Shapley Value-based Congestion Attribution: A Practical Multiagent Reinforcement Learning for Traffic Signal Control

Yixuan Li[1,†], Jiajun Li[1,†], Xiao Liu[1], Weiwei Wu[1], Wanyuan Wang[1,*]

***Abstract*— Intelligent traffic signal control is a crucial research area in automation and artificial intelligence. By modeling traffic signals as agents, Multi-Agent Reinforcement Learning (MARL) enables coordinated control of traffic signals to mitigate congestion. However, most MARL methods rely on synchronous policy updates, facing challenges in scalability and non-stationarity in large-scale traffic networks. Asynchronous updates are proposed to offer flexibility and scalability but can suffer from instability due to improper update orders. To address these limitations, we propose a congestion attribution-guided MARL optimization framework, prioritizing signal control policies based on their contribution to congestion. This framework is applicable to both synchronous and asynchronous updates, optimizing MARL training through *Partial Control* and *Sequential Control*, respectively. We introduce the Shapley value to quantify the impact of signal control policies on congestion, providing an interpretable attribution mechanism. Experimental results demonstrate the effectiveness of Shapley value-based congestion attribution and demonstrate the superior performance of our framework in enhancing both training efficiency and overall traffic efficiency compared to baseline methods.**

I. Introduction

Intelligent Transportation Systems (ITS) represent a pivotal research area at the intersection of artificial intelligence, automation, and robotics [1], [2], with traffic signal control serving as a critical component. Effective signal control, capable of dynamically optimizing vehicle behavior at intersections based on real-time traffic information, plays a key role in alleviating urban congestion [3] and has broader implications for intelligent multi-robot systems [4], [5].

With the advancement of AI technology and the availability of traffic big data, data-driven learning methods have demonstrated their potential to significantly improve road network efficiency without any manual intervention. Of particular interest is the multi-agent reinforcement learning (MARL) approach [6], which models traffic signals as agents and leverages the real-time interactions between intersections to achieve cooperative optimization of the overall traffic flow.

In terms of policy update mechanisms during training, most methods adopt synchronous updates, where all agents update their policies simultaneously [7], [8]. However, as the number of traffic signals increases, high computational complexity and poor scalability can be encountered when centralized training is employed [9], [10], [11]. In addition, synchronous updates may also introduce the non-stationarity problem as the system scales in complexity [12]. Asynchronous updates allow agents to update policies independently, enhancing flexibility and scalability. However, Wang et al. point out that incorrect update order can degrade the performance of asynchronous updates, leading to instability and slower convergence [13]. Current research on update orders mainly relies on random, expert experience, or greedy methods [14], making them less suitable for complex and dynamic traffic scenarios.

To address these limitations, we propose an attribution-guided MARL optimization framework that updates control policies based on the contribution of traffic signal policies to congestion. For synchronous updates, we propose *Partial Control Optimization* to select the most critical subset of traffic signals for MARL control, maintaining a smaller number of agents to enhance efficiency and mitigate the drawbacks of synchronous updates in large-scale networks. For asynchronous updates, we propose *Sequential Control Optimization* to guide the agent update order based on their contribution to congestion, thereby enhancing training stability and promoting monotonic improvement.

Current traffic congestion attribution analysis mainly focuses on identifying the root cause at a specific intersection through expert experience [15], heuristics [16], or AI methods [17], neglecting the cooperative nature of intersections and lacking quantitative analysis of how signal policies influence congestion. Moreover, data-driven attribution methods often suffer from poor interpretability. Shapley value decomposition from game theory [18], provides a fair estimation of each player's contribution [19] and has been proven to be the only possible explanation model that satisfies all desirable properties of feature attribution methods [20]. Therefore, we introduce the Shapley value to measure the contribution of signal policies to congestion and employ Monte Carlo sampling to comprehensively consider the traffic condition and improve computational efficiency.

In summary, our contributions are as follows:

1) We propose to leverage the Shapley value for congestion attribution, providing an interpretable and quantitative assessment of the impact of signal policies on traffic congestion.
2) We design a general attribution-guided MARL optimization framework that leverages priority to optimize both synchronous and asynchronous update strategies.
3) Through experiments, we validate the effectiveness of the Shapley value for congestion attribution and demonstrate that our optimization framework outperforms baselines in enhancing overall traffic efficiency.

[†]These authors contributed equally. *Corresponding Author.

[1]The authors are with Southeast University, Nanjing, China. {yixuanli,jiajun_li,xliu0206,weiweiwu,wywang}@seu.edu.cn

II. Related Works

A. Traffic Signal Control Methods

Traditional traffic signal control methods, such as fixed-time control [21], rely on pre-defined phase sequences and timing plans, making them inflexible to real-time traffic fluctuations. While adaptive control methods, including fuzzy logic control [22], genetic algorithms [23], and model predictive control (MPC) [24], [25], [26], can dynamically adjust signal timings based on real-time traffic data. However, they often rely on pre-defined rules or models and struggle to handle complex and unpredictable traffic scenarios. In contrast, reinforcement learning (RL) methods do not require pre-defined rules and can learn optimal policies directly from interactions with the environment, offering greater adaptability and robustness [27], [28], [29]. Moreover, as traffic systems involve multiple intersections requiring coordinated control, multi-agent reinforcement learning (MARL) can leverage the complex interactions between intersections to optimize overall traffic flow [30], [31], [32], [33], [34]. However, few existing methods differentiate the priority of traffic signals based on their contribution to congestion, nor do they offer targeted optimization strategies based on this priority. In addition, while the Shapley value has been employed in some MARL approaches [35], [36], its application primarily focuses on decomposing and optimizing reward functions, rather than explicitly guiding the design and optimization of the macroscopic MARL framework.

B. Traffic Congestion Attribution Methods

Traffic congestion attribution analysis aims to identify the various factors contributing to traffic congestion [37]. Our focus lies in the quantitative analysis of the responsibility for congestion within the existing road network structure. Traditional methods rely on qualitative analysis combined with expert experience [15]. With the widespread adoption of technologies such as traffic big data collection, more research has shifted towards data-driven approaches, utilizing heuristic methods [16] or artificial intelligence methods [17], [38] to analyze the causes of traffic congestion. However, current AI-based attribution methods lack interpretability and fail to adequately account for the cooperative relationships among traffic signals.

C. Policy Update Mechanisms in MARL

Multi-agent reinforcement learning can be categorized into synchronous and asynchronous updates based on the policy update mechanism during training. Most methods adopt synchronous updates, where all agents update their policies simultaneously [7], [8]. However, as the system scales in complexity, this approach prevents agents from observing the changes of others and introduces the non-stationarity problem, i.e., the environment dynamics change from one agent's perspective as other agents also adapt their policies [12], [39]. Asynchronous updates allow agents to update policies independently, enhancing flexibility, but Wang et al. point out that incorrect update order can degrade the performance of asynchronous updates, leading to instability and slower convergence [13]. However, current research on update orders mainly relies on expert experience or greedy methods [14]. Our proposed method aims to optimize both synchronous and asynchronous MARL update mechanisms by addressing agent selection and policy update order, respectively.

III. Model and Problem Formulation

A. Problem Definition and Introduction

This paper addresses the problem of multi-intersection traffic signal control, aiming to reduce congestion in urban road networks. To clarify the research object, we first define the relevant concepts:

Road network. The intersections in the road network are denoted as $\mathbf{I} = \{I_i\}_{i=1}^{n}$, and a road $R_{i,j} \in R$ is an edge connecting intersections I_i and I_j. Each road consists of multiple branch lanes l_i.

Traffic movement. Traffic movement (l_i, l_j) is defined as the movement of vehicles from lane l_i to lane l_j.

signal phase. The signal phase SP_i represents a group of allowable traffic movements at intersection I_i.

Vehicle Density. The vehicle density of a lane is defined as $\frac{x(l)}{x_{\max}(l)}$, where $x(l)$ is the actual number of vehicles in lane l, and $x_{\max}(l)$ is the maximum allowable number.

Traffic Pressure. The pressure of a movement is defined as the difference of vehicle density between the incoming lane and the outgoing lane:

$$w(l_i, l_j) = \frac{x(l_i)}{x_{\max}(l_i)} - \frac{x(l_j)}{x_{\max}(l_j)}, \tag{1}$$

the pressure of intersection ρ_i is defined as the sum of the absolute values of all traffic flow pressures at the intersection I_i, denoted as:

$$\rho_i = |\sum_{(l_i,l_j)\in i} w(l_i, l_j)|. \tag{2}$$

B. Multi-agent Modeling

Given a road network consisting of multiple intersections $\mathbf{I} = \{I_i\}_{i=1}^{n}$, we define the multi-agent traffic signal control problem as a Decentralized Partially Observable Markov Decision Process (Dec-POMDP). This problem can be represented by a tuple $\langle S, A, P, r, N, \gamma \rangle$. The environment's state is denoted as S, and each agent $i \in N := 1, 2, \ldots, N$ independently selects an action $a^i \in A^i$ transitioning the state s to a new state s', based on the transition function $P(s' \mid s, a) : S \times A^N \times S \to [0, 1]$. The environment provides a global reward, defined by the reward function $r(s, a) : S \times A^N \to \mathbb{R}$ and the ultimate goal of all agents is to maximize the long-term cumulative reward $\sum_{i=0}^{\infty} \gamma^i r(s^i, a^i)$, where $\gamma \in [0, 1]$ is the discount factor. Each agent independently adopts a signal control policy $\{\pi^1, \ldots, \pi^n\}$. We define the state value function and the state-action value function: $V_\pi(s) = \mathbb{E}[\sum_{t=0}^{\infty} \gamma^t r_t \mid s_0 = s]$ and $Q_\pi(s, a) = \mathbb{E}[\sum_{t=0}^{\infty} \gamma^t r_t \mid s_0 = s, a_0 = a]$. The advantage function can be denoted as $A_\pi(s, a) = Q_\pi(s, a) - V_\pi(s)$.

State. At time step $t \in \mathbb{N}$, each agent i receives partial observations s_t^i, which include the vehicle density of the

incoming and outgoing lanes, and the current traffic signal phase SP_t^i at intersection i. The global state $S_t = \prod_{i=1}^{n} s_t^i$ is obtained by aggregating the partial observations from all agents.

Actions. The signal control decision-making interval is set to 10 seconds, meaning that every 10 time steps, each agent selects a signal phase SP_t^i as its action a_t^i. The joint action of all agents is denoted as $u_t = \{a_t^1, a_t^2, \ldots, a_t^n\}$.

Reward. Intersection pressure The imbalance between the incoming and outgoing lanes reflects, thus causing traffic congestion. The reward function is designed based on the intersection pressure, and the overall reward is computed as the sum of the local rewards for each agent:

$$R = \sum_{i=1}^{n} \rho^i. \tag{3}$$

IV. Shapley Value-Based Analysis of Traffic Signal Contributions to Congestion

In this section, we propose a congestion attribution method based on Shapley value decomposition, aiming to quantify the contribution of each traffic signal to the potential congestion in the road network. This approach enables us to gain a deeper understanding of the impact of individual signals on traffic congestion, providing a basis for subsequent optimization strategies. To calculate the Shapley value, we model the traffic signal control system as a cooperative game, where individual signal agents work together to maximize traffic efficiency. Subsequently, we derive the Shapley value calculation formula for each traffic signal and employ Monte Carlo sampling to comprehensively consider various traffic flow scenarios and improve computational efficiency.

A. Game Modeling and Shapley Value Decomposition

Specifically, we model the traffic signal control system as a cooperative game $G = (N, v)$, where the set of players $N = \{1, 2, ..., n\}$ represents the n traffic signal agents in the system. The characteristic function $v : 2^N \to \mathbb{R}$ denotes the traffic efficiency achievable by any subset of traffic signals $U \subseteq N$ working together. It is defined as the average travel time of the entire road network when the traffic signals in set U adopt their respective phase control strategies. The objective of the game is to minimize $v(U)$.

To compute the Shapley value, we need to measure the marginal contribution of each intersection to the cooperative control. In the context of intelligent signal control, this marginal contribution can be calculated by assessing the change in system congestion level before and after an intersection participates in the cooperative control. Specifically, ϕ_i quantifies the contribution of traffic signal agent i to the overall traffic efficiency:

$$\phi_i = \sum_{U \subseteq N \setminus \{i\}} \frac{|U|!(n - |U| - 1)!}{n!} [v(U \cup \{i\}) - v(U)], \tag{4}$$

where $U \subseteq N \setminus \{i\}$ represents the subset of traffic signals that does not include agent i, and $|U|$ denotes the number of elements in set U. $v(U \cup \{i\}) - v(U)$ represents the improvement in traffic efficiency (i.e., the reduction in congestion) brought about by agent i joining set U. $\frac{|U|!(n-|U|-1)!}{n!}$ represents the probability of set U appearing in all possible combinations of traffic signals.

B. Estimation via Monte Carlo Sampling

To conduct a comprehensive analysis of each traffic system, we need to consider the marginal contributions of intersection agents under various traffic flow conditions. However, directly computing the Shapley value requires traversing all possible coalition combinations, which involves a large joint space and incurs high computational costs[40]. Therefore, we employ Monte Carlo sampling to consider potential traffic flow states and sample different coalition states for each traffic flow state to calculate the comprehensive Shapley value for each intersection. The calculation formula is as follows:

$$\hat{\phi}_i = \frac{1}{M} \sum_{m=1}^{M} \frac{1}{K} \sum_{k=1}^{K} [v(U_{m,k} \cup \{i\}) - v(U_{m,k})], \tag{5}$$

where $\hat{\phi}_i$ is the estimated Shapley value of agent i (i.e., a traffic signal at an intersection). M is the number of traffic flow states sampled using Monte Carlo. K is the number of coalition states sampled for each traffic flow state. $U_{m,k}$ is the k-th sampled coalition state (excluding agent i) under the m-th traffic flow state. $v(U_{m,k} \cup \{i\}) - v(U_{m,k})$ is the improvement in traffic efficiency (i.e., the reduction in congestion) brought about by agent i joining coalition $U_{m,k}$.

We jointly train all traffic signal agents to convergence by MARL to estimate the cooperative phase strategies for each state. When an agent exits a coalition, the corresponding traffic signal reverts to its original phase control strategy.

V. Congestion Attribution-Guided MARL Optimization Framework

From the perspective of policy update mechanisms, synchronous updates suffer from reduced training efficiency and scalability under centralized training. In addition, synchronous updates hinder agents from observing each other's changes and exacerbate the non-stationarity problem as the system scales in complexity. Asynchronous updates improve efficiency and scalability, but the incorrect update order may lead to training instability due to inconsistent policies. To address these limitations, we propose an attribution-guided MARL optimization framework. First, we utilize the Shapley value to assess the congestion responsibility of each intersection in a given traffic system. Then, we integrate this priority of responsibility with synchronous/asynchronous update strategies to guide the MARL training process. Specifically, we introduce two optimization approaches:

1) *Partial Synchronous Control Optimization:* Inspired by previous research [41], [42], the hybrid control method are effective and widely applied in the automated systems. We leverage the Shapley value to select the top-k intersections with the greatest impact on congestion for multi-agent modeling, and apply reinforcement learning control only to these critical traffic signals. This approach maintains the

advantages of synchronous updates under applicable scale while reducing the state-action space dimension and training overhead, making it more suitable for large-scale traffic networks. Furthermore, partial control facilitates integration with traditional signal control methods, enhancing the feasibility of real-world deployment.

2) *Sequential Asynchronous Control Optimization:* Building upon asynchronous update strategies, we determine the order of agent policy updates based on the Shapley value. This allows the important agents to have priority in updating their policies, and their policy changes and impact on the environment can provide a better learning environment for other agents, thereby accelerating the overall learning process and improving algorithm stability and performance. Previous research [14], [13] has indicated that improper update order can lead to algorithm instability, whereas our method offers a more reasonable update order guidance through the Shapley value.

Note that both of the aforementioned optimization approaches can be combined with any MARL algorithm, no matter whether the training is centralized or distributed. Moreover, the introduction of the Shapley value provides interpretability to the optimization process, aiding in understanding the impact of each traffic signal on traffic congestion.

A. Partial Synchronous Control

In this section, we propose a partially synchronous control optimization method based on Shapley value (MATCS), which aims to leverage the advantages of policy synchronization while overcoming scalability and non-stationarity issues by scaling down the control number.

Specifically, we utilize the Shapley value to select the top-k agents for control optimization, thereby reducing training overhead. MATCS trains a global value function (critic network) for all agents, optimizing inter-agent coordination in a centralized manner. The critic network receives observation information from all agents, while each agent independently receives local observations and generates action probabilities through a shared actor network.

In this algorithm, we define that $i_{1:k} \subseteq \mathbf{I} = \text{top-}k(\{\hat{\phi}_1, ..., \hat{\phi}_n\})$. Each agent $i \in i_{1:k}$ acts according to the shared policy and generates an individual trajectory $\tau_i = \left\{s_t^i, a_t^i, r_t^i\right\}_{t=1}^T$, where r_t^i is the partial reward value for agent i. We use θ and ϕ to represent all parameters of the actor network and the critic network, respectively.

For the Actor network, the shared policy π_θ is updated by maximizing the following objective function:

$$J(\theta) = \mathbb{E}\left[\min\left(l_t^i(\theta)\hat{A}_t^i, \text{clip}\left(l_t^i(\theta), 1-\varepsilon, 1+\varepsilon\right)\hat{A}_t^i\right)\right]. \quad (6)$$

Specifically, ε is a manually set clipping factor and $l_t^i(\theta) = \frac{\pi_\theta(a_t^i|s_t^i)}{\pi_{\theta_{\text{old}}}(a_t^i|s_t^i)}$, where $\pi_{\theta_{\text{old}}}$ refers to the policy used for interacting with the environment, and π_θ represents the policy to be optimized during the current learning iteration. $\hat{A}_t^i = \sum_{l=0}^h (\gamma\lambda)^l \delta_{t+l}^i$ is the GAE (Generalized Advantage Estimator) based on the current value function $\hat{V}_\phi$, with the temporal difference defined as $\delta_t = r_t + \gamma V_\phi(s_{t+1}) - V_\phi(s_t)$ where γ is the discount factor and λ is the smooth factor.

For the Critic network, we update the parameters by minimizing the following loss function:

$$L(\phi) = \mathbb{E}\left[\left(V_\phi(s_t) - R_t\right)^2\right], \quad (7)$$

where $R_t = \sum_{t'=t}^T \gamma^{t'-t} r_{t'}$.

B. Sequential Asynchronous Control

This section describes a sequential asynchronous control optimization method based on Shapley values (SeTCS), aiming to leverage the advantages of asynchronous policy updates while optimizing the update sequence to improve algorithm stability and performance. To achieve this, we leverage Shapley values to guide the update sequence.

Specifically, in SeTCS, each agent operates with an independent actor policy network, while a shared global critic network is used to evaluate the overall system. The update sequence of each agent's policy is guided by Shapley values, iteratively improving the policies of individual agents.

Firstly, we introduce the multi-agent advantage decomposition lemma [14], denoted as:

$$A_\pi^{i_{1:m}}\left(s, a^{i_{1:m}}\right) = \sum_{j=1}^m A_\pi^{i_j}\left(s, a^{i_{1:j-1}}, a^{i_{1:j}}\right), \quad (8)$$

where $A_\pi^{i_{1:m}}$ written as $Q_\pi^{i_{1:m}}(s, a^{i_{1:m}}) - V_\pi(s)$ is the multi-agent advantage function of any subset $i_{1:m} = \{i_1, i_2, \ldots, i_m\}$. This lemma demonstrates that during the sequential update process, the joint advantage function can be decomposed into the sum of the local advantages of each agent. In other words, if we can maximize $A_\pi^{i_j}$ each term, then the maximum $A_\pi^{i_{1:m}}$ can be achieved.

Then, we denote the updated policy of agent i as $\overline{\pi}^i$ and while updating the agent i the joint policy can be denoted as:

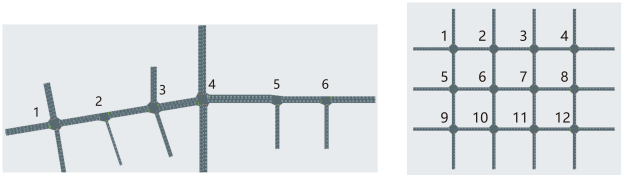
$$\hat{\pi}^i = \overline{\pi}^1 \times \cdots \times \overline{\pi}^i \times \pi^{i+1} \times \cdots \times \pi^n.$$

Inspired by [13], to avoid the distribution shift caused by preceding agents, we adopt a method of policy refinement for preceding agents, using samples collected from the joint policy π to approximate $\hat{A}^{\pi^{i-1}}$:

$$A^{\pi,\hat{\pi}^{i-1}}(s_t, a_t) = \delta_t + \sum_{k\geq 1}\left(\prod_{j=1}^k \lambda M_{t+j}^{i-1}\right)\delta_{t+k}, \quad (9)$$

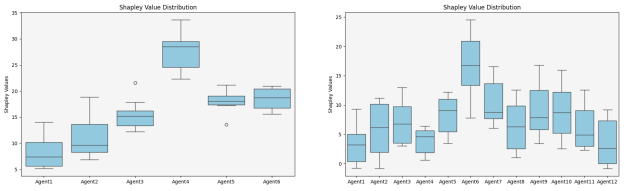
where $M_{t+j}^{i-1} = \min\left(1.0, \frac{\hat{\pi}^{i-1}(a_{t+j}|s_{t+j})}{\pi(a_{t+j}|s_{t+j})}\right)$ and $\delta_t = r(s_t, a_t) + \gamma V(s_{t+1}) - V(s_t)$.

Through this formula, we can find that during the optimization process, since the policies except π^i are fixed, updating agent with a bigger absolute value of the advantage function contributes more. For agents with greater responsibility (Shapley value) to the system, we can know that their contribution to the optimization updates is also greater. Therefore, we determine the agent selection rule based on the Shapley value. Specifically, we rank the agents based on



(a) roadnet_1 (b) roadnet_2

Fig. 1. The road network



(a) roadnet_1 (b) roadnet_2

Fig. 2. The Shapley value on the two traffic scenarios

the Shapley value, denoted as $i_s = \{i_{k_1}, i_{k_2}, \ldots, i_{k_n}\}$, and our sequential update scheme can be denoted as:

$$\hat{\pi}^{k_i} = \overline{\pi}^{k_1} \times \cdots \times \overline{\pi}^{k_i} \times \pi^{k_{i+1}} \times \cdots \times \pi^{k_n},$$

$$\pi = \hat{\pi}^{k_0} \to \hat{\pi}^{k_1} \to \cdots \to \hat{\pi}^{k_n} = \overline{\pi}.$$

Therefore, for the agent $i \in i_s$ our goal is to maximise the clipping objective of

$$L_{\hat{\pi}^{i-1}}(\theta) = \mathbb{E}\left[\min\left(l(s,a)A^{\pi}(s,a), \right.\right. \\ \left.\left. \text{clip}\left(l(s,a), 1 \pm \varepsilon^i\right) A^{\pi,\hat{\pi}^{i-1}}\right)\right], \quad (10)$$

where $l(s,a) = \frac{\overline{\pi}^i(a^i|s)}{\pi^i(a^i|s)}$.

It is worth noting that, according to the multi-agent advantage decomposition lemma [14], the idea of guiding agent policy update sequences based on Shapley values is not only limited to the specific algorithm we propose, but also can be extended to other sequential asynchronous MARL methods.

VI. Experiments

In this section, we first validate the effectiveness of the Shapley value in assessing the contribution of intersection signals to congestion levels. Subsequently, we evaluate the performance of our proposed attribution-based multi-agent reinforcement learning optimization framework under both partial control and sequential control scenarios.

A. Experimental Setup

We utilize CityFlow [43] as our experimental platform. The road networks include an irregular road network, roadnet_1, constructed based on a real-world regional road network, and a regular grid road network, roadnet_2. Each road comprises three lanes for left turns, straight-through traffic, and right turns, as illustrated in Fig. 1. The baseline methods for comparison include:

- Presslight [44]: A DQN-based MARL control approach with pressure function rewards.
- IDQL: A MARL method combining DQN and policy learning with synchronous policy updates.
- IPPO [7]: A distributed PPO-based MARL method.
- HAPPO [14]: PPO-based MARL method with asynchronous policy updates in random order.

The evaluation metrics for the algorithms include the average queue length of lanes (i.e., the number of vehicles waiting in a lane, denoted as Queue) and the average travel time of vehicles (Travel Time). The traffic congestion can be evaluated by the Traffic Congestion Rate (TCR): $\text{TCR} = \frac{1}{|R|}\sum_{i=1}^{|R|}\frac{\bar{V}_i}{V_i}$, where V_i denotes the current average speed on the i-th road, and $\bar{V}_i$ refers to the free-flow speed of vehicles on the i-th road under no external environmental constraints. Higher TCR values signify increased traffic congestion, with TCR exceeding 3.3 indicating severe congestion on the road.

B. Results on Shapley Value-Based Congestion Attribution

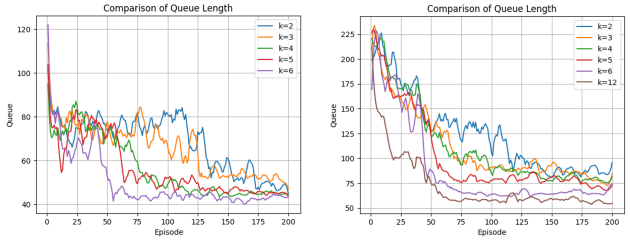
To validate the effectiveness of the Shapley value in analyzing the responsibility of traffic signals for congestion, we first perturb the phase durations of traffic signals to simulate congestion caused by suboptimal signal control strategies. Subsequently, we employ MARL to train each signal and calculate its Shapley value. By comparing the Shapley values of the congestion-causing signals with those of other signals, we aim to verify whether the Shapley value can effectively identify the signals responsible for congestion.

Specifically, the original fixed-time control strategy for the traffic signals is based on real-world signal control strategies during weekday mornings from 7:00 to 8:00. In roadnet_1, we reduce the duration of the east-west phase at intersection 4 by 30 seconds, making the average TCR increases to 4.61 to reach a severe congestion state. In roadnet_2, we increase the phase switching frequency of intersection 6 by 30 seconds, resulting in a TCR increase to 6.22. Fig. 2 presents the calculated Shapley values for both road networks.

In roadnet_1, as shown in Fig. 2(a), intersection 4 exhibits the highest average Shapley value, and its distribution is significantly higher than other intersections. The Shapley value indicates its primary responsibility for network congestion, which aligns with the experimental setup for simulating congestion. Furthermore, the Shapley values of intersections adjacent to the root cause intersection 4 are larger than those of intersections farther away (e.g., intersection 1). This is because traffic interactions between neighboring intersections necessitate adjustments in the signal phase strategies of neighboring intersections as well. This phenomenon demonstrates that the Shapley value can effectively quantify the propagation of congestion. Similar observations can be found in roadnet_2 shown in Fig. 2(b), where intersection 6 has the largest Shapley value.

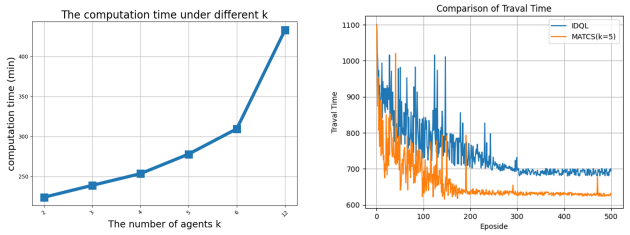
C. Results on Attribution-Guided Partial Control

In this subsection, we conduct evaluate the effectiveness of using Shapley value to guide the selection of traffic signals for partial control with synchronous policy updates.



(a) roadnet_1 (b) roadnet_2

Fig. 3. Partial synchronous control under different number of agents k



(a) Computaition time on roadnet_2 (b) Comparison on roadnet_3

Fig. 4. The performance of partial synchronous control

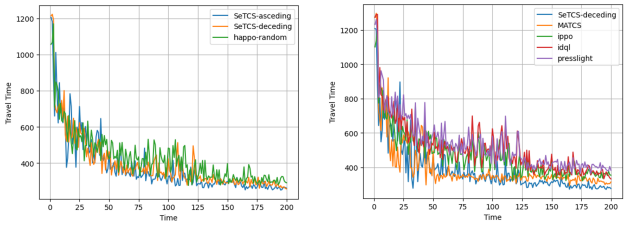
Specifically, we first calculate the Shapley values of all traffic signals under typical traffic flow scenarios and select the top-k critical intersection signals as agents for joint training in descending order of their Shapley values. Fig. 3 illustrates the performance of MARL on two road networks with varying numbers of agents. Due to the limited number of agents in roadnet_1, we only report the training convergence time for roadnet_2, as shown in Fig. 4(a).

Fig. 4(a) reveals that the training overhead required for model convergence increases with the number of agents. From Fig. 3(b) We can observe that by selecting intersection signals with higher congestion responsibility as critical agents, when $k = 6$, the model's performance is close to that achieved when all agents are trained jointly ($k = 12$), with only a 2.3% difference, while saving approximately 28.49% of training time.

Subsequently, we maintain the road network structure and expand roadnet_2 to a larger 5x5 intersection roadnet_3 scenario to test the performance in a large-scale setting. Fig. 4(b) compares the performance of our method (denoted as MATCS) with $k = 5$ agents against the IDQL method. The comparison with other signal control methods is presented in the next subsection. The experimental results demonstrate that for large-scale road networks, selecting only a subset of critical intersection signals as agents for training yields significantly better results than training all agents jointly using the IDQL method. This is because synchronous policy updates with a large number of agents can lead to increased training difficulty and potential convergence to local optima.

D. Results on Attribution-Guided Sequential Control

In this subsection, we conduct experiments to analyze the effectiveness of utilizing Shapley value to guide the agent update order in asynchronous policy updates. We first compare the performance of our method with the asynchronous



(a) Comparison on update order (b) Comparison on other baselines

Fig. 5. The performance of sequential asynchronous control

updated baseline HAPPO in different update orders. Specifically, we arrange the agents in descending order (SeTCS-descending) and ascending order (SeTCS-ascending) based on their Shapley values, and compare them with the random updated HAPPO algorithm (denoted as happo-random).

Fig. 5(a) illustrates the performance of the three algorithms. It can be observed that the random ordering method exhibits larger fluctuations in travel time, indicating instability and slightly inferior performance. Both the SeTCS-ascending and SeTCS-descending methods, based on Shapley value ordering, achieve better results. This demonstrates that prioritizing policy updates based on contribution levels benefits training stability and algorithm performance. The similar performance of ascending and descending orders suggests a potential symmetry, which we will further analyze in future work.

Fig. 5(b) presents the performance comparison of all algorithms. It is evident that both of our proposed control optimization approaches (partial control and sequential control) significantly outperform other baseline methods, with the Shapley value-guided asynchronous update method achieving the best performance. Furthermore, compared to the baseline methods, our proposed methods converge faster and exhibit stronger stability, validating the effectiveness of the congestion attribution-guided MARL framework. This framework can be combined with both synchronous and asynchronous policy update schemes to leverage their respective advantages. Considering that the Shapley values are approximated through Monte Carlo sampling, the promising results further highlight the potential of our algorithm.

VII. Conclusion

This paper optimizes the MARL-based traffic signal control methods. We propose a novel congestion attribution-based MARL optimization framework that leverages Shapley values to quantify the contribution of each traffic signal to congestion in an explainable manual. This framework guides both synchronous and asynchronous policy updates, enhancing training efficiency and stability. For synchronous updates, we introduce *Partial Control* to select critical traffic signals for partial control. For asynchronous updates, we propose *Sequential Control* to prioritize agent updates based on congestion contribution. Experimental results validate the effectiveness of Shapley values in congestion attribution and demonstrate the performance of our optimization framework in improving overall traffic efficiency.

REFERENCES

[1] Y. Lin, P. Wang, and M. Ma, "Intelligent transportation system (its): Concept, challenge and opportunity," in *2017 ieee 3rd international conference on big data security on cloud (bigdatasecurity), ieee international conference on high performance and smart computing (hpsc), and ieee international conference on intelligent data and security (ids)*. IEEE, 2017, pp. 167–172.

[2] D. S. Sarwatt, Y. Lin, J. Ding, Y. Sun, and H. Ning, "Metaverse for intelligent transportation systems (its): A comprehensive review of technologies, applications, implications, challenges and future directions," *IEEE Transactions on Intelligent Transportation Systems*, 2024.

[3] D. Zhao, Y. Dai, and Z. Zhang, "Computational intelligence in urban traffic signal control: A survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 4, pp. 485–494, 2011.

[4] J. Guo, L. Cheng, and S. Wang, "Cotv: Cooperative control for traffic light signals and connected autonomous vehicles using deep reinforcement learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 10, pp. 10 501–10 512, 2023.

[5] M. Villarreal, B. Poudel, J. Pan, and W. Li, "Mixed traffic control and coordination from pixels," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 4488–4494.

[6] Y. Liu, G. Luo, Q. Yuan, J. Li, L. Jin, B. Chen, and R. Pan, "Gplight: Grouped multi-agent reinforcement learning for large-scale traffic signal control." in *IJCAI*, 2023, pp. 199–207.

[7] C. S. De Witt, T. Gupta, D. Makoviichuk, V. Makoviychuk, P. H. Torr, M. Sun, and S. Whiteson, "Is independent learning all you need in the starcraft multi-agent challenge?" *arXiv preprint arXiv:2011.09533*, 2020.

[8] C. Yu, A. Velu, E. Vinitsky, J. Gao, Y. Wang, A. Bayen, and Y. Wu, "The surprising effectiveness of ppo in cooperative multi-agent games," *Advances in Neural Information Processing Systems*, vol. 35, pp. 24 611–24 624, 2022.

[9] T. Chu, J. Wang, L. Codecà, and Z. Li, "Multi-agent deep reinforcement learning for large-scale traffic signal control," *IEEE transactions on intelligent transportation systems*, vol. 21, no. 3, pp. 1086–1095, 2019.

[10] H. Wei, N. Xu, H. Zhang, G. Zheng, X. Zang, C. Chen, W. Zhang, Y. Zhu, K. Xu, and Z. Li, "Colight: Learning network-level cooperation for traffic signal control," in *Proceedings of the 28th ACM international conference on information and knowledge management*, 2019, pp. 1913–1922.

[11] H. Gu, S. Wang, X. Ma, D. Jia, G. Mao, E. G. Lim, and C. P. R. Wong, "Large-scale traffic signal control using constrained network partition and adaptive deep reinforcement learning," *IEEE Transactions on Intelligent Transportation Systems*, 2024.

[12] P. Hernandez-Leal, M. Kaisers, T. Baarslag, and E. M. De Cote, "A survey of learning in multiagent environments: Dealing with non-stationarity," *arXiv preprint arXiv:1707.09183*, 2017.

[13] X. Wang, Z. Tian, Z. Wan, Y. Wen, J. Wang, and W. Zhang, "Order matters: Agent-by-agent policy optimization," in *The Eleventh International Conference on Learning Representations*, 2023.

[14] J. Kuba, R. Chen, M. Wen, Y. Wen, F. Sun, J. Wang, and Y. Yang, "Trust region policy optimisation in multi-agent reinforcement learning," in *ICLR 2022-10th International Conference on Learning Representations*. The International Conference on Learning Representations (ICLR), 2022, p. 1046.

[15] S. Chawla, Y. Zheng, and J. Hu, "Inferring the root cause in road traffic anomalies," in *2012 IEEE 12th International Conference on Data Mining*. IEEE, 2012, pp. 141–150.

[16] W.-H. Lee, S.-S. Tseng, J.-L. Shieh, and H.-H. Chen, "Discovering traffic bottlenecks in an urban network by spatiotemporal data mining on location-based services," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1047–1056, 2011.

[17] Y. Chen, C. Li, W. Yue, H. Zhang, and G. Mao, "Root cause identification for road network congestion using the gradient boosting decision trees," in *GLOBECOM 2020-2020 IEEE Global Communications Conference*. IEEE, 2020, pp. 01–06.

[18] E. Winter, "The shapley value," *Handbook of game theory with economic applications*, vol. 3, pp. 2025–2054, 2002.

[19] S. K. MISHRA, "Shapley value regression and the resolution of multicollinearity," *Journal of Economics Bibliography*, vol. 3, no. 3, p. 498, 2016.

[20] S. M. Lundberg and L. Su-In, "A unified approach to interpreting model predictions. nips 2017," *arXiv preprint arXiv:1705.07874*, 2017.

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

[22] G. H. Kulkarni and P. G. Waingankar, "Fuzzy logic based traffic light controller," in *2007 International Conference on Industrial and Information Systems*. IEEE, 2007, pp. 107–110.

[23] P. W. Shaikh, M. El-Abd, M. Khanafer, and K. Gao, "A review on swarm intelligence and evolutionary algorithms for solving the traffic signal control problem," *IEEE transactions on intelligent transportation systems*, vol. 23, no. 1, pp. 48–63, 2020.

[24] R. Iglesias, F. Rossi, K. Wang, D. Hallac, J. Leskovec, and M. Pavone, "Data-driven model predictive control of autonomous mobility-on-demand systems," in *2018 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2018, pp. 6019–6025.

[25] B.-L. Ye, W. Wu, K. Ruan, L. Li, T. Chen, H. Gao, and Y. Chen, "A survey of model predictive control methods for traffic signal control," *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 3, pp. 623–640, 2019.

[26] A. Kamenev, L. Wang, O. B. Bohan, I. Kulkarni, B. Kartal, A. Molchanov, S. Birchfield, D. Nistér, and N. Smolyanskiy, "Predictionnet: Real-time joint probabilistic traffic prediction for planning, control, and simulation," in *2022 International Conference on Robotics and Automation (ICRA)*, 2022, pp. 8936–8942.

[27] S. S. Mousavi, M. Schukat, and E. Howley, "Traffic light control using deep policy-gradient and value-function-based reinforcement learning," *IET Intelligent Transport Systems*, vol. 11, no. 7, pp. 417–423, 2017.

[28] H. Wei, G. Zheng, H. Yao, and Z. Li, "Intellilight: A reinforcement learning approach for intelligent traffic light control," in *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, 2018, pp. 2496–2505.

[29] H. Wei, G. Zheng, V. Gayah, and Z. Li, "Recent advances in reinforcement learning for traffic signal control: A survey of models and evaluation," *ACM SIGKDD Explorations Newsletter*, vol. 22, no. 2, pp. 12–18, 2021.

[30] Z. Zhang, J. Yang, and H. Zha, "Integrating independent and centralized multi-agent reinforcement learning for traffic signal network optimization," in *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*, 2020, pp. 2083–2085.

[31] Z. Yu, S. Liang, L. Wei, Z. Jin, J. Huang, D. Cai, X. He, and X.-S. Hua, "Macar: Urban traffic light control via active multi-agent communication and action rectification," in *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, 2021, pp. 2491–2497.

[32] Q. JIANG, M. QIN, S. SHI, W. S. SUN, and B. ZHENG, "Multi-agent reinforcement learning for traffic signal control through universal communication method.(2022)," in *Proceedings of the 31st International Joint Conference on Artificial Intelligence (IJCAI-ECAI'22): Vienna, July*, 2022, pp. 23–29.

[33] Y. Bie, Y. Ji, and D. Ma, "Multi-agent deep reinforcement learning collaborative traffic signal control method considering intersection heterogeneity," *Transportation Research Part C: Emerging Technologies*, vol. 164, p. 104663, 2024.

[34] Y. Zhang, G. Zheng, Z. Liu, Q. Li, and H. Zeng, "Marlens: understanding multi-agent reinforcement learning for traffic signal control via visual analytics," *IEEE transactions on visualization and computer graphics*, 2024.

[35] S. G. Rizzo, G. Vantini, and S. Chawla, "Reinforcement learning with explainability for traffic signal control," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, 2019, pp. 3567–3572.

[36] J. Liu, S. Qin, M. Su, Y. Luo, Y. Wang, and S. Yang, "Multiple intersections traffic signal control based on cooperative multi-agent reinforcement learning," *Information Sciences*, vol. 647, p. 119484, 2023.

[37] W. Yue, C. Li, Y. Chen, P. Duan, and G. Mao, "What is the root cause of congestion in urban traffic networks: Road infrastructure or signal control?" *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 8662–8679, 2021.

[38] M. Wang, Y. Yuan, H. Yan, H. Sui, F. Zuo, Y. Liu, Y. Li, and D. Jin, "Discovering causes of traffic congestion via deep transfer clustering," *ACM Transactions on Intelligent Systems and Technology*, vol. 14, no. 5, pp. 1–24, 2023.

[39] I. Gemp, C. Chen, and B. McWilliams, "The generalized eigenvalue problem as a nash equilibrium," *arXiv preprint arXiv:2206.04993*, 2022.

[40] M. Coeckelbergh, "Artificial intelligence, responsibility attribution, and a relational justification of explainability," *Science and engineering ethics*, vol. 26, no. 4, pp. 2051–2068, 2020.

[41] L. Chong, M. M. Abbas, A. M. Flintsch, and B. Higgs, "A rule-based neural network approach to model driver naturalistic behavior in traffic," *Transportation Research Part C: Emerging Technologies*, vol. 32, pp. 207–223, 2013.

[42] A. Likmeta, A. M. Metelli, A. Tirinzoni, R. Giol, M. Restelli, and D. Romano, "Combining reinforcement learning with rule-based controllers for transparent and general decision-making in autonomous driving," *Robotics and Autonomous Systems*, vol. 131, p. 103568, 2020.

[43] H. Zhang, S. Feng, C. Liu, Y. Ding, Y. Zhu, Z. Zhou, W. Zhang, Y. Yu, H. Jin, and Z. Li, "Cityflow: A multi-agent reinforcement learning environment for large scale city traffic scenario," in *The world wide web conference*, 2019, pp. 3620–3624.

[44] H. Wei, C. Chen, G. Zheng, K. Wu, V. Gayah, K. Xu, and Z. Li, "Presslight: Learning max pressure control to coordinate traffic signals in arterial network," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 1290–1298.