Medicon Engineering Themes
Volume 3 Issue 6 December 2022
Article Type: Review Article
ISSN: 2834-7218

Exploring Reinforcement Learning Environment for User-centric Applications in VANET

V Padmapriya* and D N Sujatha

Department of Computer Applications, B. M. S. College of Engineering, Bangalore, India

***Corresponding Author:** V Padmapriya, Department of Computer Applications, B. M. S. College of Engineering, Bangalore, India.

Received: October 26, 2022; **Published:** November 19, 2022

DOI: 10.55162/MCET.03.092

Abstract

The past decade has identified Vehicular Ad Hoc Networks (VANET) as one of the promising technology for intelligent communication among vehicles. VANET supports the dissemination of safety, warning, and infotainment messages. Today, there is a growing demand for location-specific and infotainment messages among urban travelers. The class of applications that disseminates infotainment messages are called *user centric applications*. The task of dissemination is expected to be remunerative to promote cooperation among the vehicles. A vehicle that disseminates these user-centric messages earns a reward in the form of an *incentive*. Generally, the incentive bring in greed with a threat of malicious behavior in the network. Previously, several incentive-based approaches have been proposed that handle malicious behavior and maintain the equilibrium of rewards with the perspective of the Vehicular Network (VN). However, the Reinforcement Learning (RL) paradigm with its intelligent algorithms combined with vehicular networks is capacitated to handle several challenges in the incentive-based approaches. In this paper, we explore how RL environments can be adopted for the rewarding techniques in VANET. The paper concludes with open research challenges in this area.

Keywords: Applications; Incentive-based; Reinforced Learning; Rewards; User-centric; Vehicular Networks

Introduction

Rapid advancements in wireless communication and upgrades in automobiles have paved the way to intelligent mobility. Modern vehicles are embedded with intelligence supporting the Intelligent Transportation System (ITS). ITS is prevalent in urban areas. The primary objective of ITS is to provide safe travel, facilitate efficient traffic management, avert traffic jams and road congestions, and periodically alert drivers and passengers on road and weather conditions, etc. This transportation system leverages its services through the cooperative, self-organizing, and ad hoc network of the vehicle that forms a Vehicular Ad Hoc Network (VANET). Thus, VANET supports the building of safer, greener, more efficient, better informed, and well-managed urban road traffic management. VANET includes an ad-hoc group of cooperative and connected vehicles with varying speeds and dynamic topology. The vehicles in this network play a key role in information exchange supported by various applications. The application-assisted VANET communication is enabled by vehicle-bound devices like On Board Units (OBUs), positioning and navigating service providers, Global Positioning Systems (GPS) trackers, smart phones, and other wireless devices. VANET employs Vehicle-to-Vehicle (V2V), Vehicle-to-Infrastructure (V2I), and Vehicle-to-Anything (V2X) modes of communication to support various applications. VANET uses a range of communication protocols to disseminate information to the nodes in the network [1].

The Vehicular Networks are powered by various short-range, long-range wireless, and cellular technologies like 3G / LTE and 5G technology. These wireless technologies are expected to enable a new set of applications covering the safety, warning and comfort services and the additional services of improving traffic efficiency, optimizing routes, and thus reducing carbon footprints. However, the infotainment applications that are the most-sought category in urban areas are categorized into *thin* and *rich* variants. The non-real-time traffic and real-time traffic without any stringent QoS requirements are categorized as *thin infotainment services*. However, the good and fine-grained QoS support for real-time traffic is supported by the rich infotainment services category [2].

Short messages in the form of advertisements are popular services as part of comfort or infotainment services in VN. Advertisement forwarding is an effective way to reach a large customer base. Further, this act of forwarding would promise to reach a wider audience, if it's rewarding. The rewards are earned in the form of cashless incentives. The advertisements are considered comfort messages that are specific to the user category. So, in this paper, comfort messages are included as part of infotainment applications and are termed *user-centric applications* [3].

Complementing the technology included in the automobile industry, the urban areas are witnessing a tremendous rise in the vehicular population. With this, there is also an increasing trend in connected vehicles. Connected vehicles are a huge source of data. This huge data accumulation in vehicular networks makes the data-driven methods favorable for analysis and helps in the decision-making process.

With the emergence of new technologies and the inclusion of these latest advancements in the automobile and transportation sector, there is a new wave of services that needs human and human-like intelligence embedded in the vehicles. Also, it is anticipated that urban roads will soon be occupied by semiautonomous vehicles and the levels of autonomy are likely to increase in the future days to come. Thus, a new wave of services to control the systems are budding too. These control systems need intelligence to be designed on machine-based intelligence.

Today, we have a new paradigm like Artificial Intelligence (AI) that controls systems with minimum human intervention. The combination of AI in VN is anticipated to provide an effective solution to urban transportation problems. Thus, designing an adaptive vehicular network that includes learning-based AI algorithms would be a promising approach to solving urban travelers' hardships. In addition, Machine Learning (ML) which is a subcategory of Artificial Intelligence (AI) uses algorithms to learn insights and recognize patterns from data, hence applying learning methods helps in decision making. In this work, we focus on how user-centric messages are forwarded with the intelligence using a class of machine learning algorithms: Reinforcement Learning (RL). RL is a machine learning paradigm that deals with taking suitable action to maximize rewards for a particular event. It provides a list of paths or behavior that leads to better decisions. There is an increasing interest among researchers in RL-driven control mechanisms like vehicle route management, energy management, traffic management, autonomous driving, etc. The RL paradigm is based on Markov Decision Process (MDP), which includes both rewarding and penalizing criteria. The combination of MDP and the rewarding scheme employed for disseminating information in VANET is perceived to provide a feasible solution for the fair distribution of rewards to the vehicles [4].

Motivation

The last decade has witnessed several advancements in the urban transportation space with improved road infrastructure, technology inclusion in automobiles, and wireless connectivity. Wireless networks have devices, sensors, and vehicles forming complex systems. This complicated network facilitates the exchange of data among heterogeneous devices thus generating a huge amount of data. This data has to be smartly exploited as it is a significant source of information. The effective utilization of the data generated in this network opens doors for the implementation of several modern techniques. Several researchers have found that machine learning algorithms are the best fit to handle, analyze and provide suitable predictions from this huge data.

The predictions and decision-making abilities are boon for analyzing various issues in a vehicular network. Thus, incorporating ML techniques in VN is a favorable solution for smart data management. The safety and warning messages are part of mandatory messages in the VANET facilitating safe travel. However, the infotainment messages dissemination could be remunerative to the system. The remuneration could be in the form of rewards or penalties. In this work, we plan to explore the rewards aspects of infotainment messages disseminated in the VANET handled by the ML paradigm.

Objective

The objective of this paper is to present a theoretical overview of ML & RL algorithms and explore various RL paradigms for user-centric message distribution in VANET. Further, investigate the impact of incentivizing or rewarding the self-interest group in VANET using the RL paradigms.

The rest of this paper is organized as follows: Section 2 presents the theoretical overview of machine learning and reinforcement learning algorithms. Section 3 describes the proposed model for incentivizing vehicles using RL paradigms. Section 4 concludes the paper with the open research challenges and future work.

Overview of Machine Learning (ML) algorithms in VANET

Artificial Intelligence and Machine Learning

With the emergence of new technologies and the inclusion of the latest advancements in automobiles, the land transportation sector is witnessing a new wave of services. Artificial Intelligence is a new paradigm that has entered the dashboards of vehicles. The combination of AI and VN aims to provide an effective solution to urban transportation problems. Soon, semi-autonomous vehicles are expected to occupy the roads, and the in future, the roads are expected to be occupied by fully autonomous vehicles. Designing an adaptive vehicular network to support urban traffic systems with the learning-based AI algorithm would be a novel approach for sustainable, safe, and comfortable travel. Today Machine Learning (ML) is considered a substitute for AI. AI and ML are often used interchangeably.

However, AI refers to the general ability of computing devices to emulate human thoughts and perform tasks in a real-world environment. While ML refers to the technologies and algorithms that enable the system to identify patterns, make decisions, and improve itself through experience and data. Thus, ML is a subcategory of AI that uses algorithms to automatically learn insights and recognize patterns from data, applying learning that makes increasing better decisions [5].

Machine learning paradigms

Traditionally, the research in wireless networks uses simulation and mathematical models with a prior knowledge about the environment. The vehicular network uses wireless mode to communicate which includes other vehicles, roadside infrastructural elements, and other communicating entities like smart phones, smart watches, etc. Generally, the movement of vehicles in an urban area is steered by the mobility model. Further, vehicular movement is constrained by road conditions, route and driving-related information, speed limit, type of road, and so on. Additionally, the roads are associated with certain dimensions like the highway roads are considered one-dimensional, urban roads with streets and lanes as two-dimensional and urban roads with bridges, flyovers, and underpass are categorized under three dimensional ones.

The vehicular network is characterized by vehicles moving at varying speeds and topology. This highly dynamic network makes it very challenging to estimate the wireless channel and signal parameters using the traditional system design techniques that were employed in VANET. Today, we have contemporary solutions like Machine Learning paradigms that helps extract the patterns of movement of vehicles and signal parameters from real-time observations and historical data. Thus ML paradigms are well suited for learning and adapting to the uncertainties in the VN environment without any prior knowledge making ML a suitable partner for the VN.

The three classes of ML algorithms

The Machine Learning paradigms are categorized as Supervised Learning (SL), Unsupervised Learning (UL), and Reinforcement Learning (RL).

Supervised Learning (SL) is a learning paradigm that includes a function of a model that is fed with the input & desired output and finds the patterns and connections between the input and the output. This category of learning includes labeled data sets. SL is further categorized into *classification* and *regression*. Classification is a process of finding a model or a function that helps in separating the data into discrete values. Whereas, the regression is a process of finding a model or a function that distinguishes the data into a continuous real time value instead of classes or discrete values. Thus, the goal of learning in the SL paradigm is to map the input feature space to the output decision space.

Unsupervised Learning (UL) is a category of the ML paradigm that handles unlabeled data. This paradigm deals with the training of machines using information that is neither classified nor labeled. The aim here is to find an efficient representation of the data samples. The algorithm groups the unsorted information based on the similarities, patterns, and differences without any prior training of the data.

The third ML paradigm is Reinforcement Learning (RL). This category of learning maps the situation to actions by interacting with the environment. It uses a trial-and-error method to maximize the reward. The rewards are calculated using the Markov Decision Process (MDP). In turn, the MDP uses a classic model-free learning approach termed as Q function.

Overview of Reinforcement Learning (RL)

Human beings learn many things by interacting with nature. Learning from the interaction is a foundation ideal for all theories of learning and intelligence. Reinforcement Learning (RL) is a category of learning that tells what to do, and how to map situations to actions so that the rewards are maximized. The learner here is not told which action to take, rather the learner is allowed to discover which action yields the most reward by trying them. The interesting fact about this learning is that actions not only affect the immediate reward but also have an impact on the next situation; further subsequent rewards. So *trial-and-error search* and *delayed reward* are the two most distinguishing features of this reinforcement learning. Importantly the RL is not defined by characterizing a learning algorithm but by characterizing a learning problem. Reinforced learning is a popular trend in AI and ML that combines optimization, statistics, mathematics, and control theory. Today, the healthcare, robotics, image processing manufacturing, marketing, and wireless domains are extensively adopting the RL paradigms. Further, any algorithm that is well suited for solving a problem is considered to be a reinforcement learning problem. The RL algorithms begin by identifying an agile, interactive, goal-oriented agent that monitors and environment and takes corrective actions.

RL is different from Supervised Learning (SL). In SL the knowledge is provided by some knowledgeable external supervisor, learning is good but does not involve learning with interactions. Generally, the interactive problem is often impractical to obtain examples of desired behavior that are both correct and representative of all the situations in which an agent has to act. This means that an agent should learn from his own experiences.

The RL algorithms are based on two pivotal factors namely *exploration* and *exploitation*. In this RL technique, the agent has to exploit what is already known to obtain a reward, but it has to explore the use case to make a better selection of actions in the future. This means the agents must try a variety of actions and progressively favor those that appear to be best. Further, it implies that neither exploration nor exploitation are mutually exclusive and cannot be achieved by failing a task. The agent here operates in an uncertain environment but has explicit goals, and can choose actions to influence their environment.

Reinforced Learning is one of the machine learning paradigms which takes a sequence of actions that are guarded by the Markov Decision Process (MDP). This process leads to either rewarding or penalizing the agents. Today the RL combined with the Deep Learning form a special class of learning termed Deep Reinforced Learning or Deep RL. The Deep RL is currently the state-of-the-art learning framework used in several control systems. Generally, it is found that Deep Learning algorithms are used to handle non-linear functions that are derived from a complex data set, whereas, Reinforced Learning algorithms help solve problems with complex control systems [6].

General RL approach

RL is a novel learning tool driven by an agent. This learning benefits the agent in maximizing its rewards or minimizing the penalty. The agent learns from the environment, it sets the policy for maximizing or minimizing the rewards. Once the agent learns about the environment, it takes some action and seeks feedback from the environment. This feedback is taken iteratively and policies are correspondingly updated. Thus, the RL learns from its experience in the environment. The environment in RL is modeled using Markov Decision Process (MDP). The MDP introduces the rewards and the penalty to a Markov process. The state transition and the rewards are determined only by the current state of an agent and its selection action. This method of learning is generally seen in human beings are is based on the trial-and error method of learning. Computationally, the RL is driven by the data that is iteratively computed, applying the optimal control policy to arrive at a feasible and optimal solution. The goal 'G' of RL is to find a policy that takes action to maximize future rewards shown in equation 1:

$$G(t) = R_{t+1}, \Upsilon R_{t+2}, \ldots\ldots. = R_{t+1} + \Upsilon G(t+1) \text{ - } (1)$$

Where Υ is the discount factor and R_t is the reward at each step at a time 't'.

Components of learning in the RL paradigm

The objective of the RL paradigm is to learn from the environment, design or serach the suitable procedure for various conditions to maximize the benefit from the environment. The pivotal componets of RL paradigm are described below:

1. *Agent*: An agent is a component that decides what action to be taken in an environment. To make decisions, the agent is allowed to observe the environment. Further, these agents are goal-directed in a uncertain environment.
2. *Algorithm*: An agent is controlled by an algorithm.
3. *Action*: An action is performed by the agent by observing the environment.
4. *State*: The state is an observation an agent does in an environment after the action is taken. The environment gives an agent a state.
5. *Reward*: Feedback an agent receives after the action. Positive feedback is called reward and negative feedback is termed punishment. The learning process continues till the agent reaches the goal or meets some other condition. Figure 1 depicts the block diagram of RL, that includes the interaction of the agent with the environment, initatting action and earning rewards.

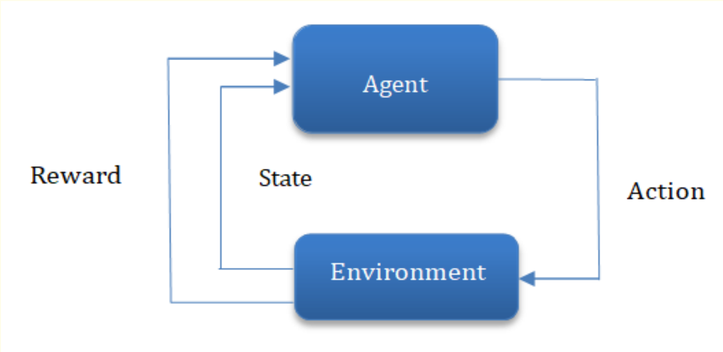

Figure 1: Block Diagram of Reinforcement Learning.

6. Every agent has a state 's_t', at a given time 't'. Agent receives reward 'r', for an action 'a_t', from the environment. The current policy for the agent is represented as 'π'. After 's_t', the system moves to the state 's_{t+1}'. After the interaction with the environment, the RL agent updates the knowledge [7].

How are these ML paradigms suitable for Wireless Networks?

The machine learning paradigms have a profound impact on wireless networks. For instance, in the case of the supervised learning algorithm, the classification algorithms in wireless networks help in securing the network from intruders, facilitate anomaly detection, and identify the malfunctioning components of the network. On contrary, the regression algorithms help in handling the network parameters like throughput and channel prediction. Similarly, in the case of the unsupervised paradigm, clustering is used in wireless networks to group nodes for energy management. However, the k-means, hierarchical, spectral clustering, and Dirichlet process are widely used clustering algorithms. Normally the cluster heads in a wireless network aggregate the data before it transmits to the infrastructural components like the Road Side Unit (RSU). This act of aggregation reduces the communication cost, the reduction methods like linear and nonlinear methods of the Unsupervised Learning (UL) helps in this context. The RL algorithms are applied to the VN to handle the temporal variations in the wireless networks.

Machine Learning and Reinforced Learning in Vehicular Networks

VANET was primarily designed to provide safety of travel, advanced warning messages to prevent road accidents, and alerts about the road, traffic, and weather conditions to travelers. But, one of the threats to message dissemination is the instability and dynamic topology of the network. The vehicular networks are an abundant source of data, however, the instability in the network leads to overload and congestion at the supporting roadside infrastructure. Numerous research articles highlight the conventional and standard protocols to handle traffic management, routing, safety parameters, etc [8]. However, not many concepts are implemented using Machine Learning algorithms. Several popular techniques like the Support Vector Machine (SVM), k-nearest neighbor, k-means clustering, Naive Bayes, etc. help in analyzing the data in VANET. A systematic study by Abdul-Halim et al. [9] identifies routing, data forwarding, road safety, and traffic management as the categories for prediction-based protocols.

VANET is considered a major source of information for travelers, however, one of the major applications is predicting the route from a source 'A' to destination 'B' through route 'C'. One of the major tasks, in this case, is to find the shortest path from source to destination and to predict the alternative route for vehicles to prevent congestion. Generally, dynamic mobility of the vehicles leads to link failure and hence contributes to route breakages. Frequent route failures need considerable time to repair and reconstruct the new route. The existing routing protocols that are widely used in mobile ad hoc networks are reactive. This means that the network waits till the existing route break and then the new route is constructed. On the other hand, it is preferable to use proactive protocols which avoid delay in the reconstruction of the route. The authors [10] have used a Prediction-Based Routing (PBR) protocol that predicts the lifetime of a route to preemptively create a new route before the existing route fails. The accuracy and the prediction of the link breakage & congestion are intelligently handled by using machine learning techniques. The authors Huang et al. [11] have proposed an enhancement mechanism to handle mobility issues using practical swarm optimization and fuzzy logic. Along the same lines, location-based routing for highways and city environments with various prediction-based algorithms like MORA, AGPm, and Park are proposed by various researchers [12-14]. The authors in [15-18] have developed prediction techniques combined with geographic, trajectory-based multicast, and position prediction multicast routing using Kalman filter in VANET to handle the geographic / geocast/ multicast / broadcast-based routing with relevant prediction techniques.

Data forwarding or message exchange is the essence of communication in VANET. The data is sent in various forms namely text, audio, video, images, animated images, short notifications, advertisements, etc that caters to safety, warning and infotainment applications of VANET. Due to the dynamic nature, and high mobility accompanied by frequent disconnections, the most acceptable way to route data packets is perceived by using the greedy strategy. The authors Shou- Chic & Wei Kun [16] have proposed several data forwarding strategies that are based on the prediction of neighboring and destination nodes. Whereas, the delay-tolerant and predictive data dissemination protocols with the ability to prevent flooding were designed in the research work by Tom et al. [17]. The authors proved that their designed protocol works both in the urban and highway environments integrated with GPS and local maps. Similarly, an efficient, reliable, and prediction-based data forwarding strategy with optimized packet routing and minimized hops in high-density vehicular networks is proposed by Wanting et al. [18].

The relay nodes play a significant role in routing and a study found that it is possible to find an optimal path when the knowledge of the future node traces is available but is an NP-hard problem [19, 20]. Thus, the knowledge of future vehicular trajectories plays a key role in optimal data delivery. The history of the vehicular traces are captured by vehicular mobility patterns. These patterns help to develop an accurate trajectory prediction using a multi-order Markov Chain proposed by Yanmin, Yachen & Bo [21]. Their proposed trajectory-based routing algorithm assures to provide a higher order delivery ratio at a lower cost compared to existing algorithms.

The communication in the vehicular network spans various edge devices like in-vehicle communicating devices, smartphones, etc. These edge devices are in turn are connected to the backbone network through the wired or wireless medium. Further, this connectivity with the increased channels and access points poses several security challenges. Vehicular networks are vulnerable to several types of attacks. Traditional network protection mechanisms like key-based authentication, password protection, and biometric techniques are not strong to retaliate against the modern categories of attacks. These techniques do not provide high accuracy. Hence, recently several ML techniques/algorithms are proven to be a suitable technology to handle security in wireless networks.

An extensive review by Anum et al. [22] brings out various security challenges and requirements for vehicular networks. Their study presents an in-depth state-of-the-art ML algorithm to solve security issues. A comprehensive survey concerning the security and privacy of communication in V2X is presented by Jiaqi et al [23].

Cooperation is key to successfully implementing data dissemination in vehicular networks. Additionally, cooperation enforcement adds security parameters and helps in averting security attacks. Panagiotis et al. [24] have proposed an advanced cooperative path prediction algorithm that provides position, velocity, acceleration, and several measurements collected from all the vehicles in the cooperative vehicular network. These parameters help the vehicles to calculate and predict the future path. Accident prevention and warning systems are boon to urban and highway driving. The Cooperative Collison Warning System by Gongjun et al. [25] provides mobility parameters and other driving parameters that alert drivers ahead of the collision. With the inclusion of ML algorithms in the VN, the VANET promises to revolutionize security, privacy, cooperative cruising, and accident or incident detection with intelligence.

Ideating the rewarding techniques for user-centric applications in Vehicular Networks

Conventionally, the agent in an RL-based environment is goal-oriented and constantly monitors the environment to take corrective actions. The agent exploits the environment rather than exploring it, but there is a tradeoff between these two. The environment enters an equilibrium state if an agent is rewarded for actions that involve both exploration and exploitation. The two primary aspects of agent is to get the information and use the information to make the next decision. This action helps the agent to earn an incentive.

RL settings for user-centric applications in VN

Urban travelers urge for infotaimnet messages on the go. The infotainment messages are disseminated by the user-centric applications loaded on the OBUs. These are non-safety messages that have commercial value. Here the agent gets incentivized to forward the messages to other agents in its neighborhood. These agents are self-interested groups that wishes to earn incentives and maximize the rewards.

Methodology

In this section, we apply the RL algorithm to the Secure Incentive Based Advertisement Distribution (SIBAD) approach. This is an incentive-based approach for distributing location-specific, commercial infotainment messages in the form of advertisements [26]. Normally, for disseminating the infotainment message in the VANET, the agent has to explore and exploit the environment to earn incentives/rewards. The SIBAD approach leverages the reward predominantly to the agents that explore the environment and then exploits the environment to disseminate messages. The agent interacts with the neighboring agent in the environment and chooses an action. The action which an agent chooses has a continuous effect. The agent performs an action and those moves to the next state. This becomes a iterative process that leads to rewards.

In this work, we apply model-based RL algorithms on the SIBAD.

The RL has two modes of learning: *model-based* and *model-free reinforcement learning*. The model-based learning has a model which guides the agent in an environment. The model guides the agent to earn rewards. It sets the policy for earning the rewards. The agents are kept away from the areas that have low rewards. The model identifies the malicious behavior of the agent. On the contrary, in model-free Reinforcement learning, an agent has no information about the environment. But the agent can learn to interact with the environment and earn a reward with a basic understanding of the reward policy. Figure 2 depicts the interaction of the agent with the environment and iteratively learning from the SIBAD approach.

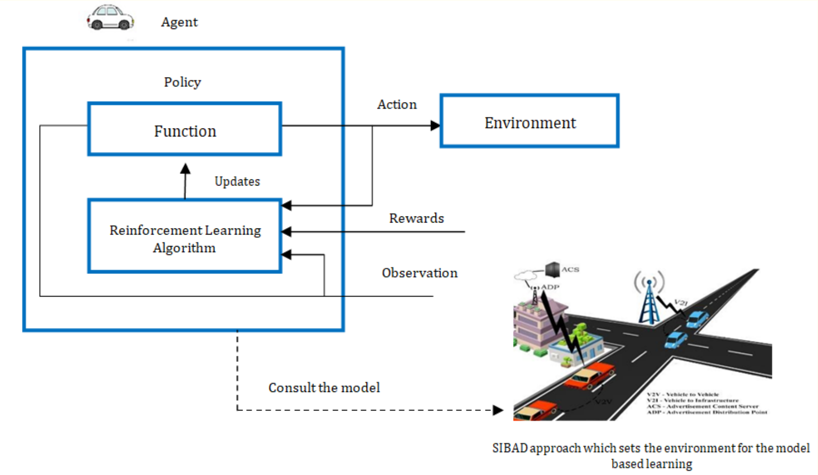
Agent
Policy
Function
Action
Environment
Updates
Reinforcement Learning Algorithm
Rewards
Observation
Consult the model
ACS
ADP
V2I
V2V
V2V - Vehicle to Vehicle
V2I - Vehicle to Infrastructure
ACS - Advertisement Content Server
ADP - Advertisement Distribution Point
SIBAD approach which sets the environment for the model based learning

Figure 2: Model-based RL applied with SIBAD approach.

The steps for applying model-based learning on SIBAD are shown below:

1. A commercial business house like a restaurant, fuel station with refreshment areas, amusement park, mall, movie house, entertainment hub, or any heritage center that wants to advertise contacts the advertisement distributors. Normally, these places are frequently visited by people with local fleet management services or by self-driving. The travelers would seek location-based services. These business hosts all the advertisements with the Advertisement Content Server (ACS). These servers authenticate the advertisers, check for the legitimacy and permission issues of the content, and the lifetime of the advertisements.
2. The advertisements are then moved to Advertisement Distribution Points (ADP) which are generally the Road Side Infrastructure (RSU) that are housed in the commercial buildings. These RSUs are enabled to distribute the advertisements to the *vehicles* that are considers are *agents*. The agents authenticate themselves with the ADP. After the handshake with the ADP, the agents are authorized to distribute or forward the advertisements to their neighboring vehicles and thus earn incentives.
3. A cooperative group of vehicles that wishes to participate in the advertisement distribution sets up the environment for the RL. Here the infotainment messages that are forwarded are termed advertisements. The vehicle that agrees to cooperate and participate in the message forwarding is regarded as an agent. The act of forwarding of advertisement is called *advertisement distribution*. The location-specific advertisements are loaded on the agents. The agents *explore* the environment. Then, they disseminate

legitimate advertisement.

4. All agents that agree to participate in this advertisement distribution are authenticated and authorized for the act of forwarding by a standard body like the Certification Authority (CA).
5. Agents observe the environment, and use other agents to disseminate the messages. The act of forwarding is remunerative to the agents as they are rewarded for every forward of advertisement.
6. The authorized agent receives an advertisement at a time ‘t’. If an agent forwards the message to the neighboring agents, then it is rewarded else they are punished. This forwarding activity sets the state of the vehicle. The agents pass through the start, authenticate, wait, forward and stop states. For every forward of an advertisement by an agent, it receives a *reward* ‘r’. If the agents are available, then agent ‘a_i’ sends an authentication message to agent ‘a_j’. If ‘a_j’ is authenticated, then ‘a_i’ sends an advertisement to ‘a_j’ with the timestamp. This act of dissemination earns ‘a_i’ a reward ‘r_i’ The reward is in form of virtual currency and it gets credited to the a_i‘s account. Else ‘a_i’ continues driving towards its destination. The cycle of explore-observe-act-reward continues and the agent keeps learning. The goal of the agent is to use reinforcement learning algorithms to learn and identify the best policy to earn rewards from the environment. It would be optimal to take action. To achieve this we need an optimal policy. But it is difficult to arrive at optimal policy as the environment is dynamic. Thus, changing the policy based on the actions the agent takes and the observations from the environment leads to optimal rewards.
7. The reward policy is set by the ACS. This needs a function that sets the profitable rewards. Traditionally, Linear Quadratic Regulator (LQR) is used to determine the cost function as a quadratic equation. However, the RL paradigms impose no such restriction on the reward function. Rewards are anticipated to be sparse at some point in the day, dense in the evenings, on weekends, or on general public holidays. The problem with the sparse reward is that the sending agent may not be able to reach other agents and; hence receives no rewards. But this situation is noted in the knowledge base of learning. Whereas, the dense situation is remunerative. However, this state may push the agents to behave maliciously. A group of agents exchange advertisements amongst themselves and earn a reward. This leads to malicious and greedy behavior sparks trust issues in the model. The malicious insider activity to earn incentives and the threat of incentive based earning were studied by the authors in their work [27, 28]. The agents in this approach get to know the policy and the probability of winning the reward. It identifies and implements the detailed actions that result in rewards. This puts the agent in a beneficial state. This method of learning is given by the State-Action-Reward-State-Action (SARSA) algorithm of the RL paradigm.
8. The SIBAD approach was initially designed based on the traditional vehicular network parameters and machine learning paradigms were not applied to this approach. Now using the model-based RL, the agents in the approach learns the incentive mechanism and the rewarding policies. This learning helps the agent to understand the environment, the reward policies, and the methods and modes of redeeming the rewards. This learning also enhances the knowledge. Effects of malicious and greedy behavior and the methods used by the model to arrest these security issues.

   These rewards are calculated using a non-linear function.
9. Finally, the model-based RL approach provides a way to balance exploration and exploitation; however, it is observed that an agent explores more things at the start of the learning and gradually transits to more exploitation later.

Open research challenges

In this subsection, we present several open research challenges that are interesting to be explored by the research community. Recently, the field of AI and MI are at their peaks in academia, industry, and the research community. It is perceived that ML algorithms are the panacea/cure-all for all the problems, especially with the significant developments in Deep Learning and Reinforcement Learning domains. However, the use of existing Machine Learning algorithm for Vehicular Networks is insufficient due to its distinguishing characteristics. The existing paradigm of Machine Learning is not an exact fit for all modes of communication in VANET. Thus, in this section, we highlight the challenges that need further investigation.

Challenges in Vehicular Network

The primary feature of a VN is to form, deform or create and destroy and validate the data in the network. This process of form, deform, and validation is a difficult task in VN as the network is dynamic. Further, the network has to ensure that there is a timely, fast, and reliable delivery of messages to all vehicles. The safety messages need reliability, and speed, whereas the comfort messages require improving passengers traveling experience and have appreciable bandwidth requirements. There are several algorithms in the VN domain to handle this issue, however, extensive research is not found in ML with the VN domain. However, several bio-inspired methods have been combined with the VN to address route decision optimization, but there is a lot of scope for further research [29].

Need for Standardization

VANET supports a myriad of applications that are suitable for the current transportation system. These applications need to focus on what route to follow to travel from point A to B. The following questions are still unanswered: How to change routes during heavy traffic or bad weather situations? How can the traffic signal be optimized? How to disseminate data based on the demand of the location?

One of the challenges in developing these algorithms is the lack of standardized algorithms in VANET. Previously, there were efforts made by companies like Open AI that have taken appreciable steps towards standardization of the environment in the RL domain. Similarly, there are standardization efforts made by the two popular transportation frameworks by *Flow* and *Simulation* of *Urban Mobility* (SUMO) in the VN domain. Use of these two environments with standardized protocols will aid in the development of algorithms. The combination algorithm would help to simulate real-world use cases. Although there are several research works published in a study on two frameworks in the respective domain, there is a lot of scope for research in the standardization area [30].

Challenges in Routing and Handover

Previously, vehicular networks banked on wireless communication such as WAVE / IEEE 802.11p, Cellular Long Term Evolution (LTE). However, enabling seamless mobility across various access networks without interruption remains a fundamental issue. Most recently the 5G system are geared up to deliver safety and traffic management infotainment messages. Hence, a 5G-enabled vehicular environment facilitates vehicles with high bandwidth-intensive applications like real-time traffic updates, video-streaming, commercial advertisement dissemination services, etc. The 5G technologies provide high bandwidth with low latencies, but, this technology suffers from the issues of mobility management, back-haul networking, usage of the air interface, and traffic safety. Mobility management includes handover and location management [31].

Conclusion

This paper presents the theoretical overview of the machine learning paradigms. Further, it highlights the need to adapt the learning algorithms in vehicular networks. The paper presents the idea of using RL using model-based learning for advertisement distribution in VANET. The paper concludes with open research challenges in this domain of RL in VN.

There are several dimensions for future work. At the first, we would want to evaluate the learning of the agent in a vehicular environment. Next we shall develop a learning mode that experimentally proves our claim of applying RL paradigms for incentivizing agents in the VN. The follow-up task would be to arrive at the appropriate rewarding policy to handle sparse and dense vehicular networks. All these future works will help in investigating further into the domain of RL-based learning in vehicular networks.

References

1. Botkar SP., et al. VANET: Challenges and Opportunities (1st ed.). CRC Press (2021).
2. Cheng Ho Ting, Hangguan Shan and Weihua Zhuang. “Infotainment and road safety service support in vehicular networking: From a communication perspective”. Mechanical systems and signal processing 25.6 (2011): 2020-2038.
3. Omar Hassan Aboubakr, Ning Lu and Weihua Zhuang. “Wireless access technologies for vehicular network safety applications”. IEEE Network 30.4 (2016): 22-26.

4. Xiao Liang., et al. “Learning-based VANET communication and security techniques”. Springer International Publishing (2019).
5. Ye Hao., et al. “Machine learning for vehicular networks”. arXiv preprint arXiv (2017).
6. Tan, Kang, et al. “Machine learning in vehicular networking: An overview”. Digital Communications and Networks (2021).
7. Sutton Richard S and Andrew G Barto. “Reinforcement learning: An introduction”. MIT press (2018).
8. Khatri Sahil., et al. “Machine learning models and techniques for VANET based traffic management: Implementation issues and challenges”. Peer-to-Peer Networking and Applications 14.3 (2021): 1778-1805.
9. Abdel-Halim, Islam Tharwat and Hossam Mahmoud Ahmed Fahmy. “Prediction-based protocols for vehicular Ad Hoc Networks: Survey and taxonomy”. Computer Networks 130 (2018): 34-50.
10. Sutton Richard S and Andrew G Barto. “Reinforcement learning: An introduction”. MIT press (2018).
11. Huang Chenn-Jung., et al. “A mobility-aware link enhancement mechanism for vehicular ad hoc networks”. EURASIP Journal on Wireless Communications and Networking (2008): 1-10.
12. Granelli Fabrizio, Giulia Boato and Dzmitry Kliazovich. “MORA: A movement-based routing algorithm for vehicle ad hoc networks”. IEEE Workshop on Automotive Networking and Applications (AutoNet 2006), San Francisco, USA (2006).
13. Yan SHI, Xiao-ye JIN and Shan-zhi CHEN. “AGP: an anchor-geography based routing protocol with mobility prediction for VANET in city scenarios”. The Journal of China Universities of Posts and Telecommunications 18 (2011): 112-117.
14. Park Kyeongdeuk, Hyundong Kim, and Sukyoung Lee. “Mobility state based routing method in vehicular ad-hoc network”. 2015 IEEE International Conference on Mobile Services. IEEE (2015).
15. Zhu Yanmin., et al. “Geographic routing based on predictive locations in vehicular ad hoc networks”. EURASIP Journal on Wireless Communications and Networking 1 (2014): 1-9.
16. Lo Shou-Chih and Wei-Kun Lu. “Design of data forwarding strategies in vehicular ad hoc networks”. VTC Spring 2009-IEEE 69th Vehicular Technology Conference. IEEE (2009).
17. Nikolovski Tomo and Richard W Pazzi. “Delay Tolerant and Predictive Data Dissemination Protocol (DTP-DDP) for urban and highway vehicular ad hoc networks (VANETs)”. Proceedings of the 6th ACM Symposium on Development and Analysis of Intelligent Vehicular Networks and Applications (2016).
18. Zhu Wanting, Deyun Gao and Chuan Heng Foh. “An efficient prediction-based data forwarding strategy in vehicular ad hoc network”. International Journal of Distributed Sensor Networks 11.8 (2015): 128725.
19. Jain Sushant, Kevin Fall and Rabin Patra. “Routing in a delay tolerant network”. Proceedings of the 2004 conference on Applications, technologies, architectures, and protocols for computer communications (2004).
20. Burgess John., et al. “MaxProp: Routing for Vehicle-Based Disruption-Tolerant Networks”. Infocom 6 (2006).
21. Zhu, Yanmin, Yuchen Wu, and Bo Li. “Trajectory improves data delivery in urban vehicular networks”. IEEE Transactions on Parallel and Distributed Systems 25.4 (2013): 1089-1100.
22. Talpur Anum and Mohan Gurusamy. “Machine learning for security in vehicular networks: A comprehensive survey”. IEEE Communications Surveys & Tutorials (2021).
23. Huang Jiaqi., et al. “Recent advances and challenges in security and privacy for V2X communications”. IEEE Open Journal of Vehicular Technology 1 (2020): 244-266.
24. Lytrivis Panagiotis., et al. “An advanced cooperative path prediction algorithm for safety applications in vehicular networks”. IEEE Transactions on Intelligent Transportation Systems 12.3 (2011): 669-679.
25. Yan Gongjun., et al. “Cooperative collision warning through mobility and probability prediction”. 2010 IEEE Intelligent Vehicles Symposium. IEEE (2010).
26. V Padmapriya and D N Sujatha. “A Futuristic approach for Secure Incentive Based Advertisement Distribution (SIBAD) in VANET”. Indian Technology Congress (ITC-2015).
27. V Padmapriya., et al. “Impact of Malicious Node on Secure Incentive Based Advertisement Distribution (SIBAD) in VANET”. 2017 IEEE 7th International Advance Computing Conference (IACC) (2017): 226-232.
28. V Padmapriya, DN Sujatha and KR Venugopal. “Threat on incentive based earning in VANET”. 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET) (2017): 123-128.

29. A Gopalakrishnan, P Manju Bala and T Ananth Kumar. “An Advanced Bio-Inspired Shortest Path Routing Algorithm for SDN Controller over VANET”. 2020 International Conference on System, Computation, Automation and Networking (ICSCAN) (2020): 1-5.
30. Teixeira Lincoln Herbert, and A rpa d Husza k. “Reinforcement Learning Environment for Advanced Vehicular Ad Hoc Networks Communication Systems”. Sensors 22.13 (2022): 4732.
31. Skondras Emmanouil, Angelos Michalas and Dimitrios D Vergados. “Mobility management on 5g vehicular cloud computing systems”. Vehicular Communications 16 (2019): 15-44.

Volume 3 Issue 6 December 2022