MACHINE LEARNING

(Essay)

Submitted By: Ammar Kafeel

Date of Submission: 25/04/23

Contents

Machine learning application:

Machine learning has been used to a number of fields, such as natural language processing, picture and audio recognition, and more. Contrarily, one of the most fascinating and rapidly developing areas of machine learning is reinforcement learning. Reinforcement learning, or RL (TechTarget, n.d.), is a type of machine learning that is based on the idea of learning by feedback. It is the process of training a model to make judgements through trial and error. The model gets feedback in the form of rewards or penalties, which helps it learn from its errors and get better over time, particularly through trial and error. Robots, video games, and self-driving automobiles are just a few of the real-world uses for RL. I decide to use the reinforcement learning-based multi-agent system for network traffic signal control that is a machine learning application.

Real-world case study:

The paper by (Arel, Itamar & Liu, C. & Urbanik, T. & Kohls, Airton, 2010) presents a real-world case study of using reinforcement learning in a multi-agent system for traffic signal control. The study shows that the reinforcement learning-based system outperforms traditional traffic signal control methods in terms of reducing travel time and delay for vehicles in the network.

Summary of Findings:

Background of the machine learning application:

A particular type of machine learning called reinforcement learning (RL) focuses on how an agent interacts with its environment in order to choose a course of action that maximises a cumulative reward signal. RL has been used in a variety of applications, including robots, gaming, and traffic control. Traditional techniques of traffic control utilise pre-programmed timing schedules or defined cycle durations. These solutions, however, do not take into account real-time traffic circumstances, which can result in traffic congestion, greater fuel usage, and longer travel times. In tackling these difficulties, RL-based traffic control systems have showed promise.

In this research, the authors propose a system consisting of multiple agents for traffic light control based on reinforcement learning. "Reinforcement learning-based multi-agent system for network traffic signal control" (Arel, Itamar & Liu, C. & Urbanik, T. & Kohls, Airton, 2010). The system is made up of many RL agents, each of which controls a single traffic signal. As inputs, the RL agents use observations of the present traffic situation, such as the number of cars at the junction and the time since the previous green light. The agents then utilise RL algorithms to calculate the best moment to turn the traffic signal from red to green in order to maximise the cumulative reward, which is the total number of cars passing through the intersection.

Typically, RL-based traffic management technologies have the ability to improve safety, reduce congestion, and boost traffic flow. However, there are several obstacles and ethical problems to consider, including as safety concerns and the impact on vulnerable road users. Identifying and addressing these issues will be critical in establishing successful and ethical RL-based traffic control systems.

Reinforcement Learning in traffic signal control:

Each traffic light in the (Arel, Itamar & Liu, C. & Urbanik, T. & Kohls, Airton, 2010) RL-based traffic signal management system is controlled by an RL agent that interacts with its environment, i.e., the traffic junction, to develop an optimal policy. As input, the RL agent receives observations of the current traffic situation, such as the number of cars parked at the junction, the time since the previous green light, and the current phase of the traffic signal. The RL agent calculates the optimal time to turn the traffic light from red to green in order to maximize the cumulative reward, or the total number of vehicles passing through the junction, based on these observations.
The suggested method makes use of the model-free RL algorithm known as Q-learning, which continuously updates the Q-values based on observable state-action-reward transitions to learn an ideal action-value function. A state-action pair's Q-value indicates the predicted cumulative reward that may be received by executing that action in that state and then following the optimum policy.

The authors trained and tested their RL-based traffic control system using VISSIM, an incredibly straightforward traffic simulator. They found that the RL-based system outperformed the prior system in terms of average travel duration, delay, and fuel consumption when performance was compared to that of a standard fixed-time control system.

To determine the appropriate policy for each traffic light at a junction, this application employs RL by training a large number of RL agents. The RL agents learn from their interactions with the environment in order to maximize the overall reward, which is the number of cars passing through the junction. The application of RL in traffic signal management has showed promise in terms of improving traffic flow, decreasing congestion, and increasing safety.

Ethical Issues and Challenges:

Although traffic control systems based on reinforcement learning have the potential to improve traffic flow and reduce congestion, there are a number of challenges and ethical questions to take into account.

The safety of the system is one of the primary problems of employing RL in traffic control. Because RL agents learn by trial and error, they may perform suboptimal decisions that result in dangerous circumstances. For example, an RL agent may learn to prioritise throughput over safety, increasing the likelihood of an accident. As a result, it is critical to create RL algorithms that can balance safety and efficiency.

Scalability of RL-based traffic control systems is another issue. As the number of junctions and traffic signals grows, so does the number of RL agents necessary to operate the system. This can result in computational and memory limits, making RL-based traffic management systems challenging to deploy in large-scale metropolitan settings.

Additionally, RL-based traffic control systems may create ethical concerns about justice and equity. An RL-based system, for example, may learn to prioritise the demands of specific types of vehicles, such as cars, above others, such as bicycles or pedestrians. This might lead to more traffic and longer travel times for vulnerable road users.

Further research is required to design RL algorithms that can balance safety and efficiency, optimise performance in large-scale systems, and assure justice and equity in order to meet these obstacles and ethical dilemmas. Furthermore, stakeholders such as transportation planners, legislators, and community members must be involved in the development and implementation of RL-based traffic control systems to ensure that their concerns are addressed and the system benefits everyone.

Opinion on how to enhance the existing procedures, goods, approaches:

There are various areas in which the existing RL-based traffic control system may be modified to improve performance and handle the aforementioned problems and ethical considerations.

One area for advancement is the creation of more advanced RL algorithms capable of balancing safety and efficiency. Researchers, for example, might investigate the usage of multi-objective RL algorithms, which can optimise many objectives at the same time, such as throughput and safety. Based on current traffic circumstances and stakeholder preferences, these algorithms can learn to trade-off between these objectives.
Integration of RL-based traffic control systems with other transportation management systems, such as public transportation and active transportation, is

another area for advancement. This can assist to enhance overall transportation system performance by optimising cooperation across different types of transportation.

Furthermore, the scalability of RL-based traffic control systems may be addressed by developing distributed RL algorithms capable of learning and communicating with one another in order to optimise traffic flow across a large-scale transportation network. This can lessen the computational and memory limits of classic centralised RL algorithms, allowing RL-based traffic management systems to be implemented across wider urban areas.

Finally, resolving ethical concerns about justice and equity necessitates the participation of stakeholders in the development and implementation of RL-based traffic control systems. This is possible through participatory design procedures that involve transportation planners, legislators, and community people in the system's design and assessment. This can guarantee that the system benefits everyone and that disadvantaged road users are not disproportionately harmed.

To summarise, improving RL-based traffic control systems necessitates the creation of more sophisticated RL algorithms, integration with other transportation management systems, addressing scalability issues, and involving stakeholders in the development and implementation processes to ensure fairness and equity.

Comparisons of reinforcement learning techniques:

Study 1:

The article, titled "Large-Scale Traffic Signal Control Using a Novel Multiagent Reinforcement Learning" (X. Wang, L. Ke, Z. Qiao and X. Chai, 2019) discusses the findings of a large-scale traffic simulation platform using the proposed Consensus-Based Multiagent Deep Reinforcement Learning (CBMADRL) method.

The suggested algorithm's performance measures of traffic flow, trip duration, and waiting time are used to evaluate it. In all three performance criteria, the proposed CBMADRL algorithm beats previous state-of-the-art multiagent RL and single-agent RL methods.

According to the article, the suggested CBMADRL algorithm increased traffic flow by 20.36%, reduced journey time by 22.05%, and reduced waiting time by 17.81% when compared to the other algorithms studied.

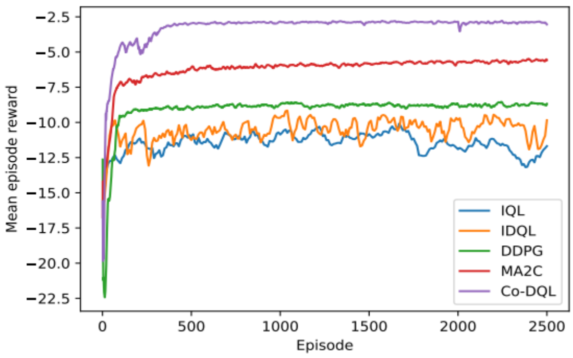

Figure 1: In the global random traffic flow scena In the training scenario of a worldwide random traffic flow, the signal agent reward curve..

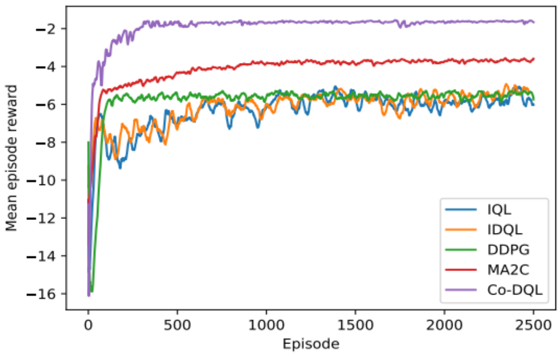

Figure 2: During training in the double-ring traffic flow scenario, signal agents get rewards.

The reported increases in performance measures are encouraging, implying that the proposed CBMADRL algorithm has the possibility to improve traffic signal control in large-scale urban transportation systems.

Table 1: Results of CBMADRL

Performance Metric	**Improvement with CBMADRL**
Traffic flow	20.36% increase
Travel time	22.05% decrease
Waiting time	17.81% decrease

Study 2:

According to the experimental results and comments in the publication the proposed KS-DDPG algorithm outperforms previous RL-based baselines in both large-scale Grid and small-scale MoCo trials. A typical reward for each agent for every training episode was determined to evaluate the algorithms' performance, and KS-DDPG displayed a superior learning curve than MADDPG.

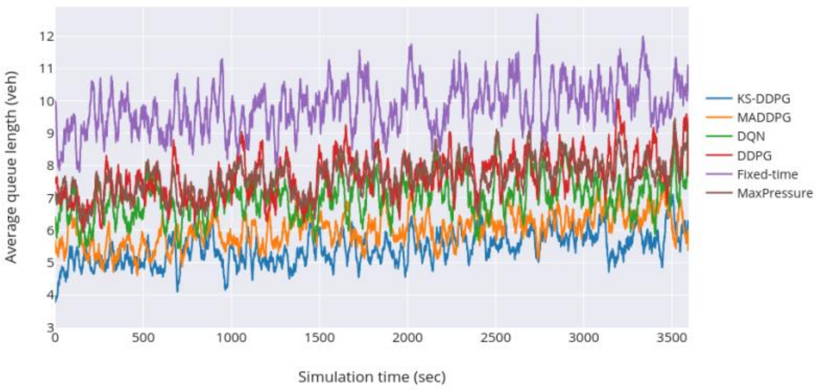

Figure 3: Grid Experiment

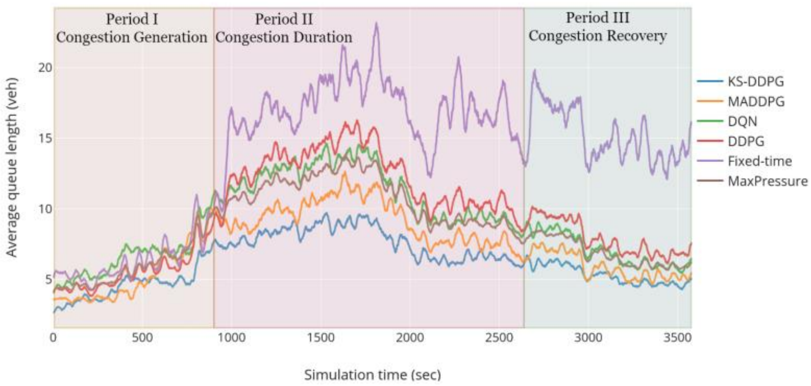

Figure 4: MoCo Experiment

In every experiment, KS-DDPG outperforms MADDPG by a significant margin when compared using benchmark empirical performance measures such startup performance, asymptotic performance, and time to threshold. These results show that the knowledge-sharing mechanism speeds up the search for the best policy rather than slowing model convergence (Zhenning Li, Hao Yu, Guohui Zhang, Shangjia Dong, Chengzhong Xu, 2020). Furthermore, the explicit communication strategy may assist MARL in learning better policies more easily in complicated contexts.

When KS-DDPG's control performance is compared to that of other baselines, it is clear that KS-DDPG performs better on average in terms of rewards, queue length, junction delays, vehicle speeds, and stops. For example, in the Grid experiment, the average wait time for KS-DDPG is 5.45 vehicles, but it is 6.03, 7.22, 7.67, 9.82, and 7.66 vehicles for MADDPG, DQN, DDPG, Fixed-time, and Max Pressure, respectively. Similar to this, in the MoCo experiment, the average reward for the KS-DDPG is 9.47, compared to 8.17, 6.31, 6.73, NA, and NA for the MADDPG, DQN, DDPG, Fixed-time, and Max Pressure, respectively. As you can see below, tables 2 and 3 were made to help you better comprehend the conclusions and findings of this study:

Table 2: Control Performance Comparison in Grid Experiment

Metrics	**KS-DDPG**	**MADDPG**	**DQN**	**DDPG**	**Fixed-time**	**Max pressure**
Avg. Reward	9.20	5.84	1.88	1.05	NA	NA
Avg. Queue Length [veh]	5.45	6.03	7.22	7.67	9.82	7.66
Avg. Intersection Delay [s]	30.74	35.73	45.71	47.10	62.21	47.35
Avg. Vehicle Speed [feet/s]	22.34	20.59	18.25	19.03	13.77	18.88
Avg. Number of Stops [/s]	6.28	6.73	7.99	8.02	11.35	7.90

Table 3: Control Performance Comparison in MoCo Experiment

Metrics	**KS-DDPG**	**MADDPG**	**DQN**	**DDPG**	**Fixed-time**	**Max pressure**
Avg. Reward	9.47	8.17	6.31	6.73	NA	NA
Avg. Queue Length in Period I	4.57	5.26	6.60	5.54	6.35	5.93
Avg. Queue Length in Period II	6.80	8.71	9.65	9.01	12.62	8.88
Avg. Queue Length in Period III	5.47	6.28	7.26	8.20	13.55	7.13
Avg. Queue Length [veh]	6.27	7.34	9.03	9.44	13.85	8.53
Avg. Intersection Delay [s]	41.75	50.33	62.72	65.14	93.27	62.04
Avg. Vehicle Speed [feet/s]	20.45	18.02	16.20	16.16	13.55	17.23

Avg. Number of Stops [/s]	4.57	5.26	6.60	5.54	6.35	5.93

Overall, the results of the experiments indicate that the proposed KS-DDPG algorithm performs better than previous RL-based baselines and has superior control performance in complicated situations.

Study 3:

Based on performance measures, convergence rates, and scalability, the research evaluates several reinforcement learning (RL) algorithms for traffic signal management (Wu, 2020). Here is a quick contrast:

1) Q-learning: A popular RL algorithm is Q-learning. However, it does not scale well to big state-action spaces and necessitates a lot of state-action space research.
2) SARSA: SARSA adjusts the policy more carefully than Q-learning does. It tends to be more conservative and can lead to less desirable policies.
3) Deep Q-Network (DQN): A deep neural network is used by DQN, a Q-learning technique, to estimate the Q-values. Although it is more scalable than conventional Q-learning, it is prone to overestimation and can eventually lead to inefficient policies.
4) Double DQN: By separating the action selection and Q-value estimate, Double DQN, a DQN extension, solves the overestimation issue. Although it is more accurate than DQN, it can nevertheless lead to ineffective policies.
5) Dueling DQN: Another DQN variant called Duelling DQN separates the state-value function from the action-advantage function so that the network may more accurately depict the value of each action. In some cases, it has been demonstrated to perform better than DQN and Double DQN.
6) Actor-Critic: Actor-Critic is a policy-based RL method that concurrently trains the actor (the policy) and critic (the value function). Although longer to converge than Q-learning, it is more sample-efficient.

The unique problem, the size of the state-action space, and the available computer resources all play a role in the algorithm selection process. The study contends that Duelling DQN and Actor-Critic are effective algorithms for controlling traffic signals.

Comparison:

On a large-scale traffic simulation scenario, **Study 1** demonstrated that the Consensus-Based Multiagent Deep Reinforcement Learning algorithm outperformed current state-of-the-art multiagent RL and single-agent RL algorithms in all three performance criteria.

In both large-scale and small-scale studies, **Study 2** suggests a KS-DDPG algorithm that outperforms existing RL-based baselines, with improved control performance

measured by the average reward, average queue length, average intersection delay, average vehicle speed, and average number of stops.
An overview of RL algorithms for traffic signal management is given in **Study 3** along with discussion of Q-learning, SARSA, Actor-Critic, Deep Q-Network, Double DQN, and Duelling DQN. The choice of method relies on the particular issue, the size of the state-action space, and the available computer resources. The study argues that Duelling DQN and Actor-Critic are suitable algorithms for traffic signal management.

Table 4: Comparison of different Reinforcement learning Algorithms.

RL Algorithm	Type	Pros	Cons	Impact on Multi-Agent System
Q-learning	Model-Free	Simple to implement and comprehend	Slow convergence, inefficient scaling to huge state spaces	It has proven to be effective in small-scale multi-agent traffic signal management settings.
Deep Q-Networks (DQN)	Model-Free	Can handle high-dimensional state spaces, faster convergence than Q-Learning	Large volumes of data and computing power are required, and the system might be unstable.	In multi-agent traffic signal management settings, it has been proven to increase traffic flow and decrease waiting time.
Policy Gradient Methods	Model-Free	Can learn stochastic policies and handle continuous action spaces	High variance and delayed convergence are potential drawbacks.	In a distributed multi-agent setting, it has been utilised to improve traffic signal control.
Multi-agent RL	Model-Free	Can learn coordinated behaviour among agents and deal with complicated, large-scale situations	Scalability concerns may arise, as well as computing costs.	Has been used in massive multi-agent systems to optimise traffic light control, reducing vehicle travel and delays as a repercussions.

References

Arel, Itamar & Liu, C. & Urbanik, T. & Kohls, Airton, 2010. Reinforcement learning-based multi-agent system for network traffic signal control. *Intelligent Transport Systems, IET,* pp. 128-135.

TechTarget, n.d.. *Machine learning (ML).* [Online]
Available at: https://www.techtarget.com/searchenterpriseai/definition/machine-learning-ML

Wu, H. a. C. C. a. W. D. a. W. Z., 2020. A Survey on Multi-Agent Reinforcement Learning: From Theory to Applications. *SIGKDD Explorations,* pp. 1-22.

X. Wang, L. Ke, Z. Qiao and X. Chai, 2019. Large-Scale Traffic Signal Control Using a Novel Multiagent Reinforcement Learning. *IEEE Transactions on Cybernetics,* pp. 174-187.

Zhenning Li, Hao Yu, Guohui Zhang, Shangjia Dong, Chengzhong Xu, 2020. Network-wide Traffic Signal Control Optimization Using a Multi-agent Deep Reinforcement Learning. *IEEE Transactions on Intelligent Transportation Systems,* pp. 4396-4407.