A Distributed Double Deep Q-Learning Method for Object Redundancy Mitigation in Vehicular Networks

Imed Ghnaya, Hasnaâ Aniss, Toufik Ahmed, Mohamed Mosbah

A Distributed Double Deep Q-Learning Method for Object Redundancy Mitigation in Vehicular Networks

Imed Ghnaya
Univ. Bordeaux, Bordeaux INP
CNRS, LaBRI, UMR5800
F-33400 - Talence, France
imed.ghnaya@u-bordeaux.fr

Hasnaâ Aniss
Gustave Eiffel University
COSYS-ERENA Lab
F-33067 - Bordeaux, France
hasnaa.aniss@univ-eiffel.fr

Toufik Ahmed
Univ. Bordeaux, Bordeaux INP
CNRS, LaBRI, UMR5800
F-33400 - Talence, France
tad@labri.fr

Mohamed Mosbah
Univ. Bordeaux, Bordeaux INP
CNRS, LaBRI, UMR5800
F-33400 - Talence, France
mohamed.mosbah@u-bordeaux.fr

Abstract—The use of Cooperative Perception (CP) enables Connected and Autonomous Vehicles (CAVs) to exchange objects perceived from onboard sensors (e.g., radars, lidars, and cameras) with other CAVs via CP messages (CPMs) through Vehicle-to-Vehicle (V2V) communication technologies. However, the same objects in the driving environment may simultaneously appear in the line of sight of multiple CAVs. Consequently, this leads to much irrelevant and redundant information being exchanged in the V2V network. This overloads the communication channel and reduces the CPM delivery to CAVs, thereby decreasing CP awareness. To address this issue, we mathematically formulate CP information usefulness as a maximization problem in a multi-CAV environment and introduce a distributed multi-agent deep reinforcement learning approach based on the double deep Q-learning algorithm to solve it. This approach allows each CAV to learn an optimal CPM content selection policy that maximizes the usefulness of surrounding CAVs as much as possible to reduce redundancy in the V2V network. Simulation results highlight that the proposal effectively mitigates object redundancy and improves network reliability, ensuring increased awareness at short and medium distances of less than 200 m compared to state-of-the-art approaches.

Keywords—connected and autonomous vehicles, cooperative perception, redundancy mitigation, multi-agent system, deep reinforcement learning.

I. INTRODUCTION

Autonomous vehicles rely on onboard sensors such as radars, lidars, and cameras to detect road objects, including other vehicles, obstacles, and pedestrians, in order to improve their driving safety. However, these sensors have a limited line of sight and may not detect all objects due to obstructions caused by buildings and other road users. This can negatively impact the vehicle's perception capacity, reducing driving safety and efficiency. Vehicle-to-Everything (V2X) communications [1], which enable a Connected Autonomous Vehicle (CAV) to communicate with other vehicles, infrastructure, and even pedestrians, have helped to overcome this limitation. This communication is performed by exchanging information via wireless communication technologies using Vehicle-to-Vehicle (V2V), Vehicle-to-Infrastructure (V2I), and Vehicle-to-Pedestrian (V2P) scenarios. In this regard, the European Telecommunications Standards Institute (ETSI) has standardized the ITS-G5 [2], an IEEE 802.11p-based communication profile specifically for CAVs to share their status information (e.g., speed, position, heading, etc.) through Cooperative Awareness Messages (CAMs) [3].

The development of Cooperative Perception (CP) [4] has been the key concept in addressing the limitations of onboard sensors. Fundamentally, the CP is utilized to extend the limited horizons of CAVs beyond their restricted Field of View (FoV). To achieve this goal, CAVs can share sensory information with other CAVs and the road infrastructure in the form of high-level descriptions of tracked objects, such as speed, position, height, and width, via CP Messages (CPMs), allowing them to access information that they would not be able to perceive on their own.

While this paper focuses exclusively on the V2V-based exchange of CPMs, the same objects in the driving environment may simultaneously appear in the FoV of multiple CAVs and, as a consequence, this leads to a large number of irrelevant and redundant information being exchanged in the V2V network. This overloads the communication channel and reduces the CPM delivery for CAVs, thereby decreasing CP awareness. It is, therefore, essential to explore ways to optimize the V2V exchange of CPMs, such as filtering out redundant information, to improve the network's overall performance.

The exchange of CPMs between CAVs is crucial for ensuring fresh and timely useful perception information. However, deriving the optimal message is challenging as it requires assessing all possible combinations of perceived objects by the transmitter CAV. This becomes computationally complex and impractical in dense driving scenarios where CAVs continuously perceive large amounts of information. Reinforcement Learning (RL) [5] has been developed as a powerful automatic decision-making method to tackle this challenge. RL enables an agent to interact with a stateful environment and learn by taking sequential actions to maximize a reward value. In this case, Q-Learning [5], a popular RL-based algorithm, can be used to allow a CAV to learn a state-action value function to find an optimal CPM content selection policy that maximizes the usefulness of the receiver CAVs. However, the main limitation of RL-based algorithms is the curse of dimensionality, which arises from the exponential growth of the state and action spaces due to the number of perceived objects. To overcome this limitation, Deep RL (DRL) represents an excellent alternative by using Deep Neural Networks (DNNs) as function approximators to approximate the optimal value function in large state and action spaces.

This paper proposes a distributed deep reinforcement learning method for a multi-CAV setting based on the Double Deep Q-Learning (DDQN) algorithm [6]. The primary objective of this method is to enable each CAV to learn a CPM content selection policy that maximizes the receiver CAVs' usefulness to mitigate the number of redundant objects in the V2V network. The obtained results indicate that the proposed method effectively reduces object redundancy and improves network reliability without the use of static thresholds, resulting in an increased perception awareness at short and medium distances of up to 200 meters, compared to current state-of-the-art works.

The main contributions of this paper are summarized as follows:

- We derive a mathematical function for determining the usefulness of objects perceived by onboard sensors and V2V communications in a multi-CAV environment based on various factors, such as distance, object size, viewing angle, and occlusions caused by road users, to adapt as much as possible to the driving environment. This function is then used to formulate the maximization problem.
- As CAVs move through the driving environment and store information about past actions to improve future decisions, it is not possible to use a positional representation for the training process. To address this, we propose a scalable position-independent design for the state and action spaces that allows CAVs to leverage past experiences from different regions and at any time during the training process.
- We develop a distributed DDQN-based algorithm in a multi-CAV environment to address the CPM content selection problem.
- We implement and evaluate the proposed method using PyTorch library and through advanced discrete-event network and road traffic simulators. We then present simulations and evaluations to show its performance compared to the state-of-the-art approaches.

The rest of this paper is structured as follows. Section II provides an overview of recent related works. The design and algorithm of the proposed method are detailed in Section III. Simulation results are presented in Section IV. Finally, the paper is concluded in Section V.

II. Related works

Sharing sensory information between CAVs is an active research topic in vehicular networks due to the limited resource available for the V2V network. Based on the earlier work in [7], the ETSI has proposed a CPM format and a set of generation rules [4] to balance the amount of perception data exchanged in the network with the channel load. Briefly, the CPM format includes a set of information containers to describe the transmitter CAV, its onboard sensors, and its perceived objects. On the other hand, the CPM generation rules determine when a CAV generates and transmits a CPM and what information it should include. Specifically, a CAV generates and transmits a CPM if one of the following conditions is satisfied. (*i*) It detects a new object. (*ii*) Its position or speed has changed by 4 m (meters) or 0.5 m/s (meters per second), respectively, since the latest information included in its CPM. (*iii*) The last time the detected object was included in a CPM was 1 (or more) seconds ago. If none of the above conditions are met, the CAV still generates a CPM every 1 s.

According to a recent study in [8], the CPM generation rules can lead to excessive information redundancy in the V2V network as CAVs do not take into account information received from their neighbors in the environments. A recent dynamics-based redundancy mitigation technique is proposed in [9], where each CAV analyzes the most recent CPMs received from other CAVs and excludes perceived objects that exceed predefined position or speed thresholds. This technique allows CAVs to adjust the number of updates to each perceived object independently; for instance, a fast-moving object will receive more updates than a slow-moving object. In addition, authors in [10] have proposed redundancy control schemas based on channel status, number, and type of V2X stations that have also provided the same perceived information. The main objective is to adapt the number of V2X stations transmitting data about the same object to the channel load while maintaining CP awareness close to the default CPM generation rules. However, these techniques rely on static thresholds that may not be suitable for varying driving environments and vehicular densities. To address this, authors in [11] have proposed omitting redundant objects based on their usefulness, which is modeled as a RL reward based on the distance from the perceived object to the receiving CAV. However, this approach does not take into account other factors that may influence object usefulness, such as object size and road occlusions. There is a need for a more comprehensive approach that considers various perception contexts and their impact on CP awareness in the V2V network.

III. Distributed DDQN for object redundancy mitigation

This section introduces the problem formulation and the proposed design and learning algorithm.

A. Problem formulation

We formulate a mathematical function to determine the usefulness of perceived information from onboard sensors and V2V communications as a maximization problem in a multi-CAV environment. This function considers multiple perception contexts, including position, distance, object size, viewing angle, and occlusions caused by other road users to adapt as closely as possible to the driving environment.

We capture a snapshot of the driving environment illustrated in Fig. 1 at time t. Each CAV is modeled as a rectangle, $v_i(t) = (c_i(t), l_i, w_i)$, where $c_i(t)$ is the geometric center of the rectangle represented by its X-position $x_i(t)$ and Y-position $y_i(t)$ at time t on a global 2-dimensional plane, its length l_i and width w_i. As an assumption, all CAVs in the driving environment possess the same capabilities, which include: (*i*) the use of Global Positioning System (GPS° and Global Navigation Satellite System (GNSS) devices to provide real-time information about its location; (*ii*) the inclusion of 360° sensors such as radars and lidars to perceive its surroundings. The 360° sensing coverage is defined as a circle with a radius m, representing the maximum sensing range of all the sensors; (*iii*)

the installation of V2V wireless communication devices to share information with other CAVs.

Every t =100 milliseconds (ms), each CAV engages in a Cooperative Awareness Service, which uses GPS/GNSS and other data to generate and transmit information about its status to other CAVs via CAMs through V2V communication [3]. This information, known as the CAV's state, $h_i(t)$, can include various features such as position, speed, length, and width. For the purpose of simplicity, this study will only focus on a CAV's position, length, and width on a global 2D plane. Therefore, the state of a CAV, $h_i(t)$, can be represented by a set of features, $\{x_i(t), y_i(t), l_i, w_i\}$. As a result of this exchange of information, each CAV receives the states of all other CAVs, $H_i(t) = \{h_j(t), j \neq i\}$, via V2V communication.

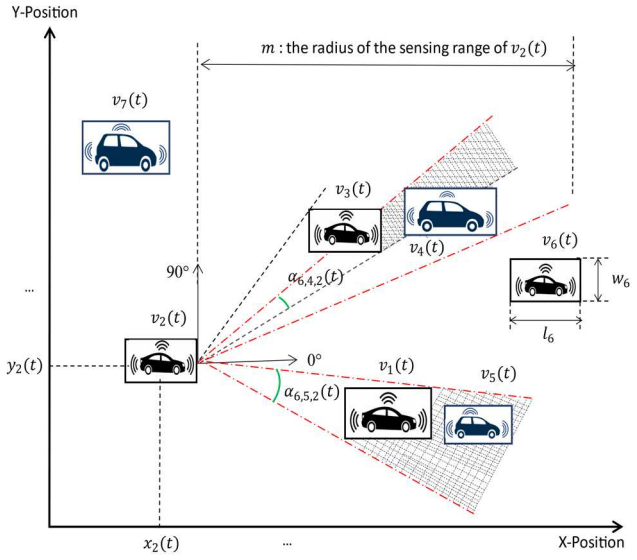

Fig. 1. A geometric representation of the driving environment at time t

We consider another time instance, t', when CAVs perceive and share objects through CPMs using V2V communications. This process begins with each CAV using its onboard sensors to perform a local perception phase, allowing it to perceive the road environment locally. Following that, the CAV fuses locally perceived objects with received objects from other CAVs and provides as an output the final list of objects to be included in a CPM and broadcasted in the V2V network. We define this final list, $T_i(t') = \{o_{i,1}(t'), o_{i,2}(t'), \dots, o_{i,k}(t')\}$, by the list of perceived features, including the position, length, and width of each object, $o_{i,j}(t') = \{x_j(t), y_j(t), l_j, w_j\}$, perceived by the i-th CAV. However, without any prior intelligence, the exchange of perceived objects between CAVs could result in a large amount of unnecessary and redundant information in the V2V network. To address this issue, we propose the CPM content selection strategy that allows each CAV to select and broadcast only useful objects to its surrounding CAVs. This selection strategy adapts to the driving environment and does not rely on static thresholds. This is done by employing received CAM and CPM information to create a geometric representation of its surroundings in the driving environment (e.g., provides as an output a list of the position, length, and width for each surrounding element.

To that end, the maximization problem at the i-th CAV is formulated as follows:

$$\underset{P_i(t')}{\text{maximize}} \quad 1 - \left(\frac{1}{n'n''} \sum_{\substack{k=1 \\ k \neq i}}^{n'} \sum_{\substack{j=1 \\ j \neq k}}^{n''} f_{i,j,k}(t') * g_{i,j,k}(t') \right), \tag{1}$$

Subject to:

$$t < t' < 2t, \tag{2}$$

where,

$P_i(t') = \{o_{i,1}(t'), o_{i,2}(t'), \dots, o_{i,n''}(t');\ o_{i,j}(t') \in T_i(t')\}$, is the list of n'' useful objects to be included in a CPM. n' is the number of CAVs in the communication coverage of the ith CAV. Equation (2) ensures that CAVs broadcast and receive CAMs before sharing CPMs and that the environment is unchanged between t and t'.

$$f_{i,j,k}(t') = \begin{cases} 0, & d_{i,j,k}(t') > m \\ 1 - \frac{d_{i,j,k}(t')}{m}, & otherwise \end{cases}, \tag{3}$$

Equation (3) denotes a distance-related factor, $f_{i,j,k}(t')$, that ranges between 0 and 1 and represents the degree to which the j-th object, $o_{i,j}(t')$, perceived by the i-th CAV appears in the sensing coverage of the k-th CAV. Specifically, as the Euclidean distance, $d_{i,j,k}(t')$, between $o_{i,j}(t')$ and the k-th CAV gets closer to m, the factor approaches zero, indicating that $o_{i,j}(t')$ is becoming less perceptible to the k-th CAV and, therefore, more useful for it.

$$g_{i,j,k}(t') = \begin{cases} 0, & \phi_{i,j,k}(t') = \sum \alpha_{i,j,k}(t') > \rho_{i,j,k}(t') \\ 1 - \frac{\phi_{i,j,k}(t')}{\rho_{i,j,k}(t')}, & otherwise \end{cases}, \tag{4}$$

Equation (4) represents an occlusion-related, $g_{i,j,k}(t')$, which also ranges from 0 to 1, and signifies the degree to which $o_{i,j}(t')$ is directly in the LoS of the k-th CAV. This factor is proportional to the sum of angles (e.g., $\alpha_{i,j,k}(t')$) that overlap and occlude the viewing angle $\rho_{i,j,k}(t')$ from the k-th CAV to $o_{i,j}(t')$. In other words, the more this sum of angles approaches the viewing angle, the more occluded the LoS to the object becomes. As an illustration, in Fig. 1, for $i = 6$, $j = 4$, and $k = 2$, $\alpha_{6,4,2}(t')$ is the only occlusion angle that occludes the viewing angle $\rho_{6,4,2}(t')$ from the 2nd CAV, $v_2(t')$, to the 4th CAV, $v_4(t')$, making it impossible for the 2nd CAV to detect the latter.

B. *System design*

Each CAV in the driving environment acts as an agent that chooses and transmits perception information through CPM at each CP-related time. The main goal is to maximize the usefulness of its CPM over its communication coverage.

Therefore, we define the following RL essential components: environment state, action, and reward.

Environment state. As CAVs move within the driving environment and collect experienced state-action to improve their future decisions, a positional representation cannot be used to conduct the training process. Therefore, we introduce a flexible position-independent representation of the environment state that allows an i-th CAV to draw on its past experiences in various locations at any given time, as follows:

$$s_i(t') = \{(d_{i,j}(t'), \beta_{i,j}(t'), l_j, w_j); \forall\, j \neq i\}, \quad (5)$$

where $d_{i,j}(t')$ and $\beta_{i,j}(t')$ are the distance and viewing angle from the ith CAV to the j-th CAV. l_j and w_j are the length and width of the j-th CAV.

Action. Given that CAVs are constantly navigating in a dynamic environment, the perception of their surroundings can vary over time. Therefore, the action to take should be independent of changing characteristics because CAVs store state-action at each timestep. To achieve this, we propose a cell-based scheme that divides the circular FoV of each CAV into p pistes and s sectors. Hence, the FoV of the i-th CAV at t' can be represented by a vector of cells of size $s * p$ as follows:

$$FoV_i(t') = \left[C^0_{1,1}(t'), \dots, C^1_{1,s-1}(t'), C^2_{1,0}(t'), \dots, C^{p*s-1}_{,p-1,s-1}(t')\right], \quad (6)$$

where $C^j_{p',s'}(t')$ is representing the cell of p′-th piste and s′-th sector indexed by $j \in [0, p * s - 1]$. To that end, we define the action space by the power set of $FoV_i(t')$ in order to cover all the unique possible combinations of cells with a complexity of $\Theta(2^{p*s} - 1)$. The action $a_i(t)$ can be represented as a natural number in the range $[0,\ 2^{p*s} - 1\ ]$. This number is then converted to a binary string $(b_0 b_1 \dots\ b_{p*s-1})_2$ of length $p * s$. Following that, the objects that appear in $C^j_{p',s'}(t')$ will be included in the current CPM only if the corresponding bit in the binary string, b_j, is set to 1. The CPM will include all perceived objects if all bits are set to 1, and will be empty if all bits are set to 0.

Reward. The reward $r_i(t')$ of the i-th CAV at timestep t is the usefulness of the objects generated from the selected action over its communication coverage. The mathematical expression for this reward function is represented in (1).

C. Learning algorithm

The goal of using RL on the i-th CAV is to find a CPM content selection policy that maximizes the usefulness of the receiving CAVs. As illustrated in Algorithm 1, the i-th CAV, at each CP-related time step t', builds the state $s_i(t')$ of its environment, takes an action $a_i(t')$, gets a reward $r_i(t')$, and observes the new state $s_i(t' + 1)$. The cumulative discounted reward at t' can be given by $R_i(t') = \sum_{\tau=0}^{t'} \gamma^{t'-\tau} r_i(\tau)$, where $\gamma \in [0,1]$ is a discounted factor given to earlier rewards to reduce their impact on the current output. The well-known Q-learning algorithm [5] is one of the most used RL-based algorithms that can be applied. The Q-learning in this context can lead the i-th CAV to learn a state-action value function, $Q^{\pi}_i(s_i(t'), a_i(t')) = R_i(t')$, when it takes $a_i(t')$ in state $s_i(t')$ following its CPM content selection policy π_i. The Q-learning algorithm uses a lookup table that stores the state-action value functions of all actions in the action space. This method can be impractical for time-sensitive applications, such as the CP, due to its high consumption of time and memory when dealing with large state and action spaces. However, recent advancements in DQN [5] have provided a solution to this issue. DQN utilizes a DNN, parametrized by θ, to approximate the state-action value function of the i-th CAV. This approximation, represented by $Q_i(s_i(t'), a_i(t'), \theta) \approx Q^*_i(s_i(t'), a_i(t'))$, allows for a more efficient and effective means of finding the optimal action value function given by the optimal policy.

Algorithm 1: The distributed DDQN-based algorithm for object redundancy mitigation in the V2V network

Inputs:
Number of episodes N_{ep}, slots per episode N_s, ϵ-greedy $\epsilon \in]0,1[$
Output:
An optimal policy for each CAV
Begin
Initialize θ_i, $\theta'_i \leftarrow \theta_i$, and D_i for each CAV i
$ep \leftarrow 0$
while $ep < N_{ep}$ **do**
  $t \leftarrow 0$
  while $t < N_s$ **do**
    each available CAV i :
      Builds $s_i(t)$ according to (5)
      Samples c from Uniform (0,1)
      if $c \leq \epsilon$ **then**
        Performs $a_i(t)$
      else
        $a_i(t) = arg \max_{a_i\ (t)} Q_i(s_i(t), a_i^{'}\ (t), \theta)$
      endif
      Gets $r_i(t)$, and observe $s_i(t + 1)$
      Stores $e_i(t) \leftarrow (s_i(t), a_i(t), r_i(t), s_i(t + 1))$ into D_i
    $t \leftarrow t + 1$ // increase the slots by one CPM gen. interval
  end while
  each available CAV i :
    Samples a minibatch from D_i
    Updates θ_i according to (7) and (8)
  $ep \leftarrow ep + 1$
  Following each number of episodes N_{init}, set $\theta'_i \leftarrow \theta_i$ for each CAV i
end while
End

In particular, the DQN utilized by the i-th CAV utilizes two deep neural networks, known as the train network and the target network, which are denoted by θ_i and θ'_i, respectively. These networks are employed to maintain a stable performance during training. The train network is updated every fixed number of time steps during an episode, while the target network is updated to match the train network after a set number of episodes. The well-used experience replay technique is employed to overcome learning stability. The i-th CAV stores experiences, represented by $e_i(t') = (s_i(t'), a_i(t'), r_i(t'), s_i(t' + 1))$, in a replay buffer $D_i[e_i(0), e_i(1), \dots, e_i(t')]$. The train network is updated by minimizing the squared loss, which is calculated from a randomly selected mini-buffer from D_i, as follows:

$$loss(\theta_i) = E_{(s,a,r,s')\sim U(D_i)}\left[y_i^{DQN} - Q_i(s, a, \theta)\right], \quad (7)$$

where $y_i^{DQN} = r_i' + \gamma \max_{a_i} Q_i(s_i^i, a_i^i, \theta_i')$, s_i', a_i', and r_i' are the state, action, and reward of the next time step. However, the same values are used to select and evaluate an action in the max operator defined in (7), which may conduct the training process efficiently. To overcome this issue, a DDQN-based algorithm is proposed in this paper to conduct the training process of each CAV in the driving environment. The proposal follows the same DQN process described above and only considers

$$y_i^{DDQN} = r_i' + \gamma Q_i\left(s_i', arg\max_{a_i'} Q_i(s_i', a_i^i, \theta_i), \theta_i'\right), \quad (8)$$

instead of y_i^{DQN}. Note that according to (8), the selection of an action depends only on the train network. Meanwhile, the target network is being used to evaluate the value of the policy.

IV. Performance Evaluation

In this section, we evaluate the proposal's performance based on simulations.

A. Simulation setup and hyperparameters

We use Artery [12] and SUMO [13] simulators to simulate information exchange and node mobility. We have chosen a 10 km² area of Bordeaux, France, obtained from OpenStreetMap, to represent different road traffic scenarios, including the city center and highways with different situations, such as ramps and T-junctions. Each CAV is equipped with GPS/GNSS and 360° radar and lidar sensors with a maximum sensing range of 100 m. The CAVs exchange CAMs and CPMs every 0.1s and 0.15s, respectively, using the ETSI ITS-G5 communication protocol within a coverage range of 500 m. We implemented the proposal based on the PyTorch library. The simulation runs for 70000 time slots, with each time slot representing a CPM generation interval. The training phase includes $N_{ep} = 6000$ episodes, with updates made up every $N_s = 10$ time slots. Furthermore, we consider the RMSProp optimizer with a learning rate $\alpha = 10^{-3}$, a minibatch size of 64, a discount factor $\gamma = 0.99$, and a buffer size $|\mathcal{D}| = 10^6$ for each CAV to conduct the training process. For complexity reasons, we divide the FoV of each CAV into 3 pistes and 3 sectors, resulting in 9 distinct cells.

B. Evaluation of training convergence

Fig. 2 illustrates the variation of the average reward for the proposed method as a function of the number of episodes. For comparison purposes, we also evaluated a multi-CAV DQN (MDQN)-based approach, in which each CAV uses a DQN model that updates based on y_i^{DQN} to learn a CPM content selection policy. The proposed method shows superior performance in maximizing the average reward over 6000 episodes compared to the MDQN-based method. This suggests that using two separate DNNs for action selection and evaluation in the proposed method allows CAVs to learn more effective policies that optimize the usefulness of objects in V2V networks.

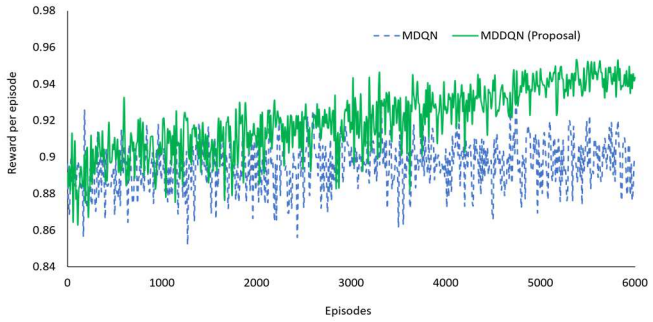

Fig. 2. The average reward variation as a function of episodes.

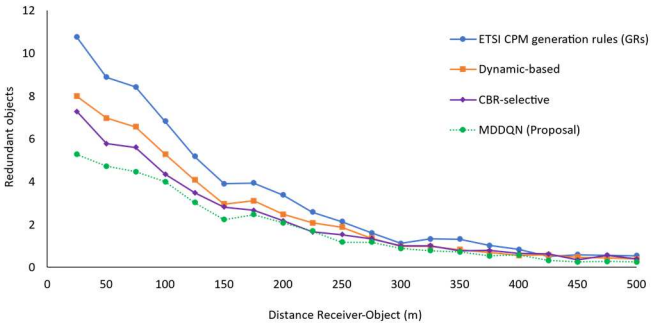

Fig. 3. Object redundancy as a function of the distance between the perceived object and the CAV receiving it.

C. Evaluation of network performance

After the training process reaches its maximum number of episodes, we assess the network performance of the proposed method. To compare its performance, we also take into account the ETSI CPM generation rules [4] (GRs), as well as the dynamics-based [9] technique, and the CBR-selective [10] scheme, which are outlined in Section II. For the dynamic-based technique and the CBR-selective scheme, we set a redundancy threshold of 5, where an object is considered redundant if its speed or position value difference is less than 0.5 m/s and 4 m, respectively. Furthermore, we set a CBR threshold of 0.6 for the CBR-selective scheme.

Fig. 3 illustrates how object redundancy (OR) changes depending on the distance between the perceived object and the CAV receiving the information. This metric measures the number of times a CAV receives identical information about the same object over the selected CP-related time interval. It is observed that OR is high at shorter distances as multiple CAVs detect and share information about the same object simultaneously. However, as the distance increases, the perception of the object becomes more difficult, resulting in a decrease in OR. The figure demonstrates that the dynamic-based technique and the CBR-selective scheme have slightly better performance in reducing OR at short and medium distances of up to 300 meters. In contrast, the proposal shows more efficient OR mitigation at shorter distances of up to 150 meters than the other approaches.

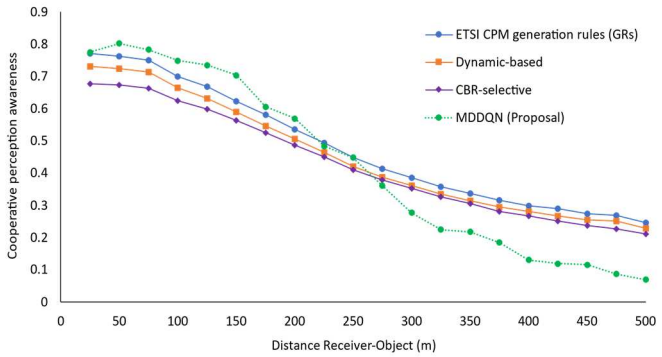

Fig. 4. The CP awareness as a function of the distance between the perceived object and the CAV receiving it.

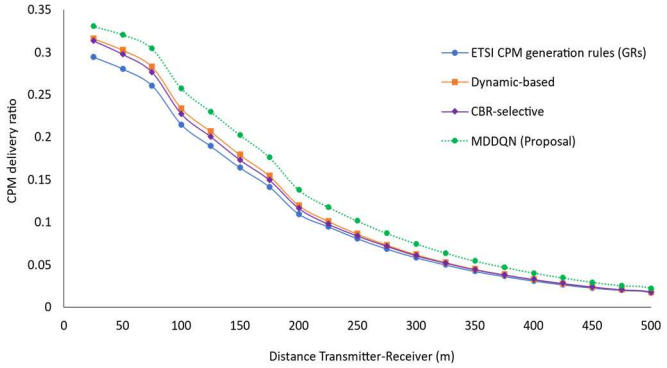

Fig. 5. The CPM delivery ratio as a function of the distance transmitter-receiver

The performance achieved by the proposal in mitigating OR shall maintain CP awareness in the V2V network close to GRs, particularly at short distances that are critical for the safety of CAVs. This is shown in Fig. 4, which depicts the CP Awareness (CPA) level reached by each method as a function of the distance between the detected object and the CAV receiving it. In this context, CPA identifies the probability of the number of unique objects perceived by a CAV using onboard sensors and V2V communication to the total number of objects within its coverage. Fig. 4 shows that the dynamics-based technique and the CBR-based scheme achieve almost the same CPA as GRs at distances larger than 100 m; however, their performance degrades by about 5% and 10%, respectively, at distances less than 100 m. On the other hand, the proposed solution significantly improves CPA compared to the other methods over distances less than 200 m, which is critical for the safety of CAVs. This improvement is achieved by enhancing the reliability of V2V communication, enabling CAVs to receive additional useful information that is lost or not transmitted by the other methods. We measure this reliability using the CPM delivery ratio (CDR) shown in Fig. 5. CDR is presented as a function of the distance between the transmitter and the receiving CAVs and highlights the increased CPM reception ratio achieved by the proposed solution compared to the other approaches over distances less than 200 m.

V. Conclusion

This paper presents a novel method for optimizing the selection of perception information by connected autonomous vehicles to share in the vehicle-to-vehicle network. The approach utilizes a distributed multi-agent double deep Q-Learning algorithm to learn the optimal information selection policy for each CAV, maximizing the usefulness of receiving CAVs to reduce redundancy and save network reliability. The proposal's effectiveness was evaluated through simulations, which showed that the method significantly reduces redundant information and improves V2V communication reliability. This leads to an increased level of cooperative perception awareness at safety-critical distances of less than 200 meters compared to existing state-of-the-art approaches methods.

References

[1] S. Chen, J. Hu, Y. Shi, Y. Peng, J. Fang, R. Zhao, and L. Zhao, "Vehicle-to-everything (v2x) services supported by LTE-based systems and 5G," IEEE Commun. Standards Mag., vol. 1, no. 2, 2017, pp. 70–76, https://doi.org/10.1109/MCOMSTD.2017.1700015.

[2] ETSI EN 302 663 V1.3.11, Intelligent Transport Systems (ITS); ITS-G5 Access layer specification for Intelligent Transport Systems operating in the 5 GHz frequency bands, 2020.

[3] ETSI EN 302 637-2 V1.3.2, Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Part 2: Specification of Cooperative Awareness Basic Service, 2014.

[4] ETSI TR 103 562-V2.1.1, Intelligent Transport System (ITS); Vehicular Communications. Basic Set of Applications; Analysis of the Collective Perception Service (CPS); Release 2, 2019.

[5] R. S. Sutton and A. G. Barto, "Reinforcement Learning: An Introduction," Cambridge, MA: MIT Press, 1998,
https://doi.org/10.1017/S0263574799271172.

[6] H. van Hasselt, A. Guez, D. Silver, "Deep Reinforcement Learning with Double Q-Learning," Proceedings of the AAAI Conference on Artificial Intelligence, 2016, https://doi.org/10.1609/aaai.v30i1.10295.

[7] H. -J. Günther, B. Mennenga, O. Trauer, R. Riebl, L. Wolf, "Realizing collective perception in a vehicle," 2016 IEEE Vehicular Networking Conference (VNC), 2016, pp. 1-8,
https://doi.org/10.1109/VNC.2016.7835930.

[8] G. Thandavarayan, M. Sepulcre, J. Gozalvez, "Analysis of Message Generation Rules for Collective Perception in Connected and Automated Driving," 2019 IEEE Intelligent Vehicles Symposium (IV), 2019, pp. 134-139, https://doi.org/10.1109/IVS.2019.8813806.

[9] G. Thandavarayan, M. Sepulcre, J. Gozalvez, "Redundancy Mitigation in Cooperative Perception for Connected and Automated Vehicles," 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring), 2020, pp. 1-5, https://doi.org/10.1109/VTC2020-Spring48590.2020.9129445.

[10] A. Chtourou, P. Merdrignac, O. Shagdar, "Context-aware content selection and message generation for collective perception services," Electronics, vol. 10, no. 20, 2021,
https://doi.org/10.3390/electronics10202509.

[11] B. Jung, J. Kim, S. Pack, "Deep Reinforcement Learning-based Context-Aware Redundancy Mitigation for Vehicular Collective Perception Services," 2022 International Conference on Information Networking (ICOIN), 2022, pp. 276-279,
https://doi.org/10.1109/ICOIN53446.2022.9687254.

[12] R. Riebl, H. Günther, C. Facchi and L. Wolf, "Artery: Extending Veins for VANET applications," 2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), 2015, pp. 450-456, https://doi.org/10.1109/MTITS.2015.7223293.

[13] Krajzewicz, Daniel & Erdmann, Jakob & Behrisch, "Michael & Bieker-Walz, Laura. (2012)," Recent Development and Applications of SUMO - Simulation of Urban Mobility," International Journal On Advances in Systems and Measurements. 3&4