International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT206458

A Survey on Deep Reinforcement Learning Network for Traffic Light Cycle Control

V. Indhumathi, Dr. K. Kumar M.E., Ph.D.
Department of Computer Science and Engineering, Government College of Technology,Coimbatore, Tamilnadu, India

Article Info
Volume 6, Issue 4
Page Number: 285-293
Publication Issue :
July-August-2020

Article History
Accepted : 25 July 2020
Published : 30 July 2020

ABSTRACT

A Traffic signal control is a challenging problem and to minimize the travel time of vehicles by coordinating their movements at the road intersections. In recent years traffic signal control systems have on over simplified information and rule-based methods and we have large amounts of data, more computing power and advanced methods to drive the development of intelligent transportation. An intelligent transport system to use the machine learning methods likes reinforcement learning and to explain the acknowledged transportation approaches and a list of recent literature in traffic signal control. In this survey can foster interdisciplinary research on this important topic.

Keywords : Traffic signal, Deep reinforcement learning, Cycle control.

I. INTRODUCTION

The intersection management for busy or major roads is primarily done through the traffic lights, whose inefficient control may cause numerous problems, such as long delay of travelers and huge waste of energy. Even worse, it may also incur vehicular accidents [1], [2]. The Existing traffic light control neither deploys fixed programs without considering the real-time traffic or considering the traffic to a very limited degree [3]. The fixed programs set traffic signals equal time duration in every cycle, or different time duration based on historical information. Some control programs take inputs from various sensors such as underground inductive loop detectors for detecting the existence of vehicles in front of traffic lights. However, the inputs are processed in a very coarse way to determine the duration of green/red lights. In some cases, existing traffic light control systems work, through only at a low efficiency. However, in many other cases, such as a football event or a more common high traffic hour scenario, the traffic light control systems become paralyzed. Instead, we often witness an experienced policeman directly manages the intersection by waving signals. In high traffic scenarios, a human operator observes the real time traffic condition in the intersecting roads and smartly determines the duration of the allowed passing time for each direction using his/her long-term experience and understanding about the intersection, which is very effective. This observation motivates us to propose a smart intersection traffic light management system which can take real-time traffic condition as input and learn how to manage the intersection just like the human operator. To implement such a system, we need 'eyes' to watch the real-time road condition and 'a brain' to process it. For the former, recent advances in sensor and networking technology enables taking real-time traffic information as input, such as the number of vehicles, the locations of vehicles, and their waiting time [4]. For the 'brain' part, reinforcement learning, as a type of machine learning techniques, is a promising way to solve the problem. A reinforcement learning system's goal is to make an action agent learn the optimal policy through interacting with the environment to maximize the reward, e.g., the minimum waiting time in our intersection control scenario. It usually contains three components: states of the environment, action space of the agent, and reward from every action [5]. A well-known application of reinforcement learning is AlphaGo [6], followed by AlphaGo Zero [7]. AlphaGo, acting as the action agent in a Go game (environment), first observes the current image of the chessboard (state), and takes the image as the input of a reinforcement learning model to determine where to place the optimal next playing piece 'stone' (action). Its final reward is to win the game or to lose. Thus, the reward may not be obvious during the playing process but becomes clear when the game is over. When applying reinforcement learning to the traffic light control problem, the key point is to define the three components at an intersection and quantify them to be computable. Some previous works propose to dynamically control the traffic lights using reinforcement learning. Some define the states by the number of waiting vehicles or the waiting queue length [4], [8]. But real traffic situation cannot be accurately captured by only the number of waiting vehicles or queue length [9]. With the popularization of vehicular networks and sensor networks, more accurate on-road traffic information can be extracted, such as vehicles' speed and waiting time [10]. However, rich information causes the number of states to increase dramatically. When the number of states increases, the complexity in a traditional reinforcement learning system grows exponentially. With the rapid development of deep learning [11], deep neural networks have been employed to deal with the large number of states, which constitutes a deep reinforcement learning model [12]. A few recent studies have proposed to apply deep reinforcement learning in the traffic light control problem [13], [14]. But there are two main limitations in existing studies: (1) the traffic signals are usually split into fixed-time intervals, and the duration of green/red lights can only be a multiple of this fixed-length interval, which is not efficient in many situations; (2) the traffic signals are designed to change in a random sequence, which is not a safe or comfortable way for drivers. In this paper, we study the problem on how to control the traffic light signal duration in a cycle based on the extracted information from vehicular networks or sensor networks. The general idea is to mimic experienced operator to control the signal duration in every cycle based upon information gathered from vehicular networks. To implement such an idea, the operation of the experienced operator is modeled as a Markov Decision Process (MDP). The MDP is high-dimension model, which consists of time duration for every phase. The system learns the control strategy based on the MDP by trial and error in a deep reinforcement learning model. To fit a deep reinforcement learning model, we divide the whole intersection into grids and build a matrix, each element of which is the vehicles' information in the corresponding grid collected by vehicular networks or extracted from cameras via image processing. The matrix is defined as the states and the reward is the cumulative waiting time difference between two cycles. In our model, a convolutional neural network is employed to match the states and expected future rewards. Note that, every traffic light's action produced from our model affects the environment.

When the traffic flow changes dynamically, the environment becomes unpredictable. To solve this problem, we employ a series of state of-the-art techniques in our model to improve the performance, including dueling network [15], target network [12], double Q-Learning network [16], and prioritized experience replay [17]. Our contribution of the paper includes 1) We are the first to combine dueling network, target network, double Q network and prioritized experience replay into one framework to solve the traffic light control problem, which can be easily applied into other problems. 2) We propose a control system to decide the phases' time duration in a whole cycle instead of dividing the time into segments. 3) Extensive experiments on a traffic micro-simulator, Simulation of Urban Mobility (SUMO) [18], show the effectiveness and high-efficiency of our model. The reminder of this paper is organized as follows. The literature review is presented in Section II. The model and problem statement are introduced in Section IV. The background on reinforcement learning is introduced in Section III. Section V details our reinforcement learning model in the traffic light control system. Section VI extends the reinforcement learning model into a deep learning model to handle the complex states in our system. The model is evaluated in Section VII. Finally, paper is concluded in Section VIII.

II. LITERATURE REVIEW

S. NO	**LITERATURE REFERRED**	**TECHNIQUE**	**LIMITATIONS**
1.	P. Balaji, X. German, and D. Srinivasan, “Urban traffic signal control using reinforcement learning agents,” IET Intel. Transp. Syst., vol. 4, no. 3, pp. 177–188, Sep. 2010.	1.Online Q-learning 2.Reinforcement learning	1.delay 2.Time.
2.	S. Chiu and S. Chand, 1993, “Adaptive- traffic-signal-control-using-fuzzy-logic,” in Proc. 1st IEEE Regional Conf. Aerosp. Control Syst, pp. 1371–1376.	1.Fuzzy logic	1.Time 2.Long delay.
3.	L. Li, Y. Lv, and F.-Y. Wang, “Traffic-signal-timing-via-deep -reinforcement-learning,” IEEE/CAA	1.Dueling network 2.Deep Q network	1.Accidents 2.Long delay 3. Wastage of energy.
4	3, no. 3, pp. 247–254, Jul. 2016 W. Genders and S. Razali, “Using a deep reinforcement learning agent for traffic signal control,” unpublished paper, 2016. [Online]. Available: https://arxiv.org/abs/1611.01142	1.Reinforcement learning	1. long delay 2. Wastage of energy 3.Accidents
5	I. Arel, C. Liu, T. Urbanik, and A. Kohls, “Reinforcement	1.**Deep q learning.**	1.Average delay 2.Congest

	learning-based multi-agent system for network traffic signal control," IET Intell. Transp. Syst., vol. 4, no. 2, pp. 128–135, Jun. 2010.		ion 3.Likelihood of intersection cross blocking.
6	L. Zhu, Y. He, F. R. Yu, B. Ning, T. Tang, and N. Zhao, "Communication-based-train-control-system-performance-optimization-using Deep-Reinforcement-Learning," IEEE Trans. Veh. Technol., vol. 66, no. 12, pp. 10705–10717, Dec. 2017	1.Deep Q network 2.Deep Reinforcement learning	1.Long communication delay 2.Time

III. REINFORCEMENT LEARNING

Reinforcement Learning (RL) is a type of algorithms in machine learning. It interacts with the environment to learn better actions to maximize the objective reward function in the long run through trial and error. In reinforcement learning, an agent, the action executor, takes an action and the environment returns a numerical reward based on the action and the current state. A four-tuple S, A, R, T can be used to define the reinforcement learning model:

S: the possible state space.

A: the possible action space.

R: the reward space.

T: the transition function space among all states, which represents the probability of the transition from one state to another.

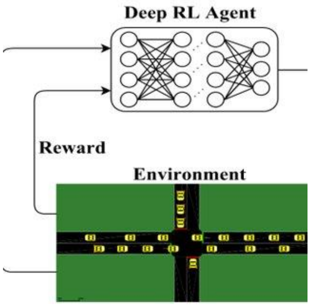

Fig. 1: Deep reinforcement learning agent of traffic signal control [21].

Action Set is used to control traffic signal phases, it defines a set of possible actions A = {North/South Green (NSG), East/West Green (EWG)}. NSG allows the vehicles to pass from North to South and vice versa, and also indicates the vehicles on East/West route should stop and not proceed through the intersection.

Reward Function typically an immediate reward rt ∈ R is a scalar value which the agent receives after taking the chosen action in the environment at each time step. We set the reward as the difference between the total cumulative delays of two consecutive actions.

IV. NETWORK-WIDE TRAFFIC CONTROL WITH VEHICULAR COMMUNICATIONS

A. Network-Wide Traffic Control with Real-Time information on vehicle

The isolated intersection control, highway ramping control and urban road network control also require accurate vehicle position information.

For example, different algorithms were proposed in the last few years to estimate the vehicle queue lengths at metered onramps [09], [10] and the queue lengths and the travel times for congested signalized arterials [11]–[15] and for a road network [16], [17]. All these studies used certain a priori knowledge of traffic flow dynamics to infer/predict the required traffic flow parameters (flow rate, occupancy, speed, etc.) at the locations with no measurements. The parameters are flow rate, occupancy, speed, etc. However, if the position and movement information of all vehicles can be achieved via vehicular communications, such difficulties will be solved neatly. changes indeed reflect a transition of design philosophies for traffic control systems. As pointed out in many literatures [08]–[10], most existing traffic control systems conform to the concept of feedback control, because they specify the control rules in response to the current values of state variables. In many recent approaches, researchers have begun to integrate both feedback and feedforward characters to build traffic control systems. When traffic demands can be measured or effectively predicted before they enter the current system, we can take a pre-emptive action to optimize the traffic efficiency.

In such systems, we can formulate a new optimization problem (6), with control u(k) determined upon future states x(k + i) and demands d(k + i), where i may equal to (1, 2, ...). Although the dimensions of variables are much larger than those that had been considered for isolated intersections, their instinct natures the same. Different preferences on choosing control u(k) will be discussed in Section IV.

B. Network-Wide Cooperative Driving

Suppose we divide the studied road network into several nonoverlapped segments (nodes) and define a graph to model the connection properties of these segments. Further assume that each vehicle has a specified route from its origin to its destination and will pass a few segments sequentially, the desired trajectory for any vehicle can be then roughly sketched as a series of time slices when the vehicle enters the selected segment. The control design problem becomes finding a set of trajectories that allow vehicles reach their destination nodes in the shortest time. It is apparent that such a discrete-time graph-scheduling problem is much more complex than the simple tree scheduling problem formulated for isolated intersections. Even if we omit the detailed driving plans of any vehicle in the segments, the solution space will expand quickly with the increasing number of intersections. Currently, knowledge on the feasibility and benefit of the network-wide cooperative driving is very meagre. To the authors' knowledge, only [13] had proposed a greedy search strategy.

V. DIFFERENT PREFERENCES

A. Model-Based Versus Simulation-Based Predictive Controls

How to utilize the rich information collected via vehicular communications is a key problem in future traffic control system designs. One representative approach in this direction is based on the model predictive control (MPC) theory [14]– [17]. In such approaches, the dynamics of traffic flows at different locations (nodes) are abstractly described by a set of difference equations, such as (3) previously. When the current states of the nodes are known or at least partly known, we can foretell the future states of traffic flows by recursively solving this set of difference equations with desired control actions. Searching the solution space for control actions, we finally adopt the control actions that will lead to the best future states of traffic flows.

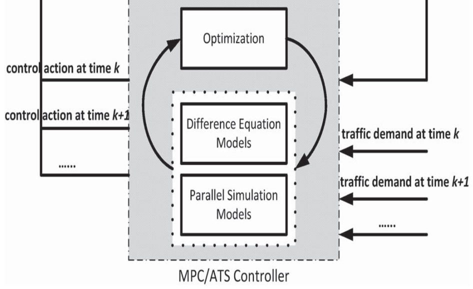
Traffic Systems
traffic measurement at time k
Optimization
control action at time k
control action at time k+1
......
Difference Equation Models
Parallel Simulation Models
traffic demand at time k
traffic demand at time k+1
......
MPC/ATS Controller

Fig. 4. Schematic view of MPC and ATS control [20]

]

Notice that the difference equations may not be able to accurately characterize the time-varying stochastic traffic flow dynamics; researchers now show increasing interests in the parallel simulation of traffic systems. In such approaches, the so-called artificial transportation systems (ATSs) [08]–[12] were built to model and analyses traffic flow dynamics. Through the parallel interactions of an actual transportation system and its corresponding ATS, we can evaluate the effectiveness of different traffic strategies under various conditions. Both MPC and ATS control use online optimization to design control actions. Their difference is that MPC uses an explicit prediction model, whereas ATS control uses an implicit prediction model. Usually, both MPC and ATS control simultaneously schedule the control inputs u (0), u (1), . . ., u(K) for a relatively long-time horizon to find a global optimal solution for J. In addition, both of them allow modifying inappropriately scheduled control inputs in the following time intervals. Compared with difference-equation-based MPC, parallel simulation- based ATS control provides more flexible and living ontology to represent and organize knowledge of transportation systems (see Fig. 4). This enables us to choose even better control strategies by using ATSs. However, the computation costs of ATS methods are much higher, too [10]. Obviously, intervehicle communication can serve as a key component in all these new traffic control systems, since we can capture the variation of traffic demands in advance. However, the best way to fuse the predicted/simulated traffic states conveniently and promptly with the sampled states still need further discussions. We are expecting the shift of research interests into this promising area in the near future.

B. Planning-Based Versus Self-Organization-Based Controls

Whether to apply global planning-based control or local self-organization- based control is another interesting and important problem. Generally, global planning-based approaches refer to control city-wide road networks that may contain tens of intersections or on/off-ramps via long-term scheduled control actions [13], [14], whereas local self-organization approaches refer to build short-term changeable control actions [15], [16]. Intuitively, global approaches seem better, because more information will be used to obtain an even better nongreedy solution. However, the temporal–spatial size of an independent traffic control system is restricted by many factors in practice. The first constraint lies in the performance limit of vehicular communications. The packet drops rates, end-to-end packet delays, and network throughputs all influence the amount of information that can be correctly delivered in time and thus limit the temporal–spatial size of traffic control systems [3]–[7].

The second constraint comes from the possible vulnerability of control systems. It was argued in [15], and [16] that many man-made systems become unstable and create uncontrollable consequences, as the complexity and interaction strengths in a networked subsystem increase, even when decisions are well planned. Noticing that all the measurements may be distorted or inaccurate due to various reasons (e.g., transmission errors in wireless communications), applying local self-organization-based control is believed by many researchers as a better choice.

C. Big-Data-Based Versus Concise-Data-Based Controls

Similar to studies in many other fields, there are two diversions in employing the amount of traffic information to design traffic control systems. One is to use the rich information, as what had been discussed earlier [15]. The other is to use concise information that is necessary.

A representative example of the second kind of approach is the traffic control system based on urbanscale macroscopic fundamental diagrams (MFDs) [15]–[18]. Since modelling the traffic flow dynamics of each link and intersection in a large urban network is a complex task, such approaches aim to capture the primary characteristics between network-wide vehicle densities and network-wide space-mean flow rates. That is, we consider the collective behaviours of a lot of vehicles rather than the movements of individual vehicles. Then, parsimonious control rules can be designed for the whole road network, based on the measurements collected at sparsely located sensors.

A typical example of MFD-based control is perimeter control for a city-wide network with complicated structures. Here, perimeter control means the access metering to maintain the mobility of cars at a stabilized level. The detailed traffic dynamics in the studied region are not studied. Instead, we describe the average degree of congestion for the region by means of average vehicle densities and space-mean flow rates estimated by a few fixed detectors and floating vehicle probes. To prevent overcrowding, traffic flow toward a congested region is restricted, whereas traffic flow toward an underutilized area is facilitated. Although we do not know the evolution details of traffic flow at every part of a region, the overall traffic is under control. Differently, the aforementioned vehicle-coordination-based approach will track every vehicle in this region, analyses their traveling plans, and set up the signal timing plan for each intersection within this region to make the overall traffic smoother. MFD-based approaches have many merits, such as a simpler design algorithm that is relatively robust to traffic demand disturbances and much lower implementation costs. However, the drawbacks of MFD-based approaches, including inaccurate estimation of performance indexes (e.g., queue length at every intersection), are apparent, too.

V. CONCLUSION

Due to the ever-increasing need for more efficient transport, vehicle-to-vehicle communications are introduced into traffic control systems to better coordinate vehicles and traffic signals nowadays. This change promotes new research frontiers to be further explored. Constrained by the length limit, we just focus on a few questions on the advance of control systems in this paper.

First, the performance limits of vehicle coordination are left untouched in this survey. It was estimated in [21] that the benefit-to-cost ratio of retiming conventional traffic signal systems was typically 40: 1. We believe the potential benefit of the intelligent vehicle coordination might be even higher. However, this technology cannot dramatically eliminate traffic congestion when all the roads are crowded. The estimation of performance limits needs further investigations. Second, the achievements of any intelligent traffic control system previously mentioned are rooted in a successful integration of lots of sensors, controllers, operations software, and hardware [17], [18]. The failure of any component in this integrated system will result in performance degradation [19], [20] or even severe traffic accidents. How to identify failure (maybe at individual vehicle level) in time and tolerate faults of some components (maybe at regional control system level) also needs to be carefully studied.

VI. REFERENCES

[1]. T. L. Willke, P. Tientrakool, and N. F. Maxemchuk, “A survey of intervehicle communication protocols and their applications,” IEEE Commun. Surveys Tuts., vol. 11, no. 2, pp. 3–20, 2nd Quart., 2009.

[2]. F. Qu and F.-Y. Wang, “Intelligent transportation spaces: Vehicles, traffic,

communications, and beyond," IEEE Commun. Mag., vol. 48, no. 11, pp. 136–142, Nov. 2010.

[3]. H. Hartenstein and K. P. Laberteaux, "A tutorial survey on vehicular ad hoc networks," IEEE Commun. Mag., vol. 46, no. 6, pp. 164–171, Jun. 2008.

[4]. T. Sukuvaara and P. Nurmi, "Wireless traffic service platform for combined vehicle-to-vehicle and vehicle-to-infrastructure communications," IEEE Wireless Commun., vol. 16, no. 6, pp. 54–61, Dec. 2009.

[5]. G. Korkmaz, E. Ekici, and F. Özgüner, "Supporting real-time traffic in multihop vehicle-to-infrastructure networks," Transp. Res. C, Emerging Technol., vol. 18, no. 3, pp. 376–392, Jun. 2010.

[6]. Y. Bi, L. X. Cai, X. Shen, and H. Zhao, "Efficient and reliable broadcast in inter-vehicle communications networks: A cross layer approach," IEEE Trans. Veh. Technol., vol. 59, no. 5, pp. 2404–2417, Jun. 2010.

[7]. T. H. Luan, X. Ling, and X. Shen, "Provisioning QoS controlled media access in vehicular to infrastructure communications," Ad Hoc Netw., vol. 10, no. 2, pp. 231–242, Mar. 2012.

[8]. R. Horowitz and P. Varaiya, "Control design of an automated highway system," Proc. IEEE, vol. 88, no. 7, pp. 913–925, Jul. 2000.

[9]. J. A. Misener and S. E. Shladover, "PATH investigations in vehicleroadside cooperation and safety: A foundation for safety and vehicleinfrastructure integration research," in Proc. IEEE Conf. Intell. Transp. Syst., 2006, pp. 9–16.

[10].R. Rajamani, H. Tan, B. Law, and W. Zhang, "Demonstration of integrated longitudinal and lateral control for the operation of automated vehicles in platoons," IEEE Trans. Control Syst. Technol., vol. 8, no. 4, pp. 695–708, Jul. 2000.

[11].V. Milanés, J. Godoy, J. Villagrá, and J. Pérez, "Automated on-ramp merging system for congested traffic situations," IEEE Trans. Intell. Transp. Syst., vol. 12, no. 2, pp. 500–508, Jun. 2011.

[12].J. Little, "The synchronization of traffic signals by mixed-integer linear programming," Oper. Res., vol. 14, no. 4, pp. 568–594, Jul./Aug. 1966.

[13].M. Papageorgiou, C. Diakaki, V. Dinopoulou, A. Kotsialos, and Y. Wang, "Review of road traffic control strategies," Proc. IEEE, vol. 91, no. 12, pp. 2043–2067, Dec. 2003.

[14].A. G. Sims and K. W. Dobinson, "The Sydney coordinated adaptive traffic (SCAT) system—Philosophy and benefits," IEEE Trans. Veh. Technol., vol. VT-29, no. 2, pp. 130–137, May 1980.

[15].P. B. Hunt, D. I. Robertson, R. D. Bretherton, and M. C. Royle, "The SCOOT on-line traffic signal optimisation technique," Traffic Eng. Control, vol. 23, no. 4, pp. 190–199, Apr. 1982.

[16].P. Mirchandani and F.-Y. Wang, "RHODES to intelligent transportation systems," IEEE Intell. Syst., vol. 20, no. 1, pp. 10–15, Jan./Feb. 2005.

[17].A. Gaur and P. Mirchandani, "Method for real-time recognition of vehicle platoons," Transp. Res. Rec., no. 1748, pp. 8–17, 2002.

[18].Y. Jiang, S. Li, and D. E. Shamo, "A platoon-based traffic signal timing algorithm for major–minor intersection types," Transp. Res. B, Methodol., vol. 40, no. 7, pp. 543–562, Aug. 2006.

[19].X.-F. Xie, G. J. Barlow, S. F. Smith, and Z. B. Rubinstein, "Platoon-based self-scheduling for real-time traffic signal control," in Proc. IEEE Conf. Intell. Transp. Syst., 2011, pp. 879–884.

[20].Li, Li, Ding Wen, and Danya Yao. "A survey of traffic control with vehicular communications." IEEE Transactions on Intelligent Transportation Systems 15, no. 1 (2013): 425-432.

[21].Wei, H., Zheng, G., Gayah, V. and Li, Z., 2019. A survey on traffic signal control methods. arXiv preprint arXiv:1904.08117.

Cite this article as :

V. Indhumathi, Dr. K. Kumar , "A Survey on Deep Reinforcement Learning Network for Traffic Light Cycle Control", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 6 Issue 4, pp. 285-293, July-August 2020. Available at doi : https://doi.org/10.32628/CSEIT206458
Journal URL : http://ijsrcseit.com/CSEIT206458