Regret Minimisation and System-Efficiency in Route Choice

Gabriel de O. Ramos[1], **Ana L. C. Bazzan**[1] **(advisor), Bruno C. da Silva**[1] **(coadvisor)**

[1]Instituto de Informática – Universidade Federal do Rio Grande do Sul – Brazil

{goramos,bazzan,bsilva}@inf.ufrgs.br

Abstract. *Traffic congestions present a major challenge in large cities. Considering the distributed, self-interested nature of traffic we tackle congestions using multiagent reinforcement learning (MARL). In this thesis, we advance the state-of-the-art by delivering the first MARL convergence guarantees in congestion-like problems. We introduce an algorithm through which drivers can learn optimal routes by locally estimating the regret associated with their decisions, which we prove to converge to an equilibrium. In order to mitigate the effects of selfishness, we also devise a decentralised tolling scheme, which we prove to minimise traffic congestion levels. Our theoretical results are supported by an extensive empirical evaluation on realistic traffic networks.*

1. Introduction

Efficient urban mobility plays a major role in modern societies. Notwithstanding, the fast-growing demand for mobility associated with the lack of appropriate investments has compromised the efficiency of traffic systems, as evidenced by the increasing number (and intensity) of traffic congestions. In fact, the cost imposed by traffic congestions on the economy of Brazil in 2013 was higher than 1% of its GDP [Cintra 2014]. In the USA and the UK, such an impact accounted for 0.7% of the GDP of these countries, and it is yet expected to increase by more than 50% until 2030 [CEBR 2014].

Traditional approaches for dealing with arising traffic congestions include expanding the physical capacity of existing traffic infrastructure. Nonetheless, such approaches have proven unsustainable from many perspectives (e.g., economic, environmental) and may even deteriorate the traffic performance. Against this background, ways of making a more efficient use of the existing infrastructure have shown necessary [Lee et al. 2017].

In this thesis, we aim at minimising traffic congestion by approaching the problem from the drivers' perspective. Specifically, traffic networks can be modelled as multiagent systems, where drivers represent self-interested agents that are all competing for a common resource. We then consider the *route choice problem* in particular, which concerns how commuting drivers choose routes to travel between their origins and destinations everyday, while trying to minimise some kind of cost (e.g., travel time) associated with their trips. In this context, traffic is typically described using the notions of user equilibrium (UE, where no driver benefits from unilaterally changing its route) and system optimality (SO, where the average travel time—or congestion level—is minimum) [Wardrop 1952].

At first, a possible approach towards minimising congestions could be to have a central authority computing an optimal traffic assignment and then explicitly saying which route each driver should take. However, dictating what each driver should do is unrealistic and even pointless, since no rational driver would accept to take a socially-desirable route (i.e., one that improves social welfare) in detriment of its own performance.

Building upon the above challenges, we are interested in understanding how drivers can effectively learn to make their own decisions based on previous experience. We can then approach the problem from the perspective of multiagent reinforcement learning, where driver agents need to concurrently learn their optimal routes while relying exclusively on their own experience. By proceeding this way, we look forward to delivering effective traffic solutions that, in a near future, could be easily deployed to help drivers choosing their routes and, thus, to minimise traffic congestions in large cities.

1.1. Challenges

Multiagent reinforcement learning (MARL) is challenging because self-interested agents need to adapt to each others' decisions, which makes learning an optimal policy a moving target [Tuyls and Weiss 2012]. Due to such dynamics, no convergence guarantees exist for the general MARL setting, i.e., for an arbitrary number of players and actions. In order to overcome such limitations, we exploit the structure of a problem in particular (namely the route choice problem) and deliver convergence guarantees (both to the UE and to the SO) of MARL in congestion-like problems.

In order to deliver the aforementioned guarantees, we need to approach the intended solution concepts in particular. Towards the user equilibrium, we consider the class of regret-minimising learning algorithms. Regret measures how much worse an agent performs on average in comparison to the best fixed action (i.e., route) in hindsight. By minimising their regret, agents converge to the user equilibrium. However, employing regret in realistic settings is challenging because (i) regret can only be computed by a central authority with full knowledge about the current state of the system, and (ii) regret does not distinguish how much each action contributes to the regret formulation. The challenge here thus refers to enabling agents to compute regret-based metrics using only their locally available information (i.e., the actually perceived travel times).

Next, towards system optimal guarantees, it is necessary to enforce drivers to take socially-desirable routes. Recall that, drivers' behaviour is intrinsically self-interested, meaning that any driver would prefer routes that minimise its own costs (regardless how much that choice may deteriorate the performance of other drivers). In this context, we employ the notion of marginal-cost tolls, where a route's cost is given by a combination of travel time and money expenses. Marginal-cost tolls are known to mitigate the effects of selfishness. The drawback here, however, is that this notion typically assumes that tolls are charged at every road/street and that their values are computed by a central authority. Dropping these assumptions, while still ensuring that the selfishness level is minimised, constitutes the main challenge here.

2. Contributions of this Thesis

This thesis advances the state-of-the-art in Artificial Intelligence and Traffic Engineering by introducing two novel multiagent reinforcement learning (MARL) algorithms. Firstly, we devise **Regret-Minimising Q-Learning** (Section 2.1), which allows driver agents to learn their optimal routes by locally estimating the regret associated with their decisions. We prove that this algorithm minimises regret and guarantees convergence to the user equilibrium. Secondly, we present **Toll-Based Q-Learning** (Section 2.2), which introduces a decentralised tolling scheme to neutralise agents' selfishness. This second algorithm is proven to converge to a user equilibrium that is aligned to the system optimum.

To the best of our knowledge, these are the first MARL algorithms to guarantee convergence to optimal solutions for an arbitrary number of players and actions in the context of congestion-like games. Accordingly, this thesis was approved with distinction by the PhD examination committee (which was awarded to less than 20% of the PhD theses defended at INF-UFRGS in 2018).

Due to space limitations, this paper presents only the main contributions and results of the thesis. The complete text of the thesis (including the complete algorithms, theorems, proofs, and experiments) is available at [Ramos 2018].

2.1. Regret-Minimising Q-Learning Towards User Equilibrium

Regret-minimisation provides a powerful paradigm for self-interested agents (drivers) to minimise the costs associated with their actions (routes). We investigate how agents can estimate their regret based exclusively on local information (i.e., the rewards actually observed by them). The idea underlying our approach is that, if agents can *estimate* the regret associated with *particular* actions, then such information could be used to guide their learning process. After all, minimising regret in route choice can be intuitively seen as choosing the best routes.

As the first contribution, we introduce the notion of *action regret*, which measures how much a single action of an agent contributes to its overall performance. The Q-value of an action can then be updated using the corresponding regret (rather than reward) as reinforcement signal.

Building upon the concept of action regret, we devise the **Regret-Minimising Q-Learning** algorithm, which allows agents to estimate the regret associated with each of their actions based exclusively on their previous experiences (i.e., travel time of actually taken routes). Our algorithm then uses the action regret to update the Q-values of the corresponding actions, so that agents end up choosing their optimal actions. In this sense, we not only eliminate the typical assumption of full information made in the literature, but also prove that our regret estimates converge to their true values in the limit.

Next, we present the main theorems derived from our algorithm. We begin with Theorem 1, which establishes an upper bound on the maximum regret obtained by any driver agent using our algorithm. Such a bound tends to zero in the limit.

Theorem 1. *The regret achieved by Regret-Minimising Q-Learning up to time T is upper bounded by $O\left(\left(\frac{K-1}{TK}\right)\left(\frac{\mu^{T+1}-\mu}{\mu-1}\right)\right)$, where K is the number of available actions, and μ is the decay factor of the exploration rate.*

Building upon the above theorem, we can derive Theorem 2, which guarantees that, in the limit, agents converge to a ϕ-approximated user equilibrium, where ϕ is the regret bound of the algorithm. Intuitively, Theorem 2 says that no driver can increase its reward by more than ϕ by unilaterally its route, which tends to zero in the limit.

Theorem 2. *Regret-Minimising Q-Learning converges to a ϕ-approximated user equilibrium in the limit, where ϕ is the regret bound of the algorithm.*

We performed several experiments on realistic road networks. On average, our algorithm approximated the user equilibrium by 99.994%. As compared to other algorithms, this represents a decrease of 48% of the gap to the equilibrium, and a decrease of

19.7% of the external regret, on average. These results are consistent with our theoretical bounds on the regret and equilibrium, which corroborate with our theoretical analysis.

2.2. Toll-Based Q-Learning Towards System-Efficient Equilibrium

Although appealing from the drivers viewpoint, the user equilibrium cannot be said reasonable from the social perspective. In fact, the average travel time under user equilibrium can be considerably higher than the system optimum (where the average travel time is minimum) [Koutsoupias and Papadimitriou 1999]. Studies on real-world road networks have shown that drivers waste on average 30% extra time due to lack of coordination [Youn et al. 2008]. This is a direct consequence of drivers' rationality. The point is that rational agents aim at minimising their own costs, meaning that they will always choose the least cost routes, regardless of their impact on the social welfare. Hence, in practice, drivers cannot be assumed to behave altruistically.

The need for system efficient traffic has motivated this second front of the thesis. We tackle the problem by punishing selfish behaviour (or, equivalently, the lack of altruism) using tolls. In particular, we measure how much agents' decisions affect the performance of other agents. Then, we formulate a tolling scheme that charges/penalises agents proportionally to the cost imposed on others, thus dissuading agents to act selfishly. This is equivalent to the so-called marginal-cost tolling.

As the first contribution of this front, we present a generalisation of marginal-cost tolling that applies to the class of univariate, homogeneous polynomial travel time functions, as shown in the next proposition. This formulation comprises the most commonly-used travel time functions in the literature. Moreover, it allows for the toll values to be computed by agents themselves, using only locally available information.

Proposition 1. *The marginal-cost toll value τ_l on any link l with a univariate, homogeneous polynomial travel time function is $\beta(p_1 x_l^{\beta})$, where β and p_1 represent constants specific to the problem instance. This toll formulation can be computed locally by agents using their observed travel times and additional constants.*

Building upon the above formulation, we introduce **Toll-Based Q-Learning**, which computes the toll dues by the end of the trips, and uses such information in the reward formulation so that agents can learn their optimal routes. The algorithm computes the toll values using only locally available information (i.e., the travel time observed by agents). In this sense, our algorithm eliminates the need for having electronic toll booths spread over the road network. Using our algorithm, drivers converge to the user equilibrium. The main contribution, however, is that our toll formulation aligns drivers' objectives to the social welfare. As a consequence, the user equilibrium achieved by drivers has minimum average travel time. In other words, drivers end up converging to a system optimum from which no driver benefits from unilaterally deviating. This is shown in the next theorem. We call this the system-efficient equilibrium.

Theorem 3. *Toll-Based Q-Learning converges to a system-efficient equilibrium.*

We remark that tolls are designed to penalise undesired (i.e., selfish) behaviour. If toll values do not correspond to the true marginal costs, then such tolls may end up penalising the wrong agents (i.e., those that are *not* acting selfishly). In contrast to other works, we avoid this issue because drivers are only charged once they complete their trips.

Theorem 4. *By charging tolls at the end of trips, Toll-Based Q-Learning ensures that agents are charged exactly their marginal costs (i.e., the cost they impose on others).*

Our theoretical results are supported by extensive experimentation on realistic road networks. Results show that our algorithm approximated the system optimum by 99.95%. These correspond to system-efficient equilibria, where no agent benefits from unilaterally changing route. Hence, the experimental results corroborate our theoretical findings, showing that our algorithm effectively minimises traffic congestions.

3. Publications

The results of this thesis were published among the most prestigious conferences and journals in the areas of Artificial Intelligence and Traffic Engineering. These research papers are listed below, in order of relevance.

- **[Ramos et al. 2017a]** ***AAMAS conference*** **(Qualis A1)**. Introduced Regret-Minimising Q-Learning, together with theoretical and experimental analyses.
- **[Ramos et al. 2019a]** ***Knowledge Engineering Review journal*** **(Qualis B1, conditionally accepted) and [Ramos et al. 2018b]** ***ALA workshop*****.** Devised Toll-Based Q-learning algorithm, together with theoretical and experimental analyses.
- **[Ramos et al. 2019b]** ***ALA workshop*****.** Presented Generalised Toll-Based Q-learning algorithm, together with theoretical and experimental analyses.
- **[Ramos et al. 2018a]** ***Transportation Research Part C journal*** **(Qualis B2 for Computer Science and A1 for Traffic Engineering)**. Performed an analysis on the impact of providing travel information to learning drivers.
- **[Ramos 2017]** ***AAMAS doctoral consortium*** **(Qualis A1)**. Presented a summary of the thesis proposal.
- **[Ramos and Bazzan 2015]** ***GECCO conference*** **(Qualis A1) and [Ramos and Bazzan 2016a]** ***CEC conference*** **(Qualis A1)**. Presented efficient local search heuristics for solving the route choice problem centrally.
- **[Ramos and Bazzan 2016b]** ***ATT workshop***. Presented an initial sketch of our Regret-Minimising Q-Learning algorithm.
- **[Ramos et al. 2017b]** ***ALA workshop*** **and [Grunitzki et al. 2014]** ***ANPET conference***. Introduced traffic simulators partially used to run our experiments.

Apart of the papers above, we have two working papers to be submitted to the *AAAI Conference on Artificial Intelligence* (AAAI, Qualis A1) and to the *Journal of Artificial Intelligence Research* (JAIR, Qualis A1).

References

CEBR, Centre for Economics and Business Research. (2014). The future economic and environmental costs of gridlock in 2030. Technical report, CEBR, London.

Cintra, M. (2014). Os custos dos congestionamentos na cidade de são paulo. Technical report, Fundação Getúlio Vargas, São Paulo.

Grunitzki, R., Ramos, G. d. O., and Bazzan, A. L. C. (2014). Uma ferramenta para alocação de tráfego e aprendizagem de rotas em redes viárias. In *Anais do XXVIII Congresso de Pequisa e Ensino em Transportes (ANPET 2014)*.

Koutsoupias, E. and Papadimitriou, C. (1999). Worst-case equilibria. In *Annual Symposium on Theoretical Aspects of Computer Science*, pages 404–413. Springer.

Lee, M., Barbosa, H., Youn, H., Holme, P., and Ghoshal, G. (2017). Morphology of travel routes and the organization of cities. *Nature communications*, 8(1):2229.

Ramos, G. de. O. (2017). Minimising regret in route choice (doctoral consortium). In *Proc. of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2017)*, pages 1855–1856, São Paulo. IFAAMAS.

Ramos, G. de. O. (2018). *Regret Minimisation and System-Efficiency in Route Choice*. PhD thesis, Universidade Federal do Rio Grande do Sul, Porto Alegre.

Ramos, G. de. O. and Bazzan, A. L. C. (2015). Towards the user equilibrium in traffic assignment using GRASP with path relinking. In *Proceedings of the 2015 on Genetic and Evolutionary Computation Conference*, GECCO '15, pages 473–480. ACM.

Ramos, G. de. O. and Bazzan, A. L. C. (2016a). Efficient local search in traffic assignment. In *Congress on Evolutionary Computation (CEC)*, pages 1493–1500. IEEE.

Ramos, G. de. O. and Bazzan, A. L. C. (2016b). On estimating action regret and learning from it in route choice. In *Proceedings of the Ninth Workshop on Agents in Traffic and Transportation (ATT-2016)*, pages 1–8, New York. CEUR-WS.org.

Ramos, G. de. O., Bazzan, A. L. C., and da Silva, B. C. (2018a). Analysing the impact of travel information for minimising the regret of route choice. *Transportation Research Part C: Emerging Technologies*, 88:257–271.

Ramos, G. de. O., da Silva, B. C., and Bazzan, A. L. C. (2017a). Learning to minimise regret in route choice. In *Proc. of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2017)*, pages 846–855. IFAAMAS.

Ramos, G. de. O., da Silva, B. C., Rădulescu, R., and Bazzan, A. L. C. (2018b). Learning system-efficient equilibria in route choice using tolls. In *Proceedings of the Adaptive Learning Agents Workshop 2018 (ALA-18)*, Stockholm.

Ramos, G. de. O., da Silva, B. C., Rădulescu, R., Bazzan, A. L. C., and Nowé, A. (2019a). Toll-based reinforcement learning for efficient equilibria in route choice. *Knowledge Engineering Review*. Conditionally accepted.

Ramos, G. de. O., Lemos, L. L., and Bazzan, A. L. C. (2017b). Developing a python reinforcement learning library for traffic simulation. In *Proceedings of the Adaptive Learning Agents Workshop 2017 (ALA2017)*, ALA2017, São Paulo.

Ramos, G. de. O., Rădulescu, R., and Nowé, A. (2019b). A budged-balanced tolling scheme for efficient equilibria under heterogeneous preferences. In *Proceedings of the Adaptive Learning Agents Workshop 2019 (ALA-19)*, Montreal.

Tuyls, K. and Weiss, G. (2012). Multiagent learning: Basics, challenges, and prospects. *AI Magazine*, 33(3):41–52.

Wardrop, J. G. (1952). Some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers, Part II*, 1(36):325–362.

Youn, H., Gastner, M. T., and Jeong, H. (2008). Price of anarchy in transportation networks: efficiency and optimality control. *Physical review letters*, 101(12):128701.