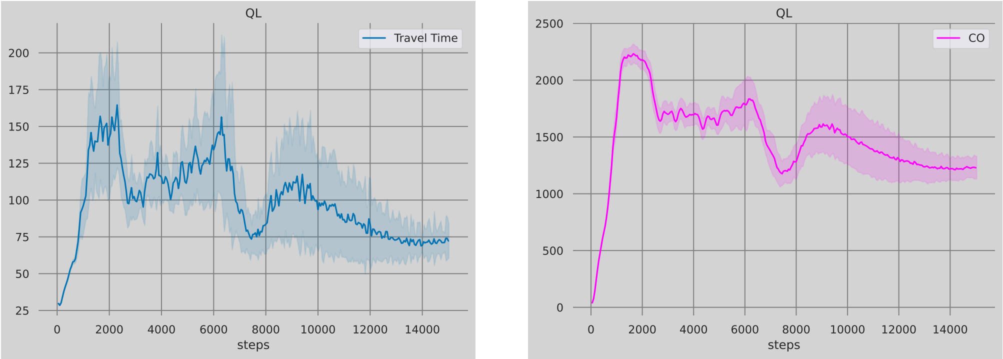
As for the emission of CO, we recall that, when using QL, drivers only optimize for travel time, as QL does not handle more than one reward value, except if they are somehow combined in a function. Despite this, the use of the virtual graph (that corresponds to similarities among links using several attributes, including CO), also leads to reduction of CO emission. This can be seen by comparing Fig. 3b, Fig. 4b, and Fig. 5b.



(a) Travel time (average over all links).

(b) CO emission in mg/s (average over all links).

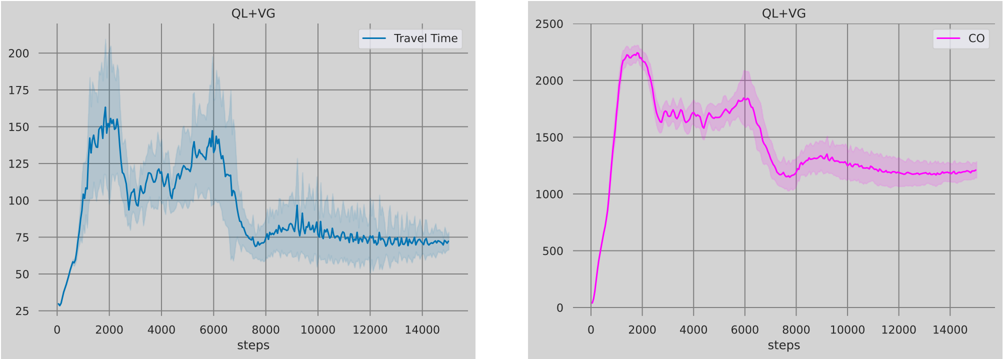
Fig. 3: No learning.



(a) Travel time (average over all links).

(b) CO emission in mg/s (average over all links).

Fig. 4: QL, no virtual graph.



(a) Travel time (average over all links).

(b) CO emission in mg/s (average over all links).

Fig. 5: QL plus virtual graph.

5. CONCLUSIONS AND FUTURE WORK

The use of new communication technologies in urban mobility is turning more and more important. MARL is an attractive method for route choice, as it mimics the way drivers perform experimentation in their daily commuting. The present paper presented a method that combines MARL with V2I communication to allow the road infrastructure to collect and use non-local information, and form a virtual neighborhood, where links that have similar patterns regarding attributes such as travel time and emission of gases are virtual neighbors. Such augmented vision is then passed to vehicles for their decision-making about which link to follow next. We compared our approach with two others, and showed that it improves the efficiency of the learning process.

As future work, we intend to investigate the use of a multiobjective RL approach, which would include further attributes in the reward function.

REFERENCES

Auld, J., Verbas, O., and Stinson, M. Agent-based dynamic traffic assignment with information mixing. *Procedia Computer Science* vol. 151, pp. 864–869, 2019.

Bazzan, A. L. C. and Grunitzki, R. A multiagent reinforcement learning approach to en-route trip building. In *2016 International Joint Conference on Neural Networks (IJCNN)*. pp. 5288–5295, 2016.

Cui, K., Tahir, A., Ekinci, G., Elshamanhory, A., Eich, Y., Li, M., and Koeppl, H. A survey on large-population systems and scalable multi-agent reinforcement learning, 2022.

Grunitzki, R. and Bazzan, A. L. C. Combining car-to-infrastructure communication and multi-agent reinforcement learning in route choice. In *Proceedings of the Ninth Workshop on Agents in Traffic and Transportation (ATT-2016)*, A. L. C. Bazzan, F. Klügl, S. Ossowski, and G. Vizzari (Eds.). CEUR Workshop Proceedings, vol. 1678. CEUR-WS.org, New York, 2016.

Huanca-Anquise, C. A. *Multi-objective reinforcement learning methods for action selection: dealing with multiple objectives and non-stationarity.* M.S. thesis, Instituto de Informática, UFRGS, Porto Alegre, Brazil, 2021.

Lopez, P. A., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flötteröd, Y.-P., Hilbrich, R., Lücken, L., Rummel, J., Wagner, P., and Wiessner, E. Microscopic traffic simulation using SUMO. In *The 21st IEEE International Conference on Intelligent Transportation Systems*, 2018.

Mahmassani, H. S. Autonomous vehicles and connected vehicle systems: Flow and operations considerations. *Transp. Sci.* 50 (4): 1140–1162, 2016.

Maimaris, A. and Papageorgiou, G. A review of intelligent transportation systems from a communications technology perspective. In *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*. pp. 54–59, 2016.

Ortúzar, J. d. D. and Willumsen, L. G. *Modelling transport.* John Wiley & Sons, Chichester, UK, 2011.

Ramos, G. de. O., da Silva, B. C., and Bazzan, A. L. C. Learning to minimise regret in route choice. In *Proc. of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2017)*, S. Das, E. Durfee, K. Larson, and M. Winikoff (Eds.). IFAAMAS, São Paulo, pp. 846–855, 2017.

Ramos, G. de. O. and Grunitzki, R. An improved learning automata approach for the route choice problem. In *Agent Technology for Intelligent Mobile Services and Smart Societies*, F. Koch, F. Meneguzzi, and K. Lakkaraju (Eds.). Communications in Computer and Information Science, vol. 498. Springer, pp. 56–67, 2015.

Santos, G. D. dos. and Bazzan, A. L. C. Accelerating learning of route choices with C2I: A preliminary investigation. In *Proc. of the VIII Symposium on Knowledge Discovery, Mining and Learning.* SBC, Rio Grande, pp. 41–48, 2020.

Santos, G. D. dos. and Bazzan, A. L. C. Sharing diverse information gets driver agents to learn faster: an application in en route trip building. *PeerJ Computer Science* vol. 7, pp. e428, March, 2021.

Santos, G. D. dos. and Bazzan, A. L. C. A multiobjective reinforcement learning approach to trip building. In *Proc. of the 12th International Workshop on Agents in Traffic and Transportation (ATT 2022)*, A. L. Bazzan, I. Dusparic, M. Lujak, and G. Vizzari (Eds.). Vol. 3173. CEUR-WS.org, pp. 160–174, 2022.

Santos, G. D. dos., Bazzan, A. L. C., and Baumgardt, A. P. Using car to infrastructure communication to accelerate learning in route choice. *Journal of Information and Data Management* 12 (2), 2021.

Yu, Y., Han, K., and Ochieng, W. Day-to-day dynamic traffic assignment with imperfect information, bounded rationality and information sharing. *Transportation Research Part C: Emerging Technologies* vol. 114, pp. 59–83, 2020.

Zhou, B., Song, Q., Zhao, Z., and Liu, T. A reinforcement learning scheme for the equilibrium of the in-vehicle route choice problem based on congestion game. *Applied Mathematics and Computation* vol. 371, pp. 124895, 2020.