INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
Special issue on fuzzy logic dedicated to the centenary of the birth of Lotfi A. Zadeh (1921-2017)
Online ISSN 1841-9844, ISSN-L 1841-9836, Volume: 16, Issue: 1, Month: February, Year: 2021
Article Number: 4123, https://doi.org/10.15837/ijccc.2021.1.4123

V2V Routing in VANET Based on Fuzzy Logic and Reinforcement Learning

W. L. Zhang, X. Y. Yang, Q. X. Song, L. Zhao

Wanli Zhang, Xiaoying Yang*, Qixiang Song
School of Information Engineering
Suzhou University
234000 Suzhou, China
*Corresponding author: yangxiaoying@ahszu.edu.cn

Liang Zhao
School of Computer Science
Shenyang Aerospace University
110136 Shenyang, China

Abstract

To ensure the transmission quality of real-time communications on the road, the research of routing protocol is crucial to improve effectiveness of data transmission in Vehicular Ad Hoc Networks (VANETs). The existing work Q-Learning based routing algorithm, QLAODV, is studied and its problems, including slow convergence speed and low accuracy, are found. Hence, we propose a new routing algorithm FLHQRP by considering the characteristics of real-time communication in VANETs in the paper. The virtual grid is introduced to divide the vehicle network into clusters. The node's centrality and mobility, and bandwidth efficiency are processed by the Fuzzy Logic system to select the most suitable cluster head (CH) with the stable communication links in the cluster. A new heuristic function is also proposed in FLHQRP algorithm. It takes cluster as the environment state of heuristic Q-learning, by considering the delay to guide the forwarding process of the CH. This can speed up the learning convergence, and reduce the impact of node density on the convergence speed and accuracy of Q-learning. The problem of QLAODV is solved in the proposed algorithm since the experimental results show that FLHQRP has many advantages on delivery rate, end-to-end delay, and average hops in different network scenarios.

Keywords: VANETs, V2V routing, fuzzy logic, clustering, heuristic Q-learning.

1 Introduction

As the core of intelligent transportation system (ITS), VANETs [27][5] has been a research topic that attracted lots of efforts [10][19]. VANET takes the vehicles and communication infrastructure as the nodes to realize vehicle-to-vehicle communication (V2V) [4, 26] and vehicles-to-infrastructure (V2I) communication [8], thus forming a real-time communication network composed of vehicles and road infrastructure. It improves not only the safety and management efficiency of road traffic but also increases the comfort of driving. The schematic diagram of VANET road communication is shown in Figure 1. VANET is a wireless mobile network with many unique features such as temporary,

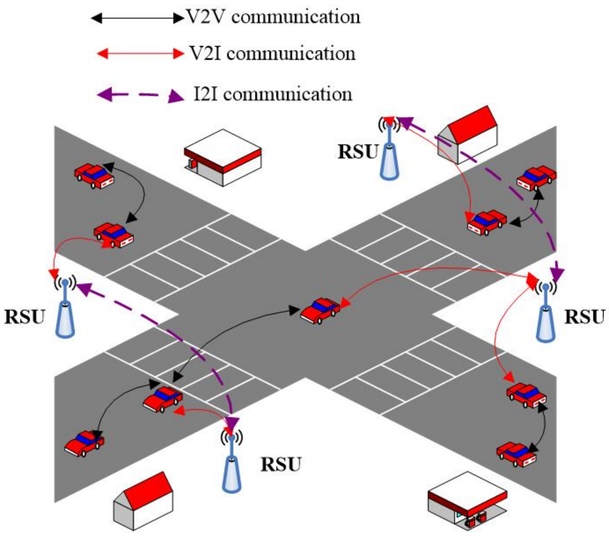

Figure 1: Schematic diagram of VANET road communication

self-organizing, and highly dynamic distributed [11][30][16][31]. It has the characteristics as follows, (1) lots of moving nodes with high-speed (2) node movement with a certain regularity, (3) frequent disconnection of node links, (4) network link is unreliable because of intermittent network connection, (5) complex communication scenarios, (6) rich external auxiliary equipment support. These characteristics make the routing protocols in traditional mobile ad hoc networks (MANETs) cannot be adopted directly in VANETs. Therefore, designing a routing protocol suitable for VANETs regarding the distinguishing characteristics in order to maximize the performance of VANETs is of great significance to give impetus to application of VANETs and the development of ITS for refining traffic problems.

Based on this, a novel routing algorithm FLHQRP based on Fuzzy Logic and Heuristic Q-learning, is proposed in this paper. In our proposal, the virtual grid is introduced to partition the vehicle network scene into clusters with three indexes, including available link bandwidth, node centrality, and node mobility. The indexes are processed by the Fuzzy Logic system for selecting the optimal Cluster Head (CH) node. The cluster is taken as Q-learning environment, in which the heuristic function is introduced to guide the forwarding action of CH nodes combined with the delay information between nodes. The main contributions of this paper are as below:

(1) A virtual grid is used to divide a VANET into clusters to deal with the problem of QLAODV is greatly affected by network topology changes.

(2) When choose CH, node centrality, node mobility, and bandwidth efficiency are processed by the Fuzzy Logic system to select the most suitable CH node to maintain stable communication with vehicles in the cluster.

(3) The heuristic function is introduced to guide the forwarding action of CH nodes combined with the delay information between nodes to accelerate the convergence speed of Q-learning.

The organization structure remainder of the paper is as below. Section 2 presents the related work. In Section 3, an improved routing algorithm–FLHQRP is introduced. In Section 4, the effectiveness of FLHQRP is evaluated by simulation experiments. Section 5 is the conclusion of this paper.

2 Related work

At present, the classification of routing protocols are mainly partitioned into the following three categories: location-based routing protocol [24][21][29], cluster-based routing protocol [1, 2, 6, 13, 15, 20, 25], and topology-based routing protocol [3, 7, 9, 22, 23, 28]. This paper briefly reviews the main VANETs routing methods. Besides, because two concepts are combined in our proposed scheme, clustering and Q-learning based routing for VANETs are studied. In [1], A VANET routing protocol based cluster was proposed, which adopted a novel addressing plan. Each node obtained an address according to its mobile mode and used Hamming distance technology to partition the network with the address as the center. In [13], a clustering-based routing protocol was proposed, which used an imperialist competition algorithm to cluster the nodes according to the node degree and vehicle speed, and the CHs were selected according to the size of the idle cushion space and the wished transmission times by using the radial primary function neural network algorithm. In [2], a reliable routing algorithm based on clustering was proposed, in which the simulated annealing algorithm was used to cluster the nodes correctly, a primary basis function neural network was used to choose the CH. In [20], a QoS based clustering algorithm was raised, with considering the compromise between QoS needs and swift move restraints, but did not consider the stability of clusters, so it does not apply in agreement for routing protocols in VANETs. In [25], an adaptive connection aware routing protocol was raised, in which the road was partitioned into several clusters, and the best route with the best network transmission quality was adaptively selected according to the vehicle density in each cluster. In [6], a new mobility-based clustering plan for VANETs was proposed, which used an affinity propagation algorithm to form clusters in a distributed manner. The algorithm considered the mobility of nodes in the process of cluster formation and produced high stability clusters, which was also robust to channel errors. In [15], a clustering approach that integrates message about way deploy vehicle mobility and chain mass as proposed.

In [22], a routing protocol QLAODV in VANET, which is suitable for unicast applications in a high mobility environment, was proposed. Q-learning algorithm was used to deduce network condition message, and the govern packet was used to inspect the usability of path in true-time. In [23], a portable VANET routing protocol that learned the best routing by fuzzy restraint Q-learning algorithm was proposed. In [28], RSAR based on the heuristic algorithm was proposed, which achieved good performance in VANET by combining reliability parameters and adjusting heuristic functions. In [9], a hierarchical protocol QGrid based on reinforcement learning was proposed, which used the minimum possible delay and hops to improve the message delivery rate. Therefore, the transmission rate has been dramatically improved. In [7], RFLQGeo, a reward function learning scheme based on Q-Learning geographical routing was proposed with strong multi-feature organization ability and improved network performance and reduced communication overhead.

In a word, these algorithms mainly center at the constancy of computing clusters or use Q-learning alone for routing selection. Different from these achievement, our plan offers a routing algorithm suitable for VANETs, which is composed of an effective clustering method and a routing algorithm. The algorithm takes into account the available link bandwidth, node centrality, and node mobility to use the Fuzzy Logic to select the best CH node and uses heuristic Q-learning to select effective routing.

3 FLHQRP routing algorithm

3.1 Clustering in FLHQRP

1. Clustering

The cluster is divided by introducing the idea of a virtual grid, which vehicle network scene is regarded as a plane and the vehicle network scene is divided into $N * N$ square grids of the same size with the side $d = \frac{\sqrt{2}}{2}R$, where R is the transmission range of the node. A cluster is represented by each grid and is assigned an ID. The vehicle node in the scene obtains its own coordinate position periodically according to the GPS global positioning device, and determines the cluster of its node. The cluster model is shown in Figure 2.

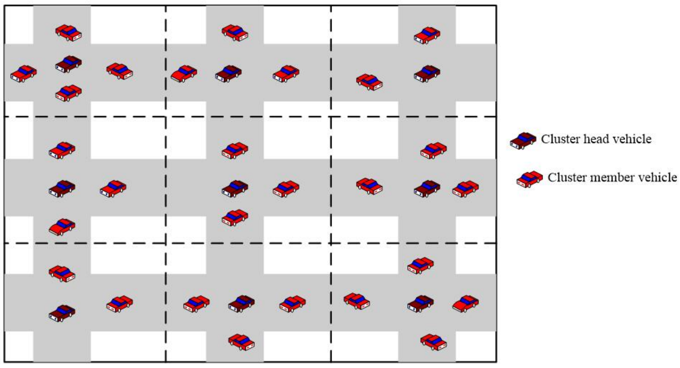

Figure 2: The cluster model

The advantage of using the partitioned grid as a cluster is that the cluster structure formed is fixed, which avoids the merging and separation process of clusters due to the movement of clusters, thus avoiding the corresponding maintenance overhead. In the ideal state, the CH node should be the node with the longest lifetime and the best connectivity in the grid. The CH node has the most nodes directly connected to it, in which the largest available bandwidth, and the CH node does not move. However, the cluster structure in the real scene may not be the same as that in the ideal state, so when selecting the CH, the algorithm tries to make the selected CH closer to the ideal state.

2. Select CH

Multimedia and other real-time services need to transmit a large number of data services, and are sensitive to delay. To ensure the transfer quality of service of this kind of real-time service, three factors, including node centrality, node mobility, and bandwidth efficiency are considered when selecting CHs. If we directly use these three parameters to establish a mathematical model for CH selection, it is hard to establish a mathematics model. It cannot adapt to the changes in the network environment because of the dynamic changes in the vehicle network topology. However, the Fuzzy Logic method can effectively simplify the process. Therefore, this paper uses the method based on Fuzzy Logic to consider these three indicators to evaluate the suitability of the CH.

Every node accesses its one-hop neighbor to ascertain which node may be the CH. The accession is performed through algorithms based on Fuzzy Logic. For each adjacent vehicle, the CH fitness of each node is calculated by the following steps, which are shown in Figure 3.

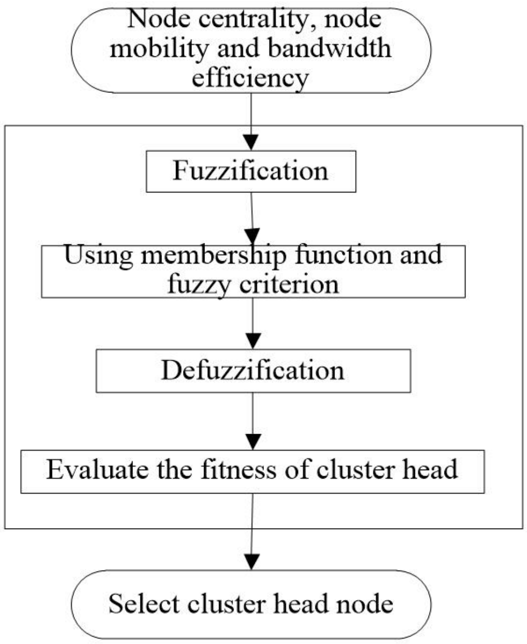

Figure 3: Steps of selecting a CH based on Fuzzy Logic

(1) Calculating Fuzzy Logic factors

• Node centrality

Vehicle nodes have social attributes, so a node centrality C is defined. The stronger the centrality of a node is, the more active it is, and the more nodes directly connected to it in the network, which implies that it is more suitable for CH and has greater connectivity. In this paper, the degree centrality proposed in reference [3] is used as a measure of node centrality. Periodically count the number of neighbor nodes C of each node in a period of time t, is the centrality of the node. If the value of C is larger, the stronger the centrality is. The calculation formula is as below:

$$C_k(i) = \sum_{j=1}^{N} a_{k,j}(i) \tag{1}$$

In Eq.(1), $C_k(i)$ express the centrality of node k at time i, whether node j is the neighbor node of node k, if it is neighbor node, then $a_{k,j}(i) = 1$, otherwise $a_{k,j}(i) = 0$. N is the total number of nodes in the cluster, and its value is updated periodically. For increasing the precision of node centrality, refer to the data $C_k(i - \Delta t)$ of the last time.

$$C_k(i) = wC_k(i) + (1 - w)C_k(i - \Delta t) \tag{2}$$

In Eq.(2), w is the weight coefficient, and its value varies with the moving speed of nodes, which is determined by the simulation environment. In order to be used as the fuzzy input value, it is normalized by Eq.(3):

$$C_{normal-k}(i) = 1 - \frac{C(i)}{N} \tag{3}$$

• Node mobility

The mobile speed of nodes has a significant effect on the change of network topology in vehicular network [17][14]. The higher the velocity of the node, the shorter the time in the cluster, so nodes with high mobility should be avoided to become CHs. When receiving the greeting message from neighbor x, the node counts the moving factor MF, as shown in Eq.(4). MF denotes the mobility standard of neighbor nodes. The bigger of MF is, the lower the mobility of the node. MF is initially 0. In Eq.(5), an index shift mean is employed on account of short-term errors are to be eliminated. β is the smoothing factor, whose value is 0.7.

$$MF(s, x) = \frac{\mid v(x) \mid - min_{y \in N_S} \mid v(y) \mid}{max_{y \in N_S} \mid v(y) \mid} \tag{4}$$

$$MF(s, x) \leftarrow (1 - \beta)MF_{i-1}(s, x) + \beta MF_i(s, x) \tag{5}$$

In Eq.(4), $v(x)$ denotes the moving rate of vehicle x, N_s is the neighbor set of nodes s. In Eq.(5), $MF_{i-1}(s, x)$ is the value of MF in the previous time, and $MF_i(s, x)$ is the value of MF in the current time.

• Bandwidth efficiency

The larger the available bandwidth, the more packets it can transmit [**?**][12], and the better for the CH node. The available bandwidth of the link can be accurately estimated by using the periodic carrier to detect the idle channel time.

$$BW_{available} = (1 - K)(1 - P)(1 - ADK)(1 - RD)\frac{T_{si}}{T}\frac{T_{di}}{T}D \tag{6}$$

In Eq.(6), K denotes the idle duty cycle of the channel consumed by the waiting backoff process, P denotes packet collision probability, ADK refers to Idle channel duty cycle of acknowledgment information consumption, RD is the idle channel duty cycle consumed by $\frac{RTS}{CTS}$ during data transmission, denote the idle time of the sensing channel of the endogenous node and the destination node in the monitoring period T respectively, and D is the channel capacity, which is normalized by Eq.(7). In order to eliminate the error caused by selecting the monitoring period, Eq.(8) is used to calculate the normalized available bandwidth of neighbor node link.

$$BW = \frac{BW_{available}}{BW_{max}} \tag{7}$$

$$BW_{s,k}(i) = \delta BW_{s,k}(i) + (1-\delta)BW_{s,k}(i-1) \tag{8}$$

where, s is the sending node, k is its one-hop neighbor node, $BW_{s,k}(i)$ is the normalized available bandwidth in the ith monitoring period, $BW_{s,k}(i-1)$ is the normalized available bandwidth of the $(i-1)th$ monitoring cycle, and δ is the influence factor, which is set according to the actual simulation environment, with value of 0.8 in this paper.

(2) Fuzzyfication

Fuzzy is defined as the procedure of changing numerical value into fuzzy value by using the Fuzzy Membership Function (FMF), which is shown in Fig. 3. The three normalized parameters of node centrality, node mobility, and bandwidth efficiency are calculated by using Eqs.(3),(5) and (8). According to the characteristics of real-time services such as multimedia and vehicular ad hoc network, the three parameters are defined by fuzzy sets.

According to the different network environment of the city, different fuzzy membership functions are determined. It is enough to select the best link from multiple links and use triangle or trapezoid membership function. Node centrality factor Low, Medium, High, the FMF of which is shown in Fig. 4. Node mobility factor slow, medium, fast, and FMF of which is shown in Fig. 5. The value of available bandwidth small, medium, large, and the FMF of which is shown in Fig. 6, nodes with small available bandwidth should be avoided as far as possible. If the available bandwidth of a node is greater than or equal to 0.5, the membership degree of mapping available bandwidth to large is 1.

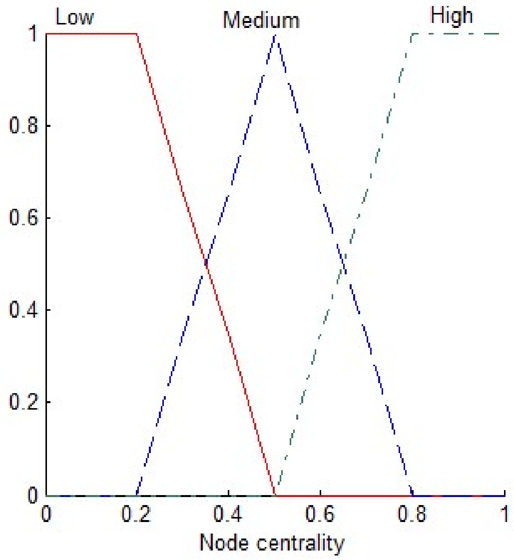

Figure 4: FMF of node centrality

(3) Fuzzy Logic criterion mapping

According to the characteristics of real-time services such as multimedia and vehicular network, combined with the Fuzzy Logic criteria in Table 1, each node calculates the rank of vehicle node as CH. Use very bad; bad; not accept; accept; good; perfect to represent the level of the mapped node as the CH. Since multiple rules can be applied simultaneously, the min-max method is used to make up their assessment outcomes. In min-max method, for each rule, the minimum value of antecedent is used as final degree. When combining different rules, you need to use the maximum value of the result.

For example, if there is a neighbor node with mobility factor Slow:0.75, Medium:0.25, Fast:0, node centrality factor high:0.75, medium:0.25, low:0, bandwidth efficiency factor large:0.5, medium:0.5, small:0, Then the level of the node as the CH is mapped to the criteria 1,2,4,5 in Table 1, and the membership degree of the fuzzy level is respectively perfect:0.5, good:0.5, accept:0.25. The specific process is shown in Figure 7.

Table 1: Fuzzy Logic criteria

	Centrality	Mobility	Bandwidth efficiency	Level
Rule 1	High	Slow	Large	Perfect
Rule 2	High	Slow	Medium	Good
Rule 3	High	Slow	Small	Not accept
Rule 4	High	Medium	Large	Good
Rule 5	High	Medium	Medium	Accept
Rule 6	High	Medium	Small	Bad
Rule 7	High	Fast	Large	Not accept
Rule 8	High	Fast	Medium	Bad
Rule 9	High	Fast	Small	Very bad
Rule 10	Medium	Slow	Large	Good
Rule 11	Medium	Slow	Medium	Accept
Rule 12	Medium	Slow	Small	Bad
Rule 13	Medium	Medium	Large	Accept
Rule 14	Medium	Medium	Medium	Not accept
Rule 15	Medium	Medium	Small	Bad
Rule 16	Medium	Fast	Large	Bad
Rule 17	Medium	Fast	Medium	Bad
Rule 18	Medium	Fast	Small	Bad
Rule 19	Low	Slow	Large	Not accept
Rule 20	Low	Slow	Medium	Bad
Rule 21	Low	Slow	Small	Very bad
Rule 22	Low	Medium	Large	Bad
Rule 23	Low	Medium	Medium	Bad
Rule 24	Low	Medium	Small	Very bad
Rule 25	Low	Fast	Large	Bad
Rule 26	Low	Fast	Medium	Very bad
Rule 27	Low	Fast	Small	Very bad

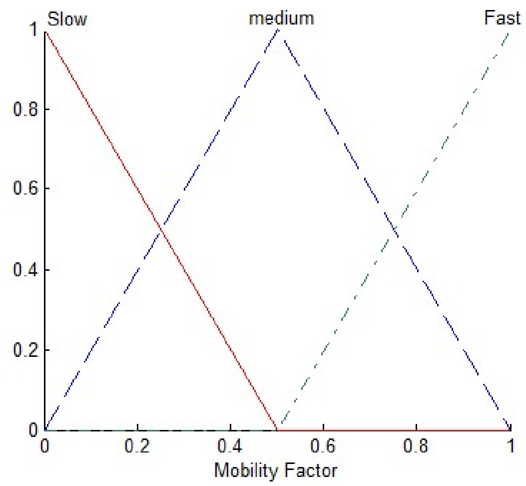

Figure 5: FMF of moving factor

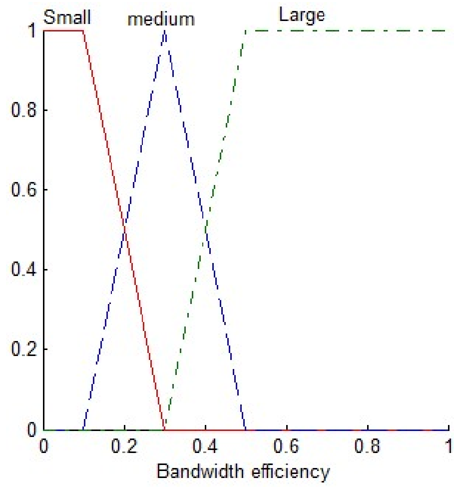

Figure 6: FMF of bandwidth efficiency

(4) Defuzzyfication

Defuzzification is a procedure of creating numerical outcomes on account of the OMF and the relevant membership degree. The definition of the Output Membership Function (OMF) is shown in Fig. 8. In this paper, the barycenter method is used to defuzzify the mapping results. That is, according to the corresponding degree, the membership function can be output by horizontal WEDM (as shown in Fig. 9), and the top can be removed. The horizontal axis value of the center of gravity of the shadow part is obtained, which is the fitness of the vehicle as the CH node. If $\mu(x)$ is applied to express the outcome function, and x is applied to express the horizontal axis, the center of gravity will be counted as below:

$$COG = \frac{\int \mu(x)x\mathrm{d}x}{\int \mu(x)\mathrm{d}x} \tag{9}$$

3. Formation of cluster

The formation of the cluster starts from the initialization of the network. The nodes in the network can obtain the message of the neighbor nodes by the reciprocal process of hello packets, and use this information to determine the relationship between the nodes in the cluster. hello package is the basis of cluster structure, and its main properties are shown in Table 2.

Neighbor list is the infrastructure of cluster head selection and cluster maintenance. When a node in the network receives a hello packet periodically broadcast by a neighbor node, it updates the corresponding table entries in the neighbor list according to the information in the Hello packet. The main properties of its structure are shown in Table 3.

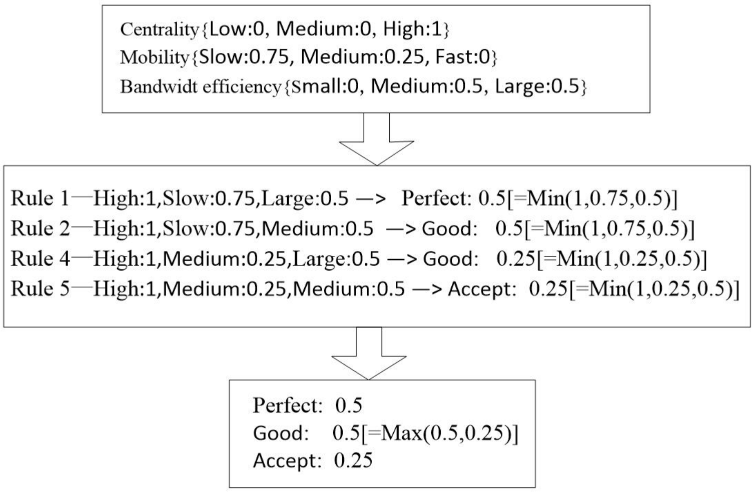

Figure 7: Example of Fuzzy Logic rule mapping

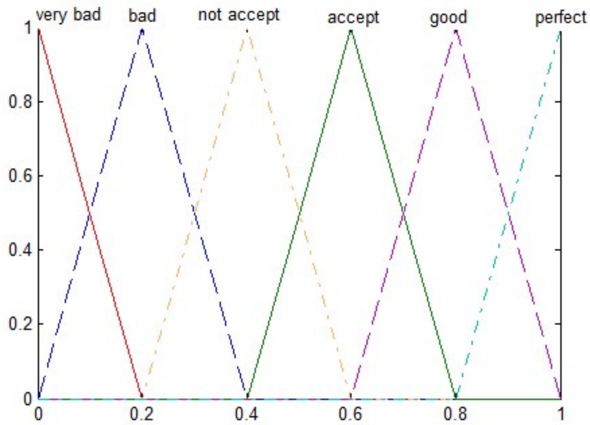

Figure 8: The OMF

The cluster formation steps in *FLHQRP* routing algorithm are as below.

Step 1: When the network is initialized, each vehicle node sets its node role as a cluster member node, which interacts with neighbor nodes through *Hello* package, including node status, node location, node speed, node CH fitness, neighbor list, and neighbor cluster table information.

Step 2: After receiving *Hello* packets sent by neighbor nodes, the node updates its neighbor list and neighbor cluster table. If without getting *Hello* packets from a neighbor node within a certain time, the information about the neighbor node in the neighbor list and neighbor cluster table will be deleted.

Step 3: Each node periodically calculates its CH fitness and compares it with the neighbor nodes with the same cluster number in the neighbor list. If the CH of this node has the highest fitness, the node state will be changed to the CH node, and the CH notification message will be broadcast to the neighbor nodes; otherwise, the CH node will wait for the CH node to broadcast the CH notification message and keep the cluster member node status.

Step 4: The node gets the CH notification information and sets the corresponding node status in the neighbor list as the CH node.

4. Cluster maintenance

Table 2: Structure of hello package

Node IP	Node Role	Cluster ID	Position	Velocity	Weight	Neighbor Table	Neighbor Cluster

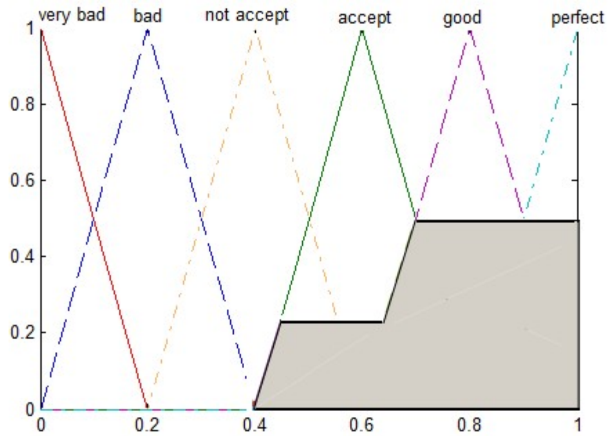

Figure 9: Example of barycenter method

Table 3: Structure of Neighbor list

Neighbor Node IP	Node Role	Cluster ID	Position	Velocity	Weight

Because the structure of the cluster of the algorithm proposed in the paper is fixed, it avoids that the CH competition event caused by the adjacent CHs due to the movement of vehicle nodes. All nodes in the cluster only need to periodically compare their CH fitness with that of other nodes with the same cluster ID. As the CH node discovers that its CH fitness is the highest, it maintains the CH state without any action; when the CH node finds that its fitness is not the highest, it abandons the CH status, waits for the CH notification message, and forwards the Q value table saved by itself to the new CH node; when the cluster member node discovers that its CH fitness is the largest, it updates the node status to the CH node, and broadcasts the CH notification message; when the cluster member node's fitness is not the highest, the cluster member state remains unchanged. When the cluster member node receives the new CH notification message, the role status of the corresponding node in the neighbor list is set as the CH node, and the role state of the original CH node in the neighbor list is set as the cluster member node.

3.2 Method modeling of heuristic Q-learning in FLHQRP algorithm

In the proposed algorithm, the vehicular network environment is regarded as the heuristic Q-learning environment, and the cluster is regarded as state of the agent. The routing exploration process is carried out through the information interaction between CH nodes. For this reason, the proxy's quad is redefined as follows:

$< S, A, R, V >$

State space S: all clusters constitute state space.

Action set A: the set of adjacent clusters that can be selected in packet forwarding.

Immediate reward R: an immediate reward for forwarding packets to adjacent clusters.

Objective function V: the discount cumulative return when the adjacent cluster is selected for forwarding.

The online learning process of heuristic Q-learning is completed by CH. As the management role of the cluster, the CH knows the coordinate position and motion state of all cluster members. According to the information, the quality of the cluster is calculated, and the calculated value is reflected in the learning rate, which affects the online learning process of Q-learning. The Q value update formula of the algorithm is as follows:

$$Q_s(d,x) \leftarrow (1-\alpha)Q_s(d,x) + \alpha\{R + \gamma \max_{y\in\tau(x)} Q_s(d,y)\} \tag{10}$$

In Eq.(10),s is the current cluster, $Q_s(d,x)$ is the corresponding Q value that the current cluster

Table 4: Fuzzy Logic criteria

Neighborhood cluster	Target cluster	Target cluster	Target cluster	Target cluster
	d_1	d_2	...	d_n
x_1	$Q(d_1, x_1)$	$Q(d_2, x_1)$	...	$Q(d_n, x_1)$
x_2	$Q(d_1, x_2)$	$Q(d_2, x_2)$	...	$Q(d_n, x_2)$
...	...	...	...	...
x_n	$Q(d_1, x_n)$	$Q(d_2, x_n)$	...	$Q(d_n, x_n)$

s selects the adjacent cluster x as the next forwarding cluster reaching the destination cluster d, and $\tau(x)$ denotes the adjacent cluster set of clusters x. γ is the discount rate, which value is 0.7. R is the immediate reward obtained when selected adjacency cluster x by current cluster s as the next forwarding cluster for arriving at destination cluster d, which is defined as follows:

$$R = \begin{cases} 1, if \quad s \in N_d \\ 0, otherwise \end{cases} \tag{11}$$

In Eq.(11), N_d represents the adjacency cluster set of destination cluster d. If the adjacent cluster is just the target cluster, the reward value is 1, otherwise, it is 0.

The calculation of learning rate α mainly considers the traffic density in the cluster. The higher the traffic flow density is, the better the connectivity of the nodes in the cluster is, the faster the learning speed is, and it is more suitable to participate in the forwarding of data packets. The calculation formula is as follows:

$$\rho = 1 - \frac{v_a}{v_{max}} \tag{12}$$

In Eq.(12), v_{max} is the maximum running speed of the vehicle, v_a is the average speed of all vehicle nodes in the cluster. Each CH node in the vehicular network saves a Q-value table to assist in finding the optimal path in the routing discovery process, as shown in Table 4.

Each node of the CH adds the maximum Q-value information to all the destination clusters in its Q-value table to the packet, and broadcasts it to the surrounding neighbor clusters. The head of the current cluster node gets the *Hello* packet from the adjacent CH node, and updates the Q-table according to Eq.(10) to realize the online learning process of heuristic Q-learning.

3.3 Route discovery process of FLHQRP algorithm

In FLHQRP algorithm, when selecting the forwarding group of the neighbor cluster, the neighbor cluster forwarding group corresponding to the maximum Q-value is not directly selected, but the greedy rule is used, as shown in Eq.(13):

$$a = \begin{cases} arg \max\limits_{x \in \tau(s)} [Q_s(d,x) + \varepsilon H_s(d,x)], if \quad q \leq p \\ \qquad\qquad\qquad\qquad a_{random}, otherwise \end{cases} \tag{13}$$

where, $Q_s(d, x)$ represents the Q-value of that the current CH s selects the adjacent CH x as the next forwarding CH and forwards the packet to the head of the target CH d. $H_s(d, x)$ denotes the heuristic function that inspires the current optimal action. a_random represents randomly selecting a neighbor CH to forwarding packets. ε is a real variable which is used to weigh the influence of heuristic function. p is the proportion of exploration and utilization, which means that CH s selects the next CH by using Q-value with probability p, that is to implement utilization strategy, and randomly selects the next CH with a probability of $1 - p$, that is to implement exploration strategy. The larger the value of p is, the smaller the probability of random selection is, with value of 0.9 in this paper. q is a random number of [0,1]. Each CH adopts the utilization exploration balance strategy, implements the utilization strategy or the exploration strategy, and forwards the packet to the next CH until the packet reaches the target CH d. Simultaneously, FLHQRP records the delay of each CH node in each packet, and determines the current optimal action of each CH node on the path according to the delay information between these CH nodes, as shown in Eq.(14):

$$TD(s,d,x) = \sum_{n=x}^{d-1} T(n,n+1) \tag{14}$$

where, $TD(s,d,x)$denotes the delay of transmitting packets to destination CH d by selecting x as the next CH starting from CH s.

When using the balance strategy of utilization-exploration, the group will arrive at the destination CH node along different paths, and the feedback information will return to the source CH node along different paths. For CH node s, it may receive multiple feedback information from different paths of the same target CH node. Among them, the feedback information with the shortest delay is the current optimal path from CH node s to target CH node, and the next CH corresponding to the path is the current optimal action of grouping from CH node s to the target CH node. As shown in Eq.(15):

$$a_{optimal} = \min_{x \in \tau(s)} TD(s,d,x) \tag{15}$$

where, $a_{optimal}$ represents the optimal action, that is, CH s transmits packets to the destination CH d with current optimal next CH node. $\tau(s)$ is the set of adjacent CHs of CH s. By Eq.(15), the current optimal action of current CH node s is established. Whenever feedback information reaches s, it recalculates the current optimal action.

When the optimal next CH node is determined, the heuristic function $H_s(d,x)$ inspires the current CH node to choose present best next CH node and update the corresponding Q-value. The value of the heuristic function $H_s(d,x)$ affects the choice of actions and must be as low as possible in order to minimize errors. It is defined as follows:

$$H_s(d,x) = \begin{cases} \max\limits_{y \in \tau(s)} [Q_s(d,y) - Q_s(d,x) + \eta], if \quad x = a_{optimal} \\ \qquad\qquad\qquad\qquad\qquad 0, otherwise \end{cases} \tag{16}$$

where, $Q_s(d,x)$ represents the Q-value of current CH s reaching the destination CH through adjacent CH x. $\max_{y \in \tau(s)} Q_s(d,y)$ denotes the maximum Q-value of all adjacent CHs of CH s to reach the destination CH. Eq.(16) shows that if the adjacent CH x is the current optimal next CH node, it is given an appropriate heuristic value to guide the current CH node to select the present optimal next neighbor CH node; otherwise, the heuristic value of adjacent CH x is 0. η is a very small positive real number, which is generally taken as 0.01.

The specific route discovery process is as follows:

Step 1: As the source node dispatches packets to the target node, it first determines whether there is routing information, and if so, starts to transmit data. Otherwise, judge the role of the node itself. If it is a cluster member node, the routing discovery process is started, generating RREQ and broadcast the packet, and go to step 2. If it is a CH node, select and send *RREQ* to the next adjacent CH node according to the utilization exploration balance strategy shown in Eq.(13), and then turn to step 3.

Step 2: The intermediate node receives the *RREQ* packet. If it is a cluster member node, it will directly discard the packet. If it is a CH node, it will go to step 3.

Step 3: When the CH node gets the *RREQ* packet, the delay between it and the previous CH node is calculated and recorded in the packet. Judge if the target node is a neighbor node. If yes, the packet is forwarded to target node and go to step 4; otherwise, select and dispatch it to the next adjacent CH node according to the utilization exploration balance strategy shown in Eq.(13), and repeat this step.

Step 4: After receiving the RREQ packet, if the delay information among nodes shows that the packet takes less time than the packet in the previous period, the *RREQ* is generated by the destination node, and the delay information between the CH nodes of the packet and the Q-value of the destination

Table 5: Simulation parameter setting

Parameter	Value
Simulation scene	2000*2000
Maximum node speed (m/s)	15
MAC Protocol	IEEE 802.11
Transmission range	250m
Simulation time	200s
Size of CBR packet (byte)	512
Data rate (packet/s)	10
Pause probability	0.2
Max pause	10s
Number of nodes	100-300

CH node are added to *RREQ* and transferred to step 5; otherwise, the *RREQ* is sent to the CH node and transferred to step 5.

Step 5: After receiving the *RREQ*, the CH node establishes a reverse route to determine if the source node s is a neighbor node. If yes, it forwards the RREP to the source node and turns to step 7; otherwise, the CH node determines the current optimal action by the delay information of *RREQ* with Eq.(15) in the first. Then Eq.(16) is used to calculate the heuristic value of the current optimal action, and then Eq.(10) is used to update its Q-value table. Finally, the Q-value of feedback information is modified to the maximum Q-value from itself to the destination CH, and the optimal action is chosen as the next forwarding CH, and RREP is forwarded to the CH node of the cluster through the gateway node.

Step 6: When the gateway node adjacent to the CH receives the *RREQ*, it forwards the *RREQ* to the adjacent CH, establishes the reverse route, and goes to step 5.

Step 7: After the source node receives *RREQ*, the route is established and starts to transmit data.

4 Experimental simulation and result analysis

1 Experimental parameter setting

To assess the performance of the algorithm, the Manhattan simulation model is used as the simulation scenario, and the node data flow is generated randomly by well-known NS2. Three routing protocols with *QLOADV*, *CBRP*, and *FLHQRP* are simulated, respectively. The configuration of basic network parameters is shown in Table 6.

2. Simulation results and experimental analysis

In order to accurately evaluate the performance of the algorithm, the improved FLHQRP routing algorithm is compared with QLAODV algorithm and CBRP algorithm under different vehicle node numbers. The experimental metrics include packet delivery rate, average end-to-end delay, average hops and routing control overhead.

(1) Performance comparison of packet delivery rate

Packet delivery rate represents the number of successful packets obtained by the target node under the effective time. The higher the packet delivery rate, the better the quality of data transmission.

As found out from Figure 10, FLHQRP has many advantages in the packet delivery rate performance. This is mainly because the QLAODV algorithm takes nodes as Q-learning state, and its convergence rate is affected by the number of nodes. The online learning results of Q-learning cannot reflect the network topology at that time. The FLHQRP algorithm takes the cluster as the state, which can reduce the influence of the node number on the convergence speed and accuracy of the algorithm. Through the exploration-utilization balance strategy, it explores the path with shorter delay and more stable and determines the current optimal next CH node, which can inspire the node to choose the present optimal next CH node to forward packets and update the Q-value. The convergence speed of Q-learning is improved for quickly responding to the topology changes, which is conducive to the smoothly delivery of packets to the destination node, therefore improves the communication performance. Compared with CBRP algorithm, FLHQRP algorithm has certain advantages in the packet delivery rate. The main reason is that CBRP algorithm does not take the centrality and mobility of nodes in the cluster into account, which may lead to the failure to choose the proper next node caused by poor connectivity of intermediate nodes in the routing or frequent disconnection of routes caused by moving fast. The FLHQRP algorithm comprehensively considers these factors and selects the more stable nodes as CH, which can maintain the stability of data transmission between the CH and the members of the cluster, which improve the quality of the transmission link. Moreover, through the online learning process of heuristic Q-learning, it can ensure that a more stable and better route is selected.

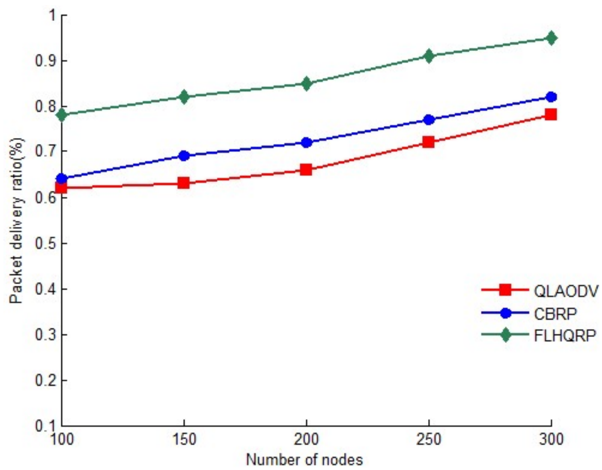

Figure 10: Comparison of packet delivery rates

(2) Performance comparison of mean end to end delay

The mean end-to-end delay expresses the mean time taken for data packets to be transmitted from the sending node to target one, which reflects the transmission efficiency of the routing protocol. The lower it is, the better the quality of data transmission.

As found in Figure 11, the time delay of FLHQRP algorithm is shorter than the other two algorithms. This is mainly because QLAODV algorithm takes nodes as Q-earning state. As node number is large, the algorithm's accuracy is low, and the established route is not necessarily the optimal path, which leads to an increase of delay. The FLHQRP algorithm takes cluster as the state space of Q-learning, the CH node is guided to use path with the shortest transmission delay to transmit packets through the heuristic function of heuristic Q-learning, which effectively improves the convergence rate and precision of the algorithm. FLHQRP algorithm has more advantages in delay performance than CBRP, which is because the CH selected by adopting the Fuzzy Logic scheme in FLHQRP algorithm can maintain stable communication with vehicles in the cluster, by reducing the number of route maintenance processes due to route interruption. It also reduces the transmission delay with precondition is guaranteeing the transmission quality, to improve delay performance.

(3) Performance comparison of average hops

The average hop count represents the times a packet forwarded from source to target.

A comparison of the average hops in different scenarios with the different number of vehicle nodes among FLHQRP, QLAODV, and CBRP is shown in Figure 12. The average hop count of FLHQRP is lower because FLHQRP uses feedback information to inspire nodes to select routing paths with smaller end-to-end delay.

(4) Performance comparison of routing cost

Routing cost represents the routing control packet amount need to be generated to transmit a group of data successfully. The lower the routing overhead is, the fewer link resources are consumed, and the better the packet transmission is.

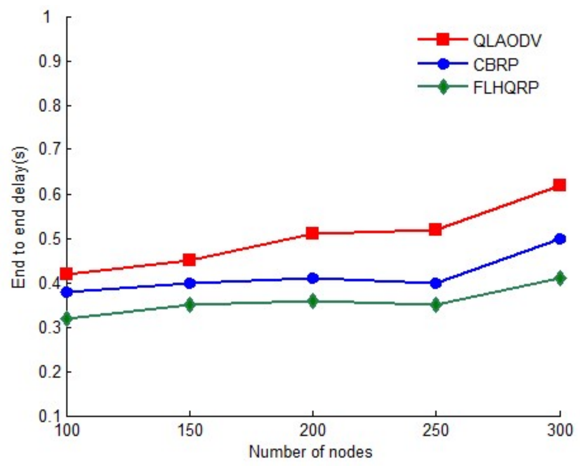

Figure 11: Comparison of average end to end delay

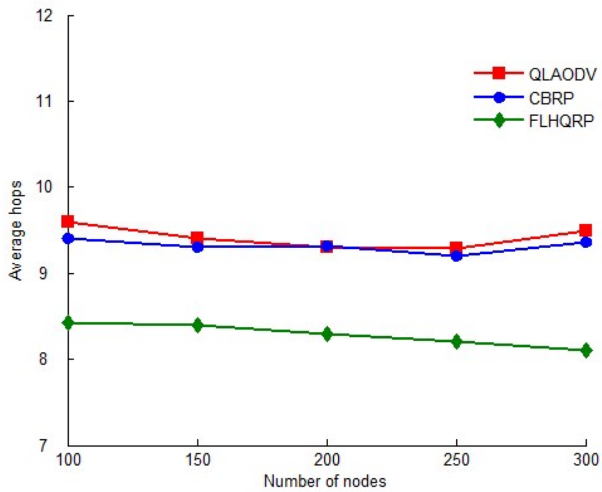

Figure 12: Comparison of average hops

The routing control cost performance of the three routing algorithms is shown in Figure 13. FLHQRP algorithm has a significant advantage in the performance of routing control overhead. This is because FLHQRP algorithm does not broadcast routing control packets with flooding in the whole scenario like that of QLAODV and CBRP algorithms but broadcasts RREQ between gateway nodes and CHs, which effectively reducing the routing control overhead. Besides, the FLHQRP algorithm clusters the network according to the grid, and CH nodes number is fixed, so the routing control overhead performance is less affected by the network size. Moreover, nodes amount in the cluster, node speed, and bandwidth efficiency are comprehensively considered in choosing CH nodes in FLHQRP which can ensure that the routing is established in the cluster with more robust connectivity and stability, and reduces the routing maintenance cost caused by the routing interruption.

(5) Performance comparison of routing protocols under different vehicle maximum moving speed

In VANET, Vehicles' speed has a certain influence on the performance of the routing. With the increase of speed, the number of suitable nodes and paths participating in forwarding decreases. Therefore, the link transmission node is easy to be disconnected, resulting in the loss of transmission data and the decrease of successful packet transmission rate, the performance comparison of packet delivery ratio of three routing are shown in Figure 14.

As can be seen from Figure 14, with the increase of node speed, the packet delivery rates of the three algorithms show a downward trend. At the same speed, FLHQRP algorithm considers the mobility of nodes when selecting cluster head nodes, so the packet transmission rate is the highest.

It can be seen from Figure 15, with the increase of node speed, the end-to-end delay of the three algorithms increases gradually. At the same speed, the end-to-end delay of FLHQRP algorithm is

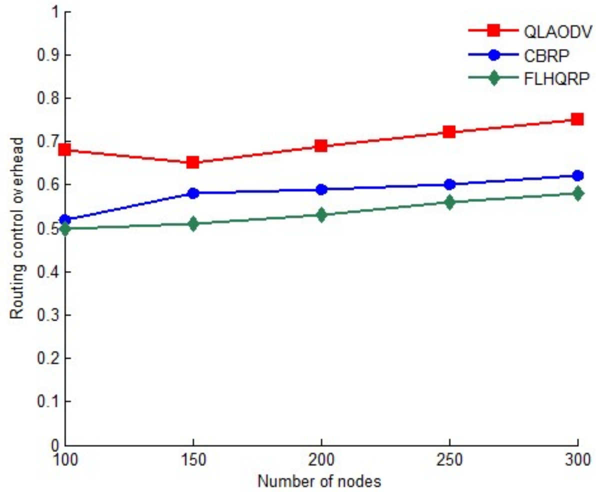

Figure 13: Comparison of routing overhead

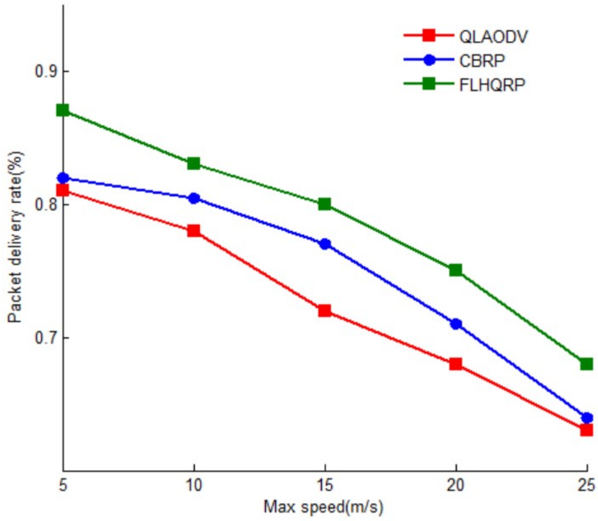

Figure 14: Comparison of packet delivery rates

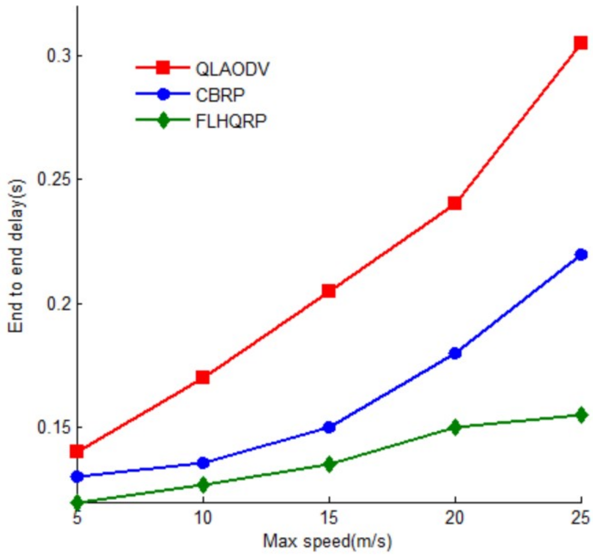

Figure 15: Comparison of average end to end delay

the minimum, because the cluster head selected by FLHQRP algorithm using fuzzy logic scheme can maintain stable communication with the vehicles in the cluster, reduce the number of route maintenance process after route interruption, and reduce the transmission delay on the premise of ensuring the transmission quality.

5 Conclusion

Aiming at the problem that the state space of QLAODV algorithm is too large in the vehicular network scene with dense nodes, the convergence rate and algorithm's precision are reduced. The virtual grid is introduced to divide the cluster, and the CH selection method is improved. The cluster is used as the state space of heuristic Q-learning, and the routing algorithm FLHQRP based on heuristic Q-learning and clustering is obtained which cut down the influence of convergence rate and precision by nodes amount. The proposed CH selection algorithm comprehensively considers vehicle centrality, mobility, and bandwidth efficiency, and chooses the best CH node based on fuzzy output ensure to choose a more stable route. The simulation results show FLHQRP is better than QLAODV and CBRP algorithm in performance, which indicates that FLHQRP can effectively deal with the vehicular network environment with frequent changes in the network topology. As the algorithm proposed in this paper is based on virtual grid clustering, the cluster structure size is fixed. The cluster size can be adaptive changed with the density of nodes in the vehicular network scene, which can make follow-up improvement to this problem.

Funding

This work was supported by the third batch of reserve candidates for academic and technical leaders (2018XJHB07), Teaching research project of Suzhou University (szxy2020jyxm06), Key curriculum construction project (szxy2018zdkc19), Large scale online open course (MOOC) demonstration project (2019mooc300, 2019mooc318), Professional leader of Suzhou University (2019XJZY22), Anhui province's key R&D projects include Dabie Mountain and other old revolutionary base areas, Northern Anhui and poverty-stricken counties in 2019 (Grant No.: 201904f06020051), Key scientific research project of Suzhou University (2019yzd04).

Author contributions

The authors contributed equally to this work.

Conflict of interest

The authors declare no conflict of interest.

References

[1] Ardakani, S. P. (2018). ACR: A Cluster-based routing protocol for VANET, *Journal of Wireless and Mobile Networks*, 10(2), 39–53, 2018.

[2] Bagherlou, H.; Ghaffari, A. (2018). A routing protocol for vehicular ad hoc networks using simulated annealing algorithm and neural networks, *The Journal of Supercomputing*, 74(6), 2528–2552, 2018.

[3] Bof, N.; Baggio, G.; Zampieri, S. (2016). On the role of network centrality in the controllability of complex networks, *IEEE Transactions on Control of Network Systems*, 4(3), 643–653, 2016.

[4] Gustafson, C.; Mahler, K.; Bolin, D.; Tufvesson, F. (2020). The COST IRACON geometry-based stochastic channel model for vehicle-to-vehicle communication in intersections, *IEEE transactions on vehicular technology*, 69(3), 2365–2375, 2020.

[5] Hammood, O. A.; Nizam, M.; Nafaa, M.; Hammood, W. A. (2019). RESP: Relay suitability-based routing protocol for video streaming in vehicular Ad hoc networks, *International Journal of Computers Communications and Control*, 14(1), 21–38, 2019.

[6] Hassanabadi, B.; Shea, C.; Zhang, L.; Valaee, S. (2014). Clustering in vehicular ad hoc networks using affinity propagation, *Ad Hoc Networks*, 13, 535–548, 2010.

[7] Jin, W.; Gu, R.; Ji, Y. (2019). Reward function learning for q-learning-based geographic routing protocol, *IEEE Communications Letters*, 23(7), 1236–1239, 2019.

[8] Kanumalli, S. S.; Chinta, A.; Chandra Murty, P. S. R. (2019). Isolation of wormhole attackers in IOV using WPWP packet, *Revue d'Intelligence Artificielle*, 33(1), 9–13, 2019.

[9] Li, F.; Song, X.; Chen, H.; Li, X.; Wang, Y. (2018). Hierarchical routing for vehicular ad hoc networks via reinforcement learning, *IEEE Transactions on Vehicular Technology*, 68(2), 1852–1865, 2018.

[10] Lipman, T.; Rodier, C.; Shaheen, S.; Finson, R. (2010). Intelligent transportation systems, *Plant Physiology*, 66(1), 93–96, 2010.

[11] Liu, Q.; He, X.; Guan, F. W.; Zhao, Y. C.; Jiang, F.; Tian, F. X.; Wang, S. X. (2019). Method and implementation of improving the pointing accuracy of an optical remote sensor using a star sensor, *Traitement du Signal*, 36(4), 311–317, 2019.

[12] Lu, K. J.; Nguyen, T.; Tran, N.; Karacolak, T. (2020). Parasitic spirals for enhancing bandwidth of a simultaneous transmit and receive patch antenna, *Microsystem Technologies*, 1–6, 2020.

[13] Mohammadnezhad, M.; Ghaffari, A. (2019). Hybrid routing scheme using imperialist competitive algorithm and RBF neural networks for VANETs, *Wireless Networks*, 25(5), 2831-2849, 2019.

[14] Pourghebleh, B.; Navimipour, N. J. (2019). Towards efficient data collection mechanisms in the vehicular ad hoc networks, *International Journal of Communication Systems*, 32(5), 1–20, 2019.

[15] Rivoirard, L.; Wahl, M.; Sondi, P.; Berbineau, M.; Gruyer, D. (2018). Chain-Branch-Leaf: A clustering scheme for vehicular networks using only V2V communications, *Ad Hoc Networks*, 68, 70–84, 2018.

[16] Ren, J.; Huang, S. Y.; Song, W.; Han, J. (2019). A novel indoor positioning algorithm for wireless sensor network based on received signal strength indicator filtering and improved Taylor series expansion, *Traitement du Signal*, 36(1), 103–108, 2019.

[17] Sachan, V.; Kumar, I.; Shankar, R.; Mishra, R. K. (2018). Analysis of transmit antenna selection based selective decode forward cooperative communication protocol, *Traitement du Signal*, 35(1), 47–60, 2018.

[18] Shimada, R.; Gassho, A.; Matsubara, N. (2019). Comparison between the impression of hot-cold and warm-cool in unspecific scale on the evaluation oftheexperimental combined environment, *Journal of Environmental Engineering* , 84(766), 1041–1050, 2019.

[19] Sodhro, A. H.; Obaidat, M. S.; Abbasi, Q. H.; Pace, P.; Pirbhulal, S.; Fortino, G.; Qaraqe, M. (2019). Quality of service optimization in an IOT-driven intelligent transportation system, *IEEE Wireless Communications*, 26(6), 10–17, 2019.

[20] Wahab, O. A.; Otrok, H.; Mourad, A. (2013). VANET QoS-OLSR: QoS-based clustering protocol for vehicular ad hoc networks, *Computer Communications*, 36(13), 1422–1435, 2013.

[21] Wang, F. F.; Hu, H. F. (2019). An energy-efficient unequal clustering routing algorithm for wireless sensor network, *Revue d'Intelligence Artificielle*, 33(3), 249–254, 2012.

[22] Wu, C.; Kumekawa, K.; Kato, T. (2010). Distributed reinforcement learning approach for vehicular ad hoc networks, *IEICE transactions on communications*, 93(6), 1431–1442, 2010.

[23] Wu, C.; Ohzahata, S.; Kato, T. (2013). Flexible, portable, and practicable solution for routing in VANETs: A fuzzy constraint Q-learning approach, *IEEE Transactions on Vehicular Technology*, 62(9), 4251–4263, 2013.

[24] Xiao, D. G.; Peng, L. X.; Song, D. (2012). Improved GPSR routing algorithm in hybrid VANET environment, *Journal of Software*, 33(3), 249–254, 2012.

[25] Yang, Q.; Lim, A.; Li, S.; Fang, J.; Agrawal, P. (2010). ACAR: Adaptive connectivity aware routing for vehicular ad hoc networks in city scenarios, *Mobile Networks and Applications*, 15(1), 36–60, 2010.

[26] Yang, X. Y.; Zhang, W. L.; Lu, H. M.; Zhao, L. (2020). V2V Routing in VANET Based on Heuristic Q-Learning, *International Journal of Computers Communications & Control*, 15(5), 3928, 2020.

[27] Zang, H. J.; Huang, Y.; Cao, H. B.; Li, C. C. (2019). A novel privacy protection protocol for vehicular ad hoc networks based on elliptic curve bilinear mapping, *Ingenierie des Systemes d'Information*, 24(4), 397–402, 2019.

[28] Zhang, D.; Zhang, T.; Liu, X. (2019). Novel self-adaptive routing service algorithm for application in VANET, *Applied Intelligence*, 49(5), 1866–1879, 2019.

[29] Zhang, X. L.; Qian, Z.; Zhang, T.(2012). Improved GPSR-SD routing protocol for VANET, *Journal of Highway and Transportation Research and Development*, 11(4), 98–103, 2017.

[30] Zhao, L.; Han, G.; Li, Z.; Shu, L. (2020). Intelligent digital twin-based software-defined vehicular networks, *IEEE Network*, 34(5), 178–184, 2020.

[31] Zhao, L.; Liu, Y.; Al-Dubai, A.Y.; Zomaya, A.; Hawbani, A. (2020). A novel generation adversarial network-based vehicle trajectory prediction method for intelligent vehicular networks, *IEEE Internet of Things Journal*, in press, 2020.

Cite this paper as:
Zhang, W. L.; Yang, X. Y.; Song, Q. X. ; Zhao, L. (2021). V2V Routing in VANET Based on Fuzzy Logic and Reinforcement Learning, *International Journal of Computers Communications & Control*, 16(1), 4123, 2021. https://doi.org/10.15837/ijccc.2021.1.4123