How to cite: *F.Kiani, O.F. Sarac, A Novel Intelligent Traffic Recovery Model for Emergency Vehicles Based on Context-Aware Reinforcement Learning,* ***Inofrmation Sciences (2022)****, doi:* ***10.1016/j.ins.2022.11.057***

A Novel Intelligent Traffic Recovery Model for Emergency Vehicles Based on Context-Aware Reinforcement Learning

Farzad Kiani[1], Ömer Faruk Saraç[2*]

[1] Computer Engineering Dept., Engineering Faculty, Fatih Sultan Mehmet Vakif University
[2] Computer Engineering Dept., Engineering and Natural Sciences Faculty, Istanbul Sabahattin Zaim University
[*] Corresponding Author: ofsarac@gmail.com

Abstract-Management of traffic emergencies has become very popular in recent years. However, timely response to emergencies and recovering from an emergency is an important problem in itself. The strategies in the current studies merely suggest that after an emergency vehicle passes, the state should iterate to the next phase. Therefore, this paper proposes a novel approach for recovering from an emergency situation at an intersection based on real scenarios. The proposed method is a combination of context-aware and Reinforcement Learning (RL) models that predicts better alternatives for different states rather than just iterating to the next phase. In this regard, a new algorithm, named Interrupt Algorithm, is proposed to predict proper actions for recovering the emergency situation. This algorithm uses a Q-learning-based model that learns from traffic context for an emergency situation and chooses viable action from an action set. The recovery actions are categorized as *max*, *min*, and *avg*, respectively. Test results show that our proposed model outperforms traffic flow over than standard single choice recovering action-based approach by approximately 80%. Based on this, it may be more beneficial to choose different actions and therefore, proposed algorithm with the help of RL presents a more dynamic emergency recovery model.

Keywords- Reinforcement Learning, Q-Learning, Intelligent Traffic Management, Emergency Situation, Traffic Recovery

1. INTRODUCTION

1.1. Background

High demand on life standards drives people to search for practical and efficient solutions for everyday issues. Traffic-related problems such as longer waiting hours, traffic jams, green environment issues, and noise complaints are among these issues. Given traffic issues, maintaining effective traffic light management could be a potential candidate for a better solution to this matter. Problems may arise if traffic light timing, phase order, and other parameters are not set properly in terms of environmental demands. On the contrary, if there is a well-defined policy, it can be said that efficiency is more likely to be acceptable by nature. Effective traffic policy implementation focuses on handling these issues and improving performance. Besides, the critical issue is to understand how accurate and good a policy is. The basic definition of a good traffic policy suggests that demand should be met with lesser traffic jams, and fewer delays and waiting times such as waiting queues [1, 2]. Therefore, defining and managing a suitable model based on a well-defined policy is essential. In other words, a good evaluation method is needed to assess the model performance in terms of expectancy.

Because problems exist for a reasonable amount of time in a highly dynamic environment, some options and approaches are made to improve traffic flow and increase the ability of traffic flow demand. Different techniques like trying a time sequence for traffic lights, using statistical models based on some performance measures, fuzzy approaches, and machine learning/reinforcement learning-based methods are used to increase traffic flow [1-5].

Within these possible techniques, machine learning-based intelligent methods come forward [6, 7]. Briefly, the best approach to optimize these systems is still an open question for researchers, but a promising approach is learning-based AI techniques [8]. Traditional traffic light management policies are based on assumptions or educated guess processes related to statistical data. However, machine learning-based techniques rely on data and experience. Because they learn from the environment and generate relevant results based on existing values without further statistical or empirical studies, these approaches have an advantage over others in terms of preprocessing, policy definition, and learning curve [9]. Once machines have learned, they can predict solutions for similar environments with reasonable success. Among the machine learning methods, Reinforcement Learning (RL)-based methods may be the most suitable for solving this problem. Because there are dynamic environmental conditions as well as many parameters in traffic light management problems. In this regard, RL methods can present ideal approaches due to their working mechanism. As RL can learn and improve over time with experience, it has great potential for intelligent traffic light management problems [10]. The RL approach ensures that the model learns through selecting the most appropriate actions, and the environment is appropriately monitored [11]. Because it learns from its experience, in an environment, by taking action for a state based on an exploring/exploiting mechanism and generates a degree of reward [12]. As time progresses, it learns according to the reward/penalty mechanism, creates a model, and becomes experienced. A reward is a critical measurement for model efficiency. Various potential candidates can be considered as reward parameters like the number of cars waiting for injunction/light, the throughput of junction or combined/grid intersection, waiting time as well as a combination of these [6, 7]. Besides, there may be different definitions for the efficiency term. For example, time derivative parameters such as waiting time are better when they are lower. Whereas the number of passing cars or throughput is better when higher. Briefly, the RL-based methods are suitable for real-time traffic, which can tolerate contingencies more efficiently. In addition, it does not require external supervision and prior knowledge of the working environment [13].

When we consider traffic flow, it is not just about handling traffic demand in a given context. Traffic flow has emergency issues or some kind of unexpected, interrupt-like actions as well. These issues are outlier elements of the traffic system in this context because they usually need exceptional state and action definitions. As in high demand traffic flow, emergency service response time is a critical issue – both in dispatch timing and arrival at the event site. Because emergency vehicles have priority in traffic flow, choosing a good way of response system for emergency situations is important. A traffic light management system should detect this kind of priority and adjust traffic light orders or states to respond to the emergency issue. Signaling intersection light management, placing vehicles in more efficient locations in a covered area, and route planning are some of the potential improvements and considerations. Similarly, recovering from this exceptional state should also be considered. Because when traffic flow is interrupted and the system is set an unstable state, at least for some time, recovering from such a state should be seen as important in traffic flow and performance. Indeed, we need to respond to what happens after the emergency vehicle passes/exits the system or intersection. Dealing with emergency issues is highly important, but what happens after the emergency has been cleared is also significant.

1.2. Motivation and Contribution

In literature numerous studies exist on emergency issues in traffic flow or urban traffic situations [14-17]. Although they cover various aspects of response issues, they lack strategies on what to do afterwards. Strategies that involve events right after the emergency response and remediation have not received academic attention. That means, although there are important ideas and approaches on how to deal with the emergency issue itself, restoring balance or a stable state should also be covered by researchers. In our study, recovering from an emergency state is worth considering as well as responding to the event itself. In this paper, a novel algorithm for dealing with the aftermath of emergencies is proposed. This is called recovery algorithm and it is based on reinforcement learning. The model learns from traffic flow, then simulates emergency issues/states, and acts accordingly based on the reinforcement learning process. It should be noted that current systems are usually just implementing the next stage/cycle configuration [17-19]. It means that the system continues with the next state defined after the given amount of time for the green light. Searching for a different and possibly better solution is the main proposal of this study. Considering proper usage with more options in what to do after an emergency vehicle passes an intersection can increase traffic flow performance. This paper suggests that there can be more than one option and reinforcement learning-based traffic light system gain experience from the environment and act accordingly. In this regard is defined three options in action selection: (1) Using the remaining interrupted red time so the traffic light configuration will resume the red time for the line which interrupt

event occurred. (2) Using the median of remaining red time, so similarly in action 1, the traffic light configuration will resume the red time for the line which interrupt event occurred, but in this case, the duration will be half of the remaining time. (3) Iterating to the next phase.

As mentioned above, the classic approach uses iterating to the next phase. However, this study is aimed to train and create an appropriate model, so that more accurate and efficient action choices will be made. Therefore, better options can be selected in terms of recovery. In other words, our proposed model helps such interrupts to be handled more smoothly. Because the line that the emergency vehicle passed has a red-light configuration, it gets green time before the desired period and may affect other lines of the intersection. Therefore, adjusting the best possible action can decrease the side effects of interrupting. Existing studies do not directly address these afterward events, so this is a relatively new research topic. In brief, our contributions in this research suggest that, with the help of reinforcement learning methods, traffic recovery scenarios that take place after the emergency situation can vary, and there can be a better option for the desired context. In other words, the addressed issue can be defined as researching a suitable and effective model for recovering from an emergency passing event in an intersection. This study shows that the traffic flow can be resumed and recovered to a better state.

The rest of the paper is as follows: related works about the topic are given with brief discussions in Section 2. Proposed methods are explained in detail in Section 3. Simulation results are given and discussions about these results are presented in Section 4. The last section concludes the paper and provides future work.

2. RELATED WORKS

In recent years, many studies have been carried out with different approaches in reinforcement learning-based Intelligent Traffic Systems (ITS). Improving intelligence of light management can be seen as a frontier topic in this research area [4]. Because traffic flow is mainly managed with traffic light configuration systems, most of the studies are focused on light management topics. In this regard, some of these studies focused on traffic light timing [1, 20], while others concentrated on phase ordering [3] and other similar issues [8-10]. Also single or multi-agent approaches can be seen. Some of these studies propose their solutions for a single intersection, while some of them consider multiple and connected intersections.

2.1. Factors in Traffic Light Configurations

Traffic light order or phase order, cycle length, green light duration, phase sequence, and variance are significant parameters in traffic light management. These elements are modified so that desired traffic flow performance could be reached. In order to observe and assess whether performance is increased, some other measurements are required. Naturally, queue length, average speed, waiting time, delay, throughput, and the number of stops are used as configuration and solution factors in studies. Also, assigning a better phase timing and phase order to an intersection is the main goal of the traffic light configuration process. Therefore, attention should be paid to investigating the timing policy as well as the order of the lights.

2.2. Application of RL method in Traffic Light Management

Reinforcement learning has the ability to learn from experience [32]. This approach does not require a predefined model or set of outcomes, and the reward definition and state-action relationship are enough to act upon experience [1, 3]. Because traffic is a dynamic environment and it is relatively complex to define a good policy to manage, RL is a good choice to consider. In this regard, researchers have focused more on light management problems with the use of RL.

As mentioned earlier, traffic light management focuses on parameters such as light timing, phase sequence, and cycle length. The timing of a traffic light and the duration of green light are considered in many studies [1-8]. In this approach, the main purpose is to find the optimum traffic light duration for a light. It could be one value for every light in the cycle or different values for every light. Phase order is another important parameter that attracts researchers. In [3, 4, 8], the order of green lights and variances of these lights are studied. A cycle could cover every green phase for lights or it could be used in one of many cycles [10, 37]. For this, Q-learning and deep learning-based methods are becoming widespread [3, 8]. The Q-learning is used in mainly single intersection problems or non-grid-based multi-

agent environments. Other approaches are integrated like deep learning if the problem domain addresses bigger connected intersection networks [8, 11].

Traffic light management also includes different movement options and other traffic-related entities such as pedestrians and small mobility vehicles. In some studies, these factors also are considered [35]. Free turn for a direction has a different learning effect on RL-based models and traffic flow considered based on these free turn lines. Also, pedestrians could affect traffic flow because in some cases, intersections have the functionality to intercept defined policy. They do that with an interaction point, such as a movement request button, and the intersection gives them proper crossing time. This changes intersection policies that's why some studies consider pedestrians whereas most of them do not [11, 35].

Reinforcement learning consists of 3 main mechanisms which are state, action, and reward, as stated earlier. These elements are combined and aligned properly with traffic light configuration choices. Mostly, one parameter in these factors is used in the state or reward definition instead of combining multiple parameters [4, 8, 10]. Action definitions basically include 2 options: iterate to the next state or extend the existing state duration [3].

Table 1. Traffic Light Configuration Factors in RL Based Approaches

Factor	Definition	Performance Expectation
Queue length	Total number of waiting cars in line	Lower queue length
Waiting time	Total duration of non-moving time for every car	Lower waiting times
Delay	Expected travel time compared to actual travel time	Minimized delay
Throughput	Total number of passed cars at certain time	Maximized throughput
Average speed	Vehicle speed percentage by speed limit	Maximized average speed
Number of stops	Total number of stop-non moving action for every car	Lower number of stops

2.3. Emergency Response and Recovery

Besides normal flow, emergency vehicles and their relations with a traffic light system are also relatively broad research areas because this is also an essential part of ITS. Most of the current studies focus on solving routing issues or guiding emergency vehicles through intersections because it is the first problem to assess [15, 21]. Reaching the event site on time which can be named response time, is crucial in emergency issues so finding the most optimum route under every possible condition is significant. For that matter, some of the researchers try to improve existing strategies in the intelligent traffic light management context as well. For example, [17, 20, 22] has addressed how to signal the intersection system or change the light state properly before the vehicle approaches the intersection. In [23, 24] has been focused on locating an emergency vehicle effectively in geographically better places in order to improve response times when an event occurs. The methods and findings of literature studies are generally as follows. It must be stated that existing studies do not directly give a solution to what happens after an emergency vehicle passes an intersection. Rather, some state that the model will be iterated to the next state defined in the policy. Because there are no examples to consider and compare directly in the literature for our knowledge, related works are given in emergency handling strategies manner, and if there is a reference to traffic recovery, it is also reported.

Moroi and Takami proposed a method for reducing travel time [18]. They used the Vehicular Ad-hoc Network (VANET) structure for messaging between cars and roadside units. When an emergency occurs, the emergency path and direction are transferred by a message through cars and roadside units, which are carrier beacons for messaging. When an emergency vehicle travels through a route, vehicles in the environment receive an emergency issue message and defined route so they can make a proper maneuver to make way. VANET acts like a connected network in the environment and these messages are transferred to the next block of vehicles. The authors state that sending route and incoming emergency vehicle position decrease the travel time to the location in which the event occurred. However, recovering from the emergency issue is not mentioned in this study. In [14], researchers have used the traveling salesman problem-based shortest path algorithm to find an optimum route for an emergency vehicle. The shortest and strongest path is the preferred route in this research because of safety and traffic density issues. They calculate weights for possible paths and choose the best option in terms of travel time and findings to show that a good performance can be achieved. Weighted route sections help emergency vehicles to select the best option possible for the event site. Although the shortest path is the desired route, the authors state that traffic density and dynamic flow may cause the

route to be a poor option. State iterates back to normal after the emergency vehicle gets service from intersections, but detailed information on how it was done was not provided.

Linders has proposed a general approach in his relatively old research which is an interesting paper to evaluate for future reference correlation [15]. He suggests a method in which every aspect of traffic flow can be digitalized and connected through some form of network, including maps, vehicle positioning, road situations in a time interval, etc. He stated that such a network could improve dispatching and routing performance for emergency vehicles because there can be a lot of proper information to use. Although this paper is relatively old, its focus on networking and using different layers of information to sort out path-finding problems is a good example of a common approach for solving traffic emergency issues and is still a valid approach. This work could be stated as a general framework for intelligent traffic management, but recovery scenarios are not mentioned.

Al-Ostath et al. have proposed a Radio Frequency (RF) based signaling unit for sending and receiving emergency messages to intersections [16]. They put an RF signaling unit in emergency vehicles which can send messages when an event occurs and needs to clear path. At the intersection, they also used an RF unit that can receive a message from an emergency vehicle and then change the phase to green for the road emergency vehicle. This is the main sketch in many types of research, which is signaling and adjusting intersection green phase status. The difference in their study is in communication level. They claim that this type of communication is secure and faster so traveling time can be reduced. There is not any reference to restoring balance or what to do after the vehicle passes the intersection. Almuraykhi and Akhlaq implemented a multi-layer signaling intelligent traffic light system to arrange a green phase for the proper path [17]. In their paper, they defined a physical layer for gathering signals and sending messages and a middleware to dispatch messages properly for the application. They have calculated the shortest path to the event site and signal traffic lights in that route to set to green before the vehicle approaches the intersection by relying on Maps services Message Queuing Telemetry Transport (MQTT) and Google Maps services. They benefit from multi-layer structure and middleware communication and they state that data transfer and orientation mechanism helps response time significantly. They show that the arrival time is significantly reduced by giving three-layer intersections to organize. The model returns to normal flow again, but detailed information on how and under what circumstances is not given. Ye et al. presented an algorithm for positioning emergency vehicles in a mapped network [22]. Different from other emergency situation response research, their focus is mainly on finding proper locations to consider for emergency vehicles. They stated that covering every site on a map is very important in order to reduce response time, so they proposed an algorithm that puts vehicles in different locations on a map. They calculate a vertex coverage weight for every possible intersection and try to find proper locations to place vehicles. They use a genetic algorithm-based coverage approach with demand ratios and show that for proper maturity level-which can be set on the model, very high percentages can be reached. The authors show in simulation results that the response times can be reduced based on their suggested algorithm. However, there is not any reference or mentioned option about recovering from the emergency situation.

In another study, Palle et al. The Arduino unit proposes an emergency response situation in terms of road clearance using a signaling system consisting of InfraRed (IR), a sound warning system, transmitters, and other sensors [23]. They are also used in various systems such as wireless sensot networks [24]. The primary purpose is to allocate road sections to increase travel time and reduce the time required to reach the emergency site. They propose that using dynamic line changer equipment and sending a move action signal will distribute traffic flow to more lines, and one of the lines will be dedicated to the ambulance. They tested this proposition with a simple led-based environment and stated that the proposed method decreases the time used to get to the destination. Again, there is no mention of an emergency recovery policy in this study. Mouhcine et al. propose a method for path selection in a distributed manner based on the ant colony algorithm [25]. They have used environmental variables like traffic density, speed limit, and availability of ambulances. The proposed method includes three agents, called routing, ambulance, and emergency center. These three agents interact with each other to find an optimum weighted path. The emergency center agent receives calls and initiates processes and manages other agents. This agent sends messages to other units to calculate the cost of possible paths. Every agent is responsible for designated tasks and performs action according to communication between agents. The routing agent is the core structure that calculates the density-based proposition for the ambulance. The authors have shown that the results in their study are better than other known approaches. However, there is no mention of an emergency recovery in their model.

Another study on traffic management was studied by Feroz et al. [26]. The authors have introduced a comprehensive fog computing-based system for a smart and connected traffic environment. In this study, a new type of messaging system has been proposed in which messages are carried mostly on a fog network and using VANET interactions, broadcasted to vehicles. Cutting network package travel time and using cars as connected agents helps reduce messaging time therefore, travel times can also be reduced due to faster response actions. Because traffic has dynamic and life flow conditions, cutting communication costs may reduce the time to take action. They stated that travel time is significantly lower than other agent-based full network models. In this model, it is said that the system state iterates back to normal coınfiguration but there is no detailed info.

In [27], Li et al. proposed a method for solving clearance problems in heavy traffic conditions. This study covers the solution for automated and connected environments due to simplifying human interaction problems. The authors suggested a type of cluster head approach as emergency vehicle approaches to a block of vehicles. When this model is used, the connected cars in front of the emergency vehicle move accordingly and give way. This is held in the shortest time and with the least loss of speed. Relevant mechanism continues until the crash site has been reached. They calculate trajectories in online and offline options. Because a connected environment can act faster than a human-oriented decision system, automated vehicle response times are also better. The authors state that with a connected and automated environment, clearance can be held in an online manner which in turn decreases travel time. There is not any reference or mention of emergency recovery that could be found in this research. In [28], the authors proposed an Internet of Things (IoT)-based method consisting of multiple layers and components interacting with each other. They use a mobile application and Global System for Mobile communications (GSM). When an emergency occurs, optimal route selection is calculated based on current location and traffic density. When a route has been selected, green wave conditions, and lights set to green in an ordered fashion are implemented via sensor and GSM modules. There are no statistical results in this study as it is a preliminary approach. However, the authors specify that by using connected internet-based applications and beacons, the system should perform better than Radio Frequency Identification (RFID) and such old technologies. Also, there is no mention of what happens after the emergency.

Shamsi et al. use deep reinforcement learning to reduce the delay that is emergency vehicles pass the intersections on the path at a better timing value [29]. They use accumulated total waiting time as a reward function. The simulation environment consists of a 4-way intersection, and the episode time is around 30 minutes. Their findings suggest that a 27% to 40% decrease in average delay can be reached. This study shows that teaching a model what to do in an emergency situation can increase response performance. However, there is not any recovery scenario in this study. In [30], researchers have studied about adaptive emergency traffic management with a multi-agent-based convolution neural network structure. They use a delay matrix to improve decisions based on actions. The authors tested the model in average traffic density as well as high-demand traffic. This paper has a similar approach to Shamsi et al. but they use a delay matrix rather than an accumulated waiting time. The results show that the proposed model is better than some of the existing benchmarking studies and methods. In this paper, no reference can be found about recovering after an emergency situation. In [31] Nama et al. have studied opportunities and challenges in machine learning-based intelligent traffic management systems. Their study gives a comprehensive background in different ITS aspects, including emergency vehicle routing. In this regard, data collection, analysis and using proper algorithms for desired solutions have been examined. Finally, the authors stated that shortest path algorithms such as Dijkstra can be used for emergency vehicle routing. This study gives an overall look at the topic, but there is no mention of an emergency recovery strategy or method.

A brief comparison of the studies discussed and reviewed is presented in a concrete form in Table 2. In all studies, there seems to be no specific work or focus on what to do after the emergency-interrupt event occurs. Only some of the papers mention models could go back to their normal state. It has been inferred from that explanation that the model should iterate to the next phase in this scenario. This shows that there is a serious gap worth investigating. Because it is unclear how back to normal approach could be accomplished, we considered these mentions to simply iterate to the next defined phase. As known, these studies mostly focused only on how to deal with emergencies. Hence, the findings show that proposed models respond well to traffic flow and emergencies. Naturally, interest in improving response and travel time and how to configure the environment in an emergency situation is vital but so is what happens after. We present a new research model by considering the latter issue. The motivation behind our study is to examine whether a different strategy can be applied to the traffic light configuration after an emergency event. If

there is, we aim to devise a solution that helps traffic management systems take more proper actions to improve traffic flow after the interrupt event.

Table 2. Comparison of current studies based on emergency issues

Study	Goal	Method/Approach	Recovery Scenario
[14]	Optimal path selection	Calculating shortest and strongest path	Not mentioned
[15]	Reduce response time	Construct a traffic network DB with different parameters available	Back to normal
[16]	Reduce response time	Arrange traffic lighs with sender-receiver RF based communications	Not mentioned
[17]	Reduce travel time	Multi layer smart traffic light system with mqtt and maps integration	Back to normal
[18]	Reduce travel time	VANET based messaging with vehicles and road beacons	Not mentioned
[22]	Optimal location selection	Calculate map coverage for vertexes and edges	N/A
[23]	Reduce travel time	Dynamic road lane allocation with signalization	N/A
[25]	Optimal path selection	Ant colony based distributed path selection	Not mentioned
[26]	Reduce travel time	Fog computing-based communication	Back to normal
[27]	Clearance for emergency vehicle	Kinematics based platoon movement calculations	Not mentioned
[28]	Reduce response time	IoT based communication via mobile apps and sensors	Not mentioned

3. PROPOSED METHODS

Emergency situations are important in traffic flow scenarios where they should be dealt with immediately. Hence, managing traffic flow and intersection light configurations are possible actions to take in an emergency situation for ITS scenarios. Because there will be an interrupt, the traffic system faces an abnormality, and it also has to be dealt with. As stated before, this study suggests that with the help of reinforcement learning methods, traffic recovery scenarios that occur after the emergency situation can vary, and there can be a better option for the desired context. Since RL methods do not need a ready-made model and learn from experience, they can be more effective and better modeled in observing the traffic environment in a particular context [32, 33]. We can benefit from this mechanism without the need for preprocessing. Due to the nature of the problem under consideration, other machine learning methods cannot offer very useful and realistic results, because neither a ready-made model can be used, nor the results are predetermined. The reinforcement Learning method can model dynamic, variable, and continuous-based problems because it supports model-free algorithms. One of the methods from this method family and suitable for the problem is the Q-Learning algorithm. The Q-learning can interact with an environment based on the reward model, so the model does not need to be defined precisely. That means we could focus on defining a good performance measurement, rather than a well-defined model. So we can overcome this problem with the help of Q-learning. This study suggests that recovering from an emergency situation could be improved in terms of traffic flow performance. A Q-learning-based model is proposed to interact with the environment and learn from experience to assess which option could be better. Hence, one of our important contributions to the topic is defining multiple action choices. These options are based on possible actions related to traffic light timing in which an emergency vehicle passes. Therefore, this study will shed light on future studies.

3.1. An overview of the proposed model

Segmentation of the emergency event should be considered in order to better and comprehensively understand the problem area. Three sections could be named as approach, event, and recovery, respectively, as shown in Figure 1. The approach can be defined as signaling-about to be period of the event. In this section, a distributed model/traffic

system is about to be aware of an interrupt action. The event section means the actual passing of the emergency vehicle from the intersection. Third, recovery represents the action to be taken after the emergency vehicle exits the intersection, and this is the focus of our proposed model. In the approach section, an emergency vehicle enters an intersection or intersection system. At this point, there must be a way to inform the traffic light infrastructure by communicating with the right-of-way rules. As shown in Figure 1, action named 1 is the event of entering the system in which an emergency vehicle sends a message to the traffic light control system. After that, the traffic light control unit configures the lights accordingly to make a proper green light setting in action named 2. This is an exception for underlying traffic light infrastructure, which means the line which is the emergency vehicle approaching should be turned to green as soon as possible, if not immediately. When the approach section event is complete, an emergency vehicle can cross the light and the intersection, named as interrupt event section. No blocking configurations should be allowed in this stage according to traffic rules [2]. The emergency vehicle passes from the intersection in this section without any concern of blocking traffic in action namely 3. After the interruption section is completed, the emergency vehicle exists at the intersection and in the system. This is the period where recovery occurs. In action number 4, the emergency vehicle signals the traffic light infrastructure that the request to pass through the intersection is complete. That means the emergency vehicle has passed the intersection, and the need for interruption is complete. After that, taking proper action according to learned values is implemented as in number 5, based on reinforcement learning. Learned action for the state is applied to intersection light configuration accordingly. If learning mode is on and the model is still learning from the environment, newly acquired values are updated for the state.

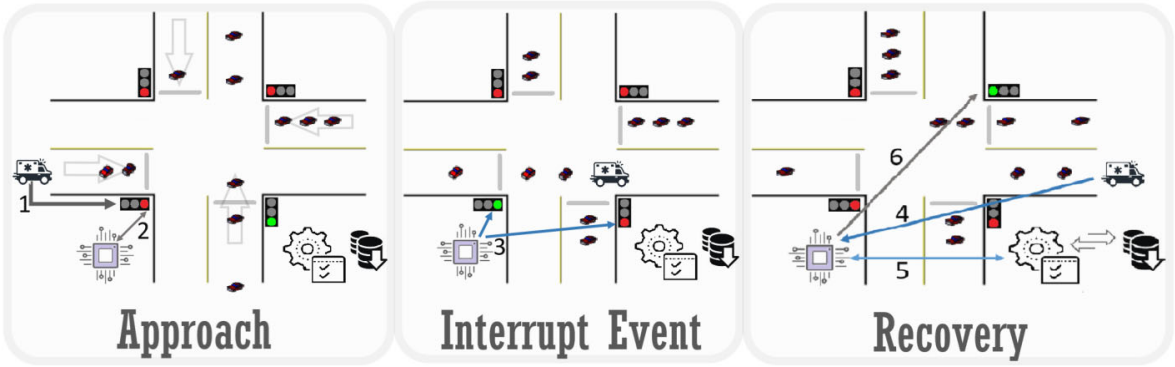

Figure 1. High level view of model interactions

Approach and event sections are prior stages for recovery and they do not play role in the learning of our proposed model. In the recovery section, an action selection, simulation, statistics collection and learning are performed. This is an important phase where data has meaningful value for the model. In other words, data collected from the recovery section is used in learning. At actions named 4, 5, and 6, the model learns what to do in time. This is done with reinforcement learning, which is essentially a mechanism that learns from observation and experience-reward feedback. Defining which data could be useful or considered valid in the environment is in the context-awareness domain. When an emergency vehicle enters a traffic light/intersection system, it must be arranged so that it can cross the intersection as soon as possible. Since these types of vehicles have the right of way, the traffic light configurations should be changed in case of need. As mentioned before, after the vehicle passes the intersection and exists the system, what should be implemented as the next action is not detailed in existing studies.

Our question is: is there a better alternative that can be learned from experience to choose rather than just iterating to the next state approach? This is why we propose an interrupt algorithm responsible for arranging the light status as well as actions after emergency vehicle passing. Once an interrupt occurred, the system has to act accordingly and give way as in standard secure approaches and previous works. After that, the policy/system can arrange settings to resume the current light state within the remaining duration or some other choice and iterate to another phase. Recovering after an emergency state is our definition of after the action period in this model. The proposed model states that the timing of the interrupt can affect recovery options. If the interrupt takes place in the last seconds of light, for example, it might be useful to continue with the next state but in some alternatives, re-running the existing state could be better. When we consider the interrupt event itself, it pulls forward the green cycle for the lines in which

the event occurred. Density then becomes another important aspect to consider in recovering situations. This paper's reinforcement learning-based context-aware model learns from these context values as experience and selects the best-known option for recovery. The results of this study, which are presented in section 4, show that this model outperforms.

3.2. Revisiting Reinforcement Learning and Context-Awareness

The RL has three main blocks: state, action, and reward [34]. In general, the RL model learns from experience based on some reward evaluation. The state represents the existing situation based on the environment and problem domain. Action is the procedure to take on that state based on some reward. There are rule-based, model-defined algorithms and model-free, off-policy algorithms, as shown in Figure 2. In every algorithm, state, action, and reward are defined regardless but with minor differences in computation methods. The state always represents environment/domain definition and action defines what could be the proper set of behavior. The reward is the measurement of state change performance, and it is crucial as it is stated above because the reward definition gives an option to the RL method to learn from experience. Performance measurements and learning feedback are held by the reward function in the reinforcement learning method. In other words, the core learning mechanism of these types of algorithms is strongly related to reward definition. The reward can be defined as a tabular relation that could be given by an expert in the domain as well as some function of the environment. The general schema of the reinforcement learning agent and environment architecture is shown in Figure 3.

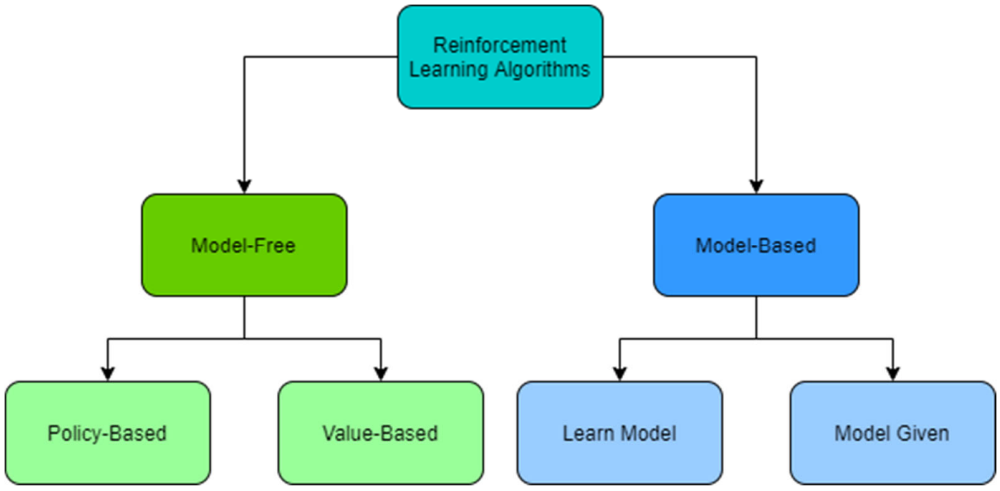

Figure 2. Taxonomy of reinforcement learning algorithms [35]

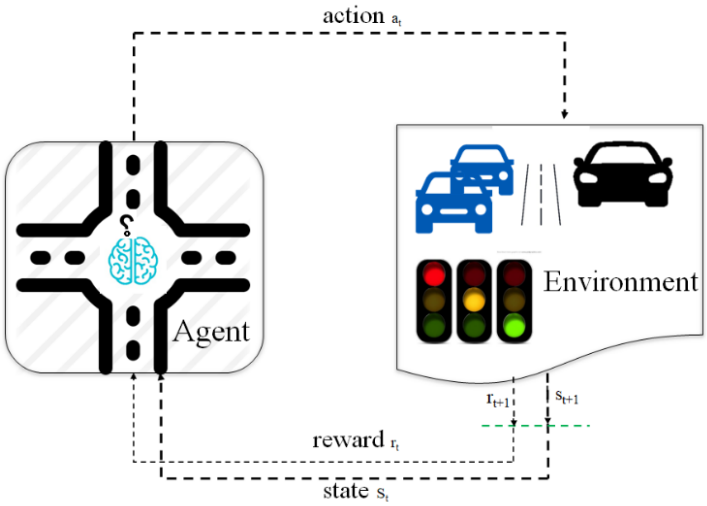

Figure 3. Elements of reinforcement learning

Q-learning, which is used in this study, is one of the algorithms that can be used in the realm of reinforcement learning. This is because the Q-learning algorithm is very suitable for working in a coordinated manner in this type of ambiguous problem. Q-learning is an incremental reinforcement learning method that can be performed online without the need for an environment model, which further assists in solving the focused problem of this study. It embodies parameters like learning rate and discount factor, and uses a reward calculation methodology, either a tabular form or some function to learn from experience. Since Q-learning can be used off-policy, it can also be easily modeled. In other words, a model can be constructed easily without defining a well-structured policy in the Q-learning algorithm [34]. As it has shown in Equation 1, Q-learning process combines the current reward value and some degree of next state value, which can be obtained from the discount factor. $Q_{(t+1)}(s_t,a_t)$ is the state that is going to be updated. Q_t (s_t, a_t) is the existing value for the state and $MaxQ_t$ (s_t+_1,a) is the maximum Q value for current state-action pairs. γ is discount factor, which can be set from 0 to 1 interval. The discount factor shows how important a future reward is for the model. If it is set to a higher value, current rewards have less weight and future rewards are considered more important. It is set to a lower value and as it is nearer to 0, the model tends to be short sighted-that is current rewards are considered more [36]. λ is learning rate, which can be set at a value between 0 to 1 as well. If it is set to near 0, newly acquired information is omitted and model might not learn the newly acquired information, whereas a higher value means that new experience is considered valuable, and Q-values are updated accordingly [37]. In this case, previous experiences may be overshadowed. Therefore, it is important to stay in a balanced place by decreasing from 1.

$$Q(t+1)\,(st\,,at\,) \leftarrow Qt\,(st\,,at\,) + \lambda[rt+1 + \gamma\,MaxQt\,(st+1,a) - Qt\,(st\,,at\,)] \quad (1)$$

The Q-learning algorithm allows us to build a simple yet efficient model if relevant parameters are selected accordingly. The algorithm gains experience and updates its knowledge with newly acquired data in the process. This process includes exploring/exploiting strategies. Explore means that the algorithm randomly chooses an action in the action set, with greedy or other approaches. This is the mechanism to achieve new information about the environment. Exploit is the option in which the algorithm selects the best action in the state based on the previously recorded reward value at this point. Defining a good balance between these two strategies is essential. Most of the implementations use more exploration at the beginning of the learning process or iterations. When the model becomes experienced, the exploit takes more action in the learning scenario.

Context-awareness can be seen as a domain-specific knowledge method, and it suggests that a model or system can handle states and changes in these states according to its context and environment rules [20]. It means that context-awareness makes environment interaction more meaningful. Thinking with reinforcement learning, which is essentially a model based on environment observation, context-aware models can be helpful in solving such problems efficiently. Although reinforcement learning uses observing, models are generally domain-independent generic approaches. Therefore, combining context-awareness with a reinforcement learning approach is potentially good practice to pursue. Reinforcement learning methods require a good understanding of the context, and interaction with context is therefore important. Context-aware models need a good definition of what is being observed, not the knowledge-based desired [24, 35]. We can feed context-aware domain parameters to the reinforcement learning environment, and ergo this will be a good combination. Traffic light state, green time and current time interval, number of cars in the system, traffic density, and such values can be used in context-aware traffic light management [6, 23]. These parameters then easily are used in the learning process of reinforcement learning.

3.3. Interrupt Recovery Algorithm

When an emergency vehicle needs to pass an intersection, it interrupts the current state, and this can be executed in different ways [16, 17]. With this interrupt request, the designated signal turns accordingly, which is red to green in this case. Because emergency vehicles have right-to-pass, non-blocking action must be taken in traffic light configuration. For evaluation criteria, the timing of interrupt is the most proper parameter to use in our model. That means the event time of interrupt is the most important value in our proposed model. It can be defined as categorical values such as beginning, middle, and end. These values define when the interrupt occurred within a light duration or phase.

Action parameters are defined as potential usage of the remaining time. The proposed method assumes that the interrupt occurs in red light states for simplification purposes although it can occur in a green light state. Three categorical values, *min*, *max*, and *avg*, are defined for the action parameters. When the option (action) is *max*, then the

remaining red light time is used. That means the state resumed from the point/second where interrupt occurs first. The next stage is used for the *min* option, which means the model continues with the next phase. This is the standard mechanism that is applied in intersections and some of the studies mentioned above. *Avg* option is the middle point of the other two selections, using the average time of the remaining red phase but not the whole period. As an example, for 40 seconds of the red-light phase, if the interrupt occurs at the 10th second and the remaining time is 30 seconds, then the model will execute avg option for the 15-second red timing phase. During the interrupt period, approaching and passing the intersection will take some time. This means that our modem can be used in real scenarios. Because emergency situation should be dealt with before the actual passing from the intersection, the state turns in the green phase for that lane. Then emergency vehicle passes through the intersection; this action is referred to as clearance in this paper. The clearance time for the model is predefined for a reasonable time period such as 10 or 15 seconds, which is normal for an emergency vehicle expected to pass an intersection in that time period.

As mentioned earlier, since Q-learning has the ability and flexibility to learn from experience without a well-defined model, the proposed model in this study was inspired by this algorithm for the intelligent traffic light system. Because it is relatively simple to implement and has a good performance in learning, we also benefit from the Q-learning algorithm in a mobile context. Without a need to define a complex/complete traffic signal model, the off-policy Q-learning approach still has good coverage over the ITS problem domain [7]. The proposed model uses a model-free and off-policy approach as well. However, due to some features of RL-based algorithms, a new method is presented by combining it with the context-aware-based approach. The context awareness helps the proposed model for addressing related data. What to observe and how to combine observed data into the proposed model could be assisted by a context awareness approach. At an intersection, traffic demand per time, traffic flow directions, queue lengths, and other known parameters could be used to gather and observe the environment. Because Q-learning is model-free and non-dependent on the environment, because it interacts with the environment via a reward mechanism, a strategy is needed to infer meaningful information. Interrupt event, as detailed above, includes many parameters to define and interact with the intersection mechanism so context-aware, self-aware observation is included in the learning process. In this study, the terms of state and action, which are constantly used, are defined as follows.

State definition consists of a couple of parameters, including lane name with labeling light, light durations for every lane, current phase, and current green lights. Interacting with a context-aware observing approach, the current state and values for each parameter can be set. For interruption events, it was assumed that an emergency vehicle could interact with the intersection system, and transmit information to make configurations accordingly. The state can be expressed as a vector, *S={current_phase, interrupt_period, duration}*. The *current_phase* is the number representation of the phase order that interrupts the action that occurred. The *interrupt_period* is the categorical value of emergency action timing and *duration* is the current green light duration. The *interrupt_period* can be beginning, middle or end, respectively. In our method, besides the definition of state, action must also be defined. ***Action*** options were defined as remaining time usage strategy, *min*, *max*, and *avg* in which they form an action set, *A={min:next_cycle, max:use_remaining, avg:use_median_remaining},* as presented in Equation 2. As the name suggests, model iterates to next phase in *min* selection. Max sets the remaining time to be used as action, and avg is the median of the remaining time. A basic comparative representation of actions is given in Figure 4 with the timeline approach.

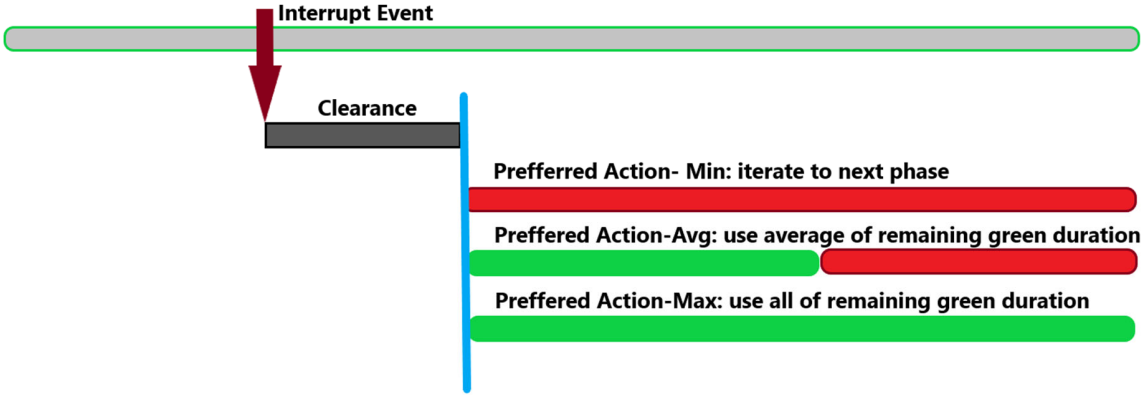

Figure 4. Prefferred actions given in timeline

$$remaining\_time = \begin{cases} (duration_{light} - interrupt\_time)/2, & action = avg \\ duration_{light} - interrupt\_time, & action = max \\ 0, & action = min \end{cases} \quad (2)$$

One of the most important and difficult elements in reinforcement learning is to define the reward mechanism [6]. Choosing an appropriate evaluation method is very important for measuring performance. Many studies use a single type of parameter such as total cars passed, average waiting time, throughput, etc. [6-17]. But in addition to that, we use efficiency as a reward, as a total of cars that go through green stages each time (Equation 3). The total number difference of passing cars in a given time is the measurement in short, for this reward calculation.

$$r = 3600.\left(\frac{N-1}{t_{N-1}-t_L}\right) \quad (3)$$

Where N is the total number of passing vehicles, t_N is the time period of the last passed vehicle (measured in seconds), and t_L is the initial time period of passed vehicle (also measured in seconds), and finally, r is the reward value calculated.

In brief, the proposed model calculates all three stages of an emergency situation because these are all parts of an emergency situation. As stated above, these stages consist of approach, event, and recovering, and every number of vehicles is counted as input. As expected, the model uses these parameters and learns from traffic flow simulations. Categorical values are used in our proposed model, and for simplification purposes, we use three for each. When we consider real traffic situations, much more values are probably required; even some form of continuous parameters should be used to improve traffic flow. However, examining whether the proposed model has potential or not and if there is, how efficient that has directed us to make simplification. Recovering from an emergency situation is a relatively new aspect and challenge of intelligent traffic management. Therefore, simplifications and dividing the problems into small examples can provide an opportunity to solve the problem effectively and to work in this field.

The core functionality of the proposed method relies on context parameters. Knowledge about events and the environment makes new and potential multiple actions. If the model could learn from experience labeled with evet timing, the lane that the emergency vehicle needs to pass as well as the direction, actions could be more specific. The proposed model uses 3 categorical values in a way that did not use before, therefore different and potential better traffic flow results are possible. As the model evolves and learns from experience, based on the Q-learning algorithm, given state and action definitions become more valuable. The existing strategy does not count on any observation about the environment right after the emergency event, whereas our proposed model's novelty suggests that it is eminent.

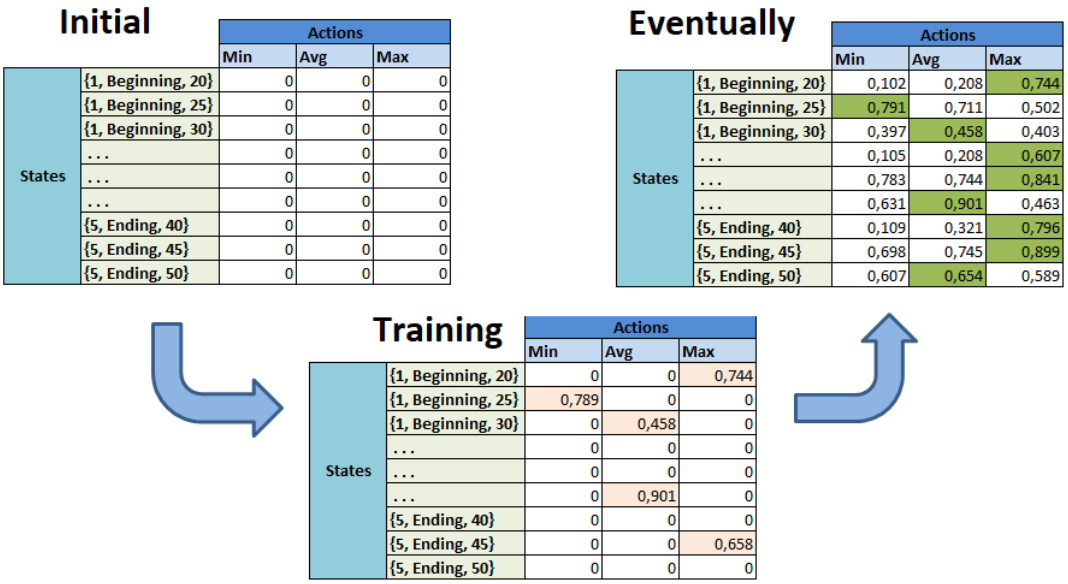

Figure 5. Q-Value update process for Q-Table

After learning is complete, the model can choose the best-known action on any interrupted event. Max Q-values are the ones for every state that must be selected when an interrupt occurs. Q-Table has been constructed with these experiments and includes the relation between state and actions. That means Q-Table is the resulting policy chart for what to do in a state. The learning process is determined according to equation 1.

In Figure 5, a simplified representation of the learning process and effects on the Q-Table are shown. At the initial stage, Q-values are set to zero. In the learning episodes, reward values are updated according to state-action pairs. Eventually, when the learning is completed, the Q-table stores max values for every state-action pair. These max pairs are preferred state-action pairs. The proposed Q-Learning-based Interrupt Recovery algorithm is introduced as pseudocode in Algorithm 1. In addition, the working mechanism of the proposed method is presented in Figure 6 as a flowchart.

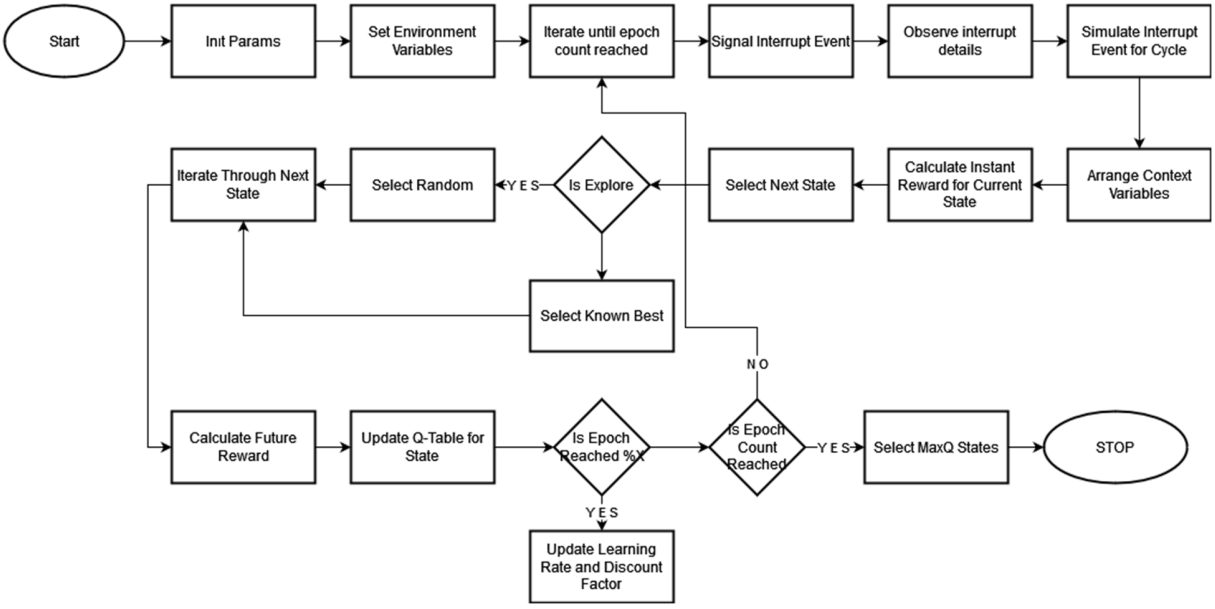

Figure 6. Flowchart of learning process

Algorithm 1. Interrupt Recovery
Input: traffic state observation ***Output***: best states for interrupt events

Init parameters
$Q_{(t)}(s_t, a_t) \leftarrow 0$ //*Initialize Q-table for Interrupt*
Set environment variables
for all episodes **do**
for all light_duration_config **do** //*Iterate through light-duration configuration*
interrupt(state) //*Signal interrupt to system*
observe(environment) //*Observe interrupt details*
environment: Light name, phase, interrupt second, interrupt phase
run(interrupt_cycle) //*Simulate interrupt cycle*
interrupt_cycle: Simulate interrupt period for selected action
stats(traffic) //*Observe traffic stats*
Transfer context values for every phase change
Calculate instant reward
select(next_state) //*Select next state*

```
15          Explore/exploit using e-greedy approach
16          Transfer context values for next cycle
17        state ← next_state //Iterate through next state
18        Calculate reward for next_state
19        Q(t+1)(St, at) ← Qt(St, at) + λ[rt+1 + γ MaxQt(st+1, a) − Qt(St, at)] //Update Q-Table
20      end for //Iterate through light-duration configuration
21    end for //Episodes
22    Select max reward solutions as interrupt-value-mapping
```

4. SIMULATION RESULTS AND DISCUSSION

In this section, we present and analyze the numerical results for the proposed model. Simulations have been held on a PC with an Intel i5 processor with 2.2GHz and 8 GB of RAM. Several simulations were made to obtain a good measurement because the Q-learning process has randomness in it. To obtain a healthy measurement, multiple runs must be made. For the Q-learning parameters, the learning rate was set to 0.7 and the discount factor to 0.9; furthermore, 5000 rounds were also used for learning epoch count. The reward function is defined as throughput, which is based on the number of passing cars in time as mentioned above. The case study is based on real data from the Bayrampaşa district of Istanbul, Turkey. Traffic data was gathered from Istanbul Municipality within Traffic Management Department's statistical analysis and monitoring process. Also, traffic data was used in a study in which researchers work on a queue mechanism for improving traffic flow [38]. Used real data is also shared as a supplementary file. Based on the traffic data of this district, the simulation environment in this paper we use has an 8-way intersection, and light durations can vary from 10 to 90 seconds with 5-second intervals, such as 15, 20, 25, etc. All road lines and lights that manage these lines are attached to state definition as in context variable. No free turns are allowed in this type of intersection. The traffic density of this intersection varies from 245 to 1492 cars per hour. Lines named 3 and 7 have a higher density than others, and 4 and 8 are the lines that the least cars pass. As seen in Figure 7, there are 8 lights and 8 roads. Every light period in a single phase is equal. It means in a given phase, every light consumes the same amount of time whether its state is red or green. The traffic cycle consists of 5 phases. In every phase, 2 lines are allowed to be green. Green lines in every phase are as follows: *1={3, 7}, 2={1, 5}, 3={3, 4}, 4={7, 8},* and *5={2, 6}*. Light duration is predefined and fixed for every phase and set to values from 30 to 60.

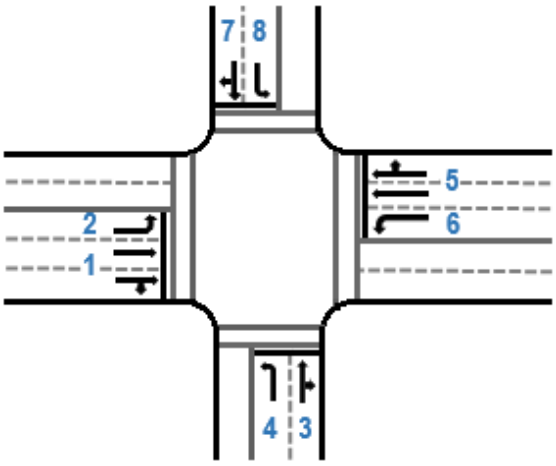

Figure 7. Intersection configuration

Table 3. Simulation parameters

Parameter	Value
Learning Rate (α)	0.7 to 0.5 gradually
Discount Factor (γ)	0.9 to 0.5 gradually
Epoch	5000 times

Min Traffic Rate	0.002 seconds per car
Max Traffic Rate	0.42 seconds per car
Phase Count	5
Green Light Duration	30 to 60 with 5 second intervals
Interrupt Period-Beginning	25% of current light duration
Interrupt Period-Middle	50% of current light duration
Interrupt Period-Ending	75% of current light duration

Traffic data is gathered by related organizations in İstanbul Municipality, as mentioned above. Data gathering is made by sensor usage at an intersection. There are sensors at every road for the intersection which are located around 100 meters away from the lights. Also, there are sensors at the lights as well. When incoming traffic is observed by input sensors, these are recorded as entry time slots. When a vehicle passes through a light, sensors on the light also observe that movement and mark it as exit action. In this way, the total number of vehicles observed and elapsed time can be calculated. Figure 8 schematically shows how traffic data is gathered from an intersection. The speed limit is set to standard urban rules and there are not any specific restrictions or conditions regarding traffic flow weather and road quality conditions, so these are not considered in the model.

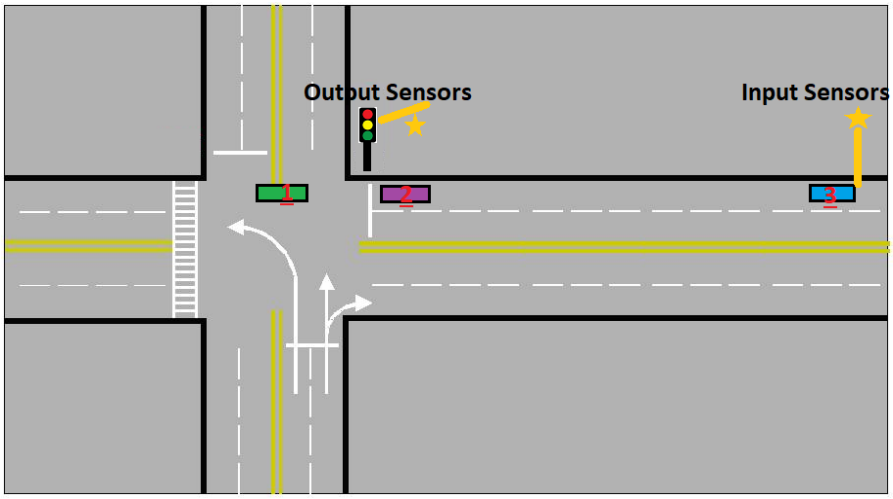

Figure 8. Schema of Traffic Data Gathering Positions

The parameters and environmental values, which are used in this study for analysis and comparison, are given in Table 3. As episodes increased, the learning rate and discount factor decreased. Since a higher learning rate indicates that the model mostly learns from new experiences, decreasing the learning rate in higher epochs helps the model to use existing best strategies because the model would have experience. Exploring from the early stages is the preferred method, but after some epochs, the model should use existing best-known states. Otherwise, the model will just memorize, which in return, could cause false output predictions. Because the discount factor measures how much the model learns from future gains, this value should also be considered as a candidate to have lesser value as episodes progress. In other words, lower rates tend to act more myopic where rewards are calculations of the current state in every epoch. Starting with 70% of epoch count, rates are rearranged accordingly, which ends with a value of 0.5 for both discount factor and learning rate. Traffic rates are measured in seconds; for example, every 0.42 seconds, a car enters the intersection on a higher traffic density road. Interrupt period values are numerical assignments for categorical state definitions. The proposed model assumes an interrupt event can occur in 3 time periods; in the learning process (as well as testing), a given amount of light duration is passed before the event. That means, if the interrupt in ending states and light duration is 40 seconds, the model assumes that the interrupt event occurs at 30 seconds.

A sample of simulation results for the interrupt algorithm is given in Table 4. An interrupt line is the way an emergency vehicle comes from. The period is the timing of the interrupt, and the phase is when the event occurred in the cycle. Preference is what we propose, what to choose in that state. Cycle length is the total time calculated after the whole cycle is completed; this includes usage of remaining time plus clearance for an emergency vehicle. The average

waiting time is in seconds, and the total passing cars is the number of cars passed after the whole cycle has been completed. The total number of passing cars is a cumulative total, including every green phase in the cycle, and the interrupt period. As mentioned before, the timing of the phase that an emergency occurred is important. If it is in the beginning seconds, we observed that the model predicts to use of an average of remaining time or continuing with the next phase more frequently. For the end of the phase, if the line has a high density, the model suggests that resetting the current phase is more favorable.

Table 4. A sample of simulation results

Interrupt line	**Interrupt occurance period**	**Current Phase**	**Interrupt Action Preference**	**Cycle Lenght** with **Interrupt (sec)**	**Average Waiting Time (sec)**	**Total Num of Passed Cars**
1	Middle	2	Avg	128	29,14	217
5	Beginning	2	Avg	125	29,71	207
7	Beginning	2	Avg	125	28,77	248
7	Middle	2	Min	122	28,97	243
1	Middle	2	Max	135	29,55	225
5	Beginning	2	Max	135	30,79	217
7	Beginning	2	Max	135	29,35	259
5	End	2	Max	135	30,80	213

In the study, an interrupt signal is sent 50 different times for every road line. That means simulations are made for 50 epochs. In Table 5, simulation results are also shown in comparison to different outcomes. Because learning involves some degree of randomization due to exploring possible actions, results vary. The worst-case column represents the lowest number of cars passed in the selected action, and the best is the opposite: the most cars passed in this case. Beginning, middle, and ending columns are time interval labels that interrupt events that occurred. As stated before, we model our algorithm to detect interrupt signals as categorical values so that model understands in which phase an emergency vehicle issued an interrupt. Phases are labeled as 0 to 4, and in every phase, there are at least two lights that are set to green, and others are red.

Table 5. Number of passed cars in interrupt cycle for worst, average and best cases for recovery algorithm

		Best			**Avg**			**Worst**		
Line	**Phase**	**Beginning**	**Middle**	**Ending**	**Beginning**	**Middle**	**Ending**	**Beginning**	**Middle**	**Ending**
	0	241	236	231	240	235	230	220	221	223
	2	227	225	223	226	224	222	216	217	219
	3	231	229	227	229	228	226	220	221	223
1	4	225	225	227	224	225	226	224	225	226
	0	227	221	215	227	220	214	227	206	207
	1	217	214	208	214	209	207	208	207	205
	2	220	217	214	219	216	214	209	209	214
2	3	221	218	215	220	217	214	210	210	211
	1	235	234	234	234	233	233	226	227	230
	3	240	239	238	239	238	238	229	231	238
3	4	238	239	242	237	239	241	237	239	241
	0	228	222	216	226	221	215	207	207	208
	1	217	214	208	214	209	207	208	207	205
4	3	219	216	213	218	215	212	208	208	209

	4	209	208	209	208	208	208	208	208	208
	0	229	223	219	228	221	218	209	209	211
	2	217	214	213	216	213	212	207	207	209
	3	218	215	214	217	214	213	208	208	210
5	4	210	209	211	209	209	211	209	209	211
	0	228	222	216	225	220	215	207	207	208
	1	217	214	208	214	208	207	208	207	205
	2	219	216	213	218	215	212	208	208	209
6	3	219	216	213	218	214	212	208	208	209
	1	247	246	247	246	245	246	238	239	243
	2	259	258	258	257	256	257	248	243	254
7	4	254	253	254	253	253	253	253	253	253
	0	228	222	216	226	221	215	207	207	208
	1	217	214	208	213	210	207	208	207	205
	2	219	216	213	219	214	212	219	208	209
8	4	209	208	209	208	208	208	208	208	208

In addition to these, Table 6 lists non-model/standard min value simulation results. As it can be seen from the values, there is exactly one outcome for any given state, so average, worst, and best of values do not apply. Because for every simulation iteration, there can be only one possible action due to exactly one set of options. When the results are examined, it is determined that more vehicles pass for the interruptions at the end of the stage. This is understandably meaningful because more cars would pass through when we use more green lights in the intersection. In some cases, the number of passing cars does not change during the interrupt period because in these lines, traffic density is low and no new cars could be passed from the intersection. As we stated before, a standard run is an action where the traffic light state iterates to the next phase after an emergency vehicle passes the intersection. That means, regardless of the interrupt event period, traffic lights change to the next state configuration after an emergency vehicle passing; named as min action in our proposed model.

Table 6. Number of passed cars in interrupt cycle for standard run

Line	Phase	Beginning	Middle	Ending
	0	201	208	216
	2	206	210	215
	3	210	214	219
1	4	224	225	226
	0	188	194	200
	1	198	201	205
	2	200	203	206
2	3	201	204	207
	1	215	220	227
	3	219	224	230
3	4	237	239	241
	0	188	194	201
	1	197	200	205
4	3	198	201	205

	4	208	208	208
	0	190	196	205
	2	197	200	206
	3	198	201	207
5	4	209	209	211
	0	188	194	201
	1	197	200	205
	2	198	201	205
6	3	198	201	205
	1	227	232	240
	2	238	243	250
7	4	253	253	253
	0	188	194	201
	1	197	200	205
	2	198	201	205
8	4	208	208	208

In addition, the BoxPlot method was used to examine the performance of our method in detailed analyzes and comparisons made in various situations. In this regard, Figures 9, 10, and 11 are boxplot representations of what we have shown in the tables 4 and 5. It should be remembered that these values include each stage and simulation results. The variance and value distributions are good criteria to compare with the standard approach. In general, the least difference between the minimum and maximum values for a box is considered a good measurement. Likewise, higher weighted average values are also good because this favors high passed cars. When we use the t-test of these simulation results, the difference is very low, and it can be said that there is an overlapping pattern. For example, in the case of beginnings, the t-test value is 7.45778E-10 which is very low. When we examine results in the group, there are similarities between passed car values for lights and phases, and curves are similar with one difference; our model has a higher value range which means it has better values. In Figure 9, simulation results are shown for interrupt events that occurred in the beginning period of light duration. The vertical line represents the total number of passing cars in the intersection, whereas the horizontal line is the number of interrupt lines/lights in which the event occurs. Obviously, for every light scenario, our proposed model has a better value distribution and generally less variance or variance with tends to maximum values. Lights 3 and 7 have high density so variations between phases are much higher than the other lines. Lines with fewer traffic densities generally have less variance, and their min-max passed cars numbers are close.

Figure 10 shows that in the middle section of light duration, the difference between simulation results is less than beginning section events. As seen in the figure, in some cases, the standard model has less variance this time; however, the proposed model has the advantage of better/higher passed car results. As in the figure 9, higher density lines have higher values and variances as well. In the middle section of traffic light duration, we can see that more cars can pass through intersections. This is because cars benefit from green light duration more than beginning phases and queues are less. In Figure 11, simulation results for the ending period of light duration are shown. When we look at the figures altogether, in every case, the proposed model has the values of the better-passed cars. Lights control the traffic line, which has higher traffic density; for example, lights 3 and 7 are showing more variance in every case, as it has mentioned. This time, the total numbers of passing cars are higher than both the beginning and middle periods of the interrupt. This is also understandable and expected because traffic light state configurations are made to gain balanced traffic flow, and any unexpected change from that course affects delays. The later the interrupt event occurs, the less effect on traffic flow should be experienced.

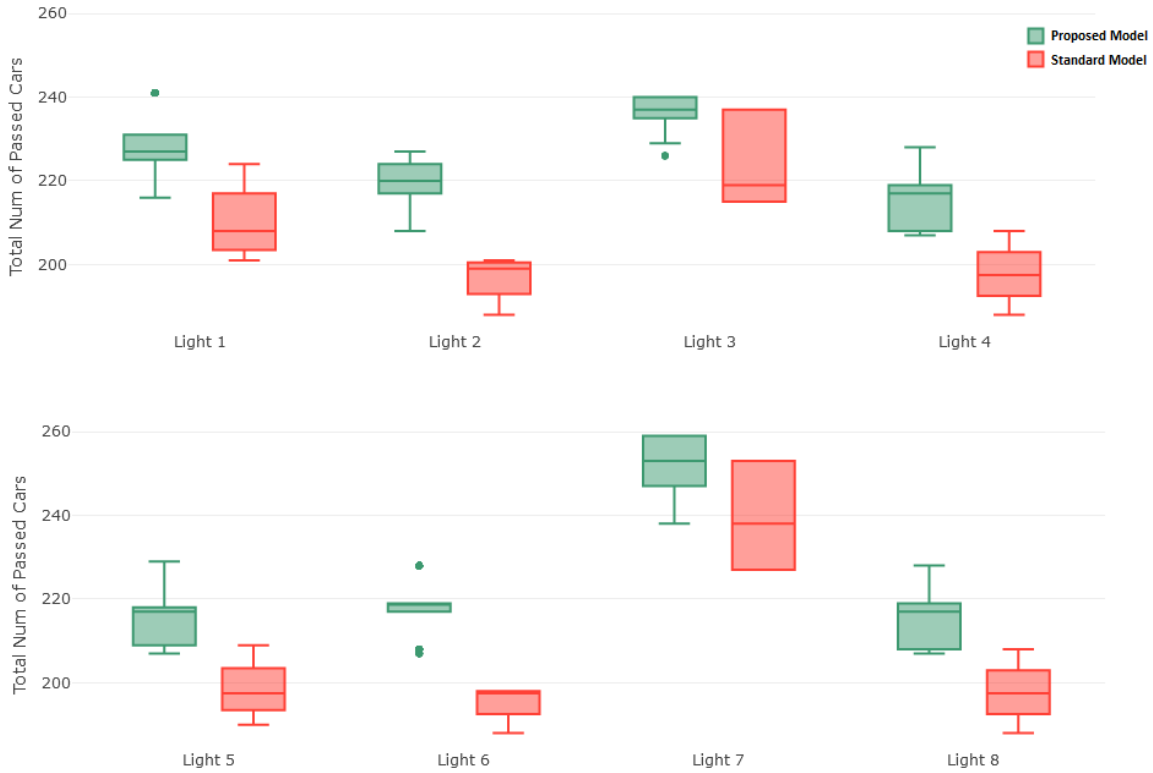

Figure 9. BoxPlot of results for beginning periods in proposed and standard models

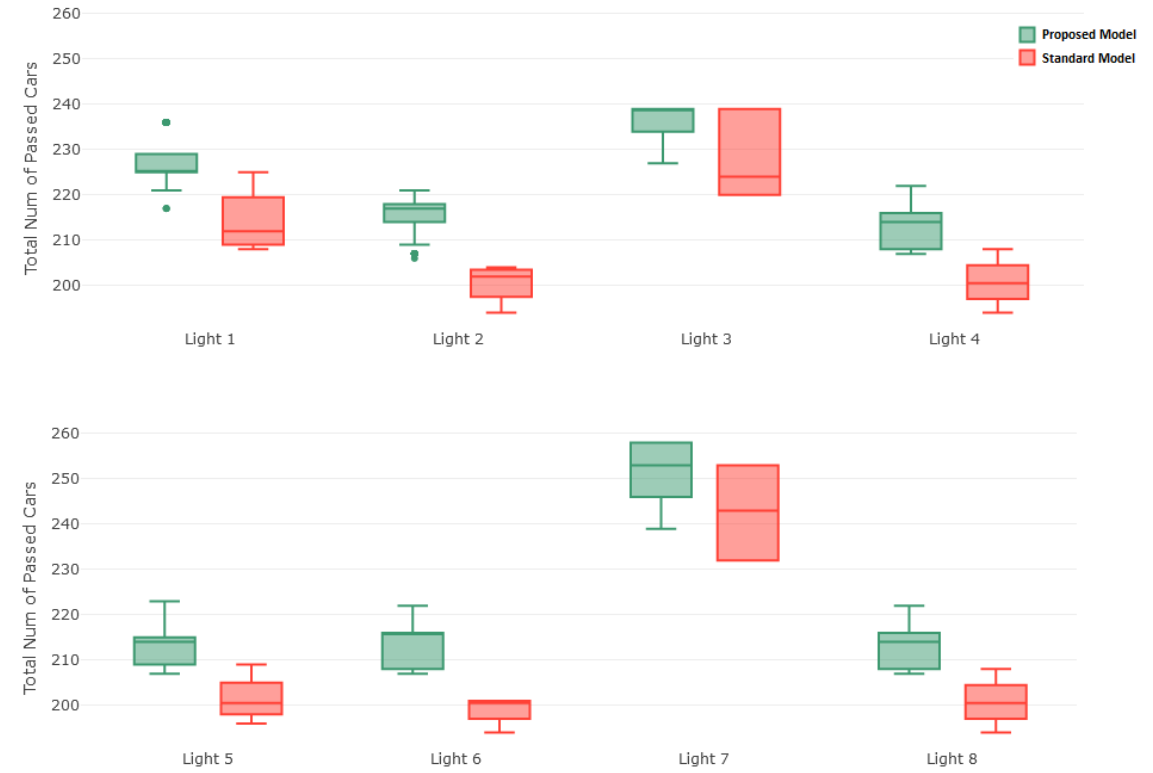

Figure 10. BoxPlot of results for middle periods in proposed and standard models

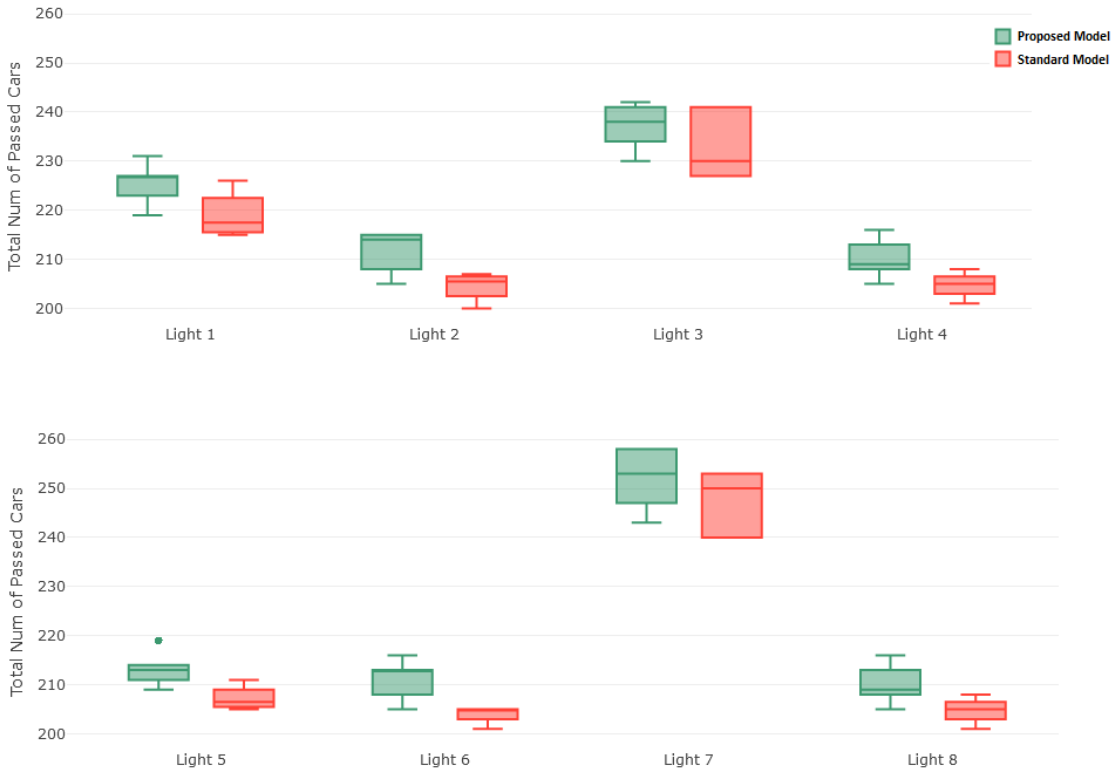

Figure 11. BoxPlot of results for ending periods in proposed and standard models

Besides these, in Figure 12, the distribution of actions is given in total numbers. This figure shows the models' choices for all simulation cases. There are 4500 test cases: 50 simulation runs, 8 lights, a minimum of 3, or a maximum of 4 phases according to light configuration. We can examine how many cases the model selects different actions rather than the standard min value approach. In most cases, our model was able to successfully select the maximum value for the desired action. There is a slightly smaller number of choices for min and avg actions. Considering a total number of simulations, in approximately %80 runs, our model predicts that there can be different valid and better performance options (as max and avg) in recovering situations. The total number of avg and max cases have a count of 3702, and the min count is 798. For every avg and max case, the total number of passing cars is higher than the standard min model. That's why the proposed model performs around %80 better.

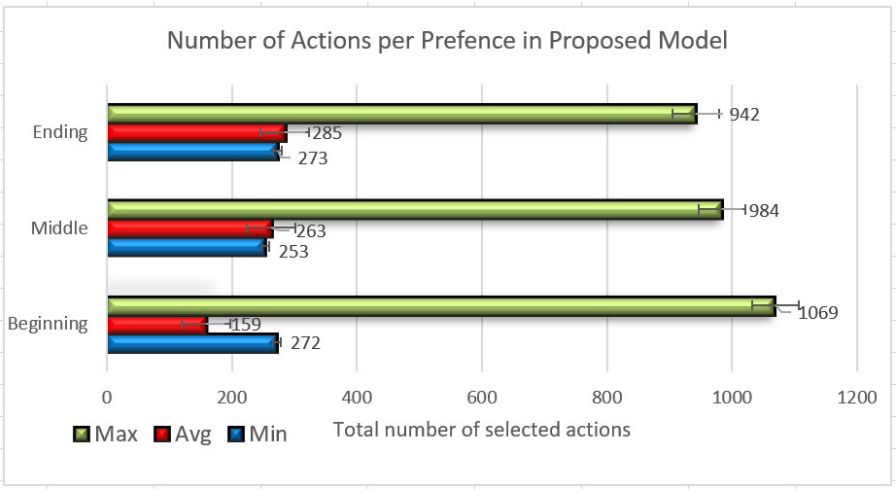

Figure 12. Number of actions per prefence in proposed model

As another simulation result, the approach of model prediction and comparing assumed behavior actions are discussed (Figures 13 and 14). Simulation results for lines named 3 and 7 are given in detail. Because these lines have the highest traffic density and passed car variances are higher than other lines, model behavior and effect can be observed clearly. The vertical line represents the highest simulation results for every interrupt period. First, three of these simulations are maximum values for beginning periods, and the other three elements are the same for middle and ending periods of interrupt.

We compared our proposed model with the existing standard approach, which is iterating to the next phase, as stated above. However, in order to have some further insight, we adopted Abdoo's method to compare whether the proposed method also has any advantage over well-known traffic light configuration studies [39]. Because Abdoo's method is used as a comparison in many studies, we also preferred this work. Although Abdoo's method does not consider directly interrupting recovery issue, it gives a standpoint on traffic light management. Mostly, our model performs well in terms of traffic throughput. The red dotted result set named as the standard model represents min value actions, which is the existing strategy in the traffic light management environment. The horizontal line is the max outcome in 50 simulation runs for the related phase. The most cases, our model outperforms the existing min value strategy. In some cases, like simulation number 6 in Figure 14, the model selects the *min* value as the action which is also acceptable.

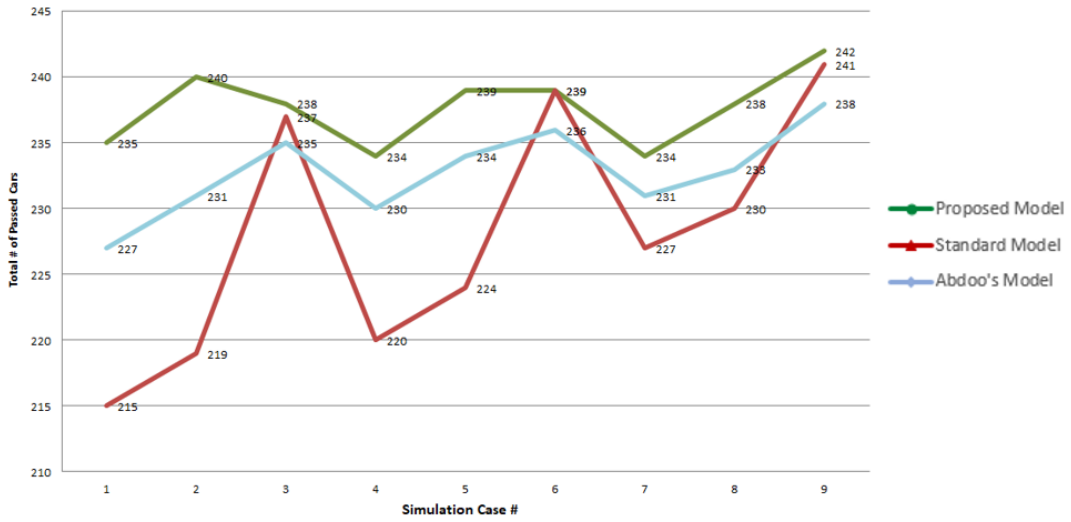

Figure 13. Intersection road line # 3 comparison in proposed model

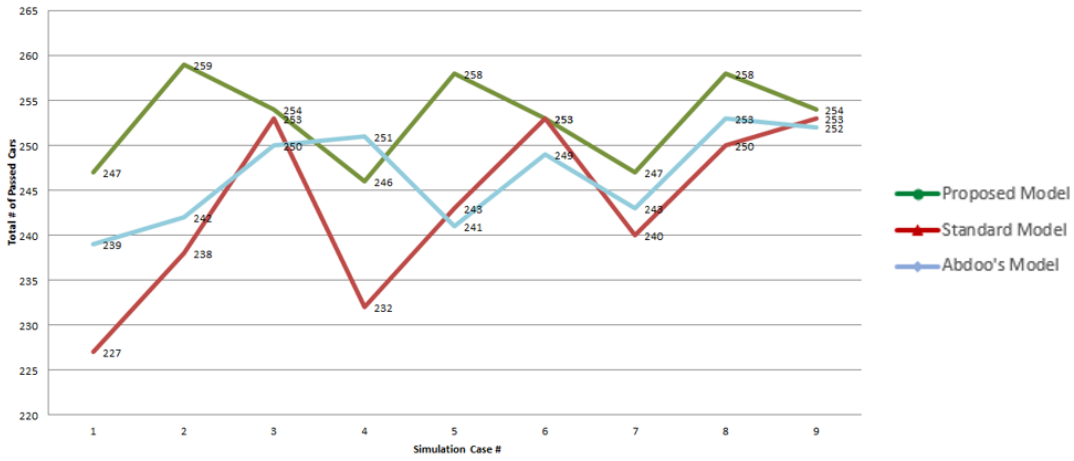

Figure 14. Intersection road line # 7 comparison in proposed model

Figure 15 shows interrupt cycle results for passed car numbers. Y-axis is the total number of cars passed in the cycle, which has an interrupt event. The X-axis is a distinct simulation result for every line that has been subtracted from all simulation results. The emergency event, interrupt action timing is color-coded in the graph: green for beginning, light blue for middle, and light orange for ending, respectively. Values are best-case results for every simulation for every phase. For comparison reasons, we also added Abdoo's model approach. As seen from the graph, for most of the cases, the model predicts a better option than just transiting to the next phase. Abdoo's model results are also mostly better than transitioning to the next phase however it falls behind our proposed model. In some cases, model performance is equal to standard run, which is also acceptable. In all the cases, the model at least chooses min action for the preferred option. As was mentioned before, in every avg and max action, the proposed model results in higher passed car numbers. In min cases, which is essentially the name of the existing approach, model and standard run are equal, as expected. In conclusion, the proposed model chooses the better option for around 80% of cases. Based on these results, in the proposed model, it is observed that the traffic recovery problems have been improved.

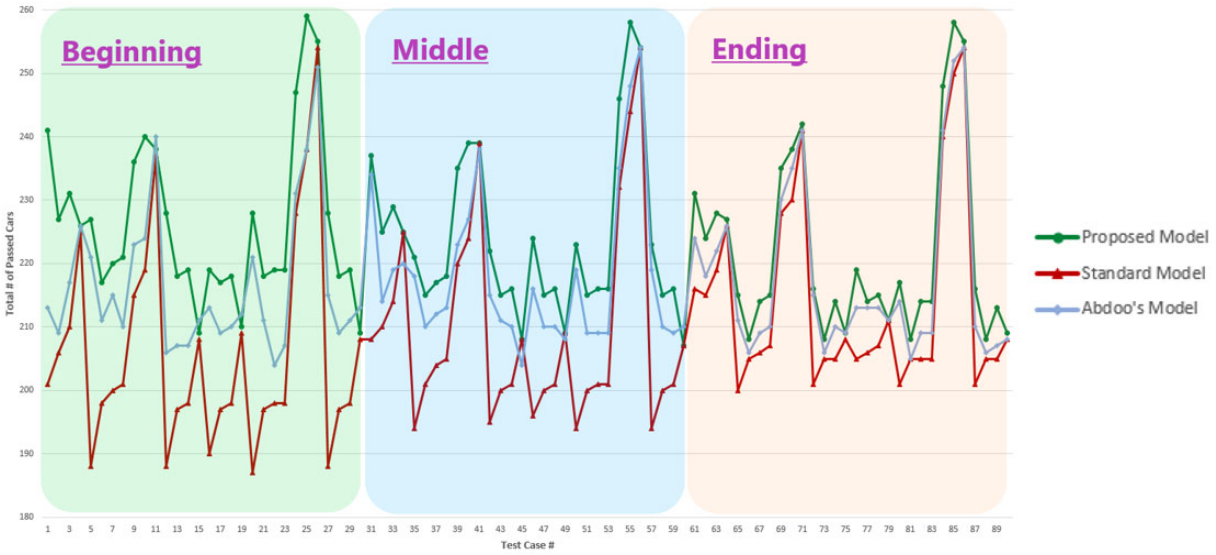

Figure 15. Interrupt outputs in proposed model

When we look at the results in-depth, just adding two options in an interrupt event afterward makes a significant difference in traffic flow performance. Once it is not bound to one option, some sort of intelligent learning mechanism can predict a better way to overcome interrupted recovery issues with the help of the proposed method. We can see that especially when the interrupt event occurs at the beginnings of phases in the higher traffic density lines, the model tends to choose avg or max actions more which is a new option for the existing context. When it is thought about that behavior because most of the waiting cars are passed during the clearance period and other lines have waiting queues, it is understandable that passing cars are considered an effective implementation for the intersection by the model. If we can give more time to a high-density road line, total throughput in the intersection could be better, which can be deduced from the simulation results. Also, if the interrupt occurs at the end of the phase or in lines with less traffic density, the model mostly chooses min value action. Because demand is low and other higher density phases are in order, this is also understandably correct.

In summary, an intelligent context-aware traffic light management system can learn and predict good options. Therefore, the proposed model suggests that different and better actions for recovering from interrupting situations can occur to benefit traffic flow. This model can also make suggestions for different emergency lines so that configuration for that intersection may vary. Hence, for every simulation result, we are confident that our model at least predicts the known-best option-min value transition or better as in avg and max action.

5. CONCLUSIONS AND FUTURE WORKS

Traffic flow has emergency issues which then cause unexpected interrupt actions. Managing these kinds of situations is crucial, especially in terms of responsiveness. Also, recovering from these actions is a real essential to plan. In this paper, we proposed a novel approach for recovering an emergency vehicle pass afterward scenario. Therefore, this paper proposes a novel approach for recovering from an emergency situation at an intersection based on real scenarios. The proposed method is a combination of context-aware and Reinforcement Learning (RL) models that predicts better alternatives for different states rather than just iterating to the next phase. We proposed that there can be more than one option for what to do after an emergency vehicle passes from an intersection, and we also modeled it as an interrupt recovery algorithm. Q-learning used as a model and test results for one intersection show that different approaches make traffic flow better in some cases. In our model, we defined categorical values for both interrupt event timing and action. Increasing categorical levels or even using numeric values as intervals will make the model more efficient and compatible with real-life scenarios. Categorical values are sufficient in order to examine model suggestions, but they could be expanded with numerical representations of state and action. Our findings show that there can be better options to follow after the emergency issue is held. As shown in simulation results, in some cases using some of the remaining red time may help increase traffic flow performance. For simplification purposes, we used a single intersection and fixed time cycle for time management in this paper. More complex and connected intersection scenarios may differ and offer some other potential aspects of the situation. Especially for connected intersections, with routing definitions and approaches, our model can make a more significant difference. Therefore, RL-based mechanisms may be good solutions for this kind of problem as well. In this study, parameters such as fuel and time will be used more efficiently. In addition, it will be possible to work in an adaptable way to adverse situations of environments. At the same time, early interventions can also be made for patients transported by ambulances. Therefore, the proposed model can contribute to traffic management in smart cities and green communications.

Our limitations in the proposed model are the usage of categorical values and testing on a single intersection scenario. As stated in the text, our state and action counts were assumed to be 3, but in a real-world scenario, this can be much more. Since recovery from an emergency situation is a relatively new field of intelligent traffic light management, we focused on structuring the proposed model so that the idea of a better choice for recovery is a potential research topic. Naturally, an emergency issue could affect multiple intersections. A vehicle should reach the event site probably passing through multiple intersections. In this study, we provided a novel and generic solution to emergencies afterward however multiple-connected intersection scenarios also could be listed as limitations. Aside from these limitations, we are confident that emergency recovery options should be considered as an improvement subject within ITS.

For future works, we plan to extend model abilities for multiple intersections. We assume that in the multi-intersection scenario, an interrupt algorithm will affect traffic flow better because the route of an emergency vehicle includes more than one line and recovering for these lines could enhance throughput. An emergency vehicle might be in need to pass several intersections in real-world scenarios, so enhancing the proposed model for a multi-agent environment is a good objective. Also integrating a continuous action event set also should be considered as future work. With the help of multi-agent methods, the proposed model can be extended to a real-life problem solving-method and implemented in Traffic Light Management (TLM) systems.

SUPPLEMENTARY FILES

The supplementary file of this study includes the datasets and is accessible at https://github.com/ofsarac/bayrampasatrafficdata

REFERENCES

[1] Bakker, B., Whiteson, S., Kester, L.J., & Groen, F.C. (2010). Traffic Light Control by Multiagent Reinforcement Learning Systems. Interactive Collaborative Information Systems.

[2] Traffic Signal Preemption for Emergency Vehicles: A Cross- Cutting Study, FHWA-JPO-05-010, a report by U.S. Department of Transportation, January 2006.

[3] Balaji, P., German, X., & Srinivasan, D. (2010). Urban traffic signal control using reinforcement learning agents. Iet Intelligent Transport Systems, 4, 177-188.

[4] Abdulhai, B., Pringle, R., & Karakoulas, G.J. (2003). Reinforcement learning for true adaptive traffic signal control. Journal of Transportation Engineering-asce, 129, 278-285.

[5] Wiering, M. (2000). Multi-Agent Reinforcement Learning for Traffic Light Control. 1151-1158.

[6] El-Tantawy, S., & Abdulhai, B. (2011). Comprehensive Analysis of Reinforcement Learning Methods and Parameters for Adaptive Traffic Signal Control. Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction. MIT press.

[7] El-Tantawy, S., & Abdulhai, B. (2012). Multi-Agent Reinforcement Learning for Integrated Network of Adaptive Traffic Signal Controllers (MARLIN-ATSC). 2012 15th International IEEE Conference on Intelligent Transportation Systems, 319-326.

[8] Haydari, A., & Yılmaz, Y. (2022). Deep Reinforcement Learning for Intelligent Transportation Systems: A Survey, IEEE Transactions on Intelligent Transportation Systems, 23(1), 11-32.

[9] Bazzan, A.L.C. Opportunities for multiagent systems and multiagent reinforcement learning in traffic control. Auton Agent Multi-Agent Syst 18, 342 (2009). https://doi.org/10.1007/s10458-008-9062-9.

[10] Abdulhai, B. & Kattan, L. (2011). Reinforcement learning: Introduction to theory and potential for transport applications. Canadian Journal of Civil Engineering. 30. 981-991. 10.1139/l03-014.

[11] Kok-Lim A.Y., Qadir, K., Khoo, H.L., Ling, M.H. & Komisarczuk, P. (2018). "A Survey on Reinforcement Learning Models and Algorithms for Traffic Signal Control". ACM Computing Survey, 50(3), 34, 1-38.

[12] Touhbi S., Babram M.A., Nguyen-Huu T., Marilleau N., Hbid M.L., Cambier C., Stinckwich S. Adaptive Traffic Signal Control: Exploring Reward Definition for Reinforcement Learning, 2017

[13] Araghi, S., Khosravi, A. & Creighton, D. (2015). "A review on computational intelligence methods for controlling traffic signal timin", Expert Systems with Applications, 42(3), 1-13.

[14] Wen-xue, C., & Zihui, Z. (2010). Path selection evaluation for emergency transport vehicles. 2010 International Conference on Logistics Systems and Intelligent Management (ICLSIM), 2, 1050-1054.

[15] Linders, J. (1989). The use of structured digital road network data bases for dispatching and routeing of emergency service. Conference Record of papers presented at the First Vehicle Navigation and Information Systems Conference (VNIS '89), A54-A59.

[16] Al-Ostath, N., Selityn, F., Al-Roudhan, Z., & El-Abd, M. (2015). Implementation of an emergency vehicle to traffic lights communication system. 2015 7th International Conference on New Technologies, Mobility and Security (NTMS), 1-5.

[17] Almuraykhi, K.M., & Akhlaq, M. (2019). STLS: Smart Traffic Lights System for Emergency Response Vehicles. 2019 International Conference on Computer and Information Sciences (ICCIS), 1-6.

[18] Moroi, Y., & Takami, K. (2015). A method of securing priority-use routes for emergency vehicles using inter-vehicle and vehicle-road communication. 2015 7th International Conference on New Technologies, Mobility and Security (NTMS), 1-5.

[19] Oliveira, L. F. P., Manera L. T., & Luz, P. D. (2021). Development of a Smart Traffic Light Control System with Real-Time Monitoring," in IEEE Internet of Things Journal, vol. 8, no. 5, pp. 3384-3393, doi: 10.1109/JIOT.2020.3022392.

[20] Arel, I., Liu, C., Urbanik, T., & Kohls, A.G. (2010). Reinforcement learning-based multi-agent system for network traffic signal control.

[21] Kiani, F. (2016). A novel channel allocation method for time synchronization in wireless sensor networks. Int. J. Numer. Model. Electron. Netw. Devices Fields 29, 805–816.

[22] Ye, Q., Song, J., Yang, Z., & Wang, L. (2011). Emergency vehicle location model and algorithm under uncertainty. 2011 2nd IEEE International Conference on Emergency Management and Management Sciences, 1-4.

[23] Palle, S., Vibha, H., Sriraksha, B.M., & Yeshashwini, A. (2019). Implementation of Smart Movable Road Divider and Ambulance Clearance using IoT. 2019 4th International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT), 1345-1348.

[24] Seyyedabbasi A, Dogan G, Kiani F (2020) HEEL: a new clustering method to improve wireless sensor network lifetime. IET Wirel Sens Syst 10:130–136. https://doi.org/10.1049/IET-WSS. 2019.0153

[25] Mouhcine, E., Karouani, Y., Mansouri, K., & Mohamed, Y. (2018). Toward a distributed strategy for emergency ambulance routing problem. 2018 4th International Conference on Optimization and Applications (ICOA), 1-4.

[26] Feroz, B., Mehmood, A., Maryam, H., Zeadally, S., Maple, C., & Shah, M.A. (2021). Vehicle-Life Interaction in Fog-Enabled Smart Connected and Autonomous Vehicles. IEEE Access, 9, 7402-7420.

[27] Li, B., Zhang, Y., Jia, N., Zhou, C., Ge, Y., Liu, H., Meng, W., & Ji, C. (2017). Paving green passage for emergency vehicle in heavy traffic: Real-time motion planning under the connected and automated vehicles environment. 2017 IEEE International Symposium on Safety, Security and Rescue Robotics (SSRR), 153-158.

[28] Colorado, L.A., Ibañez, J.F., & Martinez-Santos, J.C. (2020). Leveraging Emergency Response System Using the Internet of Things. A Preliminary Approach. 2020 17th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE), 1-6.

[29] Shamsi, Mahboubeh & Rasouli Kenari, Abdolreza & Aghamohammadi, Roghayeh. (2021). Reinforcement learning for traffic light control with emphasis on emergency vehicles. The Journal of Supercomputing. 10.1007/s11227-021-04068-w.

[30] Louati, Ali & Louati, Hassen & Li, Zhaojian. (2021). Deep learning and case-based reasoning for predictive and adaptive traffic emergency management. The Journal of Supercomputing. 77. 10.1007/s11227-020-03435-3.

[31] Nama, Mahima & Nath, Ankita & Bechra, Nancy & Bhatia, Jitendra & Tanwar, Sudeep & Chaturvedi, Manish & Sadoun, Balqies. (2021). Machine Learning-based Traffic Scheduling Techniques for Intelligent Transportation System: Opportunities and Challenges. International Journal of Communication Systems. 34. 10.1002/dac.4814.

[32] Ramazani A. & Vahdat-Nejad H. (2017). CANS: context-aware traffic estimation and navigation system. IET Intelligent Transport Systems, 11(6), pp.326-333.

[33] Yang. J. et al. (2021). Automatic generation of optimal road trajectory for the rescue vehicle in case of emergency on mountain freeway using reinforcement learning approach. IET Intelligent Transport Systems, 15, pp.1142-1152.

[34] Araghi, S., Khosravi, A., Johnstone, M., & Creighton, D.C. (2013). Q-learning method for controlling traffic signal phase time in a single intersection. 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013), 1261-1265.

[35] Seyyedabbasi, A., Aliyev, R., Kiani, F., Gulle, M. U., Basyildiz, H., & Shah, M. A. (2021). Hybrid algorithms based on combining reinforcement learning and metaheuristic methods to solve global optimization problems. Knowledge-Based Systems, 223, 107044.

[36] Sutton, R.S., & Barto, A.G. (2018). Reinforcement Learning: An Introduction, second ed., The MIT Press, Cambridge, Massachusetts, London, England, 119-140.

[37] Wei, H., Zheng, G., Gayah, V.V., & Li, Z.J. (2019). A Survey on Traffic Signal Control Methods. ArXiv, abs/1904.08117.

[38] Gunes, F., Bayrakli, S., Zaim A. H. (2021). Smart Cities and Data Analytics for Intelligent Transportation Systems: An Analytical Model for Scheduling Phases and Traffic Lights at Signalized Intersections. Appl. Sci. 2021 (11), 6816.

[39] Abdoos, M., Mozayani, N., Bazzan, A. (2011). Traffic Light Control in Non-stationary Environments based on Multi Agent Q-learning. IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC. 10.1109/ITSC.2011.6083114.