Exploring Antifragility in Urban Road Networks: Anticipating Disturbances with Reinforcement Learning

Linghang Sun[a, *], Michail A. Makridis[a], Alexander Genser[a], Cristian Axenie[b], Margherita Grossi[c], Anastasios Kouvelas[a]

[a] *Institute for Transport Planning and Systems, ETH Zürich, 8092 Zurich, Switzerland*

[b] *Computer Science Department, Nuremberg Institute of Technology, 90489 Nuremberg, Germany*

[c] *Intelligent Cloud Technologies Lab, Huawei Munich Research Center, 80992 Munich, Germany*

Abstract

Transport networks in the real world are often susceptible to disturbances, while the established control methods based on control theory focus rather on optimizing the exact timing to implement control, they lack the adaptiveness to cope with disturbances well. Hence, transport systems call for solutions that are more antifragile. Antifragility is a property of systems in which they can increase in performance to thrive as a result of stressors, shocks and volatility. In this work, to exploit and strengthen the learning capability of Reinforcement Learning (RL) and achieve antifragility, we propose a deep RL algorithm to regulate perimeter control in a two-region urban network. Based on the principle of first and second order derivatives, we investigate the impact of various additional state and reward terms on the algorithm to learn from disturbances. In comparison with other methods, the proposed RL algorithm has proven its antifragility and effectiveness.

Key-words: Antifragility, Reinforcement Learning, Traffic Disturbance, Perimeter Control

Introduction

With the ever-growing population in the cities and urbanization, traffic systems have gained both in volume and in complexity. The growth of traffic volume also leads to more incidents and more severe peak hour congestions. Therefore, new challenges and requirements have been posed to the urban road networks and particularly to their control strategies as they need to secure a decent level of service even when challenged by various disturbances. As big data techniques continue to rapidly advance, the concepts of smart cities and intelligent transportation systems are being realized through numerous real-world examples. Consequently, contemporary traffic control methods must possess the ability to learn and adapt towards antifragility in the face of traffic disruptions.

Findings in Geroliminis and Daganzo (2008) have proven the existence of Macroscopic Fundamental Diagram (MFD) with empirical data, which is a relationship between average flow and density on a regional level. Based on MFD and the indicated critical density, multiple control methods targeting a better overall performance of the road network have been proposed since then, e.g., pricing (Genser and Kouvelas, 2022) and perimeter control (Keyvan-Ekbatani et al., 2012). Perimeter control refrains the inflowing vehicles from adjacent regions into a protected region, the traffic density remains below the critical density and thus increasing the trip completion rate. Geroliminis et al., (2013) proposed a perimeter control based on Model Predictive Control (MPC) and proved its superior effectiveness and robustness regarding trip completion in comparison with a greedy controller.

Nowadays, with the abundance of data collected through various electronic approaches, an emerging amount of transport research has been shifting towards data-driven approaches, including the application of deep Reinforcement Learning (RL) in traffic light control, delay management, etc. For perimeter control, there are also some recent works associated with RL, for example, Zhou and Gayah (2021), show that RL can achieve satisfactory performance. Through interacting with a given environment over time, the RL algorithm learns to maximize the cumulative reward. The benefit of RL lies in its capability of dealing with multivariate nonlinearities.

* Corresponding author (linghang.sun@ivt.baug.ethz.ch)

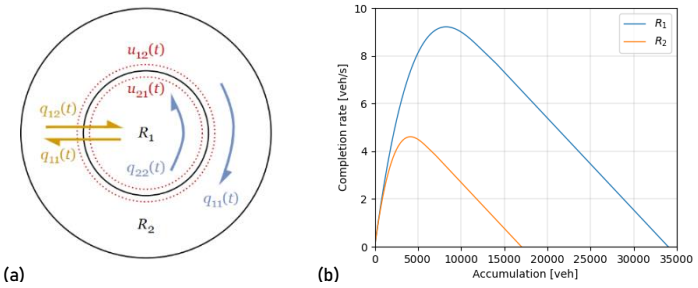
The concept of antifragility, first being introduced in the book *Antifragile: Things that gain from disorder* (Taleb, 2014), provides insights on the definition of antifragile and systems that have logarithmic response to unexpected disturbances. Since then, antifragility has become a popular concept in many disciplines, such as economy and medicine (Axenie et al., 2022). Using antifragility as the concept to optimize and control a transport system is still a novel idea. The basic principles of being antifragile in Taleb (2012) can be summarized as follows: to better withstand black swan events, one should focus on the trend of how things develop and maintain buffers for unexpected events.

In this work, we simulate a cordon shape urban network scenario with periodic disturbances, and based on the concept of antifragility, rather than using critical density or critical trip completion rate as static indicators of whether a perimeter control strategy should be applied, we use a deep-RL algorithm with additional state and reward terms based on the first and second derivatives so that the algorithm can learn from past disturbances and exhibit the property of being less fragile.

Problem formulation and Methodology

A cordon shaped urban network is studied with the inner region representing a city center, as shown in Figure 1 (a). $q_{ij}(t)$ represents the new traffic demand from origin i to destination j at time step t. Based on the number of vehicles $n_{ij}(t)$, the trip completion $M_i(t)$ can be determined based on MFD, as illustrated in Figure 1 (b). The perimeter controller variable $0 \leq u_{ij}(t) \leq 1$ regulates the percentage of transfer flow between the 2 regions. The simulation time is set to 2 hours, with the first hour allowing for the inflow of vehicles and the second hour for clearing the vehicles from the network.

Figure 1. (a). A cordon-shape urban network: (b). MFD for the inner and outer regions.



Source: own elaborations

For the details of the dynamic constraints of the model, please refer to Geroliminis *et al.* (2013). The objective function J in this paper can be mathematically formulated as:

$$J = \max_{u_{12}(t),u_{21}(t)} \int_{t_0}^{t_1} \left[\sum_{i=1,2} M_{ii}(t) + \epsilon(t) \right] dt \tag{1}$$

where $M_{ii}(t)$ is the number of vehicles completed their trips in a specific region i while $M_{ij}(t)$ $(i \neq j)$ is the interregional transfer flow from region i to j. The term $\epsilon(t)$ acts as an additional term based on the concept of the first and second order derivatives specifically used in the RL-based algorithms.

$$\epsilon(t) = \omega_k \sum_{i=1,2} \alpha_i(t) \cdot k_i(t) + \omega_{\Delta k} \sum_{i=1,2} \Delta k_i(t) \tag{2}$$

$$\alpha_i(t) = \{1,\ if\ n_i(t) \geq n_i(t-1),\ -1,\ if\ n_i(t) < n_i(t-1)\} \tag{3}$$

$$k_i(t) = (M_i(t) - M_i(t-1))/(n_i(t) - n_i(t-1)) \tag{4}$$

$$\Delta k_i(t) = k_i(t) - k_i(t-1) \tag{5}$$

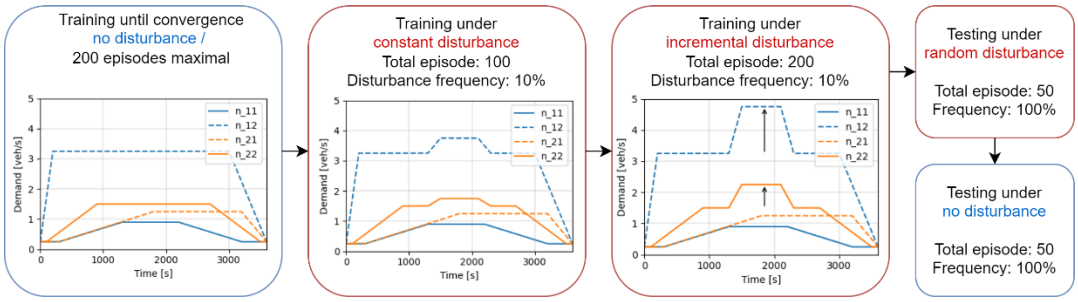
The binary variable $\alpha_i(t)$ is a used to reward the agent when it's moving towards the desired direction on the MFD. Intuitively, $\omega_k \sum_{i=1,2} \alpha_i(t) \cdot k_i(t)$ rewards the agent more when it's moving towards the critical density to maximize its trip completion rate, while $\omega_{\Delta k} \sum_{i=1,2} \Delta k_i(t)$ gives a penalty when the traffic states quickly moves towards the critical density.

Several control strategies have been evaluated, including MPC, a RL algorithm with $n_{ij}(t)$ as state, a RL algorithm based on Zhou and Gayah (2021) $[n_{ij}(t), q_{ij}(t)]$, a RL with $[n_{ij}(t), dn_{ij}(t), dn_{ij}^2(t)]$ and a RL with $[n_{ij}(t), dn_{ij}(t)]$ as state as well as the additional reward term $\epsilon(t)$.

The RL algorithm applied in this work is a Deep Determinist Policy Gradient (DDPG) algorithm (Lillicrap et al., 2019). As opposed to the Deep Q-Network algorithm, which can only be applied in an environment with a discrete action space, DDPG has the ability to manage a continuous action space, i.e., allowing the perimeter control variables to be continuous instead of only choosing from a limited set of discrete values. The DDPG algorithm can be divided into two main components, namely the actor and the critic, which are updated at each step through policy gradient and Q-value, respectively.

Early experiments showed that the algorithm could not effectively learn from disturbances if they are fully randomized, so curriculum learning is introduced, which is a common technique in deep learning and in RL (Narvekar et al., 2020). By gradually increasing the magnitude of disturbances, as shown in Figure 2, the algorithm can learn with steadier progress, and the comparison between each RL algorithm is more self-explanatory.

Figure 2. Scheme of curriculum learning

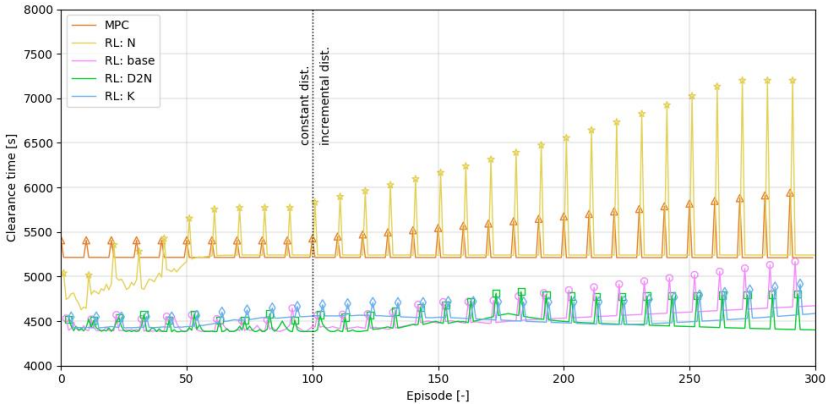


Source: own elaborations

Results

Figure 3 summarizes the performance of each method in terms of clearance time with a threshold of 80% of the vehicles. Only the process with constant and incremental disturbances is shown to provide the most essential information. Generally speaking, the performance of these methods can be classified into two groups. MPC and RL with $n_{ij}(t)$ as state perform relatively inferior, with their clearance time always higher than the other three methods. To illustrate the change in performance during the 300 episodes, we summarize the clearance time under normal situations (270 episodes) and under disturbance (30 episodes) in Figure 4.

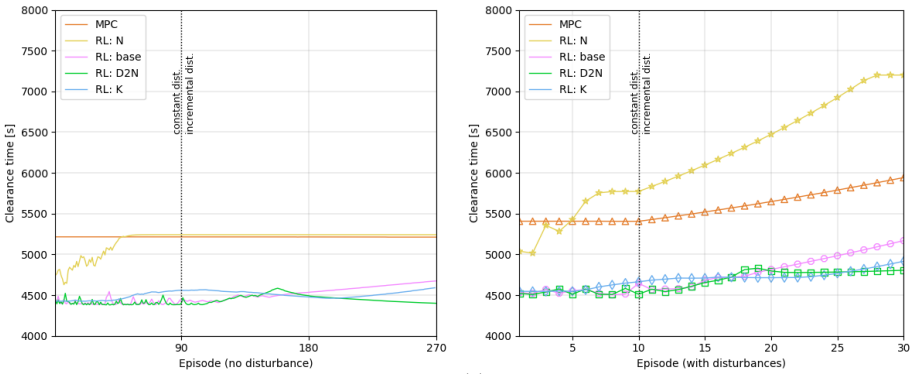
Figure 3. Clearance time over simulation episodes



Source: own elaborations

In Figure 4 (a), it is evident that with periodic disturbances, RL-based algorithms tend to perform worse over simulation episodes even under no disturbance. This is because the network parameter learned after disturbances would favor the algorithm to better deal with such occasions at the cost of reducing performance in normal situations. For instance, the RL algorithm with only $n_{ij}(t)$ as state is strongly undermined, whereas the algorithms with additional state and reward terms are less affected and perform even better compared to MPC. The RL algorithm with $[n_{ij}(t), dn_{ij}(t), dn_{ij}^2(t)]$ achieves the most stable performance and outperforms the baseline method.

Figure 4. (a). Clearance time under normal scenario; (b). Clearance time under disturbances



(a) (b)

Source: own elaborations

Figure 4(b) illustrates the clearance time under constant and incremental disturbances, which indicates the ability of the algorithm to demonstrate antifragility and learn from disturbances. MPC, RL with $n_{ij}(t)$ and baseline RL with $[n_{ij}(t), q_{ij}(t)]$ exhibit a certain degree of nonlinear response to increasing disturbances, demonstrating fragility. While the responses from the other two RL algorithms are harder to determine due to learning curve oscillations, they are clearly less fragile than the previous three methods. However, based on the clearance time with constant disturbances, there is no clear pattern showing that the RL-based algorithm can learn from multiple interactions with a constant periodic disturbance.

Conclusion and Discussion

In this work, we propose and validate the concept of antifragility in the field of urban road networks. Drawing on the antifragile concept of first and second order derivatives, by adding additional terms in the state and reward of an RL algorithm, we demonstrated that, our proposed RL algorithms with $[n_{ij}(t), dn_{ij}(t), dn_{ij}^2(t)]$ as state or additional term as reward are less fragile compared to the other methods. Although it may come at a cost of worsening performance under no disturbance scenarios, it can still perform with relatively high stability and outperform the baseline RL-based methods.

References

Axenie, C., D. Kurz, and M. Saveriano, 'Antifragile Control Systems: The Case of an Anti-Symmetric Network Model of the Tumor-Immune-Drug Interactions', *Symmetry*, Vol. 14, No. 10, October 2022, p. 2034.

Genser, A., and A. Kouvelas, 'Dynamic Optimal Congestion Pricing in Multi-Region Urban Networks by Application of a Multi-Layer-Neural Network', *Transportation Research Part C: Emerging Technologies*, Vol. 134, January 1, 2022, p. 103485.

Geroliminis, N., and C.F. Daganzo, 'Existence of Urban-Scale Macroscopic Fundamental Diagrams: Some Experimental Findings', *Transportation Research Part B: Methodological*, Vol. 42, No. 9, November 1, 2008, pp. 759–770.

Geroliminis, N., J. Haddad, and M. Ramezani, 'Optimal Perimeter Control for Two Urban Regions With Macroscopic Fundamental Diagrams: A Model Predictive Approach', *IEEE Transactions on Intelligent Transportation Systems*, Vol. 14, No. 1, March 2013, pp. 348–359.

Keyvan-Ekbatani, M., A. Kouvelas, I. Papamichail, and M. Papageorgiou, 'Exploiting the Fundamental Diagram of Urban Networks for Feedback-Based Gating', *Transportation Research Part B: Methodological*, Vol. 46, No. 10, December 1, 2012, pp. 1393–1403.

Lillicrap, T.P., J.J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, 'Continuous Control with Deep Reinforcement Learning', arXiv, July 5, 2019.

Narvekar, S., B. Peng, M. Leonetti, J. Sinapov, M.E. Taylor, and P. Stone, 'Curriculum Learning for Reinforcement Learning Domains: A Framework and Survey', *The Journal of Machine Learning Research*, Vol. 21, No. 1, January 1, 2020, p. 181:7382-181:7431.

Taleb, N.N.N., *Antifragile: Things That Gain from Disorder*, Reprint edition., Random House Publishing Group, New York, 2014.